

Patent No. 12,138,276
Petition for Post Grant Review

EXHIBIT 1001



US012138276B2

(12) **United States Patent**
Hagel et al.

(10) **Patent No.:** **US 12,138,276 B2**
(45) **Date of Patent:** ***Nov. 12, 2024**

(54) **HALOGENATED PSILOCYBIN
DERIVATIVES AND METHODS OF USING**

(71) Applicant: **Enveric Biosciences Canada Inc.,**
Calgary (CA)

(72) Inventors: **Jillian M. Hagel**, Calgary (CA); **Peter
J. Facchini**, Calgary (CA)

(73) Assignee: **Enveric Biosciences Canada Inc.,**
Calgary (CA)

(*) Notice: Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 0 days.

This patent is subject to a terminal dis-
claimer.

(21) Appl. No.: **18/655,793**

(22) Filed: **May 6, 2024**

(65) **Prior Publication Data**

US 2024/0299427 A1 Sep. 12, 2024

Related U.S. Application Data

(63) Continuation of application No. 17/903,080, filed on
Sep. 6, 2022, now Pat. No. 11,998,557, which is a
(Continued)

(51) **Int. Cl.**

A61K 31/675 (2006.01)

A61K 31/4045 (2006.01)

(Continued)

(52) **U.S. Cl.**

CPC **A61K 31/675** (2013.01); **A61K 31/4045**
(2013.01); **C12N 9/0083** (2013.01);

(Continued)

(58) **Field of Classification Search**

CPC **A61K 31/4045**; **A61K 31/675**; **C07D**
209/16; **C07F 9/5728**; **C07K 2319/21**;

(Continued)

(56)

References Cited

U.S. PATENT DOCUMENTS

5,688,807 A 11/1997 Audia et al.

2008/0171779 A1 7/2008 De Bruin et al.

(Continued)

FOREIGN PATENT DOCUMENTS

WO 2004000845 A1 12/2003

WO 2018064465 A1 4/2018

(Continued)

OTHER PUBLICATIONS

Pratuangdejkul et al.(Current Medicinal Chemistry, 2005, 12, 2393-
2410 (Year: 2005).*

(Continued)

Primary Examiner — Savitha M Rao

(74) *Attorney, Agent, or Firm* — SMART & BIGGAR
LP; Michael Fenwick

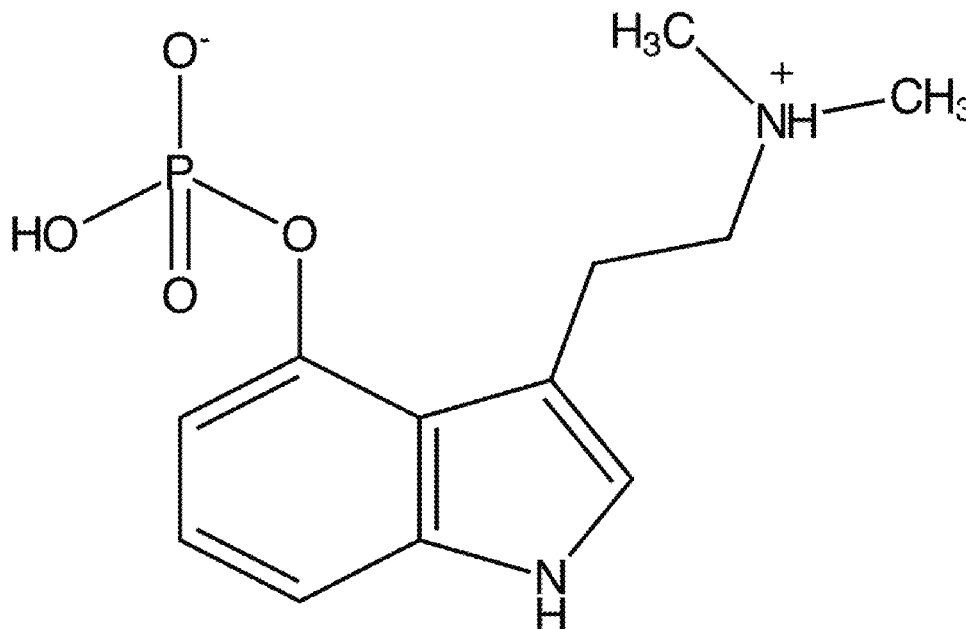
(57)

ABSTRACT

Disclosed are novel halogenated psilocybin derivative com-
pounds and pharmaceutical and recreational drug formula-
tions containing the same. The compounds may be produced
using a biosynthetic system which comprises cells compris-
ing a psilocybin biosynthetic enzyme complement.

29 Claims, 36 Drawing Sheets

Specification includes a Sequence Listing.



Related U.S. Application Data

- continuation of application No. PCT/CA2021/051209, filed on Sep. 1, 2021.
- (60) Provisional application No. 63/073,104, filed on Sep. 1, 2020.
- (51) **Int. Cl.**
C12N 9/02 (2006.01)
C12N 9/10 (2006.01)
C12N 9/12 (2006.01)
C12N 15/52 (2006.01)
C12P 17/10 (2006.01)
- (52) **U.S. Cl.**
 CPC *C12N 9/1007* (2013.01); *C12N 9/1205* (2013.01); *C12N 15/52* (2013.01); *C12P 17/10* (2013.01)
- (58) **Field of Classification Search**
 CPC C07K 2319/40; C07K 2319/41; C07K 2319/42; C07K 2319/43; C12N 15/52; C12N 9/0071; C12N 9/0083; C12N 9/1007; C12N 9/1029; C12N 9/1205; C12N 9/88; C12P 17/10; C12Y 402/0102
 See application file for complete search history.

(56) References Cited**U.S. PATENT DOCUMENTS**

2009/0093513 A1 4/2009 Hamann et al.
 2020/0397752 A1 12/2020 Perez Castillo et al.

FOREIGN PATENT DOCUMENTS

WO 2019173797 A1 9/2019
 WO 2021011913 A1 1/2021
 WO 2021108911 A1 6/2021
 WO 2021168082 A1 8/2021
 WO 2022051670 A1 3/2022

OTHER PUBLICATIONS

Daniel, J. et al. Clinical potential of psilocybin as a treatment for mental health conditions. *Mental Health Clin.*, 2017;7 (1): 24-28.

Grob, C. et al. Pilot study of psilocybin treatment for anxiety in patients with advanced-stage cancer. *Arch. Gen. Psychiatry*, 2011, 68(1) 71-78.

Cathart-Harris, R.L. et al. Psilocybin with psychological support for treatment-resistant depression: an open-label feasibility study. *Lancet Psychiatry*, 2016, 3: 619-627.

Needleman and Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, 1970, 48: 443.

Smith and Waterman. *Adv. Appl. Math.*, 1981, 2: 482.

Carillo and Lipton. *Siam J. Applied Math.*, 1988, 48:1073.

Romeo, B. et al. Clinical and biological predictors of psychedelic response in the treatment of psychiatric and addictive disorders: A systematic review. *J Psychiatr Res* 137: 273, 2021.

Devereux et al., A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.*, 1984, 12: 387.

Altschul et al., *J. Mol. Biol.*, 1990:215:403.

Thompson, J D, Higgins, D G and Gibson T J, 1994, *Nucleic Acid Res* 22(22): 4673-4680.

S. Kawai et al., Transformation of *Saccharomyces cerevisiae* and other fungi: methods and possible underlying mechanism. *Bioeng Bugs*. Nov.-Dec. 2010; 1(6):395-403.

Henikoff S & Henikoff, J G, 1992, *Proc. Natl. Acad. Sci. USA* 89: 10915-10919.

Roy A.D et al. Development of fluorescent aryltryptophans by Pd mediated cross-coupling of unprotected halotryptophans in water. *Chem Commun (Camb)*. Oct. 21, 2008;(39):4831-3.

Corr M.J. et al. Sonogashira diversification of unprotected halotryptophans, halotryptophan containing tripeptides; and generation of a new to nature bromo-natural product and its diversification in water. *Chem Sci*. Mar. 1, 2017;8 (3):2039-2046.

Finnin, B. and Morgan, T.M., Transdermal penetration enhancers: applications, limitations, and potential. *J Pharm Sci*. Oct. 1999;88(10):955-8.

Ross, S. et al. Acute and Sustained Reductions in Loss of Meaning and Suicidal Ideation Following Psilocybin-Assisted Psychotherapy for Psychiatric and Existential Distress in Life-Threatening Cancer. *ACS Pharmacol Transl Sci*. Apr. 9, 2021; 4(2): 553-562.

Schnepel C. et al. Enzymatic Halogenation: A Timely Strategy for Regioselective C—H Activation. *Chemistry*. Sep. 7, 2017;23(50):12064-12086.

Durak L.J. et al. Late-Stage Diversification of Biologically Active Molecules via Chemoenzymatic C—H Functionalization. *ACS Catal*. Mar. 4, 2016; 6(3): 1451-1454.

Runguphan W. et al. Diversification of monoterpene indole alkaloid analogs through cross-coupling. *Org Lett*. Jun. 7, 2013;15(11):2850-3.

Ding, L. et al. Discovery and Structure-Based Optimization of 6-Bromotryptamine Derivatives as Potential 5-HT_{2A} Receptor Antagonists. *Molecules*, 2015, 20:17675-17683.

Shoubhye, J. et al. Hybrid molecules inhibiting myeloperoxidase activity and serotonin reuptake: a possible new approach of major depressive disorders with inflammatory syndrome. *Royal Pharmaceutical Society, Journal of Pharmacy and Pharmacology*, 2014, 66, pp. 1122-1132.

Cregg et al., Recombinant protein expression in *Pichia pastoris*. *Mol Biotechnol*. (2000) 16(1): 23-52.

Mattanovich et al., Recombinant protein production in yeasts. *Methods Mol. Biol.*, 2012, 824:329-58.

Romanos et al., Foreign gene expression in yeast: a review. *Yeast*. Jun. 1992;8(6):423-88.

Andorfer M.C., et al. Directed Evolution of RebH for Catalyst-Controlled Halogenation of Indole C—H Bonds. (2016) *Chem. Sci*. 7: 3720.

Glenn W.S., et al. Reengineering a tryptophan halogenase to preferentially chlorinate a direct alkaloid precursor (2011) *J. Am. Chem. Soc.* 133: 19346.

Campos et al. A general synthesis of substituted indoles from cyclic enol ethers and enol lactones 2004, *Org. Lett* 6: 79-82.

Dastmalchi et al., 2019, *Nat. Chem Biol.* 15: 384-390.

Chen et al., A pathogenesis-related 10 protein catalyzes the final step in thebaïne biosynthesis. *Nat. Chem Biol.* 14: 738-743, 2018.

Chang, L. et al. Isolation and Characterization of O-methyltransferases Involved in the Biosynthesis of Glaucine in *Glaucium flavum* L. *Plant Physiology*, Oct. 2015, vol. 169, pp. 1127-1140.

Jones et al. ePathOptimize: A Combinatorial Approach for Transcriptional Balancing of Metabolic Pathways. 2015, *Sci Rep*. 5: 11301.

Pyrgiotakis G. et al., 2009, *Ann. Biomed. Eng.* 37: 1464-1473.

Rickli A. et al., 2016, *Europ. Neuropsychopharmacol.*, 26: 1326-1337.

Lombardo et al. A tandem isomerization/prins strategy: iridium(III)/Brønsted acid cooperative catalysis 2013, *Angew. Chem. Int. Ed.* 52: 12910-12914.

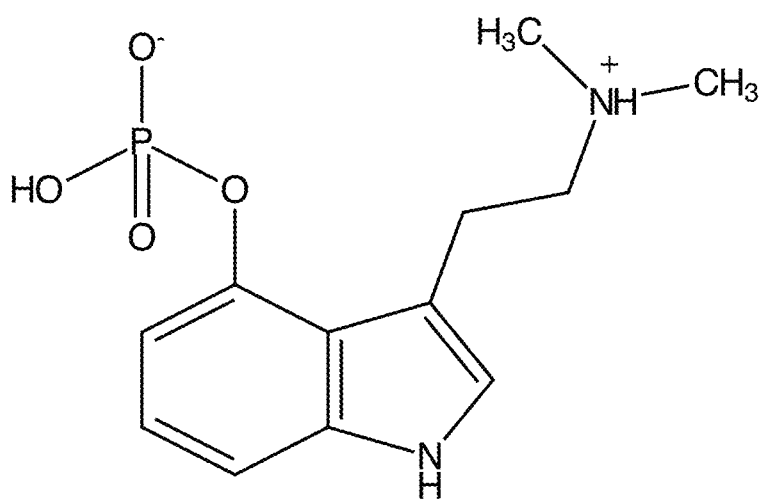
Go et al. Amino derivatives of indole as potent inhibitors of isoprenylcysteine carboxyl methyltransferase 2010, *J. Med. Chem.* 53: 6838-6850.

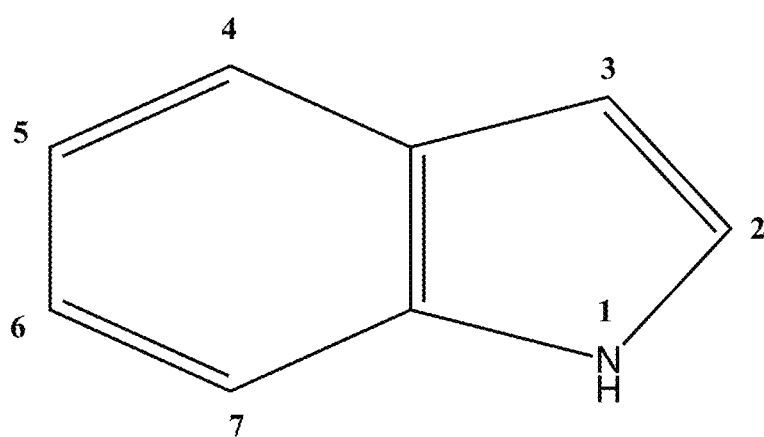
Winkelblech et al. (Organic and Biomolecular Chemistry, 2016, 14, 9883-9895) (Year: 2016).

Williams et al. (Foye's Principles of Medicinal chemistry, fifth edition, 2002) (Year: 2002).

Patani et al. (Chem. Rev. 1996, 96, 3147-3176) (Year: 1996).

* cited by examiner

**FIG. 1**

**FIG. 2**

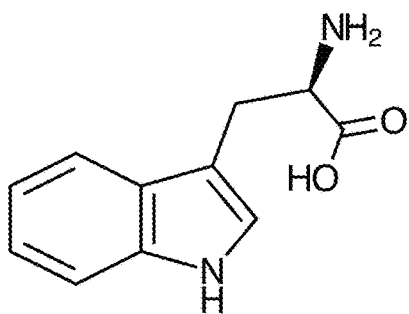


FIG. 3A

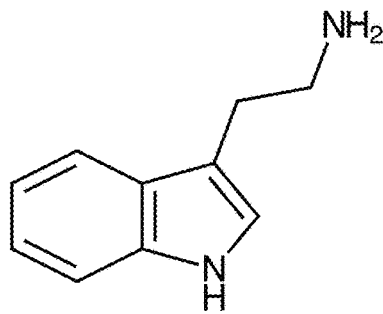


FIG. 3B

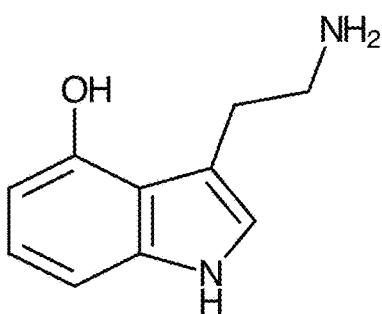


FIG. 3C

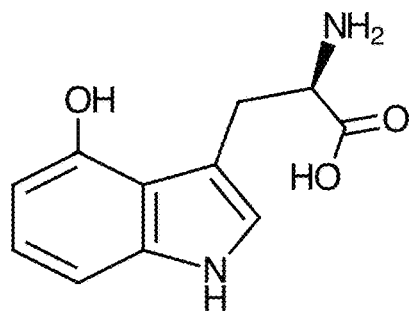


FIG. 3D

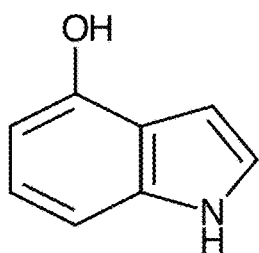


FIG. 3E

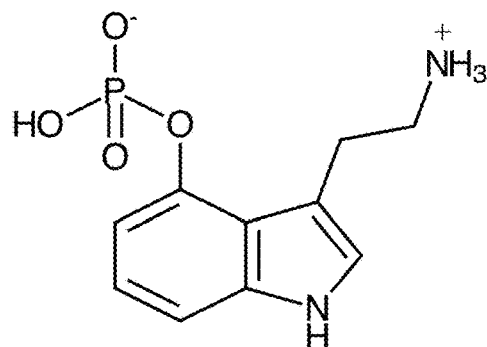


FIG. 3F

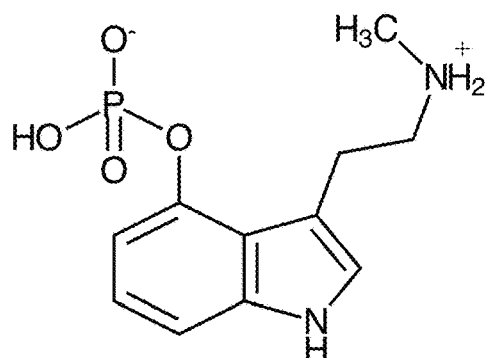


FIG. 3G

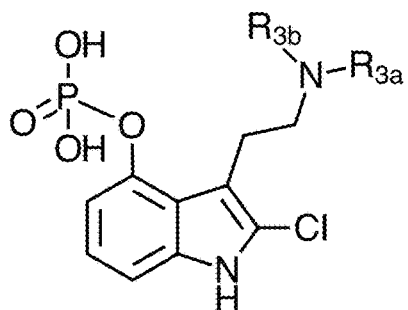


FIG. 4A

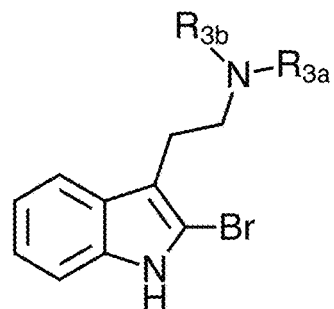


FIG. 4B

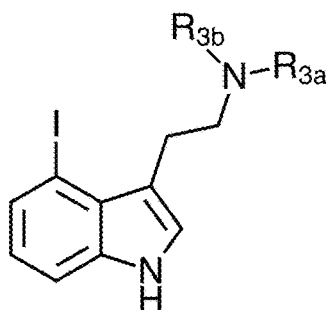


FIG. 4C

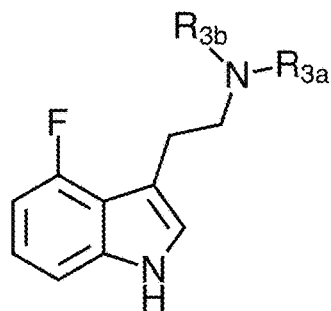


FIG. 4D

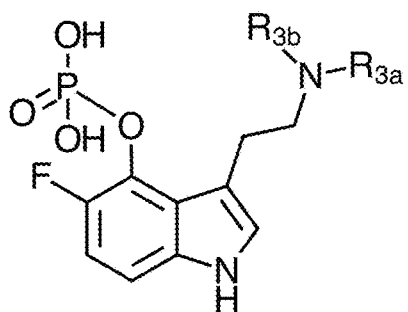


FIG. 4E

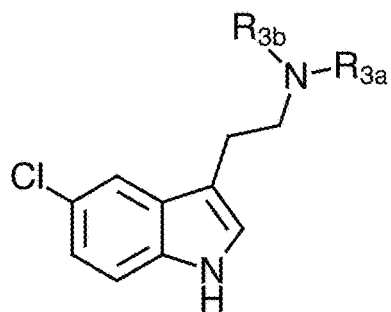


FIG. 4F

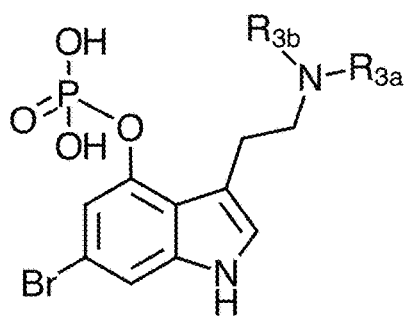


FIG. 4G

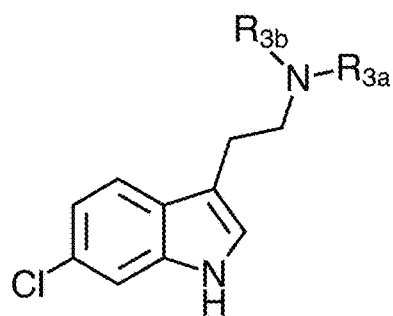


FIG. 4H

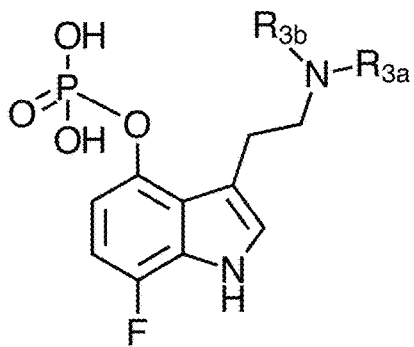


FIG. 4I

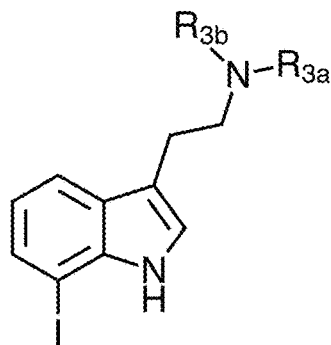


FIG. 4J

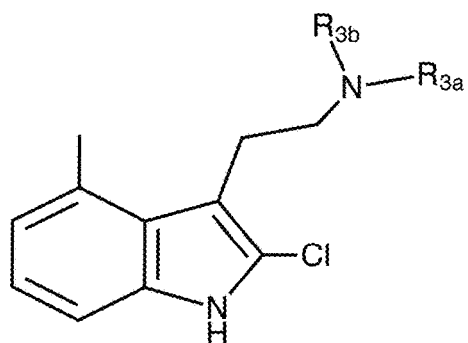


FIG. 4K

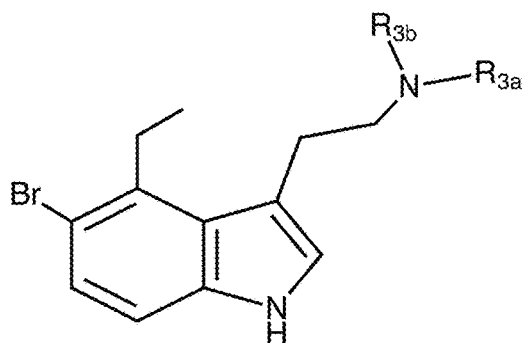


FIG. 4L

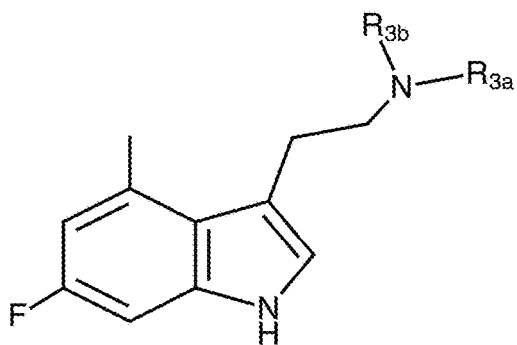


FIG. 4M

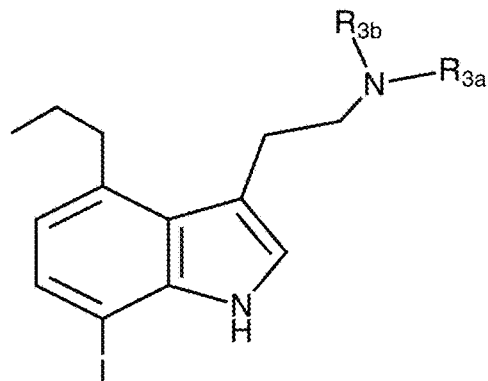
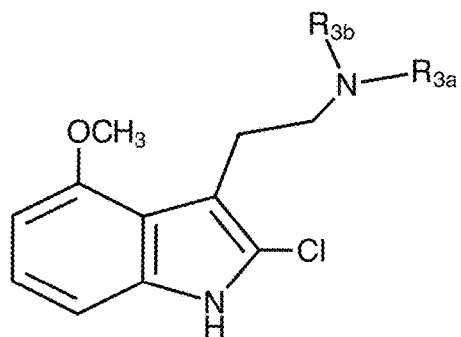
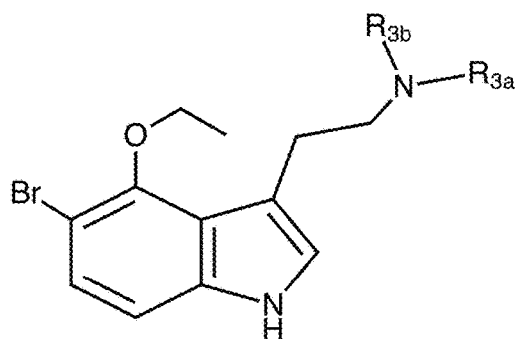
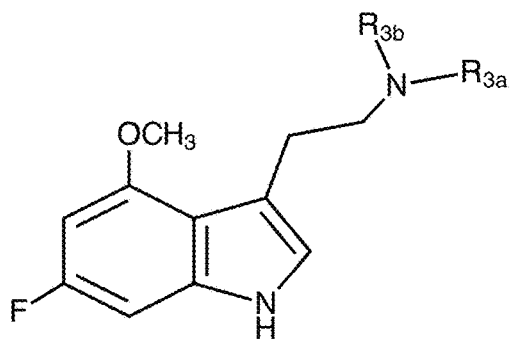
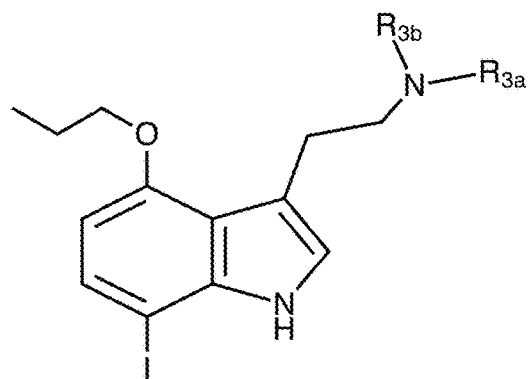


FIG. 4N

**FIG. 4O****FIG. 4P****FIG. 4Q****FIG. 4R**

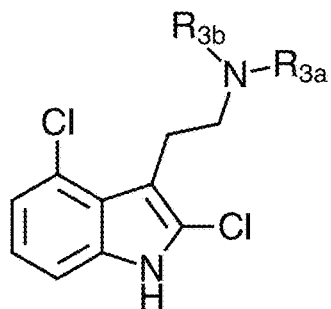


FIG. 5A

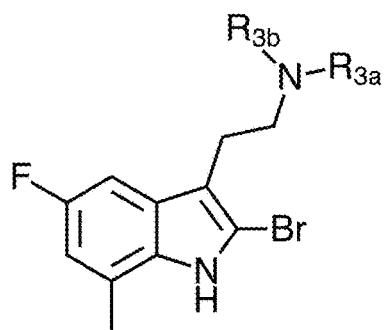


FIG. 5B

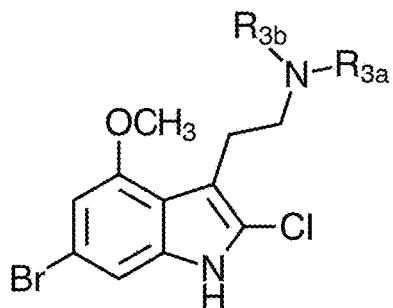


FIG. 5C

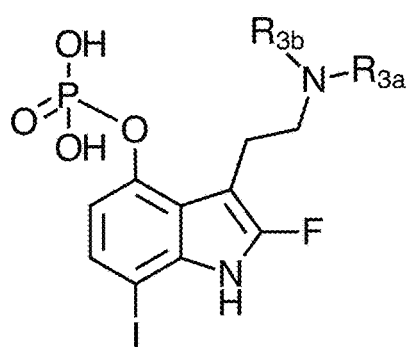


FIG. 5D

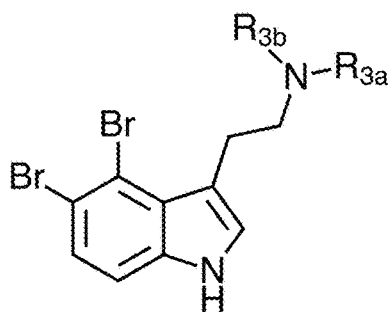


FIG. 5E

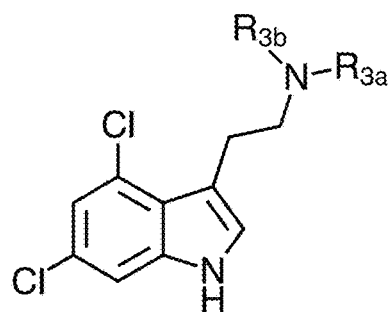


FIG. 5F

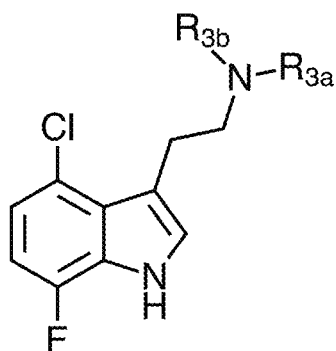


FIG. 5G

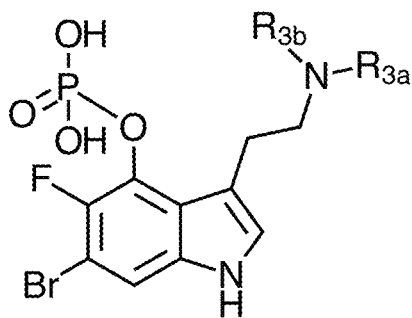


FIG. 5H

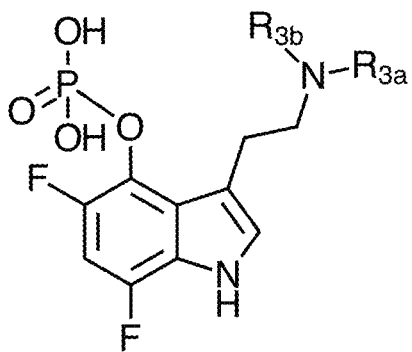


FIG. 5I

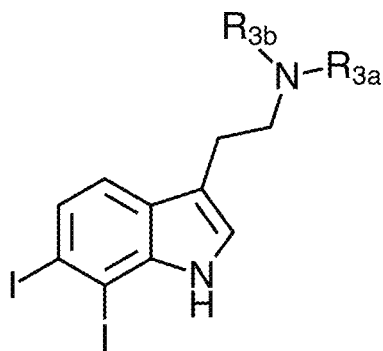


FIG. 5J

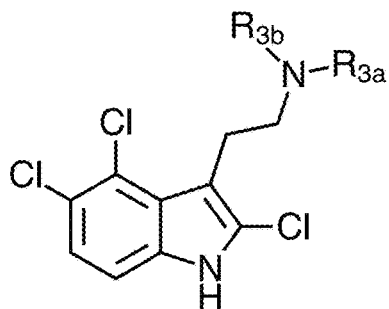


FIG. 6A

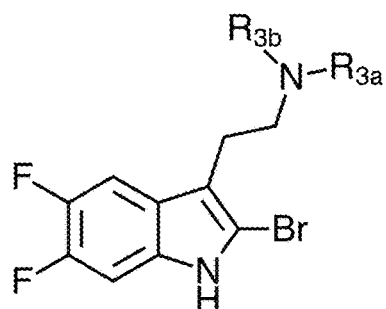


FIG. 6B

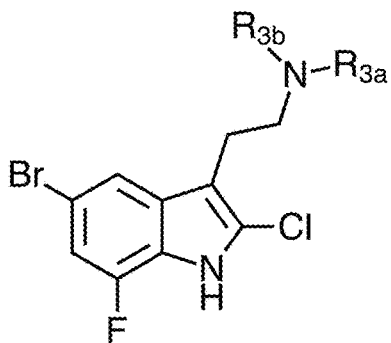


FIG. 6C

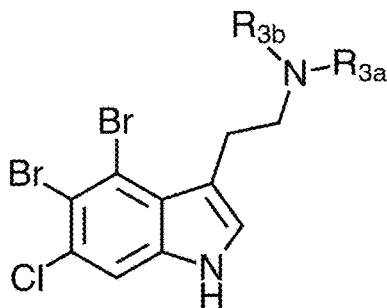


FIG. 6D

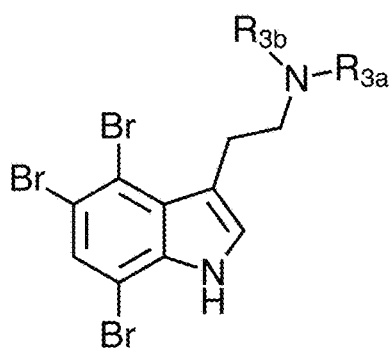


FIG. 6E

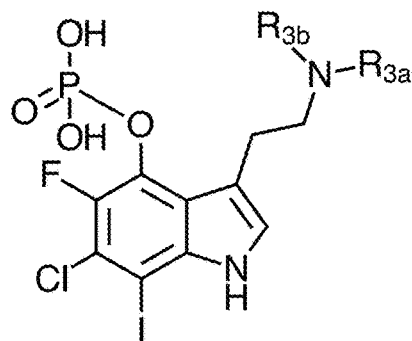


FIG. 6F

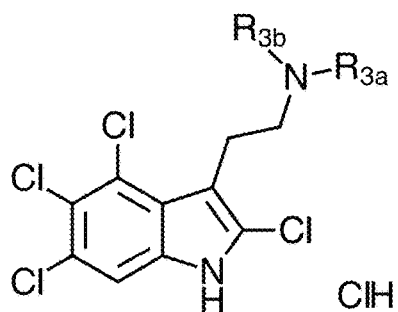


FIG. 7A

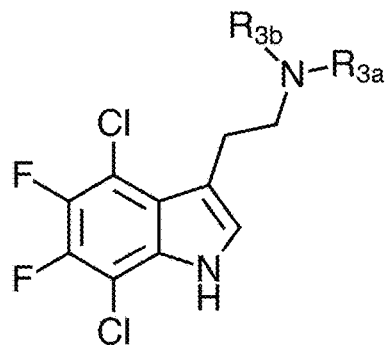


FIG. 7B

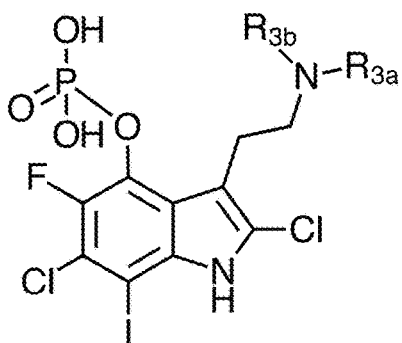


FIG. 7C

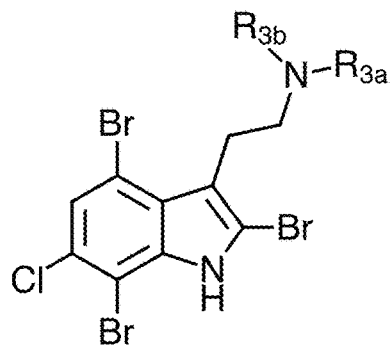


FIG. 7D

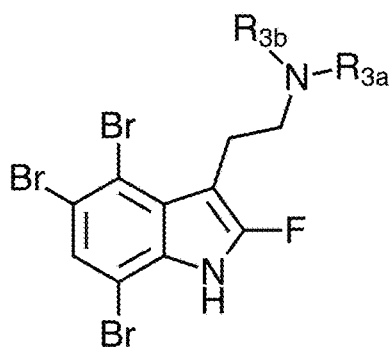
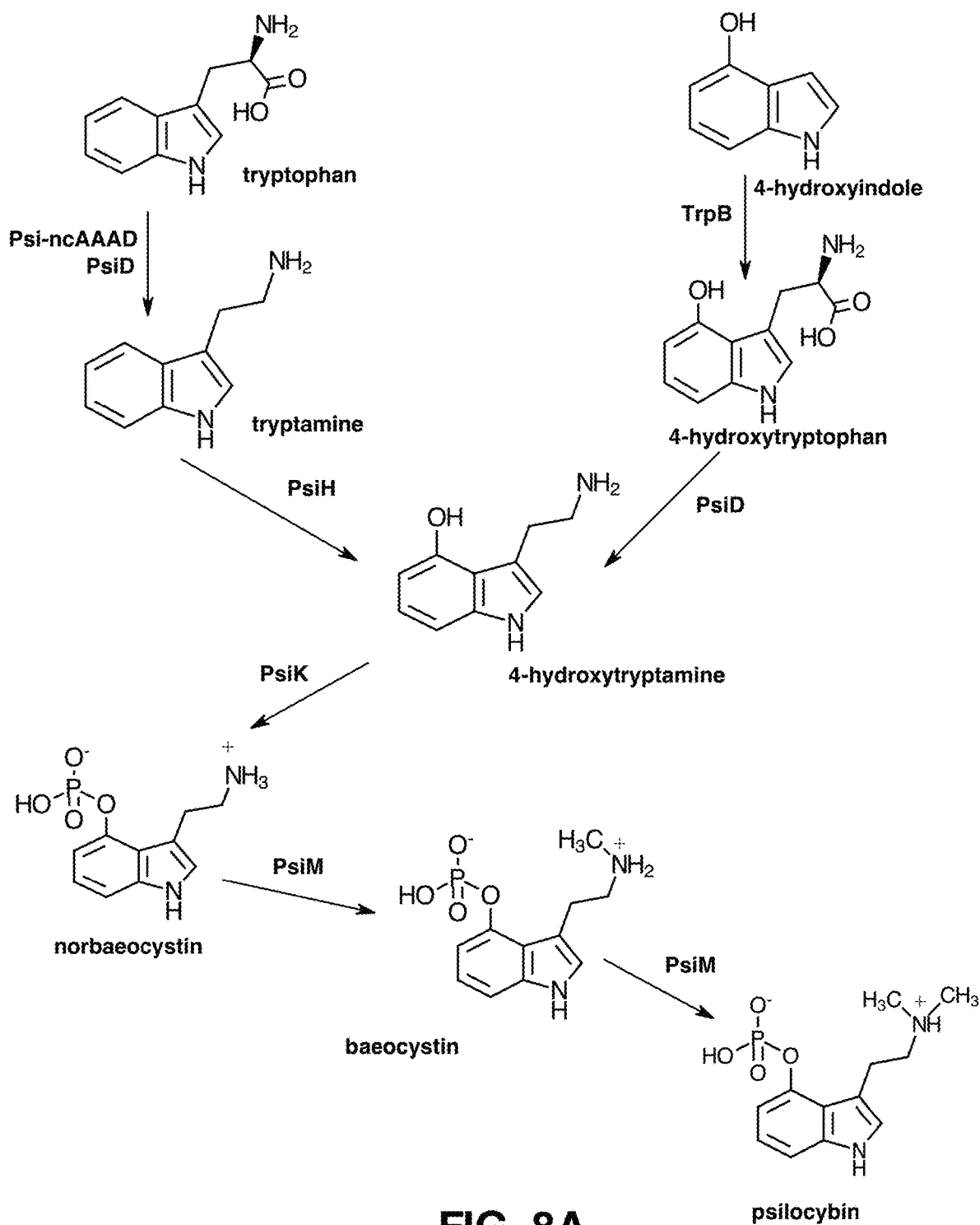
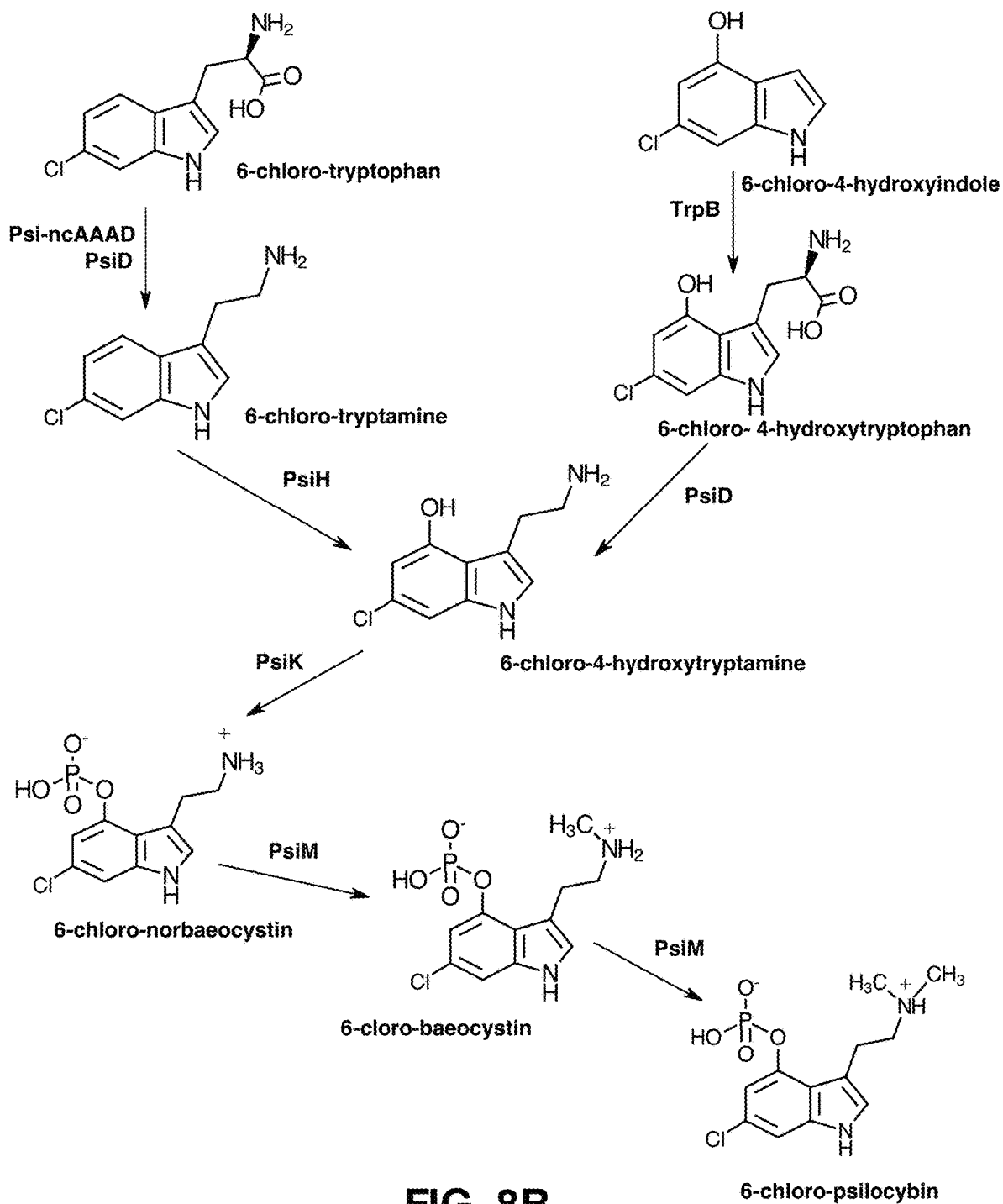


FIG. 7E





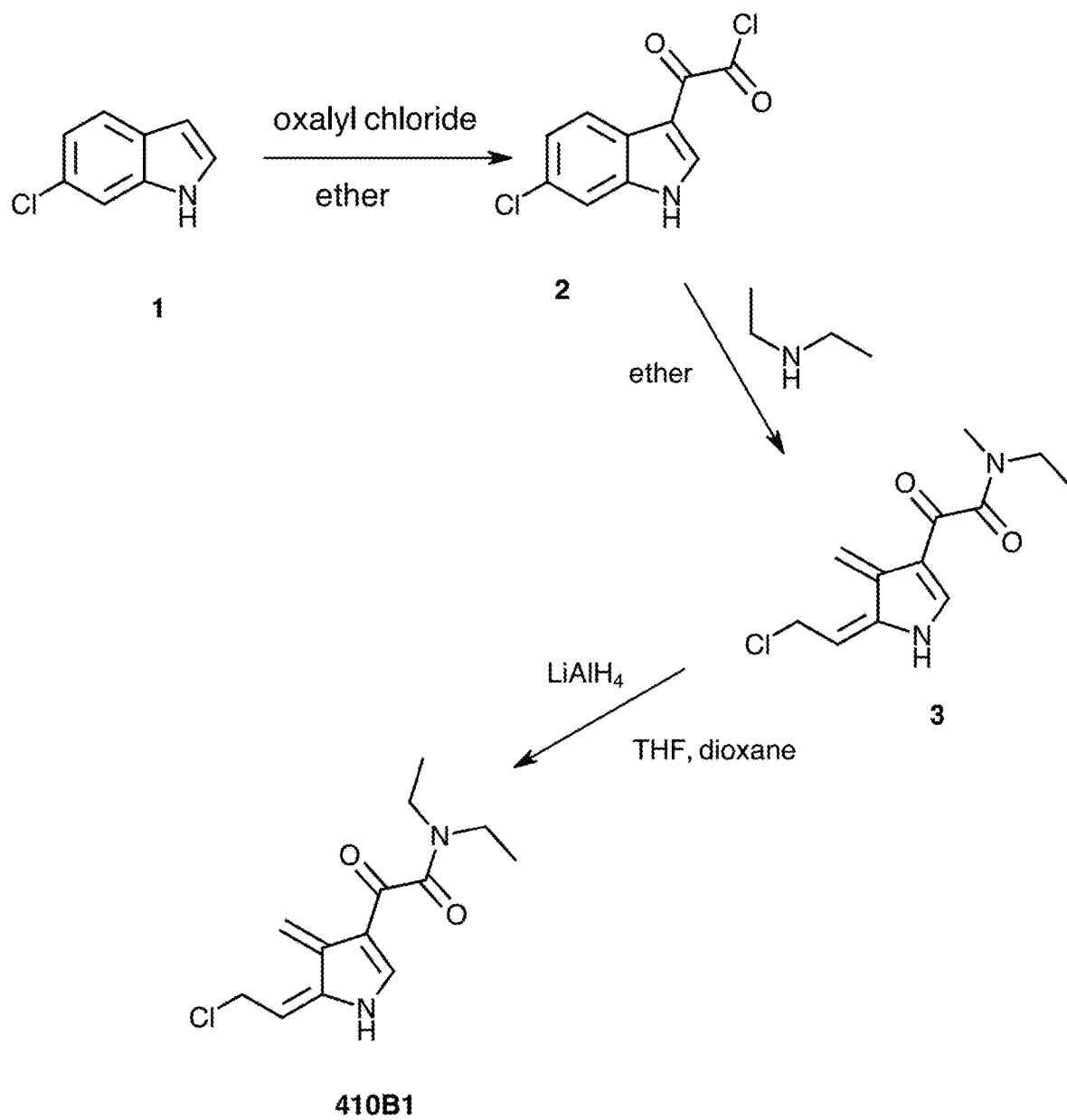


FIG. 9A

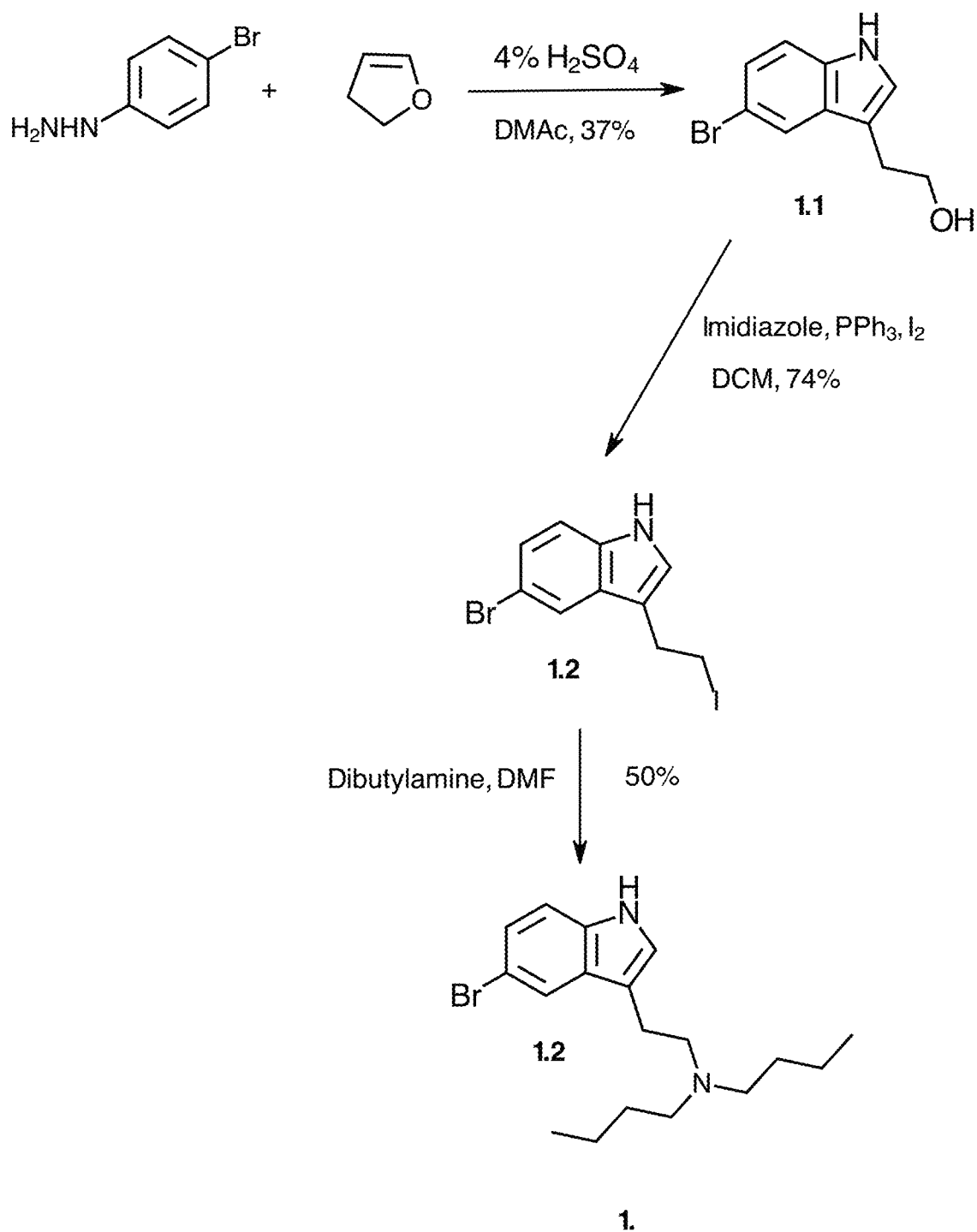
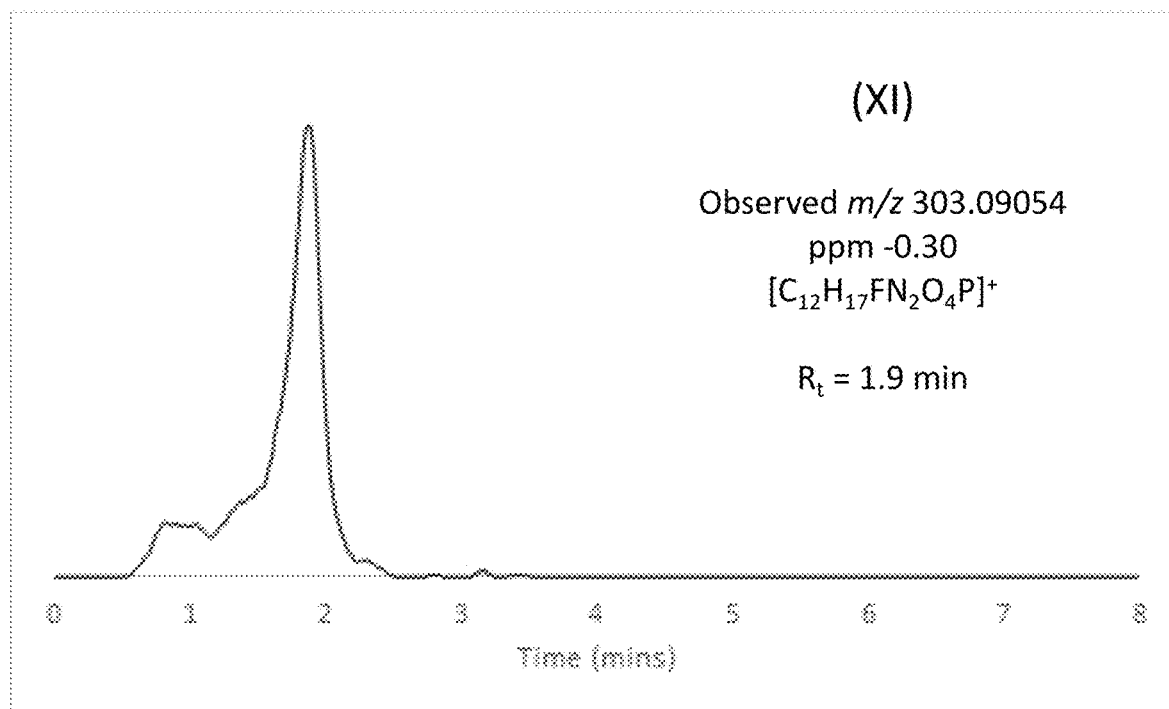
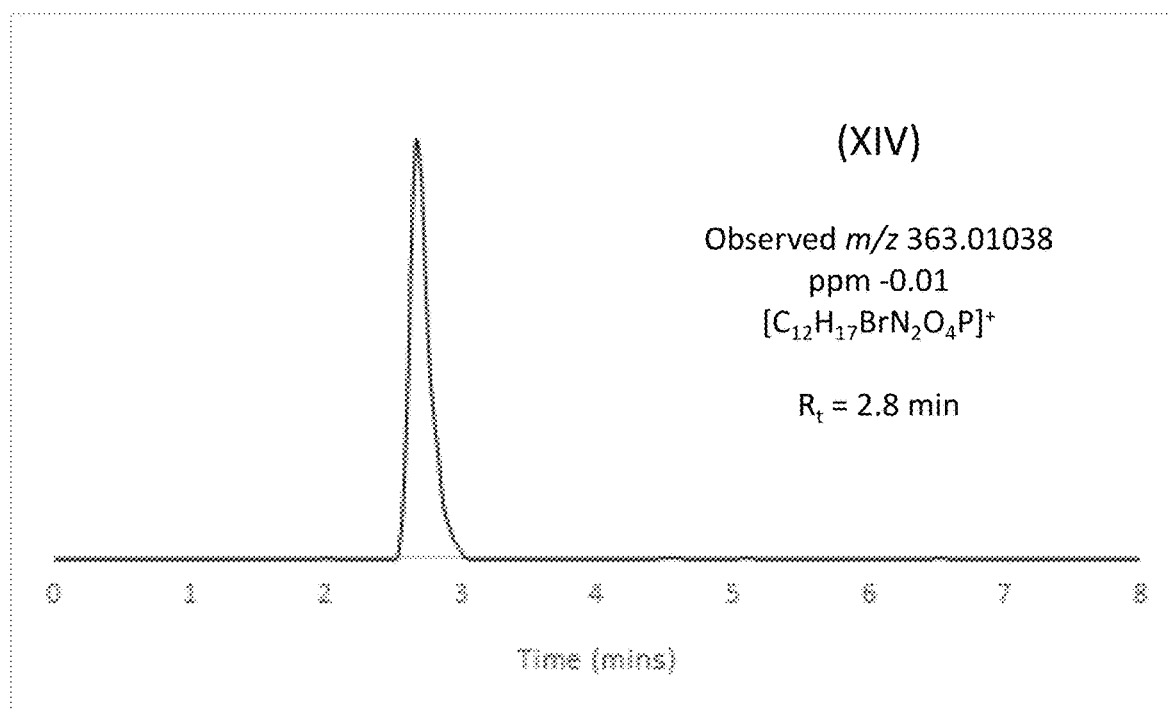
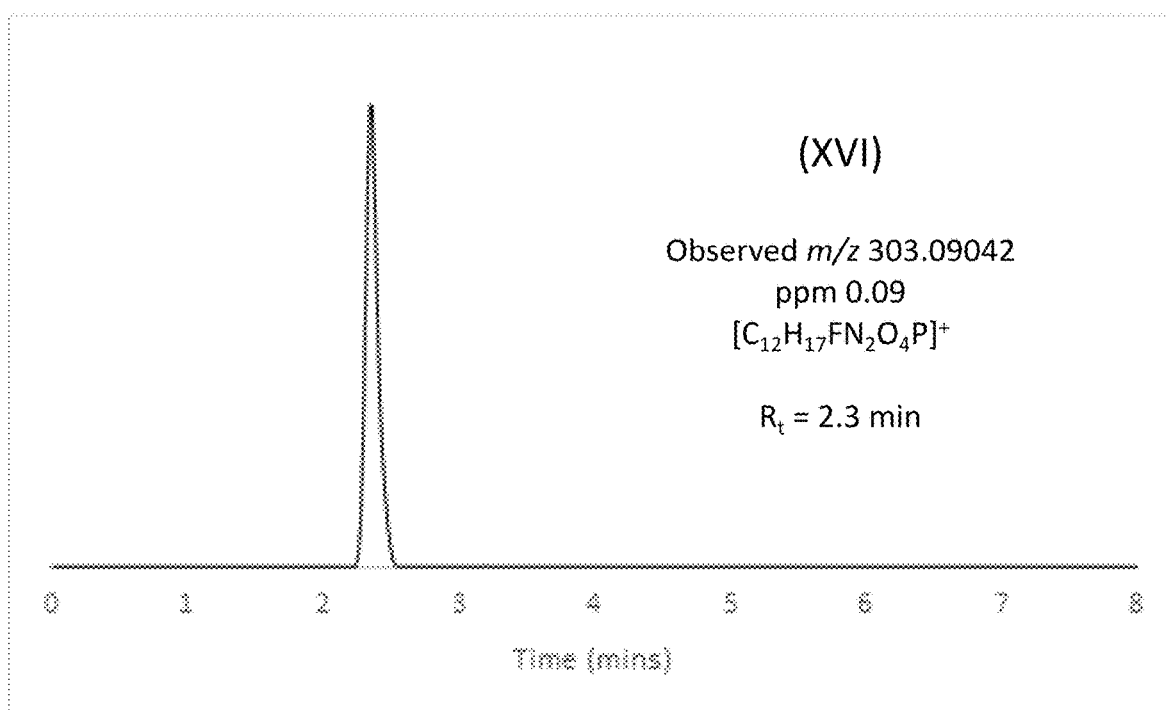
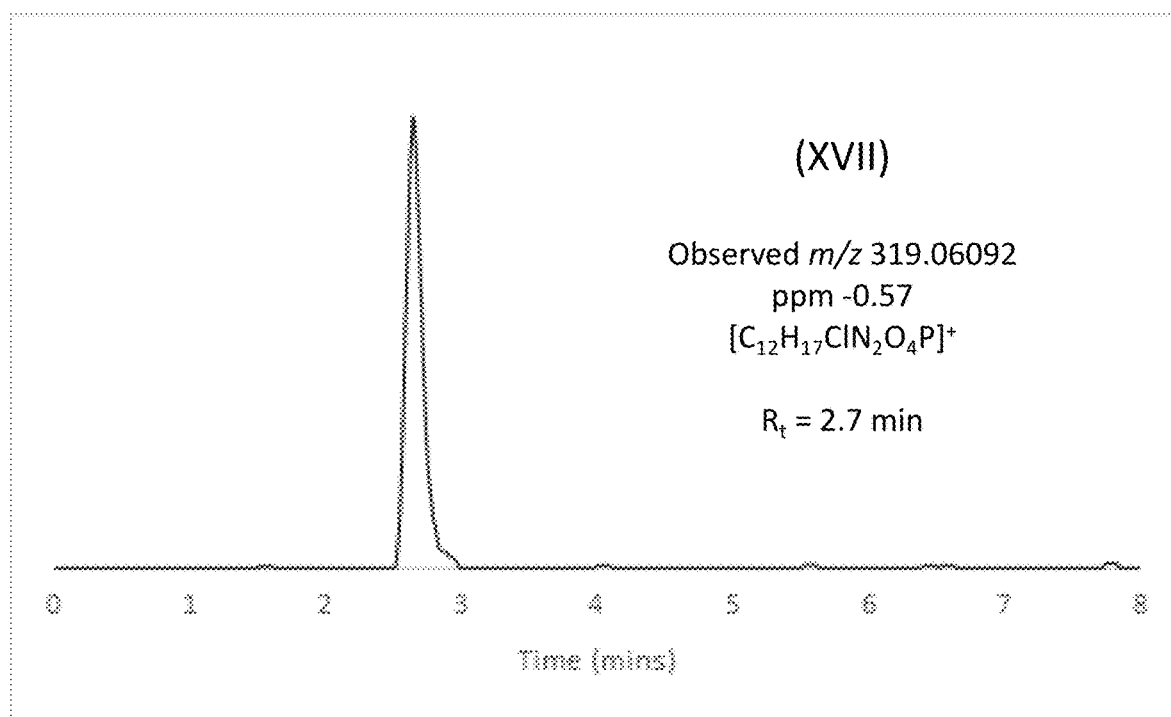


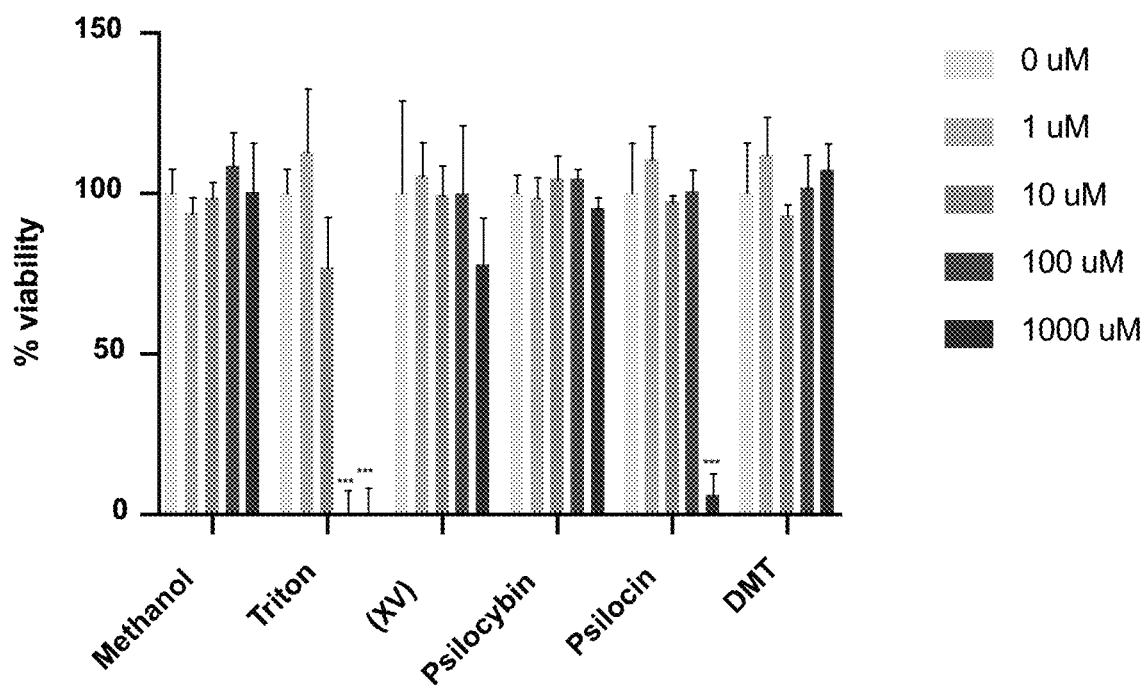
FIG. 9B

**FIG. 10**

**FIG. 11**

**FIG. 12**

**FIG. 13**

**FIG. 14**

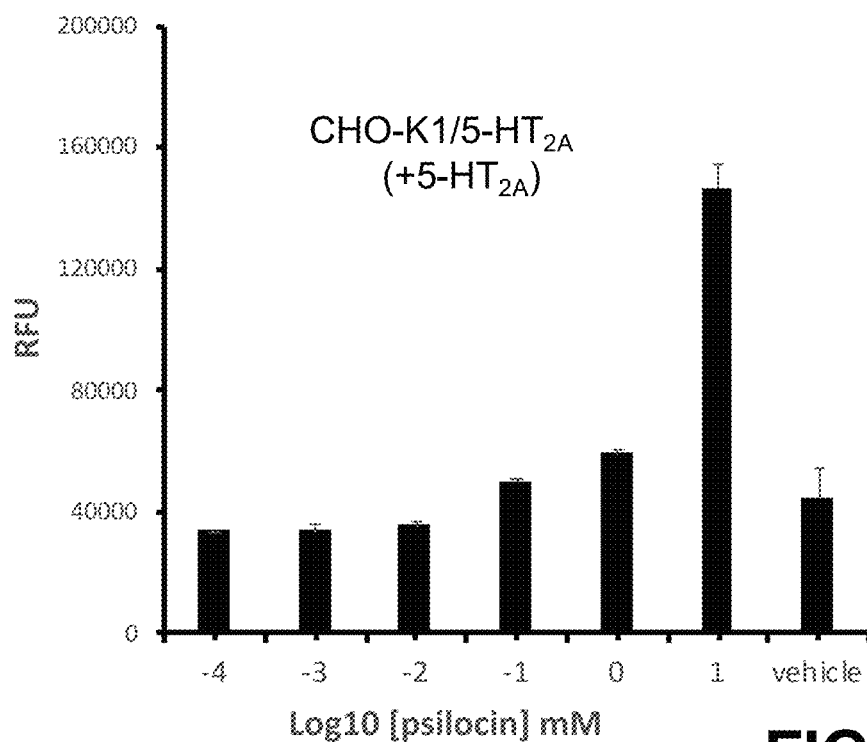


FIG. 15A

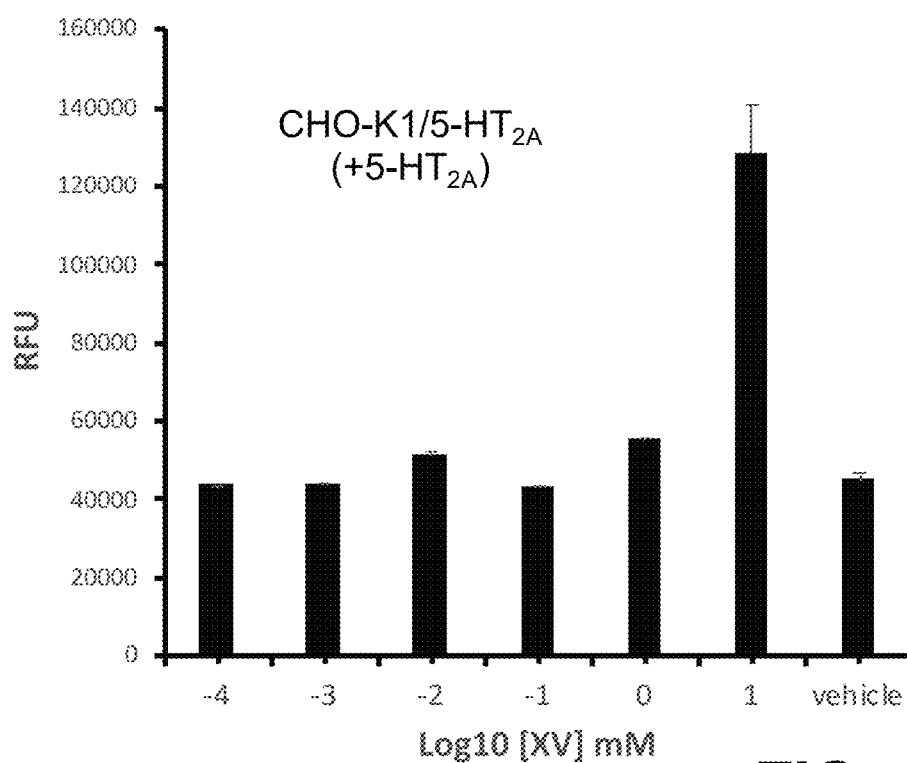


FIG. 15B

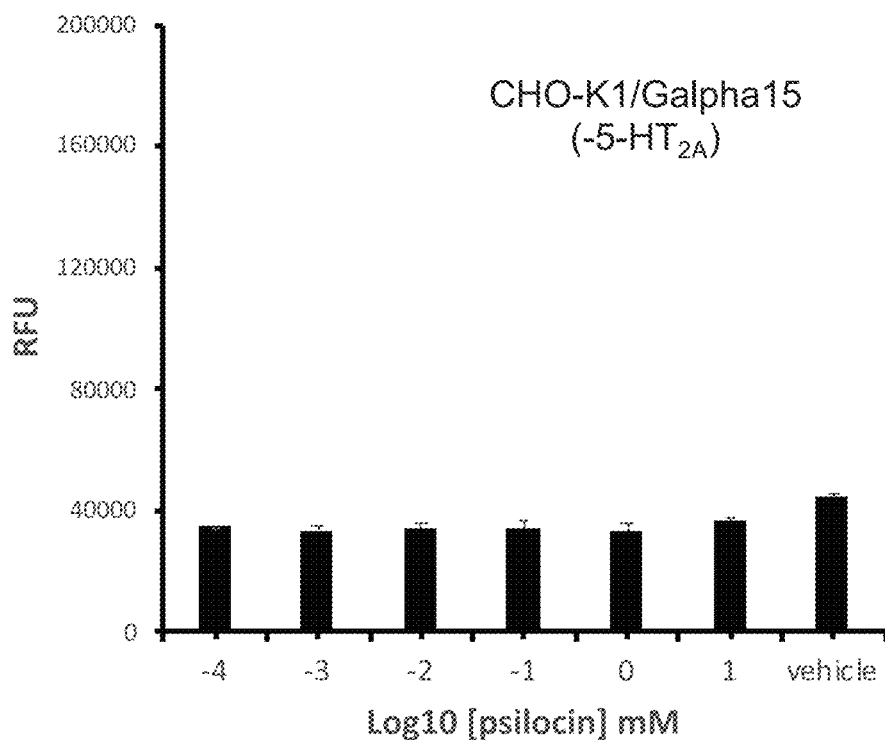


FIG. 16A

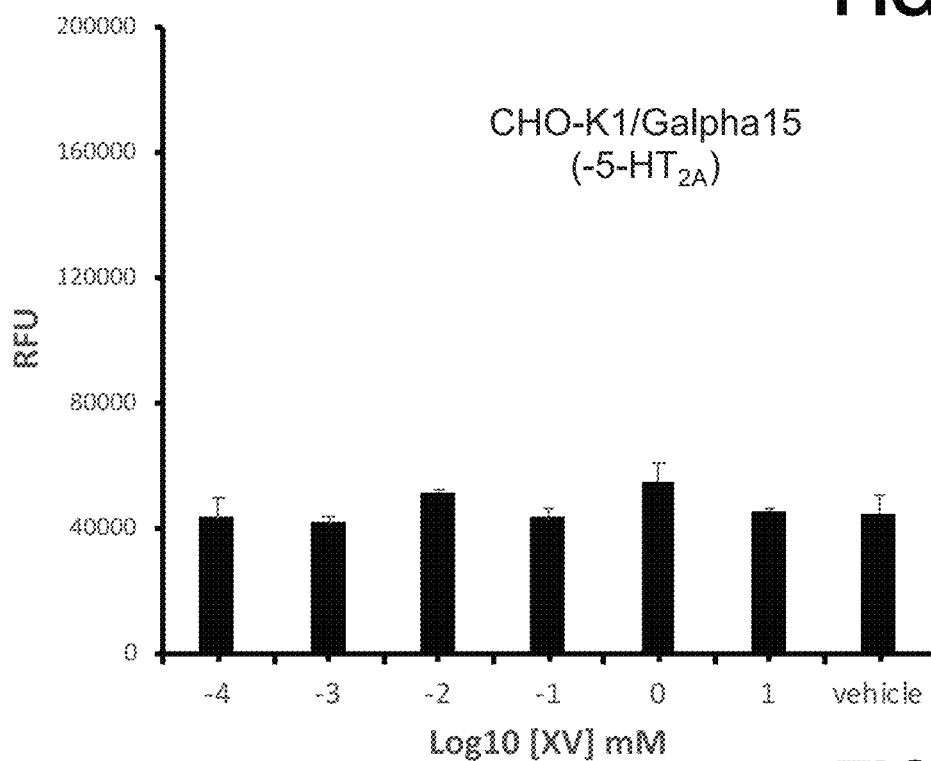
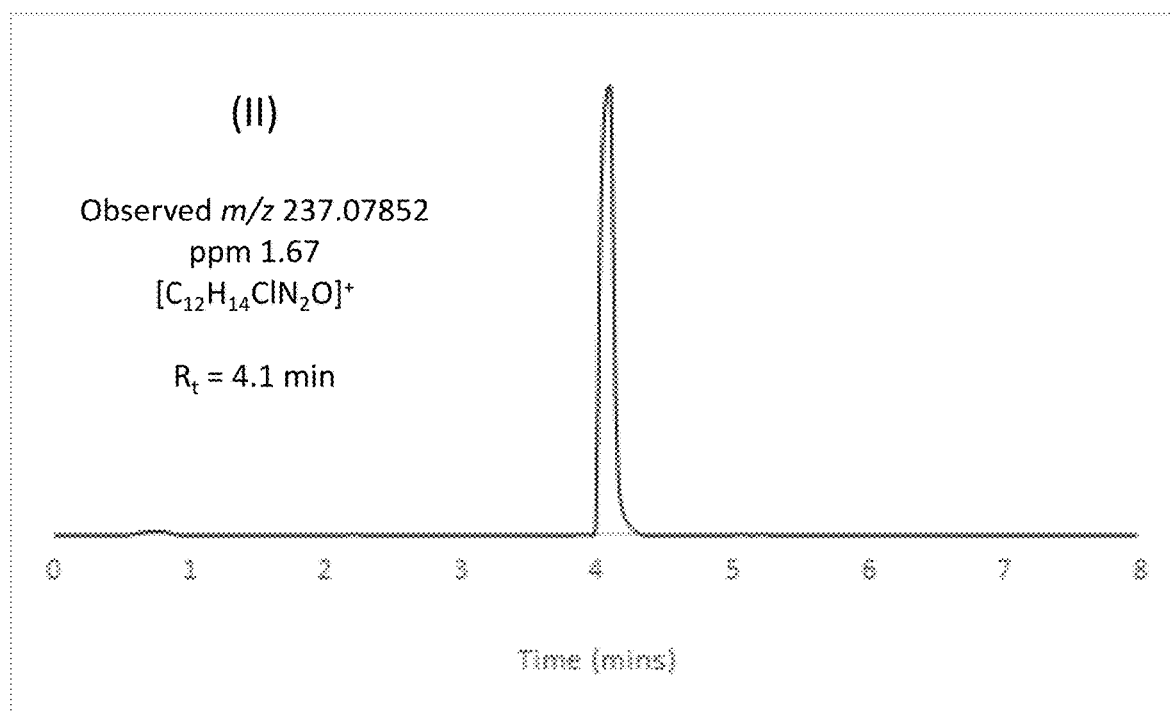
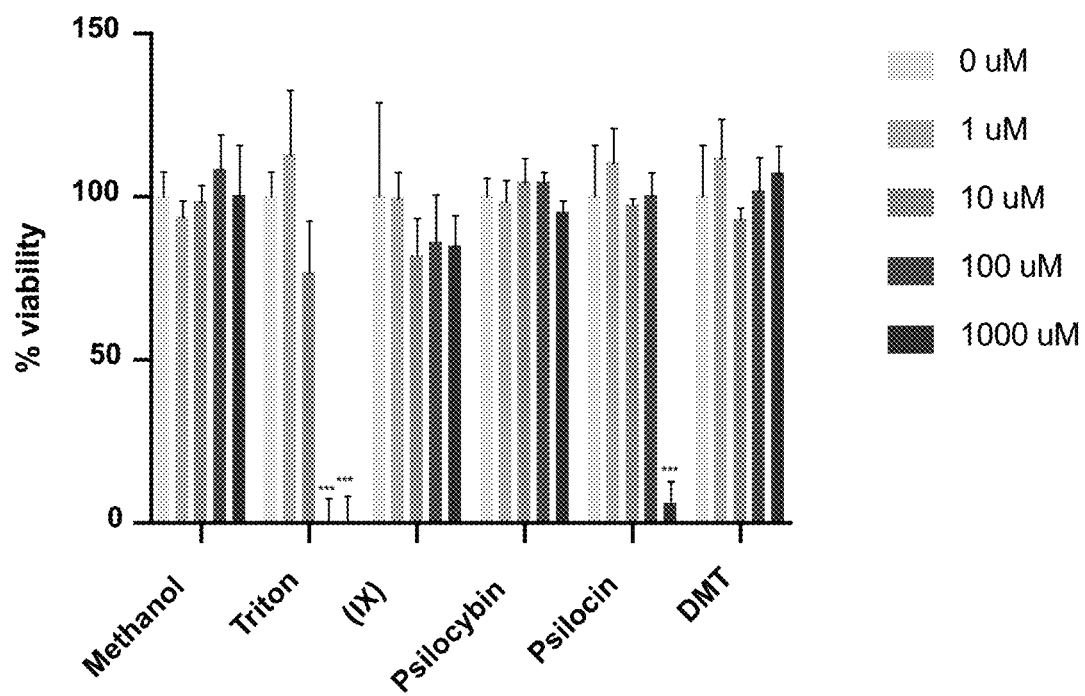
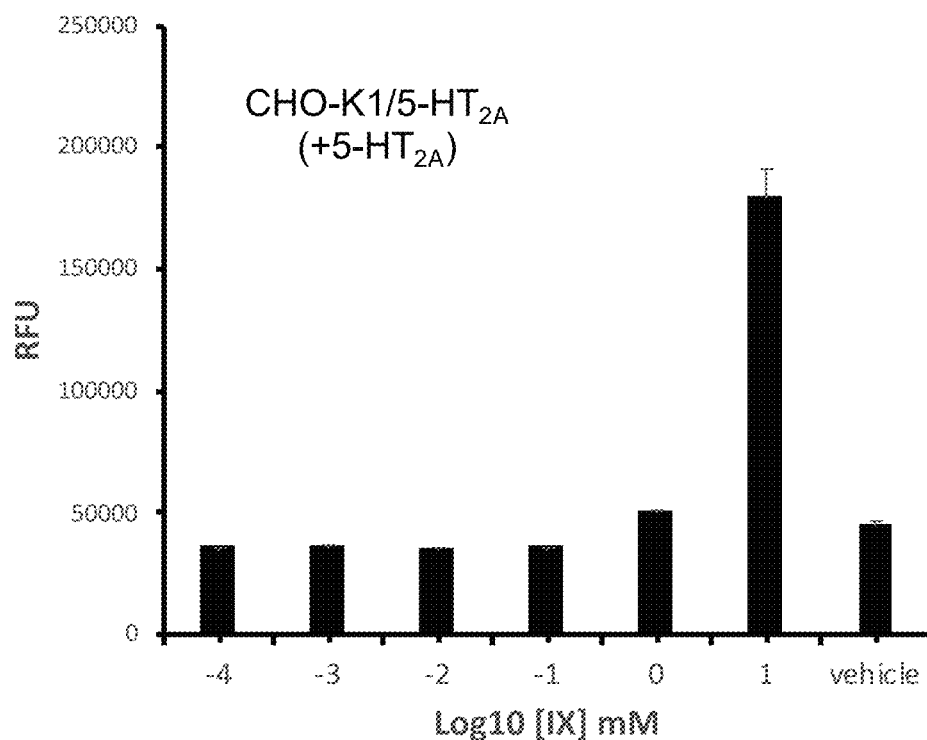
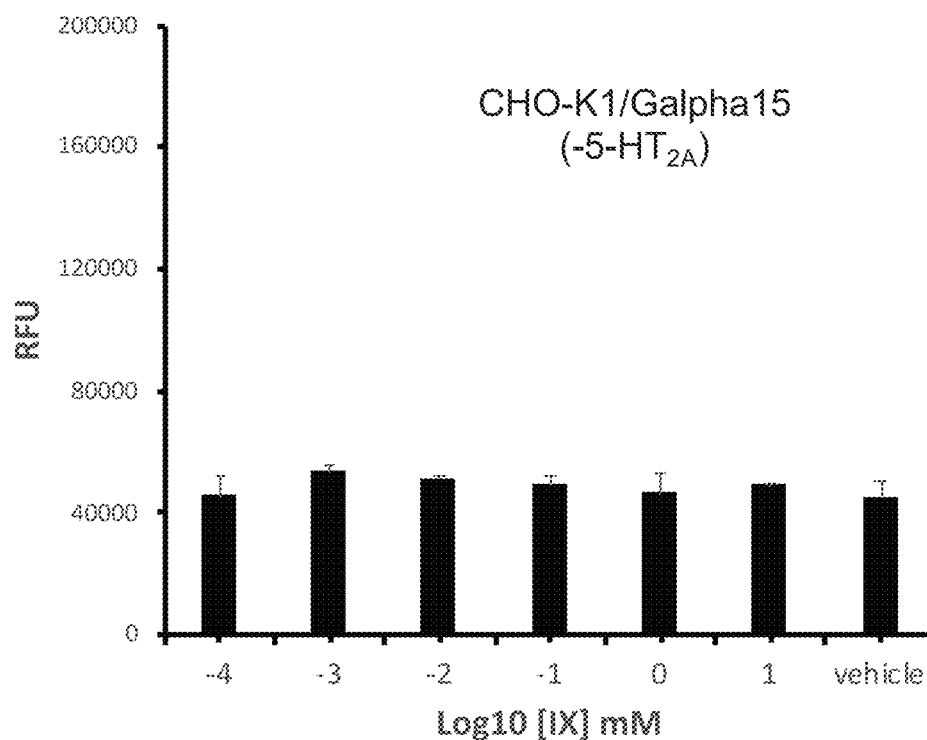
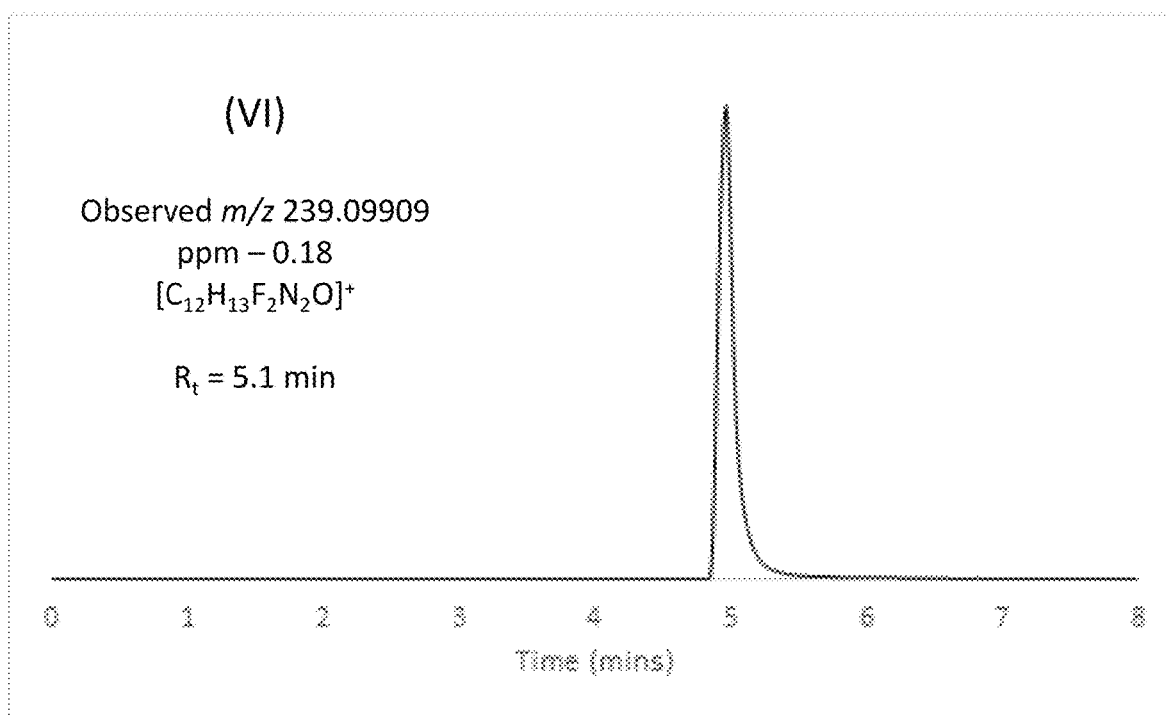


FIG. 16B

**FIG. 17**

**FIG. 18**

**FIG. 19A****FIG. 19B**

**FIG. 20**

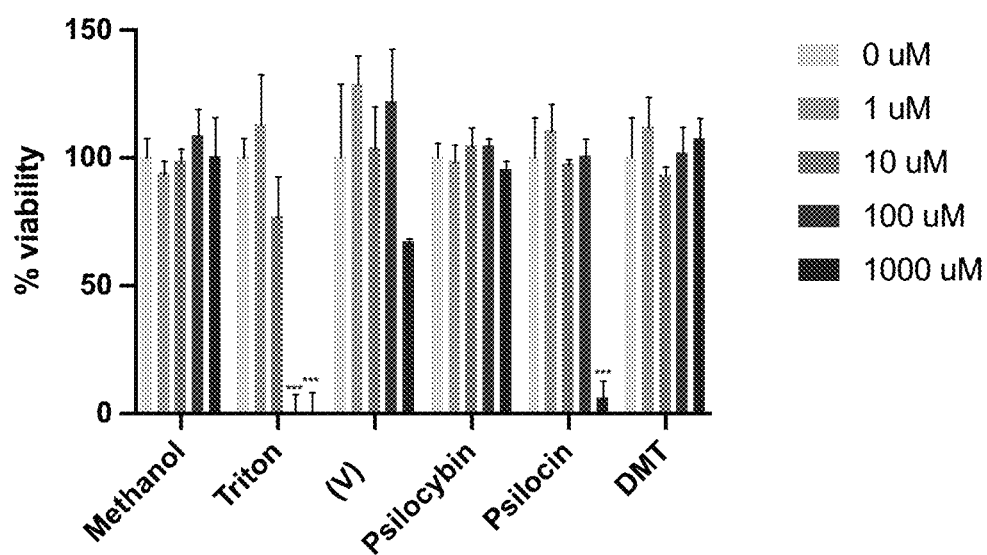


FIG. 21

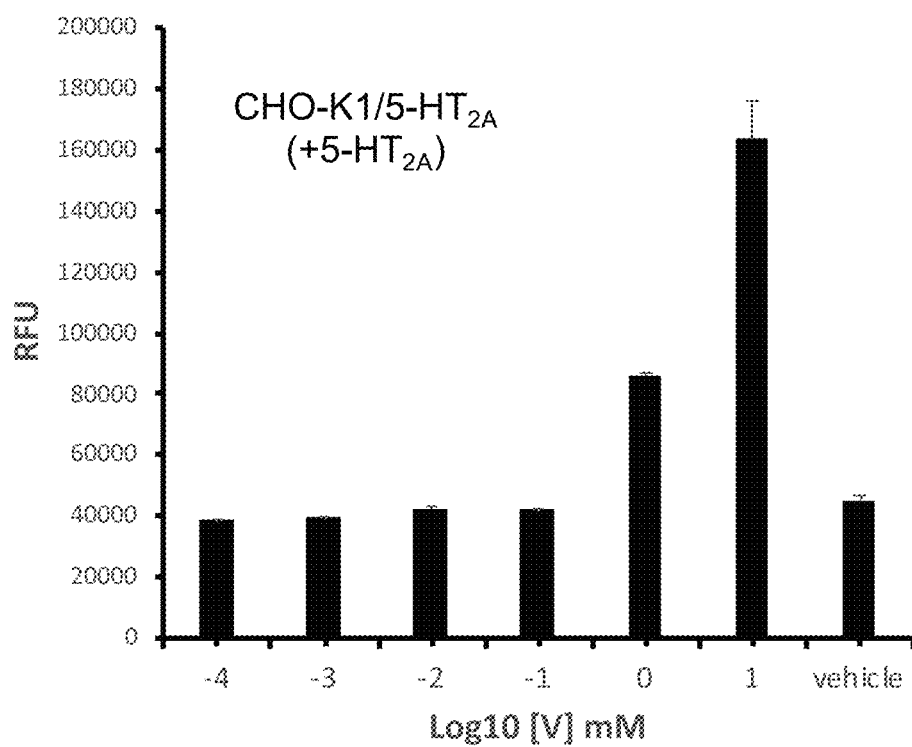


FIG. 22A

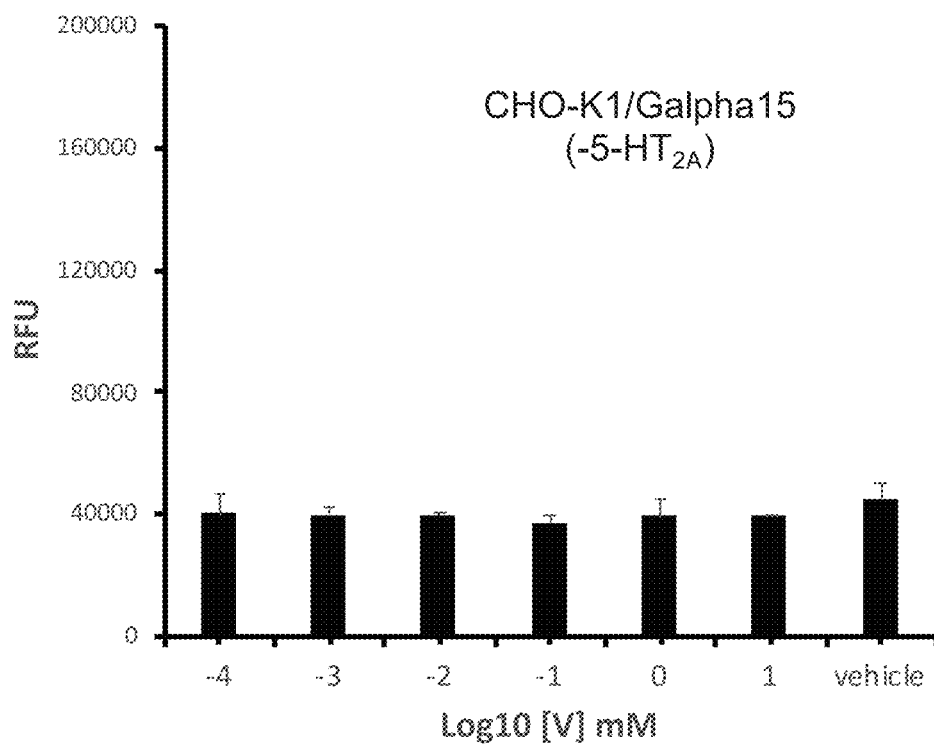
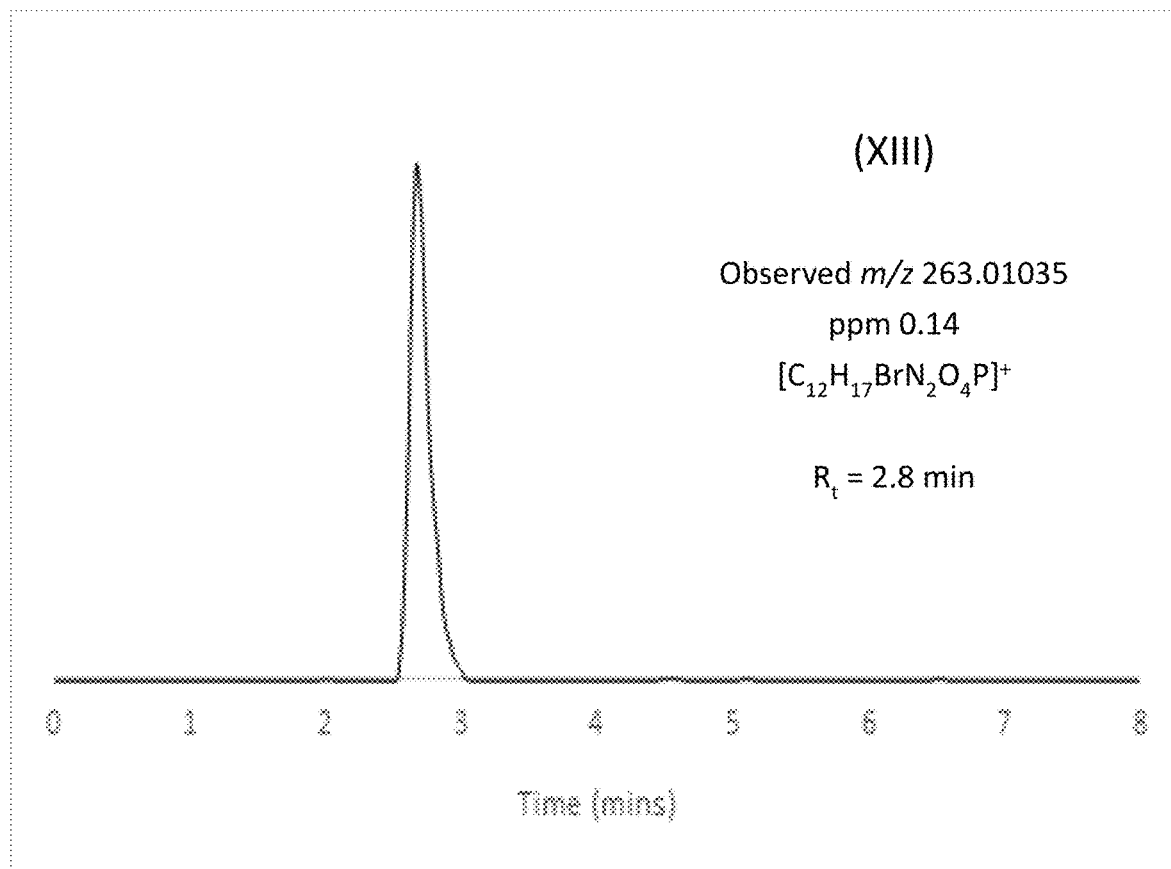
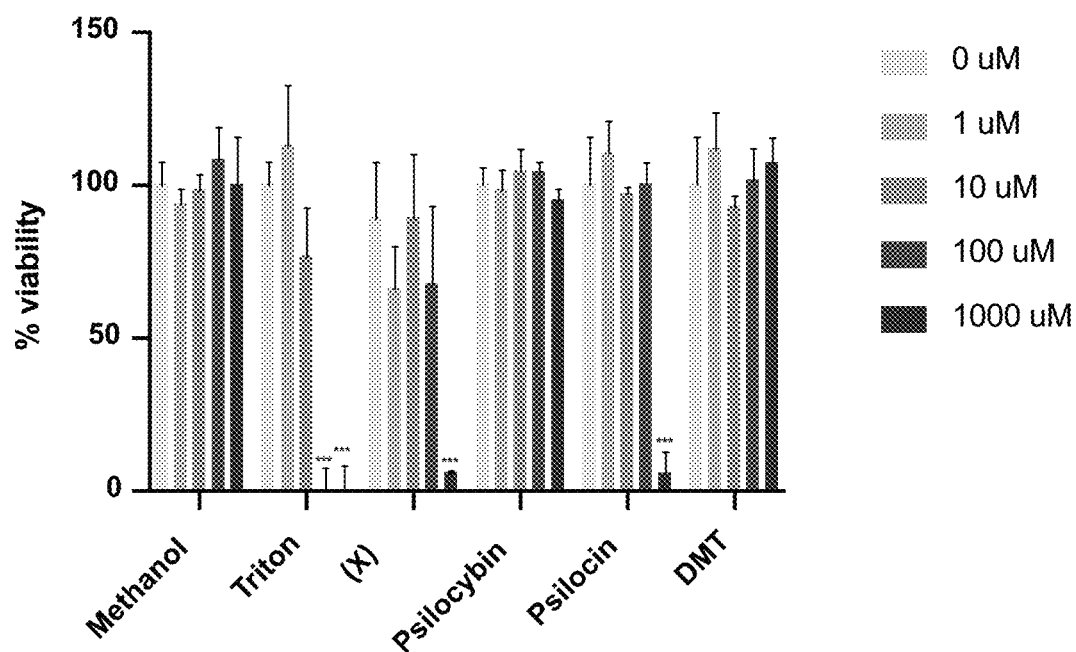
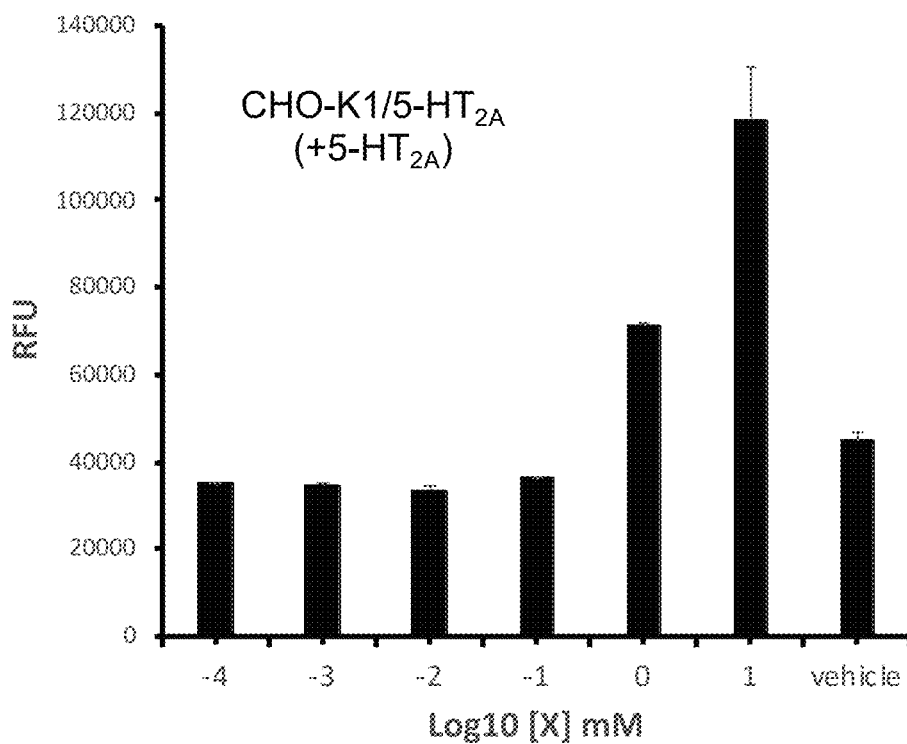
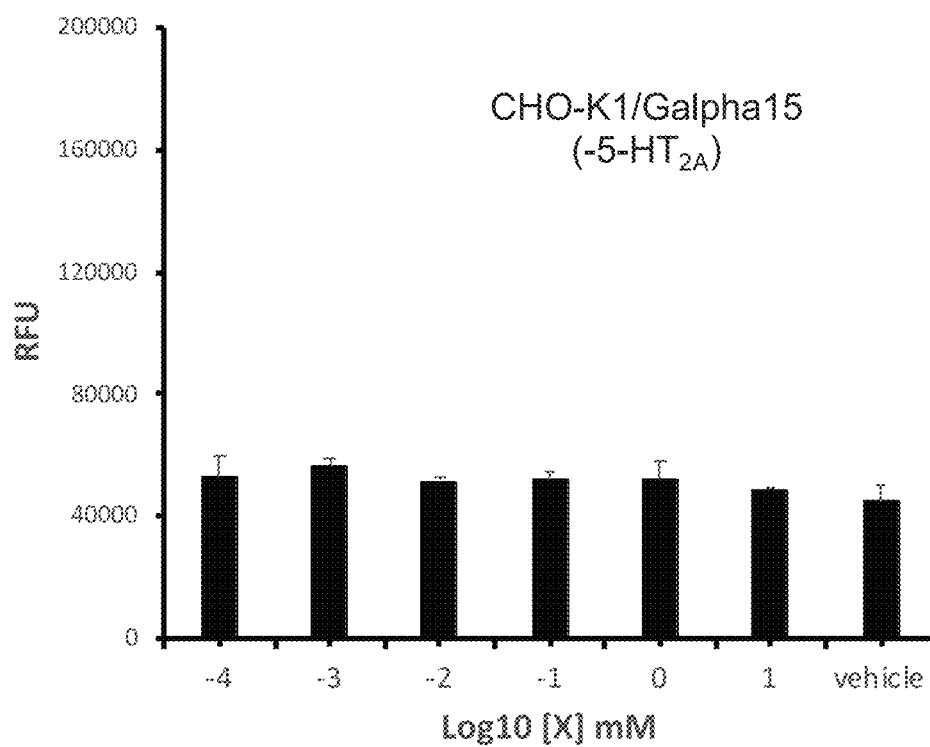
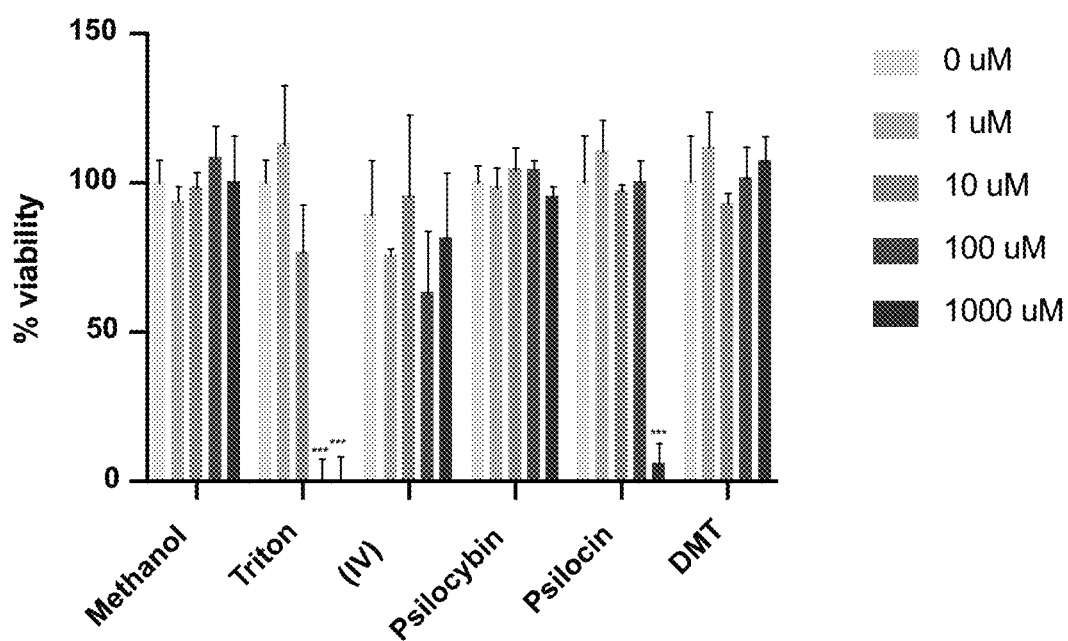


FIG. 22B

**FIG. 23**

**FIG. 24**

**FIG. 25A****FIG. 25B**

**FIG. 26**

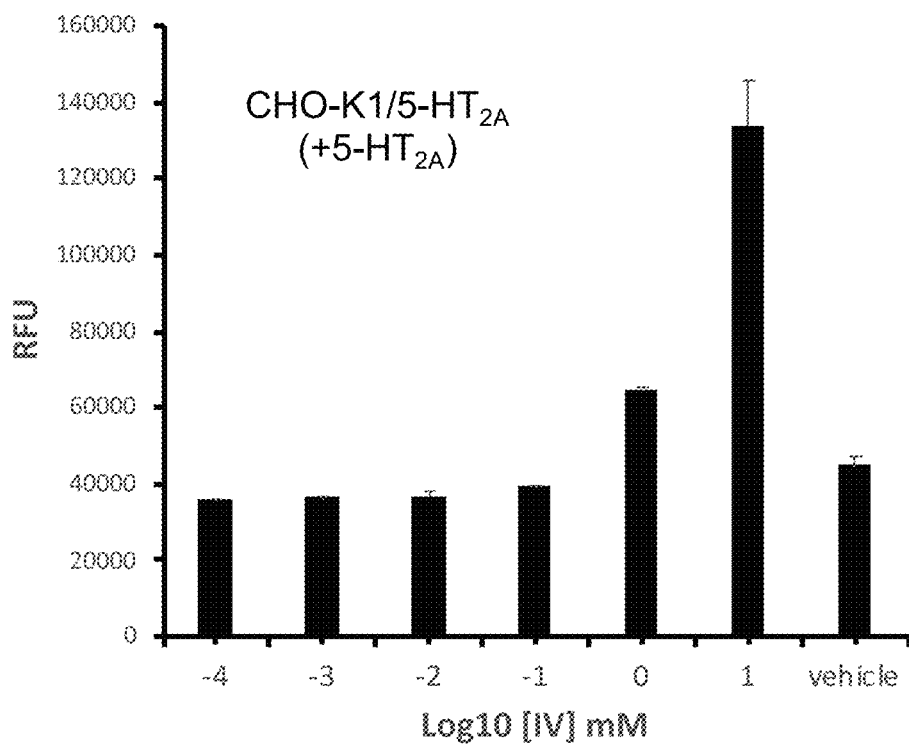


FIG. 27A

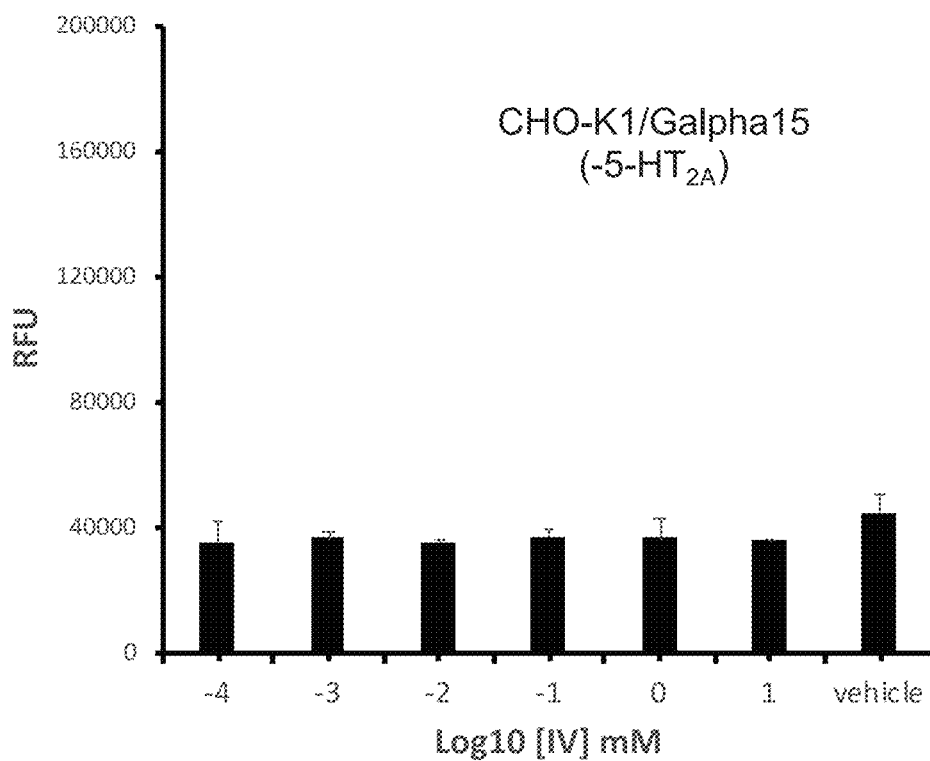


FIG. 27B

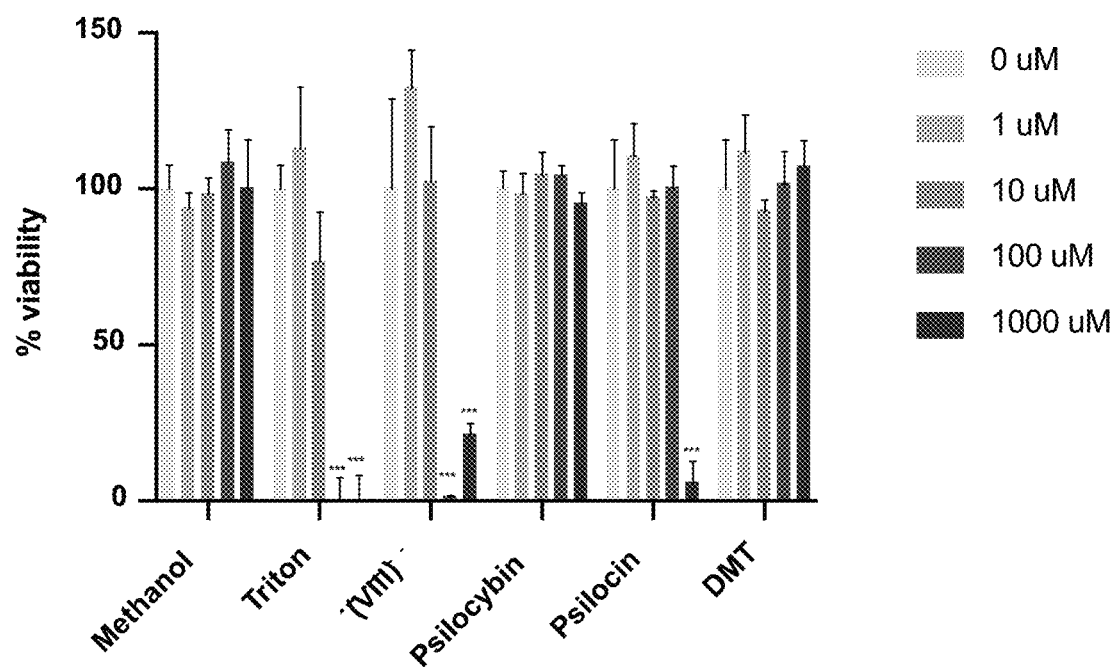


FIG. 28

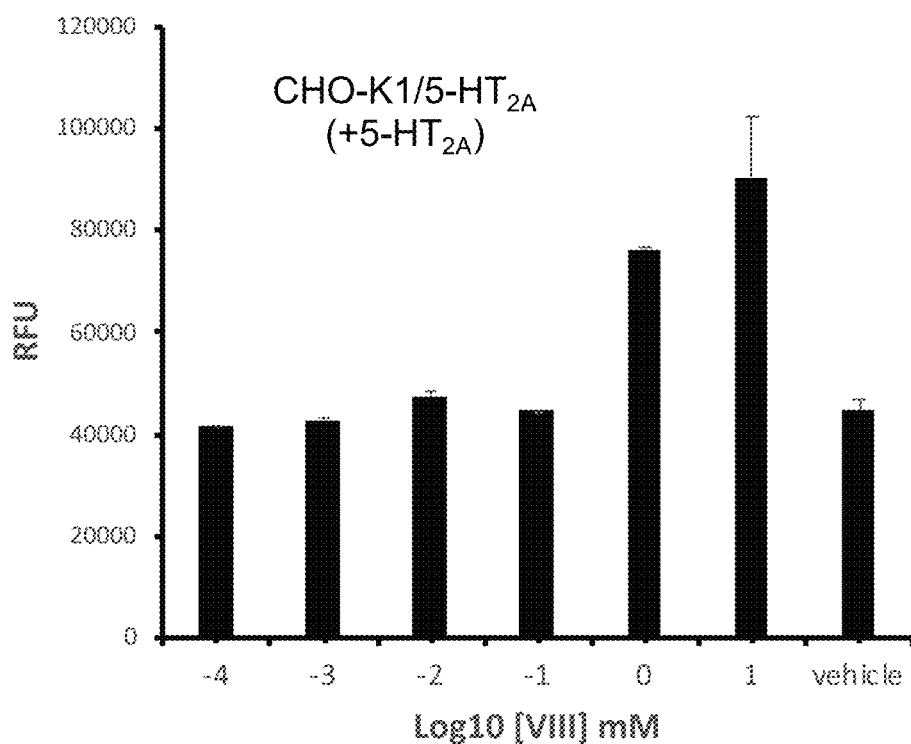


FIG. 29A

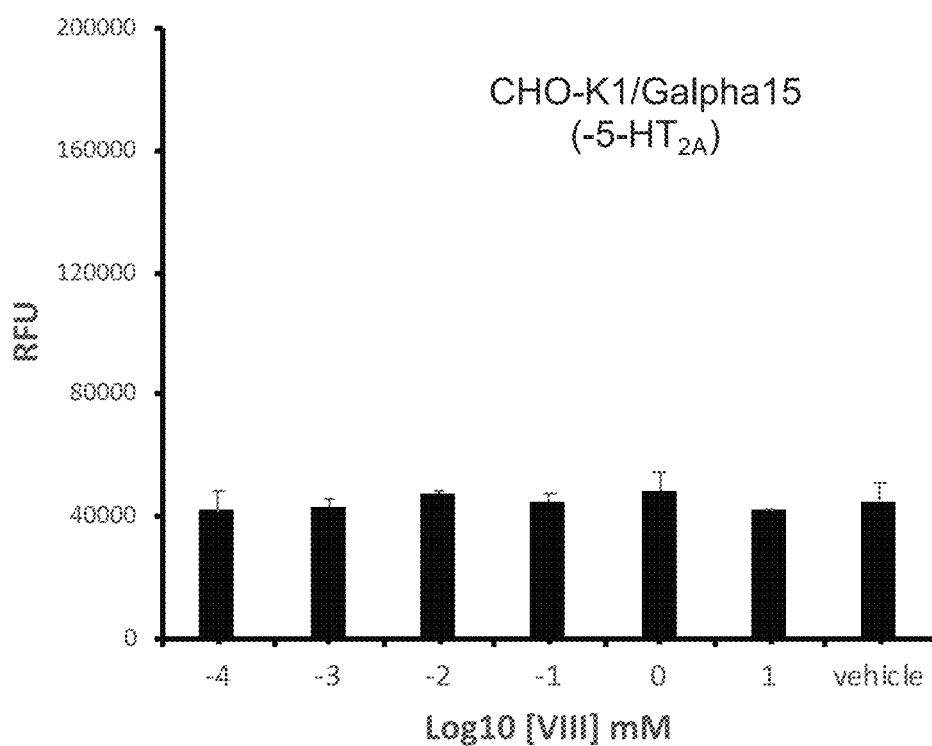


FIG. 29B

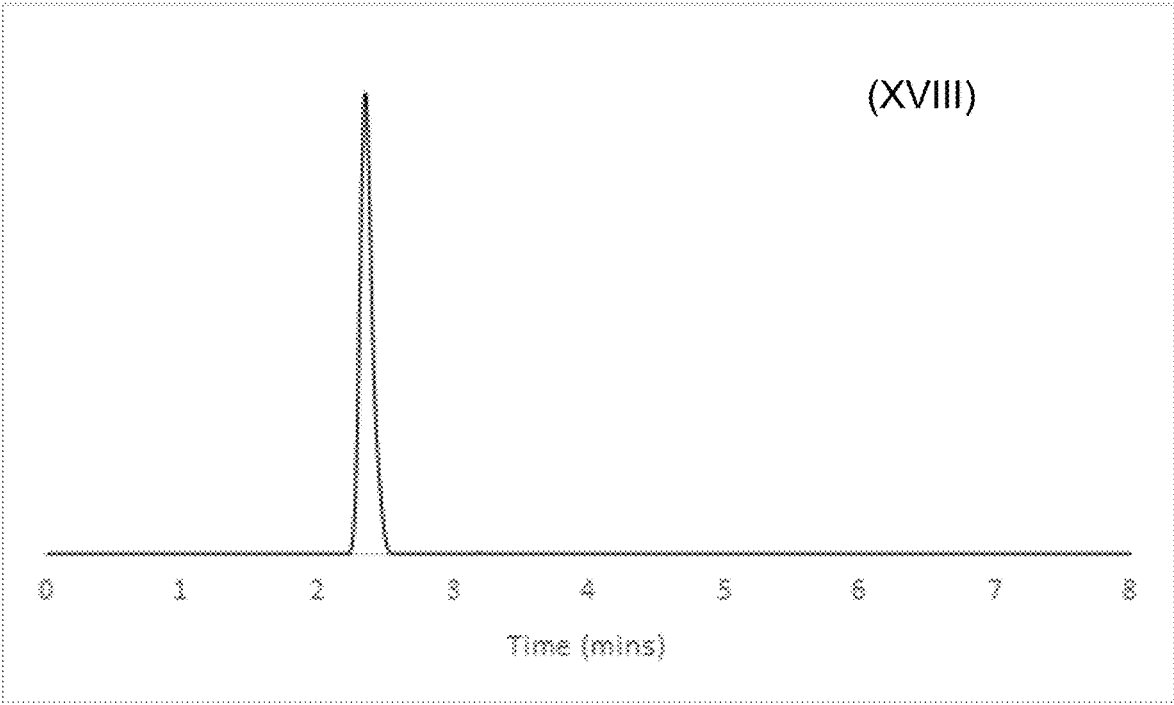


FIG. 30

1

HALOGENATED PSILOCYBIN DERIVATIVES AND METHODS OF USING

RELATED APPLICATION

This application is a continuation of U.S. application Ser. No. 17/903,080 filed Sep. 6, 2022 which is a continuation of PCT Application No. PCT/CA2021/051209 filed Sep. 1, 2021, which claims the benefit of U.S. Provisional Application No. 63/073,104 filed Sep. 1, 2020; the entire contents of which are hereby incorporated by reference.

INCORPORATION OF SEQUENCE LISTING

A computer readable form of the Sequence Listing “29664-P62051 US03_SequenceListing.xml” (216,574 bytes), submitted via Patent Center and created on May 6, 2024, is herein incorporated by reference.

FIELD OF THE DISCLOSURE

The compositions and methods disclosed herein relate to a chemical compound known as psilocybin. Furthermore, the compositions and methods disclosed herein relate in particular to halogenated forms of psilocybin.

BACKGROUND OF THE DISCLOSURE

The following paragraphs are provided by way of background to the present disclosure. They are not however an admission that anything discussed therein is prior art or part of the knowledge of a person of skill in the art.

The biochemical pathways in the cells of living organisms may be classified as being part of primary metabolism, or as being part of secondary metabolism. Pathways that are part of a cell's primary metabolism are involved in catabolism for energy production or in anabolism for building block production for the cell. Secondary metabolites, on the other hand, are produced by the cell without having an obvious anabolic or catabolic function. It has long been recognized that secondary metabolites can be useful in many respects, including as therapeutic compounds.

Psilocybin, for example, is a secondary metabolite that is naturally produced by certain mushrooms which taxonomically can be classified as belonging to the Basidiomycota division of the fungi kingdom. Mushroom species which can produce psilocybin include species belonging to the genus *Psilocybe*, such as *Psilocybe azurescens*, *Psilocybe semilanceata*, *Psilocybe serbica*, *Psilocybe mexicana*, and *Psilocybe cyanescens*, for example. The interest of the art in psilocybin is well established. Thus, for example, psilocybin is a psychoactive compound and is therefore used as a recreational drug. Furthermore, psilocybin is used as a research tool in behavioral and neuro-imaging studies in psychotic disorders, and has been evaluated for its clinical potential in the treatment of mental health conditions (Daniel, J. et al. *Mental Health Clin.*, 2017; 7(1): 24-28), including to treat anxiety in terminal cancer patients (Grob, C. et al. *Arch. Gen. Psychiatry*, 2011, 68(1) 71-78) and to alleviate symptoms of treatment-resistant depression (Cathart-Harris, R. L. et al. *Lancet Psychiatry*, 2016, 3: 619-627).

Although the toxicity of psilocybin is low, adverse side effects, including, for example, panic attacks, paranoia and psychotic states, sometimes together or individually referred to as “a bad trip”, are not infrequently experienced by recreational psilocybin users.

2

There exists therefore a need in the art for improved psilocybin compounds.

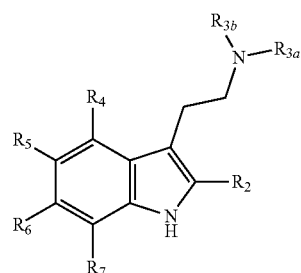
SUMMARY OF THE DISCLOSURE

The following paragraphs are intended to introduce the reader to the more detailed description, not to define or limit the claimed subject matter of the present disclosure.

In one aspect, the present disclosure relates to psilocybin and derivative compounds.

In another aspect, the present disclosure relates to halogenated psilocybin derivative compounds and methods of making and using these compounds.

Accordingly, in one aspect, the present disclosure provides, in accordance with the teachings herein, in at least one embodiment, a chemical compound or salts thereof having the formula (I):



(I)

wherein, at least one of R₂, R₄, R₅, R₆ or R₇ is a halogen atom, wherein each non-halogenated R₂, R₅, R₆, or R₇ is a hydrogen atom or an alkyl group or O-alkyl group, wherein R₄ when it is not halogenated is a phosphate group, hydrogen atom or an alkyl or O-alkyl group, and wherein R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group.

In at least one embodiment, in an aspect, the halogen atom can be a bromine atom.

In at least one embodiment, in an aspect, the halogen atom can be a chlorine atom.

In at least one embodiment, in an aspect, the halogen atom can be a fluorine atom.

In at least one embodiment, in an aspect, the halogen atom can be an iodine atom.

In at least one embodiment, in an aspect, R₂ can be a halogen atom and R₅, R₆ and R₇ can each be a hydrogen atom or an alkyl group or O-alkyl group, and R₄ can be a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, R₄ can be a halogen atom and R₂, R₅, R₆ and R₇ can each be a hydrogen atom or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, R₅ can be a halogen atom and R₂, R₆ and R₇ can each be a hydrogen atom or an alkyl group or O-alkyl group, and R₄ can be a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, R₆ can be a halogen atom and R₂, R₅ and R₇ can each be a hydrogen atom or an alkyl group or O-alkyl group, and R₄ can be a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, R₇ can be a halogen atom and R₂, R₅ and R₆ can each be a hydrogen

3

atom or an alkyl group or O-alkyl group, and R₄ can be a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, at least two of R₂, R₄, R₅, R₆ or R₇ can be a halogen atom.

In at least one embodiment, in an aspect, the at least of two of R₂, R₄, R₅, R₆ or R₇ that are halogen atoms can be at least two identical halogen atoms.

In at least one embodiment, in an aspect, the at least of two of R₂, R₄, R₅, R₆ or R₇ that are halogen atoms can at least two non-identical halogen atoms.

In at least one embodiment, in an aspect, R₂ and R₄ can be a halogen atom, and R₅, R₆ and R₇ can be a hydrogen atom or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, R₂ and R₅ can be a halogen atom, R₆ and R₇ can be a hydrogen atom or an alkyl group or O-alkyl group, and R₄ can be a hydrogen atom, a phosphate group or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, R₂ and R₆ can be a halogen atom, R₅ and R₇ can be a hydrogen atom or an alkyl group or O-alkyl group, and R₄ can be a hydrogen atom, a phosphate group or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, R₂ and R₇ can be a halogen atom, R₅ and R₆ can be a hydrogen atom or an alkyl group or O-alkyl group, and R₄ can be a hydrogen atom, a phosphate group or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, R₄ and R₅ can be a halogen atom, and R₂, R₆ and R₇ can be a hydrogen atom or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, R₄ and R₆ can be a halogen atom, and R₂, R₅ and R₇ can be a hydrogen atom or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, R₄ and R₇ can be a halogen atom, and R₂, R₅ and R₆ can be a hydrogen atom or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, R₅ and R₆ can be a halogen atom, R₂ and R₇ can be a hydrogen atom or an alkyl group or O-alkyl group, and R₄ can be a hydrogen atom, a phosphate group or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, R₅ and R₇ can be a halogen atom, R₂ and R₆ can be a hydrogen atom or an alkyl group or O-alkyl group, and R₄ can be a hydrogen atom, a phosphate group or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, R₆ and R₇ can be a halogen atom, R₂ and R₅ can be a hydrogen atom or an alkyl group or O-alkyl group, and R₄ can be a hydrogen atom, a phosphate group or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, R₄ when it is not halogenated can be a hydrogen atom.

In at least one embodiment, in an aspect, R₄ when it is not halogenated can be an alkyl group.

In at least one embodiment, in an aspect, R₄ when it is not halogenated can be an O-alkyl group.

In at least one embodiment, in an aspect, R₄ when it is not halogenated can be a phosphate group.

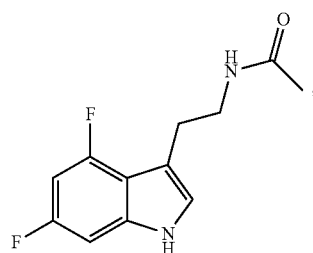
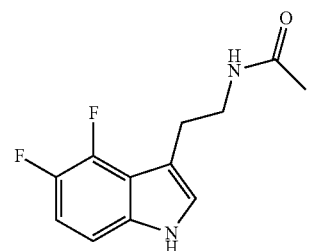
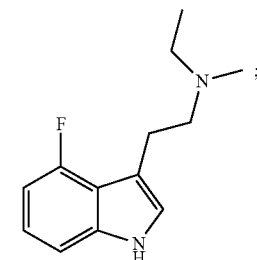
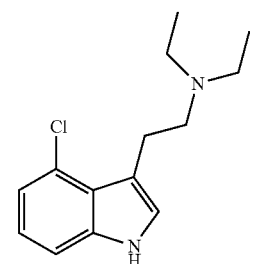
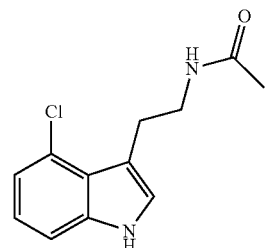
In at least one embodiment, in an aspect, three, four or all five of R₂, R₄, R₅, R₆ or R₇ can be a halogen atom.

In at least one embodiment, in an aspect, the three, four, or five of R₂, R₄, R₅, R₆ or R₇ that are halogen atoms can all be identical halogen atoms.

In at least one embodiment, in an aspect, the three, four or five of R₂, R₄, R₅, R₆ or R₇ that are halogen atoms, can include at least two non-identical halogen atoms.

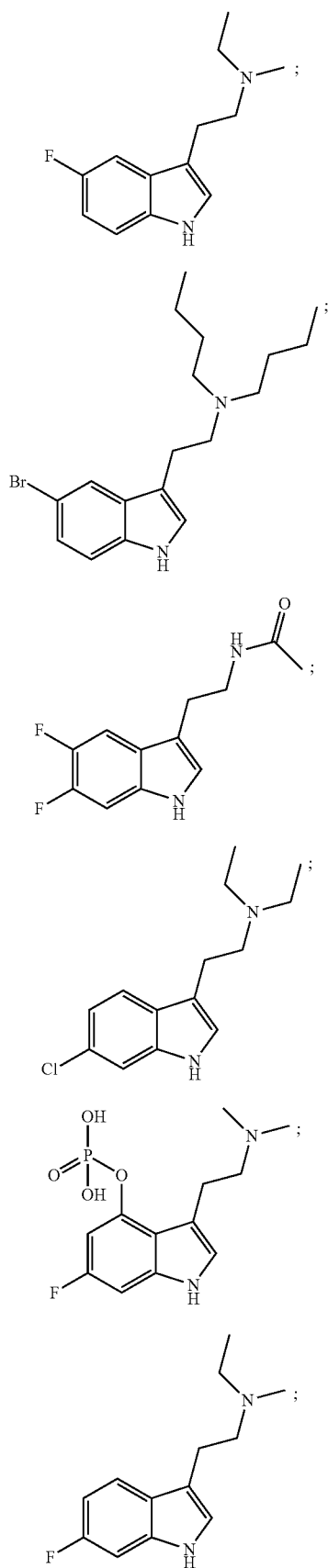
4

In at least one embodiment, in an aspect, the chemical compound can be selected from the group consisting of compounds having formulas (II); (III); (IV); (V); (VI); (VII); (VIII); (IX); (X); (XI); (XII); (XIII); (XIV); (XV); (XVI); and (XVII):

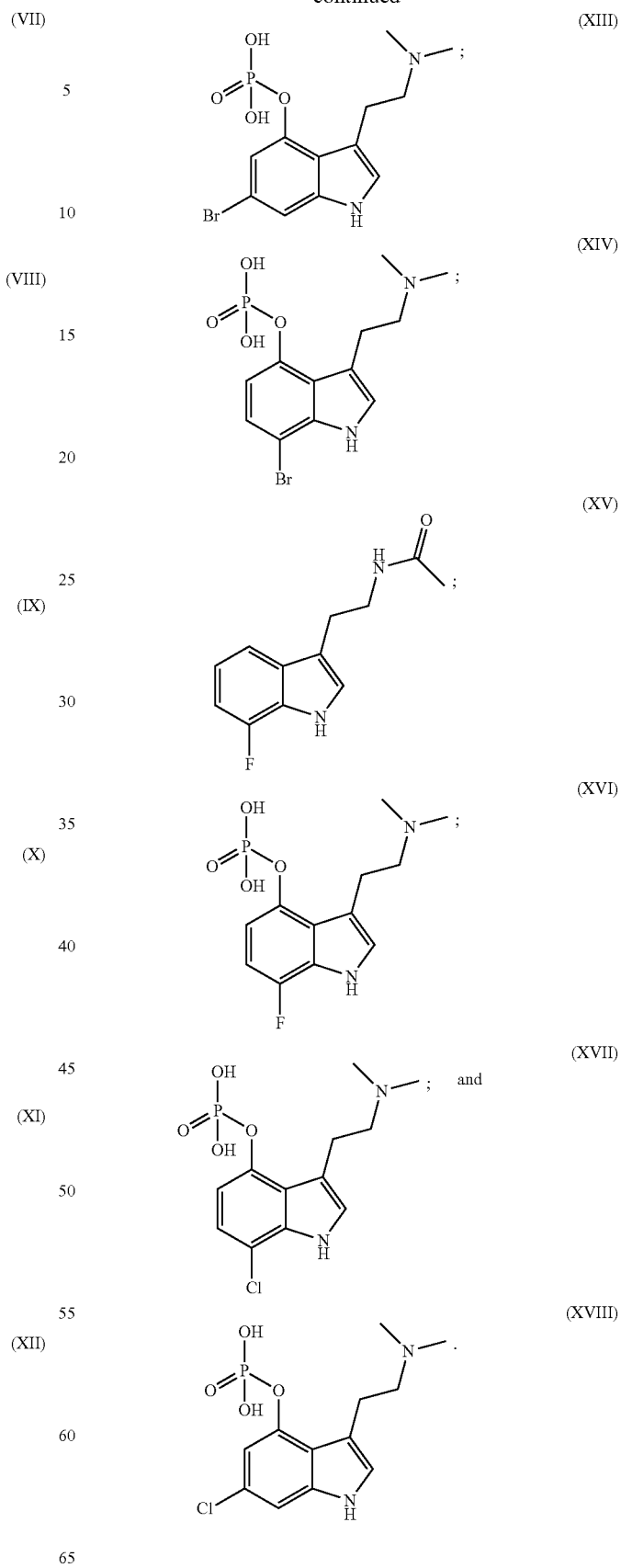


5

-continued

**6**

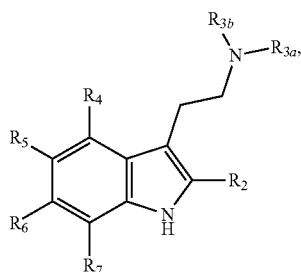
-continued



In at least one embodiment, the chemical compound is at least about 95% (w/w) pure.

7

In another aspect, the present disclosure relates to pharmaceutical and recreational drug formulations comprising halogenated psilocybin derivatives. Accordingly, in one aspect, the present disclosure provides in at least one aspect, a pharmaceutical or recreational drug formulation comprising an effective amount of a chemical compound or salts thereof having the formula (I):



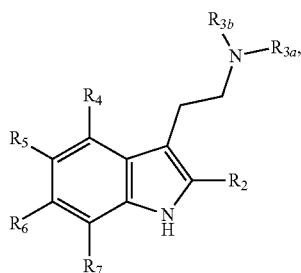
(I) 10

15

20

wherein, at least one of R_2 , R_4 , R_5 , R_6 or R_7 is a halogen atom, wherein each non-halogenated R_2 , R_5 , R_6 , or R_7 is a hydrogen atom or an alkyl group or O-alkyl group, wherein R_4 when it is not halogenated is a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group, and wherein R_{3A} and R_{3B} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group, together with a pharmaceutically acceptable excipient, diluent or carrier.

In another aspect, the present disclosure relates to methods of treatment of psychiatric disorders. Accordingly, the present disclosure further provide a method for treating a psychiatric disorder, the method comprising administering to a subject in need thereof a pharmaceutical formulation comprising a chemical compound or salts thereof having the formula (I):



(I) 40

45

50

wherein, at least one of R_2 , R_4 , R_5 , R_6 or R_7 is a halogen atom, wherein each non-halogenated R_2 , R_5 , R_6 , or R_7 is a hydrogen atom or an alkyl group or O-alkyl group, wherein R_4 when it is not halogenated is a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group, and wherein R_{3A} and R_{3B} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group, wherein the pharmaceutical formulation is administered in an effective amount to treat the psychiatric disorder in the subject.

In at least one embodiment, in an aspect the disorder can be a 5-HT_{2A} receptor mediated disorder.

In at least one embodiment, in an aspect, a dose can be administered of about 0.001 mg to about 5,000 mg.

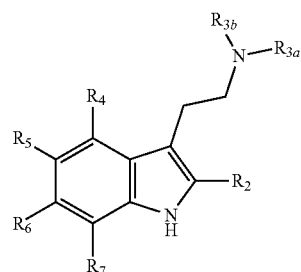
In another aspect, the present disclosure relates to methods of making halogenated psilocybin derivatives. Accordingly, in one aspect, the present disclosure provides in at

8

least one aspect, a method of making a halogenated psilocybin derivative the method comprising:

- (a) contacting a halogenated psilocybin precursor compound with a host cell comprising a psilocybin biosynthetic enzyme complement; and
- (b) growing the host cell to produce a halogenated psilocybin derivative or salts thereof having the formula (I):

(I)



wherein, at least one of R_2 , R_4 , R_5 , R_6 or R_7 is a halogen atom, wherein each non-halogenated R_2 , R_5 , R_6 , or R_7 is a hydrogen atom or an alkyl group or O-alkyl group, wherein R_4 when it is not halogenated is a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group, and wherein R_{3A} and R_{3B} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group.

In at least one embodiment, in an aspect, the halogenated psilocybin precursor compound can be contacted with the host cell by including the psilocybin precursor compound in a growth medium for the host cell.

In at least one embodiment, in an aspect, the halogenated psilocybin precursor compound can be formed by contacting the host cell with a non-halogenated psilocybin precursor compound and a halogen, the host cell further comprising a halogenase capable of halogenating the non-halogenated psilocybin compound and forming the halogenated psilocybin precursor compound.

In at least one embodiment, in an aspect, R_2 can be a halogen atom and R_5 , R_6 and R_7 can each be a hydrogen atom, and R_4 can be a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, R_4 can be a halogen atom and R_2 , R_5 , R_6 and R_7 can each be a hydrogen atom or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, R_5 can be a halogen atom and R_2 , R_6 and R_7 can each be a hydrogen atom or an alkyl group or O-alkyl group, and R_4 can be a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, R_6 can be a halogen atom and R_2 , R_5 and R_7 can each be a hydrogen atom or an alkyl group or O-alkyl group, and R_4 can be a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, R_7 can be a halogen atom and R_2 , R_5 and R_6 can each be a hydrogen atom or an alkyl group or O-alkyl group, and R_4 can be a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, at least two of R_2 , R_4 , R_5 , R_6 or R_7 can be a halogen atom.

In at least one embodiment, in an aspect, the at least of two of R_2 , R_4 , R_5 , R_6 or R_7 that are halogen atoms can be two identical halogen atoms.

9

In at least one embodiment, in an aspect, the at least of two of R_2 , R_4 , R_5 , R_6 or R_7 that are halogen atoms can be two non-identical halogen atoms.

In at least one embodiment, in an aspect, R_2 and R_4 can be a halogen atom, and R_5 , R_6 and R_7 can be a hydrogen atom or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, R_2 and R_5 can be a halogen atom, R_6 and R_7 can be a hydrogen atom or an alkyl group or O-alkyl group, and R_4 can be a hydrogen atom, an alkyl group or O-alkyl group or a phosphate group.

In at least one embodiment, in an aspect, R_2 and R_6 can be a halogen atom, R_5 and R_7 can be a hydrogen atom or an alkyl group or O-alkyl group, and R_4 can be a hydrogen atom, an alkyl group or O-alkyl group or a phosphate group.

In at least one embodiment, in an aspect, R_2 and R_7 can be a halogen atom, R_5 and R_6 can be a hydrogen atom or an alkyl group or O-alkyl group, and R_4 can be a hydrogen atom, an alkyl group or O-alkyl group or a phosphate group.

In at least one embodiment, in an aspect, R_4 and R_5 can be a halogen atom, and R_2 , R_6 and R_7 can be a hydrogen atom or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, R_4 and R_6 can be a halogen atom, and R_2 , R_5 and R_7 can be a hydrogen atom or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, R_4 and R_7 can be a halogen atom, and R_2 , R_5 and R_6 can be a hydrogen atom or an alkyl group or O-alkyl group.

In at least one embodiment, in an aspect, R_5 and R_6 can be a halogen atom, R_2 and R_7 can be a hydrogen atom or an alkyl group or O-alkyl group, and R_4 can be a hydrogen atom, an alkyl group or O-alkyl group or a phosphate group.

In at least one embodiment, in an aspect, R_5 and R_7 can be a halogen atom, R_2 and R_6 can be a hydrogen atom or an alkyl group or O-alkyl group, and R_4 can be a hydrogen atom, an alkyl group or O-alkyl group or a phosphate group.

In at least one embodiment, in an aspect, R_6 and R_7 can be a halogen atom, R_2 and R_5 can be a hydrogen atom or an alkyl group or O-alkyl group, and R_4 can be a hydrogen atom, an alkyl group or O-alkyl group or a phosphate group.

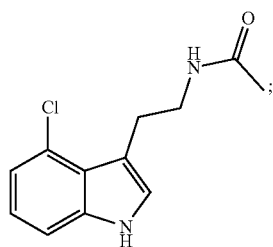
In at least one embodiment, in an aspect, R_4 when it is not halogenated can be a hydrogen atom.

In at least one embodiment, in an aspect, R_4 when it is not halogenated can be an alkyl group.

In at least one embodiment, in an aspect, R_4 when it is not halogenated can be an O-alkyl group.

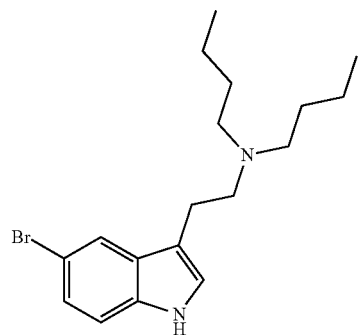
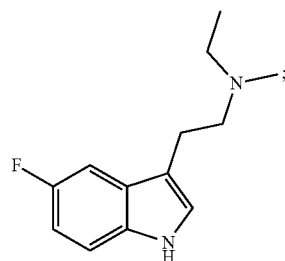
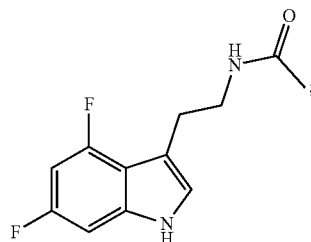
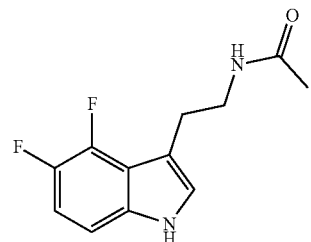
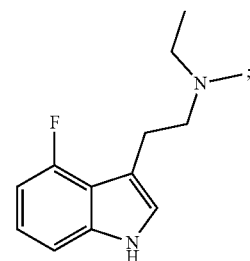
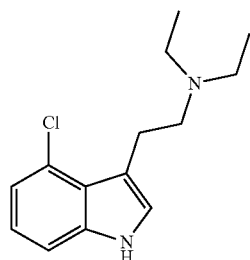
In at least one embodiment, in an aspect, R_4 when it is not halogenated can be a phosphate group.

In at least one embodiment, in an aspect, the halogenated psilocybin derivative can be a derivative selected from the group consisting of derivatives having formulas (II); (III); (IV); (V); (VI); (VII); (VIII); (IX); (X); (XI); (XII); (XIII); (XIV); (XV); (XVI); and (XVII):



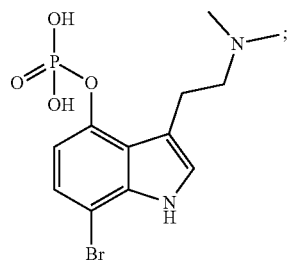
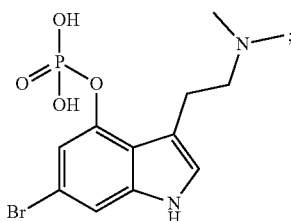
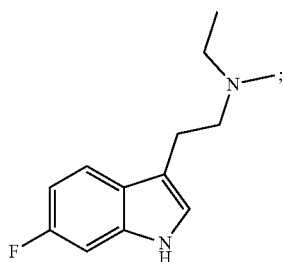
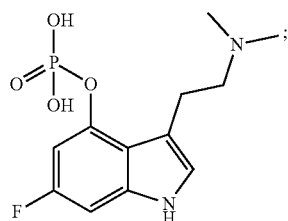
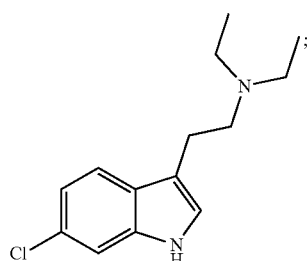
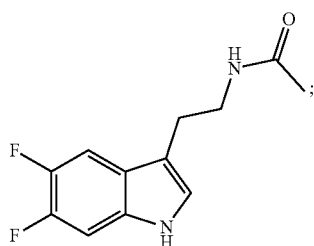
10

-continued



11

-continued

**12**

-continued

(IX)

5

10

(X)

15

20

(XI)

25

30

(XII)

35

40

(XIII)

45

(XIV)

50

55

60

65

(XV)

(XVI)

(XVII)

(XVIII)

In at least one embodiment, in an aspect, the halogenated psilocybin precursor compound can be selected from the group of halogenated compounds consisting of halogenated tryptophan, halogenated tryptamine, halogenated 4-hydroxytryptamine, halogenated 4-hydroxy-indole, halogenated 4-hydroxytryptophan, halogenated norbaeocystin, and halogenated baecocystin.

In at least one embodiment, in an aspect, the psilocybin biosynthetic enzyme complement can comprise at least one enzyme encoded by a nucleic acid selected from:

- (a) SEQ.ID NO: 1, SEQ.ID NO: 3, SEQ.ID NO: 5, SEQ.ID NO: 7, SEQ.ID NO: 9, SEQ.ID NO: 11, SEQ.ID NO: 36, SEQ.ID NO: 38, SEQ.ID NO: 40, SEQ. ID NO: 42, SEQ.ID NO: 44, SEQ.ID NO: 46, SEQ.ID NO: 48, SEQ.ID NO: 50, SEQ.ID NO: 52, SEQ.ID NO: 54, SEQ. ID NO: 56, SEQ.ID NO: 58, SEQ.ID NO: 60, SEQ.ID NO: 64, SEQ.ID NO: 66, SEQ.ID NO: 68, SEQ. ID NO: 70, and SEQ.ID NO: 72;
- (b) a nucleic acid sequence that is substantially identical to any one of the nucleic acid sequences of (a);
- (c) a nucleic acid sequence that is substantially identical to any one of the nucleic acid sequences of (a) but for the degeneration of the genetic code;

13

(d) a nucleic acid sequence that is complementary to any one of the nucleic acid sequences of (a);

(e) a nucleic acid sequence encoding a polypeptide having any one of the amino acid sequences set forth in SEQ.ID NO: 2, SEQ.ID NO: 4, SEQ.ID NO: 6, SEQ.ID NO: 8, SEQ.ID NO: 10, SEQ.ID NO: 12, SEQ.ID NO: 37, SEQ.ID NO: 39, SEQ.ID NO: 41, SEQ.ID NO: 43, SEQ.ID NO: 45, SEQ.ID NO: 47, SEQ.ID NO: 49, SEQ.ID NO: 51, SEQ.ID NO: 53, SEQ.ID NO: 55, SEQ.ID NO: 57, SEQ.ID NO: 59, SEQ.ID NO: 61, SEQ.ID NO: 65, SEQ.ID NO: 67, SEQ.ID NO: 69, SEQ.ID NO: 71, and SEQ.ID NO: 73;

(f) a nucleic acid sequence that encodes a functional variant of any one of the amino acid sequences set forth in SEQ.ID NO: 2, SEQ.ID NO: 4, SEQ.ID NO: 6, SEQ.ID NO: 8, SEQ.ID NO: 10, SEQ.ID NO: 12, SEQ.ID NO: 37, SEQ.ID NO: 39, SEQ.ID NO: 41, SEQ.ID NO: 43, SEQ.ID NO: 45, SEQ.ID NO: 47, SEQ.ID NO: 49, SEQ.ID NO: 51, SEQ.ID NO: 53, SEQ.ID NO: 55, SEQ.ID NO: 57, SEQ.ID NO: 59, SEQ.ID NO: 61, SEQ.ID NO: 65, SEQ.ID NO: 67, SEQ.ID NO: 69, SEQ.ID NO: 71, and SEQ.ID NO: 73; and

(g) a nucleic acid sequence that hybridizes under stringent conditions to any one of the nucleic acid sequences set forth in (a), (b), (c), (d), (e) or (f).

In at least one embodiment, in an aspect, the halogenase can be encoded by a nucleic acid selected from:

(a) SEQ.ID NO: 13, SEQ.ID NO: 15, SEQ.ID NO: 17, SEQ.ID NO: 19, SEQ.ID NO: 21, SEQ.ID NO: 23, SEQ.ID NO: 25, and SEQ.ID NO: 27;

(b) a nucleic acid sequence that is substantially identical to any one of the nucleic acid sequences of (a);

(c) a nucleic acid sequence that is substantially identical to any one of the nucleic acid sequences of (a) but for the degeneration of the genetic code;

(d) a nucleic acid sequence that is complementary to any one of the nucleic acid sequences of (a);

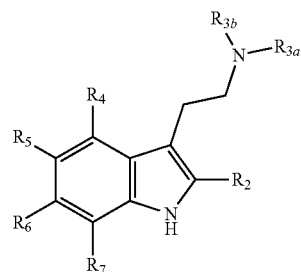
(e) a nucleic acid sequence encoding a polypeptide having any one of the amino acid sequences set forth in SEQ.ID NO: 14, SEQ.ID NO: 16, SEQ.ID NO: 18, SEQ.ID NO: 20, SEQ.ID NO: 22, SEQ.ID NO: 24, SEQ.ID NO: 26, SEQ.ID NO: 28, SEQ.ID NO: 29, SEQ.ID NO: 30, SEQ.ID NO: 31, SEQ.ID NO: 32, SEQ.ID NO: 33, SEQ.ID NO: 34, and SEQ.ID NO: 35;

(f) a nucleic acid sequence that encodes a functional variant of any one of the amino acid sequences set forth in SEQ.ID NO: 14, SEQ.ID NO: 16, SEQ.ID NO: 18, SEQ.ID NO: 20, SEQ.ID NO: 22, SEQ.ID NO: 24, SEQ.ID NO: 26, SEQ.ID NO: 28, SEQ.ID NO: 29, SEQ.ID NO: 30, SEQ.ID NO: 31, SEQ.ID NO: 32, SEQ.ID NO: 33, SEQ.ID NO: 34, and SEQ.ID NO: 35; and

(g) a nucleic acid sequence that hybridizes under stringent conditions to any one of the nucleic acid sequences set forth in (a), (b), (c), (d), (e) or (f).

In at least one embodiment, in an aspect, the host cell can further comprise an acetyl transferase capable of acetylating R_{3a} or R_{3b} , wherein the cell comprises a compound having chemical formula (I):

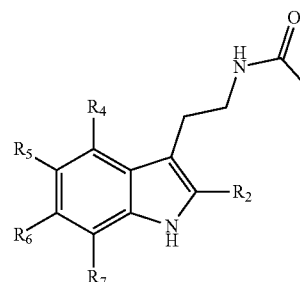
14



(I)

wherein, at least one of R_2 , R_4 , R_5 , R_6 or R_7 is a halogen atom, wherein each non-halogenated R_2 , R_5 , R_6 , or R_7 is a hydrogen atom or an alkyl group or O-alkyl group, wherein R_4 when it is not halogenated is a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group, and wherein at least one of R_{3a} and R_{3b} is a hydrogen atom,

to form a compound having the chemical formula (XIX):



(XIX)

wherein, at least one of R_2 , R_4 , R_5 , R_6 or R_7 is a halogen atom, wherein each non-halogenated R_2 , R_5 , R_6 , or R_7 is a hydrogen atom or an alkyl group or O-alkyl group, wherein R_4 when it is not halogenated is a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group.

In at least one embodiment, the acetylase can be encoded by a nucleic acid selected from:

(a) SEQ.ID NO: 62;

(b) a nucleic acid sequence that is substantially identical to the nucleic acid sequence of (a);

(c) a nucleic acid sequence that is substantially identical to the nucleic acid sequence of (a) but for the degeneration of the genetic code;

(d) a nucleic acid sequence that is complementary to the nucleic acid sequence of (a);

(e) a nucleic acid sequence encoding a polypeptide having an amino acid sequences set forth in SEQ.ID NO: 63;

(f) a nucleic acid sequence that encodes a functional variant of the amino acid sequence set forth in SEQ.ID NO: 63; and

(g) a nucleic acid sequence that hybridizes under stringent conditions to any one of the nucleic acid sequences set forth in (a), (b), (c), (d), (e) or (f).

In at least one embodiment, in an aspect,

(a) the halogenated psilocybin precursor compound can be a halogenated indole, and the a psilocybin biosynthetic enzyme complement can comprise (i) a nucleic acid encoding TrpB selected from SEQ.ID NO: 12, SEQ.ID NO: 65, SEQ.ID NO: 67, SEQ.ID NO: 69,

15

- SEQ.ID NO: 71, and SEQ.ID NO: 73, or a nucleic acid sequence encoding a polypeptide sequence substantially similar thereto; (ii) a nucleic acid encoding PsiD selected from SEQ.ID NO: 2, SEQ.ID NO: 10, SEQ.ID NO: 37, SEQ.ID NO: 39, SEQ.ID NO: 41, SEQ.ID NO: 43, SEQ.ID NO: 45, SEQ.ID NO: 47, SEQ.ID NO: 49, SEQ.ID NO: 51, SEQ.ID NO: 53, SEQ.ID NO: 55, and SEQ.ID NO: 57, or a nucleic acid sequence encoding a polypeptide sequence substantially similar thereto; (iii) a nucleic acid encoding PsiK encoded by SEQ.ID NO: 6, or a nucleic acid sequence encoding a polypeptide sequence substantially similar thereto; and (iv) a nucleic acid encoding PsiM selected from SEQ.ID NO: 8, SEQ.ID NO: 59 and SEQ.ID NO: 61 or a nucleic acid sequence encoding a polypeptide sequence substantially similar thereto; or
- (b) the halogenated psilocybin precursor compound can be a halogenated tryptophan, and the psilocybin biosynthetic enzyme complement can comprise (i) a nucleic acid encoding PsiD selected from SEQ.ID NO: 2, SEQ.ID NO: 10, SEQ.ID NO: 37, SEQ.ID NO: 39, SEQ.ID NO: 41, SEQ.ID NO: 43, SEQ.ID NO: 45, SEQ.ID NO: 47, SEQ.ID NO: 49, SEQ.ID NO: 51, SEQ.ID NO: 53, SEQ.ID NO: 55, and SEQ.ID NO: 57, or a nucleic acid sequence encoding a polypeptide sequence substantially similar thereto; (ii) a nucleic acid encoding PsiK encoded by SEQ.ID NO: 6, or a nucleic acid sequence encoding a polypeptide sequence substantially similar thereto; and (iii) a nucleic acid encoding PsiM selected from SEQ.ID NO: 8, SEQ.ID NO: 59 and SEQ.ID NO: 61 or a nucleic acid sequence encoding a polypeptide sequence substantially similar thereto; or
- (c) the halogenated psilocybin precursor compound can be a halogenated tryptamine, and the psilocybin biosynthetic enzyme complement can comprise (i) a nucleic acid encoding PsiK encoded by SEQ.ID NO: 6, or a nucleic acid sequence encoding a polypeptide sequence substantially similar thereto; and (ii) a nucleic acid encoding PsiM selected from SEQ.ID NO: 8, SEQ.ID NO: 59 and SEQ.ID NO: 61 or a nucleic acid sequence encoding a polypeptide sequence substantially similar thereto; or
- (d) the halogenated psilocybin precursor compound can be a halogenated norbaeocystin or a halogenated baeocystin and the psilocybin biosynthetic enzyme complement can comprise a nucleic acid encoding PsiM selected from SEQ.ID NO: 8, SEQ.ID NO: 59 and SEQ.ID NO: 61 or a nucleic acid sequence encoding a polypeptide sequence substantially similar thereto; or
- (e) the halogenated psilocybin precursor compound can be a halogenated tryptophan, and the psilocybin biosynthetic enzyme complement can comprise (i) a nucleic acid encoding PsiD selected from SEQ.ID NO: 2, SEQ.ID NO: 10, SEQ.ID NO: 37, SEQ.ID NO: 39, SEQ.ID NO: 41, SEQ.ID NO: 43, SEQ.ID NO: 45, SEQ.ID NO: 47, SEQ.ID NO: 49, SEQ.ID NO: 51, SEQ.ID NO: 53, SEQ.ID NO: 55, and SEQ.ID NO: 57, or a nucleic acid sequence encoding a polypeptide sequence substantially similar thereto; (ii) a nucleic acid encoding PsiH encoded by SEQ.ID NO: 4, or a nucleic acid sequence encoding a polypeptide sequence

16

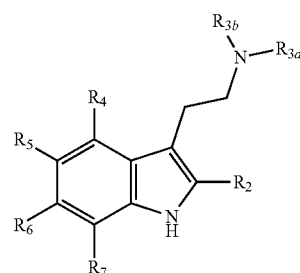
- substantially similar thereto; (iii) a nucleic acid encoding PsiK encoded by SEQ.ID NO: 6, or a nucleic acid sequence encoding a polypeptide sequence substantially similar thereto; and (iv) a nucleic acid encoding PsiM selected from SEQ.ID NO: 8, SEQ.ID NO: 59 and SEQ.ID NO: 61 or a nucleic acid sequence encoding a polypeptide sequence substantially similar thereto; or
- (f) the halogenated psilocybin precursor compound can be, a halogenated tryptamine, and psilocybin biosynthetic enzyme complement can comprise (i) a nucleic acid encoding PsiH encoded by SEQ.ID NO: 4, or a nucleic acid sequence encoding a polypeptide sequence substantially similar thereto; (ii) a nucleic acid encoding PsiK encoded by SEQ.ID NO: 6, or a nucleic acid sequence encoding a polypeptide sequence substantially similar thereto; and (iii) a nucleic acid encoding PsiM selected from SEQ.ID NO: 8, SEQ.ID NO: 59 and SEQ.ID NO: 61 or a nucleic acid sequence encoding a polypeptide sequence substantially similar thereto.

In at least one embodiment, in an aspect, the method can further include a step comprising isolating the halogenated psilocybin derivative.

In at least one embodiment, in an aspect, the host cell can be a microbial cell.

In at least one embodiment, in an aspect, the host cell can be a bacterial cell or a yeast cell.

In another aspect, the present disclosure provides, in at least one embodiment, a method for modulating a 5-HT_{2A} receptor, the method comprising contacting a 5-HT_{2A} receptor with a chemical compound or salts thereof having the formula (I):



(I)

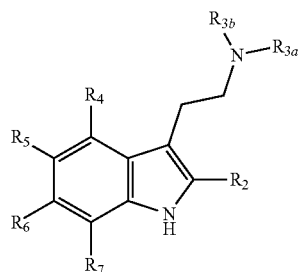
wherein, at least one of R₂, R₄, R₅, R₆ or R₇ is a halogen atom, wherein each non-halogenated R₂, R₅, R₆, or R₇ is a hydrogen atom or an alkyl group or O-alkyl group, wherein R₄ when it is not halogenated is a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group, and wherein R_{3A} and R_{3B} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group under reaction conditions sufficient to thereby modulate receptor activity.

In at least one embodiment, in an aspect, the reaction conditions can be in vitro reaction conditions.

In at least one embodiment, in an aspect, the reaction conditions can be in vivo reaction conditions.

In another aspect, the present disclosure provides, in at least one embodiment, a use of a chemical compound or salts thereof having the formula (I):

17

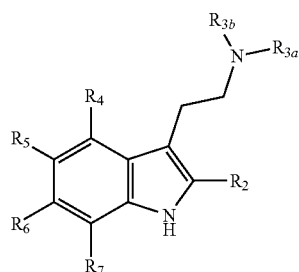


wherein, at least one of R_2 , R_4 , R_5 , R_6 or R_7 is a halogen atom, wherein each non-halogenated R_2 , R_5 , R_6 , or R_7 is a hydrogen atom or an alkyl group or O-alkyl group, wherein R_4 when it is not halogenated is a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group, and wherein R_{3A} and R_{3B} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group in the manufacture of a pharmaceutical or recreational drug formulation.

In at least one embodiment, in an aspect, the manufacture of a pharmaceutical or recreational drug formulation can comprise formulating the chemical compound with an excipient, diluent or carrier.

In at least one embodiment, the manufacture of a pharmaceutical or recreational drug formulation can comprise derivatizing the chemical compound having the formula (I) by substituting the halogen atom with another atom or chemical group.

In another aspect, the present disclosure provides, in at least one embodiment, a use of a chemical compound or salts thereof having the formula (I):



wherein, at least one of R_2 , R_4 , R_5 , R_6 or R_7 is a halogen atom, wherein each non-halogenated R_2 , R_5 , R_6 , or R_7 is a hydrogen atom or an alkyl group or O-alkyl group, wherein R_4 when it is not halogenated is a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group, and wherein R_{3A} and R_{3B} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group together with a diluent, carrier or excipient, as a pharmaceutical or recreational drug formulation.

Other features and advantages will become apparent from the following detailed description. It should be understood, however, that the detailed description, while indicating preferred implementations of the disclosure, are given by way of illustration only, since various changes and modifications within the spirit and scope of the disclosure will become apparent to those of skill in the art from the detailed description.

18

BRIEF DESCRIPTION OF THE DRAWINGS

The disclosure is in the hereinafter provided paragraphs described, by way of example, in relation to the attached figures. The figures provided herein are provided for a better understanding of the example embodiments and to show more clearly how the various embodiments may be carried into effect. The figures are not intended to limit the present disclosure.

FIG. 1 depicts the chemical structure of psilocybin.

FIG. 2 depicts a certain prototype structure of a psilocybin precursor compound, namely an indole. Certain carbon and nitrogen atoms may be referred herein by reference to their position within the indole structure, i.e. N_1 , C_2 , C_3 etc. The pertinent atom numbering is shown.

FIGS. 3A, 3B, 3C, 3D, 3E, 3F and 3G depict the chemical structures of example psilocybin precursor compounds, notably L-tryptophan (FIG. 3A), tryptamine (FIG. 3B), 4-hydroxytryptamine (FIG. 3C), L-4-hydroxytryptophan (FIG. 3D), 4-hydroxyindole (FIG. 3E), norbaeocystin (FIG. 3F), and baecocystin (FIG. 3G).

FIGS. 4A, 4B, 4C, 4D, 4E, 4F, 4G, 4H, 4I, 4J, 4K, 4L, 4M, 4N, 4O, 4P, 4Q and 4R depict the chemical structures of certain example halogenated psilocybin derivative compounds, notably a 2-chloro-4-phospho psilocybin derivative (FIG. 4A); a 2-bromo psilocybin derivative (FIG. 4B); a 4-iodo psilocybin derivative (FIG. 4C); a 4-fluoro psilocybin derivative (FIG. 4D); a 4-phospho-5-fluoro psilocybin derivative (FIG. 4E); a 5-chloro psilocybin derivative (FIG. 4F); a 4-phospho-6-bromo psilocybin derivative (FIG. 4G); a 6-chloro psilocybin derivative (FIG. 4H); a 4-phospho-7-fluoro psilocybin derivative (FIG. 4I); a 7-iodo psilocybin derivative (FIG. 4J); a 2-chloro-4-methyl psilocybin derivative (FIG. 4K); a 4-ethyl-5-bromo psilocybin derivative (FIG. 4L); a 4-methyl-6-fluoro psilocybin derivative (FIG. 4M); a 4-propyl-7-iodo psilocybin derivative (FIG. 4N); a 2-chloro-4-O-methyl psilocybin derivative (FIG. 4K); a 4-O-ethyl-5-bromo psilocybin derivative (FIG. 4L); a 4-O-methyl-6-fluoro psilocybin derivative (FIG. 4M); and a 4-O-propyl-7-iodo psilocybin derivative (FIG. 4N). It is noted that in each of FIGS. 4A, 4B, 4C, 4D, 4E, 4F, 4G, 4H, 4I, 4J, 4K, 4L, 4M, 4N, 4O, 4P, 4Q and 4R R_{3a} and R_{3b} can be a hydrogen atom, an alkyl group, an aryl group, or an acyl group.

FIGS. 5A, 5B, 5C, 5D, 5E, 5F, 5G, 5H, 5I and 5J depict the chemical structures of certain example halogenated psilocybin derivative compounds, notably a 2,4-di-chloro psilocybin derivative (FIG. 5A); a 2-bromo-5-fluoro-7-methyl psilocybin derivative (FIG. 5B); a 2-chloro-O-methyl-6-bromo psilocybin derivative (FIG. 5C); a 2-fluoro-4-phospho-7-iodo psilocybin derivative (FIG. 5D); a 4,5-di-bromo psilocybin derivative (FIG. 5E); a 4,6-di-chloro psilocybin derivative (FIG. 5F); a 4-chloro-5-fluoro psilocybin derivative (FIG. 5G); a 4-phospho-5-fluoro-6-bromo psilocybin derivative (FIG. 5H); a 4-phospho-5,7-di-fluoro psilocybin derivative (FIG. 5I); and a 6,7-di-iodo psilocybin derivative (FIG. 5J). It is noted that in each of FIGS. 5A, 5B, 5C, 5D, 5E, 5F, 5G, 5H, 5I and 5J R_{3a} and R_{3b} can be a hydrogen atom, an alkyl group, an aryl group, or an acyl group.

FIGS. 6A, 6B, 6C, 6D, 6E, and 6F depict the chemical structures of certain example halogenated psilocybin derivative compounds, notably a 2,4,5-tri-chloro psilocybin derivative (FIG. 6A); a 2-bromo-5,6-di-fluoro psilocybin derivative (FIG. 6B); a 2-chloro-5-bromo-7-fluoro psilocybin derivative (FIG. 6C); a 4,5-di-bromo-6-chloro psilocybin derivative (FIG. 6D); a 4,5,7-tri-bromo psilocybin

19

derivative (FIG. 6E); and a 4-phospho-5-fluoro-6-chloro-7-iodo psilocybin derivative (FIG. 6F). It is noted that in each of FIGS. 6A, 6B, 6C, 6D, 6E, and 6F R_{3a} and R_{3b} can be a hydrogen atom, an alkyl group, an aryl group, or an acyl group.

FIGS. 7A, 7B, 7C, 7D, and 7E depict the chemical structures of certain example halogenated psilocybin derivative compounds, notably a 2,4,5,6-tetra-chloro psilocybin derivative (FIG. 7A); a 4,7-di-chloro-5,6-di-fluoro psilocybin derivative (FIG. 7B); a 2-chloro-4-phospho-5-fluoro-6-chloro-7-iodo psilocybin derivative (FIG. 7C); a 2,4,7-tri-bromo-6-chloro psilocybin derivative (FIG. 7D); and a 2-fluoro-4,5,7-tri-bromo psilocybin derivative (FIG. 7E). It is noted that in each of FIGS. 7A, 7B, 7C, 7D, and 7E R_{3a} and R_{3b} can be a hydrogen atom, an alkyl group, an aryl group, or an acyl group.

FIGS. 8A and 8B depict example biosynthesis pathway relating to the synthesis of a halogenated psilocybin derivatives.

FIGS. 9A and 9B depict example chemical synthesis processes for the synthesis of certain example halogenated psilocybin derivatives, notably the compound denoted as 410B01 (FIG. 9A) and corresponding with a halogenated psilocybin derivative compound having the chemical formula (IV) set forth herein, and the compound denoted as 1 (FIG. 9B) and corresponding with a halogenated psilocybin derivative compound having the chemical formula (VIII) set forth herein.

FIG. 10 depicts a representation of mass spectrometry data in the form of a chromatogram, notably a chromatogram obtained in the performance of an experiment to synthesize an example halogenated psilocybin derivative compound having the chemical formula (XI) set forth herein.

FIG. 11 depicts a representation of mass spectrometry data in the form of a chromatogram, notably a chromatogram obtained in the performance of an experiment to synthesize an example halogenated psilocybin derivative compound having the chemical formula (XIV) set forth herein.

FIG. 12 depicts a representation of mass spectrometry data in the form of a chromatogram, notably a chromatogram obtained in the performance of an experiment to synthesize an example halogenated psilocybin derivative compound having the chemical formula (XVI) set forth herein.

FIG. 13 depicts a representation of mass spectrometry data in the form of a chromatogram, notably a chromatogram obtained in the performance of an experiment to synthesize an example halogenated psilocybin derivative compound having the chemical formula (XVII) set forth herein.

FIG. 14 depicts a graph obtained in the performance of an experimental assay to evaluate the efficacy of an example halogenated psilocybin derivative, notably a cell viability assay involving an example halogenated psilocybin derivative compound having the chemical formula (XV) set forth herein.

FIGS. 15A and 15B depict graphs obtained in the performance of an experimental assay to evaluate the efficacy of an example halogenated psilocybin derivative, notably a 5-HT_{2A} receptor modulation assay involving +5-HT_{2A} cells and an example halogenated psilocybin derivative compound having the chemical formula (XV) set forth herein (FIG. 15B) and a control (psilocin) (FIG. 15A).

FIGS. 16A and 16B depict graphs obtained in the performance of an experimental assay to evaluate the efficacy of

20

an example halogenated psilocybin derivative, notably a 5-HT_{2A} receptor modulation assay involving -5-HT_{2A} cells and an example halogenated psilocybin derivative compound having the chemical formula (XV) set forth herein (FIG. 16B) and a control (psilocin) (FIG. 16A).

FIG. 17 depicts a representation of mass spectrometry data in the form of a chromatogram, notably a chromatogram obtained in the performance of an experiment to synthesize an example halogenated psilocybin derivative compound having the chemical formula (II) set forth herein.

FIG. 18 depicts a graph obtained in the performance of an experimental assay to evaluate the efficacy of an example halogenated psilocybin derivative, notably a cell viability assay involving an example halogenated psilocybin derivative compound having the chemical formula (IX) set forth herein.

FIGS. 19A and 19B depict graphs obtained in the performance of an experimental assay to evaluate the efficacy of an example halogenated psilocybin derivative, notably a 5-HT_{2A} receptor modulation assay involving +5-HT_{2A} cells (FIG. 19A) and -5-HT_{2A} cells (FIG. 19B) and an example halogenated psilocybin derivative compound having the chemical formula (IX) set forth herein.

FIG. 20 depicts a representation of mass spectrometry data in the form of a chromatogram, notably a chromatogram obtained in the performance of an experiment to synthesize an example halogenated psilocybin derivative compound having the chemical formula (VI) set forth herein.

FIG. 21 depicts a graph obtained in the performance of an experimental assay to evaluate the efficacy of an example halogenated psilocybin derivative, notably a cell viability assay involving an example halogenated psilocybin derivative compound having the chemical formula (V) set forth herein.

FIGS. 22A and 22B depict graphs obtained in the performance of an experimental assay to evaluate the efficacy of an example halogenated psilocybin derivative, notably a 5-HT_{2A} receptor modulation assay involving +5-HT_{2A} cells (FIG. 22A) and -5-HT_{2A} cells (FIG. 22B), and an example halogenated psilocybin derivative compound having the chemical formula (V) set forth herein.

FIG. 23 depicts a representation of mass spectrometry data in the form of a chromatogram, notably a chromatogram obtained in the performance of an experiment to synthesize an example halogenated psilocybin derivative compound having the chemical formula (XIII) set forth herein.

FIG. 24 depicts a graph obtained in the performance of an experimental assay to evaluate the efficacy of an example halogenated psilocybin derivative, notably a cell viability assay involving an example halogenated psilocybin derivative compound having the chemical formula (X) set forth herein.

FIGS. 25A and 25B depict graphs obtained in the performance of an experimental assay to evaluate the efficacy of an example halogenated psilocybin derivative, notably a 5-HT_{2A} receptor modulation assay involving +5-HT_{2A} cells (FIG. 25A) and -5-HT_{2A} cells (FIG. 25B), and an example halogenated psilocybin derivative compound having the chemical formula (X) set forth herein.

FIG. 26 depicts a graph obtained in the performance of an experimental assay to evaluate the efficacy of an example halogenated psilocybin derivative, notably a cell viability assay involving an example halogenated psilocybin derivative compound having the chemical formula (IV) set forth herein.

21

FIGS. 27A and 27B depict graphs obtained in the performance of an experimental assay to evaluate the efficacy of an example halogenated psilocybin derivative, notably a 5-HT_{2A} receptor modulation assay involving +5-HT_{2A} cells (FIG. 27A) and -5-HT_{2A} cells (FIG. 27B), and an example halogenated psilocybin derivative compound having the chemical formula (IV) set forth herein.

FIG. 28 depicts a graph obtained in the performance of an experimental assay to evaluate the efficacy of an example halogenated psilocybin derivative, notably a cell viability assay involving an example halogenated psilocybin derivative compound having the chemical formula (VIII) set forth herein.

FIGS. 29A and 29B depict graphs obtained in the performance of an experimental assay to evaluate the efficacy of an example halogenated psilocybin derivative, notably a 5-HT_{2A} receptor modulation assay involving +5-HT_{2A} cells (FIG. 29A) and -5-HT_{2A} cells (FIG. 29B), and an example halogenated psilocybin derivative compound having the chemical formula (VIII) set forth herein.

FIG. 30 depicts a representation of mass spectrometry data in the form of a chromatogram, notably a chromatogram obtained in the performance of an experiment to synthesize an example halogenated psilocybin derivative compound having the chemical formula (XVIII) set forth herein.

The figures together with the following detailed description make apparent to those skilled in the art how the disclosure may be implemented in practice.

DETAILED DESCRIPTION

Various compositions, systems or processes will be described below to provide an example of an embodiment of each claimed subject matter. No embodiment described below limits any claimed subject matter and any claimed subject matter may cover processes, compositions or systems that differ from those described below. The claimed subject matter is not limited to compositions, processes or systems having all of the features of any one composition, system or process described below or to features common to multiple or all of the compositions, systems or processes described below. It is possible that a composition, system or process described below is not an embodiment of any claimed subject matter. Any subject matter disclosed in a composition, system or process described below that is not claimed in this document may be the subject matter of another protective instrument, for example, a continuing patent application, and the applicant(s), inventor(s) or owner(s) do not intend to abandon, disclaim or dedicate to the public any such subject matter by its disclosure in this document.

As used herein and in the claims, the singular forms, such as “a”, “an” and “the” include the plural reference and vice versa unless the context clearly indicates otherwise. Throughout this specification, unless otherwise indicated, “comprise,” “comprises” and “comprising” are used inclusively rather than exclusively, so that a stated integer or group of integers may include one or more other non-stated integers or groups of integers.

Various compositions, systems or processes will be described below to provide an example of an embodiment of each claimed subject matter. No embodiment described below limits any claimed subject matter and any claimed subject matter may cover processes, compositions or systems that differ from those described below. The claimed subject matter is not limited to compositions, processes or

22

systems having all of the features of any one composition, system or process described below or to features common to multiple or all of the compositions, systems or processes described below. It is possible that a composition, system or process described below is not an embodiment of any claimed subject matter. Any subject matter disclosed in a composition, system or process described below that is not claimed in this document may be the subject matter of another protective instrument, for example, a continuing patent application, and the applicant(s), inventor(s) or owner(s) do not intend to abandon, disclaim or dedicate to the public any such subject matter by its disclosure in this document.

When ranges are used herein for physical properties, such as molecular weight, or chemical properties, such as chemical formulae, all combinations and sub-combinations of ranges and specific embodiments therein are intended to be included. Other than in the operating examples, or where otherwise indicated, all numbers expressing quantities of ingredients or reaction conditions used herein should be understood as modified in all instances by the term “about.” The term “about” when referring to a number or a numerical range means that the number or numerical range referred to is an approximation within experimental variability (or within statistical experimental error), and thus the number or numerical range may vary between 1% and 15% of the stated number or numerical range, as will be readily recognized by context. Furthermore any range of values described herein is intended to specifically include the limiting values of the range, and any intermediate value or sub-range within the given range, and all such intermediate values and sub-ranges are individually and specifically disclosed (e.g. a range of 1 to 5 includes 1, 1.5, 2, 2.75, 3, 3.90, 4, and 5). Similarly, other terms of degree such as “substantially” and “approximately” as used herein mean a reasonable amount of deviation of the modified term such that the end result is not significantly changed. These terms of degree should be construed as including a deviation of the modified term if this deviation would not negate the meaning of the term it modifies.

Unless otherwise defined, scientific and technical terms used in connection with the formulations described herein shall have the meanings that are commonly understood by those of ordinary skill in the art. The terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention, which is defined solely by the claims.

All publications, patents and patent applications are herein incorporated by reference in their entirety to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety.

Terms and Definitions

The term “psilocybin”, as used herein, refers to a chemical compound having the structure set forth in FIG. 1, and further includes salts, thereof, such as a sodium salt, a potassium salt, etc.

The term “psilocybin precursor compound”, as used herein, refers to a chemical compound that may serve as a precursor compound in the synthesis or biosynthesis of psilocybin, and includes compounds comprising an indole prototype structure, including, for example, tryptophan, tryptamine, 4-hydroxyindole, 4-hydroxytryptophan, 4-hy-

droxytryptamine, norbaeocystin, and baeocystin, and further includes halogenated derivatives and salts of any of the foregoing.

The term "indole prototype structure", as used herein, refers to the chemical structure shown in FIG. 2. It is noted that specific carbon atoms and a nitrogen atom in the indole prototype structure are numbered. Reference may be made to these carbon and nitrogen numbers herein, for example C₂, C₄, N₁, and so forth. Furthermore, reference may be made to chemical groups attached to the indole prototype structure in accordance with the same numbering, for example R₄ and R₆ reference chemical groups attached to the C₄ and C₆ atom, respectively. In addition, R_{3A} and R_{3B}, in this respect, reference chemical groups extending from the ethyl-amino group extending in turn from the C₃ atom of the prototype indole structure.

The term "tryptophan", as used herein, refers to a chemical compound having the structure set forth in FIG. 3A and further includes its D-enantiomeric form (not shown).

The term "tryptamine", as used herein, refers to a chemical compound having the structure set forth in FIG. 3B.

The term "4-hydroxytryptamine", as used herein, refers to a chemical compound having the structure set forth in FIG. 3C.

The term "4-hydroxytryptophan", as used herein, refers to a chemical compound having the structure set forth in FIG. 3D and further includes its D-enantiomeric form (not shown).

The term "4-hydroxyindole", as used herein, refers to a chemical compound having the structure set forth in FIG. 3E.

The term "norbaeocystin", as used herein, refers to a chemical compound having the structure set forth in FIG. 3F.

The term "baeocystin", as used herein, refers to a chemical compound having the structure set forth in FIG. 3G.

The term "phosphate group", as used herein, is a molecule containing one atom of phosphorus, covalently bound to four oxygen atoms (three single bonds and one double bond). Of the four oxygen atoms one oxygen atom may be a hydroxy group, and one of the non-hydroxylated oxygen atom may be chemically bonded to another entity.

The terms "hydroxy group", and "hydroxy", as used herein, refers to a molecule containing one atom of oxygen bonded to one atom of hydrogen, and having the formula —OH. A hydroxy group through its oxygen atom may be chemically bonded to another entity.

The terms "halogen", "halogenated" and "halo-", as used herein, refer to the class of chemical elements consisting of fluorine (F), chlorine (Cl), bromine (Br), and iodine (I). Accordingly, halogenated compounds can refer to "fluorinated", "chlorinated", "brominated", or "iodinated" compounds.

The term "halogenated psilocybin derivative", as used herein, refers to a psilocybin derivative in which a halogen been bonded to psilocybin, or a derivative thereof. Reference may be made to specific carbon atoms of the psilocybin derivative which may be halogenated, for example, 7-halo-psilocybin refers to a halogenated psilocybin in which carbon atom number 7 (as identified in the indole prototype structure) is halogenated, or, similarly, 2-bromo-psilocybin refers to a brominated psilocybin in which carbon atom number 2 (as identified in the indole prototype structure) is brominated, and so on. The term further includes salts of halogenated psilocybin derivatives, such as sodium salt, potassium salt, or a chlorine salt, for example. Salts may be formed by virtue of a charged atom or group included in the

psilocybin derivative, notably a charged nitrogen atom, and/or a charged phosphate group.

The term "halogenated psilocybin precursor compound", as used herein, refers to a psilocybin precursor compound possessing a halogen. Reference may be made to specific carbon atoms of the psilocybin precursor compound which may be halogenated, for example, 6-halo-baeocystin refers to a halogenated baeocystin in which carbon atom number 6 (as identified in the indole prototype structure) is halogenated, or, similarly, 2-chloro-4-hydroxytryptamine refers to a chlorinated 4-hydroxytryptamine in which carbon atom number 2 (as identified in the indole prototype structure) is chlorinated. Furthermore, halogenated psilocybin precursor compounds include, without limitation, halogenated tryptophan, e.g. 7-bromo tryptophan, 6-chloro tryptophan, 5,6-difluoro tryptophan, halogenated 4-hydroxy indole e.g. 4-hydroxy-7-bromo indole, 4-hydroxy-5-chloro indole, 6,7-di-fluoro-indole

The term "alkyl", as used herein, refers to a straight and/or branched chain, saturated alkyl radical containing from one to "p" carbon atoms ("C₁-C_p-alkyl") and includes, depending on the identity of "p", methyl, ethyl, propyl, isopropyl n-butyl, s-butyl, isobutyl; t-butyl, 2,2-dimethylbutyl, n-pentyl, 2-methylpentyl, 3-methylpentyl, 4-methylpentyl, n-hexyl and the like, where the variable p is an integer representing the largest number of carbon atoms in the alkyl radical. Alkyl groups further include hydrocarbon groups arranged in a chain having the chemical formula —C_nH_{2n+1}, including, without limitation, methyl groups (—CH₃), ethyl groups (—C₂H₅), propyl groups (—C₃H₇), and butyl groups (—C₄H₉).

The term "O-alkyl", as used herein, refers to a hydrocarbon group arranged in a chain having the chemical formula —O—C_nH_{2n+1}. O-alkyl groups include, without limitation, O-methyl groups (—O—CH₃), O-ethyl groups (—O—C₂H₅), O-propyl groups (—O—C₃H₇) and O-butyl groups (—O—C₄H₉).

The term "acyl", as used herein, refers to a carbon atom double bonded to an oxygen and single bonded to an alkyl group. The carbon atom further can be bonded to another entity. An acyl group can be described by the chemical formula: —C(=O)—C_nH_{2n+1}.

The term "aryl", as used herein, refers to a monocyclic, bicyclic or tricyclic aromatic ring system containing, depending on the number of atoms in the rings, for example, from 6 to 14 carbon atoms (C₆-C₁₄-aryl) or from 6 to 10 carbons (C₆-C₁₀-aryl), and at least 1 aromatic ring and includes phenyl, naphthyl, anthracenyl, 1,2-dihydronaphthyl, 1,2,3,4-tetrahydronaphthyl, fluorenyl, phenanthrenyl, biphenylenyl, indanyl, indenyl and the like,

The term "5-HT_{2A} receptor", as used herein, refers to a subclass of a family of receptors for the neurotransmitter and peripheral signal mediator serotonin. 5-HT_{2A} receptors can mediate a plurality of central and peripheral physiologic functions of serotonin. Central nervous system effects can include mediation of hallucinogenic effects of hallucinogenic compounds.

The term "modulating 5-HT_{2A} receptors", as used herein, refers to the ability of a compound disclosed herein to alter the function of 5-HT_{2A} receptors. A 5-HT_{2A} receptor modulator may activate the activity of a 5-HT_{2A} receptor, may activate or inhibit the activity of a 5-HT_{2A} receptor depending on the concentration of the compound exposed to the 5-HT_{2A} receptor, or may inhibit the activity of a 5-HT_{2A} receptor. Such activation or inhibition may be contingent on the occurrence of a specific event, such as activation of a signal transduction pathway, and/or maybe manifest only in

particular cell types. The term “modulating 5-HT_{2A} receptors,” also refers to altering the function of a 5-HT_{2A} receptor by increasing or decreasing the probability that a complex forms between a 5-HT_{2A} receptor and a natural binding partner to form a multimer. A 5-HT_{2A} receptor modulator may increase the probability that such a complex forms between the 5-HT_{2A} receptor and the natural binding partner, may increase or decrease the probability that a complex forms between the 5-HT_{2A} receptor and the natural binding partner depending on the concentration of the compound exposed to the 5-HT_{2A} receptor, and or may decrease the probability that a complex forms between the 5-HT_{2A} receptor and the natural binding partner. Furthermore, the term includes allosteric modulation of the receptor 5-HT_{2A}, i.e. modulation of the 5-HT_{2A} receptor through interaction with the 5-HT_{2A} receptor that is topographically different than the orthosteric site recognized by the cell’s endogenous agonist, such modulation further including positive allosteric modulation (PAM), negative allosteric modulation (NAM) and silent allosteric modulation (SAM).

The term “5-HT_{2A} receptor-mediated disorder”, as used herein, refers to a disorder that is characterized by abnormal 5-HT_{2A} receptor activity. A 5-HT_{2A} receptor-mediated disorder may be completely or partially mediated by modulating 5-HT_{2A} receptors. In particular, a 5-HT_{2A} receptor-mediated disorder is one in which modulation of 5-HT_{2A} receptors results in some effect on the underlying disorder e.g., administration of a 5-HT_{2A} receptor modulator results in some improvement in at least some of the subjects being treated.

The term “pharmaceutical formulation”, as used herein, refers to a preparation in a form which allows an active ingredient, including a psychoactive ingredient, contained therein to provide effective treatment, and which does not contain any other ingredients which cause excessive toxicity, an allergic response, irritation, or other adverse response commensurate with a reasonable risk/benefit ratio. The pharmaceutical formulation may contain other pharmaceutical ingredients such as excipients, carriers, diluents, or auxiliary agents.

The term “recreational drug formulation”, as used herein, refers to a preparation in a form which allows a psychoactive ingredient contained therein to be effective for administration as a recreational drug, and which does not contain any other ingredients which cause excessive toxicity, an allergic response, irritation, or other adverse response commensurate with a reasonable risk/benefit ratio. The recreational drug formulation may contain other ingredients such as excipients, carriers, diluents, or auxiliary agents.

The term “effective for administration as a recreational drug”, as used herein, refers to a preparation in a form which allows a subject to voluntarily induce a psychoactive effect for non-medical purposes upon administration, generally in the form of self-administration. The effect may include an altered state of consciousness, satisfaction, pleasure, euphoria, perceptual distortion, or hallucination.

The term “effective amount”, as used herein, refers to an amount of an active agent, pharmaceutical formulation or recreational drug formulation, sufficient to induce a desired biological or therapeutic effect, including a prophylactic effect, and further including a psychoactive effect. Such effect can include an effect with respect to the signs, symptoms or causes of a disorder, or disease or any other desired alteration of a biological system. The effective amount can vary depending, for example, on the health condition, injury stage, disorder stage, or disease stage, weight, or sex of a subject being treated, timing of the administration, manner

of the administration, age of the subject, and the like, all of which can be determined by those of skill in the art.

The terms “treating” and “treatment”, and the like, as used herein, are intended to mean obtaining a desirable physiological, pharmacological, or biological effect, and includes prophylactic and therapeutic treatment. The effect may result in the inhibition, attenuation, amelioration, or reversal of a sign, symptom or cause of a disorder, or disease, attributable to the disorder, or disease, which includes mental and psychiatric diseases and disorders. Clinical evidence of the prevention or treatment may vary with the disorder, or disease, the subject and the selected treatment.

The term “pharmaceutically acceptable”, as used herein, refers to materials, including excipients, carriers, diluents, or auxiliary agents, that are compatible with other materials in a pharmaceutical or recreational drug formulation and within the scope of reasonable medical judgement suitable for use in contact with a subject without excessive toxicity, allergic response, irritation, or other adverse response commensurate with a reasonable risk/benefit ratio.

The term “psilocybin biosynthetic enzyme complement”, as used herein, refers to one or more polypeptides which alone or together are capable of facilitating the chemical conversion of a psilocybin precursor compound and form another psilocybin precursor compound or psilocybin or a halogenated form thereof. A psilocybin biosynthetic enzyme complement can include, for example, PsiD, PsiH, PsiK, PsiM, Psi-ncAAAD and TrpB.

The term “PsiD” as used herein, refers to any and all enzymes comprising a sequence of amino acid residues which is (i) substantially identical to the amino acid sequences constituting any PsiD polypeptide set forth herein, including, for example, SEQ.ID NO: 2, SEQ.ID NO: 37, SEQ.ID NO: 39, SEQ.ID NO: 41, SEQ.ID NO: 43, SEQ.ID NO: 45, SEQ.ID NO: 47, SEQ.ID NO: 49, SEQ.ID NO: 51, SEQ.ID NO: 53, SEQ.ID NO: 55, and SEQ.ID NO: 57 or (ii) encoded by a nucleic acid sequence capable of hybridizing under at least moderately stringent conditions to any nucleic acid sequence encoding any PsiD set forth herein, but for the use of synonymous codons.

The term “PsiH” as used herein, refers to any and all enzymes comprising a sequence of amino acid residues which is (i) substantially identical to the amino acid sequences constituting any PsiH polypeptide set forth herein, including, for example, SEQ.ID NO: 4, or (ii) encoded by a nucleic acid sequence capable of hybridizing under at least moderately stringent conditions to any nucleic acid sequence encoding any PsiH set forth herein, but for the use of synonymous codons.

The term “PsiK” as used herein, refers to any and all enzymes comprising a sequence of amino acid residues which is (i) substantially identical to the amino acid sequences constituting any PsiK polypeptide set forth herein, including, for example, SEQ.ID NO: 6, or (ii) encoded by a nucleic acid sequence capable of hybridizing under at least moderately stringent conditions to any nucleic acid sequence encoding any PsiK set forth herein, but for the use of synonymous codons.

The term “PsiM” as used herein, refers to any and all enzymes comprising a sequence of amino acid residues which is (i) substantially identical to the amino acid sequences constituting any PsiM polypeptide set forth herein, including, for example, SEQ.ID NO: 8, and SEQ.ID NO: 59 or (ii) encoded by a nucleic acid sequence capable of hybridizing under at least moderately stringent conditions to any nucleic acid sequence encoding any PsiM set forth herein, but for the use of synonymous codons.

The term “Psi-ncAAAD” as used herein, refers to any and all enzymes comprising a sequence of amino acid residues which is (i) substantially identical to the amino acid sequences constituting any Psi-ncAAAD polypeptide set forth herein, including, for example, SEQ.ID NO: 10, or (ii) encoded by a nucleic acid sequence capable of hybridizing under at least moderately stringent conditions to any nucleic acid sequence encoding any Psi-ncAAAD set forth herein, but for the use of synonymous codons.

The term “TrpB” as used herein, refers to any and all enzymes comprising a sequence of amino acid residues which is (i) substantially identical to the amino acid sequences constituting any TrpB polypeptide set forth herein, including, for example, SEQ.ID NO: 12, SEQ.ID NO: 65, SEQ.ID NO: 67, SEQ.ID NO: 69, SEQ.ID NO: 71, and SEQ.ID NO: 73, or (ii) encoded by a nucleic acid sequence capable of hybridizing under at least moderately stringent conditions to any nucleic acid sequence encoding any TrpB set forth herein, but for the use of synonymous codons.

The term “halogenase”, as used herein, refers to any and all enzymes comprising a sequence of amino acid residues which is (i) substantially identical to the amino acid sequences constituting any halogenase polypeptide set forth herein, including, for example, SEQ.ID NO: 14, or (ii) encoded by a nucleic acid sequence capable of hybridizing under at least moderately stringent conditions to any nucleic acid sequence encoding any halogenase set forth herein, but for the use of synonymous codons.

The term “acetyl transferase”, as used herein, refers to any and all enzymes comprising a sequence of amino acid residues which is (i) substantially identical to the amino acid sequences constituting any acetyl transferase polypeptide set forth herein, including, for example, SEQ.ID NO: 63, or (ii) encoded by a nucleic acid sequence capable of hybridizing under at least moderately stringent conditions to any nucleic acid sequence encoding any acetyl transferase set forth herein, but for the use of synonymous codons.

The terms “nucleic acid sequence encoding PsiD”, and “nucleic acid sequence encoding a PsiD polypeptide”, as may be used interchangeably herein, refer to any and all nucleic acid sequences encoding a PsiD polypeptide, including, for example, SEQ.ID NO: 1, SEQ.ID NO: 36, SEQ.ID NO: 38, SEQ.ID NO: 40, SEQ.ID NO: 42, SEQ.ID NO: 44, SEQ.ID NO: 46, SEQ.ID NO: 48, SEQ.ID NO: 50, SEQ.ID NO: 52, SEQ.ID NO: 54, and SEQ.ID NO: 56. Nucleic acid sequences encoding a PsiD polypeptide further include any and all nucleic acid sequences which (i) encode polypeptides that are substantially identical to the PsiD polypeptide sequences set forth herein; or (ii) hybridize to any PsiD nucleic acid sequences set forth herein under at least moderately stringent hybridization conditions or which would hybridize thereto under at least moderately stringent conditions but for the use of synonymous codons.

The terms “nucleic acid sequence encoding PsiH”, and “nucleic acid sequence encoding a PsiH polypeptide”, as may be used interchangeably herein, refer to any and all nucleic acid sequences encoding a PsiH polypeptide, including, for example, SEQ.ID NO: 3. Nucleic acid sequences encoding a PsiH polypeptide further include any and all nucleic acid sequences which (i) encode polypeptides that are substantially identical to the PsiH polypeptide sequences set forth herein; or (ii) hybridize to any PsiH nucleic acid sequences set forth herein under at least moderately stringent hybridization conditions or which would hybridize thereto under at least moderately stringent conditions but for the use of synonymous codons.

The terms “nucleic acid sequence encoding PsiK”, and “nucleic acid sequence encoding a PsiK polypeptide”, as may be used interchangeably herein, refer to any and all nucleic acid sequences encoding a PsiK polypeptide, including, for example, SEQ.ID NO: 5. Nucleic acid sequences encoding a PsiK polypeptide further include any and all nucleic acid sequences which (i) encode polypeptides that are substantially identical to the PsiK polypeptide sequences set forth herein; or (ii) hybridize to any PsiK nucleic acid sequences set forth herein under at least moderately stringent hybridization conditions or which would hybridize thereto under at least moderately stringent conditions but for the use of synonymous codons.

The terms “nucleic acid sequence encoding PsiM”, and “nucleic acid sequence encoding a PsiM polypeptide”, as may be used interchangeably herein, refer to any and all nucleic acid sequences encoding a PsiM polypeptide, including, for example, SEQ.ID NO: 7, SEQ.ID NO: 58 and SEQ.ID NO: 61. Nucleic acid sequences encoding a PsiM polypeptide further include any and all nucleic acid sequences which (i) encode polypeptides that are substantially identical to the PsiM polypeptide sequences set forth herein; or (ii) hybridize to any PsiM nucleic acid sequences set forth herein under at least moderately stringent hybridization conditions or which would hybridize thereto under at least moderately stringent conditions but for the use of synonymous codons.

The terms “nucleic acid sequence encoding Psi-ncAAAD”, and “nucleic acid sequence encoding a Psi-ncAAAD polypeptide”, as may be used interchangeably herein, refer to any and all nucleic acid sequences encoding a PsiD polypeptide, including, for example, SEQ.ID NO: 9. Nucleic acid sequences encoding a Psi-ncAAAD polypeptide further include any and all nucleic acid sequences which (i) encode polypeptides that are substantially identical to the Psi-ncAAAD polypeptide sequences set forth herein; or (ii) hybridize to any Psi-ncAAAD nucleic acid sequences set forth herein under at least moderately stringent hybridization conditions or which would hybridize thereto under at least moderately stringent conditions but for the use of synonymous codons.

The terms “nucleic acid sequence encoding TrpB”, and “nucleic acid sequence encoding a TrpB polypeptide”, as may be used interchangeably herein, refer to any and all nucleic acid sequences encoding a TrpB polypeptide, including, for example, SEQ.ID NO: 11, SEQ.ID NO: 64, SEQ.ID NO: 66, SEQ.ID NO: 68, SEQ.ID NO: 70, and SEQ.ID NO: 72. Nucleic acid sequences encoding a TrpB polypeptide further include any and all nucleic acid sequences which (i) encode polypeptides that are substantially identical to the TrpB polypeptide sequences set forth herein; or (ii) hybridize to any TrpB nucleic acid sequences set forth herein under at least moderately stringent hybridization conditions or which would hybridize thereto under at least moderately stringent conditions but for the use of synonymous codons.

The terms “nucleic acid sequence encoding a halogenase”, and “nucleic acid sequence encoding a halogenase polypeptide”, as may be used interchangeably herein, refer to any and all nucleic acid sequences encoding a halogenase polypeptide, including, for example, SEQ.ID NO: 13. Nucleic acid sequences encoding a halogenase polypeptide further include any and all nucleic acid sequences which (i) encode polypeptides that are substantially identical to the halogenase polypeptide sequences set forth herein; or (ii) hybridize to any halogenase nucleic acid sequences set forth herein under at least moderately stringent hybridization

conditions or which would hybridize thereto under at least moderately stringent conditions but for the use of synonymous codons.

The terms "nucleic acid sequence encoding an acetyl transferase", and "nucleic acid sequence encoding an acetyl transferase polypeptide", as may be used interchangeably herein, refer to any and all nucleic acid sequences encoding an acetyl transferase polypeptide, including, for example, SEQ.ID NO: 62. Nucleic acid sequences encoding an acetyl transferase polypeptide further include any and all nucleic acid sequences which (i) encode polypeptides that are substantially identical to the acetyl transferase polypeptide sequences set forth herein; or (ii) hybridize to any acetyl transferase nucleic acid sequences set forth herein under at least moderately stringent hybridization conditions or which would hybridize thereto under at least moderately stringent conditions but for the use of synonymous codons.

The terms "nucleic acid", or "nucleic acid sequence", as used herein, refer to a sequence of nucleoside or nucleotide monomers, consisting of naturally occurring bases, sugars and intersugar (backbone) linkages. The term also includes modified or substituted sequences comprising non-naturally occurring monomers or portions thereof. The nucleic acids of the present disclosure may be deoxyribonucleic acids (DNA) or ribonucleic acids (RNA) and may include naturally occurring bases including adenine, guanine, cytosine, thymidine and uracil. The nucleic acids may also contain modified bases. Examples of such modified bases include aza and deaza adenine, guanine, cytosine, thymidine and uracil, and xanthine and hypoxanthine. A sequence of nucleotide or nucleoside monomers may be referred to as a polynucleotide sequence, nucleic acid sequence, a nucleotide sequence or a nucleoside sequence.

The term "polypeptide", as used herein in conjunction with a reference SEQ.ID NO, refers to any and all polypeptides comprising a sequence of amino acid residues which is (i) substantially identical to the amino acid sequence constituting the polypeptide having such reference SEQ.ID NO, or (ii) encoded by a nucleic acid sequence capable of hybridizing under at least moderately stringent conditions to any nucleic acid sequence encoding the polypeptide having such reference SEQ.ID NO, but for the use of synonymous codons. A sequence of amino acid residues may be referred to as an amino acid sequence, or polypeptide sequence.

The term "nucleic acid sequence encoding a polypeptide", as used herein in conjunction with a reference SEQ.ID NO, refers to any and all nucleic acid sequences encoding a polypeptide having such reference SEQ.ID NO. Nucleic acid sequences encoding a polypeptide, in conjunction with a reference SEQ.ID NO, further include any and all nucleic acid sequences which (i) encode polypeptides that are substantially identical to the polypeptide having such reference SEQ.ID NO; or (ii) hybridize to any nucleic acid sequences encoding polypeptides having such reference SEQ.ID NO under at least moderately stringent hybridization conditions or which would hybridize thereto under at least moderately stringent conditions but for the use of synonymous codons.

By the term "substantially identical" it is meant that two amino acid sequences preferably are at least 70% identical, and more preferably are at least 85% identical and most preferably at least 95% identical, for example 96%, 97%, 98% or 99% identical. In order to determine the percentage of identity between two amino acid sequences the amino acid sequences of such two sequences are aligned, using for example the alignment method of Needleman and Wunsch (J. Mol. Biol., 1970, 48: 443), as revised by Smith and Waterman (Adv. Appl. Math., 1981, 2: 482) so that the

highest order match is obtained between the two sequences and the number of identical amino acids is determined between the two sequences. Methods to calculate the percentage identity between two amino acid sequences are generally art recognized and include, for example, those described by Carillo and Lipton (SIAM J. Applied Math., 1988, 48:1073) and those described in Computational Molecular Biology, Lesk, e.d. Oxford University Press, New York, 1988, Biocomputing: Informatics and Genomics Projects. Generally, computer programs will be employed for such calculations. Computer programs that may be used in this regard include, but are not limited to, GCG (Devereux et al., Nucleic Acids Res., 1984, 12: 387) BLASTP, BLASTN and FASTA (Altschul et al., J. Mol. Biol., 1990: 215:403). A particularly preferred method for determining the percentage identity between two polypeptides involves the Clustal W algorithm (Thompson, J D, Higgins, D G and Gibson T J, 1994, Nucleic Acid Res 22(22): 4673-4680 together with the BLOSUM 62 scoring matrix (Henikoff S & Henikoff, J G, 1992, Proc. Natl. Acad. Sci. USA 89: 10915-10919 using a gap opening penalty of 10 and a gap extension penalty of 0.1, so that the highest order match obtained between two sequences wherein at least 50% of the total length of one of the two sequences is involved in the alignment.

By "at least moderately stringent hybridization conditions" it is meant that conditions are selected which promote selective hybridization between two complementary nucleic acid molecules in solution. Hybridization may occur to all or a portion of a nucleic acid sequence molecule. The hybridizing portion is typically at least 15 (e.g. 20, 25, 30, 40 or 50) nucleotides in length. Those skilled in the art will recognize that the stability of a nucleic acid duplex, or hybrids, is determined by the T_m , which in sodium containing buffers is a function of the sodium ion concentration and temperature ($T_m = 81.5^\circ \text{C} - 16.6 (\log_{10} [\text{Na}^+]) + 0.41(\% (\text{G} + \text{C}) - 600/\text{I})$, or similar equation). Accordingly, the parameters in the wash conditions that determine hybrid stability are sodium ion concentration and temperature. In order to identify molecules that are similar, but not identical, to a known nucleic acid molecule a 1% mismatch may be assumed to result in about a 1°C . decrease in T_m , for example if nucleic acid molecules are sought that have a >95% identity, the final wash temperature will be reduced by about 5°C . Based on these considerations those skilled in the art will be able to readily select appropriate hybridization conditions. In preferred embodiments, stringent hybridization conditions are selected. By way of example the following conditions may be employed to achieve stringent hybridization: hybridization at $5\times$ sodium chloride/sodium citrate (SSC)/ $5\times$ Denhardt's solution/1.0% SDS at T_m (based on the above equation) -5°C ., followed by a wash of $0.2\times$ SSC/0.1% SDS at 60°C . Moderately stringent hybridization conditions include a washing step in $3\times$ SSC at 42°C . It is understood however that equivalent stringencies may be achieved using alternative buffers, salts and temperatures. Additional guidance regarding hybridization conditions may be found in: Current Protocols in Molecular Biology, John Wiley & Sons, N.Y., 1989, 6.3.1.-6.3.6 and in: Sambrook et al., Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989, Vol. 3.

The term "functional variant", as used herein in reference to polynucleotides or polypeptides, refers to polynucleotides or polypeptides capable of performing the same function as a noted reference polynucleotide or polypeptide. Thus, for example, a functional variant of the polypeptide set forth in SEQ.ID NO: 2, refers to a polypeptide capable of perform-

31

ing the same function as the polypeptide set forth in SEQ.ID NO: 2. Functional variants include modified a polypeptide wherein, relative to a noted reference polypeptide, the modification includes a substitution, deletion or addition of one or more amino acids. In some embodiments, substitutions are those that result in a replacement of one amino acid with an amino acid having similar characteristics. Such substitutions include, without limitation (i) glutamic acid and aspartic acid; (i) alanine, serine, and threonine; (iii) isoleucine, leucine and valine, (iv) asparagine and glutamine, and (v) tryptophan, tyrosine and phenylalanine. Functional variants further include polypeptides having retained or exhibiting an enhanced psilocybin biosynthetic bioactivity.

The term “chimeric”, as used herein in the context of nucleic acids, refers to at least two linked nucleic acids which are not naturally linked. Chimeric nucleic acids include linked nucleic acids of different natural origins. For example, a nucleic acid constituting a microbial promoter linked to a nucleic acid encoding a plant polypeptide is considered chimeric. Chimeric nucleic acids also may comprise nucleic acids of the same natural origin, provided they are not naturally linked. For example a nucleic acid constituting a promoter obtained from a particular cell-type may be linked to a nucleic acid encoding a polypeptide obtained from that same cell-type, but not normally linked to the nucleic acid constituting the promoter. Chimeric nucleic acids also include nucleic acids comprising any naturally occurring nucleic acids linked to any non-naturally occurring nucleic acids.

The terms “substantially pure” and “isolated”, as used herein, as may be used interchangeably herein describe a compound, e.g., a secondary metabolite, psilocybin or a psilocybin derivative, polynucleotide or a polypeptide, which has been separated from components that naturally accompany it. Typically, a compound is substantially pure when at least 60%, more preferably at least 75%, more preferably at least 90%, 95%, 96%, 97%, or 98%, and most preferably at least 99% of the total material (by volume, by wet or dry weight, or by mole percent or mole fraction) in a sample is the compound of interest. Purity can be measured by any appropriate method, e.g., in the case of polypeptides, by chromatography, gel electrophoresis or HPLC analysis.

The term “recovered” as used herein in association with an enzyme, protein, a secondary metabolite or a chemical compound, refers to a more or less pure form of the enzyme, protein, secondary metabolite, or chemical compound.

General Implementation

As hereinbefore mentioned, the present disclosure relates to psilocybin derivatives. In particular, the present disclosure provides novel halogenated psilocybin derivatives. In general, the herein provided compositions exhibit functional properties which deviate from the functional properties of psilocybin. Thus, for example, the psilocybin derivatives, can exhibit pharmacological properties which deviate from psilocybin. The halogenated psilocybin derivatives in this respect are useful in the formulation of pharmaceutical and recreation drug formulations. Furthermore, the halogenated psilocybin derivatives of the present disclosure may be used as a feedstock material for deriving further psilocybin derivatives. In one embodiment, the halogenated psilocybin derivatives of the present disclosure can conveniently be biosynthetically produced. The practice of this method avoids the extraction of psilocybin from mushrooms and the performance of subsequent chemical reactions to achieve halogenation. Furthermore, the growth of mushrooms can be

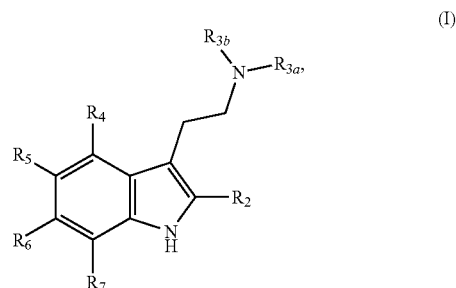
32

avoided thus limiting the dependence on climate and weather, and potential legal and social challenges associated with the cultivation of mushrooms containing psychoactive compounds. The method can efficiently yield substantial quantities of halogenated psilocybin derivatives.

In what follows selected embodiments are described with reference to the drawings.

Initially various halogenated psilocybin derivatives will be described. Thereafter methods example methods of using and making the halogenated psilocybin derivatives will be described

In at least one aspect, and in at least one example embodiment, the present disclosure provides a chemical compound or salts thereof having the formula (I):



wherein, at least one of R₂, R₄, R₅, R₆ or R₇ is a halogen atom, wherein each non-halogenated R₂, R₅, R₆, or R₇ is a hydrogen atom, wherein R₄ when it is not halogenated is a phosphate group or a hydrogen atom, and wherein R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group or an acyl group.

It is noted that in reference to the indole prototype structure shown in FIG. 2 carbon atoms C₂, C₄, C₅, C₆ and C₇ are bonded to R₂, R₄, R₅, R₆ and R₇, respectively. Furthermore, R_{3a} and R_{3b} reference chemical groups extending from the ethyl-amino group extending in turn from carbon atom C₃ of the prototype indole structure.

Referring to the chemical compound having formula (I), In one embodiment, one of R₂, R₄, R₅, R₆ and R₇ can be a halogen atom. Thus, in one embodiment, R₂ can be a bromine, chlorine, fluorine or iodine atom, each of R₅, R₆ and R₇ can be a hydrogen atom or an alkyl group or O-alkyl group, and R₄, when it is not a halogen atom, can be a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivatives shown in FIG. 4A (R₂ is chlorine; R₄ is a phosphate group; R₅, R₆ and R₇ are each a hydrogen atom; R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group); FIG. 4B (R₂ is bromine; R₄, R₅, R₆ and R₇ are each a hydrogen atom; R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group); FIG. 4K (R₂ is chlorine; R₄ is a methyl group; R₅, R₆ and R₇ are each a hydrogen atom; R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group); and FIG. 4O (R₂ is chlorine; R₄ is an O-methyl group; R₅, R₆ and R₇ are each a hydrogen atom; R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

Continuing to refer to the chemical compound having formula (I), in one embodiment, R₄ can be a bromine, chlorine, fluorine or iodine atom, and each of R₂, R₅, R₆ and R₇ can be a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivatives

shown in FIG. 4C (R_4 is iodine; R_2 , R_5 , R_6 and R_7 are each a hydrogen atom; R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group); FIG. 4D (R_4 is fluorine; R_2 , R_5 , R_6 and R_7 are each a hydrogen atom; R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group))

Continuing to refer to the chemical compound having formula (I), in one embodiment, R_5 can be a bromine, chlorine, fluorine or iodine atom, and each of R_2 , R_6 and R_7 can be a hydrogen atom or an alkyl group or O-alkyl group, and R_4 can be a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivatives shown in FIG. 4E (R_5 is fluorine; R_4 is a phosphate group; R_2 , R_6 and R_7 are each a hydrogen atom; R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group); FIG. 4F (R_5 is chlorine; R_2 , R_4 , R_6 and R_7 are each a hydrogen atom; R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group); FIG. 4L (R_5 is bromine; R_4 is an ethyl group; R_2 , R_6 and R_7 are each a hydrogen atom; R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group); and FIG. 4P (R_5 is bromine; R_2 , R_6 and R_7 are each a hydrogen atom; R_4 is an O-ethyl group; R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

Continuing to refer to the chemical compound having formula (I), in one embodiment, R_6 can be a bromine, chlorine, fluorine or iodine atom, and each of R_2 , R_5 and R_7 can be a hydrogen atom or an alkyl group or O-alkyl group, and R_4 can be a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivatives shown in FIG. 4G (R_6 is bromine; R_4 is a phosphate group; R_2 , R_5 and R_7 are each a hydrogen atom; R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group); FIG. 4H (R_6 is chlorine; R_2 , R_4 , R_5 and R_7 are each a hydrogen atom; R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group); FIG. 4M (R_6 is fluorine; R_4 is a methyl group; R_2 , R_5 and R_7 are each a hydrogen atom; R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group); and FIG. 4Q (R_6 is fluorine; R_2 , R_5 and R_7 are each a hydrogen atom; R_4 is an O-methyl group; R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

Continuing to refer to the chemical compound having formula (I), in one embodiment, R_7 can be a bromine, chlorine, fluorine or iodine atom, and each of R_2 , R_5 and R_6 can be a hydrogen atom or an alkyl group or O-alkyl group, and R_4 can be a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivatives shown in FIG. 4I (R_7 is fluorine; R_4 is a phosphate group; R_2 , R_5 and R_6 are each a hydrogen atom; R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group); FIG. 4J (R_7 is iodine; R_2 , R_4 , R_5 and R_6 are each a hydrogen atom; R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group); FIG. 4N (R_7 is iodine; R_4 is a propyl group; R_2 , R_5 and R_6 are each a hydrogen atom; R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group); and FIG. 4R (R_7 is iodine; R_2 , R_5 and R_6 are each a hydrogen atom; R_4 is an O-propyl group; R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

Still, continuing to refer to the chemical compound having formula (I), in one embodiment, two of R_2 , R_4 , R_5 , R_6 and R_7 can be a halogen atom, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom, and wherein each non-halogenated R_2 , R_5 ,

R_6 and R_7 is a hydrogen atom or an alkyl group or O-alkyl group, and wherein R_4 , when it is not a halogen atom, is a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group.

Thus, continuing to refer to the chemical compound having formula (I), in one embodiment, R_2 and R_4 can be halogen atoms, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom, and R_5 , R_6 and R_7 can be a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivative shown in FIG. 5A (R_2 is a chlorine atom; R_4 is a chlorine atom; R_5 , R_6 and R_7 are each a hydrogen atom; and R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

Continuing to refer to the chemical compound having formula (I), in one embodiment, R_2 and R_5 can be halogen atoms, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom, and R_6 and R_7 can be a hydrogen atom or an alkyl group or O-alkyl group and R_4 can be a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivative shown in FIG. 5B (R_2 is a bromine atom; R_5 is a fluorine atom; R_4 , and R_6 are each a hydrogen atom and R_7 is a methyl group; and R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

Continuing to refer to the chemical compound having formula (I), in one embodiment, R_2 and R_6 can be halogen atoms, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom, and R_5 and R_7 can be a hydrogen atom or an alkyl group or O-alkyl group, and R_4 can be a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivative shown in FIG. 5C (R_2 is a chlorine atom; R_6 is a bromine atom; R_5 and R_7 are each a hydrogen atom; R_4 is an O-methyl group; and R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

Continuing to refer to the chemical compound having formula (I), in one embodiment, R_2 and R_7 can be halogen atoms, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom, and R_5 and R_6 can be a hydrogen atom or an alkyl group or O-alkyl group and R_4 can be a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivative shown in FIG. 5D (R_2 is a fluorine atom; R_7 is an iodine atom; R_4 is a phosphate, R_5 and R_6 are each a hydrogen atom; and R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

Continuing to refer to the chemical compound having formula (I), in one embodiment, R_4 and R_5 can be halogen atoms, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom, and R_2 , R_6 and R_7 can be a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivative shown in FIG. 5E (R_4 and R_5 are each bromine atoms; R_2 , R_6 and R_7 are each a hydrogen atom; and R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

Continuing to refer to the chemical compound having formula (I), in one embodiment, R_4 and R_6 can be halogen atoms, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom, and R_2 , R_5 and R_7 can be a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivative shown in FIG. 5F (R_4 and R_6 are each chlorine

35

atoms; R₂, R₅ and R₇ are each a hydrogen atom; and R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

Continuing to refer to the chemical compound having formula (I), in one embodiment, R₄ and R₇ can be halogen atoms, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom, and R₂, R₅ and R₆ can be a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivative shown in FIG. 5G (R₄ is a chlorine atom; R₇ is a fluorine atom; R₂, R₅ and R₆ are each a hydrogen atom; and R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

Continuing to refer to the chemical compound having formula (I), in one embodiment, R₅ and R₆ can be halogen atoms, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom, and R₂ and R₇ can be a hydrogen atom or an alkyl group or O-alkyl group and R₄ can be a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivative shown in FIG. 5H (R₅ is a fluorine atom; R₆ is a bromine atom; R₄ is a phosphate group, R₂ and R₇ are each a hydrogen atom; and R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

Continuing to refer to the chemical compound having formula (I), in one embodiment, R₅ and R₇ can be halogen atoms, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom, and R₂ and R₆ can be a hydrogen atom or an alkyl group or O-alkyl group and R₄ can be a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivative shown in FIG. 5I (R₅ and R₇ are each fluorine atoms; R₄ is a phosphate group, R₂ and R₆ are each a hydrogen atom; and R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

Continuing to refer to the chemical compound having formula (I), in one embodiment, R₆ and R₇ can be halogen atoms, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom, and R₂ and R₅ can be a hydrogen atom or an alkyl group or O-alkyl group and R₄ can be a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivative shown in FIG. 5J (R₆ and R₇ are each iodine atoms; R₂, R₄ and R₅ are each a hydrogen atom; and R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

In one embodiment, two of R₂, R₄, R₅, R₆ and R₇ can be a halogen atom, and the halogen atoms are identical, the remaining of R₂, R₅, R₆, and R₇ being a hydrogen atom or an alkyl group or O-alkyl group, and R₄ when it is not a halogen atom, being a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group (see: e.g. the halogenated psilocybin derivatives shown in FIGS. 5A, 5E, 5F, 5I and 5J).

In one embodiment, two of R₂, R₄, R₅, R₆ and R₇ can be a halogen atom, and the halogen atoms are non-identical, the remaining of R₂, R₅, R₆, and R₇ being a hydrogen atom or an alkyl group or O-alkyl group, and R₄ when it is not a halogen atom, being a phosphate, a hydrogen atom or an alkyl group or O-alkyl group (see: e.g. the halogenated psilocybin derivatives shown in FIGS. 5B, 5C, 5D, 5G and 5H).

Still continuing to refer to the chemical compound having formula (I), In one further embodiment, three of R₂, R₄, R₅, R₆ and R₇ can be a halogen atom, wherein the halogen atoms

36

are independently selected from a bromine, chlorine fluorine or iodine atom, and wherein the non-halogenated R₂, R₅, R₆, or R₇ is a hydrogen atom or an alkyl group or O-alkyl group, and wherein R₄, when it is not a halogen atom, is a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group.

Thus, continuing to refer to the chemical compound having formula (I), in one embodiment R₂, R₄, and R₅ can be halogen atom, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom, and R₆ and R₇ can be a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivative shown in FIG. 6A (R₂, R₄ and R₅ are each chlorine atoms; R₆ and R₇ are each a hydrogen atom; and R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

Continuing to refer to the chemical compound having formula (I), in one embodiment R₂, R₅, and R₆ can be halogen atom, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom and R₇ can be a hydrogen atom or an alkyl group or O-alkyl group, and R₄ can be a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivative shown in FIG. 6B (R₂ is a bromine atom, R₅ and R₆ are each fluorine atoms; R₄ and R₇ are each a hydrogen atom; and R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

Continuing to refer to the chemical compound having formula (I), in one embodiment R₂, R₅, and R₇ can be halogen atom, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom and R₆ can be a hydrogen atom or an alkyl group or O-alkyl group, and R₄ can be a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivative shown in FIG. 6C (R₂ is a chlorine atom, R₅ is a bromine atom, and R₇ is a fluorine atoms; R₄ and R₆ are each a hydrogen atom; and R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

Continuing to refer to the chemical compound having formula (I), in one embodiment R₄, R₅, and R₆ can be halogen atom, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom, and R₂ and R₇ can be a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivative shown in FIG. 6D (R₄ and R₅ are each a bromine atom, R₆ is a chlorine atom; R₂ and R₇ are each a hydrogen atom; and R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

Continuing to refer to the chemical compound having formula (I), in one embodiment R₄, R₅, and R₇ can be halogen atom, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom, and R₂ and R₆ can be a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivative shown in FIG. 6E (R₄, R₅ and R₇ are each a bromine atom; R₂ and R₆ are each a hydrogen atom; and R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

Continuing to refer to the chemical compound having formula (I), in one embodiment R₅, R₆, and R₇ can be halogen atom, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom and R₂ can be a hydrogen atom or an alkyl group or O-alkyl group, and R₄ can be a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivative shown in FIG. 6F (R₅ is a

fluorine atom, R₆ is a chlorine atom, and R₇ is an iodine atom; R₄ is a phosphate group; R₂ is a hydrogen atom; and R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

In one embodiment, three of R₂, R₄, R₅, R₆ and R₇ can be a halogen atom, wherein at least two halogen atoms are identical, the remaining R₂, R₅, R₆, and R₇ being a hydrogen atom or an alkyl group or O-alkyl group, and R₄ when it is not a halogen atom, being a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group (see: e.g. the halogenated psilocybin derivatives shown in FIGS. 6A, 6B and 6E).

In one embodiment, three of R₂, R₄, R₅, R₆ and R₇ can be a halogen atom, and at least two halogen atoms are non-identical, the remaining R₂, R₅, R₆, and R₇ being a hydrogen atom or an alkyl group or O-alkyl group, and R₄ when it is not a halogen atom, being a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group (see: e.g. the halogenated psilocybin derivatives shown in FIGS. 6B, 6C, 6D and 6F).

In one embodiment, three of R₂, R₄, R₅, R₆ and R₇ can be a halogen atom, and all three halogen atoms are identical, the remaining R₂, R₅, R₆, or R₇ being a hydrogen atom or an alkyl group or O-alkyl group, and R₄ when it is not a halogen atom, being a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group (see: e.g. the halogenated psilocybin derivatives shown in FIGS. 6A and 6E).

Still continuing to refer to the chemical compound having formula (I), in one embodiment, four of R₂, R₄, R₅, R₆ and R₇ can be a halogen atom, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom, and wherein the non-halogenated R₂, R₅, R₆, or R₇ is a hydrogen atom or an alkyl group or O-alkyl group, and wherein R₄, when it is not a halogen, is a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group.

Thus, continuing to refer to the chemical compound having formula (I), in one embodiment R₂, R₄, R₅ and R₆ can be halogen atom, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom, and R₇ can be a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivative shown in FIG. 7A (R₂, R₄, R₅ and R₆ are each chlorine atoms; R₇ is a hydrogen atom; and R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

Continuing to refer to the chemical compound having formula (I), in one embodiment R₄, R₅, R₆ and R₇ can be halogen atom, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom and R₂ can be a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivative shown in FIG. 7B (R₄ and R₇ are chlorine atoms, R₅ and R₆ are each fluorine atoms; R₂ is a hydrogen atom; and R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

Continuing to refer to the chemical compound having formula (I), in one embodiment R₂, R₅, R₆ and R₇ can be halogen atom, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom and R₄ can be a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivative shown in FIG. 7C (R₂ and R₆ are each chlorine atoms, R₅ is a fluorine atom, R₇ is an iodine atom, R₄ is a phosphate group; and R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

Continuing to refer to the chemical compound having formula (I), in one embodiment R₂, R₄, R₆ and R₇ can be halogen atom, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom and R₅ can be a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivative shown in FIG. 7D (R₂, R₄ and R₇ are each bromine atoms, R₆ is a chlorine atom, R₅ is a hydrogen atom; and R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

In one embodiment R₂, R₄, R₅ and R₇ can be halogen atom, wherein the halogen atoms are independently selected from a bromine, chlorine fluorine or iodine atom and R₆ can be a hydrogen atom or an alkyl group or O-alkyl group (see: the example halogenated psilocybin derivative shown in FIG. 7E (R₄, R₅ and R₇ are each bromine atoms, R₂ is a fluorine atom, R₆ is a hydrogen atom; and R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group)).

In one embodiment, four of R₂, R₄, R₅, R₆ and R₇ can be a halogen atom, wherein at least two halogen atoms are identical, the remaining R₂, R₅, R₆, and R₇ being a hydrogen atom or an alkyl group or O-alkyl group, and R₄ when it is not a halogen atom, being a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group (see: e.g. the halogenated psilocybin derivatives shown in FIGS. 7A, 7B, 7C, 7D and 7E).

In one embodiment, four of R₂, R₄, R₅, R₆ and R₇ can be a halogen atom, and at least two halogen atoms are non-identical, the remaining R₂, R₅, R₆, and R₇ being a hydrogen atom or an alkyl group or O-alkyl group, and R₄ when it is not a halogen atom, being a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group (see: e.g. the halogenated psilocybin derivatives shown in FIGS. 7B, 7C, 7D and 7E).

In one embodiment, four of R₂, R₄, R₅, R₆ and R₇ can be a halogen atom, and all four halogen atoms are identical, the remaining R₂, R₅, R₆, or R₇ being a hydrogen atom or an alkyl group or O-alkyl group, and R₄ when it is not a halogen atom, being a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group (see: e.g. the halogenated psilocybin derivative shown in FIG. 7A).

Still continuing to refer to the chemical compound having formula (I), in one embodiment, all five of R₂, R₄, R₅, R₆ and R₇ can be a halogen atom, the halogen selected from bromine, chlorine, fluorine or iodine.

In one embodiment, all five of R₂, R₄, R₅, R₆ and R₇ can be a halogen atom, and at least two halogen atoms are identical.

In one embodiment, all five of R₂, R₄, R₅, R₆ and R₇ can be a halogen atom, and at least three halogen atoms are identical.

In one embodiment, all five of R₂, R₄, R₅, R₆ and R₇ can be a halogen atom, and at least four halogen atoms are identical.

In one embodiment, all five of R₂, R₄, R₅, R₆ and R₇ can be a halogen atom, and all five halogen atoms are identical.

In one embodiment, all five of R₂, R₄, R₅, R₆ and R₇ can be a halogen atom, and the chemical compound comprises at least two non-identical halogen atoms.

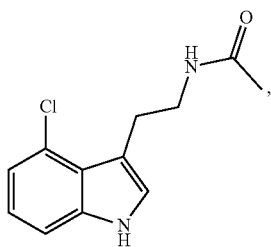
In one embodiment, all five of R₂, R₄, R₅, R₆ and R₇ can be a halogen atom, and the chemical compound comprises at least three non-identical halogen atoms.

In one embodiment, all five of R₂, R₄, R₅, R₆ and R₇ can be a halogen atom, and four halogen atoms are non-identical.

39

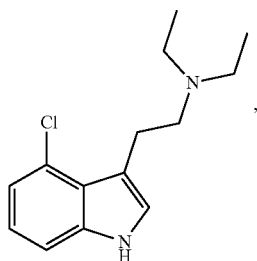
Still continuing to refer to the chemical compound having formula (I), it is noted that, in a further aspect hereof, R_{3A} and R_{3B} can be a hydrogen atom, an alkyl group, an aryl group or an acyl group. Thus, for example, R_{3A} and R_{3B} can each be a hydrogen atom, or R_{3A} and R_{3B} can each be an alkyl group, such as a methyl group, ethyl group, propyl group, or longer chain alkyl group, or R_{3A} and R_{3B} can each be an aryl group, such as a phenyl group or a naphthyl group, or R_{3A} and R_{3B} can each be an acyl group, such as an acetyl group. Furthermore, one of R_{3A} and R_{3B} can be a hydrogen atom, and one of R_{3A} and R_{3B} can be an alkyl group, and aryl group, or an acyl group. Furthermore, R_{3A} and R_{3B} can be an aryl group and an alkyl group, an aryl group and an acyl group, or an alkyl group and an acyl group.

Furthermore, in one embodiment, a psilocybin derivative according to the present disclosure can be a chemical compound having the formula (II):



or, a chemical compound having the formula (II), wherein the chlorine is substituted for a bromine, fluorine or iodine.

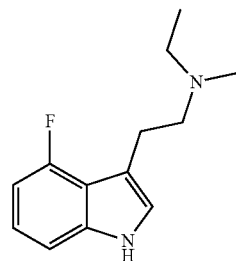
Furthermore, in one embodiment, a psilocybin derivative according to the present disclosure can be a chemical compound having the formula (III):



or, a chemical compound having the formula (III), wherein the chlorine is substituted for a bromine, fluorine or iodine.

Furthermore, in one embodiment, a psilocybin derivative according to the present disclosure can be a chemical compound having the formula (IV):

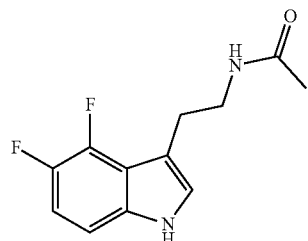
40



(IV)

or, a chemical compound having the formula (IV), wherein the fluorine is substituted for a bromine, chlorine or iodine.

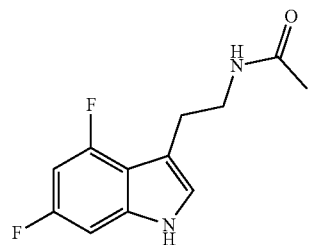
Furthermore, in one embodiment, a psilocybin derivative according to the present disclosure can be a chemical compound having the formula (V):



(V)

or, a chemical compound having the formula (V), wherein one or both of the fluorines are substituted for a bromine, chlorine or iodine.

Furthermore, in one embodiment, a psilocybin derivative according to the present disclosure can be a chemical compound having the formula (VI):

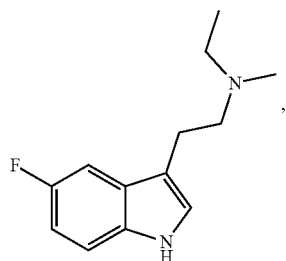


(VI)

or, a chemical compound having the formula (VI), wherein one or both of the fluorines are substituted for a bromine, chlorine or iodine.

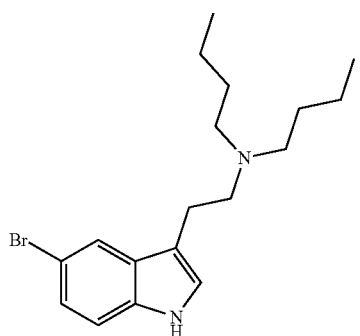
Furthermore, in one embodiment, a psilocybin derivative according to the present disclosure can be a chemical compound having the formula (VII):

41



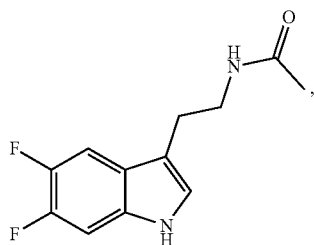
or, a chemical compound having the formula (VII), wherein the fluorine is substituted for a bromine, chlorine or iodine.

Furthermore, in one embodiment, a psilocybin derivative according to the present disclosure can be a chemical compound having the formula (VIII):



or, a chemical compound having the formula (VIII), wherein the bromine is substituted for a, chlorine, fluorine or iodine.

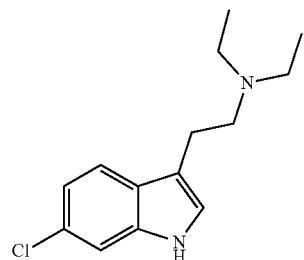
Furthermore, in one embodiment, a psilocybin derivative according to the present disclosure can be a chemical compound having the formula (IX):



or, a chemical compound having the formula (IX), wherein one or both of the fluorines are substituted for a bromine, chlorine or iodine.

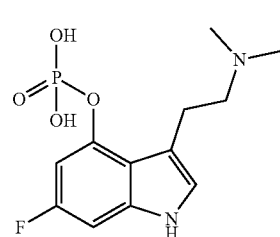
Furthermore, in one embodiment, a psilocybin derivative according to the present disclosure can be a chemical compound having the formula (X):

42



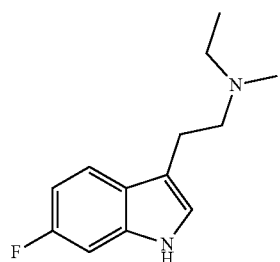
or, a chemical compound having the formula (X), wherein the chlorine is substituted for a bromine, chlorine or iodine.

Furthermore, in one embodiment, a psilocybin derivative according to the present disclosure can be a chemical compound having the formula (XI):



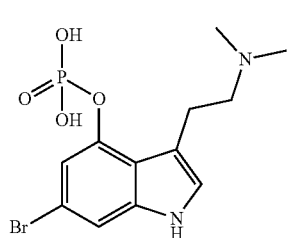
or, a chemical compound having the formula (XI), wherein the fluorine is substituted for a bromine, chlorine or iodine.

Furthermore, in one embodiment, a psilocybin derivative according to the present disclosure can be a chemical compound having the formula (XII):



or, a chemical compound having the formula (XII), wherein the fluorine is substituted for a bromine, chlorine or iodine.

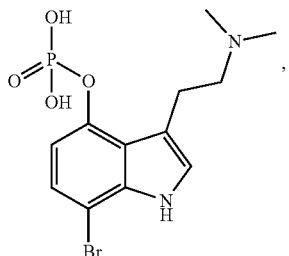
Furthermore, in one embodiment, a psilocybin derivative according to the present disclosure can be a chemical compound having the formula (XIII):



43

or, a chemical compound having the formula (XIII), wherein the bromine is substituted for a chlorine, fluorine or iodine.

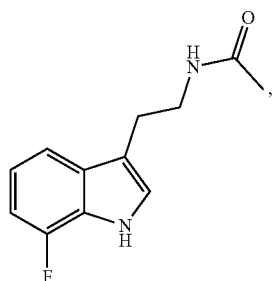
Furthermore, in one embodiment, a psilocybin derivative according to the present disclosure can be a chemical compound having the formula (XIV):



(XIV)

or, a chemical compound having the formula (XIV), wherein the bromine is substituted for a chlorine, fluorine or iodine.

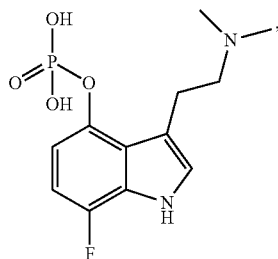
Furthermore, in one embodiment, a psilocybin derivative according to the present disclosure can be a chemical compound having the formula (XV):



(XV)

or, a chemical compound having the formula (XV), wherein the fluorine is substituted for a bromine, chlorine or iodine.

Furthermore, in one embodiment, a psilocybin derivative according to the present disclosure can be a chemical compound having the formula (XVI):

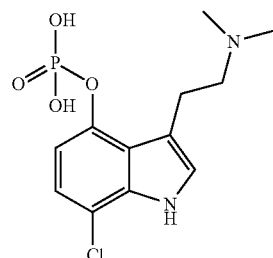


(XVI)

or, a chemical compound having the formula (XV), wherein the fluorine is substituted for a bromine, chlorine or iodine.

Furthermore, in one embodiment, a psilocybin derivative according to the present disclosure can be a chemical compound having the formula (XVII):

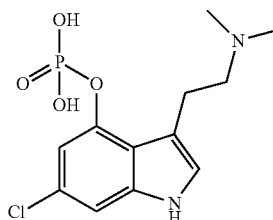
44



(XVII)

or, a chemical compound having the formula (XVII), wherein the fluorine is substituted for a bromine, chlorine or iodine.

Furthermore, in one embodiment, a psilocybin derivative according to the present disclosure can be a chemical compound having the formula (XVII):

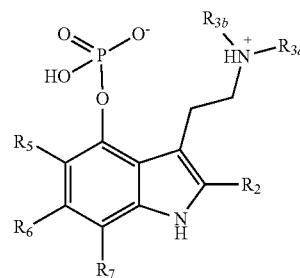


(XVII)

or, a chemical compound having the formula (XVII), wherein the fluorine is substituted for a bromine, chlorine or iodine.

Furthermore, it is noted that the halogenated psilocybin derivatives of the present disclosure include salts thereof. Thus, the nitrogen atom of the ethyl-amino group extending in turn from the C₃ atom may be protonated, and the positive charge may be balanced by, for example, chloride or sulfate ions, to thereby form a chloride salt or a sulfate salt. Furthermore, in compounds wherein R₄ is a phosphate group, the phosphate group may be de-protonated, and the negative charge may be balanced by, for example, sodium ions or potassium ions, to thereby form a sodium salt or a potassium salt.

Furthermore, it is noted that when R₄ is a phosphate group, the term halogenated psilocybin derivative also includes compounds having the formula (XVIII):



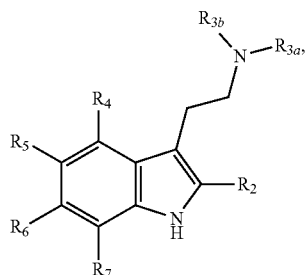
(XVIII)

wherein, at least one of R₂, R₅, R₆, or R₇ is a halogen atom or an alkyl group or O-alkyl group, and wherein any R₂, R₅, R₆, or R₇ which are not a halogen is a hydrogen atom, and

45

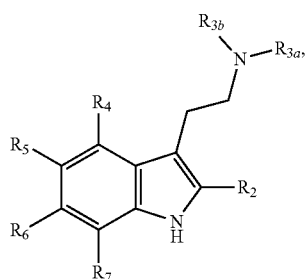
wherein R_{3A} and R_{3B} are a hydrogen atom, an alkyl group, and aryl group or an acyl group. The term further includes salts, including pharmaceutically acceptable salts, of halogenated psilocybins having the formula (XVIII), such as a sodium salt, a potassium salt etc.

Thus, to briefly recap, the present disclosure provides halogenated psilocybin derivatives. The disclosure provides, in particular, a chemical compound or salts thereof having formula (I):



wherein in an aspect, at least one of R_2 , R_4 , R_5 , R_6 , or R_7 is a halogen atom. In an aspect, in formula (I), each non-halogenated R_2 , R_5 , R_6 , or R_7 is a hydrogen atom or an alkyl group or O-alkyl group. In a further aspect, in formula (I), R_4 when it is not halogenated is a hydrogen atom, an alkyl group or O-alkyl group, a hydroxy group, or a phosphate group. Yet in a further aspect, R_{3A} and R_{3B} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group.

In one embodiment of the disclosure, a chemical compound or salts thereof having formula (I) is included:



wherein at least one of R_2 , R_4 , R_5 , R_6 , or R_7 is a halogen atom, and wherein each non-halogenated R_2 , R_5 , R_6 , or R_7 is a hydrogen atom or an alkyl group or O-alkyl group, wherein R_4 when it is not halogenated is a hydrogen atom, alkyl group or O-alkyl group, a hydroxy group, or a phosphate group, and wherein R_{3A} and R_{3B} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group.

In one embodiment, at least one of R_2 , R_4 , R_5 , R_6 , or R_7 is a halogen atom, and wherein each non-halogenated R_2 , R_5 , R_6 , or R_7 is a hydrogen atom or a (C_1-C_{20}) -alkyl group or (C_1-C_{20}) -O-alkyl group. In another embodiment, each non-halogenated R_2 , R_5 , R_6 , or R_7 is a hydrogen atom, a methyl group, ethyl group, a propyl group, an O-methyl group, an O-ethyl group, or an O-propyl group.

In another embodiment, each non-halogenated R_2 , R_5 , R_6 , or R_7 is a hydrogen atom or a (C_1-C_{10}) -alkyl group or (C_1-C_{10}) -O-alkyl group. In another embodiment, each non-halogenated R_2 , R_5 , R_6 , or R_7 is a hydrogen atom, a

46

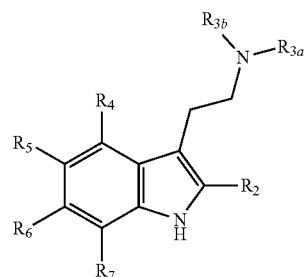
methyl group, ethyl group, a propyl group, an O-methyl group, an O-ethyl group, or an O-propyl group.

In another embodiment, each non-halogenated R_2 , R_5 , R_6 , or R_7 is a hydrogen atom or a (C_1-C_6) -alkyl group or (C_1-C_6) -O-alkyl group. In another embodiment, each non-halogenated R_2 , R_5 , R_6 , or R_7 is a hydrogen atom, a methyl group, ethyl group, a propyl group, an O-methyl group, an O-ethyl group, or an O-propyl group.

In another embodiment, when R_4 is not halogenated, R_4 is a hydrogen atom, a (C_1-C_{20}) -alkyl group or (C_1-C_{20}) -O-alkyl group, a hydroxy group, or a phosphate group. In another embodiment, when R_4 is not halogenated, R_4 is a hydrogen atom, a (C_1-C_{10}) -alkyl group or (C_1-C_{10}) -O-alkyl group, a hydroxy group, or a phosphate group. In another embodiment, when R_4 is not halogenated, R_4 is a hydrogen atom, a (C_1-C_6) -alkyl group or (C_1-C_6) -O-alkyl group, a hydroxy group, or a phosphate group. In another embodiment, when R_4 is not halogenated, R_4 is a hydrogen atom, a methyl group, an ethyl group, a propyl group, a phosphate group, an O-methyl group, an O-ethyl group, or an O-propyl group.

In another embodiment, R_{3A} and R_{3B} are a hydrogen atom, a (C_1-C_{20}) -alkyl group, a (C_6-C_{14}) -aryl group, or a $-C(=O)(C_1-C_{20})$ -alkyl group. In another embodiment, R_{3A} and R_{3B} are a hydrogen atom, a (C_1-C_{10}) -alkyl group, a (C_6-C_{10}) -aryl group, or a $-C(=O)(C_1-C_{10})$ -alkyl group. In another embodiment, R_{3A} and R_{3B} are a hydrogen atom, a (C_1-C_6) -alkyl group, a phenyl group, or a $-C(=O)(C_1-C_6)$ -alkyl group. In another embodiment, R_{3A} and R_{3B} are a hydrogen atom, a methyl group, an ethyl group, a propyl group, a phenyl group, $-C(=O)-CH_3$, $-C(=O)-CH_2CH_3$, or $-C(=O)-CH_2CH_2CH_3$.

In one embodiment of the disclosure, a chemical compound or salts thereof having formula (I) is included:



wherein

R_2 , R_5 , R_6 , and R_7 are independently or simultaneously H, an alkyl group or O-alkyl group or a halogen atom, R_{3A} and R_{3B} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group; and R_4 is hydrogen atom, alkyl group or O-alkyl group, a halogen, a hydroxy group, or a phosphate group; wherein at least one of R_2 , R_4 , R_5 , R_6 , and R_7 is a halogen.

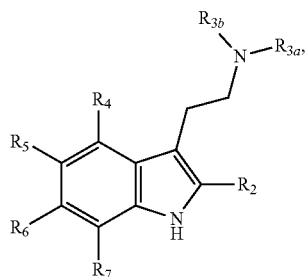
In one embodiment, R_2 , R_5 , R_6 , and R_7 are independently or simultaneously H, (C_1-C_{20}) -alkyl group or (C_1-C_{20}) -O-alkyl group or a halogen atom. In one embodiment, R_2 , R_5 , R_6 , and R_7 are independently or simultaneously H, (C_1-C_{10}) -alkyl group or (C_1-C_{10}) -O-alkyl group or a halogen atom. In one embodiment, R_2 , R_5 , R_6 , and R_7 are independently or simultaneously H, (C_1-C_6) -alkyl group or (C_1-C_6) -O-alkyl group or a halogen atom. In one embodiment, R_2 , R_5 , R_6 ,

and R_7 are independently or simultaneously H, methyl, ethyl, propyl, O-methyl, O-ethyl, O-propyl, or a halogen atom.

In one embodiment, R_4 is H, (C_1-C_{20}) -alkyl group or (C_1-C_{20}) -O-alkyl group, a halogen atom or a phosphate group. In one embodiment, R_4 is H, (C_1-C_{10}) -alkyl group or (C_1-C_{10}) -O-alkyl group, a halogen atom or a phosphate group. In one embodiment, R_4 is H, (C_1-C_6) -alkyl group or (C_1-C_6) -O-alkyl group, a halogen atom, a hydroxy group, or a phosphate group. In one embodiment, R_4 is H, methyl, ethyl, propyl, O-methyl, O-ethyl, O-propyl, a halogen atom, a hydroxy group, or a phosphate group.

In another embodiment, R_{3A} and R_{3B} are a hydrogen atom independently or simultaneously a (C_1-C_{20}) -alkyl group, a (C_6-C_{14}) -aryl group, or a $-C(=O)(C_1-C_{20})$ -alkyl group. In another embodiment, R_{3A} and R_{3B} are independently or simultaneously a hydrogen atom, a (C_1-C_{10}) -alkyl group, a (C_6-C_{10}) -aryl group, or a $-C(=O)(C_1-C_{10})$ -alkyl group or O-alkyl group. In another embodiment, R_{3A} and R_{3B} are independently or simultaneously a hydrogen atom, a (C_1-C_6) -alkyl group, a phenyl group, or a $-C(=O)(C_1-C_6)$ -alkyl group. In another embodiment, R_{3A} and R_{3B} are independently or simultaneously a hydrogen atom, a methyl group, an ethyl group, a propyl group, a butyl group, a phenyl group, $-C(=O)-CH_3$, $-C(=O)-CH_2CH_3$, or $-C(=O)-CH_2CH_2CH_3$.

The halogenated psilocybin derivatives of the present disclosure may be used to prepare a pharmaceutical or recreational drug formulation. Thus in one embodiment, the present disclosure further provides in another aspect, pharmaceutical and recreational drug formulations comprising halogenated psilocybin derivatives. Accordingly, in one aspect, the present disclosure provides in a further embodiment a pharmaceutical or recreational drug formulation comprising a chemical compound or salts thereof having the formula (I):



wherein, at least one of R_2 , R_4 , R_5 , R_6 or R_7 is a halogen atom, wherein each non-halogenated R_2 , R_5 , R_6 , or R_7 is a hydrogen atom or an alkyl group or O-alkyl group, wherein R_4 when it is not halogenated is a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group, and wherein R_{3A} and R_{3B} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group, together with a diluent, carrier or excipient. The pharmaceutical or recreational drug formulations may be prepared as liquids, tablets, capsules, microcapsules, nanocapsules, trans-dermal patches, gels, foams, oils, aerosols, nanoparticulates, powders, creams, emulsions, micellar systems, films, sprays, ovules, infusions, teas, decoctions, suppositories, etc. and include a pharmaceutically acceptable salt or solvate of the halogenated psilocybin compound together with an excipient. The term "excipient" as used herein means any ingredient other

than the chemical compound of the disclosure. As will readily be appreciated by those of skill in art, the selection of excipient may depend on factors such as the particular mode of administration, the effect of the excipient on solubility of the chemical compounds of the present disclosure and methods for their preparation will be readily apparent to those skilled in the art. Such compositions and methods for their preparation may be found, for example, in "Remington's Pharmaceutical Sciences", 22nd Edition (Pharmaceutical Press and Philadelphia College of Pharmacy at the University of the Sciences, 2012).

The dose when using the compounds of the present disclosure can vary within wide limits, and as is customary and is known to those of skill in the art, the dose can be tailored to the individual conditions in each individual case. The dose depends, for example, on the nature and severity of the illness to be treated, on the condition of the patient, on the compound employed or on whether an acute or chronic disease state is treated or prophylaxis is conducted, on the mode of delivery of the compound, or on whether further active compounds are administered in addition to the compounds of the present disclosure. Representative doses of the present disclosure include, but are not limited to, about 0.001 mg to about 5000 mg, about 0.001 mg to about 2500 mg, about 0.001 mg to about 1000 mg, about 0.001 mg to about 500 mg, about 0.001 mg to about 250 mg, about 0.001 mg to about 100 mg, about 0.001 mg to about 50 mg, and about 0.001 mg to about 25 mg. Representative doses of the present disclosure include, but are not limited to, about 0.0001 to about 1,000 mg, about 10 to about 160 mg, about 10 mg, about 20 mg, about 40 mg, about 80 mg, or about 160 mg. Multiple doses may be administered during the day, especially when relatively large amounts are deemed to be needed, for example 2, 3 or 4, doses. Depending on the subject and as deemed appropriate from the patient's physician or care giver it may be necessary to deviate upward or downward from the doses described herein.

The pharmaceutical and drug formulations comprising the halogenated psilocybin derivatives of the present disclosure may be administered orally. Oral administration may involve swallowing, so that the compound enters the gastrointestinal tract, or buccal or sublingual administration may be employed by which the compound enters the blood stream directly from the mouth. Formulations suitable for oral administration include both solid and liquid formulations.

Solid formulations include tablets, capsules (containing particulates, liquids, microcapsules, or powders), lozenges (including liquid-filled lozenges), chews, multi- and nano-particulates, gels, solid solutions, liposomal preparations, microencapsulated preparations, creams, films, ovules, suppositories and sprays.

Liquid formulations include suspensions, solutions, syrups and elixirs. Such formulations may be employed as fillers in soft or hard capsules and typically comprise a carrier, for example, water, ethanol, polyethylene glycol, propylene glycol, methylcellulose, or a suitable oil, and one or more emulsifying agents and/or suspending agents. Liquid formulations may also be prepared by the reconstitution of a solid, for example, from a sachet.

Binders are generally used to impart cohesive qualities to a tablet formulation. Suitable binders include microcrystalline cellulose, gelatin, sugars, polyethylene glycol, natural and synthetic gums, polyvinylpyrrolidone, pregelatinised starch, hydroxypropyl cellulose and hydroxypropyl methylcellulose.

Tablets may also contain diluents, such as lactose (monohydrate, spray-dried monohydrate, anhydrous and the like), mannitol, xylitol, dextrose, sucrose, sorbitol, microcrystalline cellulose, starch and dibasic calcium phosphate dihydrate.

Tablets may also optionally comprise surface active agents, such as sodium lauryl sulfate and polysorbate 80. When present, surface active agents may comprise from 0.2% (w/w) to 5% (w/w) of the tablet.

Tablets may further contain lubricants such as magnesium stearate, calcium stearate, zinc stearate, sodium stearyl fumarate, and mixtures of magnesium stearate with sodium lauryl sulphate. Lubricants generally comprise from 0.25% (w/w) to 10% (w/w), from 0.5% (w/w) to 3% (w/w) of the tablet.

In addition to the halogenated psilocybin derivative, tablets may contain a disintegrant. Examples of disintegrants include sodium starch glycolate, sodium carboxymethyl cellulose, calcium carboxymethyl cellulose, croscarmellose sodium, crospovidone, polyvinylpyrrolidone, methyl cellulose, microcrystalline cellulose, lower alkyl-substituted hydroxypropyl cellulose, starch, pregelatinised starch and sodium alginate. Generally, the disintegrant will comprise from 1% (w/w) to 25% (w/w) or from 5% (w/w) to 20% (w/w) of the dosage form.

Other possible auxiliary ingredients include anti-oxidants, colourants, flavouring agents, preservatives and taste-masking agents.

For tablet dosage forms, depending on the desired effective amount of the chemical compound, the chemical compound of the present disclosure may make up from 1% (w/w) to 80% (w/w) of the dosage form, more typically from 5% (w/w) to 60% (w/w) of the dosage form.

Exemplary tablets contain up to about 80% (w/w) of the chemical compound, from about 10% (w/w) to about 90% (w/w) binder, from about 0% (w/w) to about 85% (w/w) diluent, from about 2% (w/w) to about 10% (w/w) disintegrant, and from about 0.25% (w/w) to about 10% (w/w) lubricant.

The formulation of tablets is discussed in "Pharmaceutical Dosage Forms: Tablets", Vol. 1-Vol. 3, by CRC Press (2008).

The pharmaceutical and recreational drug formulations comprising the halogenated psilocybin derivatives of the present disclosure may also be administered directly into the blood stream, into muscle, or into an internal organ. Thus, the pharmaceutical and recreational drug formulations can be administered parenterally (for example, by subcutaneous, intravenous, intraarterial, intrathecal, intraventricular, intracranial, intramuscular, or intraperitoneal injection). Parenteral formulations are typically aqueous solutions which may contain excipients such as salts, carbohydrates and buffering agents (in one embodiment, to a pH of from 3 to 9), but, for some applications, they may be more suitably formulated as a sterile non-aqueous solution or as a dried form to be used in conjunction with a suitable vehicle such as sterile water.

Formulations comprising the halogenated psilocybin derivatives of the present disclosure for parenteral administration may be formulated to be immediate and/or modified release. Modified release formulations include delayed-, sustained-, pulsed-, controlled-, targeted and programmed release. Thus the chemical compounds of the disclosure may be formulated as a solid, semi-solid, or thixotropic liquid for administration as an implanted depot providing modified release of the active compound. Examples of such formu-

lations include drug-coated stents and poly(dl-lactic-coglycolic)acid (PGLA) microspheres.

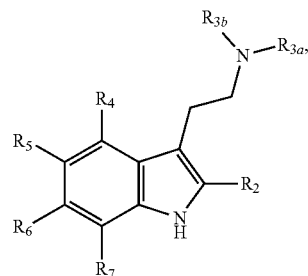
The pharmaceutical or recreational drug formulations of the present disclosure also may be administered topically to the skin or mucosa, i.e. dermally or transdermally. Example pharmaceutical and recreational drug formulations for this purpose include gels, hydrogels, lotions, solutions, creams, ointments, dusting powders, cosmetics, oils, eye drops, dressings, foams, films, skin patches, wafers, implants, sponges, fibres, bandages and microemulsions. Liposomes may also be used. Example carriers include alcohol, water, mineral oil, liquid petrolatum, white petrolatum, glycerin, polyethylene glycol and propylene glycol. Penetration enhancers may be incorporated (see: for example, Finnin, B. and Morgan, T. M., 1999 J. Pharm. Sci, 88 (10), 955-958).

Other means of topical administration include delivery by electroporation, iontophoresis, phonophoresis, sonophoresis and microneedle or needle-free (e.g., Powderject™, Bioject™, etc.) injection.

Pharmaceutical and recreational drug formulations for inhalation or insufflation include solutions and suspensions in pharmaceutically acceptable, aqueous or organic solvents, or mixtures thereof, and powders. The liquid or solid pharmaceutical compositions can contain suitable pharmaceutically acceptable excipients. In some embodiments, the pharmaceutical compositions are administered by the oral or nasal respiratory route for local or systemic effect. Pharmaceutical compositions in pharmaceutically acceptable solvents can be nebulized by use of inert gases. Nebulized solutions can be inhaled directly from the nebulizing device or the nebulizing device can be attached to a face mask tent, or intermittent positive pressure breathing machine. Solution, suspension, or powder pharmaceutical compositions can be administered, e.g., orally or nasally, from devices that deliver the formulation in an appropriate manner.

In further embodiments, in which the halogenated psilocybin compounds of present disclosure are used as a recreational drug, the compounds may be included in compositions such as food or food product, a beverage, a food seasoning, a personal care product, such as a cosmetic, perfume or bath oil, or oils (both for topical administration as massage oil, or to be burned or aerosolized). The chemical compounds of the present disclosure may also be included in a "vape" product, which may also include other drugs, such as nicotine, and flavorings.

The pharmaceutical formulations comprising the chemical compounds of the present disclosure may be used to treat a subject, and in particular to treat a psychiatric disorder in a subject. Accordingly, the present disclosure includes in a further embodiment, a method for treating a psychiatric disorder, the method comprising administering to a subject in need thereof a pharmaceutical formulation comprising a chemical compound or salts thereof having the formula (I):—



(I)

51

wherein, at least one of R_2 , R_4 , R_5 , R_6 or R_7 is a halogen atom, wherein each non-halogenated R_2 , R_5 , R_6 , or R_7 is a hydrogen atom or an alkyl group or O-alkyl group, wherein R_4 when it is not halogenated is a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group, and wherein R_{3A} and R_{3B} are a hydrogen atom, or an alkyl group, an aryl group, or an acyl group, wherein the pharmaceutical formulation is administered in an effective amount to treat the psychiatric disorder in the subject.

Psychiatric disorders that may be treated include, for example, neurodevelopmental disorders such as intellectual disability, global development delay, communication disorders, autism spectrum disorder, and attention-deficit hyperactivity disorder (ADHD); bipolar and related disorders, such as mania, and depressive episodes; anxiety disorder, such as generalized anxiety disorder (GAD), agoraphobia, social anxiety disorder, specific phobias (natural events, medical, animal, situational, for example), panic disorder, and separation anxiety disorder; stress disorders, such as acute stress disorder, adjustment disorders, post-traumatic stress disorder (PTSD), and reactive attachment disorder; dissociative disorders, such as dissociative amnesia, dissociative identity disorder, and depersonalization/derealization disorder; somatoform disorders, such as somatic symptom disorders, illness anxiety disorder, conversion disorder, and factitious disorder; eating disorders, such as anorexia nervosa, bulimia nervosa, rumination disorder, pica, and binge-eating disorder; sleep disorders, such as narcolepsy, insomnia disorder, hypersomnolence, breathing-related sleep disorders, parasomnias, and restless legs syndrome; disruptive disorders, such as kleptomania, pyromania, intermittent explosive disorder, conduct disorder, and oppositional defiant disorder; depressive disorders, such as disruptive mood dysregulation disorder, major depressive disorder, persistent depressive disorder (dysthymia), premenstrual dysphoric disorder, substance/medication-induced depressive disorder, postpartum depression, and depressive disorder caused by another medical condition, for example, psychiatric and existential distress within life-threatening cancer situations (ACS Pharmacol Transl Sci 4: 553-562; J Psychiatr Res 137: 273); substance-related disorders, such as alcohol-related disorders, cannabis related disorders, inhalant-use related disorders, stimulant use disorders, and tobacco use disorders; neurocognitive disorders, such as delirium; schizophrenia; compulsive disorders, such as obsessive compulsive disorders (OCD), body dysmorphic disorder, hoarding disorder, trichotillomania disorder, excoriation disorder, substance/medication induced obsessive-compulsive disorder, and obsessive-compulsive disorder related to another medical condition; and personality disorders, such as antisocial personality disorder, avoidant personality disorder, borderline personality disorder, dependent personality disorder, histrionic personality disorder, narcissistic personality disorder, obsessive-compulsive personality disorder, paranoid personality disorder, schizoid personality disorder, schizotypal personality disorder.

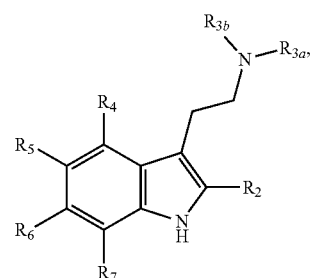
In an aspect, the compounds of the present disclosure may be used to be contacted with a 5-HT_{2A} receptor to thereby modulate the 5-HT_{2A} receptor. Such contacting includes bringing a compound of the present disclosure and 5-HT_{2A} receptor together under in vitro conditions, for example, by introducing the compounds in a sample containing a 5-HT_{2A} receptor, for example, a sample containing purified 5-HT_{2A} receptors, or a sample containing cells comprising 5-HT_{2A} receptors. In vitro conditions further include the conditions described in Example 3 hereof. Contacting further includes bringing a compound of the present disclosure and 5-HT_{2A}

52

receptor together under in vivo conditions. Such in vivo conditions include the administration to an animal or human subject, for example, of a pharmaceutically effective amount of the compound of the present disclosure, when the compound is formulated together with a pharmaceutically active carrier, diluent or excipient, as hereinbefore described, to thereby treat the subject. Upon having contacted the 5-HT_{2A} receptor, the compound may activate the 5-HT_{2A} receptor or inhibit the 5-HT_{2A} receptor.

Thus, in a further aspect, the condition that may be treated in accordance herewith can be any 5-HT_{2A} receptor mediated disorder. Such disorders include, but are not limited to schizophrenia, psychotic disorder, attention deficit hyperactivity disorder, autism, and bipolar disorder.

The chemical compounds of the present disclosure may also be used as a feedstock material for other psilocybin derivatives. Thus in one embodiment, the chemical compounds of the present disclosure may be in used manufacture of a pharmaceutical or recreational drug formulation, wherein the manufacture may comprise derivatizing a chemical compound or salts thereof having the formula (I):



(I)

wherein, at least one of R_2 , R_4 , R_5 , R_6 or R_7 is a halogen atom, wherein each non-halogenated R_2 , R_5 , R_6 , or R_7 is a hydrogen atom or an alkyl group or O-alkyl group, wherein R_4 when it is not halogenated is a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group, and wherein R_{3A} and R_{3B} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group.

In order to use the compound having formula (I) as a feedstock, one or more halogen atoms may be replaced by any atoms or groups, for example hydrocarbon groups. Those of skill in the art will be generally familiar with methods that may be used to substitute halogen groups. In this respect, guidance may be found in Schnepel C. et al. (2017) Chem. Eur. J. 23:12064-12086; Durak L. J. et al. (2016) ACS Catal. 6: 1451; Runguphan W. et al. (2013) Org Lett 15: 2850; Corr M. J. et al. (2017) Chem. Sci. 8: 2039; and Roy A. D. et al. Chem. Comm. 4831.

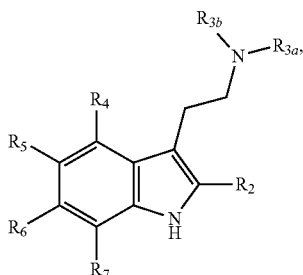
Turning now to methods of making the halogenated psilocybin derivatives, it is noted that the psilocybin compounds of the present disclosure may be prepared in any suitable manner, including any organic chemical synthesis methods, biosynthetic methods, or a combination thereof. Synthesis generally may involve selecting a psilocybin precursor compound, and modifying the psilocybin precursor compound to form a halogenated psilocybin derivative. In this respect, it is noted that a non-halogenated psilocybin derivative may be selected and modified to form psilocybin, which subsequently may be halogenated, or, alternatively, a halogenated psilocybin derivative may be selected to subsequently form a halogenated psilocybin. Suitable psilocybin precursor compounds include compounds comprising an

53

indole prototype structure (see: FIG. 2), including, for example tryptophan, tryptamine, 4-hydroxyindole, 4-hydroxytryptophan, 4-hydroxytryptamine, norbaeocystin and baeocystin, and halogenated forms thereof, notably, with respect to the indole prototype structure, 2-, 5-, 6-, 7-halogenated forms thereof. The halogenated psilocybin precursor compounds may be provided in a more or less chemically pure form, for example, in the form of a psilocybin precursor preparation having a purity of at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least 99.9%. The psilocybin precursor compounds may be chemically synthesized, or obtained from a fine chemical manufacturer.

In one embodiment of the present disclosure the halogenated psilocybin derivatives may be formed biosynthetically. Accordingly, the present disclosure further includes in one embodiment, a method of making a halogenated psilocybin derivative the method comprising:

- (a) contacting a halogenated psilocybin precursor compound with a host cell comprising a psilocybin biosynthetic enzyme complement; and
- (b) growing the host cell to produce a halogenated psilocybin compound or salts thereof having the formula (I):



wherein, at least one of R_2 , R_4 , R_5 , R_6 or R_7 is a halogen atom, wherein each non-halogenated R_2 , R_5 , R_6 , or R_7 is a hydrogen atom or an alkyl group or O-alkyl group, wherein R_4 when it is not halogenated is a phosphate group, hydrogen atom or an alkyl group or O-alkyl group, and wherein R_{3A} and R_{3B} are a hydrogen atom, or an alkyl group, an aryl group, or an acyl group of the chemical compound.

Implementation of the foregoing example embodiment initially involves providing halogenated psilocybin precursor compounds and host cells having a psilocybin biosynthetic enzyme complement. Accordingly, next, exemplary halogenated psilocybin precursor compounds and example host cells that may be selected and used in accordance with the present disclosure will be described. Thereafter, example methodologies and techniques will be described to contact and use the halogenated psilocybin precursor compounds and cells to produce example halogenated psilocybin compounds.

A variety of halogenated psilocybin precursor compounds may be selected, prepared and used. In some embodiments, for example, the halogenated psilocybin precursor compound is a compound comprising a halogenated indole prototype structure. Examples of such compounds are halogenated tryptophan, halogenated tryptamine, halogenated 4-hydroxyindole, halogenated 4-hydroxytryptophan, halogenated 4-hydroxytryptamine, halogenated norbaeocystin, and halogenated baeocystin, and in particular the 2-halogenated,

54

5-halogenated, 6-halogenated and 7-halogenated forms (with respect to the indole prototype structure) of any of the foregoing.

Turning now to the host cells that can be used in accordance with the present disclosure, it is initially noted that a variety of host cells may be selected in accordance with the present disclosure, including microorganism host cells, plant host cells, and animal host cells.

In accordance herewith the host cell includes a psilocybin biosynthetic enzyme complement. Such cells can be obtained in at least two ways. First, in some embodiments, host cells may be selected in which a psilocybin biosynthetic enzyme complement is naturally present. Generally cells naturally producing psilocybin for example, cells of fungal species belonging to the genus *psilocybe*, are suitable in this respect. Second, in some embodiments, a host cell that not naturally produces psilocybin may be modulated to produce a psilocybin biosynthetic enzyme complement. Thus, for example, a nucleic acid sequence encoding a psilocybin biosynthetic enzyme complement may be introduced into a host cell, and upon cell growth the host cells can make the psilocybin biosynthetic enzyme complement.

Typically a nucleic acid sequence encoding one or more enzymes constituting a psilocybin biosynthetic enzyme complement further includes one or more additional nucleic acid sequences, for example, a nucleic acid sequences controlling expression of the one or more enzymes, and these one or more additional nucleic acid sequences together with the nucleic acid sequence encoding the one or more enzymes can be said to form a chimeric nucleic acid sequence.

A host cell which upon cultivation expresses the chimeric nucleic acid can be selected and used in accordance with the present disclosure. Suitable host cells in this respect include, for example, microbial cells, such as bacterial cells, yeast cells, for example, and algal cells or plant cells. A variety of techniques and methodologies to manipulate host cells to introduce nucleic acid sequences in cells and attain expression exists and are well known to the skilled artisan. These methods include, for example, cation based methods, for example, lithium ion or calcium ion based methods, electroporation, biolistics, and glass beads based methods. As will be known to those of skill in the art, depending on the host cell selected, the methodology to introduce nucleic acid material in the host cell may vary, and, furthermore, methodologies may be optimized for uptake of nucleic acid material by the host cell, for example, by comparing uptake of nucleic acid material using different conditions. Detailed guidance can be found, for example, in Sambrook et al., *Molecular Cloning, a Laboratory Manual*, Cold Spring Harbor Laboratory Press, 2012, Fourth Ed. It is noted that the chimeric nucleic acid is a non-naturally occurring chimeric nucleic acid sequence and can be said to be heterologous to the host cell.

In some embodiments, the one or more enzymes constituting a psilocybin enzyme complement can be selected from by a nucleic acid sequence selected from the nucleic acid sequences consisting of:

- (a) SEQ.ID NO: 1, SEQ.ID NO: 3, SEQ.ID NO: 5, SEQ.ID NO: 7, SEQ.ID NO: 9, SEQ.ID NO: 11, SEQ.ID NO: 36, SEQ.ID NO: 38, SEQ.ID NO: 40, SEQ.ID NO: 42, SEQ.ID NO: 44, SEQ.ID NO: 46, SEQ.ID NO: 48, SEQ.ID NO: 50, SEQ.ID NO: 52, SEQ.ID NO: 54, SEQ.ID NO: 56, SEQ.ID NO: 58, SEQ.ID NO: 60, SEQ.ID NO: 64, SEQ.ID NO: 66, SEQ.ID NO: 68, SEQ.ID NO: 70, and SEQ.ID NO: 72;

- (b) a nucleic acid sequence that is substantially identical to any one of the nucleic acid sequences of (a);
- (c) a nucleic acid sequence that is substantially identical to any one of the nucleic acid sequences of (a) but for the degeneration of the genetic code;
- (d) a nucleic acid sequence that is complementary to any one of the nucleic acid sequences of (a);
- (e) a nucleic acid sequence encoding a polypeptide having any one of the amino acid sequences set forth in SEQ.ID NO: 2, SEQ.ID NO: 4, SEQ.ID NO: 6, SEQ.ID NO: 8, SEQ.ID NO: 10, SEQ.ID NO: 12, SEQ.ID NO: 37, SEQ.ID NO: 39, SEQ.ID NO: 41, SEQ.ID NO: 43, SEQ.ID NO: 45, SEQ.ID NO: 47, SEQ.ID NO: 49, SEQ.ID NO: 51, SEQ.ID NO: 53, SEQ.ID NO: 55, SEQ.ID NO: 57, SEQ.ID NO: 59, SEQ.ID NO: 61, SEQ.ID NO: 63, SEQ.ID NO: 65, SEQ.ID NO: 67, SEQ.ID NO: 69, SEQ.ID NO: 71, and SEQ.ID NO: 73;
- (f) a nucleic acid sequence that encodes a functional variant of any one of the amino acid sequences set forth in SEQ.ID NO: 2, SEQ.ID NO: 4, SEQ.ID NO: 6, SEQ.ID NO: 8, SEQ.ID NO: 10, SEQ.ID NO: 12, SEQ.ID NO: 37, SEQ.ID NO: 39, SEQ.ID NO: 41, SEQ.ID NO: 43, SEQ.ID NO: 45, SEQ.ID NO: 47, SEQ.ID NO: 49, SEQ.ID NO: 51, SEQ.ID NO: 53, SEQ.ID NO: 55, SEQ.ID NO: 57, SEQ.ID NO: 59, SEQ.ID NO: 61, SEQ.ID NO: 63, SEQ.ID NO: 65, SEQ.ID NO: 67, SEQ.ID NO: 69, SEQ.ID NO: 71, and SEQ.ID NO: 73; and
- (g) a nucleic acid sequence that hybridizes under stringent conditions to any one of the nucleic acid sequences set forth in (a), (b), (c), (d), (e) or (f).

Thus any of the nucleic acid sequence set forth in (a), (b), (c), (d), (e), (f) or (g) may be selected and introduced into a host cell. In general, however the nucleic acid sequence is selected in conjunction with the selected psilocybin precursor compound, as hereinafter further discussed in reference to FIGS. 8A and 8B.

One example host cell that conveniently may be used is *Escherichia coli*. The preparation of the *E. coli* vectors may be accomplished using commonly known techniques such as restriction digestion, ligation, gel electrophoresis, DNA sequencing, the polymerase chain reaction (PCR) and other methodologies. A wide variety of cloning vectors is available to perform the necessary steps required to prepare a recombinant expression vector. Among the vectors with a replication system functional in *E. coli*, are vectors such as pBR322, the pUC series of vectors, the M13 mp series of vectors, pBluescript etc. Suitable promoter sequences for use in *E. coli* include, for example, the T7 promoter, the T5 promoter, tryptophan (trp) promoter, lactose (lac) promoter, tryptophan/lactose (tac) promoter, lipoprotein (lpp) promoter, and λ phage PL promoter. Typically, cloning vectors contain a marker, for example, an antibiotic resistance marker, such as ampicillin or kanamycin resistance marker, allowing selection of transformed cells. Nucleic acid sequences may be introduced in these vectors, and the vectors may be introduced in *E. coli* by preparing competent cells, electroporation or using other well known methodologies to a person of skill in the art. *E. coli* may be grown in an appropriate medium, such as Luria-Broth medium and harvested. Recombinant expression vectors may readily be recovered from cells upon harvesting and lysing of the cells.

Another example host cell that may be conveniently used is a yeast cell. Example yeast host cells that can be used are yeast cells belonging to the genus *Candida*, *Kluyveromyces*, *Saccharomyces*, *Schizosaccharomyces*, *Pichia*, *Hansenula*, and *Yarrowia*. In specific example embodiments, the yeast

cell can be a *Saccharomyces cerevisiae* cell, a *Yarrowia lipolytica* cell, or *Pichia pastoris* cell.

A number of vectors exist for the expression of recombinant proteins in yeast host cells. Examples of vectors that may be used in yeast host cells include, for example, Yip type vectors, YEp type vectors, YRp type vectors, YCp type vectors, pGPD-2, pAO815, pGAPZ, pGAPZa, pHIL-D2, pHIL-S1, pPIC3L5K, pPIC9K, pPICZ, pPICZa, pPIC3K, pHWO10, pPUZZLE and 2 μ m plasmids. Such vectors are known to the art and are, for example, described in Cregg et al., Mol Biotechnol (2000) 16(1): 23-52. Suitable promoter sequences for use in yeast host cells are also known and described, for example, in Mattanovich et al., Methods Mol. Biol., 2012, 824:329-58, and in Romanos et al., 1992, Yeast 8: 423-488. Examples of suitable promoters for use in yeast host cells include promoters of glycolytic enzymes, like triosephosphate isomerase (TPI), phosphoglycerate kinase (PGK), glyceraldehyde-3-phosphate dehydrogenase (GAPDH or GAP) and variants thereof; lactase (LAC) and galactosidase (GAL), *P. pastoris* glucose-6-phosphate isomerase promoter (PPGI), the 3-phosphoglycerate kinase promoter (PPGK), the glycerol aldehyde phosphate dehydrogenase promoter (PGAP), translation elongation factor promoter (PTEF), *S. cerevisiae* eno/ase (ENO-1), *S. cerevisiae* galactokinase (GAL1), *S. cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH1, ADH2/GAP), *S. cerevisiae* triose phosphate isomerase (TPI), *S. cerevisiae* metallothionein (CUP1), and *S. cerevisiae* 3-phosphoglycerate kinase (PGK), and the maltase gene promoter (MAL). Marker genes suitable for use in yeast host cells are also known to the art. Thus, antibiotic resistance markers, such as ampicillin resistance markers, can be used in yeast, as well as marker genes providing genetic functions for essential nutrients, for example, leucine (LEU2), tryptophan (TRP1 and TRP2), uracil (URA3, URA5, URA6), histidine (HIS3), and the like. Methods for introducing vectors into yeast host cells can, for example, be found in S. Kawai et al., 2010, Bioeng. Bugs 1(6): 395-403.

Further, guidance with respect to the preparation of expression vectors and introduction thereof into host cells, including in *E. coli* cells, yeast cells, and other host cells, may be found in, for example: Sambrook et al., Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory Press, 2012, Fourth Ed.

Thus, to briefly recap, a host cell comprising a chimeric nucleic acid comprising (i) a nucleic acid sequence controlling expression in a host cell and (ii) a nucleic acid sequence encoding a psilocybin biosynthetic enzyme complement, can be prepared in accordance with the present disclosure.

In accordance herewith, host cells are grown to multiply and to express a chimeric nucleic acid. Expression of the chimeric nucleic acid results in the biosynthetic production in the host cell of a psilocybin biosynthetic enzyme complement. Growth media and growth conditions can vary depending on the host cell that is selected, as will be readily appreciated to those of ordinary skill in the art. Growth media typically contain a carbon source, one or several nitrogen sources, essential salts including salts of potassium, sodium, magnesium, phosphate and sulphate, trace metals, water soluble vitamins, and process aids including but not limited to antifoam agents, protease inhibitors, stabilizers, ligands and inducers. Example carbon sources are e.g. mono- or disaccharides. Example nitrogen sources are, e.g. ammonia, urea, amino acids, yeast extract, corn steep liquor and fully or partially hydrolyzed proteins. Example trace metals are e.g. Fe, Zn, Mn, Cu, Mo and H₃BO₃. Example water soluble vitamins are e.g. biotin, pantothenate, niacin,

thiamine, p-aminobenzoic acid, choline, pyridoxine, folic acid, riboflavin and ascorbic acid. Further, specific example media include liquid culture media for the growth of yeast cells and bacterial cells including, Luria-Bertani (LB) broth for bacterial cell cultivation, and yeast extract peptone dextrose (YEPD or YPD), for yeast cell cultivation. Further media and growth conditions can be found in Sambrook et al., *Molecular Cloning, a Laboratory Manual*, Cold Spring Harbor Laboratory Press, 2012, Fourth Ed.

In order for the host cells to produce the halogenated psilocybin compound, the cells are provided with a psilocybin precursor compound. Thus in accordance herewith, host cells may be contacted with a psilocybin precursor compound. In some embodiments, a psilocybin precursor compound can be exogenously supplied, for example, by including a psilocybin precursor compound in the growth medium of the host cells, and growing the host cells in a medium including the psilocybin precursor compound.

Referring next to FIGS. 8A and 8B, shown therein is an example natural biosynthetic pathway showing the conversion of example psilocybin precursor compounds to form psilocybin (FIG. 8A). Thus, as can be appreciated from FIG. 8A, various psilocybin precursor compounds may be selected and prepared in halogenated form, in conjunction with a psilocybin biosynthetic enzyme complement.

Thus, by way of example, referring to FIG. 8B, a chlorinated psilocybin precursor compound, for example, 6-chloro-4-hydroxy indole may be selected and contacted with a host cell comprising a psilocybin biosynthetic enzyme complement comprising PsiD, (e.g. SEQ.ID NO: 2) PsiH (e.g. SEQ.ID NO: 4), PsiK (e.g. SEQ.ID NO: 6) and PsiM (e.g. SEQ.ID NO: 8), and upon growth of the cells 6-chloro-psilocybin can be formed. Or by way of another example, 6-chloro-tryptophan may be selected and contacted with a host cell comprising a psilocybin biosynthetic enzyme complement comprising TrpB (SEQ.ID NO: 12) PsiD, (e.g. SEQ.ID NO: 2), PsiK (e.g. SEQ.ID NO: 6) and PsiM (e.g. SEQ.ID NO: 8), and upon growth of the cells 6-chloro-psilocybin can be formed. It will be clear to those of skill in the art that by referring to FIGS. 8A and 8B, other halogenated psilocybin precursors may be selected (e.g. 2-chloro tryptophan, 4-chloro-tryptophan, 5-chloro-tryptophan, 7-chlorotryptophan; 6-bromo-tryptophan, 6-fluoro-tryptophan, 6-iodotryptophan; and 6-chlorotryptamine, 6-chloro-4-hydroxytryptophan, 6-hydroxy-4-chloro-tryptamine, 6-chloro-norbaeocystin or 6-chloro-baeocystin), and contacted with a host cell comprising an appropriate psilocybin biosynthetic enzyme complement to produce a halogenated psilocybin derivative.

Thus, in one further example embodiment, a halogenated indole, such as 6-chloro-4-hydroxy indole, for example (see: FIG. 8B), may be selected and contacted with a host cell comprising a psilocybin biosynthetic enzyme complement comprising (i) a nucleic acid encoding TrpB selected from SEQ.ID NO: 12, SEQ.ID NO: 65, SEQ.ID NO: 67, SEQ.ID NO: 69, SEQ.ID NO: 71, and SEQ.ID NO: 73, or a nucleic acid sequence encoding a polypeptide sequence substantially similar to any of the aforementioned sequences; (ii) a nucleic acid encoding PsiD selected from SEQ.ID NO: 2, SEQ.ID NO: 10, SEQ.ID NO: 37, SEQ.ID NO: 39, SEQ.ID NO: 41, SEQ.ID NO: 43, SEQ.ID NO: 45, SEQ.ID NO: 47, SEQ.ID NO: 49, SEQ.ID NO: 51, SEQ.ID NO: 53, SEQ.ID NO: 55, and SEQ.ID NO: 57, or a nucleic acid sequence encoding a polypeptide sequence substantially similar to any of the aforementioned sequences; (iii) a nucleic acid encoding PsiK encoded by SEQ.ID NO: 6, or a nucleic acid sequence encoding a polypeptide sequence substantially

similar thereto; and (iv) a nucleic acid encoding PsiM selected from SEQ.ID NO: 8, SEQ.ID NO: 59 and SEQ.ID NO: 61 or a nucleic acid sequence encoding a polypeptide sequence substantially similar to any of the aforementioned sequences to produce a halogenated psilocybin derivative.

Thus, in one further example embodiment, a halogenated tryptophan, such as 6-chloro-4-hydroxy tryptophan, for example (see: FIG. 8B), may be selected and contacted with a host cell comprising a psilocybin biosynthetic enzyme complement comprising (i) a nucleic acid encoding PsiD selected from SEQ.ID NO: 2, SEQ.ID NO: 10, SEQ.ID NO: 37, SEQ.ID NO: 39, SEQ.ID NO: 41, SEQ.ID NO: 43, SEQ.ID NO: 45, SEQ.ID NO: 47, SEQ.ID NO: 49, SEQ.ID NO: 51, SEQ.ID NO: 53, SEQ.ID NO: 55, and SEQ.ID NO: 57, or a nucleic acid sequence encoding a polypeptide sequence substantially similar to any of the aforementioned sequences; (ii) a nucleic acid encoding PsiK encoded by SEQ.ID NO: 6, or a nucleic acid sequence encoding a polypeptide sequence substantially similar thereto; and (iii) a nucleic acid encoding PsiM selected from SEQ.ID NO: 8, SEQ.ID NO: 59 and SEQ.ID NO: 61 or a nucleic acid sequence encoding a polypeptide sequence substantially similar to any of the aforementioned sequences to produce a halogenated psilocybin derivative.

Thus, in one further example embodiment, a halogenated tryptamine, such as 6-chloro-4-hydroxy tryptamine, for example (see: FIG. 8B), may be selected and contacted with a host cell comprising a psilocybin biosynthetic enzyme complement comprising (i) a nucleic acid encoding PsiK encoded by SEQ.ID NO: 6, or a nucleic acid sequence encoding a polypeptide sequence substantially similar thereto; and (ii) a nucleic acid encoding PsiM selected from SEQ.ID NO: 8, SEQ.ID NO: 59 and SEQ.ID NO: 61 or a nucleic acid sequence encoding a polypeptide sequence substantially similar to any of the aforementioned sequences to produce a halogenated psilocybin derivative.

Thus, in one further example embodiment, a halogenated norbaeocystin, such as 6-chloro-norbaeocystin, or a halogenated baeocystin may, such as 6-chlorobaecocystin, for example (see: FIG. 8B), may be selected and contacted with a host cell comprising a psilocybin biosynthetic enzyme complement comprising a nucleic acid encoding PsiM selected from SEQ.ID NO: 8, SEQ.ID NO: 59 and SEQ.ID NO: 61 or a nucleic acid sequence encoding a polypeptide sequence substantially similar to any of the aforementioned sequences to produce a halogenated psilocybin derivative.

Thus, in one further example embodiment, a halogenated tryptophan, such as 6-chloro-4-hydroxy tryptophan, for example (see: FIG. 8B), may be selected and contacted with a host cell comprising a psilocybin biosynthetic enzyme complement comprising (i) a nucleic acid encoding PsiD selected from SEQ.ID NO: 2, SEQ.ID NO: 10, SEQ.ID NO: 37, SEQ.ID NO: 39, SEQ.ID NO: 41, SEQ.ID NO: 43, SEQ.ID NO: 45, SEQ.ID NO: 47, SEQ.ID NO: 49, SEQ.ID NO: 51, SEQ.ID NO: 53, SEQ.ID NO: 55, and SEQ.ID NO: 57, or a nucleic acid sequence encoding a polypeptide sequence substantially similar to any of the aforementioned sequences; (ii) a nucleic acid encoding PsiH encoded by SEQ.ID NO: 4, or a nucleic acid sequence encoding a polypeptide sequence substantially similar thereto; (iii) a nucleic acid encoding PsiK encoded by SEQ.ID NO: 6, or a nucleic acid sequence encoding a polypeptide sequence substantially similar thereto; and (iv) a nucleic acid encoding PsiM selected from SEQ.ID NO: 8, SEQ.ID NO: 59 and SEQ.ID NO: 61 or a nucleic acid sequence encoding

a polypeptide sequence substantially similar to any of the aforementioned sequences to produce a halogenated psilocybin derivative.

Thus, in one further example embodiment, a halogenated tryptamine, such as 6-chloro-tryptamine, for example (see: FIG. 8B), may be selected and contacted with a host cell comprising a psilocybin biosynthetic enzyme complement comprising (i) a nucleic acid encoding PsiH encoded by SEQ.ID NO: 4, or a nucleic acid sequence encoding a polypeptide sequence substantially similar thereto; (ii) a nucleic acid encoding PsiK encoded by SEQ.ID NO: 6, or a nucleic acid sequence encoding a polypeptide sequence substantially similar thereto; and (iii) a nucleic acid encoding PsiM selected from SEQ.ID NO: 8, SEQ.ID NO: 59 and SEQ.ID NO: 61 or a nucleic acid sequence encoding a polypeptide sequence substantially similar to any of the aforementioned sequences to produce a halogenated psilocybin derivative.

In some embodiments, the psilocybin can be a halogenated psilocybin precursor compound which is exogenously supplied to a host cell, for example by inclusion in the host cell's growth medium. Thus, for example, referring to FIG. 8, it will be understood that in accordance herewith, for example, 6-chloro-tryptophan or 6-chloro-4-hydroxy indole, may be included in the growth medium of a host cell comprising a psilocybin biosynthetic enzyme complement.

It is noted that in some embodiments, the host cells may in addition to the psilocybin biosynthetic enzyme complement include a halogenase. In such embodiments the cell may be contacted with a non-halogenated psilocybin precursor compound, and halogenation may occur in vivo in the host cells. Thus, referring again to FIG. 8, by way of example, non-halogenated tryptophan may be selected and contacted with a host cell comprising a psilocybin biosynthetic enzyme complement comprising PsiD, (e.g. SEQ.ID NO: 2) PsiH (e.g. SEQ.ID NO: 4), PsiK (e.g. SEQ.ID NO: 6) and PsiM (e.g. SEQ.ID NO: 8), the host cells further including a halogenase, and upon growth of the host cells in the presence of a halogen, for example in the form of a halogen salt, sodium chloride (NaCl) or sodium bromide (NaBr) halogenated psilocybin derivatives can be formed.

Suitable halogenases that may be used in accordance herewith include, for example, a PyrH halogenase of *Streptomyces rugosporus* (SEQ.ID NO: 13; SEQ.ID NO: 14), which may, for example, be contacted with non-halogenated tryptophan to form 5-halo-tryptophan; a KtZR halogenase of *Kutzneria* sp. 744 (SEQ.ID NO: 15; SEQ.ID NO: 16), which may be contacted with non-halogenated tryptophan to form 6-halo-tryptophan; a SstH halogenase of *Streptomyces toxytricini* (SEQ.ID NO: 17; SEQ.ID NO: 18), which may be contacted with non-halogenated tryptophan to form 6-halo-tryptophan; a Thal halogenase of *Streptomyces albogriseolus* (SEQ.ID NO: 19; SEQ.ID NO: 20), which may be contacted with non-halogenated tryptophan to form 6-halo-tryptophan; a BorH halogenase (see: SEQ.ID NO: 21; SEQ.ID NO: 22), which may be contacted with non-halogenated tryptophan to form 6-halo-tryptophan; a KtzQ halogenase of *Kutzneria* sp. 744 (SEQ.ID NO: 23; SEQ.ID NO: 24), which may be contacted with non-halogenated tryptophan to form 7-halo-tryptophan; a PrnA halogenase of *Pseudomonas fluorescens* (SEQ.ID NO: 25; SEQ.ID NO: 26), which may be contacted with non-halogenated tryptophan and other psilocybin precursor compounds to form 7-halo-tryptophan or 2-halo-tryptophan and other 7-halo or 2-halo psilocybin precursor compounds; a RebH halogenase of *Lentzea aerocolonigenes* (SEQ.ID NO: 27; SEQ.ID NO: 28), which may be contacted with non-halogenated trypto-

phan to form 7-halo-tryptophan; and a SatH halogenase of *Streptomyces albus* (SEQ.ID NO: 29), which may be contacted with non-halogenated tryptophan to form 6-halo-tryptophan.

It is noted that halogenases comprising amino acid sequences substantially similar to any of the aforementioned halogenase sequences may be prepared and used to modulate the substrate specificity and/or the enzyme's catalytic efficiency. Thus, for example, halogenases comprising amino acid sequences substantially similar to SEQ.ID NO: 14; SEQ.ID NO: 16; SEQ.ID NO: 18; SEQ.ID NO: 20; SEQ.ID NO: 22; SEQ.ID NO: 24; SEQ.ID NO: 26; SEQ.ID NO: 28, and SEQ.ID NO: 29, may be prepared and used to obtain 2-halo, 5-halo, 6-halo or 7-halo tryptophan, or halogenases comprising amino acid sequences substantially similar to SEQ.ID NO: 14; SEQ.ID NO: 16; SEQ.ID NO: 18; SEQ.ID NO: 20; SEQ.ID NO: 22; SEQ.ID NO: 24; SEQ.ID NO: 26; SEQ.ID NO: 28, and SEQ.ID NO: 29 may be prepared and used to halogenate other non-halogenated psilocybin precursor compounds, such as tryptamine, 4-hydroxyindole, 4-hydroxytryptophan, 4-hydroxytryptamine, norbaeocystin, and baecocystin, for example, to form halogenated forms thereof.

Thus, for example, the native polypeptide sequence of a RebH halogenase *Lentzea aerocolonigenes* (SEQ.ID NO: 27; SEQ.ID NO: 28) may be modulated to thereby modulate the substrate specificity. Thus, for example, a modulated *Lentzea aerocolonigenes* RebH halogenase, which may be referred to as a RebH-Y455W mutant, set forth in SEQ.ID NO: 30 prefers tryptamine over tryptophan as a substrate and when contacted with tryptamine forms 7-halo-tryptamine.

By way of a further example, a modulated *Lentzea aerocolonigenes* RebH halogenase, which may be referred to as a RebH-N470S mutant, set forth in SEQ.ID NO: 31 prefers tryptamine over tryptophan as a substrate and when contacted with tryptamine forms 7-halo-tryptamine.

By way of a further example, a modulated *Lentzea aerocolonigenes* RebH halogenase, which may be referred to as a RebH-8F mutant, set forth in SEQ.ID NO: 32 prefers tryptamine over tryptophan as a substrate and when contacted with tryptamine forms 6-halo-tryptamine.

By way of a further example, a modulated *Lentzea aerocolonigenes* RebH halogenase, which may be referred to as a RebH-10F mutant, set forth in SEQ.ID NO: 33 prefers tryptamine over tryptophan as a substrate and when contacted with tryptamine forms 5-halo-tryptamine.

By way of a further example, a modulated *Lentzea aerocolonigenes* RebH halogenase, which may be referred to as a RebH-3SS mutant, set forth in SEQ.ID NO: 34 has an altered substrate acceptance profile and converts tricyclic tryptoline derivatives.

By way of yet a further example, a modulated *Lentzea aerocolonigenes* RebH halogenase, which may be referred to as a RebH-3SS mutant set forth in SEQ.ID NO: 35 has an altered substrate acceptance profile and converts a broad range of large indoles and carbazoles.

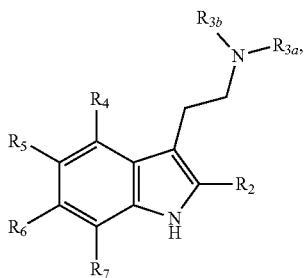
The foregoing examples of halogenases exhibiting modulated substrate specificity and/or catalytic efficiency are known to the art and are documented in, for example, Andorfer M. C., et al. (2016) Chem. Sci. 7: 3720; Andorfer M. C. et al. (2017) Catalysis 7: 1897; Glenn W. S., et al. (2011) J. Am. Chem. Soc. 133: 19346; and Payne J. T. et al. (2015) Angew. Chem. Int. Ed. 54: 4226.

Accordingly, in one embodiment the halogenase can be encoded by a nucleic acid selected from:

61

- (a) SEQ.ID NO: 13, SEQ.ID NO: 15, SEQ. ID NO: 17, SEQ.ID NO: 19, SEQ. ID NO: 21, SEQ.ID NO: 23, SEQ.ID NO: 25, and SEQID NO: 27;
- (b) a nucleic acid sequence that is substantially identical to any one of the nucleic acid sequences of (a);
- (c) a nucleic acid sequence that is substantially identical to any one of the nucleic acid sequences of (a) but for the degeneration of the genetic code;
- (d) a nucleic acid sequence that is complementary to any one of the nucleic acid sequences of (a);
- (e) a nucleic acid sequence encoding a polypeptide having any one of the amino acid sequences set forth in SEQ.ID NO: 14, SEQ.ID NO: 16, SEQ. ID NO: 18, SEQ.ID NO: 20, SEQ. ID NO: 22, SEQ.ID NO: 24, SEQ.ID NO: 26, SEQ.ID NO: 28, SEQ.ID NO: 29, SEQ.ID NO: 30, SEQ.ID NO: 31, SEQ.ID NO: 32, SEQ.ID NO: 33, SEQ.ID NO: 34, and SEQ.ID NO: 35;
- (f) a nucleic acid sequence that encodes a functional variant of any one of the amino acid sequences set forth in SEQ.ID NO: 14, SEQ.ID NO: 16, SEQ. ID NO: 18, SEQ.ID NO: 20, SEQ. ID NO: 22, SEQ.ID NO: 24, SEQ.ID NO: 26, SEQ.ID NO: 28, SEQ.ID NO: 29, SEQ.ID NO: 30, SEQ.ID NO: 31, SEQ.ID NO: 32, SEQ.ID NO: 33, SEQ.ID NO: 34, and SEQ.ID NO: 35; and
- (g) a nucleic acid sequence that hybridizes under stringent conditions to any one of the nucleic acid sequences set forth in (a), (b), (c), (d), (e) or (f).

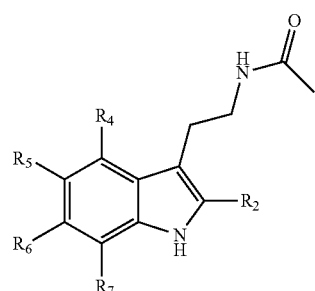
In one further embodiment, the host cell can comprise an acetyl transferase capable of acetylating R_{3a} or R_{3b} , wherein the cell comprises a compound having chemical formula (I):



wherein, at least one of R_2 , R_4 , R_5 , R_6 or R_7 is a halogen atom, wherein each non-halogenated R_2 , R_5 , R_6 , or R_7 is a hydrogen atom or an alkyl group or O-alkyl group, wherein R_4 when it is not halogenated is a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group, and wherein at least one of R_{3a} and R_{3b} is a hydrogen atom,

62

to form a compound having the chemical formula (XIX):



(XIX)

wherein, at least one of R_2 , R_4 , R_5 , R_6 or R_7 is a halogen atom, wherein each non-halogenated R_2 , R_5 , R_6 , or R_7 is a hydrogen atom or an alkyl group or O-alkyl group, wherein R_4 when it is not halogenated is a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group.

In some embodiments, R_{3a} and R_{3b} are each a hydrogen atom.

Suitable acyl transferases that may be used in accordance herewith include a PsmF acetyl transferase obtained from *Streptomyces griseofuscus*, including a PsmF N-acetyltransferase comprising polypeptide sequence SEQ.ID NO: 63 set forth herein, and further including a PsmF polypeptide encoded by a nucleic acid selected from:

- (a) SEQ.ID NO: 62;
- (b) a nucleic acid sequence that is substantially identical to the nucleic acid sequence of (a);
- (c) a nucleic acid sequence that is substantially identical to the nucleic acid sequence of (a) but for the degeneration of the genetic code;
- (d) a nucleic acid sequence that is complementary to the nucleic acid sequence of (a);
- (e) a nucleic acid sequence encoding a polypeptide having an amino acid sequences set forth in SEQ.ID NO: 63;
- (f) a nucleic acid sequence that encodes a functional variant of the amino acid sequence set forth in SEQ.ID NO: 63; and
- (g) a nucleic acid sequence that hybridizes under stringent conditions to any one of the nucleic acid sequences set forth in (a), (b), (c), (d), (e) or (f).

It will be clear to those of skill in the art that a significant variety of different halogenated and non-halogenated psilocybin precursor compounds may be selected. FIGS. 8A and 8B in this respect provide guidance and allow a person of skill in the art to select appropriate psilocybin precursor compounds and a matching a psilocybin biosynthetic enzyme complement.

Upon production by the host cells of the halogenated psilocybin compounds in accordance with the methods of the present disclosure, the halogenated psilocybin compounds may be extracted from the host cell suspension, and separated from other constituents within the host cell suspension, such as media constituents and cellular debris. Separation techniques will be known to those of skill in the art and include, for example, solvent extraction (e.g. butane, chloroform, ethanol), column chromatography based techniques, high-performance liquid chromatography (HPLC), for example, and/or countercurrent separation (CCS) based systems. The recovered halogenated psilocybin compounds may be obtained in a more or less pure form, for example, a preparation of halogenated psilocybin compounds of at least about 60% (w/w), about 70% (w/w), about 80% (w/w),

about 90% (w/w), about 95% (w/w), about 96% (w/w), about 97% (w/w), about 98% (w/w) or about 99% (w/w) purity may be obtained. Thus, in this manner, halogenated psilocybin derivatives in more or less pure form may be prepared.

Similarly, other methods of making the halogenated psilocybin compounds that may be used in accordance herewith may yield preparations of halogenated compounds of at least about 60% (w/w), about 70% (w/w), about 80% (w/w), about 90% (w/w), about 95% (w/w), about 96% (w/w), about 97% (w/w), about 98% (w/w), or about 99% (w/w) purity.

It will now be clear from the foregoing that novel halogenated psilocybin derivatives are disclosed herein. The halogenated psilocybin compounds may be formulated for use as a pharmaceutical drug or recreational drug. The halogenated psilocybin compounds may also be used as a feedstock to produce other psilocybin derivatives.

Summary of Sequences

SEQ.ID NO: 1 sets forth a *Psilocybe cubensis* nucleic acid sequence encoding a PsiD polypeptide.

SEQ.ID NO: 2 sets forth a deduced amino acid sequence of a *Psilocybe cubensis* PsiD polypeptide.

SEQ.ID NO: 3 sets forth a *Psilocybe cubensis* nucleic acid sequence encoding a PsiH polypeptide.

SEQ.ID NO: 4 sets forth a deduced amino acid sequence of a *Psilocybe cubensis* PsiH polypeptide.

SEQ.ID NO: 5 sets forth a *Psilocybe cubensis* nucleic acid sequence encoding a PsiK polypeptide.

SEQ.ID NO: 6 sets forth a deduced amino acid sequence of a *Psilocybe cubensis* PsiK polypeptide.

SEQ.ID NO: 7 sets forth a *Psilocybe cubensis* nucleic acid sequence encoding a PsiM polypeptide.

SEQ.ID NO: 8 sets forth a deduced amino acid sequence of a *Psilocybe cubensis* PsiM polypeptide.

SEQ.ID NO: 9 sets forth a *Psilocybe cubensis* nucleic acid sequence encoding a Psi-ncAAAD polypeptide.

SEQ.ID NO: 10 sets forth a deduced amino acid sequence of a *Psilocybe cubensis* Psi-ncAAAD polypeptide.

SEQ.ID NO: 11 sets forth a *Psilocybe cubensis* nucleic acid sequence encoding a TrpB polypeptide.

SEQ.ID NO: 12 sets forth a deduced amino acid sequence of a *Psilocybe cubensis* TrpB polypeptide.

SEQ.ID NO: 13 sets forth a *Streptomyces rugosporus* nucleic acid sequence encoding a PyrH halogenase polypeptide.

SEQ.ID NO: 14 sets forth a deduced amino acid sequence of a *Streptomyces rugosporus* PyrH halogenase polypeptide.

SEQ.ID NO: 15 sets forth a *Kutzneria* sp. 744 nucleic acid sequence encoding a KtzR halogenase polypeptide.

SEQ.ID NO: 16 sets forth a deduced amino acid sequence of a *Kutzneria* sp. 744 KtzR halogenase polypeptide.

SEQ.ID NO: 17 sets forth a *Streptomyces toxytrinic* nucleic acid sequence encoding a SttH halogenase polypeptide.

SEQ.ID NO: 18 sets forth a deduced amino acid sequence of a *Streptomyces toxytrinic* SttH halogenase polypeptide.

SEQ.ID NO: 19 sets forth a *Streptomyces albogriseolus* nucleic acid sequence encoding a Thal halogenase polypeptide.

SEQ.ID NO: 20 sets forth a deduced amino acid sequence of a *Streptomyces albogriseolus* Thal halogenase polypeptide.

SEQ.ID NO: 21 sets forth a nucleic acid sequence encoding a BorH halogenase polypeptide of an unidentified soil bacterium.

SEQ.ID NO: 22 sets forth a deduced amino acid sequence of a BorH halogenase polypeptide of an unidentified soil bacterium.

SEQ.ID NO: 23 sets forth a *Kutzneria* sp. 744 nucleic acid sequence encoding a KtzQ halogenase polypeptide.

SEQ.ID NO: 24 sets forth a deduced amino acid sequence of a *Kutzneria* sp. 744 KtzQ halogenase polypeptide.

SEQ.ID NO: 25 sets forth a *Pseudomonas fluorescens* nucleic acid sequence encoding a PrnA halogenase polypeptide.

SEQ.ID NO: 26 sets forth a deduced amino acid sequence of a *Pseudomonas fluorescens* PrnA halogenase polypeptide.

SEQ.ID NO: 27 sets forth a nucleic acid sequence encoding a *Lentzea areocolonigenes* RebH halogenase polypeptide.

SEQ.ID NO: 28 sets forth a deduced amino acid sequence of a *Lentzea aerocolonigenes* RebH halogenase polypeptide.

SEQ.ID NO: 29 sets forth a deduced amino acid sequence of a *Streptomyces albus* SatH halogenase polypeptide.

SEQ.ID NO: 30 sets forth a deduced amino acid sequence of a mutated *Lentzea aerocolonigenes* RebH halogenase polypeptide.

SEQ.ID NO: 31 sets forth a deduced amino acid sequence of another mutated *Lentzea aerocolonigenes* RebH halogenase polypeptide.

SEQ.ID NO: 32 sets forth a deduced amino acid sequence of another mutated *Lentzea aerocolonigenes* RebH halogenase polypeptide.

SEQ.ID NO: 33 sets forth a deduced amino acid sequence of another mutated *Lentzea aerocolonigenes* RebH halogenase polypeptide.

SEQ.ID NO: 34 sets forth a deduced amino acid sequence of another mutated *Lentzea aerocolonigenes* RebH halogenase polypeptide.

SEQ.ID NO: 35 sets forth a deduced amino acid sequence of another mutated *Lentzea aerocolonigenes* RebH halogenase polypeptide.

SEQ.ID NO: 36 sets forth a nucleic acid sequence encoding a *Bacillus atrophaeus* BaTDC tryptophan decarboxylase polypeptide.

SEQ.ID NO: 37 sets forth a deduced amino acid sequence of a *Bacillus atrophaeus* BaTDC tryptophan decarboxylase polypeptide.

SEQ.ID NO: 38 sets forth a nucleic acid sequence encoding a *Ruminococcus gnavus* RgnTDC tryptophan decarboxylase polypeptide.

SEQ.ID NO: 39 sets forth a deduced amino acid sequence of a *Ruminococcus gnavus* RgnTDC tryptophan decarboxylase polypeptide.

SEQ.ID NO: 40 sets forth a nucleic acid sequence encoding a *Streptomyces griseofuscus* SgTDC tryptophan decarboxylase polypeptide.

SEQ.ID NO: 41 sets forth a deduced amino acid sequence of a *Streptomyces griseofuscus* SgTDC tryptophan decarboxylase polypeptide.

SEQ.ID NO: 42 sets forth a nucleic acid sequence encoding a *Ceriporiopsis subvermispota* CsTDC tryptophan decarboxylase polypeptide.

SEQ.ID NO: 43 sets forth a deduced amino acid sequence of a *Ceriporiopsis subvermispota* CsTDC tryptophan decarboxylase polypeptide.

SEQ.ID NO: 44 sets forth a nucleic acid sequence encoding a *Clostridium sporogenes* ClostSporTDC tryptophan decarboxylase polypeptide.

SEQ.ID NO: 45 sets forth a deduced amino acid sequence of a *Clostridium sporogenes* ClostSporTDC tryptophan decarboxylase polypeptide.

SEQ.ID NO: 46 sets forth a nucleic acid sequence encoding a *Homo sapiens* HsDDC DOPA decarboxylase polypeptide.

SEQ.ID NO: 47 sets forth a deduced amino acid sequence of a *Homo sapiens* HsDDC DOPA decarboxylase polypeptide.

SEQ.ID NO: 48 sets forth a nucleic acid sequence encoding an *Ophiorrhiza pumila* OpTDC tryptophan decarboxylase polypeptide.

SEQ.ID NO: 49 sets forth a deduced amino acid sequence of an *Ophiorrhiza pumila* OpTDC tryptophan decarboxylase polypeptide.

SEQ.ID NO: 50 sets forth a nucleic acid sequence encoding an *Oryza sativa* OsTDC1 tryptophan decarboxylase polypeptide.

SEQ.ID NO: 51 sets forth a deduced amino acid sequence of an *Oryza sativa* OsTDC1 tryptophan decarboxylase polypeptide.

SEQ.ID NO: 52 sets forth a nucleic acid sequence encoding another *Oryza sativa* OsTDC2 tryptophan decarboxylase polypeptide.

SEQ.ID NO: 53 sets forth a deduced amino acid sequence of another *Oryza sativa* OsTDC2 tryptophan decarboxylase polypeptide.

SEQ.ID NO: 54 sets forth a nucleic acid sequence encoding a *Xenorhabdus doucetiae* XdTDC decarboxylase polypeptide.

SEQ.ID NO: 55 sets forth a deduced amino acid sequence of a *Xenorhabdus doucetiae* XdTDC decarboxylase polypeptide.

SEQ.ID NO: 56 sets forth a nucleic acid sequence encoding a *Camptotheca acuminata* CaTDC tryptophan decarboxylase polypeptide.

SEQ.ID NO: 57 sets forth a deduced amino acid sequence of a *Camptotheca acuminata* CaTDC tryptophan decarboxylase polypeptide.

SEQ.ID NO: 58 sets forth a nucleic acid sequence encoding an *Oryctolagus cuniculus* OcINMT indolethylamine N-methyltransferase polypeptide.

SEQ.ID NO: 59 sets forth a deduced amino acid sequence of an *Oryctolagus cuniculus* OcINMT indolethylamine N-methyltransferase polypeptide.

SEQ.ID NO: 60 sets forth a nucleic acid sequence encoding a *Hordeum vulgare* HvNMT N-methyltransferase polypeptide.

SEQ.ID NO: 61 sets forth a deduced amino acid sequence of a *Hordeum vulgare* HvNMT N-methyltransferase polypeptide.

SEQ.ID NO: 62 sets forth a nucleic acid sequence encoding a *Streptomyces griseofuscus* PsmF N-acetyltransferase polypeptide.

SEQ.ID NO: 63 sets forth a deduced amino acid sequence of a *Streptomyces griseofuscus* PsmF N-acetyltransferase polypeptide.

SEQ.ID NO: 64 sets forth a nucleic acid sequence encoding a mutated *Thermotoga maritima* TmTrpB-2F3 tryptophan synthase subunit B polypeptide.

SEQ.ID NO: 65 sets forth a deduced amino acid sequence of a mutated *Thermotoga maritima* TmTrpB-2F3 tryptophan synthase subunit B polypeptide.

SEQ.ID NO: 66 sets forth a nucleic acid sequence encoding a mutated *Pyrococcus furiosus* PfTrpB-0A9 tryptophan synthase subunit B polypeptide.

SEQ.ID NO: 67 sets forth a deduced amino acid sequence of a mutated *Pyrococcus furiosus* PfTrpB-0A9 tryptophan synthase subunit B polypeptide.

SEQ.ID NO: 68 sets forth a nucleic acid sequence encoding another mutated *Pyrococcus furiosus* PfTrpB-2A6 tryptophan synthase subunit B polypeptide.

SEQ.ID NO: 69 sets forth a deduced amino acid sequence of a mutated *Pyrococcus furiosus* PfTrpB-2A6 tryptophan synthase subunit B polypeptide.

SEQ.ID NO: 70 sets forth a nucleic acid sequence encoding another mutated *Pyrococcus furiosus* PfTrpB-7E6 tryptophan synthase subunit B polypeptide.

SEQ.ID NO: 71 sets forth a deduced amino acid sequence of a mutated *Pyrococcus furiosus* PfTrpB-7E6 tryptophan synthase subunit B polypeptide.

SEQ.ID NO: 72 sets forth a nucleic acid sequence encoding another mutated *Pyrococcus furiosus* PfTrpB-Quat tryptophan synthase subunit B polypeptide.

SEQ.ID NO: 73 sets forth a deduced amino acid sequence of a mutated *Pyrococcus furiosus* PfTrpB-Quat tryptophan synthase subunit B polypeptide.

SEQ.ID NO: 74 sets forth a nucleic acid sequence used as integration cassette XII-4::TADH1-PsiH-HA-PPGK1-PTDH3-CPR-c-myc-TCYC1 for homologous recombination-based genome editing of *Saccharomyces cerevisiae* strain Sc-1.

SEQ.ID NO: 75 sets forth a nucleic acid sequence used as integration cassette XII-5::TADH1-PsiK-V5-PPGK1-PTDH3-PsiM-FLAG-TCYC1 for homologous recombination-based genome editing of *Saccharomyces cerevisiae* strain Sc-1.

SEQ.ID NO: 76 sets forth a nucleic acid sequence of modified pMM1 vector pMM1-PTDH3-ClostSporTDC-His-TCYC1 for engineering of *Saccharomyces cerevisiae* strain Sc-1.

SEQ.ID NO: 77 sets forth a nucleic acid sequence for *Saccharomyces cerevisiae* promoter PGK1.

SEQ.ID NO: 78 sets forth a nucleic acid sequence for *Saccharomyces cerevisiae* promoter TDH3.

SEQ.ID NO: 79 sets forth a nucleic acid sequence for *Saccharomyces cerevisiae* promoter CLN1.

SEQ.ID NO: 80 sets forth a nucleic acid sequence for *Saccharomyces cerevisiae* promoter UGA1.

SEQ.ID NO: 81 sets forth a nucleic acid sequence encoding a synthetic HA epitope tag polypeptide

SEQ.ID NO: 82 sets forth deduced amino acid sequence of a synthetic HA epitope tag polypeptide

SEQ.ID NO: 83 sets forth a nucleic acid sequence encoding a synthetic c-MYC epitope tag polypeptide

SEQ.ID NO: 84 sets forth deduced amino acid sequence of a synthetic c-MYC epitope tag polypeptide

SEQ.ID NO: 85 sets forth a nucleic acid sequence encoding a synthetic FLAG epitope tag polypeptide

SEQ.ID NO: 86 sets forth deduced amino acid sequence of a synthetic FLAG epitope tag polypeptide

SEQ.ID NO: 87 sets forth a nucleic acid sequence encoding a synthetic V5 epitope tag polypeptide

SEQ.ID NO: 88 sets forth deduced amino acid sequence of a synthetic V5 epitope tag polypeptide

SEQ.ID NO: 89 sets forth a nucleic acid sequence encoding a synthetic HIS epitope tag polypeptide

SEQ.ID NO: 90 sets forth deduced amino acid sequence of a synthetic HIS epitope tag polypeptide

SEQ.ID NO: 91 sets forth a nucleic acid sequence of pMM1 vector

SEQ.ID NO: 92 sets forth a nucleic acid sequence of pCDM4 vector

SEQ.ID NO: 93 sets forth a nucleic acid sequence of pETM6 vector

67

SEQ.ID NO: 94 sets forth a nucleic acid sequence of pET28a(+) vector

SEQ.ID NO: 95 sets forth a nucleic acid sequence encoding a *Psilocybe cubensis* PcCPR cytochrome P450 reductase polypeptide.

SEQ.ID NO: 96 sets forth a deduced amino acid sequence of a *Psilocybe cubensis* PcCPR cytochrome P450 reductase polypeptide.

SEQ.ID NO: 97 sets forth a nucleic acid sequence of pETM6-H10 vector.

SEQ.ID NO: 98 sets forth a nucleic acid sequence of pRSM3 vector.

SEQUENCE LISTING

SEQ. ID NO: 1
 ATGCAGGTGATACCCGCGTGCAACTCGGCAGCAATAAGATCACTATGT
 CCTACTCCCGAGTCTTTTAGAAACATGGGATGGCTCTCTGTCAGCGAT
 GCGGTCTACAGCGAGTTCATAGGAGAGTTGGCTACCCGCGCTTCCAAT
 CGAAATTACTCCAACGAGTTCGGCTCATGCAACCTATCCAGGAATTC
 AAGGCTTTTATTGAAAGCGACCCGGTGGTGACCAAGAATTTATTGAC
 ATGTTTCGAGGGCATTTCAGGACTCTCCAAGGAATTATCAGGAACATGT
 AATATGTTTCAACGATATCTTTCGAAAGCTCCCGTCTACGAGACCTT
 GGCCCTCCCGTTTATATGATTATGGCCAAATTAATGAACACCCGAGCG
 GGCTTCTCTGCATTACAGAGACAAAGGTTGAACCTTCACTTCAAAAAA
 CTTTTCGATACCTGGGGATTGTTCTGTCTTCGAAAGATTCTCGAAAT
 GTTCTTGTGGCCGACGAGTTCGACGACAGACATTGCGGCTGGTTGAAC
 GAGCGGGCCTTGTCTGCTATGGTTAAACATTACAATGGACGCGCATTT
 GATGAAGTCTTCTCTCGGATAAAAAATGCCCCATACTACGCTTCAAC
 TCTTACGACGACTCTTTAATCGCAGATTTGAAACCGAGATATCGAC
 CGACCTGTAGTCGGTGGAGTTAACACACCCCTCATTCTGTCTGCT
 TGCGAATCACTTTCTACAACGCTCTCTTATGACGTCCAGTCTCTCGAC
 ACTTTAGTTTTCAAAGGAGAGACTTATTCGCTTAAGCATTGTCTGAAT
 AATGACCCCTTTCACCCACAATTTCGAGCATGGGAGTATCTACAAGGA
 TTCTTGAACGTCACCGCTTACCACCGATGGCAGCACCCGTCATGGG
 ACAATCGTCAAAATCATCAACGTTCCAGGTACCTACTTTGCGCAAGCC
 CCGAGCACGATTGGCGACCCCTATCCCGGATAACGATTACGACCCACCT
 CCTTACCTTAAGTCTCTTGTCTACTTCTCTAATATTGCGCAAGGCAA
 ATTATGTTTATTGAAGCCGACACAAGGAAATTGGCCTCATTTTCTCTT
 GTGTTTCATCGGCATGACCGAAATCTCGACATGTGAAGCCACGGTGTCC
 GAAGGTCAACACGTCATCGTGGCGATGACTTGGGAATGTTCCATTTCT
 GGTGGTCTCTCGTTCGCGCTTGGTCTGAGGAAGGATTGCAGGGCAGAG
 ATCGTTGAAAAGTTTACCGAACCCGGAACAGTGATCAGAATCAACGAA
 GTCGTCGCTGCTCTAAAGGCTTAG

SEQ. ID NO: 2
 MQVIPACNSAAIRSLCPTPESFRNMGLSVSDAVYSEF IGELATRASN
 RNYSEFGLMQPIQEFKAFIESDPVHVEFIDMFEIGIDSPRNYQELC
 NMFNDIFRKAPVYGLGPPVYIMAKLMNTRAGFSAFTRQRLNLHFKK

68

-continued

LFDTWGLFLSSKDSRNLVADQFDDRHCGWLNERALSAMVKHYNGRAF
 DEVFLCDKNAPYYGFNSYDDFFNRRFRNRDIDRPVVGVMNTTLLISAA
 5 CESLSYNVSYDVQSLDTLVFKGETYSLKHLNNDFPTPQFEHGSILQG
 FLNVTAYHRWHAPVNGTIVKI INVPGTYFAQAPSTIGDPIPDNDYDPP
 PYLKSLVYFSNIAARQIMFIEADNKEIGLIFLVFIGMTEISTCEATVS
 10 EGQHVNRGDDLGMFHFSGSSFALGLRKDCRAEIVEKFTPEPGTVIRINE
 VVAALKA

SEQ. ID NO: 3
 ATGATCGCTGTACTATTCTCCTTCGTTCATTGCAGGATGCATATACTAC
 15 ATCGTTTCTCGTAGAGTGAGCGGTGCGCTTGCCACCAGGGCCGCT
 GGCATTCTTATCCCTTCATTGGGAACATGTTTGATATGCCTGAAGAA
 TCTCCATGGTTAACATTTCTACAATGGGGACGGGATTACAGTCTGTCT
 20 TGCCCGGTGACTTCTAATATATGAACAGCTAATATATTGTGAGACAC
 CGATATTCTCTACGTGGATGCTGGAGGACAGAAATGGTTATTCTTAA
 CACGTTGGAGACCATTACCGATCTATTAGAAAAGCGAGGGTCCATTTA
 25 TTCTGGCCGGTGAGCTGATGTGAGTTTTTTGCAATTGAATTTGTGGT
 CACACGTTTCCAGACTTGAGAGTACAATGGTCAACGAACCTTATGGGGT
 GGGAGTTTGACTTAGGGTTCATCACATACGGCGACAGGTGGCGCGAAG
 30 AAAGGCGCATGTTTCGCCAAGGAGTTCAGTGAGAAGGGCATCAAGCAAT
 TTCGCCATGCTCAAGTGAAGCTGCCCATCAGCTTGTTCCAACAGCTTA
 CCAAAACGCCAGACCGCTGGGCACAACATATTGCCAGTAAGTACTAC
 35 TTGAGGAAAATAGCGTACGCTTCGCTGACCCGCTCCGTACATCAAAGTC
 AGATAGCGGCAATGTCACTGGATATTGGTTATGGAATTGATCTTGACG
 AAGACGACCCCTTGGCTGGAAGCGACCCATTTGGCTAATGAAGCCCTCG
 40 CCATAGCATCAGTGCCGGGCAAAATTTGGGTGATTGTTTCCCTTCTC
 GTGAGCATCCTTCTTCTATGTAGGAAGGAAGGAGTCTAACAAGTGT
 AGTAAAATACCTTCTGCTTGGTTCCAGGTGCTGTCTTCAAGCGCAA
 AGCGAAGGTCTGGCGAGAAGCCGCCGACCATATGGTTGACATGCCTTA
 45 TGAAACTATGAGGAAATTAGCAGTTAGTCAAATGCGTTCTCCCGTAT
 TTTTCAATACTCTAACTTCAGCTCACAGCCTCAAGGATTGACTCGTC
 CGTCGTATGCTTCAGCTCGTCTGCAAGCCATGGATCTCAACGGTGACC
 50 TTGAGCATCAAGAACACGTAATCAAGAACACAGCCGACAGGTTAATG
 TCGGTAAGTCAAAAGCGTCCGTGCGCAATTCAAAATTCAGGCGCTAAA
 GTGGGTCTTCTCACCAAGGTGGAGGCGATACTGTAAGGATTCTCAAT
 55 CGTTAGAGTATAAGTGTCTAATGCAGTACATACTCCACCAACCAGAC
 TGTCTCTGCTATGCTGCGTTTCTTGGCCATGGTGAAGTACCTTGA
 GGTCCAGCGAAAGGTTCAAGCGAGCTTGATGCTCTGACCAATAACGG
 60 CCAAAATCTCTGACTATGACGAAGAAGATGACTCTTGCCATACCTCAC
 CGCATGTATCAAGGAGCTTTTCCGGTGAATCAAATCGCACCCCTCGC
 TATACCGCACAAATTAATGAAGGACGAGCTGTACCGCGGTATCTGAT
 65 TCCCAAGAACACTCTAGTCTTCGCAACACCTGGTGAGGCTGTCCATT

69

-continued

CATTCTTAGTACATCCGTTGCCCCACTAATAGCATCTTGATAACAGGG
CAGTATTAACAGATCCAGAAGTCTATCCAGATCCCTCTGTGTTCCGCC
CAGAAAGATATCTTGGTCTGACGGGAAGCCTGATAACACTGTACGCG
ACCCACGTAAAGCGGCATTGGCTATGGACGACGAAATTGGTAAGTGC
GCTTTCAGAACCCCCCTCCGTTGACTAGTGCCATGCGCGCATACAA
TATCGCTATTGATCTGATATAACTTCCCTGCGGCATTTATTTGGCAT
TCCTTTAGTCCCGGAATTCATCTAGCGCAGTCGACGGTTTGGATTGCA
GGGGCAACCCCTTTATCAGCGTTCAATATCGAGCGACCTGTGCGATCAG
AATGGGAAGCCCATTGACATACCGGCTGATTTTACTACAGGATTCTTC
AGGTAGCTAATTTCCGTCTTGTGTGCATAATACCCCTAACGACGCAC
GTTTACCTTTTGTAAAGACCCAGTGCCTTTCAGTCGAGGTTTGT
TCCTCGAACAGAGCAAGTCTCAGTCGGTATCCGGACCCCTGA

SEQ. ID NO: 4

MIAVLFSFVIAGCIYYIVSRVRRLPPGPGIPIPIGNMFDMPPEE
SPWLTFLOWGRDYNLDILYVDAGGTEMVILNTLETITDLEKRGSIYS
GRLESTMVNELMGWFDLGFITYGRWRERRMFAKEPSEKGIKQFRH
AQVKAHQVLVQQLTKTPDRWAQHHRQIAAMSLDIGYGIDLAEDDPWL
EATHLANEGLAIASVPGKFWVDSFPSLKYLPAWFPFAVFKRKAQVWRE
AADHMVMDMPYETMRKLAPOGLTRPSYASARLQAMDNLGDLEHQEHVIK
NTAAEVNVGGDDTTVSAMSAFILAMVKYPEVQRKVAELDALTNNGOI
PDYDEEDDSLPLYTACIKELFRWNQIAPLAIPHKLKDDVYRGYLIPK
NTLVFANTWAVLNDPEVYPDPVSFRPERYLGPDKPDNTVRDPRKAAF
GYGRNCPGIHLAQSTVWIAGATLLSAFNIERPVDQNGKPIDIPADFT
TGFFRHPVPFQCRFVPRTEQVSQSVSGP

SEQ. ID NO: 5

ATGGCGTTCGATCTCAAGACTGAAGACGGCCTCATCACATATCTCACT
AAACATCTTTCTTTGGACGTCGACACGAGCGAGTGAAGCGCCTTAGC
GGAGGCTTTGTCAATGTAACTGGCGCATTAAGCTCAATGCTCCTTAT
CAAGGTCATACGAGCATCATCCTGAAGCATGCTCAGCCGCACATGTCT
ACGGATGAGGATTTAAGATAGGTGTAGAAGTTCGGTTTACGAATAC
CAGGCTATCAAGCTCATGATGGCCAATCGGGAGGTTCTGGGAGGCGTG
GATGGCATAGTTTCTGTGCCAGAAGGCCTGAACTACGACTTAGAGAAT
AATGCATTGATCATGCAAGATGTCGGGAAGATGAAGACCCTTTATAGAT
TATGTCACCGCCAAACCGCCACTTGCGACGGATATAGCCCGCTTGTT
GGGACAGAAATTGGGGGTTCTGTTGCCAGACTCCATAACATAGGCCGC
GAGAGGCGAGACGATCCTGAGTTCAAATTTCTCTGGAATATTGTCT
GGAAGGACGACTTCAGACCAGTGTATCAAACCATCATACCAACGCA
GCGAAATATGGCGTCGATGACCCCTTGCTGCCTACTGTGGTTAAGGAC
CTTGTGGACGATGTGATGCACAGCGAAGAGACCCCTTGTCATGGCGGAC
CTGTGGAGTGGAAATATTCTTCTCCAGTTGGAGGAGGAAACCCATCG
AAGCTGCAGAAGATATATATCCTGGATTGGGAACCTTGCAAGTACGGC
CCAGCGTCGTTGGACCTGGGCTATTCTTGGGTGACTGCTATTTGATA

70

-continued

TCCCGCTTTCAAGACGAGCAGGTCCGTACGACGATGCGGCAAGCCTAC
TTGCAAAGCTATGCGCGTACGAGCAAGCATTCGATCAACTACGCCAAA
5 GTCCTGCGAGGTATTGCTGCTCATATTGTGATGTGGACCGACTTTATG
CAGTGGGGGAGCGAGGAAGAAAGGATAAAATTTGTGAAAAAGGGGTA
GCTGCCTTTCACGACGCCAGGGGCAACAACGACAATGGGGAAATTACG
10 TCTACCTTACTGAAGGAATCATCCACTGCGTAA

SEQ. ID NO: 6

MHIRNPYRTPIDYQALSEAFPPLKPFVSVNADGTSSVDLTIPEAQRAF
15 TAALLHRDFGLTMTIPEDRLCPTVPNRLNYLVWIEDIFNYTNKTLGLS
DDRPIKGV DIGTGASAIY PMLACARFKAWSMVGTEVERKCIDTARLNV
VANNLQDRLSILETSIDGPILVPIFEATEEYEFYETMCNPPFYDGAAD
MQTSDAAKGFPGVGVGAPHS GTVIEMSTEGGESAFVQMVRESLKLRT
20 CRWYTSNLGKLSLKEIVGLLKELEISNYAINEYVQGSTRRYAVAWSF
TDIQLPEELSRPSNPELSSLF

SEQ. ID NO: 7

ATGCATATCAGAAATCCTTACCGTACACCAATTGACTATCAAGCACTT
TCAGAGGCCTTCCCTCCCTCAAGCCATTGTGTCTGTCAATGCAGAT
GGTACCAGTTCTGTTGACCTCACTATCCCAGAAGCCAGAGGGCGTTC
30 ACGGCCGCTCTTCTTCATCGTGACTTCGGGCTCACCATGACCATACCA
GAAGACCGTCTGTGCCCCAAGTCCCCAATAGGTGAACTACGTCTCTG
TGGATTGAAGATATTTTCAACTACACGAACAAAACCCCTCGGCTGTCTG
GATGACCGTCTTATTAAAGCGGTGTATATTGGTACAGGAGCCTCCGCA
35 ATTTATCCTATGCTTGCTGTGCTCGGTTCAAGGCATGGTCTATGGTT
GGAACAGAGGTCGAGAGGAAGTGCATTGACACGCGCCCGCTCAATGTC
GTGCGGAACAATCTCCAAGACCGTCTCTCGATATTAGAGACATCCATT
GATGGTCTTATTCTCGTCCCCATTTTCGAGGCGACTGAAGAATACGAA
TACGAGTTTACTATGTGTAACCCCTCCATTCTACGACGGTGTGCGCGAT
45 ATGCAGACTTCGGATGCTGCCAAGGATTTGGATTGGCGTGGGCGCT
CCCCATTCTGGAACAGTCATCGAATGTGCGACTGAGGGAGGTGAATCG
GCTTTCGTGCTCAGATGGTCCGTGAGAGCTTGAAGCTTCGAACACGA
TGCAGATGGTACACGAGTAACTTGGGAAAGCTGAAATCCTTGAAAGAA
ATAGTGGGGCTGCTGAAAGAACTTGAGATAAGCAACTATGCCATTAAAC
GAATACGTTACGGGGTCCACACGTCGTTATGCCGTTGCGTGGTCTTTC
ACTGATATTCAACTGCCTGAGGAGCTTTCTCGTCCCTCTAACCCCGAG
55 CTCAGCTCTCTTTCTAG

SEQ. ID NO: 8

MHIRNPYRTPIDYQALSEAFPPLKPFVSVNADGTSSVDLTIPEAQRAF
60 TAALLHRDFGLTMTIPEDRLCPTVPNRLNYLVWIEDIFNYTNKTLGLS
DDRPIKGV DIGTGASAIY PMLACARFKAWSMVGTEVERKCIDTARLNV
VANNLQDRLSILETSIDGPILVPIFEATEEYEFYETMCNPPFYDGAAD
MQTSDAAKGFPGVGVGAPHS GTVIEMSTEGGESAFVQMVRESLKLRT
65 CRWYTSNLGKLSLKEIVGLLKELEISNYAINEYVQGSTRRYAVAWSF

71

-continued

TDIQLPEELSRPSNPPELSSLF

SEQ. ID NO: 9

ATGCCTTCCAGTCAACCTCACATTACTCATCGCTATCGGGTTCCTTCG
 AGTGACGACCATGAACGTATATCTGCTCTGTTCTTGGGTCCCAAAGCA
 GAAAATGCCGCATTCTCCAGCAATGGTTGACCACGGTCGTCGCACAG
 CAAAAGGCTGCCCGCATGCATACTTCCCGGATGACAATGCTTTTATT
 ACTACAGACATGCAAACTTCCCCCGCCTTGTCTCAGACTACTAAAGTA
 ATCGCTTCCAATCTCACGAATTATTGACTGCACCTCGGTGAAAGGTCTG
 ATTCCTTTCTTCTCACCTCGGTACAGCGCCATATGTCTGTGGACCAA
 AGTCTACCTGCCATTCTCGGATTCTTATCGACCACATTTTATAATCCT
 AACAATGTTCCTTCGAGGCTAGTCCATTACGACCCCTCATCGAGGAA
 GAAGTTGGCTTGCAACTCTCTGAAATGTGGGTTATAATCGGCTAAAT
 AACACCGAGAAACCTCTCGCCTGGGGACATATTGCATCAGGTGGAAC
 GTTGCAAACCTTGGAAGCGATGTGGCGGCGCGAAACCTCAAGTTTAC
 CCTCTCTCACTCCGTGATGCTTCAGCCGAAGCGCAGAGATGGAATTC
 ATTCGTGACACATTCTCCGTCAAAACCTGTGTGTGGTGACAAAAATTA
 TTAAAGGATTGCGACCCATGGGAACCTCCTCAATTTGCATGTTTCTACT
 ATCTTAGACATGCCCGACCGCTGTCACGACGAGTACAATATTTACCT
 CAGTTCCTCGAAAAGGTTATGCGAAAGTATATCATCCAGTCTACCAAC
 AAAGACACGTTGATGCAGCGTTGGGGACTTACCAACACCTGTCTGTT
 TTATCCCCGAGCACAAACCATATTCTCGGCAAGGCTGCAGCTGTG
 CTCGGTATTGGCTCAGACAACCTTCGCAACGTCCAGTAGACATCCAA
 GCCCACATGGACATAAAGCACTCGATCGTATGTTAAAAATTTGCTTG
 GACGAGGAGACGCCAGTATATCAAGTAGTTGCTGTTATCGGTACCACC
 GAAGAGGGCGGTGTGATCGCATACGGAGATCCTGAAGCTGCGCCAA
 AAGTATGAAGCTTTGGGCTGTCTTTTGCCATCCATGCAGATGCTGCT
 TGGGAGGCTATTTTGCAACCATGTACCCAAAGATACATTGGGTGCA
 AACCGGACTAGGCTTCCCAAAGAGGACACTACCTCGGGCTTTGTCCT
 CACGTGCTGTCGCGAGGAGAGCGGTTACAACCTACGCATATAAAG
 TATGCCGATTCTATTACTATCGACCCGACAAAGGAGGCTATGTTCT
 TACCCCGCTGGGGCACTCTGTTATCGCGACGGAAGAATGAGGTACCTG
 CTTACATGGTCCGCCCTACCTTGCCCAAGCAACGAGGCCAAAGT
 ATCGGAATATACGGGATCGAAGGAAGCAAACTGGTGACGACGATCC
 GCGGTATTATGGCGCACGAAACCATTTGGCCTGACTCCTTCTGGATAC
 GGGAACTTCTTGGCCAGGCAATGTTACATGTCGCCGATACGCTGCT
 CACTGGTCTGCAATGTCAACGATACTACCAGTTTCACTGTACCCCCG
 TTCAATCCTATCCCTGCTGACATCGACCCCAACGCTGACCCCGCAAAG
 GTCGAAGAGCAAAAACAGTTTATCATGAGATCGTATCTTGTTCAAATCG
 AACGAGGAAATATACAACGATTCTGAGGCTATGGAACCTTGACACAA
 CTTGGGTCCGATCTCAATATCAACGTTTTCGCATGCAACTTCCGCGAC
 CGCGATAAATCTCAACACCGACGTCGAGGAAGCCAACTGGCTCAAT

72

-continued

AACCGTATTTTCCAACGCTTTTCTGTTACAAGTGTCTGAGGAGAACCCA
 TTGGAAACGCCATTCTTCTCAGCTCAACTACATTGAAACAATCCGAA
 TACGGCGTCTGCGCAACCGAAGTAAAGAGACGCATGGGACTTGTGGT
 GACCAGGATGTTATAGTCTGAGGAACGTCGTTATGTCTCCATTACT
 ACAACGAACGACTTTGTGGGAACCTGCGCAACACCTTCCAAAAGATC
 GTTGAGGAGGAGGTGAGTATGCACGGATCCGCAACGATATGAAACCT
 AGCATTACACCTTCTCTTCTCATGGTTGAGGAGAGCAATACTATCTT
 GTCCACACCCCAACGATCCATATGGCCAGCGGCGTCCGCAAAATCATC
 CTTTCAGTAAATGTTGAAGGCCAAGTTCGGCAGGCGATACATGCCCAT
 GAAAGAGTTGAAGCAGTGATTGTACATAAACTGTGCCCTCCGCTT
 GACGAAATCGTTGACGGAGGATCATTGACGGCATACTACCATCGGA
 AAGAGGAAAACCTAGTTTCAAAGTGAAGATTTCAAACATTAAAGTAGTC
 AAGAAGCGCTCTCTGATGACTGAGGACCTGGAATCTGCGTACCCATCG
 TTGATGCCATTCTATTTCTACGGGACTCAAGGACACGCTCATCTCGAC
 CATGTCATTACTGTCTGTTCTAACATCCATCTGAGTGTGCGGAAATA
 CAGTACAAATTCGACGACGAGGTGTCAAGCGAGGACCTCGCCAAAGGGC
 CTCATTGTTGTTGCTGAGAACGTACACGAGGCATCCATGCAGCCCTTC
 CCGCTCATGAAAGATTTCAAGATCACCACCAATTTCTTCTCAGCTCC
 GGGCAAATACTCCGCGTCAAAGTGTACAGAGATCCATACCCGGCATCG
 ACAATGGATCCCATCCCTCTCCACGACATCAAGAACCAGCCCGTCGTG
 ACACAAGGCACCATCACGCTCGTCGGAATATTTACGTCGATTCTGAT
 GCGCTCAACGTCGCTTCCGAGCCTACTGCCGACGAAGACGCGGCGCAT
 GTTCCTCACGCTCGCAACATGTACGGCGAGATGACCGCTGGAACGATC
 AAAGGCTGGCAAAACGCTGTTCTGTCATTTCCACAACAAATTGGAGACT
 GTTGCTCCGACGAAGTAG

SEQ. ID NO: 10

MPSSHPIHTRYRVPSSDDHERISALFLGPKAENAAFLQWLTTVVAQ
 QKAARDAYFPDDNAFITTDMTSPAPAQTKKVIASNLTELLTALGERS
 IPFFSPRYSGHMSVDQSLPAILGLFLSTTFYNPNVAFESPFTTLIEE
 EVGLQLSEMLGYNRLNNTKPLAWGHIASGGTVANLEAMWAARNLKFY
 PLSLRDASAEGAEMEFIRDTFVSVKTCVGDKKLLKDCSPWELLNLHVST
 ILDMPDRLHDEYNISPQFLEKVMRYIIQSTNKDTLMQRWGLTQQPVV
 LSPSTNHYSWPKAAAVLIGSDNLRNVPVDIAQHMDINELDRMLKICL
 DEETPVYQVVAVIGTTEEGVDRIEILKLRQKYEALGLSFAIHADAA
 WGGYFATMLPKDTLGRNRLPKEDTTSGFVPHVGLREESALQLSHIK
 YADSIIDPHKAGVVPYAGALCYRDGRMYLLTWSAPYLAQNEGQS
 IGIYIGIEGSKPGAAASAVFMAHETIGLTPSGYGNLLGQAMFTCRRYAA
 HWSAMSTDTSFTVTPFNP IPADIDPNADPAKVEBQKFIRDRILFKS
 NEEIYNDSEAMELLHQLGSDLNINVFACNFRDRNLDVEEANWLN
 NRIFQRFSVTSAEENPLETPFFLSSTTLKQSEYGVCADEVKRRMGLVG
 DQDVIVLRNVMSPTTTNDFVGTLANTFQKIVEEVEYARIRNDMKP

73

-continued

SIHTFLHGSGEQYYLVHTPTIHMASGRRQIILSVNVEGQVRQAIHAH
 ERVEAIVHNTPVLRLEIVDGGSGFDILITIGKRKTSFKVKISNIKVV
 KKRSMLTEDLESAYPSLMPFFYFGTQGHALDHDVITVVPNIHLSAGEI
 QYKFDEVSSEDLAKGLIVVAENVHEASMOPPLMKDFKITNQFFSS
 GQILRVKVYRDPYPASTMDPIPLHDIKNQPVVTQGTITLVGNIYVDS
 ALNVASEPTADEDAHVPHARNMYGEMTAGTIKGWQNAVHRHFNKLET
 VAPTK

SEQ. ID NO: 11
 ATGGAGGGCTATCAAAAAGGTTTTGAGAACAAAAGCGGAGGGCATT
 CCTGTGTTGGTGACCTTTGTTACTGCAGGATATCCTCGTCCCGAAGAT
 ACTGTTCCCATCTTGTCTGGCCATGGAGGCCGGTGGTGTGATATCATC
 GAGCTTGGTATGCCATTTTCAGACCCAATTGCAGATGGTCTGTCTATC
 CAGGAAACGAACACAATCGCCGTGCAAACAGGTAGATTATACCACT
 GTTCTCGGACAACTTCGGGAAGCCCGCAACAAGGGCTCAAGGCACCC
 GTTCTTCTGATGGGATATTATAACCCCATATTGGCTTACGGAGAAGAC
 AGATCTATTCAAGATGCGGCTGAAGCTGGAGCCAATGGGTTTATTATG
 GTCGACCTTCCACCCGAGGAGGCTGTGCTTTTCGAGAGAAATGTATC
 AAATCCAACCTCTCATATGTTCTCTAATTGCACCTCAACGACTCTG
 TCGCGTATAAAGTTCTCTCAACAATTGCAGACACGTTCTATCTATGTC
 GTGTCTAAATGGGAACCAACCGGATCCTCAGAGAAGGTTGCCATGAAT
 AACGCCCTTCCACCATCATCGATCGTATTGCGAGTACGCTGAAGTT
 CCTTTAGCAGTCGGATTTGGAGTCGCCACTCGGGCTCACTTCAACTAC
 GTCGCCGATTCCGGTGTGATGGTGTGCTTATTGGCACAAACTCGTT
 AACGTTATTAAAGAGTCACCGCAAGGGGAAGCACCAAAATGTTGAG
 GCATACTGCCGTGAGATGAGCCAAAGGGAGAAACAAATCGCGTCAAA
 TCTCCACCAACTGCCGTGCTGCCAGCTCCGAATCAATCTCTGTTGTT
 GTTCTTCTGTTCTCCCGCACGTTTCGAGAATTCGGAGGACAATAC
 GTTCCCGAAGCTCTTGTGATGTTCTGTTGAACTAGAAGAAGCTCAC
 AAATCTGCCATGGCTGATCCTGAATTCAGAAGGAACACAAATCGCAT
 GCCGATATGCAAAATCGTCTTCACAAATATACCTCGCCGAAAATCTC
 ACCAAGGATGCTGGGGGTGCAAAATATTGGTTGAAACGTGAAGATTG
 AACCACACAGGTTCCCACAAAATCAATAACGCTTTGGGACAAATCTG
 CTTGCCCGGAGAATCGAAAGACCAAGATTATCGCAGAAACAGGTGCC
 GGCCAGCATGGTGTGCAACAGCGACTGTTTGGCTAAGTTTGAATG
 GAATGTGTTATCTACATGGGCGCAGAAGATGTGCGACGGCAAGCTCTA
 AATGTATTAGGATTGAGATGCTAGGAGCAAAAGTTGTTCTGTTACT
 TCAGGATCATGCACATTGAAGGACGCTGTAACGAGGCCCTTCCGTGAC
 TGGGTGACAAACCTTTCTACGACGATTATTGTTGGTCTGTGAATT
 GGACCTCATCCCTTCCCCACCATTTGTCGAGATTTCAAAAGGTCATT
 GGTCAAGAGATCAAGGCTCAGATGTTGGCCGCCCGCGGCAAACTTCCT
 GATGTCGTCTGCTTGTGTTGGTGGAGGAAGCAATGCTATCGGTACG

74

-continued

TTCTATGATTTTATTGGCGACAAGAGTGACGTCTAGTTGGGGTGGAA
 GCAGGAGGAGAAGGTATTGACGGAGACCGACATAGCGCCACACTTTCG
 5 ATGGGGCAACCGGGAGTACTTCACGGTGTAGAACATATATCTACAA
 GACAAGGCCGGTCAAATCATCGAGACGCACTCAATCAGCGCTGGATTG
 GATTATCCCGGCGTTGGACCAGAACATGCTTGGCTAAAGGACTCTAAA
 10 AGAGCAGAATATGTTGTGCGCCACAGACGAAGAAGCACTTCGCGGTTTC
 CGTATGCTAACACAAAGGGAGGAATTATTCTGCCCTGAATCTTCC
 CATGCGATCTGGGAGGCTGTCAGGATTGCCCGCACCATGTCGAAGGAC
 15 CAGGATCTTGTGTGTGTTGTCTGCGCGAGGTGATAAAGACGTTGAG
 CAAATTTCTCAACTTCTTCCCAAGTGGGCGGATATTCTAGACTGGCAT
 GTTTCTTCCCATGCCGTTGGACACACAACAAATCTTAA

SEQ. ID NO: 12
 MEAIKKVPENKKAEGIPVLVTFVTAGYPRPEDTVPILLAMEAGGADII
 ELGMPFSDPIADGPVIQETNTIAVANQVDYTTVLGQLREARKQGLKAP
 VLLMGYYNPILAYGEDRSIQDAAEAGANGFIMVDLPPEEAVAFREKCI
 25 KSNLSYVPLIAPSTTSLRIKFLSTIADTFIYVSKMGTGSSSEKVMAN
 NALPTIIDRIEYAEVPLAVGFGVATRAHFNYPVADSGADGVVIGTKLV
 NVIKESPGGEAPKNVEAYCREMSQKGETNRVKSPTARAASSESIPIVV
 30 VPSVLPARFGEFGGQYVPEALVDCLVELEEAKHSAMADPEFQKELQSH
 AGYANRPSQIYLAENLTKDAGGANIWLKREDLNHTGSHKINNALGQIL
 LARRIGKTRIIAETGAGQHGVATATVCAKFGMECVIYMGAEVDVRQAL
 35 NVFRIEMLGAKVVPVTSGSCTLKDAVNEAFRDWNTLSTTHYLVGSVI
 GPHFPPTIVRDFQKVIQEIKAQMLAARGKLPDVVAVCGGGSNAIGT
 FYDFIGDKSVRLVGVEAGGEGIDGRHSATLSMGQPGVLHGVRTYILQ
 40 DKAGQIIETHSISAGLDYPGVPEHAWLKDSKRAEYVATDEEALRGF
 RMLTQREGIIPALESSHAIWEAVRIARTMSKDQDLVVCLSGRGDKDVE
 QISQLLPKWADILDWHVSSHAVGHTTKF

SEQ. ID NO: 13
 ATGGAAAGGCGGAAGCGTGAGCGTCTTGGCTCACTCGGCCGACCGACC
 AAAAAGGAGCTCCGCATGATCCGATCTGTGGTGATCGTGGGTGGTGGC
 ACGGCGGGCTGGATGACCGCCTCTACCTCAAGCCGCGCTTCGACGAC
 50 CGCATCGACGTAACGCTCGTGGAGTCAGGGAACGTCAGGCGGATCGGG
 GTCGGCGAAGCGACCTTCAGCACGGTCCGCCACTTCTTCGACTACCTG
 GGCTTCGACGAGCGCGAGTGGCTGCCCGCTGCGCCGGCGGCTACAAG
 55 CTCGGCATCCGCTTCGAGAACTGGAGCGAGCCGGCGAGTACTTCTAC
 CACCCGTTTCGAGCGCCTGCGCGTCTGCGACGGCTTCAACATGGCCGAG
 TGGTGGCTCGCGGTGCGCGACCGCAGGACGCTCTTCAGCGAGGCGCTGC
 60 TATCTCACGCACCGGCTGTGCGAGGCCAAGCGGGCGCCCCGCATGCTC
 GACGGCTCGCTCTTCGCCTCCAGGTGGACGAGTCGCTCGGCCGCTCG
 ACCCTGGCCGAGCAGCGCGCCAGTTCCTGACGCTACCACTTCGAC
 65 GCCGACGAGGTGCGCCGCTACCTGTCGAGTACGCCATCGCCCGCGGC

75

-continued

GTCCGCCACGTGGTCGACGACGTGCAGCACGTTCGGCCAGGACGAGCGC
 GGCTGGATCAGCGCGCTCCACACCAAGCAGCAGCGGAGATCAGCGGC
 GACCTGTTTCGTCGACTGCACCGGCTTCCGCGGCCCTGCTCATCAACCAG
 ACGCTGGGCGGCAGGTTCCAGTCTTCTCCGACGTGCTGCCCAACAAC
 CGGGCGGTGCGCTGCGCGTCCCGCGGGAGAACGACGAGGACATGCGG
 CCGTACACGACGGCGACCGCATGAGCGCCGCTGGATGTGGACGATC
 CCGCTGTTCAAGCGCGACGGCAACGGCTACGTCTACTCCGACGAGTTC
 ATCTCGCCGAGGAGGCGAGCGCGAGCTGCGGTCCACCGTCGCCCCC
 GGCCGCGACGACCTGGAGGCCAACACATCCAGATGCGGATCGGCAGG
 AACGAGCGCACCTGGATCAACAACGCGTCCGCTCGGCCTGTCCGCC
 GCCTTCGTCGAGCCGCTGGAGTCGACCGGCATCTTCTTATCCAGCAC
 GCCATCGAGCAGCTCGTGAAGCACTTCCCCGGCGAGCGCTGGGACCCG
 GTGCTGATCAGCGCGTACAACGAGCGCATGGCGCACATGGTCGACGGG
 GTCAAGGAGTTCCTCGTCTCCACTACAAGGGCGCCAGCGCGAGGAC
 ACCCCCTACTGGAAGGCCGCCAAGACCAGGGCCATGCCCGACGGCCTC
 GCGCGCAAGCTGGAGCTGTCCGCTCCACCTGCTGGACGAGCAGACG
 ATCTACCCCTACTACCAGCGCTTCGAGACCTATTCTGGATCACCATG
 AACCTCGGTCTCGGCATCGTCCCCGAGCGCGCGTCCCGCGCTCCTG
 CACATGGACCCGGCGCCCGCTGGCCGAGTTCGAACGGCTCAGGCGC
 GAGGGCGACGAGCTGATCGCCGCCCTGCCAGCTGCTACGAGTACCTC
 GCCAGCATCCAATGA

SEQ. ID NO: 14

MERRKRERLGLRPTKKELRMIRSVVIVGGTAGWMTASYLKAAPDD
 RIDVTLVESGNVRRIGVGEATFSTVRHFFDYLGLDEREWLPRCAGGYK
 LGIRFENWSEPEGYFYHPFERLRVDGFMMAEWLAVGDRRTSFSEAC
 YLTHRLCEAKRAPRMLDGLSLFASQVDESLGRSTLAEQRAQFPYAYHFD
 ADEVARYLSEYAIARGVRHVDDVQHVQDERGWISGVHTKQHGIEISG
 DLFVDCTGFRGLLINQTLGGRFQFSFSDVLPNNRAVALRVPRENDEDMR
 PYTTATAMSAGWMWTIPLFKRDGNGYVYSEFISPEEAERELRSTVAP
 GRDDLEANHIQMRIGRNERNTWINNCVAVGLSAAFVEPLESTGIFFIQH
 AIEQLVKHFPGERWDPVLI SAYNERMAHMDGVKEFLVLHYKGAQRED
 TPYWKAATRAMPDGLARKLELSASHLLDEQTTYPYHGFETYSWITM
 NLGLGIVPERPRPALLHMDPAPALAEFERLRREGDELIAALPSCYEYL
 ASIQ

SEQ. ID NO: 15

ATGACCGCGCCTACTGAAGACCGGCTTCGCGACCGGCTTTCGATC
 ACCGTCGTGGAATCGAGCAGAATCGGGACCATCGCGCTCGGTGAGGCG
 ACCTTCAGCGACATCCAGCACTTCTCCAGTTCCTCAACCTCCGGGAA
 CAGGACTGGATGCCGCGCTGCAACGCCACCTACAAGCTGGGCATCCGC
 TTCGAGAACTGGCGCATGTGCGCCATCATTTCTACCAGCGTTCGAG
 CAGATCCGCGCGGTCTACGGCTTTCGCTGACCGACTGGTGGCTGCAC
 GACGCGCCGACCGCGCTTCGACACCGACTGCTTCGTATGCCCAAC

76

-continued

CTGTGCGAGGCCGGGCGCAGCCCCGCGCCACCTCGACGGCAGCTGGCC
 GACGAGGACTTCGTGGAGGAGGGCGCAGAGCTCGCCAACCGCACCATG
 TCCGAGCACCAGGGCAAGTCGCAGTTCCTCGTACGCCTACCCTTCGAG
 GCGGCGCTGCTGGCCAAGTTCCTCACCGGGTACGCGGTGGACCGCGGC
 GTCGAGCACGTCGTCGACGACGTCTTGACGTGCGGCTGGACCAGCGC
 GGGTGGATCGAGCACGTGGTCACCGCCGAGCAGCGGAGATCCACGGC
 GACCTGTTTCGTCGACTGCACCGGTTTCCGGGGACTGTGCTGAACAAG
 GCGCTCGGTGTGCCGTTCTGTGCTCTATCAGGACACTCTGCCCAACGAC
 AGCGCGGTCCGCCCTCAGGTCCGCTGGACATGCAGCGCGCGCGCATC
 GTGCCGAACACCACGGCCACCGCGCGGGAGGCGGCTGGATCTGGACC
 ATCCCGCTGTTTCGGTTCGGGTCGGCACCGGGTACGTCTACGCCAAGGAC
 TACCTCTCGCCCGAGGAGGCCGAGCGCACGCTGCGCGAGTTCTGTCGGC
 CCGGCGGCGCGGACGTGGAGGCCAACACATCCGCATGCGGATCGGC
 CGCAGCCAGGAGTCTTGCGGAACAACGCGTGGCCATCGGCCTGTTC
 AGCGGCTTCGTCGAGCCGCTGGAGTCGACCGGGATCTTCTTATCCAC
 CACGCCATCGAGCAGCTGGTGAAGCACTTCCCGCGCGCGGACTGGAAC
 CCGAAGTCCCGGACATGTACAACCTCCGCGGTGCGCGACGTGATGGAC
 GGCATCCGCGAGTTCCTGGTGATCCACTACCGGGCGCGCGCGGGCC
 GACAACCAGTACTGGCGCGACACCAAGACCCGCGCTGCCGAGCGGC
 CTGGCCGAGCGCATCGAGTGTGGCAGCGCAGCTGCCGACACCGAG
 ACGATCTACCCGTACTACCACGGGCTGCGGCCCTACTCTATACGTGC
 ATCTGTATGGGCGCGGGGCCATCCGAGCGCCGCTCGGCCGCGTTG
 GCCCTGACCGACCGGGGGCGCGCAGAAGGAGTTCGCCCGCGTCCGG
 GACCGGGCGCGCAGCTGCGCGACACGCTGCCAGCCACTACGAGTAC
 CTCGCGCGGATGCGTGGTCTGGATGTCTGA

SEQ. ID NO: 16

MTAAYLKTAFGDRLSITVVESSRIGTIGVGEATFSDIQHFFQFLNRE
 QDWMPACNATYKLGIRFENWRHVGHFYPPEQIRPVYGFPLTDWWLH
 DAPTRDFTDCFVMPNLCAGRSRPHLDGTLADEDFVEEGDELANRTM
 SEHQKGSQFPYAYHFEAALLAKFLTGYAVDRGVEHVDDVLDVRLDQR
 GWIEHVVTAEHGEIHGDLFVDCTGFRGLLLNKALGVFPVSYQDTLPND
 SAVALQVPLDMQRRGIVPNTTATAREAGWIWTIPLFGRVGTGYVYAKD
 YLSPEEAERTLREFVGPAAADVEANHIRMIRIGRSQESWRNNCVAIGLS
 SGFVEPLESTGIFFIHHAIEQLVKHFPADWNPKSRDMYNSAVAHVMD
 GIREFLVIHYRGAARADNQYWRDTKTRPLQKEFAAVRDRAAQLRDLTP
 SHYEYLARMRLDV

SEQ. ID NO: 17

ATGAACACGAGAAATCCGACAAGGTCGTATCGTCGGCGCGGCACG
 GCGGGATGGATGACCGCTCGTACCTGAAGAAGGCGTTCCGGCAGCGC
 GTGTCCGTGACCCCTCGTGAATCGGGCACCATCGGCACCGTCGGCGTC
 GCGGAGGCGACCTTCAGCGACATCCGGCACTTCTTCGAATTCTCGAC

77

-continued

CTGCGGGAGGAGGAATGGATGCGCGCCTGCAACGCCACGTACAAGCTG
 GCCGTCCGCTTCCAGGACTGGCAGCGCCCCGGCCACCCTTCTACCAC
 CCGTTCGAGCAGATGCGCTCCGTAGACGGGTTCCCCTGACCGACTGG
 TGGCTGCAGAACGGGCCGACCGACCGGTTTCGACAGGGACTGCTTCGTG
 ATGGCGTCGCTGTGCGACGCGGGGCGCAGCCCGCTATCTCAACGGT
 TCCCTGCTGCAGCAGGAGTTGACGAGCGCGGAGGAGCCGGCGGGG
 CTCACCATGTCCGAGCACCAGGGCAAGACGCAGTTCCCCACGCCCTAC
 CACTTCGAGGCGGCCCTGCTCGCCGAATTCCTTTCTGGGTACTCCAAG
 GACCGCGCGCTGAAGCAGTGGTTCGACGAGGTGCTGGAGGTGAAGCTC
 GACGACCGCGGATGGATCAGCCACGTTGTACCAAGGAGCAGGCGAC
 ATCGGCGCGCAGCTGTTGCTGACTGCACCGGTTTCCGCGCGTACTG
 CTGAACCAAGCGCTGGGTGTGCCCTTCGTCTCTTACCAGGACACCCTG
 CCCAATGACAGCGCGGTGCGCCTCCAGGTGCCGCTCGACATGGAGGCG
 CGGGGCATCCCGCGTACACCGGGCCACCGCCAAGGAGGCCGGGTGG
 ATCTGGACGATCCCGCTGATCGGCCGGATCGGCACCGGTTACGTGTAC
 GCCAAGGACTACTGCTCGCCGAGGAGGCCGAGCGCACCTGCGCGAG
 TTCGTGCTCGGAGGCGCGGACGCTCGAGGCCAACACATCCGCATG
 CGGATCGGGCGCAGCGAGCAGTCTGGAAGAACAACCTGCGTCGCGATC
 GGGCTCTCCAGTGGCTTCGTGCAACCCCTCGAGTCGACCGGCATCTTC
 TTCATCCACCACGCCATCGAGCAGTTGGTGAAGCACTTCCCGCGGGC
 GACTGGCACCCGAGCTGCGCGCCGGCTACAACCTCGGCCGTCGCAAC
 GTCATGGACGGGTCCGGGAGTTCCTCGTCTGCACTACCTGGGCGCC
 GCGCGCAACGACACCCGCTACTGGAAGGACACCAAGACCCGGGCGGTT
 CCGGATGCCCTGGCGGAGCGCATCGAGCGCTGGAAGGTGACGCTGCC
 GACTCCGAGAACGTCTCCCTACTACCACGGCTGCCGCCGTAATCG
 TACATGGCCATCTGCTCGGACCGGCGGATCGGGCTGCGACCGTCTG
 CCGGCGCTCGCACTGGCCGACCCGGCGCGCGGAGAAGGAGTTACCC
 GCGATCCGGGACAGGGCGCGGTTCTTGGTGCACCCCTGCCGAGCCAG
 TACGAGTACTTCCGCGCATGGGGCAGCGTGTCTGA

SEQ. ID NO: 18

MNTRNPDKVVIVGGGTAGWMTASYLKAFGERVSVTLVESGTIGTVGV
 GEATFSDIRHFFELDLREEWMPACNATYKLAVRFQDWQRPGHHFYH
 PFEQMRSVDGFPLTDWWLQNGPTDRFDRDCFVMASLCDAGRSPLYLNG
 SLLQQEFDERAEAPGLTMSHQGTQFPYAYHFEAALLAEFLSGYSK
 DRGVKHVVDEVLEVLKDDRGWISHVVTKEHGDIGDGLFVDCTGFRGVL
 LNQALGVPFVSQDTLPNDSAVALQVPLDMEARGIPPYTRATAKEAGW
 IWTIPLIGRIGTYVYAKDYCSPEEAERTLREFVGPEADVEANHIRM
 RIGRSEQSWKNNCVAIGLSSGFVEPLESTGIFFIHHAIEQLVKHPPAG
 DWHPQLRAGYNSAVANVMDGVREFVLVHLYLGAARNDRYWKDTKTRAV
 PDALAERIERWKVQLPDSENVFPYHGLPPYSYMAILLGTGAIGLRPS
 PALALADPAAAEKFTAIRDRARFLVDTLPSQYEFYFAAMGQRV

78

-continued

SEQ. ID NO: 19

ATGGACAATCGAATCAAGACAGTCGTGATCTGGGCGGCGGCACCGCC
 GGCTGGATGACTGCGGCGTATCTCGGCAAGGCGCTGCAGAACACGGTG
 5 AAGATCGTGGTCTCTGGAGGCGCCACCATCCCGCGGATCGGTGTGGGC
 GAAGCCACCGTCCCCAACCTGCAGCGCGCTTCTTCGACTACCTCGGC
 ATCCCCGAGGAGGAGTGGATGCGCGAGTGCAACGCCAGTTACAAGATG
 10 GCGGTGAAGTTCATCAACTGGCGCACACCGGGGAGGGCAGCCCGGAT
 CCCCCGACGCTCGACGACGGTCACACCACACCTTCCACCACCCCTTC
 GGACTGCTGCCGTCCGCCGACAGATACCGCTCTCCCACTACTGGGCG
 15 GCCAAGCGGCTCCAGGGCGAGACCAGAGAACTTCGACGAGGCGTGT
 TTCGCCGACACCGCGATCATGAACGCCAAGAAGGCGCCCGCTTCTCTC
 GACATGCGGCGGGGACCAACTACGCCTGGCACTTCGACGCCAGCAAG
 20 GTCGCCGCGTTCCTGCGCAACTTCGCGCTCACCAGCAGGCGGTGAG
 CACGTCGAGGACGAGATGACCGAGGTCTCACCAGCAGAACGGGGCTTC
 ATCACCGCCCTGCGCACCAAGTCGGGACGGATCTGCAGGGTGACCTC
 25 TTCGTGCTGCTGCTCCGGCTTCGCGGGCTGCTGATCAACAAGGCCATG
 GAGGAGCCGTTTCATCGACATGAGCGACCACTCTGTGCAACAGCGCG
 GTCGCCACGGCCGTACCGCACGACGACGAGAAGAACGGTGTGAGCCG
 30 TACACCTCTCTGATCGCCATGGAGGCCGGCTGGACCTGGAAGATCCC
 ATGCTGGGCGGTTTCGGCAGCGGTACGCTCTACTCCGACCACTTCGCC
 ACCCAGGACGAGGCCACCTGGCCTTCTCGAAGCTGTGGGGCTCGAC
 35 CCGGACAACACCGAGTTCAACCACGTCGCTTCCGGGTGCGCCGAAC
 CGCAGGGCTGGGTGCGCAACTGCGTCAGCGTCGCGCTCGCGTCTGTC
 TTCGTGGAGCCGCTGGAGTCCAGCGCATCTACTTCTATCTACGCGGCC
 40 ATCCACATGCTGGCCAAGCACTTCCCCGACAAGACCTTCGACAAGGTG
 CTCGTGACCCGTTCAACCGGGAGATCGAGGAGATGTTTCGACGACACC
 CGGGACTTCTCCAGGCGCACTACTTCTCGCCGCGCGTCGACACG
 CCGTTCCTGGCGGCGAACAAGGAAGTGAAGCTGGCCGACTCCATCAAG
 45 GACAAGGTGAGACGTACCGGGCCGGGTGCGGTCACCTGCCGGTC
 ACCGACGAGGGGACCTACTACGGCAACTTCGAGGCCGAGTTCGGGAAC
 TTCTGGACCAACGGCAGCTACTACTGCATCTTCGCGCGCTCGGCTG
 50 ATGCCGCGCAACCGCTGCCGCGCTCGCCTACAAGCCGAGTCGATC
 GCGGAGGCCGAGCTGCTGTTCCGCCGACGTCAAGCGCAAGGGGGACACC
 CTGGTCGAGTCGCTGCCGTCGACCTATGACCTGCTGCGTCAGTTGAC
 55 GGTGCGCTGTGA

SEQ. ID NO: 20

MDNRKTVVILGGGTAGWMTAAYLGKALQNTVKIVVLEAPTIPRIGVG
 EATVPNLQRAFFDYLGIPPEEWMRECNASYKMAVKFINWRTPEGSPD
 60 PRTLDDGHDTDFHHFPGLLPADQIPLSHYWAAKRLQGETDENFDEAC
 FADTAIMNAKKAPRFLDMRRATNYAWHFDASKVAFLRNFAVTKQAVE
 HVEDEMTEVLTDERGFI TALRTKSGRI LQGDLFVDCSGFRGLLINKAM
 65 BEPFDMSDHLLCNSAVATAVPHDDEKNGVEPYTSSIAMEAGWTWKIP

79

-continued

MLGRFGSGHVYSDFHATQDEATLAFSKLWGLDPDNTFHNHVRFRVGRN
 RRAWVRNCVSVGLASCFVEPLESSGIYFIYAAIHMLAKHFPDKTFDKV
 LVDRFNREIEEMFDDTRDFLQAHYFSPRVDTPFWRANKELKLADSIK
 DKVETYRAGLPVNLPTDEGTYYGNFEAEFRNFWTNGSYCYIFAGLGL
 MPRNPLPALAYKQSIABAEELLFADVVRKRGDTLVESLPSTYDLLRQLH
 GAS

SEQ. ID NO: 21

ATGGACAACCGCATCAACCGGATCGTCATCCTCGCGCGGGACCGCG
 GGATGGATGACCGCTCTACTGCGCAAGCGCTCGGTGACACCGTC
 ACGATCACCTGCTGGAAGCACCGCGCATCGGCCGATCGGGTCGGC
 GAGGCCACCGTGCCCACTGCAGCGCTGTCTTCTGACTTCTCTCGC
 CTCCGCGAGGAGGAGTGGATGCGCGAGTGCAACGCGCGTTCAAGACG
 GCCGTGAAGTTTCATCACTGGCGTACCCCGGGCCCGCGAGGCCAAG
 GCGCGACGATCGACGCGCGCGGACCACTTCTACACCGCTTCGGG
 CTGCTGCCCCGAGCAGCGCCAGGTGCGCGTGTGCGACTACTGGCCTAC
 AACCAGGCGCGCCGCGCACCGGACGAGCGGTTCGACTACGCTGTTC
 GCGGAGACCGCCCATGGACGCGGTCCGCGCGCCCAAGTGGCTCGAC
 GGGCGGCCCGCCACCGGTATGCTGGCACTTCGACGCGCCACCTGGTC
 GCCGAGTTCCTGCGCGCGCACGCCACCGAGCGGCTCAACGTGGAGCAC
 GTGACGGGTGAGATGCAGCAGGTGCTGCGGACGAGCGGGCTTCATC
 ACCGCCCTGCGCACGTCGAGGGCCGGGACCTCGAGGGTGACCTGTT
 ATCGACTGCTCCGGCTTCGCGGCTGTGATCAACAAGGCCATGGAG
 GAGCCCTTCATCGACATGAACGACAGCTGTGTGCAACCGCGCGGTG
 GCCACGGCGATCAAGCAGCAGACGACGCGCAGGGGTGAGCCGTAC
 ACCTCGGCGATCGCCATGCGGTCCGGCTGGAGCTGGAAGATCCCGATG
 CTGGCGCGCTTCGGGACCGGTACGTCTACTCCAGCCGTTTCGCCGAG
 AAGGACGAGGCCACCTCGACTTCTGCGGATGTGGGCCTGGACCG
 GAGAACACCCCGCTCAACAGGTGGCTTCGGGTTCGGCGCAACCGC
 CGGGCTGGGTGAAGAACTGCGTGAGCATCGGCTGGCGTGTGCTTC
 CTGGAGCCACTGGAGTCCACCGGCATCTACTTCATCACCGCGGCATC
 TACCAGCTGACTCAGCACTTCCCGACCGCACCTTCGCGCTCGCGTTG
 AGCGACGCGTTCAACACGAGATCGAAGCCATGTTCGACGACACCGG
 GACTTCATCCAGGCGCACTTCTACGTGTGCGCGCACCGACACGCGG
 TTCTGGAAGGCGAACAAGGACCTGCACCTGCCGAGCAGATGCGCGAG
 AAGATCGCGATGTACAAGGCCGCTGCGGATCAACGCGCGGTACAC
 GACGAGTCGACCTACTACGGCAGGTTCGAGGCGGAGTTCCGCAACTTC
 TGGACCAACGGCAGCTACTACTGCATCTTCGCGGGCTCGGGCTGCGC
 CCGGACAACCGCTGCCGATGTGCGACACCGTCCGGAGCAGGTCCGC
 GAGGCGCAGGCGCTGTTGCGCGGGTGAAGGACAAGCAGCGGAGTTG
 GTGGAGACGCTGCCGAGCAACCTGGAGTTCCTGCGCAGCTGCACGGC
 AAGTAG

80

-continued

SEQ. ID NO: 22

MDNRINRIVILGGGTAGWMTASYLAKALGDTVTTITLLEAPAIGRIGVG
 5 EATVPNLQRVFFDFLGLREEWMPECNAAFKTAVKFINWRTPGPGEAK
 ARTIDGRPDHFYHPFGLLEHGQVPLSHYWAYNRAAGTTDEPFYACF
 AETAAMDAVRAPKWL DGRPATRYAWHFD AHLVAEFLRRHATERLNVEH
 10 VQGEQQVLRDERGFITALRTVEGRDLEGLDFIDCSGFRGLLINKAME
 EPFIDMNDQLLCNRAVATAIKHDDDAHGVEPYTSAIAMRSGWSWKIPM
 LGRFGTGYVSSRFAEKDEATLDFCRMWGLDPENTPLNQVAFVRGRNR
 15 RAWKNCVSI GLASCFLEPLESTGIYFITAIIYQLTQHFPDRTPALAL
 SDAFNHEIEAMFDDTRDFIQAHFYVSPRTDTPFWKANKDLHLPEQMRE
 KIAMYKAGLPINAPVTDESTYYGRFEAEFRNFWINGSYCYIFAGLGLR
 PDNPLPMLRHRPEQVREQAQALFAGVKDKQRELVELTPSNLEFLRLSHG
 20 K

SEQ. ID NO: 23

ATGGATGACAATCGAATTCGGAGCATCCTTGTCTTGGCGCGCGCACG
 25 GCCGGCTGGATGTCGCGCTGTACTGAGCAAGCGCTCGGGCCCGGC
 GTCGAGGTACCGTGCTCGAGGCGCCCTCCATCTCGCGCATCCGGGT
 GCGGAGGCCACCATTCCCAACCTGCACAAGGTCTTCTTCGACTTCCTG
 30 GGCATCGCGGAGGACGAGTGGATGCGGGAGTGAACGCCAGCTACAAG
 GCCGCGGTCCGGTTCGTCACTGGCGGACGCGCGGCGCAGGCCAGGCC
 ACGCCGCGGCGGCGTCCGGACGCGCGCCCGGACCACTTCGACCACCTG
 35 TTCGGCCAGCTGCGGAGCACGAGAACCTGCGCTGTGCGAGTACTGG
 GCGCACCGGCGCTCAACGGCTGACCGACGAACCTTCGACCGCTCC
 TGCTACGTGACGCGGAGCTGCTGGACCGCAAGCTCTCGCCGAGGTTG
 40 ATGGACGGCACGAACTGGCCAGTACGCTTGGCACTTCGACGCGGAC
 CTGGTGGCGGACTTCTCTGCGGTTTCGCGTGCAGAAGCTGAACGTG
 ACCCACGTCCAGGACGTGTTACGCGATGCCGACCTCGACCAGCGCGGC
 45 CACATCACGGCGCTCAACACGAGTCCGGCGCACGCTGGCGCGCGAC
 CTGTTTCATCGACTGCAGCGGCTTCGCGAGCGTGCTCATGGGCAAGGT
 ATGCAGGAGCCGTTCTGACATGAGCAAGCACCTGCTCAACGACCGC
 GCGGTGGCGCTGATGCTCCCGCACGACGACGAGAAGGTGGGCATCGAG
 50 CCGTACACCTCGTGTGCGCATGCGGTGGGCTGGTGTGGAAGATC
 CCGCTGCTCGGCCGTTTCGGCTCCGGCTACGTCTACTCCAGCCAGTTT
 ACCTCCCAGGACGAGGCGGCGGAGGAGCTCTGCCGATGTGGGACGTC
 55 GACCCGCGGAGCAGACGTTCACCAACGTCCGGTTCGGGTTCGGCGCG
 AGCCGCGGGCTGGGTGCGCAACTGCGTGCACATCGGCGTGTCCGCC
 ATGTTCTGTGAGCGCGCTGGAGTCGACCGGCTGTACTTCAGTTACGCC
 60 TCGCTCTACAGCTGGTGAAGCACTTCGCGACAAGCGGTTCCGGCGG
 ATCCTGGCCGACCGGTTCAACCGGAGGTGGCGACCATGTACGACGAC
 ACCCGCGACTTCTCTCAGGCGCACTTCAGCTGTGCGCGGTGACGAC
 65 TCCGAGTTCTGGCGGCGCTGCAAGGAGCTGCCGTTTCGCGGACGGGTT

81

-continued

GCCGAGAAGGTCGAGATGTACAGGGCCGGCTGCCGGTCGAAC TGCCG
 GTCACCATCGACGACGGGCACTACTACGGCAATTTGAGGCCGAGTTC
 CGCAACTTCTGGACCAACTCGAACTACTACTGCATCTTCGCCGGGCTC
 GGTTCCTGCCCCGAGCATCCGCTGCCGGTGCTCGAATTCGCCCGGAG
 GCCGTGATCGCGCGGAGCCGGTGTCGCCCGGTGCGCCGGCGCACG
 GAGGAGCTGGTCGCCACCGCCCGACCATGCAGGCCTACCTGCGGCGC
 CTGCACCAGGGCACGTAG

SEQ. ID NO: 24

MDDNRIRSI LVLGGGTAGWMSACYLSKALGPGVEVTVLEAPSI SRIRV
 GEATIPNLHKVFFDFLGLIAEDWMECNASYKAAVRFVNWRTPGDQA
 TPRRRPDGRPDHDFHLFGLPEHENLPLSQYWAHRRLLNGLTDEPFDRS
 CYVQPELLDRKLSPLRMDGTKLASAWHFDADLVADFLCRFAVQKLN
 THVQDVFTHADLDQRGHITAVNTESGRTLAADLFIDCSGFRSVLMGKV
 MQEPFLDMSKHLLNDRVALMLPHDDEKVGIEPYTSSLAMRSGWSKI
 PLLGRFGSGYVYSSQFTSQDEAAEELCRMWDVDPAEQTFNNVRFVRGR
 SRRWVRNCVAIGVSAMFVEPLESTGLYFSYASLYQLVKHFPDKRFRP
 ILADRFNREVATMYDDTRDFLQAHFSLSPRDDSEFWRACKELPFADGF
 AEKVMYRAGLPVELPVITIDGHYYGNFEAEFRNFWTNSNYICIFAGL
 GFLPEHPLPVLEFRPEAVDRAEPVFAAVRRRTEELVATAPTQAYLRR
 LHQGT

SEQ. ID NO: 25

ATGAACAAGCCGATCAAGAATATCGTCATCGTGGCGCGGTACTGCG
 GGCTGGATGGCCGCTCGTACCTCGTCCGGGCCCTCCAACAGCAGGCG
 AACATTACGCTCATCGAATCTGCGGCGATCCCTCGGATCGCGTGGC
 GAAGCGACCATCCCAAGTTTGAGAAGGTGTTCTTCGATTCTCTCGG
 ATACCGGAGCGGGAATGGATGCCCCAAGTGAAAGCGCGTTCAGGCC
 GCGATCAAGTTCGTGAATTGGAGAAAGTCTCCGACCCCTCGCGCGAC
 GATCACTTCTACCATTTGTTCGGCAACGTGCCGAACGCGACGCGTG
 CCGCTTACCCACTACTGGCTGCGCAAGCGCGAACAGGGCTTCCAGCAG
 CCGATGGAGTACGCGTGCTACCCGAGCCCGGGGCACTCGACGGCAAG
 CTGGCACCGTGCTGTCCGACGGCACCCGCCAGATGTCCACGCGTGG
 CACTTCGACGCGCACCTGGTGGCCGACTTCTTGAAGCGCTGGGCCGTC
 GAGCGCGGGTGAAACCGGTGGTCGATGAGGTGGTGACGTTTCGCTG
 AACAAACCGCGCTACATCTCCAACCTGCTCACAAGGAGGGCGGACG
 CTGGAGGCGGACCTGTTTCATCGACTGCTCCGGCATGCGGGGGCTCCTG
 ATCAATCAGGCGCTGAAGGAACCTTCATCGACATGTCCGACTACCTG
 CTGTGCGACAGCGCGTGCACAGCGCGTGCACCAACGACGACGCGCGC
 GATGGGGTCGAGCCGTACACCTCTCGATCGCCATGAACTCGGGATGG
 ACCTGGAAGATTCCGATGCTGGGCCGGTTCGGCAGCGCTACGCTTTC
 TCGAGCCATTTACCTCGCGCGACAGGCCACCGCGACTTCTCTCAA
 CTCTGGGGCTCTCGGACAATCAGCCGCTCAACCAGATCAAGTTCGG
 GTCGGGCGCAACAAGCGGGCGTGGGTCAACAACCTGCGTCTCGATCGG

82

-continued

CTGTCTGCTGCTTTCTGGAGCCCCCTGGAATCGACGGGGATCTACTTC
 ATCTACGCGGCGCTTTACCAGCTCGTGAAGCACTTCCCCGACACCTCG
 5 TTCGACCCGCGGCTGAGCGACGCTTTCAACGCCGAGATCGTCCACATG
 TTCGACGACTGCCGGGATTTCGTCCAAGCGCACTATTTCACCACGTCG
 CGCGATGACACGCGCTTCTGGCTCGCGAACCGGCACGACCTGCGGCTC
 10 TCGGACGCCATCAAAGAGAAGGTTACGCGCTACAAGCGGGGCTGCCG
 CTGACCACCACTGCTTCGACGATTCCACGTACTACGAGACCTTCGAC
 TACGAATTCAAGAATTTCTGGTTGAACGGCAACTACTACTGCATCTTT
 15 GCCGGCTTGGGCATGCTGCCGACCGGTGCTGCCGCTGTTGACGAC
 CGACCGGAGTCGATCGAGAAAGCCGAGGCGATGTCGCGCAGCATCCG
 CGCGAGGCGGAGCGTCTGCGCACCAGCTGCCGACAACTACGACTAC
 20 CTGCGGTGCTGCTGACGCGCAGCGGGGCTGTGCGCGGCCAGCGT
 GGGCCGAAGCTCGCAGCGCAGGAAAGCCTGTAG

SEQ. ID NO: 26

MNKPINKNIVIVGGGTAGWMAASYLVRLAQQAANTLIESAAIPRIGV
 25 EATIPSLQKVFFDFLGIPEWMPQVNGAFKAAIKFVNWRKSPDP
 DHFYHLFGNVNCDGVPLTHYWLKREQGFQQPMEYACYPQFALD
 LAPCLSDGTRQMSHAWHFDHLVADFLKRWAVRGVNRVDEVDVRL
 30 NNRYISNLLTKEGRTLEADLFIDCSGMRGLLINQALKEPFIDMSDYL
 LCDSAVASAVPNDDARDGVEPYTSSIAMNSGWTWKIPMLGRFGSGYV
 SSHFTSRDQATADFLKLWGLSDNQPLNQIKFRVGRNKRWVNNVCSIG
 35 LSSCFLEPLESTGIYFIYAALYQLVKHFPDTSFDPRLSDAFNAEIVHM
 FDDCRDFVQAHYFTTSRDDTPFWLANRHLRLSDAIKEKVQRYKAGLP
 LTTTSFDDSTYYETFDYEFKNFWLNGNYICIFAGLMPLDRSLPLLQ
 40 RPESIEKAEAMFASIRREAERLRTSLPTNYDYLRLSDGDAGLSRGQR
 GPKLAAQESL

SEQ. ID NO: 27

ATGTCCGGCAAGATTGACAAGATCCTCATCGTCGCGCGCGCACCGCC
 45 GGATGGATGGCCGCTCTATCTCGGCAAGGCCCTGCAGGGCACCGCG
 GACATCACACTGCTGCAGGCACCCGACATCCGACGCTCGGGGTCGGC
 GAGGCCACGATCCCCAATCTGCAGACGGCGTCTCTCGACTTCTCTCGA
 50 ATCCCCGAGGACGAGTGATGCGGGAGTGCAACGCGAGCTACAAGT
 GCCATCAAGTTCATCAACTGGCGCACCGCGGGCGAGGGGACGTCCGAG
 GCGCGGAGCTCGACGGAGGGCCGACCACTTCTACCACTCTCTCGGT
 55 CTGCTCAAGTACCACGAGCAGATTCCGCTGTCGCACTACTGGTTCGAC
 CGTTCTGACCGGGGAAGACCGTCGAGCCGTTGACTACGCCTGCTAC
 AAGGAACCGCTCATCTCGACGCCAACAGGTACCGCGCAGGCTCGAC
 60 GGTTCGAAGGTACGAACACGCGTGGCACTTCGACGCGCACCTCGTC
 GCCGACTTCTGCGCCGGTTCGCCACCGAGAAGCTCGGCGTGCGCCAC
 GTCGAGGACCGCGTTCGACACGTCAGCGCGACGCCAACCGCAACATC
 65 GAGTCGGTTCGACCGCAACGGGGCGTGTCTTCGATGCCGACCTCTTC

83

-continued

GTCGACTGCTCGGGCTTCCGCGGGCTGCTGATCAACAAGGCGATGGAG
 GAGCCCTTCTCGACATGAGCGATCACCTGCTCAACGACAGCGCCGTC
 GCCACCCAGGTGCCGACGACGACGACGCGAACGGTGTGGAACCGTTC
 ACCTCGGCGATCGCCATGAAGTCGGGCTGGACGTGGAAGATCCCGATG
 CTCGGCAGGTTCCGCACCGGTACTCTACTCGAGCCGGTTCGCCACC
 GAGGACGAGGCGGTGCGCGAGTTCTGCGAGATGTGGCACCTCGACCCG
 GAGACCCAGCCCCCTCAACAGGATCCGGTTCGGGTCGGCCGCAACCGG
 CGCGCGTGGTTCGGCAACTGCGTCAGCATCGGCACGTCGTCGTGCTTC
 GTGGAACCACTGGAGTCGACGGGCATCTACTCTGCTACGCCGCGCTG
 TACCAGCTGGTGAAGCACTTCCCCGACAAGAGCCTCAACCCCGTGCTG
 ACCGCCAGGTTCAACCGCGAGATCGAGACGATGTTGACGACACGCGC
 GACTTCATCCAGGCGCACTTCTACTCTCGCCGCGCACGGACACCCCG
 TTCTGGAGGGCCAAACAAGGAGCTGCGCCTGGCGGACGGCATGCAGGAG
 AAGATCGACATGTACCGCGCGGCATGGCGATCAACGCGCCCGCTCC
 GACGACGCGCCAGCTCTACTACGGCAACTTCGAGGAGGAGTTCCGCAAC
 TTCTGGAACAACAGCACTACTACTGCGTGTGCGCCGCGCTCGGTCTG
 GTGCCGACGACCCCTACACAGCCIGCGCACATGCCACAGCGGACG
 GAGTCGGTGGACGAGGTCTTCGGCGCCGTCAGGACCGGACGCGAAC
 CTGCTCGAGACCCCTGCCGAGCCTCCACGAGTTCCTGAGGCAACAGCAC
 GGCCGCTGA

SEQ. ID NO: 28

MSGKIDKILIVGGGTAGWMAASYLGKALQGTADITLLQAPDIPTLGVG
 EATIPNLQTAFFDFLGIPEDEWMREC NASYKVAIKFINWRTAGEGTSE
 ARELDGGPDH FYHSFGLLKYHEQIPLSHYWFDRSYRGKTVEPFYACY
 KEPVILDANRSPRRLDGSKVTNYAWHFD AHLVAKAMEEPFLDMSDHL
 NDSAVATQVPHDDDANGVEPFTSAIAMKSGWTWKIPMLGRFGTGYVYS
 SRFATEDEAVREFCEMWHLDPETQPLNRIRFRVGNRRRAWVGNVCSIG
 TSSCFVEPLESTGIYFVYAALYQLVKHFPDKSLNPVLTARFNREIETM
 FDDTRDFIQAHFYFSPRTDTPFWRANKELRLADGMQEKIDMYRAGMAI
 NAPASDDAQLYYGNFEEEFNRFWNNSNYCVLAGLGLVPDAPSPRLAH
 MPQATESVDEVFGAVKDRQRNLETLPSLHEFLRQQHGR

SEQ. ID NO: 29

MLNNVIVGGGTAGWMTASYFKAAGGERINITLVESGNIGAVGVGEAT
 FSDIRHFFDLGLKEKDWPACNATYKLAVRFENWREEGHYFYHPFEQ
 MRSVNGFPLADWWLREKPTDRFDKDCFVMPISIDAGLSPRHLDGSLID
 QPFVEGAEMQGLTMSHQGKTQFPYAYQFEAALLAKYLTTSYVERGV
 KHIVDDVQVNLDERGWIRNVTTAEHGEIGGDLFIDCTGFRGLLNKA
 LEEPFISFQDTLPNDSAVALQVPMDEMERRGILPCTTATAQDAGWIWTI
 PLMGVGTGYVYAKDYISPEEAERTLREFVGPAADAEEANHIKMRVGR
 SENSWVKNCVAIGLSSGFVEPLESTGIFFIHHAIEQLAKNFPGEDWNP
 VQRNLYNDSVAHVMGDVREFVLVHYVAAKRSDTQYWRDTKTRKIPDAL

84

-continued

AERIEKWKTLQLPDTEVFPYHGLPPYSYHCVLLGMGGIDVKPSPALA
 LGDSAAAEREFAEIREKTRHLTSVLPKAYDYFSRMR
 5 MSGKIDKILIVGGGTAGWMAASYLGKALQGTADITLLQAPDIPTLGVG
 SEQ. ID NO: 30
 EATIPNLQTAFFDFLGIPEDEWMREC NASYKVAIKFINWRTAGEGTSE
 ARELDGGPDH FYHSFGLLKYHEQIPLSHYWFDRSYRGKTVEPFYACY
 10 KEPVILDANRSPRRLDGSKVTNYAWHFD AHLVAKAMEEPFLDMSDHL
 NDSAVATQVPHDDDANGVEPFTSAIAMKSGWTWKIPMLGRFGTGYVYS
 SRFATEDEAVREFCEMWHLDPETQPLNRIRFRVGNRRRAWVGNVCSIG
 15 TSSCFVEPLESTGIYFVYAALYQLVKHFPDKSLNPVLTARFNREIETM
 FDDTRDFIQAHFYFSPRTDTPFWRANKELRLADGMQEKIDMYRAGMAI
 NAPASDDAQLYYGNFEEEFNRFWNNSNYCVLAGLGLVPDAPSPRLAH
 20 MPQATESVDEVFGAVKDRQRNLETLPSLHEFLRQQHGR

SEQ. ID NO: 31

MSGKIDKILIVGGGTAGWMAASYLGKALQGTADITLLQAPDIPTLGVG
 25 EATIPNLQTAFFDFLGIPEDEWMREC NASYKVAIKFINWRTAGEGTSE
 ARELDGGPDH FYHSFGLLKYHEQIPLSHYWFDRSYRGKTVEPFYACY
 KEPVILDANRSPRRLDGSKVTNYAWHFD AHLVADFLRRFATEKLGVRH
 30 VEDRVEHVQRDANGNIESVRTATGRVFDADLFVDCSGFRGLLINKAME
 EPFLDMSDHLNDSAVATQVPHDDDANGVEPFTSAIAMKSGWTWKIPM
 LGRFGTGYVYSSRFATEDEAVREFCEMWHLDPETQPLNRIRFRVGNR
 35 RAWVGNVCSIGTSSCFVEPLESTGIYFVYAALYQLVKHFPDKSLNPVL
 TARFNREIETMFDDTRDFIQAHFYFSPRTDTPFWRANKELRLADGMQ
 KIDMYRAGMAINAPASDDAQLYYGNFEEEFNRFWNNSNYCVLAGLGL
 40 VPDAPSPRLAHMPQATESVDEVFGAVKDRQRNLETLPSLHEFLRQQH
 GR

SEQ. ID NO: 32

MSGKIDKILIVGGGTAGWMAASYLGKALQGTADITLLQAPDIPTLGVG
 45 EATMPNLQTAFFDFLGIPEDEWMREC NASYKVAIKFINWRTAGEGTSE
 ARELDGGPDH FYHFPGLLKYHEQIPLSHYWFDRSYRGKTVEPFYACY
 KEPVILDANRSPRRLDGSKVTSYAWHFD AHLVAKAMEEPFLDMSDHL
 50 NDSAVATQVPHDDDANGVEPFTSAIAMKSGWTWKIPMLGRFGTGYVYS
 SRFATEDEAVREFCEMWHLDPETQPLNRIRFRVGNRRRAWVGNVCSIG
 TSSCFVEPLESTGIYFVYAALYQLVKHFPDKSFNPVLTARFNREIETM
 55 FDDTRDFIQAHFYFSPRTDTPFWRANKELRLADGMQEKIDMYRAGMAI
 NAPAPDDAQLYYGNFEEEFNRFWNNSNYCVLAGLGLVPDAPSPRLAH
 MPRATESVDEVFGAVKDQRNLETLPSLHEFLRQQHGR

SEQ. ID NO: 33

MSGKIDKILIVGGGTAGWMAASYLGKALQGTADITLLQAPDIPTLGVG
 EATHPNLQTAFFDFLGIPEDEWMREC NASYKVAIKFINWRTAGEGTSE
 ARELDGGPDH FYHSFGLLKYHEQIPLSHYWFDRSYRGKTVEPFYACY
 65 KEPVILDANRSPRRLDGSKVTNYAWHFD AHLVAKAMEEPFLDMSDHL

85

-continued

NDSAVATQVPHDDDANGVEPFTSAIAMKSGWTWKIPMLGRFGTGYVYS
 SRFATEDEAVREFCEMWHLDPEQTPLNRIRFRVGRNRRRAWVGNVCSIG
 TSSCFVEPLESTGIYFVYAALYQLVKHFPDKSFPVLTARFNREIETM
 FDDTRDFIQAHFYFSPRTDTPFWRANKELRLADGMQEKIDMYRAGMAI
 NAPASDDAQLYYGNFEEEFRCNWNSSYYCVLAGLGLVPDAPSPRLAH
 MPRATESVDEVFGAVKDQQRNLLLETPLSLHEFLRQQHGR

SEQ. ID NO: 34

MSGKIDKILIVGGGTAGWMAASYLGKALQGTADITLQAPDIPTLGVG
 EATIPNLQTAFFDFLGIPEDEWMRECNASYKVAIKFINWRTAGEGTSE
 ARELDGGPDHFYHSFSLKLYHEQIPLSHYWFDRSYRGKTVEFPDYACY
 KEPVILDANRSPRRLDGSKVTNYAWHFD AHLVAKAMEEPFLDMSDHL
 NDSAVATQVPHDDDANGVEPFTSAIAMKSGWTWKIPMLGRFGTGYVYS
 SRFATEDEAVREFCEMWHLDPEQTPLNRIRFRVGRNRRRAWVGNVCSIG
 TSSCFVEPLESTGIYFVYAALYQLVKHFPDKSLNPVLTARFNREIETM
 FDDTRDFIQAHFYFSPRTDTPFWRANKELRLADGMQEKIDMYRAGMAI
 NAPASDDAQLYYGNFEEEFRCNWNSSYYCVLAGLGLVPDAPSPRLAH
 MPQATESVDEVFGAVKDRQRNLLLETPLSLHEFLRQQHGR

SEQ. ID NO: 35

MSGKIDKILIVGGGTAGWMAASYLGKALQGTADITLQAPDIPTLGVG
 EATIPNLQTAFFDFLGIPEDEWMRECNASYKVAIKFINWRTAGEGTSE
 ARELDGGPDHFYHSFGLKLYHEQIPLSHYWFDRSYRGKTVEFPDYACY
 KEPVILDANRSPRRLDGSKVTNYAWHFD AHLVADFLRRFATEKLGV
 VEDRVEHVQRDANGNIESVRTATGRVFDADLFVDCSGFRGLLINKAME
 EPFLDMSDHLNDSAVATQVPHDDDANGVEPFTSAIAMKSGWTWKIPM
 LGRFGTGYVYSSRFATEDEAVREFCEMWHLDPEQTPLNRIRFRVGRNR
 RAWVGNVCSIGTSSCFVEPLESTGIYFVYAALYQLVKHFPDKSLNPVL
 TARFNREIETMFDDTRDFIQAHFYFSPRTDTPFWRANKELRLADGMQE
 KIDMYRAGMVINAPASDDAQLYYGNFEEEFRCNWNSSYYCVLAGLGL
 VPDAPSPRLAHMPQATESVDEVFGAVKDRQRNLLLETPLSLHEFLRQQH
 GR

SEQ. ID NO: 36

ATGATGTCTGAAAATTTGCAATTGTGAGCTGAAGAAATGAGACAATTG
 GGTTACCAAGCAGTTGATTGTATCATCGATCACATGAACCATTGTAAG
 TCTAAGCCAGTTTCAGAAACAATCGATTCTGATATCTTGAGAAATAAG
 TTGACTGAATCTATCCCAGAAAATGGTTCAGATCCAAAGGAATTGTTG
 CATTTCTTGAACAGAAACGTTTTTAATCAAATTACACATGTTGATCAT
 CCACATTTCTTGGCTTTTGTTCAGGTCCTAAATAATTACGTTGGTGTT
 GTTGACAGATTTCTTGGCTTCTGTTTTTAATGTTTTTCCAACGTCATGG
 ATTGCTGGTGCAGGTGCTGAACAAATCGAATTGACTACAATTAATTGG
 TTGAAATCTATGTTGGGTTTTCCAGATTGAGCTGAAGGTTTATTTGTT
 TCTGGTGGTTCAATGGCAAATTTGACAGCTTTGACTGTTGCAAGACAG
 GCTAAGTTGAACAACGATATCGAAAATGCTGTTGTTTACTTCTCTGAT

86

-continued

CAAACACATTTCTCAGTTGATAGAGCATTGAAGGTTTTAGGTTTTAA
 CATCATCAAATCTGTAGAATCGAAACAGATGAACATTTGAGAATCTCT
 5 GTTTCAGCTTTGAAGAAACAAATTAAAGAAGATAGAATAAGGGTAAA
 AAGCCATTCTGTGTTATTGCAATGCTGGTACTACAAATTGTGGTGCT
 GTTGATTCTTTGAACGAATTAGCAGATTTGTGTAAACGATGAAGATGTT

10 TGGTTCATGCTGATGGTCTTATGGTCTCCAGTATCTTGTCTGAA
 AAGGGTTCAGCTATGTTGCAAGGTATTCTAGAGCAGATTCTTTGACT
 TTAGATCCACATAAGTGGTGTTCACACATACGATGTTGGTGTGTT

15 TTGATCAGAACTCTCAATATTTGTCAAAGACTTTTAGAATGATGCCA
 GAATACATCAAGGATTCAGAACTAACGTTGAAGGTGAAATTAATTTT
 GGTGAATGTGGTATCGAATTGTCAAGAAGATTGAGAGCTTTGAAGGTT

20 TGGTGTCTTTTAAAGTTTTCGGTGTGCTGCTTTTAGACAAGCAATC
 GATCATGGTATCATGTTAGCAGAACAGTTGAAGCATTTTGGGTAAA
 GCAAAGATTGGGAAGTTGTTACACCAGCTCAATTGGGTATCGTTACT

25 TTTAGATACATTCATCTGAATTGGCATCAACAGATACTATTATAGAA
 ATTAATAAGAAATGGTTAAGGAAATCACACATAGAGGTTTCGCTATG
 TTATCTACTACAGAATTGAAGAAAAGTTGTTATTAGATTGTGTTCA

30 ATTAATCCAAGAACTACAACGAAGAAATGTTGCAATCATGATGAAG
 ATTAAGCATTGGCTGAAGAAAGTTTCTATTTCATACCCATGTGTTGCT
 GAATAA

SEQ. ID NO: 37

35 MMSENQLSAAEMRQLGYQAVDLIIDHMHNLKSKPVSETIDSIDILRNK
 LTESIPENGSDPKELLHFLNRNVFNQITHVDHPHFLAFVPGPNVYGV
 VADFLASGFNVFPTAWIAGAGAEQIELTTINWLKSMGLFPDSAEGLFV

40 SGGSMANLTALTVARQAKLNNDIENAVVYFSDQTHFSDRALKVLGFK
 HHQICRIETDEHLRISVSALKKQIKEDRTKGKPCFVIANAGTTFPQY
 DVGCVLIRNSQYLSKTFRMMPEYIKDSETNVEGEINFGECGIELSRRF

45 RALKVWLSFKVFGVAAPRQADHGMILAEQVEAFGLKAKDWEVVTAPQ
 LGIVTFRYIPSELASTDTINEINKLVKEITHRGFAMLSSTTELKEKVV
 IRLCSINPRTTTEMLQIMMKIKALAEVVISISYPCVAE

SEQ. ID NO: 38

50 ATGTCTCAAGTGATCAAAAAAGAGAAACACTTTCATGATCGGTACCG
 AATATATATTGAACCTACCCAATTAGAAGAGGCTATTAAATCTTTTCG
 TCCATGATTTCTGTGCCGAAAAGCACGAAATTCAGACCAACCAGTTG

55 TCGTTGAAGCTAAGGAACACCAAGAAGATAAGATTAAGCAAATCAAAA
 TTCCAGAAAAGGTCGTCCAGTCAATGAAGTTGTTTCTGAAATGATGA
 ACGAAGTCTACAGATACAGAGGTGATGCCAACCATCCAAGATTCTTCT

60 CTTCGTTCCAGGTCCAGCTTCTCTGTCTCCTGGTTGGGTGACATCA
 TGACTTCTGCTTACAACATCCACGCTGGTGGTTCTAAATTGGCCCCAA
 TGGTTAACTGTATCGAACAAGAAGCTTTGAAATGGCTCGCAAAGCAAG

65 TTGGTTTCACCGAAAACCTGGTGGTGTTCGTTTCCGGTGGTTCTA
 TGGCTAACATTACTGCTTTGACAGCCGCTAGAGACAACAGCTTACCG

87

-continued

ATATTAACCTGCACTTGGGTACTGCTTACATTTCCGACCAAAACCCACT
CTTCAGTTGCTAAGGGTTTACGTATCATTGGTATCACTGACTCCAGAA
TCAGAAGAATTCCAATACTCCCACTTCCAAATGGACACCACTAAGT
TGGAAGAAGCTATCGAAACCGACAAGAAGTCCGGTTACATCCCATTG
TTGTTATCGGTACTGCCGGTACCACCAACTGGTTCATTGACCCAT
TGACTGAAATTTCCGCTTTGTGTAAGAAGCATGACATGTGGTTCCACA
TTGATGGTGCCTACGGTCTTCAGTTTGTGTCTCCAAAGTACAAGT
CTTTGTTAACAGGTACCGGTTTGGCTGACTCTATCTCTTGGGATGCTC
ACAAGTGGTTATTTCAAACCTACCGTTGTGCTATGGTTTGGTCAAGG
ATATTAGAAACTGTGTTTATTCTTCCACGTCATCCAGAATACTTGA
AGGATCTGGAACACGACATCGATAACGTTAACACCTGGGACATCGGCA
TGGAATTGACCAGACCAGCCAGAGGTTTGAAGTTGGTTGACTTTGC
AAGTTCAGGTCTGATCTTATTGGTTCTGCCATTGAACACGGTTTCC
AATTGGCTGTCTGGGCTGAAGAAGCCTTGAATCCAAAGAAGGACTGGG
AAATTGCTCTCTCTGCTCAAATGGCTATGATCAACTCAGATATGCTC
CAAAGGACTTGACCAAGGAAGAACAAGATATCTTGAACGAAAAGATCT
CTCACCGTATTTTGAATCCGGTTACGCTGCTATCTTCACTACCGTCC
TAAACGGTAAGACTGTCTTGAGAATCTGCGCTATTACCCAGAAGCTA
CTCAAGAAGACATGCAACACACCATTTGATTGTTGGACCAATACGGTA
GAGAAATCTCACTGAAATGAAGAAGGCTTAG

SEQ. ID NO: 39

MSQVIKKRNTFMIGTEYILNSTQLEEAIKSFVHDFCAEKHEIHDPV
VVEAKEHQEDKIKIKIPEKGRPVNEVSEMMNEVYRYRGDANHPRFF
SFVPGPASSVSWLGDIMTSAYNIHAGGSKLAPMVNICEQEVKLWLAKQ
VGFTENPGGVFVSGGSMANITALTAARDNKLTIDINHLGTAYISDQTH
SSVAKGLRIIGITDSRIRRIPTNSHFQMDTTKLEEAIEDKKSGYIPF
VVIGTAGTTNTGSIDLPEISALCKKHDWPHIDGAYGASVLLSPKYK
SLLTGTGLADSIWDHAKWLFQTYGCAMVLVKDIRNLFHSPHVNPEYL
KDLENDIDNVNTWDIGMELTRPARGKLWLTQLVGLSDLIGSAIEHGF
QLAVWAEALNPKKDWEIVSPAQMAMINFRYAPKDLTKEEQDILNEKI
SHRILESYAAIFTTVLNGKTVLRICAIHPEATQEDMQHTIDLLDQYG
REIYTEMKKA

SEQ. ID NO: 40

ATGAAGCCAGCTGAAGCTAAACCACCTCACATGGACCATGATACTTTC
AGATCCTTGGGTACCAAGCTATTGACTGGATTGCTGATTACTGGCAA
CACTTGGCTGACAGACCAGTTGCTCCATTGGTTGAACCAGGTACTGTT
AGAGCCCAATTGCCTACCACTCCACCAGAGTACGGTGAAGACTTCCCA
GCTTTGTTATCTGACTTGGAAGAATTGTCTTGCCAGGCTTGTTCAC
TGGCAACACCCAAGATTCTTCGGTTACTTCCAGCTAACGCTTCCGGT
CCAGCCGTTTTAGCTGAATTATTATCTGCTGGTCTAGGTGTCCAAGT
ATGAACCTGGAACACTTCTCCAGCTTGTACCGAAGTTGAACAACAAATG

88

-continued

TTGGACTGGTTCGTTTCATTGTTAGGGTTGCCAGAACACTTCAGAGGT
GGTGGTGTATCCAAGACACTGCCTCCTCCGCCGTTTTTGGTCGCTCTA
5 TTGACTGCTTTGCACAGAGCTTCTGCTGGTAGAACAGAGATCACGGT
ACTGGTGAATGTGGTTACAGGGTTTACCTAACCGCTGAAACTCACTCC
GCTGCTAGAAAGGCTGCCGTCATCACCGTTTTGGGTTTGCCTGCTGTT
10 TGTGAAGTCGCAACTGATGCTGACGGTCTATGGATGCCGCTGACTTG
GCCGCTCGTATTCAAGCCGATTGGCTGCCGGTTTGACCCCATTTGATG
GTTGTTGCTACCCGTGGTACCACCTCTGATTTGCTCTTCGATCCATTG
15 GAAGACATCGGTCCAGTCTGTAGAAGACATGGTGTCTGGTTACACGTT
GACGCTGCTTATGCTGGTGTGCTGCAGTGTGCGATGAATTGCATTGG
GTTAATGATGGTGTGTAGATACGCTGATTCTTACTGTACTAACCCACAC
AAGTGGTTGTTGACCAACTTCGACTGTGATTATTATGGGTTGCCGAC
20 CCAGAAGCTTTGGTCAACGCTTTGTCCGTTTTGCTGCTTACTTGAGA
AACCACGCTTCTGAATCAGGTAGAGTTACTGACTACAGACACTGGCAA
GTTCCATTGGGCAGAAGATTCAGAGCCTTGAAGTGTGGTCTGCTCTTA
25 AGATGGTACGGTGCCGAAGTTTGGATCTCACATCAGAACTGGTGTC
AGACACGCCAGATTGTTGCTGACCTGGTCAGAGAAGATGACCGTTTC
ACCTTGTTGGCGCCAGCTACTTTGGGTTTGGTCACTTTCCTGTCAGCT
30 GGTCTGACACCGACAACAGAACTTGTGTCAGCTATCAACGCCGAA
GGTACAACCTTTTTTAAGTATAGTAAAAGAACGGTACCTTTTTCTTG
AGATTTGCTACCGGTGGTACCCACACTGAAGACCATCACGCTCCAGAA
35 GCTTGGAGAGCTGTCCAAATGCCGTCCAAAACGCCGTCCTCAAGAGCC
CAACATGCCACCGATGGTACTGCCGATGCTTTGAACGGTTAA

SEQ. ID NO: 41

MKPAAEKPPHMDHDTFRSLGHQAIDWIADYQHLADRPVAPLVEPGTV
40 RAQLPTTPPEYGEDFPALLSDLERIVLPLGLHWQHPFFGYFPANASG
PAVLAELLSAGLVQGMNWNTPACTEVEQQMLDWFVHLLGLPEHFRG
GGVIQDTASSAVLVALLTALHRASAGTRDHTGECGRVYLTAETHS
45 AARKAAVITGLGLRAVEVATDADGAMDAADLAARIQADLAAGLTPLM
VVATRGTTSDLSFDPLEDIGPVCRRHGVWLHVDAAYAGVAAVDELHW
VNDGVRYADSYCTNPHKWLLTNFDCDLLWVADPEALVNALSVLPAYLR
50 NPASESGRVTDYRHWQVPLGRRFRALKLWSVLRWYGAEGRLSHIRTGV
RHARLFADLVREDDRFTLLAPATLGLVTFRHAGSDTDNRNLLHAINAE
GTTFLTHSEKNGTFFLRFATGGTHTEHHVHEAWRAVQNAVQNAVRA
55 QHATDGTADALNG

SEQ. ID NO: 42

ATGGATATTGAAGCTTTTAGAAAAGCAGGTACCAAGCTATCGATAGA
ATTTGTGACTACTACTCTTTGCAAAATAGACCAGTTGTTCCATCA
60 GTTCAACACGAGTTATTTGTAGATGCATTACCAGATTCTCCACCAGAA
CAAGGTGAAGATTTCACTGTTATTGCTGATGATTACCAAAAGTACATC
TTGCCAGGTTTGACACATTGGCAACATCCATCTTTCTTTGCATACTTT
65 CCAACAGCTTGTAAGTTTGAAGGTATTTGGGTGACTTGATTCTACA

89

-continued

TCAACTGCAATCCAGGTTTTAATGGTTGGCTTCTCCAGCATGTACT
 GAATTAGAAATGGTTGTTATGGATTGGTCTGCTAAGTTGTTAGGTTTG
 TCAGAACATTTCTTGCAATCTTCAGGTAAAGGTGGTGGTTATTCAA
 ACTACAGCATCAGAATTGGCTTTAGTTGTTGTTGTTGCTGCAAGAGAA
 AGATACTTGAGAATTCATCCAGATGCTAAGGCAGATGAATTGGTTATA
 TATACTACAACCTAAACACATCTCTTGGGTGTTAAAGCTGGTTTAGTT
 TTTGGTATGGAATGTAGAGCATTGGAAGTTAAAGCTGAAGATGCTTAC
 GCATTAAGAGGTGCAACTTTGAAGTCAGCTTTAGAAGAAGATGAAAAG
 AGAGGTAAAGACCTTTTATTTTGGTTGCAACAGTTGGTACAACCTTCT
 TCAGGTGCTATTGATAGATTGGATGAAATTGGTCAAGTTTCTGAAGAT
 TATCCATCATTATGGATTTCATGTTGATGCTGCATGGGCAGGTGTTACT
 TTGGCTTGTCCAGAATACAGAGGTACAGCTCAATTAGAAAACATCAAC
 GCTTATGCAACATCTTTCGGTACTAAGTTCCATAAGTGGGGTTTGGTT
 AACTTTGATGCTGCATTGTTGTGGGTTAAGGATAGAAAGGATTGACA
 GATGCTTTAGATGTTACTCCAGAATCTTGAGAACAAAACAAGGTGAC
 GCTGGTGCAGTTGTTGATTTTGAAGATTGGCATTGGGTTTGGGTAGA
 AGATTTCAGATCTTTGAAGGTTTGGTTCGTTTGGATCATAACGGTGT
 GAAGGTTTGTAGAACTACATCAGACAAGGTATTAATTAAGTGAACAT
 TTTACTTCTTTAATAGAGCTTCTTTGATTTTTCATTAGTTACAGCA
 CCATCATTGCTTTGACAGTTTTTAGATTGACTCCAGCTGGTGCATCT
 TTGACTGGTTTGAAGTGAACGAATTAACAGAGCTTCTACGCAAGA
 TTATCTTCAAGACATGATATCATGTTGACACAACTGTTTGAACGGT
 GTTTTCTGTATCAGATTGTCAGTTGGTGTGCAAGAACTCAACAAGAA
 CATATTGATACAGCTTGGGATTTGTTACAACAAGAAGGTGCTGTTGCA
 GTTCAAGAATACATGCAAAAGATTTCTAAGGATGTTCAATAA

SEQ. ID NO: 43

MDIEAFRKAGYQAIIDRICDYYSYLQNRPVVPSVQPGYLLDALPDSPE
 QGEDFTVIADDDYQKYLPLTHWQHPSFFAYFPTACTFEGILGDLYST
 STANPGFNWLSPACTELEMVMDWSAKLLGLSEHFLHSSGKGGVQ
 TTASELALVVVVAARERYLRHPDAKADELVIYTTTQTHSLGVKAGLV
 FGMECRALEVKAEADAYALRGATLKSALAEDEKRGKRPFILVATVGTT
 SGAIIDRLDEIQVSEDPYSLWIHVDAWAGVTLACPEYRGTAQLENIN
 AYATSFGTNFHKWGLVNFDAALLWVKDRKDLTDALDVTPEFLRTKQGD
 AGAVVDFRNWHLGLGRRFRSLKVFVFLRSYGVGFRNYIRQGIKLNH
 FTSLIRASLDVSLVTAPSFALTVFRLTPAGASLTGSELNENLRAFYAR
 LSSRHDIMLTQTVLNGVFCIRFAVGAARTQQEHIDTAWDLLQQEGAVA
 VQEYMQKISKDVQ

SEQ. ID NO: 44

ATGAAGTTCTGGAGAAAGTACACACAAGAAATGGATGAAAAGATT
 ACTGAATCTTTGGAAGAACTTTGAAGTACGATAACACTAAGACAATC
 GGTATTCCAGGTACTAAGTTGGATGATACAGTTTCTATGATGATCAT

90

-continued

TCTTTCGTTAAGCATTCACCATCTTGAGAACTTTTATTCAAAACCCA
 AACCATATCGGTTGTCTACTTATGATAAGGCTGATATCTTGTTCGGT
 5 GGTACATTGATATCGAAAGAGAATTAATCCAATTGTTAGCAATCGAT
 GTTTTGAACGGTAACGATGAAGAATTTGATGGTTACGTTACTCAAGGT
 GGTACAGAAGCTAACATCCAAGCAATGTGGGTTTACAGAACTACTTC
 10 AAGAAAGAAAGAAAGGCTAAGCATGAAGAAATCGTATCATCACTTCA
 GCAGATACACATTACTCTGCATACAAAGGTTAGATTGTTGAACATC
 GATATTATTAAGGTTCCAGTTGATTTTATTCAAGAAAAATCAAGAA
 15 AATACATTGGATTCAATTGTTAAGAAGCTAAAGAAATGGTAAAAAG
 TACTTCATCGTTATCTCTAACATGGGTACTACAATGTTTGGTTAGTT
 GATGATCCAGATTGTTACGCTAACATCTTCGATAAGTACAATTTGGAA
 TACAAAATTCATGTTGATGGTGCATTGGTGGTTTATATATCCAATT
 20 GATAATAAGGAATGTAAAGTATTTCTCTAATAAGAACGTTTCTTCA
 ATCACATTAGATGGTCATAAGATGTTGCAAGCTCCATACGGTACTGGT
 ATCTTCGTTTCAAGAAAGATTTGATCCATAACACTTTGACAAAGGAA
 25 GCAACTTACATCGAAAATTTGGATGTTACATTGTCTGGTTCAAGATCT
 GGTTCAAATGCTGTTGCAATTTGGATGGTTTGTAGCTTCTTATGTTCCA
 TACGGTTGGATGGAAGGATTAATAAGTTGAGAAATAGAATAAATGG
 30 TTGTGTAAGCAATTGAACGATATGAGAATTAATATTACAAAGAAGAT
 TCAATGAATATTGTTACAATTGAAGAACAAATGTTAATAAGGAAATC
 GCTGAAAAGTACTTTTTAGTTCCAGAAGTTTATAACCACTAACAAC
 35 TGGTACAAGATCGTTGTTATGGAACATGTTGAATTGGATATCTTGAA
 TCTTTGGTTTACGATTTGAGAAGTTTAAATAAGGAACATTTGAAGGCA
 ATGTAA
 40 SEQ. ID NO: 45
 MKFWRKYQQEMDEKITESLEKTLNNDNTKTIIGIPGTLDDTVFYDDH
 SFVKHSPYLRTFIQPNHIGCHTYDKADILFGGTFDIERELIQLLAID
 45 VLNGNDEEFDGYVTQGGTEANIQAMWVYRNYFKKERKAKHEIAIITS
 ADTHYSAYKSDLLNIDIIKVPVDFYSRKIQENTLDSIVKEAKEIGKK
 YFIVISNMGTTMFGSVDDPDLYANIFDKYNLEYKIHVDGAFGGFIYPI
 DNKECKTDFSNKNVSSITLDGHKMLQAPYGTGIFVSRKNLIHNTLTKE
 50 ATYIENLDVTLSGSRSGSNAVAIWMVLASYGPYGMKINKLRNRTKW
 LCKQLNDMRIKYKEDSMNIVTIEEQYVNKEIAEKYFLVPEVHNPTNN
 WYKIVMEHVELDILNSLVYDLRKFNKEHLKAM
 55 SEQ. ID NO: 46
 ATGAATGCATCTGAATTGAGAAGAGGTAAGAAATGGTTGATTAC
 ATGGCTAACTACATGGAAGGTATTGAAGGTAGACAAGTTTATCCAGAT
 GTTGAACCAGGTTACTTGAGACCATTAAATCCAGCTGCAGCTCCACAA
 60 GAACCAGATACATTGAAGATATTATTAACGATGTTGAAAAGATTATC
 ATGCCAGGTGTTACACATTGGCATTACCATATTTCTTTGCATACCTT
 CCAACTGCTTCTTATATCCAGCAATGTTGGCTGATATGTTATGTGGT
 65 GCAATTGGTTGATTGGTTTTTCTTGGGCAGCTTACCAGCTTGTACA

91

-continued

GAATTGGAACATGTTATGATGGATTGGTTAGGTAAATGTTGGAATTA
 CCAAAGGCATTTTTGAACGAAAAGGCTGGTGAAGTGGTGGTGTATT
 CAAGGTTCTGCATCAGAAGCTACATTAGTTGCTTTGTTAGCAGCTAGA
 ACTAAAGTTATTCATAGATTGCAAGCAGCTTCTCCAGAATTAACACAA
 GCAGCTATCATGGAATAATGGTTGCATATTCTTCAGATCAAGCTCAT
 TCTTCAGTTGAAAGAGCTGGTTTGATCGGTGGTGTAAAGTTGAAGGCA
 ATCCCATCTGATGGTAATTTTCTATGAGAGCATCAGCTTTGCAAGAA
 GCATTAGAAAGAGATAAAGCAGCTGGTTTGATTCATTTTTTCATGGTT
 GCTACTTTAGGTACTACAACTTGTGTCTTTTCGATAATTTGTTAGAA
 GTTGGTCCAATTTGTAATAAGGAAGATATTTGGTTGCATGTTGATGCA
 GCTTACGCAGGTTACGCTTTTATTTGTCCAGAATTCAGACATTTGTTG
 AACGGTGTGAATTCGCTGATTCCTTTTAAATTTCAATCCACATAAATGG
 TTGTTAGTTAACTTCGATTGTTTCAGCAATGTGGGTTAAGAAAAGAACA
 GATTTGACTGGTGCTTTTAGATTAGATCCAACATATTTGAAGCATTCCT
 CATCAAGATTACGCTTTGATCACTGATTACAGACATTGGCAAATTCCTCA
 TTGGGTAGAAGATTAGATCTTTGAAGATGTGGTTCGTTTTAGAAATG
 TACGGTGTAAAGGGTTTGCAAGCTTACATCAGAAAGCATGTTCAATTG
 TCTCATGAATTCGAATCATTAGTTAGACAAGATCCAAGATTCGAAATC
 TGTGTTGAAGTTATTTTGGGTTAGTTTGTGTTTCAGATTGAAGGGTTCT
 AATAAGGTTAACGAAGCATTGTTGCAAGAATTAATTCAGCTAAGAAA
 ATTCATTTGGTTCCATGTCAATTTGCGTGATAAGTTTCGTTTTGAGATTC
 GCTATTTGTTCAAGAACTGTTGAATCAGCACATGTTCAAAGAGCTTGG
 GAACATATCAAGGAATTGGCAGCTGATGTTTTGAGAGCTGAAAGAGAA
 TAA

SEQ. ID NO: 47

MNASEFRRRGKEMVDYMANYMIEGRQVYPDVEPGYLRPLIPAAAPQ
 EPDTFEDIINDVEKIIIMPVTHWHSPYFFAYFPTASSYPAMLADMLCG
 AIGCIGFSWAASPACTELETVMMDWLKMLELPKAFLEKAGEGGGVI
 QGSASEATLVALLAARTKVIHRLQAASPELTQAAIMEKLVAISSDQAH
 SSVERAGLIGGVKLKAIPSDGNFAMRASALQEALERDKAAGLIPFFMV
 ATLGTTCSSFDNLLEVGPICNKEDIWLHVDAAYAGSAFICPEFRHLL
 NGVEFADSFNPNPHKWLNVNFDCSAMWVKKRTDLTGAFRLDPTYLKHS
 HQDSGLITDYRHWQIPLGRRFRSLKMWVFRMYGVKGLQAYIRKHVQL
 SHEFESLVRQDPRFEICVEVILGLVCFRLKGSNKVNEALLQRINSACK
 IHLVPCHLRDKFVLRFACSRIVESAHVQRAWEHIKELAADVLRAERE

SEQ. ID NO: 48

ATGGGTTCTATCTCAGAAACTGTGATGATTCTATTTCATTAGCTGCA
 CCTTTTAGACCATTGGAACAGAAGAAATTCAGAAAGCAAGCTCATGTT
 ATGGTTGATTTCATCGCAGATTACTACAAGAACATTGAAAATTATCCA
 GTTTTGTACAAAGTTGAACAGGTTATTTGAAAAATAGATTGCCAGAA
 ACTGCTCCACATTTGCCAGAATCTTTTGAAACAATCTTGAAGGATATT

92

-continued

AAGAAAGATATCGTTCCTGGTATGACTAATTGGTTGTACCAAATTTTC
 TTTGCTTATTTTCCAGCAACTGTTTCTTCAGCTGCATTGTGGTGAA
 5 ATGTTGTGTACAGGTTTTAATCTGTTGGTTTTAATGGTTAGCTTCA
 CCAGCATCTACAGAATTGGAATGGTTGTTATTGATTGGTTAGCTAAC
 ATGTTGAAGTTGCCAAAGTCTTTTATGTTTCATGGTACTGGTGGTGGT
 10 GTTATTCAAGGTACTACATCTGAAGCTATCTTGTGTACATTGATTGCT
 GCAAGAGATGGTGCTTTGGAAAAGATTGGTATGGAAAACGTTGGTAAA
 TTAGTTGTTTACGGTTCAGATCAAACCTATTCTTTCTTTCAAAGACT
 15 TGTAAGGTTGCTGGTATCTTCCCATGTAACATCAAGTTGATCCCAACT
 ACAAGAGAAGATAACTTCTCTATGTCACCAATCGCATTGAGAGAACAA
 ATTGAAGCTGATGTTGCAGATGGTTAGTTCCAAATTTCTTGTGTACT
 ACAGTTGGTACTACATCAACTGCTGCAATTGATCCAGTTTCTGAAGTT
 20 GCTAAGGTTGCAAACGATTTCAATATTTGGGTTTCATGTTGATGCTGCT
 TATGCTGGTCTGCTGATGTTTGTCCAGAATTGAGACAATACTTAGAT
 GGTATCGAATTGGTTGATTCTTTTTCATTGTCTCCACATAAGTGGTTG
 25 TTGTGTTTCTTGGATTGTTGTTGTTTGTGGTTGAAGAAACACATTTG
 ATGGTTAAGGCTTTTGTCAACTAACCCAGAATATTTGAGAAATAAGAGA
 TCTGAATTCGATGGTGTGTTGATTCAAGGATTGGCAAATCGGTACA
 30 GGTAAGAAGATTCAAAGCTTTGAGATTATGGTTGGTTATGAGATCATA
 GGTGTTGAAAATTTGAAGAGACATATCTGTCTGATGTTCAAATGGCT
 AAGATGTTGCAAGGTTTAGTTAAATCTGATCCAAGATTGCAAAATCATC
 35 GTTCCAAGAGCTTTCGCATTAGTTTGTGTTTCAGATTGAACCTGGTAAA
 GGTATGATGATGAAATCGATAAGGAAATCTGAATAAGGAATTGTTA
 GATTTGATTAATTCAACAGGTAGAGCTTACATGACTCATACAAAGCA
 40 GGTGGTATCTATATGTTGAGATTTGCTGTTGGTACTACATTGACTGAA
 GAACATCATGTTTACGCTGCATGGGAATTGATTAAAGAATGTACAGAT
 GCATCTTTGACTAAAACAAATATTATTGAATAA

SEQ. ID NO: 49

MGSISENCDDSSISLAAPFRPLEPEFRKQAHVMVDFIADYYKNIENYP
 VLSQVEPGYLNRLPETAPHLPEPFETILKDIKKDIVPGMTNWLSPNF
 FAYFPATVSSAAFVGEMLCTGFNSVGFNWLSPASTELEMVVIDWLAN
 50 MLKLPKSFMFHGTGGGVIQGTTSAILCTLIAARDGALEKIGMENVGK
 LVVYGSDDQTHSFFQKTCKVAGIFPCNIKLIPTTREDNFSMSPIALREQ
 IEADVADGLVPIFLCTTVGTTSTAIDPVSEVAKVANDFNIVWHVDAA
 55 YAGSACICPEFRQYLDGIELVDSFSLSPHKWLLCFDCCCLWLKPHL
 MVKALSTNPEYLRNKRSEFDGVVDFKDWQIGTGRRFKALRLWLVMRSY
 GVENLKRHILSDVQMAKMEGLVKSDPRFEIIVPRAFALVCFRLNPGK
 60 GYDDEIDKEILNKELLDLINSTGRAYMTHTKAGGIYMLRFAVGTTLTE
 EHHVYAAWELIKECTDASLTKTNIIE

SEQ. ID NO: 50

ATGGGTTTCATTAGATACAAATCCAACTGCTTTTCTGCATTTCCAGCT
 65 GGTGAAGGTGAAACTTTTCAACCATTGAACGCAGATGATGTTAGATCA

93

-continued

TATTTGCATAAGGCTGTTGATTTCATCTCTGATTACTACAAATCAGTT
 GAATCTATGCCAGTTTTGCCAAACGTTAAGCCAGGTTATTTGCAAGAT
 GAATTAAGAGCATCACCACCACTTACTCTGCTCCATTGATGTTTACA
 ATGAAGGAATTGAGATCTTCAGTTGTTCTGCTGATGACTCATTGGGCA
 TCTCCAAATTTCTTTGCTTTCTTTCCATCAACAAATTTGCTGCAGCT
 ATTGCTGGTGACTTGATTGCATCAGCTATGAATACAGTTGGTTTTACT
 TGGCAAGCTTCTCCAGCAGCTACTGAAATGGAAGTTTGGCATTAGAT
 TGGTTGGCTCAAATGTTGAATTTGCCAACTTCTTTTATGAACAGAACT
 GGTGAAGGTAGAGGTACTGGTGGTGGTGTATTTTGGGTACTACATCT
 GAAGCTATGTTGGTTACATTAGTTGCAGCTAGAGATGCAGCTTTGAGA
 AGATCTGGTTCAGATGGTGGTGCAGGTTGCATAGATTAGCTGTTTAT
 GCAGCTGATCAACACATTCAACTTTCTTTAAGGCATGTAGATTGGCT
 GGTTTTGATCCAGCAAATATTAGATCTATTCCAACAGGTGCTGAACT
 GATTATGGTTTAGATCCAGCAAGATTGTTAGAAGCTATGCAAGCAGAT
 GCTGATGCAGGTTTAGTTCCAACATACGTTTGTGCAACTGTTGGTACT
 ACATCTTCAAATGCTGTTGATCCAGTTGGTGTCTGTTGCAGATGTTGCA
 GCTAGATTTCAGCTTGGGTTTCATGTTGATGCAGCTTACGCAGGTTCA
 GCTTGTATTTGTCAGAATTCAGACATCATTTGGATGGTGTGAAAGA
 GTTGATTCTATCTCAATGTCTCCACATAAGTGGTTGATGACATGTTTA
 GATTGTACTTGTTTGTACGTTAGAGATACACATAGATTGACTGGTTCT
 TTAGAAACTAACCCAGAATACTTGAAAATCATGCTTCAGATTCTGGT
 GAAGTTACAGATTTGAAAGATATGCAAGTTGGTGTGGTAGAAGATTC
 AGAGGTTTGAAGTTGTGGATGGTTATGAGAACTTATGGTGTGCTAAG
 TTGCAAGAACATATCAGATCAGATGTTGCTATGGCAAAGTTTTTGAA
 GATTTGGTTAGAGGTGACGATAGATTGCAAGTTGTTGTTCCAAGAAAC
 TTCGCTTTGGTTTGTTCAGAATTAGAGCTGGTGCAGGTGCAGCTGCA
 GCTACAGAAGAAGATGCAGATGAAGCTAACAGAGAATTGATGGAAGA
 TTGAATAAGACTGGTAAAGCATACGTTGCTCATAAGTTGTTGGTGGT
 AGATTTGTTTTGAGATTTGCAGTTGGTCTTCATTACAAGAAGACAT
 CATGTTAGATCTGCTTGGGAATTGATTAAGAAAACACAACGTAAATG
 ATGAACTAA

SEQ. ID NO: 51

MGLSDTNPTAFSAFPAPEGETFPQLNADDVRSYLHKAVDFISDYKSV
 ESMPVLPNVKPGYLQDELRAFPPTYSAPFDVTMKELRSSVVPGMTHWA
 SPNFFAFFPSTNSAAAIAAGDLIASAMNTVGFTWQASPAATEMEVLALD
 WLAQMLNLPTSMFMRNTGEGRGTTGGVILGTTSEAMLVTLVAARDAALR
 RSGSDGVAGLHRLAVYAADQTHSTFFKACRLAGFDPANIRSIPTGAET
 DYGLDPARLLEAMQADADAGLVPTVYCATVGTSSNAVDPVGAVADV
 ARFAAWVHVDAAYAGSACICEFRHHLGDGVERVDSISMSPHKWLMTCL
 DCTCLYVRDTHRLTGSLETNPEYLKNHASDSGEVTDLKDMQVGVGRRF
 RGLKLWMVMRTYGVAKLQEHIRSDVAMAKVFEDLVRGDDRFVGVVPRN

94

-continued

FALVCFRIRAGAGAAAAATEDEANRELMERLNKTGKAYVAHTVVG
 RFVLRFAVGSSLQEEHHVRSAWELIKKTTTEMMN
 5 ATGGAAGGTGTTGGTGGTGGTGGTGGTGGTGAAGAATGGTTGAGACCA
 SEQ. ID NO: 52
 ATGGATGCAGAACAAATTAAGAGAATGTGGTCATAGAATGGTTGATTTT
 10 GTTGCAGATTACTACAAGTCAATCGAAGCTTTTCCAGTTTGTCTCAA
 GTTCAACCAGGTTATTTGAAAGAAGTTTGGCCAGATTGAGCTCCAAGA
 CAACCATGATCTTTTAGATTCTTTGTTTCGATGATATCCAACAAAAGATT
 15 ATCCCAGGTGTTACACATTGGCAATCACCAACTACTTCGCATACTAC
 CCATCTAATTCTTCTACTGCTGGTTTCTTGGGTGAAATGTTGTCTGCT
 GCTTTTAATATCGTTGGTTTTTCATGGATCACTTCTCCAGCTGCAACA
 GAATTGGAAGTTATTGTTTTAGATTGGTTCGCAAGATGTTGCAATTG
 20 CCATCACAATTCTTATCTACTGCTTTGGGTGGTGGTGTATTCAAGGT
 ACAGCTTCAGAAGCAGTTTGGTTGCATTGTTAGCTGCAAGAGATAGA
 GCTTTGAAGAAACATGGTAAACATTCTTTAGAAAAATGGTTGTTTAC
 25 GCATCAGATCAAACCTATTCTGCTTTACAAAAAGCATGTCAAATCGCT
 GGTATTTTCTCTGAAAACGTTAGAGTTGTTATTGCTGATTGTAATAAG
 AACTACGCTGTTGCACCAGAAGCAGTTTCAGAAGCTTTGTCTATCGAT
 30 TTGTCTTCAGGTTTGATCCCATTTTTTCAATTTGTGCAACAGTTGGTACT
 ACATCTTCATCTGCTGTTGATCCATTACCAGAATTGGGTCAAATCGCT
 AAGTCAAACGATATGTGGTTCCATATCGATGCTGCTTATGCTGTTTCT
 35 GCATGTATTTGTCCAGAATACAGACATCATTTGAATGGTGTGGAAGAA
 GCAGATTCTTTTAATATGAATGCTCATAAGTGGTCTTGACTAACTTC
 GATTGTTCAATTGTTGTGGGTTAAGGATAGATCATTTTTGATCCAATCA
 40 TTATCTACAAATCCAGAATCTTGAAAAATAAGGCTTCACAAGCAAAC
 TCTGTTGTTGATTTCAAGGATTGGCAAATCCCATTGGGTAGAAGATTC
 AGATCATGGAAGTTGTGGATGGTTTGGAGATTGTACCGTGTGATAAT
 45 TTGCAATCTTACATCAGAAAGCATATCCATTTGGCTGAACATTTCGAA
 CAATTGTTGTTGTGATTCAGATTCAAGATTGCAAGTTGTTACTCCAAGAACA
 TTTTCATTGGTTTGTTCAGATTAGTTCCACCAACTTCTGATCATGAA
 50 AACGGTAGAAAGTTGAACTACGATATGATGGATGGTGTAAATTCATCT
 GGTAATAATTTTCTTGTACATACAGTTTGTCTGGTAAATTCGTTTTG
 AGATTTGCTGTTGGTGCACCATTAACTGAAGAAAGACATGTTGATGCT
 55 GCATGGAAGTTGTTGAGAGATGAAGCTACAAAGGTTTTAGGTAAATG
 GTTTAA

SEQ. ID NO: 53

MEGVGGGGGGEWLRPMDAEQLRECGHRMVFVADYYKSI EAFPVLSQ
 60 VQPGYLKEVLPSAPRQPDTLDSLFDIQQKIIPGVTHWQSPNYFAYY
 PSNSSTAGFLGEMLSAARNIVGFSWITSAPAATELEVIVLDWFAKMLQL
 PSQFLSTALGGGVIQGTASEAVLVALLAARDRAKHKHKSLEKLKVVY
 65 ASDQTHSALQKACQIAGIFSENVRVVIADCNKNYAVAPEAVSEALSID

95

-continued

LSSGLIPFFICATVGTSSSAVDPLPELGQIAKSNDMWFHIDAAYAGS
 ACICPEYRHHLNGVEADSFNMNNAHKWFLTNFDCSLLWVKDRSFLIQS
 LSTNPEFLKNKASQANSVVDKDWQIPLGRRFRSLKLWMVLRLYGVDN
 LQSYIRKHIHLAEHFEQLLLSDSRFEVVTPTFSLVCFRLVPPTSDHE
 NGRKLNNDMDGVNSSGKIFLSHTVLSGKFVLRFAVGAPLTEERHVDA
 AWKLLRDEATKVLGKMV

SEQ. ID NO: 54

ATGACAAAGCCATTCCATCCAAATTGAACGTTGATGCATTGTTTTTA
 GGTCCAAAGTCTGAAAACGCTGTTTTCTTTAGAGAAATGATGGATTAT
 GCAGTTTCTGAACACATGCATTGGAGATCAGGTTTTTCATCCAGAAGAT
 TCTACATTGGTTACTTCAGTTGATAGATACGAACATAACTACAGAGAA
 ACATTGTACAGAACTGAAGGTATTTTGAATCAATTATCTGCTAAGTTG
 AAAAATCTTCAATCCCATTCTCTCTCCAAGATACTTAGGTCATATT
 TCTTCAGATACTTTGATGGTTCTAATTGGCTTATGTTATGGCTATG
 ATGTACAACCCAAACAACCTGTTCTTATGAAGCTTCACCAACTACAAC
 GAATTGGAATTAGAATCTGGTTTGGATTGTGTAGAAATGTCGGTTAC
 AATCCACAACAATCTTGGGTCATATTACATCAGGTGGTACTGTTGCA
 AATTACGAAGGTTTGTGGGTTGCTAGAAATTTGAAGACATTGTCTTTC
 GCAATCTCACAACTCCAAAGGCTAAGGATTGTGTCGAACATAAGACT
 CCAAAGCAATTGAAAAATATGCCAACATCTGAATTGTTAGATTGATC
 ACTGAATTACAAAAGAGAGGTTTGTTCGAAGAAATCAGAGATATGACA
 TGTAGAGGTACTGGTGTAAAGGCTCAAATTTTAGGTAATTTGTTAGTT
 CCACAATCTAAGCATTATTTCATGGATGAAAGCTGCAGATATCTTCGGT
 ATCGGTCAAGAAAACATCATCCATTGCCAGTTAATGATCATTACCAA
 ACAGATGTTGCTCAAATGAGAAAGATCACTTTCGGTTTGATTGAACAA
 GGTGAACCAATTTTAGCAATGATTGCTGTTGTTGTTGTTGTTGTTGTT
 GGTGCAATCGATAGAATCGATGAAGTTATTGCTTTAAGAAAAGAATGT
 GAAGAAAGATACGGTGAATCTTTCTACATCCATGTTGATGCTGCTTAT
 GCTGTTTACGCTTGTTCATGTTGTTGAACGAACAGGGTGAATTCGTT
 GAATACGATGAATTGGCAAACTACCATAGATCTGCTGGTATTATGCCA
 GAAAATATTTTCATGGCCAAAACAGAAGTTTACCAATCTTTTAAAGCT
 TTGAAAGATGTTGATTCAATTACAGTTGATCCACATAAGGTTGGTTTT
 ATTCAATATGCTGCAGGTGCAATTTGTATGAAGGATAAGAGAATTTG
 GAATTAATTTCTTCACATGCTGCATACGTTTGTGAAGGTCATGCTGAT
 AAGACATCTACTAAAAATAGAGGTATTTTGGGTTCTTCAATTATGGAA
 GGTCTTAAATCAGGTGCAACAGCTGCAGCTTTGTGGGCAGCTCATAGA
 TTGTTACCATTGAACATCAACGGTTACGGTAAAGTTATTGACAGCTGGT
 ATCGTTACTGCACAATCTTTGTTGAATAAGTTGGCTAATTTGCCACCA
 ATCGAAATCGGTAACATAGATTGGAAGTTCATATTTTCCATCACCA
 GATTTCCACATGATCAACTTCACTTTTAAAGAAGTTGGTAACGAATCT
 TTGTCAAACCATAACGCTTTGAATAAGAGATTGTATGAATTATGTTCT

96

-continued

TACTCAGCAGGTAGAGCTTACACAAATGATTTCTTGACATCTTCAACT
 ATTTTAGATTACAAGGAATACGGTGACACTCCTTGGCATCATGCTCAA
 5 AAATGTGGTTTTTCAAGATCAGAATGGGAAAAAGTTAGAAACATCTAT
 GTTTTGAGAGCAGCTGTTATGACATACTGTTTGAGAGATGAAGAACAT
 TTCGAAGAATCTCGGGAACAATTGCAAGCTATCTTCGTTAAGAAATTG
 10 ACTCAAATCGTTGATGAAGAAGATAAGAAAGCAAGATTGAGATTAGAA
 CCAGATGTTCTCTTTTATTGGTTAA

SEQ. ID NO: 55

MTKPFHPNLNVDALFLGPKSENAVFFREMMDYAVSEHMHWRSGFHPED
 15 STLVTSDVREHNYRETLYRTEGILNQLSAKLKNSSIPFFSPRYLGHI
 SSDTLMVSNLAYVMAMMYPNNCSYEASPTTTELEESGLDLCRMFGY
 NPQQSWGHITSGGTVANYEGLWVARNLKTLSFAISQHPKAKDLIEHKT
 20 PKQLKNMPTSELLDLITELQKRGLEFEEIRDMTCRGTGVKAQILGKLLV
 PQSKHYSWMKAADIFGIGQENIILPVLNDHYQTDVAQMRKITFGLIEQ
 GEPILAMIAVVGTTVEGAIDRIDEVIALRKECEERYGESFYIHVDAAY
 25 AGYACSMMLNEQGEFVEYDELANYHRSAGIMPENISWPKEVYQSFKA
 LKDVSITVDPHKVGFIQYAAGAIMKDKRILELISSHAAYVFEGHAD
 KTSTKNRGILGSSIMEGSKSGATAAALWAAHRLPLNINYGKVI AAG
 30 IVTAQSLNKLNLNLPPIEIGKHFVHIFPSPDFHMINFTPFKEVGNES
 LSNHNALNKRLYELCSYSAGRAYTNDPLTSTILDYKEYGDTPWHAQ
 KCGFSRSEWEKVRNIYVLR AAVMTYCLRDEEHFEFWEQLQAFVKKL
 35 TQIVDEEDKKARLRLEPDVPPFIG

SEQ. ID NO: 56

ATGGGTTCTTTGGATTCAAACACTACGATACTGAATCTCCAGCTTCAGTT
 GGTCAATTCAATCCATTAGATCCAGAAGAATTGAGAAAGCAAGCTCAT
 40 TGTATCGTTGATTTCATCGCAGATTACTACAAGAACATTGAATCTTAT
 CCAGTTTGTGACAAGTTGATCCAGGTTACAGACATTCAAGATTGGGT
 AAAAATGCTCCATACAGATCTGAACCATTGCAATCAATCTTGAAGGAT
 45 GTTCAAAAGGATATCATCCCTGGTATGACTCATTGGATGCTCTCAAAT
 TTCTTTGCTCATTTTCCAGCAACTGTTTCTTCAGCTGCATTGTTGGT
 GAAATGTTGTGTACATGTTTCAACTCTGTTGGTTTTAATTGGTTAGCA
 50 TCACCAGCTGCAACAGAATTGGAATGGTTGTTATTGATTGGTTAGCT
 AACATGTTGAAGTTGCCAAAGTCTTTTATGTTTTTCAGGTACTGGTGGT
 GGTGTTTTGCAAGGTACTACATCTGAAGCTATCTGTGTACATTGATT
 55 GCTGCATCACCAATGCATTTGCAAAATCGTTGGTGTAAAGACTTCTACA
 TCATTTCGTTGTTTATGGTTCTGATCAAACCTATTCAACATACGCTAAA
 GCATGTAAATTAGCAGGTATTTTGCCATGTAAACATCAGATCTATTCCA
 ACTACAGCTGATTCAAACCTCTCTGTTTACCATTGTTGTTGAGAAGA
 60 GCTATTGAAGCAGATAAAGCTGCAGGTATGGTTCCATTGTATATTTGT
 GCTACAGTTGGTACTACATCTACTACAGCAATTGATCCATTATCTTCA
 TTGGCTGATGTTGCAAAATGATTATGGTGTGTTGGTTTCATGTTGATGCT
 65 GCATACGCTGGTTCTGCATGTATTGTCCAGAATTACAGACATTATTTG

97

-continued

GATGGTATCGAAAGAGCAGATTCTTTGTGTCATTGTCTCCACATAAGTGG
TTGTTGTCTTACTTAGATTGTTGTTGTTGTTGGGTTAAGTCACCATCT
TTATTGGTTAAGGCTTTATCAACTGATCCAGAATATTTGAAAAATCAA
CCATCAGAATCTAAGTCAGTTGTTGATTACAAGGATTGGCAAGTTGGT
ACAGGTAGAAGATTCAAAGCTTTGAGATTGTGGTTCGTTATGAGATCT
TATGGTGTGCAAATTTGCAATCACATATCAGAACTGATGTTCAAATG
GCTAAGATGTTCAAGGTTTCGTTAAGTCTGATCCAAGATTGCAAAAT
TTGGTCCAAGAGTTTTCTCTTTGGTTTGTTCAGATTGAACCCAATC
TCTGGTTCAGATCCAACGGTACAGAAGCTTTGAACAGAAAGTTATTG
GATTGGGTTAATTCTACAGGTAGAGTTTATATGACTCATACAAAAGTT
GGTGGTATCTATATGTTGAGATTGCTGTTGGTGCAACTTTAACAGAA
AAGAGACATGTTTCTTCAGCTTGAAGTTGATTAAAGAAGGTGCAGAT
GTTTTGTTGAAGGAAGATTAA

SEQ. ID NO: 57

MGSLDSNYDTESPAVSGQFNPLDPEEFRKQAHCIVDFIADYKNIESY
PVLQVDPGYRHSRLGKNAPYRSEPFESILKDVQKDIIPGMTHWMSPN
FFAHFPATVSSAAFVGEMLCTCFNSVGFNWLASPAATELEMVVIDWLA
NMLKLPKSFMSFGTGGGVLQGTSEAILCTLIAASPMHFEIVGVTST
SFVVYGSQDQTHSTYAKACKLAGILPCNIRSIPPTADSNFVSPLLLRR
AIEADKAAGMVPLYICATVGTSTTTAIDPLSSLADVANDYGVWFHVD
AYAGSACICEFRHYLDGIERADSLSLSPHKWLLSYLDCCCLWVKSPS
LLVKALSTDPEYLKNQPSSEKSVVDYKDQVGTGRRFKALRLWFVMRS
YGVANLQSHIRTDVQMAKMFEGFVKSDPRFEILVPRVFLVCFRLNPI
SGSDPTGTEALNRKLLDWNSTGRVYMTHTKVGGIYMLRFAVGATLTE
KRHVSSAWKLIKEGADVLLKED

SEQ. ID NO: 58

ATGGAAGGTGGTTTTACTGGTGGTGACGAATACCAAAAGCATTCTTG
CCAAGAGATTACTTAAACACATACTACTCATTCCAATCTGGTCCATCA
CCAGAAGCTGAAATGTTGAAGTTTAAATTTGGAATGTTGCATAAGACT
TTCGGTCCAGGTGGTTTGCAAGGTGACACTTTAATTGATATTGGTTCT
GGTCCAACAATCTATCAAGTTTTGGCTGCATGTGAATCTTTTAAAGAT
ATCACTTTGTGAGATTTTACAGATAGAAAATAGAGAAGAATTGGCTAAA
TGTTTGAAGAAAGAACAGGTGCATACGATTGGACTCCAGCTTTGAAG
TTCGCATGTGAATTAGAAGGCAACTCTGGTAGATGGCAAGAAAAAGCT
GAAAAGTTGAGAGCAACAGTTAAGAGAGTTTGAAGTGTGATGCTAAT
TTGTCAAATCCATTAACTCCAGTTGTTTGGCCACCAGCAGATTGTGTT
TTGACATTGTTAGCTATGGAATGTGCATGTGTTCTTTGGATGCTTAT
AGAGCTGCATTGAGAAAATTGGCATCATGTTAAACCAGGTGGTCAT
TTGGTTACTACAGTTACTTTGCAATTATCTTCATACATGGTTGGTGAA
AGAGAATTTCTTGTTGTGCTTTGGAAAAAGAAGATTGAACAAGCT
GTTTTAGATGCAGGTTTCGATATCGAACAATTGTTGTACTCTCCACAA

98

-continued

TCTTACTCAGCTTCTACAGCACCAAAATAGAGGTGTTTGTCTTCTGGTT
GCTAGAAAGAAAACCAGGTTCATAA

5 SEQ. ID NO: 59
MEGGFTGGDEYQKHFLPRDYLNITYYSFQSGSPPEAEMLKFNLECLHKT

FGPGGLQGDTLIDIGSGPTIYQVLAACESFKDITLSDFTDRNREELAK
WLKKEPGAYDWTPLKFACELEGNSGRWQEKAEKLRATVKRVLKCDAN
10 LSNPLTPVVLPPADCVLTLAMECACCSLDAYRAALRNLASLLKPGGH
LVTTVTQLQLSSYMGVEREFSCVALEKEEVEQAVLDAGFDIEQLLYSPQ
SYSASTAPNRGVCFLVARKKPGS

15 SEQ. ID NO: 60
ATGGATAAGATCTCTGCACCATTCTCTGGTACTTCACCAGCTGCA

GCTTCAGTTGCTGGTGTGATGAAGATGATAGATTGTGTTTCCAAGCT
CAAGAATTGATGTTTCGCATACAACATCTCTATGTTTTGAGAGCAGCT
20 ATTCAAATTAGTTTTGTTAGATGCTTTGTGAGCAGCTGGTGGTAAAGCA
TTGACTCCAAACGAATTAGTTGAAAACGTTGAAACATCTTCAAATAAG
GCTGAAGCAGCTGCAGCTGTTGATAGAATTTTGAGATATTTGTCTTGT
25 TTCAACGTTGTTACTTGTCTTCAGAAGCAGCTGGTCCAGATGGTACA
TTAGTTAGAAGATACACTACAGGTCCATTGTGTAGATGGTTAACTAAA

GATAGAGGTGACGGTACATTATCTCCATTGCTGTTTTTGTGTTGAT
30 CCAGATCATTGTTTCTTGGCATCATATTGCAGAAGCTGTACTGCT
GGTGGTCCATCAGCATTGAAAGAACACAAAAGTGGCCATACTACGAA
TACATGGGTAAAAATCAAAGATTGGGTACTTTGTTCGATAACGCAATG
35 GCTCAACATTCTGTTATTTTGGTTACAAAGATGTTAGAAAGATTCAAA
GGTTTTGATGGTGTTCAAAGATTAGTTGATGTTGGTGGTGGTACTGGT

40 TCTACATTGGGTATGATCACTTCAAAGTACAAGCACATGACAGGTATT
AATTACGATTTGCCACATGTTATTGCTCAAGGTTTGCCATTACCAGGT
GTTGAACATGTTGCAGGTGACATGTACGAATCTATTCCAACGTGTGAC

45 GCTGTTTTGTTACAATGGATTACATTGATGTTGAACGATGATGAATTC
GTTAAGATCTTGTCAAATGTGATAATGCTTTGCCAAAAGATGGTAAA
GTTATTGTTGTTGATGGTATTTTACCAGAAAAACCAGATTCTTCATTG
50 ACTGCAAGAGATGCTTTTACATTAGATATCATCATGTTTCGTTTTGTTT
AAAGGTGCAAGCAAGAACTGAAAAGGAATTGCAAGATTAGCTAAG
CAAGCAGGTTTTTACTGGTGGTATTAGAAAACTTATATTTCTTTAAT

55 TTTTACGCTTTTGAATTCATAAATAA

SEQ. ID NO: 61

MDKISAPFFSGTSPAAASVAGVDEDDRLCFQAQELMFAYNISMLRLAA
IQLGLLDALSAAGGKALTPNELVENVETSSNKAEEAAAVDRLILRYLSC
60 FNVVTCSSAAGPDGTLVRRYTTGPLCRWLTKDRGDGTLSPFAVFVVD
PDHLPFWHHIAEAVTAGGPSAFERTQKWPYEYMGKNQRLGTLFDNAM
AQHSVILVTKMLERFKGFDGVQRLVDVGGGTGSLGMITSKYKHMTGI
65 NYDLPHVIAQGLPLPGVEHVAGDMYESIPTGDAVLLQWITLMLNDEF

99

-continued

VKILSNCHNALPKDGKIVVVDGILPENPDSSLTARDAFTLDIIMFVLF
 KGAKQRTEKEFARLAKQAGFTGGIKKTYIFFNFYALEFTK
 SEQ. ID NO: 62
 ATGAACACCTTCAGAACAGCCACTGCCAGAGACATACCTGATGTAGCA
 GCAACTCTTACGGAAGCCTTCGCAACTGATCCACCCACGCAGTGGGTG
 TTCCCCGACGGTACTGCCGCCGTGAGCAGGTTCTTTACACATGTTGCA
 GATAGGGTTACACGGCCGGTGGTATTGTTGAGCTACTACCAGACAGA
 GCCGCCATGATTGCATTGCCACCACACGTGAGGCTGCCAGGAGAAGCT
 GCCGACGGAAGGCAGGCGGAAATTCAGAGAAGGCTGGCAGACAGGCAC
 CCGCTGACACCTCACTACTACCTGCTGTTTTACGGAGTTAGAACGGCA
 CACCAGGTTTCGGGATTGGCGGAAGAATGCTGGCCAGATTAACTAGC
 AGAGCTGATAGGACAGGTTGGGTACATATACTGAGGCATCCACCTGG
 CGTGGCGCTAGACTGATGCTGAGACATGGATTCCATGCTACAAGGCCA
 CTAAGATTGCCAGATGGACCCAGCATGTTTCCACTTTGGAGAGATCCA
 ATCCATGATCATTCTGATTAG
 SEQ. ID NO: 63
 MNTFRTATARDIPDVAATLLEAFATDPPTQWVFPDGTAAVSRFFTHVA
 DRVHTAGGIVELLPDRAAMIALPPHVRLPGEAADGRQAEIQRRLLADRH
 PLTPHYLLFYGVRTAHQSSGLGRMLARLTSRADRDVGTYTEASTW
 RGARLMLRHGFHATRPLRLPDGSPMFPLWRDPIHDHSD
 SEQ. ID NO: 64
 ATGAAAGGATATTTTCGGACCATACGGTGGCCAGTACGTACCAGAAATA
 TTAATGGGTGCCTTAGAGGAGTTAGAGGCAGCATACGAGGAGATTATG
 AAGGATGAGAGCTTCTGGAAGGAGTTCAACGATCTACTGAGGATTAC
 GCAGGCAGACCAACGCCATTGTACTTTGCCAGGAGATTGTCTGAGAAG
 TACGGCGCCCGTGTTTACTTGAAGCGTGAGGATCTGCTGCACACTGGA
 GCACACAAGATAAATAACGCTATCGGACAGGTTTTATTGGCCAAATTA
 ATGGGGAAGACACGTATCATAGCCGAGACGGGAGCTGGCAGCATGGA
 GTCGCTACTGCTACCGCTGCTGCCCTGTTTCGGAATGGAATGTGTGATC
 TACATGGGTGAAGAGGACACAATCAGACAGAAGTTGAACGTGGAGCGT
 ATGAAATTATTAGGGGCTAAAGTTGTCCCTGTTAAGTCTGGCAGTAGG
 ACCTTGAAGGATGCGATAGACGAGGCTTTGAGAGACTGGATTACTAAT
 TTACAGACAACATATTATGTTATCGGATCTGTTGTTGGTCCCCACCT
 TACCAATTATCGTAAGGAATTTCCAGAAGGTTATCGGTGAGGAGACC
 AAGAAGCAAATACCAGAAAAGGAAGGTCGTTTGCCAGACTATATAGTT
 GCCTGCGTAGGCGGCGGTAGCAATGCCGAGGTATATTTTACCCATT
 ATAGACTCTGGAGTAAAGCTGATAGGTGTTGAGGCAGGTGGCAGGGA
 TTGGAGACAGGTAACACGCAGCCTCGTTATTAAAGGGTAAATTTGGC
 TATTTACATGGATCGAAGACCTTTGTTCTACAAGATGACTGGGGTCAA
 GTCCAAGTGAGCATTTCGGTGTGAGCTGGTCTTGACTATTGAGGAGTA
 GGACCTGAGCATGCTTATTGGAGAGAGACAGGAAGGTTCTGTACGAC
 GCAGTGACTGACGAAGAGGCTTTGGACGCATTTATAGAGTTATCAAGA

100

-continued

CTAGAGGGCATTATACCCGCTTTAGAGTCATCGCATGCTCTAGCATAT
 5 TTGAAGAAGATAAATATAAAAAGGTAAGGTTGTGGTGGTCAACCTATCA
 GGGAGAGGGGATAAAGACCTGGAGTCAGTCTTAAACCATCCATACGTG
 AGAGAAAGAATTAGATGA
 SEQ. ID NO: 65
 10 MKGYFGPYGGQYVPEILMGALEEELEAAYEIMKDESFWKEFNLLRDY
 AGRPTPLYFARRLSEKYGARVYLKREDDLHTGAHKINNAIGQVLLAKL
 MGKTRIIAETGAGQHGVATATAALFGMECVIYMGEEDTIRQKLNVER
 15 MKLLGAKVVPVKSGSRTLKDAIDEALRDWITNLQTTYVIGSVVGP
 YPIIVRNQKVIQEETKKQIPEKEGRLPDYIVACVGGGSNAAGIFYPF
 IDSGVKLIGVEAGGEGLETGKHAASLLKGKIGLYHGSKTFLQDDWQG
 20 VQVSHSVSAGLDYSGVGPHEYWRETGKVLDAVDEEALDAFIELSR
 LEGIIPALESSHALAYLKKINIKGKVVVNLSSRGDKDLESVLNHPYV
 RERIR
 SEQ. ID NO: 66
 25 ATGTGGTTCGGTGAGTTTGGTGGACAATATGTGCCAGAGACTTTAGTG
 GGTCTCTTAAGGAATTGGAAAAGGCATATAAAAGGTTCAAGGACGAT
 GAGGAGTTCAACAGGCAACTAACTATTATTGAAGACATGGGCCGGT
 30 AGACCAACGCCCTTGTTATTGCTAAGAGGTTAACTGAAAAGATTGGC
 GGCGCGAAAAGTGATCTGAAAAGAGAAGACCTAGTTCATGGTGGAGCA
 CACAAGACAAATAATGCCATTGGACAAGCACTATTGGCAAAGCTAATG
 35 GGTAAACTAGATTGATAGCTGAGACAGGAGCGGGTCAACATGGGGTC
 GCGACAGCGATGGCTGGTGCACTACTGGGGATGAAGGTAGATATTTAC
 ATGGGTGCTGAGGACGTTGAGCGTCAGAACTAAATGTCTTCAGGATG
 40 AAGCTATTAGGTGCCAATGTTATACCTGTAAATTCTGGCTCAAGAACA
 CTAAGGACGCCCTTCGACGAGGCTCTTAGAGACTGGGTGGCACTTTTC
 GAGTATACTCATTACTTGATCGGTTCACTGGTGGACCATCCATAC
 45 CCAACCATCGTTAGGGACTTTTCAGAGCGTGATTGGTAGAGAGGCTAAG
 GCACAGATCTTAGAAGCAGAGGGACAGCTACCTGACGTCATAGTTGCC
 TCGCTCGGCGGTGGCTCTAACGCAATGGGTATATCTATCCATTCTGT
 50 AATGACAAGAAGGTTAAATTAGTAGGAGTCGAAGCTGGCGGAAAGGGG
 TTAGAGTCGGGTAAACACTCAGCAAGCTTAAATGCAGGACAGGTAGGG
 GTGTCCACGGCATGTTGTGCTATTTCTTGCAAGACGAGGAAGGTGAG
 55 ATAAAGCCAAGTCATTCAATTGCTCCAGGCCTTGACCACCCCGGTGTT
 GGTCCAGAGCACGCTTACTTAAAGAAGATTCAAAGGGCCGAGTACGTC
 GCTGTAACAGACGAAGAGGCATTGAAAGCTTTCCATGAGCTATCCAGA
 60 ACTGAGGGGATTATACCCGCCCTTGAGTCTGCCCATGCTGTGGCGTAC
 GCCATGAAGTTAGCTAAAGAGATGTCCCGTGACGAAATCATCATTGTA
 AATCTATCAGGGAGAGGAGACAAGGATTGGACATTGTATTGAAGGCA
 65 AGCGGAAATGTTTGA

101

-continued

SEQ. ID NO: 67
MWFGEFGGQYVPETLVGPLKELEKAYKRFKDDEEFNRQLNYYLKTWAG
RPTPLYAKRLTEKIGGAKVYLKREDLVHGGAHKTNNAIGQALLAKLM
GKTRLIAETGAGQHGVATAMAGALLGMKVDIYMGAEDVERQKLNVFRM
KLLGANVIVPNSGSRTLKDAFDEALRDWVATFEYTHYLIGSVVGGPHPY
PTIVRDFQSVIGREAKAQILEAEGQLPDVIVACVGGGSNAMGIYFPFV
NDKKVKLVGVEAGGKLGESGKHSASLNAGQVGVSHGMLS YFLQDEEGQ
IKPSHSIAPGLDHPGVGPEHAYLKKIQRAEYVAVTDEEALKAFHELRSR
TEGIIPALESAAHAYAMKLAKEMSRDEIIIVNLSGRGDKDLDIVLKA
SGNV

SEQ. ID NO: 68
ATGTGGTTTGGGGAATTCGGTGGTCAATACGTGCCAGAAACCTTAGTT
GGGCCCTTTAAGGAGCTTGAGAAAGCTTACAAAAGATTTAAGGACGAC
GAGGAGTTTAAATAGCAATTAACTATTACTTGAAGACTTGGGCAGGT
AGACCAACACCACTTTATTACGCGAAGAGACTTACAGAGAAGATCGGT
GGCGCTAAGSTTTATTTGAAGCGTGAGGATTTGGTTCATGGTGGAGCA
CATAAGACAAACAATGCCATCGGGCAAGCCTTGTGGCCAAATTTGATG
GGTAAGACTAGGTTGATAGCTGGCACAGGTGCCGGCCAGCATGGCGTT
GCAACAGCAATGGCTGGTGCTCTTTTAGGGATGAAGGTGGATATCTAT
ATGGGCGCCGAGGACGTTGAAAGACAAAAGCTTAACGTATTGAGAAATG
AAGTTATTAGGTGCGAATGTTATTCCAGTGAATTCAGGTTCCAGGACG
CTGAAGGACGCAATAGACGAAGCCTTGAGAGACTGGGTGCTACTTTT
GAGTACACACATTATTTGATTGGTTCGGTGGTGGACCCACCCGTAC
CCTACAATAGTTAGAGATTTTCAATCTGTTATCGGTAGAGAGCGAAG
GCTCAAATTCAGAGGCCGAGGGACAGCTACCCGATGTTATAGTCGCC
TGCGTTGGAGGAGGATCAAATGCGATGGGAATCTTCTACCCCTTTTGTA
AATGATAAGAAAGTTAAGTTAGTCGGAGTTGAGGCCGGTGGTAAAGGG
CTAGAAAGCGGCAACACTCGGCATCTCTTAATGCCGGTCAAGTAGGA
GTAAGTCACGGGATGTTATCTTACTTCTGCAGGATGAGGAAGGTCAA
ATTAAGCCATCTCATTCTATAGCCCCAGGATTAGACTATCCCGGTGTT
GGGCTGAGCACGCTTATCTGAAGAAGATACAGCGTGCTGAGTACGTT
GCAGTAACGTATGAAGAGGCTTTGAAGGCTTTTCACGAGTTATCAAGA
ACTGAGGGAATAATTCCTGCCTTGAATCGGCACATGCGGTTGCCTAC
GCCATGAAGCTTGCTAAGGAATGTCAAGAGACGAAATCATTATAGTG
AATTTATCGGGAAGGGCGATAAGGACTTGACATTTGTTTTAAAGCA
AGTGGCAATGTATAA

SEQ. ID NO: 69
MWFGEFGGQYVPETLVGPLKELEKAYKRFKDDEEFNRQLNYYLKTWAG
RPTPLYAKRLTEKIGGAKVYLKREDLVHGGAHKTNNAIGQALLAKLM
GKTRLIAETGAGQHGVATAMAGALLGMKVDIYMGAEDVERQKLNVFRM
KLLGANVIVPNSGSRTLKDAIDEALRDWVATFEYTHYLIGSVVGGPHPY
PTIVRDFQSVIGREAKAQIPEAEGQLPDVIVACVGGGSNAMGIYFPFV

102

-continued

NDKKVKLVGVEAGGKLGESGKHSASLNAGQVGVSHGMLS YFLQDEEGQ
IKPSHSIAPGLDYPGVGPEHAYLKKIQRAEYVAVTDEEALKAFHELRSR
TEGIIPALESAAHAYAMKLAKEMSRDEIIIVNLSGRGDKDLDIVLKA
SGNV

SEQ. ID NO: 70
10 ATGTGGTTTGGTGAGTTTGGTGGTCAATACGTGCCTGAGACGCTAGTT
GGGCCACTGAAGGAGTTAGAGAAAGCGTACAAGAGATTCAAAGATGAC
GAGGAGTTTAAACCGTCAGTTGAATTATTATCTTAAGACGTGGGCTGGT
15 AGGCCGACCCCTCTATACTACGCTAAGCGTTTAAACGAAAAGATAGGT
GGCGCCAAGATCTATCTGAAGAGAGAGGACCTAGTCCACGGCGGGGCT
CACAAAACCTAACAACGCAATTGGTCAGGCACCATTAGCAAAGTTAATG
GGCAAAACACGCTCTAATTGCCGAACTGGTGCTGGTCAACATGGAGTT
20 GCTACGGCAATGGCGGGCGCACTGTTAGGTATGAAAGTGGATATCTAT
ATGGGAGCAGAAGACGTGGAGAGACAGAAGATGAACGTGTTCCGTATG
AAACTATTGGGAGCTAATGTTATCCCGGTTAACTCTGGCTCAAGAACT
25 GCAAAGGACGCCATAAAGCAAGCTTTGAGGGATTGGGAAGCCACATTC
GAGTACACACACTACCTTATAGGCTCAGTCGTAGGCCCGCATCCCTAC
CCCACCATAGTTCTGTGACTTCCAGAGTGTTATTGGCCGTGAGGCCAAG
GCACAGATCTTGAAGCTGAAGGTCAATTGCCTGACGTAATAGTTGCA
TGTGTGGGTGGCGGTTCTAACGCTATGGGTATATTCTACCCCTTCGTG
35 AACGATAAGAAAGTAAAGCTAGTTGGAGTGGAAGCAGGTGGCAAAGGG
TTGGAGTCGGGCAAGCACTCAGCCTCATTAACGCAAGTCAGGTAGGC
GTTCTGCACGGTATGTTATCTATTTCTTACAGGACGAGGAAGGCCAA
ATCAAACCATCGCACTCAATCGCACCTGGTTTAGATTACCCAGGAGTC
40 GGTCCCGAGCACGCATATTTAAAGAAAATCCAGAGACGAGTACGTC
ACGGTGACGGATGAGGAAGCCCTTAAAGCGTTTCACGAGCTGTGAGG
ACCGAGGGCATAATACCTGCTCTTGAAGTGCCACGCAAGTGCCTAT
45 GCTATGAAGTTGGCCAAGGAGATGAGTAGAGACGAGATTATCATAGTA
AATCTGTCAGGCAGAGGTGACAAGGATCTTGACATAGTCCTTAAGGCG
50 AGCGGTAAATGTTTTGGAGTAA

SEQ. ID NO: 71
MWFGEFGGQYVPETLVGPLKELEKAYKRFKDDEEFNRQLNYYLKTWAG
RPTPLYAKRLTEKIGGAKIYKREDLVHGGAHKTNNAIGQAPLAKLM
GKTRLIAETGAGQHGVATAMAGALLGMKVDIYMGAEDVERQKMNVFRM
KLLGANVIVPNSGSRTAKDAINEALRDWEATFEYTHYLIGSVVGGPHPY
PTIVRDFQSVIGREAKAQILEAEGQLPDVIVACVGGGSNAMGIYFPFV
60 NDKKVKLVGVEAGGKLGESGKHSASLNAGQVGVSHGMLS YFLQDEEGQ
IKPSHSIAPGLDYPGVGPEHAYLKKIQRAEYVTVTDEEALKAFHELRSR
TEGIIPALESAAHAYAMKLAKEMSRDEIIIVNLSGRGDKDLDIVLKA
SGNVLE

103

-continued

SEQ. ID NO: 72
 ATGTGGTTCGGAGAGTTCGGCGGGCAGTACGTCCAGAAACGCTAGTC
 GGCCCACTGAAGGAGCTGGAGAAAGCATACAAAAGATTCAAGGATGAC
 GAGGAGTTCAACAGGCAATTAACTATTACCTAAGACCTGGGCTGGT
 AGGCCACCCCTTTATACTATGCAAAACGTTTACGAGAGAAATCGGT
 GGCGCAAAAGTGACTTGAAGAGAGAAGACTTAGTCCACGGTGGTGCT
 CATAAACTAACAACGCTATAGGGCAGGCACTTTTAGCTAAGTTGATG
 GGTAAACGAGGTTGATTGCGGGTACCGGTGCCGGTCAACATGCGGTT
 GCTACGGCAATGGCGGGAGCCCTATTGGGCATGAAGGTTGACATTTAT
 ATGGGCGCTGAGGACGTGGAACGTCAGAACTGAACGTCTATCGTATG
 AAATTGCTTGGTGCCAACGTTATACCACTCAATCCGGCTCGCGTACG
 GTTAAAGACGCTTTTCGATGAGGCATATGTGATAGAGTTGCCACTTTC
 GAATACACCCACTATCTTATTGGTACTGTCTGGGGTCCACATCCGTAT
 CCAACTATTGTAAGGGAAGTTCAGACTGTTATCGGAAGAGAGGCAAAG
 GCCCAATATTAGAGGCTGAGGGAAGGTTACCCGACGCCATTGTTGCA
 TCGGTAGGCGGTGGTAGTAACGCAATGGGTATTTCTATCCCTTCGTT
 AACGATAAGAAGGTTAAGTTAGTGGGTGTCGAGGCCGGTGGTAAGGGG
 CTTGAGTCCGGTAAACATTCCGCATCTTTGAATGCTGGGCAAGTAGGT
 GTCAGTCATGGAATGTTATCATACTTTCTACAGGACGAGGAAGGACAG
 ATAAACCTTCTCATTCTATAGCGCCAGGTCTTGACCACCTGGCGTT
 GGACCAGAGCATGCTTATTGAAGAAGATCCAGAGAGCTGAATATGTA
 GCTGTTACGGATGAGGAAGCTCTAAAGGCGTTCCACGAGTTAAGTAGG
 ACTGAGGGAATTATACCGCATTAGAGTCTGCCACGCTGTAGCTTAC
 GCAATGAAGCTGGCCAAAGAGATGAGCCGTGACGAGATCATCATCGTA
 AACTTGTCTGGCAGAGGCGATAAGGACCTAGACATCGTATTAAAGGCC
 TCGGGGAACGCTCTAGAGTAG

SEQ. ID NO: 73
 MWFGEFGGQYVPETLVGPLKELEKAYKRFKDEEFNRQLNYYLKTWAG
 RPTPLYYAKRLTEKIGGAKVYLKREDLVHGGAHKTNNAIGQALLAKLM
 GKTRLIAGTGAGQHG VATAMAGALLGMKVDIYMGAE DVERQKLN VYRM
 KLLGANV I PVNSGSR TVKDAFDEALCDRVATFEYTHYLIGTVWGHPHY
 PTIVRDFQTVIGREAKA QILEA EGRLPDAIVACVGGGSNAMGI FYPFV
 NDKKVKLVGVEAGGKGLES GKHSASLNAGQVGVSHGMLS YFLQDEEGQ
 IKPSHSIAPGLDHPGVGPEHAYLKKIQRAEYVAVTDEEALKAFHEL SR
 TEGI I PALESAHAVAYAMKLAKEMSRDEI I IVNLSGRGDKDL DIVLKA
 SGNVLE

SEQ. ID NO: 74
 GTATCCGGCTGTTCCCTTCATAGCCCTTTCAATGAACGTTGCAGCCCTT
 TGAAGATTGGCCATTTTGTGAGGACTCGAGCCTGACAGTTGGACCAAC
 GCAACTTTAATTTTTTGTGAAAGAAATCTTGAAGCACTCATACTGGCG
 ATCTTCACGCCCTCCTGCTATTACAAAAGCTGTGTTTTTACAAGAATC
 AAATTAAGTTAGCAAGATATTATACAACATTATTGATAATTTCAATAT

104

-continued

CGTGTTCTGACCTGATGACGTATCTGTGCATTGATAAGGCCCGCATGG
 TTTTCAGAAAGCAGAGCGGAACGATTCCAAATTAGTGGCCTTGTGCTTT
 GCATGTCAATTGTGTACCTTCAGCTCGTGGATTGTGTTTTATCAATAC
 ACAGTCTACAGTCAAGAATTTTTTTTATCAAATTTTGC GTTCGAGCGT
 ATAAAAATAGCCGCTGTAGCTACTTAAGTTCCTGTTCAGCGATAGTTTT
 TTTCCATCACACGTACTATGGCAATTAAGTCCTCAGCGAGCTCGCATG
 GAATGCGTGCATGAGCGACCTCATGCTATACCTGAGAAAGCAACCTG
 ACCTACAGGAAAGAGTTACTCAAGAATAAGAATTTTCGTTTTAAACC
 TAAGAGTCACTTTAAATTTGTATACACTTATTTTTTTTATAACTTAT
 TTAATAATAAAAATCATAAATCATAAGAAATTCGCTTATTTAGAAGTG
 TCAACAACGTATCTACCAACGGAATGCGTGCATTTCAGGCGTAGTCTG
 GAACGTCGTATGGGTATGGACCAGAGACGGATTGAGAACTGTTCGG
 TTCTTGGGACGAATCTACATTGGAATGGAAC TGGGTGCTTGAAGAAAC
 CAGTGGTGAAATCAGCTGGGATGTCAATTGGCTTACC GTTTTGGTCAA
 CTGCTCTTTCATGTTGAAAGCAGACA CAAAGTAGCACC GGCAATCC
 AAACAGTAGATTGAGCTAGGTGAATACCTGGG CAGTTCTCTTACC GT
 AACC GAAAGCAGCCTTTCTTGGGTCTCTAACAGTGTGTCTGGTTTAC
 CATCAGGACCC AAGTATCTTTCTGGACGAAAACGGATGGATCTGGAT
 AGACTTCTGGGTCAATCAAAC TGGCCAGGTGTTAGCAAAAACCAATG
 TGTTCTTTGGAATCAAATAACCTCTGTAAACATCATCCTTCATCAATT
 TATGAGGGATGGCTAATGGAGCAATTTGGTTCATCTGAATAATTCTCT
 TGATACAAGCGGTCAAATATGGGAGTGAGTCGTCTTCTCATCGTAGT
 CAGGGATTTGACCGTTGTTGGTCAAAGCATCCAATT CAGCTTGAACCT
 TTCTTTGCACTTCTGGGTATTTAACCATGGCCAATATGAAAGCGGACA
 TAGCAGAGACGGTAGTGTGCGCCACCACCAAGTTAACTTCAGCAGCAG
 TGTTCTTGATAACATGTTCTTGGTGTCTTAAATCACC GTTCAAGTCCA
 TAGCTTGTAACTAGCAGAAGCGTAAGATGGTCTGGTTAGACCTTGTG
 GAGCCAACCTTCTCATAGTTTCGTATGGCATGTCAACCATGTGGTCAG
 CGGCTTCTCTCCAGACCTTGGCCTTACGCTTGAAGACGGCACCTGGGA
 ACCAAGCTGGCAAGTACTTCAAAGATGGGAAGGAGTCAACCAGAACT
 TACCTGGAACACTAGCAATGGCCAAACCTTCGTTGGCTAAGTGGGTAG
 CTTCTAACCATGGGT CATCTTCGGCTAGATCGATACCGTAACCAATGT
 CCAAAGACATAGCAGCGATTTGATGTCTGATGTGTTGAGCCCACCGGT
 CCGGAGTCTTGGTCAATTGTTGAACCAATTTGGTGAGCAGCCTTGACTT
 GAGCGTGTCTGAATTGCTTGATACCCCTTTTCAGAGAATTCCTTAGCAA
 ACATTCTTCTTTCTTCTCTCCAACGGTCACCGTAAGTGATAAAACCCA
 GATCAAATTTCCCAACCCATCAATTCAATTGACCATAGTGGATTCCAATC
 TACCGGAGTAGATGGAACCACGTTTTTCCAAGAGATCCGTGATAGTTT
 CCAAGGTGTTTTAAATGACCATTTTCGTTACCACCGATCTACGTACA
 AGATATCAGTGTGTAGTCACGACCCCATTGCAAGAAAGTCAACCATG

105

-continued

GAGATTCTTCTGGCATGTCGAACATGTTACCAATAAATGGGATTGGAA
 TTCCTGGTGGACCAGGTGGCAATCTGGATCTACGAACCTCTCTGGAGA
 CGATGTAGTAAATACAACCAGCAATGACGAAAGAGAACAAGACAGCAA
 TCATTGTTTTATATTTGTTGTAAGGATAGATAATTACTTCCTTGATG
 ATCTGTAAAAAGAGAAAAAGAACATCTAAGAACTTGAAAACTAC
 GAATTAGAAAAGACAAATATGTATTTCTTGCAATGACCAATTTATGC
 AAGTTTATATATATGTAATGTAAGTTTCACGAGGTTCTACTAACTA
 AACCACCCCTTGGTTAGAAGAAAAGAGTGTGTGAGAACAGGCTGTTG
 TTGTACACGATTTCGACAATCTGTGTTGAAAGAGAGAGATAACAGT
 ACGATCGAACGAACCTTGTCTGGAGATCACAGTGGGCATCATAGCAT
 GTGGTACTAAACCTTTCCCGCCATTCCAGAACCTTCGATTGCTTGTT
 ACAAACCTGTGAGCCGTCGCTAGGACCTTGTGTGTGACGAAATGG
 AAGCTGCAATCAATAGGAAGACAGGAAGTCGAGCGTGTCTGGGTTTTT
 TCAGTTTTGTTCTTTTTCGAAACAACAGTTTATTCTTGGCATCCACTA
 AATATAATGGAGCCCGCTTTTAAAGCTGGCATCCAGAAAAAAGAGAA
 TCCCAGCACCAAAATATGTTTTCTTACCAACCATCAGTTCATAGGT
 CCATTCTCTTAGCGCAACTACAGAGAACAGGGGCACAACAGGCAGAAA
 AACGGGCACAACCTCAATGGAGTGATGCAACCTGCCTGGAGTAAATGA
 TGACACAAGGCAATTGACCCACGCATGTATCTATCTCATTTTCTTACA
 CCTTCTATTACCTTCTGCTCTCTCTGATTTGGAAAAAGCTGAAAAAA
 AGGTTGAAACAGTTCCCTGAAATATTCCCTACTTGACTAATAAGT
 ATATAAGACGGTAGGTATTGATTGTAATCTGTAAATCTATTTCTTA
 AACTTCTTAAATTTACTTTTATAGTTAGTCTTTTTTTTAGTTTTTAA
 ACACCAAGAACTTAGTTTTGCAATAAACACACATAAACAAACAAATGG
 CTTCTAGTTCTTCCGATGTCTTCGTTTTGGGTCTAGGTGTTGTTTTGG
 CTGCCTTGTATATCTTCAGAGACCAATATTTCGCTGCTTCTAAGCCAA
 AGGTGGCTCCAGTTTCCACTACGAAGCCTGCCAACGGTTCGCTAACC
 CAAGAGACTTCATCGCCAAGATGAAACAAGGTAAGAAGAGAATCGTAA
 TCTTCTACGGTTCTCAAACCTGACCGCTGAAGAATATGCTATTTCGTT
 TGGCTAAGGAAGCTAAGCAAAAGTTTCGGTCTAGCCTCCTTGTTTGTG
 ATCCAGAAGAATACGATTTTGAAAGTTGGACCAATTGCCAGAAGATT
 CTATTGCTTTCTTCGTCGTGTCTACCTATGGTGAAGGTGAACCTACAG
 ACAACGCTGTCCAATTGTTGCAAACTTGCAAGATGAAAGCTTCGAAT
 TCTCCTCTGGTGAGAGAAAGTTGTGAGGTTTGAAGTACGTTGTTTTG
 GTCTGGGTAAACAGACCTACGAACATTACAACCTCATTGGGAGAACTG
 TTGACGCTCAATTGGCCAAGATGGGTGCTATCAGAATCGGTGAAAGAG
 GTGAAGGTGATGATGACAAGTCCATGGAAGAAGACTACTTGAATGGA
 AGGATGGTATGTGGGAAGCGTTTGCCACTGCTATGGGTGTTGAAGAAG
 GTCAAGGTGGTGACTCCGCTGATTTTCGTCGTTTCCGAATTGGAATCTC
 ACCCACCAGAAAAGGTTTACCAAGGTGAATTTTCTGCTAGAGCTTTAA
 CCAAAACCAAGGTTATTACGACGCTAAGAATCCTTTTGTCTGCTCCAA

106

-continued

TTGCGGTTGCTAGAGAATTGTTCCAATCTGTTGTCGATAGAAAAGTGT
 TCCACGTCGAATTCACATTGAAGGCTCTGGTATCACCTATCAACACG
 5 GTGACCACGTTGGTTTGTGGCCATTGAATCCAGATGTTGAAGTCGAAC
 GGTTGTGTGTGTTTTAGGTTTAGCTGAAAAGAGAGATGCTGTCATCT
 CCATTGAATCCTTAGACCCGGCTTTGGCTAAGGTTCCATTCCCAGTCC
 10 CAACTACTTACGGTGTGTGTGAGACACTACATTGACATCTCTGCTG
 TCGCCGCTAGACAAATCTGGGTACTTTGTCCAAATTCGCTCCAACCC
 CAGAAGCTGAAGCTTTCTTGAGAACTTGAACACTAACAAGGAAGAAT
 15 ACCACAACGTCGTCGCTAACGGTTGTTTGAATTTGGGTGAAATTTTGC
 AAATCGCTACCGGTAACGACATTACTGTCCCACTACTGCTCAACA
 CCACCAATGGCCAATTCCATTGACATCATTGTTTTCTGCCATCCCAA
 20 GATTGCAACCAAGATACTACTCTATCTCTTCTTCCCCAAAATTCATC
 CAAACACCATCCACGCTACCGTTGTTGTGCTCAAATACGAAAACGTTT
 CAACCGAACCAATCCCAAGAAAGTGGGTTTACGGTGTGCGTAGTAAT
 25 TCTTGTGAATTTAAAGTACGCTGTTAAACAAGGAACAGTTCATACA
 TCACTCAAAATGGCGAACAAAGAGTCGGTGTCCCGGAATACTTGATTG
 CTGGTCCACGTGGTTCTTACAAGACTGAATCTTCTACAAGGCTCCAA
 30 TCCATGTTAGACGTTCTACTTTCCGTTTGCCAACCAACCCAAAGTCTC
 CAGTCATCATGATTGGTCCAGGTACTGGTGTGCCCCATTGAGAGGCT
 TCGTTCAAGAAAGAGTTGCCCTGGCCAGAAGATCCATCGAAAAGAACG
 35 GTCTGACTCTTTGGCTGACTGGGTCGTATTTCTCTGTTCTACGGTT
 GTAGAAGATCCGACGAAGACTTCTGTGACAAGGACGAATGGCCACAAT
 ACGAAGCTGAGTTGAAGGTAAGTTCAAGTTGCACTGTGCTTTCTCCA
 40 GACAAAACCTACAAGCCAGACGTTCTAAGATTACGTCCAAGATTGTA
 TCTGGGAAGACAGAGAACACATTGCCGATGCCATCTTAAACGGTAAGG
 GTTACGCTCTACATCTGCGGTGAAGCTAAGTCCATGTCTAAACAAGTTG
 45 AAGAAGTTCTAGCCAAGATCTTGGGCGAAGCCAAAGGTGGTTCCGGTC
 CAGTTGAAGGTGTTGCTGAAGTCAAGTTACTGAAGGAACGGTCCAGAT
 TGATGTTGGATGTCTGGTCTGAACAAAAGTTAATTTCTGAAGAAGATT
 TGGAATGAATCGCGTGCATTATCCGCTCTAACCGAAAAGGAAGGAT
 50 TAGACAACCTGAAGTCTAGGTCCCTATTTATTTTTTATAGTTATGTT
 AGTATTAAGAACGTTATTTATATTTCAAATTTTTCTTTTTTTCTGTA
 CAGACGCGTGTACGCATGTAACATTATACTGAAAACCTTGCTTGAGAA
 55 GGTTTTTGGGACGCTCGAAGATCGCGTCCCAATTGCGCCTATAGTGAGT
 CGTATTACGCGCGCTCACTGGCCGTCGTTTTACAACGTCGTGACTGGG
 AAAACCTGGCGTTACCCCTGCAGGACTAGTGTGAGGCATTAATAGT
 60 TTACTCAATTCTGAAGCCAATTTGTACAATTTCCCATTAGAGTCAAA
 TAAAAGGATGCCTCACGAGGTATGTTACCCGCGCTATTTACATGGC
 TCATTGAATTAGAGGTGGAATTTGGTGTACCTCCCTCCTCATCTGA
 65 TGAAGTAGTGATCCGACAATCTTAAAGTTGTAGACATTACTTTTAC

107

-continued

CACCAACTAAGTTGTATTTATATTGCTACCCCTATCCTTTTATATCTA
 ACTAGCGCTCATAAGGTGGGGCAATACTAAACGTGTCTTATTCA
 ACTCATTAATACGTGGCAGTACGTACCCCTATTAGAAACAATAGGAAA
 CAGCAGAGTCGGAAGAAGCCAAATGCCAGATTGGAAGTCCAAACCTT
 GTCAAGCCAATCTTTGGGAGCGGTATTCCCTCCAGAAATTGTGTACCA
 AATACTTACATACCAGTTTAGGGATTGTGTAAGAAATGACCATCCAGG
 TACGGCAGAAA

SEQ. ID NO: 75

CAATCTGGCGCTTGAGTTCTCAACATGTTTTATTTTTACTTATATT
 GCTGGTAGGGTAAAAAATATAACTCCTAGGAATAGGTTGTCTATATG
 TTTTGTCTTGCTTCTATAATTGTAACAAACAAGGAAAGGAAAAATAC
 TGGGTGTAAAAGCCATTGAGTCAAGTTAGGTCATCCCTTTTATACAAA
 ATTTTCAATTTTTTTTCCAAGATTCTGTACGATTAATTATTTTTTT
 TTTGCGTCTACAGCGTGATGAAAATTTGCGCTGTGCAAGATGAGCG
 GGAACGGGCGAAATGTGCACGCGCACAACCTTACGAAACGCGGATGAGT
 CACTGACAGCCACCGCAGAGTTCTGACTCTACTGAGCTCTATTGGA
 GGTGGCAGAACCGGTACCGGAGGAGGCCGCTATAACCGGTTGAATTT
 ATTGTACAGTGTACATCAGCATTAAAGTCCTCAGCGAGCTCGCATGG
 AATGCGTGCATGAGCGACCTCATGCTATACCTGAGAAAGCAACCTGA
 CCTACAGGAAAGAGTTACTCAAGAATAAGAATTTTCGTTTTAAACCT
 AAGAGTCACTTTAAATTTGTATACACTTATTTTTTTTATAACTTATT
 TAATAATAAAAAATCATAATCATAAGAAATTCGCTTATTTAGAAGTGT
 CAACAACGTATCTACCAACGGAATGCGTGCGATTGAGTGGAGTCCAA
 ACCCAACAAAGGATTTGGAATTGGCTTACCGGCAGTGGAGGATTCCTT
 TAGCAACGTGGAAGTGATTTACCGGTTGTCGTTGTACCTCTAGCATC
 GTGGAAGCAGCAACACCTTCTTAACAAAGTTGATTCTTTCTTCTTC
 AGAACCCCATTGCTGAAGTCAGTCCACATAACAATGTGAGCGCGAT
 ACCAGCGGTAACCTTGGCGTAGTTGATGGAATGCTTGAAGTACGGGC
 GTAAGATTGCAAGTAAGCTTGCTCATGGTTGTACCAACTTGTTCGTC
 TTGGAATCTGCTAATCAAGTAACAGTCACCCAAGAAGTAACCCAAATC
 CAATGAAGCTGGACCGTACTTACACAATCCCAGTCTAAGATGTAGAT
 CTTTTGCAACTTAGATGGGTTACCTTCTCAAGTTGCAACAAGATGTT
 CCCAGACCACAAGTCAGCCATGACCAAAGTTTCTTCGGAGTGCATAAC
 ATCGTCAACTAGATCCTTGACAACAGTTGGCAACAATGGATCATCGAC
 GCCGTATTTAGCGCGTTTGGGATAATAGTTTGGTACAATTGGTCAGA
 GGTGGTCTTACCGACAATGTTACCAGAGAAGAACTTGAATCTGGGTC
 GTCTCTTCTTCTCTACCTATGTTGTGCAATCTGGCGACGAAACCACC
 AATCTCGGTACCAACCAATCTAGCAATATCGGTAGCCAAAGTGGCTT
 AGCAGTAACGTAGTCTAATAAGGTCTTCATTTTACCACATCTTGCAT
 AATCAAAGCATGTTTTTCCAAGTCATAGTTGAGACCTTCTGGAACAGA
 GACAATACCATCAACACCACCCAAAACCTTCTCTGTTAGCCATCATCAA

108

-continued

CTTGATAGCTTGGTATTCGTAGACAGAACGTTCAACACCGATTTTGAA
 ATCTTCATCAGTAGACATGTGTGGTTGAGCGTGCTTCAAAATGATAGA
 5 AGTGTGACCTTGGTATGGAGCGTTCAATTTAATACGCCAGGTGACGTT
 AACGAAACCCCGGATAATCTCTTGACACCAGAAGTGTCAACATCTAA
 AGACAAATGCTTGGTCAAATAGGTGATTAAACCGTCTTCAGTCTTCAA
 10 GTCAAAGGCCATTGTTTTATATTGTTGTGTAAGTAGATAATTACTT
 CCTTGATGATCTGTAAAAAGAGAAAAAGAAAGCATCTAAGAACTTGA
 AAAACTACGAATTAGAAAAGACCAATATGTATTTCTTGCATTGACCA
 15 ATTTATGCAAGTTTATATATATGTAAATGTAAGTTTCACGAGGTTCTA
 CTAAACTAAACCACCCCTTGGTTAGAAGAAAAGAGTGTGTGAGAACA
 GGCTGTTGTTGTACACGATTTCGGACAATTCTGTTTGAAGAGAGAGA
 20 GTAACAGTACGATCGAACGAACCTTGCTCTGGAGATCACAGTGGGCAT
 CATAGCATGTGGTACTAAACCCCTTCCCGCCATTCCAGAACCTTCGAT
 TGCTTGTTACAAAACCTGTGAGCCGTGCTAGGACCTTGTGTGTGAC
 GAAATTGGAAGCTGCAATCAATAGGAAGACAGGAAGTCGAGCGTGTCT
 25 GGGTTTTTTCAGTTTTGTTCTTTTGCACAAACAGTTTATTCCTGGC
 ATCCACTAAATATAATGGAGCCCGCTTTTAAAGCTGGCATCCAGAAA
 AAAAGAATCCCAGCACCAAAATATTGTTTTCTTCAACACCATCAGT
 30 TCATAGGTCCATTCTCTTAGCGCAACTACAGAGAACAGGGGCACAAAC
 AGGCAAAAAACGGGCACAACCTCAATGGAGTGATGCAACCTGCCTGGA
 GTAAATGATGACACAAGGCAATTGACCCACGATGATCTATCTCATT
 35 TTCTTACACCTTCTATTACCTTCTGCTCTCTCTGATTGGAAAAAGCT
 GAAAAAAAGGTTGAAACAGTTCCTCGAAATATTCCCTACTTGAC
 TAATAAGTATATAAGACGGTAGGTATTGATTGTAATCTGTAATCT
 40 ATTTCTTAAACTTCTTAAATCTACTTTTATAGTTAGTCTTTTTTTTA
 GTTTTAAACACCAAGAACCTTAGTTTCGAATAAACACACATAAACAAA
 CAAATGCATATCAGAAACCATATAGAACTCCAATTGACTACCAAGC
 45 TTTGTCTGAAGCTTTCCACCATGGAAGCCATTTGTTTCCGTTAAGC
 TGATGGTACCTCCTCAGTTGACTTGACCATTCAGAAGCCCAAGAGC
 TTTTACCGCTGCCCTTTTGCACAGAGACTTCGGCTTGACTATGACTAT
 50 CCCAGAAGATCGTTTGTGTCCAACCGTTCCAAACAGATTGAACTACGT
 TTTGTGGATTGAAGACATTTTCACTACACCAACAAGACTTTGGGTTT
 ATCTGACGACCGTCCAATCAAGGGTGTGATATCGGTACCGGTGCTTC
 55 TGCCATTTACCCAATGTTGGCTTGCAGCAGATTCAAGGCTTGGTCCAT
 GGTGGTACTGAAGTTGAAAGAAAGTGTATCGACACTGCTAGATTAAA
 CGTTGTTGCTAACAACCTTGCAAGATCGTCTATCCATCTTGAAACCTC
 TATTGACGGTCCAATTTTAGTCCCAATTTTGAAGCTACCGAAGAATA
 60 CGAATACGAATTCACCATGTGTAACCCACCTTTCTACGATGGTGCCGC
 TGACATGCAACTAGCGATGCTGCAAGGGTTTTGGTTTCGGTGTCCG
 TGCTCCACACTCTGGTACAGTCATCGAAATGTCTACTGAAGGTGGTGA
 65 ATCCGCTTTCGTGGCTCAAATGGTTAGAGAATCTCTTAAGTTGAGAAC

109

-continued

CAGATGTAGATGGTACACTTCTAACTTAGGTAAGTTGAAATCTTTGAA
 GGAAATTGTCGGTTTGTGTAAGGAATTAGAAATCTCTAATTACGCCAT
 CAACGAATATGTCCAAGGTTCCACTAGAAAGATACGCTGTCGCTTGGAG
 TTTCAC TGATATCCAATTGCCAGAAGAATTGCCAGACCATCTAATCC
 TGAATTGTCCTCTTTGTTGCGACTACAAGGATGACGATGACAAATGAAT
 CGCGTGCAATTCATCCGCTCTAACCGAAAAGGAAGGAGTTAGACAACCT
 GAAGTCTAGGTCCTTATTTATTTTTTATAGTTATGTTAGTATTAAGA
 ACGTTATTTATATTTCAAATTTTTCTTTTTTTCTGTACAGACGCGTG
 TACGCATGTAACATTATACTGAAAACCTTGCTTGAGAAGGTTTTGGGA
 CGCTCGAAGATCGCGTACCCCAATTGCGCCTATAGTGAGTCGTATTACG
 CGCGCTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTG
 GCGTTACCCCTGCAGGACTAGTGCTGAGGCATTAATGCAACTCAGAAG
 TTTGACAGCAAGCAAGTTCATCATTCGAACTAGCCTTATTGTTTTAGT
 TCAGTGACAGCGAACTGCCGTACTCGATGCTTTATTTCTCACGCTAGA
 GCGGAAGAACAGATAGGGGCGAGCTGAGAAGAGTTAGAAAGTAAATTT
 TTATCACGCTGGAAGTATTCTTATTCATAGGAAATTTGCAAGGTTTT
 TTAGCTCAATAACGGGCTAAGTTATATAAGGTGTTACGCGATTTTCT
 TGTATTGTATACCTCTTCTCTGAGGAATGGTACTACTGTCCTGATGTA
 GGCTCCTTAAATTGGTGGGCAAGAATACTTATCGATATTTGTATAT
 TGGTCTTGGAGTTCACCACGTAATGCCTGTTTAAGACCATCAGTTAAC
 TCTAGTATTATTTGGTCTTGCTACTGGCCGTTTGTCTATTATCAAGT
 CTTTTGTGCTTCCCGTCGGGTAAGGGAGTTATTTAGGGATACAGAAT
 CTAACGAAAATAAATCTCAATGATTAACTCTATTTAATCCTTTTTTG
 AAAGGCAAAAGAGGTCCCTTGTTCACTTACAACGTTCTTAGCCAAATT
 CGCTTATCACTTACTACTTACGATATACAGAAGTAAAAACATATAAA
 AAGATGCTCTGTTTGTTTAGCCATCAGAAAAGGTATCGCAG

SEQ. ID NO: 76

TGTTTTATATTTGTTGTAAGGATAGATAATTACTTCTTGTATGATCT
 GTAAAAAGAGAAAAAGAAAGCATCTAAGAAGTTGAAAACTACGAAT
 TAGAAAAAGACCAATATGTATTTCTTGCAATTGACCAATTTATGCAAGT
 TTATATATATGTAATGTAAGTTTACGAGGTTCTACTAACTAAACC
 ACCCCCTTGGTTAGAAGAAAAGAGTGTGTGAGAACAGGCTGTTGTTGT
 CACACGATTCGGACAATTCTGTTTGAAAGAGAGAGTAACAGTACGA
 TCGAACGAACTTTGCTCTGGAGATCACAGTGGGCATCATAGCATGTGG
 TACTAAACCCCTTCCCGCCATTCCAGAACCCTCGATTGCTTGTTACAA
 AACCTGTGAGCCGTCGCTAGGACCTTGTTGTGTGACGAAATTGGAAGC
 TGCAATCAATAGGAAGACAGGAAGTCGAGCGTGTCTGGGTTTTTTCAG
 TTTTGTCTTTTGTCAAACAACAGTTTATTTCTGGCATCCACTAAATA
 TAATGGAGCCCGCTTTTTAAGCTGGCATCCAGAAAAAAAAGAATCCC
 AGCACCAAAATATTGTTTTCTTACCAACCATCAGTTTCATAGGTCCAT
 TCTCTTAGCGCACTACAGAGAACAGGGGCACAAACAGGCCAAAAACG

110

-continued

GGCACAACCTCAATGGAGTGATGCAACCTGCCTGGAGTAAATGATGAC
 ACAAGGCAATTGACCCACGCATGTATCTATCTCATTTTCTTACACCTT
 CTATTACCTTCTGCTCTCTCTGATTGGAAAAAGCTGAAAAAAGGT
 TGAAACCAGTTCCTTGAAATTATCCCCTACTTGACTAATAAGTATAT
 AAAGACGGTAGGTATTGATTGTAATTCGTAAATCTATTTCTTAAACT
 TCTTAAATCTACTTTTATAGTTAGTCTTTTTTTTAGTTTTTAAACAC
 CAAGAACTTAGTTTCGAATAAACACACATAAAACAAACAAAGGATCCAT
 GATGAAGTTCTGGAGAAAGTACACACAACAAGAAATGGATGAAAAGAT
 TACTGAATCTTTGGAAAAGACTTTGAACTACGATAACACTAAGACAAT
 CGGTATTCAGGTACTAAGTTGGATGATACAGTTTTCTATGATGATCA
 TTCTTTCGTTAAGCATTACCATACTTGAGAACTTTTATTCAAACCC
 AAACCATATCGGTTGTCTACTTTATGATAAGGCTGATATCTTGTTCCG
 TGGTACATTCGATATCGAAAGAGAATTAATCCAATTGTTAGCAATCGA
 TGTTTTGAACGGTAACGATGAAGAATTTGATGGTTACGTTACTCAAGG
 TGGTACAGAAGCTAACATCCAGCAATGTGGGTTTACAGAACTACTT
 CAAGAAAGAAAGAAAGGCTAAGCATGAAGAATCGCTATCATCACTTC
 AGCAGATACACATTACTCTGCATACAAAGGTTGAGATTTGTTGAACAT
 CGATATTATTAGGTTCCAGTTGATTTTTATTCAAGAAAAATTCAGA
 AAATACATTGGATTCAATTGTTAAAGAAGCTAAAGAAATGGTAAAAA
 GTACTTCATCGTTATCTCTAACATGGGTACTACAATGTTTGGTTCAGT
 TGATGATCCAGATTTGTACGCTAACATCTTCGATAAGTACAATTTGGA
 ATACAAAATTCATGTTGATGGTGCATTTGGTGGTTTTATATATCCAAT
 TGATAATAAGGAATGTAAACTGATTTCTCTAATAAGAACGTTTCTTC
 AATCACATTAGATGGTCATAAGATGTTGCAAGCTCCATACGGTACTGG
 TATCTTCGTTTCAAGAAAGAATTTGATCCATAACACTTTGACAAAGGA
 AGCAACTTACATCGAAAATTTGGATGTTACATTGCTGGTTCAAGATC
 TGGTTCAAATGCTGTTGCAATTTGGATGGTTTTAGCTTCTTATGGTCC
 ATACGGTTGGATGGAAAAGATTAATAAGTTGAGAAATAGAATAAATG
 GTTGTGTAAGCAATTGAACGATATGAGAATTAAATATTACAAAGAAGA
 TTCAATGAATATTGTTACAATTGAAGAACAATATGTTAATAAGGAAAT
 CGCTGAAAAGTACTTTTTAGTTCCAGAAGTTTCATAACCCAATAACAA
 CTGGTACAAGATCGTTGTTATGGAACATGTTGAATTGGATATCTTGAA
 CTCTTTGGTTTACGATTTGAGAAAGTTTAATAAGGAACATTGAAGGC
 AATGCATCATCATCATCATTAACCGCGCTAGCTAAGATCCGCTC
 TAACCGAAAAGGAAGGAGTTAGACAACCTGAAGTCTAGGTCCTATTT
 ATTTTTTTTAGTTATGTTAGTATTAAGAAGCTTATTTATATTTCAA
 TTTTCTTTTTTTTCTGTACAGACGCGGTGACGCATGTAACATTATAC
 TGAAAACCTTGCTTGAGAAGGTTTTGGGACGCTCGAAGATCCAGCTGC
 ATTAATGAATCGGCCAACGCGCGGGGAGAGCGGTTTTCGCTATTGGGC
 GCTCTTCCGCTTCTCGCTCACTGACTCGCTCGCTCGGTCGTTCCGC

111

-continued

TCGCGCGAGCGGTATCAGCTCACTCAAAGGCGGTAAATACGGTTATCCA
 CAGAATCAGGGGATAACGACAGGAAAGAACATGTGAGCAAAAGGCCAGC
 AAAAGGCCAGGAACCGTAAAAAGGCGCGTTGCTGGCGTTTTTCCATA
 GGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGA
 GGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTG
 GAAGCTCCCTCGTGCCTCTCCTGTCCGACCCTGCCGCTTACCGGAT
 ACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCT
 CACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCCGCTCAAGCTGG
 GCTGTGTGCACGAACCCCCGTTACGCCGACCGCTGCGCCTTATCCG
 GTAACATATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCAC
 TGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCG
 GTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAA
 GAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAA
 AAAGAGTTGGTAGCTCTTGATCCGGCAACAAACACCGCTGGTAGCG
 GTGGTTTTTTTGTGTAAGCAGCAGATTACGCGCAGAAAAAAGGAT
 CTAAGAAGATCCTTTGATCTTTCTACGGGTCTGACGCTCAGTGGAA
 ACGAAAACTCAGTTAAGGATTTTGGTCATGAGATTATCAAAAAGGA
 TCTTCACCTAGATCCTTTTAAATAAAAATGAAGTTTAAATCAATCT
 AAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAATCA
 GTGAGGCACCTATCTCAGCGATCTGTCTATTTCTGTTTATCCATAGTTG
 CCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCAT
 CTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTC
 CAGATTTATCAGCAATAAACACGCCAGCCGGAAGGGCCGAGCGCAGAA
 GTGGTCCTGCAACTTTATCCGCTCCATCCAGTCTATTAATTGTTGCC
 GGGAGCTAGAGTAAGTAGTTCCGCCAGTTAATAGTTTGCACAACGTTG
 TTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCTTGGTATGG
 CTTTATTAGCTCCGGTTCCCAACGATCAAGCGAGTTACATGATCCC
 CCATGTTGTGCAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGTTG
 TCAGAAGTAAGTTGGCCGCGTGTATCACTCATGGTTATGGCAGCAC
 TGCATAATTCTCTTACTGTATGCCATCCGTAAGATGCTTTTCTGTGA
 CTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGAC
 CGAGTTGCTCTTGCCGCGCGTCAATACGGGATAATACCGCGCCACATA
 GCAGAACTTTAAAGTGCTCATCATTTGAAACGTTCTTCGGGGCGAA
 AACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTTCGATGTAACCCA
 CTCGTGCAACCACTGATCTTCAGCATCTTTTACTTTTACCAGCGTTT
 CTGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAATAA
 GGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATT
 ATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTG
 AATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCC
 GAAAAGTGCCACCTGAACGAAGCATCTGTGCTTCATTTGTAGAACAA
 AAATGCAACGCGAGAGCGCTAATTTTTCAACAAAGAATCTGAGCTGC

112

-continued

ATTTTTACAGAACAGAAATGCAACGCGAAAGCGCTATTTTACCAACGA
 AGAATCTGTGCTTCATTTTTGTAAAACAAAATGCAACGCGAGAGCGC
 5 TAATTTTTCAACAAAGAATCTGAGCTGCATTTTTACAGAACAGAAAT
 GCAACGCGAGAGCGCTATTTTACCAACAAAGAATCTATACTTCTTTTT
 TGTTCTACAAAAATGCATCCGAGAGCGCTATTTTTCTACAAAGCAT
 10 CTTAGATTACTTTTTTCTCCTTTGTGCGCTCTATAATGCAGTCTCTT
 GATAACTTTTTGCACTGTAGGTCCGTTAAGGTTAGAAGAAGGCTACTT
 TGGTGTCTATTTTCTCTTCCATAAAAAAGCCTGACTCCACTTCCCGC
 15 GTTTACTGATTACTAGCGAAGCTGCGGGTGCAATTTTTTCAAGATAAAG
 GCATCCCGATTATATTCTATACCGATGTGGATTGCGCATACTTTGTG
 AACAGAAAGTGATAGCGTTGATGATTCTTCATTGGTCAGAAAATTATG
 20 AACGGTTTCTTCTATTTTGTCTCTATATACTACGTATAGGAAATGTTT
 ACATTTTCGTATTGTTTTCGATTCACTCTATGAATAGTTCTTACTACA
 ATTTTTTTGTCTAAAGAGTAATACTAGAGATAAACATAAAAAATGTAG
 25 AGGTCGAGTTTAGATGCAAGTCAAGGAGCGAAAGGTGGATGGGTAGG
 TTATATAGGGATATAGCACAGAGATATATAGCAAGAGATACTTTTGA
 GCAATGTTTGTGGAAGCGGTATTTCGCAATATTTTAGTAGCTCGTTACA
 30 GTCCGGTGCCTTTTTGGTTTTTTGAAAGTGCCTTTCAGAGCGCTTTT
 GGTTTTCAAAAGCGCTCTGAAGTTCCTATACTTTCTAGAGAATAGGAA
 CTTCCGGAATAGGAACCTCAAAGCGTTTCCGAAAACGAGCGCTTCCGAA
 35 AATGCAACGCGAGCTGCGCACATACAGCTCACTGTTCACTGCGCACCT
 ATATCTGCGTGTGCTGTATATATATATACATGAGAAGAACGCGCATA
 GTGCGTGTTTATGCTTAAATGCGTACTTATATGCGTCTATTTATGTAG
 40 GATGAAAGGTAGTCTAGTACCTCCTGTGATATTATCCCATTCATGCG
 GGGTATCGTATGCTTCCTTCAGCACTACCCCTTAGCTGTTCTATATGC
 TGCCACTCCTCAATTGGATTAGTCTCATCCTTCAATGCTATCATTTCC
 TTTGATATTGGATCATACTAAGAAACATTATTATCATGACATTAAAC
 45 TATAAAAAATAGCGGTATCACGAGGCCCTTTCGTCTCGCGCTTTCCGT
 GATGACGGTGAAAACCTCTGACACATGCAGCTCCCGAGAGCGGTCACA
 GCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGCGCGC
 50 TCAGCGGGTGTGGCGGGTGTGCGGGCTGGCTTAACTATGCGGCATCA
 GAGCAGATTGTACTGAGAGTGCACCATATTTACACCGCATAGATCCG
 TCGAGTTCAAGAGAAAAAAGAAAAAGCAAAAAGAAAAAGGAAAG
 55 CGCGCCTCGTTCAGAAATGACACGTATAGAATGATGCATTACCTTGTC
 TCTTCAGTATCATACTGTTGTTATACATACTTACTGACATTATAGGT
 ATACATATATACACATGTATATATATCGTATGCTGCAGCTTTAAATAA
 60 TCGGTGTCACTACATAAGAACACCTTTGGTGGAGGGAACATCGTTGGT
 ACCATTGGGCGAGGTGGCTTCTCTTATGGCAACCGCAAGAGCCTTGAA
 CGCACTCTCACTACGGTGTGATGATCATTTCTGCCTCGCAGACAATCAAC
 65 GTGGAGGGTAATTCTGCTAGCCTCTGCAAGCTTTCAAGAAAAATGCGG

113

-continued

GATCATCTCGAAGAGAGATCTCCTACTTTCTCCCTTTGCAAACCAAG
 TTCGACAACCTGCGTACGGCCTGTTTCGAAAGATCTACCACCGCTCTGGA
 AAGTGCCTCATCCAAAGGCGCAAACTCTGATCCAAACCTTTTACTCC
 ACGCACGGCCCTAGGGCCTCTTTAAAAGCTTGACCGAGAGCAATCCC
 GCAGTCTTCAGTGGTGTGATGGTCGTCTATGTGAAGTCACCAATGCA
 CTCAACGATTAGCGACCAGCGGAATGCTTGCCAGAGCATGTATCAT
 ATGGTCCAGAAACCCTATACCTGTGTGGACGTTAATCACTTGCGATTG
 TGTGGCCTGTTCTGCTACTGCTTCTGCCTCTTTTCTGGGAAGATCGA
 GTGCTCTATCGCTAGGGGACCACCTTTTAAAGAGATCGCAATCTGAAT
 CTTGGTTTCATTGTGAATACGCTTTACTAGGGCTTTCTGCTCTGTCAT
 CTTTGCCCTCGTTTATCTTGCTGCTCATTTTTTAGTATATTCTTCGA
 AGAAATCACATTACTTTATATAATGTATAATTCATTATGTGATAATGC
 CAATCGCTAAGAAAAAAAAGAGTCATCCGCTAGGTGAAAAAAA
 ATGAAATCATTACCGAGGCATAAAAAATATAGAGTGTACTAGAGGA
 GGCCAAAGAGTAATAGAAAAAGAAATTGCGGGAAAGGACTGTGTATG
 ACTTCCCTGACTAATGCCGTGTTCAAACGATACCTGGCAGTGACTCCT
 AGCGCTCACCAAGCTCTTAAAACGGGAATTTATGGTGCACCTCTCAGTA
 CACGCGCCAGATCTGTTTAGCTTGCTCGTCCCGCCGGGTCACCCGG
 CCAGCGACATGGAGGCCAGAATACCTCCTTGACAGTCTTGACGTGC
 GCAGCTCAGGGGCATGATGTGACTGTGCGCCGTACATTTAGCCCATAC
 ATCCCATGTATAATCATTGTCATCCATACATTTTGATGGCGCACGG
 CGCGAAGCAAAATACGGCTCCTCGCTGCAGACCTGCGAGCAGGGAA
 ACGTCTCCCTCACAGACGCGTGAATTGTCCCGACGCGCGCCCTGT
 AGAGAAATATAAAGGTTAGGATTTGCCACTGAGGTTCTTCTTTCATA
 TACTTCTTTTAAATCTTGCTAGGATACAGTCTCACATCACATCCG
 AACATAAACCAACTGGGTAAAGGAAAGACTCACGTTTCGAGGCCGCG
 ATTAATTTCCAACTGGATGCTGATTTATATGGGTATAAATGGGCTCG
 CGATAATGTGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAA
 GCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGC
 CAATGATGTTACAGATGAGATGGTCAGACTAACTGGCTGACGGAATT
 TATGCCCTCTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGC
 ATGTTACTCACCACTGCGATCCCCGGCAAAACAGCATTCCAGGTATT
 AGAAGAATATCCTGATTAGGTGAAAAATTTGTTGATGCGCTGGCAGT
 GTTCTGCGCCGGTTGCATTGATTCCTGTTTGTAAATGTCTTTTAA
 CAGCGATCGCGTATTTCTGCTCGCTCAGGCGCAATCAGAAATGAATAA
 CGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCC
 TGTGAACAAGTCTGGAAAGAAATGCATAAGCTTTTGCCATTCTCACC
 GGATTTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTT
 TGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAAT
 CGCAGACCGATACCAGGATCTTGCCATCCTATGGAACCTGCCTCGTGGA
 GTTTTCTCCTTACATACAGAAACGGCTTTTCAAAAATATGGTATTGA

114

-continued

TAATCTGATATGAATAAATTGCAGTTTCATTGTAGTCTGATGAGTT
 TTTCTAATCAGTACTGACAATAAAAAAGATTCTGTTTTCAAGAACTTG
 5 TCATTTGTATAGTTTTTTTATATTGTAGTTGTTCTATTTTAATCAAAT
 GTTAGCGTGATTTATATTTTTTTTCGCTCGACATCATCTGCCAGAT
 GCGAAGTTAAGTGCAGCAAGTAATATCATGCGTCAATCGTATGTGA
 10 ATGCTGGTTCGTATACTGCTGTCGATTGCTACTAACCGCCCATCCA
 GTGTGCAATTGCCATTAGGCTGCGCAACTGTTGGGAAGGCGGATCG
 GTGCGGGCCTCTTCGCTATTACGCCAGCTGAATTGGAGCGACCTCATG
 15 CTATACCTGAGAAAGCAACCTGACCTACAGGAAAGAGTTACTCAAGAA
 TAAGAAATTTTCGTTTAAACCTAAGAGTCACCTTAAATTTGTATAC
 ACTTATTTTTTTTATACTTATTTAATAATAAAAAATCATAAATCATAA
 20 GAAATTCGCTTATTTAGAAGTGTCAACAACGTATCTACCAACGATTTG
 ACCCTTTTCCATCTTTTCGTAAATTTCTGGCAAGGTAGACAAGCCGAC
 AACCTTGATTGGAGACTTGACCAACCTCTGGCGAAGAATTGTTAATT
 25 AAGAGCTCAGATCTTTTTCGCGCCGC
 SEQ. ID NO: 77
 GTTTGCAAAAAGAACAAAACCTGAAAAACCCAGACACGCTCGACTTCC
 TGTCTTCTTATTGATTGTCAGCTTCCAATTTTCGTACACAACAAGGTCC
 30 TAGCGACGGCTCACAGGTTTTGTAAACAGCAATCGAAGGTTCTGGAAT
 GGCGGGAAGGGTTTAGTACCACATGCTATGATGCCCACTGTGATCTC
 CAGAGCAAGTTTCGTTTCGATCGTACTGTTACTCTCTCTCTTCAACA
 35 GAATTGTCCGAATCGTGTGACAACAACAGCCTGTTCTCACACACTCTT
 TTCTTCTAACCAAGGGGGTGGTTTAGTTTAGTAGAACCTCGTGAACT
 TACATTTACATATATATAAACTTGCAATAAATGGTCAATGCAAGAAAT
 40 ACATATTTGGTCTTTTCTAATTCGTAGTTTTTCAAGTTCTTAGATGCT
 TTCTTTTCTCTTTTTTACAGATCATCAAGGAAGTAATTATCTACTTT
 TTACAACAAATATAAACA
 SEQ. ID NO: 78
 45 ACAGTTTATTCCTGGCATCCACTAAATATAATGGAGCCCGCTTTTAA
 GCTGGCATCCAGAAAAAAGAAATCCAGCACCAAAATATTGTTTTTC
 TTCACCAACCATCAGTTCATAGGTCCATTCTCTTAGCGCAACTACAGA
 50 GAACAGGGGCACAAACAGGCAAAAAACGGGCACAACTCAATGGAGTG
 ATGCAACCTGCCTGGAGTAAATGATGACACAAGGCAATTGACCCACGC
 ATGTATCTATCTCATTTTCTTACACCTTCTATTACCTTCTGCTCTCTC
 55 TGATTTGGAAAAAGCTGAAAAAAGGTTGAAACAGTTCCCTGAAAT
 TATTCCCTACTTGACTAATAAGTATATAAAGACGGTAGGTATTGATT
 GTAATTCTGTAATCTATTTCTTAACTTCTTAAATCTACTTTTATA
 60 GTTAGTCTTTTTTTTAGTTTTTAAACACCAAGAACTTAGTTTCGAATA
 AACACACATAAACAAACAAA
 SEQ. ID NO: 79
 65 GCTGGATTGAGCTGAATGGTGCCAGGTGAGGCTGGGAGGAGACTAA
 CTCGAAAGTGACGAAGACTCGAAAATAAAAAAGATACTGCAGAA

115

-continued

GGCAAGATTGAGAATGGAGTAAAGGCAGCGTGGGTCCCTGTGGAAAC
CGCAGTTTTCTGCGCCAAGTGGTACCGGTGCGAGTGCAGCAATTAAT
CTCTCGATATTTCTTAGTATCTCTTTTATATAAGAATATATTTTGG
AATTGGTAATGCTTATCTTCAATAGTTTCTTAGTTGAATGCACACTTA
AGAGCAAATTGGCCAAGGAGTTCCTCGTTCGCTTTAATTTATTTCTG
GTTATTGTCAATTTATTCATCCCATCTCCCCAGGATAGAAGAAATTAG
TGTAATTTTGCTGACAATACATTTTAAACGACGATAACAATAATAGCAA
TTAAATAAAATAGCACTACCACCACTCCACTGCTCGTTAGCTATTCT
GTAAATAAATAAAAAGATC

SEQ. ID NO: 80

ATTCGCGCTATCTCGATTTCTACCTATATAGTTAATCTCTGTACAAA
ACAATCTTTCCAATATCCATTAATCATAGTATATTATCAGCGTCGGC
GATTTTACCACGCTTGACAAAAGCCGCGGGCGGGATTCTGTGGGTAG
TGGCACCGGCAGTTAATCTAATCAAAGGCGCTTGAAGGAAGAGATAGA
TAATAGAACAAAGCAATCGCCGCTTTGGACGGCAATATGTTTATCCA
TTGGTGCGGTGATTGGATATGATTTGTCTCCAGTAGTATAAGCAAGCG
CCAGATCTGTTTACTGTAAAATTAAGTGAGTAATCTCGCGGGATGTAA
TGATTTAAGGGAATCTGGTTGAGTTTTACATATATTTGTATATAAG
GCCATTTGTAATTTCAATAGTTTGTAGATTTTCTCTCTCCAAAATA
CTCACTTACTGTGTACATTACAGAAAGAACAGACAAGAAACCGTCAA
TAAGAAATATACTAAGAACA

SEQ. ID NO: 81

TACCATACGACGTTCCAGACTACGCC

SEQ. ID NO: 82

YPYDVDPYA

SEQ. ID NO: 83

GAACAAAAGTTAATTTCTGAAGAAGATTGGAA

SEQ. ID NO: 84

EQKLISEEDL

SEQ. ID NO: 85

GACTACAAGGATGACGATGACAAA

SEQ. ID NO: 86

DYKDDDDK

SEQ. ID NO: 87

GGTAAGCCAATTCCAAATCCTTTGTTGGGTTTGGACTCCACC

SEQ. ID NO: 88

GKPIPNPLGLDST

SEQ. ID NO: 89

CATCATCATCATCAT

SEQ. ID NO: 90

HHHHHH

SEQ. ID NO: 91

CAGGTGGCACTTTTCGGGGAATGTGCGGGAACCCCTATTGTTTAT
TTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCT
GATAAATGCTTCAATAATATTGAAAAGGAAGAGTATGAGTATTTCAAC
ATTTCCGTGTCGCCCTTATTCCTTTTTTGCGGCATTTTGCCTTCCTG

116

-continued

TTTTTGCTCACCAGAAACGCTGGTGAAAGTAAAGATGCTGAAGATC
AGTTGGGTGCACGAGTGGGTTACATCGAAGTGGATCTCAACAGCGGTA
5 AGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCA
CTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCG
GGCAAGAGCAACTCGGTGCGCGCATACACTATTCTCAGAATGACTTGG
10 TTGAGTACTCACCAGTACAGAAAAGCATCTTACGGATGGCATGACAG
TAAGAGAATTATGCAAGTGTGCCATAACCATGAGTGATAACACTGCGG
CCAACCTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTT
15 TTTTGCACAACATGGGGGATCATGTAACCTCGCCTTGATCGTTGGGAAC
CGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGC
CTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTAC
20 TTAGTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATA
AAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTA
TTGCTGATAAATCTGGAGCCGTTGAGCGTGGGTCTCGCGGTATCATTG
CAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACA
25 CGACGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTG
AGATAGGTGCCTCACTGATTAGCATTTGGTAAGTGTGACAGCAAGTTT
ACTCATATATACTTTAGATTGATTTAAAACCTCATTTTTAAATTTAAAA
30 GGATCTAGGTGAAGATCCTTTTGTATAATCTCATGACCAAAATCCCTT
AACGTGAGTTTTCTGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCA
AAGGATCTTCTTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGC
35 AAACAAAAAACCCCGCTACCAGCGGTGGTTTGTGTTGCCGGATCAAG
AGCTACCAACTCTTTTTCCGAAGTAAGTGGCTTACGAGAGCGCAGA
TACCAATACTGTTCTTCTAGTGTAGCCGATGTTAGGCCACCACTTCA
40 AGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTAC
CAGTGGCTGCTGCCAGTGGCGATAAGTCTGTCTTACCGGTTTGGACT
CAAGACGATAGTTACCGGATAAGGCGCAGCGGTGCGGCTGAACGGGG
45 GTTCGTGCACACAGCCAGCTTGGAGCGAACGACCTACACCGAACTGA
GATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGA
GAAAGGCGGACAGGTATCCGGTAAGCGCGAGGTCGGAACAGGAGAGC
50 GCACGAGGGAGCTTCCAGGGGAAACGCTGGTATCTTTATAGTCTCTG
TCGGGTTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCGT
CAGGGGGCGGAGCCTATGGAACAAACGCGAGCAACGCGCCTTTTAC
55 GGTTCTTGGCCTTTTGTGCTGGCCTTTTGTCTCACATGTTCTTCTGCGT
TATCCCTGATCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTG
ATACCGCTCGCCGCGAGCCGAACGACCGAGCGCAGGATCAGTGAGCG
60 AGGAAGCGGAAGAGCGCCCAATACGCAAAACCGCTTCTCCCGCGCGTT
GGCCGATTCAATATGCAAGTGGATCTTTCGAGCGTCCCAAAACCTTCT
CAAGCAAGGTTTTCAGTATAATGTTACATGCGTACACGCGTCTGTACA
GAAAAAAGAAAAATTTGAAATATAATAACGTTCTTAATACTAACA
65 TAACATAAAAAATAAATAGGGACCTAGACTTCAGGTTGTCTAACTC

117

-continued

CTTCCTTTTCGGTTAGAGCGGATCTTAGCTAGCCGCGGTACCAAGCTG
 GTGGATCCTTTGTTTGTATTATGTGTGTTATTTCGAACTAAGTTCTTG
 GTGTTTTAAACTAAAAAAGACTAATATAAAGTAGAATTTAAGA
 AGTTTAAGAAATAGATTTACAGAATTACAATCAATACCTACCGTCTTT
 ATATACTTATTAGTCAAGTAGGGGAATAATTTAGGGAACTGGTTTCA
 ACCTTTTTTTTCAGCTTTTCCAAATCAGAGAGAGCAGAAGGTAATAG
 AAGGTGTAAGAAAATGAGATAGATACATGCGTGGGTCAATTGCCTTGT
 GTCATCATTTACTCCAGGCAGGTTGCATCACTCCATTGAGGTTGTGCC
 CGTTTTTTTGCCTGTTTGTGCCCTGTTCTCTGTAGTTGCGCTAAGAGA
 ATGGACCTATGAACGTAGGTTGGTGAAGAAAACATATTTTGGTGCT
 GGGATTCTTTTTTTTTCTGGATGCCAGCTTAAAAAGCGGGCTCCATTA
 TATTTAGTGGATGCCAGGAATAAAGTGTGTTTGCAAAAGAACAAA
 CTGAAAAAACCCAGACACGCTCGACTTCTGTCTTCCTATTGATTGCA
 GCTTCCAATTTTCGTACACAACAAGGTCCTAGCGACGGCTCACAGGTT
 TTGTAACAAGCAATCGAAGGTTCTGGAATGGCGGAAAGGTTTAGTA
 CCACATGCTATGATGCCCACTGTGATCTCCAGAGCAAAGTTCGTTCGA
 TCGTACTGTTACTCTCTCTCTTTCAAACAGAATTGTCCGAATCGTGTG
 ACAACAACAGCCTGTTCTCACACACTCTTTCTTCTAACCAGGGGTT
 GGTTTAGTTTAGTAGAACCTCGTGAACCTTACATTTACATATATATAA
 ACTTGCATAAATGGTCAATGCAAGAAATACATATTTGGTCTTTTCTA
 ATTCGTAGTTTTTCAAGTCTTAGATGCTTTCTTTTCTCTTTTTTAC
 AGATCATCAAGGAAGTAATTATCTACTTTTTACAACAAATATAAAACA
 GCGGCCGCAAAAGATCTGAGCTCTTAATTAACAATTCTTCGCCAGAGG
 TTTGGTCAAGTCTCCAATCAAGGTTGTGCGCTTGCTACCTTGCCAGA
 AATTTACGAAAAGATGGAAGGTTCAAATCGTTGGTAGATACGTTGT
 TGACACTTCTAAATAAGCGAATTTCTTATGATTTATGATTTTATTAT
 TAAATAAGTTATAAAAAAATAAGTGTATACAAATTTAAAGTGACTC
 TTAGGTTTTAAACGAAAATTCTTATTCTTGAGTAACCTTTCTCTGTA
 GGTGAGGTTGCTTTCTCAGGTATAGCATGAGGTCGCTCCAATTGAGCT
 GGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCAACAGTTGC
 GCAGCCTGAATGGCGAATTTCGACACTGGATGGCGCGTTAGTATCGAA
 TCGACAGCAGTATAGCGACCAGCATTACATACGATTGACGCATGATA
 TTACTTTCTGCGCACTTAACTTCGCATCTGGGCAGATGATGTCGAGGC
 GAAAAAAAATATAAATCAGCTAACATTTGATTAAAAATAGAACAACTA
 CAATATAAAAAAATATACAAATGACAAGTTCTTGAAAAACAAGATCT
 TTTTATTGTGCTAGTACTGATTAGAAAACTCATCGAGCATCAAATGAAA
 CTGCAATTTATTATCATAGGATTATCAATACCATATTTTGAAGAAAG
 CCGTTTCTGTAATGAAGGAGAAAACTCACCAGGCAGTTCCATAGGAT
 GGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAAT
 ACAACCTATTAATTTCCCTCGTCAAAAATAAGGTTATCAAGTGAGAA

118

-continued

ATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTATG
 CATTTCCTTTCCAGACTTGTTCACAGGCCAGCCATTACGCTCGTCATC
 5 AAAATCACTCGCATCAACCAAACCGTTATTCATTGCTGATTGCGCCTG
 AGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGG
 AATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATT
 10 TTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTGGCC
 GGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAA
 ATGCTTGATGGTCGGAAGAGGCATAAATTCGCTCAGCCAGTTTAGTCT
 15 GACCATCTCATCTGTAACATCATTTGGCAACGCTACCTTTGCCATGTTT
 CAGAAACAACCTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGT
 CGCACCTGATTGCCCCGACATTATCGCGAGCCCATTTATACCATATAA
 ATCAGCATCCATGTTGGAATTTAATCGCGGCTCGAAACGTGAGTCTT
 20 TTCTTACCCATGGTTGTTTATGTTCCGATGTGATGTGAGAACTGTAT
 CCTAGCAAGATTTTAAAAGGAAGTATATGAAAGAAGAACCCTCAGTGGC
 AAATCCTAACCTTTTATATTTCTTACAGGGGCGCGCGTGGGGACAA
 25 TTCAACGCGTCTGTGAGGGGAGCGTTTCCCTGCTCGCAGGTCTGCAGC
 GAGGAGCCGTAATTTTGCCTTCGCGCCGTGCGGCCATCAAATGTATG
 GATGCAAATGATTATACATGGGGATGTATGGGCTAAATGTACGGGCGA
 30 CAGTCACATCATGCCCTGAGCTGCGCACGTCAAGACTGTCAAGGAGG
 GTATTCTGGCCTCCATGTGCTGGCCGGGTGACCCGCGGGGACGAG
 GCAAGCTAAACAGATCTGGCGCGTGTACTGAGAGTGCAACCATAAATTC
 35 CCGTTTTAAGAGCTTGGTGAGCGCTAGGAGTCACTGCCAGGTATCGTT
 TGAACACGCGATTAGTCAGGGAAGTCATAACACAGTCCTTTCCCGCAA
 TTTTCTTTTCTATTACTCTTGGCCTCCTCTAGTACACTCTATATTTT
 40 TTTATGCCCTCGGTAATGATTTTCATTTTTTTTTTTCCACCTAGCGGAT
 GACTCTTTTTTTTCTTAGCGATTGGCATTATCACATAATGAATTATA
 CATTATATAAAGTAATGTGATTTCTTGAAGAATATACTAAAAAATGA
 45 GCAGGCAAGATAAACGAAGGCAAGATGACAGAGCAGAAAGCCCTAGT
 AAAGCGTATTACAAATGAAACCAAGATTGAGATTGCGATCTCTTTAA
 GGGTGGTCCCCTAGCGATAGAGCACTCGATCTTCCGAAAAAAGAGGC
 50 AGAAGCAGTAGCAGAACAGGCCACACAATCGCAAGTGATTAAAGTCCA
 CACAGGTATAGGGTTTCTGGACCATATGATACATGCTCTGGCCAAGCA
 TTCCGGCTGGTCGCTAATCGTTGAGTGCAATTGGTGACTTACACATAGA
 55 CGACCATCACACCACTGAAGACTGCGGGATTGCTCTCGGTCAAGCTTT
 TAAAGAGGCCCTAGGGGCGGTGCGTGGAGTAAAAAGGTTTGGATCAGG
 ATTTGCGCCTTTGGATGAGGCACCTTCCAGAGCGGTGGTAGATCTTTC
 GAACAGGCCGTACGCAGTTGTGCAACTTGGTTTGCAAGGGGAGAAAGT
 60 AGGAGATCTCTCTTTCGAGATGATCCCGCATTTTCTTGAAAGCTTTGC
 AGAGGCTAGCAGAATTACCTCCACGTTGATTGTCTGCGAGGCAAGAA
 TGATCATCACCGTAGTGAGAGTGCCTTCAAGGCTCTTTCGGTTGCCAT
 65 AAGAGAAGCCACCTCGCCCAATGGTACCAACGATGTTCCCTCCACCAA

119

-continued

AGGTGTTCTTATGTAGTGACACCGATTATTTAAAGCTGCAGCATACGA
 TATATATACATGTGTATATATGTATACCTATGAATGTGAGTAAGTATG
 TATACGAACAGTATGATACTGAAGATGACAAGGTAATGCATCATTCTA
 TACGTGTCAATTCTGAACGAGGCGCGCTTTCCTTTTTTCTTTTGTCTT
 TTCTTTTTTTTTCTCTTGAACCTGACGGATCTATGCGGTGTGAAATAT
 GGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGC
 CCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGC
 TCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCA
 TGTGTGAGAGTTTTACCGTCATACCGAAACGCGGAGACGAAAGG
 GCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAAATAGG
 TTTCTTAGTATGATCCAATATCAAAGGAAATGATAGCATTGAAGGATG
 AGACTAATCCAATTGAGGAGTGGCAGCATATAGAACAGCTAAAGGTA
 GTGCTGAAGGAAGCATACGATACCCCGCATGGAATGGGATAATATCAC
 AGGAGGTACTAGACTACCTTTCATCCTACATAAATAGACGCATATAAG
 TACGCATTTAAGCATAAACACGCACATATGCCGTTCTTCTCATGTATAT
 ATATATACAGGCAACACGCAGATATAGGTGCGACGTGAACAGTGAGCT
 GTATGTGCGCAGCTCGCGTTGCATTTTCGGAAGCGCTCGTTTTCGGAA
 ACGCTTTGAAGTTCCTATTCGGAAGTTCCTATTCTCTAGAAAGTATAG
 GAACTTCAGAGCGCTTTTGAAAACCAAAGCGCTCTGAAGACGCACTT
 TCAAAAAACCAAACGACCGGACTGTAACGAGCTACTAAAATATTG
 CGAATACCGCTTCCACAAACATTGCTCAAAAGTATCTCTTTGCTATAT
 ATCTCTGTGCTATATCCCTATATAACCTACCCATCCACCTTTCGCTCC
 TTGAACCTTGCACTCTAACTCGACCTCTACATTTTATGTTTATCTCT
 AGTATTACTCTTTAGACAAAAAATGTAGTAAGAACTATTATAGAG
 TGAATCGAAAAACAATACGAAATGTAAACATTTCTATACGTAGTATA
 TAGAGACAAAATAGAAGAAACCGTTTATAATTTTCTGACCAATGAAGA
 ATCATCAACGCTATCACTTCTGTTTCAAAAGTATGCGCAATCCACAT
 CGGTATAGAATATAATCGGGGATGCCTTTATCTTGAAAAATGCACCC
 GCAGCTTCGCTAGTAAATCAGTAAACGCGGGAAGTGGAGTCAGGCTTTT
 TTTATGGAAGAGAAAAATAGACACCAAAGTAGCCTTCTTCTAACCTTAA
 CGGACCTACAGTGCAAAAAGTTATCAAGAGACTGCATTATAGAGCGCA
 CAAAGGAGAAAAAAGTAATCTAAGATGCTTTGTTAGAAAAATAGCGC
 TCTCGGGATGCATTTTGTAGAACAAAAAGAAGTATAGATTCTTTGT
 TGGTAAAAATAGCGCTCTCGCGTTGCATTTCTGTTCTGTAAAAATGCAG
 CTCAGATTCTTTGTTTGA AAAATAGCGCTCTCGCGTTGCATTTTGT
 TTTACAAAAATGAAGCAGAGATTCTTCGTTGGTAAAAATAGCGCTTTG
 CGTTGCATTTCTGTTCTGTAAAAATGCAGCTCAGATTCTTTGTTTGA
 AAATTAGCGCTCTCGCGTTGCATTTTGTCTACAAAAATGAAGCACAG
 ATGCTTCGTT

120

-continued

SEQ. ID NO: 92

GCTAACAGCGCGATTGTGCTGGTGACCCAATGCGACCAGATGCTCCACG
 5 CCCAGTCCGCTACCGTCTTCATGGGAGAAAAATAACTGTTGATGGGT
 GTCTGGTCAGAGACATCAAGAAATAACGCCGAACATTAGTGCAGGCA
 GCTTCCACAGCAATGGCATCTGGTCATCCAGCGGATAGTTAATGATC
 10 AGCCCACTGACGCGTTGCGCGAGAAGATTGTGACCCGCGCTTTACAG
 GCTTCGACGCGCTTCGTTCTACCATCGACACCACCGCTGGCACCC
 AGTTGATCGCGCGAGATTTAATCGCCGCGACAATTTGCGACGCGCGC
 15 TGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTG
 CCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCC
 ATCGCCGCTTCCACTTTTTCCCGCTTTTCGCGAGAAAGTGGCTGGCC
 20 TGGTTCACACGCGGAAACGGTCTGATAAGAGACACCGGCATACTCT
 GCGACATCGTATAACGTTACTGGTTTACATTCAACACCTGAATTGA
 CTCTCTTCGCGCGCTATCATGCCATACCGGAAAGGTTTTGCGCCAT
 TCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCAT
 25 TAGGAAGCAGCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCG
 AAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGCCAC
 GGGGCTTGCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAA
 30 GTGGCGAGCCCGATCTTCCCATCGGTGATGTGCGGATATAGGCGCC
 AGCAACCGCACCTGTGGCGCCGATGTCGCGCCACGATGCGTCCGCG
 GTAGCCTAGGATCGAGATCGATCTCGATCCCGCGAAATTAACGACT
 35 CACTATAGGGGAATTGTGAGCGGATAACAATCCCCCTTAGAAATAAT
 TTTGTTTAACTTTAAGAAGGAGATATACATATGCGAGATCTCAATTGG
 ATATCGGCCGCGCACGCGATCGCTGACGTGCGTACCTCGAGTCTGGT
 40 AAAGAAACCGCTGCTGCGAAATTTGAACGCCAGCACATGGACTCGTCT
 ACTAGTCGCAGCTTAATTAACCTAAACTGCTGCCACCGCTGAGCAATA
 ACTAGCATAACCCCTTGGGGCTCTAAACGGGTCTTGAGGGGTTTTTT
 45 GCTAGCGAAAGGAGAGTCGACACTGCTTCCGGTAGTCAATAAACCGG
 TAAACCAGCAATAGACATAAGCGGCTATTTAACGACCTGCGCTGAAC
 CGACGACCGGGTCATCGTGGCCGGATCTTGGCGGCCCTCGGCTTGAAC
 50 GAATTGTTAGACATTATTTGCCGACTACCTGGTGATCTCGCCTTTCA
 CGTAGTGGACAAATCTTCCAACTGATCTGCGCGCGAGGCCAAGCGAT
 CTTCTTCTGTCCAAGATAAGCCTGTCTAGCTTCAAGTATGACGGGCT
 55 GATACTGGGCCGGCAGGCGCTCCATTGCCAGTCGGCAGCGACATCCT
 TCGGCGCGATTTTGCCGTTACTGCGCTGTACCAATGCGGGACAACG
 TAAGCACTACATTTGCTCATCGCCAGCCAGTCGGGCGCGAGTTCC
 ATAGCGTTAAGGTTTCATTTAGCGCTCAAATAGATCCTGTTCAAGAA
 60 CCGGATCAAAGAGTTCTCCGCGCTGGACCTACCAAGGCAACGCTAT
 GTTCTCTTGTCTTTGTGCAAGATAGCCAGATCAATGTCATCGTGG
 CTGGCTCGAAGATACCTGCAAGATGTCAATGCGCTGCCATTCTCCAA
 65 ATTGCACTTCGCGCTTAGCTGGATAACGCCACGGAATGATGTCGTCTG

121

-continued

GCACAACAATGGTGACTTCTACAGCGCGGAGAATCTCGCTCTCTCCAG
 GGGGAAGCCGAAGTTTCCAAAAGGTCGTTGATCAAAGCTCGCCGCGTTG
 TTTCATCAAGCCTTACGGTCACCGTAACAGCAAATCAATATCACTGT
 GTGGCTTCAGGCCCATCCACTGCGGAGCCGTACAAATGTACGGCCA
 GCAACGTCGGTTCGAGATGGCGCTCGATGACGCCAACTACCTCTGATA
 GTTGAGTCGATACTTCGGCGATACCGCTTCCCTCATACTCTTCTTT
 TCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGAT
 ACATATTTGAATGTATTTAGAAAAATAACAAATAGCCAGCTCACTCG
 GTCGTACGCTCCGGGCGTGAGACTGCGGCGGGCGCTGCGGACATA
 CAAAGTATCCCAAGATTCCGTGGATAAGCAGGGGACTAACATGTGAG
 GCAAAACAGCAGGGCCGCGCCGGTGGCGTTTTTCCATAGGCTCCGCCC
 TCCTGCCAGAGTTCACATAAACAGACGCTTTTCCGGTGCACTGTGGG
 AGCCGTGAGGCTCAACCATGAATCTGACAGTACGGGCGAAACCCGACA
 GGACTTAAAGATCCCCACCGTTTCCGGCGGGTCTCCTCTTGCGCT
 CTCCTGTTCCGACCTCGCGTTTACCGGATACCTGTTCCGCTTTCTC
 CCTTACGGGAAGTGTGGCGCTTCTCATAGCTCACACACTGGTATCTC
 GGCTCGGTGTAGGTCTGCTCCAAGCTGGGCTGTAAGCAAGAACTC
 CCCGTTACGCCCAGCTGCTGCGCTTATCCGGTAACTGTTCACTTGAG
 TCCAACCCGGAAGAACGCGTAAACGCCACTGGCAGCAGCCATTGGT
 AACTGGGAGTTCGAGAGGATTTGTTTAGCTAAACACGCGTTGCTCT
 TGAAGTGTGCGCCAAAGTCCGGCTACACTGGAAGGACAGATTTGGTTG
 CTGTGCTCTGCGAAAGCCAGTTACACGGTTAAGCAGTTCCCAACTG
 ACTTAACCTTCGATCAAACCACCTCCCAGGTGGTTTTTTCGTTTACA
 GGGCAAAAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTG
 ATCTTTTCTACTGAACCGCTCTAGATTTTCACTGCAATTTATCTCTCA
 AATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTAATCTCA
 TGTTAGTCATGCCCGCGCCACCGGAAGGAGCTGACTGGTTGAAGG
 CTCTCAAGGGCATCGGTGAGATCCCGGTGCTTAATGAGTGAGCTAAC
 TTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACC
 TGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGAGAGGCG
 GTTTGCGTATTGGGCGCCAGGTGGTTTTTCTTTTACCAGTGAGACG
 GGCAACAGCTGATTGCCCTTACCAGCTGGCCCTGAGAGAGTTGCAGC
 AAGCGGTCCACGCTGGTTTGGCCCAGCAGGCGAAATCCTGTTTGATG
 GTGGTTAACGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTAT
 CCCACTACCGAGATGTCGCCACCAACGCGCAGCCCGGACTCGGTAATG
 GCGCGCATTGCGCCAGCGCATCTGATCGTTGGCAACCAGCATCGCA
 GTGGGAACGATGCCCTCATTAGCATTTGCAATGGTTTGTGAAAACCG
 GACATGGCACTCCAGTCGCCCTCCCGTTCCGCTATCGGCTGAATTTGA
 TTGCGAGTGAGATATTTATGCCAGCGAGCCAGACGAGACGCGCCGAG
 ACAGAACTTAATGGGCC

122

-continued

SEQ. ID NO: 93

GGGGAATTGTGAGCGGATAACAATTCCCTCTAGAAATAATTTGTTT
 5 AACTTTAAGAAGGAGATATACATATGGCAGATCTCAATTGGATATCGG
 CCGGCCACGCGATCGCTGACGTGCGTACCTCTGAGTCTGGTAAAGAAA
 CCGCTGCTGCGAAATTTGAACGCCAGCACATGGACTCGTCTACTAGTC
 GCAGCTTAATTAACCTAAACTGCTGCCACCGCTGAGCAATAACTAGCA
 10 TAACCCCTTGGGGCTCTAAACGGGTCTTGAGGGGTTTTTTGCTAGCG
 AAAGGAGGAGTCTGACTATATCCGGATTGGCGAATGGGACGCGCCCTGT
 AGCGGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACC
 15 GCTACACTTGCCAGCGCCCTAGCGCCGCTCCTTTCTGCTTTCTTCCCT
 TCTTTTCTCGCCACGTTTCGCCGGCTTTCCCGCTCAAGCTCTAAATCGG
 GGGCTCCCTTTAGGGTTCGATTTAGTGCTTTACGGCACCTCGACCCC
 20 AAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGA
 TAGACGGTTTTTTCGCCCTTGACGTTGGAGTCCACGTTCTTTAATAGT
 GGACTCTTGTTCCAACTGGAACAACACTCAACCTATCTCGGTCTAT
 25 TCTTTTGATTTATAAGGGATTTTGCCGATTTTCGCCCTATTGGTTAAAA
 AATGAGCTGATTTAACAAAAATTTAACGCGAATTTAACAAAAATATTA
 ACGTTTACAATTTCTGGCGGCAGATGGCATGAGATTATCAAAAAGGA
 30 TCTTACCTAGATCCTTTTAAATTAATAAGTGTAAATCAATCT
 AAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCA
 GTGAGGCACCTATCTCAGCGATCTGTCTATTTCTGTTTATCCATAGTTG
 35 CCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCAT
 CTGGCCCCAGTGCTGCAATGATACCGCAGACCCACGCTCACCGGCTC
 CAGATTTATCAGCAATAAACAGCCAGCCGGAAGGGCCGAGCGCAGAA
 40 GTGGTCTGCAACTTTATCCGCTCCATCCAGTCTATTAATGTGTGCC
 GGGGAAGCTAGAGTAAGTAGTTTCGCCAGTTAATAGTTTTCGCAACGTTG
 TTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGG
 45 CTTTATTAGCTCCGGTTCCTAACGATCAAGCGAGTTACATGATCCC
 CCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGTTG
 TCAGAAAGTAAGTTGGCCGAGTGTATCACTCATGGTTATGGCAGCAC
 TGCATAATTCTCTTACTGTATGCCATCCGTAAGATGCTTTTCTGTGA
 50 CTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGCGCAC
 CGAGTTGCTCTTGCCCGGCTCAATACGGGATAATACCGCGCCACATA
 GCAGAACTTTAAAGTGCTCATATTGGAACAGTTCTTCGGGGCGAA
 55 AACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTGATGTAACCCA
 CTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTTACCAGCGTTT
 CTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAATAA
 60 GGGCGACACGGAAATGTTGAATACTACACTCTTCTTTTCAATCAT
 GATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTT
 GAATGTATTTAGAAAAATAACAAATAGGTCATGACCAAAATCCCTTA
 65 ACGTGAGTTTTCTGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAA

123

-continued

AGGATCTTCTTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCA
 AACAAAAAACACCGCTACAGCGGTGGTTTGTGGCGGATCAAGA
 GCTACCAACTCTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCAGAT
 ACCAAATACTGTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAA
 GAACTCTGTAGCACCGCTACATACCTCGCTCTGCTAATCTGTGTACC
 AGTGGCTGCTGCCAGTGGCGATAAGTCTGTCTTACCGGGTTGGACTC
 AAGACGATAGTTACCGGATAAGGCGCAGCGGTGGGGCTGAACGGGGGG
 TTCGTGCACACAGCCAGCTTGGAGCGAACGACCTACACCGAACTGAG
 ATACCTACAGCGTAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAG
 AAAGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCG
 CACGAGGGAGCTTCCAGGGGGAACGCGCTGTATCTTTATAGTCTGT
 CGGGTTTCCGCACTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTC
 AGGGGGGCGGAGCCTATGAAAAACGCCAGCAACGCGGCCTTTTACG
 GTTCTTGCCCTTTTGTGCGCTTTTGTCTACATGTTCTTCTCGCTT
 ATCCCTGATTCTGTGGATAACCGTATTACCGCTTTGAGTGAGCTGA
 TACCGCTCGCCGACGCCGAACGCCAGCGCAGCGAGTCACTGAGCGA
 GGAAGCGGAAGAGCGCCTGATGCGGTATTTCTCTCTACGCATCTGTG
 CGGTATTTACACCGCATATATGGTGCACCTCTCAGTACAATCTGCTCT
 GATGCGGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACT
 GGGTCATGGCTGCGCCCGACACCCGCCAACCCGCTGACGCGCCCT
 GACGGGCTTGCTGCTCCCGCATCCGCTTACAGACAAGCTGTGACCG
 TCTCCGGGAGCTGCATGTGTGAGAGGTTTACCGTCATCACCGAAAC
 GCGCAGGCAGCTGCGCTAAAGCTCATCAGCGTGGTCGTGAAGCGATT
 CACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCA
 GAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGG
 TTTTTCTCTGTTTGGTCACTGATGCTCCGTGTAAGGGGATTCTGT
 TCATGGGGTAAATGATACCGATGAAACGAGAGAGGATGCTCACGATAC
 GGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTA
 AACAACTGGCGGTATGGATGCGGCGGGACAGAGAAAAATCACTCAGG
 GTCAATGCCAGCGCTTCGTAAATACAGATGTAGGTGTTCCACAGGGTA
 GCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCG
 CTGACTTCGCGTTTCCAGACTTTACGAAACCGGAAACCGAAGACCA
 TTCATGTTGTTGCTCAGGTGCGAGACGTTTTGCGAGCAGCTCGCTTC
 ACGTTGCTCGCTATCGGTGATTCTCTGCTAACAGTAAGGCAAC
 CCCGCCAGCTAGCCGGTCTCAACGACAGGAGCACGATCATGCTAG
 TCATGCCCCGCCCCACCGGAAGGAGCTGACTGGGTGAAGGCTCTCA
 AGGGCATCGGTGAGATCCCGGTGCCATAATGAGTGAGCTAACTTACAT
 TAATTGCGTTGCGCTCACTGCCCGCTTCCAGTGGGAAACCTGTGCT
 GCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGC
 GTATTGGGCGCCAGGGTGGTTTTCTTTTACCAGTGAGACGGGCAAC
 AGCTGATTGCCCTTACCAGCTGCGCTGAGAGAGTTGCGAGCAAGCGG

124

-continued

TCCACGCTGGTTTGCCCCAGCAGGCGAAAACTCTGTTTGATGGTGGTT
 AACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACT
 5 ACCGAGATGTCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGC
 ATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGA
 ACGATGCCCTCATTCAGCATTTGCATGGTTTGTGAAAAACCGGACATG
 10 GCATCCAGTCGCCTTCCCGTTCGCTATCGGCTGAATTTGATTGCGA
 GTGAGATATTTATGCCAGCCAGCCAGACGACGCGCCGAGACAGAA
 CTTAATGGGCCCCGTAACAGCGCGATTTGCTGGTGACCCAATGCGACC
 15 AGATGCTCCACGCCAGTCGCGTACCGTCTTCATGGGAGAAAAATA
 CTGTGTATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACA
 TTAGTGACGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGA
 20 TAGTTAATGATCAGCCACTGACGCGTTGCGCGAGAAGATTGTGCACC
 GCGCTTTACAGGCTTCGACGCGCTTCGTTCTACCATCGACACCACC
 ACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATT
 25 TGCAGCGGCGCTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGC
 AACGACTGTTTGCCCGCAGTTGTTGTGCCACGCGGTTGGGAATGTAA
 TTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTTCGAGAA
 30 ACGTGGCTGGCTGGTTACACGCGGGAACGGTCTGATAAGAGACA
 CCGGCATACTCTGCGACATCGTATAACGTTACTGTTTACATTACC
 ACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATAACCGGAAAG
 35 GTTTTGCGCCATTTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATG
 CGACTCTGCTAGGAAGCAGCCAGTAGTAGTTGAGGCGGTTGAG
 CACCGCCGCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAG
 40 TCCCCCGGCCACGGGGCTGCCACCATAACCCACGCCGAAACAGCGCT
 CATGAGCCCCAAGTGGCGAGCCCGATCTTCCCATCGGTGATGTGCGC
 GATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGTGCGGCCAC
 45 GATGCGTCCGCGTAGCCTAGGATCGAGATCGATCTCGATCCCGCGAA
 ATTAATACGACTCACTATA
 SEQ. ID NO: 94
 ATCCGATATAGTTCTCTCTTTCAGCAAAAAACCCCTCAAGACCCGTT
 50 TAGAGGCCCCAAGGGTTATGCTAGTTATTGCTCAGCGGTGGCAGCAG
 CCAACTCAGCTTCCTTTTCGGGCTTGTGTAGCAGCCGATCTCAGTGGT
 GGTGGTGGTGGTGTGCTGAGTGCAGCCGCAAGCTTGTGACGAGACTCG
 55 AATTCCGATCCGCGACCCATTTGCTGTCCACCAGTCATGCTAGCCATA
 TGGCTGCCGCGCGCACAGGCGGCTGCTGTGATGATGATGATGATGG
 CTGCTGCCCATGGTATATCTCTTCTTAAAGTTAAACAAAATTATTTTC
 TAGAGGGGAATTGTTATCCGCTCACAATTCCCTATAGTGAGTCGTAT
 60 TAATTTGCGGGATCGAGATCTCGATCCTCTACGCCGACGCATCGTG
 GCCGCATCACCGCGCCACAGGTGCGGTTGCTGCGCCCTATATCGCC
 GACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGC
 65 GCTTGTTCGCGGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTG

125

-continued

TTGGGCGCCATCTCCTTGCATGCACCATTCCTTGC GCGCGGCGGTGCTC
 AACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCAT
 AAGGGAGAGCGTCGAGATCCCGACACCATCGAATGGCGCAAAACCTT
 TCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGT
 GAATGTGAAACCAGTAACGTTATACGATGTCGAGAGTATGCCGGTGT
 CTCTTATCAGACCGTTTCCCGCGTGGTGAACAGGCCAGCCACGTTTC
 TCGCAAAACGCGGGAAAAAGTGAAGCGCGCATGGCGGAGCTGAATTA
 CATTCCCAACCGCGTGGCACAACAACTGGCGGGCAACAGTCGTTGCT
 GATTGGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCGTCGCAAT
 TGTGCGGCGGATTAAATCTCGCGCCGATCAACTGGGTGCCAGCGTGGT
 GGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGT
 GCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCC
 GCTGGATGACCAGGATGCCATTGCTGTGAAGCTGCCTGCACTAATGT
 TCCGCGCTTATTTCTTGATGTCTCTGACCAGACACCCATCAACAGTAT
 TATTTTCTCCCATGAAGACGCTACGCGACTGGGCGTGGAGCATCTGGT
 CGCATTGGGTCAACAGCAAATCGCGCTGTTAGCGGGCCCATTAAGTTC
 TGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCG
 CAATCAAATTCAGCCGATAGCGGAACGGAAGGCGACTGGAGTGCCAT
 GTCCGGTTTTCAACAAACCATGCAAAATGCTGAATGAGGGCATCGTTCC
 CACTGCGATGCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAATGCG
 CGCCATTACCGAGTCCGGGCTGCGCGTTGGTGGCATATCTCGTAGT
 GGGATACGACGATACCGAAGACAGCTCATGTTATATCCCGCGTTAAC
 CACCATCAACAGGATTTTTCGCTGCTGGGGCAACAGCGTGGACCG
 CTTGTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTT
 GCCCCCTCTCACTGGTGAAAGAAAAACACCTGGCGCCCAATACGCA
 AACCGCCTCTCCCGCGCGTTGGCCGATTCTAATGCAGCTGGCAG
 ACAGGTTTCCCGACTGGAAGCGGGCAGTGAGCGCAACGCAATTAATG
 TAAGTTAGCTCACTCATAGGACCGGATCTCGACCGATGCCCTTGA
 GAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTA
 TCGTCGCGCACTTATGACTGTCTTCTTATCATGCAACTCGTAGGAC
 AGGTGCCGGCAGCGCTCTGGGTCAATTTTCGGCGAGGACCGCTTTCGCT
 GGAGCGCGACGATGATCGGCTGTGCTTGGCGTATTCGGAATCTTGC
 ACGCCCTCGCTCAAGCCTTCGTCAGTGGTCCCGCACCAACAGTTTCG
 GCGAGAAGCAGGCCATTATCGCCGCGATGGCGGCCCCACGGGTGCGCA
 TGATCGTGCTCCTGTGTTGAGGACCCGGCTAGGCTGGCGGGGTGGC
 TTACTGGTTAGCAGAATGAATCACCGATACGCGAGCGAACGTGAAGCG
 ACTGTGCTGCAAAACGCTGCGACCTGAGCAACACATGAATGCTCT
 TCGGTTTCCGTGTTTCGTAAAGTCTGGAACGCGGAAGTCAGCGCCCT
 GCACCATATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCCCT
 GTGGAACACCTACATCTGTATTAACGAAGCTTACCCTCACAACGTTCC

126

-continued

AGTAACCGGGCATGTTTCATCATCAGTAACCCGTATCGTGAGCATCCTC
 TCTCGTTTCATCGGTATCATTACCCCATGAACAGAAATCCCCCTTAC
 5 ACGGAGGCATCAGTGACCAAACAGGAAAAACCGCCCTTAACATGGCC
 CGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAACTCAACGAG
 CTGGACGCGGATGAACAGGCAGACATCTGTGAATCGCTTCAGGACCAC
 10 GCTGATGAGCTTTACCGAGCTGCCTCGCGGCTTTCGGTGATGACGGT
 GAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTG
 TAAGCGGATGCCGGGAGCAGACAAGCCGTCAGGGCGCGTCAGCGGGT
 15 GTTGGGGGGTGTGGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGC
 GGAGTGTATACTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGA
 GAGTGCAACATATATGCGGTGTGAATAACCGCACAGATGCGTAAGGAG
 20 AAAATACCGCATCAGGCGCTCTTCCGCTTCTCGCTCACTGACTCGCT
 GCGCTCGGTGCTTCCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGC
 GGTAATACGGTTATCCACAGAATCAGGGGATAACGAGGAAGAATCAT
 GTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCTT
 25 GCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAA
 TCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAGATA
 CCAGGCGTTTTCCCCCTGAAGCTCCCTCGTGCGCTCTCCTGTTCCGAC
 30 CCTGCGGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGT
 GCGCTTTCTCATAGCTCAGCTGTAGGTATCTCAGTTCCGTGTAGGT
 CGTTGCTCCCAAGCTGGGCTGTGTGACGAACCCCGCTTACGCCGA
 35 CCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAG
 ACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAG
 AGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCTTAA
 40 CTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAA
 GCCAGTTACCTTCGGAAGAGTTGGTAGCTCTTGATCCGGCAACA
 AACCACGCTGGTAGCGGTGGTTTTTTTGGTTGCAAGCAGCAGATTAC
 45 GCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTCTACGGG
 GTCTGACGCTCAGTGAACGAAACTCAGTTAAGGGATTTTGGTCAT
 GAACAATAAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTA
 50 TGAGCCATATTCAACGGGAAACGCTCTGCTCTAGGCCGCGATTAAATT
 CCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATG
 TCGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATG
 55 CGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATG
 TTACAGATGAGATGGTCAGACTAACTGGCTGACGGAATTTATGCCTC
 TTCCGACCATCAAGCATTTTATCCGTACTCTGATGATGCATGTTTAC
 TCACCACTGCGATCCCGGGAAAAACAGCATTCAGGTATTAGAAGAAT
 60 ATCCTGATTACAGGTGAAATATTGTTGATGCGCTGGCAGTGTTCTCTGC
 GCCGGTTGCATTTCGATTCTGTTTGTAAATGTCCTTTTAAACAGCGATC
 GCGTATTTCTGCTCTGCTCAGGCGCAATCACGAATGAATAACGGTTTGG
 65 TTGATGCGAGTGATTTTGTATGACGAGCGTAATGGCTGGCCTGTTGAAC

127

-continued

AAGTCTGGAAGAAATGCATAAACTTTTGCCATTCTACCGGATTTCAG
 TCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTGTACGAGG
 GGAAATTAATAGGTGTATTGATGTTGGACGAGTCGGAATCGCAGACC
 GATACCAGGATCTTGCCATCTATGGAAGTGCCTCGGTGAGTTTCTC
 CTTCAATTACAGAAACGGCTTTTCAAAAATATGGTATTGATAATCCTG
 ATATGAATAAATTGCGATTTTCATTGATGCTCGATGAGTTTCTAAG
 AATTAATTGATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAA
 CAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGAAATT
 GTAAACGTTAATATTTTGTAAAAATCGCGTTAAATTTTGTAAATC
 AGCTCATTTTAAACCAATAGGCGGAAATCGGCAAAATCCCTTATAAA
 TCAAAAGAATAGACCGAGATAGGGTTGAGTGTGTTCCAGTTTGGAAC
 AAGAGTCCACTATTAAAGAAGTGGACTCCAACGCTAAGGGCGGAAAA
 ACCGCTCATCAGGGCGATGGCCACTACGTGAACCATCACCTAATCA
 AGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCTTAAA
 GGGAGCCCCGATTAGAGCTTGACGGGAAAGCGGCGAAGTGGCG
 AGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGTAGGGCGCTGGCA
 AGTGTAGCGGTACGCTGCGCGTAACCACCACACCCGCGCGCTTAAT
 GCGCGCTACAGGGCGCGTCCCATTCGCCA

SEQ. ID NO: 95

ATGGCTTCTAGTTCTTCCGATGTCTTCGTTTGGGTCTAGGTGTTGTT
 TTGGCTGCCTTGATATCTTCAGAGACCAATTATTCGCTGCTTCTAAG
 CCAAAGGTGGCTCCAGTTTCCACTACGAAGCCTGCCAACGGTTCCGCT
 AACCCAGAGACTTTCATCGCCAAGATGAAACAAGGTAAGAAGAGAATC
 GTAATCTTCTACGGTTCTCAAACCTGGTACCGCTGAAGAATATGCTATT
 CGTTTGGCTAAGGAAGCTAAGCAAAAGTTCCGTCTAGCCTCCTTGGTT
 TGTGATCCAGAAGAATACGATTTTGAAAGTTGGACCAATTGCCAGAA
 GATTCTATTGCTTTCTTCGTCGTTGCTACCTATGGTGAAGGTGAACCT
 ACAGACAACGCTGTCCAATTGTTGCAAACTTGCAAGATGAAAGCTTC
 GAATTCTCCTCTGGTGAGAGAAAGTTGTGAGTTTGAAGTACGTTGTT
 TTTGGTCTGGGTAACAAGACCTACGAACATTACAACCTCATTGGGAGA
 ACTGTTGACGCTCAATTGGCCAAGATGGGTGCTATCAGAATCGGTGAA
 AGAGGTGAAGGTGATGATGACAAGTCCATGGAAGAAGACTACTTGAA
 TGGAAGGATGGTATGTGGAAGCGTTTGCCACTGCTATGGGTGTTGAA
 GAAGGTCAAGGTGGTGAAGTCCGCTGATTTCTGCTGTTTCCGAATTGGAA
 TCTCAACCCACAGAAAAGTTTACCAAGGTGAATTTCTGCTAGAGCT
 TTAACCAAAACCAAGGTTATTCACGACGCTAAGAATCCTTTTGTGCT
 CCAATTGCGGTTGCTAGAGAATTGTTCCAATCTGTTGTCGATAGAAAC
 TGTGTCCACGTCGAATTCAACATTGAAGGCTCTGGTATCACCTATCAA
 CACGGTGACCAAGTGGTTTGTGGCCATTGAATCCAGATGTTGAAGTC
 GAACGGTTGTTGTGTTGTTTAGGTTTAGCTGAAAAGAGAGATGCTGTC
 ATCTCCATTGAATCCTTAGACCCGCTTTGGCTAAGGTTCCATTCCTCA

128

-continued

GTCCCAACTACTTACGGTGCTGTGTTGAGACACTACATTGACATCTCT
 GCTGTGCGCCGTAGACAAATCTTGGGTACTTTGTCCAAATTCGCTCCA
 5 ACCCCAGAAGCTGAAGCTTTCTTGAGAACTTGAACACTAACAGGAA
 GAATACCACAACGTCGTCGCTAACGGTTGTTTGAATTTGGGTGAAATT
 TTGCAATCGCTACCGGTAACGACATTACTGTCCACCAACTACTGCC
 10 AACACCACCAATGGCCAATTCCATTGCACATCATTGTTTCTGCCATC
 CCAAGATTGCAACCAAGATACTACTCTATCTCTTCTTCCCCAAAAATT
 CATCCAAACACCATCCACGCTACCGTTGTTGTGCTCAAATACGAAAAAC
 15 GTTCCAACCGAACCAATCCCAAGAAAGTGGGTTTACGGTGTGCGTAGT
 AACTCTTGTGTAATTTAAAGTACGCTGTTAACAAGGAACCAAGTTCCA
 TACATCACTCAAAATGGCGAACAAAGAGTCGGTGTCCCGGAATACTTG
 20 ATTGCTGGTCCACGTGGTTCTTACAAGACTGAATCTTTCTACAAGGCT
 CCAATCCATGTTAGACGTTCTACTTTCCGTTTGCCAACCAACCCAAAG
 TCTCCAGTCATCATGATTGGTCCAGGTACTGGTGTGCGCCCATTCAGA
 25 GGCTTCGTTCAAGAAAGAGTTGCCTTGCCAGAAAGATCCATCGAAAAG
 AACGGTCTGACTCTTTGGCTGACTGGGGTCGTATTTCTTGTCTTAC
 GGTGTGTAAGATCCGACGAAGACTTCTTGTAAGGACGAATGGCCA
 30 CAATACGAAGCTGAGTTGAAGGTAAGTTCAAGTTGCACTGTGCTTTC
 TCCAGACAAAACCTACAAGCCAGACGGTTCTAAGATTTACGTCCAAGAT
 TTGATCTGGGAAGACAGAGAACACATTGCCGATGCCATCTTAAACGGT
 35 AAGGGTTACGTCTACATCTGCGGTGAAGCTAAGTCCATGTCTAAACAA
 GTTGAAGAAGTTCTAGCCAAGATCTTGGGCGAAGCCAAAGGTGTTCC
 GGTCCAGTTGAAGGTGTTGCTGAAGTCAAGTTACTGAAGGAACGGTCC
 40 AGATTGATGTTGGATGTCTGGTCTTGA

SEQ. ID NO: 96

MASSSDVFLGLGVLAALYIFRDQLFAASKPKVAPVSTTKPANGSA
 NRPDFAIKMKQKKRIVIFYGSQTGAEEYAIRLAKAKQKQFGLASLV
 45 CDPEEYDFEKLQDLPEDSIAFFVVATYGEGETDNAVQLQLQDSEF
 EFSSGERKLSGLKYVVFGLGNKTYEHYNLIGRTVDAQLAKMGAIRIGE
 RREGDDDKSMEEDYLEWKDGMWEAFATAMGVEEGQGGDSADFVSELE
 50 SHPPEKVYQGEFSARALTKGIHDAKNPFAAPIAVARELQSVVDRN
 CVHVEFNIEGSGITYQHDHVLWPLNPDVEVERLLCVLGLAEKRDAV
 ISIESLDPALAKVPPVPTTYGAVLRHYIDISAVAGRQILGLTKFAP
 55 TPFAEAFRLNLTNKEEYHNVVANGCLKLGEILQIATGNDITVPPTTA
 NTKWPIPFDIIVSAIPRLQPRYSSISSPKIHNPITHATVVVLKYEN
 VPTEPIPRKWVYGVGSNFLNLKYAVNKEPVYITQNGEQRVGVPEYL
 60 IAGPRGSYKTESFYKAPIHVRRTFRLPTNPKSPVIMIGPGTVAPFR
 GFVQERVALARRSIEKNGPDSLADWGRISLFYGCRRSDEDFLYKDEWP
 QYEAEKLGKFKLHCAFSRQNYKPDGSKIYVQDLIWEDEHRIADAILNG
 KGYVYICGEAKSMSKQVEEVLAAILGEAKGSGPVEGVAEVKLLKERS
 65 RLMLDVWS

129

-continued

SEQ. ID NO: 97
GAAGAATTGTGAGCGGATAACAATTTCCCTCTAGAAATAATTTGTTT
AACTTTAAGAAGGAGATATACATATGGCAGATCTCAATTGGATATCGG
CCGGCCACGCGATCGCTGACGTCGGTACCCTCGAGTCTGGTAAAGAAA
CCGCTGCTGCGAAATTTGAACGCCAGCACATGGACTCGTCTACTAGTC
GCAGCTTAATTAACCTAAACTGCTGCCACCGCTGAGCAATAACTAGCA
TAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTGTCTAGCG
AAAGGAGGAGTCGACTATATCCGATTGGCGAATGGGACGCGCCCTGT
AGCGGCGCATTAAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACC
GCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCT
TCCTTTCTCGCCACGTTTCGCCGGCTTTCCCGTCAAGCTCTAAATCGG
GGGCTCCCTTTAGGGTTCGATTAGTGCTTTACGGCACCTCGACCCC
AAAAAATTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGA
TAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGT
GGACTCTTGTTCACAACTGGAACAACACTCAACCTATCTCGGTCTAT
TCTTTTGATTATAAGGGATTTTGGCGATTTCGGCCTATTGGTTAAAA
AATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTA
ACGTTTACAATTTTCGGCGGCACGATGGCATGAGATTATCAAAAAGGA
TCTTCACCTAGATCCTTTTAAATAAAAATGAAGTTTTAAATCAATCT
AAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAATCA
GTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTTCATCCATAGTTG
CCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCAT
CTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTC
CAGATTTATCAGCAATAAACAGCCAGCCGAAGGGCCGAGCGCAGAA
GTGGTCTGCAACTTTATCCGCCCTCATCCAGTCTATTAATTGTTGCC
GGGAAGCTAGAGTAAGTAGTTCCGCAGTTAATAGTTTGCACAACGTTG
TTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTGGTTGGTATGG
CTTCATTGAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCC
CCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTG
TCAGAAGTAAGTTGGCGCAGTGTTATCACTCATGGTTATGCGAGCAC
TGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGA
CTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGAC
CGAGTTGCTCTTGCCGCGCTCAATACGGGATAATACCGCGCCACATA
GCAGAACTTTAAAGTGCTCATCATTTGGAACCGTTCTTCGGGGCGAA
AACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCA
CTCGTGCAACCACTGATCTTCAGCATCTTTTACTTTTACCAGCGTTT
CTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAATAA
GGGCGACACGGAATGTTGAATACTCATACTCTTCTTTTCAATCAT
GATTGAAGCATTATCAGGGTTATTGTCTCATGAGCGGATACATATTT
GAATGTATTTAGAAAAATAACAAATAGGTATGACCAAAATCCCTTA
ACGTGAGTTTTCTGTTCACTGAGCGTCAGACCCCGTAGAAAAGATCAA

130

-continued

AGGATCTTCTTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCA
AACAAAAAAACCACCGCTACCAGCGGTGGTTTGGTTGCCGGATCAAGA
GCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGAT
ACCAAATACTGTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAA
GAACTCTGTAGCACCGCCTACATACTCGCTCTGCTAATCCTGTTACC
AGTGGCTGCTGCCAGTGGCGATAAGTCGTGCTTACCAGGTGGACTC
AAGACGATAGTTACCGGATAAGGCGCAGCGGTGGGGTGAACGGGGG
TTCGTGCACACAGCCAGCTTGAGCGAACGACCTACACCGAACTGAG
ATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAG
AAAGCGGACAGGTATCCGGTAAGCGGAGGGTCGGAACAGGAGAGCG
CACGAGGGAGCTTCCAGGGGAAACGCTGGTATCTTTATAGTCCTGT
CGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTC
AGGGGGGCGAGCCTATGAAAAACGCCAGCAACCGCGCTTTTACG
GTTCTCGGCTTTTGTGCGCTTTTGTCTCACATGTTCTTCTCGCTT
ATCCCCGTATCTGTGGATAACCGTATTACCGCTTTGAGTGAGCTGA
TACCGCTCGCCGACGCCAAGCAGCAGCGCAGGTCAGTGAGCGA
GGAAGCGGAAGAGCGCTGATGCGGTATTTCTCCTTACGCATCTGTG
CGGTATTTACACCCGATATATGGTGCACTCTCAGTACAATCTGCTCT
GATGCCGATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACT
GGGTGATGGCTGCGCCCCGACACCCGCCAACCCGCTGACGCGCCCT
GACGGGCTTGTCTGCTCCCGCATCCGCTTACAGACAAGCTGTGACCG
TCTCCGGGAGCTGCATGTGTGAGAGTTTTTACCCTCATCACCGAAAC
GCGCGAGGACGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATT
CACAGATGTCTGCTGTTCATCCGCTCCAGCTCGTTGAGTTTCTCCA
GAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGG
TTTTTCTGTTTGGTCACTGATGCCCTCCGTGTAAGGGGGATTCTGT
TCATGGGGTAAATGATACCGATGAAACGAGAGAGGATGCTCACGATAC
GGGTACTGATGATGAACATGCCCGTTACTGGAACGTTGTGAGGGTA
AACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGG
GTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTA
GCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGAGGGCG
CTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAACCGAAGACCA
TTCATGTTGTTGCTCAGGTGCGAGACGTTTTCGAGCAGCAGTCGCTTC
ACGTTTCGCTCGCGTATCGGTGATTCACTTCTGCTAACCAAGTAAGCAAC
CCCGCCAGCCTAGCCGGTCTCAACGACAGGAGCACGATCATGCTAG
TCATGCCCCGCGCCACCGGAAGGAGCTGACTGGGTGAAGGCTCTCA
AGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACAT
TAATTGCGTTGCGCTCACTGCGCGCTTTCCAGTCGGGAAACCTGTGCT
GCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGCGGTTTGC
GTATTGGGCGCCAGGGTGGTTTTTCTTTTACCAGTGAGACGGGCAAC

131

-continued

AGCTGATTGCCCTTCACCGCTGGCCCTGAGAGAGTTGCAGCAAGCGG
 TCCACGCTGGTTTGCCCCAGCAGCGGAAATCCTGTTGATGGTGGTT
 AACGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACT
 ACCGAGATGTCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGC
 ATTGCGCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGA
 ACGATGCCCTCATTAGCATTTGCATGGTTTGTGAAAACCGGACATG
 GCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGA
 GTGAGATATTTATGCCAGCCAGCCAGACGCGAGACGCGCGAGACAGAA
 CTTAATGGGCCCGCTAACAGCGCGATTGCTGGTGACCAATGCGACC
 AGATGCTCCACGCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATA
 CTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACA
 TTAGTGACAGCGAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGA
 TAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACC
 GCCGCTTTACAGGCTTCGACGCGCGTTCGTTCTACCATCGACACCACC
 ACGCTGGCACCAGTTGATCGCGCGAGATTTAATCGCCGCGACAATT
 TGCAGCGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGC
 AACGACTGTTTGCCCGCAGTTGTTGTGCCACGCGTTGGGAATGTAA
 TTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTTCGAGAA
 ACGTGGCTGGCCTGGTTCAACACGCGGAAACGGTCTGATAAGAGACA
 CCGGCATACTCTCGACATCGTATAACGTTACTGGTTTCACATTACCC
 ACCCTGAATTGACTCTCTTCCGGCGCTATCATGCCATACCGCGAAAG
 GTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATG
 CGACTCCTGCATTAGGAAGCAGCCAGTAGTAGGTTGAGGCCGTTGAG
 CACCGCGCGCGCAAGGAATGGTGATGCAAGGAGATGGCGCCCAACAG
 TCCCCCGGCCACGGGGCTGCCACCATAACCCACGCCGAAACAAGCGCT
 CATGAGCCCGAAGTGGCGAGCCGATCTTCCCATCGGTGATGTCCGC
 GATATAGGCGCCAGCAACCGCACCTGTGGCGCGGTGATGCCGCCAC
 GATGCGTCCGGCTAGCCTAGGATCGAGATCGATCTCGATCCCGCGAA
 ATTAATACGACTCACTACG

SEQ. ID NO: 98

CGCTAACAGCGCGATTGCTGGTGACCAATGCGACCAGATGCTCCAC
 GCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAACTGTTGATGGG
 TGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGACGGC
 AGCTTCCACAGCAATGGCATCTCGGTATCCAGCGGATAGTTAATGAT
 CAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCGCTTTACA
 GGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACC
 CAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCACGGCGC
 GTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTT
 GCCCCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGC
 CATCGCCGCTTCCACTTTTTCCCGCGTTTTTCGAGAAACGTGGCTGGC
 CTGGTTCAACACGCGGAAACGGTCTGATAAGAGACACCGGCATACTC

132

-continued

TGCGACATCGTATAACGTTACTGGTTTCACATTCAACCACCTGAATTG
 ACTCTCTTCCGGGCGCTATCATGCCATAACCGGAAAGGTTTTGCGCCA
 5 TTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCTGCA
 TTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCCGCCCG
 CAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCGGCCA
 10 CGGGGCTGCCACCATAACCCACGCCGAAACAAGCGCTCATGAGCCGA
 AGTGGCGAGCCCGATCTTCCCATCGGTGATGTGGCGATATAGGCGC
 CAGCAACCGCACCTGTGGCGCCGGTGTGCGGCCACGATCGCTCCGG
 15 CGTAGCCTAGGATCGAGATCGATCTCGATCCCGGAAATTAATACGAC
 TCACTATAGGGGAATTGTGAGCGGATAACAATTCCCTCTAGAAATAA
 TTTTGTTTAACTTTAAGAAGGAGATATACATATGCGAGATCTCAATTG
 GATATCGGCCGCGCCACGCGATCGCTGACGTGGTACCCTCGAGTCTGG
 20 TAAAGAAACCGCTGCTGCGAAATTTGAACGCCAGCACATGGACTCGTC
 TACTAGTCGCAGCTTAATTAACCTAACTGCTGCCACCGCTGAGCAAT
 AACTAGCATAACCCCTTGGGGCTCTAAACGGGTCTTGGGGGTTTTT
 25 TGCTAGCGAAAGGAGAGTCGACAAGCTGACGACCGGGTCTCCGCAAG
 TGGCACTTTTGGGGAAATGTGCGCGGAACCCCTATTGTGTTATTTTT
 CTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAA
 30 ACTCATCGAGCATCAAATGAAACTGCAATTTATTATATCAGGATTAT
 CAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACT
 CACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGA
 35 TTCCGACTCGTCCACATCAATACAACCTATTAAATTTCCCTCGTCAA
 AAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCG
 GTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCACAG
 40 GCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAACCGT
 TATTCAATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGTCTGCTGT
 TAAAGGACAATTACAAACAGGAATCGAATGCAACCGCGCAGGAACA
 45 CTGCCAGCGCATCAACAATATTTTACCTGAATCAGGATATTCTTCTA
 ATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAATCAACATG
 CATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAA
 50 ATTCGCTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGG
 CAACGCTACCTTTGCCATGTTTCAGAAACAACCTCTGGCGCATCGGGCT
 TCCCATACAATCGATAGATTGTGCGACCTGATTGCCCACATTATCGC
 55 GAGCCCATTTATACCATATAAAATCAGCATCCATGTTGGAATTTAATC
 GCGGCCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATACTCTTCC
 TTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCG
 GATACATATTTGAATGTATTTAGAAAAATAAAACAATAGGCATGCAGC
 60 GCTCTTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTGCTTCGAC
 TGCGCGAGCGGTGTGAGCTCACTCAAAAGCGGTAATACGGTTATCCA
 CAGAATCAGGGGATAAAGCGGAAAGAACATGTGAGCAAAAAGCAAAG
 65 CACCGGAAGAAGCCAACGCCGCGAGCGTTTTTTCATAGGCTCCGCCCC

133

-continued

CCTGACGAGCATCACAAAATCGACGCTCAAGCCAGAGGTGGCGAAAC
 CCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTC
 GTGCGCTCTCTGTTCCGACCTGCGCTTACCGGATACCTGTCCGCC
 TTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTGG
 TATCTCAGTTCGGTGTAGGTGCTTCGCTCCAGCTGGGCTGTGTGCAC
 GAACCCCCGTTTCAGCCGACCGCTGCGCTTATCCGGTAACATATCGT
 CTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCC
 ATTGGTAACTGATTTAGAGGACTTTGTCTTGAAGTTATGCACCTGTTA
 AGGCTAACTGAAAGAACAGATTTTGGTGTAGTGGTCTCCAAACCCA
 CTTACCTTGGTTCAAAGAGTTGGTAGCTCAGCGAACCTTGAGAAAACC
 ACCGTTGGTAGCGGTGGTTTCTTTATTTATGAGATGATGAATCAAT
 CGGTCTATCAAGTCAACGAACAGCTATTCGGTTACTCTAGATTTTCAGT
 GCAATTTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGATA
 TAAGTTGTAATTTCTCATGTTAGTCATGCCCCGCGCCACCGGAAGGAG
 CTGACTGGGTGAAGGCTCTCAAGGCGATCGGTTCGAGATCCCGGTGCC
 TAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCT
 TTCCAGTCGGGAAACCTGTCTGTCAGCTGCATTAATGAATCGGCCAA
 CGCGCGGGAGAGGCGGTTTTCGTATTGGGCGCCAGGGTGGTTTTTCT
 TTTCACCACTGAGACGGGCAACAGCTGATTGCCCTTACCCTGGGCC
 CTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGGCCAGCAGGCG
 AAAATCCTGTTTGTAGTGGTGTAAACGGCGGATATAACATGAGCTGTC
 TTCGTATCGTCGTATCCCACTACCGAGATGTCGCACCAACGCGCAG
 CCCGACTCGGTAATGGCGCGCATTGCGCCAGCGCCATCTGATCGTT
 GGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTGACGATTTGCAT
 GGTGTTGTTGAAAACCGGACATGGCACTCCAGTCGCTTCCGCTTCCGC
 TATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAG
 ACGCAGACGCGCCGAGACAGAACTTAATGGGCC

Hereinafter are provided examples of specific implementations for performing the methods of the present disclosure, as well as implementations representing the compositions of the present disclosure. The examples are provided for illustrative purposes only, and are not intended to limit the scope of the present disclosure in any way.

EXAMPLES

Example 1—Process for Making a First Halogenated Psilocybin Derivative from Halogenated Tryptophan Feedstock

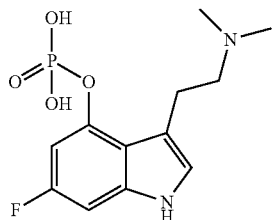
Yeast (*Saccharomyces cerevisiae*) was genetically engineered to enable bioconversion of commercially obtained, halogenated, simple indole, tryptophan, or tryptamine feedstock to generate final product. The parent yeast (*Saccharomyces cerevisiae*) strain was CEN.PK with genotype Mato; ura3-52; trp1-289; leu2-3,112; his3Δ1; MAL2-8C; SUC2. The parent strain was engineered to include three enzymatic steps through chromosomal homologous recombination of integration cassettes as described previously

134

(Dastmalchi et al., 2019, Nat. Chem Biol. 15: 384-390; Chen et al., 2018, Nat. Chem Biol. 14: 738-743). The three steps were comprised of PsiH (SEQ.ID NO: 4), PsiK (SEQ.ID NO: 6) and PsiM (SEQ.ID NO: 8), encoded by SEQ.ID NO: 3, SEQ.ID NO: 5 and SEQ.ID NO: 7, respectively, with addition of in-frame, C-terminal HA (SEQ.ID NO: 81, SEQ.ID NO: 82), V5 (SEQ.ID NO: 87, SEQ.ID NO: 88), and FLAG (SEQ.ID NO: 85, SEQ.ID NO: 86) epitope tags, respectively. DNA (SEQ.ID NO: 95) encoding a fifth enzyme, PcCPR (SEQ.ID NO: 96) was also integrated following PCR-based, C-terminal, in-frame addition of epitope tag c-MYC (SEQ.ID NO: 83, SEQ.ID NO: 84) to support functionality of PsiH. Integration cassettes were built using yeast promoter sequences amplified from *S. cerevisiae* genomic DNA as described (Dastmalchi et al. 2019; Chen et al. 2018) enabling constitutive gene expression. Amplified promoters included PGK1 (SEQ.ID NO: 77), TDH3 (SEQ.ID NO: 78), CLN1 (SEQ.ID NO: 79), and UGA1 (SEQ.ID NO: 80). Two integration cassettes were assembled: the first (SEQ.ID NO: 74) harbored tagged PsiH and PcCPR, and the second (SEQ.ID NO: 75) harbored tagged PsiK and PsiM. Successive genomic integration of these cassettes was performed as described previously (Chen et al. 2018). Following stable integration of these two cassettes, the strain was further manipulated by transformation with a yeast episomal vector encoding a promiscuous decarboxylase, ClostSporTDC (pMM1-pTDH3-ClostSporTDC-His-tCYC1) (SEQ.ID NO: 76) enabling efficient decarboxylation of halogenated tryptophan intermediates and/or feedstocks. For this Example, heterologous expression of a non-native or engineered TrpB gene was not necessary, as endogenous tryptophan synthase activity proved sufficient. The final engineered strain was called Sc-1. For scaled-up production of halogenated product, culturing was performed as follows. Seed cultures were inoculated in SD-drop-out medium overnight. The overnight culture was then divided into two flasks containing 500 ml each of SD-drop-out medium containing 2% (w/v) glucose, 0.3% (w/v) KH₂PO₄, 0.05% (w/v) MgSO₄·7H₂O, 0.5% (w/v) (NH₄)₂SO₄ plus 500 μM 6-fluorotryptophan (ThermoFisher Scientific) for conversion by Sc-1. Yeast cultures were grown for 48 h. Cultures were then centrifuged (10,000 g×5 minutes) to remove cellular content, and culture broth containing secreted halogenated product was stored at -80° C. until further processing. To determine successful biosynthesis of 6-fluoro-psilocybin (compound with formula (XI)), culture broth was analyzed using high-resolution LC-HESI-LTQ-Orbitrap-XL MS (Thermo Fisher Scientific), employing a modified version of a method described previously (Chang et al., 2015, Plant Physiol. 169: 1127-1140), with the exception that liquid chromatography was carried out using an UltiMate 3000 HPLC (Thermo Fisher Scientific) equipped with a Poroshell 120 SB-C18 column (Agilent Technologies) instead of an Accela HPLC system (Thermo Fisher Scientific) equipped with a Zorbax C18 column (Agilent Technologies). Briefly, 10 microliters of sample was injected at a flow rate of 0.5 mL/min and a gradient of solvent A (water with 0.1% of formic Acid) and solvent B (ACN with 0.1% formic Acid) as follows: 100% to 0% (v/v) solvent A over 5 min; isocratic at 0% (v/v) for 1 min; 0% to 100% (v/v) over 0.1 min; and isocratic at 100% (v/v) for 1.9 min. Total run time was 8 minutes. Heated ESI source and interface conditions were operated in positive ion mode as follows: vaporizer temperature, 400° C.; source voltage, 3 kV; sheath gas, 60 au, auxiliary gas, 20 au; capillary temperature, 380° C.; capillary voltage, 6 V; tube lens, 45 V. Instrumentation was performed as a single, HR scan event

135

using Orbitrap detection of m/z in the range of 100-500 m/z. Ion injection time was 300 ms with scan time of 1 s. External and internal calibration procedures ensured <2 ppm error to facilitate elemental formulae predictions. Singly protonated product with exact m/z and expected elemental formula matching 6-fluoro-psilocybin having chemical formula (XI):

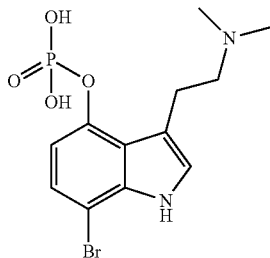


(XI)

eluted at 1.9 minutes (EIC, see: FIG. 10).

Example 2—Process for Making a Second Halogenated Psilocybin Derivative from Halogenated Tryptophan Feedstock

The same yeast strain (Sc-1) and procedures described in Example 1 were used to biosynthesize a halogenated psilocybin derivative with chemical formula (XIV), with the following exception: in place of 6-fluorotryptophan, 500 μ M 7-bromotryptophan (Combi-Blocks, www.combi-blocks.com) was supplied as feedstock for bioconversion. Singly protonated product with exact m/z and expected elemental formula matching 7-bromo-psilocybin having chemical formula (XIV):



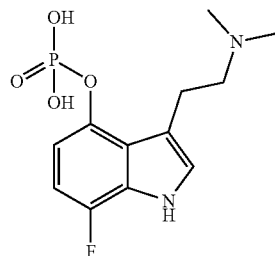
(XIV)

eluted at 2.8 minutes (EIC, see: FIG. 11).

Example 3—Process for Making a Third Halogenated Psilocybin Derivative from Halogenated Indole Feedstock

The same yeast strain (Sc-1) and procedures described in Example 1 were used to biosynthesize a halogenated psilocybin derivative with chemical formula (XVI), with the following exception: in place of 6-fluorotryptophan, 500 μ M 7-fluoroindole (Combi-Blocks, www.combi-blocks.com) was supplied as feedstock for bioconversion. Singly protonated product with exact m/z and expected elemental formula matching 7-fluoro-psilocybin compound having chemical formula (XVI):

136

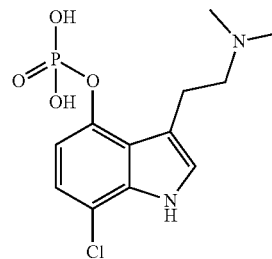


(XVI)

eluted at 2.3 minutes (EIC, see: FIG. 12).

Example 4—Process for Making a Fourth Halogenated Psilocybin Derivative from Halogenated Indole Feedstock

The same yeast strain (Sc-1) and procedures described in Example 1 were used to biosynthesize a halogenated psilocybin derivative with formula (XVII), with the following exception: in place of 6-fluorotryptophan, 500 μ M 7-chloroindole (Combi-Blocks, www.combi-blocks.com) was supplied as feedstock for bioconversion. Singly protonated product with exact m/z and expected elemental formula matching 7-chloro-psilocybin having formula (XVII):



(XVII)

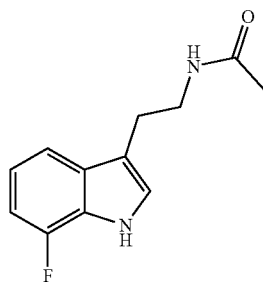
eluted at 2.7 minutes (EIC, see: FIG. 13).

Example 5—Process for Making a Fifth Halogenated Psilocybin Derivative from Halogenated Indole Feedstock

E. coli strain Ec-1 was constructed as follows. For plasmid cloning, Top10 or XL1-blue strains were used depending on antibiotic markers. Standard LB media was used for culturing. For gene expression and feeding experiments, the parent host strain employed was BL21 (DE3). For this Example, heterologous expression of a non-native or engineered TrpB gene was not necessary, as endogenous tryptophan synthase activity proved sufficient. An in frame, C-terminally His-tagged (SEQ.ID NO: 89, SEQ.ID NO: 90) BaTDC (SEQ.ID NO: 36, SEQ.ID NO: 37) was generated by PCR, followed by cloning into the NdeI and XhoI sites of the plasmid pCDM4 (SEQ.ID NO: 92) where the expression of the gene was driven by a T7 promoter. In a second plasmid, an in-frame, C-terminally His-tagged (SEQ.ID NO: 89, SEQ.ID NO: 90) PsmF (SEQ.ID NO: 62, SEQ.ID NO: 63) was cloned into the NdeI and XhoI sites of pET28a(+) (SEQ.ID NO: 94) where the expression of the gene was also driven by a T7 promoter. The *E. coli* host strain BL21 (DE3) was transformed with both plasmids and selected for using streptomycin and kanamycin antibiotics. This strain (Ec-1,

137

F-ompT hsdSB (rB-, mB-) gal dcm (DE3), pCDM4-BaTDC-His, pET28a(+)-PsmF-His) enabled production of N-acetylated products from modified indole feedstock. Further, the use of promiscuous BaTDC (SEQ.ID NO: 36, SEQ.ID NO: 37) allowed decarboxylation of a wide variety of halogenated tryptophan intermediates. Scaled-up culturing of engineered *E. coli* was conducted as follows: seed cultures were inoculated in AMM (Jones et al., 2015, Sci. Rep. 5: 11301) medium overnight. The overnight culture was then divided into two flasks containing 500 mL each of AMM medium additionally containing 0.5% (w/v) serine, 1M IPTG, 50 ug/L kanamycin and streptomycin, and 100 mg/L 7-fluoroindole (Combi-Blocks, www.combi-blocks.com) for conversion by the *E. coli* strain. Cultures were grown for 24 h. Cultures were then centrifuged (10,000 g x 5 minutes) to remove cellular content, and culture broth containing secreted derivative was stored at -80 C until further processing. The N-(2-(7-difluoro-1H-indol-3-yl)ethyl)acetamide product having chemical formula (XV):



(XV)

contained in 1 L of *E. coli* culture was extracted by ethyl acetate (3x300 ml). The organic layer was dried over Na₂SO₄, followed by concentration under reduced pressure. The residue was purified by column chromatography on silica gel using ethyl acetate-hexane (50→80%) as eluent to give the compound as a light yellow solid (22 mg). Following purification, high-resolution MS (HRMS), ¹H NMR, and selective ¹³C NMR were performed to assess purity, estimate total quantity, and confirm molecular structure. ¹H NMR (400 MHz, CDCl₃): δ=1.96 (s, 3H), 2.97 (t, J=6.8 Hz, 2H), 3.59 (dt, J=6.7, 6.1 Hz, 2H), 5.7 (br, s, 1H), 6.91 (dd, J=11.2, 7.8 Hz, 1H), 7.03 (m, 2H), 7.35 (d, J=7.9 Hz, 1H), 8.6 (br, s, 1H). ¹³C NMR (100 MHz, CDCl₃): δ=23.3, 25.3, 39.8, 106.9 (d, J_{C,F}=16.1 Hz), 113.7, 114.4 (d, J_{C,F}=3.4 Hz), 119.6 (d, J_{C,F}=6.1 Hz), 122.8, 124.7, 131.1 (d, J_{C,F}=5.3 Hz), 149.7 (d, J_{C,F}=244.2 Hz), 170.3. HRMS (ESI) m/z: calcd. for C₁₂H₁₃FN₂O [M+H]⁺ 221.1090, found 221.1087. The compound having chemical formula (XV) was 95% (w/w) pure. It is noted that the acetyl group of the compound having chemical formula (XV) is included therein by virtue of the acetyl transferase having SEQ.ID NO: 63 in *Escherichia coli* strain Ec-1.

Cell lines for pharmacology assays. CHO-K1/Galpa15 (GenScript, M00257) (-5-HT_{2A}) and CHO-K1/5-HT_{2A} (GenScript, M00250) (+5-HT_{2A}) cells lines were used in both toxicology/growth inhibition (MTT) and calcium release assays. Briefly, CHO-K1/Galpa15 is a control cell line that constitutively expresses Galpa15 which is a promiscuous Gq protein. It is engineered as a host cell, allowing transfected receptor(s) to signal through the Gq signal transduction pathway and mobilize intracellular calcium from the endoplasmic reticulum (ER). These control cells lack any transgene encoding 5-HT_{2A} receptors, thus prevent-

138

ing calcium mobilization in response to 5-HT_{2A} activation. Conversely, CHO-K1/5-HT_{2A} cells stably express 5-HT_{2A} receptor in the CHO-K1 host background. This design enables Gq-11 expressed in CHO-K1 cells to mobilize intracellular calcium changes when 5-HT_{2A} receptors are activated by ligands.

Cell lines were maintained in Ham's F12 media plus 10% FBS in the presence of 100 ug/ml hygromycin for CHO-K1/Ga15 or 400 ug/ml G418 for CHO-K1/5-HT_{2A} unless indicated otherwise for specific assays. Cell maintenance was carried out as recommended by the cell supplier. Briefly, vials with cells were removed from the liquid nitrogen and thawed quickly in 37° C. water bath. Just before cells were completely thawed, vial exteriors were decontaminated with 70% ethanol spray. Cell suspension was then retrieved from the vial and added to warm (37° C.), 'complete' (non-dropout) growth media, and centrifuged at 1,000 rpm for 5 minutes. The supernatant was discarded and the cell pellet was then resuspended in another 10 ml of complete growth media, and added to a 10 cm cell culture dish (Greiner Bio-One #664160). The media was changed every third day until the cells reached ~90% confluence. The ~90% confluent cells were then split 10:1, and used either for maintenance or pharmacological study.

Assessment of cell viability upon treatment of glycosylated psilocybin derivatives. To establish suitable ligand concentrations for the calcium release assays, MTT (3-(4, 5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium) assays were first performed. Results of these assays were conducted using both control ligands (e.g. psilocybin, psilocin, DMT) and novel derivatives, in part as a pre-screen for any remarkable toxic effects on cell cultures up to concentrations of 1 mM. A known cellular toxin (Triton X-100, Pyrgiotakis G. et al., 2009, Ann. Biomed. Eng. 37: 1464-1473) was included as a general marker of toxicity. Modified Chinese Hamster Ovary cells (CHO-K1/Ga15) were cultured using standard procedures using the manufacturer's protocols (Genscript, M00257). Briefly, cells were cultured in Ham's F12 medium supplemented with 10% fetal bovine serum and 100 mg/ml Hygromycin B, and grown at 37° C. in the presence of 5% CO₂. To test the various compounds with the cell line, cells were seeded in a clear 96-well culture plate at 10,000 cells per well. After allowing cells to attach and grow for 24 hours, assay compounds were added at 1 μM, 10 μM, 100 μM, and 1 mM final concentrations. Methanol concentrations used are 0.001, 0.01, 0.1, and 1%. Triton concentrations used are 0.0001, 0.001, 0.01 and 0.1%. Cells were incubated with compounds for 48 hours before accessing cell viability with the MTT assay following the manufacturer's protocol (MTT Cell Growth Assay Kit; Millipore Sigma, CTO2). MTT reagent was added to cells and allowed to incubate for 4 hours before solubilization with isopropanol plus 0.04 N HCl. Absorbance readings were performed at 570 nm with the reference at 630 nm on a SpectraMax iD3 plate reader. Non-treated cells were assigned 100% viability. Results of the cell viability assays are shown in FIG. 14. Bar graphs show the mean±SD (n=3). Significance (P<0.0001), as indicated by (***) was determined using 2-way ANOVA with Dunnett's multiple comparisons test. The results using compound with formula (XV) are indicated as "(XV)" on the x-axis.

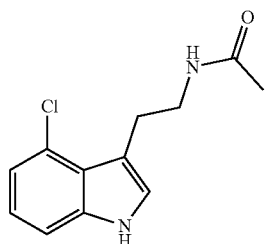
Increase in cytosolic calcium concentration by 5-HT_{2A} activation. Changes in intracellular calcium concentration due to the treatment with assay compounds was measured using Fluo-8 dye (Abcam, #ab112129) according to the manufacturer's instructions. Briefly, CHO-K1 cells stably expressing 5-HT_{2A} (Genscript #M00250) (+5-HT_{2A}) or lack-

139

ing 5-HT_{2A} (Genscript, M00257) (–5-HT_{2A}) were seeded on black walled clear bottom 96-well plates (Thermo Scientific #NUNC165305), allowing 70,000 cells/well in 100 ul media (HAM's F12, GIBCO #11765-047) with 1% FBS (Thermo Scientific #12483020). Cultures were maintained in a humidified incubator at 37° C. and 5% CO₂. Fluo-8 dye was loaded into the cultures for 30 min at 37° C., followed by 30 min additional incubation at room temperature. Next, different dilutions of novel molecules and controls were prepared in serum-free culture media and added to the cells. Fluorescence (ex 490 nm/em 525 nm) obtained after the addition of molecules was expressed relative to values obtained before addition of the molecules (relative Fluo-8 fluorescence=Fmax/F0, where Fmax=maximum fluorescence and F0=baseline fluorescence). Fluorescence intensities were measured using a Spectramax ID3 plate reader (www.moleculardevices.com). Relative fluorescence (RFU) at increasing concentrations of compound was determined, illustrating concentration-dependent calcium flux. Psilocin, a known agonist with binding activity at 5-HT_{2A} (Rickli A. et al., 2016, Europ. Neuropsychopharmacol., 26: 1326-1337) was used as a positive control to establish assay functionality. The Example compound (XV) was then evaluated. Results are shown in FIGS. 15A-15B and 16A-16B. The results using different concentrations of psilocin and compound with formula (XV) (indicated as “(XV)”) in +5-HT_{2A} cell cultures are shown in FIG. 15A and FIG. 15B, respectively. The results using different concentrations of psilocin and compound with formula (XV) (indicated as “(XV)”) in –5-HT_{2A} cell cultures are shown in FIG. 16A and FIG. 16B, respectively. Results for negative control treatment with vehicle containing neither psilocin nor derivative are shown in FIGS. 15A-15B and 16A-16B. Error bars show +/–SD about the mean (n=3).

Example 6—Process for Making a Sixth
Halogenated Psilocybin Derivative from
Halogenated Indole Feedstock

Escherichia coli strain Ec-1 was used to biosynthesize halogenated tryptamine derivative having chemical formula (II) from halogenated indole feedstock. The construction of Ec-1 is described in Example 5. Scaled-up culturing and material storage of engineered *E. coli* was conducted as described in Example 5, except that 4-chloroindole (Combi-Blocks, www.combi-blocks.com) was used in place of 7-fluoroindole. To assess product, high-resolution LC-HESI-LTQ-Orbitrap-XL MS analysis was conducted as described in Example 1. Singly protonated product with exact m/z and expected elemental formula matching 4-chloro-N-acetyl-tryptamine having chemical formula (II):



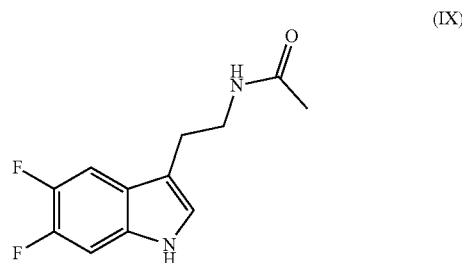
Eluted at 4.1 minutes (EIC, see: FIG. 17). It is noted that the acetyl group of the compound having chemical formula (II)

140

is included therein by virtue of the acetyl transferase having SEQ.ID NO: 63 in *Escherichia coli* strain Ec-1.

Example 7—Process for Making a Seventh
Halogenated Psilocybin Derivative from
Halogenated Indole Feedstock

Escherichia coli strain Ec-1 was used to biosynthesize halogenated tryptamine derivative having chemical formula (IX) from halogenated indole feedstock. The construction of Ec-1 is described in Example 5. Scaled-up culturing and material storage of engineered *E. coli* was conducted as described in Example 5, except that 5,6-difluoroindole (Combi-Blocks, www.combi-blocks.com) was used in place of 7-fluoroindole. N-(2-(5,6-difluoro-1H-indol-3-yl)ethyl)acetamide product having Chemical formula (IX):



contained in 0.55 L of *E. coli* culture was extracted by ethyl acetate (3×300 ml). The organic layer was dried over Na₂SO₄, followed by concentration under reduced pressure. The residue was purified by column chromatography on silica gel using ethyl acetate-hexane (50→80%) as eluent to give the compound as a light yellow solid (11 mg). Following purification, high-resolution MS (HRMS), ¹H NMR, and selective ¹³C NMR were performed to assess purity, estimate total quantity, and confirm molecular structure. ¹H NMR (400 MHz, CDCl₃): δ=1.70-1.79 (m, 1H), 1.99-2.06 (m, 1H), 2.29 (s, 6H), 2.33-2.38 (m, 1H), 2.40-2.48 (m, 1H), 3.22 (t, 1H, J=8.0 Hz), 3.28-3.36 (m, 1H), 3.68 (t, 1H, J=8.4 Hz), 3.94 (br, 1H), 6.65 (d, 1H, J=8.4 Hz), 6.74 (t, 1H, J=7.6 Hz), 7.05 (t, 1H, J=7.6 Hz), 7.11 (d, 1H, J=7.6 Hz). ¹³C NMR (100 MHz, CDCl₃): δ=23.4, 25.2, 39.7, 99.1 (d, J_{C, F}=21.7 Hz), 105.4 ((d, J_{C, F}=20.0 Hz), 123.3 (d, J_{C, F}=3.7 Hz), 150.6 (dd, J_{C, F}=237.4, 17.2 Hz), 148.9 (dd, J_{C, F}=238.3, 18.1 Hz), 170.1. HRMS (ESI) calcd. for C₁₂H₁₂F₂N₂O 239.0996 [M+H]⁺, found 239.0989. The compound having chemical formula (IX) was 95% (w/w) pure. It is noted that the acetyl group of the compound having chemical formula (IX) is included therein by virtue of the acetyl transferase having SEQ.ID NO: 63 in *Escherichia coli* strain Ec-1.

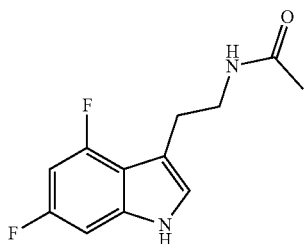
Efficacy testing was carried out as described in Example 5. MTT assay results are shown in FIG. 18, and calcium mobility assay results are shown in FIG. 19A (+5-HT_{2A} cells) and FIG. 19B (–5-HT_{2A} cells). In both cases, the example compound having the chemical formula (IX) is shown as “(IX)”.

Example 8—Process for Making an Eighth
Halogenated Psilocybin Derivative from
Halogenated Indole Feedstock

Escherichia coli strain Ec-1 was used to biosynthesize halogenated tryptamine derivative having chemical formula

141

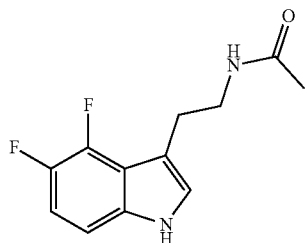
(VI) from halogenated indole feedstock. The construction of Ec-1 is described in Example 5. Scaled-up culturing and material storage of engineered *E. coli* was conducted as described in Example 5, except that 4,6-difluoroindole (Combi-Blocks, www.combi-blocks.com) was used in place of 7-fluoroindole. To assess product, high-resolution LC-HESI-LTQ-Orbitrap-XL MS analysis was conducted as described in Example 1. Singly protonated product with exact m/z and expected elemental formula matching 4,6-difluoro-N-acetyltryptamine having chemical formula (VI):



eluted at 5.1 minutes (EIC, see: FIG. 20).

Example 9—Process for Making a Ninth Halogenated Psilocybin Derivative from Halogenated Indole Feedstock

Escherichia coli strain Ec-1 was used to biosynthesize halogenated tryptamine derivative having chemical formula (V) from halogenated indole feedstock. The construction of Ec-1 is described in Example 5. Scaled-up culturing and material storage of engineered *E. coli* was conducted as described in Example 5, except that 4,5-difluoroindole (Combi-Blocks, www.combi-blocks.com) was used in place of 7-fluoroindole. N-(2-(4,5-difluoro-1H-indol-3-yl)ethyl)acetamide product having chemical formula (V):



contained in 1 L of *E. coli* culture was extracted by ethyl acetate (3×600 ml). The organic layer was dried over Na₂SO₄, followed by concentration under reduced pressure. The residue was purified by flash chromatography on silica gel using ethyl acetate-hexane (50→80%) as eluent, followed by crystallization from EtOAc/Hexane to give the compound as a white solid (20 mg). Following purification, high-resolution MS (HRMS), ¹H NMR, and selective ¹³C NMR were performed to assess purity, estimate total quantity, and confirm molecular structure. ¹H NMR (400 MHz, CDCl₃): δ=1.96 (s, 3H), 3.04 (t, J=6.9 Hz, 2H), 3.60 (dt,

142

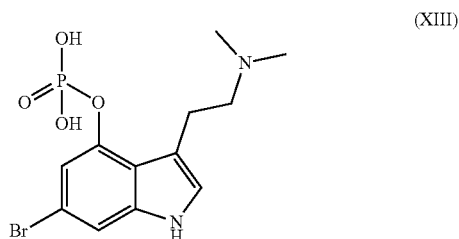
J=6.6, 6.4 Hz, 2H), 5.64 (br, s, 1H), 7.01 (m, 3H), 8.41 (br, s, 1H). ¹³C NMR (100 MHz, CDCl₃): δ=23.4, 26.3, 40.5 (d, J_{C, F}=2.0 Hz), 106.6 (dd, J_{C, F}=7.7, 4.2 Hz), 111.7 (d, J_{C, F}=21.7 Hz), 124.0, 143.9 (dd, J_{C, F}=246.4, 14.5 Hz), 143.8 (dd, J_{C, F}=246.4, 14.5 Hz), 170.2. HRMS (ESI) m/z: calcd. for C₁₂H₁₂F₂N₂O [M+H]⁺ 239.0996, found 239.0989. The compound having chemical formula (V) was 95% (w/w) pure. It is noted that the acetyl group of the compound having chemical formula (II) is included therein by virtue of the acetyl transferase having SEQ.ID NO: 63 in *Escherichia coli* strain Ec-1.

Efficacy testing was carried out as described in Example 5. MTT assay results are shown in FIG. 21, and calcium mobility assay results are shown in FIGS. 22A and 22B. In both cases, the example compound having the chemical formula (V) is shown as “(V)”.

Example 10—Process for Making a Tenth Halogenated Psilocybin Derivative from Halogenated Indole Feedstock

E. coli strain Ec-2 was constructed as follows. For plasmid cloning, Top10 or XL1-blue strains were used depending on antibiotic markers. Standard LB media was used for culturing. For gene expression and feeding experiments, the parent host strain employed was BL21 (DE3). For this Example, heterologous expression of a non-native or engineered TrpB gene was not necessary, as endogenous tryptophan synthase activity proved sufficient. From plasmid pCDM4 (SEQ.ID NO: 92), the plasmid pCDM4-PsiD-FLAG was created by inserting an in-frame, C-terminally FLAG-tagged (SEQ.ID NO: 85) PsiD gene (SEQ.ID NO: 1) into the NdeI/XhoI site of pCDM4. The plasmid pETM6-H10-PsiK-V5-PsiM-FLAG was created by first cloning the in-frame, C-terminally V5-tagged (SEQ.ID NO: 87) PsiK (SEQ.ID NO: 5) into the NdeI/XhoI site of pETM6-H10 (SEQ.ID NO: 97) to create pETM6-H10-PsiK-V5. This intermediate plasmid was digested with SpeI and SalI, and in-frame, C-terminally FLAG tagged (SEQ.ID NO: 85) PsiM (SEQ.ID NO: 7) was cloned into the site with XbaI and SalI, nullifying the SpeI restriction site. In this setup, the T7 polymerase was able to drive the expression of the polycistronic DNA containing both PsiK and PsiM. The two target plasmids pCDM4-PsiD-FLAG and pETM6-H10-PsiK-V5-PsiM-FLAG were transformed into BL21 (DE3) cells, and antibiotics ampicillin plus streptomycin were used to select for the correct clones containing both plasmids. Scaled-up culturing of engineered *E. coli* was conducted as follows: seed cultures were inoculated in AMM (Jones et al. 2015, Sci Rep. 5: 11301) medium overnight. The overnight culture was then divided into two flasks containing 500 mL each of AMM medium additionally containing 0.5% (w/v) serine, 1M IPTG, 50 ug/L ampicillin and streptomycin, and 100 mg/L 4-hydroxy-6-bromoindole (BLDPharm, www.bldpharm.com) for conversion by Ec-2. Cultures were grown for 24 h. Cultures were then centrifuged (10,000 g×5 minutes) to remove cellular content, and culture broth containing secreted derivative was stored at -80 C until further processing. To assess product, high-resolution LC-HESI-LTQ-Orbitrap-XL MS analysis was conducted as described in Example 1. Singly protonated product with exact m/z and expected elemental formula matching 6-bromo-psilocybin having chemical formula (XIII):

143

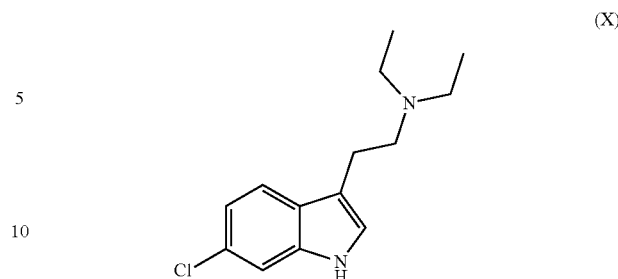


eluted at 2.8 minutes (IC, see: FIG. 23).

Example 11—Process for Making an Eleventh
Halogenated Psilocybin Derivative from
Halogenated Indole Feedstock

In a flame-dried round bottom flask under argon, 6-chloroindole (compound 1, see: FIG. 9A) (210 mg, 1.39 mmol, 1.00 eq) was dissolved in diethyl ether (4.0 mL) and cooled to 0° C. To this solution, oxalyl chloride (369 mg, 2.91 mmol, 2.10 eq) in diethyl ether (2.0 mL) was added dropwise over a period of 10 minutes. The reaction mixture was allowed to stir at room temperature for 4 hours. Additional ether (5.0 mL) was added to the reaction mixture which was then cooled to 0° C. and subsequently vacuum-filtered to yield a bright yellow solid (compound 2) (211 mg, 0.872 mmol, 63%). Compound 2 (see: FIG. 9A) was washed with cold ether (10 mL) and allowed to briefly dry before being resuspended in diethyl ether (3.0 mL) and cooled to 0° C. Diethylamine (510 mg, 6.97 mmol, 8.00 eq) was diluted with diethyl ether (3.0 mL) and added dropwise to the reaction flask, immediately precipitating out a white solid. This mixture was allowed to stir for 10 minutes at 0° C., warmed to room temperature, and stirred for an additional 1.5 hours. The visible precipitate was vacuum filtered to yield a solid which was washed with cold diethyl ether (10 mL). This solid was suspended in deionized water (50 mL), stirred for 10 minutes, vacuum-filtered, and allowed to dry overnight, yielding compound 3 (see: FIG. 9A) as a white powder (124 mg, 0.445 mmol, 51%). No further purification was required. ¹H NMR (400 MHz, CDCl₃): δ (ppm)=9.56 (s, 1H), 8.24 (d, J=8.5 Hz, 1H), 7.74 (d, J=3.0 Hz, 1H), 7.36 (dd, J=1.8, 0.6 Hz, 1H), 7.28 (d, J=1.8 Hz, 1H), 3.54 (q, J=7.1 Hz, 2H), 3.36 (q, J=7.1 Hz, 2H), 1.27 (t, J=7.2 Hz, 3H), 1.19 (t, J=7.1 Hz, 3H). A THF solution of lithium aluminum hydride (2.20 mL, 2.22 mmol, 5.00 eq, 1.0 M) was added to a 3-neck, flame-dried, round-bottom flask under argon. Anhydrous dioxane (3.0 mL) was added to the reaction mixture. In a separate vial, compound 3 (124 mg, 0.445 mmol) was dissolved in a THF (2.0 mL)/dioxane (1.0 mL) solution and added dropwise to the reaction vessel. This mixture was then allowed to reflux at 95° C. for 1.5 hours. The reaction mixture was cooled to 0° C. and quenched through dropwise addition of a water (1.0 mL)/THF (2.0 mL) solution, yielding a flocculent precipitate. This precipitate was removed through vacuum-filtration and the filter-cake was washed with diethyl ether (20 mL). The filtrate was collected, dried with MgSO₄, and solvent removed to yield a pale-yellow oil. Purification through column chromatography on silica gel (2% MeOH/DCM+1% NEt₃) yielded compound 410B01 (see: FIG. 9A), corresponding to compound with formula (X):

144



as a clear, colourless oil (34 mg, 0.136 mmol, 30%). ¹H NMR (400 MHz, CDCl₃): δ (ppm)=8.37 (s, 1H), 7.49 (d, J=8.5 Hz, 1H), 7.30 (d, J=1.8 Hz, 1H), 7.07 (dd, J=8.5, 1.8 Hz, 1H), 6.98 (dd, J=2.5, 1.1 Hz, 1H), 2.90 (high order m, 2H), 2.79 (high order m, 2H), 2.74-2.62 (q, J=7.1 Hz, 4H), 1.15-1.04 (t, J=7.1 Hz, 6H). ²⁵ ¹³C NMR (100 MHz, CDCl₃): δ (ppm)=136.6, 127.8, 126.2, 122.2, 119.9, 119.7, 114.8, 111.0, 70.5, 53.5, 46.9, 22.8, 11.7. The compound having chemical formula (IX) was 95% (w/w) pure.

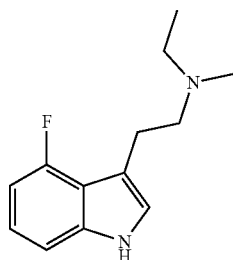
Efficacy testing was carried out as described in Example 5. MTT assay results are shown in FIG. 24, and calcium mobility assay results are shown in FIGS. 25A and 25B. In both cases, the example compound having chemical formula (X) is shown as “(X)”.

Example 12—Process for Making a Twelfth
Halogenated Psilocybin Derivative from
Halogenated Indole Feedstock

A similar process as illustrated in FIG. 9A (Example 11) was followed. In a flame-dried round bottom flask under argon, 4-fluoroindole (compound 1) (200 mg, 1.48 mmol, 1.00 eq) was dissolved in diethyl ether (4.0 mL) and cooled to 0° C. To this solution, oxalyl chloride (394 mg, 3.11 mmol, 2.10 eq) in diethyl ether (2.0 mL) was added dropwise over a period of 10 minutes. The reaction mixture was allowed to stir at room temperature for 4 hours. Additional ether (5.0 mL) was added to the reaction mixture which was then cooled to 0° C. and subsequently vacuum-filtered to yield a bright yellow solid (compound 2) (264 mg, 1.17 mmol, 79%). Compound 2 was washed with cold ether (10 mL) and allowed to briefly dry before being resuspended in diethyl ether (3.0 mL) and cooled to 0° C. N-Methylethylamine (574 mg, 9.72 mmol, 8.00 eq) was diluted with diethyl ether (3.0 mL) and added dropwise to the reaction flask. This mixture was allowed to stir for 10 minutes at 0° C., warmed to room temperature, and stirred for an additional 1.5 hours, yielding a suspension of light-pink gooeey solid. The diethyl ether was decanted, and the solid dissolved in dichloromethane (10 mL) and washed with water (3×10 mL). The organic phase was dried with MgSO₄ and solvent removed to yield a pale-pink glassy solid (128 mg, 0.516 mmol, 43%). ¹H NMR (400 MHz, CDCl₃): δ (ppm)=10.72 (s, 1H), 7.72 (d, J=3.8 Hz, 1H), 7.10-7.00 (m, 2H), 6.84 (ddd, J=11.1, 7.2, 1.6 Hz, 1H), 3.55 (q, J=7.2 Hz, 1H), 3.35 (q, J=7.1 Hz, 1H), 3.05 (s, 2H), 2.98 (s, 1H), 1.23 (t, J=7.2 Hz, 1H), 1.15 (t, J=7.1 Hz, 2H). A THF solution of lithium aluminum hydride (2.58 mL, 2.58 mmol, 5.00 eq, 1.0 M) was added to a 3-neck, flame-dried, round-bottom flask under argon. Anhydrous dioxane (3.0 mL) was added to the reaction mixture. In a separate vial, compound 3 (128 mg, 0.516 mmol) was dissolved in an anhydrous THF (2.0 mL)/dioxane (1.0 mL) solution and added dropwise to the reaction vessel. This mixture was then allowed to reflux at

145

95° C. for 2.5 hours. The reaction mixture was cooled to 0° C. and quenched through dropwise addition of a water (1 mL)/THF (2 mL) solution, yielding a flocculent precipitate. This precipitate was removed through vacuum-filtration and the filter-cake was washed with diethyl ether (20 mL). The filtrate was collected, dried with MgSO₄, and solvent removed to yield a pale-orange oil. Purification through column chromatography on silica gel (1% MeOH/CH₂Cl₂+ 1% NEt₃) yielded compound 456B01, corresponding to compound having chemical formula (IV):

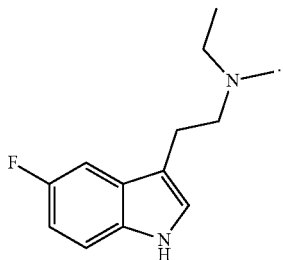


as clear, colourless oily crystal (86 mg, 0.390 mmol, 76%). ¹H NMR (400 MHz, CDCl₃): δ (ppm)=8.68 (s, 1H), 7.12-6.99 (m, 2H), 6.91 (d, J=2.3 Hz, 1H), 6.73 (ddd, J=11.2, 6.9, 1.6 Hz, 1H), 3.04 (high order m, 2H), 2.75 (high order m, 2H), 2.57 (q, J=7.2 Hz, 2H), 2.38 (s, 3H), 1.12 (t, J=7.2 Hz, 3H). ¹³C NMR (100 MHz, CDCl₃): δ (ppm)=157.2 (d, J_{C-F}=245.6 Hz), 139.2 (d, J_{C-F}=11.8 Hz), 122.2 (d, J_{C-F}=8.2 Hz), 121.9, 107.2 (d, J_{C-F}=3.6 Hz), 104.2 (d, J_{C-F}=19.7 Hz), 58.5, 51.2, 41.5, 24.1, 12.1. The compound having chemical formula (IV) was 95% (w/w) pure.

Efficacy testing was carried out as described in Example 5. MTT assay results are shown in FIG. 26, and calcium mobility assay results are shown in FIG. 27A and FIG. 27B. In both cases, the example compound having the chemical formula (IV) is shown as “(IV)”.

Example 13—Process for Making a Thirteenth
Halogenated Psilocybin Derivative from
Halogenated Indole Feedstock

A thirteenth halogenated psilocybin derivative compound was synthesized according to the method of Example 11, except using 5-fluoroindole and N-methylethylamine as starting materials. Analysis indicated the isolation of a compound having chemical formula (VII):

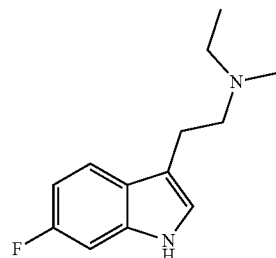


¹H NMR (400 MHz, CDCl₃): δ (ppm)=8.09 (s, 1H), 7.29-7.20 (m, 2H), 7.06 (d, J=2.4 Hz, 1H), 6.96-6.87 (m, 1H), 2.93-2.86 (m, 2H), 2.72-2.64 (m, 2H), 2.53 (q, J=7.2 Hz, 2H), 2.35 (s, 3H), 1.11 (t, J=7.2 Hz, 3H).

146

Example 14—Process for Making a Fourteenth
Halogenated Psilocybin Derivative from
Halogenated Indole Feedstock

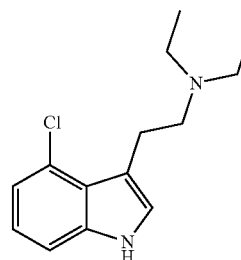
A fourteenth halogenated psilocybin derivative was synthesized according to the method of Example 11, except using 6-fluoroindole and N-methylethylamine as starting materials. Analysis indicated the isolation of a compound having chemical formula (XII):



¹H NMR (400 MHz, CDCl₃): δ (ppm)=8.05 (s, 1H), 7.50 (ddt, J=8.7, 5.3, 0.6 Hz, 1H), 7.02 (ddd, J=9.7, 2.3, 0.5 Hz, 1H), 7.00-6.97 (m, 1H), 6.88 (ddd, J=9.6, 8.6, 2.3 Hz, 1H), 2.96-2.89 (m, 2H), 2.74-2.66 (m, 2H), 2.53 (q, J=7.2 Hz, 2H), 2.35 (s, 3H), 1.11 (t, J=7.2 Hz, 3H).

Example 15—Process for Making a Fifteenth
Halogenated Psilocybin Derivative from
Halogenated Indole Feedstock

A fifteenth halogenated psilocybin derivative was synthesized according to the method of Example 11, except using 4-chloroindole as starting material. Analysis indicated the isolation of a compound having chemical formula (III):



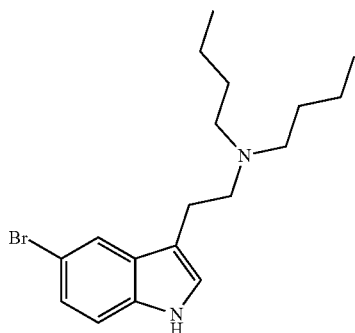
¹H NMR (400 MHz, CDCl₃): δ (ppm)=8.28 (s, 1H), 7.26-7.19 (m, 1H), 7.08-7.01 (m, 3H), 3.18-3.10 (m, 2H), 2.86-2.77 (m, 2H), 2.67 (q, J=7.1 Hz, 4H), 1.09 (t, J=7.1 Hz, 6H).

Example 16—Process for Making a Sixteenth
Halogenated Psilocybin Derivative from
Halogenated Indole Feedstock

The initial reaction to generate compound 1.1 (see: FIG. 9B) followed a previously reported procedure with modifications (Campos et al. 2004, Org. Lett 6: 79-82). To a solution of 4-bromophenylhydrazine hydrochloride (300 mg, 1.34 mmol) in a mixture of N,N-dimethylacetamide (1.9 mL) and 4% H₂SO₄ (1.9 mL), was added 2,3-dihydrofuran (0.10 mL, 1.33 mmol). The reaction mixture was heated at

147

100° C. for 16 hours. Once the mixture was cooled to room temperature, ethyl acetate (15 mL) was added, and the organic solution was washed with water (2×6 mL), and dried over anhydrous MgSO₄. The organic solvent was removed in vacuo. The product (compound 1.1, (see: FIG. 9B)) was purified by flash chromatography on silica gel (eluted with a gradient of hexanes-EtOAc, 100:0 to 0:100) to afford an orange oil. Yield: 37%. The characterization of this first product agreed with previously reported procedure (Go et al. 2010, J. Med. Chem. 53: 6838-6850). ¹H NMR (400 MHz, CDCl₃) δ 8.13 (bs, 1H), 7.74 (d, J=1.7 Hz, 1H), 7.28 (dd, J=8.6, 1.9 Hz, 1H), 7.23 (d, J=8.6 Hz, 1H), 7.08 (d, J=2.3 Hz, 1H), 3.89 (t, J=6.3 Hz, 2H), 2.98 (t, J=6.4 Hz, 2H). The second reaction (see: FIG. 9B) followed a previously reported procedure (Lombardo et al. 2013, Angew. Chem. Int. Ed. 52: 12910-12914). Imidazole (73 mg, 1.1 mmol), triphenyl phosphine (140 mg, 0.54 mmol), and iodine (140 mg, 0.54 mmol) were added to a solution of compound 1.1 (see: FIG. 9B) (119 mg, 0.49 mmol) in anhydrous dichloromethane (5.0 mL) cooled in an ice-water bath. The reaction was stirred at room temperature for 6 hours. The mixture was quenched with saturated sodium thiosulfate (10 mL), extracted in dichloromethane (3×10 mL), and the combined organic solution was dried over anhydrous MgSO₄, and concentrated under reduced pressure. The crude product was purified by flash chromatography on silica gel (eluted with a gradient of hexanes-EtOAc, 100:0 to 50:50) to afford an orange oil. Yield of compound 1.2 (see: FIG. 9B): 74%. ¹H NMR (400 MHz, CDCl₃) δ 8.07 (s, 1H), 7.71-7.69 (m, 1H), 7.29 (dd, J=8.6, 1.7 Hz, 1H), 7.25 (d, J=8.6 Hz, 1H), 7.10 (d, J=2.4 Hz, 1H), 3.44-3.39 (m, 2H), 3.30 (t, J=7.0 Hz, 2H). To a solution of compound 1.2 (129 mg, 0.37 mmol) in anhydrous DMF (1.0 mL) was added di-n-butylamine (0.25 mL, 1.48 mmol). The reaction mixture was heated at 60° C. for 16 hours. The reaction was quenched with saturated sodium bicarbonate solution, extracted in ethyl acetate (3×20 mL), and the combined organic solution was washed with water (2×40 mL), brine (2×40 mL), dried over anhydrous MgSO₄ and concentrated under reduced pressure. The crude mixture was purified by flash column chromatography on silica gel (eluted with a gradient of hexanes-EtOAc, 100:0 to 0:100) to afford an orange oil. Yield of compound 1, with structure as indicated by compound (VIII):



(VIII)

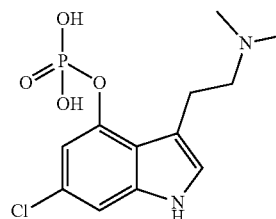
148

50%. ¹H NMR (400 MHz, CDCl₃) δ 8.16 (s, 1H), 7.72 (s, 1H), 7.28-7.22 (m, 1H), 7.22-7.18 (m, 1H), 7.03-6.99 (m, 1H), 2.89-2.82 (m, 2H), 2.81-2.73 (m, 2H), 2.54 (t, J=7.6 Hz, 4H), 1.52-1.43 (m, 5H), 1.33 (h, J=7.0 Hz, 4H), 0.93 (t, J=7.4 Hz, 6H) (Figure X, NMR). ¹³C NMR (100 MHz, CDCl₃) b 134.95, 129.58, 124.81, 122.83, 121.65, 114.83, 112.62, 112.57, 54.75, 54.01, 29.42, 22.87, 20.95, 14.27. HRMS (ESI, positive) m/z for C₁₈H₂₈BrN₂ [M+H]⁺ calcd. 351.1430, found 351.1427. The compound having chemical formula (VIII) was 95% (w/w) pure.

Efficacy testing was carried out as described in Example 5. MTT assay results are shown in FIG. 28, and calcium mobility assay results are shown in FIGS. 29A and 29B. In both cases, the example compound having the chemical formula (VIII) is shown as "(VIII)".

Example 17—Process for Making a Seventeenth Halogenated Psilocybin Derivative from Halogenated Indole Feedstock

E. coli strain Ec-3 was constructed as follows. For plasmid cloning, Top10 or XL1-blue strains were used depending on antibiotic markers. Standard LB media was used for culturing. For gene expression and feeding experiments, the parent host strain employed was BL21 (DE3). For this Example, heterologous expression of a non-native or engineered TrpB gene was not necessary, as endogenous tryptophan synthase activity proved sufficient. Plasmids pCDM4-PsiD-FLAG and pETM6-H10-PsiK-V5-PsiM-FLAG are described in Example 10. A third plasmid, pRSM3-ThaI-HIS, was created by cloning an in-frame, C-terminally HIS-tagged (SEQ.ID NO: 89) ThaI gene (SEQ.ID NO: 19), encoding a halogenase having SEQ.ID NO: 20, into the NdeI/XhoI site of pRSM3 (SEQ.ID NO: 98). The three target plasmids pCDM4-PsiD-FLAG, pETM6-H10-PsiK-V5-PsiM-FLAG, and pRSM3-ThaI-HIS were transformed into BL21 (DE3) cells and antibiotics ampicillin plus streptomycin plus kanamycin were used to select for the correct clones containing all three plasmids. Scaled-up culturing of engineered *E. coli* was conducted as described in Example 10. To assess product, high-resolution LC-HESI-LTQ-Orbitrap-XL MS analysis was conducted as described in Example 1. Singly protonated product with exact m/z and expected elemental formula matching 6-chloro-psilocybin having chemical formula (XVIII):



(XVIII)

eluted at 2.3 minutes (IC, see: FIG. 30).

SEQUENCE LISTING

Sequence total quantity: 98
 SEQ ID NO: 1 moltype = DNA length = 1320
 FEATURE Location/Qualifiers

-continued

```

source          1..1320
                 mol_type = genomic DNA
                 organism = Psilocybe cubensis

SEQUENCE: 1
atgcagggtga  taccgcgctg  caactcggca  gcaataagat  cactatgtcc  tactcccgag  60
tcttttagaa  acatgggatg  gctctctgtc  agcgatgcgg  tctacagcga  gttcatagga  120
gagttggcta  cccgcgcttc  caatcgaaat  tactccaacg  agttcggcct  catgcaacct  180
atccaggaat  tcaaggcttt  catgaaaagc  gaccgggtgg  tgcaccaaga  atttattgac  240
atgttcgagg  gcattcagga  ctctccaagg  aattatcagg  aactatgtaa  tatgttcaac  300
gatatcttcc  gcaaagctcc  cgtctacgga  gaccttgccc  ctcccgttta  tatgattatg  360
gccaaattaa  tgaacacccg  agcgggcttc  tctgcattca  cgagacaaaag  gttgaacctt  420
cacttcaaaa  aacttttcga  tacctgggga  ttgttccgtg  cttcgaaaga  ttctcgaaat  480
gttcttgtgg  cgaaccagtt  cgacgacaga  cattgcggct  ggttgaacga  gcgggccttg  540
tctgctatgg  ttaaacatta  caatggacgc  gcatttgatg  aagtcttcc  ctgcgataaa  600
aatgccccat  actacggctt  caactcttac  gacgacttct  ttaatcgcat  atttcgaaac  660
cgagatatcg  accgacctgt  agtcggtgga  gttacaaca  ccacctcat  ttctgctgct  720
tgcgaaatca  ttctctacaa  cgtctcttat  gacgtccagt  ctctcgacac  tttagttttc  780
aaaggagaga  cttattcgct  taagcatttg  ctgaataatg  accctttcac  cccacaattc  840
gagcatggga  gtattctaca  aggattcttg  aacgtcaccg  cttaccaccg  atggcacgca  900
cccgtaaatg  ggacaatcgt  caaaatcacc  aacgttccag  gtacctactt  tgcgcaagcc  960
ccgagcacga  ttggcgaccc  tatcccgcat  aacgattacg  acccaacctc  ttaccttaag  1020
tctcttgtct  acttctctaa  tattgcgcga  aggcaaatga  tgtttattga  agccgacaa  1080
aaggaaattg  gcctcatttt  ccttggttcc  atcggcatga  ccgaaatctc  gacatgtgaa  1140
gccacggtgt  ccgaaggtea  acacgtcaat  cgtggcgatg  acttggaagt  gttccatttc  1200
ggtggttctt  cgttcgcgct  tggctgagg  aaggattgca  gggcagagat  cgttgaaaag  1260
ttcacccgac  ccgaacagat  gatcagaatc  aacgaagtcg  tcgctgctct  aaaggetttag  1320

SEQ ID NO: 2      moltype = AA  length = 439
FEATURE          Location/Qualifiers
source          1..439
                 mol_type = protein
                 organism = Psilocybe cubensis

SEQUENCE: 2
MQVIPACNSA  AIRSLCPTPE  SFRNMGWLSV  SDAVYSEFIG  ELATRASNRN  YSNEFGLMQP  60
IQEFKAFIES  DPVVHQEFID  MFEGIQDSRP  NYQELCNMFN  DIPRKAPVYG  DLGPPVYMIM  120
AKLMNTRAGF  SAFTRQRLNL  HFKKLFDTWG  LFLSSKDSRN  VLVADQFDDR  HCGWLNERAL  180
SAMVKHYNGR  AFDEVFLCDK  NAPYYGFNSY  DDFNRRFRFN  RDIDRPVVG  VNNNTLISAA  240
CESLSYNVSY  DVQSLDLTVF  KGETYSLKHL  LNNDPPTPQF  EHGSILQGFL  NVTAYHRWHA  300
PVNGTIVKII  NVPPTYFAQA  PSTIGDPIPD  NDYDPPPYLK  SLVYFSNIAA  RQIMFIEADN  360
KEIGLIFLVF  IGMTIEISTC  ATVSEGOHVN  RGDDLGMFHF  GGSSPALGLR  KDCRAEIVEK  420
FTEPGTVIRI  NEVVAALKA  439

SEQ ID NO: 3      moltype = DNA  length = 2155
FEATURE          Location/Qualifiers
source          1..2155
                 mol_type = genomic DNA
                 organism = Psilocybe cubensis

SEQUENCE: 3
atgacgcgtg  tactattctc  cttcgtcatt  gcaggatgca  tatactacat  cgtttctcgt  60
agagtgaggg  ggtgcgcgtt  gccaccaggg  cgccttgcca  ttctatttcc  cttcattggg  120
aacatgtttg  atatgcctga  agaattccca  tgggttaaat  ttctacaatg  gggacgggat  180
tacagtctgt  ttgcccgcgt  tgacttctaa  tatatgaaca  gctaataat  tgccagacac  240
cgatattctc  tacgtggatg  ctggaggggc  agaaatgggt  attcttaaca  cgttggagac  300
cattaccgat  ctattagaaa  agcaggggtc  cattattctc  ggccggtgag  ctgatgttga  360
gattttttga  attgaatttg  tggtcacacg  ttccagact  tgagagtaca  atggtcaacg  420
aaccttatgg  gtgggagttt  gccttaggg  tcatcacata  cggcgacagg  tggcgcaag  480
aaaggcgcat  gttcgccaag  gagttcagtg  agaagggcac  caagcaattt  cgccatgctc  540
aagtgaaga  tgcccatcag  cttgtccaa  agcttaccaa  aacgccagac  cgctgggac  600
aacatattcg  ccagtaagta  ctacttgagg  aaaatagcgt  acgcttcgct  gaccgggtcc  660
tacatcaaa  tcagatagcg  gcaatgtcac  tggatattgg  ttatggaatt  gatcttcgag  720
aagacgaccc  ttggctggaa  gccacccatt  tggctaata  agcctcgc  atagcatcag  780
tgccggggca  attttgggtc  gattcgttcc  cttctcgtga  gcatccttct  tctatgtagg  840
aagggaagga  gtctaacaag  tgttagtaaa  ataccttcc  gcttggttcc  cagggtgctg  900
cttcaagcgc  aaagcgaagg  tctggcgaga  agccggcgac  catatgggtg  acatgcctta  960
tgaaactatg  aggaatttag  cagttagtca  aatgcgttct  ccccgatttt  tttcaatact  1020
ctaacttcag  ctcacagcct  caaggattga  ctgcctcgct  gtatgcttca  gctcgtctgc  1080
aagccatgga  tctcaacggt  gaccttgagc  atcaagaaca  cgtaataca  aacacagccg  1140
cagagggtta  tgtcggtta  tcaaaagcgt  ccgtcggcaa  ttcaaaattc  aggcgctaaa  1200
gtgggtcttc  tcaccaaggt  ggaggcgata  ctgtaaggat  ttctcaatcg  tttagagtata  1260
agtgttctaa  tgcagtacat  actccaccaa  ccagactgtc  tctgctatgt  ctgcgttcat  1320
cttggccatg  gtgaagtacc  ctgaggtcca  gcgaagggtt  caagcggagc  ttgatgctct  1380
gaccaataac  ggccaatttc  ctgactatga  cgaagaagat  gactccttgc  catacctcac  1440
cgcatgtatc  aaggagcttt  tccggtggaa  tcaaatcgca  cccctcgcta  taccgcacaa  1500
attaatgaag  gtagcagtg  accgcgggta  tctgattccc  aagaacactc  tagtcttcgc  1560
aaacacctgg  tgaggctgtc  cattcattcc  tagtacatcc  gttgccccac  taatagcatc  1620
ttgataacag  ggcagtatta  aacgatccag  aagtctatcc  agatccctct  gtgttccgcc  1680
cagaaagata  tcttggctcc  gacgggaagc  ctgataaac  tgtacgcgac  ccacgtaaa  1740
cggcatttgg  ctatggacga  cgaatttggt  aagtgcgctt  tcagaacccc  ccttccggt  1800
gactagtgcc  atgcgcgcat  acaatatcgc  tattgatctg  atataacttc  cctgcggcat  1860

```

-continued

ttattttggc	attcctttag	tcccggaatt	catctagcgc	agtcgacggt	ttggattgca	1920
ggggcaaccc	tcttatcagc	gttcaatata	gagcgacctg	tcgatcagaa	tggaagccc	1980
attgacatac	cggctgattt	tactacagga	ttcttcaggt	agctaatttc	cgtctttgtg	2040
tgataatac	ccctaacgac	gcacgtttac	ctttttgtaa	agacacccag	tgcccttcca	2100
gtgcagggtt	gttctcgaa	cagagcaagt	ctcacagtcg	gtatccggac	cctga	2155

SEQ ID NO: 4 moltype = AA length = 508
 FEATURE Location/Qualifiers
 source 1..508
 mol_type = protein
 organism = *Psilocybe cubensis*

SEQUENCE: 4

MIAVLFSFVI	AGCIYYIVSR	RVRRSRLPPG	PPGIPIPFIG	NMFDMPESP	WLTFLQWGRD	60
YNTDILYVDA	GGTEMVILNT	LETITDLLEK	RGSIYSGRLE	STMVNELMGW	EFDLGFITYG	120
DRWREERMF	AKEFSEKGIK	QFRHAQVKAA	HQLVQQLTKT	PDRWAQHHRH	QIAAMSLDIG	180
YGIDLAEDDP	WLEATHLANE	GLAIASVPGK	FWVDSFPLK	YLPWFPGAV	FKRKAKVWRE	240
AADHMVDMPY	ETMRKLAPQG	LTRPSYASAR	LQAMDLNGDL	EHQEHVIKNT	AAEVNVGGGD	300
TTVSAMSAFI	LAMVKYPEVQ	RKVQAEALDAL	TNNGQIPDYD	EEDDSLPLYL	ACIKELFRWN	360
QIAPLAIPHK	LMKDDVYRGY	LIPKNTLVFA	NTWAVLNDPE	VYPDPVSFRP	ERYLGPDKPK	420
DNTVRDPRKA	AFGYGRNCP	GIHLAQSTVW	IAGATLLSAF	NIERPVDQNG	KPIDIPADFT	480
TGFFRHPVPF	QCRFVPRTEQ	VSQSVSGP				508

SEQ ID NO: 5 moltype = DNA length = 1089
 FEATURE Location/Qualifiers
 source 1..1089
 mol_type = genomic DNA
 organism = *Psilocybe cubensis*

SEQUENCE: 5

atggcggtcg	atctcaagac	tgaagacggc	ctcatcacat	atctcaactaa	acatctttct	60
ttggacgtcg	acacgagcgg	agtgaagcgc	cttagcggag	gctttgtcaa	tgtaacctgg	120
cgcattaagc	tcaatgctcc	ttatcaaggt	catcacgaca	tcctcctgaa	gcattgctcag	180
cgcacatagt	ctacggatga	ggatttttaag	ataggtgtag	aacgttcggg	ttacgaatac	240
caggctatca	agctcatgat	ggccaatcgg	gaggttctgg	gaggcgtgga	tgccatagtt	300
tctgtgccag	aaggcctgaa	ctacgactta	gagaataatg	cattgatcat	gcaagatgtc	360
gggaagatga	agaccctttt	agattatgtc	accgccaac	cgccacttgc	gacggatata	420
gcccgccttg	ttgggacaga	aattgggggg	ttcgttgcca	gactccataa	cataggccgc	480
gagaggcgag	acgatcctga	gttcaaatc	ttctctggaa	atattgtcgg	aaggacgact	540
tcagaccagc	tgtatcaaac	catcataccc	aacgcagcga	aatatggcgt	cgatgacccc	600
ttgctgccta	ctgtggttaa	ggaccttgtg	gacgatgtca	tgacacagca	agagaccctt	660
gtcatggcgg	acctgtggag	tggaatatatt	cttctccagt	tggaggaggg	aaacccatcg	720
aagctgcaga	agatatatat	cctggattgg	gaacttttga	agtacggccc	agcgtcgctg	780
gacctgggct	atttcttggg	tgaactgctat	ttgatataccc	gctttcaaga	cgagcaggtc	840
ggtacgcaga	tgccggcaagc	ctacttgcaa	agctatgcgc	gtacgagcaa	gcattcgatc	900
aactacgcca	aagtcactgc	aggtatttgc	gctcatattg	tgatgtggac	cgactttatg	960
cagtggggga	gcgaggaaga	aaggataaat	tttgtgaaaa	agggggtagc	tgccctttcac	1020
gacgccaggg	gcaacaacga	caatggggaa	attacgtcta	ccttactgaa	ggaatcatcc	1080
actgcgtaa						1089

SEQ ID NO: 6 moltype = AA length = 309
 FEATURE Location/Qualifiers
 source 1..309
 mol_type = protein
 organism = *Psilocybe cubensis*

SEQUENCE: 6

MHIRNPYRTP	IDYQALSEAF	PPLKPFVSVN	ADGTSSVDLT	IPEAQRAFTA	ALLHRDFGLT	60
MTIPEDRLCP	TVPNRLNLYL	WIEDIFNYTN	KTLGLSDDRP	IKGVDIGTGA	SAIYPMLACA	120
RPKAWSMVGT	EVERKCIDTA	RLNVVANLQ	DRLSILETSI	DGPILVPIFE	ATEEYEFYFT	180
MCNPPFYDGA	EDMKDSDAK	GFGFGVGAPH	SGTVIEMSTE	GGESAFVAQM	VRESLKLRT	240
CRWYTSNLGK	LKSLKEIVGL	LKELEISNYA	INEYVQGST	RYAVAWSFTD	IQLPEELSRP	300
SNPELSSLF						309

SEQ ID NO: 7 moltype = DNA length = 930
 FEATURE Location/Qualifiers
 source 1..930
 mol_type = genomic DNA
 organism = *Psilocybe cubensis*

SEQUENCE: 7

atgcataatca	gaaatcctta	cgtacacca	attgactatc	aagcactttc	agaggccttc	60
cctcccccaca	agccatttgt	gtctgtcaat	gcagatggta	ccagttctgt	tgacctcaact	120
atcccagaag	cccagagggc	gttcacggcc	gctcttcttc	atcgtgactt	cgggctcacc	180
atgaccatac	cagaagaccg	tctgtgccca	acagtcccca	atagggtgaa	ctacgttctg	240
tgattgaag	atattttcaa	ctacacgaac	aaaaccctcg	gcctgtcgga	tgaccgtcct	300
attaaaggcg	ttgatattgg	tacaggagcc	tccgcaattt	atcctatgct	tgctgtgct	360
cggttcaagg	catggtctat	ggttggaaca	gaggtcgaga	ggaagtgcac	tgacacggcc	420
cgcctcaatg	tcgtcgcgaa	caatctccaa	gaccgtctct	cgatattaga	gacatccatt	480
gatggtccta	ttctcgtccc	cattttcgag	gcgactgaag	aatacgaata	cgagtttact	540
atgtgttaacc	ctccattcta	cgacggtgct	gccgatatgc	agacttcgga	tgctgccaaa	600
ggatttggat	ttggcggtgg	cgctccccc	tctggaacag	tcacgaaat	gtcgactgag	660
ggaggatgaat	cggctttcgt	cgctcagatg	gtccgtgaga	gcttgaagct	tcgaacacga	720

-continued

tgcagatggt	acacgagtaa	cttgggaaag	ctgaaatcct	tgaagaagaat	agtggggctg	780
ctgaaagAAC	ttgagataag	caactatgcc	attaacgaat	acgttcaggg	gtccacacgt	840
cgttatgccg	ttgcgtggtc	tttactgat	attcaactgc	ctgaggagct	ttctcgtccc	900
tctaaccocg	agctcagetc	tcttttctag				930

SEQ ID NO: 8 moltype = AA length = 309
 FEATURE Location/Qualifiers
 source 1..309
 mol_type = protein
 organism = *Psilocybe cubensis*

SEQUENCE: 8

MHIRNPYRTP	IDYQALSEAF	PPLKPFVSVN	ADGTSSVDLT	IPEAQRAFTA	ALLHRDFGLT	60
MTIPEDRLCP	TVPNRLNYVL	WIEDIFNYTN	KTLGLSDDRP	IKGVDIGTGA	SAIYPMLACA	120
RPKAWSMVGT	EVERKCIDTA	RLNVVANNLQ	DRLSILETSI	DGPILVPIPE	ATEEYEFYFT	180
MCNPPFYDGA	ADMQTSDAK	GFGFGVGAPH	SGTVIEMSTE	GGESAFVAQM	VRESLKLRT	240
CRWYTSNLGK	LKSLKEIVGL	LKELEISNYA	INEYVQGST	RYAVAWSFTD	IQLPEELSRP	300
SNPELSSSLF						309

SEQ ID NO: 9 moltype = DNA length = 3042
 FEATURE Location/Qualifiers
 source 1..3042
 mol_type = genomic DNA
 organism = *Psilocybe cubensis*

SEQUENCE: 9

atgccttcca	gtcaccctca	cattactcat	cgctatcg	ttccttcgag	tgacgacccat	60
gaacgtatat	gtctctgtgt	cttgggtccc	aaagcagaaa	atgccgcatt	tctccagcaa	120
tggttgacca	cggtcgtcgc	acagcaaaag	gctgcccgcg	atgcatactt	cccggatgac	180
aatgctttta	ttactacaga	catgcaaaact	tcccgcgcct	ttgctcagac	tactaaagta	240
atcgcttcca	atctcacoga	attattgact	gcaactcggtg	aaaggtcgat	tcttttcttc	300
tcacctcggt	acagcggcca	tatgtctgtg	gaccaaagtc	tacctgccat	tctcggattc	360
ttatcgacca	cattttataa	tcctaacaat	gttgccctcg	aggctagtcc	attcacgacc	420
ctcatcgagg	aagaagtgtg	cttgcaactc	tctgaaatgc	tgggttataa	tgggctaaat	480
aacaccgaga	aacctctcgc	ctggggacat	attgcatcag	gtggaactgt	tgcaaaactg	540
gaagcgatgt	ggcgggcgcg	aaacctcaag	ttttaccctc	tctcaactcg	tgatgcttca	600
gccgaaggcg	cagagatgga	attcattcgt	gacacattct	ccgtcaaaac	ctgtgttgtt	660
gacaaaaaat	tattaaagga	ttgcagccca	tgggaaactc	tcaatttgca	tgtttctact	720
atcttagaca	tgcccgaacg	tctgcacgac	gagtagaata	tttcaactca	gttctcga	780
aagggttatg	gaaagtatat	catccagctc	accaacaaag	acacgttgat	gcagcgttgg	840
ggacttacc	aacaacctgt	cgttttatcc	ccgagcacaa	accattatcc	ctggcctaa	900
gctgcagctg	tgctcggtat	tggtcagac	aaccttcgca	acgtccagct	agacatccaa	960
gcccacatgg	acataaacga	actcgatcgt	atgttaaaaa	tttgcttgga	cgaggagacg	1020
ccagtatatc	aagttagttgc	tgttatcggt	accacggaag	agggcggtgt	cgatcgcat	1080
acggagatcc	tgaagctg	gaagctttgg	ggctgtcttt	tgccatccat		1140
gcagatgctg	cttggggagg	ctattttgca	accatgctac	ccaaagatac	attgggtcga	1200
aaccgggacta	ggcttcccaa	agaggacact	acctcggtct	ttgtccctca	cgctcggtctg	1260
cgcgaggaga	gcgcgcatat	actcagccat	ataaagtatg	ccgattctat	tactatcgac	1320
ccgcacaagg	caggctatgt	tccttacc	gctggggcac	tctgttatcg	cgacggaaga	1380
atgaggtacc	tgcttacatg	gtccgcgc	tacctgtccc	aaggcaacga	gggcccgaagt	1440
atcggaaat	acgggatcga	aggaagcaaa	cctggtgcag	cagcatccgc	ggattcctatg	1500
gcgcacgaaa	ccattggcct	gactccttct	ggatacggga	accttcttgg	ccaggcaatg	1560
tttcatgtc	gccgatacgc	tgctcactgg	tctgcaatgt	caacgggatac	taccagtttc	1620
actgtcaccc	acttcaatcc	tatccctgct	gacatcgacc	ccaacgctga	cccgcgaag	1680
gtcgaagagc	aaaaacagtt	catcagagat	cgtatcttgt	tcaaatcgaa	cgaggaaata	1740
tacaacgatt	ctgaggctat	ggaactcttg	caccaacttg	ggctccgatct	caatatcaac	1800
gttttcgcat	gcaacttcgc	gcaccgcgat	aataatctca	acaccgacgt	cgagggaagcc	1860
aactgggtca	ataaacgctat	tttccaacgc	ttttctgtta	caagtgtctga	ggagaaccca	1920
ttggaaacgc	cattcttctc	cagctcaact	acattgaaac	aatccgaata	cggcgtctgc	1980
gcaaccgaag	taaagagacg	catgggactt	gttggtgacc	aggatgttat	agtcctgagg	2040
aacgtcgtaa	tgtctccatt	tactacaacg	aacgactttg	tgggaactct	ggcaaacacc	2100
ttccaaaaga	tcgttgagga	ggaggtcgag	tatgcacgga	tcgcgaacga	tatgaaacct	2160
agcattcaca	ccttcttctc	tcatggttca	ggagagcaat	actatcttgt	ccacacccca	2220
acgatccata	tggccagcgg	ccgtcgccaa	atcatccttt	cagtaaatgt	tgaaggccaa	2280
gttcggcagg	cgatacatgc	ccatgaaaga	gttgaagcag	tgattgtaca	taacactgtg	2340
cccccccgcc	ttgacgaaat	ctgtgacgga	ggatcatttg	acggcatact	caccatcgga	2400
aagaggaaaa	ctagtttcaa	agtgaagatt	tcaaacatta	aagtgtctca	gaagcgctct	2460
ctgatgactg	aggacctgga	atctgcgtac	ccatcggtga	tgccattcta	tttctacggg	2520
actcaaggac	acgctcatct	cgacctgtc	attactgtcg	ttcctaaccat	ccatctgagt	2580
gctggcgaaa	attcgacaca	gaggtgtcaa	gcgaggacct	cgccaaagggc		2640
ctcattgttg	ttgctgagaa	cgtacacgag	gcattccatgc	agcccttccc	gctcatgaaa	2700
gatttcaaga	tcaccaacca	attcttcttc	agctccgggc	aaataactccg	cgtaaaagtg	2760
tacagagatc	catacccgcc	atcgacaatg	gatcccatcc	ctctccacga	catcaagaac	2820
cagcccgctg	tgacacaagg	caccatcacg	ctcgtcgga	atatttacgt	cgattctgat	2880
gcgctcaacg	tcgcttccga	gctactgcc	gacgaagacg	cgggcgatgt	tcctcacgct	2940
cgaacacatg	acggcgagat	gaccgctgga	acgatcaaa	gctggcaaaa	cgctgttcgt	3000
catttccaca	acaaattgga	gactgttgtc	ccgacgaagt	ag		3042

SEQ ID NO: 10 moltype = AA length = 1013
 FEATURE Location/Qualifiers
 source 1..1013

-continued

```

mol_type = protein
organism = Psilocybe cubensis

SEQUENCE: 10
MPSSSPHITH RYRVPSDDH ERISALFLGP KAENAAFLQQ WLTTVVAQQK AARDAYFPDD 60
NAFITTDMQT SPAPAQTKV IASNLTELLT ALGERSIPFF SPRYSGHMSV DQSLPAILGF 120
LSTTFYNPNM VAFEASPTT LIEEEVGLQL SEMLGYNRLN NTEKPLAWGH IASGGTVANL 180
EAMWAARNLK FYPLSLRDAS AEGAEMEFIR DTFSVKTCVG DKLLKDCSP WELLNLHVST 240
ILDMPDRLHD EYNISPQFLE KVMRKYIIQS TNKDTLMQRW GLTQQPVVLS PSTNHYSWPK 300
AAAVLGIGSD NLRNVPVDIQ AHMDINELDR MLKICLDEET PVYQVVAVIG TTEEGVDRI 360
TEILKLQKY EALGLSFAIH ADAAWGGYFA TMLPKDTLGR NTRLPKEDT TSGFVPHVGL 420
REESALQLSH IKYADSITID PHKAGYVPYP AGALCYRDGR MRYLLTWSAP YLAQGNEGQS 480
IGIYIGIESK PGAAASAVFM AHETIGLTPS GYGNNLGQAM FTCCRYYAAHW SAMSTDTSF 540
TVTPFNPIPA DIDPNADPAK VEEQKQFIRD RILFKSNEEI YNDSEAMELL HQLGSDLNIN 600
VFACNFRDRD NNLNTDVEEA NWLNNRIFQR FSVTSAEENP LETPFFLSST TLKQSEYGV 660
ATEVKRRMGL VGDQDVIVLR NVVMSPTTT NDFVGLTANT FQKIVEEEVE YARIRNDMKP 720
SIHTFLHGS GEQYYLVHTP TIHMASGRRO IILSVNVEGQ VRQAIHAHER VEAIVHNTV 780
PLRLDEIVDG GSGFDGILTIG KRKTSFKVKI SNIKVVKKRS LMTEDLESAY PSLMPFFYFG 840
TQGHAAHLDHV ITVVPNIHLS AGEIQYKFD D EVSSEDLAKG LIVVAENVHE ASMOPFPLMK 900
DFKITNQPFV SSGILRVKV YRDPYPASTM DPIPLHDIKN QPVVTQGTIT LVGNIVVDS 960
ALNVASEPTA DEDAAHVPHA RNMYGEMTAG TIKGWQNAVR HFHNKLETV PTK 1013

```

```

SEQ ID NO: 11      moltype = DNA length = 2103
FEATURE           Location/Qualifiers
source            1..2103
                  mol_type = genomic DNA
                  organism = Psilocybe cubensis

```

```

SEQUENCE: 11
atggaggcta tcaaaaagggt tttgagaac aaaaaggcgg agggcattcc tgtgttggtg 60
acctttgtta tgcagagata tcctcgtccc gaagatactg ttcccatctt gctggccatg 120
gaggccgggt gtcctgatat catcgagctt ggtatgccat ttccagacc aattgcagat 180
ggctctgtca tccaggaaac gaacacaatc gccgttgcaa accaggtaga ttataccact 240
gtttctcgga aacttcggga agcccgcaaa caagggtcca aggcaccgtg tcttctgatg 300
ggatattata acccatattt ggcttacgga gaagacagat ctattcaaga tgcggctgaa 360
gctggagcca atgggtttat tatggtcgac ctccaccgg aggaggtgtg cgcttttcga 420
gagaaatgta tcaaatccaa cctctcatat gtctctctaa ttgcaccctc aacgactctg 480
tcgcgtataa agttcctctc aacaattgca gacacgttca tctatgtcgt gtctaaaatg 540
ggaaccaccc gatcctcaga gaagggtgcc atgaataacg ccttccacc catcatcgat 600
cgtattccgc agtacgctga agttccttta gcagtcggat ttggagtcgc cactcgggct 660
cacttcaact acgtcgccga ttccggtgct gatggtgtcg ttattggcac caaactcggt 720
aacgttatta aagagtcacc gcaaggggaa gcacccaaaa atgttgaggc atactgccgt 780
gagatgagcc aaaagggaga aacaaatcgc gtcaaatctc caccaaactgc ccgtgctgcc 840
agctccgaat caattcctgt tgtgttctct tctgttctcc ccgcacgttt cgagaaattc 900
ggaggacaat acgttcccca agctcttgct gattgtctgg ttgaactaga agaagctcac 960
aaatctgcca tggctgatcc tgaattccag aaggaactac aatcgcatgc cggatatgca 1020
aatcgtcctt cacaataata cctcgccgaa aatctcacc aggatgctgg ggggtgcaat 1080
atttggttga aacgtgaaga ttgaaccac acaggttccc acaaaatcaa taacgctttg 1140
ggacaaattc tgcttgcccg gagaatcgga aagaccagaa ttatcgaga aacaggtgcc 1200
ggccagcatg gtgttgcaac agcgactgtt tgcgcataag ttggaatgga atgtgttatc 1260
tacatgggag cagaagatgt gcgagggcaa gctctaaatg tatcaggat tgagatgcta 1320
ggagcaaaag ttgttctctg tacttcagga tcatgcacat tgaaggacgc tgtaaacgag 1380
gccttcctgt actgggtgac aaacctttct acgacgcatt atttggttgg ctctgtaatt 1440
ggacctcatc ccttcccacc cattgtccga gatttccaaa aggtcaattg tcaagagatc 1500
aaggctcaga tgttgccgc cccgcccga cttctctgat tgcgtcgcgc ttgtgttggt 1560
ggaggaagca atgctatcgg tacgttctat gattttattg gcgacaagag tgtacgtcta 1620
gttgggttgg aagcaggagc agaaggtatt gacggagacc gacatagcgc cacactttcg 1680
atggggcaac cgggagtact tcacgtgtt agaacatata ttctacaaga caaggccggt 1740
caaatcatcg agacgcactc aatcagcgct ggattggatt atcccggcgt tggaccagaa 1800
catgcttggc taaaggactc taaaagagca gaatatgttg tgcacacaga cgaagaagca 1860
cttcgcggtt tccgtatgct aacacaaaag gaggaatta ttctgcccct tgaatcttcc 1920
catgcatctc gggaggctgt caggattgcc cgcaccatgt cgaaggacca ggatcttgtt 1980
gtgtgtttgt ctggcccgag gtgataaagc gttgagcaaa tttctcaact tcttcccaag 2040
tgggaggata ttctagactg gcatgtttct tcccatgccg ttggacacac aacaaaattc 2100
taa 2103

```

```

SEQ ID NO: 12      moltype = AA length = 700
FEATURE           Location/Qualifiers
source            1..700
                  mol_type = protein
                  organism = Psilocybe cubensis

```

```

SEQUENCE: 12
MEAIKKVFEN KKAEGIPVLV TFFTGYPRP EDTVPILLAM EAGGADIIEI GMPFSDPIAD 60
GPVIQETNTI AVANQVDYTT VLQQLREARK QGLKAPVLLM GYNNPILAYG EDRSIQDAE 120
AGANGFIMVD LPPEEAVAFR EKCIKSNLSY VPLIAPSTTL SRIKFLSTIA DTFIYVVKM 180
GTTGSSEKVA MNNALPTIID RIREYAEVPL AVGFGVATRA HFNVYADSGA DGVIIGTKLV 240
NVIKESPQGE APKNVEAYCR EMSQKGETNR VKSPPTARAA SSESIPVVVP SVLPARFGEF 300
GGQYVPEALV DCLVELEEAH KSAMADPEFQ KELQSHAGYA NRPSQIYLA NLTKDAGGAN 360
IWLKREDLNH TGSCHKINNAL GQILLARRIG KTRIIAETGA GQHG VATATV CAKFGMECVI 420
YMGAEDEVRRQ ALNVFRIEML GAKVVPVTS G SCTLKDAVNE AFRDVTNLS TTHYLVGSEI 480
GPHPPPTIVR DFQKVIQOEI KAQMLAARGK LPDVVVACVG GGSNAIGTFY DFIGDKSVRL 540

```

-continued

VGVEAGGEGE	DGDRHSATLS	MGQPGVLHGV	RTYILQDKAG	QIIETHSISA	GLDYPGVGPE	600
HAWLKDSKRA	EYVVATDEEA	LRGFRMLTQR	EGTIIPALESS	HAIWEAVRIA	RTMSKDQDLV	660
VCLSGRGDKD	VEQISQLLPK	WADILDWHVS	SHAVGHTTKF			700

SEQ ID NO: 13 moltype = DNA length = 1599
 FEATURE Location/Qualifiers
 source 1..1599
 mol_type = genomic DNA
 organism = Streptomyces rugosporus

SEQUENCE: 13

atggaaaggc	ggaagcgtga	gcgtcttggc	tcaactcgcc	gaccgaccaa	aaaggagctc	60
cgcatgatcc	gatctgtggg	gatcgtgggt	ggtggcacgg	cgggctggat	gaccgcctcc	120
tacctcaagg	cgcccttcga	cgaccgcac	gacgtaacgc	tcgtggagtc	agggaacgtc	180
aggcggatcg	gggtcggcga	agcgaccttc	agcacgggtc	gccacttctt	cgactacctg	240
ggcctcgacg	agcgcgagtg	gctgccccgc	tgcgcggcgc	gctacaagct	cggcacccgc	300
ttcagaagac	ggagcgagcc	gggcgagtag	ttctaccacc	cggtcgagcg	cctgcgcgtc	360
gtcgacggct	tcaacatggc	cgagtgggtg	ctcgcgggtc	gcgaccgcag	gacgtccttc	420
agcgaggcct	gctatctcac	gcaccggctg	tgcgaggcca	agcgggcccgc	ccgcatgctc	480
gacggctcgc	tcttcgcctc	ccagggtggac	gagtcgctcg	gccgctcgac	cctggccgag	540
cagcgccgcc	agttcccgtg	cgccctacc	ttcgacgcgc	acgaggtcgc	ccgctacctg	600
tcggagtagc	ccatcgcccg	cggcgctccg	cacgtgggtc	acgacgtgca	gcacgtcggc	660
caggacgagc	gcggctggat	cagcgccgtc	cacaccaagc	agcacggcga	gacacggcgc	720
gacctgttcg	tccgtctcac	cggtctccgc	ggcctgctca	tcaaccagac	gctgggcccgc	780
aggttccagt	ccttctccga	cgtgctgccc	aacaacccgg	cggctcgctc	gcgcgtcccg	840
cgggagaacg	acgaggacat	gcggccgtac	acgacggcga	ccgcatgag	gcgggctgg	900
atgtggagca	tcccgctgtt	caagcgcgac	ggcaacggct	acgtctactc	cgacgagttc	960
atctcgccgc	aggaggccga	gcgcgagctg	cgggtccacc	tcgcccccg	ccgcgacgac	1020
ctggaggcca	accacatcca	gatgcggatc	ggcaggaacg	agcgccacct	gatcaacaac	1080
tgcgtcgccg	tcggcctgtc	cgccgccttc	gtcgagccgc	tggagtcgac	cggcattctc	1140
ttcatccagc	acgcccacga	gcagctcgtg	aagcacttcc	ccggcgagcg	ctgggaccgc	1200
gtgctgatca	gcgcgtacaa	cgagcgcac	gcgcacatgg	tcgacggggg	caaggagttc	1260
ctcgtccctc	actacaagg	cgcccagcgc	gaggacaccc	cctactggaa	ggccgccaag	1320
accagggcca	tgcccagcgc	cctcgccccg	aagctggagc	tgtccgcctc	ccacctgctg	1380
gacgagcaga	cgatctaccc	ctactaccac	ggcttcgaga	cctattcgtg	gatcaccatg	1440
aacctcggtc	tcggcatcgt	gcccgcgagg	cgcgctccgc	cgctcctgca	catggaccgc	1500
gcgcccgcgc	tggccgagtt	cgaacggctc	aggcgcgagg	gcgacgagct	gatcgccgcc	1560
ctgcccagct	gctacgagta	cctcgccagg	atccaatga			1599

SEQ ID NO: 14 moltype = AA length = 532
 FEATURE Location/Qualifiers
 source 1..532
 mol_type = protein
 organism = Streptomyces rugosporus

SEQUENCE: 14

MERRKRERLG	SLGRPTKKEL	RMIRSVVIVG	GGTAGWMTAS	YLKAAFDDRI	DVTLVESGNV	60
RRIGVGEATF	STVRHFFDYL	GLDEREWLPR	CAGGYKLGIR	FENWSEPGY	FYHPFERLRV	120
VDGFNMAEWW	LAVGDRRTSF	SEACYLTHRL	CEAKRAPRML	DGSLFASQVD	ESLGRSTLAE	180
QRAQFPYAYH	FDADDEVARYL	SEYAIARGVR	HVVDVQHV	QDERGWISGV	HTKQHGIEISG	240
DLFVDTCTGFR	GLLINQTLGG	RFQSFSDVLP	NNRAVALRVP	RENDEDMRPY	TTATAMSAGW	300
MWTIPLFKRD	NGYVYVDEF	ISPPEAEREL	RSTVAPGRDD	LEANHIQMRI	GRNERTWINN	360
CVAVGLSAAF	VEPLESTGIF	FIQHAIEQLV	KHFPGERWDP	VLI SAYNERM	AHMVDGVKEF	420
LVLHYKGAQR	EDTPYKAAK	TRAMPDGLAR	KLELSASHLL	DEQTIYPYH	GFETYSWITM	480
NLGLGIVPER	PRPALLHMDP	APALAEFERL	RREGDELIAA	LPSCYEYLAS	IQ	532

SEQ ID NO: 15 moltype = DNA length = 1518
 FEATURE Location/Qualifiers
 source 1..1518
 mol_type = genomic DNA
 organism = Kutzneria sp. 744

SEQUENCE: 15

atgaccgccc	ctacactgaa	gaccgcgttc	ggcgaccggc	tttcgatcac	cgctcgtgaa	60
tcgagcagaa	tcgggacccat	cggcgtcggg	gaggcgacct	tcagcgacat	ccagcacttc	120
ttccagttcc	tcaacctccg	ggaacaggac	tggatgccgg	cctgcaacgc	caactacaag	180
ctgggcaccc	ggttcgagaa	ctggcgccat	gtcgccatc	atttctacca	gccgttcgag	240
cagatccgcc	gggtctacgg	ctttccgctg	accgactggg	ggctgcacga	cgcgccgacc	300
gaccgcttcg	acaccgactg	cttcgtccat	cccaacctgt	gcgaggccgg	gcgcagcccg	360
cgccacctcg	acggcacgct	ggccgacgag	gacttcgtgg	aggagggcga	cgagctcgcc	420
aaccgcacca	tgtccgagca	ccagggaag	tcgcagttcc	cgtacgccta	ccacttcgag	480
gcggcgctgc	tggccaagtt	cctcacccgg	tacgcggtgg	accgcggcgt	cgagcacgtc	540
gtcgacgacg	tcctggacgt	gcggctggac	cagcgccggg	ggatcgagca	cgtggtcacc	600
gccgagcacg	gcgagatcca	cggcgacctg	ttcgtcgact	gcaccggttt	ccggggagctg	660
ctgctgaaca	aggcgctcgg	tgtgccgttc	gtgtctctac	aggacactct	gccccaacgac	720
agcgcggtcg	ccctccaggt	cccgctggac	atgcagcgcc	gcggcatcgt	gccgaacacc	780
acggccaccc	cgggggaggg	cggtcggttc	tggaccatcc	cgctgttcgg	tcgggtcgcc	840
accgggtacg	tctacgcccc	ggactacctc	tcgcccagg	aggccgagcg	cacgtgcgc	900
gagttcgtcg	gcccgccggc	cgcggacgtg	gaggccaacc	acatccgcat	gcggatcgcc	960
gcgagccagg	agtcctggcg	gaacaactgc	gtggccatcg	gcctgtccag	cggcttcgtc	1020
gagcgctggg	agtcgacgg	gactctcttc	atccaccacg	ccatcgagca	gctggtagag	1080
cacttcccg	cgcccgactg	gaacccgaag	tcccgggaca	tgtacaactc	cgcggtcgcg	1140

-continued

```

cacgtgatgg acggcatccg cgagttctctg gtgateccact accggggcgcg ggcgcggggcc 1200
gacaaccagt actggcgcgga caccaagacc cgcccgctgc cggacggcct ggcgcgagcgc 1260
atcgagtgct ggcagacgca gctgccggac accgagacga tctacccgta ctaccacggg 1320
ctgccgcctt actcctacat gtgcatectg atggggcgcg gggccatccg gacgcgggcc 1380
tcggcccgct tggccctgac cgaccagggg gccgcgcaga aggagttcgc cgcggtccgg 1440
gaccgggccc cgcagctgcg cgacacgctg cccagccact acgagtaact cgcgcgggatg 1500
cgtggtctgg atgtctga                                     1518

```

```

SEQ ID NO: 16      moltype = AA length = 505
FEATURE           Location/Qualifiers
source            1..505
                  mol_type = protein
                  organism = Kutzneria sp. 744

```

```

SEQUENCE: 16
MTAAYLKTAFA GDRLSITVVE SSRIGTIGVG EATFSDIQHF FQFLNLRQED WMPACNATYK 60
LGIRFENWRH VGHHFYOPFE QIRPVYGFPL TDWWLHDAPT DRFDTDCEVM PNLCEAGRSP 120
RHLDTGLADE DFVEEGDELA NRTMSEHQGK SQFPYAYHFE AALLAKFLTG YAVDRGVEHV 180
VDDVLDVRLD QRGWIEHVVT AEHGEIHGDL FVDCTGFRGL LLNKALGVPF VSYQDTLPND 240
SAVALQVPLD MQRRIIVPNT TATAREAGWI WTIPLFGRVG TGVYAKDYL SPEEAERTLR 300
EFVGPAAADV EANHIRMIRG RSQESWRNNC VAIGLSSGFV EPLESTGIFF IHHAIEQLVK 360
HPPAADWNPK SRDMYNSAVA HVMGDIREFL VIHYRGAARA DNQYWRDTKT RPLPDGLAER 420
IECWQTQLPD TETIYPYHGH LPPYSYMCIL MGGGAIRTPA SAALALTDQG AAQKEFAAVR 480
DRAAQLRDTL PSHYEYLARM RGLDV                                     505

```

```

SEQ ID NO: 17      moltype = DNA length = 1572
FEATURE           Location/Qualifiers
source            1..1572
                  mol_type = genomic DNA
                  organism = Streptomyces toxytrinic

```

```

SEQUENCE: 17
atgaacacga gaaatccgga caaggtcgct atcgctcgcg cgcgacggcg gggatggatg 60
accgcctcgt acctgaagaa ggcgctcgcg gagecgctgt ccgtgacctt cgtcgaaatcg 120
ggcaccatcg gcaccgtcgg cgtcgcgcgag cgcaccttca cgcacatccg gcactttctt 180
gaattctcgg acctgcggga ggaggaatgg atgccggcct gcaacgccac gtacaagctg 240
gcgcgtcgct tccaggactg gcagcgcccc ggccaccact tctaccaccc gttcgagcag 300
atgcgctcgg tagacgggtt cccctcgacc gactgggtgg tgcagaacgg gccgaccgac 360
cgggttcgaca gggactgctt cgtgatggcg tcgctgtgcg acgcggggcg cagcccgcgc 420
tatctcaacg gttccctgct gcagcaggag ttcgacgagc gcgcggagga gccggcgggg 480
ctcaccatgt ccagagcaca gggcaagacg cagttccctt acgcctacca cttcgaggcg 540
gccctgctcg ccgaattcct ttctgggtac tccaaggacc gcgcggtgaa gcacgtggtc 600
gacgaggtgc tggaggtgaa gctcgacgac cgcggatgga tcagccacct tgtcaccacg 660
gagcaccggc acatcgcgcg cgacctgttc gtcgactgca ccggtttccg cggcgctactg 720
ctgaaccagg cgctgggtgt gcccttcgtc tctaccagg acaccctgcc caatgacagc 780
gcggtcgccc tccagggtgc gctcgacatg gaggcgcggg gcaccccgcc gtacaccggg 840
gccaccgcca aggaggcgcg gtggatctgg acgatcccg tgatcgcccg gatcggcacc 900
gggtacgtgt acgccaaagg ctactgctcg ccgagaggag ccgagcgcac cctgcgcgag 960
ttcgtcggtc cggaggcgcg cgacgtcgag gccaaaccac tccgcatgcg gatcgggcg 1020
agcgagcagt cctggaagaa caactgcgtc gcgatcgggc tctccagtgg ctctgctgaa 1080
cccccgagt cgaccggcat ctcttctcat caccacgcca tcgagcagtt ggtgaagcac 1140
ttcccgcgcg gcgactggca cccgcagctg cgcgcgggct acaactcgcc cgtcgccaac 1200
gtcatggacg gggctccggga gttcctcgtc ctgcactacc tgggcgcgcg gcgcaacgac 1260
accgggtact ggaagggcac caagaccggg gcggttcggg atgcccggg ggagcgcatc 1320
gagcgctgga aggtgcagct gcccgactcc gagaacgtct tcccctacta ccacggcctg 1380
ccgcgctact cgtacatggc catcctgctc gggaccggcg cgatcggggt gcgaccgctg 1440
ccgcgctcgc cactggccga cccggcgcgcg gcggagaagg agttcaccgc gatccgggac 1500
agggcgcggt tcctggctga caccctgcgg agccagtagc agtacttcgc cgcgatgggg 1560
cagcgtgtct ga                                     1572

```

```

SEQ ID NO: 18      moltype = AA length = 523
FEATURE           Location/Qualifiers
source            1..523
                  mol_type = protein
                  organism = Streptomyces toxytrinic

```

```

SEQUENCE: 18
MNTRNPDKVV IVGGGTAGWM TASYLKKAFG ERVSVTLVES GTIGTVGVGE ATFSDIRHFF 60
EFLDLREEEW MPACNATYKL AVRFQDWQRP GHFHYHPFEQ MRSVDGFFLT DWLQNGPTD 120
RFDRDCFVMA SLCDAGRSRPR YLNGSLQQE FDERAEPPAG LTMSEHQGKT QFPYAYHFEA 180
ALLAEFLSGY SKDRGVKHVV DEVLEVKLDD RGWISHVVTG EHGDIIGDLF VDCGFRGVL 240
LNQALGVFPV SYQDTLPNDS AVALQVPLDM EARGIPPYTR ATAKEAGWIW TIPLIGRIGT 300
GYVYAKDYCS PEEAERTLRE FVGPEAADVE ANHIRMIRIG SEQSWKNNCV AIGLSSGFVE 360
PLESTGIFFI HHATEQLVKH FPAGDWHQPQ RAGYNSAVAN VMDGVREFLV LHYLGAARNR 420
TRYWKDTRKT AVPDALAERI ERWKVQLPDS ENVFPYYHGL PPYSYMAILL GTGAIGLRPS 480
PALALADPAA AEKEFTAIRD RARFLVDTLP SQYEYFAAMG QRV 523

```

```

SEQ ID NO: 19      moltype = DNA length = 1596
FEATURE           Location/Qualifiers
source            1..1596
                  mol_type = genomic DNA
                  organism = Streptomyces albobogriseolus

```

-continued

SEQUENCE: 19

atggacaatc	gaatcaagac	agtctgtgac	ctgggaggcg	gcaccgccc	ctggatgact	60
gcggcggtac	tccgcaaggc	gctgcagAAC	accgtgaaga	tctgtgctcc	ggaggcgccc	120
accatcccc	ggatcggtgt	gggcgaagcc	accgtcccca	acctgcagcg	cgcgttcttc	180
gactacctcg	gcatacccca	ggaggagtg	atgcgcgagt	gcaacgccag	ttacaagatg	240
gcggtgaagt	tcatcaactg	gcgcacaccg	ggggagggca	gcccggatcc	ccggacgctc	300
gacgacggtc	acaccgacac	cttccaccac	cccttcggac	tgctgcccgc	cgcgaccag	360
ataccgctct	cccactactg	ggcgcccaag	cggctccagg	gcgagaccga	cgagaacttc	420
gacgaggcgt	gtttcgccga	caccgcgac	atgaacgcc	agaaggccc	ccgcttcttc	480
gacatgccc	gggcgaccaa	ctacgcctgg	cacttcgacg	ccagcaaggt	cgcgcgcttc	540
ctgcgcaact	tccgctgcac	caagcaggcg	gtcgagcacg	tcgaggacga	gatgaccgag	600
gtcctcaacc	acgaacgggg	cttcatcacc	gccctgcgca	caaagtcggg	acggatcctg	660
cagggtgacc	tcttcgtcga	ctgctccggc	ttcccgggcg	tgctgatcaa	caaggccatg	720
gaggagccgt	tcatcgacat	ctccgaccac	ctcctgtgca	acagccgggt	cgcacggccc	780
gtaccgcacg	acgacgagaa	gaacggtgtc	gagccgtaca	cctcctcgat	cgcctaggag	840
gccggctgga	cctggaagat	tcccatgctg	ggccggttcg	gcagcgggtc	cgtctactcc	900
gaccactctc	cgaaccagga	caggcccaac	ctggccttct	cgaagctgtg	gggcctcgac	960
ccggacaaca	ccgagttcaa	ccacgtccgc	ttccgggtcg	gccccaaccg	cagggccttg	1020
gtgcgcaact	gcgtcagcgt	cggcctcgcg	tcgtgcttcg	tgagccgctc	ggagtcacgc	1080
ggcatctact	tcatctacgc	ggccatccac	atgctggcca	agcaactccc	cgacaagacc	1140
ttcgacaagg	tgctcgtcga	ccggttcaac	cgggagatcg	aggagatggt	cgacgacacc	1200
cgggacttcc	tccaggcgca	ctactacttc	tcgcccgcgc	tcgacacgcc	gttctggcgg	1260
gcgaacaagg	aactgaagct	ggccgactcc	atcaaggaca	aggctcgagc	gtaccggggc	1320
gggctgccc	tcaacctgcc	ggtcaccgac	gaggggacct	actacggcaa	cttcgaggcc	1380
gagttccgga	actctggac	caacggcagc	tactactgca	tcttcgccc	cctcgccctg	1440
atgccgcgca	accctgtgcc	cgcctcgccc	tacaagccgc	agtcgacgc	ggaggccgag	1500
ctgctgttcg	ccgacgtcaa	gcgcaagggg	gacaccctcg	tcgagtcgct	gccgtcgacc	1560
tatgacctgc	tgcgtcagtt	gcacggtgcg	tcgtga			1596

SEQ ID NO: 20 moltype = AA length = 531
 FEATURE Location/Qualifiers
 source 1..531
 mol_type = protein
 organism = Streptomyces albogriseolus

SEQUENCE: 20

MDNRKIKTVVI	LGGGTAGWMT	AAYLGKALQN	TVKIVVLEAP	TIPRIGVGEA	TVPNLQRAFF	60
DYLGIPPEEW	MRECNASYKM	AVKFINWRTP	GEKSPDPRTL	DDGHTDTFHH	PFGLLPADQ	120
IPLSHYWAAK	RLQGETDENF	DEACFADTAI	MNAKKAPRFL	DMRRATNYAW	HFDASKVAAP	180
LRNFAVTKQA	VEHVEDEMTE	VLTDERGFIT	ALRTKSGRIL	QGDLEFVDCSG	FRGLLINKAM	240
EPFFIDMSDH	LLCNSAVATA	VPHDDEKNGV	EPYTSIAME	AGWTWKIPML	GRFGSGHVYS	300
DHFATQDEAT	LAFSKLWGLD	PDNTEFNHVR	FRVGRNRRAW	VRNCVSVGLA	SCFVEPLESS	360
GIYFYIAAIH	MLAKHFPDKT	FDKVLVDRFN	REIEEMFDDT	RDFLQAHYYF	SPRVDTFPFWR	420
ANKELKLADS	IKDKVETYRA	GLPVNLPTVD	EGTYYGNFEA	EFNRNFWTNGS	YYCIFAGLGL	480
MPRNPLPALA	YKPQSIABAE	LLFADVKRKG	DTLVESLPST	YDLLRQLHGA	S	531

SEQ ID NO: 21 moltype = DNA length = 1590
 FEATURE Location/Qualifiers
 source 1..1590
 mol_type = genomic DNA
 note = BorH halogenase
 organism = unidentified

SEQUENCE: 21

atggacaacc	gcatacaacc	gatcgtcacc	ctcggcgccg	ggaccgccc	atggatgacc	60
gcctcctacc	tggcgaaggc	gctcggtgac	accgtcaccg	tcaccctgct	ggaagcaccg	120
gcgacgcgcc	ggatcggggt	cgccgagccc	accgtgccca	acctgcagcg	cgtgttcttc	180
gacttcctcg	gcctccgcga	ggaggagtg	atgcggaggt	gcaacgcgcg	gttcaagacg	240
gccgtgaagt	tcatcaactg	gcgtaccccc	ggggccggcg	aggccaaggc	ccgcacgacg	300
gacggccggc	cggaccactt	ctaccaccgc	ttcgggctgc	tgcccagagc	cggccagggtg	360
ccgctgtcgc	actactgggc	ctacaaccgc	gcggccggca	ccaccgacga	gccgttcgac	420
tacgcctggt	tcgccgagac	gcgcccatg	gacgcgggtc	gcgcgcccaa	gtggctcgac	480
ggcgccggcc	ccaccctgta	tgccctggac	ttcgacgccc	acctggtcgc	cgagttctctg	540
cgccggcgac	ccaccgagcg	gctcaacgtg	gagcacgtgc	agggtgagat	gcagcagggtg	600
ctgcgcgacg	agcggggctt	catcaccgcc	ctgcgcacgg	tcgaggcccg	ggacctcgag	660
ggtagacctg	tcatcgactg	ctccgcttc	cgcggcctgc	tgatcaacaa	ggccatggag	720
gagcccttca	tcgacatgaa	cgaccagctg	ctgtgcaacc	gcgcgggtgg	cacggcgatc	780
aagcacgacg	acgacgcgca	cggggtcgag	ccgtacacct	cggcgatcgc	catgcccgtc	840
ggctggagct	ggaagatccc	gatgctgggc	cgcttcggga	ccggctacgt	ctactccagc	900
cggttcgccg	agaaggacga	ggccaccctc	gacttctgcc	ggatgtgggg	cctgggacccg	960
gagaacaccc	cgctcaacca	ggtggccttc	cgggtcggcc	gcaaccgccg	ggcctgggtg	1020
aagaactcgg	tgagcatcgg	cctggcgctg	tgcttcctcg	agccactgga	gtccaccggc	1080
atctacttca	tcaccgcgcc	catctaccag	ctgactcagc	acttcccggg	ccgcaccttc	1140
gcgctcgcgt	tgagcgacgc	gttcaaccac	gagatcgaag	ccatgttcga	cgacaccggg	1200
gacttcctcc	agcgcgactt	ctacgtgtcg	ccgcgcacgg	acacgcggtt	ctggaaggcg	1260
aacaaggacc	tgacacctgc	cgagcagatg	cgcgagaaga	tcgagatgta	caaggccggc	1320
ctgcgcgaca	acgcgcgggt	caccgacgag	tcgacctact	acggcagggt	cgaggccgag	1380
ttccgcaact	tctggaccaa	cggcagctac	tactgcactc	tcgcccggct	cgggctgcgc	1440
ccggacaacc	cgtgcgcgat	gctgcgacac	cgtccggagc	aggctccgca	ggcgcaggcg	1500
ctgttcgcgc	gggtgaagga	caagcagcgg	gagttggtgg	agacgctgcc	gagcaacctg	1560
gagttcctgc	gcagcctgca	cggcaagtag				1590

-continued

SEQ ID NO: 22 moltype = AA length = 529
 FEATURE Location/Qualifiers
 source 1..529
 mol_type = protein
 note = BorH halogenase
 organism = unidentified

SEQUENCE: 22
 MDNRINRIVI LGGGTAGWMT ASYLAKALGD TVTITLLEAP AIGRIGVGEA TVPNLQRVFF 60
 DFLGLREEEW MPECNAAFKT AVKFINWRTP GPGEAKARTI DGRPDHFYHP FGLLPEHGQV 120
 PLSHYWAYNR AAGTTDEFPD YACFAETAAM DAVRAPKWLD GRPATRYAWH FDAHLVAEFL 180
 RRHATERLNV EHVQEMQOV LRDERGFITA LRTVEGRDLE GDLFIDCSGF RGLLINKAME 240
 EPFIDMNDQL LCNRAVATAI KHDDDAHGVE PYTSAIAMRS GWSWKIPMLG RFGTGVVYSS 300
 RFAEKDEATL DFCRMWGLDP ENTPLNQVAF RVGRNRRRAW KNCVSI GLAS CFLEPLESTG 360
 IYFITAAIYQ LTQHFPDRTF ALALSDAFNH EIEAMFDDTR DFIQAHFYVS PRDTPFWKA 420
 NKDLHLPEQM REKIAMYKAG LPINAPVTDE STYYGRFEAE FRNFWTNGSY YCIFAGLGLR 480
 PDNPLPMLRH RPEQVREAQA LFAGVKDKQR ELVETLPSNL EFLRSLHGK 529

SEQ ID NO: 23 moltype = DNA length = 1602
 FEATURE Location/Qualifiers
 source 1..1602
 mol_type = genomic DNA
 organism = Kutzneria sp. 744

SEQUENCE: 23
 atggatgaca atcgaattcg gagcatcctt gtccttgccg gcggcacggc cggtcggtatg 60
 tccgcctgct acctgagcaa ggcgctcggg cccggcgctg aggtcacggt gctcgaggcg 120
 ccctccatct cgcgcacccg ggtcggcgag gccaccattc ccaacctgca caaggtcttc 180
 ttcgaactcc tgggcacatcg cgaggacgag tggatgcggg agtgcaacgc cagctacaag 240
 gccgcgggtc ggttcgtcaa ctggcgagcg cggggcgacg gccaggccac gccgcggcgg 300
 cgtccggacg gccgccccga ccacttcgac cactgtgctc gccagctgcc cgagcacgag 360
 aacctgcgcg tgtcgagta ctgggcgcac cggcgctcga acggcctgac cgacgaaccc 420
 ttcgaccgct cctgctacgt gcagcccgag ctgctggacc gcaagctctc gccgaggttg 480
 atggacggca cgaaactggc cagctacgcc tggcacttcg acgcccagct ggtggccgac 540
 ttcctctgcc ggttcgctgt gcagaagctg aacgtgaccc acgtccagga cgtgttcacg 600
 catgccgacc tcgaccagcg cggccacatc acggccgctc acaccgagtc cggccgcacg 660
 ctggccgccc acctgttcat cgaactgcagc ggtctccgca gcgtgctcat gggcaaggtc 720
 atgcaggagc cgttcctgga catgagcaag cacctgctca acgaccgcgc ggtggcgctg 780
 atgtctccgc acgacgacga gaaggtgggc atcgagccgt acacctcgct gctggccatg 840
 cggtcgggct ggtcgtggaa gatcccgctg ctccgcccgt tcggtctccg ctacgtctac 900
 tccagccagt tcacctccca ggaagaggcg gccgaggagc tctgccgat gtgggacgtc 960
 gaccgcggcg agcagacgtt caacaacgtc cgggtccggg tcggcccgac cgcgggggccc 1020
 tgggtgcgca actgcgtcgc catcgccgtg tccgccatgt tcgtggagcc gctggagtcg 1080
 accggcctgt acttcagtta cgctcgctc taccagctgg tgaagcactt cccggacaag 1140
 cgggttcggc cgatcctggc cgaccgggtc aaccgcgagg tggcgaccat gtacgacgac 1200
 accgcgact tcctccaggc gcaactcagc ctgtcgccgc gtgacgactc cgagtctcgg 1260
 cgggcctgca aggagctgcc gttcgccgac ggggttcgcc agaaggtcga gatgtacagg 1320
 gccgggctgc cggtcgaact gccggtcacc atcgacgacg ggcactacta cggcaatttc 1380
 gaggccgagt tccgcaactt ctggaccaac tcgaactact actgcatctt cgcggggctc 1440
 ggtttcctgc ccgagctccc gctgcgggtg ctgcaattcc gccggaggcg cgtcgatcgc 1500
 gccggagccg tgttcgcccg ggtgcgccgg cgcacggagg agctggctgc caccgccccg 1560
 accatgcagg cctacctgcg gcgcctgcac cagggcacgt ag 1602

SEQ ID NO: 24 moltype = AA length = 533
 FEATURE Location/Qualifiers
 source 1..533
 mol_type = protein
 organism = Kutzneria sp. 744

SEQUENCE: 24
 MDDNRIRISIL VLGGGTAGWM SACYLSKALG PGVEVTVLEA PSISRIRVGE ATIPNLHKVF 60
 FDFLGIAEDE WMRECNASYK AAVRFVNWRT PGDQATPRR RPDGRPDHFD HLFGLPEHE 120
 NLPLSQYWAH RRLNGLTDEP FDRSCYVQPE LLDRKLSPLR MDGTKLASYA WHFDADLVAD 180
 FLCRFVAVQKL NVTHVQDVFT HADLDQRGHI TAVNTESGRT LAADLFIDCS GFRSVLMGKV 240
 MQEPFLDMSK HLLNDRAVAL MLPHDDEKVG IEPYTSSLAM RSGWSWKIPL LGRFGSGYVY 300
 SSQFTSQDEA AEELCRMWDV DPAEQTFNNV RFRVGRSRRR WVRNCVAIGV SAMFVEPLES 360
 TGLYFSYASL YQLVKHFPDK RFRPILADRF NREVATMYDD TRDFLQAHFS LSPRDDSEFW 420
 RACKELPFAD GFAEKVEMYR AGLPVLPVLT IDGHYYGNF EAEFRNFWTN SNYYCIFAGL 480
 GFLPEHPLPV LEFRPEAVDR AEPVFAAVRR RTEELVATAP TMQAYLRLRH QGT 533

SEQ ID NO: 25 moltype = DNA length = 1617
 FEATURE Location/Qualifiers
 source 1..1617
 mol_type = genomic DNA
 organism = Pseudomonas fluorescens

SEQUENCE: 25
 atgaacaagc cgatcaagaa tatcgctcgc gtggcgccgg gtactgcggg ctggatggcc 60
 gcctcgctacc tcgtccgggc cctccaacag caggcgaaca ttacgctcat cgaatctgcg 120
 gcgatccctc ggatcgccgt gggcggaagc accatcccaa gtttgacaga ggtgttcttc 180
 gatttcctcg ggataccgga gcgggaatgg atgccccaa tgaacggcgc gttcaaggcc 240
 gcgatcaagt tcgtgaattg gagaaagtct cccgaccctc cgcgcgacga tcacttctac 300

-continued

```

catttggttcg gcaacgtgcc gaactgcgac ggcgtgccgc ttaccacta ctggctgcgc 360
aagcgcgaaac agggcttoca gcagcgcgat gctaccgcga gcccggggca 420
ctcgacggca agctggcacc gtgcctgtcc gacggcacc ccagatgtc ccacgcgtgg 480
cacttcgacg cgcacctggg ggcgcacttc ttgaagcgct gggcgctcga gcgcgggggtg 540
aaccgcgtgg tcgatgaggt ggtggacgtt cgcctgaaca accgcggtta catctccaac 600
ctgctcacc aaggaggggc gacgctggag gcggacctgt tcatcgactg ctccgcatg 660
cggggggtcc tgatcaatca ggcgtgaa gaaaccttca tcgacatgtc cgactacctg 720
ctgtgcgaca gcgcggctgc cagcgccgtg cccaacgacg acgcgcgcga tggggctcag 780
ccgtacacct cctcgatcgc catgaactcg ggatggacct ggaagattcc gatgctgggc 840
cggttcggca gcgcgtacgt ctctcgagc catttcacct cgcgcgacca ggccaccgcc 900
gacttcctca aactctgggg cctctcggc aatcagccgc tcaaccagat caagtcccg 960
gtcggggcga acaagcgggc gtgggtcaac aactgcgtct cgatcgggct gtcgtcgtgc 1020
tttctggagc ccttggaatc gacggggatc tacttcatct acgcggcgct ttaccagctc 1080
gtgaagcaact tcccgcaac cctgctcgac cgcgggtga gcgacgctt caacgcgcag 1140
atcgctccaca tgttcgacga ctgcccggat ttcgtccaag cgcactattt caccacgtcg 1200
cgcgatgaca cgcggttctg gctcgcgaac cggcacgacc tgcggctctc ggacgccatc 1260
aaagagaagg ttacgcgcta caaggcggg ctgcccgtga ccaccacgtc gttcgacgat 1320
tccacgtact acgagacctt cgactacgaa ttcaagaatt tctggttgaa cggcaactac 1380
tactgcatct ttgcccgtt gggcatgctg cccgaccggt cgtgcgcgt gttgcagcac 1440
cgaccggagt gatcgagaa agccgaggcg atgttcgcca gcaccggcg cgaggccgag 1500
cgtctgcgca ccagcctgcc gacaaactac gactacctgc ggtcgtgcg tgacggcgac 1560
gcggggctgt cgcgcggcca gcgtgggccc aagctcgag cgaggaaa cctgtag 1617

```

```

SEQ ID NO: 26      moltype = AA length = 538
FEATURE
source            Location/Qualifiers
                  1..538
                  mol_type = protein
                  organism = Pseudomonas fluorescens

```

```

SEQUENCE: 26
MNKPIKNIVI VGGGTAGWMA ASYLVRALQQ QANITLIESA AIPRIGVGEA TIPS LQKVFF 60
DPLGIPEREW MPQVNGAFKA AIKFVNWRKS PDPSRDDHFF HLFGNVPNCD GVLPTHWLR 120
KREQGFQQPM EYACYPOPGA LDGKLAPCLS DGTRQMSHAW HFD AHLVADF LKRWAVERGV 180
NRVDEVVDV RLNNRGYISN LLTKEGRTLE ADLFIDCSGM RGLLINQALK EPFIDMSDYL 240
LCDSAVASAV PND DARGVE PYTSSIAMNS GWTWKIPMLG RFGSGYVFS HFTSRDQATA 300
DPLKLWGLSD NQPLNQIKFR VGRNKRAWVN NCVSIGLSSC FLEPLESTGI YFIYAALYQL 360
VKHFPDTSFD PRLSDAFNAE IVHMFDDCRD FVQAHYFTTS RDTTPFWLAN RHD LRLSDAI 420
KEKVQRYKAG LPLTTTSFDD STYYETFDYE FKNFWLNGNY YCIFAGLGML PDRSLPLLQH 480
RPESIEKAEAF MFASIRREAE RLRTSLPTNY DYLRSLRDGD AGLSRGQRGP KLAAQESL 538

```

```

SEQ ID NO: 27      moltype = DNA length = 1593
FEATURE
source            Location/Qualifiers
                  1..1593
                  mol_type = genomic DNA
                  organism = Lentzea areocolonigenes

```

```

SEQUENCE: 27
atgtccggca agattgacaa gatcctcatc gtcggcggcg gcaccgcggc atggatggcc 60
gcgtcctatc tcggcaaggc cctgcagggc accgcggaca tcacactgct gcaggcacc 120
gacatcccg a cgtcggggg cggcgaggcc acgatcccca atctgcagac ggcgttcttc 180
gacttcctc gaatcccgga ggaagagtgg atgcgggagt gcaacgcgag ctacaaggtc 240
gccatcaagt tcatcaactg gcgcaccgcg ggcgagggga cgtccgaggg ccgcgagctc 300
gaecgagggc ccgaccactt ctaccactcc ttcggtctgc tcaagtacca cgagcagatt 360
ccgctgtcgc actactgggt gcaccgttcg taccggggga agaccgtcga gccgttcgac 420
tacgctgtct acaaggaaac cgtcctctc gacgccaaca ggtcaccgcg caggctcgac 480
ggttccaagg tgacgaacta cgcgtggcac ttcgacgcgc acctcgtcgc cgacttctc 540
cgccgggttc ccaccagaaa gctcggcggt cgcacgtcgc aggaaccgct cgagcagctc 600
cagcgcgacg ccaacggcaa catcgagtcg gttcgcacgg caacggggcg tgtcttcgat 660
gccgacctct tcgctgactg ctcgggcttc cgcggggtgc tgatcaacaa ggcgatggag 720
gagcccttcc tcgacatgag cgatcacctg ctcaacgaca gcgcgctcgc caccagggtg 780
ccgcacgacg acgacgcgaa cgtgtgtgaa ccgttcacct cggcgatcgc catgaagtcg 840
ggctggagct ggaagatccc gatgctcggc aggttcggca ccgggtacgt ctactcgagc 900
cggttcggca ccgaggacga ggcggtgcgc gatgtctgcg agatgtggca cctcgaccgc 960
gagacccagc cctccaacag gatccgggtc cgggtcggcc gcaaccggcg cgcgtgggtc 1020
ggcaactcgc tcagcatcgg cacgtcgtcg tgcttcgtgg aacctgga gtcgacgggc 1080
atctacttcg tctacgcgc gctgtaccag ctggtgaagc acttcccga caagagcttc 1140
aaccgccgtg tgaccgccag gttcaaccgc gagatcgaga cgatgttcga cgacacgcgc 1200
gacttcctcc aggcgcactt ctactctcgc ccgcgcacgg acaccccggt ctggaggggc 1260
aacaaggagc tgcgcctggc ggacggcatg caggagaaga tcgacatgta ccgcgcgggc 1320
atggcgatca acgcgcccg gtcgacgac gccagctct actacggcaa cttcgaggag 1380
gagttccgca actctggaa caacagcaac tactactgcg tgctggccgg cctcggtctg 1440
gtgcccgacg caccctcacc acgcctggcg cacatgccac aggcgacgga gtcggtggac 1500
gaggtcttcg gcgcgccgca ggaccggcag cggaacctgc tcgagacctt gccgagcctc 1560
cacgagttcc tgaggcaaca gcacggccgc tga 1593

```

```

SEQ ID NO: 28      moltype = AA length = 530
FEATURE
source            Location/Qualifiers
                  1..530
                  mol_type = protein
                  organism = Lentzea areocolonigenes

```

SEQUENCE: 28

-continued

```

MSGKIDKILI VGGGTAGWMA ASYLGKALQG TADITLLQAP DIPTLGVGEA TIPNLQTAF 60
DPLGIPEDW MRECNASYKV AIKFINWRTA GEGTSEAREL DGGPDHFIYS FGLLKYHEQI 120
PLSHYWFDRL YRGKTVEPFD YACYKEPVIL DANRSPRRLD GSKVTNYAWH FDAHLVADFL 180
RRFATEKLG V RHVEDRVEHV QRDANGNIES VRTATGRVFD ADLFVDCSGF RGLLINKAME 240
EPFLDMSDHL LNDASAVATQV PHDDDANGVE PFTSAIAMKS GWTWKIPMLG RFGTGYVYSS 300
RFATEDEAVR EFCEMWHLDP ETQPLNRIRF RVGRNRRRAW GNCVSIPTSS CFVEPLESTG 360
IYFVYAALYQ LVKHFPDKSL NPVLTARFNR EIETMFDDTR DFIQAHFYFS PRTDTPFWRA 420
NKLRLADGM QEKIDMYRAG MAINAPASDD AQLYYGNFEE EFRNFWNNNS YYCVLAGLGL 480
VPDAPSPRLA HMPQATESVD EVFGAVKDRQ RNLETLPSL HEFLRQQHGR 530

```

```

SEQ ID NO: 29      moltype = AA  length = 516
FEATURE           Location/Qualifiers
source            1..516
                  mol_type = protein
                  organism = Streptomyces albus

```

```

SEQUENCE: 29
MLNNVVI VGG GTAGWMTASY FKAAPGERIN ITLVESGNIG AVGVGEATFS DIRHFFDFLG 60
LKEKDWMPC NATYKLAVRF ENWREEGHYF YHPFEQMRVS NGFPLADWWL REKPTDRFDK 120
DCFMPSIID AGLSPRLDGL SLIDQPFVEG AEEMOGLTMS EHQKTKQFPY AYQFEAALLA 180
KYLTTYSVER GVKHIVDDVQ QVNLDERGWI RNVTTAEHGE IGGDLFIDCT GFRGLLLNKA 240
LEEFISFQD TLPNDASVAL QVPMDMERRG ILPCTTATAQ DAGWIWITPL MGRVGTGYVY 300
AKDYISPEEA ERTLREFVGP AADAEANHI KMRVGRSENS WVKNCVAIGL SSGFVEPLES 360
TGIFPIHHAI EQLAKNFPGE DWNPVQRNLY NDSVAHVMDG VREFVLVHYV AAKRSDTQYW 420
RDTKTRKIPD ALAERIEKWK TQLPDTETVF PYYHGLPPYS YHCVLLGMGG IDVKPSPALA 480
LGDSAAAERE PAEIREKTRH LTSVLPKAYD YFSRMR 516

```

```

SEQ ID NO: 30      moltype = AA  length = 530
FEATURE           Location/Qualifiers
source            1..530
                  mol_type = protein
                  organism = Lentzea aerocologinenes

```

```

SEQUENCE: 30
MSGKIDKILI VGGGTAGWMA ASYLGKALQG TADITLLQAP DIPTLGVGEA TIPNLQTAF 60
DPLGIPEDW MRECNASYKV AIKFINWRTA GEGTSEAREL DGGPDHFIYS FGLLKYHEQI 120
PLSHYWFDRL YRGKTVEPFD YACYKEPVIL DANRSPRRLD GSKVTNYAWH FDAHLVADFL 180
RRFATEKLG V RHVEDRVEHV QRDANGNIES VRTATGRVFD ADLFVDCSGF RGLLINKAME 240
EPFLDMSDHL LNDASAVATQV PHDDDANGVE PFTSAIAMKS GWTWKIPMLG RFGTGYVYSS 300
RFATEDEAVR EFCEMWHLDP ETQPLNRIRF RVGRNRRRAW GNCVSIPTSS CFVEPLESTG 360
IYFVYAALYQ LVKHFPDKSL NPVLTARFNR EIETMFDDTR DFIQAHFYFS PRTDTPFWRA 420
NKLRLADGM QEKIDMYRAG MAINAPASDD AQLYWGNFEE EFRNFWNNNS YYCVLAGLGL 480
VPDAPSPRLA HMPQATESVD EVFGAVKDRQ RNLETLPSL HEFLRQQHGR 530

```

```

SEQ ID NO: 31      moltype = AA  length = 530
FEATURE           Location/Qualifiers
source            1..530
                  mol_type = protein
                  organism = Lentzea aerocologinenes

```

```

SEQUENCE: 31
MSGKIDKILI VGGGTAGWMA ASYLGKALQG TADITLLQAP DIPTLGVGEA TIPNLQTAF 60
DPLGIPEDW MRECNASYKV AIKFINWRTA GEGTSEAREL DGGPDHFIYS FGLLKYHEQI 120
PLSHYWFDRL YRGKTVEPFD YACYKEPVIL DANRSPRRLD GSKVTNYAWH FDAHLVADFL 180
RRFATEKLG V RHVEDRVEHV QRDANGNIES VRTATGRVFD ADLFVDCSGF RGLLINKAME 240
EPFLDMSDHL LNDASAVATQV PHDDDANGVE PFTSAIAMKS GWTWKIPMLG RFGTGYVYSS 300
RFATEDEAVR EFCEMWHLDP ETQPLNRIRF RVGRNRRRAW GNCVSIPTSS CFVEPLESTG 360
IYFVYAALYQ LVKHFPDKSL NPVLTARFNR EIETMFDDTR DFIQAHFYFS PRTDTPFWRA 420
NKLRLADGM QEKIDMYRAG MAINAPASDD AQLYWGNFEE EFRNFWNNNS YYCVLAGLGL 480
VPDAPSPRLA HMPQATESVD EVFGAVKDRQ RNLETLPSL HEFLRQQHGR 530

```

```

SEQ ID NO: 32      moltype = AA  length = 530
FEATURE           Location/Qualifiers
source            1..530
                  mol_type = protein
                  organism = Lentzea aerocologinenes

```

```

SEQUENCE: 32
MSGKIDKILI VGGGTAGWMA ASYLGKALQG TADITLLQAP DIPTLGVGEA TMPNLQTAF 60
DPLGIPEDW MRECNASYKV AIKFINWRTA GEGTSEAREL DGGPDHFIYP FGLLKYHEQI 120
PLSHYWFDRL YRGKTVEPFD YACYKEPVIL DANRSPRRLD GSKVTSYAWH FDAHLVADFL 180
RRFATEKLG V RHVEDRVEHV QRDANGNIES VRTATGRVFD ADLFVDCSGF RGLLINKAME 240
EPFLDMSDHL LNDASAVATQV PHDDDANGVE PFTSAIAMKS GWTWKIPMLG RFGTGYVYSS 300
RFATEDEAVR EFCEMWHLDP ETQPLNRIRF RVGRNRRRAW GNCVSIPTSS CFVEPLESTG 360
IYFVYAALYQ LVKHFPDKSL NPVLTARFNR EIETMFDDTR DFIQAHFYFS PRTDTPFWRA 420
NKLRLADGM QEKIDMYRAG MAINAPAPDD AQLYWGNFEE EFRNLWNNSS YYCVLAGLGL 480
VPDAPSPRLA HMPRATESVD EVFGAVKQQQ RNLETLPSL HEFLRQQHGR 530

```

```

SEQ ID NO: 33      moltype = AA  length = 530
FEATURE           Location/Qualifiers
source            1..530
                  mol_type = protein
                  organism = Lentzea aerocologinenes

```

-continued

```

SEQUENCE: 33
MSGKIDKILI VGGGTAGWMA ASYLGKALQG TADITLLQAP DIPTLGVGEA THPNLQTAF 60
DPLGIPEDW MRECNASYKV AIKFINWRTA GEGTSEAREL DGGPDHXYHS FGLLKYHEQI 120
PLSHYWFDRS YRGKTVEPFD YACYKEPVIL DANRSPRRLD GSKVTNYAWH FDAHLVADFL 180
RRFATEKLG V RHVEDRVEHV QRDANGNIES VRTATGRVFD ADLFVDCSGF RGLLINKAME 240
EPFLDMSDHL LNDASAVATQV PHDDDANGVE PFTSAIAMKS GWTWKIPMLG RFGTGYVYSS 300
RFATEDEAVR EFCMWHLDP ETQPLNRIRF RVGRNRRRAWV GNCVSIPTSS CFVEPLESTG 360
IYFVYAALYQ LVKHFPDKSF NPVLTARFNR EIETMFDDTR DFIQAHFYFS PRDTPFWRA 420
NKLRLADGM QEKIDMYRAG MAINAPASDD AQLYYGNFEE EFRNCWNNSS YYCVLAGLGL 480
VPDAPSPRLA HMPRATESVD EVFGAVKQQ RNLLETPLSL HEFLRQQHGR 530

```

```

SEQ ID NO: 34      moltype = AA length = 530
FEATURE           Location/Qualifiers
source            1..530
                  mol_type = protein
                  organism = Lentzea aerocologinenes

```

```

SEQUENCE: 34
MSGKIDKILI VGGGTAGWMA ASYLGKALQG TADITLLQAP DIPTLGVGEA TIPNLQTAF 60
DPLGIPEDW MRECNASYKV AIKFINWRTA GEGTSEAREL DGGPDHXYHS FGLLKYHEQI 120
PLSHYWFDRS YRGKTVEPFD YACYKEPVIL DANRSPRRLD GSKVTNYAWH FDAHLVADFL 180
RRFATEKLG V RHVEDRVEHV QRDANGNIES VRTATGRVFD ADLFVDCSGF RGLLINKAME 240
EPFLDMSDHL LNDASAVATQV PHDDDANGVE PFTSAIAMKS GWTWKIPMLG RFGTGYVYSS 300
RFATEDEAVR EFCMWHLDP ETQPLNRIRF RVGRNRRRAWV GNCVSIPTSS CFVEPLESTG 360
IYFVYAALYQ LVKHFPDKSL NPVLTARFNR EIETMFDDTR DFIQAHFYFS PRDTPFWRA 420
NKLRLADGM QEKIDMYRAG MAINAPASDD AQLYYGNFEE EFRNFWNNSS YYCVLAGLGL 480
VPDAPSPRLA HMPQATESVD EVFGAVKDRQ RNLLETPLSL HEFLRQQHGR 530

```

```

SEQ ID NO: 35      moltype = AA length = 530
FEATURE           Location/Qualifiers
source            1..530
                  mol_type = protein
                  organism = Lentzea aerocologinenes

```

```

SEQUENCE: 35
MSGKIDKILI VGGGTAGWMA ASYLGKALQG TADITLLQAP DIPTLGVGEA TIPNLQTAF 60
DPLGIPEDW MRECNASYKV AIKFINWRTA GEGTSEAREL DGGPDHXYHS FGLLKYHEQI 120
PLSHYWFDRS YRGKTVEPFD YACYKEPVIL DANRSPRRLD GSKVTNYAWH FDAHLVADFL 180
RRFATEKLG V RHVEDRVEHV QRDANGNIES VRTATGRVFD ADLFVDCSGF RGLLINKAME 240
EPFLDMSDHL LNDASAVATQV PHDDDANGVE PFTSAIAMKS GWTWKIPMLG RFGTGYVYSS 300
RFATEDEAVR EFCMWHLDP ETQPLNRIRF RVGRNRRRAWV GNCVSIPTSS CFVEPLESTG 360
IYFVYAALYQ LVKHFPDKSL NPVLTARFNR EIETMFDDTR DFIQAHFYFS PRDTPFWRA 420
NKLRLADGM QEKIDMYRAG MVINAPASDD AQLYYGNFEE EFRNFWNNSS YYCVLAGLGL 480
VPDAPSPRLA HMPQATESVD EVFGAVKDRQ RNLLETPLSL HEFLRQQHGR 530

```

```

SEQ ID NO: 36      moltype = DNA length = 1446
FEATURE           Location/Qualifiers
source            1..1446
                  mol_type = genomic DNA
                  organism = Bacillus atrophaeus

```

```

SEQUENCE: 36
atgatgtctg aaaatttgca attgtcagct gaagaaatga gacaattggg ttaccaagca 60
gttgatttga tcatcgatca catgaacccat ttgaagtcta agccagtttc agaaacaatc 120
gattctgata tcttgagaaa taagttgact gaatctatcc cagaaaatgg ttcagatcca 180
aagggaattgt tgcatttctt gaacagaaac gtttttaatc aaattacaca tgttgatcat 240
ccacattctt tggcttttgt tccaggtcca aataaattacg ttggtgttgt tgcagatttc 300
ttggcttctg gttttaatgt ttttccaaact gcatggattg ctggtgcagg tgctgaacaa 360
atcgaattga ctacaattaa ttgggtgaaa tctatgttgg gttttccaga ttcagctgaa 420
gggtttattt tttctggtgg ttcaatggca aatttgacag ctttgactgt tgcaagacag 480
gctaagttga acaacgatat cgaaaatgct gttgtttact tctctgatca aacacatttc 540
tcagttgata gagcattgaa ggttttaggt tttaaacatc atcaaatctg tagaatcgaa 600
acagatgaac atttgagaat ctctgtttca gctttgaaga acaaaattaa agaagataga 660
actaagggtga aaaagccatt ctgtgttatt gcaaatgctg gtactacaaa ttgtggtgct 720
gttgattctt tgaacgaatt agcagatttg tgtaacgatg aagatgtttg gttgcatgct 780
gatggttctt atggtgctcc agctatcttg tctgaaaagg gttcagctat gttgcaaggt 840
attcatagag cagattcttt gacttttagat ccacataagt ggtgtgtcca accatcagat 900
gttggttgtg ttttgatcag aaactctcaa tatttgtcaa agacttttag aatgatgcca 960
gaatacatca aggattcaga aactaacgtt gaaggtgaaa ttaatttcgg tgaatgtggt 1020
atcgaattgt caagaagatt cagagctttg aaggtttggt tgtcttttaa agttttcggg 1080
gttgctgctt tttagacaagc aatcgatcat ggtatcatgt tagcagaaca agttgaagca 1140
tttttgggta aagcaaaaga ttgggaagtt gttacaccag ctcaattggg tatcggtact 1200
tttagataca ttccatctga attggcatca acagatacta ttaatgaaat taataagaaa 1260
ttggttaagg aaatcacaca tagaggtttc gctatgttat ctactacaga attgaagcaa 1320
aaggttgtta ttagattgtg ttcaattaat ccaagaacta caactgaaga aatgttgcaa 1380
atcatgatga agattaaagc attggctgaa gaagtttcta tttcataccc atgtgttgct 1440
gaataaa 1446

```

```

SEQ ID NO: 37      moltype = AA length = 481
FEATURE           Location/Qualifiers
source            1..481
                  mol_type = protein

```

-continued

```

organism = Bacillus atrophaeus

SEQUENCE: 37
MMSENLQLSA EEMRQLGYQA VDLIIDHMH LKSKPVSSETI DSDILRNKLT ESIPENGSDP 60
KELLHFLNRN VFNQITHVDH PHFLAFVPGP NNYVGVVADF LASGFNVFPT AWIAGAGAEQ 120
IELTTINWLK SMLGFPDSAE GLFVSGGSMA NLALTVARQ AKLNNDIENA VVYFSDQTHF 180
SVDRALKVLG FKHQICRIE TDEHLRISVS ALKKQIKEDR TKGKKPFCVI ANAGTTNCGA 240
VDSLNELADL CNDEDVWLHA DGSYGAPAIL SEKGSAMLQG IHRADSLTLD PHKWLFPQYD 300
VGCVLIRNSQ YLSKTFRMP EYIKDSETNV EGEINFGECE IELSRFRFRAL KWLSPKVPFG 360
VAAFRAIDH GIMLAEQVEA FLGKAKDWEV VTPAQLGIPT FRYIPSELAS TDTINEINKK 420
LVKEITHRGF AMLSTTELKE KVVIRLCSIN PRTTTEMLQ IMMUKALAE EVSISYPCVA 480
E 481

```

```

SEQ ID NO: 38      moltype = DNA length = 1473
FEATURE           Location/Qualifiers
source            1..1473
                  mol_type = genomic DNA
                  organism = Ruminococcus gnavus

```

```

SEQUENCE: 38
atgtctcaag tcatcaaaaa aaagagaaac actttcatga tcggtaccga atatatattg 60
aactctaccc aattagaaga ggctattaaa tcttctgtcc atgatttctg tgccgaaaag 120
cacgaaattc acgaccaacc agttgtcggt gaagctaagg aacaccaaga agataagatt 180
aagcaaatca aaattccaga aaagggtcgt ccagtcgaatg aagttgtttc tgaaatgatg 240
aacgaagtct acagatacag aggtgatgcc aacctccaa gattcttctc ttctgttcca 300
gggtccagctt cctctgtctc ctggttggtg gacatcatga cttctgttta caacatccac 360
gctggtggtt ctaaatgggc cccaattggt aactgtatcg aacaagaagt cttgaaatgg 420
ctcgcaaaagc aagttgtttt caccgaaaaa cctggtggtg tttctgtttc cggtggtttc 480
atggctaaca ttactgtctt gacagccgct agagacaaca agcttaccga tattaacttg 540
cacttgggta ctgcttacct ttccgaccaa accactctt cagttgctaa gggtttacct 600
atcattggta tcactgactc cagaatcaga agaattccaa ctaactccca cttccaaatg 660
gacaccacta agttggaaga agctatcgaa accgacaaga agtccggtta catccattt 720
gttgttatcg gtagtgccgg taccaccaac actggttcca ttagccatt gactgaaatt 780
tccgctttgt gtaagaagca tgacatgtgg ttccacattg atggtgccta cggtgcttca 840
gttttgtgtg ctccaaagta caagtctttg ttaacaggtg ccggtttggc tgactctatc 900
tcttgggatg ctcaagaagt gttatttcaa acttacggtt gtgctatggt tttggtcaag 960
gatattagaa acttgtttca ttctctccac gtcaatccag aatacttga ggatctgga 1020
aacgacatcg ataactgtta cactctggac atcggtcagg aattgaccag accagccaga 1080
ggtttgaagt ttggttgtag ttgcaaggt ctaggttctg atcttattgg ttctgcaatt 1140
gaacacgggt tccaattggc tgtctgggct gaagaagcct tgaatccaaa gaaggactgg 1200
gaaattgtct ctctgtctca aatggctatg atcaacttca gatagctcc aaaggacttg 1260
aacaaggaag aacaagatgt ctggaacgaa aagatctctc accgtatttt ggaatccggt 1320
tacgctgcta tcttacttac cgtctaaaac ggtaagactg tcttgagaat ctgcgctatt 1380
caccagaaag ctactcaaga agacatgcaa cacaccattg atttggtgga ccaatacgg 1440
agagaaatct aactgaaat gaagaaggct tag 1473

```

```

SEQ ID NO: 39      moltype = AA length = 490
FEATURE           Location/Qualifiers
source            1..490
                  mol_type = protein
                  organism = Ruminococcus gnavus

```

```

SEQUENCE: 39
MSQVIKKRN TFMIGTEYIL NSTQLEAAIK SFVHDFCAEK HEIHDQPVV EAKEHQEDKI 60
KQIKIPEKGR PVNEVVSEMM NEVYRYRGDA NHPRFFSFVP GPASSVSWLG DIMTSAYNIH 120
AGGSKLAPMV NCIEQEVLEK LAKQVGFTEP PGGVVFVSGS MANITALTAA RDNKLTIDINL 180
HLGTAYISDQ THSVVAKGLR IIGITDSRIR RIPTNSHFQM DTTKLEAAIE TDKKSGYIPF 240
VVIGTAGTIN TGSIDPLTEI SALCKKHDMM RHIDGAYGAS VLLSPKYKSL LTGTGLADSI 300
SWDAHKWLFQ TYGCAMVLVK DIRNLFHSPH VNPEYLDKLE NDIDNVNTWD IGMELTRPAR 360
GLKLWLTLQV LGSDILGSAI EHGFLAVWA EEALNPKKDW EIVSPAQMAM INFYAPKDL 420
TKEEQDILNE KISHRILESG YAAIFTTVLN GKTVLRCIAI HPEATQEDMQ HTIDLLDQYG 480
REIYTEMKKA 490

```

```

SEQ ID NO: 40      moltype = DNA length = 1482
FEATURE           Location/Qualifiers
source            1..1482
                  mol_type = genomic DNA
                  organism = Streptomyces griseofuscus

```

```

SEQUENCE: 40
atgaagccag ctgaagctaa accacctcac atggaccatg atactttcag atccttgggt 60
caccagcta ttgactggat tgctgattac tggcaacact tggctgacag accagttgct 120
ccattggttg aaccaggtag tgttagagcc caattgccta cactccacc agagtacggt 180
gaagacttcc cagctttggt atctgacttg gaaagaattg tcttgccagg cttgttgac 240
tggtcaacacc caagattctt cggttacttc ccagctaacc cttccggtcc agcogtttta 300
gctgaattat tatctgctgg tctaggtgtc caaggtatga actggaacac ttctccagct 360
tgtaccgaag ttgaacaaca aatgttggtg tgggtcgttc atttgtagg gttgcccaga 420
cacttcagag tgggtggtgt tatccaagac actgcctcct ccgctgtttt ggtcgctcta 480
ttgactgctt tgcacagagc ttctgctggt agaaccagag atcaccgtac tgggtgaatg 540
ggttacaggg tttacctaac cgtgaaact cactccgctg ctagaaaggc tgcgctcatc 600
accggtttgg gtttgctgct tgtttgtgaa gtcgcaactg atgtgacgg tgctatggat 660
gcccgtgact tggcgcctcg tattcaagcc gatttggtg ccggtttgac cccattgatg 720
gttgttgcta cccgtggtac cacctctgat ttgtccttcg atccattgga agacatcggt 780

```

-continued

```

ccagttctgta gaagacatgg tgtctgggta cacgttgacg ctgcttatgc tgggtgttgc 840
gcagtggtcg atgaattgca ttgggttaat gatggtgta gatacgctga ttcttactgt 900
actaaccac acaagtgggt gttgaccaac ttogactgtg atttattatg ggttgccgac 960
ccagaagctt tgggtcaacgc tttgtccgtt ttgctgctt acttgagaaa cccagcttct 1020
gaatcaggtg gagttactga ctacagacac tggcaagttc cattgggcag aagattcaga 1080
gccttgaagt tgtggtctgt cttaaagatg tacggtgccc aaggtttgag atctcacatc 1140
agaactgggt tcagacacgc cagattgttc gctgacctgg tcagagaaga tgacctttc 1200
accttgttgg cgccagctac tttgggtttg gtcactttcc gtcacgctgg ttctgacacc 1260
gacaacagaa acttgttgca cgctatcaac gccgaaggta caactttttt aactcatagt 1320
gaaaagaacg gtaccttttt tttagagatt gctaccgggt gtaccacac tgaagaccat 1380
cacgtccacg aagcttggag agctgtccaa aatgccgtcc aaaacgccc cccaagagcc 1440
caacatgcc ccatggttac tgcgatgct ttgaacggtt aa 1482

```

```

SEQ ID NO: 41      moltype = AA  length = 493
FEATURE           Location/Qualifiers
source            1..493
                  mol_type = protein
                  organism = Streptomyces griseofuscus

```

```

SEQUENCE: 41
MKPAEAKPPH MDHDTFRSLG HQAIDWIADY WQHLADRPVA PLVEPGTVRA QLPTTPPEYG 60
EDFPALLSDL ERIVLPGLLH WQHPRFPGYF PANASGPVL AELLSAGLV QGMNWNTPA 120
CTEVEQQMLD WFWHLLGLPE HFRGGGVIQD TASSAVLVAL LTALHRASAG RTRDHTGEC 180
GYRVYLTAET HSAARKAAVI TGLGLRAVCE VATDADGAMD AADLAARIQA DLAAGLTPLM 240
VVATRGTTSD LSFDPLEDIG PVCRRHGVWL HVDAAYAGVA AVCDLHWVN DGVRYADSYC 300
TNPHKWLTLN PDCDLLNVAD PEALVNALSV LPAYLRNPAS ESGRVTYRHR WQVPLGRRFR 360
ALKLWSVLRW YGAEGLRSHI RTGVRHARLF ADLVREDDRF TLLAPATLGL VTFRHAGSDT 420
DNRNLLHAIN AEGTTFLLTHS EKNGTFFLRF ATGGTHTEDH HVHEAWRAVQ NAVQNAVPR 480
QHAATDGTADA LNG 493

```

```

SEQ ID NO: 42      moltype = DNA  length = 1482
FEATURE           Location/Qualifiers
source            1..1482
                  mol_type = genomic DNA
                  organism = Ceriporiopsis subvermispora

```

```

SEQUENCE: 42
atggatatgg aagcttttag aaaagcaggt taccaagcta tcgatagaat ttgtgactac 60
tactactctt tgcaaaatag accagttgtt ccatcagttc aaccagggtta ttgttagat 120
gcattaccag attctccacc agaacaaggt gaagatttca ctgttattgc tgatgattac 180
caaaagtaca tcttgccagg ttgacacatc tggcaacatc catctttctt tgcatacttt 240
ccaacagctt gtacttttga aggtattttg ggtgacttgt attctacatc aactgcaaat 300
ccaggtttta attggttggc ttctccagca tgtactgaat tagaaatggt tgttatggat 360
tggctcgtca agttgttagg ttgtgcagaa cattcttgc attctcagg taaaggtggt 420
ggtgtttatt aaactacagc atcagaattg gctttagtgt ttgttgttgc tgcaagagaa 480
agatacttga gaattcatcc agatgctaag gcagatgaat tggttatata tactacaact 540
caaacacatt ctttgggtgt taaagctggt ttagtttttg gtatggaatg tagagcattg 600
gaagttaaag ctgaagatgc ttacgcatta agaggtgcaa ctttgaagtc agctttagaa 660
gaagatgaaa agagaggttaa aagacctttt attttggttg caacagttgg tacaacttct 720
tcaggtgcta ttgatagatt ggaatgaatt ggtcaagttt ctgaagatta tccatcatta 780
tggattcatg ttgatgtctg atgggcaggt gttacttttg cttgtccaga atacagaggt 840
acagctcaat tagaaaacat caacgcttat gcaacatctt tcggtactaa cttccataag 900
tggggtttgg ttaactttga tctgcattg ttgtgggtta aggatagaaa ggatttgaca 960
gatgctttag atgttactcc agaattcttg agaacaaaac aaggtgacgc tgggtcagtt 1020
gttgatttta gaaattggca tttgggtttg ggtagaagat tcagatcttt gaaggttttg 1080
ttcgttttga gatcacacg tgttgaaggt tttagaacct acatcagaca aggtattaaa 1140
ttgaatgaac attttacttc ttttaattga gcttcttttg atttttcatt agttacagca 1200
ccatcattcg ctttgacagt ttttagattg actccagctg gtgcattctt gactggttca 1260
gaattgaacg aattaaacag agctttctac gcaagattat cttcaagaca tgatcatcg 1320
ttgacacaaa ctgtttttga ccggtttttc tgtatcagat ttgacagttg tgctgcaaga 1380
actcaacaag aacatattga tacagcttgg gatttgttac aacaagaagg tgctgttgca 1440
gttcaagaat acatgcaaaa gatttctaag gatgttcaat aa 1482

```

```

SEQ ID NO: 43      moltype = AA  length = 493
FEATURE           Location/Qualifiers
source            1..493
                  mol_type = protein
                  organism = Ceriporiopsis subvermispora

```

```

SEQUENCE: 43
MDIEAFRKAG YQAIIDRICDY YYSQNRPVV PSVQPGYLLD ALPDSPEQGG EDFTVIADDDY 60
QKYILPGLTH WQHPSPFFAYF PTACTFEGIL GDLYSTSTAN PGFNWLASPA CTELEMVMD 120
WSAKLLGLSE HFLHSSGKGG GVIQTASEL ALVVVVAARE RYLRHPDAK ADELVIYTTT 180
QTHSLGVKAG LVFGMECRAL EVKAEDAYAL RGATLKSAL EDEKRGKRP ILVATVGTT 240
SGAIDRLDEI QOVSEYPSL WIHVDAAWAG VTLACPEYRG TAQLENINAY ATSFGTNPHK 300
WGLVNFDAAL LWVKDRDLT DALDVTPEFL RTKQGDAGAV VDFRNWHLGL GRRFRSLKVV 360
FVLRSYGVEG FRNVIROGKI LNEHFTSLIR ASLDFSLVTA PSFALTVERL TPAGASLTGS 420
ELNELNRAFY ARLSSRDHIM LTQTVLNGVF CIRFAVGAAR TQQEHIDTAW DLLQQEGAVA 480
VQEYMQKISK DVQ 493

```

```

SEQ ID NO: 44      moltype = DNA  length = 1254
FEATURE           Location/Qualifiers

```

-continued

```

source          1..1254
                mol_type = genomic DNA
                organism = Clostridium sporogenes

SEQUENCE: 44
atgaagttct ggagaaagta cacacaacaa gaaatggatg aaaagattac tgaatctttg 60
gaaaagacttt tgaactacga taacactaag acaatcggta ttccaggtag taagttggat 120
gatacagttt tctatgatga tcattctttc gttaaagcatt caccatactt gagaactttt 180
attcaaaacc caaacccatat cggttgtcat acttatgata aggcgtgatat cttgttcggt 240
gggtacattcg atatcgaaag agaattaatc caattgttag caatcgatgt tttgaacggt 300
aacgatgaag aatttgatgg ttacgttact caaggtggta cagaagctaa catccaagca 360
atgtgggttt acagaaaacta cttcaagaaa gaaagaaaagg ctaagcatga agaaatcgct 420
atcatcactt cagcagatac acattactct gcatacaaa gttcagattt gttgaacatc 480
gatattatta aggttccagt tgatttttat tcaagaaaaa ttcaagaaaa tacattggat 540
tcaattgtta aagaagctaa aaaaagtact tcatcgttat ctctaacatg 600
gggtactacaa tgtttggttc agttgatgat ccagatttgc acgctaacat cttcgataag 660
tacaatttgg aatacaaaat tcattgtgat ggtgcatttg gtggttttat atatccaatt 720
gataataagg aatgtaaaac tgatttctct aataagaacg tttcttcaat cacattagat 780
gggtcataaga tgttgcaagc tccatcgggt actggtatct tcgtttcaag aaagaatttg 840
atccataaca ctttgcaaaa ggaagcaact tacatcgaaa atttggatgt tacattgtct 900
ggttcaagat ctggttcaaa tgctgttgca atttggatgg ttttagcttc ttagtggtcca 960
tacggttgga tggaaaagat taataagttg agaaatagaa ctaaatgggt gtgtaagcaa 1020
ttgaacgata tgagaattaa atattacaaa gaagattcaa tgaatattgt tacaattgaa 1080
gaacaatatg ttaataagga aatcgctgaa aagtactttt tagttccaga agttcataac 1140
ccaactaaca actggtacaa gatcgttgtt atggaacatg ttgaattgga tatcttgaac 1200
tctttggttt acgatttgag aaagtttaat aaggaacatt tgaaggcaat gtaa 1254

SEQ ID NO: 45      moltype = AA length = 417
FEATURE           Location/Qualifiers
source            1..417
                  mol_type = protein
                  organism = Clostridium sporogenes

SEQUENCE: 45
MKFWRKYTQQ EMDEKITESL EKTLYNDNTK TIGIPGTKLD DTVFYDDHSF VKHSPYLRTF 60
IQNPNIHIGCH TYDKADILFG GTFDIERELI QLLAIDVLNG NDEEFDGYVT QGGTEANIQA 120
MWVYRNYFKK ERKAKHEEIA IITSADTHYS AYKGSDDLNI DIKVPVDFY SRKIQENTLD 180
SIVKEAKEIG KKYFIVISNM GTTMFGSVDD PDLYANIFDK YNLEYKIHVD GAFGGFIYPI 240
DNKECKTDFS NKNVSSITLD GHKMLQAPYG TGIFVSRKNL IHNTLTKEAT YIENLDVTL 300
GSRSGSNAVA IWMVLASYPG YGWMEKINKL RNRTKWLCQ LNDMRIKYYK EDSMNIVTIE 360
EQYVNKEIAE KYFLVPEVHN PTNNWYKIVV MEHVELDILN SLVYDLRKFN KEHLKAM 417

SEQ ID NO: 46      moltype = DNA length = 1443
FEATURE           Location/Qualifiers
source            1..1443
                  mol_type = genomic DNA
                  organism = Homo sapiens

SEQUENCE: 46
atgaatgcat ctgaattcag aagaagaggt aaagaaatgg ttgattacat ggctaactac 60
atggaaggta ttgaaggtga acaagtttat ccagatgttg aaccaggtta cttgagacca 120
ttaattccag ctgcagctcc acaagaacca gatacatttg aagatattat taacgatggt 180
gaaaagatta tcatgccagg tggtacacat tggcattcac catatttctt tgcatacttt 240
ccaactgctt cttcatatcc agcaatgttg gctgatatgt tatgttggtc aattggttgt 300
attggttttt cttgggcagc ttaccagctg tgtacagaaat tggaaactgt tatgatggat 360
tggttaggta aaattggtgga attaccaaaag gcatttttga acgaaaaggc tggtagaagg 420
gggtggtgta ttcaagggtc tgcacagaaa gctacattag ttgctttggt agcagctaga 480
actaaaagta ttcatagatt gcaagcagct tctccagaaat taacacaagc agctatcatg 540
gaaaaaattg ttgcataatc ttccagatcaa gctcattctt cagttgaaag agctgggttg 600
atcgggtggt taaagttgaa ggcaatccca tctgatggta attttgctat gagagcatca 660
gctttgcaag aagcattaga aagagataaa gcagctgggt tgattccatt tttcatggtt 720
gctactttag gtactacaac ttgttgttct ttcgataatt tggtagaagt tgggtccaatt 780
tgtaataaag aagatatttg gttgcatgtt gatgcagctt acgcaggttc agcttttatt 840
tgtccagaat tcagacattt gtggaacggt gttgaattcg ctgattcttt taatttcaat 900
ccacataaat ggttggttagt taacttcgat tgttcagcaa tgtgggttaa gaaaagaaca 960
gatttgactg gtgcttttag attagatcca acatatttga agcatttcca tcaagattca 1020
ggtttgatca ctgattacag acattggcaa attccattgg tgaagaagat cagatctttg 1080
aagatgtggt tcgtttttag aatgtacggt gttaaggggt tgcaagctta catcagaaaag 1140
catgttcaat tgtctcatga attcgaatca ttagttagac aagatccaag attcgaatc 1200
tgtgttgaag ttattttggg tttagtttgt ttcagattga agggttctaa taaggttaac 1260
gaagcattgt tgcaaaagaa taattcagct aagaaaaattc atttgggttc atgtcatttg 1320
cgtgataaag tcgttttgag attcgtattt tgttcaagaa ctgttgaatc agcacatggt 1380
caaagagctt ggggaacatat caaggaattg gcagctgatg ttttgagagc tgaagagaaa 1440
taa 1443

SEQ ID NO: 47      moltype = AA length = 480
FEATURE           Location/Qualifiers
source            1..480
                  mol_type = protein
                  organism = Homo sapiens

SEQUENCE: 47
MNASEFRRRG KEMVDYMAN MEGIEGRQVY PDVEPGYLRP LIPAAAPQEP DTFEDIINDV 60

```

-continued

EKIIMPGVTH	WHSPLYFFAYF	PTASSYPAML	ADMLCGAIGC	IGFSWAASPA	CTELETVMMD	120
WLGKMLELPK	AFLNEKAGEG	GGVIQGSASE	ATLVALLAAR	TKVIHRLQAA	SPELTQAAIM	180
EKLVAISSDQ	AHSSVERAGL	IGGVKLKAIP	SDGNFAMRAS	ALQEALERDK	AAGLIPIFFMV	240
ATLGTTTCCS	FDNLLEVGP	CNKEDIWLHV	DAAYAGSAFI	CPEPRHLNNG	VEFADSFNFN	300
PHKWLLVNF	CSANMWVKRT	DLTGAFRLDP	TYLKHSHQDS	GLITDYRHWQ	IPLGRRFRSL	360
KMWFVFRMYG	VKGLQAYIRK	HVQLSHEFES	LVRQDPRFEI	CVEVILGLVC	FRLKGSNKVN	420
EALLQRINSA	KKIHLVPCHL	RDKFVLRF	CSRTVESAHV	QRAWEHIKEL	AADVLAERE	480

SEQ ID NO: 48 moltype = DNA length = 1521
 FEATURE Location/Qualifiers
 source 1..1521
 mol_type = genomic DNA
 organism = *Ophiorrhiza pumila*

SEQUENCE: 48

atgggttcta	tctcagaaaa	ctgtgatgat	tctatttcat	tagctgcacc	ttttagacca	60
ttggaaccag	aagaattcag	aaagcaagct	catgttatgg	ttgatttcat	cgcagattac	120
tacaagaaca	ttgaaaatta	tccagttttg	tcacaagttg	aaccaggtta	tttgaaaaat	180
agattgcagc	aaactgctcc	acatttgcca	gaatctttcg	aaacaatctt	gaaggatatt	240
aagaagata	tcgttctcgg	tatgactaat	tggttgctac	caaatttctt	tgcttatttt	300
ccagcaactg	tttcttcagc	tgcatattgt	ggtgaaatgt	tggtgacagg	ttttaattct	360
gttggtttta	attggtttagc	ttcaccagca	tctacagaat	tggaaatggt	tggtattgat	420
tggttagcta	acatgttgaa	gttgccaaag	tcttttatgt	ttcatggtac	tggtggtggt	480
gttattcaag	gtactacatc	tgaagctatc	ttgtgtacat	tgattgctgc	aagagatggt	540
gctttggaag	agatttggtat	ggaacacgtt	ggtaaatagg	ttgtttacgg	ttcagatcaa	600
actcattctt	tctttcaaaa	gacttgtaag	gttgctggta	tcttcccatg	taacatcaag	660
ttgatcccaa	ctacaagaga	agataaactc	tctatgtcac	caatcgcat	gagagaacaa	720
attgaagctg	atgttgacaga	tggttttagt	ccaattttct	tggtgactac	agttggtact	780
acatcaactg	ctgcaattga	tccagtttct	gaagttgcta	aggttgcaaa	cgatttcaat	840
atttgggttc	atgtttgatgc	tgcttatgct	ggttctgcat	gtatttctgc	agaattcaga	900
caatacttag	atggtatcga	attggttgat	tctttttcat	tgctccaca	taagtgggtg	960
ttgtgtttct	tggattgttg	ttgtttgtgg	ttgaagaaac	cacatttgat	ggttaaggct	1020
ttgtcaacta	accagaata	tttgagaaat	aagagatctg	aattcgatgg	tggtgttgat	1080
ttcaaggatt	ggcaaatcgg	tacaggtaga	agattcaaa	ctttgagatt	atgggtggtt	1140
atgagatcat	acggtgttga	aaatttgaag	agacatatct	tgctgatgt	tcaaatggct	1200
aagatgttcg	aaggttttagt	taaatctgat	ccaagattcg	aatcatcgt	tccaagagct	1260
ttcgcattag	tttgtttcag	attgaaccct	ggtaaaaggt	atgatgatga	aatcgataag	1320
gaaatcttga	ataaggaatt	gttagatttg	attaattcaa	caggtagagc	ttacatgact	1380
catacaaaag	caggtggtat	ctatatgttg	agatttgctg	ttggtactac	attgactgaa	1440
gaacatcatg	tttacgctgc	atgggaattg	attaaagaat	gtacagatgc	atctttgact	1500
aaaacaaata	ttattgaata	a				1521

SEQ ID NO: 49 moltype = AA length = 506
 FEATURE Location/Qualifiers
 source 1..506
 mol_type = protein
 organism = *Ophiorrhiza pumila*

SEQUENCE: 49

MGSISENCDD	SISLAAPFRP	LEPEEFRKQA	HVMVDFIADY	YKNIENYPVL	SQVEPGYLKN	60
RLPETAPHL	ESFETILKDI	KKDIVPGMTN	WLSNPFYF	PATVSSAAV	GEMLCGTGNS	120
VGFNWLSPA	STELEMVVID	WLANMLKLPK	SFPMFHGTGGG	VIQTTTSEAI	LCTLIAARDG	180
ALEKIGMEN	GKLVVYSDQ	THSFFQKTC	VAGIFPCNIK	LIPTTREDNF	SMSPIALREQ	240
IEADVADGL	PIFLCTTVGT	TSTAIDPVS	EVAKVANDFN	IWVHVDAAYA	GSACICEPR	300
QYLDGIELVD	SFSLSPHKWL	LCFLDCCCLW	LKKPHLMVKA	LSTNPEYLRN	KRSEFDGVVD	360
FKDWQIGTR	RFKALRLWL	MRSYGVENLK	RHILSDVQMA	KMFEGVLKSD	PRFEIIVPRA	420
FALVCFRLNP	RKGYDDIDK	EILNKELDL	INSTGRAYMT	HTKAGGIYML	RFVAGTTLTE	480
EHHVYAAWEL	IKECTDASLT	KTNIE				506

SEQ ID NO: 50 moltype = DNA length = 1545
 FEATURE Location/Qualifiers
 source 1..1545
 mol_type = genomic DNA
 organism = *Oryza sativa*

SEQUENCE: 50

atgggttcat	tagatacaaa	tccaactgct	ttttctgcat	ttccagctgg	tgaaggtgaa	60
acttttcaac	cattgaacgc	agatgatgtt	agatcatatt	tgcataaggc	tggtgatctc	120
atctctgatt	actacaatc	agttgaatct	atgccagttt	tgccaaacgt	taagccaggt	180
tatttgcaag	atgaattaa	agcatcacca	ccaacttact	ctgctccatt	cgatgttaca	240
atgaaggaat	tgagatcttc	agttgttctc	ggtagtactc	attgggcatc	tccaaatttc	300
tttgctttct	ttccatcaac	aaattctgct	gcagctattg	ctggtgactt	gattgcatca	360
gctatgaata	cagttgggtt	tacttgccaa	gcttctccag	cagctactga	aatggaagtt	420
ttgcatcatt	attggttgcc	tcaaatgttg	aatttgccaa	cttcttttat	gaacagaact	480
gggtgaagta	gaggtactgg	tggtggtggt	attttgggta	ctacatctga	agctatgttg	540
gttacattag	ttgcagctag	agatgcagct	ttgagaagat	ctggttcaga	tggtgttgca	600
gggttgcata	gattgactgt	ttatgcagct	gatcaaacac	attcaacttt	ctttaaggca	660
tgtagattgg	ctggttttga	tccagcaaat	attagatcta	ttccaacagg	tgtgaaact	720
gattatgggt	tagatccagc	aagattgtta	gaagctatgc	aagcagatgc	tgatgcaggt	780
ttagtcccaa	catagctttg	tgcaactgtt	ggtactacat	cttcaaatgc	tggtgatcca	840
gttggtgctg	ttgcagatgt	tgcaactaga	tttgagctt	gggttcatgt	tgatgcagct	900
tacgcaggtt	cagcttgat	ttgtccagaa	ttcagacatc	atttgatggt	tggtgaaaga	960

-continued

```

gttgattcta tctcaatgtc tccacataag tgggtgatga catgttttaga ttgtacttgt 1020
ttgtacgtta gagatacaca tagattgact gggtcttttag aaactaacc agaatacttg 1080
aaaaatcatg cttcagattc tgggtgaagt acagatttga aagatatgca agttgggtgt 1140
ggtagaagat tcagaggttt gaagtgtgtg atggttatga gaacttatgg tgttgctaag 1200
ttgcaagaac atatcagatc agatgttgct atggcaaaag tttttgaaga tttgggttaga 1260
ggtgacgata gattcgaagt tgttgttcca agaaacttcg ctttggtttg tttcagaatt 1320
agagctgggt caggtgcagc tgcagctaca gaagaagatg cagatgaagc taacagagaa 1380
ttgatggaaa gattgaaata gactggtaaa gcatacgttg ctcatacagt tgttggtggt 1440
agatttggtt tgagatttgc agttggttct tcattacaag aagaacatca tgttagatct 1500
gcttgggaat tgattaagaa aactacaact gaaatgatga actaa 1545

```

```

SEQ ID NO: 51      moltype = AA length = 514
FEATURE           Location/Qualifiers
source            1..514
                  mol_type = protein
                  organism = Oryza sativa

```

```

SEQUENCE: 51
MGSLDTNPTA FSAPPAGEGE TFQPLNADDV RSYLHKAVDF ISDYYKSVES MPVLPNVKPG 60
YLQDELRSAP PTYSAPPDVT MKELRSSVVP GMTHWASPNF FAFPPSTNSA AAIAGDLIAS 120
AMNTVGFTWQ ASPAATEMEV LALDWLAQML NLPTSFMNRT GEGRGTGGGV ILGTTSEAML 180
VTLVAARDAA LRRSGSDGVA GLHRLAVYAA DQTHSTFFKA CRLAGFDPAN IRSIPTGAET 240
DYGLDPARLL EAMQADADAG LVPTYVCATV GTSSNAVDV VGAVADVAA FAAWVHVDAA 300
YAGSACICPE FRHHLDGVER VDSISMSPHK WLMTCLDCTC LYVRDTHRLT GSLETNPEYL 360
KNHASDSGEV TDLKDMQGVG GRRFRGLKLW MVMRTYGVAK LQEHIRSDVA MAKVFEDLVR 420
GDDRFEVVVP RNFALVCFRI RAGAGAAAAT EEDADEANRE LMERLNKTKG AYVAHTVVVG 480
RFVLRFAVGS SLQEEHHVRS AWELIKKTTT EMMN 514

```

```

SEQ ID NO: 52      moltype = DNA length = 1494
FEATURE           Location/Qualifiers
source            1..1494
                  mol_type = genomic DNA
                  organism = Oryza sativa

```

```

SEQUENCE: 52
atggaaggtg ttggtgggtg tgggtgggtg gaagaatggt tgagaccaat ggatgcagaa 60
caattaagag aatgtggtca tagaatggtt gatttcgttg cagattacta caagtcaatc 120
gaagcttttc cagttttgtc tcaagtccaa ccaggttatt tgaaagaagt tttgccagat 180
tcagctccaa gacaaccaga tactttagat tctttgttcg atgatatcca acaaaagatt 240
atcccagggtg ttacacattg gcaatcacca aactacttcg catactaccc atctaattct 300
tctactgctg gtttctggg tgaaatggtg tctgctgctt ttaatatcgt tggtttttca 360
tggatcactt ctcagctgct aacagaattg gaagttattg ttttagattg gttcgcaaa 420
atggtgcaat tgcacatcac attcttatct actgctttgg gtggtgggtg tattcaaggt 480
acagcttcag aagcagtttt ggttgcatgt ttagctgcaa gagatagagc tttgaagaaa 540
catggttaaac attctttaga aaaattgggt gtttacgcat cagatcaaac tcattctgct 600
ttacaaaaag catgtcaaat cgtggttatt ttctctgaaa acgttagagt tgttattgct 660
gattgtaata agaactacgc tgttgacca gaagcagttt cagaagcttt gtcctatcgt 720
ttgtcttcag gtttgatccc atttttoatt tgtgcaacag ttggtactac atcttcattc 780
gctgttgatc cattaccaga attgggtcaa atcgctaagt caaacgatat gtggttccat 840
atcgatgctg cttatgctgg ttctgcatgt attgtgccag aatacacaga tcatttgaat 900
gggtgtgaag aagcagattc tttaatatg aatgctcata agtgggtctt gactaaactc 960
gattgttcat tgtgtgtggg taaggataga tcatttttga tccaatcatt atctacaaa 1020
ccagaattct tgaaaaataa ggcttcacaa gcaaaactctg ttgttgattt caaggattgg 1080
caaatcccat tgggtagaag attcagatca ttgaagtgtg ggatggtttt gagattgtac 1140
gggtgttgata atttgcaatc ttacatcaga aagcatatcc atttgggtga acatttcgaa 1200
caattgttgt tgcagattc aagattcgaa gttgttactc caagaacatt ttcattggtt 1260
tgtttcagat tagttccacc aactctgtat catgaaaacg gttagaaagt gaactacgat 1320
atgatggatg gtgttaattc atctggtaaa attttcttgt cacatacagt tttgtctggt 1380
aaattcggtt tgagatttgc tgttggtgca ccattaactg aagaaagaca tgttgatgct 1440
gcattggaagt tgttgagaga tgaagctaca aaggttttag gtaaaatggt ttaa 1494

```

```

SEQ ID NO: 53      moltype = AA length = 497
FEATURE           Location/Qualifiers
source            1..497
                  mol_type = protein
                  organism = Oryza sativa

```

```

SEQUENCE: 53
MEGVGGGGGG EEWLRPMDAE QLRECGHRMV DFVADYYKSI EAFPVLSQVQ PGYLKEVLDP 60
SAPRQPDITD SLFDIIQQKI IPGVTHWQSP NYFAYYPSNS STAGFLGEML SAAFNIIVGFS 120
WITSPAATEL EVIVLDWFAK MLQLPSQFLS TALGGGVIGQ TASEAVLVAL LAARDRALKK 180
HGKHSLEKLV VYASDQTHSA LQKACQIAGI FSENVRVVIA DCNKNYAVAP EAVSEALSID 240
LSSGLIPFFI CATVGTSSSS AVDPLPELGQ IAKSNDMWFH IDAAAYAGSAC ICPEYRHHLN 300
GVEEADSFNM NAHKWFITNF DCSLLWVKDR SFLIQSLSTN PEFLKNKASQ ANSVVDFKDW 360
QIPLGRRFRS LKLWMLRLRY GVDNLQSYIR KHIHLAEHFE QLLLSDSRFE VVTPRTFSLV 420
CFRLVPPTSD HENGRKLNLD MMDGVNSSGK IFLSHTVLSG KFVLRFAVGA PLTEERHVDA 480
AWKLLRDEAT KVLGKMV 497

```

```

SEQ ID NO: 54      moltype = DNA length = 1944
FEATURE           Location/Qualifiers
source            1..1944
                  mol_type = genomic DNA

```

-continued

organism = *Xenorhabdus doucetiae*

SEQUENCE: 54

atgacaaagc	cattccatcc	aaatttgaac	gttgatgcac	tgtttttagg	tccaaagtct	60
gaaaacgctg	ttttcttttag	agaaatgatg	gattatgcag	ttctgaaca	catgcattgg	120
agatcaggtt	ttcatccaga	agattctaca	ttggttactt	cagttgatag	atacgaacat	180
aactacagag	aaacattgta	cagaactgaa	ggtattttga	atcaattatc	tgctaagttg	240
aaaaattctt	caatcccat	tttctctcca	agatacttag	gtcatatttc	ttcagatact	300
ttgatgggtt	ctaatttggc	ttatgttatg	gctatgatgt	acaacccaaa	caactgttct	360
tatgaagcct	caccaactac	aactgaattg	gaattagaat	ctggtttggg	tttgtgtaga	420
atgttcgggt	acaatccaca	acaactcttg	ggtcatatta	catcaggtgg	tactgttgca	480
aattacgaag	tttgttgggt	tgctagaaat	ttgaagacat	tgtctttcgc	aatctcacia	540
catccaaagg	ctaaggattt	gatcgaacat	aagactccaa	agcaattgaa	aaatatgcc	600
acatctgaat	ctgttagatt	gatcactgaa	ttacaaaaga	gaggtttgtt	cgaagaatc	660
agagatatga	catgtagagg	tactggtgtt	aaggctcaaa	ttttaggtaa	attgttagtt	720
ccacaatcta	agcattatct	atggatgaaa	gctgcagata	tcttcggtat	cggccaagaa	780
aacatctccc	cattgccagt	taatgatcat	taccaaacag	atgttgctca	aatgagaaag	840
atcacttttg	gtttgattga	acaaggtgaa	ccaatttttag	caatgattgc	tgttgttggt	900
acaactgaag	ttggtgcaat	cgatagaatc	gatgaagtta	ttgctttaag	aaaagaatgt	960
gaagaaagat	acggtgaaat	ttctacatc	catgttgatg	ctgcttatgc	tggttacgct	1020
tggtcaatgt	tggtgaacga	acaggttgaa	ttcgttgaat	acgatgaatt	ggcaaacctac	1080
catagatctg	ctggtatttat	gccagaaaat	atctcatggc	caaaaccaga	agtttacc	1140
tcttttaag	ctttgaaaga	ggttgattca	attacagttg	atccacataa	ggttggtttt	1200
attcaatatg	ctgcaggtgc	aatttgtatg	aaggataaga	gaatttttga	attaatttct	1260
tcacatctg	catcagtttt	tgaaggtcat	gctgataaga	catctactaa	aaatagaggt	1320
atcttgggtt	cttcaattat	ggaaggttct	aaatcaggtg	caacagctgc	agctttgttg	1380
gcagctcata	gattgttacc	attgaacatc	aacggttacg	gtaaagttaa	tgcaagctgg	1440
atcgttactg	cacaatcttt	gttgataaag	ttggctaatt	tgccaccaat	cgaatccggt	1500
aaacatagat	tcgaagttca	tatttttcca	tcaccagatt	tcacatgat	caacttcaact	1560
tttaagaag	ttggttaacga	atctttgtca	aaccataacg	ctttgaataa	gagattgtat	1620
gaattatggt	cttactcagc	aggtagagct	tacacaaatg	atctcttgac	atcttcaact	1680
atcttagatt	acaaggaata	cggtagacact	ccttggtcgc	atgctcaaaa	atgttggtttt	1740
tcaagatcag	atggggaaaa	agttagaac	atctatgttt	tgagagcagc	tgttatgaca	1800
tactgtttga	gagatgaaga	acatttcgaa	gaattctggg	aacaattgca	agctatcttc	1860
gttaagaagt	tgactcaaat	ctgttgatgaa	gaagataaga	aagcaagatt	gagattagaa	1920
ccagatgttc	cttttattgg	ttaa				1944

SEQ ID NO: 55 moltype = AA length = 647
 FEATURE Location/Qualifiers
 source 1..647
 mol_type = protein
 organism = *Xenorhabdus doucetiae*

SEQUENCE: 55

MTKPFHPNIN	VDALFLGPKS	ENAVFFREMM	DYAVSEHMH	W RSGFHPEDST	LVTSDVRYEH	60
NYRETLRYTE	GILNQLSAKL	KNSSIPFFSP	RYLGHISSDT	LMVSNLAYVM	AMMYNPNMCS	120
YEASPTTTEL	ELESGLDLCR	MFGYNPQQSW	GHITSGGTVA	NYEGLWVARN	LKTLSPAISQ	180
HPKAKDLIEH	KTPKQLKNMP	TSELLDLITE	LQKRGLFEEI	RDMTCRGTGV	KAQILGKLLV	240
PQSKHYSWMK	AADIFGIGQE	NIIPLPVNDH	YQTDVAQMRK	ITFGLIEQGE	PILAMIAVVG	300
TTEVGAIARI	DEVIALRKEC	EERYGESFYI	HVDAAYAGYA	CSMLLNQGE	FVEYDELANY	360
HRSAGIMPEN	ISWPKPEVYQ	SFKALKDVDS	ITVDPHKVGF	IQYAAGAICM	KDKRILELIS	420
SHAAVVFEGH	ADKTSTKNRG	ILGSSIMEGS	KSGATAAALW	AAHRLPLINI	NGYGKVIAG	480
IVTAQSLNKK	LANLPPIEIG	KHRFEVHIIP	SPDPHMINFT	FKEVGNESLS	NHNALNKRLY	540
ELCSYSAGRA	YTNDFLTSST	ILDYKEYGDT	PWHHAQKCGF	SRSEWEKVRN	IYVLRRAVMT	600
YCLRDEEHFE	EFWEQLQAI	F VKKLTVQVDE	EDKKARLRLE	PDVPFIFG		647

SEQ ID NO: 56 moltype = DNA length = 1509
 FEATURE Location/Qualifiers
 source 1..1509
 mol_type = genomic DNA
 organism = *Camptotheca acuminata*

SEQUENCE: 56

atgggttctt	tgatttcaaa	ctacgatact	gaatctccag	cttcagttgg	tcaattcaat	60
ccattagatc	cagaagaatt	cagaaagcaa	gctcattgta	tcgttgattt	catcgcagat	120
tactacaaga	acattgaaat	ttatccagtt	ttgtcacaa	ttgatccagg	ttacagacat	180
tcaagattgg	gtaaaaatgc	tcacatcaga	tctgaacct	tcgaatcaat	cttgaaggtat	240
gttcaaaagg	atatcatccc	tggtatgact	cattggatgt	ctccaaattt	ctttgtctcat	300
tttcagcaaa	ctgtttcttc	agctgcattt	gttggtgaaa	tggtgtgtac	atgtttcaac	360
tctgttggtt	ttaatttggt	agcatcacca	gctgcaacag	aattggaat	ggttggtatt	420
gattgggttg	ctaacatggt	gaagttgcca	aagtctttta	tggtttcagg	tactggtggt	480
gggtgtttgc	aagggtactac	atctgaagct	atctgtgtga	cattgattgc	tgcatcacca	540
atgcatttcg	aaatcgttgg	tgtaagact	tctacatcat	tcgttggtta	tggttctgat	600
caactcatt	caacatacgc	taaagcatgt	aaattagcag	gtattttgcc	atgtaacatc	660
agatctatct	caactacagc	tgattcaaac	ttctctgttt	caccattggt	gttgagaaga	720
gctattgaag	cagataaagg	tcagggtatg	gttccattgt	atatttgtgc	tacagttggt	780
actacatcta	ctacagacaat	tgatccatta	tcttcattgg	ctgatgttgc	aaatgattat	840
gggtgtttgt	ttcatgttga	tgctgcatac	gctggttctg	catgtatttg	tccagaattc	900
agacattatt	tggtatgtat	cgaagagca	gattctttgt	cattgtctcc	acataagtg	960
ttgttgtctt	acttagattg	ttgtgttttg	tggtgtaagt	caccatcttt	attggttaag	1020
gctttatcaa	ctgatccaga	atattgaaa	aatcaacct	cagaatctaa	gtcagttggt	1080
gattacaagg	attggcaagt	tggtacaggt	agaagattca	aagcttttag	attgtggttc	1140

-continued

```

gttatgagat cttatggtgt tgcacaaattt caatcacata tcagaactga tgttcaaatg 1200
gctaagatgt tcgaaggttt cgttaagtct gatccaagat tcgaaatttt ggttccaaga 1260
gtttttctctt tggtttgttt cagattgaac ccaatctctg gttcagatcc aactggtaca 1320
gaagctttga acagaaagt attgattgg gtttaattcta caggtagagt ttatatgact 1380
catacaaaag ttggtggtat ctatatgttg agatttgctg ttggtgcaac tttaacagaa 1440
aagagacatg tttcttcagc ttggaagttg attaagaag gtgcagatgt tttgttgaag 1500
gaagattaa 1509

```

```

SEQ ID NO: 57      moltype = AA length = 502
FEATURE
source            Location/Qualifiers
                  1..502
                  mol_type = protein
                  organism = Camptotheca acuminata

```

```

SEQUENCE: 57
MGS L D S N Y D T   E S P A S V G Q F N   P L D P E E F R K Q   A H C I V D F I A D   Y Y K N I E S Y P V   L S Q V D P G Y R H   60
S R L G K N A P Y R   S E P F E S I L K D   V Q K D I I P G M T   H W M S P N F F A H   F P A T V S S A A F   V G E M L C T C F N   120
S V G F N W L A S P   A A T E L E M V V I   D W L A N M L K L P   K S F M F S G T G G   G V L Q G T T S E A   I L C T L I A A S P   180
M H F E I V G V K T   S T S F V V Y G S D   Q T H S T Y A K A C   K L A G I L P C N I   R S I P T T A D S N   F S V S P L L L R R   240
A I E A D K A A G M   V P L Y I C A T V G   T T S T T A I D P L   S S L A D V A N D Y   G V W F H V D A A Y   A G S A C I C P E F   300
R H Y L D G I E R A   D S L S L S P H K W   L L S Y L D C C C L   W V K S P S L L V K   A L S T D P E Y L K   N Q P S E S K S V V   360
D Y K D W Q V G T G   R R F K A L R L W F   V M R S Y G V A N L   Q S H I R T D V Q M   A K M F E G F V K S   D P R F E I L V P R   420
V F S L V C F R L N   P I S G S D P T G T   E A L N R K L L D W   V N S T G R V Y M T   H T K V G G I Y M L   R F A V G A T L T E   480
K R H V S S A W K L   I K E G A D V L L K   E D                                     502

```

```

SEQ ID NO: 58      moltype = DNA length = 792
FEATURE
source            Location/Qualifiers
                  1..792
                  mol_type = genomic DNA
                  organism = Oryctolagus cuniculus

```

```

SEQUENCE: 58
a t g g a a g g t g   g t t t a c t g g   t g g t g a c g a a   t a c c a a a g c   a t t t c t t g c c   a a g a g a t t a c   60
t t a a a c a c a t   a c t a c t c a t t   c c a a t c t g g t   c c a t c a c c a g   a a g c t g a a a t   g t t g a a g t t t   120
a a t t t g g a a t   g t t t g c a t a a   g a c t t t c g g t   c c a g g t g g t t   t g c a a g g t g a   c a c t t t a a t t   180
g a t a t t g g t t   t c g g t c c a a c   a a t c t a t c a a   g t t t t g g c t g   c a t g t g a a t c   t t t t a a g a t   240
a t c a c t t t g t   c a g a t t t t a c   a g a t a g a a a t   a g a g a a g a a t   t g g c t a a a t g   g t t g a a g a a a   300
g a a c c a g g t g   c a t a c g a t t g   g a c t c c a g c t   t t g a a g t t c g   c a t g t g a a t t   a g a a g g c a a c   360
t c t g g t a g a t   g g c a a g a a a a   a g c t g a a a a g   t t g a g a g c a a   c a g t t a a g a g   a g t t t t g a a g   420
t g t g a t g c t a   a t t t g t c a a a   t c c a t t a a c t   c c a g t t g t t t   t g c c a c c a g c   a g a t t g t g t t   480
t t g a c a t t g t   t a g c t a t g g a   a t g t g c a t g t   t g t t c t t t g g   a t g c t t a t a g   a g c t g c a t t g   540
a g a a a t t t g g   c a t c a t t g t t   a a a a c c a g g t   g g t c a t t t g g   t t a c t a c a g t   t a c t t t g c a a   600
t t a t c t t c a t   a c a t g g t t g g   t g a a a g a g a a   t t t t c t t g t g   t t g c t t t g g a   a a a a g a a g a a   660
g t t g a a c a a g   c t g t t t t a g a   t g c a g g t t t c   g a t a t c g a a c   a a t t g t t g t a   c t c t c c a c a a   720
t c t t a c t c a g   c t t a c t a c a g c   a c c a a a t a g a   g g t g t t t g t t   t c t t g g t t g c   t a g a a a g a a a   780
c c a g g t t c a t   a a                                     792

```

```

SEQ ID NO: 59      moltype = AA length = 263
FEATURE
source            Location/Qualifiers
                  1..263
                  mol_type = protein
                  organism = Oryctolagus cuniculus

```

```

SEQUENCE: 59
M E G G F T G G D E   Y Q K H F L P R D Y   L N T Y Y S F Q S G   P S P E A E M L K F   N L E C L H K T F G   P G G L Q G D T L I   60
D I G S G P T I Y Q   V L A A C E S F K D   I T L S D F T D R N   R E E L A K W L K K   E P G A Y D W T P A   L K F A C E L E G N   120
S G R W Q E K A E K   L R A T V K R V L K   C D A N L S N P L T   P V V L P P A D C V   L T L L A M E C A C   C S L D A Y R A A L   180
R N L A S L L K P G   G H L V T T Y T L Q   L S S Y M V G E R E   F S C V A L E K E E   V E Q A V L D A G F   D I E Q L L Y S P Q   240
S Y S A S T A P N R   G V C F L V A R K K   P G S                                     263

```

```

SEQ ID NO: 60      moltype = DNA length = 1131
FEATURE
source            Location/Qualifiers
                  1..1131
                  mol_type = genomic DNA
                  organism = Hordeum vulgare

```

```

SEQUENCE: 60
a t g g a t a a g a   t c t c t g c a c c   a t t t t t c t c t   g g t a c t t c a c   c a g c t g c a g c   t t c a g t t g c t   60
g g t g t t g a t g   a a g a t g a t a g   a t t g t g t t t c   c a a g c t c a a g   a a t t g a t g t t   c g c a t a c a a c   120
a t c t c t a t g g   t t t t g a g a g c   a g c a t t a t c a a   t t a g g t t t g t   t a g a t g c t t t   g t c a g c a g c t   180
g g t g g t a a a g   c a t t g a c t c c   a a a c g a a t t a   g t t g a a a a c g   t t g a a a c a t c   t t c a a a t a a g   240
g c t g a a g c a g   c t g c a g c t g t   t g a t a g a a t t   t g t c t t g t t t   c a a c g t t g t t   300
a c t t g t t c t t   c a g a a g c a g c   t g g t c c a g a t   g g t a c a t t a g   t t a g a a g a t a   c a c t a c a g g t   360
c c a t t g t g t a   g a t g g t t a a c   t a a a g a t a g a   g g t g a c g g t a   c a t t a t c t c c   a t t t g c t g t t   420
t t t g t t g t g   a t c c a g a t c a   t t t g t t c c t   t g g c a t c a t a   t t g c a g a a g c   t g t t a c t g c t   480
g g t g g t c c a t   c a g c a t t c g a   a a g a a c a c a a   a a g t g g c c a t   a c t a c g a a t a   c a t g g g t a a a   540
a a t c a a a g a t   t g g g t a c t t t   g t t c g a t a a c   g c a a t g g c t c   a a c a t t c t g t   t a t t t t g g t t   600
a c a a a g a t g t   t a g a a g t t t t   c a a a g g t t t t   g a t g g t g t t c   a a a g a t a g t   t g a t g t t g g t   660
g g t g g t a c t g   g t t c t a c a t t   g g g t a t g a t c   a c t t c a a a g t   a c a a g c a c a t   g a c a g g t a t t   720
a a t t a c g a t t   t g c a c a t a t   t a t t g c t c a a   g g t t t g c c a t   t a c c a g g t g t   t g a a c a t g t t   780
g c a g g t g a c a   t g t a c g a a t c   t a t t c c a a c t   g g t g a c g c t g   t t t t g t t a c a   a t g g a t t a c a   840
t t g a t g t t g a   a c g a t g a t a   a t c g t t a a g   a t c t t g t c a a   a t t g t c a t a a   t g c t t t g c c a   900
a a a g a t g g t a   a a g t t a t t g t   t g t t g a t g g t   a t t t t a c c a g   a a a a c c c a g a   t t c t t c a t t g   960

```

-continued

```

actgcaagag atgctttttac attagatatc atcatgttcg ttttgtttaa aggtgcaaaag 1020
caaagaactg aaaaggaatt cgcaagatta gctaagcaag caggttttac tgggtgtatt 1080
aagaaaactt atatttttctt taattttttac gcttttgaat tcaactaaata a 1131

```

```

SEQ ID NO: 61      moltype = AA  length = 376
FEATURE           Location/Qualifiers
source            1..376
                  mol_type = protein
                  organism = Hordeum vulgare

```

```

SEQUENCE: 61
MDKISAPFFS GTSPAAASVA GVDEDDRLCF QAQELMFAYN ISMVLRAAIQ LGLLDALSAA 60
GGKALTPNEL VENVETSSNK AEAAAAVDRI LRYLSCFNVV TCSSEAAGPD GTLVRRYTTG 120
PLCRWLTKDR GDGTLSPFAV FVVDPDHLFP WHHIAEAVTA GGPSAFERTQ KWPLYEYMGK 180
NQRLGTLFDN AMAQHSVILV TKMLERFKGF DGVQRLVDVG GGTGSTLGMI TSKYKHMTGI 240
NYDLPHVIAQ GLPLPGVEHV AGDMYESIPT GDAVLLQWIT LMLNDDFVK ILSNCHNALP 300
KDGKVIIVDG ILPENPDSSL TARDATLDI IMFVLFKGAK QRTKEFARL AKQAGFTGGI 360
KTTYIFFNFY ALEFTK 376

```

```

SEQ ID NO: 62      moltype = DNA  length = 549
FEATURE           Location/Qualifiers
source            1..549
                  mol_type = genomic DNA
                  organism = Streptomyces griseofuscus

```

```

SEQUENCE: 62
atgaacacct tcagaacagc cactgccaga gacatacctg atgtagcagc aactcttacg 60
gaagccttcg caactgatcc acccacgcag tgggtgttcc ccgacggtac tgccgccgtc 120
agcaggttct ttacacatgt tcgagatagg gttcacacgg ccggtggtat tgttgagcta 180
ctaccagaca gagccgccaat gattgcattg ccaccacacg tgaggctgcc aggagaagct 240
gccgacggaa ggcagggcga aattcagaga aggctggcag acagggcaccc gctgacacct 300
cactactacc tgctgtttta cggagttaga acggcacacc aggggtcggg attggggcga 360
agaatgctgg ccagattaac tagcagagct gatagggaca ggggtgggtac atatactgag 420
gcatccacct ggcgtggcgc tagactgatg ctgagacatg gattccatgc tacaaggcca 480
ctaagattgc cagatggacc cagcatgttt ccactttgga gagatccaat ccatgatcat 540
tctgattag 549

```

```

SEQ ID NO: 63      moltype = AA  length = 182
FEATURE           Location/Qualifiers
source            1..182
                  mol_type = protein
                  organism = Streptomyces griseofuscus

```

```

SEQUENCE: 63
MNTFRTATAR DIPDVAATLT EAFATDPPTQ WVFPDGTAAV SRFFTHVADR VHTAGGIVEL 60
LPDRAAMIAL PPHVRLPGEA ADGRQAEIQR RLADRHLPTP HYLLFYGVR TAHQSGSLGG 120
RMLARLTSRA DRDRVGYTTE ASTWRGARLM LRHGFHATRP LRLPDGPSMF PLWRDPIHDH 180
SD 182

```

```

SEQ ID NO: 64      moltype = DNA  length = 1170
FEATURE           Location/Qualifiers
source            1..1170
                  mol_type = genomic DNA
                  organism = Thermotoga maritima

```

```

SEQUENCE: 64
atgaaaggat atttcggacc atacgggtggc cagtacgtac cagaaatatt aatgggtgcc 60
ttagaggagt tagaggcagc atacgaggag attatgaagg atgagagctt ctggaaggag 120
ttcaacgatc tactgaggga ttacgcaggc agaccaacgc cattgtactt tgccaggaga 180
ttgtctgaga agtacggcgc ccgtgtttac ttgaagcgtg aggatctgct gcacactgga 240
gcacacaaga taaataacgc tatcggacag gttttattgg ccaaatatatt ggggaagaca 300
cgatatcatg ccgagacggg agctgggcag catggagtcg ctactgctac cgctgctgcc 360
ctgttcggaa tggaaatgtg gatctacatg ggtgaagagg acacaatcag acagaagttg 420
aacgtggagc gtagtgaatt attaggggct aaagtgtgcc ctgttaagtc tggcagtagg 480
accttgaagg atcgcataga cgaggctttg agagactgga ttactaatat acagacaaca 540
tattatgtta tcggatctgt tgttgggtccc cacccttacc caattatcgt aaggaatttc 600
cagaaggtta tcggtgagga gaccaagaag caaataccag aaaaggaagg tcgtttgcca 660
gactatatag ttgctctcgt agcgcgcggt agcaatgccg caggtatatatt ttaccattc 720
atagactctg gagtaaaagt gataggtgtt gaggcagggtg gcgagggatt ggagacaggt 780
aaacacgcag cctcgttatt aaagggtaaa attggctatt tacatggatc gaagaccttt 840
gttctacaag atgactgggg tcaagtccaa gtgagccatt cggtgtcagc tggctctgac 900
tattcaggag taggacctga gcatgcttat tggagagaga cagggaagggt tctgtacgac 960
gcagtgactg acgaagaggg tttggacgca tttatagagt tatcaagact agagggcatt 1020
ataccgcgtt tagagtcatc gcatgtctta gcatatttga agaagataaa tataaaagggt 1080
aagggttggtg tggccaacct atcagggaga ggggataaag acctggagtc agtcttaaac 1140
catccatacg tgagagaaag aattagatga 1170

```

```

SEQ ID NO: 65      moltype = AA  length = 389
FEATURE           Location/Qualifiers
source            1..389
                  mol_type = protein
                  organism = Thermotoga maritima

```

```

SEQUENCE: 65

```

-continued

MKGYFGPYGG QYVPEILMGA LEELEAAYEE IMKDESFWE FNDLLRDYAG RPTPLYFARR 60
 LSEKEYGARVY LKREDLLHTG AHKINNAIGQ VLLAKLMGKT RIIAETGAGQ HGVATATAAA 120
 LFGMECVIYM GEEDTIRQKL NVERMKLLGA KVVVPKSGSR TLKDAIDEAL RDWITNLQTT 180
 YVIGSVVGP HPYPIIVRNF QKVIQEETKK QIPEKEGRLP DYIVACVGGG SNAAGIFYPF 240
 IDSGVKLLGV EAGGEGLETG KHAASLLKKG IGYLHGSKTF VLQDDWGQVQ VSHSVSAGLD 300
 YSGVGPEHAY WRETGKVLVD AVTDEEALDA FIELSRLEGI IPALESSHAL AYLKKINIKG 360
 KVVVVNLSGR GDKDLESVLN HPYVRERIR 389

SEQ ID NO: 66 moltype = DNA length = 1167
 FEATURE Location/Qualifiers
 source 1..1167
 mol_type = genomic DNA
 organism = Pyrococcus furiosus

SEQUENCE: 66
 atgtgggttcg gtagtattgg tggacaatat gtgccagaga ctttagtggg tcctcttaag 60
 gaattggaag aggcataata aaggttcaag gacgatgagg agttcaacag gcaactaaac 120
 tattatttga agacatgggc cggtagacca acgcccttgt attatgctaa gaggttaact 180
 gaaaagattg gcggcgcgaa agtgatatctg aaaagagaag acctagtcca tgggtggagca 240
 cacaagacaa ataatgccat tggacaagca ctattggcaa agctaattggg taaaactaga 300
 ttgatagctg agacaggagc ggggtcgcga cagcgatggc tgggtgcacta 360
 ctggggatga aggtagatat ttacatgggt gctgaggacg ttgagcgcca gaaactaaat 420
 gtcttcagga tgaagctatt aggtgccaat gttatacctg taaattctgg ctcaagaaca 480
 ctaaaggacg ccttcgacga ggcctcttaga gactgggttg ccactttcga gtatactcat 540
 tacttgatcg gttcagtggt tggaccacat ccatacccaa ccatcgtagg ggactttcag 600
 agcgtgattg gtagagaggc taaggcacag atcttagaag cagagggaca gctacctgac 660
 gtcatagttg cctgcgtcgg cgggtgctct aacgcaatgg gtataattca tccattcgtt 720
 aatgacaaga aggttaaat agtaggagtc gaagctggcg gaaaggggtt agagtcgggt 780
 aaacactcag caagcttaaa tcaggacag gtaggggtgt ccacggcat gttgtcgtat 840
 ttcttgcaag acgaggaagg tcagataaag ccaatgctcc caattgctcc aggccttgac 900
 caccocgggt ttggtccaga gcacgcttac ttaaagaaga ttcaaaagggc cgagtacgac 960
 gctgtaacag acgaagaggc attgaaagct ttccatgagc tatccagaac tgaggggatt 1020
 ataccggccc ttgagtcgtg ccatgctgtg gcgtacgcca tgaagttagc taaagagatg 1080
 tcccgtgacg aaatcatcat tgtaaatcta tcaggagagag gagacaagga ttggacatt 1140
 gtattgaagg caagcggaag tgtttga 1167

SEQ ID NO: 67 moltype = AA length = 388
 FEATURE Location/Qualifiers
 source 1..388
 mol_type = protein
 organism = Pyrococcus furiosus

SEQUENCE: 67
 MWFGEFGGQY VPETLVGPLK ELEKAYKRFK DDEEFNRQLN YYLKTWAGRP TPLYYAKRLT 60
 EKIGGAKVYL KREDLVHGGG HKTNNNAIGQA LLAKLMGKTR LIAETGAGQH GVATAMAGAL 120
 LGMKVDIYMG AEDVERQKLN VFRMKLLGAN VIPVNSGSRT LKDAFDEALR DWVATFEYTH 180
 YLIGSVVGPV PYPTIVRDFQ SVIGREAKAQ ILEAEGQLPD VIVACVGGGS NAMGIFYPFV 240
 NDKKVKLVGV EAGGKGLESG KHSASLNAGQ VGVSHGMLSY FLQDEEGQIK PSHSIAPGLD 300
 HPGVGPEHAY LKKIQRAEYV AVTDEEALKA FHELSTRTEGI IPALESAAHV AYAMKLAKEM 360
 SRDEIIIVNL SGRGDKDLDI VLKASGNV 388

SEQ ID NO: 68 moltype = DNA length = 1167
 FEATURE Location/Qualifiers
 source 1..1167
 mol_type = genomic DNA
 organism = Pyrococcus furiosus

SEQUENCE: 68
 atgtgggtttg ggggaattcgg tgggtcaatac gtgccagaaa ccttagtggg gcctttaaag 60
 gagcttgaga aagcttacaa aagatttaag gacgacgagg agtttaatag gcaattaaac 120
 tattacttga agacttgggc aggttagacca acaccacttt attacgcgaa gagacttaca 180
 gagaagatcg gtggcgctaa ggtttatttg aagcgtgagg atttggttca tgggtggagca 240
 cataagacaa acaatgccat cgggcaagcc ttgttgccca aattgatggg taagactagg 300
 ttgatagctg gcacagtgct cggccagcat ggcgttgcaa cagcaatggc tgggtgctctt 360
 ttagggatga aggtggatat ctatatgggc gccgaggacg ttgaaagaca aaagcttaac 420
 gtattcagaa tgaagtatt aggtgcgaat gttattccag tgaattcagg ttccaggacg 480
 ctgaaggacg caatagacga agccttgaga gactgggttg ctacttttga gtacacacat 540
 tatttgattg gttccgtggt tggacccccc ccgtacccta caatagttag agattttcaa 600
 tctgttatcg gtagagaggc gaaggctcaa attccagagg ccgagggaca gctaccggat 660
 gttatagtcg cctgcgttgg aggaggatca aatgcgatgg gaattcttca cccttttcta 720
 aatgataaga aagtttaagtt agtcggagtt gaggcggtg gtaaaaggct agaaagcggc 780
 aaacactcgg catctcttaa tgcgggtcaa gtaggagtaa gtcacgggat gttatcttac 840
 ttcttgacag atgaggaagg tcaaaattaag ccatctcatt ctatagcccc aggattagac 900
 tatcccggtg ttggggcctga gcacgcttat ctgaagaaga tacagcgctg tgagtacggt 960
 gcagtaactg atgaagaggc tttgaaggct tttcacgagt tatcaagaac tgaggggaata 1020
 attcctgcct tggaaatcgg acatcggtt gcctacgcca tgaagcttgc taaggaaatg 1080
 tcaagagacg aaatcattat agtgaattta tcgggaaggg gcgataagga cttggacatt 1140
 gttttaaaag caagtggcaa tgtataa 1167

SEQ ID NO: 69 moltype = AA length = 388
 FEATURE Location/Qualifiers
 source 1..388

-continued

```

mol_type = protein
organism = Pyrococcus furiosus

SEQUENCE: 69
MWFGEFGGQY VPETLVGPLK ELEKAYKRFK DDEEFNRQLN YYLKTWAGRP TPLYYAKRLT 60
EKIGGAKVYL KREDLVHGGG HKTNNNAIGQA LLAKLMGKTR LIAGTGAGQH GVATAMAGAL 120
LGMKVDIYMG AEDVERQKLN VFRMKLLGAN VIPVNSGSRT LKDAIDEALR DWVATFEYTH 180
YLIGSVVGPY PYPTIVRDFQ SVIGREAKAQ IPEAEGQLPD VIVACVGGGS NAMGIFYPFV 240
NDKKVKLVGV EAGGKGLESG KHSASLNAGQ VGVSHGMLSY FLQDEEGQIK PSHSIAPGLD 300
YPGVGPEHAY LKKIQRAEYV AVTDEEALKA FHELRSRTEGI IPALESABAV AYAMKLAKEM 360
SRDEIIIVNL SGRGDKDLDI VLKASGNV 388

SEQ ID NO: 70      moltype = DNA length = 1173
FEATURE            Location/Qualifiers
source              1..1173
                    mol_type = genomic DNA
                    organism = Pyrococcus furiosus

SEQUENCE: 70
atgtggttgg gtgagtttgg tggtcagtat gtgcctgaga cgctagttgg gccactgaag 60
gagtttagaga aagcgtacaa gagattcaaa gatgacgagg agtttaaccg tcagttgaat 120
tattatctta agacgtgggc tggtaggcgg acccctctat actacgctaa gcgtttaacc 180
gaaaagatag gtggcgccaa gatctatctg aagagagagg acctagtcca cggcggggct 240
cacaanaacta acaacgcaat tggtaggca ccattagcaa agttaatggg caaaacacgt 300
ctaattggcg aaactgggtg tggtaacat ggagttgcta cggcaatggc gggcgccactg 360
ttaggtatga aagtggatat ctatatggga gcagaagacg tggagagaca gaagatgaac 420
gtgttcctga tgaactattt gggagctaat gttatcccg ttaactctgg ctcaagaact 480
gcaaaggacg ccataaagca agctttgagg gattgggaag ccacattcga gtacacacac 540
taccttatag gctcagtcgt agggcccgcat cctaccccca ccatagttcg tgacttccag 600
agtgttattg gccgtgaggg caaggcacag atcttggaag ctgaagggtca attgcctgac 660
gtaattagtt catgtgtggg tggcggttct aacgctatgg gtatatctta ccccttcgtg 720
aacgataaga aagtaaaagt agttggagtg gaagcaggtg gcaaagggtt ggagtcgggc 780
aagcactcag cctcattaaa gcaggtcag gtaggcggtt tgcacggtat gttatcctat 840
ttcttacagg acgaggaagg ccaaatcaaa ccatcgcat caatcgacc tggttttagat 900
taccaggag tcggtcccga gcacgcata ttaagaaaaa tccagagagc agagtacgtc 960
acggtgacgg atgaggaagc ccttaagcgt ttccagagc tgtcgaggac cgagggcata 1020
atacctgttc ttgaaagtgc ccacgcagtt gcctatgcta tgaagttggc caaggagatg 1080
agtagagacg agattatcat agtaaatctg tcaggcagag gtgacaagga tcttgacata 1140
gtccttaagg cgagcggtaa tgttttgagg taa 1173

SEQ ID NO: 71      moltype = AA length = 390
FEATURE            Location/Qualifiers
source              1..390
                    mol_type = protein
                    organism = Pyrococcus furiosus

SEQUENCE: 71
MWFGEFGGQY VPETLVGPLK ELEKAYKRFK DDEEFNRQLN YYLKTWAGRP TPLYYAKRLT 60
EKIGGAKIYL KREDLVHGGG HKTNNNAIGQA PLAKLMGKTR LIAETGAGQH GVATAMAGAL 120
LGMKVDIYMG AEDVERQKMN VFRMKLLGAN VIPVNSGSRT AKDAINEALR DWEATFEYTH 180
YLIGSVVGPY PYPTIVRDFQ SVIGREAKAQ ILEAEGQLPD VIVACVGGGS NAMGIFYPFV 240
NDKKVKLVGV EAGGKGLESG KHSASLNAGQ VGVLHGMLSY FLQDEEGQIK PSHSIAPGLD 300
YPGVGPEHAY LKKIQRAEYV TVTDEEALKA FHELRSRTEGI IPALESABAV AYAMKLAKEM 360
SRDEIIIVNL SGRGDKDLDI VLKASGNVLE 390

SEQ ID NO: 72      moltype = DNA length = 1173
FEATURE            Location/Qualifiers
source              1..1173
                    mol_type = genomic DNA
                    organism = Pyrococcus furiosus

SEQUENCE: 72
atgtggttgg gagagttcgg cgggcagtag gtcccagaaa cgctagtcgg cccactgaag 60
gagctggaga aagcatacaa aagattcaag gatgacgagg agttcaacag gcaattaaac 120
tattacctaa agacctgggc tggtaggcgg acccctttat actatgcaaa acgtttaacg 180
gagaaaaatcg gtggcgcaaa agtgtagctg aagagagaag acttagtcca cggtggtgct 240
cataaaacta acaacgctat agggcaggca cttttagcta agttgatggg taaaacgagg 300
ttgattgcgg gtaccgtgct gcgtcaacat ggcgttgcta cggcaatggc gggagcccta 360
ttgggcatag aggttgacat ttatatgggc gctgaggacg tggaaactga gaaactgaac 420
gtctatcgta tgaatttctt tgggtccaac gttataccag tcaattccgg ctgcgctacg 480
gttaaagacg ctttcgatga ggcactatgt gatagagttg ccactttcga atacaccac 540
tatcttattg gtactgtctg gggtcacat ccgtatccaa ctattgtaag ggacttccag 600
actgttatcg gaagagaggg aaaggcccaa atattagagg ctgagggaag gttaccggac 660
gccattgttg catgcgtagg cgggtgtagt aacgcaatgg gtattttcta tcccttcggt 720
aacgataaga aggttaagtt agtgggtgct gaggcgggtg gtaaggggct tgagtcgggt 780
aaacattccg catctttgaa tgctgggcaa gtaggtgtca gtcattggaat gttatcatac 840
ttctacagg acgaggaagg acagataaaa ccttctcatt ctatagcgcc aggtcttgac 900
caccctggcg ttggaccaga gcatgcttat ttgaagaaga tccagagagc tgaatatgta 960
gctgttacgg atgaggaagc tctaaaggcg ttccacgagt taagtaggac tgaggggaat 1020
ataccggcat tagagctgct ccacgctgta gcttacgcaa tgaagctggc caaagagatg 1080
agcgtgacg agatcatcat cgtaaacttg tctggcagag gcgataagga cctagacatc 1140
gtattaaagg cctcggggaa cgtccctagag tag 1173

```

-continued

SEQ ID NO: 73 moltype = AA length = 390
 FEATURE Location/Qualifiers
 source 1..390
 mol_type = protein
 organism = Pyrococcus furiosus

SEQUENCE: 73

MWFGFEGGQY	VPETLVGLPK	ELEKAYKRFK	DDEEFNRQLN	YYLKTWAGRP	TPLYYAKRLT	60
EKIGGAKVYL	KREDLVHGGG	HKTNNNAIGQA	LLAKLMGKTR	LIAGTGAGQH	GVATAMAGAL	120
LGMKVDIYMG	AEDVERQKLN	VYRMKLLGAN	VIPVNSGSR	VKDAFDEALC	DRVATFEYTH	180
YLIGTVWGP	PYPTIVRDFQ	TVIGREAKAQ	ILEAEGRLPD	AIVACVGGGS	NAMGIYFPFV	240
NDKKVKLVGV	EAGGKGLSE	KHSASLNAGQ	VGVSHGMLSY	FLQDEEGQIK	PSHSIAPGLD	300
HPGVGPEHAY	LKKIQRAEYV	AVTDEEALKA	FHELSRTEGI	IPALESAAHV	AYAMKLAKEM	360
SRDEIIIVNL	SGRGDKDLDI	VLKASGNVLE				390

SEQ ID NO: 74 moltype = DNA length = 6347
 FEATURE Location/Qualifiers
 source 1..6347
 mol_type = genomic DNA
 organism = Saccharomyces cerevisiae

SEQUENCE: 74

gtatccggct	gttcccttca	agccctttca	atgaacgttg	cagccctttg	aagattggcc	60
attttgtcag	gactcgagcc	tgacagttag	accaaagcaa	ctttaatttt	ttgtgaaaga	120
atcttcgaag	cactcatact	ggcgatcttc	acgccctcct	gctattacaa	aagctgtgtt	180
tttacaagaa	tcaaatcaag	ttagcaagat	attatacaac	attattgata	atttcaatat	240
cgtgttcgta	ctgatgagc	tatctgtgca	ttgataaggc	cgcgatgggt	tcagaaagca	300
gagcggaacg	attccaaatt	agtggccttg	tgctttgcat	gtcaattgtg	ttaccttcag	360
ctcgtggatt	tgttttatca	atacacagtc	tacagtcaag	aatttttttt	atcaaatatt	420
gcgttcgagc	gtataaaata	gccgctgtag	ctacttaagt	tcctgttcag	cgatagtatt	480
tttccatcac	acgtactatg	gcaattaaag	cctcagcgag	ctcgcagtga	atcgctgcga	540
tgagcgacct	catgctatac	ctgagaaaag	aacctgacct	acaggaagaa	gttactcaag	600
aataagaatt	ttcgtttttt	aaacctaaag	tcactttaaa	atttgataac	acttattttt	660
tttataaatt	atttaataat	aaaaatcata	aatcataaga	aattcgctta	tttagaagtg	720
tcaacaacgt	atctaccaac	ggaatgcgtg	cgattcagcg	gtagtctgga	acgtcgtatg	780
ggatgggacc	agagacggat	tgagaaactt	gttcggttct	tgggacgaat	ctacattgga	840
atggaactgg	gtgcctgaag	aaaccagtg	tgaaatcagc	tgggatgtca	attggcttac	900
cggttttggtc	aactggctct	tcaatgttga	aagcagacaa	caaagtagca	ccggcaatcc	960
aaacagtaga	ttgagctagg	tgaataacct	ggcagtttct	tctaccgtaa	ccgaaagcag	1020
cctttctctg	gtctcttaac	gtgttgctcg	gtttaccatc	aggacccaag	tatctttctg	1080
gacggaaaac	ggatggatct	ggatagactt	ctgggtcatt	caaaactgcc	cagggtgttag	1140
caaaaaccac	ttgtgttttt	ggaatacaat	aacctctgta	aacatcatcc	ttcatcaatt	1200
tatgagggat	ggctaattga	gcaatttggt	tccatctgaa	taattccttg	atacaagcgg	1260
tcaaatatgg	gagtgagtcg	tcttctctat	cgtagtccag	gatttgaccg	ttgttggtca	1320
aagcatccaa	ttcagcttga	acctttcttt	gcaattcttg	gtatttaacc	atggccaata	1380
tgaaagcgga	catagcagag	acggtagtgt	cgcaccacc	aacgttaact	tcagcagcag	1440
tggtcttgat	aacatgttct	tggtgttcta	aatcacccgt	caagtccata	gcttgtaatt	1500
tagcagaagc	gtaagatggg	tcggttagac	cttgtggagc	caactttctc	atagtttctg	1560
atggcatgtc	aacctagtgg	tcagcggcct	ctctccagac	cttggcctta	cgttggaaga	1620
cggcacctgg	gaaccaagct	ggcaagtact	tcaaagatgg	gaaggagtag	accagaagct	1680
tacctggaac	actagcaatg	gccaaaacct	cgttgggtaa	gtgggtagct	tctaaccatg	1740
ggatcatctc	ggctagatcg	ataccgtaac	caatgtccaa	agacatagca	gcgatttgat	1800
gtctgatgtg	ttgagccac	cgggtccggg	tcttggtcaa	ttgttgaaac	aattggtgag	1860
cagccttgac	ttgagcgtgt	tgataccctt	ttcagagaat	tccttagcaa		1920
acattctctc	ttcttctctc	caacggtcac	cgtaagtgtg	aaaaccagaa	tcaaatctcc	1980
aacctcatca	ttcattgacc	atagtggtat	ccaatctacc	ggagtagatg	gaaccacgtt	2040
ttccaagag	atccgtgata	gtttccaagg	tgtttaaaat	gacctttctg	gtaccaccag	2100
catctacgta	caagatatca	gtgttgtagt	cacgacccca	ttgcaagaaa	gtcaaccatg	2160
gagattcttc	tggcatgtcg	aacatgttac	caataaatgg	gattggaatt	cctgggtggac	2220
caggtggcaa	ctgggatcta	cgaactcttc	tggagacgat	gtagtaataa	caaccagcaa	2280
tgacgaaaga	gaacaagaca	gcaatcattg	ttttatattt	gttgtaaaaa	gtagataatt	2340
acttccttga	tgatctgtaa	aaaagagaaa	aagaagcat	ctaagaactt	gaaaaactac	2400
gaattagaaa	agacaaaata	tgtatttctt	gcattgacca	atttatgcaa	gtttatatat	2460
atgtaaatgt	aagtttccag	aggttctact	aaactaaacc	acccctcttg	ttagaagaaa	2520
agagtgtgtg	agaacaggct	gttggtgtca	cacgattcgg	acaattctgt	ttgaaagaga	2580
gagagtaaca	gtacgatcga	acgaactttg	ctctggagat	cacagtgggc	atcatagcat	2640
gtggtagtaa	acctttctcc	gccattccag	aaccttcgat	tgcttggtac	aaaacctgtg	2700
agcgtcgctc	aggaccttgt	tgtgtgacga	aattggaagc	tgcaatcaat	aggaagacag	2760
gaagtcgagc	gtgtctgggt	tttttcagtt	ttgttctttt	tgcaaacac	agtttatctc	2820
tcggcatcac	taaatataat	ggagcccgct	ttttaagctg	gcattccagaa	aaaaaaagaa	2880
tcccagcacc	aaaatattgt	ttcttccacc	aacctcagtg	tcataggctc	attctcttag	2940
cgcaactaca	gagaacaggg	gcacaaacag	gcaaaaaacg	ggcacaacct	caatggagtg	3000
atgcaacctg	cctggagtaa	atgatgacac	aaggcaattg	accacgcgat	gtatctatct	3060
cattttctta	caccttctat	taccttctgc	tctctctgat	ttggaaaaag	ctgaaaaaaa	3120
aggttgaaac	cagttccctg	aaattattcc	cctacttgac	taataagtat	ataaagacgg	3180
taggtattga	tgtaatctct	gttaaatctt	ttcttaaaact	tcttaaaatt	tacttttata	3240
gttagtcttt	tttttagttt	taaaacacca	agaacttagt	ttcgaataaa	cacacataaa	3300
caaacaaaat	ggcttctagt	tcttccgatg	tcttcgtttt	gggtctaggt	gttggttttg	3360
ctgccttgta	tatcttcaga	gaccaattat	tcgctgtctc	taagccaaag	gtggctccag	3420
tttccactac	gaagcctgcc	aacggttcgg	ctaaccacag	agacttcac	gccaagatga	3480
aacaaggtaa	gaagagaatc	gtaattctct	acggttctca	aactggtacc	gctgaagaat	3540

-continued

```

atgctattcg tttggctaag gaagctaagc aaaagttcgg tctagcctcc ttggtttgtg 3600
atccagaaga ataccgatttt gaaaagttag accaattgcc agaagattct attgctttct 3660
tcgtcgttgc tacctatgggt gaaggtgaac ctacagacaa cgctgtccaa ttgttgcaaa 3720
acttgcaaga tgaagacttc gaattctcct ctggtgagag aaagtgtgca ggtttgaagt 3780
acgttggttt tggctcgggt aacaagacct acgaacatta caacctcatt gggagaactg 3840
ttgacgctca attggccaag atgggtgcta tcagaatcgg tgaagagggt gaagggtgatg 3900
atgacaagtc catggaagaa gactacttgg aatggaagga tggatatgtg gaagcgtttg 3960
ccactgctat ggggtgtgaa gaaggtaaac gtggtgactc cgtcgatttc gtcgtttccg 4020
aattggaatc tccccacca gaaaagggtt accaagggtga attttctgct agagctttaa 4080
ccaaaacca ggggtattc acgcgtaaga atccttttgc tgcctcaatt gcggttgcta 4140
gagaattggt ccaatctggt gtcgatagaa actgtgtcca cgtcgatttc aacattgaag 4200
gctctggtat cacctatcaa caggtgacc acgttggttt gtggccattg aatccagatg 4260
ttgaagtcga acggttggtt tgtgttttag gtttagctga aaagagagat gctgtcatct 4320
ccattgaaat ccttagaccg gctttggcta aggttccatt ccagtcacca actactacg 4380
gtgctgtggt gagacactac attgacatct ctgctgtcgc cggtagacaa atcttgggta 4440
ctttgtccaa attcgtcca accccagaag ctgaagcttt cttgagaaac ttgaacacta 4500
acaaggaaag ataccacaac gtcgtcgcta acggttggtt gaaattgggt gaaattttgc 4560
aaatcgctac cggtaacgac attactgtcc caccaactac tgccaacacc accaaatggc 4620
caattccatt cgacatcatt gtttctgcca tcccaagatt gcaaccaaga tactactcta 4680
tctcttcttc cccaaaaaatt catccaaaca ccatccacgc taccgttgtt gtgctcaaat 4740
acgaaaaagt tccaacgaa ccaatcccaa gaaagtgggt ttacggtgtc ggtagtaact 4800
tctgttgtaa tttaaagtac gctgttaaca aggaaccagt tccatacatc actcaaatg 4860
gcgaacaaag agtcggtgtc ccggaatact tgattgtcgg tccacgtggt tcttacaaga 4920
ctgaatcttt ctacaaggct ccaatccatg ttagacgttc tactttccgt ttgccaacca 4980
acccaaagtc tccagtcate atgattggtc caggtactgg tgcgcacca ttcagaggct 5040
tcgttcaaga aagagtgtgc ttggccagaa gatccatcga aaagaacggt cctgactctt 5100
tggctgactg gggctcgtatt ccctgttct acggtttag aagatccgac gaagacttct 5160
tgtacaagga cgaatggcca caatacgaag ctgagttgaa gggtaagttc aagttgcact 5220
gtgctttctc cagacaaaac tacaagccag acggttctaa gatttacgtc caagatttga 5280
tctgggaaga cagagaacac attgcccgtg ccatcttaaa cggtaagggt tacgtctaca 5340
tctgcggtga agctaaagtc atgtctaacc aagttgaaga agttctagcc aagatcttgg 5400
gcgaagccaa aggtggttcc ggtccagttg aaggtgtgac tgaagtcaag ttactgaagg 5460
aacggtccag attgatgttg gatgtctggt ctgaacaaaa gtttaattct gaagaagatt 5520
tggaatgaat cgcgtgctct catccgctct aaccgaaaa gaaggagtta gacaacctga 5580
agctaggttc cctatttatt tttttatagt tatgttagta ttaagaacgt tatttatatt 5640
tcaaatTTTT ttttttttct tgtacagacg cgtgtacgca tgtaacatta tactgaaaac 5700
cttgcttagc aaggtttttg gacgctcgaa gatcgctcc caattcgccc tatagttagt 5760
cgtattacgc gcgctcactg gcgctcgttt tacaacgtcg tgactgggaa aaccctggcg 5820
ttacccttgc aggactagtg ctgaggcatt aatagtttac tcaattcttg aagccaattt 5880
gtcaaatcc ccatagagtg caaataaaa gatgcctcac ggaggtagt taccgcgct 5940
atthcacatg gctcattgaa tttagaggtg aatttggtgt accctccct cctcatctga 6000
tgaagtagtg atccgacaat tcttaaaagt tgtagacatt actttacca ccaactaagt 6060
tgtatttata ttgtctacct tatcctttta tatctaacta gcgctcataa ggttggggca 6120
atactaaaac tgtgttctta ttcaactcat taaatacgtg gcagtagcta ccctattaga 6180
aacaatagga aacagcagag tcggaagaag ccaaatgcca gatttgaagt ccaaaacctt 6240
gtcaagccaa tctttgggag cggctattcc tccagaaatt gtgtaccaaa tacttacata 6300
ccagtttagg gatttggtta gaaatgacca tccaggtacg gcagaaa 6347

```

```

SEQ ID NO: 75      moltype = DNA length = 4792
FEATURE            Location/Qualifiers
source              1..4792
                    mol_type = genomic DNA
                    organism = Saccharomyces cerevisiae

```

```

SEQUENCE: 75
caatctggcg gcttgagttc tcaacatggt ttatttttta cttatattgc tggtagggta 60
aaaaaatata gctcctagga ataggttgtc tatatgtttt tgtcttgctt ctataattgt 120
aacaaaacaa gaaagggaaa atactgggtg taaaagccat tgagtcaagt taggtcatcc 180
cttttataca aaatttttca atttttttct caagattctt gtacgattaa ttattttttt 240
tttgctgctc acagcgtgat gaaaatttcg cctgctgcaa gatgagcggg aacggggcga 300
atgtgcacgc gcacaactta cgaaacgcgg atgagtcact gacagccacc gcagagggtc 360
tgactcctac tgagctctat tggaggtggc agaaccggta cggaggagg cgcgtataac 420
cggtttgaat ttattgtcac agtgtcacat cagcattaa tccctacgca gctcgcattg 480
aatgcgtgcg atgagcgacc tcatgctata cctgagaaag caacctgacc tacaggaaag 540
agttactcaa gaataagaat ttctgtttta aaacctaaaga gtcactttaa aatttgtata 600
cacttatttt ttttataact tatttaataa taaaaatcat aaatcataag aaattcgctt 660
atthagaagt gtcacaacg tatctaccaa cggaatcgtg gcgattcagg tggagtccaa 720
acccaacaaa ggatttggaa ttggcttacc ggcagtgagg gatctcttta gcaacgtgga 780
agtgatttca ccgttgctgt tgttacctct agcatcgtgg aaagcagcaa cacccttctt 840
aacaagttg attcttctt cttcagaacc ccattgcatg aagtcagtc acataacaat 900
gtgagcggcg ataccagcgg taaccttggc gtagttagtg gaatgcttgg aagtagggc 960
gtaagattgc aagtaagctt gtctcatggt tgtaccaact tgttctgctt ggaatcgtc 1020
aatcaagtaa cagtcaccca agaagtaacc caaatccaat gaagctggac cgtacttaca 1080
caattccag tctaaagtgt agatcttttg caacttagat ggggtacctt cttcaagttg 1140
caacaagatg tccccagacc caagtcagc catgacaaa gtttcttgg agtgcataac 1200
atcgtaact agatccttga caacagttg caacaatgga tcatcgacgc cgtatttagc 1260
ggcgtttggg ataatagttt ggtacaattg gtcagagggt gttctacga caatgttacc 1320
agagaagaac tgaattctg ggtcgtctct tcttctctca cctatgttgt gcaatctggc 1380
gacgaaacca ccaatctcgg taccaccaa tctagcaata tcggtagcca aagggtggct 1440
agcagtaacg tagtctaata aggtcttcat tttaccgaca tcttgcataa tcaaagcatt 1500

```

-continued

```

gttttccaag tcatagtgtg gaccttctgg aacagagaca ataccatcaa caccacccaa 1560
aacttctctg ttagccatca tcaacttgat agcttgggtat tcgtagacag aacgttcaac 1620
accgattttg aaatcttcat cagtagacat gtgtgggtga gcgtgcttca aaatgataga 1680
agtgtgacct tggatggag cgttcaattt aatacgccag gtgacgttaa cgaaccacc 1740
ggataaatctc ttgacaccag aagtgtcaac atctaaagac aaatgcttgg tcaaataggt 1800
gattaaacgc tcttcagttc tcaagtcaaa ggccattggt ttatatttgt tgtaaaaagt 1860
agataaattac ttcttctgat atctgtaaaa aagagaaaaa gaaagcatct aagaacttga 1920
aaaactacga attagaaaaa accaaatatg tatttcttgc attgaccaat ttatgcaagt 1980
ttatatatat gtaaatgtaa gtttcacgag gttctactaa actaaaccac ccccttgggt 2040
agaagaaaaa agtgtgtgag aacaggctgt tgtgtgcaca cgattcggac aattctgttt 2100
gaaagagaga gagtaacagt acgatcgaac gaactttgct ctggagatca cagtgggcat 2160
catagcatgt ggtactaaac cctttccgcg cattccagaa ccttcgattg cttgttcaaa 2220
aacctgtgag ccgtcgctag gacctgtgtg tgtgacgaaa ttggaagctg caatcaatag 2280
gaagacagga agtcggcggt gtctgggttt ttccagtttt gttctttttg caaacacag 2340
tttattctctg gcattccacta aatataatgg agcccgcttt ttaagctggc atccagaaaa 2400
aaaaagaatc ccagcaccaa aatattgttt tcttcaccaa ccacagttc atagggtccat 2460
tctcttagcg caactacaga gaacaggggc aaaaacgggc aaaaacgtca cacaacctca 2520
atggagtgat gcaacctgcc tggagttaat gatgacacaa ggcaattgac ccacgcatgt 2580
atctatctca tttcttata ccttctgctc tctctgattt ggaaaaagct 2640
gaaaaaaaag gttgaaacca gttccctgaa attattcccc tacttgacta ataagtatat 2700
aaagacggta ggtattgatt gtaattctgt aaatctattt cttaaacctc ttaaatctta 2760
cttttatagt tagtcttttt tttagtttta aaacaccaag aacttagttt cgaataaaca 2820
cacataaaca aacaaaaatgc atatcagaaa cccatataga actccaattg actaccaagc 2880
tttgtctgaa gcttctccac catggaagcc atttgtttcc gttaacgctg atgggtacct 2940
ctcagttgac ttgaccttcc cagaagccca aagagctttt accgtgccc tttgacag 3000
agacttcggc ttgactatga ctatccaga agatcggttg tgtccaaccc ttccaacacg 3060
attgaactac gttttgtgga ttgaagacat tttcaactac accaacaaga ctttgggttt 3120
atctgacgac cgtccaatga aggggtgttg tatcggtacc ggtgcttctg ccatttacc 3180
aatgttggct tgcgcacgat tcaaggcttg gtccatgggt ggtagtgaag ttgaaagaaa 3240
gtgtatcgac actgctagat taaacgttgt tgcatacaac ttgcaagatc gtctatccat 3300
cttggaacc tctattgacg gtccaatttt agtcccaatt ttcgaagcta ccgaagaata 3360
cgaatacgaa taccacatgt gtaaccacc tttctacgat ggtgccgctg acatgcaaac 3420
tagcgatgct gcaaaagggt ttgggttcgg tgcgggtgct ccacactctg gtacagtcac 3480
cgaaatgtct actgaagggt gtgaatccgc tttcgtggct caaatggtta gagaatctct 3540
taagttgaga accagatgta gatggtacac ttctaactta ggtaagttga aatctttgaa 3600
ggaaattgtc ggtttgttga aggaattaga aatctctaat tacgccatca acgaatatgt 3660
ccaagggttc actagaagat acgtgtcgc ttggagtttc actgatatcc aattgccaga 3720
agaattgtcc agaccatcta atcctgaatt gtccctcttg ttcgactaca aggatgacga 3780
tgacaaatga atcgcggtga ttcatccgct ctaaccgaaa aggaaggagt tagacaacct 3840
gaagtctagg tccctatttta ttttttata gttatgtag tattaagaac gttatttata 3900
tttcaaattt ttcttttttt tctgtacaga cgcgtgtacg catgtaacat tatactgaaa 3960
accttgcctg agaaggtttt gggacgctcg aagatcgctg acccaattcg ccttatagtg 4020
agtcgtatta tgcgcgctca ctggccgctg ttttacaacg tcgtgactgg gaaaaacctg 4080
gcgttacccc gtcaggacta gtgctgagc attaatgcaa ctcagaagtt tgacagcaag 4140
caagttcatc attcgaacta gccttattgt ttagtttcag tgacagcgaa ctgccgtact 4200
cgatgcttta tttctcacg tagagcggaa gaacagatag gggcagcgtg agaagagtta 4260
gaaagtaaat ttttatcag tctgaagtat tcttattcat aggaaatttt gcaaggtttt 4320
ttagctcaat aacgggctaa gttatataag gtgttcacgc gattttcttg ttatgtatac 4380
ctctctctcg aggaatggta ctactgtcct gatgtaggct ccttaaattg gtgggcaaga 4440
ataacttata gattatttgt atatttgtct tggagttcac cacgtaatgc ctgttttga 4500
ccatcagtta actctagtat tattttgtct tggctactgg ccgtttgtcta ttattcaagt 4560
cttttgtgcc ttcccgctgg gtaaggaggt tatttaggga tacagaatct aacgaaaact 4620
aaatctcaat gattaactct atttaactct tttttgaaa gcaaaaagag tcccttgttc 4680
acttacaacg ttcttagcca aattcgctta tcaacttacta cttcacgata tacagaagta 4740
aaaacatata aaaagatgtc tgtttgttta gccatcacaa aaggtatcgc ag 4792

```

```

SEQ ID NO: 76      moltype = DNA length = 9145
FEATURE            Location/Qualifiers
source              1..9145
                    mol_type = genomic DNA
                    organism = Saccharomyces cerevisiae

```

```

SEQUENCE: 76
tggtttatat ttgtgttaaa aagtagataa ttacttctct gatgatctgt aaaaaagaga 60
aaaagaaaagc atctaaagac ttgaaaaaact acgaattaga aaagacaaaa tatgtatttc 120
ttgcattgac caatttatgc aagtttatat atatgtaaat gtaagtttca cgaggttcta 180
ctaaactaaa cccccctctt ggttagaaga aaagagtgtg tgagaacagg ctgttgttgt 240
cacacgattc ggacaattct gtttgaaga gagagagtaa cagtacgac gaacgaactt 300
tgctctggag atcacagtgg gcatcatagc atgtgggtact aaacctttc ccgccattcc 360
agaacctctg attgcttggt acaaaaacctg tgagccgtcg ctaggacctt gttgtgtgac 420
gaaattggaa gctgcaatca ataggaagac aggaagtoga gcgtgtcttg gttttttcag 480
ttttgttctt ttgcaaaaa acagtttatt cctggcatcc actaaatata atggagcccg 540
ctttttaagc tggcatccag aaaaaaaaag aatcccagca ccaaaatatt gttttcttca 600
ccaacctca gttcataggt ccattctctt agcgcaacta cagagaacag gggcacaac 660
aggcaaaaaa cgggcacaac ctcaatggag tgatgcaacc tgctggagt aaatgatgac 720
acaaggcaat tgaccacgc atgtatctat ctcatcttct tacaccttct attaccttct 780
gctctctctg atttggaaaa agctgaaaaa aaaggttgaa accagttccc tgaaattatt 840
cccctacttg actaataagt atataaagac ggtaggtatt gattgtaatt ctgtaaatct 900
atctcttaaa cttcttaaat tctactttta tagttagctt ttttttagt tttaaaacac 960
caagaactta gtttcaata aacacacata aacaacaaaa ggatccatga tgaagttctg 1020

```

-continued

gagaaagtac	acacaacaag	aaatggatga	aaagattact	gaatctttgg	aaaagacttt	1080
gaactacgat	aacactaaga	caatcgggat	tccagggtact	aagttggatg	atacagtttt	1140
ctatgatgat	cattctttcg	ttaagcattc	accatacttg	agaactttta	ttcaaaaccc	1200
aaaccataac	gggtgtcata	cttatgataa	ggctgatatc	ttgttcggtg	gtacattcga	1260
tatcgaaaga	gaattaatcc	aattgttagc	aatcgatggt	ttgaacggta	acgatgaaga	1320
atthgatggt	tacgttactc	aaggtgggtac	agaagctaac	atccaagcaa	tgtgggttta	1380
cagaaactac	tcaagaaag	aaagaaaggc	taagcatgaa	gaaatcgcta	tcatcacttc	1440
agcagataca	cattactctg	catacaaaag	ttcagatttg	ttgaacatcg	atattattaa	1500
ggttccagtt	gattttttatt	caagaaaaat	tcaagaaaaat	acattggatt	caattgttaa	1560
agaagctaaa	gaaattggta	aaaagtactt	catcgttata	tctaactatg	gtactacaat	1620
gtttgggtca	gttgatgatc	cagattttga	cgctaacttc	ttcgataagt	acaatttgga	1680
atacaaaatt	catgttgatg	gtgcattttg	tgtgtttata	tatccaattg	ataataagga	1740
atgtaaaact	gatttctcta	ataagaacgt	ttcttcaatc	acattagatg	gtcataagat	1800
gttgcaagct	ccatacggta	ctgggtatct	cggttcaaga	aagaatttga	tccataaacac	1860
tttgacaaag	gaagcaactt	acatcgaaaa	tttggatggt	acattgtctg	gttcaagatc	1920
tgggtcfaat	gctgttgcaa	tttggatggt	tttagcttct	tatgggtccat	acggttggtg	1980
ggaaaagatt	aataagttga	gaaatagaac	ttaaagggtg	tgaagcaaat	tgaacgatata	2040
gagaattaaa	tattacaaga	aagattcaat	gaatattggt	acaattgaag	aacaatatgt	2100
taataaggaa	atcgctgaaa	tgtacttttt	agttccagaa	gttcataacc	caactaacia	2160
ctgggtacaa	atcgtttgta	tggaacatgt	tgaattggat	atcttgaact	ctttggttta	2220
cgatttgaga	aagttttaata	aggaacattt	gaaggcaatg	catcatcatc	atcatcatta	2280
accgcggcta	gctaagatcc	gctctaaccg	aaaagggaag	agttagacaa	cctgaagtct	2340
aggtccctat	tatttttttt	atagttatgt	tagtattaa	aacgttattt	atatttcaaa	2400
tttttctttt	ttttctgtac	agacgcgtgt	acgcgatgaa	cattatactg	aaaaccttgc	2460
ttgagaaggt	tttgggacgc	tccaagatcc	agctgcatta	atgaatcgcc	caacgcgcgg	2520
ggagaggcgg	tttgcgtatt	gggcgcctct	cgcttctctc	gctcaactgac	tcgctgcgct	2580
cggtcggtcg	gctgcggcga	gcgggtatcag	ctcactcaaa	ggcggtaata	cggttatcca	2640
cagaatcagg	ggataacgca	ggaaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	2700
accgtaaaaa	ggcgcggttg	ctggcggttt	tccataggct	ccgccccctc	gacgagcatc	2760
acaaaaatcg	acgctcaagt	cagaggtggc	gaaacccgac	aggactataa	agataccagg	2820
cgtttccccc	tggaaagctc	ctcgtgcgct	ctcctgttcc	gacctgcgc	cttaccggat	2880
acctgtccgc	ctttctccct	tccgggaagc	tggcgcttct	tcatagctca	cgctgtaggt	2940
atctcagttc	gggtgtaggtc	gttcgctcca	agctgggctg	tgtgcacgaa	ccccccgttc	3000
agcccgcagc	ctgcgcctta	tccggtaact	atcgtcttga	gtccaacccg	gtaagacacg	3060
acttatcgcc	actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtaggcg	3120
gtgtctacaga	gttcttgaag	tgtgtggccta	actacggcta	cactagaaga	acagtatttg	3180
gtatctgcgc	ttgctgtga	ccagttacct	tccgaaaaag	agttggtagc	tcttgatccg	3240
gcaaaacaa	caccgctgggt	agcgggtggt	tttttgtttg	caagcagcag	attacgcgca	3300
gaaaaaaagg	atctcaagaa	gatcctttga	tctttctctc	ggggtctgac	gctcagtgga	3360
acgaaaaactc	acgttaagg	attttggta	tgagattatc	aaaaaggatc	ttcaactaga	3420
tccttttaaa	ttaaaaatga	agtttttaaa	caatctaag	tatatatgag	taaaacttgg	3480
ctgacagtta	ccaatgctta	atcagtggag	cacctatctc	agcgatctgt	ctatttcgtt	3540
catccatagt	tgctcgactc	cccgctgtgt	agataactac	gatccgggag	ggcttaacct	3600
ctggccccc	tgctgcaatg	ataccgcgag	acccacgctc	accggctcca	gattttatcag	3660
caataaacca	gccagccgga	aggggccgag	gcgaagtggt	tccgtgcaat	ttatccgcct	3720
ccatccagtc	tatttaattgt	tgccgggaag	ctagagtaag	tagtttcgca	gttaaatggt	3780
tgccgcaact	tgttgccatt	gctacaggca	tcgtgggtgc	acgctcgtcg	tttgggtatg	3840
cttcattcag	ctccggttcc	caacgatcaa	ggcgagttac	atgatcccc	atgttgtgca	3900
aaaaagcgg	tagctccttc	ggtcctccga	tcgttctcag	aagtaagttg	gccgcagttg	3960
tatcactcat	gggttatggca	gcactgcata	attctcttac	tgctatgcca	tcctgtaagt	4020
cgctttctgt	gactgggtgag	tactcaacca	agtcattctg	agaatagtgt	atgcggcgac	4080
cgagttgctc	ttgcccggcg	tcaataccgc	gccacatagc	agaactttta		4140
aagtgctcat	cattggaaaa	cggtcttctg	ggcgaaaact	ctcaaggatc	ttaccgctgt	4200
tgagatccag	ttcgatgtaa	cccactcgtg	cacccaactg	atcttcagca	tcttttactt	4260
tcaccagcgt	ttctgggtga	gcaaaaaaac	gaaggcaaaa	tgccgcaaaa	aagggaataa	4320
ggggcagacg	gaaatgttga	atactcatac	tcttctcttt	tcaatattat	tgaagcattt	4380
atcaggggtta	ttgtctcatg	agcggataca	tatttgaatg	tatttagaaa	aataaaciaa	4440
taggggttcc	gcccacattt	ccccgaaaag	tgccacctga	acgaagcatc	tgtgcttcat	4500
ttgttagaac	aaaaatgcaa	cgcgagagcg	ctaatttttc	aaacaaagaa	tctgagctgc	4560
atthttacag	aacagaaatg	caacgcgaaa	gcgctatttt	accaacgaag	aatctgtgct	4620
tcatttttgt	aaaacaaaaa	tgcaacgcga	gagcgctaat	ttttcaaaaa	aagaattctga	4680
gctgcatttt	tacagaacag	aaatgcaacg	cgagagcgct	atthttacca	caaagaatct	4740
atacttcttt	tttgttctac	aaaaatgcat	cccagagcg	ctatttttct	aacaaagcat	4800
cttagattac	tttttttctc	ctttgtgcgc	tctataatgc	agctctctga	taactttttg	4860
cactgtaggt	ccgttaaggt	tagaagaagg	ctacttttgt	gtctattttc	tcttccataa	4920
aaaaagcctg	actccacttc	ccgcgtttac	tgattactag	cgaagctgcg	gggtgatttt	4980
ttcaagataa	aggcatcccc	gatttatatt	tataccgatg	tggattgcgc	atactttgtg	5040
aacagaaagt	gatagcgttg	atgtattctc	attggtcaga	aaattatgaa	cggtttcttc	5100
tattttgtct	ctatatacta	cgtataggaa	atgtttacat	tttcgtattg	ttttcgattc	5160
actctatgaa	tagttcttat	tacaattttt	ttgtctaaag	agtaatacta	gagataaaca	5220
taaaaaatgt	agaggtcgag	tttagatgca	agttcaagga	gcgaagggtg	gatgggttag	5280
ttatataggg	atatagcaca	gagatatata	gcaaaagagat	acttttgagc	aatgtttgtg	5340
gaagcgggtat	tcgcaatatt	tttagtagtc	gttacagctc	gggtgcgttt	tgggtttttg	5400
aaagtgcgtc	ttcagagcgc	ttttgttttt	caaaagcgct	ctgaagttcc	tataactttc	5460
agagaatagg	aacttcggaa	taggaacttc	aaagcgtttc	cgaaaaacgag	cgcttccgaa	5520
aatgcaacgc	gagctgcgca	catacagctc	actgttcacg	tcgcacctat	atctgcgtgt	5580
tgctgtata	tatatataca	tgagaagaac	ggcatagtcg	gtgtttatgc	ttaaatgcgt	5640
acttatatgc	gtctattttt	gtaggatgaa	gtacctctcg	tgatattatc		5700
ccattccatg	cggggtatcg	tatgcttctc	tcagcactac	ccttttagctg	ttctatatgc	5760

tgcactcct	caattggatt	agtcctatcc	ttcaatgcta	tcatttcctt	tgatattgga	5820
tcatactaag	aaaccattat	tatcatgaca	ttaacctata	aaaataggcg	tatcacgag	5880
cctcttcgtc	tcgcgcgttt	cgggtgatgc	ggtgaaaacc	ttctgacact	gcagctcccg	5940
gagacggtca	cagcttgtct	gttaagcgat	gcggggagca	gacaaacccg	tacggggcgc	6000
tcagcgggtg	ttggcgggtg	tcggggctgg	cttaactatg	cgcgcacaga	gcagattgta	6060
ctgagagtgc	accataattc	acaccgcata	gatccgtcga	gttcaagaga	aaaaaaaaa	6120
aaaaagaaaa	aaaaaaaagg	caagcgcgcc	tcgttgcaga	tgacagctat	agaatgatgc	6180
attaccttgt	catcttcagt	atcatactgt	tcgtatacat	acttactgac	attcataggt	6240
atacatatat	acacatgtat	atatatcgta	tgcgtgcagct	ttaaaatact	gggtctacta	6300
cataagaaca	ctcttgggtg	agggaacctc	ttggtgacca	tgtggcgatc	tgtgtctctc	6360
tatggcaacc	gcaagagctc	tgaacgcact	ctcacactcg	tgaatgatcat	tcttgcctcg	6420
gcacaatca	acgtggaggg	taattcttgt	agcctctgca	aagctttcaa	gaaaatcgcg	6480
gatcatctcg	cagtagagat	ctctctactt	ctccctttgc	aaaccaagtt	gcacaactgc	6540
gtccggcgtc	ttcgaagaat	ctaccacgcg	tctggaaagt	gcctcatcca	aaggcgcaaa	6600
tctgatccca	aaccttttta	ccaccgcgac	ggccctcagg	gcctctttaa	aagcttgacc	6660
gagagcaatc	cgcagctctt	cagtgggtgc	atggtcgctc	atgtgtaagt	ccaccaatga	6720
ctcaacgatt	agcgaccagc	cggaaatgct	ggccagagca	tgtatcatat	ggtcacagaa	6780
ccctataact	gtgtgcagct	taataccttg	cgatttgtgt	gctgtgtctg	cttactgctc	6840
tgcctctttt	tctgggaaga	tcgagtgctc	tatcgctagg	ggaaccacct	ttaaagagat	6900
cgcaactctg	atcttgggtt	cattttgtaat	acgcttttact	agggctttct	gctctgtcat	6960
ctttgccttc	gtttatcttg	ctgcctcatt	ttttagtata	ttcttcgga	aaatcacatt	7020
actttatata	gtgtataatt	catttatgtg	taatgccaat	cgctaagaaa	aaaaaagagt	7080
catccgctag	gtggaaaaaa	aaaaatgaaa	atcatatccg	aggcataaaa	aaatatagag	7140
tgtactagag	gagggcaaga	gtaataagaa	aagaaaattg	cgggaaagga	ctgtgttatg	7200
acttccctga	ctaatcgctg	gttcaaacga	ctctggcgag	tgactctcag	cgctcaccaa	7260
gctcttaaaa	cgggaattta	tgggtgcactc	tcagtagaacg	cgcagatctg	gtttatgtcg	7320
ctctgctccc	gcggggctac	ggcgccagcg	acatggaggc	ccagaataac	ctccttgaca	7380
gtcttgacgt	gcgcagctca	ggggctgatg	gtgactgtgc	cccgctacatt	tagcccatcac	7440
atccccatgt	ataatacttt	gcactcatcac	attttgatgt	cgcgcacgcg	cgaagcaaaa	7500
attacggctc	ctcgcgtcag	acctggcagca	agggaaaacg	tcacctccca	gacgcgttga	7560
attgtcccca	cgcgcgcgcc	ctgtagagag	atataaaagg	ttaggatttg	ccactgaggt	7620
tctcttttca	tatacttctc	tttaaaactc	tgtcaggata	cagtttctcac	atcacatccg	7680
aacataaaaa	accattgggt	aggaaaaagc	tcacgtttct	agcgccgat	taaatccaaa	7740
catggatgct	gatttatatg	gggtataaag	ggctcgcgat	aatgctgggc	aatcaggtgc	7800
gcgaatctgt	cgaattgatg	ggagagccga	tgcgcagagc	tgttttctga	aacatggcaa	7860
aggtacgctt	gcgaatgatg	ttacagatga	gatggctcaga	ctaaactcgt	tgacggaaat	7920
tatgcctctt	ccgaccatca	agcattttat	ccgtactcct	gatgatgcac	gggtactcac	7980
cactgcgcat	cccggcacaaa	cagcatctca	ggatatagaa	gaatatcctg	attcagtgta	8040
aaatatttgt	gatgcgctgg	caggtgttct	gcgcgggttg	catctgatct	ctgtttgtta	8100
ttgtcctttt	aacacgcgat	gcgtattttg	tctcgctcag	gcgcaatcac	gaatgaataa	8160
cggtttgggt	gatgcgagtg	attttgatga	cgcagcgtct	ggctggcctc	ttgcaacaag	8220
ctggaaagaa	atgcataa	ttttgcctac	ctcaccggat	tcagctgccta	ctcatgggtg	8280
tttctcactt	gataacctta	tttttgacga	ggggaaaata	ataggttgta	ttgatgttgg	8340
acgagtccga	atcgcagacc	gatcacagga	ttctgcatat	ctatggaact	gcctcgggtg	8400
gttttctctc	tcaattacga	aacgcgcttt	tcaaaaatct	ggbtattgata	atctcgatat	8460
gaataaaatt	cagtttcatt	tgatgctcga	tgagtttttc	taatcagtac	tgacaataaa	8520
aagattcttg	ttttcaagaa	cttgcctatt	tttatgtttt	tttatattgt	agtttgttcta	8580
ttttaatcca	atgttagcgt	gatttatatt	tttttccgc	tcgacatcat	ctgccagagt	8640
gcgaagttta	gtgcgcagaa	agtaatatca	tgcgtcaatc	gtatgtgaat	gctggctgct	8700
atactcgtg	cgaatcgata	ctaacgcgcg	catccagttg	cgaattccgc	attcaggtcg	8760
cgcaactggt	gggaaggcgc	atcgtgtcgg	gcctcttcgc	tatcacgca	cgtaatttgg	8820
agcgacctca	tgctatacct	gagaaaagca	cctgacctac	aggaaagagt	tactcaagaa	8880
taagaatttt	cgttttaaaa	ctcaagagtc	actataaaat	ttgtatacac	ttattttttt	8940
tataaactat	tttaataaaa	aaatcataaaa	tctaagaaaa	tttgccttat	tagaaggtgc	9000
aacaacgtat	ctaccaacga	tttgaccctt	ttccatcttt	tcgtaaaatt	ctggcaaggt	9060
agacaagcgc	acaaccttga	ttggagactt	gaccaaaact	ctggcgaaaga	attgttaatt	9120
aaagactcag	atctttttcc	gcgc				9180

```
SEQ ID NO: 77      moltype = DNA length = 499
FEATURE           Location/Qualifiers
source            1..499
                  mol_type = genomic DNA
                  organism = Saccharomyces cerevisiae
```

SEQUENCE: 77						
gtttgcacaaa	agcaacaaaac	tgaaaaaaac	cagacacgct	cgacttctctg	tcttctctatt	60
gattgcagct	tccaatttctg	tcacacaaca	aggctctagc	gacggctcac	aggttttcta	12
atcagcaatc	gaagggtctg	gaatgcgcgg	aaagggttta	tgaccacatg	ctatgatgcc	18
cactgtgtac	tccagacgaa	agttcgttcg	atcgtatctg	tactctctct	ctttcaaaaa	24
gaattgtctc	aatcgttgtg	caacaacacg	ctggtctctc	acactctttt	cttctaacca	30
agggggttgt	ttagttttag	agaacctcgt	ggaacttcac	tttaccatata	tataaacttg	36
cataaattgg	tcaatgcaag	aaatacatat	ttggctcttt	ctaattcgta	gtttttcaag	42
ttcttagatg	ctttctcttt	ctctttttta	cagatcatca	aggaagtaat	tatctacttt	48
ttacaacaaa	tataaaaaa					49

```
SEQ ID NO: 78      moltype = DNA  length = 500
FEATURE            Location/Qualifiers
source             1..500
                  mol_type = genomic DNA
                  organism = Saccharomyces cerevisiae
```

-continued

SEQUENCE: 78
acagttttatt cctggcatcc actaaatata atggagcccg ctttttaagc tggcatccag 60
aaaaaaaaag aatcccagca ccaaaatatt gttttcttca ccaaccatca gttcataggt 120
ccattctctt agcgcaacta cagagaacag gggcacaaac aggcacaaaa cgggcacaaac 180
ctcaatggag tgatgcaacc tgcctggagt aaatgatgac acaaggcaat tgaccacacg 240
atgtatctat ctcatcttct tacaccttct attaccttct gctctctctg atttggaaaa 300
agctgaaaaa aaaggttgaa accagttccc tgaaattatt cccctacttg actaataagt 360
atataaagac ggtaggtatt gattgtaatt ctgtaaatct atttcttaaa cttcttaaat 420
tctactttta tagttagtct tttttttagt tttaaaacac caagaactta gtttcgaata 480
aacacacata aacaaacaaa 500

SEQ ID NO: 79 moltype = DNA length = 500
FEATURE Location/Qualifiers
source 1..500
 mol_type = genomic DNA
 organism = *Saccharomyces cerevisiae*

SEQUENCE: 79
gctggattga gctgaatggt gccaggtcga ggctgggagg gagactaact cgaaagtgc 60
gaagactcga aaattaaaaa aaaagatact gcagaaggca agattgagaa tggagtaaag 120
gcagcgtggg tccctctgtg aaaccgcagt ttctctgcgc caagtgggtac cggtgcgagt 180
gcagcaatta atctctcgat attttcttag tatctctttt tatataagaa tatatttttg 240
aattggtaat ccttatcttc aatagtttct tagttgaatg cacacttaag agcaaattgg 300
ccaaggagtt gttcgttcgc tttaatattt ttcctgggta ttgtcaattt attcatccca 360
tctcccagg atagaagaaa ttagtgaat tttgctgaca atacatttta acgacgataa 420
caataatagc aattaaataa aatagcacta ccaccactcc actgctcgtt agctatttct 480
gtaaaataaa taaaaagatc 500

SEQ ID NO: 80 moltype = DNA length = 501
FEATURE Location/Qualifiers
source 1..501
 mol_type = genomic DNA
 organism = *Saccharomyces cerevisiae*

SEQUENCE: 80
attcgcgcta tctcgatttc tacctatata gttaatctct gtacaaaaac aatctttcca 60
actatccatt aatcatagta tattatcagc gtcggcgatt ttaccacgct tgacaaaagc 120
cgcgggcggg attcctgtgg gtatggcac cggcagttaa tctaatcaaa ggcgcttgaa 180
ggaagagata gataatagaa caaagcaatc gccgctttgg acggcaataa tgtttatcca 240
ttggtgcggt gattggatat gatttgtctc cagtagtata agcaagcgcc agatctgttt 300
actgtaaaaa taagttagta atctcgcggg atgtaatgat ttaagggaa ctgggtcagg 360
ttttcacata tattgtgata taaggccatt tgtaatttca atagttttag gatttttctc 420
tctcccaaaa tactcactta ctgtgttaca ttacagaaag aacagacaag aaaccgtcaa 480
taagaaatat aactaagaac a 501

SEQ ID NO: 81 moltype = DNA length = 27
FEATURE Location/Qualifiers
misc_feature 1..27
 note = Synthetic HA epitope tag polypeptide
source 1..27
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 81
taccatacag acgttccaga ctacgcc 27

SEQ ID NO: 82 moltype = AA length = 9
FEATURE Location/Qualifiers
REGION 1..9
 note = Synthetic HA epitope tag polypeptide
source 1..9
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 82
YPYDVPDYA 9

SEQ ID NO: 83 moltype = DNA length = 33
FEATURE Location/Qualifiers
misc_feature 1..33
 note = Synthetic c-MYC epitope tag polypeptide
source 1..33
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 83
gaacaaaagt taatttctga agaagatttg gaa 33

SEQ ID NO: 84 moltype = AA length = 10
FEATURE Location/Qualifiers
REGION 1..10
 note = Synthetic c-MYC epitope tag polypeptide
source 1..10
 mol_type = protein

-continued

```

organism = synthetic construct

SEQUENCE: 84
EQKLISEEDL                                     10

SEQ ID NO: 85      moltype = DNA  length = 24
FEATURE            Location/Qualifiers
misc_feature        1..24
                    note = Synthetic FLAG epitope tag polypeptide
source              1..24
                    mol_type = other DNA
                    organism = synthetic construct

SEQUENCE: 85
gactacaagg atgacgatga caaa                      24

SEQ ID NO: 86      moltype = AA  length = 8
FEATURE            Location/Qualifiers
REGION              1..8
                    note = Synthetic FLAG epitope tag polypeptide
source              1..8
                    mol_type = protein
                    organism = synthetic construct

SEQUENCE: 86
DYKDDDDK                                           8

SEQ ID NO: 87      moltype = DNA  length = 42
FEATURE            Location/Qualifiers
misc_feature        1..42
                    note = Synthetic V5 epitope tag polypeptide
source              1..42
                    mol_type = other DNA
                    organism = synthetic construct

SEQUENCE: 87
ggtaagccaa ttccaaatcc ttgttgagggt ttgactcca cc    42

SEQ ID NO: 88      moltype = AA  length = 14
FEATURE            Location/Qualifiers
REGION              1..14
                    note = Synthetic V5 epitope tag polypeptide
source              1..14
                    mol_type = protein
                    organism = synthetic construct

SEQUENCE: 88
GKPIPNPLLG LDST                                   14

SEQ ID NO: 89      moltype = DNA  length = 18
FEATURE            Location/Qualifiers
misc_feature        1..18
                    note = Synthetic HIS epitope tag polypeptide
source              1..18
                    mol_type = other DNA
                    organism = synthetic construct

SEQUENCE: 89
catcatcatc atcatcat                               18

SEQ ID NO: 90      moltype = AA  length = 6
FEATURE            Location/Qualifiers
REGION              1..6
                    note = Synthetic HIS epitope tag polypeptide
source              1..6
                    mol_type = protein
                    organism = synthetic construct

SEQUENCE: 90
HHHHHH                                           6

SEQ ID NO: 91      moltype = DNA  length = 7882
FEATURE            Location/Qualifiers
misc_feature        1..7882
                    note = pMM1 vector
source              1..7882
                    mol_type = other DNA
                    organism = synthetic construct

SEQUENCE: 91
caggtggcac ttttcgggga aatgtgcgcg gaaccacctat ttgtttatct ttctaaatac 60
attcaaatat gtatccgctc atgagacaat aacctgata aatgcttcaa taatattgaa 120
aaaggaagag tatgagtatt caacatttcc gtgtcgccct tattcccttt ttgctggcat 180
tttgccctcc tgtttttgct caccagaaa cgctgggtgaa agtaaaagat gctgaagatc 240
agttgggtgc acgagtgggt tacatcgaa tgatctcaa cagcggtaag atccttgaga 300
gttttcgccc cgaagaacgt ttccaatga tgagcacttt taaagttctg ctatgtggcg 360
cggattatc ccgatttgac gccgggcaag agcaactcgg tcgccgcata cactattctc 420

```

-continued

agaatgactt	ggttgagtag	tcaccagtc	cagaaaagca	tcttacggat	ggcatgacag	480
taagagaatt	atgcagtgct	gccataacca	tgagtataaa	cactgcggcc	aacttacttc	540
tgacaacgat	cggaggaccg	aaggagctaa	ccgctttttt	gcacaacatg	ggggatcatg	600
taactcgct	tgatcggttg	gaaccggagc	tgaatgaagc	cataccaaac	gacgagcgtg	660
acaccacgat	gcctgtagca	atggcaacaa	cggtgcgcaa	actattaact	ggcgaactac	720
ttactctagc	ttcccgcaa	caattaatag	actggatgga	ggcggataaa	gttgacggac	780
cactcttcgc	ctcgccctct	ccggctggct	ggtttattgc	tgataaatct	ggagccgggtg	840
agcgtgggtc	tcgcggtatc	attgcagcac	tggggccaga	tggtaaagcc	tcccgatcgc	900
tagttatcta	cacgacgggg	agtccaggca	ctatggatga	acgaaataga	cagatcgctg	960
agatagggtc	ctcactgatt	aagcattggg	aactgtcaga	ccaagtttac	tcatatatac	1020
tttagattga	tttaaaactt	catttttaat	ttaaaaggat	ctagggtgaag	atcctttttg	1080
ataatctcat	gaccaaaatc	ccttaacgtg	agttttcgtt	ccactgagcg	tcagaccccg	1140
tagaaaagat	caaaggatct	cttgagatc	ctttttttct	gcgcgtaatc	tgctgcttgc	1200
aaacaaaaaa	accaccgcta	ccagcgggtg	ttgttttgcc	ggatcaagag	ctaccaactc	1260
tttttcggaa	ggtaactggc	ttcagcagag	cgcagatacc	aaataactgt	cttctagtgt	1320
agccgtagtt	aggccaccac	ttcaagaact	ctgtagcacc	gcctacatac	ctcgtcttgc	1380
taatcctgtt	accagtggtc	gctgccagtg	gcgataagtc	gtgtcttacc	gggttggact	1440
caagacgata	gttaccggat	aaggcgcagc	ggtcgggctg	aacggggggt	tcgtgcacac	1500
agcccgactt	ggagcgaaag	acctacaccg	aactgagata	cctacagcgt	gagctatgag	1560
aaagcgccac	gcttcccga	gggagaaaag	cggacaggtg	tccggtaagc	ggcagggctg	1620
gaacaggaga	gcgcacgagg	gagcttcocag	ggggaaacgc	ctggtatctt	tatagtcctg	1680
tcgggtttcg	ccacctctga	cttgagcgtc	gatttttgtg	atgctcgtca	ggggggcgga	1740
gacctatggaa	aaacgcagc	aacgcggcct	ttttacgggt	cctggccttt	tgctggcctt	1800
ttgctcacat	gttctttcct	gcgttatccc	ctgattctgt	ggataaacgt	attaccgcct	1860
ttgagttagc	tgataccgct	gcgcgagcc	gaacgaccca	gcgcgagcag	tcagttagcg	1920
aggaagcgga	agagcgccca	atcgcacaaa	cgctctccc	cgcgcttggt	ccgattcatt	1980
aatgcagctg	gatcttcgag	cgctccaaaa	ccttctcaag	caaggttttc	agtataatgt	2040
tacatgcgta	cacgcgtctg	tacagaaaaa	aaagaaaaat	tgaaatata	aataacgttc	2100
ttataactaa	cataactata	aaaaaataaa	tagggaccta	gacttcaggt	tgtctaaact	2160
cttcttttct	ggtttagagc	gatcttagct	agcccgcgta	ccaagctggt	ggatcctttg	2220
ttgttttatg	tgtgtttatt	gaaactaag	ttcttggtgt	tttaaaacta	aaaaaaagac	2280
taactataaa	agtagaat	aagaaagtta	agaaatagat	ttacagaatt	acaatcaata	2340
cctaccgctc	ttatataact	attagtcag	taggggaata	atttcaggga	actggtttca	2400
accttttttt	tcagcttttt	ccaaatcaga	gagagcagaa	ggtaatagaa	ggtgtaagaa	2460
aatgagatag	atacatcgct	gggtcaattg	ccttgtgtca	tcattttact	caggcaggtt	2520
gcatcactcc	attgaggttg	tgcccgtttt	ttgctgtttt	gtgcccctgt	tcctctgtagt	2580
tgcgctaaga	gaattggacct	atgaactgat	ggttggtgaa	gaaaacaata	ttttggtgct	2640
gggattctct	ttttttctgg	atgccagctt	aaaaagcggg	ctccattata	tttagtggat	2700
gccaggaata	aactgtgtgt	tgcaaaaaa	acaaaaactg	aaaaaccag	acacgctcga	2760
cttctgtctc	tcctattgtc	tgcaagcttc	aatttcgtca	cacaacaagg	tcctagcgac	2820
ggctcacagg	ttttgttaaa	agcaatcgaa	ggttctggaa	tgccgggaaa	gggttttagta	2880
ccacatgcta	tgatgcccac	tgtgatctcc	agagcaaaag	tcgttcgatc	gtactgttac	2940
ttctctctct	tcaaacgaaa	ttgtcgaat	cgtgtgacaa	caacagcctg	ttctcacaca	3000
ctcttttctt	ctaaccagg	gggtgggtta	gttttagtag	acctcgtgaa	acttacattt	3060
acatatatat	aaacttgcat	aaattgggtc	atgcaagaaa	tacatatatt	gtcttttcta	3120
attcgtagtt	tttcaagttc	tttagtgctt	ttttttctct	ttttttacag	atcatcaagg	3180
aagttaattat	ctacttttta	caacaaatat	aaaacagcgg	ccgcaaaaga	tcctgagctct	3240
taatttaacaa	ttcttcgcga	gaggttttgt	caagtctcca	atcaagggtg	tcggcttgtc	3300
taccttgcca	gaaattttacg	aaagatgga	aaagggtcaa	atcgttggtg	gatacgttgt	3360
tgacacttct	aaataagcga	atttcttatg	attttatgatt	ttttattata	aataagttat	3420
aaaaaaaata	agtgatatac	aattttaaag	tgactcttag	gttttaaaac	gaaaattctt	3480
attcttgagt	aactctttcc	tgtaggtcag	gttgctttct	caggtatagc	atgaggtcgc	3540
tccaattcag	ctggcgtaat	agcgaagagg	ccgcacccga	tcgcccctcc	caacagttgc	3600
gcagcctgaa	tgccgaattc	gacactggat	ggcggcgcta	gtatcgaatc	gacagcagta	3660
tagcgaccac	cattcacata	catttagacgc	atgatattac	tttctgcgca	cttaacttcg	3720
catctgggca	gatgatgtcg	aggcgaaaaa	aaatataaat	cacgctaaca	tttgattaaa	3780
atagaacaac	tacaatatata	aaaaactata	caaatgacaa	gttcttgaaa	acaagaatct	3840
ttttattgtc	agtactgatt	agaaaaactc	atcagcgcac	aaatgaaact	gcaatttatt	3900
catatcagga	ttatcaatac	catatttttg	aaaaagccgt	ttctgtaatg	aaggagaaaa	3960
ctcaccgagg	cagttccata	ggatggcaag	atcctgggtat	cggtctcgca	ttccgactcg	4020
tccaacatca	atacaacctc	tttaatttcc	ctcgtcaaaa	ataagggtat	caagtggaga	4080
atcaccatga	gtgacgactg	aatccgggtg	gaatggcaaa	agcttatgca	tttctttcca	4140
gacttggtca	acaggccagc	cattacgctc	gtcatcaaaa	tcactcgcat	caaccaaacc	4200
gttattcatt	cgtgatctgc	cctgagcgag	acgaaatacg	cgatcgctgt	taaaaggaca	4260
attacaaaca	ggaatcgaat	gcaaccggcg	caggaaactc	gccagcgcat	caacaataat	4320
ttcactgaa	tcaggatatt	cttctaatac	ctggaatgct	gttttgccgg	ggatcgagct	4380
ggtgagtaac	catgcacatc	caggagtagc	gataaaatgc	ttgatggctg	gaagaggcat	4440
aaattccgtc	agccagttta	gtctgacctc	ctcatctgta	acatcattgg	caacgctacc	4500
tttgccatgt	ttcagaaaac	actctggcgc	atcgggcttc	ccatacaatc	gatagattgt	4560
cgccactgat	tgcccgacat	tatcgcgagc	ccattttatac	ccatataaat	cagcatccat	4620
gttggaaatt	aatcgcgggc	tcgaaaacgtg	agtcttttcc	ttacccatgg	ttgtttatgt	4680
tcggatgtga	tgtgagaact	gtatcctagc	aagattttta	aaggaaagtat	atgaaagaag	4740
aacctcagtg	gcaaatccta	acctttttata	ttctctacac	ggggcgcggc	gtggggacaa	4800
ttcaacgcgt	ctgtgagggg	agcgttttcc	tgctcgcagg	tcgcagcgca	ggagccgtaa	4860
tttttgcttc	gcgcgctggc	gccatcaaaa	tgatgggatg	caaatgatta	tacatgggga	4920
tgatgggct	aaatgtacgg	gcgacagtc	catcatgccc	ctgagctgcg	cacgtcaaga	4980
ctgtcaagga	gggtattctg	ggcctccatg	tcgctggccg	ggtgaccccg	cggggacgag	5040
gcaagctaaa	cagatctggc	gcgtgtactg	agagtgcacc	ataaattccc	gttttaagag	5100
cttggtgagc	gctaggagtc	actgccaggt	atcgtttgaa	cacggcatta	gtcagggaag	5160

-continued

```

tcataaacaca gtccttttccc gcaatttttct ttttctatta ctcttggect cctctagtag 5220
actctatatt tttttatgcc tcggtaaatga ttttctattt tttttttcca cctagcggat 5280
gactcttttt tttttctagc gatttgcoatt atcacataat gaattatata ttatataaag 5340
taatgtgatt tcttcgaaga atatactaaa aaatgagcag gcaagataaa cgaaggcaaa 5400
gatgacagag cagaaaagccc tagtaaaagcg tattacaaat gaaaccaaga ttcagattgc 5460
gatctcttta aagggtgggc ccctagcgat agagcactcg atcttcccag aaaaagaggc 5520
agaagcagta gcagaacagg ccacacaatc gcaagtattt aacgtccaca caggtatagg 5580
gtttctggac catatgatac atgctctggc caagcattcc ggctgggtgc taatcggtga 5640
gtgcattggg gacttacaca tagacgacca tcacaccact gaagactcgc ggattgctct 5700
cgggtcaagct tttaaagagg ccctaggggc cgtgcgtgga gtaaaaaggt ttggatcagg 5760
atttgcgcct ttggatgagg cactttccag agcgggtgga gatctttcga acaggcgcta 5820
cgagttgttc gaacttggtt tgcaaaagga gaaagtagga gatctctctt gcgagatgat 5880
cccgcatttt cttgaaagct ttgcagaggc tagcagaatt accctccacg ttgattgtct 5940
gcagggcaag aatgatcatc accgtagtga gagtgcgttc aaggctcttg cggttgccat 6000
aagagaagcc accctcgcca atggtagcaa cgatgttccc tccaccaag gtgttcttat 6060
gtagtgcac cgattattta aagctgcagc atacgatata tatcatgtg tatatatgta 6120
tacctatgaa tgtcagtaag tatgtatacg aacagtatga tactgaagat gacaaggtaa 6180
tgcatcattc tatcgtgtc attctgaacg aggcgcgctt tcttttttct tttttgcttt 6240
ttcttttttt ttctcttgaa ctgcacggat ctatgcggtg tgaaatatgg tgcactctca 6300
gtacaatctg ctctgatgcc gcatagttaa gccagccccg acacccgcca acacccgctg 6360
acgcgccttg acgggcttgt ctgctccccg catccgctta cagacaagct gtgacgctct 6420
ccgggagctg catgtgtcac aggttttcac cgtcatcacc gaaacgcgcg agacgaaagg 6480
gcctcggtat acgctattt ttatagggtt atgtcatgat aataatgggt tcttagtatg 6540
atccaatatc aaaggaaatg atagcattga aggatgagac taatccaatt gaggagtggc 6600
agcatataga acagctaaag ggtagtgtg aaggaagcat acgatcccc gcattggaatg 6660
ggataaatatc acaggagggtt ctgactacc ttctatccta cataaataga cgcataataa 6720
tacgatttta agcataaaca cgcactatgc cgttcttctc atgtatatat atatacaggc 6780
aacacgcaga tatagggtgc agctgaacag tgagctgtat gtgcgcagct cgcgttgcat 6840
tttcgggaagc gctcgttttc ggaaacgctt tgaagttcct attccgaagt tctatttctc 6900
tagaaagtat aggaacttca gagcgtttt gaaaaccaa agcgtctctg agacgcactt 6960
tcaaaaaacc aaaaacgcac cggactgtaa cgagctacta aaatatgctg aataccgctt 7020
ccacaaacat tgctcaaaag tatctctttg ctatatatct ctgtgctata tccctatata 7080
acctaccatc ccactcttcg ctcttgaaac ttgcatctaa actcgacctc tacatttttt 7140
atgtttatct ctagtattac tctttagaca aaaaaattgt agtaagaact attcatagag 7200
tgaatcgaaa acaatacgaa aatgtaaaaca ttctctatct gtagtatata gagacaaaat 7260
agaagaaacc gttcataatt ttctgaccaa tgaagaatca tcaacgctat cactttctgt 7320
tcacaaagta tcgcgcaatc acatcggtat agaataaat cggggatgcc tttatcttga 7380
aaaaatgcac ccgcagcttc gctagtaatc agtaaacgcg ggaagtggag tcaggctttt 7440
tttatggaag agaaaataga caccaaagta gccttctctt aaccttaacg gacctacagt 7500
gcacaaaagtt atcaagagac tgcattatag agcgcacaaa ggagaaaaaa agtaattctaa 7560
gatgctttgt tagaaaaata gcgctctcgg gatgcatttt tgtagaacaa aaaagaagta 7620
tagattcttt gttggtaaaa tagcgtctct cgcgttgcat tctgttctgt aaaaatgcag 7680
ctcagattct ttgtttgaaa aattagcgct ctgcggttgc atttttgttt tacaaaaatg 7740
aagcacagat tcttcgttgg taaaatagcg ctttcgctgt gcatttctgt tctgtaaaaa 7800
tgacgctcag attcttgggt tgaaaaatta gcgctctcgc gttgcatttt tgttctacaa 7860
aatgaagcac agatgcttcg tt 7882

```

```

SEQ ID NO: 92      moltype = DNA length = 3810
FEATURE           Location/Qualifiers
misc_feature       1..3810
                   note = pCDM4 vector
source            1..3810
                   mol_type = other DNA
                   organism = synthetic construct

```

```

SEQUENCE: 92
gctaacacgcg cgatttgcgt gtgacccaat gcgaccagat gctccacgcc cagtgcgcta 60
ccgtcttcat gggagaaaaa aatactgttg atgggtgtct ggtcagagac atcaagaaat 120
aacgccggaa cattagtgtca ggcagcttcc acagcaatgg catcctggtc atccagcgga 180
tagttaatga tcagccact gacgcgttgc gcgagaagat tgtgcaccgc cgttttacag 240
gcttcgacgc cgcttcgttc taccatcgac accaccacgc tggcaccagg ttgatcggcg 300
cgagatttaa tcgcccgcac aatttgcgac ggccggtgca gggccagact ggaggtggca 360
acgccaatca gcaacgactg tttgcccggc agttgttgtg ccacgcgggtt gggaaatgta 420
ttcagctccg ccacgcgcgc ttccactttt tcccgcgttt tcgcagaaac gtggctggcc 480
tggttcacca cgcgggaaac ggtctgataa gagacacggc ctactctgc gacatcgta 540
aacgttactg gtttcacatt caccaccctg aattgactct cttccggggc ctatcatgcc 600
ataccgcgaa aggttttgcg ccattcgatg gtgtccggga tctcgacgct ctcccttatg 660
cgactcctgc attaggaagc agcccagtag taggttgagg ccgttgagca ccgccgcgcg 720
aaggaatggg gcatgcaagg agatggcgcc caacagtccc ccggccacgg ggcttgcac 780
cataccacgc ccgaaacaag cgctcatgag cccgaagtgg cgagcccgat cttccccatc 840
ggatgatgct gcgatatagg cgcacgcaac cgcacctgtg gcgcgggtga tgcggccac 900
gatgcgtccg gcgtagccta ggatcgagat cgatctcgat cccgcgaaat taatacgact 960
cactataggg gaattgtgag cggataacaa ttccccctta gaaataattt tgtttaactt 1020
taagaaggag atatacatat gccagatctc aattggatat cggccggcca cgcgatcgct 1080
gacgtcggtg cctcgcagtc ttgtaaaaga accgctgctg cgaatttga acgccagcac 1140
atggactcgt ctactagtgc cagcttaatt aacctaaact gctgccaccg ctgagcaata 1200
actagcataa ccccttgggg cctctaaccg ggtcttgagg ggttttttgc tagcgaaagg 1260
aggagtgcac actgcttccg gtatgtaata aaccggtaaa ccagcaatag acataagcgg 1320
ctatttaacg acctcgccct gaaccgacga ccgggtcacc gtggccggat cttgcggccc 1380
ctcggcttga acgaattggt agacattatt tgccgactac cttggtgatc tcgcttttca 1440

```

-continued

```

cgtagtggac aaattcttcc aactgatctg cgcgcgagge caagegatct tcttcttgtc 1500
caagataaagc ctgtctagct tcaagtatga cgggctgata ctgggcccgc aggcgctcca 1560
ttgcccagtc ggcagcgaca tccttcggcg cgattttgcc ggttactcgc ctgtacccaa 1620
tgccgggacaa cgtaagcact acatttcgct catcgccagc ccagtcgggc ggcgagttcc 1680
atagcggttaa gggtttcattt agcgctcaa atagatcctg ttcaggaacc ggatcaaaaga 1740
gttctctccgc cgctggacct accaaggcaa cgctatgttc tcttgccttt gtccagcaaga 1800
tagccagatc aatgtcgatc gtggctggct cgaagatacc tgcaagaatg tcattgcgct 1860
gccattctcc aaattgcagt tcgcgcttag ctggataaac ccacggaatg atgtcgtcgt 1920
gcacacaat ggtgacttct acagcgcgga gaattctcgt ctctccaggg gaagccgaag 1980
tttccaaaag gtctgtgac aaagctcgcc gcgttgttcc atcaagcctt acggtcaccg 2040
taaccagcaa atcaatatca ctgtgtggct tcaggccgcc atccactcgc gagccgtaca 2100
aatgtacggc cagcaacgct gggtcgagat ggcgctcgat gacgccaaact acctctgata 2160
gttgagtcga tacttcggcg atcacgcctt ccctcatact ctctcttttt caatattatt 2220
gaagcattta tcaggggttat gtctctatga gcggatacat atttgaatgt atttagaaaa 2280
ataaacaat agccagctca ctcggtcgct acgctccggg cgtgagactg cgcggggcgc 2340
tgccgacaca taccaaagtta ccacagatt ccgtggataa gcaggggact aacatgtgag 2400
gcaaaacagc agggcgcgcc cggtggcggt ttccatagg ctccgctctc ctgccagagt 2460
tcacataaac agacgctttt ccggtgcac tcgtgggagcc gtgaggctca accatgaatc 2520
tgacagtacg ggcgaaaccc gacaggactt aaagatcccc accgtttccg gcgggtcgct 2580
ccctcttgcg ctctcctggt ccgaccctgc cgtttaccgg atacctgttc cgcctttctc 2640
ccttacggga agtgtggcgc tttctcatag ctccacact ggatatctcg ctcggtgtag 2700
gtcgttcgct ccaagctggg ctgtaagcaa gaactcccc ttcaggccga ctgctgcgc 2760
ttatccggta actgttcaact tgagtccaac ccggaagcgc acggtaaaaa gccactggca 2820
gcagccattg gtaactggga gttcgagag gatgtgtta gctaaacacg cggttgtctc 2880
tgaagtgtgc gccaaagtcc ggctacactg gaaggacaga tttgggtgct gtgctctgcg 2940
aaagccaggtt accaggttta agcagttccc caactgactt aaccttcgat caaacacact 3000
ccccaggtgg ttttttcggt tacagggcaa aagattacgc gcagaaaaaa aggatctcaa 3060
gaagatcctt tgatcttttc tactgaaccg ctctagattt cagtgcattt tatctcttca 3120
aatgtagcac ctgaagtcag ccccatcaga tataagttgt aattctcatg ttagtcatgc 3180
cccgcgccca ccggaaggag ctgactgggt tgaaggctct caaggcgatc ggtagagatc 3240
ccggtgccta atcagtgagc taacttatc taattgcgtt gcgctcactg cccgctttcc 3300
agtcgggaaa ctgtcgtgac cagctgcatt aatgaatcgg ccaacgcgcg gggagaggcg 3360
gtttgcgtat tggcgccag ggtggttttt cttttccaca gtgagcggg caacagctga 3420
ttgcccttca ccgctggcc ctgagagagt tgcagcaagc ggtccacgct ggtttgcccc 3480
agcaggcgaa aatcctgttt gatgggtggt aacggcgga tataacatga gctgtcttcg 3540
gtatcgtcgt atccactac cgagatgtcc gcaccaacgc gcagcccgga ctcggtaatg 3600
gcgcgcatgt gcgccagcgc catctgatcg ttggcaacca gcctgcaggt gggaaacgatg 3660
ccctcattca cgtattgcat ggtttgttga aaaccggaca tggcactcca gtcgcttcc 3720
cgttccgcta tcgctgaat ttgattgcca gtgagatatt tatgccagcc agccagacgc 3780
agacgcgcgc agacagaact taatgggccc 3810

```

```

SEQ ID NO: 93      moltype = DNA length = 5203
FEATURE            Location/Qualifiers
misc_feature        1..5203
                    note = pETM6 vector
source              1..5203
                    mol_type = other DNA
                    organism = synthetic construct

```

```

SEQUENCE: 93
ggggaattgt gagcgataa caattcccct ctagaataaa ttttgtttaa ctttaagaag 60
gagatataca tatggcagat ctcaatttga tatcgccgg ccacgcgatc gctgacgtcg 120
gtaccctcga gtcctggttaa gaacacgcgt ctgcgaaatt tgaacgccag cacatggact 180
cgtctactag tgcagcttta attaacctaa actgctgcca ccgctgagca ataactagca 240
taacccttcg ggcctctaa acgggtcttg aggggttttt tgctagcgaa aggaggagtc 300
gactatctc ggattggcga atgggacgcg cctgtagcgc gcgcattaa gcgcggcggt 360
gtgggtggtta gcgcgacgct gaccgctaca cttgccagcg ccctagcgcc cgtctctttc 420
gctttcttcc ctctctttct cgcacgcttc gccggttttc ccgctcaagc tctaaatcgg 480
gggctccctt tagggttcgc atttagtgct ttacggcacc tcgaccccaa aaaacttgat 540
taggggtgatg gttcacgtag tgggcccacg cctgtagata cgggttttctg ccctttgacg 600
ttggagtcca cgttctttaa tagtggaact ttgttccaaa ctggaacaa actcaacct 660
atctcggtct attcttttga ttataaggg attttgcga tttcgcccta ttgggtaaaa 720
aatgagctga ttttaacaaa atttaacgcg aattttaaca aaatattaac gtttacaatt 780
tctggcgcca cgatggcatg agattatcaa aaaggatctt cacttagatc cttttaaatt 840
aaaaatgaag ttttaaatca atctaaagta tatatgagta aacttggtct gacagttacc 900
aatgcttaat cagtgaaggca cctatctcag cgatctgtct atttcgttca tccatagttg 960
cctgactccc cgtcgtgtag ataactacga tacgggaggg cttaccatct ggcgccagtg 1020
ctgcaatgat accgcgagac ccacgctcac cggctccaga tttatcagca ataaaccagc 1080
cagccggaag gcccgagcgc agaagtggtc ctgcaacttt atccgctcc atccagctca 1140
ttaattgttg ccgggaagct agagtaagta gttcgccagt taatagtttg cgcaacggtg 1200
ttgccattgc tacaggcatc gtgggtgtcac gctcgtcgtt tggtaggtgt tcattcagct 1260
ccggttccca acgatcaagg cgagttacat gatccccat gttgtgcaaa aaagcgggtta 1320
gctccttcgg tctctcgatc gttgtcagaa gtaagttggc cgcagtggtta tctactatgg 1380
ttatggcagc actgcataat tctcttactg tcatgccatc cgtaagatgc ttttctgtga 1440
ctgggtgagta ctcaaacaa aatctctgag gcggcgacgc agttgctctt 1500
gcccggcgtc aatacgggat aataccgcgc cacatagcag aactttaaaa gtgctcatca 1560
ttgaaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg agatccagtt 1620
cgatgtaac cactcgtgca cccaactgat cttcagcatc ttttactttc accagcggtt 1680
ctgggtgagc aaaaacagga aggcataaat ccgcaaaaaa gggaataagg gcgacacgga 1740
aatgttgaat actcatactc ttctttttt aatcatgatt gaagcattta tcagggttat 1800

```

-continued

```

tgtctcatga gcgatacat attgaaatgt atttagaaaa ataaacaaat aggtcatgac 1860
caaaatccct taacgtgagtt ttctgttcca ctgagcgta gaccccgtag aaaagatcaa 1920
aggatottct tgagatcctt tttttctgcg cgtaatctgc tgcttgcaaa caaaaaaacc 1980
accgctacca gcggtgggtt gtttgccgga tcaagagcta ccaactcttt ttccgaaggt 2040
aactggcttc agcagagcgc agataccaaa tactgtcctt ctagtgtagc cgtagttagg 2100
ccaccacttc aagaactctg tagcacccgc tacataacct gctctgctaa tctgtttacc 2160
agtggctgct gccagtggcg ataagtctgt tcttaccggg ttggactcaa gacgatagtt 2220
accggataag gcgagcggtt cgggctgaac ggggggttcg tgcacacagc ccagcttgga 2280
gcgaacgacc tacaccgaac tgagatacct acagcgtgag ctatgagaaa ggcaccgct 2340
tcccgaaggg agaagggcgg acaggtatcc ggtaagcggc agggctcgaa caggagagcg 2400
cacgagggag cttccagggg gaaacgcctg gtatctttat agtctgtcgc ggtttcgcca 2460
cctctgactt gagcgtcgat ttttgtgatg ctgctcaggg gggcggaacc tatggaaaaa 2520
cgccagcaac gcgccctctt tacggttctt ggccttttgc tggccttttg ctcacatggt 2580
ctttctgcgc ttatccctgt attctgtgga taaccgtatt accgcctttg agtgagctga 2640
taccgctcgc cgcagccgaa cgaccgagcg cagcgagta gtgagcgagg aagcggaaga 2700
gcgctgatg cggatatttc tcttaacgca tctgtgcggt atttcacac gcataatagg 2760
tgactctca gacaactctg cctctgatgc gcatagttaa gccagtatc actccgctat 2820
cgctacgtga ctgggtcatg gctgcgcccc gacacccgcc aacacccgct gacgcgccct 2880
gacgggcttg tctgtcccg gcatccgctt acagacaagc tgtgaccgtc tccgggagct 2940
gcatgtgtca gctgttttca ccgtcatcac cgaacgcgc gaggcagctg cggtaaaagt 3000
catcagcgtg gtcgtgaagc gattcacaga tctctgcctg ttcacccgcg tccagctcgt 3060
tgagtttctc cagaagcgtt aatgtctggc ttctgataaa gcgggccatg ttaagggcgg 3120
ttttttctgc ttgggtcact gatgcctcgc tgtaaggggg atttctgttc atgggggtaa 3180
tgataccgat gaaacgagag aggatgctca cgatacgggt tactgatgat gaacatgccc 3240
ggttactgga acgttgtgag ggtaaacaaac tggcggtatg gatcgccgg gaccagagaa 3300
aaatcactca ggggtcaatgc cagcgcttcg ttaatacaga ttaggtgtt ccacagggta 3360
gccagcagca tctctgcgat cagatccgga acataatggt gcagggcgct gacttccgcg 3420
tttcagact ttacgaacga cggaaaccca agaccattca tgttgttgct caggtcgca 3480
acgtttttgca gcagcagctg cttcacgttc gctcgcgtat cggtgattca ttctgctaac 3540
cagtaaggca accccgcag cctagccggg tctcaacga caggagcacg atcatgctag 3600
tcataccccc cgcccacggg aaggagctga ctgggttgaa ggctctcaag ggcacggtc 3660
gagatccccc tgctaatga gtgagctaac ttacattaat tgcgttgccg tcactgccc 3720
ctttccagtc gggaaacctg tctgtccagc tgcattaatg aatcgccaa cgcgcgggga 3780
gaggcggttt cgtattggg gccagggtg gtttttctt taccagtgga gacgggcaac 3840
agctgattgc cttcacccgc ctggccctga gagagttgca gcaagcggtc cacgtgggt 3900
tgcccagca gggcaaaact ctgtttgatg gtggttaacg gcgggataa acatgagctg 3960
tcttcggtat gctcgtatcc cactacagag atgtccgcac caacgcgcag cccggaactc 4020
gtaattggcg cgtattgcgc cagcgcctat tgatcgttgg caaccagcat cgcagtgga 4080
acgatgcctt cattcagcat ttgcatggtt ttgtgaaaa cggacatggc actccagtcg 4140
ccttccggtt ccgctatcgg tgaatttga ttgcgagtga gatatttatg ccagccagcc 4200
agacgcagac ggcgcgagac agaaactaat gggcccgcga acagcgcgat ttgctggtga 4260
cccaatgcga ccagatgctc cagcccaggt cgcgtaccgt cttcatggga gaaaaataa 4320
ctgttgatgg gtgttggtgc agagcatca agaaataacg ccggaacatt agtgaggga 4380
gcttccacag caatggcatc ctgggtcatc agcggatagt taatgatcag cccactgacg 4440
cggtgcgcga gaagattgtg caccgcgcgt ttacaggctt cgacgcgcgt tctgtctacc 4500
atcgacacca ccacgctggc acccagttga tccgcccagc atttaatcgc cgcgacaatt 4560
tgccagggcg cgtgcagggc cagactggag gtggcaacgc caatcagcaa cgaactgttg 4620
cccgccaggt gttgtgccc cgggttgga atgtaattca gctccgcct cgcgccttc 4680
actttttccc gcgttttcgc agaaacgtgg ctggcctggt taccacgcgc ggaacggtc 4740
tgataagaga caccggcata ctctgcgaca tctgtataacg ttactggttt cacattcacc 4800
accctgaatt gactctcttc cgggctctat catgccatc cgcgaaaggt ttgcccctat 4860
tcgatgggtg ccgggacttc gacgctctcc cttatgcgac tctgcatga ggaagcagcc 4920
cagtagtagt ttgaggcgtg tagcaccgc cgcgcgaagg aatgggtgat gcaaggagat 4980
ggcgcccaac agtccccgc ccacggggcc tgccaccata cccacgcga aacaagcgt 5040
catgagccgc aagtggcgag cccgatcttc cccatcggtg atgtgcgcga tataggccgc 5100
agcaaccgca cctgtggcgc cggtagtgcc ggccacgatg cgtccggcgt agcctaggat 5160
cgagatcgat ctcgatcccg cgaattaat acgactcact ata 5203

```

```

SEQ ID NO: 94      moltype = DNA length = 5369
FEATURE            Location/Qualifiers
misc_feature        1..5369
                    note = pET28a(+) vector
source              1..5369
                    mol_type = other DNA
                    organism = synthetic construct

```

```

SEQUENCE: 94
atccggatat agttcctcct ttcagcaaaa aaccctctca gaccggttta gagggcccaa 60
gggggttatgc tagttattgc tcagcgggtg cagcagccaa ctacagcttc ttctgggctt 120
tgtagcagc cggatctcag ttgtgtgtgt ggtgtgtgct gagtgccggc gcaagcttgt 180
cgacggagct cgaattcgga tccgcgaccc atttgcgtgc caccagtcac gctagccata 240
tggttgccgc gcggcaccag gccgctgctg tgatgatgat gatgatggct gctgcccatt 300
gtatatctcc ttcttaaagt taacaaaaat tatttctaga ggggaattgt tatccgctca 360
caattccctt atagttagtc gtattaattt cgcgggtagc agatctcgat cctctacgcc 420
ggagcgcagc tggcggcgtt caccggcgcc acaggtgcgc ttgctggcgc ctatatcgcc 480
gacatcaccg atgggggaaga tcgggctcgc cacttcgggc tcatgagcgc ttgtttcggc 540
gtgggtatgg tggcaggccc cgtggccggg ggaactgttg gcgccatctc cttgcatgca 600
ccattccttg cggcggcggt gctcaacggc ctcaacctac tactgggctg cttcctaagt 660
caggagtcgc ataaggaga gcgtcgagat cccggacacc atcgaatggc gcaaacctt 720
tcgcggtatg gcatgatagc gcccggaaga gagtcaattc aggggtgtga atgtgaacc 780

```

-continued

agtaacgtta	tacgatgtcg	cagagtatgc	cgggtgtctct	tatcagaccg	tttcccgcgt	840
ggtgaaccag	gccagccacg	tttctgcgaa	aacgcgggaa	aaagtggaa	cgccgatggc	900
ggagctgaat	tacattccca	accgcgtggc	acaacaactg	gcgggcaaac	agtcgttgct	960
gattggcggt	gccacctcca	gtctggccct	gcacgcgcgc	tcgcaaattg	tcgcggcgat	1020
taaatctcgc	gccgatcaac	tgggtgccag	cgtgggtggg	tcgatggtag	aacgaagcgg	1080
cgctcgaagcc	tgtaaagcgg	cgggtgcaca	tcttctcgcg	caacgcgtca	gtgggctgat	1140
cattaactat	ccgctggatg	accaggatgc	cattgtctgtg	gaagctgcct	gcactaatgt	1200
tcggcggtta	tttcttgatg	tctctgacca	gacacccatc	aacagtatta	tttctcccca	1260
tgaagacggg	acgcgactgg	gcgtggagca	tctggtcgca	ttgggtcacc	agcaaatcgc	1320
gctgttagcg	ggccccatta	gttctgtctc	ggcgcgtctg	cgtctggctg	gctggcataa	1380
atatctcact	cgcatacaaa	ttcagccgat	agcggaaacg	gaaggcgact	ggagtgccat	1440
gtccggtttt	caacaaacca	tgcaaatgct	gaatgagggc	atcgttccca	ctgcgatgct	1500
ggttgccaac	gatcagatgg	cgctgggcgc	aatgcgcgcc	attaccgagt	ccgggctgcg	1560
cgctgggtcg	gatactcagg	tagtgggata	cgacgatacc	gaagacagct	catgttatat	1620
cccgcggtta	accaccatca	aacaggattt	tcgctcgtcg	gggcaaacca	gcgtggaccg	1680
cttgctgcaa	ctctctcagg	gccaggcggg	gaaggggcaat	cagctgttgc	ccgtctcact	1740
ggtgaaaaga	aaaaccaccc	tggcgcccaa	tacgcaaac	gcctctcccc	gcgcgttggc	1800
cgattcattt	atgcagctgg	cacgcagagt	ttcccgcactg	gaaagcgggc	agtgcgcgca	1860
acgcaattaa	tgtaaagtta	ctcactcatt	aggcaccggg	atctcgaccg	atgccttga	1920
gagccttcaa	cccagtcagc	tccttccggt	gggcgcgggg	catgactatc	gtcgcgcgac	1980
ttatgactgt	cttctttatc	atgcaactcg	taggacaggt	gccggcagcg	ctctgggtca	2040
ttttcggcga	ggaccgcgtt	gcctggagcg	cgacgatgat	cggcctgtcg	cttgcgggat	2100
tcggaaatctt	gcacgccttc	gctcaagcct	tcgtcactgg	tcgccccacc	aaacgtttcg	2160
gcgagaagca	ggccattatc	gccggcatgg	cggccccacg	ggtgcgcgat	atcgtgctcc	2220
tgctgttag	gaccgcgttc	ggctggcggg	gttgctctac	tggttagcag	aatgaatcac	2280
cgatacgcga	gcgaacgtga	agcgactgct	gctgcaaaac	gtctgcgacc	tgagcaacaa	2340
catgaatggt	cttcggtttc	cgtgtttcgt	aaagtctgga	aacgcggaag	tcagcgccct	2400
gcaccattat	gttcggcgtt	tgcatcgcag	gatgctgctg	gctaccctgt	ggaacacctt	2460
catctgtatt	aacgaagcgc	tgccattgac	cctgagtgat	ttttctctgg	tcgcccgcca	2520
tcctaccgcg	cagttgttta	ccctcacaca	gttccagtaa	ccgggcagtg	tcctcatcag	2580
taaccctgat	cgtgagcatc	ctctctcgtt	tcctcgggat	cattaccccc	atgaacagaa	2640
atccccctta	cacggagggca	tcagtgaacca	aacaggaaaa	aaccgccttc	aaactggccc	2700
gctttatcag	aagccagaca	ttaacgcttc	tggagaaact	caacgagctg	gacgcgggat	2760
aacaggcaga	catctgtgaa	tcgcttcacg	accacgctga	tgagctttac	cgcagctgcc	2820
tcgcgcgttt	cgggtgatgc	ggtgaaaaac	tctgacacat	gcagctcccc	gagacgggtca	2880
cagcttgtct	gtaagcggat	gccgggagca	gacaagcccg	tcagggcgcg	tcagcgggtg	2940
ttggcgggtg	tcggggcgca	gccatgacc	agtcacgtag	cgatagcgga	gtgtatactg	3000
gcttaactat	gcggcatcag	agcagattgt	actgagagtg	caccatatac	gcgggtgtga	3060
ataccgcaca	gatgcgtta	gagaaaaaac	cgcacagcgc	gctcttcctc	ttctcgcctc	3120
actgactcgc	tcgctcgggt	cgttcgggtg	cggcgagcgg	tatcagctca	ctcaaaaggcg	3180
gtaatacggg	tatccacaga	atcaggggat	aacgcaggaa	agaacatgtg	agcaaaaaggc	3240
cagcaaaaagg	ccaggaaacg	taaaaaaggc	cgcttgctgg	cgtttttcca	taggctccgc	3300
ccccctgacg	agcatcagaa	aaatcgacgc	tcaagtcaga	ggtggcgaaa	cccgcacagga	3360
ctataaagat	accaggcggt	tcctccctgga	agctccctcg	tcgcctctcc	tggtcccgacc	3420
ctgcgcgtta	ccggataact	gtccgccttt	ctcccttcgg	gaagcgtggc	gctttctcat	3480
agctcacgct	gtaggtatct	cagttcgggt	taggtcgttc	gctccaaagt	gggctgtgtg	3540
cacgaacccc	cggttcagcc	cgaccgctgc	gccttatccg	gtaactatcg	tcttgagtcc	3600
aaccgcgtta	gacacgactt	atcgccactg	gcagcagcca	ctggttaacag	gatttagcaga	3660
gcgaggtatg	taggcgtgtc	tacagagttc	ttgaagtggg	ggcctaacta	cggctacact	3720
agaaggacag	tatttgggtg	ctgcgctctg	ctgaagccag	ttaccttcgg	aaaaagagtt	3780
ggtagctctt	gatccggcaa	acaaaccacc	gctggtagcg	gtggtttttt	tggtttgcaa	3840
cagcagattt	cgcgcagaaa	aaaaggatct	caagaagatc	ctttgatctt	ttctacgggg	3900
tctgacgctc	agtggaaacga	aaactcacgt	taagggtatt	tggtcatgaa	caataaaaact	3960
gtctgcttac	ataaacagta	atacaagggt	tggttatgag	catattcaac	gggaaaacgtc	4020
ttgctctagg	ccgcgattaa	attccaacat	ggatgctgat	ttatatgggt	ataaatgggc	4080
tcgcgataat	gtcgggcaat	caggtgcgac	aatctatcga	ttgtatggga	agcccgatgc	4140
gccagagtgt	tttctgaaac	atggcaaaag	tagcgttgcc	aatgatgtta	cagatgagat	4200
ggtcagacta	aactggctga	cgaattttat	gcctcttcgg	accatcaagc	attttatccg	4260
tactcctgat	gatgcctggt	tactcaccac	tcgcgacccc	gggaaaacag	cattccagggt	4320
attagaagaa	tatcctgatt	caggtgaaaa	tattgttgat	gcgctggcag	tggtcctgcg	4380
ccggttgcat	tcgatttcctg	tttgaattg	tccttttaac	agcgatcgcg	tatttcgtct	4440
cgtcaggcgg	caatcacgaa	tgaataacgg	tttggttgat	gcgagtgtat	ttgatgacga	4500
gcgtaatggc	tggcctgttg	aacaagtctg	gaaagaaatg	cataaaacttt	tgccattctc	4560
accggattca	gtcgtcactc	atggtgattt	ctcacttgat	aaccttattt	ttgacgaggg	4620
gaaattaata	gggtgtattg	atgttggacg	agtcggaatc	gcagaccgat	accaggatct	4680
tgccatccta	tggaaactgcc	tcggtgagtt	ttctcctcca	ttacagaaac	ggctttttca	4740
aaaaatgggt	attgataaat	ctgatatgaa	taaattgcag	tttcatttga	tgctcgatga	4800
gtttttctaa	gaatttaattc	atgagcggat	acataattga	atgtatttag	aaaaataaac	4860
aaataggggt	tcgcgcgaca	tttccccgaa	aagtgccacc	tgaaattgta	aacgttaata	4920
ttttgttaaa	attcgcgtta	aatttttgtt	aaatcagctc	attttttaac	caataggccg	4980
aaatcgccaa	aatcccttat	aaatcaaaag	aatagaccga	gatagggttg	agtggtgttc	5040
cagtttgtaa	caagagtcca	ctattaaaga	acgtggactc	caacgtcaaa	gggcgaaaaa	5100
ccgtctatca	gggcgatggc	ccactacgtg	aaccatcacc	ctaatacagt	tttttggggg	5160
cgaggtgcgc	taagcacta	aatcggaacc	ctaaaggag	ccccgatatt	agagcttgac	5220
ggggaaagcc	ggcgaacgtg	gcgagaaagg	aagggaagaa	agcgaagga	gcgggcgcta	5280
gggcgctggc	aagtgtagcg	gtcacgctgc	gcgtaaccac	cacccccgcc	gcgcttaagt	5340
cgccgctaca	gggcgcgtcc	cattcgcca				5369

-continued

FEATURE	Location/Qualifiers
source	1..2187 mol_type = genomic DNA organism = <i>Psilocybe cubensis</i>
SEQUENCE: 95	
atggcttcta gttcttcoga tgtcttcgtt ttgggtctag gtgttgtttt ggctgccttg 60	
tatatcttca gagaccaatt attcgtctgt tctaagccaa aggtggctcc agtttccact 120	
acgaagcctg ccaacggttc cgctaacca agagacttca tcgccaagat gaaacaaggt 180	
aagaagagaa tcgtaattct ctacgggtct caaactggta ccgctgaaga atatgctatt 240	
cgtttggtta aggaagctaa gcaaaagtgc ggtctagcct ccttggtttg tgatccagaa 300	
gaatacgtatt ttgaaaagtt ggaccaattg ccagaagatt ctattgcttt cttcgtcgtt 360	
gctacctatg gtgaagggtg acctacagac aacgctgtcc aattgttgca aaacttgcaa 420	
gatgaaagct tcgaattctc ctctggtag agaaagtgtg caggtttgaa gtaagtgtgt 480	
tttggtctgg gtaacaagac ctacgaacat tacaacctca ttgggagaac tgttgacgct 540	
caattggcca agatgggtgc tatcagaatc ggtgaaagag gtgaagggtg tgatgacaag 600	
tccatggaag aagactactt ggaatggaag gatggtatgt gggaagcgtt tgccactgct 660	
atgggtgttg aagaaggtca aggtggtgac tccgtgatt tcgtcgtttc cgaattggaa 720	
tctcaccac cagaaaaggt ttaccaaggt gaattttctg cttagagcttt aacccaaacc 780	
aagggtattc acgacgtcaa gaatcctttt gctgctccaa ttgcggttgc tagagaattg 840	
ttccaatctg ttgtcgatag aaactgtgtc cacgtcgaaat tcaacattga aggcctcgtt 900	
atcacctatc aacacggtga ccacgttggg ttgtggccat tgaatccaga tgttgaagtc 960	
gaacggttgt tgtgtgtttt aggtttagct gaaaagagag atgctgtcat ctccattgaa 1020	
tccttagacc cggttttggc taaggttcca tccccagtc caactactta cgggtgctgtg 1080	
ttgagacact acattgacat ctctgctgtc gccggtagac aaactctggg tactttgtcc 1140	
aaattcgctc caaccccaga agctgaaagct ttcttgagaa acttgaacac taacaaggaa 1200	
gaataccaca acgtcgtcgc taacggttgt ttgaaattgg gtgaaatttt gcaaatcgct 1260	
accggtaacg acattactgt cccaccaact actgccaaac ccaccaaagt gccaatcca 1320	
ttcgacatca ttgtttctgc catcccaga ttgcaaccaa gatactactc tatctcttct 1380	
tccccaaaaa ttatcccaaa caccatccac gctaccgttg ttgtgtctca atacgaaaac 1440	
gttccaaaccg aaccaatccc aagaaaagtg gtttacggtg tcggtagtaa cttctgtgtg 1500	
aatttaaaagt acgctgttaa caaggaaacca gttccataca tcaactcaaaa tggcgaaaca 1560	
agagtcggtg tccccgaata cttgattgct ggtccacgtg gttcttaca gactgaatct 1620	
ttctacaagg ctccaatcca tgttagacgt tctactttcc gtttgccaac caacccaaag 1680	
tctccagtc aatcattgtg tcacaggtact ggtgtcgccc cattcagagg cttcgttcaa 1740	
gaaaagagtg ccttggccag aagatccatc gaaaagaacg gtccgtgactc tttggctgac 1800	
tggggtcgta tttccttgtt ctacggttgt agaagatccg acgaagactt cttgtacaag 1860	
gacgaatggc ccaaatagca agctgagttg aagggttaagt tcaagttgca ctgtgctttc 1920	
tccagacaaa actacaagcc agacggttct aagatttacg tccaagattt gatctgggaa 1980	
gacagagaa acattgcccga tgcattctta aacggttaagg gttacgtcta catctgcggt 2040	
gaagctaaagt ccatgtctaa acaagttgaa gaagtcttag ccaagatctt gggcgaaagcc 2100	
aaaggtggtt ccgtccagtg tgaaggtgtt gctgaagtc agttactgaa ggaacgggtc 2160	
agattgatgt tggatgtctg gtcctga 2187	
SEQ ID NO: 96	
FEATURE	Location/Qualifiers
source	1..728 mol_type = protein organism = <i>Psilocybe cubensis</i>
SEQUENCE: 96	
MASSSSDVFV LGLGVVLAAL YIFRDQLFAA SKPKVAPVST TKPANGSANP RDFIAKMKQG 60	
KKRIVIFYGS QTGTAEEYAI RLAKKAKQKF GLASLVCDPE EYDFEKLDQL PEDSIAFFV 120	
ATYGEGETPD NAVQLQNLQ DESFEFSSGE RKLGLKVVV FGLGNKTYEH YNLIGRTVDA 180	
QLAKMGAIIRI GERGEGDDDK SMEEDYLEWK DGMWEAFATA MGVEEGQGGD SADFFVSELE 240	
SHPPEKVVQG EFSARALTKT KGIHDAKNPF AAPIAVAREL FQSVVDRNCV HVEFNIEGSG 300	
ITYQHGDHVG LWPLNPVEV ERLLCVLGLA EKRDVISIE SLDPALAKVP FPVPTTYGAV 360	
LRHYIDISAV AGRQILGTLK KFAPTPEAEA FLRNLNTNKE EYHNVVANGC LKLGEILQIA 420	
TGNDITVPPT TANTTKWPIP FDIIVSAIPR LQPRYSSISS SPKIHNTIAT ATVVVLKYEN 480	
VPTEPIPRKW VYGVGSNLL NLKYAVNKEP VPYITQNGEQ RVGVPEYLIA GPRGSYKTES 540	
FYKAPIHVR STFRLPNTPK SPVIMIGPGT GVAPFRGFVQ ERVALARRSI EKNGPDSLAD 600	
WGRISLFYGC RRSDEDFLYK DEWPQYEAEL KGKFKLHCAF SRQNYKPDGS KIYVQDLIWE 660	
DREHIADAIL NGKGYVYICG EAKSMKQVE EVLAKILGEA KGGSGPVEGV AEVKLLKERS 720	
RLMLDVWS 728	
SEQ ID NO: 97	
FEATURE	Location/Qualifiers
misc_feature	1..5203 note = pETM6-H10 vector
source	1..5203 mol_type = other DNA organism = synthetic construct
SEQUENCE: 97	
gaagaattgt gagcggataa caattcccct ctagaataaa ttttgtttta ctttaagaag 60	
gagatataca tatggcagat ctcaatttga tatcgccgg ccacgcgac gctgacgtcg 120	
gtaccctcga tctctgtaaa gaaacgcgtg ctgcgaaatt tgaacgccag cacatggact 180	
cgtctactag gtcgacgtta attaacctaa actgctgcca ccgctgagca ataactagca 240	
taaccccttg gggcctctaa acgggtcttg aggggttttt tgctagcgaa aggaggagtc 300	
gactatatcc ggattggcga atgggacgcg cctgtagcgg gcgcattaag cgcggcggtg 360	
gtgggtggtta cgcgcagcgt gaccgctaca cttgccagcg ccttagcgcc cgctcctttc 420	
gctttcttcc cttcctttct cgccacgttc gccggttttc cccgtcaagc tctaatacgg 480	

-continued

gggctccctt	taggggtccg	atttagtgct	ttacggccac	tcgaccccaa	aaaacttgat	540
tagggtagtg	gttcacgtag	tgggccatcg	cctgataga	cggtttttcg	ccctttgacg	600
ttggagtcca	cggtttctaa	tagtggaactc	ttgttccaaa	ctggaacaac	actcaaccct	660
atctcggctc	attcttttga	tttataaggg	atthttgcga	tttcggccta	ttggttaaaa	720
aatgagctga	tttaacaaaa	attttaacgcg	aatttttaaca	aaatattaac	gtttacaatt	780
tctggcggca	cgatggcatg	agattatcaa	aaaggatctt	cacctagatc	cttttaaatt	840
aaaaatgaag	ttttaaatca	atctaaagta	tatatgagta	aacttggtct	gacagttacc	900
aatgcttaat	cagtgaaggca	cctatctcag	cgatctgtct	atttcgttca	tccatagttg	960
cctgactccc	cgctgtgtag	ataactacga	tacgggaggg	cttaccatct	ggccccagtg	1020
ctgcaatgat	accgcgagac	ccacgctcac	cggtccaga	tttatcagca	ataaaccagc	1080
cagccggaag	ggccgagcgc	agaagtgggtc	ctgcaacttt	atccgcctcc	atccagtcta	1140
tttaattgtg	ccgggaagct	agagtaagta	gttcgocagt	taatagtttg	cgcaacggtg	1200
ttgccattgc	tacaggcatc	gtgggtgtcac	gctcgtcgtt	tggtatggct	tcattcagct	1260
ccgggttccca	acagatcaagg	cgagttacat	gatcccccat	gttggtgcaaa	aaagcgggtta	1320
gctccttcgg	tcctccgatc	gttggtcagaa	gtaagttggc	cgaggtgta	tcactcatgg	1380
ttatggcagc	actgcataat	tccttactg	tcattgccatc	cgtaagatgc	ttttctgtga	1440
ctgggtgagta	ctcaaccag	tcattctgag	aatagtgtat	gcggcgacccg	agttgctctt	1500
gccccgggctc	aatacgggat	aataccgcgc	cacatagcag	aactttaaaa	gtgctcatca	1560
ttggaaaaacg	ttcttcgggg	cgaaaactct	caaggatctt	accgctgttg	agatccagtt	1620
cgatgtaaac	cactcgtgca	cccaactgat	cttcagcatc	ttttactttc	accagcggtt	1680
ctgggtgagc	aaaaacagga	aggcaaaatg	ccgcaaaaaa	gggaataaag	gcgacacgga	1740
aatgttgat	actcatactc	ttcctttttc	aatcatgatt	gaagcattta	tcagggttat	1800
tgtctcatga	gcggatacat	atthgaatgt	atthtagaaaa	ataaaccat	aggtcatgac	1860
caaaatccct	taacgtgagt	tttcgttcca	ctgagcgtca	gaccccgtag	aaaagatcaa	1920
aggatctctc	tgagatcctt	ttttctcgcg	cgtaactctgc	tgcttgcaaa	caaaaaaac	1980
accgctacca	gcgggtggtt	gtttgcggga	tcaagagcta	ccaactcttt	ttccgaaggt	2040
aactggcttc	agcagagcgc	agataccaaa	tactgtcctt	ctagtgtagc	cgtagttagg	2100
ccaccacttc	agaactctg	tagcacccgc	tacatacctc	gctctgctaa	tcctgttacc	2160
agtggctgct	gccagtggtg	ataagtcgtg	tcttaccggg	ttggactcaa	gacgatagtt	2220
accggataag	gcgcagcggt	cggtgtgaac	gggggggttcg	tgcacacagc	ccagcttgga	2280
gcgaacgacc	tacaccgaac	acagatacct	acagcgtgag	ctatgagaaa	gcgccacgct	2340
tcccgaaggg	agaaagcggg	acaggtatcc	ggtaagcggc	agggtcggaa	caggagagcg	2400
cacgagggag	cttccagggg	gaaacgcctg	gtatctttat	agtccgtcgc	gggttcgcca	2460
cctctgactt	gagcgtcgat	ttttgtgatg	ctcgtcaggg	gggcggagcc	tatggaaaaa	2520
cgccagcaac	gggccttttt	tacggttcct	ggccttttgc	tgcccttttg	ctcacatgtt	2580
ctttcctgcg	ttatccctcg	attctgtgga	taaccgtatt	accgcctttg	agttagctga	2640
taccgctcgc	cgacgcgaa	cgacgagcg	cagcagatca	gtgagcagg	aagcgggaaga	2700
gcgcctgatg	cggtattttc	tccttacgca	tctgtgcggt	atttcacacc	gcataatatg	2760
tgcaactcca	gtacaactcg	ctctgatgcc	gcatagttaa	gccagtatcc	actccgctat	2820
cgctacgtga	ctgggtcatg	gctgcgcgcc	gacacccgcc	aacacccgct	gacgcgcctc	2880
gacgggcttg	tctgtcccg	gcatccgctt	acagacaagc	tgtgaccgtc	tcggggagct	2940
gcattgttca	gaggttttca	ccgtcatcac	cgaaacgcgc	gaggcagctg	cggttaagct	3000
catcagcgtg	gtcgtgaagc	gatttcacaga	tgtctgcctg	ttcatccgcg	tcagctcgt	3060
tgagtttctc	cagaagcggt	aatgtctggc	ttctgataaa	gcggggccatg	tttaagggcgg	3120
ttttttcctg	tttggtcact	gatgcctccg	tgtaaggggg	atttctgttc	atgggggtaa	3180
tgataccgat	gaaacgagat	aggatgctca	cgataccggg	tactgatgat	gaacatgcc	3240
gggtactgga	acgttgtgag	ggtaaacac	tggcgggtatg	gatgcggcgg	gaccagagaa	3300
aaatcactca	gggtcaatgc	cagcgtctcg	ttaatcacaga	tgtagggtgt	ccacagggtta	3360
gccagcagca	cctgcgatg	cagatccgga	acataatggg	gcagggcgct	gacttccgcg	3420
tttccagact	ttacgaaaca	cggaaaccga	agaccattca	tgttgttgc	cagggtcgag	3480
acgttttgca	gcagcagctg	cttcacgttc	gctcgcgtat	cggtgatcca	ttctgttaac	3540
cagtaaggca	accccgccag	cctacccggg	tcctcaacga	caggagcagc	atcatgctag	3600
tcattgcccg	cgccccacgg	aaggagctga	ctgggttgaa	ggctctcaag	ggcatcggtc	3660
gagatcccg	tgccaatga	gtgagctaac	ttacattaat	tgctgtgcgc	tcactgcccg	3720
ctttccagtc	gggaaacctc	tcgtgccagc	tgcattaatg	aatcggccaa	cgccggggga	3780
gaggcggttt	gcgtattggg	cgccaggggtg	gtttttcttt	tcaccagtga	gacggggcaac	3840
agctgattgc	ccttcaccgc	ctggccctga	gagagttgca	gcaagcggtc	cacgctgggt	3900
tgccccagca	ggcgaaatcc	ctgtttgatg	gtggttaacg	gcgggatata	acatgagctg	3960
ttctcggtat	cgctgtatcc	cactaccgag	atgtccgcac	caacgcgcag	cccgactcgc	4020
gtaattggcg	gcattgcgcc	cagcgcctatc	tgatcgttgg	caaccagcat	cgcagtgga	4080
acgatgcctc	cattcagcat	ttgcattggt	tggtgaaaa	cggacatggc	actccagtcg	4140
ccttcccggt	ccgctatccg	ctgaatttga	ttgcgagtg	gatatttatg	ccagccagcc	4200
agacgcagac	gcgcgagac	agaacttaat	gggcccgccta	acagcgcgat	ttgctgggtga	4260
cccaatgcga	ccagatgctc	cagcgcagat	cgctaccgt	cttcattggga	gaaaaataa	4320
ctgttgatgg	gtgtctggtc	agagacatca	agaaataacg	ccggaacatt	agtgcaggca	4380
gcttccacag	caatggcatc	ctgggtcatcc	agcggatagt	taatgatcag	cccaactgag	4440
cggtgcgcga	gaagattgtg	caccgcgcgt	ttacaggctt	cgacgcgcgt	tcgtttctacc	4500
atcgacacca	ccacgctggc	accagttga	tcggcgcgag	atttaatcgc	cgcgacaatt	4560
tgcgacggcg	cgtgcagggc	cagactggag	gtggcaacgc	caatcagcaa	cgactgtttg	4620
cccgcagatt	gttggtccac	gcgggtggga	atgtaattca	gctccgccat	cgccgcttcc	4680
actttttccc	agcttttccg	agaaaacgtg	ctggcctggg	tcaccacgcg	ggaaacggtc	4740
tgataagaga	caccggcata	ctctgcgaca	tcgtataacg	ttactgggtt	cacattcacc	4800
accctgaatt	gactctcttc	cgggcgctat	catgccatcc	cgcgaaaggt	tttgccgcat	4860
tcgatgggtg	ccgggacttc	gacgctctcc	cttatgcgac	tcctgcatta	ggaagcagcc	4920
cagttagtagg	ttgaggccgt	tgagcaccgc	cgccgcaagg	aatgggtgcat	gcaaggagat	4980
ggcgcccaac	agtcgcccg	ccacggggcc	tgccaccata	cccacgcga	aacaagcgct	5040
catgagcccg	aagtggcgag	cccgatcttc	cccacgggtg	atgtcggcga	tataggcgcc	5100
agcaaccgca	cctgtggcgc	cggtgatgcc	ggccacgatg	cgtccggcgt	agcctaggat	5160
cgagatcgat	ctcgatcccg	cgaattaat	acgactcact	acg		5203

-continued

```

SEQ ID NO: 98      moltype = DNA length = 3777
FEATURE           Location/Qualifiers
misc_feature       1..3777
                  note = pRSM3 vector
source            1..3777
                  mol_type = other DNA
                  organism = synthetic construct

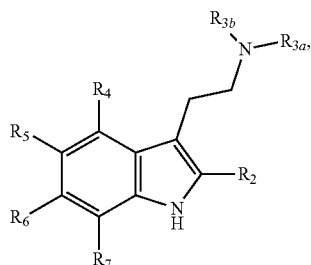
SEQUENCE: 98
cgctaacagc gcgatttgct ggtgacccaa tgcgaccaga tgcctccacgc ccagtcgcgt 60
accgtcttca tgggagaaaa taatactggt gatgggtgtc tgggtcagaga catcaagaaa 120
taacgccgga acattagtgc aggcagcttc cacagcaatg gcatoctggt catccagcgg 180
atagttaatg atcagccccc tgacgcgttg cgcgagaaga ttgtgcaccg ccgctttaca 240
ggcttcgacg ccgcttcggt ctaccatcga caccaccacg ctggcaccca gttgatcggc 300
gcgagattta atccgcgcga caatttgcca cggcgcgtgc agggccagac tggaggtggc 360
aacgccaatc agcaacgact gtttgccgc cagttgtgt gccacgcggg tgggaatgta 420
attcagctcc gccatcgccg cttccacttt ttcccgcgtt ttccgagaaa cgtggctggc 480
ctggttcacc acgcgggaaa cggtctgata agagacaccg gcatactctg cgacatcgta 540
taacgttact ggtttcacat tcaccacct gaattgactc tcttcggggc gctatcatgc 600
cataccgcga aaggttttgc gccattcgat ggtgtccggg atctcgacgc tctcccttat 660
gcgactcctg ctaggaagag cagcccagta gtaggttgag gccgttgagc accgcgcggc 720
caaggaatgg tgcattgcaa gagatggcgc ccaacagtc ccggccacg gggcctggca 780
ccataccccc gccgaaacaa gcgctcatga gcccgaaagt gcgagcccca tcttccccat 840
cgggtgatgc ggcgatatag gcgccagcaa ccgcacctgt ggccgcgggt atgccggcca 900
cgatgcgtcc ggcgtagcct aggatcgaga tcgatctcga tcccgcgaaa ttaatacgac 960
tcactatagg ggaattgtga gcggataaca attcccctct agaaataatt ttgtttaact 1020
ttaagaagga gatatacata tggcagatct caattggata tcggccggcc acgcgatcgc 1080
tgacgtcggg accctcgagt ctggtaaaaga aaccgctgct gcgaaatttg aacgccagca 1140
catggactcg gcaactagtc cgaagcttaat taacctaaac tgcggccacc gctgagcaat 1200
aactagcata accccttggg gcctctaaac gggctcttgg ggggtttttg ctagegaaag 1260
gaggagtcga caagctgacg accgggtctc cgcgaagtggc acttttcggg gaaatgtgcy 1320
cggaaccctc attgttttat tttctaaat acattcaaat atgtatccgc tcatgaatta 1380
attcttagaa aaactcatcg agcatcaaat gaaactgcaa tttattcata tcaggatat 1440
caataccata tttttgaaaa agccgtttct gtaatgaagg agaaatacca ccgaggcagt 1500
tccataggat ggcgaagatcc tggtatcggg ctgcgattcc gactcgtcca acatcaatac 1560
aacctattaa tttcccctcg tcaaaaataa ggttatcaag tgagaaatca ccatgagtga 1620
cgactgaatc cgggtgagaat ggcaaaagtt tatgcatttc tttccagact tgttcaacag 1680
gccagccatt acgctcgtca tcaaaatcac tcgcatcaac caaacggtta ttcattcgtg 1740
attgcgcctg agcgagcga aatacgcggg cgctgtttaa aggacaatta caaacaggaa 1800
tcgaatgcaa ccggcgccagg aacactgcca gcgcatcaac aatattttca cctgaatcag 1860
gatattcttc taatacctgg aatgctggtt tcccggggat cgcagtggtg agtaaccatg 1920
catcatcagg agtacggata aaatgcttga tggtcggaag aggcataaat tccgtcagcc 1980
agtttagtct gaccatctca tctgtaacat cattggcaac gctacctttg ccatgtttca 2040
gaaacaactc tggcgcatcg ggcttcccat acaatcgata gattgtcgca cctgattgac 2100
cgacattatc gcgagcccat ttatacccat ataaatcagc atccatgttg gaatttaac 2160
gcgcctaga gcaagacggt tcccgttgaa tatggctcat actcttcctt ttccaatatt 2220
attgaagcat ttatcagggt tattgtctca tgagcggata catatttgaa tgtatttaga 2280
aaaataaaca aataggcatg cagcgtctct ccgcttctc gctcaactgac tcgctacgct 2340
cggctcgttcg actcgggcga gcggtgtcag ctactcaaa agcggtaata cggttatcca 2400
cagaatcagg gataaaagcc gaaaagaaca tgtgagcaaa aagcaaacga ccggaagaag 2460
ccaacgcgcg aggcgttttt ccataggctc cgcctccctg acgagcatca caaaaatcga 2520
cgctcaagcc agaggtggcg aaacccgaca ggactataaa gataccaggc gtttccccct 2580
ggaagctccc tcgtgcgctc tctgtttccg accctgcgcg ttaccggata cctgtccgac 2640
tttctccctt cgggaagcgt ggcgctttct catagctcac gctgttggtg tctcagttcg 2700
gtgtaggtcg ttcgctccaa gctgggctgt gtgcacgaac ccccggttca gcccgaccgc 2760
tgcgccttat ccggtaaact tcgtcttgag tccaacccgg taagacacga cttatcgcca 2820
ctggcagcag ccattggtta ctgatttaga ggactttgtc ttgaagttag gcacctgtta 2880
aggctaaact gaaagaacag attttggtga gtgcgggtcct ccaacccact tacctgggtt 2940
caaagagttg gtatctcagc gaaccttgag aaaaccaccg ttggtagcgg tggtttttct 3000
ttatttatga gatgatgaat caatcggctc atcaagtcaa cgaacagcta ttcggttact 3060
ctagatttca gtgcaattta tctcttcaaa tgtagcacct gaagtacgac ccatacgata 3120
taagttgtaa ttctcatggt agtcatgccc cgcgcccacc ggaaggagct gactgggttg 3180
aaggctctca agggcatcgg tcgagatccc ggtgcctaata gagtgaagct acttacatta 3240
attgcgttgc gctcactgcc cgctttccag tcgggaaacc tgcgtgcca gctgcattaa 3300
tgaatcggcc aacgcgcggg gagaggcggg ttgcgtattg ggccgacagg tggtttttct 3360
tttaccaggt gagacgggca acagctgatt gcccttcacc gcctggccct gagagagttg 3420
cagcaagcgg tccacgctgg tttgccccag caggcgaaaa tccgttttga tgggtggttaa 3480
cggcgggata taacatgagc tgtcttcggg atcgtcgtat cccactaccg agatgtccgc 3540
accaacgcgc agcccgagct cggtaattgg cgcgatttgc cccagcgcca tctgatcggt 3600
ggcaaccagc atcgcagttg gaacgatgcc ctacttcagc atttgcagtg tttgttgaaa 3660
accggacatg gcactccagt cgccttcccg ttccgctatc ggctgaattt gattgcgagt 3720
gagatattta tgcagccag ccagacgcag acgcgccgag acagaactta atggggcc 3777

```

221

The invention claimed is:

1. A chemical compound or salt thereof having the formula (I):



wherein, at least one of R₂, R₄, R₅, R₆ or R₇ is a halogen atom, wherein each non-halogenated R₂, R₅, R₆, or R₇ is a hydrogen atom or an alkyl group or O-alkyl group, wherein R₄ when it is not halogenated is a phosphate group, a hydrogen atom or an alkyl group or O-alkyl group, and wherein R_{3a} and R_{3b} are a hydrogen atom, an alkyl group, an aryl group, or an acyl group.

2. A chemical compound according to claim 1, wherein one or two of R₄, R₅, R₆, or R₇ are a halogen atom, wherein the halogen atom is selected from fluorine (F), chlorine (Cl), and bromine (Br).

3. A chemical compound according to claim 1, wherein one or two of R₄, R₅, or R₆ are halogenated, wherein the halogen atom is selected from fluorine (F), chlorine (Cl), and bromine (Br).

4. A chemical compound according to claim 1, wherein two of R₄, R₅, and R₆ are halogenated, wherein the halogen atom is selected from fluorine (F), chlorine (Cl), and bromine (Br).

5. A chemical compound according to claim 4, wherein R₄ is halogenated, and wherein one of R₅, and R₆ is halogenated.

6. A chemical compound according to claim 4, wherein R₅ is halogenated, and wherein one of R₄, and R₆ is halogenated.

7. A chemical compound according to claim 5, wherein the halogen atoms are fluorine atoms.

8. A chemical compound according to claim 6, wherein the halogen atoms are fluorine atoms.

9. A chemical compound according to claim 2, wherein R₇ is a fluorine atom, and wherein one atom is halogenated.

10. A chemical compound according to claim 7, wherein each non-halogenated R₂, R₄, R₅, R₆, or R₇, is a hydrogen atom.

11. A chemical compound according to claim 8, wherein each non-halogenated R₂, R₄, R₅, R₆, or R₇, is a hydrogen atom.

12. A chemical compound according to claim 9, wherein each non-halogenated R₂, R₄, R₅, R₆, or R₇, is a hydrogen atom.

13. A chemical compound according to claim 3, wherein at least one of the halogen atoms is a chlorine atom.

14. A chemical compound according to claim 3, wherein R₆ is a chlorine atom.

15. A chemical compound according to claim 14, wherein R₅ is a fluorine atom.

16. A chemical compound according to claim 14, wherein each non-halogenated R₂, R₄, R₅, R₆, or R₇, is a hydrogen atom.

222

17. A chemical compound according to claim 15, wherein each non-halogenated R₂, R₄, R₅, R₆, or R₇, is a hydrogen atom.

18. A chemical compound according to claim 3, wherein R₅ is a bromine atom, and one atom is halogenated.

19. A chemical compound according to claim 18, wherein each non-halogenated R₂, R₄, R₅, R₆, or R₇, is a hydrogen atom.

20. A chemical compound according to claim 2, wherein R_{3a} and R_{3b} are each independently selected from a hydrogen atom, a (C₁-C₆)-alkyl group, and a (C₁-C₆)-acyl group.

21. A chemical compound according to claim 10, wherein R_{3a} and R_{3b} are, respectively, a hydrogen atom and a (C₁-C₆)-acyl group.

22. A chemical compound according to claim 11, wherein R_{3a} and R_{3b} are, respectively, a hydrogen atom and a (C₁-C₆)-acyl group.

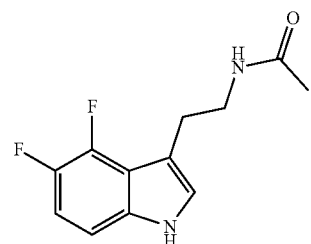
23. A chemical compound according to claim 12, wherein R_{3a} and R_{3b} are, respectively, a hydrogen atom and a (C₁-C₆)-acyl group.

24. A chemical compound according to claim 16, wherein R_{3a} and R_{3b} are each a (C₁-C₆)-alkyl group.

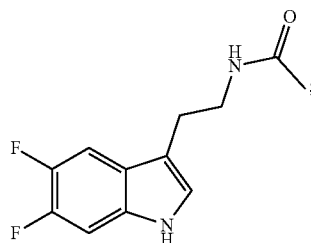
25. A chemical compound according to claim 17, wherein R_{3a} and R_{3b} are each a (C₁-C₆)-alkyl group.

26. A chemical compound according to claim 19, wherein R_{3a} and R_{3b} are each a (C₁-C₆)-alkyl group.

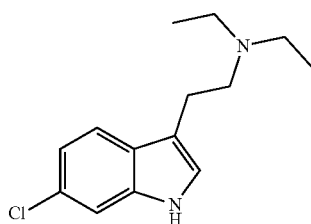
27. A chemical compound according to claim 1, wherein the chemical compound is selected from the group consisting of compounds having formulas (V); (IX); (X); (XV); (XX); (XXI); (XXII); (XXIII); and (XXIV):



(V)



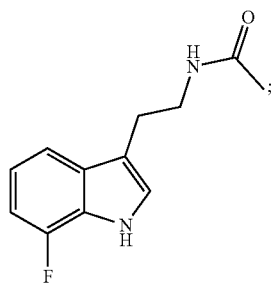
(IX)



(X)

223

-continued



5

10

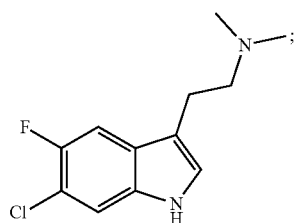
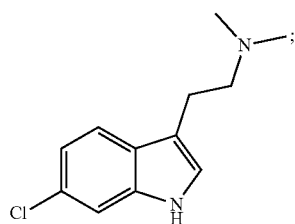
(XX)

20

25

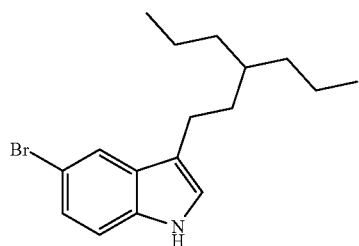
(XXI)

35

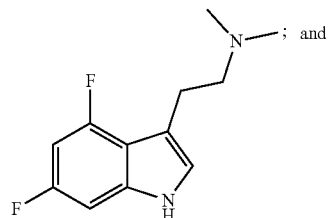
**224**

-continued

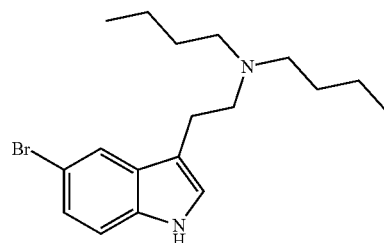
(XXII)



(XXIII)



(XXIV)



28. A pharmaceutical drug formulation comprising an effective amount of the chemical compound according to claim 1, together with a pharmaceutically acceptable excipient, diluent, or carrier.

29. A pharmaceutical drug formulation comprising an effective amount of the chemical compound according to claim 27, together with a pharmaceutically acceptable excipient, diluent, or carrier.

* * * * *