



(51) International Patent Classification:

C07K 1/02 (2006.01) C07K 14/00 (2006.01)
C07K 1/10 (2006.01) C12N 9/02 (2006.01)
C07K 7/50 (2006.01) C12N 15/09 (2006.01)
C07K 7/64 (2006.01)

(21) International Application Number:

PCT/US2018/029954

(22) International Filing Date:

27 April 2018 (27.04.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/491,632 28 April 2017 (28.04.2017) US

(71) Applicant: BOARD OF TRUSTEES OF MICHIGAN
STATE UNIVERSITY [US/US]; 450 Administration
Building, East Lansing, MI 48824 (US).

(72) Inventors: WALTON, Jonathan, D.; 810 Roxburgh Av-
enue, East Lansing, MI 48823 (US). SGAMBELLURI, R.,
Michael; 2329 Storehodge Drive, Apt. B, East Lansing, MI
48823 (US).

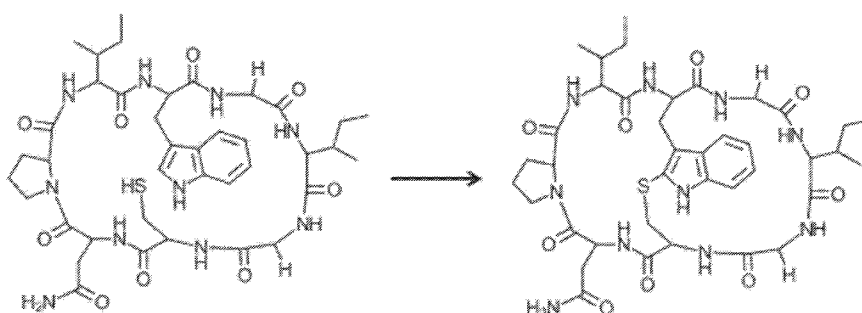
(74) Agent: PERDOK, Monique, M. et al.; Schwegman, Lund-
berg & Woessner, P.A., P.O. Box 2938, Minneapolis, MN
55402 (US).

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

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,
UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,
TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,
EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,
MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

(54) Title: TRYPTATHIONINE BIOSYNTHESIS BY FLAVIN MONO-OXYGENASE 1 (FMO1)

FIG. 1B



(57) Abstract: Flavin mono-oxygenase (FMO) enzymes, as well as expression cassettes and host cells for expressing such enzymes, and methods for using such FMO enzymes are described herein. The flavin mono-oxygenase enzymes are useful for generating peptidyl cross-bridges.



TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).

Published:

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

Tryptathionine Biosynthesis by Flavin Mono-Oxygenase 1 (FMO1)

This application claims benefit of priority to the filing date of U.S. Provisional Application Ser. No. 62/491,632, filed April 28, 2017, the contents of which are specifically incorporated herein by reference in their entirety.

Government Funding

5 This invention was made with government support under GM088274 awarded by the National Institutes of Health. The government has certain rights in the invention.

Background of the Invention

Cyclic and bicyclic peptides have been found in plants, bacteria, fungi, and animals (Arnison et al., Nat. Prod. Rep. 30:108–160 (2013); Umemura et al., Fung. Genet. Biol. 10 68:23–30 (2014)). Examples of macrocyclized peptides include the cyanobactins from marine bacteria (Donia et al., Nat. Chem. Biol. 2:729–735 (2006)) and the orbitides and cyclotides from plants (Craik & Malik, Curr. Opin. Chem. Biol. 17:546–554 (2013)). Cyclization confers several pharmacologically advantageous attributes to peptides, including increased rigidity, enhanced membrane permeability, and resistance to proteolysis and denaturation 15 (Craik & Allewell, Biol. Chem. 287:26999–27000 (2012)). As a result, many cyclic peptides have potent biological activities against a wide range of medically and ecologically significant targets (Bockus et al., Curr. Top. Med. Chem. 13:821–836 (2013); Giordanetto & Kihlberg, J. Med. Chem. 57:278–295 (2014); Walton, Plant Cell. 8:1723–1733 (1996)).

Summary

20 Described herein are compositions and methods involving expression cassettes, enzymes, and methods useful for generating cross-bridges in peptides. The expression cassettes, and methods can involve using nucleic acids encoding a flavin mono-oxygenase (FMO) enzyme. Methods are also described for generating cross-bridges in peptides within *in vitro* (e.g., cell free) reaction mixtures. The cross-bridge enzymes can be used in conjunction 25 with other enzymes such as enzymes that cyclize peptides. These cross-bridge and cyclization enzymes are useful for making a variety of crosslinked, bicyclic and/or cyclic peptides.

Described herein are expression systems that can include at least one expression cassette having a promoter operably linked to a heterologous nucleic acid segment encoding a flavin mono-oxygenase (FMO) with at least 95% sequence identity to SEQ ID NO: 2, 4, 5, 6, 30 or 7.

Host cells containing such expression systems are also described herein, and methods for using such expression systems and host cells. For example, the methods can include

incubating a host cell that has one or more expression systems that can include at least one expression cassette having a promoter operably linked to a heterologous nucleic acid segment encoding a flavin mono-oxygenase (FMO) with at least 95% sequence identity to SEQ ID NO: 2, 4, 5, 6, or 7, for a time and under conditions sufficient for generation of a

5 tryptathionine cross-bridge between a tryptophan (Trp) and a cysteine (Cys) in the peptide substrate. The expression systems, host cells, and methods can also include use of at least one expression cassette having a promoter operably linked to a heterologous nucleic acid segment encoding a prolyl oligopeptidase with at least 95% sequence identity to SEQ ID NO:11, 13, 14, or 15.

10 In some cases, tryptathionine cross-bridges between tryptophan (Trp) and cysteine (Cys) residues in the peptide substrate are formed *in vivo* (within a host cell) by FMO1 while in other cases tryptathionine cross-bridges between tryptophan (Trp) and cysteine (Cys) residues in the peptide substrate are formed *in vitro* (not within a host) by FMO1. For example, in some cases additional cellular enzymes and other factors can help optimize cross-

15 bridge formation and/or can facilitate cyclization of peptides. Accordingly, in some cases the cross-bridges are prepared *in vivo*.

However, the FMO1 enzyme can be incubated *in vitro* with a peptide that includes tryptophan (Trp) and cysteine (Cys) residues to generate a peptide with a tryptathionine cross-bridge between the Trp and Cys residues. Hence, the FMO1 proteins can be used in *in*

20 *vitro* compositions, assay mixtures, and/or enzymatic manufacturing processes.

Description of the Figures

FIG. 1A-1B illustrates cyclization of peptides. FIG. 1A illustrates conversion of a 35-amino acid precursor peptide (SEQ ID NO: 19) to a cyclic octapeptide by a prolyl oligopeptidase B from *Galerina marginata* (GmPOPB) enzyme (SEQ ID NOs: 19 and 27-

25 29). As illustrated, GmPOPB catalyzes two reactions to convert the linear 35mer precursor peptide (S1) to cyclo(IWGIGCNP, SEQ ID NO:1), an intermediate in amanitin biosynthesis. Product P3 is then cross-linked (Trp to Cys) as illustrated in FIG. 2 and hydroxylated (3-4 times) to make bicyclic peptide (see also Luo et al. Chemistry and Biology 21:1610 (2014)). FIG. 1B illustrates the reaction catalyzed by flavin mono-oxygenase (FMO1) to form

30 tryptathionine. The substrate is the product of POPB.

FIG. 2A-2B illustrates how hydroxylation of an indole ring can activate the indole for further reaction including cross-linking to cysteine. FIG. 2A illustrates 3-hydroxylation of indole as catalyzed by an FMO from *Methylophaga*. FIG. 2B illustrates reaction via Savigne-Fontana synthesis of tryptathionine from 3a-hydroxy-pyrrolo[2,3-b]-indole (Hpi), in which 3-

hydroxyTrp is an intermediate. See, Cho et al., J Struct Biol 175:39-48 (2011); May JP, Perrin DM (2007) Tryptathionine bridges in peptide synthesis. Biopolymers 88:714-724.)

FIGs. 3A-3C illustrate *in vitro* expression of FMO1 and its catalysis of a tryptathionine crossbridge product. FIG 3A shows expression of FMO1 as a 60-kDa protein fused to a myc tag. Left panel: Coomassie stained gel. Right panel: western blot of the same gel with anti-myc antibodies. FIG. 3B shows LC/MS analysis of substrate (monocyclic cyclo[IWGIGCNP], bottom) and product after reaction with FMO1 (bicyclic[IWGIGCNP], top). FIG. 3C shows a mass spectrum of the product eluting at 12.0 min as shown in FIG. 3B, illustrating the parent ion at M+H+ 839 (tallest peak) as predicted for bicyclic compound.

FIGs. 4A-4B illustrate that targeted gene disruption of *GmFMO1* abolishes amanitin biosynthesis. FIG. 4A shows an HPLC trace (UV absorbance at 280 nm) illustrating amanitin made by wild type *G. marginata* (solid trace) and not by deletion mutant *FMO1 G. marginata* (dotted trace) cells. FIG. 4B shows a Southern blot of genomic DNA from deletion mutant *FMO1 G. marginata* cells. The Southern blot shows that the *FMO1* gene has been deleted in the mutant. The results show that the FMO1 gene is important for amanitin production.

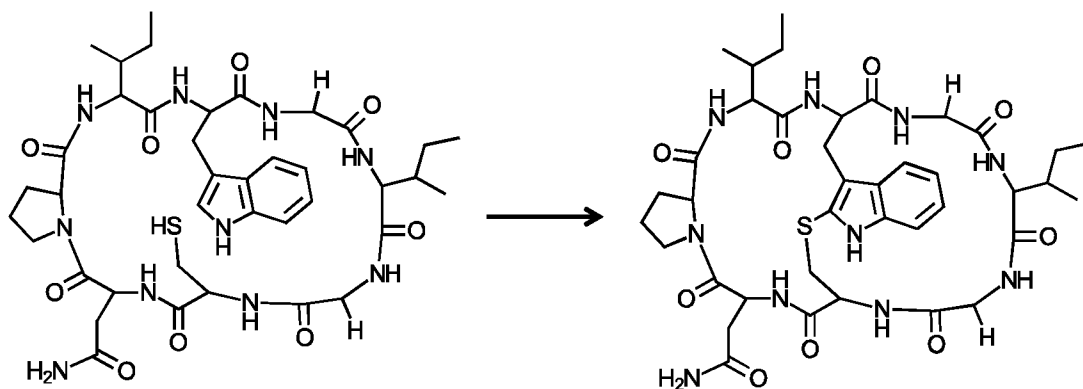
Detailed Description

Cyclic peptides, bicyclic peptides and peptides with cross-bridges are useful as drugs due to their stability, high surface area, membrane permeability, and resistance to proteases. For example, the bicyclic peptide toxins of Amanita mushrooms, the amatoxins such as amanitin and the phallotoxins such as phalloidin, include a thioether cross-bridge between tryptophan (Trp) and cysteine (Cys). This cross bridge, called tryptathionine, contributes to the activities of these powerful compounds. Tryptathionine is not present in any other natural product, and its biosynthetic pathway is unknown. Amatoxins currently cannot be chemically synthesized and Amanita cannot be cultured, so the only current source is extraction from mushrooms collected in the wild. The enzyme that catalyzes tryptathione biosynthesis can be used for *in vitro* or *in vivo* production of a new class of bicyclic peptides with potential pharmaceutical or industrial applications, ultimately including the amatoxins themselves.

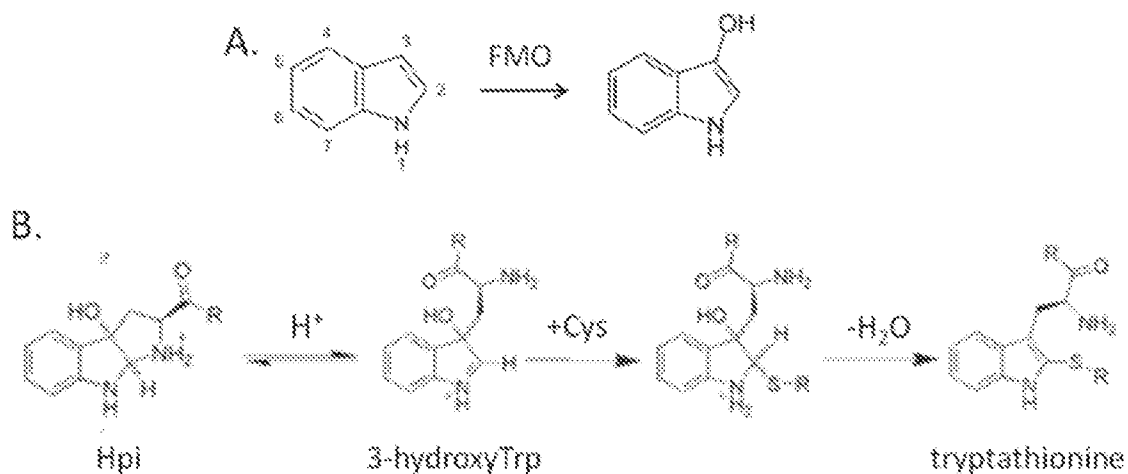
In the amanitin-producing fungus *Galerina marginata*, a gene encoding a flavin mono-oxygenase (FMO) is clustered with two genes in the amanitin biosynthetic pathway, prolyl oligopeptidase B (POPB) and AMA1. The specific versions of these three genes in *G. marginata* are also known as GmAMA1, GmFMO1, and GmPOPB. AMA1 encodes the 35-amino acid pro-protein for amanitin, and POPB encodes the macrocyclase that converts the 35mer precursor peptide to a cyclic octapeptide, cyclo (IWGIGCNP, SEQ ID NO:1) (Luo et

al., Fung Genet Biol 49:123 (2012); Luo et al. Chemistry and Biology 21:1610 (2014); Sgambelluri et al. ACS Synth Biol, 7 (1): 145–152 (2017)). Macrocyclases are enzymes that can convert a linear substrate into a circular (cyclic) product (e.g., a cyclic peptide). Genes for such enzymes are also present in *Amanita bisporigera* and *Amanita phalloides* (Pulman et al. BMC Genomics 17:1038 (2016)).

FMO1 catalyzes tryptathionine biosynthesis as shown by the following reaction.



The process by which this reaction proceeds is believed to proceed as illustrated below, where step (A) involves 3-hydroxylation of indole, like the reaction catalyzed by an FMO from *Methylophaga*. Step (B) is a reaction pathway for the Savige-Fontana synthesis of tryptathionine from the synthetic compound 3a-hydroxy-pyrrolo[2,3-b]-indole (Hpi), in which 3-hydroxyTrp is an intermediate.



After formation of the tryptathionine cross bridge by FMO1, peptides can undergo one or more hydroxylations and sulfoxidations. For example, after formation of a tryptathionine cross bridge, the remaining steps in amanitin biosynthesis include two to four hydroxylations and sulfoxidation. Identification of the tryptathionine forming enzyme FMO1 is a key missing link.

Flavin mono-oxygenase (FMO) nucleic acids and enzymes are described herein that are useful for generating crosslinked and/or cyclized peptides. For example, a nucleic acid encoding a *Galerina marginata* FMO1 (GmFMO1) enzyme was isolated as a cDNA. The *Galerina marginata* FMO1 (GmFMO1) protein encoded by such a cDNA is shown below as

SEQ ID NO:2.

```

      1 MVQIKRLLLG FLSSPSQTPL ESNHGPVPSK SIAVVGAGSA
     41 GLAMLRTLVE LEAFSRNNWE VVLYEERESV GGIWLPDNN
     81 VFPPEIPKTP LYPLLRTNTP VPSMTYPGFP FFPSTPLYPR
    121 HDHVEAYHLR YARRHNLLDF IKFDTMVEKA FWNGTPEEGY
    161 WNLTLSSKEG RMRYKTFDHL VVATGNNHIP HIPVWKGQED
    201 WLASPAHNSR KIIHSVYYRG PEAFSNQTVL IVGNCGSGRD
    241 AATQILGYAS QTFMSIRRSY GPVDDGVIVK PDISHFTEAG
    281 VVFVDGTILD PDVILLGTCY EMQKPLLSEG GELSFDPTAK
    321 DNSSVRGTLV TNGHYIFPLH RHIFSLSPRY PPNALAFI
    361 LSFIAASCPD IAQSLFAAHA ILDP SILPPR HLLLEELASY
    401 EDKARRQGLD PYLKGPIMLN NTSNDYQDEL VEYLKQKNAI
    441 PDDGKKFVEE WRREILAYHY LQRGWSRIEK LGMGPAWTEG
    481 VKTEAQWFDL MTRVNEWQKN WETENGIAFR VDLDLTG
  
```

The region that includes amino acid positions 34-105 of the SEQ ID NO:2 FMO1 protein includes a Rossmann-fold NAD(P)(+)-binding protein domain.

A nucleic acid sequence (cDNA) encoding the SEQ ID NO:2 *Galerina marginata* FMO1 is shown below as SEQ ID NO:3.

```

      1 ATGGTTCAAA TAAAACGACT CTTGCTGGGC TTTCTATCGT
     41 CTCCGTCGCA GACGCCGTTG GAGTCCAATC ATGGTCCAGT
     81 CCCTTCAAAA TCCATCGCCG TTGTTGGTGC TGGGAGTGCA
    121 GGCTTGCGCA TGCTCAGGAC TCTTGTGGAA TTAGAGGCAT
    161 TTTCTCGGAA TAATTGGGAA GTCGTTTTGT ACGAAGAACG
    201 AGAAAGTGTA GGGGGTATCT GGCTACCAGA TAACAACGAT
    241 GTCTTTCCAC CTGAGATACC AAAAATCCT CTCTATCCTC
    281 TCCTCCGTAC AAATACCCCA GTGCCTTCTA TGACATACCC
    321 TGGCTTTCCG TTCCCGCCCA GTACGCCTCT CTATCCACGT
    361 CATGACCATG TAGAAGCCTA TCATTTGCGT TACGCCCGCC
    401 GCCACAACCT TCTGGACTTC ATTAAATTCG ACACCATGGT
    441 TGAAAAGGCT TTCTGGAATG GAACACCAGA GGAAGGCTAT
    481 TGGAATCTTA CTCTTTCGAG CAAAGAGGGT CGAATGCGCT
    521 ATAAAACATT TGATCATCTT GTAGTCGCAA CTGGCAATAA
    561 CCACATACCT CACATTCCCG TTTGGAAGGG CCAAGAAGAC
  
```

25

30

35

40

45

50

Seq2 298 TGYEMQKPLLSEGGELSFDPTAKDNSSVRGTLVTNGHYIFPLHRHIFSLSPRYPPNALAF
Seq4 300 TGYEMRKPFLLSDSGELKVDSSARDNSSVOETLATNLRYLFPLYRHILSLSPRYPTNALAF

10

This related protein from *Galerina marginata* CBS 339.88 has accession number KDR84895.1 and the following sequence (SEQ ID NO:4).

20

25

30

35

40

45

50

20

25

30

35

40

45

50

Seq2 180 LVVATGNNHIPHIPVWKGQEDWLASAPANHSRKIIHSVYYRGPEAFSNQTVLIVGNGGSGR
 Seq6 179 LVVASGNNHIPRIPTWPGQKEWLTSSTG--REIVHSAWFRPEKYSNKSVLIVGYSASGQ
 ***** ** * ** * * * * * * * * * * * * * * * * *

5

Seq2 240 DAATQILGYASQTFMSIRRSYG---PVDDGVIVKPDISHFTEAGVVFVDGTILDPPDVIL
 Seq6 237 DETALIAPLAANTYVSTRSENATYQRPIPSNVIKKPAIAHFTKDSAVFVDKTEINVDSVL
 *

10

Seq2 296 LGTGYEMQKPLLSEGGELSFDPSTAKDNSSVRGTLVTNGHYIFPLHRHIFSLSPRYPPNAL
 Seq6 297 LATGYQIRKPFLEAGNGLLIKPTVRSSSESTTTLISNTKYIFPLHRHIFSLSTQYPTTAL
 *

15

Seq2 356 AFIGLLSFIA SCPSDIAQSLFAAHAILDPSILPPRHLLLEELASYEDKARRQGLDPYILKG
 Seq6 357 FIGLPSHIANCPSDTAQAIFAAHAIRNASILPSRDQMLQELAEFEQKSRDEGLDPYITG
 ***** *

Seq2 416 PIMLNNT--SNDYQDELVEYLYKQKNAIPDDGKKFVEEWRREILAYHYLQRGWSRIEKLGMG
 Seq6 417 HKLPDNTASDYQDELLDYLKEKG VIEDEGKEFVEEWRREIYSLQYLRRGWGRIEELGVG
 ** *

20

Seq2 475 PAWTEGVKTEAQWFDLMTRVNEWQKNWE
 Seq6 477 DDWVRGVETEETEEWSGVMARVNEWQKAYE
 * * * * * * * * * * * * * * * * *

25

This FMO1-related protein from *Amanita muscaria* has accession number KIL66087.1 and the following sequence (SEQ ID NO:6).

1 MVNVRHFIFS LIGTQLPAVY ESPNPEQVNE NSIAILGAGS
 41 AGLAMLKTL DFPDHNLDWI ALFEERENVG GIWLPDPDPS
 81 PPPVLPETPL YPALHTNTPV PMMTYPNFPF PAGTHLYPSH
 121 HHVEAYHVRY AHHYNLLPYI RFNHRVLITS WIGAPDDGFW
 161 NITFSNHKNE THHQAFRYLV VASGNNHIPR IPTWPGQKEW
 201 LTSSTGREIV HSAWFRPEK YSNKSVLIVG YSASGQDETA
 241 LIAPLAANTY VSTRSENATY QRPIPSNVIK KPAIAHFTKD
 281 SAVFVDKTEI NVDSVLLATG YQIRKPFLEA GNGLLIKPTV
 321 RSSESSTTTL ISNTKYIFPL HRHIFSLSTQ YPTTALSFIG
 361 LPSHIANCPS DTAQAIFAAH AIRNASILPS RDQMLQELAE
 401 FEQKSRDEGL DPYITGHKLP DNTRASDYQD ELLDYLKEKG
 441 VIEDEGKEFV EEWRREIYSL QYLRRGWGRI EELGVGDDWV
 481 RGVETEEEW S GVMARVNEWQ KAYEES

Another FMO1-related protein from *Amanita muscaria* Koide BX008 with SEQ ID NO:7 shares at least 52-53% sequence identity with the SEQ ID NO:2 protein sequence as illustrated below, where the asterisks illustrate which amino acids are the same between these two proteins.

Seq2 1 MVQIKRLLLGFLSSPSQTPLESNHEGP-VPSKSIIVGAGSAGLAMLRTLVELEAFSRNNW
 Seq7 1 MVNFRQIIFSLIGTQLPVVHQSSHPVLNKSIAIVGAGSAGLAMLKTLDFP--DHNSW
 ** * * * * * * * * * * * * * * * * *

50

Seq2 60 EVVLYEERESVGGIWLDPDNDVFPPEIPKTPLYLLRTNTPVPSMTYPGFFPFPSTPLYP
 Seq7 59 DIALFEERENVGGIWLDPDPDPNPPVLPETPLYPALHTNTPVPAMTYPNFPFPVVGTHLYP
 *

30

35

45

50

In another example, the *FMO1* nucleic acid can encode a FMO1 protein with an amino acid sequence that has at least 90%, or at least 95%, or at least 96%, or at least 97%, at least 98%, or at least 99% sequence identity to SEQ ID NO:2, 4, 5, 6, or 7. Using restriction endonucleases, the entire coding sequence for the *FMO1* nucleic acid is subcloned downstream of the promoter in a 5' to 3' sense orientation.

In some cases, the FMO1 proteins can be used in *in vitro* compositions, assay mixtures, and/or enzymatic manufacturing processes. The FMO1 proteins in these compositions, mixtures, and processes can have amino acid sequences that have at least 90%, or at least 95%, or at least 96%, or at least 97%, at least 98%, or at least 99% sequence identity to SEQ ID NO:2, 4, 5, 6, or 7.

Peptide Cyclization Enzymes

In some cases, the FMO1 nucleic acids and/or FMO1 enzymes can be used in conjunction with other enzymes. For example, the FMO1 nucleic acids and/or FMO1 enzymes can be co-expressed or combined *in vitro* with other enzymes that can cyclize peptides.

One example of an enzyme that can cyclize peptides is a prolyl oligopeptidase enzyme. As used herein, "prolyl oligopeptidase" or "POP" refers to a member of a family of enzymes classified and referred to as EC 3.4.21.26-enzymes that are capable of cleaving a peptide sequence, for example via hydrolysis of Pro-l-Xaa>> Ala-l-Xaa in oligopeptides.

Prolyl oligopeptidases are also referred to as "post-proline cleaving enzyme," "proline-specific endopeptidase," "post-proline endopeptidase," "proline endopeptidase," "endoprolol peptidase," "prolyl endopeptidase," "post-proline cleaving enzyme," "post-proline endopeptidase," and "prolyl endopeptidase." POPA enzymes are found in the majority of mushrooms. POPB enzymes generally have approximately about 50% to 60% (e.g. 55%) amino acid identity to POPA enzymes. The POPB enzymes are primarily found in *Amanita*, *Galerina*, and *Lepiota* peptide-producing mushroom species.

One example of an *Amanita bisporigera* prolyl oligopeptidase (POPA) sequence is shown below as SEQ ID NO:8.

```

1  MHRFLQPVRE  RLRSalaryF  GSRIMSSTQW  TPNMYPSARR
30  41  SDHIDTYRSE  TRGEVKVPDP  YHWLEEYSEE  TDKWTSDQEE
    81  FTRTYLDSNP  DRKKLEDAFR  KSMDYPKFSA  PFLNDDKRWY
    121 WFYNTGLQAA  TVICRSKDET  LPDFSES DYV  GETFFDPNLL
    161 SSDGTASLSM  YDFSHCGKYF  AYGISLSGSD  FSTIYVRSTS
    201 SPLAPGNNSI  RNDDGRLPDE  LRYVKFSSIS  WTKDSKGFFY
35  241 QRYPGTGTVN  GQNGIQTQGD  RDAMIYYHRI  GTSQSDDILV
    281 HEDQEHPDWV  FGAEVTEGDK  YVALYTMKDT  SRKNLLWIAD

```

321 LGQNEVGRNM KWNKICNVFD SEYDLIGNDG SLLYIRTNKA
 361 APQYKIVTLD IEKPELGFKE FIPEDPKAYL SQVKIFNKDR
 401 LALVYKRNVI GELYVYNNITG SRLMRLARDF VGSMTVTARE
 441 TEPWFFATLT GFNTPGIVCR YNIQRPEEQR WSVYRTAKVK
 5 481 GLNPNDFEAR QVWYDSYDGT KIPMFIVRHK NTQFNGTAPA
 521 IQYGYGGFNI SINPFFSPTI LTFLQKYGAI LAVPNIRGGG
 561 EFGETWHDAG IREKRANVYD DFIAATQFLV KNKYAAGGKV
 601 AINGGSNGGL LVAACVNRAR EGTFGAAIAE VGVLDLLKFP
 641 KFTIGKAWIS DYGDPEDEPRD FDYIYTHSPL HNIPKNMVL
 10 681 PTMLLTADHD DRVPMHSFK YAAMLQYTLR HNRHPLLLRV
 721 DKKAGHGGGK STEKRLQEAA DKWGFAAQSM GLAWKDRQAN
 741 L

An example of a nucleic acid (cDNA) sequence that encodes the *Amanita bisporigera* prolyl

15 oligopeptidase (POPA) protein is shown below as SEQ ID NO:9.

1 ATGCACCGTT TTTTGCAGCC CGTAAGAGAA CGCCTTCGCT
 41 CTGCTCTCGC CCGCTACTTT GGTTCGCGGA TCATGTCTTC
 81 TACACAGTGG ACACCCAACA TGTACCCTTC TGCTCGCCGT
 121 TCAGACCATA TAGACACATA CAGGAGCGAA ACGAGAGGCG
 20 161 AAGTCAAGGT GCCGGACCCG TACCACTGGC TAGAGGAATA
 201 TTCAGAAGAG ACGGACAAGT GGACGTCCGA CCAGGAGGAG
 241 TTCACGAGGA CATATTTGGA CAGCAACCCT GATCGAAAGA
 281 AGCTAGAAGA CGCATTGAGA AAGAGTATGG ATTATCCCAA
 321 GTTCTCCGCT CTTTTTTTGA ATGATGACAA GCGATGGTAT
 25 361 TGGTTTTACA ATACCGGCCT TCAAGCACAA ACAGTCATCT
 401 GCAGATCAAA GGATGAGACT CTTCCCGACT TCTCAGAGAG
 441 TGACTIONGTC GGGGAAACAT TTTTGTATCC GAACCTATTA
 481 TCCTCGGATG GCACAGCCTC GCTCTCCATG TATGATTTCT
 521 CACACTGTGG CAAATACTTC GCATATGGTA TTTCTCTTTC
 30 561 CGGGAGCGAT TTTTCAACTA TATACGTACG GTCAACTTCC
 601 TCTCCACTGG CCCCTGGCAA CAACAGCATT AGAAATGACG
 641 ACGGTAGACT TCCAGACGAG CTTAGATATG TCAAATTTTC
 681 CTCCATCAGC TGGACAAAGG ACTCCAAAGG ATTTTCTAT
 721 CAGCGCTATC CCGGTACAGG CACTGTGAAT GGACAGAATG
 35 761 GCATCCAAAC TCAAGGCGAT CGTGATGCTA TGATTTACTA
 801 TCACCGGATA GGGACATCAC AATCCGATGA TATTCTTGTC
 841 CATGAAGACC AGGAACATCC TGATTGGGTA TTTGGCGCAG
 881 AAGTCACGGA AGATGGTAAA TATGTGGCCC TGTACACAAT
 921 GAAGGACACA TCAAGGAAAA ATCTATTGTG GATTGCTGAT
 40 961 CTGGGACAAA ACGAAGTTGG ACGAAACATG AAATGGAACA
 1001 AGATTTGCAA CGTTTTTGAC TCAGAATACG ACCTAATTGG
 1041 CAACGACGGT TCATTACTAT ACATCAGAAC TAATAAAGCT
 1081 GCACCTCAAT ACAAGATTGT CACCTTAGAT ATAGAGAAAC
 1121 CAGAATTAGG GTTTAAGGAA TTCATACCGG AAGATCCCAA
 45 1161 AGCATATCTC TCTCAAGTCA AAATTTTTTAA TAAGGATAGA
 1201 CTAGCACTAG TATACAAGCG TAACGTTATA GGCGAACTCT
 1241 ACGTCTACAA TAACACTGGGTCACGACTAA TGCGCCTAGC
 1281 CCGGGACTTT GTTGGCTCCA TGACGGTGAC CGCTCGAGAA
 1321 ACGGAGCCAT GTTTTTTTGC CACTCTCACG GGCTTCAATA

1361 CCCCTGGAAT CGTATGCAGG TACAATATCC AGCGACCGGA
 1401 AGAACAGCGT TGGAGCGTAT ATCGAACTGC CAAGGTCAAG
 1441 GGTTTAAATC CGAACGATTT CGAGGCTCGA CAGGTGTGGT
 1481 ATGACAGCTA CGATGGAACA AAGATTCCAA TGTTTCATCGT
 5 1521 CCGTCACAAG AATACCCAAT TTAATGGGAC GGCGCCAGCT
 1561 ATACAATATG GTTACGGTGG CTTTAATATA TCTATAAATC
 1601 CCTTCTTTAG TCCAACGATT TTGACGTTCT TGCAAAAAGTA
 1641 TGGAGCAATT CTAGCTGTAC CTAATATCCG AGGAGGCGGC
 1681 GAGTTCGGCG AGACATGGCA TGATGCTGGT ATACGAGAGA
 10 1721 AACGAGCTAA TGTTTACGAT GATTTTCATTG CGGCAACTCA
 1761 GTTCTTGGTA AAAAACAAGT ATGCCGCGGG CGGCAAAGTG
 1801 GCCATCAACG GGGGGTCCAA TGGAGGACTT TTGGTCGCGG
 1841 CCTGTGTCAA TCGTGCACGT GAAGGAACCT TTGGAGCTGC
 1881 CATTGCTGAA GTTGGGGTCC TAGACTTGCT CAAGTTCCCC
 15 1921 AAATTTACCA TAGGCAAAGC TTGGATTAGC GACTACGGCG
 1961 ATCCAGAAGA TCCGCGTGAT TTTGATTACA TTTACACACA
 2001 TTCACCACTT CATAATATAC CAAAGAACAT GGTCTTACCT
 2041 CCGACGATGC TTCTGACAGC TGATCATGAT GACCGTGTCTG
 2081 TCCCAATGCA TTCATTTAAG TATGCTGCAA TGCTACAATA
 20 2121 CACCCTGCCG CATAATCGTC ATCCACTTCT GCTACGTGTA
 2161 GACAAGAAAG CGGGGCATGG CGGAGGAAAA TCTACTGAGA
 2201 AGAGGTTACA GGAGGCTGCC GACAAATGGG GTTTTGGCCG
 2241 GCAGTCCATG GGTCTTGCGT GGAAGGATAG ACAAGCTAAT
 2281 CTGTGA
 25

An example of a *Galerina marginata* prolyl oligopeptidase (POPA) sequence is shown below as SEQ ID NO:10.

1 MARTPWLPNA YPPARRSDHV DIYKSALRGD VRVQDPYQWL
 41 EEYTDETDKW TTAQEVFTRT YLDKNPDLPR LEKAFQACND
 30 81 YPKSYAPYLH DDNRWYWYYN SGLEPQTALY RSKDSSLPDL
 121 STADGSGGDL FFDPNALSND GTAALSTYAF SDCGKYFAYG
 161 ISFSGSDFVT IYVRLTDSPL TKDVEDAKNDK GRLPEEIKFV
 201 KFSSIGWTPD SKGFFYQRY P DTSTVTQENG PIATEGDLDA
 241 MVYYHRLGTP QSEDTLIYQD KEHRDWMFSI DVTDDGNYLL
 35 281 LYILKDSSRQ NLLWIAAFDP ANLGPNIKWQ KVFDEYHSEY
 321 EIITNKGSLF YVRTNESAPQ YRVITVDIAK GNEINELIPE
 361 TDAYLSSITS VNKG YFALVY KRN VKDEVYV YSHAGNQLAR
 401 LAEDFVGAAH VSGREKHSSF FVELNGFTSP GTIGRYKFTD
 441 PEEQRWSIYR TTKLNLNTE DFEASQVWYE SKDGTSIPMF
 40 481 IVRHKSTKFD GTAPVIQYGY GGFSISIDPF FSATILTFLQ
 521 KYGVVFALPN IRGGGEFGED WHLAGCREKK GNCFDDFIAA
 561 TQYLVKNKYA APDKVTINGG SNGGLLV SAC VNRAPECTFG
 601 CAVADVGVHD LLKFHKFTIG KAWTS DYGNP DDPNDFDFIF
 641 PISPLQNI PK DKVFPPMLLL TADHDDRVPV MHSFKLAAEL
 45 681 QYSLPHNP NP LLIRIDKKAG HGAGKSTQQK IKESADKWGF
 721 VAQSLGLVWK DSTEQPNL

The following comparison of *Amanita bisporigera* (SEQ ID NO:8) and *Galerina marginata* (SEQ ID NO:10) prolyl oligopeptidase (POPA) sequences is shown below, illustrating conserved and non-conserved sequence regions or domains, where the asterisks illustrate which amino acids are the same between these two proteins.

```

5  Seq8   25 MSSTQWTPNMYPSARRSDHIDTYRSETRGEVKVPDPYHWLEEYSEETDKWTSdqEEFTRT
   Seq10  1  MARTPWLPNAYPPARRSDHVDIYKSALRGDVRVQDPYQWLEEYTDETDKWTTAQEVFTRT
      * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

   Seq8   85 YLDSNPDRKKLEDAFRKSMDDPKFSAPFLNDDKRWYWFYNTGLQAQTVCRSKDETLPDF
10  Seq10  61 YLDKNPDLPRLEKAFQACNDYPKSYAPYLHDDNRWYWYNSGLEPQTALYRSKDSLSPLD
      *** ** * * * * * * * * * * * * * * * * * * * * * * * * * * * *

   Seq8  145 SESDYVG--ETFFDPNLLSSDGTASLSMYDFSHCGKYFAYGISLSGSDFSIYVRSTSSPL
   Seq10 121 STADGSGGDLFFDPNALSNDGTAALSTYAFSDCGKYFAYGISFSGSDFVTIYVRLTDSPL
15      * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

   Seq8  204 APGNNSIRNDDGRLPDELRYVKFSSISWTKDSKGFFYQRYPGTGTVNGQNG-IQTQGD RD
   Seq10 181 TKDVDA-KNDKGRLPPEIKFVKFSSIGWTPDSKGFFYQRYPDTSTVTQENGPIATEGDL D
      * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

20  Seq8  263 AMIYYHRIGTSQSDDILVHEDQEHDPDWVFGAEVTEDGKYVALYTMKDTSRKNLLWIADLG
   Seq10 240 AMVYYHRLGTPQSEDTLIYQDKHRDWMFSIDVTDDGNYLLLYILKDSSRQNLWLWIAAFD
      ** * * * * * * * * * * * * * * * * * * * * * * * * * * * *

25  Seq8  323 QNEVGRNMKWNKICNVFDSEYDLIGNDGSLLYIRTNKAAPQYKIVTLDIEKPELGFKEFI
   Seq10 300 PANLGPNIKWQKVFDEYHSEYEIITNKGSLFYVRTNESAPQYRVITVDIAKGN-EINELI
      * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

   Seq8  383 PEDPKAYLSQVKIFNKDRLALVYKRNVI GELYVYNN TGSRLMLRLARDFVGSMTVTARETE
30  Seq10 359 PETD-AYLSSIISV NKGYFALVYKRNKDEYVYSHAGNQLARLAEDFVGAHVSGREKH
      ** * * * * * * * * * * * * * * * * * * * * * * * * * * * *

   Seq8  443 PWFFATLTGFNTPGIVCRYNIQRPEEQRWSVYRTAKVKGLNPNDFEARQVWYDSYDGTKI
   Seq10 418 SSFFVELNGFTSPGTIGRYKFTDPEEQRWSIYRTTKLNLNTEDFEASQVWYESKDGTSI
35      ** * * * * * * * * * * * * * * * * * * * * * * * * * * * *

   Seq8  503 PMFIVRHKNTQFNGTAPAIQYGYGGFNISINPFFSPTILTF LQKYGALLAVPNIRGGGEF
   Seq10 478 PMFIVRHKSTKFDGTAPVIQYGYGGFSISIDPFFSATILTF LQKYGVVFALPNIRGGGEF
40      ***** * * * * * * * * * * * * * * * * * * * * * * * * * * * *

   Seq8  563 GETWHDAGIREKRANVYDDFIAATQFLVKNKYAAGGKVAINGGSNGG LLVAACVNRAREG
   Seq10 538 GEDWHLAGCREKKGNCFDDFIAATQYLVKNKYAAPDKVTINGGSNGG LLVSACVNRAREG
      ** * * * * * * * * * * * * * * * * * * * * * * * * * * * *

45  Seq8  623 TFGAAIAEVGVLDLLKFPKFTIGKAWISDYGD PEDPRDFDIYTHSPLHNIPKNMVL PPT
   Seq10 598 TFGCAVADVGVHDL LKFKFTIGKAWTS DYGNPDDPNDFDFI FPI SPLQNI PKDKVFP PM
      *** * * * * * * * * * * * * * * * * * * * * * * * * * * * *

   Seq8  683 MLLTADHDDRVP MHSFKYAAMLQYTLPHNRHPLLLRV DKKAGHGGGKST EKR LQEAADK
50  Seq10 658 LLLTADHDDRVP MHSFKLAELQYSLPHNPNPLLIRIDKKAGHGAGKST QQKIKESADK
      ***** * * * * * * * * * * * * * * * * * * * * * * * * * * * *

   Seq8  743 WGFAAQSMGLAWKD
   Seq10 718 WGFVAQSLGLVWKD
55      *** * * * * * * * * * * * * * * * * * * * * * * * * * * * *

```

As illustrated, the *Amanita bisporigera* (SEQ ID NO:8) and *Galerina marginata* (SEQ ID NO:10) prolyl oligopeptidase (POPA) sequences have about 65% sequence identity.

However, POPA from *Amanita bisporigera* and *Galerina marginata* do not have significant macrocyclase activity, and instead POPA enzymes often just cut small peptides at proline residues.

Prolyl oligopeptidase B (POPB) is an enzyme that can cyclize peptides. An example of an *Amanita bisporigera* prolyl oligopeptidase (POPB) sequence is shown below as SEQ ID NO:11.

```

1  MPPTPWAPHS  YPPTRRSDHV  DVYQSASRGE  VPVPDPYQWL
10  41  EENSNEVDEW  TTAQTAFTQG  YLDKNADRQK  LEEKFRASKD
    81  YVKFSAPTRL  DSGHWYWFYN  SGVQSQAVLY  RSKKPVLPDF
    121 QRGTRKVGEV  YFDPNVLSAD  GTAIMGTCRF  SPSGEYFAYA
    161 VSHLGVDYFT  IYVRPTSSSL  SQAPEAEGGD  GRLSDGVKWC
    201 KFTTITWTKD  SKGFLYQRY  ARESLVAKDR  DKDAMVCYHR
15  241 VGTTQLEDII  VQQDKENPDW  TYGTDASEDG  KYIYLVVYKD
    281 ASKQNLLWVA  EFDKDGVKPE  IPWRKVINEF  GADYHVIINH
    321 GSLIYVKTNV  NAPQYKVVTI  DLSTGEPEIR  DFIPEQKDAK
    361 LTQVKCVNKG  YFVAIYKRN  KDEIYLYSKA  GDQLSRLASD
    401 FIGVASITNR  EKQPHSFLTF  SGFNTPGTIS  RYDFTAPDTQ
20  441 RLSILRTTKL  NGLNADDFES  TQVWYKSKDG  TKVPMFIVRH
    481 KSTKFDGTAP  AIQNGYGGFA  ITADPFFSPI  MLTFMQTYGA
    521 ILAVPNIRGG  GEFGGGEWHKA  GRRETKGNTF  DDFIAAAQFL
    561 VKNKYAAPGK  VAITGASNGG  FLVCGSVVRA  PEGTFGAAVS
    601 EGGVADLLKF  NKFTGGMAWT  SEYGNPFIKE  DFDFVQALSP
25  641 VHNVPKDRVL  PATLLMTNAG  DDRVPMHSL  KFVANLQYNV
    681 PQNPHPLLIR  VDKSWLGHGF  GKTTDKHTKD  AADKWSFVAQ
    721 SLGLEWKTVD

```

An example of a nucleic acid (cDNA) sequence that encodes the *Amanita bisporigera* prolyl oligopeptidase (POPB) protein is shown below as SEQ ID NO:12.

```

1  ATGCCCCCTA  CACCATGGGC  TCCTCACAGT  TATCCTCCTA
41  CCCGTCGTTT  TGACCACGTT  GATGTATATC  AGAGCGCATC
81  CAGAGGCGAA  GTACCAGTAC  CGGACCCGTA  CCAATGGCTG
121 GAGGAGAATT  CAAATGAAGT  AGACGAATGG  ACGACGGCGC
35  161 AGACAGCTTT  CACGCAAGGC  TATCTTGATA  AGAATGCGGA
    201 TAGACAGAAG  CTCGAGGAGA  AATTTCTGTC  AAGCAAGGAC
    241 TACGTCAAGT  TTTCTGCGCC  AACTCTGCTT  GATAGTGGAC
    281 ACTGGTATTG  GTTCTACAAT  AGCGGCGTAC  AATCGCAAGC
    321 AGTCCTCTAC  CGCTCCAAGA  AACCCGTCTT  TCCTGATTTT
40  361 CAAAGAGGGA  CGAGGAAAGT  CGGCGAAGTA  TACTTCGATC
    401 CAAACGTACT  CTCTGCTGAT  GGCACCGCAA  TTATGGGCAC
    441 GTGCCGATTC  TCCCCTAGTG  GCGAGTATTT  CGCATATGCA
    481 GTGTCCCACT  TGGGAGTTGA  TTATTTTACT  ATCTATGTTC
    521 GCCCTACGAG  TTCATCATTG  TCTCAAGCTC  CGGAAGCTGA
45  561 AGGTGGGGAT  GGTCGATTGT  CGGATGGAGT  GAAATGGTGC

```

```

5   601 AAGTTTACGA CTATAACGTG GACAAAGGAC TCCAAAGGAT
    641 TTCTTTACCA GCGGTACCCT GCTCGGGAAT CTCTTGTTGGC
    681 GAAAGATCGT GATAAAGATG CTATGGTATG CTATCATAGG
    721 GTTGGAACGA CTCAATTGGA AGATATCATT GTCCAACAAG
    761 ACAAGGAGAA CCCAGACTGG ACATATGGAA CAGATGCGTC
    801 AGAGGACGGC AAATATATCT ACTTAGTGGT ATACAAGGAT
    841 GCCTCGAAGC AAAATCTTCT GTGGGTTGCA GAATTCGACA
    881 AGGACGGGGT CAAGCCGGAA ATTCCCTGGC GAAAAGTCAT
    921 CAATGAGTTT GGGGCGGATT ACCATGTTAT CACGAACCAC
10  961 GGATCTTTGA TCTATGTCAA GACTAACGTG AATGCGCCCC
    1001 AATATAAAGT TGTCACTATC GACCTTTCGA CAGGAGAACC
    1041 CGAAATTCGT GATTTTCATC CGGAACAGAA AGATGCGAAG
    1081 CTCACTCAAG TCAAATGCGT CAACAAGGGA TATTTCTGTCG
    1121 CGATCTACAA GCGCAATGTC AAAGATGAAA TATATCTTTA
15  1161 CTCCAAAGCA GCGCATCAAC TCAGTCGTCT GGCGTCGGAC
    1201 TTCATTGGCG TTGCATCTAT AACTAACAGA GAGAAACAAC
    1241 CTCATTCTTT CCTCACTTTC TCTGGATTTA ACACGCCGGG
    1281 CACCATTCTT CGCTACGATT TTACAGCTCC AGACACACAA
    1321 CGTCTCAGCA TCCTTAGGAC TACGAAGCTA AATGGTCTGA
20  1361 ATGCAGATGA CTTTGAGAGC ACACAAGTCT GGTATAAGAG
    1401 CAAAGACGGA ACGAAAGTTC CAATGTTTCAT CGTTCGTCAC
    1441 AAATCAACGA AATTTGACGG AACGGCGCCG GCGATTCAAA
    1481 ACGGTTATGG TGGTTTCGCT ATTACAGCCG ATCCATTCTT
    1521 TAGTCCCATC ATGCTCACCT TTATGCAGAC ATATGGCGCA
25  1561 ATCCTGGCTG TCCCGAACAT CAGAGGTGGA GGTGAATTCG
    1601 GCGGAGAATG GCACAAGGCA GGGAGACGAG AAACCAAGGG
    1641 AAATACTTTT GATGATTTCA TCGCTGCCGC TCAATTTCTT
    1681 GTCAAAAACA AGTACGCGGC TCCAGGCAAG GTGGCCATCA
    1721 CTGGTGCATC CAATGGCGGT TTTCTTGTCT GTGGTTCCGT
30  1761 AGTTCGGGCA CCAGAGGGAA CATTCCGGCG TGCTGTTTCC
    1801 GAAGGTGGTG TCGCGGACCT CCTAAAGTTT AATAAATTCA
    1841 CCGGGGGGAT GGCGTGGACG AGTGAATATG GAAACCCTTT
    1881 TATTAAGGAG GACTTCGACT TTGTCCAAGC ATTGTCTCCT
    1921 GTGCATAACG TACCCAAGGA TAGGGTTCTT CCTGCCACAT
35  1961 TACTTATGAC CAATGCGGGT GACGATCGTG TAGTTCCAAT
    2001 GCATTGCTTT AAGTTCGTCT CAAACCTTCA GTACAATGTG
    2041 CCTCAAAATC CTCATCCATT GCTCATCCGT GTGGATAAAT
    2081 CTTGGCTTGG TCATGGTTTT GGCAAGACAA CAGACAAGCA
    2121 TACCAAAGAT GCTGCGGACA AGTGGAGTTT CGTAGCGCAA
40  2161 TCGTTAGGGC TAGAATGGAA AACGGTTGAC TAG

```

An example of an *Galerina marginata* prolyl oligopeptidase (POPB) sequence is shown below as SEQ ID NO:13.

```

45  1  MSSVTWAPGN YPSTRSDHV DTYQSASKGE VPVPDPYQWL
    41  EESTDEVKDW TTAQADLAQS YLDQNADIQK LAEKFRASRN
    81  YAKFSAPTLL DDGHWWFYFN RGLQSQSVLY RSKEPALPDF
   121  SKGDDNVGDV FFDPNVLAAD GSAGMVLCKF SPDGKFFAYA
   161  VSHLGGDYST IYVRSTSSPL SQASVAQGVD GRLSDEVKWF
   201  KFSTIIWTKD SKGFLYQRYP ARERHEGTRS DRNAMMCYHK

```

241 VGTQEEEDII VYQDNEHPEW IYGADTSEDG KYLYLYQFKD
 281 TSKKNLLWVA ELDEDEGVKSG IHWRKVVNEY AADYNIITNH
 321 GSLVYIKTNL NAPQYKVITI DLSKDEPEIR DFIPEEKDAK
 361 LAQVNCANEE YFVAIYKRVN KDEIYLYSKA GVQLTRLAPD
 5 401 FVGAASIANR QKQTHFFLTL SGFNTPGTIA RYDFTAPETQ
 441 RFSILRTTKV NELDPDDFES TQVWYESKDG TKIPMFIVRH
 481 KSTKFDGTAA AIQYGYGGFA TSADPFFSPI ILTFLQTYGA
 521 IFAVPSIRGG GEFGEWHKG GRRETKVNTF DDFIAAAQFL
 561 VKNKYAAPGK VAINGASNGG LLVMGSIVRA PEGTFGAAMP
 10 601 EGGVADLLKF HKFTGGQAWI SEYGNPSIPE EFDYIYPLSP
 641 VHNVRTDKVM PATLITVNIG DGRVVPMSHF KFIATLQHNH
 681 PQNPHPLLIK IDKSWLGHGM GKPTDKNVKD AADKWGFAR
 721 ALGLELKTVE

15 The following comparison of *Amanita bisporigera* (SEQ ID NO:11) and *Galerina marginata*
 (SEQ ID NO:13) prolyl oligopeptidase (POPB) sequences is shown below, illustrating
 conserved and non-conserved sequence regions or domains, where the asterisks illustrate
 which amino acids are the same between these two proteins.

Seq11 1 MPPTPWAPHSYPPTRRS DHVDVYQSASRGEVPVDPYQWLEENSNEVDEWTTAQTAFQTG
 Seq13 1 MSSVTWAPGNYPSTRRS DHVDIYQSASKGEVPVDPYQWLEESTDEVDKWTAAQADLAQS
 * *** ** ***** ***** ***** ** *

 Seq11 61 YLDKNADRQKLEEKFRASKDYVKFSAPTLLDSGHWYWFYNSGVQSQAVLYRSKKPVLPDF
 Seq13 61 YLDQNADIQKLAEEKFRASRNIAKFSAPTLLDDGHWYWFYNRGLQSQSVLYRSKEPALPDF
 25 *** ** * ***** * ***** ***** * *** ***** *

 Seq11 121 QRGTRKVGVEYFDPNVLSADGTAIMGTCRFSPEGEYFAYAVSHLGVDYFTIYVRPTSSSL
 Seq13 121 SKGDDNVGDVFFDPNVLAADGSAGMVLCKFSPDGKFFAYAVSHLGGDYSTIYVRSTSSPL
 * * * ***** * * * * * ***** * * * * *
 30
 Seq11 181 SQAPEAEGGDGRLSDGVKWKCFITITWTKDSKGFYQRYPARSLVAKDRDKDAMVCYHR
 Seq13 181 SQASVAQGVDGRLSDEVKWKCFSTIITWTKDSKGFYQRYPARERHEGTRSDRNAMMCYHK
 *** * * ***** ** * * * ***** ***** * ** *

 Seq11 241 VGTQLEEDII VYQDNEHPEW IYGADTSEDG KYLYLYQFKDTSKKNLLWVAELDEDEGVKSG
 Seq13 241 VGTQEEEDII VYQDNEHPEW IYGADTSEDG KYLYLYQFKDTSKKNLLWVAELDEDEGVKSG
 ***** ***** * * * * * ***** * * * * *

 Seq11 301 IPWRKVINFEFGADYHVITNHGSLIYVKTNVNAPQYKVITIDLSGTEPEIRDFIPEEKDAK
 Seq13 301 IHWRKVVNEYAADYNIITNHGSLVYIKTNLNAPQYKVITIDLSKDEPEIRDFIPEEKDAK
 * **** ** * ***** * * * ***** ***** *****

 Seq11 361 LTQVKCVNKG YFVAIYKRVN KDEIYLYSKAGDQLSRLASDFIGVASITNREKQPHSFLTF
 Seq13 361 LAQVNCANEE YFVAIYKRVN KDEIYLYSKAGVQLTRLAPDFVGAASIANRQKQTHFFLTL
 45 * * * * ***** * * * * * * * * * *

 Seq11 421 SGFNTPGTISRYDFTAPDTQRLSILRTTKLNLNADDFESTQVWYKSKDGTKVPMFIVRH
 Seq13 421 SGFNTPGTIARYDFTAPETQRFSLRTTKVNELDPDDFESTQVWYKSKDGTKIPMFIVRH
 ***** ***** * * ***** *****
 50
 Seq11 481 KSTKFDGTAPAIQNGYGGFAITADPFFSPIMLTFMQTYGAILAVPNIRGGGEFGGEWHKA
 Seq13 481 KSTKFDGTAAAIQYGYGGFATSADPFFSPIILTFMQTYGAIFAVPSIRGGGEFGGEWHKG
 ***** ***** ***** ***** ***** *****

```

Seq11 541 GRRETkgntfddFIAAAQFLVKNKYAAPGKVAITGASNGGFLVCGSVVRAPEGTFGAAVS
Seq13 541 GRRETkVNTfDDFIAAAQFLVKNKYAAPGKVAINGASNGGLLVMGSIvRAPEGTFGAAVP
*****
5 Seq11 601 EGGVADLLKFNKFTGGMAWTSEYGNPFIKEDFDFVQALSPVHNVPKDRVLPATLLMTNAG
Seq13 601 EGGVADLLKfHKFTGGQAWISEYGNPSIPEEFdYIYPLSPVHNVRIDKVMpATLITVNIG
*****
10 Seq11 661 DDRVVPMHSLKFVANLQYNVPQNPHPLLIRVDKSWLGHGFGKTTDKHTKDAADKWSFVAQ
Seq13 661 DGRVVPMHSfKFIAIQLQHNVPQNPHPLLIKIDKSWLGHGMGKPTDKNVKDAADKwGFIAr
*****
Seq11 721 SLGLEWKTVD
Seq13 721 ALGLELKTVE
15 *****

```

As illustrated, the *Amanita bisporigera* (SEQ ID NO:11) and *Galerina marginata* (SEQ ID NO:13) prolyl oligopeptidase (POPB) sequences have at least about 76% sequence identity.

An example of an *Hypsizygus marmoreus* prolyl oligopeptidase (POP-like) sequence is shown below as SEQ ID NO:14.

```

1 MAISPTPWTP NTYPPTRRSS HVDIYKSATR GEVRVADPYQ
41 WLEENTEETD KWTTAQEEFT RSYLDKNIDR QRLEDAFRIS
25 81 TDYAKFSSPT LYEDGRWYWF YNSGLQPQPL IYRSKGKTLF
121 DFSQDDNVVG EVFFDPNLLS DDGTAALSIY DFSDCGKYFA
161 YGISFSGSDF STIYVRSTES PLAKKNSGST DDDRLSDEIK
201 HVKFSAVTWT KDSKGFFYQR YPAHENAKEG IETGGDVDAM
241 IYYHVIGTSQ SEDILVHSDK SNPEWMSID ITEDGKYLIL
30 281 YTMKDSSRKN LMWIAELSKN EIGPNIQWNK IIDVFDAEYH
321 LITNDGPILY VKTNADAPQY KLVTMDisGD KDISRDLIPE
361 DKNANLVQVD CVNRDTFAVI YKRNVKDEIY LYSKTGIQLS
401 RLASDFVGAA SISRREKQPH FFVTMTGFST PGTvaryDFG
441 APEEQRWSIY RSVKVNGLNP DDFESKQVWY ESKDGTKIPM
35 481 FIVRHKATKF DGTAPAIQYG YGGFSISINP FFSPTILTFL
521 QTYGAVLAVP NIRGGAEFGE DWHKAGTREK KGNVFDDFVA
561 ATQYLvKNKY AGECKVAING GSNGGLLVGA CINRAPEGTF
601 GAAVAEVGVM DLLKFSKFTI GKAWTSdyGD PDDPKDFDFI
641 CPLSPLHNIP TDRVLPPTML LTADHDDRvV PMHSFKHAAT
40 681 LQYTLPHNPH PLVIRIDKKA GHGAGKSTEK RIKESADKWG
721 FVAQSLGLVW QEPA

```

The following comparison of *Amanita bisporigera* (SEQ ID NO:11) and *Hypsizygus marmoreus* (SEQ ID NO:14) prolyl oligopeptidase (POP) sequences is shown below, illustrating conserved and non-conserved sequence regions or domains, and sequence identity of about 63-64% between these two proteins.

```

Seq11 3 PTPWAPHSYPPTRRSDHVDVYQSASRGEVPVDPYQWLEENSNEVDEWTTAQTAFTQGYL
Seq14 5 PTPWTpNTYPPTRRSShVDIYKSATRGEVRVADPYQWLEENTEETDKWTTAQEEFTRSYL
*****
50 *****

```

Seq11 63 DKNADRQKLEEKFRASKDYVKFSAPITLLDSGHWYWFYNSGVQSQAVLYRSKKPVLPDFQR
 Seq14 65 DKNTDRQRLEDAFRTSTDYAKFSSPTLYEDGRWYWFYNSGLQPQPLIYRSKGKTLPDFSQ
 *** **

5 Seq11 123 GTRKVGVEYFDPNVLSADGTAIMGTCRFSPSGEYFAYAVSHLGVDYFTIYVRPTSSSLSQ
 Seq14 125 DDNVVGEVFFDPNLLSDDGTAALSIYDFSDCGKYFAYGISFSGSDFSTIYVRSTESPLAK
 *** **

10 Seq11 183 APEAEGGDGRLSDGVKWKCFITITWIKDSKGFYQRYPARESL----VAKDRDKDAMVCYH
 Seq14 185 KNSGSTDDRLSDEIKHVKFSAVTWTKDSKGFFYQRYPAHENAKEGIETGGDVDAMIYYH
 * **

15 Seq11 240 RVGTTQLEDIIVQQDKENPDWIYGTDA SEDGKYIYLVVYKDASKQNLLWVAEFDKDGVPK
 Seq14 245 VIGTSQSEDILVHSDKSNPEWMSIDITEDGKYLILYTMKSSRKNLMMWIAELSKNEIGP
 ** *

20 Seq11 300 EIPWRKVINEFGADYHVITNHGSLIYVKTNVNAPQYKVVITDLSTGEPEIRDFIPEQKDA
 Seq14 305 NIQWNKIIDVFDAEYHLITNDGPILYVKTINADAPQYKLVMTDISGDKDISRDLIPEDKNA
 * * *

Seq11 360 KLTQVKCVNKGYFVAIYKRVKDEIYLYSKAGDQLSRLASDFIGVASITNREKQPHSFLT
 Seq14 365 NLVQVDCVNRDTFAVIYKRVKDEIYLYSKTGIQLSRLASDFVGAASISSREKQPHFFVT
 * **

25 Seq11 420 FSGFNTPGTISR YDFTAPDTQRLSILRTTKLNLNADDFESTQVWYKSKDGTKVPMFIVR
 Seq14 425 MTGFSTPGTVARYDFGAPEEQRWSIYRSVKVNGLNPDDEFESKQVWYESKDGTKIPMFIVR
 ** *

30 Seq11 480 HKSTKFDGTAPAIQNGYGGFAITADPFFSPIMLTFMQTYGAILAVPNIRGGGEFGGEWHK
 Seq14 485 HKATKFDGTAPAIQYGYGGFSISINPFFSPTILTFMQTYGAVLAVPNIRGGAFFGEDWHK
 ** *

35 Seq11 540 AGRRETGKNTFDDFIAAAQFLVKNKYAAPGKVAITGASNGGFLVCGSVVRAPEGTFGA AV
 Seq14 545 AGTREKKGNVDFDFAATQYL VKNKYAGEGKVAINGGSNGGLLVGACINRAPEGTFGA AV
 ** *

Seq11 600 SEG G VADLLKFNKFTGGMAWTSEYGNPFKEDEDFVQALSPVHNVPKDRVLPATILLMTNA
 Seq14 605 AEVGVMDLLKFSKFTIGKAWTSDYGD PDDPKDFDFICPLSPLHNIPTRVLPPTMLLTAD
 * **

40 Seq11 660 GDDRVPVPMHSLKEFVANLQYNVPQNPHPLLIRVDKSWLGHGFGKTTDKHKDAADKWSEVA
 Seq14 665 HDDRVPVPMHSFKHAATLQYTLPHNPHPLVIRIDKK-AGHGAGKSTEKRIKESADKWGEVA
 ***** *

45 Seq11 720 QSLGLEWK
 Seq14 724 QSLGLVWQ
 ***** *

50 As illustrated, the *Amanita bisporigera* (SEQ ID NO:11) and *Hypsizygus marmoreus* (SEQ ID NO:14) prolyl oligopeptidase (POP) sequences have at least about 64% sequence identity.

An example of a *Termitomyces* sp. J132 prolyl oligopeptidase (POP-like) sequence is shown below as SEQ ID NO:15.

1 MTPTPWTPNS YPFTRSDHV DLYKSAMKGE VRVSDPYQWL
 55 41 EEYSDETDKW TTSQETFTRT YLDKNSDRQR LEDVFRSNMD
 81 YAKFSAPTLY DDGRWYWFYN SGLQPQSTLY RSRGNTLPDF


```

Seq11 477 IVRHKSTKFDGTAPAIQNGYGGFAITADPFFSPIMLTQMOTYGAILAVPNIRGGGEFGGE
Seq15 481 IVRHKTTPFDGTAPAIQYGYGGFSISINPFFSPTILTFLOTYGAILAVPNIRGGAEFGED
      *****
5  Seq11 537 WHKAGRRETKGNTFDDFIAAAQFLVKNKYAAPGKVAITGASNGGFLVCGSVVRAPEGTFG
Seq15 541 WHKAGTREKKVNVEDDFIAATEHLVKNKYAGAGKVAINGGSNGGLLVGACINRAPEGTFG
      *****
10 Seq11 597 AAVSEGGVADLLKENKFTGGMAWTSEYGNPF IKEDFDFVQALSPVHNVPKDRVLPATLLM
Seq15 601 AAVAEVGVMDLLKFSKFT-----IDRVLPPTLLL
      ***
15 Seq11 657 TNAGDDRVPMSLKFVANLQYNVPQNPHELLIRVDKSWLGHGFGKTTDKHTKDAADKWS
Seq15 630 TADHDDRVPMSFKLAATLQYTLPHNPHLLLRVDKK-SGHGAGKSTEKRIKEAADKWG
      *
20 Seq11 717 FVAQSLGLEWK
Seq15 689 FVAQSLGLVWK
      *****

```

In some embodiments, a cDNA clone encoding a cyclization enzyme such as POPB enzyme is synthesized, isolated, and/or obtained from a selected cell. In other embodiments, cDNA clones from other species (that encode a POP-related protein) are isolated from selected fungal tissues. For example, the nucleic acid encoding an enzyme (e.g. with cyclization activity) can be any nucleic acid with a coding region that hybridizes to SEQ ID NO:11, 13, 14, or 15. Such a cyclization-related nucleic acid can encode a protein that has POP activity.

In another example, the *POP* nucleic acid can encode a POP protein with an amino acid sequence that has at least 90%, or at least 95%, or at least 96%, or at least 97%, at least 98%, or at least 99% sequence identity to SEQ ID NO:11, 13, 14, or 15. Using restriction endonucleases, the entire coding sequence for a selected *POP* nucleic acid segment is subcloned downstream of the promoter in a 5' to 3' sense orientation.

In some cases, the cyclization proteins can be used in *in vitro* compositions, assay mixtures, and/or enzymatic manufacturing processes. The cyclization proteins in these compositions, mixtures, and processes can have amino acid sequences that have at least 90%, or at least 95%, or at least 96%, or at least 97%, at least 98%, or at least 99% sequence identity to SEQ ID NO: 11, 13, 14, or 15.

Expression Cassettes

The cross-bridge and cyclization enzymes described herein can be expressed from expression cassettes. Such an expression cassette (also referred to as a transgene) can include a promoter operably linked to a nucleic acid segment that encodes any of the FMO1 or other enzymes described herein.

The terms "expression vector" and "expression cassette" refer to a recombinant nucleic acid molecule containing a desired nucleic acid coding segment and appropriate nucleic acid sequences for the expression of the operably linked nucleic acid coding segment in a particular host cell or organism. Nucleic acid sequences for expression of proteins in prokaryotes usually include a promoter, an operator (optional), and a ribosome-binding site, often along with other sequences. Eukaryotic cells can utilize promoters, enhancers, and termination and polyadenylation signals.

Nucleic acid sequences for expression of proteins can be heterologous to the nucleic acid coding segment. A "heterologous nucleic acid," "heterologous gene" and "heterologous protein" indicate a nucleic acid, gene or protein, respectively, which has come from a source other than its native source. The terms "heterologous" and "exogenous" are sometimes used interchangeably with "recombinant." Such heterologous nucleic acid, heterologous genes, and/or heterologous proteins may have been artificially supplied to the biological system or host cell. In contrast, the terms "endogenous protein," "native protein," "endogenous gene," and "native gene" refer to a protein or gene that is native to the biological system, cell, species or chromosome under study. A "native" or "endogenous" nucleic acid is a nucleic acid that does not contain nucleotides or other features normally present in the chromosome in which it is normally found in nature.

Promoters provide for expression of mRNA from the *FMOI* and other nucleic acids. In some cases, the promoter can be a *FMOI* native promoter or a promoter that expresses a native cyclization enzyme. However, the promoter can in some cases be heterologous to the *FMOI* nucleic acid segment. Cyclization enzyme can also be expressed from heterologous promoters. In other words, such a heterologous promoter may not be naturally linked to such a *FMOI* and/or other nucleic acid segment. Instead, some expression cassettes and expression vectors have been recombinantly engineered to include a *FMOI* or other nucleic acid coding segment operably linked to a heterologous promoter. A *FMOI* or other nucleic acid coding segment is operably linked to the promoter, for example, when it is located downstream from the promoter.

A variety of promoters can be included in the expression cassettes and/or expression vectors. In some cases, the endogenous *FMOI* promoter, or an endogenous promoter that expresses a native cyclization enzyme, can be employed. Promoter regions are typically found in the flanking DNA upstream from the coding sequence in both prokaryotic and eukaryotic cells. A promoter sequence provides for regulation of transcription of the downstream gene sequence and typically includes from about 50 to about 2,000 nucleotide

base pairs. Promoter sequences can also contain regulatory sequences such as enhancer sequences that can influence the level of gene expression. Some isolated promoter sequences can provide for gene expression of heterologous DNAs, that is a DNA different from the native or homologous DNA.

5 Promoters can be strong or weak, or inducible. A strong promoter provides for a high level of gene expression, whereas a weak promoter provides for a very low level of gene expression. An inducible promoter is a promoter that provides for the turning on and off gene expression in response to an exogenously added agent, or to an environmental or developmental stimulus. For example, a bacterial promoter such as the P_{lac} promoter can be
10 induced to vary levels of gene expression depending on the level of isothiopyrogalactoside added to the transformed cells. Promoters can also provide for cell specific, tissue specific or developmental regulation. A strong promoter for heterologous DNAs can be advantageous because it provides for a sufficient level of gene expression for easy detection and selection of transformed cells and provides for a high level of gene expression when desired. In some
15 cases, the promoter within such expression cassettes / vectors can be functional in a fungal cell.

Expression cassettes / vectors can include, but are not limited to, fungal promoters of genes such as isocytichrome C1 (CYC1), alcohol dehydrogenase (ADH1), transcription elongation factor (TEF), yeast glyceraldehyde phosphodehydrogenase (GAPDH), yeast
20 phosphoglycerate kinase (PGK below), yeast cytochrome b5, Yb5 promoter, or a combination thereof.

Examples of other promoters, which may be useful include fungal and bacterial promoters. Some specific nonlimiting examples include; the *aprE* promoter or a mutant *aprE* promoter (see, WO 01/51643); the *aph* promoter of the *Streptomyces fradiae* aminoglycoside
25 3'-phosphotransferase gene; an *Aspergillus niger* glucoamylase (*glaA*) promoter; the glucose isomerase (GI) promoter of *Actinoplanes missouriensis* and the derivative GI (GIT) promoter (U.S. Pat. No. 6,562,612 and EPA 351029); the glucose isomerase (GI) promoter from *Streptomyces lividans* (e.g., SEQ ID NO: 1 in WO 03/089621), the short wild-type GI promoter (SEQ ID NO: 33 in WO 03/089621), the 1.5 GI promoter (SEQ ID NO: 31 in WO
30 03/089621), the 1.20 GI promoter (SEQ ID NO: 32 in WO 03/089621), or any of the variant GI promoters (SEQ ID NOs: 9-28 in WO 03/089621) as disclosed in WO 03/089621; the *cbh1*, *cbh2*, *egl1* and *egl2* promoters from filamentous fungi and specifically the *Trichoderma reesei* cellobiohydrolase promoter (GenBank Accession No. D86235); the *lacZ* and *tac*

promoters (Bagdasarion et al., 1983, *Gene* 26:273-282); the ermE promoter (Ward et al., 1986, *Mol. Gen. Genet.* 203:468-478 and Schmitt-John et al., 1992, *Appl. Microbiol. Biotechnol.* 36:493-498); and the *Bacillus subtilis* phage ϕ 29 promoters (Pulido et al., 1986, *Gene* 49:377-382). Promoters effective in *Streptomyces* are listed in Hopwood et al.,
 5 (Hopwood et al., Regulation of Gene Expression in Antibiotic-producing *Streptomyces*. In Booth, I. and Higgins, C. (Eds) SYMPOSIUM OF THE SOCIETY FOR GENERAL MICROBIOLOGY, REGULATION OF GENE EXPRESSION, Cambridge University Press, 1986 pgs. 251-276). *Streptomyces* phage promoters are also disclosed in Labes et al., 1997, *Microbiol.* 143:1503-1512. Other promoters that can be used are described in Deuschle et al., 1986 *EMBO J.*
 10 5:2987-2994 and WO 96/00787.

Plant promoters can also be used such as the CaMV 35S promoter (Odell et al., *Nature.* 313:810-812 (1985)), or others such as CaMV 19S (Lawton et al., *Plant Molecular Biology.* 9:315-324 (1987)), *nos* (Ebert et al., *Proc. Natl. Acad. Sci. USA.* 84:5745-5749 (1987)), *Adhl* (Walker et al., *Proc. Natl. Acad. Sci. USA.* 84:6624-6628 (1987)), sucrose
 15 synthase (Yang et al., *Proc. Natl. Acad. Sci. USA.* 87:4144-4148 (1990)), α -tubulin, ubiquitin, actin (Wang et al., *Mol. Cell. Biol.* 12:3399 (1992)), *cab* (Sullivan et al., *Mol. Gen. Genet.* 215:431 (1989)), PEPCase (Hudspeth et al., *Plant Molecular Biology.* 12:579-589 (1989)) or those associated with the R gene complex (Chandler et al., *The Plant Cell.* 1:1175-1183 (1989)). Further suitable promoters include the poplar xylem-specific secondary cell wall
 20 specific cellulose synthase 8 promoter, cauliflower mosaic virus promoter, the Z10 promoter from a gene encoding a 10 kD zein protein, a Z27 promoter from a gene encoding a 27 kD zein protein, inducible promoters, such as the light inducible promoter derived from the pea *rbcS* gene (Coruzzi et al., *EMBO J.* 3:1671 (1984)) and the actin promoter from rice (McElroy et al., *The Plant Cell.* 2:163-171 (1990)). Seed specific promoters, such as the
 25 phaseolin promoter from beans, may also be used (Sengupta-Gopalan, *Proc. Natl. Acad. Sci. USA.* 83:3320-3324 (1985)).

A *FMOI* nucleic acid or another enzyme-encoding nucleic acid segment can be combined with a promoter by standard methods to yield an expression cassette or transgene, for example, as described in Sambrook et al. (MOLECULAR CLONING: A LABORATORY
 30 MANUAL. Second Edition (Cold Spring Harbor, NY: Cold Spring Harbor Press (1989); MOLECULAR CLONING: A LABORATORY MANUAL. Third Edition (Cold Spring Harbor, NY: Cold Spring Harbor Press (2000)). Briefly, a plasmid containing a promoter can be constructed as described in Jefferson (*Plant Molecular Biology Reporter* 5:387-405 (1987)) or obtained from Takara/Clontech Lab in Mountain View, California (e.g., pBII21 or

pBI221). Typically, these plasmids are constructed to have multiple cloning sites having specificity for different restriction enzymes downstream from the promoter. The *FMO1* or other enzyme encoding nucleic acids can be subcloned downstream from the promoter region using restriction enzymes and positioned to ensure that the nucleic acid segment is inserted in proper orientation with respect to the promoter so that the coding segment can be expressed as sense or antisense RNA. Once the *FMO1* or other enzyme nucleic acid is operably linked to a promoter, the expression cassette so formed can be subcloned into a plasmid or other vector (e.g., an expression vector, such as the pESC-HIS vector illustrated herein for expression of GmFMO1).

In some cases, an endogenous *FMO1* gene can be modified to generate modified host cells and the host cells can express a modified FMO1 protein. Mutations can be introduced into *FMO1*-expressing cell genomes by introducing targeting vectors, T-DNA, transposons, nucleic acids encoding TALENS, CRISPR, or ZFN nucleases, and combinations thereof into a recipient host cell to create a transformed cell.

In some cases, the endogenous *FMO1* or cyclization enzyme gene can be deleted and host cells with such a deleted endogenous *FMO1* or cyclization enzyme gene can be transformed to include a modified *FMO1* or cyclization enzyme transgene, for example, by transformation of the host cells with a *FMO1* or cyclization enzyme expression cassette or expression vector.

Accordingly, the FMO1 and/or cyclization enzyme can be encoded within expression cassettes and expressed in host cells.

The host cells can be any eukaryotic or prokaryotic cell. In some cases, a host cell can be any fungal cell or microorganism. Several types of fungi are available for use as host cells. For example, a host cell may be located in a transgenic mushroom or other fungus such as *Saccharomyces cerevisiae* (yeast). Prokaryotes include but are not limited to gram negative or positive bacterial cells. Numerous cell lines and cultures are available for use as host cells. Host cells can be obtained through the American Type Culture Collection (ATCC), an organization that serves as an archive for living cultures and genetic materials (atcc.org).

An appropriate host can be determined by one of skill in the art based on the vector nucleic acid sequence and the desired result.

Bacterial cells used as host cells for expression vector replication and/or expression include, among those listed elsewhere herein, DH5-alpha, JM109, and KC8, as well as a number of commercially available bacterial hosts such as SURE™, Competent Cells and

SOLOPACK.TM Gold Cells (Stratagene, La Jolla). Alternatively, bacterial cells such as *E. coli* LE392 can be used as host cells.

In some embodiments, a host cell is used as a recipient for heterologous nucleic acids. A host cell may be "transfected" or "transformed," which refers to a process by which
5 exogenous nucleic acid is transferred or introduced into the host cell. An expression cassette, plasmid or cosmid, for example, can be introduced into a host cell, where the expression cassette, plasmid or cosmid can be integrated into the host cell genome or autonomously replicated. A transformed cell includes the primary subject cell and its progeny.

Examples of fungal cells that can serve as hosts include yeast cells, mold cells, and
10 mushroom cells. For example, the following types of cells can serve as host cells:
Acremonium chrysogenum, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*,
Cochliobolus carbonum, *Coprinus cinereus*, *Mucor circinelloides*, *Neurospora crassa*,
Saccharomyces cerevisiae, *Tolypocladium geodes*, *Trichoderma reesei*, *Ustilago maydis*, and
the like.

15 **Peptide Substrates**

Peptides are substrates for the FMO1 and/or cyclization enzymes described herein. A variety of different peptide substrate sequences can be cross-linked and/or cyclized. For example, although the product of POPB, cyclo(IWGIGCNP), may be a natural substrate for FMO1, FMO1 expressed in a cell-free *in vitro* system exhibited nonspecific FMO activity,
20 indicating that the FMO1 enzyme can be active on a variety of peptide substrates.

Peptide substrates can be of different sizes or lengths, especially when cyclization enzymes are present with FMO1 enzymes. For example, peptide substrates can have about 6 to about 100 amino acids, or about 7 to about 75 amino acids, or about 8 to about 50 amino acids, or about 8 to about 35 amino acids, or about 10 to about 30 amino acids.

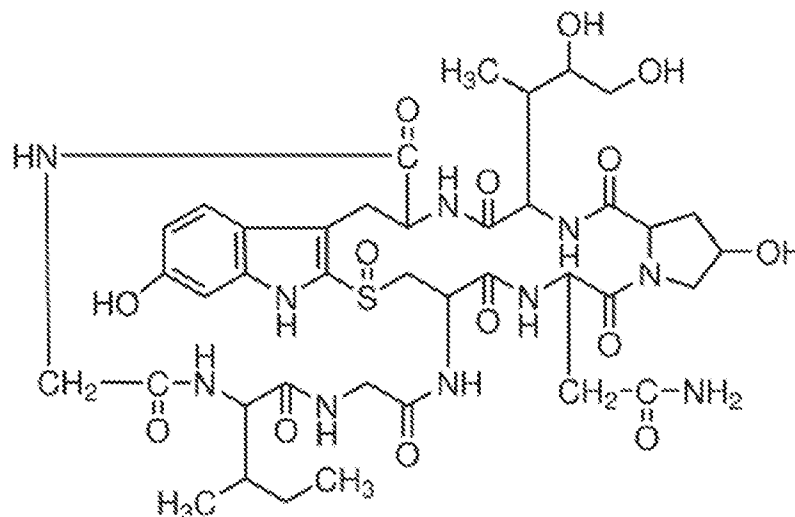
25 Peptide substrates can contain conventional and/or nonconventional amino acids. In some cases, the peptide substrates can contain hydroxylated amino acids, D-amino acids, atypical amino acids (e.g., ornithine, amino-isobutyric acid), N-methylated amino acids (e.g., N-methyl glycine), beta-amino acids (beta-alanine), or combinations thereof. For example, cyclo(IWGIGCNP) can be a substrate for FMO1 crosslinking, where the Ile, the Trp, and/or
30 the Pro are pre-hydroxylated. POPB can cyclize peptides containing hydroxylated, D-amino acids, N-methyl-amino acids, and beta-amino acids. FMO1 can also bicyclize cyclic peptides forms.

Precursor peptides that are cyclized can serve as peptide substrates for FMO enzymes. Alternatively, the precursor peptides can be linear and one or more cyclization enzymes can be used with the FMO enzymes.

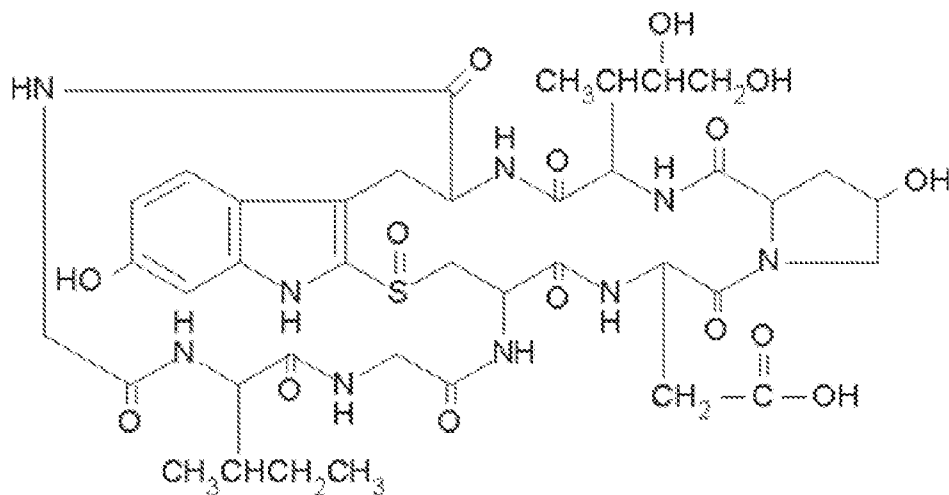
The peptide substrates can be made by biosynthesis of peptide substrate precursors in a host cell such as *E. coli*. In other cases, peptide substrates with or without unusual amino acids can be chemically synthesized.

In some cases, synthetic peptide substrates can be modified. To produce cyclized / crosslinked peptides with different sequences, peptide substrates can be synthesized by recombinant host cells that have expression cassettes with modified coding region segments. For example, selected nucleotides of synthetic *G. marginata* alpha-amanitin sequences can be changed to make other types of peptide toxins and peptides. In one example, replacing the codon AAC (Asn) with GAC (Asp) provides beta-amanitin instead of alpha-amanitin. Beta-amanitin production in *G. marginata* can readily be detected by reverse-phase HPLC.

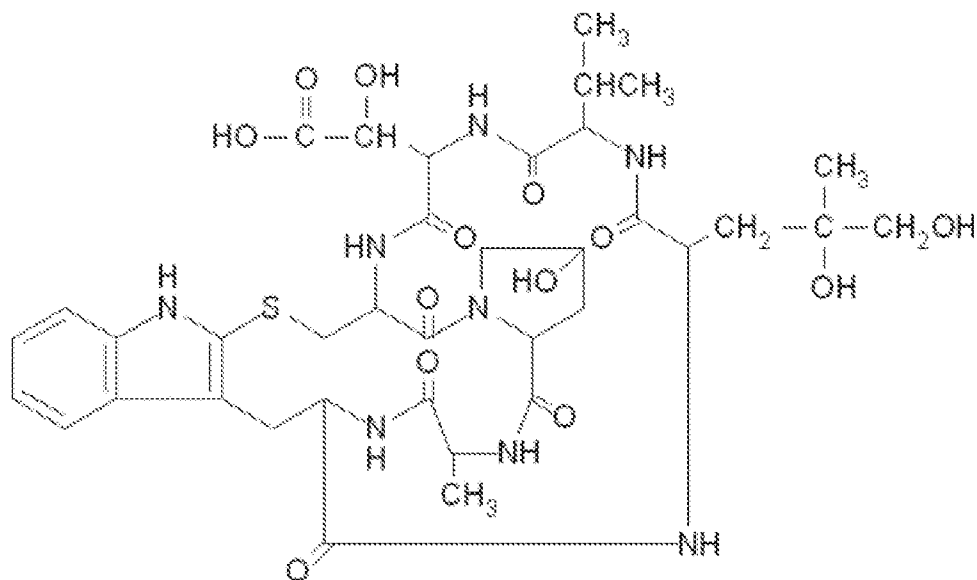
Alpha-amanitin, for example, is Asn-Hyp-Hyi-Trp-Gly-Ile-Gly-Cys [Bridge: 4-8] (also referred to as cyclic L-asparaginyl-4-hydroxy-L-prolyl-(*R*)-4,5-dihydroxy-L-isoleucyl-6-hydroxy-2-mercapto-L-tryptophylglycyl-L-isoleucylglycyl-L-cysteinyl) cyclic (4 → 8)-sulfide(*R*)-*S*-oxide), and it has the following structure.



Beta-amanitin has the following structure.



Phalloidin, for example, is cyclo(Ala-2-Mercapto-Trp-4,5-Dihydroxy-Leu-Val-*erythro*-3-Hydroxy-D-Asp-Cys-*cis*-4-Hydroxy-Pro), and it has the following structure.



5

Various amino acids in precursor peptides can be modified to make non-natural amanitin derivatives. One example of a non-natural amanitin derivative can involve replacing Gly with Ala by replacing GGT with GCT in a nucleic acid segment encoding a precursor peptide for amanitin.

10

The peptide substrates can contain Trp and Cys at positions convenient for forming tryptathionine cross-bridges (e.g., Trp and Cys may in some cases not be adjacent in the primary structure or sequence of the peptide substrates). Linear and cyclic peptides of at least six, seven, eight, nine, ten or more amino acids can be made that include the following sequence: XWXXXCXP (SEQ ID NO:16), where X is any amino acid. The proline can be

retained in these peptides for correct processing by POP, and the presence of Trp (W) and Cys (C) allow for biosynthesis of tryptathionine, a unique hallmark of the Amanita toxin peptides. Peptide sequences that include MDSTN (SEQ ID NO:17), TRIPL (SEQ ID NO:18), or other peptide sequences with prolines in conserved positions can be used in the peptide substrates, for example, when forming a cyclized peptide product.

The peptide substrates can have an amino acid sequence that includes a start site (e.g., methionine), a leader sequence (which is cleaved by POPB), a follower peptide that is removed when POPB cyclizes the core region (IWGIGCNP; SEQ ID NO:1) by transpeptidation, and a smaller peptide representing the amino acid sequence of the final peptide product (the core, IWGIGCNP; SEQ ID NO:1). One example of a peptide substrate (an exemplary precursor peptide) that includes an amino acid sequence for mature amatoxins and phallotoxins is illustrated as: MFDTNATRLP IWGIGCNPWTAEHVDQTLASGNDIC (SEQ ID NO:19), where the peptide that is cyclized is underlined. See also FIG. 1A. This is an example of an excellent peptide substrate for POPB.

Expression of synthetic peptides and peptide toxins can be monitored by assays including but not limited to, LC / MS and HPLC. Neither the precursor peptide nor the cyclic toxin peptide products are secreted. Hence, the cyclic and/or cross-linked products can be found inside host cells.

Separation of peptide products from endogenous peptides and proteins (e.g. peptides and proteins produced from genomic DNA originally contained in host cells) can be by standard techniques including but not limited to HPLC methods. Isolated peptide products can be evaluated in assays for assessing structure and/or biological activity. Peptide structures can be assessed by high resolution MS and NMR. The toxicity of peptide products such as synthetic amanitin toxins can be determined in assays that measure inhibition of transcription in eukaryotic cells, such as capability to inhibit RNA Polymerase II or other assays such as binding to actin, immunosuppression, or blocking the mitochondrial permeability transition pore. Useful peptide products can be produced in bulk (either within host cells or by reaction mixtures containing crosslinking and/or cyclization enzymes).

Table 1 illustrates exemplary sequences encoded by *AMAI* and *PHAI* genes.

Table 1: Exemplary AMAI and PHAI comparisons

Name	Proprotein	% Identity
------	------------	------------

Amanitin Asn-Hyp-Hyi-Trp-Gly-Ile-Gly-Cys [Bridge: 4-8]	MSDINATRLP <u>IWGIGCNP</u> CVGDDVTLLTRGEALC SEQ ID NO:20	100%
Phallacidin cyclo(Ala-2-Mercapto-Trp-4,5-Dihydroxy-Leu-Val- <i>erythro</i> -3-Hydroxy-D-Asp-Cys- <i>cis</i> -4-Hydroxy-Pro)	MSDINATRLP <u>AWLVDCP</u> CVGDDVNRLLTRGESLC SEQ ID NO:21	77% to Amanitin
Consensus peptide precursor re Amanitin and Phallacidin, where X is any amino acid	MSDINATRLP <u>XWXXXCXP</u> CVGDDVXXLLTRGELXC SEQ ID NO:22	74% to Amanitin
Peptide 1	MSDINATRLP <u>FNILPFMLPP</u> CVSDDVNILLTRGEDLC SEQ ID NO:23	67% to Amanitin 67% to Phallacidin
Peptide 2 Beta-amanitin	MSDINATRLP <u>IWGIGCDP</u> CIGDDVTILLTRGEALC SEQ ID NO:24	91% to Amanitin 74% to Phallacidin
Peptide 3	MSDINATRLP <u>IVGILGLPCI</u> GDDVNSTLTHGEDLC SEQ ID NO:25	71% to Amanitin 66% to Phallacidin
Peptide 4 Phalloidin	MSDINTTCLPAWLATCPCTG DDVNPTLTCGESLC SEQ ID NO:26	63% to Amanitin 76% to Phallacidin

Methods

The cyclic and/or crosslinked peptides that can be made using the enzymes described herein can have useful properties. For example, cyclic, crosslinked peptide toxins of *Amanita*

mushrooms are chemically and biologically unique among natural products. Such toxins have uses as biochemical reagents and as anti-cancer therapeutics, for example, in targeted therapeutics such as antibody-drug conjugates. The enzymes described herein can be used to biosynthesize many types of cyclic / crosslinked peptides.

5 Cyclic and/or crosslinked peptides can be made *in vivo* (e.g., in host cells or host organisms) or *in vitro* (e.g., in a reaction mixture containing one or more enzymes, a peptide substrate, and any other cofactors or other reactants that may facilitate the cyclization or crosslinking reaction).

 For *in vivo* production of cyclic / crosslinked peptides, peptide substrates can be co-
10 expressed in a host cell with a FMO1 and/or other cyclization enzyme. For example, an expression cassette can be used that has a promoter operably linked to a nucleic acid segment encoding a peptide substrate. Such peptides can then be expressed with FMO1 and/or cyclization enzymes so that cyclization and/or cross-bridge formation can occur in a host cell. The expression cassette encoding the peptide substrate can have any of the promoters
15 described herein. Similarly, the FMO1 and/or cyclization enzymes can be expressed from expression cassettes within the host cells. After incubation of the host cells in a culture medium that supports the growth and functioning of the host cells, the cyclic / crosslinked peptides can be isolated from the cell culture.

 For *in vitro* production of cyclic / crosslinked peptides, peptide substrates can be
20 mixed with one or more FMO1 and/or other cyclization enzymes in a reaction mixture. The reaction mixture can contain other factors to enhance the formation of cyclic / crosslinked peptide products. For example, the reaction can contain one or more of the following: NADPH, FAD, ATP, oxygen, dithiothreitol to maintain free sulfhydryls, buffer (to maintain the pH), or metal cofactors.

25 The reaction temperature that allows synthesis of peptide is 0 to 60° C, or 5 to 40° C. In addition, the reaction pH that allows synthesis of peptide is 6.5 to 10.5, and preferably 6.5 to 9.5.

 For example, bicyclic peptides can be produced using the enzymes and methods described herein. Host cells such as yeast, bacterial (e.g., *E. coli*), or *Galerina* cells can be
30 transformed with one or more copies of an expression cassette encoding a substrate precursor peptide useful for making a cyclic or bicyclic peptide, and one or more expression cassettes encoding FMO1 and/or cyclization enzymes. Alternatively, for example, a reaction mixture of peptide substrates for such cyclic or bicyclic peptides as well as FMO1 and/or cyclization enzymes can be incubated *in vitro* to generate the cyclic or bicyclic product.

Definitions

The use of the article "a" or "an" is intended to include one or more.

As used herein, terms defined in the singular are intended to include those terms defined in the plural and vice versa.

5 As used herein, "peptide" refers to compounds containing two or more amino acids linked by the carboxyl group of one amino acid to the amino group of another, i.e. "peptidyl linkages" to form an amino acid sequence. It is contemplated that peptides may be purified and/or isolated from natural sources or prepared by recombinant or synthetic methods. Amino acid sequences may be encoded by naturally or non-naturally occurring nucleic acid
10 sequences or synthesized by recombinant nucleic acid sequences or artificially synthesized. A peptide may be a linear peptide or a cyclopeptide, i.e. cyclic peptide including bicyclic feature.

As used herein, "cyclic peptide" or "cyclopeptide" in general refers to a peptide comprising at least one internal bond attaching nonadjacent amino acids of the peptide, such
15 as when the end amino acids of a linear sequence are attached to form a circular peptide. A "bicyclic peptide" may have at least two internal bonds forming a cyclopeptide of the present inventions, such as when the end amino acids of a linear sequence are attached to form a circular peptide in addition to another internal bond attaching two nonadjacent amino acids, for example, see e.g. FIG. 1.

20 As used herein, the term "Amanita peptide" or "Amanita toxin" or "Amanita peptide toxin" refers to any linear or cyclic peptide produced by a mushroom, not restricted to a biologically active toxin. It is not intended that the present invention be limited to a toxin or a peptide produced by an *Amanita* mushroom and includes similar peptides and toxins produced by other fungi, including but not limited to species of *Lepiota*, *Conocybe*, *Galerina*,
25 and the like. An Amanita peptide toxin resembles any of the amatoxins and phallotoxins, such as similarity of amino acid sequences, matching toxin motifs as shown herein, encoded between the conserved regions (A and B) of their proproteins, encoded by hypervariable regions of their proproteins (P) and the like. The Amanita peptides include, but are not restricted to, amatoxins such as the amanitins, and phallotoxins such as phalloidin and
30 phallacidin. For example, an exemplary Amanita peptide in one embodiment ranges from 6-17 amino acids in length. In another embodiment an Amanita peptide toxin ranges from 7-11 amino acids in length. In one embodiment, an Amanita peptide is linear. In another embodiment, an Amanita peptide is a bicyclic peptide. It is not meant to limit an Amanita peptide to a naturally produced peptide. In some embodiments, an Amanita peptide has an

artificial sequence, in other words a nucleic acid encoding an artificial peptide sequence was not naturally found in a fungus or found encoded by a nucleic acid sequence isolated from a fungus.

As used herein, "biologically active" refers to a peptide that when contacted with a cell, tissue or organ induces a biological activity, such as stimulating a cell to divide, causing a cell to alter its function (e.g., altering T cell function, causing a cell to change expression of genes, etc.)

The terms "peptide," "polypeptide," "propeptide," "propolypeptide," "prepropeptide," "prepropolypeptide," "precursor peptide", and "protein" in general refer to a primary sequence of amino acids that are joined by covalent "peptide linkages." Polypeptides may encompass either peptides or proteins. In general, a peptide consists of a few amino acids, and is shorter than a protein. "Amino acid sequence" and like terms, such as "polypeptide" or "protein" are not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule.

As used herein, the term "synthetic" or "artificial" in relation to a peptide sequence refers to a peptide made either artificially from covalently bonding amino acids, such as by Peptide Synthesizer, (for example, Applied Biosystems) or a peptide derived from an amino acid sequence encoded by a recombinant nucleic acid sequence.

As used herein, the term "toxin" refers to a substance (for example, a cyclopeptide) that is detrimental (i.e., poisonous) to cells and/or organisms, in particular a human organism. In some cases, the peptides made by the enzymes and methods described herein are not toxins.

As used herein, "amatoxin" generally refers to a family of peptide compounds, related to and including the amanitins. An amatoxin refers to any small peptide, linear and cyclic, comprising an exemplary chemical structure (e.g., as illustrated herein, see also FIG. 1) or encoded by a nucleic acid sequence. Such an amatoxin nucleic acid sequence and/or proprotein can have higher sequence homology to amanitin (or e.g., an *AMA1* nucleic acid) than to an analogous sequence of a phallotoxin (or e.g., a *PHA1* nucleic acid). In some cases, the peptides made by the enzymes and methods described herein are not amatoxins or phallotoxins. However, in other cases, the enzymes and methods described herein can efficiently make amatoxins and/or phallotoxins.

As used herein, "phallotoxin" generally refers to a family of peptide compounds, related to and including phalloidin and phalloidin. A phallotoxin refers to any small peptide encoded by nucleic acid sequences where the nucleic acid sequence and/or proprotein has a

higher sequence homology to a phallotoxin (or e.g., a *PHA1* nucleic acid) than to an analogous sequence of an amatoxin (or e.g., an *AMA1* nucleic acid).

As used herein the term "microorganism" refers to microscopic organisms and taxonomically related macroscopic organisms within the categories of algae, bacteria, fungi
5 (including lichens), protozoa, viruses, and subviral agents.

The terms "eukaryotic" and "eukaryote" are used in the broadest sense. It includes, but is not limited to, any organisms containing membrane bound nuclei and membrane bound organelles. Examples of eukaryotes include but are not limited to animals, plants, algae, diatoms, and fungi.

10 The terms "prokaryote" and "prokaryotic" are used in the broadest sense. It includes, but is not limited to, any organisms without a distinct nucleus. Examples of prokaryotes include but are not limited to bacteria, blue-green algae (cyanobacteria), archaeobacteria, actinomycetes and mycoplasma. In some cases, a host cell is any microorganism.

As used herein, the term "fungi" is used in reference to eukaryotic organisms such as mushrooms, rusts, molds and yeasts, including dimorphic fungi. "Fungus" or "fungi" also
15 refers to a group of lower organisms lacking chlorophyll and dependent upon other organisms for source of nutrients.

As used herein, "mushroom" refers to the fruiting body of a fungus.

As used herein, "fruiting body" refers to a reproductive structure of a fungus which
20 produces spores, typically comprising the whole reproductive structure of a mushroom including cap, gills and stem, for example, a prominent fruiting body produced by species of Ascomycota and Basidiomycota, examples of fruiting bodies are "mushrooms," "carpophores," "toadstools," "puffballs", and the like.

As used herein, "fruiting body cell" refers to a cell of a cap or stem which may be
25 isolated or part of the structure.

As used herein, "spore" refers to a microscopic reproductive cell or cells.

As used herein, "mycelium" refers to a mass of fungus hyphae, otherwise known as a vegetative portion of a fungus.

As used herein, "Basidiomycota" in reference to a Phylum or Division refers to a
30 group of fungi whose sexual reproduction involves fruiting bodies comprising basidiospores formed on club-shaped cells known as basidia.

As used herein, "Basidiomycetes" in reference to a class of Phylum Basidiomycota refers to a group of fungi. Basidiomycetes include mushrooms, of which some are rich in cyclopeptides and/or toxins, and includes certain types of yeasts, rust and smut fungi, gilled-

mushrooms, puffballs, polypores, jelly fungi, brackets, coral, mushrooms, boletes, puffballs, stinkhorns, etc.

As used herein, "Ascomycota" or "ascomycetes" in reference to members of a fungal Phylum or Division refers to a "sac fungus" group. Of the Ascomycota, a class

5 "Ascomycetes" includes *Candida albicans*, unicellular yeast, *Morchella esculentum*, the morel, and *Neurospora crassa*. Some ascomycetes cause disease, for example, *Candida albicans* causes thrush and vaginal infections; or produce chemical toxins associated with diseases, for example, *Aspergillus flavus* produces a contaminant of nuts and stored grain called aflatoxin, that acts both as a toxin and a deadly natural carcinogen.

10 As used herein "Amanita" refer to a genus of fungus whose members comprise poisonous mushrooms, e.g., *Amanita (A.) bisporigera*, *A. virosa*, *A. ocreata*, *A. suballiacea*, and *A. tenuifolia* which are collectively referred to as "death angels" or "Destroying Angels" and "*Amanita phalloides*" or "*A. phalloides* var. *alba*" or "*A. phalloides* var. *verna*" or "*A. verna*", referred to as "death cap." The toxins of these mushrooms frequently cause death
15 through liver and kidney failure in humans. Not all species of this genus are deadly, for example, *Amanita muscaria*, the fly agaric, induces gastrointestinal distress and/or hallucinations while others do not induce detectable symptoms.

As used herein, nonribosomal peptide synthetase (NRPS) is an enzyme that catalyzes the biosynthesis of a small (20 or fewer amino acids) peptide or depsipeptide, linear or
20 circular, and is composed of one or more domains (modules) typical of this class of enzyme. Each domain is responsible for aminoacyl adenylation of one component amino acid. NRPSs can also contain auxiliary domains catalyzing, e.g., N-methylation and amino acid epimerization (Walton. et al., in *Advances in Fungal Biotechnology for Industry, Agriculture, and Medicine*, et al., Eds. (Kluwer Academic/Plenum, N.Y., 2004, pp. 127-162; Finking, et
25 al., (2004) *Annu Rev Microbiol* 58:453-488, all of which are herein incorporated by reference). Examples are gramicidin synthetase, HC-toxin synthetase, cyclosporin synthetase, and enniatin synthetase.

As used herein, the terms "cell," "cell line," "cell population," and "cell culture" may be used interchangeably. All of these terms also include their progeny, which are any and all
30 subsequent generations. It is understood that all progeny may not be identical due to deliberate or inadvertent mutations.

As used herein, "host fungus cell" refers to any fungal cell, for example, a yeast cell, a mold cell, and a mushroom cell (such as *Neurospora crassa*, *Aspergillus nidulans*, *Cochliobolus carbonum*, *Coprinus cinereus*, *Ustilago maydis*, and the like).

As used herein, the term "Fungal expression system" refers to a system using fungi to produce (express) enzymes and other proteins and peptides. Examples of filamentous fungi which are currently used or proposed for use in such processes are *Neurospora crassa*, *Acremonium chrysogenum*, *Tolypocladium geodes*, *Mucor circinelloides*, *Trichoderma reesei*, *Aspergillus nidulans*, *Aspergillus niger*, *Coprinus cinereus*, *Aspergillus oryzae*, etc. Further examples include an expression system for basidiomycete genes (for example, Gola, et al., (2003) J Basic Microbiol. 43(2):104-12; herein incorporated by reference) and fungal expression systems using, for example, a monokaryotic laccase-deficient *Pycnoporus cinnabarinus* strain BRFM 44 (Banque de Ressources Fongiques de Marseille, Marseille, France), and *Schizophyllum commune*, (for example, Alexandra, et al., (2004) Appl Environ Microbiol. 70(11):6379-638; Lugones, et al., (1999) Mol. Microbiol. 32:681-700; Schuren, et al., (1994) Curr. Genet. 26:179-183; all of which are herein incorporated by reference).

The term "transgene" as used herein refers to a foreign gene, such as a heterologous gene, that is placed into an organism by, for example, introducing the foreign gene into cells or primordial tissue. The term "foreign gene" refers to any nucleic acid (e.g., gene segment) that is introduced into the genome of a host cell by experimental manipulations and may include gene sequences found in that cell so long as the introduced gene does not reside in the same location as does the naturally-occurring gene.

As used herein, the term "vector" is used in reference to nucleic acid molecules that transfer DNA segment(s) from one cell to another. The term "vehicle" is sometimes used interchangeably with "vector." A vector "backbone" comprises those parts of the vector which mediate its maintenance and enable its intended use (e.g., the vector backbone may contain sequences necessary for replication, genes imparting drug or antibiotic resistance, a multiple cloning site, and possibly operably linked promoter and/or enhancer elements which enable the expression of a cloned nucleic acid). The cloned nucleic acid (e.g., such as a cDNA coding sequence, or an amplified PCR product) is inserted into the vector backbone using common molecular biology techniques.

A "recombinant vector" indicates that the nucleotide sequence or arrangement of its parts is not a native configuration, and has been manipulated by molecular biological techniques. The term implies that the vector is comprised of segments of DNA that have been artificially joined.

As used herein, "recombinant nucleic acid" or "recombinant gene" or "recombinant DNA molecule" or "recombinant nucleic acid sequence" indicates that the nucleotide sequence or arrangement of its parts is not a native configuration, and has been manipulated

by molecular biological techniques. The term implies that the DNA molecule is comprised of segments of DNA that have been artificially joined together. Protocols and reagents to manipulate nucleic acids are common and routine in the art (See e.g., Maniatis et al. (eds.), Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, NY.

5 [1982]; Sambrook et al. (eds.), Molecular Cloning: A Laboratory Manual, Second Edition, Volumes 1-3, Cold Spring Harbor Laboratory Press, NY, [1989]; and Ausubel et al. (eds.), Current Protocols in Molecular Biology, Vol. 1-4, John Wiley & Sons, Inc., New York [1994]; all of which are herein incorporated by reference). Similarly, a "recombinant protein" or "recombinant polypeptide" refers to a protein molecule that is expressed from a
10 recombinant DNA molecule. Use of these terms indicates that the primary amino acid sequence, arrangement of its domains or nucleic acid elements which control its expression are not native, and have been manipulated by molecular biology techniques. As indicated above, techniques to manipulate recombinant proteins are also common and routine in the art.

The terms "exogenous" and "heterologous" are sometimes used interchangeably with
15 "recombinant." An "exogenous nucleic acid," "exogenous gene" and "exogenous protein" indicate a nucleic acid, gene or protein, respectively, that has come from a source other than its native source, and has been artificially supplied to the biological system. In contrast, the terms "endogenous protein," "native protein," "endogenous gene," and "native gene" refer to a protein or gene that is native to the biological system, species or chromosome under study.
20 A "native" or "endogenous" polypeptide does not contain amino acid residues encoded by recombinant vector sequences; that is, the native protein contains only those amino acids found in the polypeptide or protein as it occurs in nature. A "native" polypeptide may be produced by recombinant means or may be isolated from a naturally occurring source. Similarly, a "native" or "endogenous" gene is a gene that does not contain nucleic acid
25 elements encoded by sources other than the chromosome on which it is normally found in nature.

As used herein, the term "heterologous nucleic acid" or "heterologous polypeptide" refers to a nucleic acid or polypeptide that is not in its natural environment. For example, a heterologous coding region includes a coding region from one species introduced into another
30 species. A heterologous nucleic acid or heterologous polypeptide also includes a nucleic acid or polypeptide native to an organism that has been altered in some way (e.g., mutated, added in multiple copies, linked to non-native regulatory sequences, etc.). Heterologous nucleic acids or heterologous polypeptides are distinguished from endogenous nucleic acids in that the heterologous nucleic acids are typically joined to nucleic acid segments that are not found

naturally associated with the nucleic acids in the chromosome or are associated with portions of the chromosome not found in nature (e.g., coding regions expressed in loci where the corresponding gene is not normally expressed).

In addition to containing introns, genomic forms of a gene may also include
5 sequences located on both the 5' and 3' end of the sequences that are present on the RNA transcript. These sequences are referred to as "flanking" sequences or regions (these flanking sequences are located 5' or 3' to the untranslated sequences present on the mRNA transcript). The 5' flanking region may contain regulatory sequences such as promoters and enhancers that control or influence the transcription of the gene. The 3' flanking region may contain
10 sequences that direct the termination of transcription, post-transcriptional cleavage and polyadenylation.

As used herein, the term "regulatory element" refers to a genetic element that controls some aspect of the expression of nucleic acid sequences. For example, a promoter is a regulatory element that facilitates the initiation of transcription of an operably linked coding
15 region. Other regulatory elements include splicing signals, polyadenylation signals, termination signals, etc.

The terms "in operable combination," "in operable order," "operably linked" and similar phrases when used in reference to nucleic acid segments herein are used to refer to the linkage of nucleic acid segments in such a manner that a nucleic acid molecule capable of
20 directing the transcription of a given coding region and/or the synthesis of a desired protein molecule is produced. The term also refers to the linkage of amino acid sequences in such a manner so that a functional protein is produced.

A "promoter" is a control sequence that is a region of a nucleic acid sequence at which initiation and rate of transcription are controlled. It may contain genetic elements at
25 which regulatory proteins and molecules may bind such as RNA polymerase and other transcription factors. The phrases "operably linked," "operatively positioned," "operatively linked," "under control," and "under transcriptional control" mean that a promoter is in a correct functional location and/or orientation in relation to a nucleic acid segment (e.g., a nucleic acid segment encoding an enzyme) to control transcriptional initiation and/or
30 expression of that nucleic acid segment. A promoter may or may not be used in conjunction with an "enhancer," which refers to a cis-acting regulatory sequence involved in the transcriptional activation of a nucleic acid segment.

A promoter may be one naturally associated with a gene or nucleic acid segment, and may be obtained by isolating the 5' non-coding sequences located upstream of the coding

segment and/or exon. Such a promoter can be referred to as "endogenous." Similarly, an enhancer may be one naturally associated with a nucleic acid sequence, located either downstream or upstream of that sequence. Alternatively, certain advantages will be gained by positioning the coding nucleic acid segment under the control of a recombinant or

5 heterologous promoter, which refers to a promoter that is not normally associated with a nucleic acid sequence in its natural environment. A recombinant or heterologous enhancer refers also to an enhancer not normally associated with a nucleic acid in its natural environment. Such promoters or enhancers may include promoters or enhancers of other genes, and promoters or enhancers isolated from any other prokaryotic, viral, or eukaryotic

10 cell, and promoters or enhancers not "naturally occurring," e.g., containing different elements of different transcriptional regulatory regions, and/ or mutations that alter expression. In addition to producing nucleic acid sequences of promoters and enhancers synthetically, sequences may be produced using recombinant cloning and/or nucleic acid amplification technology, including PCR, in connection with the compositions disclosed herein (see U.S.

15 Pat. No. 4,683,202, U.S. Pat. No. 5,928,906, each incorporated herein by reference). It is further contemplated that control sequences that direct transcription and/or expression of sequences within non-nuclear organelles such as mitochondria, chloroplasts, and the like, can be employed as well.

A promoter and/or enhancer can be used that effectively directs the expression of a

20 nucleic acid segment (e.g., including a nucleic acid segment encoding an enzyme) in the cell type, organelle, and organism chosen for expression. Those of skill in the art of microbiology and molecular biology generally know the use of promoters, enhancers, and cell type combinations for protein expression, for example, see Sambrook et al. (1989); herein incorporated by reference. The promoters employed may be constitutive, tissue-specific,

25 inducible, and/or useful under the appropriate conditions to direct the desired level of expression of the introduced nucleic acid segment encoding a target protein (e.g., high levels of expression that are advantageous in the large-scale production of recombinant proteins and/or peptides). The promoter may be heterologous or endogenous.

Suitable control elements such as enhancers/promoters, splice junctions,

30 polyadenylation signals, etc. may be placed in close proximity to the coding region of a coding region to facilitate proper initiation of transcription and/or correct processing of the primary RNA transcript. Alternatively, the coding region utilized in the expression cassettes may contain endogenous enhancers/promoters, splice junctions, intervening sequences,

polyadenylation signals, etc. or a combination of both endogenous and exogenous control elements.

As used herein, the term "hybridization" is used in reference to the pairing of complementary nucleic acids. Hybridization can be demonstrated using a variety of hybridization assays (Southern blot, Northern Blot, slot blot, phage plaque hybridization, and other techniques). These protocols are available in the art (See e.g., Sambrook et al. (eds.), Molecular Cloning: A Laboratory Manual, Second Edition, Volumes 1-3, Cold Spring Harbor Laboratory Press, NY, [1989]; Ausubel et al. (eds.), Current Protocols in Molecular Biology, Vol. 1-4, John Wiley & Sons, Inc., New York [1994]; all of which are herein incorporated by reference).

Hybridization is the process of one nucleic acid pairing with an antiparallel counterpart that may or may not have 100% complementarity. Two nucleic acids that contain 100% antiparallel complementarity will show strong hybridization. Two antiparallel nucleic acids which contain no antiparallel complementarity (generally considered to be less than 30%) will not hybridize. Two nucleic acids which contain between 31-99% complementarity will show an intermediate level of hybridization. A single molecule that contains pairing of complementary nucleic acids within its structure is said to be "self-hybridized."

During hybridization of two nucleic acids under high stringency conditions, complementary base pairing will occur only between nucleic acid fragments that have a high frequency of complementary base sequences. Thus, conditions of "weak" or "low" stringency are often required with nucleic acids that are derived from organisms that are genetically diverse, as the frequency of complementary sequences is usually less. As used herein, two nucleic acids which are able to hybridize under high stringency conditions are considered "substantially homologous." Whether sequences are "substantially homologous" may be verified using hybridization competition assays. For example, a "substantially homologous" nucleotide sequence is one that at least partially inhibits a completely complementary probe sequence from hybridizing to a target nucleic acid under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (i.e., selective) interaction. The absence of non-specific binding may be verified by the use of a second target that lacks even a partial degree of complementarity (e.g., less than about 30% identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target. When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic DNA, the term "substantially homologous" refers to

any probe that can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of high stringency.

Hybridization and the strength of hybridization (i.e., the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementary
5 between the nucleic acids, stringency of the conditions involved, the T_m of the formed hybrid, and the G:C ratio within the nucleic acids.

As used herein, the term "stringency" is used in reference to the conditions of temperature, ionic strength, and the presence of other compounds such as organic solvents, under which nucleic acids hybridize. "Low or weak stringency" conditions are reaction
10 conditions which favor the complementary base pairing and annealing of two nucleic acids. "High stringency" conditions are those conditions which are less optimal for complementary base pairing and annealing. Several variables affect the strength of hybridization, including the length and nature of the probe and target (DNA, RNA, base composition, present in solution or immobilized, the degree of complementary between the nucleic acids, the T_m of
15 the formed hybrid, and the G:C ratio within the nucleic acids). Conditions may be manipulated to define low or high stringency conditions: factors such as the concentration of salts and other components in the hybridization solution (e.g., the presence or absence of formamide, dextran sulfate, and polyethylene glycol) as well as temperature of the hybridization and/or wash steps. Conditions of "low" or "high" stringency are specific for the
20 particular hybridization technique used.

As used herein the term "stringency" is used in reference to the conditions of temperature, ionic strength, and the presence of other compounds such as organic solvents, under which nucleic acid hybridizations are conducted. Those skilled in the art will recognize that "stringency" conditions may be altered by varying the parameters just described either
25 individually or in concert. With "high stringency" conditions, nucleic acid base pairing will occur only between nucleic acid fragments that have a high frequency of complementary base sequences (e.g., hybridization under "high stringency" conditions may occur between homologs with about 85-100% identity, preferably about 70-100% identity). With medium stringency conditions, nucleic acid base pairing will occur between nucleic acids with an
30 intermediate frequency of complementary base sequences (e.g., hybridization under "medium stringency" conditions may occur between homologs with about 50-70% identity). Thus, conditions of "weak" or "low" stringency are often required with nucleic acids that are derived from organisms that are genetically diverse, as the frequency of complementary sequences is usually less.

"High stringency conditions" when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 65° C in a solution consisting of 5xSSPE (43.8 g/l NaCl, 6.9 g/l NaH₂PO₄H₂O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5% sodium dodecyl sulfate (SDS), 5xDenhardt's reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 0.1xSSPE, 1.0% SDS at 42° C. when a probe of about 500 nucleotides in length is employed. However, the washing step is most responsible for the stringency of hybridization. Hence, "high stringency conditions" can be described as washing in 0.1xSSPE, 1.0% SDS at 42° C.

"Medium stringency conditions" when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 55° C. in a solution consisting of 5xSSPE (43.8 g/l NaCl, 6.9 g/l NaH₂PO₄H₂O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5% SDS, 5xDenhardt's reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 1.0xSSPE, 1.0% SDS at 42° C. when a probe of about 500 nucleotides in length is employed. Because the washing step is most responsible for the stringency of hybridization, "medium stringency conditions" can be described as washing in 1.0xSSPE, 1.0% SDS at 42° C.

"Low stringency conditions" comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5xSSPE (43.8 g/l NaCl, 6.9 g/l NaH₂PO₄H₂O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.1% SDS, 5xDenhardt's reagent (50xDenhardt's contains per 500 ml: 5 g Ficoll (Type 400, Pharmacia), 5 g BSA (Fraction V; Sigma)) and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 5xSSPE, 0.1% SDS at 42° C when a probe of about 500 nucleotides in length is employed. Because the washing step is most responsible for the stringency of hybridization, "low stringency conditions" can be described as washing in 5xSSPE, 1.0% SDS at 42° C.

As used herein, the term "T_m" is used in reference to the "melting temperature." The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated ("denatures") into single strands. An estimate of the T_m value may be calculated by the equation: $T_m = 81.5 + 0.41(\% G+C)$, when a nucleic acid is in aqueous solution at 1 M NaCl (see e.g., Anderson and Young, Quantitative Filter Hybridization, in Nucleic Acid Hybridization (1985)). Other references include more sophisticated computations that take structural as well as sequence characteristics into account for the calculation of T_m.

As used herein, the terms "antiparallel complementarity" and "complementarity" are synonymous. Complementarity can include the formation of base pairs between any type of nucleotides, including non-natural bases, modified bases, synthetic bases and the like.

The following definitions are the commonly accepted definitions of the terms

5 "sequence identity," "similarity" and "homology." Percent sequence identity is a measure of strict amino acid or nucleotide conservation. Percent similarity is a measure of amino acid conservation which incorporates both strictly conserved amino acids, as well as "conservative" amino acid substitutions, where one amino acid is substituted for a different amino acid having similar chemical properties (i.e. a "conservative" substitution). The term

10 "homology" can pertain to either proteins or nucleic acids. Two proteins can be described as "homologous" or "non-homologous," but the degree of amino acid conservation is quantitated by percent identity and percent similarity. Nucleic acid conservation is measured by the strict conservation of the bases adenine, thymine, guanine and cytosine in the primary nucleotide sequence. When describing nucleic acid conservation, conservation of the nucleic acid

15 primary sequence is sometimes expressed as percent sequence identity or percent homology. In the same nucleic acid, one region may show a high percentage of nucleotide sequence conservation, while a different region can show no or poor conservation. Nucleotide sequence conservation may not necessarily be inferred from an amino acid similarity score. Two proteins may show domains that in one region are homologous, while other regions of the

20 same protein are clearly non-homologous.

The following non-limiting Examples illustrate some aspects of the enzymes, enzymatic reactions, and methods provided herein.

25 **Example 1: cDNA for *Galerina marginata* FMO1**

This Example illustrates isolation of a cDNA for *Galerina marginata* (CBS 339.88) FMO1.

Four species of *Galerina* were obtained from Centraalbureau voor Schimmelcultures (CBS), Utrecht, Netherlands, including *Galerina marginata* (CBS 339.88), *Galerina badipes*

30 (CBS 268.50), *Galerina venenata* (CBS 924.72), and *Galerina hybrida* (CBS 335.88). *Galerina marginata* CBS 339.88 is monokaryotic and was confirmed to make alpha-amanitin. *G. venenata* is considered synonymous with *G. marginata* (Gulden et al., 2001, 2005). The cultures were maintained on potato dextrose agar. For DNA isolation, the isolates were cultured in liquid medium for 15–30 d with rotary shaking at 120 rpm at 23 °C. The

medium was HSV-2C, which contains (per liter) 1 g yeast extract, 2 g glucose, 0.1 g NH₄Cl, 0.1 g CaSO₄·5 H₂O, 1 mg thiamine-HCl, and 0.1 mg biotin, pH 5.2 (Muraoka and Shinozawa, J. Biosci. Bioeng. 89, 73–76 (2000)). For induction experiments, the media had the same formulation, except that the media contained high carbon (HSV-5C) and low carbon (HSV-1C) amounts of glucose (i.e. 5 g glucose and 1 g glucose, respectively) (see, Muraoka and Shinozawa, 2000).

Lyophilized fungal mycelia were ground in liquid nitrogen with a mortar and pestle. High molecular weight DNA was isolated using genomic-tip 100/G (Qiagen, Germantown, MD; catalog #10,243), RNA was extracted with TRIzol (Invitrogen, Carlsbad, CA), and cDNA was synthesized following the manufacturers' protocols.

Transformation of *G. marginata* was performed as described previously by the inventors (Luo et al., 2014). For candidate genes, 1.5 to 1.8 kb upstream and downstream sequences were used for homologous recombination. The fragments were cloned into the pHg vector (Kemppainen and Pardo, 2010) and transformed into *Agrobacterium tumefaciens* strain LBA1100 (Kemppainen et al., 2005), which were then used in the *Galerina* transformation.

Analysis of the genomic location of the *Galerina marginata* FMO1 gene indicated that this gene is tightly clustered in the genome of *Galerina* between *AMA1*, which the inventors have shown encodes the "core" 35-amino acid precursor peptide (Luo et al., 2012), and POPB, which the inventors have shown is a unique, dedicated prolyl oligopeptidase that catalyzes macrocyclization (FIG. 1; Luo et al., 2014). Genes for natural products biosynthetic pathways are typically clustered in fungi. Hence, such clustering of the FMO1 gene with known genes in a pathway is evidence that the new gene has a function in the amanitin biosynthetic pathway.

Example 2: GmFMO1 catalyzes tryptathionine biosynthesis

This Example illustrates that the *Galerina marginata* FMO1 enzyme can form a tryptathionine (sulfide) cross bridge.

One reason to believe that FMO1 catalyzes tryptathionine biosynthesis is based on reactions catalyzed by other FMOs. A bacterial FMO catalyzes Trp hydroxylation at the three (3) position (May & Perrin, Biopolymers 88:714-724 (2007)). A chemical route to tryptathionine biosynthesis can involve hydroxylation of Trp at the 3 position, which activates the 2-position for spontaneous reaction with Cys. One mechanism of FMO1 action is hydroxylation of the Trp at the 3 position (FIG. 2A), so that conjugation to Cys occurs spontaneously (FIG. 2B). In the course of the second step, the hydroxyl group leaves. Thus,

the inventors hypothesized that FMO1 catalyzes “cryptic” (i.e., transient) hydroxylation. To test the theory that FMO1 catalyzes tryptathionine biosynthesis, recombinant FMO1 was incubated with cyclo(IWGIGCNP, SEQ ID NO:1), NADPH, and FAD, and the reaction products were analyzed by LC/MS.

- 5 FMO1 was successfully expressed in a wheat germ cell free system (ENDEXT Technology, see website at fsciences.com/eg/tech.html) and it exhibited detectable activity consistent with tryptathionine biosynthesis (FIG 3A). In particular, a new product was found with a mass two daltons lower than the substrate, cyclo(IWGIGCNP) (FIG. 3B). This is the expected mass for cyclo(IWGIGCNP) (SEQ ID NO: 1) plus the Trp-Cys cross bridge.
- 10 Expression in *E. coli* and yeast is also an option.

- FIGs. 3B shows extracted ion counts (EIC) at m/z 839 of (top) product, and the (bottom) substrate. As illustrated by FIG. 3B, when FMO1 was incubated with cyclo(IWGIGCNP), NADPH, FAD, and an NADPH-regenerating system, a new product was observed with a mass 2 Da smaller than the substrate, which is consistent with oxidative
- 15 coupling of Cys and Trp. In particular, the m/z of cyclo(IWGIGCNP) is 841 ($M + H^+$), and the predicted mass of cyclo(IWGIGCNP) with a sulfide cross-bridge between Trp and Cys is 839 ($M + H^+$). The substrate, cyclo(IWGIGCNP), m/z 841, elutes at 13.0 min. FIG. 3C shows that the compound with m/z 839 is the dominant component eluting at 12.0 min. The peak at m/z 861 in FIG. 3C corresponds to the sodium adduct. The change in the UV absorption
- 20 spectrum (ratio of 295 to 280 nm) of the new, FMO1-catalyzed product was consistent with the structure being the bicyclic compound. Coupling of Cys to Trp shifts the UV absorption maximum of Trp from 280 nm to 295 nm (Wieland, 1986).

Example 3: Gene knockout in *G. marginata*

- This Example illustrates that targeted gene disruption of GmFMO1 abolishes amanitin
- 25 biosynthesis, demonstrating that FMO1 is needed for amanitin biosynthesis.

- Deletion of GmFMO1 in *G. marginata* results in total loss of α -amanitin. FIG. 4A illustrates LC-MS analysis of amanitin from wild type *G. marginata* (solid line) and mutant *G. marginata* (dashed line). FIG. 4B illustrates Southern analysis blot of *G. marginata* wild type and *G. marginata* deletion mutant genomic DNA. The probes employed were for
- 30 GmFMO1; Hyg stands for hygromycin B resistance (the selection for transformation). The southern blots show that the FMO1 gene has been deleted in the mutant. The gene disruption method employed is described in Luo et al. Chemistry and Biology 21:1610 (2014). The results show that GmFMO1 is needed for biosynthesis of amanitin.

References

- Arnison PG, et al. (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* 30:108-160. doi: 10.1039/c2np20085f. PMID: 23165928.
- 5 Bockus et al. (2013) Form and function in cyclic peptide natural products: a pharmacokinetic perspective. *Curr Top Med Chem* 13:821.
- Bouchard et al. (2014) Antibody-drug conjugates – a new wave of cancer drugs. *Bioorg Med Chem Lett* 24:5357.
- Craik & Allewell, *Biol. Chem.* 287:26999–27000 (2012)
- 10 Craik & Malik, *Curr. Opin. Chem. Biol.* 17:546–554 (2013)
- Donia et al., *Nat. Chem. Biol.* 2:729–735 (2006)
- Hallen et al (2007) Gene family encoding the major toxins of lethal *Amanita* mushrooms. *Proc Natl Acad Sci* 104:19097.
- Liu et al. (2015) TP53 loss creates therapeutic vulnerability in colorectal cancer. 15 *Nature* 520:697.
- Luo et al. Biosynthesis of alpha-amanitin in *Galerina marginata*. *Fung Genet Biol* 49:123 (2012).
- Luo H, Hong SY, Sgambelluri RM, Angelos E, Li X, Walton JD (2014) Peptide macrocyclization catalyzed by a prolyl oligopeptidase involved in α -amanitin biosynthesis. 20 *Chem Biol* 21:1610-1617. PMID: 25484237.
- May M (2014) Drug development: time for teamwork. *Nature* 509:S4-S5.
- May JP, Perrin DM (2007) Tryptathionine bridges in peptide synthesis. *Biopolymers* 88:714-724.
- May JP, Fournier P, Patrick BO, Perrin DM (2008) Synthesis, characterisation, and in vitro evaluation of Pro2-Ile3-S deoxo-amaninamide and Pro2-D-allo-Ile3-S-deoxo- 25 amaninamide. *Chemistry* 14:3410-3417.
- Pulman JA, Childs KL, Sgambelluri RM, Walton JD (2016) Expansion and diversification of the MSDIN family of cyclic peptide genes in the poisonous agarics *Amanita phalloides* and *A. bisporigera*. *BMC Genomics* 17:1038. PMID: 27978833.
- 30 Riley, R., A. et al. (2014). Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white rot/brown rot paradigm for wood decay fungi. *Proc Natl Acad Sci USA* 111:9923-9928. PMID: 24958869.
- Sgambelluri RM, Smith MO, Walton JD (2017) Versatility of prolyl oligopeptidase B in peptide macrocyclization. *ACS Synth Biol*, doi: 10.1021/acssynbio.7b00264.

Smith et al. (2013) Design, synthesis, and diversification of ribosomally derived peptide macrocycles. *Curr Opin Struct Biol* 23:571-580. Umemura et al., *Fung. Genet. Biol.* 68:23-30 (2014).

Walton, J.D. (1996) Host-selective toxins: agents of compatibility. *Plant Cell* 8: 1723-1733.

Wieland T (1986) Peptides of poisonous *Amanita* mushrooms. Springer, New York.

All patents and publications referenced or mentioned herein are indicative of the levels of skill of those skilled in the art to which the invention pertains, and each such referenced patent or publication is hereby specifically incorporated by reference to the same extent as if it had been incorporated by reference in its entirety individually or set forth herein in its entirety. Applicants reserve the right to physically incorporate into this specification any and all materials and information from any such cited patents or publications.

The following statements of the invention are intended to describe and summarize various embodiments of the invention according to the foregoing description in the specification.

Statements:

1. An expression system comprising at least one expression cassette comprising a promoter operably linked to a heterologous nucleic acid segment encoding a flavin mono-oxygenase (FMO).
2. The expression system of statement 1, wherein the flavin mono-oxygenase has at least 90%, or at least 91%, or at least 93%, or at least 95%, or at least 96%, or at least 97%, or at least 98%, or at least 99% sequence identity to SEQ ID NO: 2, 4, 5, 6, or 7.
3. The expression system of statement 1 or 2, wherein the flavin mono-oxygenase can catalyze a tryptathionine cross-bridge between a tryptophan (Trp) and a cysteine (Cys) in a peptide substrate.
4. The expression system of statement 1, 2, or 3, wherein the flavin mono-oxygenase can catalyze a tryptathionine cross-bridge between a tryptophan (Trp) and a cysteine (Cys) in a peptide substrate in the presence of nicotinamide adenine dinucleotide phosphate (NADPH), flavin adenine dinucleotide (FAD), or a combination thereof.
5. The expression system of statement 1-3 or 4, wherein the peptide substrate has about 6 to about 100 amino acids, or about 7 to about 75 amino acids, or about 8 to about 50 amino acids, or about 8 to about 35 amino acids, or about 10 to about 30 amino acids.
6. The expression system of statement 1-4 or 5, wherein the peptide substrate has conventional and/or nonconventional amino acids.

7. The expression system of statement 1-5 or 6, wherein the peptide substrate has one or more hydroxylated amino acids.
8. The expression system of statement 1-6 or 7, wherein the peptide substrate contains one or more D-amino acid, N-methyl amino acid, beta amino acids, other atypical amino acid (e.g., ornithine, amino-isobutyric acid), or combinations thereof.
9. The expression system of statement 1-7 or 8, wherein the peptide substrate has the following sequence: XWXXXCXP (SEQ ID NO:16), where X is any amino acid.
10. The expression system of statement 1-8 or 9, wherein one or more Ile, Trp, or Pro amino acids are hydroxylated in the peptide substrate.
11. The expression system of statement 1-9 or 10, further comprising at least one expression cassette comprising a promoter operably linked to a heterologous nucleic acid segment encoding a prolyl oligopeptidase that has macrocyclase activity.
12. The expression system of statement 11, wherein the prolyl oligopeptidase has at least 90%, or at least 91%, or at least 93%, or at least 95%, or at least 96%, or at least 97%, or at least 98%, or at least 99% sequence identity to SEQ ID NO:11, 13, 14, or 15.
13. A host cell comprising an expression system comprising at least one expression cassette comprising a promoter operably linked to a heterologous nucleic acid segment encoding a flavin mono-oxygenase (FMO).
14. The host cell of statement 13, wherein the flavin mono-oxygenase has at least 90%, or at least 91%, or at least 93%, or at least 95%, or at least 96%, or at least 97%, or at least 98%, or at least 99% sequence identity to SEQ ID NO: 2, 4, 5, 6, or 7.
15. The host cell of statement 13 or 14, further comprising at least one expression cassette comprising a promoter operably linked to a heterologous nucleic acid segment encoding a prolyl oligopeptidase.
16. The host cell of statement 15, wherein the prolyl oligopeptidase has at least 90%, or at least 91%, or at least 93%, or at least 95%, or at least 96%, or at least 97%, or at least 98%, or at least 99% sequence identity to SEQ ID NO:11, 13, 14, or 15.
17. The host cell of statement 13-15 or 16, further comprising at least one expression cassette comprising a promoter operably linked to a heterologous nucleic acid segment encoding a peptide substrate.
18. The host cell of statement 17, wherein the peptide substrate has about 6 to about 100 amino acids, or about 7 to about 75 amino acids, or about 8 to about 50 amino acids, or about 8 to about 35 amino acids, or about 10 to about 30 amino acids.

19. The host cell of statement 17 or 18, wherein the peptide substrate has the following sequence: XWXXXCXP (SEQ ID NO:16), where X is any amino acid.
20. A method comprising incubating the host cell of statement 13-18 or 19, for a time and under conditions sufficient for generation of a tryptathionine cross-bridge between a tryptophan (Trp) and a cysteine (Cys) in the peptide substrate.
21. The method of statement 20, wherein the conditions sufficient for generation of a tryptathionine cross-bridge comprise culture conditions for maintenance of the host cell.
22. The method of statement 20 or 21, wherein the conditions sufficient for generation of a tryptathionine cross-bridge comprise conditions where the host cell expresses the flavin mono-oxygenase and the peptide substrate.
23. A method comprising contacting a peptide substrate with a flavin mono-oxygenase that has at least 90%, or at least 91%, or at least 93%, or at least 95%, or at least 96%, or at least 97%, or at least 98%, or at least 99% sequence identity to SEQ ID NO: 2, 4, 5, 6, or 7.
24. The method of statement 23, wherein the flavin mono-oxygenase catalyzes a tryptathionine cross-bridge between a tryptophan (Trp) and a cysteine (Cys) in the peptide substrate.
25. The method of statement 23 or 24, wherein the flavin mono-oxygenase catalyzes a tryptathionine cross-bridge between a tryptophan (Trp) and a cysteine (Cys) in a peptide substrate in the presence of nicotinamide adenine dinucleotide phosphate (NADPH), flavin adenine dinucleotide (FAD), or a combination thereof.
26. The method of statement 23, 24, or 25, wherein the peptide substrate has about 6 to about 100 amino acids, or about 7 to about 75 amino acids, or about 8 to about 50 amino acids, or about 8 to about 35 amino acids, or about 10 to about 30 amino acids.
27. The method of statement 23-25 or 26, wherein the peptide substrate has conventional and/or nonconventional amino acids.
28. The method of statement 23-26 or 27, wherein the peptide substrate is one or more D-amino acid, N-methyl amino acid, beta amino acids atypical amino acid (e.g., ornithine, amino-isobutyric acid), or combinations thereof.
29. The method of statement 23-27 or 28, wherein the peptide substrate has the following sequence: XWXXXCXP (SEQ ID NO:16), where X is any amino acid.
30. The method of statement 23-28 or 29, wherein the peptide substrate has one or more hydroxylated amino acids.

31. The method of statement 23-29 or 30, performed *in vitro*.
32. The method of statement 23-30 or 31, performed in a cell-free reaction mixture.
33. The method of statement 32, wherein the reaction mixture further comprises
nicotinamide adenine dinucleotide phosphate (NADPH), flavin adenine dinucleotide
5 (FAD), or a combination thereof.
34. A flavin mono-oxygenase that has at least 90%, or at least 91%, or at least 93%, or at
least 95%, or at least 96%, or at least 97%, or at least 98%, or at least 99% sequence
identity to SEQ ID NO: 2, 4, 5, 6, or 7, and with at least one amino acid substitution.
35. The flavin mono-oxygenase of statement 34 with less than 100% sequence identity to
10 SEQ ID NO: 2, 4, 5, 6, or 7.
36. The flavin mono-oxygenase of statement 34 or 35 with less than 99.5% sequence
identity to SEQ ID NO: 2, 4, 5, 6, or 7.
37. An expression system comprising at least one expression cassette comprising a
promoter operably linked to a heterologous nucleic acid segment encoding the flavin
15 mono-oxygenase (FMO) of statement 34, 35, or 36.
38. A host cell comprising the expression system of statement 37.

The specific methods and compositions described herein are representative of preferred
embodiments and are exemplary and not intended as limitations on the scope of the invention.
20 Other objects, aspects, and embodiments will occur to those skilled in the art upon
consideration of this specification, and are encompassed within the spirit of the invention as
defined by the scope of the claims. It will be readily apparent to one skilled in the art that
varying substitutions and modifications may be made to the invention disclosed herein without
departing from the scope and spirit of the invention. The invention illustratively described
25 herein suitably may be practiced in the absence of any element or elements, or limitation or
limitations, which is not specifically disclosed herein as essential.

The methods and processes illustratively described herein suitably may be practiced in
differing orders of steps, and the methods and processes are not necessarily restricted to the
orders of steps indicated herein or in the claims. As used herein and in the appended claims,
30 the singular forms "a," "an," and "the" include plural reference unless the context clearly
dictates otherwise. Thus, for example, a reference to "a compound," "a nucleic acid" or "a
promoter" includes a plurality of such compounds, nucleic acids or promoters (for example, a
solution of compounds or nucleic acids, or a series of promoters), and so forth.

Under no circumstances may the patent be interpreted to be limited to the specific examples or embodiments or methods specifically disclosed herein. Under no circumstances may the patent be interpreted to be limited by any statement made by any Examiner or any other official or employee of the Patent and Trademark Office unless such statement is
5 specifically and without qualification or reservation expressly adopted in a responsive writing by Applicants.

The terms and expressions that have been employed are used as terms of description and not of limitation, and there is no intent in the use of such terms and expressions to exclude any equivalent of the features shown and described or portions thereof, but it is recognized that
10 various modifications are possible within the scope of the invention as claimed. Thus, it will be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the concepts herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention as defined by the appended claims and
15 statements of the invention.

What is Claimed:

1. An expression system comprising at least one expression cassette comprising a promoter operably linked to a heterologous nucleic acid segment encoding a flavin mono-oxygenase (FMO) with at least 95% sequence identity to SEQ ID NO: 2, 4, 5, 6, or 7.
2. The expression system of claim 1, further comprising at least one expression cassette comprising a promoter operably linked to a heterologous nucleic acid segment encoding a prolyl oligopeptidase with at least 95% sequence identity to SEQ ID NO:11, 13, 14, or 15.
3. The expression system of claim 1, further comprising at least one expression cassette comprising a promoter operably linked to a heterologous nucleic acid segment encoding a peptide substrate.
4. The expression system of claim 1, wherein the peptide substrate has one or more hydroxylated amino acids.
5. A host cell comprising an expression system comprising at least one expression cassette comprising a promoter operably linked to a heterologous nucleic acid segment encoding a flavin mono-oxygenase (FMO) with at least 95% sequence identity to SEQ ID NO: 2, 4, 5, 6, or 7.
6. The host cell of claim 5, where the expression system further comprises at least one expression cassette comprising a promoter operably linked to a heterologous nucleic acid segment encoding a peptide substrate.
7. The host cell of claim 5, where the expression system further comprises at least one expression cassette comprising a promoter operably linked to a heterologous nucleic acid segment encoding a prolyl oligopeptidase with at least 95% sequence identity to SEQ ID NO:11, 13, 14, or 15.
8. A method comprising incubating the host cell of claim 5, for a time and under conditions sufficient for generation of a tryptathionine cross-bridge between a tryptophan (Trp) and a cysteine (Cys) in the peptide substrate.
9. A method comprising contacting a peptide substrate with a flavin mono-oxygenase that has at least 95% sequence identity to SEQ ID NO: 2, 4, 5, 6, or 7.
10. The method of claim 9, wherein the peptide substrate has one or more hydroxylated amino acids.
11. The method of claim 9, wherein the peptide substrate is one or more D-amino acid, N-methyl amino acid, beta amino acids atypical amino acid, or combinations thereof.

12. The method of claim 9, wherein the flavin mono-oxygenase catalyzes a tryptathionine cross-bridge between a tryptophan (Trp) and a cysteine (Cys) in the peptide substrate.

FIG. 1A

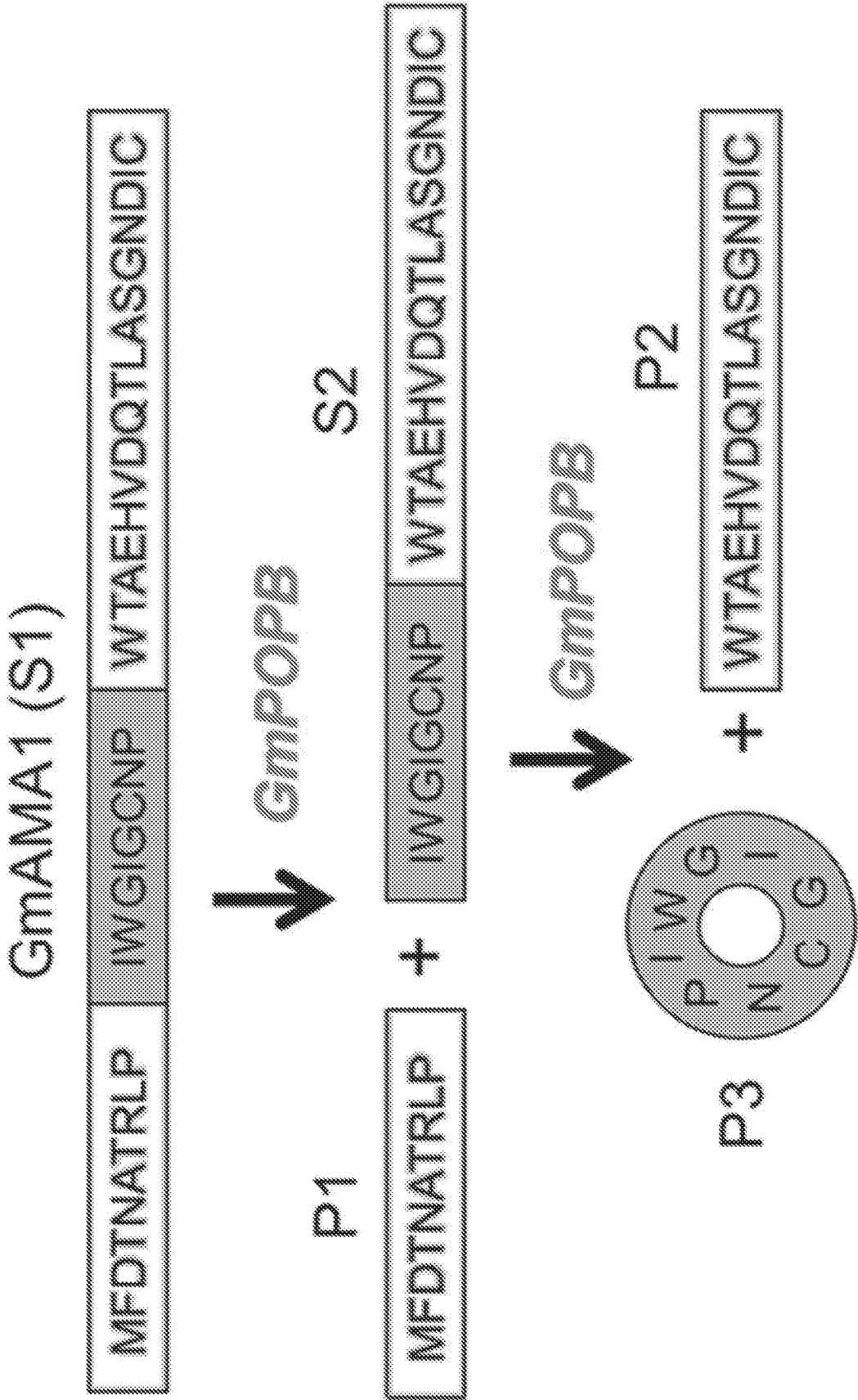
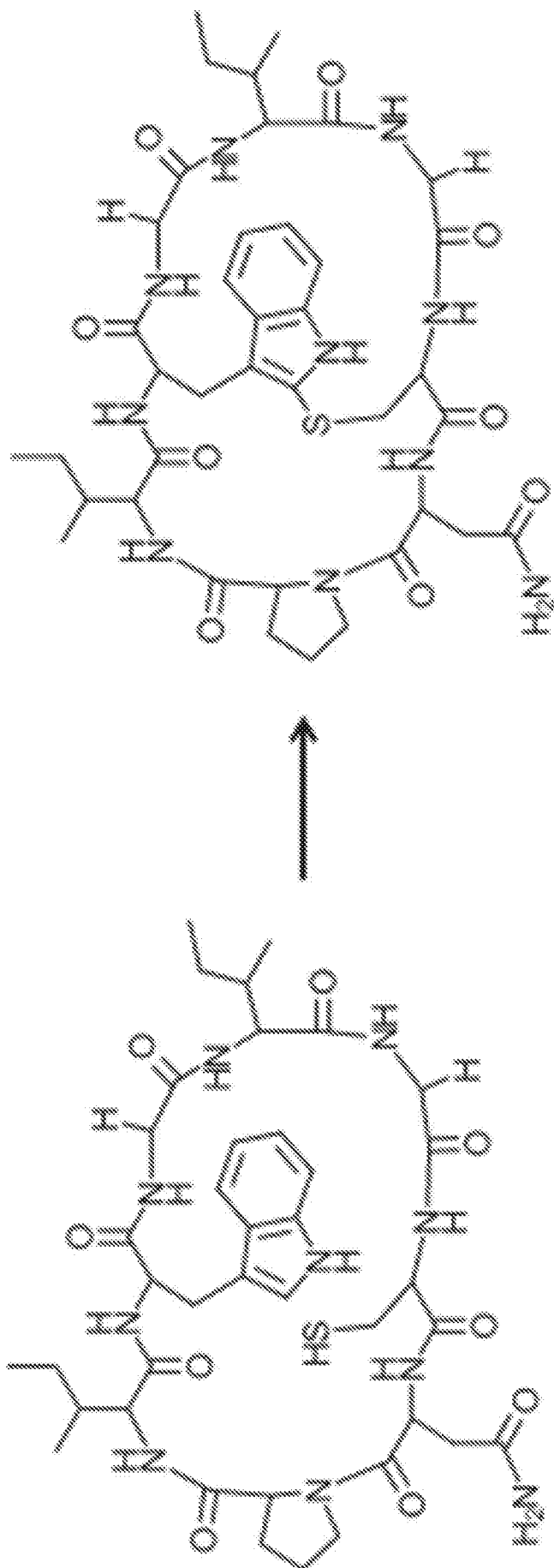


FIG. 1B

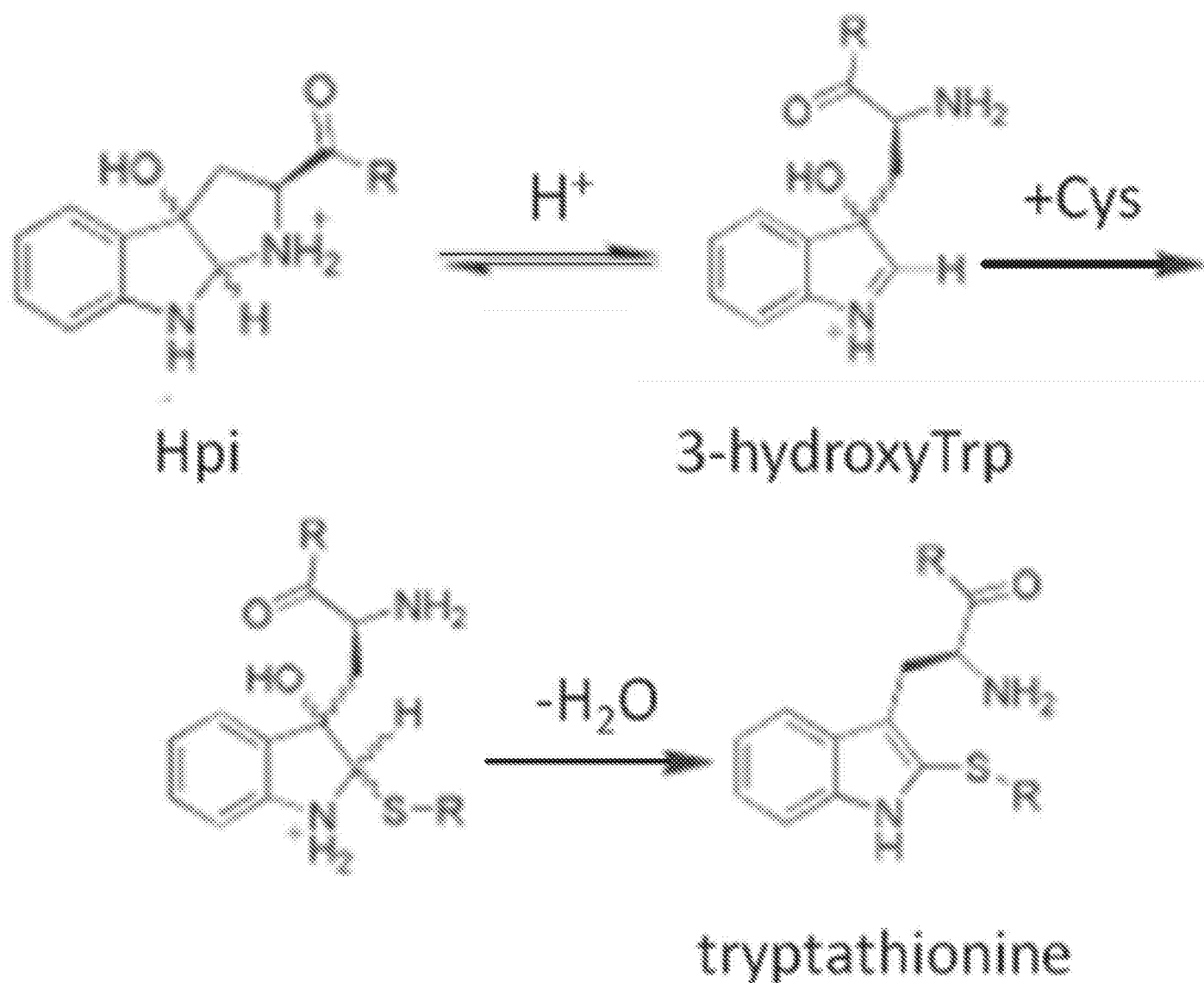


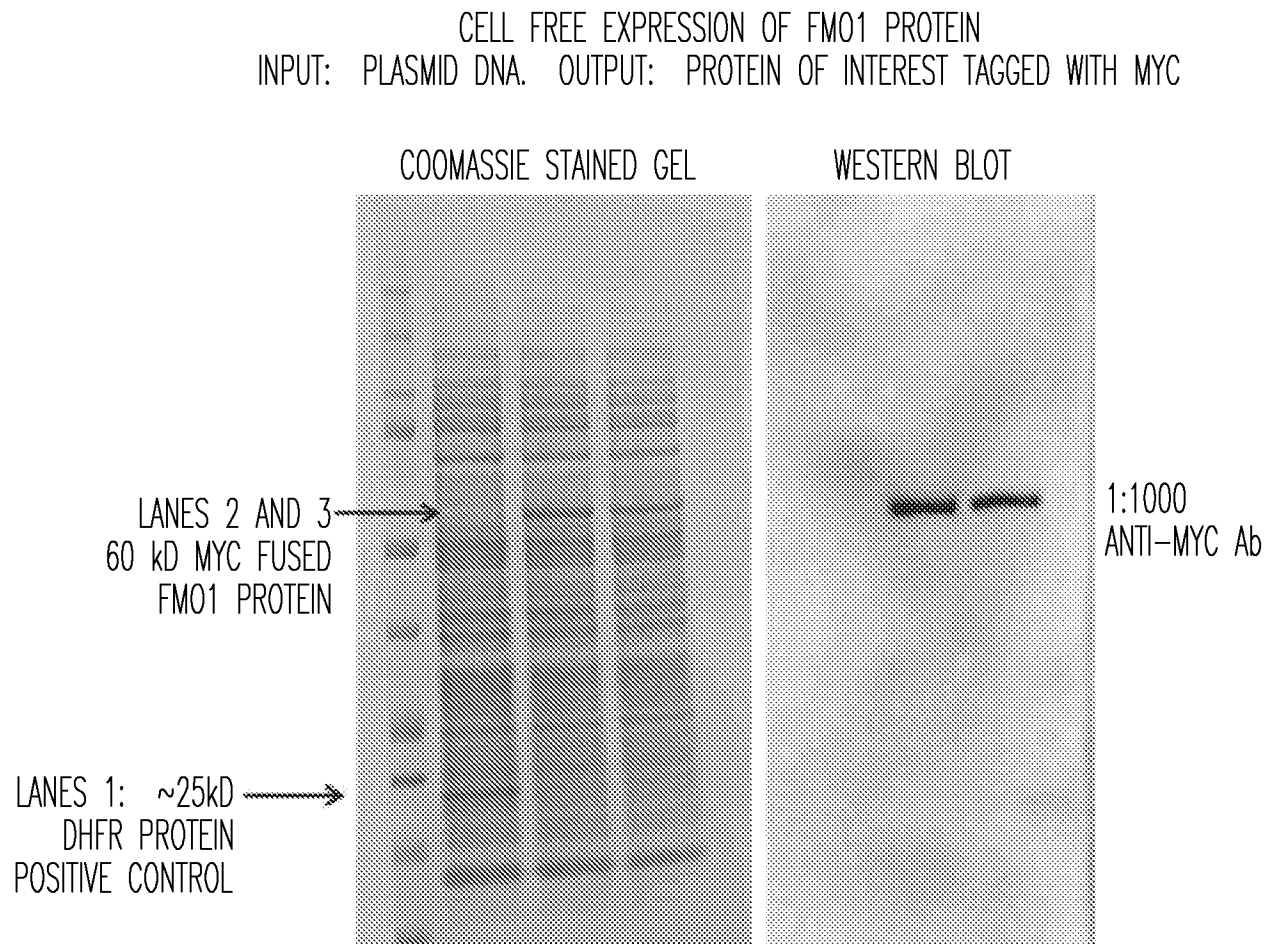
3/8

FIG. 2A



FIG. 2B



**FIG. 3A**

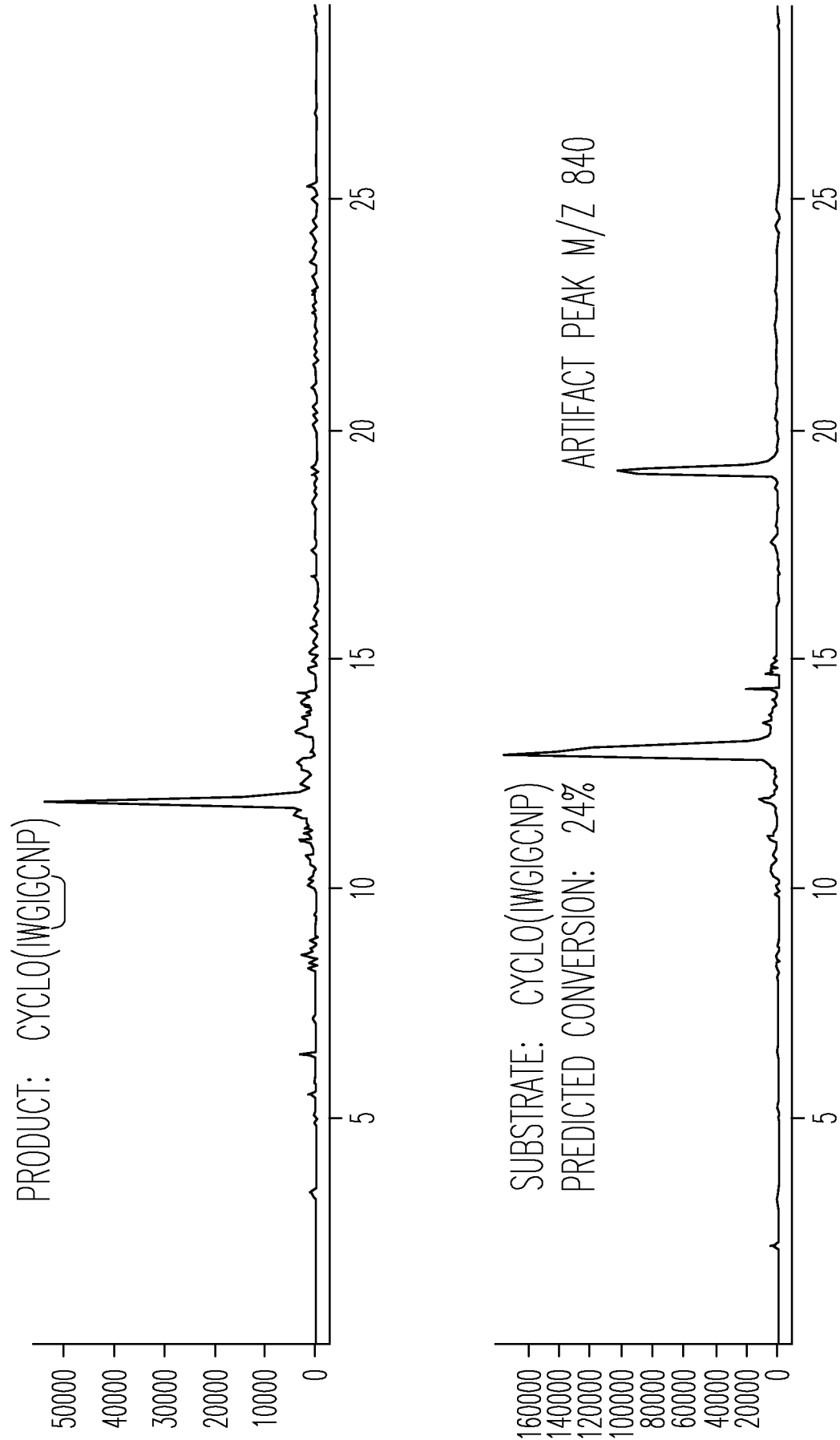


FIG. 3B

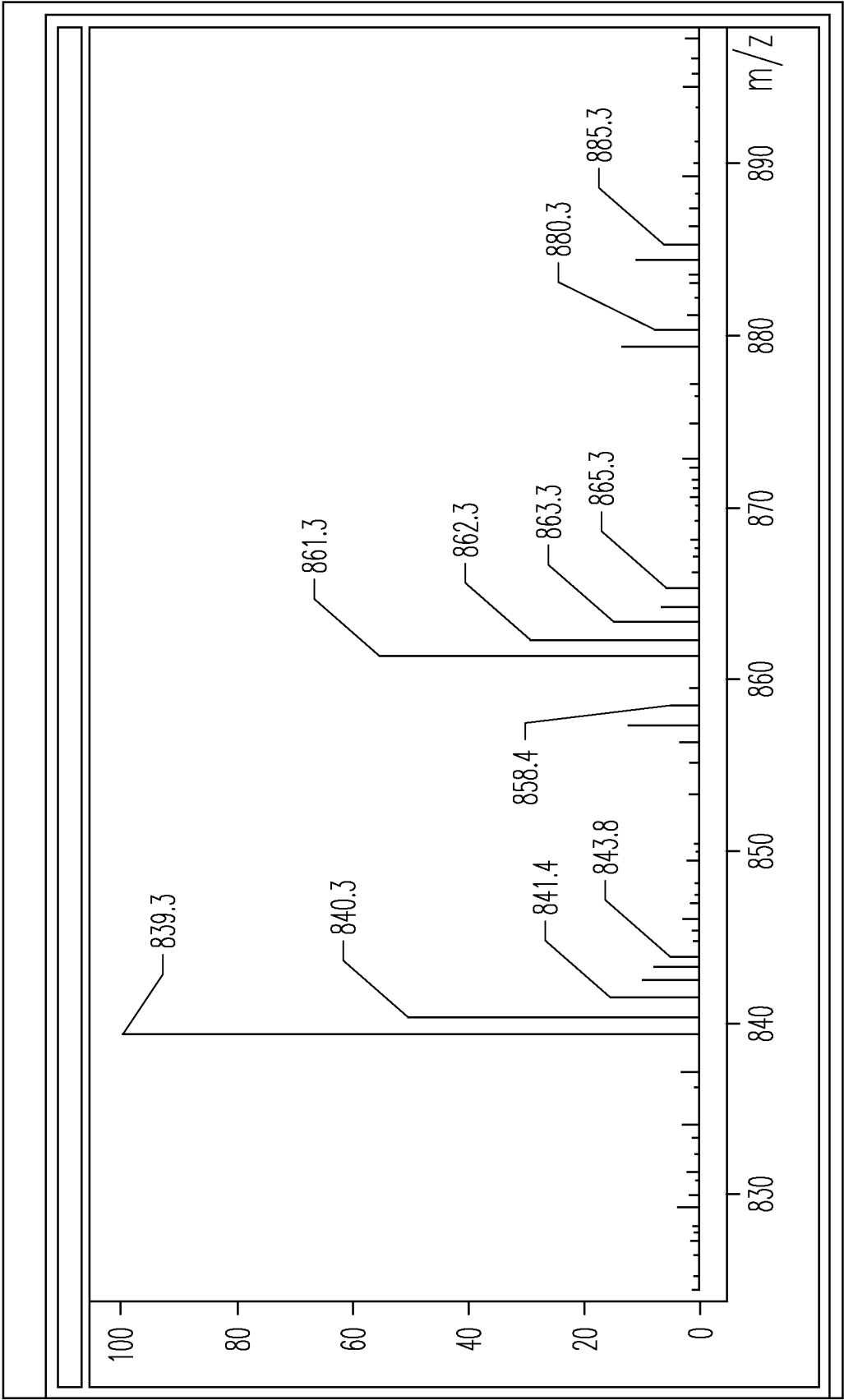


FIG. 3C

FIG. 4A

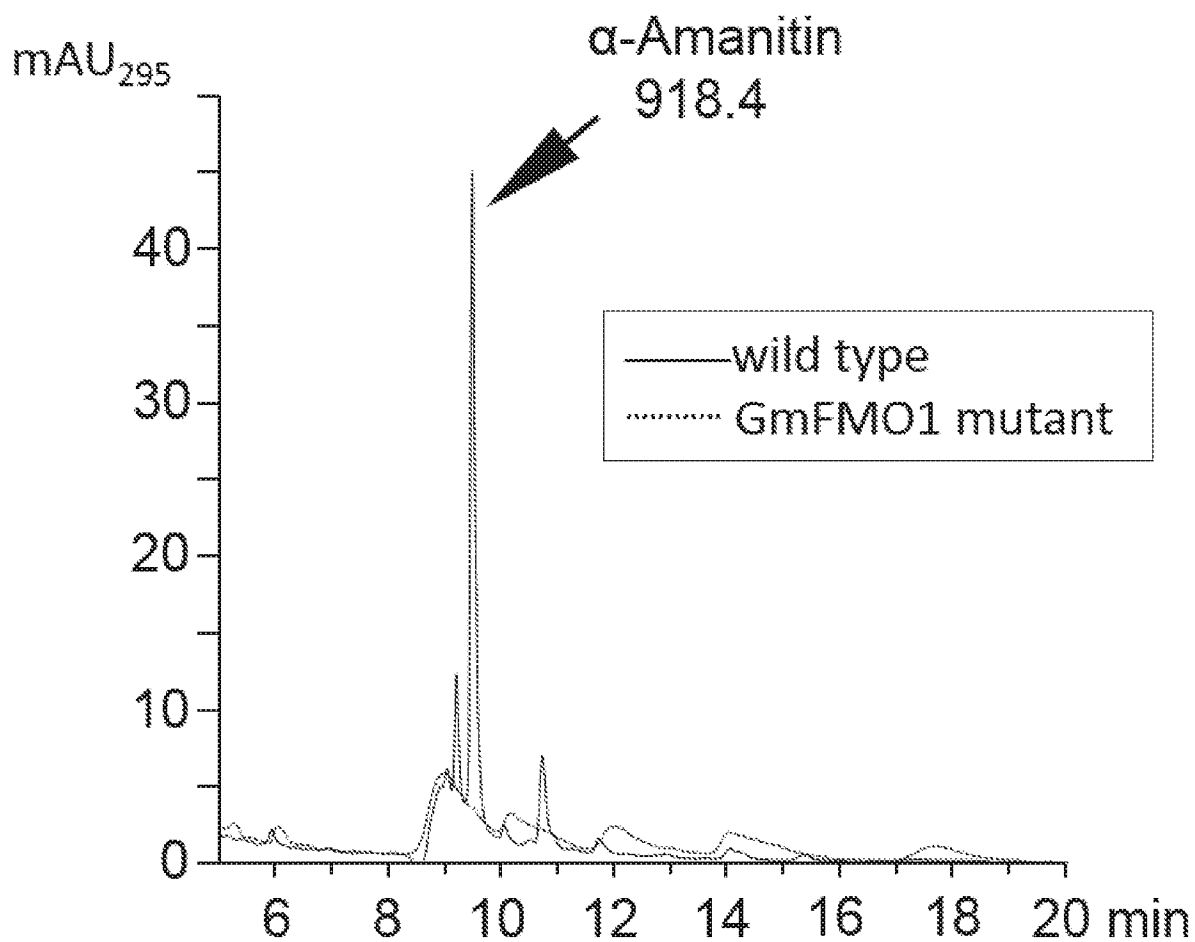
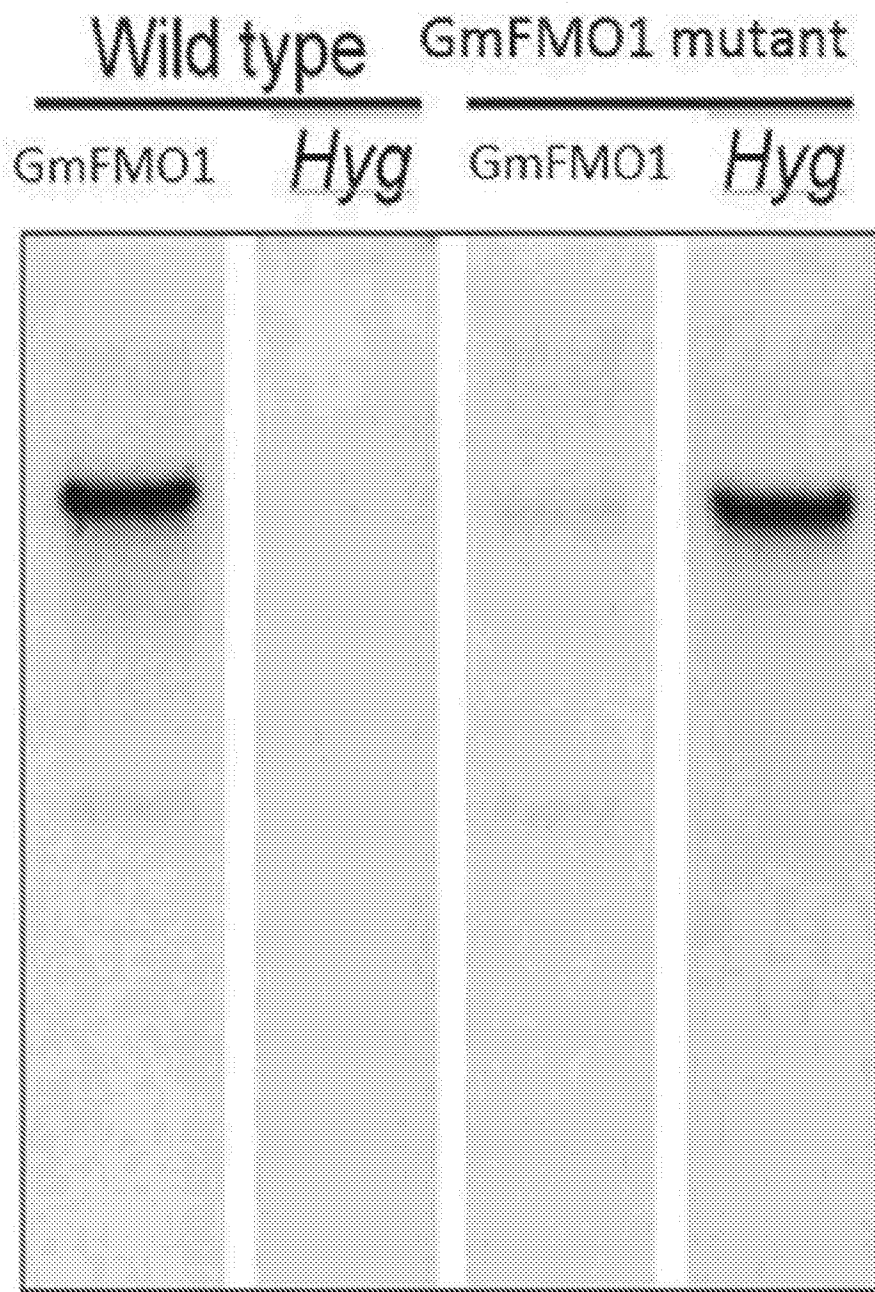


FIG. 4B



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2018/029954

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C07K 1/02; C07K 1/10; C07K 7/50; C07K 7/64; C07K 14/00; C12N 9/02; C12N 15/09 (2018.01)
 CPC - C07K 1/02; C07K 1/10; C07K 1/1072; C07K 1/1075; C07K 7/50; C07K 7/64; C12N 9/0004; C12N 9/0071; C12N 9/0073; C12N 15/09; C12N 15/66; C12Y 114/00; C12Y 114/13; C12Y 114/13008 (2018.05)

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

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

USPC - 424/94.4; 424/195.15; 435/25; 530/317; 530/323; 530/371; 530/402; 530/408; 530/823 (keyword delimited)

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

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2015/0080549 A1 (CHUGAI SEIYAKU KABUSHIKI KAISHA) 19 March 2015 (19.03.2015) entire document	1-12
A	TRUMAN, A. "Cyclisation mechanisms in the biosynthesis of ribosomally synthesised and post-translationally modified peptides," Beilstein J. Org. Chem. 20 June 2016 (20.06.2016), Vol. 12, Pgs. 1250-1268. entire document	1-12
A	WO 2012/068236 A2 (DYADIC INTERNATIONAL (USA) INC. et al) 24 May 2012 (24.05.2012) entire document	1-12
A	US 2013/0084601 A1 (PARK et al) 04 April 2013 (04.04.2013) entire document	1-12
P, A	SGAMBELLURI et al. "Biosynthesis of Cyclic Peptide Natural Products in Mushrooms," Thesis, Michigan State University, 13 December 2017 (13.12.2017), Pgs. 1-162 (only Pgs. 1-100 provided). Retrieved from the Internet: < https://d.lib.msu.edu/etd/6906/datastream/OBJ/download/Biosynthesis_of_Cyclic_Peptide_Natural_Products_in_Mushrooms.pdf > on 12 July 2018 (12.07.2018). entire document	1-12
T	WALTON, J. "Biosynthesis of the Amanita Cyclic Peptide Toxins," The Cyclic Peptide Toxins of Amanita and Other Poisonous Mushrooms, Springer, 10 May 2018 (10.05.2018), Pgs. 110-111. entire document	1-12

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

12 July 2018

Date of mailing of the international search report

06 AUG 2018

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
 P.O. Box 1450, Alexandria, VA 22313-1450
 Facsimile No. 571-273-8300

Authorized officer

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300
 PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2018/029954

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

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

a. ☒ forming part of the international application as filed:

☒ in the form of an Annex C/ST.25 text file.

☐ on paper or in the form of an image file.

b. ☐ furnished together with the international application under PCT Rule 13ter.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

☐ in the form of an Annex C/ST.25 text file (Rule 13ter.1(a)).

☐ on paper or in the form of an image file (Rule 13ter.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

SEQ ID NOs: 2, 4, 5, 6, 7, 11, 13, 14, and 15 were searched.