



US 20170342132A1

(19) **United States**

(12) **Patent Application Publication**

**Fraser et al.**

(10) **Pub. No.: US 2017/0342132 A1**

(43) **Pub. Date: Nov. 30, 2017**

(54) **SECRETION OF HEME-CONTAINING POLYPEPTIDES**

(71) Applicant: **Impossible Foods Inc.**, Redwood City, CA (US)

(72) Inventors: **Rachel Fraser**, San Francisco, CA (US); **Simon Christopher Davis**, San Francisco, CA (US); **Patrick O'Reilly Brown**, Stanford, CA (US)

(21) Appl. No.: **15/672,579**

(22) Filed: **Aug. 9, 2017**

**Related U.S. Application Data**

(63) Continuation of application No. 15/021,447, filed on Mar. 11, 2016, filed as application No. PCT/US2014/055227 on Sep. 11, 2014.

(60) Provisional application No. 61/908,689, filed on Nov. 25, 2013, provisional application No. 61/876,676, filed on Sep. 11, 2013.

**Publication Classification**

(51) **Int. Cl.**  
*C07K 14/805* (2006.01)  
*A23L 13/40* (2006.01)  
*A23J 3/22* (2006.01)  
*C12N 15/82* (2006.01)

(52) **U.S. Cl.**  
 CPC ..... *C07K 14/805* (2013.01); *C12N 15/8257* (2013.01); *A23L 13/424* (2016.08); *A23J 3/227* (2013.01)

(57) **ABSTRACT**

This disclosure provides for methods and compositions for the expression and secretion of heme-containing polypeptides.

Figure 1

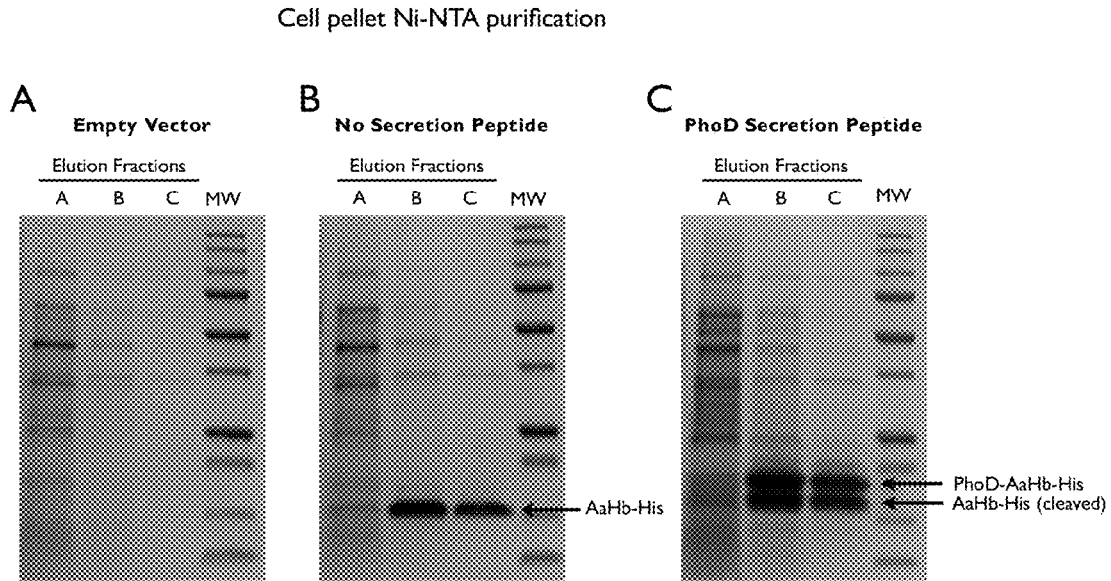


Figure 2

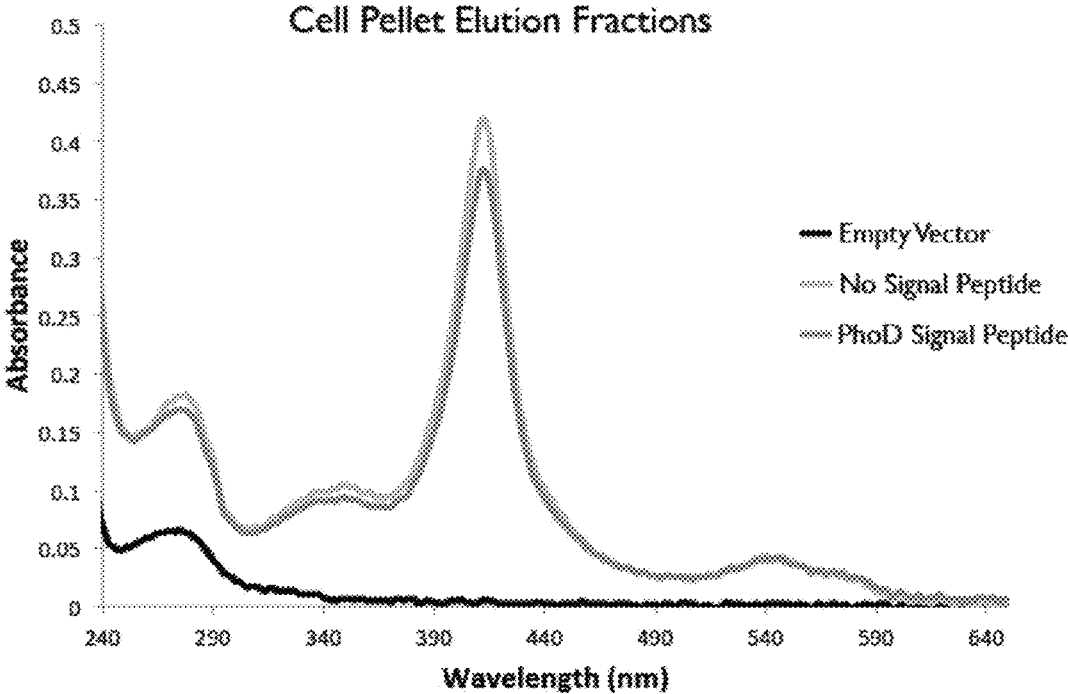


Figure 3

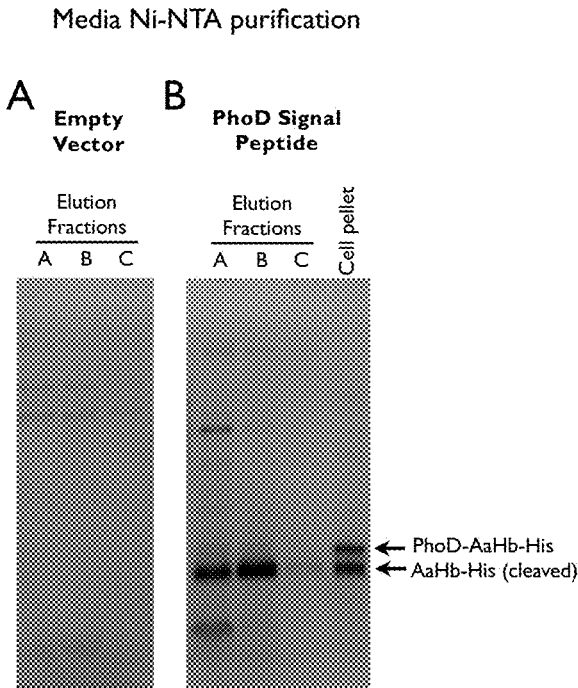


Figure 4

Full length **PhoD**-*Synthetic protease cleavage site*-AaHb-His6 sequence:

**MAYDSRFDEWVQKLKEESFQNNTFDRRKFIQGAGKIAGLSLGLTIAQSASAAGAHMELGLS**  
**EETIRVIKSTVPLLKEHGTEITARMYELLFSKYPKTKELFAGASEEQPKKLANAIIAYATYIDRLEE**  
**LDNAISTIARSHVRRNVKPEHYPLVKECLLQAIEEVLNPGEEVLKAWEEAYDFLAKTLITLEKKL**  
**YSQPRHHHHHH**

Predicted signal peptidase I (SPI) recognition site: ASA/AGAHMELGLS...

N-terminal sequencing results:

(N-term) GAHMELGLSE (C-term)

Figure 5

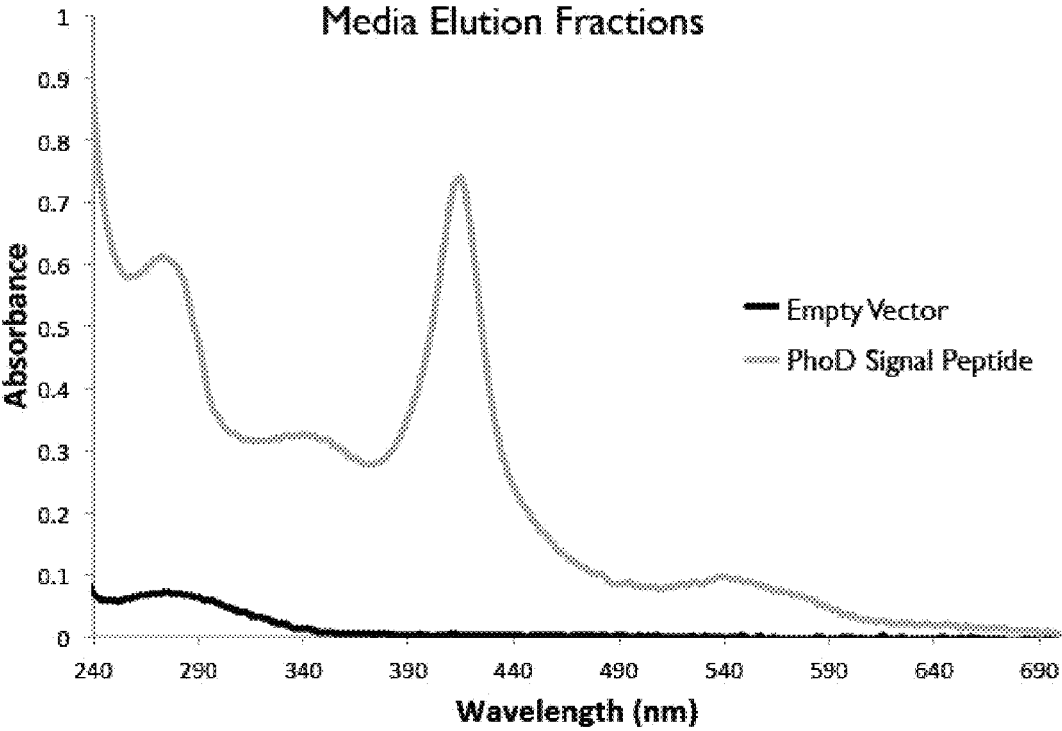


Figure 6

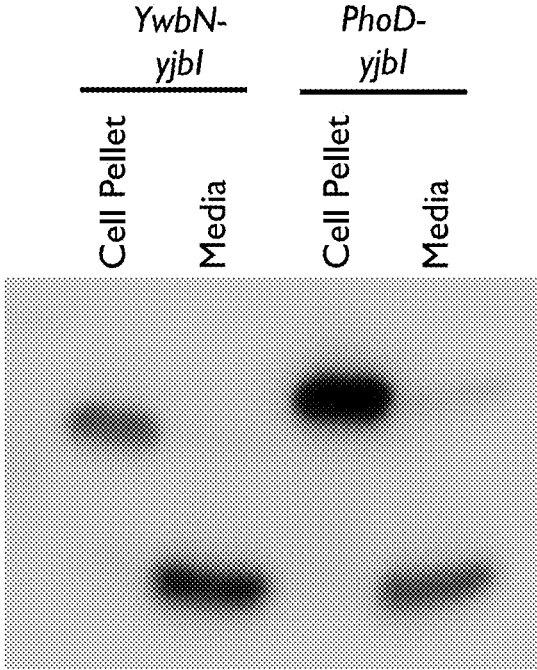


Figure 7

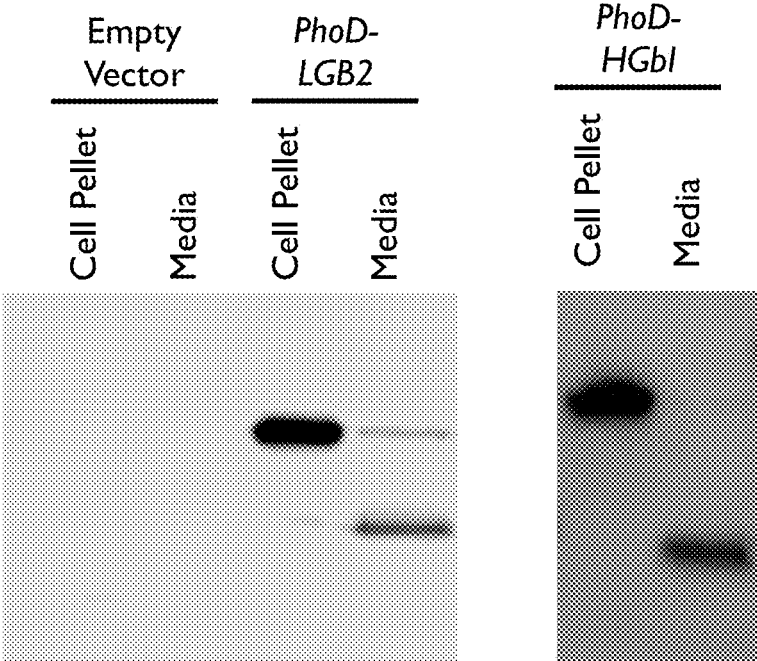




Figure 8

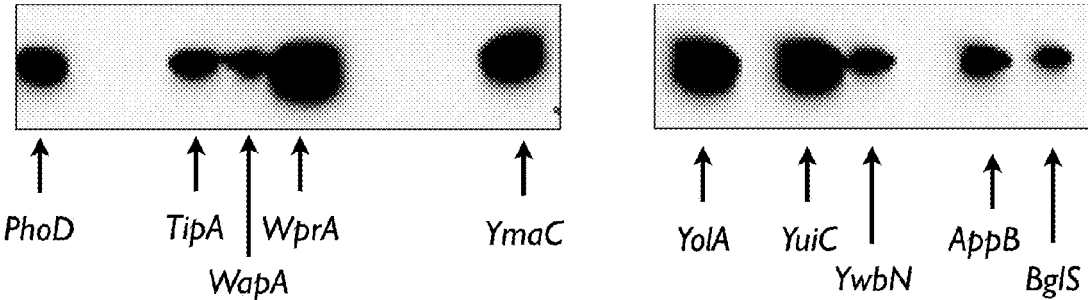


FIGURE 9

SEQ ID NO:1 *Vigna radiata*

MTTTLERGFTEEQEALVVKSWNVMMKNSGELGLKFFLKI FEIAPSAQKLF'SFLRDSTVPLEQNPKLKP  
AVSVFVMTCD SAVQLRKAGKVTVRESNLKKGATHFRTGVANEHFVTKFALLETIKEAVPEMWS PAMK  
NAWGEAYDQLVDAIKYEMKPPSS

SEQ ID NO:2 *Methylococcus thermophilus*

MIDQKEKELIKESWKRIEPNKNE IGLLFYANLFKEEPTVSVLFFQNP I SSQSRLKMQVLGILVQGI DNLE  
GLIPTLQDLGRRHKQYGVVDSHYPLVGDCLLKS IQEYLGQGFTEEAKAAWTKVYGIAAQVMTAE

SEQ ID NO:3 *Aquifex aeolicus*

MLSEETIRVIKSTVPLLKEHGTEITARMYELLSKYPKTKELFAGASEEQPKLANAI IAYATYIDRLE  
ELDNAISTARSHVRRNVKPEHYPLVKECLLQATEEVLNPGEEVLKAWEEAYDFLAKTLITLLEKKLYSQ  
P

SEQ ID NO:4 *Glycine max*

MGAFTKQEALVSSSFEAFKANI PQYSVVFYTSILEKAPAAKDLFSFLSNGVDPSNPKLTGHAEKLFGL  
VRDSAGQLKANGTVVADAALGSIHAQKAITDPQFVVVKEALLKTIKEAVGDKWSEDELSSAWEVAYDELA  
AAIKKAF

SEQ ID NO:5 *Hordeum vulgare*

MSAAEGAVVFSEEKEALVLKSWAIMKKDSANLGLRFFLKI FEIAPSARQMF'PFLRDS'DVPLETNP'PKLKT'  
HAVSVFVMTCEAAAQLRKAGKITVRETTLKRLLGTHLKYGVADGHFEVTRFALLETIKEALPADMWGPE  
MRNAWGEAYDQLVAAIKQEMKPAE

*Magnaporthe oryzae*, (SEQ ID NO:6)

1 mdgavrlidwt gldltgheih dgvpiasrvq vmvsfplfkd qhiimsskes psrksstigt  
61 strngscqad tqkgglppvg ekpkpvkenp mkkkkemsqr piptqhgdt yptekklgti  
121 qedlkhirgy dvktilamvk sklkgekikd dktmlnervm qlvarlptes kkraeitdsl  
181 inelwesidh pplnylgpeh syrtpdgsyn hpfnpqlgaa gsryarsvip tvtpgalpd  
241 pglifdsimg rtpnsyrkhp nrvssilwyw atiihdifw tdprdtntk sssyidlapi  
301 ygnsqemqds irtfkdgimk pdcyadkrla qmpgqvsvll imfnrlnhv aenlaineq  
361 grfnkpsdll egeareaaawk kynddlfava rlvteglyin itlvdyvrm vlnrvdttw  
421 tldprqdag hvgtdagaer gtgnavsaeef nlcyrwhsci sekdkfvea qfqnifgkpa  
481 sevrpdemwk gfakmeqntp adpgqrtfgg fkrqpdgkfd dddivrcise avedvagafg  
541 arnvpqamkv vetmgiqgr kwnvagnlef rkhfhkypys tfedinsdpq vaealrrlyd  
601 hpdnveiyyp lvaeedkqpm vpgvgipty tisrvvlsda vclvrgdrfy tcdftprnlt  
661 nwygkvdvd lsvnhgcvfy klfirafpnh fkqnsvyahy pmvvpnenkr ilealgradl  
721 idfeapkyip prvnitsygg aeyiletqek ykvtwhegly flmgegqkif mlsqddplha  
781 qqrkcmqaqi ykdgwteavk afyagmmeel lvksyflgn nkhrhvdiir dvgnmvvhvf  
841 asqvfglplk taknptgvft eqemygilaa ifttiffdld psksfplrtk trevcqklak

FIG. 9 CONT.

901 lveanvclin kipwsrgmfv gkpakdeplis iygktnikgl kahgisdydi awshvvptsg  
961 amvpngaqvf aqavdyylsp agmhyipeih mvalqpstpe tdaillyam egiriagtfg  
1021 syreaavddv vkedngrqvp vkagdrvfvs fvdaardpkh fpqpevvnpv rpakkyihyg  
1081 vgpshaclgrd asqiaitemf rclfrrrnvr rvpqpgqelk kvprpggfiv ymredwgqif  
1141 pipvtmrvmw dde

*Fusarium oxysporum* (SEQ ID NO:7)

1 mksatlafa lvqfsaasql vpskwdeve dllymqqgfn krgfadalrt cefqsnvpgt  
61 qntaewlrta fhdaithdak agtggldasi ywessrpenp gkafnntfgf fsgfhnprat  
121 asdltalgtv lavgacngpr ipfragrida ykagpagvpe pstnlkdtfa aftkagftke  
181 emtamvacgh aiggvhsvdf peivgikadp nodtnvpfqq dvssfhngiv teylagtstkn  
241 plvasknatf hsdkrifdnd katmkklistk agfnsncadi ltrnidtvpk svqltpvlea  
301 ydvrypyitel sinnknkinf tgsrvrvritn nirdndiai nliyvgvrdgk kvrvptqqvt  
361 fgggtsfgag evfanfefdt tmdakngitk ffiqevkpst katvthdnqk tggvkvddtv  
421 lyqiqqscav lekpnaplv vtamvrdera kdaltirvah kkpvkgsivp rfqtaitnfk  
481 atgkksqgyt gfqaktmfee qstyfdivlg gpasqvqfl teqampsqcs

*Fusarium graminearum* (SEQ ID NO:8)

1 masatrqfar aatrtrngf aiaprvvirq qrrryyssep aqkssawiw ltgaavagga  
61 gyyfyqnsas satakvfnpv kedyqkvyne iaarleekdd yddgsygpvl vrlawhasqt  
121 ydketgtggs ngatmrfafe sdhganaglia aardflqpvk ekfpwitysd lwiaqvcai  
181 qemlgpaipy rprsrdrdvs gotpdgrlpd askrgdhlry ifgrmgfndq eivalaqaha  
241 lgrchtdrsq ysgpwtfspt vltndyfrll veekwqwkvw nqpaqyedks tkslmlpsd  
301 ialiedkkfk pwvekyakdn daffkdfsnv vlrlfelgvp faqqtenqrw tfxpthqe

SEQ ID NO: 9 *Chlamydomonas eugametos*

MSLFAKLGGREAVEAAVDKFFYNKIVADPTVSTYFNSNTDMKVQRSKQFAFLAYALGGASEWKGKDMRTAH  
KDLVPHLSDVHVFQAVARHLSDTLTELGVPPEDITDAMAVVASTRTEVLNMPQQ

SEQ ID NO:10 *Tetrahymena pyriformis*

MNKPQTIYEKLGGENAMKAAVPLFYKKVLADERVKHFFKNTDMDHQTKQQTDFLTMLLGGPNHYKGNM  
TEAHKGMNLQNLHFDAI IENLAATLKELGVTDAVINEAARVIEHTRKDMLGK

FIG. 9 CONT.

SEQ ID NO:11 *Paramecium caudatum*

MSLFEQLGGQAAVQAVTAQFYANIQADATVATFFNGIDMPNQTNKTA AFLCAALGGPNAWTGRNLKEVH  
ANMGVSNAQFTTVIGHLRSALTGAGVAAAALVEQTVAVAETVRGDVVTV

SEQ ID NO:12 *Aspergillus niger*

MPLTPEQIKIKATVPVLQEYGTKITTA FYMNMSTVHPELNAVFNTANQVKGHQARALAGALFAYASHI  
DDL GALGPAVELICNKHASLYIQADEYKIVGKYLLEAMKEVLGDACTDDILD AWGAAYWALADIMINRE  
AALYKQSQG

SEQ ID NO:13 *Zea mays*

MALAEADDGAVVFGEEQEALV LKSWAVMKKDAANLGLRFFLKVF EIAPSAEQMFSFLRDS DVPLEKNPK  
LKTHAMSVFVMTCEAAAQLRKAGKVTVRETT LKRLGATHLRYGVADGHFEVTGFALLETIKEALPADMW  
SLEMKKAWAEAYSQLVAAIKREMPDA

SEQ ID NO:14 *Oryza sativa subsp. japonica*

MALVEGNNGVSGGAVSFSEEQEALV LKSWAIMKKDSANIGLRFFLKI FEVAPSASQMFSFLRNS DVPLE  
KNPKLKTHAMSVFVMTCEAAAQLRKAGKVTV RDTTLKRLGATHFKYGVGDAHFEVTRFALLETIKEAVP  
VDMWSPAMKSAWSEAYNQLVAAIKQEMKPAE

SEQ ID NO:15 *Arabidopsis thaliana*

MESEGKIVFTEEQEALVVKSW SVMKKN SAE LGLKLFIKIFEIAPT TTKMFSFLRDSPI PAEQNP K LKPH  
AMSVFVMCCESAVQLRKTGKVTVRETT LKRLGASHSKYGVVDEHFEVAKYALLETIKEAVPEMWSPEMK  
VAWGQAYDHLVAAIKAEMNLSN

FIG. 9 CONT.

SEQ ID NO:16 *Pisum sativum*

MGFTDKQEALVNSSWESFKQNLGNSILFYTIILEKAPAAKGLFSFLKDTAGVEDSPKLQAHAEQVFGL  
VRDSAAQLRRTKGEVVLGNATLGAIHVQRGVTDPHFVVVKEALLQTIKKASGNNWSEELNTAWEVAYDGL  
ATAIKKAMT

SEQ ID NO:17 *Vigna unguiculata*

MVAFSDKQEALVNGAYEAFKANI PKYSVVFYTTILEKAPAAKNLFSFLANGVDATNPKLTGHAEKLFGL  
VRDSAAQLRASGGVVADAALGAVHSQKAVNDAQFVVVKEALVKTLKEAVGDKWSDELGTAVELAYDELA  
AAIKKAY

SEQ ID NO:18 *Bos taurus*

MGLSDGEWQLVLNVAWGKVEADVAGHGQEV LIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGNTV  
LTALGGILKKKGHHEAEVKHLAESHANKHKI PVKYLEFISDAI IHVLHAKHPSDFGADAQAAMSKALEL  
FRNDMAAQYKVLGFHG

SEQ ID NO:19 *Sus scrofa*

MGLSDGEWQLVLNVWGKVEADVAGHGQEV LIRLFTGHPETLEKFDKFKHLKSEDEMASEDLKKHGNTV  
LTALGGILKKKGHHEAELTPLAQSHATKHKI PVKYLEFISEAIIQVLQSKH  
PGDFGADAQGAMSKALELFRNDMAAKYKELGFQG

SEQ ID NO: 20 *Equus caballus*

MGLSDGEWQQVLNVWGKVEADIAGHGQEV LIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGTVV  
LTALGGILKKKGHHEAELKPLAQSHATKHKI PIKYLEFISDAI IHVLHSHK  
PGDFGADAQGAMTKALELFRNDIAAKYKELGFQG

SEQ ID NO: 21 *Synechocystis PCC6803*

MSTLYEKLGTTAVDLAVDKFYERVLQDDRIKHFFADVDMAKQRAHQKAFPTYAFGGTDKY  
DGRYMREAHKELVENHGLNGEHFDVAEDLLATLKEMGVPEDLIAEVA AVAGAPAHKRDLV  
NQ

FIG. 9 CONT.

SEQ ID NO: 22: *Synechococcus sp. PCC 7335*

MDVALLEKSFEQISPRAIIEFSASFYQNLFHHPPELKPFAETSQTIQEKKLIFSLAAHENLRNPDIL  
QPALKSLGARHAEVGTIKSHYPLVGQALIEFAEYLAADWTEQLATAWVEAYDVIASMTIEGA  
DNPAAYLEPELTFYEWLDLYGEESPKVRNAIATLTHFHYGEDPQDVQRDSRG

SEQ ID NO: 23 *Nostoc commune*

MSTLYDNIGGQPAIEQVVDLHKRIATDSLLAPVFAGTDMVKQRNHLVAFLAQIFEGPKQYGG  
RPMDKTHAGLNLQPPHFDIAIAKHLGERMAVRGVSVAENTKAALDRVTNMKGAILNK

SEQ ID NO: 24: *Vitreoscilla stercoraria*

MLDQQTINII KATVPVLKEH GVTITTFYK NLFKHPPEVR PLFDMGRQES LEQPKALAMT VLAAAQNIEN  
LPAILPAVKK IAVKHCQAGV AAHYPIVGQ ELLGAIKEVL GDAATDDILD AWGKAYGVIA DVFIQVEADL YAQAVE

SEQ ID NO: 25 *Corynebacterium glutamicum ATCC 13032*

MTTSENFYDSVGGEEFSLIVHRFYEQVPNDLILGPMYPPDDFEGAEQRLKMFLSQYWGGPKD  
YQEQRGHPRLMRHVNYPIGVTAERWLQLMSNALDGVDLTAEQREAIWEHMOVRAADMLIN  
SNPDPHA

SEQ ID NO: 26 *Bacillus subtilis*

MGQSFNAPYEAIGEELLSQLVDTFYERVASHPLLKPIFPSDLTETARKQKQFLTQYLGPPPLYTE  
EHGHPMLRARHLPPFITNERADAWLSCMKDAMDHVGLEGEIREFLFGRLELTARHMVNQTEA  
EDRSS

SEQ ID NO: 27 *Bacillus megaterium*

MREKIHSPYELLGGEHTISKLVDAFYTRVGQHPPELAPIFPDNLTTETARKQKQFLTQYLGGPSLY  
TEEHGHPMLRARHLPPFEITPSRAKAWLTCMHEAMDEINLEGPERDELYHRLILTAQHMINSP  
QTDEKGFHS

SEQ ID NO: 28: *Saccharomyces cerevisiae*

MLAEKTRSI IKATVPVLEQQGTVITRTFYKNMLTEHTELLNIFNRTNQKVGQAQPNALATTVLAAAKNID  
DLSVLMDHVKQIGHKRALQIKPEHYPIVGEYLLKAIKEVLGDAATPEI INAWGEAYQAIADIFITVEK  
KMYEALWPGWKPFDITAKEYVASDIVEFTVKPKFGSGIELESLEPITPGQYITVNTHTPIRQENQYDALR  
HYSLCSASTKNGLRFVAVKMEAAARENFPAGLVSEYLHKDAKVGDEIKLSAPAGDFAINKELIHQNEVPLV  
LLSSGVGVTPLLAMLEEQVKCNPNRPI YWIQSSYDEKTAFAFKKHVDELLAECANVDKI IVHTDTEPLIN  
AAFLKEKSPAADVYTCGSLAFMQAMIGHLKELEHRDDMIHYEPFGPKMSTVQV

FIG. 9 CONT.

SEQ ID NO: 29 *Nicotina tobaccum*

MSSFSEEQEALVLSWDSMKNAGEWGLKFLKIFEIAPSAKKLFSFLKDSNVPLEQNAKPKHAKSVFVMTCEAAVQ  
LRKAGKVVVRDSTLKKLGAHFYGVADHFEVTKFALLETIKEAVPDMWSVDMKNAWGEAFDQLVNAIKTEMK

SEQ ID NO: 30 *Medicago sativa*

mgtdtkgft eeqealvvks wnamkknsae lglklflkif eiapsaqklf sflkdkvpl eqntklkpha msvflmtces avqlrksgkv  
tvresslklk ganhfkygvv dehfevtfka lletikeavp emwspamkna wgeaydqlvn aiksemkpss

SEQ ID NO 31: *Glycine max*

MTTTLERGFSEEQEALVVKSWNVMMKNSGELGLKFFLKIFEIAPSAQKLFSFLRDSTVPLEQNPCLKPHAVSVFVMTCD  
SAVQLRKAGKVTVRESNLKLGATHFRTGVANEHFEVTKFALLETIKEAVPEMWSPAMKNAWGEAYDQLVDAIKSE  
MKPPSS

## SECRETION OF HEME-CONTAINING POLYPEPTIDES

### CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation of U.S. patent application Ser. No. 15/021,447 filed on Mar. 11, 2016, which is a U.S. National Application of PCT Application No. PCT/US2014/055227 filed on Sep. 11, 2014, which claims priority to U.S. Provisional Application Ser. No. 61/876,676, filed Sep. 11, 2013, and to U.S. Provisional Application Ser. No. 61/908,689, filed Nov. 25, 2013, and is related to the following patent applications: Application Serial No. PCT/US12/46560, filed Jul. 12, 2012; Application Serial No. PCT/US12/46552, filed Jul. 12, 2013; U.S. Provisional Application Ser. No. 61/751,816, filed Jan. 11, 2013; and U.S. Application Ser. No. 61/751,818, filed Jan. 11, 2013, all of which are incorporated herein by reference in their entirety.

### TECHNICAL FIELD

[0002] This invention relates to methods and material for producing heme-containing polypeptides, and more particularly, to producing heme-containing polypeptides in recombinant bacterial cells such as *Bacillus* cells or in recombinant plants or plant cells.

### BACKGROUND

[0003] There is a continuing need for methods to produce proteins at large scale for industrial and food purposes. *Bacillus* species can be used in the production of industrial enzymes such as lipases and proteases. In addition, a number of food additives such as glucoamylase, lipases, and amylases can be produced in these hosts, providing a long history of safe use in the food industry. *Bacillus* species can secrete high levels of protein into the media surrounding the bacteria. Plant species, such as *Nicotiana tabacum* or *Glycine max* can also be used for the production of proteins.

[0004] Heme-containing polypeptides can be difficult to secrete as the cofactor must be inserted into the polypeptide and remain associated with the polypeptide throughout the secretion process in its native configuration. *Bacillus* species may use two different systems to secrete proteins (SEC and TAT). The SEC pathway unfolds the protein as they pass through the cell membrane. The TAT system can secrete the proteins in the folded state. However, it is unclear whether a recombinant hemoprotein containing a non-covalently bound heme group can be expressed, secreted and folded properly by the *Bacillus* system, until it is successfully done.

### SUMMARY

[0005] In one aspect, this document features a recombinant bacterium cell (e.g., a *Bacillus* cell such as a *Bacillus subtilis*, *Bacillus megaterium*, or *Bacillus licheniformis* cell) capable of secreting a heme-containing polypeptide. The cell includes at least one exogenous nucleic acid, the exogenous nucleic acid comprising first and second nucleic acid sequences, wherein the first nucleic acid sequence encodes a signal peptide and the second nucleic acid sequence encodes a heme-containing polypeptide, wherein the first and second nucleic acid sequences are operably linked to produce a fusion polypeptide comprising the signal peptide and the heme-containing polypeptide. The exogenous

nucleic acid also can include a third nucleic acid sequence encoding a tag such as an affinity tag. The cell can secrete the heme-containing polypeptide from the cell, and upon secretion, the signal peptide is removed from the heme-containing polypeptide. The signal peptide can comprise or consist of an amino acid sequence having at least 60% identity to a signal peptide set forth in SEQ ID NO: 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, or 93. For example, the signal peptide can comprise or consist of an amino acid sequence having at least 60% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:55 or to residues 1-52 of SEQ ID NO:55

[0006] This document also features a method for producing a heme-containing polypeptide. The method includes culturing a recombinant bacterium cell (e.g., a *Bacillus* cell such as a *Bacillus subtilis*, *Bacillus megaterium*, or *Bacillus licheniformis* cell) in a culture medium under conditions that allow the heme-containing polypeptide to be secreted into the culture medium, the recombinant bacterium cell comprising at least one exogenous nucleic acid, the exogenous nucleic acid comprising first and second nucleic acid sequences, wherein the first nucleic acid sequence encodes a signal peptide and the second nucleic acid sequence encodes a heme-containing polypeptide, wherein the first and second nucleic acid sequences are operably linked to produce a fusion polypeptide comprising the signal peptide and the heme-containing polypeptide, and wherein upon secretion of the fusion polypeptide from the cell into the culture medium, the signal peptide is removed from the heme-containing polypeptide. The method further can include recovering the heme-containing polypeptide from the culture medium. The signal peptide can comprise or consist of an amino acid sequence having at least 60% identity to a signal peptide set forth in SEQ ID NO: 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, or 93. For example, the signal peptide can comprise or consist of an amino acid sequence having at least 60% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:55 or to residues 1-52 of SEQ ID NO:55.

[0007] In another aspect, this document features a recombinant plant or plant cell (a *Glycine max*, *Zea mays*, *Hordeum vulgare*, or *Arabidopsis thaliana* plant or plant cell) producing a heme-containing polypeptide. The plant or plant cell can include at least one exogenous nucleic acid encoding a heme-containing polypeptide, wherein the plant or plant cell is from a species other than *Nicotiana*. The exogenous nucleic acid further can include a regulatory control element such as a promoter (e.g., a tissue-specific promoter such as leaves, roots, stems, or seeds). The exogenous nucleic acid also can encode a signal peptide that targets the heme-containing polypeptide to a subcellular location such as an oil body, vacuole, plastid (e.g., chloroplast), or other organelle.

[0008] This document also features a method of producing a heme-containing polypeptide. The method can include growing a recombinant plant (a *Glycine max*, *Zea mays*, *Hordeum vulgare*, or *Arabidopsis thaliana* plant), the recombinant plant comprising at least one exogenous nucleic acid encoding the heme-containing polypeptide, wherein the plant is from a species other than *Nicotiana*, and purifying the heme-containing polypeptide from a tissue of the plant.



**[0009]** In another aspect, this document features a vector that includes a polynucleotide sequence encoding a heme-containing polypeptide; and a polynucleotide sequence encoding a signal peptide, wherein the signal peptide comprises or consists of an amino acid sequence having at least 60% amino acid sequence identity to a signal peptide listed in Table 1. For example, the signal peptide can include an amino acid sequence having at least 60% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO:55 or to residues 1-52 of SEQ ID NO:55. In some embodiments, the signal peptide comprises or consists of the amino acid sequence of residues 1-52 of SEQ ID NO:55. The polynucleotide sequence encoding the heme-containing polypeptide can be operably linked to a promoter.

**[0010]** In yet another aspect, this document features a composition that includes a purified heme-containing polypeptide; and a recombinant *Bacillus* cell (e.g., *Bacillus subtilis*, *Bacillus megaterium*, or *Bacillus licheniformis* cell) or a recombinant plant cell other than a *Nicotiana* plant cell (e.g., a non-naturally occurring component of a recombinant *Bacillus* cell or a recombinant plant cell). In some instances, the heme-containing polypeptide does not naturally occur in the host cell. The plant cell can be, for example, a *Glycine max* plant cell, a *Zea mays* plant cell, or an *Arabidopsis thaliana* plant cell. The composition can include at least 1 part per billion of said component of the cell or at most 1% (w/w) of the component of the cell.

**[0011]** This document also features a vector comprising a polynucleotide sequence encoding a heme-containing polypeptide, a signal peptide; and a tag, wherein expression of the polynucleotide sequence in a host cell produces a fusion protein containing the heme-containing polypeptide, the signal peptide, and the tag (e.g., an affinity tag such as a 6-histidine tag or a detectable tag) and genetically modified organisms containing such a vector. The vector further can include a polynucleotide sequence encoding at least one of: a) an amino acid linker between the sequence encoding the tag and the sequence encoding the heme-containing polypeptide; and b) an amino acid linker between the sequence encoding the signal peptide and the sequence encoding the heme-containing polypeptide.

**[0012]** In one aspect, this document features a method for secreting a heme-containing polypeptide from a bacterium (e.g., a *Bacillus* cell such as a *Bacillus subtilis*, *Bacillus megaterium*, or *Bacillus licheniformis* cell) that includes culturing a recombinant bacterium under conditions that allow the heme-containing polypeptide to be secreted from the bacterium, the recombinant bacterium comprising an exogenous nucleic acid encoding the heme-containing polypeptide, a signal peptide, and a tag.

**[0013]** This document also features a purified fusion polypeptide that includes a heme-containing polypeptide and a tag. The polypeptide further can include a linker between the tag and the heme-containing polypeptide. In some embodiments, the tag can be located at the C-terminus of the heme-containing polypeptide either directly bound to the C-terminus or via a linker.

**[0014]** In any of the methods, compositions, recombinant bacterial cells, recombinant plants or plant cells, or vectors, the heme-containing polypeptide can be selected from the group consisting of an androglobin, a cytoglobin, globin E, globin X, globin Y, a hemoglobin, a myoglobin, a leghemoglobin, an erythrocyruorin, a beta hemoglobin, an alpha hemoglobin, a non-symbiotic hemoglobin, a flavohemoglobin,

bin, a protoglobin, a cyanoglobin, a Hell's gate globin I, a bacterial hemoglobin, a ciliate myoglobin, a histoglobin, a neuroglobin, a protoglobin, and a truncated globin (e.g., truncated 2/2 globin, HbN, HbO, or Glb3). For example, the heme-containing polypeptide can have at least 60% sequence identity to an amino acid sequence set forth in SEQ ID NOs: 1-31.

**[0015]** In one aspect, the disclosure provides for a vector comprising a polynucleotide sequence encoding a heme-containing polypeptide, e.g., a globin, and a polynucleotide sequence encoding a signal peptide. In some embodiments, the signal peptide is for a secretory pathway. In some such embodiments, the signal peptide can be referred to as a signal peptide or a secretion signal peptide. In some embodiments, the signal peptide directs said heme-containing polypeptide, e.g., a globin, into a secretory pathway. In some embodiments, the signal peptide comprises at least 60% amino acid sequence identity to a signal peptide listed in Table 1. In some embodiments, the signal peptide comprises at least 60% amino acid sequence identity to a PhoD signal peptide (e.g., at least about 65, 70, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% amino acid sequence identity). In some embodiments, the signal peptide is a PhoD signal peptide. In some embodiments, the polynucleotide sequence encoding a heme-containing polypeptide, e.g., a globin, is operably linked to a promoter. In some embodiments, the heme-containing polypeptide is selected from the group consisting of: androglobin, cytoglobin, globin E, globin X, globin Y, hemoglobin, myoglobin, leghemoglobin, erythrocyruorin, beta hemoglobin, alpha hemoglobin, non-symbiotic hemoglobin, flavohemoglobin, protoglobin, cyanoglobin, non-symbiotic hemoglobin, Hell's gate globin I, bacterial hemoglobin, ciliate myoglobin, flavohemoglobin, histoglobin, neuroglobins, protoglobin, truncated 2/2 globin, HbN, HbO, and, Glb3.

**[0016]** In some embodiments, the disclosure provides for a genetically modified organism comprising a vector comprising a polynucleotide sequence encoding a heme-containing polypeptide, e.g., a globin, and a polynucleotide sequence encoding a signal peptide. In some embodiments, the genetically modified organism is a gram positive species of bacteria. In some embodiments, the genetically modified organism is a *Bacillus* species. In some embodiments, the genetically modified organism is selected from the group consisting of *Bacillus subtilis*, *Bacillus megaterium*, and *Bacillus licheniformis*. In some embodiments, the genetically modified organism is a *Nicotiana* species. In some embodiments, the genetically modified organism is a *Nicotiana tabacum*.

**[0017]** In one aspect, the disclosure provides for a non-naturally occurring composition comprising a purified heme-containing polypeptide, e.g., a globin, and a part of a host cell. In some embodiments, the heme-containing polypeptide is a globin that does not naturally occur in said host cell. In some embodiments, the host cell is a *Nicotiana* species of plant. In some embodiments, the host cell is a *Nicotiana tabacum* species of plant. In some embodiments, the host cell is a *Glycine max* species of plant. In some embodiments, the host cell is selected from the group consisting of *Nicotiana tabacum*, *Glycine max*, *Zea mays*, and *Arabidopsis thaliana*. In some embodiments, the composition comprises at least 1 part per billion of said part of the host cell. In some embodiments, the composition com-

prises at most 1% (w/w) of said part of a host cell. In some embodiments, the heme-containing polypeptide is selected from the group consisting of: androglobin, cytoglobin, globin E, globin X, globin Y, hemoglobin, myoglobin, leghemoglobin, erythrocrurin, beta hemoglobin, alpha hemoglobin, non-symbiotic hemoglobin, flavohemoglobin, protoglobin, cyanoglobin, non-symbiotic hemoglobin, Hell's gate globin I, bacterial hemoglobin, ciliate myoglobin, flavohemoglobin, histoglobin, neuroglobins, protoglobin, truncated 2/2 globin, HbN, HbO, and, Glb3. In some embodiments, the heme-containing polypeptide is a globin. In some embodiments, the globin is leghemoglobin. In some embodiments, the globin is hemoglobin. In some embodiments, the heme-containing polypeptide comprises at least 60% amino acid sequence identity to an amino acid sequence listed in FIG. 9 (e.g., at least about 65, 70, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% amino acid sequence identity). In some embodiments, the heme-containing polypeptide comprises a tag, e.g., is covalently bound to a tag, e.g., at the C or N-terminus. In some embodiments, a meat consumable comprises a heme-containing polypeptide as described herein. In some embodiments, a meat consumable comprises at least 0.001% (w/w) of the heme-containing polypeptide. In some embodiments, the meat consumable comprises at most 10% (w/w) of the globin. In some embodiments, the meat consumable comprises a replica selected from the group consisting of: a fat replica, a connective tissue replica, and a muscle replica, or any combination thereof. In some embodiments, the meat consumable accurately recapitulates key features associated with the cooking and consumption of an equivalent meat product derived from animals. In some embodiments, the host cell is a bacterium. In some embodiments, the heme-containing polypeptide is secreted from said host cell. In some embodiments, the host cell is a *Bacillus* species of bacterium. In some embodiments, the host cell is *Bacillus subtilis*.

**[0018]** In one aspect the disclosure provides for a method for purifying a heme-containing polypeptide, e.g., a globin, from a plant cell comprising inserting a polynucleotide comprising a polynucleotide encoding a heme-containing polypeptide, e.g., a globin, into a plant cell, and purifying the heme-containing polypeptide. In some embodiments, the polynucleotide further comprises a sequence encoding a tag. In some embodiments, the plant cell is a *Nicotiana* species. In some embodiments, the plant cell is *Nicotiana tabacum*. In some embodiments, the plant cell is *Glycine max*. In some embodiments, the plant cell is selected from the group consisting of *Nicotiana tabacum*, *Glycine max*, *Zea mays*, and *Arabidopsis thaliana*. In some embodiments, the heme-containing polypeptide is selected from the group consisting of: androglobin, cytoglobin, globin E, globin X, globin Y, hemoglobin, myoglobin, leghemoglobin, erythrocrurin, beta hemoglobin, alpha hemoglobin, non-symbiotic hemoglobin, flavohemoglobin, protoglobin, cyanoglobin, non-symbiotic hemoglobin, Hell's gate globin I, bacterial hemoglobin, ciliate myoglobin, flavohemoglobin, histoglobin, neuroglobins, protoglobin, truncated 2/2 globin, HbN, HbO, and, Glb3. In some embodiments, the heme-containing polypeptide is a globin. In some embodiments, the globin is leghemoglobin. In some embodiments, the globin is hemoglobin. In some embodiments, the heme-containing polypeptide comprises at least 60% amino acid sequence identity

to an amino acid sequence set forth in FIG. 9 (e.g., at least about 65, 70, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% amino acid sequence identity). In some embodiments, the method further comprises combining a heme-containing polypeptide with a meat consumable. In some embodiments, the meat consumable comprises a replica selected from the group consisting of: a fat replica, a muscle replica, and a connective tissue replica, or any combination thereof. In some embodiments, the meat consumable comprises at least 0.001% (w/w) of said heme-containing polypeptide, e.g., a globin. In some embodiments, the meat consumable comprises at most 10% (w/w) of said heme-containing polypeptide. In some embodiments, the meat consumable accurately recapitulates key features associated with the cooking and consumption of an equivalent meat product derived from animals.

**[0019]** In one aspect the disclosure provides for a method for purifying an endogenous heme-containing polypeptide, e.g., a globin, from a plant comprising altering the expression levels of an endogenous heme-containing polypeptide in a plant, and purifying the heme-containing polypeptide from said plant. In some embodiments, the altering increases the expression levels of the endogenous heme-containing polypeptide. In some embodiments, the altering increases the expression levels of said endogenous heme-containing polypeptide in a leaf, a seed, a bean, or any combination thereof. In some embodiments, the altering comprises altering the expression levels of a protein in the pathway of production of the endogenous heme-containing polypeptide. In some embodiments, the plant is a *Nicotiana* species. In some embodiments, the plant cell is *Nicotiana tabacum*. In some embodiments, the plant cell is *Glycine max*. In some embodiments, the plant cell is selected from the group consisting of: *Nicotiana tabacum*, *Glycine max*, *Zea mays*, and *Arabidopsis thaliana*. In some embodiments, the heme-containing polypeptide is a globin selected from the group consisting of: hemoglobin, leghemoglobin, non-symbiotic hemoglobin, and, Glb3. In some embodiments, the globin is leghemoglobin. In some embodiments, the globin is hemoglobin. In some embodiments, the heme-containing polypeptide comprises at least 60% amino acid sequence identity to an amino acid sequence listed in FIG. 9 (e.g., at least about 65, 70, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% amino acid sequence identity). In some embodiments, the heme-containing polypeptide comprises a tag.

**[0020]** In one aspect the disclosure provides for a method for secreting a heme-containing polypeptide from a bacterium comprising inserting a polynucleotide comprising a polynucleotide encoding a heme-containing polypeptide and a signal peptide into a bacterium, and secreting said heme-containing polypeptide from said bacterium. In some embodiments, the bacterium is a *Bacillus* species. In some embodiments, the bacterium is *Bacillus subtilis*. In some embodiments, the heme-containing polypeptide is selected from the group consisting of: androglobin, cytoglobin, globin E, globin X, globin Y, hemoglobin, myoglobin, leghemoglobin, erythrocrurin, beta hemoglobin, alpha hemoglobin, non-symbiotic hemoglobin, flavohemoglobin, protoglobin, cyanoglobin, Hell's gate globin I, bacterial hemoglobin, ciliate myoglobin, histoglobin, neuroglobins, truncated 2/2 globin, HbN, HbO, and, Glb3. In some embodiments, the heme-containing polypeptide is a globin.

In some embodiments the globin is leghemoglobin. In some embodiments, the globin is hemoglobin. In some embodiments, the heme-containing polypeptide comprises at least 60% amino acid sequence identity to an amino acid sequence listed in FIG. 9 (e.g., at least about 65, 70, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% amino acid sequence identity). In some embodiments, the signal peptide is a signal peptide for a secretory pathway. In some embodiments, the signal peptide directs said heme-containing polypeptide into a secretory pathway. In some embodiments, the signal peptide comprises at least about 60% amino acid sequence identity to a signal peptide listed in Table 1 (e.g., at least about 65, 70, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% amino acid sequence identity). In some embodiments, the polynucleotide sequence encodes for a signal peptide comprising at least 60% amino acid sequence identity to a PhoD signal peptide (e.g., at least about 65, 70, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% amino acid sequence identity). In some embodiments, the polynucleotide sequence encodes for a PhoD signal peptide. In some embodiments, the method further comprises purifying the heme-containing polypeptide. In some embodiments, the method further comprises combining said purified heme-containing polypeptide with a meat consumable. In some embodiments, the meat consumable comprises a fat replica, a muscle replica, and a connective tissue replica, or any combination thereof. In some embodiments, the meat consumable comprises at least 0.001% (w/w) of said purified heme-containing polypeptide. In some embodiments, the meat consumable comprises at most 10% (w/w) of said purified heme-containing polypeptide. In some embodiments, the meat consumable accurately recapitulates key features associated with the cooking and consumption of an equivalent meat product derived from animals.

[0021] In one aspect, the disclosure provides for a method for secreting a heme-containing polypeptide from a bacterium comprising altering the expression levels of an endogenous heme-containing polypeptide in a bacterium, and purifying said heme-containing polypeptide from said bacterium. In some embodiments, the altering increases the expression levels of said endogenous heme-containing polypeptide. In some embodiments, the altering comprises altering the expression levels of a protein in the pathway of production of said endogenous heme-containing polypeptide. In some embodiments, the bacterium is a *Bacillus* species. In some embodiments, the bacterium is *Bacillus subtilis*. In some embodiments, the heme-containing polypeptide is selected from the group consisting of: hemoglobin, flavohemoglobin, cyanoglobin, Hell's gate globin I, bacterial hemoglobin, HbN, and HbO. In some embodiments, the heme-containing polypeptide is hemoglobin. In some embodiments, the heme-containing polypeptide comprises at least 60% amino acid sequence identity to an amino acid sequence listed in FIG. 9 (e.g., at least about 65, 70, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% amino acid sequence identity). In some embodiments, the heme-containing polypeptide comprises a tag. In some embodiments, the method further comprises purifying the heme-containing polypeptide. In some embodiments, the method further comprises combining the purified heme-containing polypeptide with a

meat consumable. In some embodiments, the meat consumable comprises a fat replica, a muscle replica, and a connective tissue replica, or any combination thereof.

#### INCORPORATION BY REFERENCE

[0022] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference. In the event of multiple definitions describing the same concept, the definition in the instant application can govern.

#### BRIEF DESCRIPTION OF THE DRAWINGS

[0023] The novel features of the disclosure are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles are utilized, and the accompanying drawings of which:

[0024] FIG. 1 contains three SDS-PAGE gels of the proteins after Ni-NTA affinity purification of the cell pellet, showing a comparison of the effect of a secretion signal peptide (PhoD) on cytosolic expression of *Aquifex aeolicus* hemoglobin (AaHb). FIG. 1A is the empty vector control. FIG. 1B is without the secretion signal peptide and FIG. 1C is with the secretion signal peptide.

[0025] FIG. 2 is a graph depicting the heme content of a cytosolically expressed polypeptide (AaHb). The line corresponding to the empty vector is the line with the lowest peak. The line corresponding to the polypeptide with no signal peptide is the line with the highest peak. The line corresponding to the polypeptide with the PhoD signal peptide is the line with the second highest peak.

[0026] FIG. 3 contains two SDS-PAGE gels of the proteins after Ni-NTA affinity purification of the media, showing a comparison of the effect of a secretion signal peptide (PhoD) on secretory expression of a polypeptide (AaHb). FIG. 3A is the empty vector control. FIG. 3B is with the secretion signal peptide.

[0027] FIG. 4 is the sequence of the fusion polypeptide containing the PhoD (in bold text)-synthetic protease cleavage site (in italics, ASAA)—AaHb (underlined)-His6 (double underlined) sequence (SEQ ID NO: 94). The predicted signal peptidase I (SPI) recognition site is shown (SEQ ID NO:95), with the cleavage site indicated. N-terminal sequencing (SEQ ID NO:96) of the secreted polypeptide present in the media after cytosolic expression of a PhoD-AaHb fusion polypeptide indicated the N-terminus corresponded to the N-terminus of the AaHb protein.

[0028] FIG. 5 depicts the heme content of a secreted polypeptide (after expression of the fusion polypeptide PhoD-AaHb). The line corresponding to the empty vector is the line with the lowest peak. The line corresponding to the secreted polypeptide after expression of the fusion polypeptide which included the PhoD signal peptide is the line with the highest peak.

[0029] FIG. 6 illustrates 1) detection of two exemplary fusion polypeptides (PhoD-yjbl and YwbN-yjbl) in the cell pellet following expression of an endogenous polypeptide (yjbl) fused to one of two different signaling peptides (PhoD

or YwbN), and 2) media detection of the polypeptides demonstrating proper cleavage of the signal peptides.

**[0030]** FIG. 7 illustrates 1) detection of two exemplary fusion polypeptides (PhoD-LGB2 and PhoD-HGN) in the cell pellet following expression of two heterologous polypeptides (LGB2 and HGB1) fused to the signaling peptide (PhoD) and 2) media detection of the polypeptides, demonstrating proper cleavage of the signal peptide.

**[0031]** FIG. 8 illustrates media detection of a polypeptide (AaHb) after fusion with a subset of a number of different exemplary secretion signaling peptides (PhoD, TipA, WapA, WprA, YmaC, YoaA, YuiC, YwbN, AppB, and BglS), expression, and cleavage of the secretion signal peptide. Labels indicate the secretion signal peptide that was fused to the 5' end of the polypeptide prior to cleavage.

**[0032]** FIG. 9 contains the amino acid sequences of exemplary heme-containing polypeptides (SEQ ID NOs: 1-31).

#### DETAILED DESCRIPTION

**[0033]** Polypeptides

**[0034]** This disclosure provides for compositions and methods for the expression of a polypeptide in a host cell (e.g., bacteria and/or plants). A polypeptide can refer to subunits or domains of a polypeptide. A polypeptide of the disclosure can be a heme-containing polypeptide. The term heme-containing polypeptide can refer to all proteins or protein subunits that are capable of covalently or noncovalently binding a heme moiety. Heme-containing polypeptides can transport or store oxygen. In some instances, the polypeptide of the disclosure can be a globin. Polypeptides can comprise the globin fold, which can comprise a series of eight alpha helices. A polypeptide can comprise an alpha globin and/or a beta globin. A polypeptide can comprise a characteristic higher structure (e.g., the "myoglobin fold") generally associated with globins. A polypeptide can be an oligomer. Polypeptides can be monomers, dimers, trimers, tetramers, and/or higher order oligomers. In some instances, a polypeptide can be an iron-containing polypeptide.

**[0035]** A polypeptide of the disclosure can include, but is not limited to, androglobin, cytoglobin, globin E, globin X, globin Y, hemoglobin, myoglobin, leghemoglobins, erythrocrucorins, beta hemoglobins, alpha hemoglobins, non-symbiotic hemoglobins, flavohemoglobins, protoglobins, cyanoglobins, cytoglobin, Hell's gate globin I, bacterial hemoglobins, ciliate myoglobins, histoglobins, neuroglobins, chlorocruorin, erythrocrucorin, protoglobin, truncated 2/2 globin, HbN, HbO, G1b3, and cytochromes, ribosomal proteins, actin, hexokinase, lactate dehydrogenase, fructose bisphosphate aldolase, phosphofructokinases, triose phosphate isomerases, phosphoglycerate kinases, phosphoglycerate mutases, enolases, pyruvate kinases, proteases, lipases, amylases, glycoproteins, lectins, mucins, glyceraldehyde-3-phosphate dehydrogenases, pyruvate decarboxylases, actins, translation elongation factors, histones, ribulose-1,5-bisphosphate carboxylase oxygenase (rubisco), ribulose-1,5-bisphosphate carboxylase oxygenase activase (rubisco activase), albumins, glycinins, conglycinins, globulins, vicilins, conalbumin, gliadin, glutelin, gluten, glutenin, hordein, prolamin, phaseolin (protein), proteinoplast, secalin, extensins, triticeae gluten, collagens, zein, kafirin, avenin, dehydrins, hydrophilins, late embryogenesis abundant proteins, natively unfolded proteins, any seed storage protein, oleosins, caloleosins, steroleosins or other oil body proteins, vegetative storage protein A, vegetative storage

protein B, moong seed storage 8S globulin, globulin, pea globulins, and pea albumins. In some instances, a polypeptide of the disclosure can comprise or can be a polypeptide listed in FIG. 9. In some instances, a polypeptide can be introduced into a host cell. For example, a polypeptide can be expressed, secreted, and/or purified from bacteria such as a species of *Bacillus*. A polypeptide can be expressed and/or purified from a plant.

**[0036]** A polypeptide listed in FIG. 9, may be expressed, but may not be properly secreted and/or folded using the methods of the disclosure. A polypeptide listed in FIG. 9 may be expressed, but may not be correctly localized in the cell using the methods of the disclosure. A polypeptide listed in FIG. 9, may be expressed, but may not retain levels of activity comparable to a wild-type polypeptide. A polypeptide listed in FIG. 9 may retain at least about 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100% activity level of a wild-type polypeptide. A polypeptide listed in FIG. 9 may retain at most about 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100% activity level of a wild-type polypeptide. A polypeptide comprising at least about 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100% amino acid sequence identity to a polypeptide listed in FIG. 9 may be expressed, but may not be properly secreted and/or folded using the methods of the disclosure. A polypeptide comprising at most about 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100% amino acid sequence identity to a polypeptide listed in FIG. 9 may be expressed, but may not be properly secreted and/or folded using the methods of the disclosure. A polypeptide comprising at most about 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100% amino acid sequence identity to a polypeptide listed in FIG. 9 may be expressed, but may not be retain activity compared to a wild-type polypeptide. A polypeptide comprising at most about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100% amino acid sequence identity to a polypeptide listed in FIG. 9 may be expressed, but may contain less heme cofactor compared to a wild-type polypeptide.

**[0037]** In some instances, a sequence of a polypeptide to be expressed in a host cell can be a sequence comprising at least about 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100% amino acid sequence identity to an endogenous polypeptide sequence, e.g., an endogenous heme-containing polypeptide, of the host cell. In some instances, a sequence of a polypeptide to be expressed in a host cell can be a sequence comprising at most about 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100% amino acid sequence identity to an endogenous polypeptide sequence of the host cell. For example, a polypeptide can be a polypeptide sequence found in an animal, a mammal, a vertebrate, an invertebrate, a plant, a fungus, a bacterium, a yeast, an alga, an archaea, a genetically modified organism such as a genetically modified bacterium or yeast. A polypeptide sequence can be chemically synthesized, and/or synthesized by in vitro synthesis.

**[0038]** A polypeptide sequence can be a sequence of a polypeptide, e.g., a heme-containing polypeptide, found in plants. Non-limiting examples of plants can include grains such as, e.g., corn, maize, oats, rice, wheat, barley, rye, triticale, teff, oilseeds including cottonseed, sunflower seed, safflower seed, crambe, camelina, mustard, rapeseed, leafy greens such as, e.g., lettuce, spinach, kale, collard greens, turnip greens, chard, mustard greens, dandelion greens, broccoli, cabbage, sugar cane, trees, root crops such as cassava, sweet potato, potato, carrots, beets, turnips, plants from the legume family, such as, e.g., clover, peas such as

cowpeas, English peas, yellow peas, green peas, beans such as, e.g., soybeans, fava beans, lima beans, kidney beans, garbanzo beans, mung beans, pinto beans, lentils, lupins, mesquite, carob, soy, and peanuts, coconut, vetch (*vicia*), stylo (*stylosanthes*), arachis, indigofera, acacia, leucaena, cyamopsis, and sesbania. Plants not ordinarily consumed by humans, including biomass crops, including, for example, switchgrass, miscanthus, tobacco, *Arundo donax*, energy cane, sorghum, other grasses, alfalfa, corn stover, kelp, or other seaweeds. Polypeptides that can be found in any organism in the plant kingdom may be used in the present disclosure. In some instances, the plant can be soy. In some instances, the plant can be barley.

**[0039]** In some instances, a polypeptide sequence can be a sequence, e.g., a heme-containing polypeptide sequence, found in metazoa. For example, a polypeptide sequence of the disclosure can be a polypeptide sequence found in mammals such as cow, pig, rat, dog, or horse. In some instances, the polypeptide sequence comes from cow. In some instances, the polypeptide sequence comes from pig. In some instances, a polypeptide sequence can be a sequence found in protists. For example, a polypeptide sequence of the disclosure can be a polypeptide sequence found in protists such as algae. In some instances, a polypeptide sequence can be a sequence found in archaea. For example, a polypeptide sequence of the disclosure can be a polypeptide sequence found in archaea such as halobacteria or pyrococcus. In some instances, a polypeptide sequence can be a sequence found in eubacteria. For example, a polypeptide sequence of the disclosure can be a polypeptide sequence found in eubacteria such as *Bacillus*, *Clostridia*, or *Escherichia*.

**[0040]** As used herein, the term “heme containing protein” includes any polypeptide that can covalently or noncovalently bind to a heme moiety. In some embodiments, the heme-containing polypeptide is a globin and can include a globin fold, which comprises a series of seven to nine alpha helices. Globin type proteins can be of any class (e.g., class I, class II, or class III), and in some embodiments, can transport or store oxygen. For example, a heme-containing polypeptide can be a non-symbiotic type of hemoglobin or a leghemoglobin. A heme-containing polypeptide can be a monomer, i.e., a single polypeptide chain, or can be a dimer, a trimer, tetramer, and/or higher order oligomers. The lifetime of the oxygenated Fe<sup>+</sup> state of a heme-containing polypeptide can be similar to that of myoglobin or can exceed it by 10%, 20%, 30%, 40%, 50%, 100% or more.

**[0041]** Non-limiting examples of heme-containing polypeptides can include an androglobin, a cytoglobin, a globin E, a globin X, a globin Y, a hemoglobin, a myoglobin, an erythrocrucorin, a beta hemoglobin, an alpha hemoglobin, a protoglobin, a cyanoglobin, a histoglobin, a neuroglobin, a chlorocruorin, a truncated hemoglobin (e.g., HbN, HbO, a truncated 2/2 globin, a hemoglobin 3 (e.g., Glb3)), a cytochrome, or a peroxidase.

**[0042]** Heme-containing polypeptides can be from mammals (e.g., farm animals such as cows, goats, sheep, pigs, ox, or rabbits), birds, plants, algae, fungi (e.g., yeast or filamentous fungi), ciliates, or bacteria. For example, a heme-containing polypeptide can be from a mammal such as a farm animal (e.g., a cow, goat, sheep, pig, ox, or rabbit) or a bird such as a turkey or chicken. Heme-containing polypeptides can be from a plant such as *Nicotiana tabacum* or *Nicotiana sylvestris* (tobacco); *Zea mays* (corn), *Arabidopsis*

*thaliana*, a legume such as *Glycine max* (soybean), *Cicer arietinum* (garbanzo or chick pea), *Pisum sativum* (pea) varieties such as garden peas or sugar snap peas, *Phaseolus vulgaris* varieties of common beans such as green beans, black beans, navy beans, northern beans, or pinto beans, *Vigna unguiculata* varieties (cow peas), *Vigna radiata* (Mung beans), *Lupinus albus* (lupin), or *Medicago sativa* (alfalfa); *Brassica napus* (canola); *Triticum* sps. (wheat, including wheat berries, and spelt); *Gossypium hirsutum* (cotton); *Oryza sativa* (rice); *Zizania* sps. (wild rice); *Helianthus annuus* (sunflower); *Beta vulgaris* (sugarbeet); *Pennisetum glaucum* (pearl millet); *Chenopodium* sp. (quinoa); *Sesamum* sp. (sesame); *Linum usitatissimum* (flax); or *Hordeum vulgare* (barley). Heme-containing polypeptides can be isolated from fungi such as *Saccharomyces cerevisiae*, *Pichia pastoris*, *Magnaporthe oryzae*, *Fusarium graminearum*, or *Fusarium oxysporum*. Heme-containing polypeptides can be isolated from bacteria such as *Escherichia coli*, *Bacillus subtilis*, *Synechocystis* sp., *Aquifex aeolicus*, *Methylococcus thermophilus*, or thermophilic bacteria such as *Thermophilus*.

**[0043]** The sequences and structure of numerous heme-containing polypeptides are known. See for example, Reedy, et al., *Nucleic Acids Research*, 2008, Vol. 36, Database issue D307-D313 and the Heme Protein Database available on the world wide web at <http://hemeprotein.info/heme.php>.

**[0044]** A non-symbiotic hemoglobin can be from a plant selected from the group consisting of soybean, sprouted soybean, alfalfa, golden flax, black bean, black eyed pea, northern, garbanzo, moong bean, cowpeas, pinto beans, pod peas, quinoa, sesame, sunflower, wheat berries, spelt, barley, wild rice, or rice.

**[0045]** Any of the heme-containing polypeptides described herein can have at least 60% (e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100%) sequence identity to the amino acid sequence of the corresponding wild-type heme-containing polypeptide or fragments thereof that contain a heme-binding motif. For example, a heme-containing polypeptide can have at least 60% sequence identity to an amino acid sequence set forth in FIG. 9, including a non-symbiotic hemoglobin such as that from *Vigna radiata* (SEQ ID NO:1), *Hordeum vulgare* (SEQ ID NO:5), *Zea mays* (SEQ ID NO:13), *Oryza sativa* subsp. *japonica* (rice) (SEQ ID NO:14), or *Arabidopsis thaliana* (SEQ ID NO:15), a Hell's gate globin I such as that from *Methylococcus thermophilus* (SEQ ID NO:2), a flavohemoprotein such as that from *Aquifex aeolicus* (SEQ ID NO:3), a leghemoglobin such as that from *Glycine max* (SEQ ID NO:4), *Pisum sativum* (SEQ ID NO:16), or *Vigna unguiculata* (SEQ ID NO:17), a heme-dependent peroxidase such as from *Magnaporthe oryzae*, (SEQ ID NO:6) or *Fusarium oxysporum* (SEQ ID NO:7), a cytochrome c peroxidase from *Fusarium graminearum* (SEQ ID NO:8), a truncated hemoglobin from *Chlamydomonas moewusii* (SEQ ID NO:9), *Tetrahymena pyriformis* (SEQ ID NO:10, group I truncated), *Paramecium caudatum* (SEQ ID NO:11, group I truncated), a hemoglobin from *Aspergillus niger* (SEQ ID NO:12), or a mammalian myoglobin protein such as the *Bos taurus* (SEQ ID NO:18) myoglobin, *Sus scrofa* (SEQ ID NO:19) myoglobin, *Equus caballus* (SEQ ID NO:20) myoglobin, a *Synechocystis* PCC6803 (SEQ ID NO:21) truncated hemoglobin, a *Synechococcus* sp. PCC 7335 (SEQ ID NO:22) truncated hemoglobin, a *Nostoc commune* (SEQ ID NO:23) hemoglobin, a *Vitreoscilla ster-*

*coraria* (SEQ ID NO:24) hemoglobin, a *Corynebacterium glutamicum* (SEQ ID NO:25) hemoglobin, a *Bacillus subtilis* (SEQ ID NO:26) truncated hemoglobin, a *Bacillus megaterium* (SEQ ID NO:27) truncated hemoglobin, a *Saccharomyces cerevisiae* (SEQ ID NO:28) flavohemoglobin, a *Nicotina tobaccum* (SEQ ID NO:29) non-symbiotic hemoglobin, a *Medicago sativa* (SEQ ID NO:30) non-symbiotic hemoglobin, or a *Glycine max* (SEQ ID NO: 31) non-symbiotic hemoglobin.

**[0046]** The percent identity between two amino acid sequences can be determined as follows. First, the amino acid sequences are aligned using the BLAST 2 Sequences (BL2seq) program from the stand-alone version of BLASTZ containing BLASTP version 2.0.14. This stand-alone version of BLASTZ can be obtained from Fish & Richardson's web site (e.g., [www.fr.com/blast/](http://www.fr.com/blast/)) or the U.S. government's National Center for Biotechnology Information web site ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). Instructions explaining how to use the BL2seq program can be found in the readme file accompanying BLASTZ. BL2seq performs a comparison between two amino acid sequences using the BLASTP algorithm. To compare two amino acid sequences, the options of BL2seq are set as follows: -i is set to a file containing the first amino acid sequence to be compared (e.g., C:\seq1.txt); -j is set to a file containing the second amino acid sequence to be compared (e.g., C:\seq2.txt); -p is set to blastp; -o is set to any desired file name (e.g., C:\output.txt); and all other options are left at their default setting. For example, the following command can be used to generate an output file containing a comparison between two amino acid sequences: C:\BL2seq -i c:\seq1.txt -j c:\seq2.txt -p blastp -o c:\output.txt. If the two compared sequences share homology, then the designated output file will present those regions of homology as aligned sequences. If the two compared sequences do not share homology, then the designated output file will not present aligned sequences. Similar procedures can be following for nucleic acid sequences except that blastn is used.

**[0047]** Once aligned, the number of matches is determined by counting the number of positions where an identical amino acid residue is presented in both sequences. The percent identity is determined by dividing the number of matches by the length of the full-length polypeptide amino acid sequence followed by multiplying the resulting value by 100. It is noted that the percent identity value is rounded to the nearest tenth. For example, 78.11, 78.12, 78.13, and 78.14 is rounded down to 78.1, while 78.15, 78.16, 78.17, 78.18, and 78.19 is rounded up to 78.2. It also is noted that the length value will always be an integer.

**[0048]** It will be appreciated that a number of nucleic acids can encode a polypeptide having a particular amino acid sequence. The degeneracy of the genetic code is well known to the art; i.e., for many amino acids, there is more than one nucleotide triplet that serves as the codon for the amino acid. For example, codons in the coding sequence for a given enzyme can be modified such that optimal expression in a particular species (e.g., bacteria or fungus) is obtained, using appropriate codon bias tables for that species.

**[0049]** Heme-containing polypeptides can be extracted from the source material (e.g., extracted from animal tissue, or plant, fungal, algal, or bacterial biomass, or from the culture supernatant for secreted proteins) or from a combination of source materials (e.g., multiple plant species). Leghemoglobin is readily available as an unused by-product

of commodity legume crops (e.g., soybean, alfalfa, or pea). The amount of leghemoglobin in the roots of these crops in the United States exceeds the myoglobin content of all the red meat consumed in the United States.

**[0050]** In some embodiments, extracts of heme-containing polypeptides include one or more non-heme-containing polypeptides from the source material (e.g., other animal, plant, fungal, algal, or bacterial proteins) or from a combination of source materials (e.g., different animal, plant, fungi, algae, or bacteria).

**[0051]** A polypeptide of the disclosure (e.g., a globin, a heme-containing polypeptide, or an iron-containing protein), can be referred to as a "purified" polypeptide. A polypeptide of the disclosure can be purified from other components of the source material (e.g., other animal, plant, fungal, algal, or bacterial proteins). A purified polypeptide can refer to a polypeptide that has been enriched in a composition, has been manipulated in some fashion to remove unwanted debris (e.g., cell debris, genomic DNA, and/or other polypeptides), and/or is removed from the host cell in which it was synthesized (e.g., transcribed/translated) (e.g., cell lysis). A "purified" polypeptide can be a polypeptide extracted from its host cell. In some embodiments, a "purified" polypeptide is at least 1% pure, e.g., at least 2%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% pure. Proteins can be separated on the basis of their molecular weight, for example, by size exclusion chromatography, ultrafiltration through membranes, or density centrifugation. In some embodiments, the proteins can be separated based on their surface charge, for example, by isoelectric precipitation, anion exchange chromatography, or cation exchange chromatography. Proteins also can be separated on the basis of their solubility, for example, by ammonium sulfate precipitation, isoelectric precipitation, surfactants, detergents or solvent extraction. Proteins also can be separated by their affinity to another molecule, using, for example, hydrophobic interaction chromatography, reactive dyes, or hydroxyapatite. Affinity chromatography can also include using antibodies having specific binding affinity for the heme-containing polypeptide, antibody to the protein, nickel NTA for His-tagged recombinant proteins, lectins to bind to sugar moieties on a glycoprotein, or other molecules which specifically binds the protein.

**[0052]** Hemoglobin

**[0053]** Hemoglobin (Hb) can be the major constituent of an erythrocyte which can carry oxygen from the lungs throughout the body. When contained in red blood cells, human Hb can exist as a tetramer structure composed of two oxygen linked  $\alpha\beta$  dimers, each having a molecular weight of about 32 kD. Each  $\alpha$  and  $\beta$  subunit of each dimer can have a protein chain and a heme molecule. Hemoglobin, or "Hb" can refer to (a) an iron-containing respiratory pigment found in vertebrate red blood cells that comprises a globin composed of four subunits (a tetramer) each of which is linked to a heme molecule, that functions in oxygen transport to the tissues after conversion to oxygenated form in the gills or lungs, and that assists in carbon dioxide transport back to the gills or lungs after surrender of its oxygen. A hemoglobin can refer to a recombinantly produced hemoglobin;  $\alpha\beta$ -dimers of hemoglobin, inter- or intramolecularly crosslinked hemoglobin, as well as modified versions of the hemoglobins provided in the disclosure, which can include but are not limited to modifications increasing or decreasing the

oxygen affinity of hemoglobin (e.g., such as substituting an alanine, valine, leucine, or phenylalanine for histidine at position E7 (e.g., position 62 of SEQ ID NO: 4). See, for example, Hargrove et al., *J. Mol. Biol.* (1997) 266, 1032-1042. All hemoglobins can be capable of binding heme. A hemoglobin can be a variant hemoglobin. Variant hemoglobins can comprise amino acid mutations, substitutions, additions, and/or deletions. Hemoglobin variants can include hemoglobin Kansas, hemoglobin S, hemoglobin C, hemoglobin E, hemoglobin D-Punjab, hemoglobin O-Arab, hemoglobin G-Philadelphia, hemoglobin Hasharon, hemoglobin Lepore, and hemoglobin M.

**[0054]** Leghemoglobin

**[0055]** In some instances, the sequence (amino acid and/or nucleic acid) of a leghemoglobin can be a plant leghemoglobin sequence. Various legumes species and their varieties, for example, Soybean, Fava bean, Lima bean, Cowpeas, English peas, Yellow peas, Lupine, Kidney bean, Garbanzo beans, Peanut, Alfalfa, Vetch hay, Clover, Lespedeza and Pinto bean, comprise nitrogen-fixing root nodules in which leghemoglobin can have a key role in controlling oxygen concentrations. Leghemoglobins from different species can be homologs and have similar color properties. Some plant species can express several leghemoglobin isoforms (for example soybean has four leghemoglobin isoforms). Minor variations in precise amino acid sequence can modify overall charge of the protein at a particular pH and can modify precise structural conformation of an iron containing heme group in leghemoglobin. In some instances, an alanine, valine, leucine, or phenylalanine can be substituted for histidine at position 62 of SEQ ID NO: 4). Differences in structural conformation of the heme group of different leghemoglobins can influence oxidation and reduction rates of the heme iron. These differences may contribute to color and flavor generation properties of different leghemoglobins.

**[0056]** In other embodiments, the sequence (amino acid and/or nucleic acid) of a heme-containing polypeptide can be from a non-plant organism, such as from animals (e.g., a cow, pig, dog, rat, or horse), fish, archaea, protists, bacteria, fungus, eubacteria, metazoa, or yeast.

**[0057]** Variants

**[0058]** A polypeptide of the disclosure can be a variant (e.g., comprise a mutation such as an amino acid substitution, e.g., a non-conservative or conservative amino acid substitution, an amino acid deletion, an amino acid insertion, or non-native sequence). In some instances, a variant polypeptide can be a variant of a polypeptide listed in FIG. 9 (see, e.g., SEQ ID NOs: 1-31). In some instances, a variant polypeptide can include at most 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, or 50 mutations. In some instances, a variant polypeptide comprises at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, or 50 or more mutations. In some instances, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, or 50% of the sequence of a polypeptide of the disclosure can be mutated. In some instances, at most 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, or 50% of the sequence of a polypeptide of the disclosure can be mutated. In some instances, a polypeptide of the disclosure can comprise at least about 10, 20, 30, 40, 50, 60, 65, 70, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% amino acid sequence identity to a naturally occurring polypeptide of the disclosure. In some instances, a polypeptide of the disclosure can comprise at most about 10, 20, 30, 40, 50, 60, 65,

70, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% amino acid sequence identity to a naturally occurring polypeptide of the disclosure.

**[0059]** In some instances, the polypeptide of the disclosure comprises a non-native sequence (e.g., a tag or a label). A tag can be covalently bound to the polypeptide sequence of the polypeptide. The tag can be bound to the N-terminus, or the C-terminus, or to an intervening amino acid. The tag can be inserted in the polypeptide sequence (e.g., in a solvent accessible surface loop). Examples of tags can include, but are not limited to, affinity tags (e.g., myc, maltose binding protein, or 6xhis, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system), and fluorescent tags (e.g., green fluorescent protein).

**[0060]** A tag can be a composition detectable by spectroscopic, photochemical, biochemical, immunochemical, chemical, or other physical means. For example, tags suitable for use in the present disclosure can include biotin, digoxigenin, or haptens as well as proteins which can be made detectable, fluorescent dyes (e.g., fluorescein isothiocyanate, Texas red, rhodamine, and the like), radiolabels (e.g., <sup>3</sup>H, <sup>125</sup>I, <sup>35</sup>S, <sup>14</sup>C, or <sup>32</sup>P), dyes (e.g., alexa, cy3 cy5), chemical conjugates (e.g., quantum dots), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold, colored glass or plastic beads (e.g., polystyrene, polypropylene, latex, etc.).

**[0061]** A tag can be detected. For example, where the tag is radioactive, means for detection can include a scintillation counter or photographic film, as in autoradiography. Where the tag is a fluorescent tag, it may be detected by exciting the fluorochrome with the appropriate wavelength of light and detecting the resulting fluorescence. The fluorescence may be detected visually by the use of electronic detectors such as charge coupled devices (CCDs) or photomultipliers and the like. Similarly, enzymatic tags may be detected by providing the appropriate substrates for the enzyme and detecting the resulting reaction product. Colorimetric or chemiluminescent tags may be detected simply by observing the color associated with the tag.

**[0062]** In some instances, a tag can be a signal peptide. A signal peptide can be a peptide sequence usually present at the N-terminal end of newly synthesized secretory or membrane polypeptides which directs the polypeptide across or into a cell membrane of the cell (the plasma membrane in prokaryotes or the endoplasmic reticulum membrane in eukaryotes). It can be subsequently removed (e.g., by a protease). In particular, the signal peptide may be capable of directing the polypeptide into a cell's secretory pathway. In some instances, the signal peptide is a secretory pathway signal peptide. In some such embodiments, the signal peptide can be referred to as a signal peptide or a secretion signal peptide.

**[0063]** Examples of signal peptides can include, but are not limited to, the signal peptides listed in Table 1 (SEQ ID NO: 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, or 93). Nucleotides in parenthesis in Table 1 may or may not be included and therefore, the signal peptide may or may not have the residues encoded by the nucleotides. For example,

in some embodiments, the nucleic acid sequence set forth in SEQ ID NO:42 does not include the last 9 nucleotides (i.e., it would only contain nucleotides 1-84 of SEQ ID NO:42) and accordingly, the signal peptide set forth in SEQ ID NO:43 does not include the last three residues (i.e., the signal peptide only would contain amino acids 1-28 of SEQ ID NO:43). For example, in some embodiments, the nucleic acid sequence set forth in SEQ ID NO:44 does not include the last 6 nucleotides (i.e., it would only contain nucleotides 1-96 of SEQ ID NO:44) and accordingly, the signal peptide set forth in SEQ ID NO:45 does not include the last two residues (i.e., the signal peptide only would contain amino acids 1-32 of SEQ ID NO:45). For example, in some embodiments, the nucleic acid sequence set forth in SEQ ID NO:54 does not include the last 12 nucleotides (i.e., it would only contain nucleotides 1-156 of SEQ ID NO:54) and accordingly, the signal peptide set forth in SEQ ID NO:55 does not include the last four residues (i.e., the signal peptide only would contain amino acids 1-52 of SEQ ID NO:55). In some instances, a signal peptide can comprise PhoD (e.g., SEQ ID NO:55 or residues 1-52 of SEQ ID NO:55). Similarly, the signal peptides of SEQ ID NOs: 59, 63, 65, 67, 71, 73, 75, 77, 79, 83, 85, 87, or 93 may lack from one to four C-terminal amino acids.

**[0064]** In some instances, a signal peptide can be a variant signal peptide. A signal peptide can be a variant of the signal

peptides listed in Table 1 (SEQ ID NO: 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, or 93). In some instances, a variant signal peptide comprises at least about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% amino acid sequence identity to a signal peptide (e.g., a signal peptide listed in Table 1). For example, a variant signal peptide can have at least about 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% sequence identity to any one of the signal peptides set forth in Table 1. In some instances, a variant signal peptide can comprise at most about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% amino acid sequence identity to a signal peptide (e.g., a signal peptide listed in Table 1). In some instances, the cleavage site between the signal peptide and the polypeptide may be derived from the signal peptide. In some instances, the cleavage site may be a synthetic protease cleavage site. Cleavage of the signal peptide can result in a polypeptide of the disclosure comprising all, some, or none of the signal peptide. Cleavage of the synthetic protease cleavage site can result in a polypeptide of the disclosure comprising all, some, or none of the synthetic protease cleavage site.

TABLE 1

Exemplary Signal Peptides					
Gene	DNA sequence	SEQ ID NO	Peptide sequence (predicted cut site underlined)	SEQ ID NO:	Uniprot
AbnA	ATGAAAAGAAAAAACATG GAAACGCTTCTTACACTTTTC GAGTGCAGCTCTGGCTGCAG GTTTGATATTCCTTCTGCTG CTCCCGCAGAGGCAG	32	MKKKKTWKRFLHFSSAALAAGLIFTSAAP <u>AEAA</u>	33	P94522
AlbB	ATGTCACCAGCACAAAGAAG AATTTACTGTATATCCTTTTC ATTTATCTTTGTGCATCGGCCG AGTCGTCTATTTTGTCAAAG CGATTATCTGTTTACGCTGAT TTTCATTGCCATTGCCATTCT GTTCCG	34	MSPAQRILL YILSFIVIG VVVYFVKSDY <u>LFTLIFIAIA</u> ILF	35	P71010
AmyX	ATGGTCAGCATCCGCCGAGC TTCGAAGCGTATGTCGATGAC ATGAATATCATTACTGTTCTG ATTCTGCTGAACAAAAGGA AATCATG	36	MVSIRRSFEA YVDDMNIITV LIPAEQKEIM	37	COSPA0
AppB	ATGGCTGCATATATTATCAGA AGAACCCTTAATGCTATCCCG ATTTTATGGGAATTACGATT TTATCATTTGTTATCATGAAA GCCGCGCCCGGAGAT	38	MAAYIIRRTL MSIPILLGIT ILSFVIMKAA PGD	39	P42062
BglC	ATGAAACGGTCAATCTCTATT TTTATTACGTGTTTATTGATTA CGTTATTGACAAATGGCGGCA TGATAGCTTCGCCGCATCAG CA	40	MKRSISIFITCLLITLLTMGGMIASPASA	41	P42403
BglS	ATGCCTTATCTGAAACGAGTG TTGCTGCTTCTTGTCACTGGA TTGTTTATGAGTTTGTGTTGCA GTCACGTCTACTGCCTCAGCT (CAA ACA GGT)	42	MPYLKRVLLL LVTGLFMSLF AVTAT <u>ASA</u> QT G	43	P04957



TABLE 1-continued

Exemplary Signal Peptides					
Gene	DNA sequence	SEQ ID NO	Peptide sequence (predicted cut site underlined)	SEQ ID NO:	Uniprot
LipA	ATGAAATTTGTA AAAAGAAG GATCATTGCACCTTGTAACAAT TTTGATGCTGCTGTACATC GCTGTTTGCCTGACGCCGTC AGCAAAAGCCGCT (GAA CAC)	44	MKFVKRRIIA LVTILMLSVT SLFALQPSAK <u>AAEH</u>	45	O32129
LytD	ATGAAAAAGAGACTAATCGC ACCTATGCTTCTATCCGCCG GTCCCTTGCCCTTTTGGCCATG TCTGGTTCTGCCAGGCAGCC GCGTAT	46	MKKRLIAPML LSAASLAFFA MSGSA <u>QAAAY</u>	47	P39848
OppA	ATGAAAAACGTTGGTCGATT GTCACGTTGATGCTCATTTTC ACTCTCGTGCTGAGCGCGTGC GGCTTTGGC	48	MKKRWSIVTL MLIFTLVLSA CGFG	49	P24141
OppB	ATGCTAAATATATCGGAAG ACGCTTAGTCTATATGATTAT CACACTATTGTGATGTAAAC TGTGACATTCTCTTAATGCA AGCAGCACCGGCGGG	50	MLKYIGRRLV YMIITLFVIV TVTFFLMQAA PGG	51	P24138
PbpX	ATGACAAGCCCAACCCGACG AAGAAGTGCAGAACGCAGAC GGAGAAAATAAATAAAGA GGAAAACGTTGTTGGTCTT TTAGCAGTGATGGTTGCATT ACGATTTGGAATGCTCTTCAT CGA	52	MTSPTRRRTA KRRRRLNKR GKLLFGLLAV MVCIT <u>W</u> NAL HR	53	O31773
PhoD	ATGGCATAACGACAGTCGTTT GATGAATGGGTACAGAAACT GAAAGAGGAAAGCTTTCAAA ACAATACGTTTGACCCCGCA AATTTATTCAAGGAGCGGG AAGATTGCAGGACTTCTCTT GGATTAACGATGCCCAGTCG GTTGGGCGCTTT (GAA GTA AAT GCT)	54	MAYDSRFEWVQKLKEESFNNTFDRRK FIQAGKIAGLSLGLTIAQSVGAPEVNA	55	P42251
QcrA	ATGGGCGGAAAACATGATAT ATCCAGACGTCAATTTTGGAA TTATACGCTCACAGGCGTAGG AGGTTTTATGGCGCTAGTAT GCTCATGCCTATGGTTCGCTT CGCACTCGACCCG	56	MGGKHDISRR QFLNYTLTG GGFMAASMLM PMVRFALDP	57	P46911
SpoIIIJ	ATGTTGTTGAAAAGGAGAAT AGGGTTGCTATTAAGTATGGT TGCGTATTATGCTTTTGGC TGGA (TGC TCG AGT GTG)	58	MLLKRRIGLL LSMVGFVMLL AG CSSV	59	Q01625
TipA	ATGAAAAAACACTCACCAC TATTCGACAGATCAATTCG AAGGAGACTTATTTCTTT CCTGCTGATCTTAATTTGTTCC GATAACCGCCCTTTCGGTTAG CGCTTATCAATCAGCAGTTGC CTCA	60	mkktl <del>tt</del> ir <del>r</del> ssiar <del>ri</del> lis fllilivpit alsvsayqsa vas	61	
WapA	ATGAAAAAAGAAAGAGGCG AAACTTTAAAAGGTTCAATTGC AGCATTTTGTAGTTGGCTTT AATGATTTTATAGTGCAGC CGATGTACTAGCA (AAA TCT ACA)	62	MKKRKRNFK RFIAAFLVLA LMISLVPADV <u>LA</u> KST	63	Q07833
WprA	ATGAAACGCAGAAAATTCAG CTCGTTGTGGCGGCGAGTCT TATTTTTCAGTATTTTCAG	64	MKRRKFSSVV AAVLIFALIF SLFSPGTKAA <u>A</u> AGA	65	P54423

TABLE 1-continued

Exemplary Signal Peptides					
Gene	DNA sequence	SEQ ID NO	Peptide sequence (predicted cut site underlined)	SEQ ID NO:	Uniprot
	CCTTTTTTCTCCGGGAACCAA AGCTGCAGCG (GCC GGC GCG)				
YdeJ	ATGAAGAAAACGCAGAAAGAT ATGTTATTGCAATACTGCCCT GCTGCTTATGATTTTGCTTGC TGGA (TGT ACG GAC AGT)	66	MKKRRKICYCNTALLLMILL AG CTDS	67	P96667
YesM	ATGAAGAAAAGAGTTGCTGG CTGGTACAGGCGGATGAAGA TTAAGGATAAGCTGTTTGTGT TTCTATCGTTGATTATGGCCG TATCCTTTCTGTTTGTATACA GCGGGTCCAGTATGCCTTTC ATGTG	68	MKKRVAGWYR RMKIKDKLFV FLSLIMAVSF LFVYSGVQA FHV	69	Q31516
YesW	ATGAGAAGGAGCTGTCTGAT GATTAGACGAAGGAAACGCA TGTTTACCGCTGTTACGTTGC TGGTCTTGTGGTGATGGGAA CCTCTGTATGTCCTGTGAAAG CTGAAGGG (GCA)	70	MRRSCLMIRR RKRMF <del>AVTL</del> LVLLVMGTSV CPVKA <del>EG</del> A	71	Q31526
YflcN	ATGAGAATACAGAAAAGACG AACACACGTGAAAACATTCT CCGTATTCTTTTGCCCCAAT TATGATACTTAGCCTAATCCT CCCAACACCACCCATTCATGC A (GAA GAA AGC)	72	MRIQRRRTHV ENILRILLPP IMILSLILPT PPIHA EES	73	034313
YhcR	ATGCTGTCTGTCGAAATGATA AGCAGACAAAATCGTTGTCTAT TATGTGTATAAGGGAGGAAA TATGATGAGGCGTATCTGCA TATTGTGTTGATCAGGCATT AATGTTC <del>TAAATGTAATGTA</del> CACGTTCCGAAGCT (GTA AAG GCA)	74	MLSVEMISRQ NRCHYVYKGG NMMRRILHIV LITALMFLNV MYT <del>FEA</del> VKA	75	P54602
YkpC	ATGCTAAGAGATTTTAGGAAG AAGAGTAGCGATCGCAGCCA TTTTAAGCGAATTATTCTTG GAGGCATGAGCATTCTTTGG CA (AAT ATG CCC)	76	MLRDLGRRVA IAAILSGIIL GGMSI <del>SLA</del> NM P	77	Q45492
YkuE	ATGAAAAAGATGTCAGAAG ACAATTTCTAAAGGAATGTT CGGCGCTCTGCTGCCGGGGC TTTAAACGGCCGGCGGGGAT ATGGCTATGCC (AGG TAT CTC)	78	MKKMSRRQFL KGMFGALAAG ALTAGGGY <del>GY</del> A RYL	79	034870
YmaC	ATGAGAAGATTTTACTAAAT GTCATATTAGTCTTAGCCATT GTCTTGTTCTTGAGATATGTT CATTACTCATTGGAACCGAA	80	MRRFLLNVIL VLAIVLFLRY VHYSLEPE	81	031789
YmzC	ATGTTTGAAAGTGAAGCAGA ACTGAGACGAATCAGGATTG CACTGTATGGATAGCTGTCT TTTTACTGTTCCGGGGCG (TGC GGG AAT)	82	MFESAEALRRIRIALVWIAV FLLFGA CGN	83	031797
YolA	ATGAAGAAGAGAATTACATA TTCCTGCTGTCTTCTTAGC AGTTGTTGCTTTCGCTTTCCT GATTTCATCAAAGCAAAGC G (GCA GAA GCA)	84	MKKRITYSL ALLAVVAF <del>AF</del> TDSSK <del>AKA</del> AE A	85	031994

TABLE 1-continued

Exemplary Signal Peptides					
Gene	DNA sequence	SEQ ID NO	Peptide sequence (predicted cut site underlined)	SEQ ID NO:	Uniprot
YubF	ATGCAGAAATATAGACGCAG AAACACGGTTGCCCTTACAGT ACTAGCTTATTTTACTTTTTTTT GCGGGAGTATTTTGTTTAGT ATCGGACTCTATAATGCTGAT AATCTGGAAC	86	MQKYRRRNTV AFTVLAYFTF FAGVFLFSIG LYNADNLE	87	O32082
YuiC	ATGATGTTGAATATGATCAGA CGTTTGCTGATGACCTGTTTA TTCTGCTTGCATTTGGCACG ACATTTTATCAGTGCAGGA ATTGAAGCG (AAG GAC TTG)	88	MMLNMIRLL MTCLFLLAFG TTFLSVSGIE A KDL	89	O32108
YvhJ	ATGGCTGAACGCGTTAGAGT GCGTGTGCGAAAAAGAAAA AGAGCAAACGTAGGAAAATT TTAAAAAGAAATATGTTATTG TTCGCCCTTGCACTATTGGTA GTTGTAGGGCTTGGCGGTAT AAACTTTAT	90	MAERVVRVVR KKKKSKRRKI LKRIMLLFAL ALLVVVVLGG YKLY	91	P96499
YwbN	ATGAGCGATGAACAGAAAA GCCAGAACAAATTCACAGAC GGGACATTTTAAAATGGGGA GCGATGGCGGGGGCAGCCGT TGCGATCGGTGCCAGCGTCT CGGCGGTCTGCTCCGCTTGT TCAGACTGCGGCT (AAG CCA)	92	MSDEQKKPEQ IHRRDILKWG AMAGAAVAIG ASGLGGLAPL VQTAA KP	93	P39597

**[0065]** Protoporphyrins

**[0066]** A polypeptide can bind to a tetrapyrrole (e.g., protoporphyrin). A polypeptide can bind to a protoporphyrin with its protoporphyrin binding portion (e.g., domain). A polypeptide can bind to a protoporphyrin as the polypeptide is being translated/folded. A polypeptide can bind to a protoporphyrin after the polypeptide is translated/folded. A polypeptide can remain bound to a protoporphyrin after it has been subcellularly localized (e.g., localized to a subcellular compartment, secreted).

**[0067]** Protoporphyrins can comprise side chains including methyl groups, propionic acid groups, and vinyl groups. Suitable protoporphyrin structures can include, but are not limited to, diiododeuteroporphyrin, mesoporphyrin, metalloporphyrins, and protoporphyrin IX. In some instances, a polypeptide can bind to more than one protoporphyrin. A polypeptide can bind to one, two, three, four, five, six, seven, eight, nine, ten or more protoporphyrins.

**[0068]** A protoporphyrin can be a protoporphyrin IX. Protoporphyrin IX (PpIX), Pheophorbide, a naturally occurring photosensitizer, can be the immediate precursor of heme in the heme biosynthetic pathway. Protoporphyrin IX can be referred to as heme. Heme can comprise a protoporphyrin ring and an iron atom, wherein the iron atom is coordinated by the members of the ring (e.g., the iron atom is inside the ring). In some instances, the protoporphyrin can be heme A, heme B, heme C, heme D, heme I, heme M, heme O or Heme S. In some instances, a protoporphyrin can coordinate an atom other than iron (i.e., metalloporphyrin). Other atoms can include for example, zinc, gadolinium, magnesium, manganese, cobalt, nickel, tin, and copper.

**[0069]** Vectors and Genetically Modified Organisms**[0070]** Exogenous Nucleic Acids

**[0071]** The disclosure can provide for an exogenous nucleic acid encoding a polypeptide of the disclosure (e.g., a heme-containing polypeptide, a globin). An exogenous nucleic acid can encode any of the heme-containing polypeptides described herein, e.g., a heme-containing polypeptide having at least about 60% identity (e.g., at least 65%, 75%, 80%, 85%, 90%, 95%, 99% identity) to one of the amino acid sequences listed in FIG. 9). An exogenous nucleic acid can be RNA or DNA, and can be single stranded, double stranded, and/or codon optimized. An exogenous nucleic acid sequence encoding a polypeptide of the disclosure can be transcribed and/or translated. The term "polynucleotide" can be used interchangeably herein with "exogenous nucleic acid."

**[0072]** The term "exogenous" as used herein with reference to a nucleic acid (or a protein) and a host refers to a nucleic acid that does not occur in (and cannot be obtained from) a cell of that particular type as it is found in nature or a protein encoded by such a nucleic acid. Thus, a non-naturally-occurring nucleic acid is considered to be exogenous to a host once in the host. It is important to note that non-naturally-occurring nucleic acids can contain nucleic acid subsequences or fragments of nucleic acid sequences that are found in nature provided the nucleic acid as a whole does not exist in nature. For example, a nucleic acid molecule containing a cDNA sequence within an expression vector is non-naturally-occurring nucleic acid, and thus is exogenous to a host cell once introduced into the host, since that nucleic acid molecule as a whole (cDNA plus vector DNA) does not exist in nature. Thus, any vector, autonomously replicating plasmid, or virus (e.g., retrovirus, adeno-

virus, or herpes virus) that as a whole does not exist in nature is considered to be non-naturally-occurring nucleic acid. It follows that genomic DNA fragments produced by PCR or restriction endonuclease treatment as well as cDNAs are considered to be non-naturally-occurring nucleic acid since they exist as separate molecules not found in nature. It also follows that any nucleic acid containing a regulatory element (e.g., a promoter sequence and/or a signal sequence) and a sequence encoding a heme-containing polypeptide (e.g., cDNA or genomic DNA) in an arrangement not found in nature is non-naturally-occurring nucleic acid. A nucleic acid that is naturally-occurring can be exogenous to a particular host (e.g., bacteria or plant). For example, an entire chromosome isolated from a cell of plant x is an exogenous nucleic acid with respect to a cell of plant y once that chromosome is introduced into a cell of plant y.

**[0073]** In contrast, the term “endogenous” as used herein with reference to a nucleic acid (e.g., a gene) (or a protein) and a host refers to a nucleic acid (or protein) that does occur in (and can be obtained from) that particular host as it is found in nature. Moreover, a cell “endogenously expressing” a nucleic acid (or protein) expresses that nucleic acid (or protein) as does a host of the same particular type as it is found in nature. Moreover, a host “endogenously producing” or that “endogenously produces” a nucleic acid, protein, or other compound produces that nucleic acid, protein, or compound as does a host of the same particular type as it is found in nature.

**[0074]** The degeneracy of the genetic code can permit variations of the nucleotide sequence, while still producing a polypeptide having the identical amino acid sequence as the polypeptide encoded by the native polynucleotide sequence. Variations in the polynucleotide sequence can be customized for any organism of interest. In some instances, a polynucleotide encoding a polypeptide can be codon optimized for expression in a bacteria (e.g., a gram positive bacteria such as *B. subtilis*). In some instances, a polynucleotide encoding a polypeptide can be codon optimized for expression in a plant (e.g., *N. tabacum*).

**[0075]** The frequency of individual synonymous codons for cognate amino acids varies widely from genome to genome among eukaryotes and prokaryotes. These differences in codon choice patterns can contribute to the overall expression levels of individual genes by modulating peptide elongation rates.

**[0076]** Bacterial Vectors

**[0077]** As described herein, an exogenous nucleic acid can include a nucleic acid encoding a signal peptide and a nucleic acid encoding a heme-containing polypeptide. In some instances, the exogenous nucleic acid is a vector. A vector can be suitable for expression in a prokaryote (e.g., bacteria).

**[0078]** The vectors of the present disclosure generally comprise regulatory control sequences, (e.g., transcriptional or translational control sequences) required for expressing the polypeptide. Suitable regulatory control sequences can include but are not limited to replication origin, promoter, enhancer, repressor binding regions, transcription initiation sites, ribosome binding sites, translation initiation sites, or termination sites for transcription and translation.

**[0079]** In some embodiments, the vector can comprise a polynucleotide sequence encoding two or more polypeptides (e.g., heme-containing polypeptides or heme biosynthesis pathway enzymes), which can be present on the same vector.

In some embodiments, when present on the same vector, the polynucleotides are arranged such that they form an operon, (i.e., transcription of the polynucleotides will generate a polycistronic messenger RNA). In some instances, the two or more polynucleotides can be arranged on the same vector such that they are operably linked to their own promoter.

**[0080]** The origin of replication (generally referred to as an ori sequence) can permit replication of the vector in a host cell. The choice of ori can depend on the type of host cells that are employed. Where the host cells are prokaryotes, the expression vector can comprise an ori directing autonomous replication of the vector within the prokaryotic cells. Non-limiting examples of this class of ori include pMB1, pUC, ColE1 as well as other bacterial origins.

**[0081]** The host cell can comprise a polynucleotide encoding a polypeptide of the disclosure for recombinant production, wherein the polypeptide can be linked to a functional signal sequence (e.g., a sequence encoding a signal peptide). Sequences encoding signal peptides can include, for example, those derived from spA, phoA, ribose binding protein, pelB, ompA, ompT, dsbA, torA, torT, and tolT, the signal peptides listed in Table 1, or signal peptides from the TAT secretion pathway in bacteria. Also included within the scope of the disclosure are signal sequences derived from eukaryotic cells that also function as signal sequences in prokaryotic host cells.

**[0082]** The vectors may comprise a selectable marker such as a gene encoding a protein necessary for the survival or growth of a host cell transformed with the vector. A marker gene can be carried on another polynucleotide sequence co-introduced into the host cell. Only those host cells into which a selectable gene has been introduced can survive and/or grow under selective conditions. Typical selection genes can encode protein(s) that (a) confer resistance to antibiotics or other toxins, (e.g., ampicillin, kanamycin, neomycin, G418, methotrexate, etc.); (b) complement auxotrophic deficiencies; or (c) supply critical nutrients not available from complex media.

**[0083]** The vector may comprise a polynucleotide sequence encoding a tag. A tag sequence can be in-frame to the coding sequence of the polypeptide of the disclosure, such that upon translation of the sequence, the polypeptide of the disclosure is covalently bound, e.g., fused, to the tag (e.g., is a fusion protein). The tag can be separated from the polypeptide of the disclosure by a linker, e.g., a polypeptide linker. A vector can comprise a polynucleotide sequence encoding a linker between the sequence encoding the tag and the sequence encoding the polypeptide of the disclosure.

**[0084]** The vectors encompassed by the disclosure can be obtained using recombinant cloning methods and/or by chemical synthesis. Recombinant cloning techniques can include PCR, restriction endonuclease digestion and ligation. Sequence data, that can be located in the public or proprietary databases, can be used to obtain a desired vector by any synthetic means. Additionally, using restriction and ligation techniques, appropriate sequences can be excised from various DNA sources and integrated in operative relationship with the exogenous sequences to be expressed in accordance with the present disclosure.

**[0085]** In some instances, the vector can comprise a regulatory element (also referred to as a regulatory control element herein). A regulatory element can include, for example, replication origin, promoters, TATA boxes, enhancers, ribosome binding sites, repressor binding

regions, transcription initiation sites, transcription termination sites, non-native sequences (e.g., tags) and untranslated regions. In some instances, the regulatory element can be a promoter. A promoter can be constitutive, inducible, and/or tissue specific. Exemplary promoters can include both constitutive promoters and inducible promoters. A natural promoter can be modified by replacement, substitution, addition or elimination of one or more nucleotides without changing its function.

**[0086]** In some embodiments, in addition to a promoter sequence, the polynucleotide sequence also includes a transcription termination region downstream of the nucleic acid sequence encoding the polypeptide to provide for efficient termination. In some embodiments, the termination region can be obtained from the same gene as the promoter sequence, while in other embodiments it can be obtained from another gene.

**[0087]** In some instances, the vector can comprise a polynucleotide sequence encoding a polypeptide of the disclosure and a tag. In some instances, the tag can be a signal peptide.

**[0088]** In some embodiments, once the desired form of a polypeptide, nucleic acid sequence, homologue, variant or fragment thereof, is obtained, it can be modified by any number of ways. Where the sequence involves non-coding flanking regions, the flanking regions may be subjected to resection, mutagenesis, etc. Thus, transitions, transversions, deletions, and insertions may be performed on the naturally occurring sequence.

**[0089]** In some preferred embodiments, a polynucleotide encoding a polypeptide can include the coding sequence for at least one polypeptide (e.g., globin), or variant(s), fragment(s) or splice variant(s) thereof: (i) in isolation; (ii) in combination with additional coding sequences; such as fusion protein or signal peptide coding sequences, where the polypeptide-coding sequence is the dominant coding sequence; and/or (iii) in combination with non-coding sequences, such as control elements, such as promoter and terminator elements or 5' and/or 3' untranslated regions, effective for expression of the coding sequence in a suitable host.

**[0090]** In some embodiments, a polynucleotide encoding a polypeptide, together with appropriate promoter and control sequences, can be introduced into bacterial host cells to permit the cells to express at least one polypeptide (e.g., globin) or variant thereof.

**[0091]** Natural or synthetic polynucleotide fragments encoding a polypeptide (e.g., a heme-containing polypeptide, a globin) may be incorporated into vectors, capable of introduction into, and replication in, a bacterial cell. Any vector may be used as long as it is replicable and viable in the cells into which it is introduced. The appropriate DNA sequence can be inserted into a plasmid or vector by any suitable method. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by recombinant molecular biology techniques.

**[0092]** In some instances, the disclosure can provide for a genetically modified organism. In some instances, a genetically modified organism can comprise a polypeptide of the disclosure. A genetically modified organism can comprise a polynucleotide encoding a heme-containing polypeptide.

**[0093]** Plant Vectors

**[0094]** The present disclosure provides for vectors for introduction of exogenous nucleic acid in a method accord-

ing to the disclosure which can find use in the expression of a nucleic acid in a plant cell, specific plant tissues such as the leaf, the root, the seed or the bean, or a specific compartment of a plant cell.

**[0095]** Transgenic plant cells and plants are provided herein comprising at least one exogenous nucleic acid. As described herein, an exogenous nucleic acid can include a nucleic acid encoding a signal peptide and a nucleic acid encoding a heme-containing polypeptide. A plant or plant cell can be transformed by having a construct integrated into its genome, i.e., can be stably transformed. Stably transformed cells typically retain the introduced nucleic acid with each cell division. A plant or plant cell also can be transiently transformed such that the construct is not integrated into its genome. Transiently transformed cells typically lose all or some portion of the introduced nucleic acid construct with each cell division such that the introduced nucleic acid cannot be detected in daughter cells after a sufficient number of cell divisions. Both transiently transformed and stably transformed transgenic plants and plant cells can be useful in the methods described herein.

**[0096]** Typically, transgenic plant cells used in methods described herein constitute part or all of a whole plant. Such plants can be grown in a manner suitable for the species under consideration, either in a growth chamber, a greenhouse, or in a field. Transgenic plants can be bred as desired for a particular purpose, e.g., to introduce a recombinant nucleic acid into other lines, to transfer a recombinant nucleic acid to other species or produce the desired polypeptide. Alternatively, transgenic plants can be propagated vegetatively for those species amenable to such techniques. Progeny includes descendants of a particular plant or plant line provided the progeny inherits the transgene. Progeny of an instant plant include seeds formed on  $F_1$ ,  $F_2$ ,  $F_3$ ,  $F_4$ ,  $F_5$ ,  $F_6$  and subsequent generation plants, or seeds formed on  $BC_1$ ,  $BC_2$ ,  $BC_3$ , and subsequent generation plants, or seeds formed on  $F_1BC_1$ ,  $F_1BC_2$ ,  $F_1BC_3$ , and subsequent generation plants. Seeds produced by a transgenic plant can be grown and then selfed (or outcrossed and selfed) to obtain seeds homozygous for the exogenous nucleic acid.

**[0097]** Transgenic plant cells growing in suspension culture, or tissue or organ culture, can be useful for extraction of polypeptides. Solid and/or liquid tissue culture techniques can be used. When using solid medium, transgenic plant cells can be placed directly onto the medium or can be placed onto a filter film that is then placed in contact with the medium. When using liquid medium, transgenic plant cells can be placed onto a floatation device, e.g., a porous membrane that contacts the liquid medium. Solid medium typically is made from liquid medium by adding agar. For example, a solid medium can be Murashige and Skoog (MS) medium containing agar and a suitable concentration of an auxin, e.g., 2,4-dichlorophenoxyacetic acid (2,4-D), and a suitable concentration of a cytokinin, e.g., kinetin.

**[0098]** In some embodiments, the disclosure provides for a method wherein *Agrobacterium* cells comprise a vector comprising sequence elements, which are essential for maintenance and replication of the plasmid in *Escherichia coli* and/or *Agrobacterium* cells, and for the transfer of a T-DNA to a plant cell, and further a T-DNA region, comprising the coding sequence of a polypeptide that is under control of regulatory elements functional in a plant and, optionally, a plant selectable marker gene.

**[0099]** In any of the transformation methods, the vector can include a plant selectable marker such as an antibiotic-based selectable marker (e.g., spectinomycin, kanamycin, streptomycin). A plant selectable marker can be a fluorescent protein marker and/or a colorimetric marker (e.g., LacZ). In some instances, a plant selectable marker can comprise peptide deformylase. Peptide deformylase can hydrolyze the N-formyl group on an initiating methionine. Peptide deformylase activity can be essential for cell viability. A peptide deformylase can originate from a number of genes including DEF1 and DEF2. Plants expressing a peptide deformylase selectable marker can be resistant to peptide deformylase inhibitors (e.g., actinonin). Selectable marker genes can be excised. Excision can occur by site-specific recombination (e.g., Cre, Int recombination), transposons (e.g., Ac/Ds family transposons), meganucleases (e.g., homing endonucleases, I-sceI), intrachromosomal homologous recombination, the Cas9/CRISPR endonucleases, and/or zinc fingers. See, for example, the below discussion re homologous recombination.

**[0100]** In some embodiments, site specific breakage in the plant genome is utilized either to greatly enhance targeted homologous-recombination-based mutation/replacement of endogenous sequences (i.e., to reprogram a globin gene) or to greatly enhance mutation or recombination rates at specific sites (e.g., promoter of globin genes or promoter of globin genes/promoter of a highly endosperm-expressed gene to retarget the globin expression to seeds or other targeted tissues). Site specific breakage can occur by TAL-ENS (plant-transcription factor derived endonucleases that exploit a simple system for designing DNA-recognition elements to create synthetic endonucleases with sufficiently high specificity to cleave a single genomic site *in vivo*) or the Cas9/Crispr system.

**[0101]** In some instances, a vector can comprise one or more of the following nucleic acid elements: a) a first nucleic acid element comprising a nucleotide sequence encoding a selectable marker (e.g., which can be functional in *Escherichia coli* and/or *Agrobacterium* species); b) a second nucleic acid element comprising a nucleotide sequence of a first origin of replication which can be functional in *Escherichia coli*; c) a third nucleic acid element comprising a nucleotide sequence encoding a replication initiator protein; and/or d) a fourth nucleic acid element comprising a nucleotide sequence of a second origin of replication, which can be different from the first origin of replication and which is functional in *Agrobacterium*, wherein the above nucleic acid elements are provided on a circular polynucleotide molecule and are separated by gap nucleotide sequences which have no function in replication, maintenance or nucleic acid transfer, and wherein said gap nucleotide sequences account for less than 20%, 25%, 30%, 35%, 40%, or 45% of the total vector size.

**[0102]** In another embodiment, the disclosure relates to a method, wherein the regulatory sequences operable in a plant or a plant cell include a promoter that can drive and/or control expression of a gene of interest. Suitable promoters can include mirabilis mosaic virus (MMV), figwort mosaic virus (FMV) or Peanut Chlorotic Streak Caulimovirus (PCLSV) promoters. Other examples of suitable promoters can include a cauliflower mosaic virus 35S promoter, a modified cauliflower mosaic virus 35S promoter, a double cauliflower mosaic virus 35S promoter, a minimal 35 S promoter, nopaline synthase promoter, a cowpea mosaic

virus promoter, a HT-CPMV promoter, a tobacco copalyl synthase CPS2p promoter, a dihydrinin promoter, a plastocyanin promoter, a 35S/HT-CPMV promoter, full length transcript (FLt) promoters, sub-genomic transcript promoters, and many other promoters that are derived from DNA viruses belonging to the Caulimoviridae virus family.

**[0103]** Many such promoters can be modified by linking multiple copies, for example two copies, of their enhancer sequence in tandem to enhance the promoter activity, such as but not limited to double CaMV 35S promoter (35S $\times$ 2), double MMV promoter (MMV $\times$ 2), or double FMV promoter (FMV $\times$ 2). Functional fragments of these promoters can be used in the vector of the disclosure. Nucleotide sequences that are at least 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to these promoter sequences and that are functional in enabling expression in plants of the operably linked nucleotide sequence can also be used in the vectors of the disclosure.

**[0104]** Plant expression vectors which can be functional in a plant cell and may be used within the method of the present disclosure in order to drive and/or control expression of a gene of interest in a plant may also comprise, if desired, a promoter regulatory region (for example, one conferring inducible or constitutive, environmentally- or developmentally-regulated, or cell- or tissue-specific expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal. The regulatory elements to be used within the methods of the disclosure may be present in a vector molecule (e.g., binary vector) operably linked to a nucleotide sequence encoding a polypeptide of the disclosure. In some embodiments, a regulatory element can be present in the T-DNA region of a binary vector (e.g., a minimally sized binary vector).

**[0105]** In some embodiments the promoters controlling gene expression do so in a tissue-dependent manner and according to the developmental stage of the plant. The transgene sequences of the disclosure driven by these type of promoters can be expressed in tissues where the transgene product is desired, leaving the rest of the tissues in the plant unmodified by transgene expression. Tissue-specific promoters may be induced by endogenous or exogenous factors. Tissue specific examples can include but are not limited to beta-amylase gene or barley hordein gene promoters (for seed gene expression), tomato pZ7 and pZ130 gene promoters (for ovary gene expression), tobacco RD2 gene promoter (for root gene expression), banana TRX promoter and melon actin promoter (for fruit gene expression). In some embodiments, the tissue specific promoters can be chosen to target expression of the polypeptide to bulky and easily harvestable parts of the plant such as the seeds, fruits, tubers and leaves.

**[0106]** Plant expression vectors may further comprise a nucleotide sequence encoding a signal peptide that can target the newly expressed protein to a subcellular location. Signal peptides that may be used within such vector molecules can be, for example, a vacuolar targeting sequence, a chloroplast targeting sequence, a mitochondrial targeting sequence, a sequence that induces the formation of protein bodies in a plant cell, a sequence that induces specific targeting of the protein fused there onto to a specific organelle within the plant or plant cell, or a sequence that induces the formation of oil bodies in a plant cell.

[0107] In some embodiments, the targeting sequence can be a signal peptide for export of a protein into the extracellular space. Signal peptides can be transit peptides that are located at the extreme N-terminus of a protein and cleaved co-translationally during translocation across a plasma membrane.

[0108] In some embodiments, the signal peptide may be a sequence that when fused to a protein results in the formation of non-secretory storage organelles in the endoplasmic reticulum.

[0109] Endogenous nucleic acids can be modified by homologous recombination techniques. For example, sequence specific endonucleases (e.g., zinc finger nucleases (ZFNs)) and meganucleases can be used to stimulate homologous recombination at endogenous plant genes. See, e.g., Townsend et al., *Nature* 459:442-445 (2009); Tovkach et al., *Plant J.*, 57:747-757 25 (2009); and Lloyd et al., *Proc. Natl. Acad. Sci. USA*, 102:2232-2237 (2005). CRISPR Cas (Xie and Yang, *Mol. Plant* 2013, 6:1975-1983) and TALEN (Zhang et al., *Plant Physiology* 2013 161:20-27) genome editing techniques also can be used to replace an endogenous nucleic acid.

[0110] In some embodiments, the vector can further comprise in the T-DNA region a site-specific recombination site for site-specific recombination. In some embodiments, the site-specific recombination site can be located downstream of the plant regulatory element. In some embodiments, the site-specific recombination site can be located upstream of the plant regulatory element. In some embodiments, the recombination site can be a LoxP site and part of a Cre-Lox site-specific recombination system. The Cre-Lox site-specific recombination system can use a cyclic recombinase (Cre), which can catalyze the recombination between specific sites (LoxP) that comprise specific binding sites for Cre.

[0111] In some embodiments, the recombination site can be a Gateway destination site. For example, polynucleotides can be cloned into a commercially available "entry vector" and subsequently recombined into a "destination vector". The destination vector can be used for the analysis of promoter activity of a given nucleic acid sequence or number of sequences, for analysis of function, for protein localization, for protein-protein interaction, for silencing of a given gene or for affinity purification experiments. The Gateway cloning technology can be purchased from Invitrogen Inc., USA.

[0112] In some embodiments targeted lesions are created by homologous recombination or other gene editing techniques in the vicinity of an endogenous nucleic acid encoding a hemoprotein or globin and in the vicinity of desired tissue specific promoters and mutants are screened for expression of the desired endogenous hemoprotein or globin in abundant, accessible tissues. In some embodiments, targeted lesions are created in the vicinity of desired tissue specific promoters and mutants are screened for the expression of the desired endogenous hemoprotein or globin in abundant, accessible tissues. Examples of methods for creating these targeted lesions include but are not limited to TALENS and the cas9/crispr system as discussed above.

[0113] In some embodiments, the resultant plants may contain more target protein than the original plant. For example, the resultant plant may express more than 2x the level of the target native protein compared to the original plant. In some embodiments, the resultant plant may express

the native target protein in a tissue, such as the seed or leaf, in which it is not normally found. The native target protein may be expressed in a tissue that it is not normally found at 1 fold, 2 fold, 3 fold, 4, fold, 5 fold, 6 fold, 7 fold, 8 fold, 9 fold, 10 fold or more higher or lower than the levels of the native target protein in the tissue in which it is normally found. In some embodiments the target protein is a hemoglobin such as the leghemoglobin of *Glycine max* (e.g., SEQ ID NO:4) or the nonsymbiotic hemoglobin of barley (SEQ ID NO:5) produced in engineered soy or barley plants respectively.

[0114] In some embodiments the resultant plant contains no foreign DNA. In some instances, lesions in the vicinity of the endogenous polypeptide of the disclosure and/or tissue specific promoters can be used for the engineering the expression of the polypeptide of the disclosure. Methods for assessing this type of engineering can include screening for production of the endogenous polypeptide.

[0115] Methods of Expression in Bacteria

[0116] The disclosure can provide for methods for expression of a polypeptide (e.g., globin) in a host cell (e.g., bacteria). Suitable bacteria for expression of a polypeptide of the disclosure can be gram negative or gram positive bacteria. For example, the bacteria can be a species of *Escherichia* (e.g., *E. coli*) or a species of *Bacillus* (e.g., *B. subtilis*, *B. licheniformis*, *B. lentus*, *B. brevis*, *B. stearothermophilus*, *B. alkalophilus*, *B. amyloliquefaciens*, *B. clausii*, *B. coagulans*, *B. circulans*, *B. lautus*, *B. megaterium*, or *B. thuringiensis*). In some instances, the bacteria suitable for expression of a polypeptide of the disclosure can be *B. subtilis*.

[0117] In some instances, expression of a polypeptide can include introducing a vector comprising a polynucleotide sequence encoding the polypeptide into the host cell, and inducing expression of the polypeptide.

[0118] In some embodiments, the methods of the disclosure provide for a host cell that comprises a stably integrated sequence of interest (i.e., polypeptide-encoding nucleic acid). However, in alternative embodiments, the methods of the present disclosure provide for maintenance of a self-replicating extra-chromosomal transformation vector.

[0119] Methods of introducing the polynucleotide into cells for expression of the polynucleotide sequence can include, but are not limited to electroporation, transformation, transduction, high velocity bombardment with DNA-coated microprojectiles, infection with modified viral (e.g., phage) nucleic acids; chemically-mediated transformation, or competence. In some embodiments, polynucleotides encoding a polypeptide of the disclosure can be transcribed in vitro, and the resulting RNA can be introduced into the host cell.

[0120] Following introduction of a polynucleotide comprising the coding sequence for a polypeptide of the disclosure, the host cell can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, and/or amplifying expression of a polypeptide-encoding polynucleotide. The culture conditions, such as temperature, pH and the like, can be those previously used for the host cell selected for expression. The progeny of cells into which such polynucleotide constructs have been introduced can be considered to comprise the polypeptide-encoding polynucleotide.

[0121] In some embodiments, the polypeptide or variant thereof can be expressed as a fusion protein by the host

bacterial cell. Although cleavage of the fusion polypeptide to release the desired protein can often be useful, it is not necessary. Polypeptides and variants thereof expressed and secreted as fusion proteins can retain their function.

**[0122]** Expression of a polypeptide of the disclosure can comprise transient expression and/or constitutive expression (e.g., developing of a stable cell line).

**[0123]** Expression of Recombinant Polypeptides

**[0124]** Expression of a polypeptide can comprise inducing the host cell to transcribe and/or translate the polypeptide encoded in the polynucleotide introduced to the host cell. Induction can occur after the host cell has been cultured for at least about 0.1, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 or 20 or more hours. Induction can occur after the host cell culture has an optical density (OD) of at least about 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, or 1.5 or more. Induction can occur after the host cell culture has an optical density (OD) of at most about 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 2, 3, 5, 10, or 20 or more. Induction may be caused by addition of chemicals such as IPTG, arabinose or in response to a limiting nutrient such as Nitrogen, phosphorus, glucose or oxygen. In some instances, the polypeptide is linked to a promoter, such as *aprE*, *liaG*, *lepA*, *cry3Aa*, or *gsiB* that leads to constitutive expression of the polypeptide.

**[0125]** In some instances, chemical agents can be added to the media. In some instances, the chemical agents can aid the stability, heme content, and/or protein folding capability of the expressed polypeptide. A chemical agent can comprise a small molecule such as a metal. Examples of suitable metals for addition to media can include iron fluorides (iron difluoride, iron trifluoride), iron dichloride, iron trichloride, iron dibromide, iron tribromide, iron diiodide, iron triiodide, iron oxide, diiron trioxide, tri-iron tetraoxide, iron sulfide, iron persulfide, iron selenide, iron telluride, di-iron nitride, iron pentacarbonyl, diiron nonacarbonyl, triiron dodecacarbonyl, iron dichloride dihydrate, iron trifluoride trihydrate, iron dibromide hexahydrate, iron dichloride tetrahydrate, iron nitrate hexahydrate, iron trichloride hexahydrate, iron difluoride tetrahydrate, iron sulphate heptahydrate, iron trinitrate nonahydrate, diiron trisulfate nonahydrate, iron chromate, iron citrate, iron gluconate, magnesium iron hexahydride, iron lactate, iron phosphate, iron pentacarbonyl, ammonium iron sulfate, ammonium ferric citrate, ferric oxalate, and triiron diphosphate octahydrate.

**[0126]** A chemical agent can be added to the media at a final concentration of at least about 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 5.0, 10.0 or more millimolar. A chemical agent can be added to the media at a final concentration of at most about 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 5.0, 10.0 or more millimolar.

**[0127]** In some instances, a chemical agent can be a heme derivative. A heme derivative can increase the heme content of the expressed polypeptide (e.g., increase the number of globin molecules that comprise a heme). Suitable heme derivatives can include delta-aminolevulinic acid, derivatives of heme A, derivatives of heme B, derivatives of heme C, derivatives of heme O, heme precursors, derivatives of heme I, derivatives of heme m, derivatives of heme D, and derivatives of heme S. A heme derivative can be added to the media at a final concentration of at least about 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 5.0, 10.0 or more millimolar. A heme derivative

can be added to the media at a final concentration of at most about 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 5.0, 10.0 or more millimolar. In some instances, no heme derivative is added to the media.

**[0128]** After inducing the polypeptide, the host cell can be cultured for a period of time favoring maximal expression levels of the polypeptide. For example, a polypeptide can be expressed for at least about 0.1 hours, 0.5 hours, 1 hour, 2 hours, 3 hours, 4 hours, 5 hours, 6 hours, 7 hours, 8 hours, 9 hours, 10 hours, 11 hours, 12 hours, 13 hours, 14 hours, 15 hours, 16 hours, 17 hours, 18 hours, 19 hours, 20 hours, 21 hours, 22 hours, 23 hours, 24 hours, 1 days, 2 days, 3 days, 4 days, 5 days, 6 days, 7 days, 1 week, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 7 weeks, 8 weeks, 9 weeks, 10 or more weeks. A polypeptide can be expressed in a host cell for at most about at least about 0.1 hours, 0.5 hours, 1 hour, 2 hours, 3 hours, 4 hours, 5 hours, 6 hours, 7 hours, 8 hours, 9 hours, 10 hours, 11 hours, 12 hours, 13 hours, 14 hours, 15 hours, 16 hours, 17 hours, 18 hours, 19 hours, 20 hours, 21 hours, 22 hours, 23 hours, 24 hours, 1 days, 2 days, 3 days, 4 days, 5 days, 6 days, 7 days, 1 week, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 7 weeks, 8 weeks, 9 weeks, or 10 or more weeks.

**[0129]** A polypeptide can be expressed at a variety of temperatures. A polypeptide can be expressed at a temperature of at least about 4, 10, 16, 18, 21, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 42, 42, 43, 44, 45, 46, 47, 48, 49, or 50° C. A polypeptide can be expressed at a temperature of at most about 4, 10, 16, 18, 21, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 42, 42, 43, 44, 45, 46, 47, 48, 49, or 50° C.

**[0130]** Accessory proteins such as thiol-disulfide oxidoreductases or chaperones may be beneficial to help fold the polypeptide into its active conformation. Thiol-disulfide oxidoreductases and protein disulfide isomerases can catalyze the formation of the correct disulfide bonds in the protein. Expression of the *bdbDC* operon in *B. subtilis* has been shown to be beneficial for the production of a polypeptide with disulfide bonds. Chaperones can help the secretory protein to fold by binding to exposed hydrophobic regions in the unfolded states and preventing unfavorable interactions and prolyl-peptidyl cis-trans isomerases assist in formation of the proper conformation of the peptide chain adjacent to proline residues. In some embodiments, the host cells can be transformed with an expression vector encoding at least one thiol-disulfide oxidoreductase or chaperone.

**[0131]** In some embodiments, the fraction of properly folded polypeptide can be increased by the addition of chemicals to the growth medium that reduce/oxidize disulfide bonds, and/or alter the general redox potential, and/or chemicals that alter solvent properties thus affecting protein conformation and aggregation. In some embodiments, a reagent that reduces disulfide bonds, such as 2-mercaptoethanol, can increase the fraction of correctly folded protein. In some embodiments and depending on the medium used, other disulfide reducing or oxidizing agents (e.g., DTT, TCEP, reduced and oxidized glutathione, cysteine, cystine, cysteamine, thioglycolate,  $S_2O_3^{2-}$ ,  $S_2O_4^{2-}$ ,  $S_2O_5^{2-}$ ,  $SO_3^{2-}$ ,  $S_2O_7^{2-}$ ,  $Cu^+$ , etc.), either used alone or in combination, can find use in the present disclosure. It can be contemplated that other adjuvants that alter solvent properties, (e.g., urea, DMSO, TWEEN®-80, etc.), either added to the growth medium alone or preferably in combination with disulfide



reducing/oxidizing agents, such as  $\beta$ ME, can also increase the fraction of correctly folded secretory protein and find use in various embodiments of the present disclosure. In some embodiments, the  $\beta$ ME can be used at concentrations ranging from 0.5 to 4 mM, or from about 0.1 mM to 10 mM.

**[0132]** The polypeptide can be recovered from the culture (e.g., by centrifugation, purification, etc.), as described below and herein.

**[0133]** Fermentation Parameters

**[0134]** The present disclosure provides for fermentation procedures for culturing bacterial species. Culturing can be accomplished in a growth medium comprising an aqueous mineral salts medium, organic growth factors, the carbon and energy source material, molecular oxygen (for aerobic and facultative bacteria), and, of course, a starting inoculum of one or more particular microorganism species to be employed.

**[0135]** In addition to the carbon and energy source, oxygen, assimilable nitrogen, and an inoculum of the microorganism, it can be necessary to supply suitable amounts in proper proportions of mineral nutrients to assure proper microorganism growth, maximize the assimilation of the carbon and energy source by the cells in the microbial conversion process, and achieve maximum cellular yields with maximum cell density in the fermentation medium.

**[0136]** Various culture media can be used. Standard bacterial culture media can be used. The media can include, in addition to nitrogen, suitable amounts of phosphorus, magnesium, calcium, potassium, sulfur, and sodium, in suitable soluble assimilable ionic and combined forms, and can also comprise certain trace elements such as copper, manganese, molybdenum, zinc, iron, boron, and iodine, and others, again in suitable soluble assimilable form.

**[0137]** In some embodiments, the fermentation reaction can involve an aerobic process in which the molecular oxygen needed can be supplied by a molecular oxygen-containing gas such as air, oxygen-enriched air, or even substantially pure molecular oxygen, provided to maintain the contents of the fermentation vessel with a suitable oxygen partial pressure effective in assisting the microorganism species to grow in a thriving fashion. In effect, by using an oxygenated hydrocarbon substrate, the oxygen requirement for growth of the microorganism can be reduced. Nevertheless, molecular oxygen can be supplied for growth of aerobic and to a lesser extent, facultative organisms.

**[0138]** Although the aeration rate can vary over a considerable range, aeration generally can be conducted at a rate that is in the range from about 0.5 to 10 or from about 0.5 to 7 volumes (at the pressure employed and at 25° C.) of oxygen-containing gas per liquid volume in the fermenter per minute. This amount can be based on air of normal oxygen content being supplied to the reactor, and in terms of pure oxygen the respective ranges can be from about 0.1 to 1.7, or from about 0.1 to 1.3, volumes (at the pressure employed and at 25° C.) of oxygen per liquid volume in the fermenter per minute.

**[0139]** The pressure employed for the microbial conversion process can range widely. Pressures generally can be within the range of about 0 to 50 psig, from about 0 to 30 psig, or at least slightly over atmospheric pressure, as a balance of equipment and operating cost versus oxygen solubility achieved. Greater than atmospheric pressures can

increase a dissolved oxygen concentration in the aqueous ferment, which in turn can help increase cellular growth rates.

**[0140]** The fermentation temperature can vary somewhat, but for most bacterial host species used in the present, the temperature generally can be within the range of from about 20° C. to 40° C., or in the range of from about 28° C. to 37° C., depending on the strain of microorganism chosen.

**[0141]** In some instances, the microorganisms may require a source of assimilable nitrogen. The source of assimilable nitrogen can be any nitrogen-containing compound or compounds capable of releasing nitrogen in a form suitable for metabolic utilization by the microorganism. While a variety of organic nitrogen source compounds, such as protein hydrolysates, can be employed, usually cheap nitrogen-containing compounds such as ammonia, ammonium hydroxide, urea, and various ammonium salts such as ammonium phosphate, ammonium sulfate, ammonium pyrophosphate, ammonium chloride, or various other ammonium compounds can be utilized. Ammonia gas itself can be convenient for large-scale operations, and can be employed by bubbling through the aqueous ferment (fermentation medium) in suitable amounts. At the same time, such ammonia can also be employed to assist in pH control.

**[0142]** The pH range in the aqueous microbial ferment (fermentation admixture) can be in the exemplary range from about 2.0 to 8.0. However, pH range optima for certain microorganisms can be dependent on the media employed to some extent, as well as the particular microorganism, and thus change somewhat with change in media.

**[0143]** The average retention time of the fermentation admixture in the fermenter can vary considerably, depending in part on the fermentation temperature and culture employed. In some embodiments, the fermentation can be conducted in such a manner that the carbon-containing substrate can be controlled as a limiting factor, thereby providing good conversion of the carbon-containing substrate to cells and avoiding contamination of the cells with a substantial amount of unconverted substrate. The latter may not be a problem with water-soluble substrates, since any remaining traces can be readily removed. It may be a problem, however, in the case of non-water-soluble substrates, and may use added product-treatment steps such as suitable washing steps. The time needed to reach this limiting substrate level may vary with the particular microorganism and fermentation process being conducted. The fermentation can be conducted as a batch or continuous operation, fed batch operation can be used for ease of control, production of uniform quantities of products, and most economical uses of all equipment.

**[0144]** If desired, part or all of the carbon and energy source material and/or part of the assimilable nitrogen source such as ammonia can be added to the aqueous mineral medium prior to feeding the aqueous mineral medium into the fermenter. Indeed, each of the streams introduced into the reactor can be controlled at a predetermined rate, or in response to a need determinable by monitoring such as concentration of the carbon and energy substrate, pH, dissolved oxygen, oxygen or carbon dioxide in the off-gases from the fermenter, cell density measurable by light transmittance, or the like. The feed rates of the various materials can be varied so as to obtain as rapid a cell growth rate as possible, consistent with efficient utilization of the carbon and energy source, to obtain as high a yield of

microorganism cells relative to substrate charge as possible, and to obtain the highest production of the desired protein per unit volume.

**[0145]** In a batch, equipment, reactor, or fermentation means, vessel or container, piping, attendant circulating or cooling devices, and the like, can be initially sterilized, usually by employing steam at about 121° C. for at least about 15 minutes. The sterilized reactor can be inoculated with a culture of the selected microorganism in the presence of all the required nutrients, including oxygen, and the carbon-containing substrate.

**[0146]** Methods of Secretion in bacteria

**[0147]** In some instances, an expressed polypeptide can be secreted from a host cell (e.g., bacteria). Secretion of a polypeptide can comprise releasing the polypeptide from a cell or subcellular compartment in a cell (e.g., nucleus, cell wall, plasma membrane). Secretion can occur through plasma membranes, which can surround cells and/or subcellular compartments. In some instances, secretion can refer to releasing a polypeptide to the cell envelope. In some instances, secretion can refer to releasing a polypeptide to the extracellular space (e.g., into the culture media).

**[0148]** A host cell of the disclosure can comprise secretory pathways, which can comprise a number of proteins that function together to secrete a protein. In some instances, the host cell can comprise a twin-arginine translocation (TAT) secretory pathway. In some instances, an organism can comprise a SEC secretory pathway. The TAT secretory pathway can comprise secretion of polypeptides (e.g., globins) in a folded state. The TAT secretory pathway can transport proteins across a plasma membrane (e.g., lipid layer, i.e., lipid bilayer).

**[0149]** The disclosure provides for secretion factors and methods that can be used in host cells to ameliorate the bottleneck to protein secretion and the production of proteins in secreted form, in particular when the polypeptides are recombinantly introduced and expressed by the host cell. The present disclosure provides the secretion factors TatC and TatA derived from *Bacillus subtilis*. In particular, the TatAdCd and TatAyCy peptide, as well as the genes encoding them are also suitable secretion factors. PhoD of *B. subtilis*, can be secreted via the twin-arginine translocation pathway. TatAdCd is of major importance for the secretion of PhoD, whereas TatAyCy may not be required for this process.

**[0150]** Expression of Polypeptides in Plants

**[0151]** A polypeptide can be expressed in monocot plants and/or dicot plants. Techniques for introducing nucleic acids into plants are known in the art, and include, without limitation, *Agrobacterium*-mediated transformation, viral vector-mediated transformation, electroporation, and particle gun transformation (also referred to as biolistic transformation). See, for example, U.S. Pat. Nos. 5,538,880; 5,204,253; 6,329,571; and 6,013,863; Richards et al., *Plant Cell. Rep.* 20:48-20 54 (2001); Somleva et al., *Crop Sci.* 42:2080-2087 (2002); Sinagawa-Garcia et al., *Plant Mol Biol* (2009) 70:487-498; and Lutz et al., *Plant Physiol.*, 2007, Vol. 145, pp. 1201-1210. In some instances, intergenic transformation of plastids can be used as a method of introducing a polynucleotide into a plant cell. In some instances, the method of introduction of a polynucleotide into a plant comprises chloroplast transformation. In some instances, the leaves and/or stems can be the target tissue of the introduced polynucleotide. If a cell or cultured tissue is

used as the recipient tissue for transformation, plants can be regenerated from transformed cultures if desired, by techniques known to those skilled in the art.

**[0152]** Other suitable methods for introduce polynucleotides include electroporation of protoplasts, polyethylene glycol-mediated delivery of naked DNA into plant protoplasts, direct gene transformation through imbibition (e.g., introducing a polynucleotide to a dehydrated plant), transformation into protoplasts (which can comprise transferring a polynucleotide through osmotic or electric shocks), chemical transformation (which can comprise the use of a polybrene-spermidine composition), microinjection, pollen-tube pathway transformation (which can comprise delivery of a polynucleotide to the plant ovule), transformation via liposomes, shoot apex method of transformation (which can comprise introduction of a polynucleotide into the shoot and regeneration of the shoot), sonication-assisted *agrobacterium* transformation (SAAT) method of transformation, infiltration (which can comprise a floral dip, or injection by syringe into a particular part of the plant (e.g., leaf)), silicon-carbide mediated transformation (SCMT) (which can comprise the addition of silicon carbide fibers to plant tissue and the polynucleotide of interest), electroporation, and electrophoresis.

**[0153]** A polypeptide can be expressed in many different plant species, including, for example, grains such as, e.g., corn, maize, oats, rice, wheat, barley, rye, triticale, teff, oilseeds including cottonseed, sunflower seed, safflower seed, crambe, camelina, mustard, rapeseed, leafy greens such as, e.g., lettuce, spinach, kale, collard greens, turnip greens, chard, mustard greens, dandelion greens, broccoli, cabbage, sugar cane, trees, root crops such as cassava, sweet potato, potato, carrots, beets, turnips, plants from the legume family, such as, e.g., clover, peas such as cowpeas, English peas, yellow peas, green peas, beans such as, e.g., soybeans, fava beans, lima beans, kidney beans, garbanzo beans, mung beans, pinto beans, lentils, lupins, mesquite, carob, soy, and peanuts, coconut, vetch (*vicia*), stylo (*stylosanthes*), arachis, indigofera, acacia, leucaena, cyamopsis, and sesbania. Plants not ordinarily consumed by humans, including biomass crops, including switchgrass, miscanthus, tobacco, *Arundo donax*, energy cane, sorghum, other grasses, alfalfa, corn stover, kelp, or other seaweeds. Polypeptides that can be found in any organism in the plant kingdom may be used in the present disclosure. In some instances, the plant can be soy (*Glycine max*). In some instances, the plant can be barley (*Hordeum vulgare*). In some instances, the plant can be *Nicotiana tabacum*. In some instances, the plant or plant cell is not a *Nicotiana* plant or plant cell.

**[0154]** In some embodiments, the *Nicotiana tabacum* variety, breeding line, or cultivar can be *N. tabacum* accession PM016, PM021, PM92, PM102, PM132, PM204, PM205, PM215, PM216 or PM217 as deposited with NCIMB, Aberdeen, Scotland, or DAC Mata Fina, P02, BY-64, AS44, RG17, RG8, HB04P, Basma Xanthi BX 2A, Coker 319, Hicks, McNair 944 (MN 944), Burley 21, K149, Yaka JB 125/3, Kasturi Mawar, NC 297, Coker 371 Gold, P02, Wislica, Simmaba, Turkish Samsun, AA37-1, B13P, F4 from the cross BU21xHoja Parado line 97, Samsun NN, Izmir, Xanthi NN, Karabagar, Denizli and P01.

**[0155]** Expression of a polypeptide of the disclosure can comprise transient expression and/or constitutive expression (e.g., developing of a stable cell line).

**[0156]** *Agrobacterium* Species and Strains

**[0157]** The disclosure can provide for *Agrobacterium* strains for use in methods for producing a polypeptide by expression of an expressible sequence (e.g., a sequence encoding a polypeptide of the disclosure). One of the *Agrobacterium* strains can be used to infiltrate a preselected plant variety in order to optimize the yield of the polypeptide. In certain embodiments, the *Agrobacterium* species that may be used in method according to the disclosure can include but are not limited to *Agrobacterium tumefaciens*, *Agrobacterium rhizogenes*, *Agrobacterium radiobacter*, *Agrobacterium rubi*, *Agrobacterium vitis*. In some embodiments, at least one *Agrobacterium* strain comprises *Agrobacterium tumefaciens*. The *Agrobacterium* species used can be a wild type (e.g., virulent) or a disarmed strain. Suitable strains of *Agrobacterium* can include wild type strains (e.g., such as *Agrobacterium tumefaciens*) or strains in which one or more genes is mutated to increase transformation efficiency, (e.g., such as *Agrobacterium* strains wherein the vir gene expression and/or induction thereof is altered due to the presence of mutant or chimeric virA or virG genes), *Agrobacterium* strains comprising an extra virG gene copies, such as the super virG gene derived from pTiBo542, linked to a multiple-copy plasmid. Other suitable strains can include, but are not limited to: *A. tumefaciens* C58C1, A136; LBA401 1, LBA4404; EHA101; EHA105; AGL1; and A281. In some embodiments, the selected *Agrobacterium* strain can be AGL1, EHA105, GV2260, GV3101, or Chry5.

**[0158]** In some embodiments, multiple suspensions of *Agrobacterium* cells, each expressing different genes can be used to produce the polypeptide, or to enhance the level of expression of a polypeptide of the disclosure. In such instances, it is contemplated that the *Agrobacterium* cells in the different suspensions of *Agrobacterium* cells can be the same strain or different strains. Alternatively, or additionally, a single *Agrobacterium* strain may comprise a plurality of sequences comprising different polynucleotides, particularly polynucleotides encoding polypeptides of the disclosure. The different genes may be comprised within a single nucleic acid molecule (e.g., a single vector) or may be provided in different vectors. A non-limiting example of a second gene that can be expressed in the host plant is a gene that encodes a suppressor of silencing, of viral origin.

**[0159]** Expression of a polypeptide in a host cell (e.g., plant cell), can occur for at least about 0.1 hours, 0.5 hours, 1 hour, 2 hours, 3 hours, 4 hours, 5 hours, 6 hours, 7 hours, 8 hours, 9 hours, 10 hours, 11 hours, 12 hours, 13 hours, 14 hours, 15 hours, 16 hours, 17 hours, 18 hours, 19 hours, 20 hours, 21 hours, 22 hours, 23 hours, 24 hours, 1 days, 2 days, 3 days, 4 days, 5 days, 6 days, 7 days, 1 week, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 7 weeks, 8 weeks, 9 weeks, 10 weeks, 15 weeks, 20 weeks, 30 or more weeks. A polypeptide can be expressed in a host cell for at most about 0.1 hours, 0.5 hours, 1 hour, 2 hours, 3 hours, 4 hours, 5 hours, 6 hours, 7 hours, 8 hours, 9 hours, 10 hours, 11 hours, 12 hours, 13 hours, 14 hours, 15 hours, 16 hours, 17 hours, 18 hours, 19 hours, 20 hours, 21 hours, 22 hours, 23 hours, 24 hours, 1 days, 2 days, 3 days, 4 days, 5 days, 6 days, 7 days, 1 week, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 7 weeks, 8 weeks, 9 weeks, or 10 weeks, 15 weeks, 20 weeks, 30 or more weeks.

**[0160]** A polypeptide can be expressed at a variety of temperatures. A polypeptide can be expressed at a temperature of at least about 4, 10, 16, 18, 21, 25, 26, 27, 28, 29, 30,

31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 42, 42, 43, 44, 45, 46, 47, 48, 49, or 50° C. A polypeptide can be expressed at a temperature of at most about 4, 10, 16, 18, 21, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 42, 42, 43, 44, 45, 46, 47, 48, 49, or 50° C.

**[0161]** Enhancement of Endogenous Polypeptides

**[0162]** In some embodiments, the disclosure provides for enhanced production of an endogenous polypeptide (e.g., an endogenous heme-containing polypeptide). Enhanced production of an endogenous polypeptide can be accomplished by modulating the pathway that produces the endogenous polypeptide. Modulation can refer to modulation of transcription, translation, subcellular localization, localization to different tissues, timing of expression, folding, affinity for binding partners, and the like. Modulation can occur at the DNA level (e.g., knock-out the gene, knock-in an enhancer/promoter element). Modulation can occur at the RNA level (e.g., silence the gene via RNA interference). Modulation can occur at the protein level (e.g., modulation by allosteric inhibitors, small molecule binders).

**[0163]** In some instances, modulation can refer to altering the activity and/or levels of the endogenous polypeptide by at least about 1 fold, 2-fold, 3-fold, 4-fold, 5-fold, 10-fold, 20-fold, 50-fold, or 100-fold or more higher or lower relative to the wild-type levels of the endogenous polypeptide. In some instances, modulation can refer to altering the activity and/or levels of the endogenous polypeptide by at most about 1 fold, 2-fold, 3-fold, 4-fold, 5-fold, 10-fold, 20-fold, 50-fold, or 100-fold or more higher or lower relative to the wild-type levels of the endogenous polypeptide. In some instances, modulation can refer to altering the activity and/or levels of the endogenous polypeptide by at least about 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100% of the wild-type levels of the endogenous polypeptide. In some instances, modulation can refer to altering the activity and/or levels of the endogenous polypeptide by at least about 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100% of the wild-type levels of the endogenous polypeptide.

**[0164]** In some instances, polypeptides in the heme biosynthesis pathway that can produce the heme cofactor can be modulated. See, for example, Tanaka and Tanaka, *Ann. Rev. Plant Biol.* 2007. 58:321-46, which describes modifications to the tetrapyrrole biosynthesis pathway in plants. The modulation of polypeptides in the heme biosynthetic pathway can be at the DNA, RNA, or protein level. In some instances, the modulation of other polypeptides in the pathway can refer to increasing the levels and/or activity of an activator of the pathway. In some instances, the modulation of other polypeptides in the pathway can refer to decreasing the levels and/or activity of a suppressor of the pathway.

**[0165]** In some instances, modulation can refer to altering the activity and/or levels of polypeptides in the heme biosynthesis pathway (including the heme cofactor that associates with a heme-containing polypeptide of the disclosure) by at least about 1 fold, 2-fold, 3-fold, 4-fold, 5-fold, 10-fold, 20-fold, 50-fold, or 100-fold or more higher or lower relative to the wild-type levels of the polypeptide in the pathway. In some instances, modulation can refer to altering the activity and/or levels of polypeptides in the heme biosynthesis pathway (including the heme cofactor) by at most about 1 fold, 2-fold, 3-fold, 4-fold, 5-fold, 10-fold, 20-fold, 50-fold, or 100-fold or more higher or lower relative to the wild-type levels of the polypeptide in the pathway. In some instances, modulation can refer to altering

the activity and/or levels of polypeptides in the heme biosynthesis pathway (including the heme cofactor) by at least about 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100% of the wild-type levels of the polypeptide in the pathway. In some instances, modulation can refer to altering the activity and/or levels of polypeptides in the heme biosynthesis pathway (including the heme cofactor) by at least about 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100% of the wild-type levels of the polypeptide in the pathway.

**[0166]** In some instances, modulation can refer to altering the expression levels of an endogenous polypeptide in a specific location of the host cell. For example, the expression levels of an endogenous polypeptide can be altered in a leaf, a seed, a bean, a stalk, a xylem, a stamen, and a petal, or any combination thereof.

**[0167]** Methods of Purification

**[0168]** An expressed and/or secreted polypeptide of the disclosure may be recovered (e.g., from the culture medium or from a plant tissue). For example, when the expressed heme-containing polypeptide is secreted from the bacterial cells, the polypeptide can be purified from the culture medium. In some embodiments, the host cells expressing the polypeptide can be removed from the media before purification of the polypeptide (e.g., by centrifugation).

**[0169]** When the expressed recombinant desired polypeptide is not secreted from a host cell, the host cell can be disrupted and the polypeptide released into an aqueous "extract" which can be the first stage of purification. The expression host cells can be collected from the media before the cell disruption. The cell disruption may be performed by using any suitable means, such as by lysozyme or beta-glucanase digestion, grinding, sonication, homogenization, milling or by forcing the cells through high pressure.

**[0170]** A recovered polypeptide may be purified. Purification may be accomplished by means of a salt (e.g., ammonium sulfate) or low pH (typically less than 3) wash/fractionation or chromatographic procedures (e.g., ion exchange chromatography, affinity chromatography, hydrophobic interaction chromatography, hydrophobic charge induction chromatography, size exclusion chromatography etc.). During purification, the cumulative abundance by mass of protein components other than the specified protein, which can be a single monomeric or multimeric protein species, can be reduced by a factor of 2 or more, 3 or more, 5 or more, 10 or more, 20 or more, 50 or more, 100 or more or 1000 or more relative to the source material from which the specified protein was isolated.

**[0171]** In some instances, a polypeptide can be recovered from a fermenter. The fermentation broth can generally comprise cellular debris, including cells, various suspended solids and other biomass contaminants, as well as the desired protein product, which can be removed from the fermentation broth. Suitable processes for such removal can include conventional solid-liquid separation techniques (e.g., centrifugation, filtration, dialysis, microfiltration, rotary vacuum filtration, or other known processes), to produce a cell-free filtrate. In some embodiments, it can be acceptable to further concentrate the fermentation broth or the cell-free filtrate prior to the purification and/or crystallization process using techniques such as ultrafiltration, evaporation and/or precipitation. In some instances, the polypeptide is further purified to reduce the cumulative abundance by mass of protein components other than the specified protein, which can be a single monomeric or multimeric protein species, by

a factor of 2 or more, 3 or more, 5 or more, 10 or more, 20 or more, 50 or more, 100 or more or 1000 or more relative to the source material from which the specified protein was isolated. Purification may be accomplished by means of a salt (e.g., ammonium sulfate) or low pH (typically less than 3) wash/fractionation or chromatographic procedures (e.g., ion exchange chromatography, affinity chromatography, hydrophobic interaction chromatography, and/or hydrophobic charge induction chromatography etc.).

**[0172]** Characterization of a Polypeptide

**[0173]** A purified polypeptide can be characterized for purity, heme content, oligomerization state, stability, degradation, binding affinity and the like. For some applications the polypeptides (e.g., globins) produced using the present disclosure can be very highly pure (e.g., having a purity of more than 99%). A purified polypeptide can be characterized for odor, taste and color.

**[0174]** The purified polypeptide can be at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% pure. The purified polypeptide can be at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% pure. The purified polypeptide can comprise at least about 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% impurities. The purified polypeptide can comprise at most about 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% impurities. The purified polypeptide can comprise at least about 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 parts per million impurities. The purified polypeptide can comprise at most about 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 parts per million impurities. The purified polypeptide can comprise at least about 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 parts per billion impurities. The purified polypeptide can comprise at most about 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 parts per billion impurities.

**[0175]** In some instances, the purified globin can be tested for activity, oligomerization state, proper protein folding, stability, secondary structure and/or heme content. Activity, oligomerization state, protein folding, and/or stability can be determined by a number of methods including spectroscopy, ELISA, binding assays, analytical ultracentrifugation, circular dichroism, x-ray crystallography, surface plasmon resonance, mass spectrometry, or NMR.

**[0176]** A polypeptide of this disclosure may have similar properties to myoglobin isolated from animal tissues. In one embodiment a group of people can be asked to rate a myoglobin isolated from an animal tissue, according to properties that describe the myoglobin. These ratings can be used as an indication of the properties of the animal tissue derived myoglobin. The polypeptide of the present invention can then be compared to the animal derived globin to determine how similar the polypeptide of this disclosure is to the animal tissue derived myoglobin. So, in some embodiments, the polypeptide is rated similar to animal tissue derived myoglobin according to human evaluation. In some embodiments the polypeptide is indistinguishable from animal tissue derived myoglobin to a human.

**[0177]** In some embodiments the polypeptides of this disclosure are compared to animal tissue derived myoglobin based upon olfactometer readings. In various embodiments the olfactometer can be used to assess odor concentration and odor thresholds, odor suprathresholds with comparison to a reference gas, hedonic scale scores to determine the degree of appreciation, or relative intensity of odors. In some embodiments the olfactometer allows the training and automatic evaluation of expert panels. In some embodiments the consumable is a product that causes similar or identical olfactometer readings. In some embodiments the similarity is sufficient to be beyond the detection threshold of human perception.

**[0178]** Gas chromatography—mass spectrometry (GCMS) is a method that combines the features of gas-liquid chromatography and mass spectrometry to separate and identify different substances within a test sample. GCMS can, in some embodiments, be used to evaluate the properties of polypeptides of this disclosure. For example, volatile chemicals can be isolated from the head space around animal tissue derived myoglobin. These chemicals can be identified using GCMS. A profile of the volatile chemicals in the headspace around animal tissue derived myoglobin is thereby created. In some instances each peak of the GCMS can be further evaluated. For instance, a human could rate the experience of smelling the chemical responsible for a certain peak. This information could be used to further refine the profile. GCMS could then be used to evaluate the properties of a polypeptide of the disclosure. The GCMS profile could be used to refine the polypeptide.

**[0179]** Heme content can refer to the percentage of polypeptide molecules that comprise the correct amount of heme moieties. For example, if a polypeptide of the disclosure binds one heme moiety, then heme content can refer to the number of polypeptides that are bound to the heme moiety. Heme content of a polypeptide can be at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 75, 80, 85, 90, 95, or 100%. Heme content of a polypeptide can be at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 75, 80, 85, 90, 95, or 100%. In some instances, heme content can be expressed as a molar ratio of polypeptide concentration to heme concentration. The molar ratio heme content can be at least about 1:1, 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:20, 1:30, or 1:40 or less. The molar ratio heme content can be at most about 1:1, 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:20, 1:30, or 1:40 or less. The heme content can be 1-fold, 2-fold, 3-fold, 4-fold, 5-fold, 10-fold, 15-fold, 20-fold, 30-fold, or 40-fold or more lower than the heme content of a full-occupied polypeptide (e.g., the polypeptide is 100% occupied with heme). The heme content can be 1-fold, 2-fold, 3-fold, 4-fold, 5-fold, 10-fold, 15-fold, 20-fold, 30-fold, or 40-fold or more higher than the heme content of a fully-unoccupied polypeptide (e.g., the polypeptide is 0% occupied with heme). Heme content can be determined by a number of methods including spectroscopy (Raman, UV-Vis), electron paramagnetic resonance (EPR), protein denaturation assays, heme stealing assays, and heme reduction assays.

**[0180]** Methods for Using a Polypeptide in a Meat Consumable

**[0181]** The disclosure provides for methods for the use of a polypeptide of the disclosure in a meat consumable. The consumables can compete with, supplement or replace animal based foods. For instance, the consumables can be meat

replicas made entirely from plant sources. The consumables can be made to mimic the cut or appearance of meat as it is currently sold. For instance, a consumable may be visually similar to or indistinguishable from ground beef or a particular cut of beef. Alternatively, the consumables can be made with a unique look or appearance. For instance, the consumable could contain patterns or lettering that is based upon the structure of the consumable. In some instances, the consumables can look like traditional meat products after they are prepared. For example, a consumable may be produced which is larger than a traditional cut of beef but which, after the consumable is sliced and cooked appears the same as a traditional cooked meat. In some embodiments the consumable may resemble a traditional meat shape in two dimensions, but not in a third. For example, the consumable may resemble a cut of meat in two dimensions (for example when viewed from the top), but may be much longer (or thicker) than the traditional cut. A meat consumable (e.g., substitute) can have similar physical characteristics as traditional meat (taste, texture, force, nutrients). In some embodiments, a meat consumable can comprise a similar cook loss characteristic as meat. In some embodiments a meat consumable can comprise a similar fat and protein content as ground beef has the same reduction in size when cooked as real ground beef. Methods of producing a meat consumable are described in PCT/US2012/046560, which is hereby incorporated by reference in its entirety.

**[0182]** In some instances, a meat consumable can comprise a polypeptide of the disclosure. A polypeptide of the disclosure can be used as a colorant or indicator of cooking of the meat consumable.

**[0183]** In some instances, the disclosure provides for a method for expressing a polypeptide (e.g., globin), in a host cell, secreting the polypeptide from the host cell, purifying the secreted polypeptide, and mixing the purified polypeptide with fats and lipids to produce a meat substitute.

**[0184]** In some instances, the disclosure provides for a method for enhancing the expression of an endogenous polypeptide (e.g., globin) in a host cell, purifying the polypeptide from the cell, and mixing the purified polypeptide with fats and lipids to produce a meat substitute.

**[0185]** In some instances, the disclosure provides for a method for expressing a polypeptide (e.g., globin), in a host cell (e.g., a plant), purifying the secreted polypeptide, and mixing the purified polypeptide with fats and lipids to produce a meat substitute.

**[0186]** Compositions

**[0187]** In some instances, the disclosure provides for a composition comprising a polypeptide of the disclosure. A composition can comprise media into which the polypeptide was secreted. A composition can comprise at least about 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10% (w/w) or more media. A composition can comprise at most about 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10% or more media. A composition can comprise at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more parts per million media. A composition can comprise at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more parts per million media. A composition can comprise at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more parts per billion media. A composition can comprise at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more parts per billion media.

**[0188]** A composition can comprise a host cell (e.g., a bacterium). A host cell of the composition can be the host

cell from which the polypeptide was expressed and/or secreted. A composition can comprise at least about 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10% (w/w) or more host cells. A composition can comprise at most about 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10% host cells. A composition can comprise at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more parts per million host cells. A composition can comprise at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more parts per million host cells. A composition can comprise at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more parts per billion host cells. A composition can comprise at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more parts per billion host cells. A composition can be at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% free of a host cell. A composition can be at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% free of a host cell.

**[0189]** A composition can comprise a component of a recombinant host cell (e.g., a recombinant bacterial cell or plant cell). A component of a host cell can include, for example, a cell wall, a subcellular compartment (e.g., Golgi complex, endoplasmic reticulum, or nucleus), nucleic acid, protein, genomic DNA, and/or a plasma membrane. For a bacterial cell, a component also can include flagella, and for a plant or plant cell, a component can be any part of the plant such as a shoot, a stem, a seed, a bean, a leaf, xylem tissue, a rosette, a root. A component of a host cell can be a component of a host cell from which the polypeptide was expressed and/or secreted. For example, a composition can include a non-naturally occurring component of a recombinant host cell such as a fusion protein or an exogenous nucleic acid. A composition can comprise at least about 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10% (w/w) or more of a component of a host cell. A composition can comprise at most about 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10% of a component of a host cell. A composition can comprise at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more parts per million of the component of a bacterium. A composition can comprise at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 parts per million of a component of a host cell. A composition can comprise at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more parts per billion of a component of a host cell. A composition can comprise at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 parts per billion of a component of a host cell. A composition can be at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% free of a component of a host cell. A composition can be at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 75, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% free of a component of a host cell.

**[0190]** Meat Consumable Compositions

**[0191]** In some instances, a composition of the disclosure can comprise a meat consumable and a host cell (e.g., bacterium, a part of a bacterium, and/or a part of a plant). In some instances, a composition can further comprise a polypeptide of the disclosure. In some instances, a composition comprises a polypeptide and a meat consumable (i.e., meat substitute) (as described in PCT/US2012/046560, which is herein incorporated by reference in its entirety).

**[0192]** A meat consumable can refer to meat-like product (e.g., a meat substitute) that is not made of meat. A meat consumable can refer to a meat substitute that is made from

non-animal products (e.g., a plant). A meat consumable can be meat replicas made entirely from plant sources. The consumables may also be made from a combination of plant based sources and animal based sources. The consumables can be made to mimic the cut or appearance of meat as it is currently sold. For instance, a consumable may be visually similar to or indistinguishable from ground beef or a particular cut of beef. In some instances, the consumables look like traditional meat products after they are prepared. The meat consumable can be substantially or entirely composed of ingredients derived from non-animal sources, yet recapitulates key features associated with the cooking and consumption of an equivalent meat product derived from animals.

**[0193]** A composition can comprise a meat consumable and a polypeptide of the disclosure. A meat consumable can comprise at least about 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100% (w/w) of one or more polypeptides of the disclosure. In some instances, a meat consumable can comprise at most about 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100% of one or more polypeptides of the disclosure. A meat consumable can comprise at least about 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100% weight/volume of one or more polypeptides of the disclosure. In some instances, a meat consumable can comprise at most about 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100% weight/volume of one or more polypeptides of the disclosure.

**[0194]** A composition can comprise a meat consumable and a host cell (e.g., bacterium). A host cell of the composition can be the host cell from which the polypeptide was expressed and/or secreted. A composition can comprise a meat consumable and at least about 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10% (w/w) or more host cells. A composition can comprise a meat consumable and at most about 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10% host cells. A composition can comprise a meat consumable and at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more parts per million host cell. A composition can comprise a meat consumable and at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 parts per million host cell. A composition can comprise a meat consumable and at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more parts per billion host cell. A composition can comprise a meat consumable and at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 parts per billion host cell. A composition can comprise a meat consumable and be at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% free of a host cell. A composition comprises a meat consumable and can be at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% free of a host cell.

**[0195]** A composition can comprise a part of a meat consumable and a component of a host cell (e.g., a part of a bacterium). A component of a host cell can include a cell wall, a subcellular compartment (e.g., Golgi complex, endoplasmic reticulum, nucleus), a flagella, nucleic acid, protein, genomic DNA, or a plasma membrane. A component of a host cell can be a part of a bacterium from which the polypeptide was expressed and/or secreted. A composition can comprise a meat consumable and at least about 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10% or more of part of a host cell. A composition can comprise a

meat consumable and at most about 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10% of a component of a host cell. A composition can comprise a meat consumable and at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more parts per million of part of a host cell. A composition can comprise a meat consumable and at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 parts per million of a component of a host cell. A composition can comprise a meat consumable and at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more parts per billion of a component of a host cell. A composition can comprise a meat consumable and at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 parts per billion of a component of a host cell. A composition can comprise a meat consumable and be at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% free of a component of a host cell. A composition can comprise a meat consumable and be at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% free of a component of a host cell.

**[0196]** A composition can comprise a meat consumable and a component of a host cell (e.g., plant, e.g., a tobacco plant, i.e., a *Nicotiana tabacum* species or a soybean plant, i.e., a *Glycine max* species). In some embodiments, the host cell is not a *Nicotiana* plant. A part of a host cell can include a cell wall, a subcellular compartment (e.g., Golgi complex, endoplasmic reticulum, nucleus), a shoot, a stem, a leaf, a seed, a bean, a xylem, a rosette, a root, nucleic acid, protein, genomic DNA, and a plasma membrane. A component of a host cell can be a part of a plant from which the polypeptide was expressed and/or secreted. A composition can comprise a meat consumable and at least about 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10% or more of a component of a host cell. A composition can comprise a meat consumable and at most about 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10% of a component of a host cell. A composition can comprise a meat consumable and at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more parts per million of a component of a host cell. A composition can comprise a meat consumable and at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 parts per million of a component of a host cell. A composition can comprise a meat consumable and at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more parts per billion of a component of a host cell. A composition can comprise a meat consumable and at most about 11, 2, 3, 4, 5, 6, 7, 8, 9, or 10 parts per billion of a component of a host cell. A composition can comprise a meat consumable and can be at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% free of a component of a host cell. A composition can comprise a meat consumable and be at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% free of a component of a host cell.

**[0197]** In some embodiments, the disclosure can provide for a consumable that can be substantially or entirely composed of ingredients derived from non-animal sources, yet recapitulates key features associated with the cooking and consumption of an equivalent meat product derived from animals. The equivalent meat product can be a white meat or a dark meat. The equivalent meat product can be derived from any animal. Non-limiting examples of animals used to derive the equivalent meat product include farmed animals such as, e.g., cattle, sheep, pig, chicken, turkey, goose, duck, horse, dog or game animals (whether wild or farmed) such

as, e.g., rabbit, deer, bison, buffalo, boar, snake, pheasant, quail, bear, elk, antelope, pigeon, dove, grouse, fox, wild pig, goat, kangaroo, emu, alligator, crocodile, turtle, groundhog, marmot, possum, partridge, squirrel, raccoon, whale, seal, ostrich, capybara, nutria, guinea pig, rat, mice, vole, any variety of insect or other arthropod, seafood such as, e.g., fish, crab, lobster, oyster, muscle, scallop, abalone, squid, octopus, sea urchin, tunicate and others. Many meat products are typically derived from skeletal muscle of an animal but it is understood that meat can also come from other muscles or organs of the animal. In some embodiments, the equivalent meat product is a cut of meat derived from skeletal muscle. In some embodiments, the equivalent meat product is an organ such as, e.g., a kidney, heart, liver, gallbladder, intestine, stomach, bone marrow, brain, thymus, lung, tongue. Accordingly, in some embodiments the compositions of the present are consumables similar to skeletal muscle or organs.

**[0198]** In some instances, the disclosure provides meat substitute products comprising one or more of a first composition comprising a muscle tissue replica, a second composition comprising an adipose tissue replica, and/or a third composition comprising a connective tissue replica, wherein the one or more compositions are combined in a manner that recapitulates the physical organization of meat. In other aspects, the present disclosure provides compositions for a muscle tissue replica (herein referred to as “muscle replica”), an adipose tissue replica (herein referred to as “fat replica”), and a connective tissue replica (herein referred to as “connective tissue replica”). In some embodiments, the compositions and meat substitute products are principally or entirely composed of ingredients derived from non-animal sources. In alternative embodiments, the muscle, fat, and/or connective tissue replica, or the meat substitute products comprising one or more of said replicas, are partially derived from animal sources but supplemented with ingredients derived from non-animal sources.

**[0199]** In some embodiments, meat products can be substantially derived from animal sources but which are supplemented with one or more of a muscle tissue replica, a fat replica, and/or a connective tissue replica, wherein the replicas can be derived substantially or entirely from non-animal sources. A non-limiting example of such a meat product is an ultra-lean ground beef product supplemented with a non-animal derived fat replica which can improve texture and mouthfeel while preserving the health benefits of a consumable low in animal fat. Such alternative embodiments result in products with properties that more closely recapitulate key features associated with preparing and consuming meat but which are less costly and associated with a lesser environmental impact, less animal welfare impact, or improved health benefits for the consumer.

**[0200]** The physical organization of the meat substitute product can be manipulated by controlling the localization, organization, assembly, or orientation of the muscle, fat, and/or connective tissue replicas described herein. In some embodiments the product is designed in such a way that the replicas described herein are associated with one another as in meat. In some embodiments the consumable is designed so that after cooking the replicas described herein are associated with one another as in cooked meat. In some embodiments, one or more of the muscle, fat, and/or connective tissue replicas are combined in a manner that recapitulate the physical organization of different cuts or prepa-

rations of meat. In an example embodiment, the replicas are combined in a manner that approximates the physical organization of natural ground meat. In other embodiments, the replicas are combined in a manner that approximates different cuts of beef, such as, e.g., rib eye, filet mignon, London broil, among others.

**[0201]** Indicators of Cooking Meat

**[0202]** In some instances, a polypeptide of the disclosure can be used in a composition of the disclosure as an indicator for cooking meat. The release of odorants upon cooking is an important aspect of meat consumption. In some embodiments, the consumable is a meat replica entirely composed of non-animal products that when cooked generates an aroma recognizable by humans as typical of cooking beef. In some embodiments, the consumable when cooked generates an aroma recognizable by humans as typical of cooking pork. In some embodiments, the consumable is a meat replica entirely composed of non-animal products that when cooked generates an aroma recognizable by humans as typical of cooking bacon. In some embodiments, the consumable is a meat replica entirely composed of non-animal products that when cooked generates an aroma recognizable by humans as typical of cooking chicken. In some embodiments, the consumable is a meat replica entirely composed of non-animal products that when cooked generates an aroma recognizable by humans as typical of cooking lamb. In some embodiments, the consumable is a meat replica entirely composed of non-animal products that when cooked generates an aroma recognizable by humans as typical of cooking fish. In some embodiments, the consumable is a meat replica entirely composed of non-animal products that when cooked generates an aroma recognizable by humans as typical of cooking turkey. In some embodiments the consumable is a meat replica principally or entirely composed of ingredients derived from non-animal sources, with an odorant that is released upon cooking. In some embodiments the consumable is a meat replica principally or entirely composed of ingredients derived from non-animal sources, with an odorant that is produced by chemical reactions that take place upon cooking. In some embodiments the consumable is a meat replica principally or entirely composed of ingredients derived from non-animal sources, comprising a polypeptide of the disclosure and mixtures of proteins, peptides, amino acids, nucleotides, sugars and polysaccharides and fats in combinations and spatial arrangements that enable these compounds to undergo chemical reactions during cooking to produce odorants and flavor-producing compounds. In some embodiments the consumable is a meat replica principally or entirely composed of ingredients derived from non-animal sources (e.g., a polypeptide of the disclosure), with a volatile or labile odorant that is released upon cooking. In some embodiments the consumable is a method for preparing a meat replica where meat replicas principally or entirely composed of ingredients derived from non-animal sources are heated to release a volatile or labile odorant.

**[0203]** Odorants released during cooking of meat are generated by reactions that can involve as reactants fats, protein, amino acids, peptides, nucleotides, organic acids, sulfur compounds, sugars and other carbohydrates. In some instances, a reactant can be a polypeptide of the disclosure (e.g., a globin, a secreted globin). In some embodiments the odorants that combine during the cooking of meat are identified and located near one another in the consumable,

such that upon cooking of the consumable the odorants combine. In some embodiments, the characteristic flavor and fragrance components are produced during the cooking process by chemical reactions involving amino acids, fats and sugars found in plants as well as meat. In some embodiments, the characteristic flavor and fragrance components are mostly produced during the cooking process by chemical reactions involving one or more amino acids, fats, peptides, nucleotides, organic acids, sulfur compounds, sugars and other carbohydrates found in plants as well as meat.

**[0204]** Some reactions that generate odorants released during cooking of meat can be catalyzed by iron, in particular the iron of heme, which may be comprised (e.g., bound) by a polypeptide of the disclosure. Thus in some embodiments, some of the characteristic flavor and fragrance components are produced during the cooking process by chemical reactions catalyzed by iron. In some embodiments, some of the characteristic flavor and fragrance components are produced during the cooking process by chemical reactions catalyzed by heme. In some embodiments, some of the characteristic flavor and fragrance components are produced during the cooking process by chemical reactions catalyzed by the heme iron in leghemoglobin. In some embodiments, some of the characteristic flavor and fragrance components are produced during the cooking process by chemical reactions catalyzed by the heme iron in a heme protein (e.g., the polypeptides listed in FIG. 9, hemoglobin, myoglobin, neuroglobin, cytoglobin, leghemoglobin, non-symbiotic hemoglobin, Hell's gate globin I, bacterial hemoglobins, ciliate myoglobins, flavohemoglobins, androglobin, cytoglobin, globin E, globin X, globin Y, myoglobin, leghemoglobins, erythrocytins, beta hemoglobins, alpha hemoglobins, non-symbiotic hemoglobins, protoglobins, cyanoglobins, Hell's gate globin I, bacterial hemoglobins, ciliate myoglobins, histoglobins and neuroglobins, etc).

**[0205]** Color Indicators

**[0206]** The color of meat is an important part the experience of cooking and eating meat. For instance, cuts of beef are of a characteristic red color in a raw state and gradually transition to a brown color during cooking. As another example, white meats such as chicken or pork have a characteristic pink color in their raw state and gradually transition to a white or brownish color during cooking. The amount of the color transition is used to indicate the cooking progression of beef and titrate the cooking time and temperature to produce the desired state of done-ness. In some aspects, the disclosure provides a non-meat based meat substitute product that provides a visual indicator of cooking progression. In some embodiments, the visual indicator is a color indicator that undergoes a color transition during cooking. In particular embodiments, the color indicator recapitulates the color transition of a cut of meat as the meat progresses from a raw to a cooked state. In some embodiments, the color indicator colors the meat substitute product a red color before cooking to indicate a raw state and causes the meat substitute product to transition to a brown color during cooking progression. In some embodiments, the color indicator colors the meat substitute product a pink color before cooking to indicate a raw state and causes the meat substitute product to transition to a white or brown color during cooking progression.

**[0207]** The main determinant of the nutritional definition of the color of meat is the concentration of iron carrying proteins in the meat. In the skeletal muscle component of



meat products, one of the main iron-carrying proteins is myoglobin. So, in some embodiments, the composition is a meat consumable (e.g., replica) which comprises an iron-carrying protein. In some embodiments, the composition comprises about 0.05%, about 0.1%, about 0.2%, about 0.3%, about 0.4%, about 0.5%, about 0.6%, about 0.7%, about 0.8%, about 0.9%, about 1%, about 1.1%, about 1.2%, about 1.3%, about 1.4%, about 1.5%, about 1.6%, about 1.7%, about 1.8%, about 1.9%, about 2%, or more than about 2% of an iron-carrying protein by dry weight or total weight. In some embodiments, the composition comprises at least about 10% of a polypeptide of the disclosure. In some embodiments, the composition comprises at most about 10% of a polypeptide of the disclosure. In some cases, the iron carrying protein has been isolated and purified from a source. In other cases, the iron carrying protein has not been isolated and purified. In some cases, the source of the iron-carrying protein is an animal source, or a non-animal source such as a plant, fungus, or genetically modified organisms such as, e.g., bacteria or yeast. In some cases, the iron-carrying protein is myoglobin. In some embodiments the composition comprises a consumable that is a plant based meat replica that has animal myoglobin added. So, for example a replica of young beef can have about 0.4-1% myoglobin. In some cases, the iron-carrying protein is leghemoglobin. In some embodiments the composition comprises a consumable that is a plant based meat replica that has leghemoglobin added. So, for example a replica of young beef can have about 0.4-1% leghemoglobin. In some cases, the iron-carrying protein is a cytochrome. In some embodiments the composition comprises a consumable that is a plant based meat replica that has a cytochrome added. So, for example a replica of young beef can have about 0.4-1% of a cytochrome. In some aspects the consumable is a plant-based meat replica containing hemoglobin. In some instances, the iron-carrying protein is a polypeptide of the disclosure (e.g., a globin).

**[0208]** Additional iron containing proteins exist in nature. In some embodiments the composition (e.g., consumable) comprises an iron containing protein that is not myoglobin. In some embodiments the composition (e.g., consumable) does not contain myoglobin. In some embodiments the compositions (e.g., consumable) does not contain hemoglobin. In some embodiments the consumable is a meat replica that comprises an iron containing protein other than myoglobin or hemoglobin (e.g., the globins listed in FIG. 9, and described herein, e.g., hemoglobin, myoglobin, neuroglobin, cytoglobin, leghemoglobin, non-symbiotic hemoglobin, Hell's gate globin I, bacterial hemoglobins, ciliate myoglobins, flavohemoglobins).

**[0209]** In some embodiments the composition comprises a consumable that is a meat replica principally or entirely composed of ingredients derived from non-animal sources, including a muscle tissue replica, an adipose tissue replica, a connective tissue replica, and leghemoglobin. In some embodiments the composition comprises a consumable that is a meat replica principally or entirely composed of ingredients derived from non-animal sources, containing a heme protein. In some embodiments the composition comprises a consumable that is a meat replica principally or entirely composed of ingredients derived from

non-animal sources, containing a member of the globin protein family. In some embodiments the composition comprises a consumable that is a meat replica principally or entirely composed of ingredients derived from non-animal sources, with a high iron content from a heme protein. In some embodiments the iron content is similar to meat. In some embodiments the consumable has the distinctive red color of meat, such color provided by leghemoglobin.

**[0210]** Leghemoglobin is, in some embodiments, used as an indicator that the consumable is finished cooking. In some embodiments of the disclosure there is a method for cooking a consumable comprising detecting leghemoglobin which has migrated from the interior of the consumable to the surface when the product is cooked. In some embodiments of the disclosure there is a method for cooking a consumable comprising detecting the change in color of from red to brown when the product is cooked.

**[0211]** The oxidation state of the iron ion in leghemoglobin can be important for its color. Leghemoglobin with the heme iron in the +2 oxidation state can appear vivid red in color, while leghemoglobin with the heme iron in the +3 oxidation state can appear brownish red. Thus, in using leghemoglobin as a source of red color in a meat replica for example, it can be desirable to reduce the heme iron from the +3 state to the +2 state. Heme iron in leghemoglobin can be switched from oxidized (+3) state to reduced (+2) state with reducing reagents.

**[0212]** A heme protein can, in some embodiments, be used as an indicator that the consumable is finished cooking. In some embodiments, there is a method for cooking a consumable comprising detecting leghemoglobin which has migrated from the interior of the consumable to the surface when the product is cooked. In some embodiments, there is a method for cooking a consumable comprising detecting the change in color of from red to brown when the product is cooked.

**[0213]** A heme protein (e.g., Hemoglobin, myoglobin, neuroglobin, cytoglobin, leghemoglobin, non-symbiotic hemoglobin, Hell's gate globin I, bacterial hemoglobins, ciliate myoglobins, flavohemoglobins), can be, in some embodiments, used as an indicator that the consumable is finished cooking. So, in some embodiments, the disclosure provides for a method for cooking a consumable comprising detecting leghemoglobin which has migrated from the interior of the consumable to the surface when the product is cooked. The disclosure can provide for a method for cooking a consumable comprising detecting the change in color of from red to brown when the product is cooked.

**[0214]** Food Products comprising Purified Polypeptide

**[0215]** In some embodiments a polypeptide of the disclosure (e.g., a heme-containing polypeptide, a globin such as leghemoglobin) is added to meat to enhance the properties of meat. See, for example, WO 2014/110532, WO 2014/110539, and WO 2013/010042, each of which is incorporated by reference in its entirety. For example, a polypeptide-containing solution can be injected into raw or cooked meat. In another example a solution comprising a polypeptide of the disclosure is dripped over meat or a consumable to enhance appearance. In some embodiments advertising, photography, or videography of food products such as meat or a meat substitute is enhanced with leghemoglobin.

**[0216]** Polypeptides, for example leghemoglobin and hemoglobin, can be combined with other plant based meat replica components. In some embodiments the polypeptides

are captured in a gel that contains other components, for example lipids and or proteins. In some aspects multiple gels are combined with non-gel based heme proteins. In some embodiments the combination of the polypeptides and the other compounds of the consumable are done to insure that the heme proteins are able to diffuse through the consumable. In some embodiments the consumable comprises a heme-protein containing solution, for instance a leghemoglobin solution. In some embodiments the consumable is soaked in a heme protein containing solution, for instance a leghemoglobin solution for 1, 5, 10, 15, 20 or 30 hours. In some embodiments the consumable is soaked in a heme containing solution, for instance a leghemoglobin solution for 1, 5, 10, 15, 30, or 45 minutes.

**[0217] Muscle Replicas**

**[0218]** A large number of meat products comprise a high proportion of skeletal muscle. Accordingly, the present disclosure provides a composition derived from non-animal sources which replicates or approximates key features of animal skeletal muscle. In another aspect, the present disclosure provides a meat substitute product that comprises a composition derived from non-animal sources which replicates or approximates animal skeletal muscle. Such a composition will be labeled herein as "muscle replica". In some embodiments, the muscle replica and/or meat substitute product comprising the muscle replica are partially derived from animal sources. In some embodiments, the muscle replica and/or meat substitute product comprising the muscle replica are entirely derived from non-animal sources.

**[0219]** Many meat products comprise a high proportion of striated skeletal muscle in which individual muscle fibers are organized mainly in an isotropic fashion. Accordingly, in some embodiments the muscle replica comprises fibers that are to some extent organized isotropically. In some embodiments the fibers comprise a protein component. In some embodiments, the fibers comprise about 1%, about 2%, about 5%, about 10%, about 15%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, about 95%, about 99% or more of a protein component. Animal skeletal muscle typically contains around 1% myoglobin, but can be as much as 7% of muscle mass in some whale muscles. In some embodiments the muscle replica comprises hemoglobins of this disclosure.

**[0220]** In some embodiments, the protein component comprises one or more isolated, purified proteins. For example, the one or more isolated, purified protein can comprise the 8S globulin from Moong bean seeds, or the albumin or globulin fraction of pea seeds. These proteins provide examples of proteins with favorable properties for constructing meat replicas because of their ability to form gels with textures similar to animal muscle or fat tissue. Examples and embodiments of the one or more isolated, purified proteins are described herein. The list of potential candidates here is essentially open and may include Rubisco, any major seed storage proteins, proteins isolated from fungi, bacteria, archaea, viruses, or genetically engineered microorganisms, or synthesized in vitro. The proteins may be artificially designed to emulate physical properties of animal muscle tissue. The proteins may be artificially designed to emulate physical properties of animal muscle tissue. In some embodiments, one or more isolated, purified proteins accounts for about 0.1%, 0.2%, 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%,

45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99% or more of the protein component by weight.

**[0221]** Skeletal muscle of animals such as beef cattle typically contain substantial quantities of glycogen, which can comprise on the order of 1%> of the mass of the muscle tissue at the time of slaughter. After slaughter, a fraction of this glycogen continues to be metabolized yielding products including lactic acid, which contributes to lowering the pH of the muscle tissue, a desirable quality in meat. Glycogen is a branched polymer of glucose linked together by alpha (1→4) glycosidic bonds in linear chains, with branch points comprising alpha (1→6) glycosidic bonds. Starches from plants, particularly amylopectins are also branched polymers of glucose linked together by alpha (1→4) glycosidic bonds in linear chains, with branch points comprising alpha (1→6) glycosidic bonds and can therefore be used as an analog of glycogen in constructing meat replicas. Thus in some embodiments, the muscle or meat replica includes a starch or pectin.

**[0222]** Additional components of animal muscle tissue include sodium, potassium, calcium, magnesium, other metal ions, lactic acid, other organic acids, free amino acids, peptides, nucleotides and sulfur compounds. Thus in some embodiments, the muscle replica can include sodium, potassium, calcium, magnesium, other metal ions, lactic acid, other organic acids, free amino acids, peptides, nucleotides and sulfur compounds. In some embodiments the concentration of sodium, potassium, calcium, magnesium, other metal ions, lactic acid, other organic acids, free amino acids, peptides, nucleotides and/or sulfur compounds in the muscle replica or consumable are within 10%> of the concentrations found in a muscle or meat being replicated.

**[0223]** In another aspect, the disclosure provides methods for making a muscle replica. In some embodiments, the composition is formed into asymmetric fibers prior to incorporation into the consumable. In some embodiments these fibers replicate muscle fibers. In some embodiments the fibers are spun fibers. In other embodiments the fibers are extruded fibers. Accordingly, the present disclosure provides for methods for producing asymmetric or spun protein fibers. In some embodiments, the fibers are formed by extrusion of the protein component through an extruder.

**[0224]** In some embodiments extrusion can be conducted using an MPF19 twin-screw extruder (APV Baker, Grand Rapids, Mich.) with a cooling die. The cooling die can cool the extrudate prior to return of the extrudate to atmospheric pressure, thus substantially inhibiting expansion or puffing of the final product. In the MPF19 apparatus, dry feed and liquid can be added separately and mixed in the barrel. Extrusion parameters can be, for example: screw speed of 200 rpm, product temperature at the die of 150 C, feed rate of 23 g/min, and water-flow rate of 11 g/min. Product temperature can be measured during extrusion by a thermocouple at the end of the extrusion barrel. Observations can be made on color, opacity, structure, and texture for each collected sample. Collected samples can be optionally dried at room temperature overnight, then ground to a fine powder (<60 mesh) using a Braun food grinder. The pH of samples can be measured in duplicate using 10% (w/v) slurries of powdered sample in distilled water.

**[0225] Fat Replica**

**[0226]** Animal fat is important for the experience of eating cooked meat. Accordingly, the present disclosure provides a composition derived from non-animal sources which reca-

pitulates key features of animal fat. In another aspect, the present disclosure provides a meat substitute product that comprises a composition derived from non-animal sources which recapitulates animal fat. Such a composition will be labeled herein as a “fat replica”. In some embodiments, the fat replica and/or meat substitute product comprising the fat replica are partially derived from animal sources.

**[0227]** In some embodiments the meat substitute product has a fat component. In some embodiments the fat content of the consumable is 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, or 60%) fat. In some embodiments, the fat replica comprises a gel with droplets of fat suspended therein. In some embodiments, the gel is a soft, elastic gel comprising proteins and optionally carbohydrates. In particular embodiments, the proteins used in the gel are plant or microbial proteins. In some embodiments, the proteins used in the fat replica might include Rubisco, any major seed storage proteins, proteins isolated from fungi, bacteria, archaea, viruses, or genetically engineered microorganisms, or synthesized in vitro. The proteins may be artificially designed to emulate physical properties of animal fat. The proteins may be artificially designed to emulate physical properties of animal fat.

**[0228]** The fat droplets used in some embodiments of the present disclosure can be from a variety of sources. In some embodiments, the sources are non-animal sources. In particular embodiments, the sources are plant sources. Non-limiting examples of oils include corn oil, olive oil, soy oil, peanut oil, walnut oil, almond oil, sesame oil, cottonseed oil, rapeseed oil, canola oil, safflower oil, sunflower oil, flax seed oil, algal oil, palm oil, palm kernel oil, coconut oil, babassu oil, shea butter, mango butter, cocoa butter, wheat germ oil, rice bran oil, oils produced by bacteria, algae, archaea or fungi or genetically engineered bacteria, algae, archaea or fungi, triglycerides, monoglycerides, diglycerides, sphingosides, glycolipids, lecithin, lysolecithin, phosphatidic acids, lysophosphatidic acids, oleic acid, palmitoleic acid, palmitic acid, myristic acid, lauric acid, myristoleic acid, caproic acid, capric acid, caprylic acid, pelargonic acid, undecanoic acid, linoleic acid, 20:1 eicosanoic acid, arachidonic acid, eicosapentanoic acid, docosohexanoic acid, 18:2 conjugated linoleic acid, conjugated oleic acid, or esters of: oleic acid, palmitoleic acid, palmitic acid, myristic acid, lauric acid, myristoleic acid, caproic acid, capric acid, caprylic acid, pelargonic acid, undecanoic acid, linoleic acid, 20:1 eicosanoic acid, arachidonic acid, eicosapentanoic acid, docosohexanoic acid, 18:2 conjugated linoleic acid, or conjugated oleic acid, or glycerol esters of oleic acid, palmitoleic acid, palmitic acid, myristic acid, lauric acid, myristoleic acid, caproic acid, capric acid, caprylic acid, pelargonic acid, undecanoic acid, linoleic acid, 20:1 eicosanoic acid, arachidonic acid, eicosapentanoic acid, docosohexanoic acid, 18:2 conjugated linoleic acid, or conjugated oleic acid, or triglyceride derivatives of oleic acid, palmitoleic acid, palmitic acid, myristic acid, lauric acid, myristoleic acid, caproic acid, capric acid, caprylic acid, pelargonic acid, undecanoic acid, linoleic acid, 20:1 eicosanoic acid, arachidonic acid, eicosapentanoic acid, docosohexanoic acid, 18:2 conjugated linoleic acid, or conjugated oleic acid.

**[0229]** In some embodiments, fat droplets are derived from pulp or seed oil. In other embodiments, the source may

be yeast or mold. For instance, in one embodiment the fat droplets comprise triglycerides derived from *Mortierella isabellina*.

**[0230]** In some embodiments plant oils are modified to resemble animal fats. The plant oils can be modified with flavoring or other agents to recapitulate the taste and smell of meat during and after cooking. Accordingly, some aspects of the disclosure involve methods for testing the qualitative similarity between the cooking properties of animal fat and the cooking properties of plant oils in the consumable.

**[0231]** In some embodiments, the fat replica comprises a protein component comprising one or more isolated, purified proteins. The purified proteins contribute to the taste and texture of the meat replica. In some embodiments purified proteins can stabilize emulsified fats. In some embodiments the purified proteins can form gels upon denaturation or enzymatic crosslinking, which replicate the appearance and texture of animal fat. Examples and embodiments of the one or more isolated, purified proteins are described herein. In particular embodiments, the one or more isolated proteins comprise a protein isolated from the legume family of plants. Non-limiting examples of legume plants are described herein, although variations with other legumes are possible. In some embodiments, the legume plant is a pea plant. In some embodiments the isolated purified proteins stabilize emulsions. In some embodiments the isolated purified proteins form gels upon crosslinking or enzymatic crosslinking. In some embodiments, the isolated, purified proteins comprise seed storage proteins. In some embodiments, the isolated, purified proteins comprise albumin. In some embodiments, the isolated, purified proteins comprise globulin. In a particular embodiment, the isolated, purified protein is a purified pea albumin protein. In another particular embodiment, the isolated, purified protein is a purified pea globulin protein. In another particular embodiment the isolate purified protein is a Moong bean 8S globulin. In another particular embodiment, the isolated, purified protein is an oleosin. In another particular embodiment, the isolated, purified protein is a caloleosin. In another particular embodiment, the isolated, purified protein is Rubisco. In some embodiments, the protein component comprises about 0.1%, 0.5%, 1%, 2%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% or more of the fat replica by dry weight or total weight. In some embodiments, the protein component comprises about 0.1-5%, about 0.5-10%, about 1-20%, about 5-30%, about 10-50%, about 20-70%, or about 30-90% or more of the fat replica by dry weight or total weight. In some embodiments, the protein component comprises a solution containing one or more isolated, purified proteins.

**[0232]** In some embodiments, the fat replica comprises cross-linking enzymes that catalyze reactions leading to covalent crosslinks between proteins. Cross-linking enzymes can be used to create or stabilize the desired structure and texture of the adipose tissue replica, to mimic the desired texture of an equivalent desired animal fat. Non-limiting examples of cross-linking enzymes include, e.g., transglutaminase, lysyl oxidases, or other amine oxidases (e.g. *Pichia pastoris* lysyl oxidase). In some embodiments, the cross-linking enzymes are isolated and purified from a non-animal source, examples and embodiments of which are described herein. In some embodiments, the fat replica comprises at least 0.0001%, or at least 0.001%, or at least 0.01%, or at least 0.1%, or at least 1% (wt/vol) of a

cross-linking enzyme. In particular embodiments, the cross-linking enzyme is transglutaminase.

**[0233]** In another aspect, the disclosure provides methods for making a fat replica. In some embodiments, the fat droplets are suspended in a gel. In some embodiments the present disclosure provides for methods for producing droplets of fat suspended in the gel. The fat can be isolated and homogenized. For example, an organic solvent mixture can be used to help mix a lipid. The solvent can then be removed. At this point the lipid can be frozen, lyophilized, or stored. So in some aspects the disclosure provides for a method for isolating and storing a lipid which has been selected to have characteristics similar to animal fat. The lipid film or cake can then be hydrated. The hydration can utilize agitation or temperature changes. The hydration can occur in a precursor solution to a gel. After hydration the lipid suspension can be sonicated or extruded to further alter the properties of the lipid in the solution.

**[0234]** In some embodiments, the fat replica is assembled to approximate the organization adipose tissue in meat. In some embodiments some or all of the components of the fat replica are suspended in a gel. In various embodiments the gel can be a proteinaceous gel, a hydrogel, an organogel, or a xerogel. In some embodiments, the gel can be thickened to a desired consistency using an agent based on polysaccharides or proteins. For example fecula, arrowroot, cornstarch, katakuri starch, potato starch, sago, tapioca, alginin, guar gum, locust bean gum, xanthan gum, collagen, egg whites, furcellaran, gelatin, agar, carrageenan, cellulose, methylcellulose, hydroxymethylcellulose, acacia gum, konjac, starch, pectin, amylopectin or proteins derived from legumes, grains, nuts, other seeds, leaves, algae, bacteria, of fungi can be used alone or in combination to thicken the gel, forming an architecture or structure for the consumable.

**[0235]** In particular embodiments, the fat replica is an emulsion comprising a solution of one or more proteins and one or more fats suspended therein as droplets. In some embodiments, the emulsion is stabilized by one or more cross-linking enzymes into a gel. In some embodiments, the one or more proteins in solution are isolated, purified proteins. In some embodiments, the isolated, purified proteins comprise a purified pea albumin enriched fraction. In some embodiments, the isolated, purified proteins comprise a purified pea globulin enriched fraction. In some embodiments, the isolated, purified proteins comprise a purified Moong bean 8S globulin enriched fraction. In some embodiments, the isolated, purified proteins comprise a Rubisco enriched fraction. In some embodiments, the one or more fats are derived from plant-based oils. In some embodiments, the one or more fats are derived from one or more of: corn oil, olive oil, soy oil, peanut oil, walnut oil, almond oil, sesame oil, cottonseed oil, rapeseed oil, canola oil, safflower oil, sunflower oil, flax seed oil, algal oil, palm oil, palm kernel oil, coconut oil, babassu oil, shea butter, mango butter, cocoa butter, wheat germ oil, rice bran oil, oils produced by bacteria, algae, archaea or fungi or genetically engineered bacteria, algae, archaea or fungi, triglycerides, monoglycerides, diglycerides, sphingosides, glycolipids, lecithin, lysolecithin, phosphatidic acids, lysophosphatidic acids, oleic acid, palmitoleic acid, palmitic acid, myristic acid, lauric acid, myristoleic acid, caproic acid, capric acid, caprylic acid, pelargonic acid, undecanoic acid, linoleic acid, 20: 1 eicosanoic acid, arachidonic acid, eicosapentanoic acid, docosohexanoic acid, 18:2 conjugated linoleic

acid, conjugated oleic acid, or esters of: oleic acid, palmitoleic acid, palmitic acid, myristic acid, lauric acid, myristoleic acid, caproic acid, capric acid, caprylic acid, pelargonic acid, undecanoic acid, linoleic acid, 20: 1 eicosanoic acid, arachidonic acid, eicosapentanoic acid, docosohexanoic acid, 18:2 conjugated linoleic acid, or conjugated oleic acid, or glycerol esters of oleic acid, palmitoleic acid, palmitic acid, myristic acid, lauric acid, myristoleic acid, caproic acid, capric acid, caprylic acid, pelargonic acid, undecanoic acid, linoleic acid, 20: 1 eicosanoic acid, arachidonic acid, eicosapentanoic acid, docosohexanoic acid, 18:2 conjugated linoleic acid, or conjugated oleic acid, or triglyceride derivatives of oleic acid, palmitoleic acid, palmitic acid, myristic acid, lauric acid, myristoleic acid, caproic acid, capric acid, caprylic acid, pelargonic acid, undecanoic acid, linoleic acid, 20: 1 eicosanoic acid, arachidonic acid, eicosapentanoic acid, docosohexanoic acid, 18:2 conjugated linoleic acid, or conjugated oleic acid. In yet even more particular embodiments, the one or more fats is a rice bran oil. In another particular embodiment, the one or more fats is a canola oil. In some embodiments, the cross-linking enzyme is transglutaminase, lysyl oxidase, or other amine oxidase. In some embodiments, the cross-linking enzyme is transglutaminase. In particular embodiments, the fat replica is a high fat emulsion comprising a protein solution of purified pea albumin emulsified with 40-80% rice bran oil, stabilized with 0.5-5% (wt/vol) transglutaminase into a gel. In some embodiments, the fat replica is a high fat emulsion comprising a protein solution of partially-purified moong bean 8S globulin emulsified with 40-80% rice bran oil, stabilized with 0.5-5% (wt/vol) transglutaminase into a gel. In some embodiments, the fat replica is a high fat emulsion comprising a protein solution of partially-purified moong bean 8S globulin emulsified with 40-80% canola oil, stabilized with 0.5-5% (wt/vol) transglutaminase into a gel. In some embodiments, the fat replica is a high fat emulsion comprising a protein solution of purified pea albumin emulsified with 40-80% rice bran oil, stabilized with 0.0001-1% (wt/vol) transglutaminase into a gel. In some embodiments, the fat replica is a high fat emulsion comprising a protein solution of partially-purified moong bean 8S globulin emulsified with 40-80% canola oil, stabilized with 0.0001-1% (wt/vol) transglutaminase into a gel.

#### **[0236]** Connective Tissue Replica

**[0237]** Animal connective tissue provides key textural features that are an important component of the experience of eating meat. Accordingly, the present disclosure provides a composition derived from non-animal sources which recapitulates key features of animal connective tissue. In another aspect, the present disclosure provides a meat substitute product that comprises a composition derived from non-animal sources which recapitulates important textural and visual features of animal connective tissue. Such a composition will be labeled herein as "connective tissue replica". In some embodiments, the connective tissue replica and/or meat substitute product comprising the connective tissue replica are partially derived from animal sources.

**[0238]** Animal connective tissue can generally be divided into fascia-type and cartilage-type tissue. Fascia-type tissue is highly fibrous, resistant against extension (has high elastic

modulus), and has a high protein content, a moderate water content (ca. 50%), and low-to-none fat and polysaccharide content. Accordingly, the present disclosure provides a connective tissue replica that recapitulates key features of fascia type tissue. In some embodiments, the connective tissue replica comprises about 50% protein by total weight, about 50% by liquid weight, and has a low fat and polysaccharide component.

**[0239]** The protein content of most fascia-type connective tissue is comprised mainly of collagen. Collagen is characterized by a high fraction of proline and alanine, and also is assembled into characteristic elongated fibrils or rod-like, flexible structures. Prolamins are one family of proteins found in non-animal sources, such as plant sources. Prolamins are highly abundant in plants and are similar in amino acid composition to collagen. Among proteins we tested for this purpose, prolamins were particularly favorable because of their low cost and their ability to readily form fibers or sheets when spun or extruded. Non-limiting examples of prolamins family proteins include, e.g., zein (found in corn), these include hordein from barley, gliadin from wheat, secalin, extensins from rye, kafirin from sorghum, avenin from oats. In fascia-type connective tissue, the prolamins family of proteins, individually or combinations thereof, demonstrates suitability for the protein component because they are highly abundant, similar in global amino acid composition to collagen (high fraction of proline and alanine), and amenable to processing into films and fibers. In addition to zein (found in corn), these include hordein from barley, gliadin from wheat, secalin, extensins from rye, kafirin from sorghum, avenin from oats. Other proteins may be necessary to supplement prolamins in order to achieve targets specifications for physicochemical and nutritional properties. The list of potential candidates here is essentially open and may include Rubisco, any major seed storage proteins, proteins isolated from fungi, bacteria, archaea, viruses, or genetically engineered microorganisms, or synthesized in vitro. The proteins may be artificially designed to emulate physical properties of animal connective tissue, animal-derived or recombinant collagen, extensins (hydroxyproline-rich glycoproteins abundant in cell walls e.g. *Arabidopsis thaliana*, monomers of which are “collagen-like” rod-like flexible molecules). The proteins may be artificially designed to emulate physical properties of animal connective tissue.

**[0240]** Methods for forming fascia-type connective tissue will be as those practiced in the art with a bias towards methods producing fibrous or fibrous-like structures by biological, chemical, or physical means, individually or in combination, serially or in parallel, before final forming. These methods may include extrusion or spinning.

**[0241]** Cartilage-type tissue can be macroscopically homogenous, resistant against compression, has higher water content (up to 80%), lower protein (collagen) content, and higher polysaccharide (proteoglycans) contents (ca. 10% each).

**[0242]** Compositionally, cartilage-type connective tissue can be very similar to fascia-type tissue with the relative ratios of each adjusted to more closely mimic ‘meat’ connective tissue.

**[0243]** Methods for forming cartilage-type connective tissue can be similar to those for fascia-type connective tissue, but with a bias towards methods producing isotropically homogenous structures.

**[0244]** The fat can be suspended in a gel. In some embodiments the present disclosure provides for methods for producing droplets of fat suspended in the proteinaceous gel. The fat can be isolated from plant tissues and emulsified. The emulsification can utilize high-speed blending, homogenization, agitation or temperature changes. The lipid suspension can be sonicated or extruded to further alter the properties of the lipid in the solution. At this point, in some embodiments other components of the consumable are added to the solution followed by a gelling agent. In some embodiments crosslinking agents (e.g. transglutaminase or lysyl oxidase) are added to bind the components of the consumable. In other embodiments the gelling agent is added and the lipid/gel suspension is later combined with additional components of the consumable. In fascia-type connective tissue, the prolamins family of proteins, individually or combinations thereof, demonstrates suitability for the protein component because they are highly abundant, similar in global amino acid composition to collagen (high fraction of proline and alanine), and amenable to processing into films. In addition to zein (found in corn), these include hordein from barley, gliadin from wheat, secalin, extensins from rye, kafirin from sorghum, avenin from oats. Other proteins may be necessary to supplement prolamins in order to achieve targets specifications for physicochemical and nutritional properties. The list of potential candidates here is essentially open and may include any major seed storage proteins, animal-derived or recombinant collagen, extensins (hydroxyproline-rich glycoproteins abundant in cell walls e.g. *Arabidopsis thaliana*, monomers of which are “collagen-like” rod-like flexible molecules).

**[0245]** In some embodiments some or all of the components of the consumable are suspended in a gel. In various embodiments the gel can be a hydrogel, an organogel, or a xerogel. The gel can be made thick using an agent based on polysaccharides or proteins. For example fucella, arrowroot, cornstarch, katakuri starch, potato starch, sago, tapioca, alginin, guar gum, locust bean gum, xanthan gum, collagen, egg whites, furcellaran, gelatin, agar, carrageenan, cellulose, methylcellulose, hydroxymethylcellulose, acadia gum, konjac, starch, pectin, amylopectin or proteins derived from legumes, grains, nuts, other seeds, leaves, algae, bacteria, of fungi can be used alone or in combination to thicken the gel, forming an architecture or structure for the consumable. Enzymes that catalyze reactions leading to covalent crosslinks between proteins can also be used alone or in combination to form an architecture or structure for the consumable. For example, transglutaminase, lysyl oxidases, or other amine oxidases (e.g. *Pichia pastoris* lysyl oxidase (PPL0)) can be used alone or in combination to form an architecture or structure for the consumable. In some embodiments multiple gels with different components are combined to form the consumable. For example, a gel containing a plant-based protein can be associated with a gel containing a plant-based fat. In some embodiments fibers or stings of proteins are oriented parallel to one another and then held in place by the application of a gel containing plant based fats.

**[0246]** The compositions of the disclosure can be puffed or expanded by heating, such as frying, baking, microwave heating, heating in a forced air system, heating in an air tunnel, and the like.

**[0247]** In some embodiments multiple gels with different components are combined to form the consumable. For example, a gel containing a plant-based protein can be

associated with a gel containing a plant-based fat. In some embodiments fibers or strings of proteins are oriented parallel to one another and then held in place by the application of a gel containing plant based fats.

**[0248]** In some embodiments the meat replica contains no animal products, less than 1% wheat gluten, no methylcellulose, no carrageenan, no caramel color and no Konjac flour, no gum Arabic, and no acacia gum. In some embodiments the meat replica contains no animal products, no wheat gluten, no methylcellulose, no carrageenan, no caramel color and no Konjac flour, no gum Arabic, and no acacia gum. In some embodiments the meat replica contains no animal products, no soy protein isolate, no wheat gluten, no methylcellulose, no carrageenan, no caramel color and no Konjac flour, no gum Arabic, and no acacia gum. In some embodiments the meat replica contains no animal products, no soy protein concentrate, no wheat gluten, no methylcellulose, no carrageenan, no caramel color and no Konjac flour, no gum Arabic, and no acacia gum. In some embodiments the meat replica contains no animal products, no soy protein, no wheat gluten, no methylcellulose, no carrageenan, no caramel color and no Konjac flour, no gum Arabic, and no acacia gum. In some embodiments the meat replica contains no animal products, no tofu, no wheat gluten, no methylcellulose, no carrageenan, no caramel color and no Konjac flour, no gum Arabic, and no acacia gum. In some embodiments the meat replica contains no animal products, no tofu, and no wheat gluten. In some embodiments the meat replica contains no animal products, no soy protein, and no wheat gluten. In some embodiments the meat replica contains no methylcellulose, no carrageenan, no caramel color, no Konjac flour, no gum Arabic, and no acacia gum. In some embodiments the meat replica contains no animal products and less than 5% carbohydrates.

**[0249]** In some embodiments the meat replica contains no animal products, no soy protein, no wheat gluten, no methylcellulose, no carrageenan, no caramel color and no Konjac flour, no gum Arabic, and no acacia gum and less than 5% carbohydrates. In some embodiments the meat replica contains no animal products, and less than 1% cellulose. In some embodiments the meat replica contains no animal products, and less than 5% insoluble carbohydrates. In some embodiments the meat replica contains no animal products, no soy protein, and less than 1% cellulose. In some embodiments the meat replica contains no animal products, no soy protein, and less than 5% insoluble carbohydrates. In some embodiments the meat replica contains no animal products, no wheat gluten, and less than 1% cellulose. In some embodiments the meat replica contains no animal products, no wheat gluten, and less than 5% insoluble carbohydrates.

**[0250]** The percentage of different components may also be controlled. For example, non-animal-based substitutes for muscle, fat tissue, connective tissue, and blood components can be combined in different ratios and physical organizations to best approximate the look and feel of meat. The various components can be arranged to insure consistency between bites of the consumable. The components can be arranged to insure that no waste is generated from the consumable. For example, while a traditional cut of meat may have portions that are not typically eaten, a meat replicate can improve upon meat by not including these inedible portions. Such an improvement allows for all of the product made or shipped to be consumed, which cuts down on waste and shipping costs. Alternatively, a meat replica

may include inedible portions to mimic the experience of meat consumption. Such portions can include bone, cartilage, connective tissue, or other materials commonly referred to as gristle, or materials included simulating these components. In some embodiments the consumable may contain simulated inedible portions of meat products which are designed to serve secondary functions. For example, a simulated bone can be designed to disperse heat during cooking, making the cooking of the consumable faster or more uniform than meat. In other embodiments a simulated bone may also serve to keep the consumable at a constant temperature during shipping. In other embodiments, the simulated inedible portions may be biodegradable.

**[0251]** In some embodiments the meat substitute compositions contain no animal protein, comprising between 10-30% protein, between 5-80% water, between 5-70% fat, comprising one or more isolated purified proteins. In particular embodiments, the meat substitute compositions comprise transglutaminase. In some embodiments the consumable contains components to replicate the components of meat. The main component of meat is typically skeletal muscle. Skeletal muscle typically consists of roughly 75 percent water, 19 percent protein, 2.5 percent intramuscular fat, 1.2 percent carbohydrates and 2.3 percent other soluble non-protein substances. These include organic acids, sulfur compounds, nitrogenous compounds, such as amino acids and nucleotides, and inorganic substances such as minerals.

**[0252]** Accordingly, some embodiments of the present disclosure provide for replicating approximations of this composition for the consumable. For example, in some embodiments the consumable is a plant-based meat replica can comprise roughly 75% water, 19% protein, 2.5% fat, 1.2% carbohydrates; and 2.3 percent other soluble non-protein substances. In some embodiments the consumable is a plant-based meat replica comprising between 60-90% water, 10-30% protein, 1-20% fat, 0.1-5% carbohydrates; and 1-10 percent other soluble non-protein substances. In some embodiments the consumable is a plant-based meat replica comprising between 60-90% water, 5-10% protein, 1-20% fat, 0.1-5% carbohydrates; and 1-10 percent other soluble non-protein substances. In some embodiments the consumable is a plant-based meat replica comprising between 0-50% water, 5-30% protein, 20-80% fat, 0.1-5% carbohydrates; and 1-10 percent other soluble non-protein substances. In some embodiments, the replica contains between 0.01% and 5% by weight of a heme protein. In some embodiments, the replica contains between 0.01% and 5% by weight of leghemoglobin. Some meat also contains myoglobin, a heme protein, which accounts for most of the red color and iron content of some meat. In some embodiments, the replica contains between 0.01% and 5% by weight of a heme protein. In some embodiments, the replica contains between 0.01% and 5% by weight of leghemoglobin. It is understood that these percentages can vary in meat and the meat replicas can be produced to approximate the natural variation in meat. Additionally, in some instances, the present disclosure provides for improved meat replicas, which comprise these components in typically unnatural percentages. For example, a meat replica can be produced with a higher than typical average fat content. The percentages of these components may also be altered to increase other desirable properties.

**[0253]** In some instances, a meat replica is designed so that, when cooked, the percentages of components are

similar to cooked meat. So, in some embodiments, the uncooked consumable has different percentages of components than uncooked meat, but when cooked the consumable is similar to cooked meat. For example, a meat replica may be made with a higher than typical water content for raw meat, but when cooked in a microwave the resulting product has percentages of components similar to meat cooked over a fire.

**[0254]** In some embodiments the consumable is a meat replica with a lower than typical water content for meat. In some embodiments the disclosures provide for methods for hydrating a meat replica to cause the meat replica to have a content similar to meat. For example, a meat replica with a water content that would be low for meat, for example 1%, 10%, 20%, 30%, 40% or 50% water, is hydrated to roughly 75% water. Once hydrated, in some embodiments, the meat replica is then cooked for human consumption.

**[0255]** While preferred embodiments of the present have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the present disclosure. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the disclosure and that methods and structures within the scope of these claims and their equivalents be covered thereby.

## EXAMPLES

### Example 1

#### Method of Producing and Characterizing a Heme-Containing Polypeptide

**[0256]** A plasmid was produced that comprised the polynucleotide sequence from *Aquifex aeolicus* encoding hemo-globin (AaHb), where the nucleotide sequence was codon-optimized for *E. coli*. The plasmid was sub-cloned into pBE-S expression vector provided in the *B. subtilis* Secretory Protein Expression System (Takara Bio). This vector contained the aprE promoter sequence to promote constitutive expression of AaHb in *B. subtilis* and a C-terminal 6-histidine tag. A polynucleotide encoding the secretion signal peptide from the Twin arginine translocation (Tat)-dependent *B. subtilis* protein PhoD was synthesized and cloned in frame at the 5' end of AaHb, replacing the aprE secretion signal peptide within the pBE-S backbone. To generate a cytosolic AaHb expression construct, the aprE secretion signal peptide was deleted from the 5' end of the AaHb open reading frame using inverse PCR followed by ligation.

**[0257]** The expression plasmids were transformed into *B. subtilis* strain RIK1285 AaHb expression was monitored by growing transformed strains in LB media, 10 µg/ml kanamycin, 0.1 mM FeCl<sub>3</sub>, and 20 µg/ml d-aminolevulinic acid. Expression was carried out at 37° C., with shaking at 200 RPM for 24 hours. After expression, the culture was collected and the secreted polypeptide was separated from the bacteria.

**[0258]** Cytosolic and secreted expression of AaHb was monitored by Ni-NTA affinity purification of the cell pellet and media supernatant followed by SDS-PAGE and coo-

massic staining of the elution fractions. Heme loading was monitored by UV-vis analysis of purified AaHb-containing fractions.

**[0259]** As shown in FIG. 1, cytosolic expression of AaHb in *B. subtilis* was compared with (FIG. 1C) and without (FIG. 1B) a secretion signal peptide. The PhoD secretion peptide did not disrupt cytosolic expression of the AaHb polypeptide. Cytosolic AaHb was tested for heme content using UV-Vis spectroscopy. As shown in FIG. 2, addition of the PhoD signal peptide did not interfere with AaHb heme binding in the cytosol.

**[0260]** As shown in FIG. 3, when the AaHb polypeptide was fused to the PhoD secretion peptide, the fusion protein was effectively secreted outside of the host cell (FIG. 3B, lanes A, B). Additionally, it was shown that the PhoD secretion peptide was properly cleaved since both forms of the polypeptide (cleaved and uncleaved) were localized in the cell pellet fraction (e.g., inside the host cell) (FIG. 3B lane "cell pellet"). However, only the cleaved version of the AaHb fusion polypeptide was secreted, which indicated proper function of the PhoD signal peptide.

**[0261]** In support of the effectiveness of the PhoD signal peptide, N-terminal protein sequencing was performed on the secreted polypeptide. FIG. 4 depicts the protein sequence of the PhoD-AaHb fusion protein sequence (which also comprises a His6 tag). A protein band was removed from the gel for N-terminal protein sequencing analysis. This band corresponded to the secreted protein because of its abundance and size on the gel. The N-terminal sequencing results indicated that the PhoD peptide was indeed cleaved (e.g., the size was not due to non-specific degradation) at the correct site because the N-terminus of the sequencing results matched the predicted cut site of the protease.

**[0262]** The secreted AaHb was further characterized for heme content by monitoring the UV-vis absorbance of AaHb purified from the media. FIG. 5 illustrates the effects of the PhoD signal peptide on the heme content of the secreted AaHb. The secreted AaHb was heme bound, as evidenced by a noticeable absorption peak at approximately 415 nm.

**[0263]** Taken together, these results suggest that the PhoD signal peptide does not interfere with cytosolic secretion of a polypeptide, is properly cleaved, enhances secretion of a polypeptide, and maintains the heme content of the secreted polypeptide.

### Prophetic Example 1

#### Method of Using a Polypeptide in a Meat Consumable

**[0264]** In some instances, a polypeptide in this disclosure will be expressed and purified as described in Example 1.

**[0265]** In some instances, a muscle tissue analog will be constructed with the polypeptide and pea vicilin protein using transglutaminase cross-linking.

**[0266]** In some instances, a muscle tissue analog will be constructed by heat/cool gel formation of purified pea vicilin proteins. The heme-containing polypeptide will be thoroughly mixed with the partially gelled muscle tissue upon cooling to room temperature.

**[0267]** In some instances, a muscle tissue analog will be constructed by co-extruding heme-containing polypeptide with purified pea vicilin proteins.

**[0268]** In some instances, a fat tissue analog will be constructed by emulsifying pea albumin proteins, coconut

oil, and lecithin through high-pressure homogenization followed by a heat/cool treatment. Heme-containing polypeptide will be thoroughly mixed with the partially gelled adipose tissue upon cooling to room temperature.

[0269] In some instances, a connective tissue analogue will be prepared with a zein protein source by extrusion or electrospinning.

[0270] In some instances, a ground beef replica (e.g., meat consumable) will be prepared by combining the muscle analog comprising a purified polypeptide of the disclosure with varying amounts of the fat tissue analog and the connective tissue analog. In some instances, the different tissues may be combined using a meat grinder. The resulting meat consumable can be cooked before eating. The cooking procedure will induce the red color of the meat consumable to change to a brown color, indicating cooking. In some instances, the red color of the meat consumable is due to the purified heme-containing polypeptide of the disclosure. The cooking procedure will catalyze the release of meat flavors and aromas. In some instances, the flavors and aromas of the meat consumable are due to the purified heme-containing polypeptide of this disclosure.

#### Example 2

##### Method to Modulate Expression and Secretion of an Endogenous Heme-Containing Polypeptide

[0271] Two plasmids were produced, transformed, and cultured as described in Example 1. Both plasmids include the polynucleotide sequence from *B. subtilis* encoding the endogenous truncated hemoglobin gene, *yjbI* (FIG. 9, SEQ ID NO:26). In the case of the first plasmid, a polynucleotide encoding the secretion signal peptide from the Twin arginine translocation (Tat)-dependent *B. subtilis* protein PhoD (Table 1) was synthesized and cloned in frame at the 5' end of the *yjbI* gene (PhoD-*yjbI*). The second plasmid included the signal peptide YwbN (Table 1), which was also synthesized and cloned in frame at the 5' end of the *yjbI* gene (YwbN-*yjbI*), as described in Example 1.

[0272] As shown in FIG. 6, after 24 hours of growth, cytosolic and secreted expression of *yjbI* was monitored. The cell pellet and media fractions were separated by SDS-PAGE, transferred onto a PVDF membrane, and probed with an anti-His6 antibody (Abcam). Specifically, the figure shows the detection of the fused polypeptides (PhoD-*yjbI* and YwbN-*yjbI*) in the cell pellet following expression of the endogenous polypeptide from the exogenous nucleic acid, and media detection of the polypeptide, which indicated proper cleavage of the signal peptide.

#### Example 3

##### Method to Modulate Expression and Secretion of a Heterologous Heme-Containing Polypeptide

[0273] Two plasmids were produced, transformed, and cultured as described in Example 1. The first plasmid included the polynucleotide sequence from *Glycine max* (soybean) encoding LGB2 (FIG. 9, SEQ ID NO:4). The second plasmid included the protein coding polynucleotide sequence from *M. inferorum*, HGbI (FIG. 9, SEQ ID NO: 2). In both plasmids, a polynucleotide encoding the secretion signal peptide from the Twin arginine translocation (Tat)-dependent *B. subtilis* protein PhoD (Table 1) was synthesized and cloned in frame at the 5' end of each polynucleotide sequence (PhoD-LGB2 and PhoD-HGbI).

[0274] As shown in FIG. 7, after 24 hours of growth, cytosolic and secreted expression of the two polypeptides was monitored. The cell pellet and media fractions were separated by SDS-PAGE, transferred onto a PVDF membrane, and probed with an anti-His6 antibody (Abcam). Specifically, the figure shows the detection of the fused polypeptides (PhoD-LGB2 and PhoD-HGbI) in the cell pellet following expression of each heterologous polypeptide, and media detection of each heterologous polypeptide, which indicated proper cleavage of the signal peptide.

#### Example 4 Method of Producing and Characterizing a Heme-Containing Polypeptide Fused to Different Signal Peptides

[0275] A series of plasmids were produced, transformed, and cultured as described in Example 1. Each plasmid included polynucleotide sequence from *Aquifex aeolicus* encoding hemoglobin (AaHb). To the 5' end of the AaHb sequence, a nucleic acid sequence encoding a secretion signaling peptide selected from AbnA, AlbB, AppB, BglS, LipA, OppA, SpoIIIJ, TipA, WapA, WprA, YkpC, YmaC, YolA, YuiC, or YwbN (Table 1) was fused as described previously.

[0276] As shown in FIG. 8, after 24 hours of growth, the media supernatants of a subset of the fusion polypeptides were separated by SDS-PAGE, transferred onto a PVDF membrane, and blotted with an anti-His6 antibody (Abcam). The figure illustrates the media detection of the endogenous polypeptide (AaHb) after fusion with a subset of a number of different exemplary secretion signaling peptides (PhoD, TipA, WapA, WprA, YmaC, YolA, YuiC, YwbN, AppB, and BglS), expression, and cleavage of the secretion signal peptide.

---

#### SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 96

<210> SEQ ID NO 1

<211> LENGTH: 161

<212> TYPE: PRT

<213> ORGANISM: *Vigna radiata*

<400> SEQUENCE: 1

Met Thr Thr Thr Leu Glu Arg Gly Phe Thr Glu Glu Gln Glu Ala Leu  
1 5 10 15



-continued

---

Val Val Lys Ser Trp Asn Val Met Lys Lys Asn Ser Gly Glu Leu Gly  
 20 25 30

Leu Lys Phe Phe Leu Lys Ile Phe Glu Ile Ala Pro Ser Ala Gln Lys  
 35 40 45

Leu Phe Ser Phe Leu Arg Asp Ser Thr Val Pro Leu Glu Gln Asn Pro  
 50 55 60

Lys Leu Lys Pro His Ala Val Ser Val Phe Val Met Thr Cys Asp Ser  
 65 70 75 80

Ala Val Gln Leu Arg Lys Ala Gly Lys Val Thr Val Arg Glu Ser Asn  
 85 90 95

Leu Lys Lys Leu Gly Ala Thr His Phe Arg Thr Gly Val Ala Asn Glu  
 100 105 110

His Phe Glu Val Thr Lys Phe Ala Leu Leu Glu Thr Ile Lys Glu Ala  
 115 120 125

Val Pro Glu Met Trp Ser Pro Ala Met Lys Asn Ala Trp Gly Glu Ala  
 130 135 140

Tyr Asp Gln Leu Val Asp Ala Ile Lys Tyr Glu Met Lys Pro Pro Ser  
 145 150 155 160

Ser

<210> SEQ ID NO 2  
 <211> LENGTH: 133  
 <212> TYPE: PRT  
 <213> ORGANISM: *Methylophilum inferorum*

<400> SEQUENCE: 2

Met Ile Asp Gln Lys Glu Lys Glu Leu Ile Lys Glu Ser Trp Lys Arg  
 1 5 10 15

Ile Glu Pro Asn Lys Asn Glu Ile Gly Leu Leu Phe Tyr Ala Asn Leu  
 20 25 30

Phe Lys Glu Glu Pro Thr Val Ser Val Leu Phe Gln Asn Pro Ile Ser  
 35 40 45

Ser Gln Ser Arg Lys Leu Met Gln Val Leu Gly Ile Leu Val Gln Gly  
 50 55 60

Ile Asp Asn Leu Glu Gly Leu Ile Pro Thr Leu Gln Asp Leu Gly Arg  
 65 70 75 80

Arg His Lys Gln Tyr Gly Val Val Asp Ser His Tyr Pro Leu Val Gly  
 85 90 95

Asp Cys Leu Leu Lys Ser Ile Gln Glu Tyr Leu Gly Gln Gly Phe Thr  
 100 105 110

Glu Glu Ala Lys Ala Ala Trp Thr Lys Val Tyr Gly Ile Ala Ala Gln  
 115 120 125

Val Met Thr Ala Glu  
 130

<210> SEQ ID NO 3  
 <211> LENGTH: 139  
 <212> TYPE: PRT  
 <213> ORGANISM: *Aquifex aeolicus*

<400> SEQUENCE: 3

Met Leu Ser Glu Glu Thr Ile Arg Val Ile Lys Ser Thr Val Pro Leu  
 1 5 10 15

Leu Lys Glu His Gly Thr Glu Ile Thr Ala Arg Met Tyr Glu Leu Leu

-continued

---

	20		25		30														
Phe	Ser	Lys	Tyr	Pro	Lys	Thr	Lys	Glu	Leu	Phe	Ala	Gly	Ala	Ser	Glu				
	35						40					45							
Glu	Gln	Pro	Lys	Lys	Leu	Ala	Asn	Ala	Ile	Ile	Ala	Tyr	Ala	Thr	Tyr				
	50						55				60								
Ile	Asp	Arg	Leu	Glu	Glu	Leu	Asp	Asn	Ala	Ile	Ser	Thr	Ile	Ala	Arg				
	65				70					75					80				
Ser	His	Val	Arg	Arg	Asn	Val	Lys	Pro	Glu	His	Tyr	Pro	Leu	Val	Lys				
				85					90					95					
Glu	Cys	Leu	Leu	Gln	Ala	Ile	Glu	Glu	Val	Leu	Asn	Pro	Gly	Glu	Glu				
		100							105				110						
Val	Leu	Lys	Ala	Trp	Glu	Glu	Ala	Tyr	Asp	Phe	Leu	Ala	Lys	Thr	Leu				
		115					120					125							
Ile	Thr	Leu	Glu	Lys	Lys	Leu	Tyr	Ser	Gln	Pro									
	130					135													

<210> SEQ ID NO 4  
 <211> LENGTH: 145  
 <212> TYPE: PRT  
 <213> ORGANISM: Glycine max

<400> SEQUENCE: 4

Met	Gly	Ala	Phe	Thr	Glu	Lys	Gln	Glu	Ala	Leu	Val	Ser	Ser	Ser	Phe				
1				5					10						15				
Glu	Ala	Phe	Lys	Ala	Asn	Ile	Pro	Gln	Tyr	Ser	Val	Val	Phe	Tyr	Thr				
			20					25					30						
Ser	Ile	Leu	Glu	Lys	Ala	Pro	Ala	Ala	Lys	Asp	Leu	Phe	Ser	Phe	Leu				
		35				40						45							
Ser	Asn	Gly	Val	Asp	Pro	Ser	Asn	Pro	Lys	Leu	Thr	Gly	His	Ala	Glu				
	50					55					60								
Lys	Leu	Phe	Gly	Leu	Val	Arg	Asp	Ser	Ala	Gly	Gln	Leu	Lys	Ala	Asn				
	65				70					75					80				
Gly	Thr	Val	Val	Ala	Asp	Ala	Ala	Leu	Gly	Ser	Ile	His	Ala	Gln	Lys				
				85					90					95					
Ala	Ile	Thr	Asp	Pro	Gln	Phe	Val	Val	Val	Lys	Glu	Ala	Leu	Leu	Lys				
			100						105					110					
Thr	Ile	Lys	Glu	Ala	Val	Gly	Asp	Lys	Trp	Ser	Asp	Glu	Leu	Ser	Ser				
		115				120						125							
Ala	Trp	Glu	Val	Ala	Tyr	Asp	Glu	Leu	Ala	Ala	Ala	Ile	Lys	Lys	Ala				
	130					135						140							

Phe  
145

<210> SEQ ID NO 5  
 <211> LENGTH: 162  
 <212> TYPE: PRT  
 <213> ORGANISM: Hordeum vulgare

<400> SEQUENCE: 5

Met	Ser	Ala	Ala	Glu	Gly	Ala	Val	Val	Phe	Ser	Glu	Glu	Lys	Glu	Ala				
1				5					10					15					
Leu	Val	Leu	Lys	Ser	Trp	Ala	Ile	Met	Lys	Lys	Asp	Ser	Ala	Asn	Leu				
			20					25					30						
Gly	Leu	Arg	Phe	Phe	Leu	Lys	Ile	Phe	Glu	Ile	Ala	Pro	Ser	Ala	Arg				

-continued

---

35	40	45
Gln Met Phe Pro Phe Leu Arg Asp Ser Asp Val Pro Leu Glu Thr Asn 50	55	60
Pro Lys Leu Lys Thr His Ala Val Ser Val Phe Val Met Thr Cys Glu 65	70	75
Ala Ala Ala Gln Leu Arg Lys Ala Gly Lys Ile Thr Val Arg Glu Thr 85	90	95
Thr Leu Lys Arg Leu Gly Gly Thr His Leu Lys Tyr Gly Val Ala Asp 100	105	110
Gly His Phe Glu Val Thr Arg Phe Ala Leu Leu Glu Thr Ile Lys Glu 115	120	125
Ala Leu Pro Ala Asp Met Trp Gly Pro Glu Met Arg Asn Ala Trp Gly 130	135	140
Glu Ala Tyr Asp Gln Leu Val Ala Ala Ile Lys Gln Glu Met Lys Pro 145	150	155
Ala Glu		160

<210> SEQ ID NO 6  
 <211> LENGTH: 1153  
 <212> TYPE: PRT  
 <213> ORGANISM: Magnaporthe oryzae

<400> SEQUENCE: 6

Met Asp Gly Ala Val Arg Leu Asp Trp Thr Gly Leu Asp Leu Thr Gly 1	5	10
His Glu Ile His Asp Gly Val Pro Ile Ala Ser Arg Val Gln Val Met 20	25	30
Val Ser Phe Pro Leu Phe Lys Asp Gln His Ile Ile Met Ser Ser Lys 35	40	45
Glu Ser Pro Ser Arg Lys Ser Ser Thr Ile Gly Gln Ser Thr Arg Asn 50	55	60
Gly Ser Cys Gln Ala Asp Thr Gln Lys Gly Gln Leu Pro Pro Val Gly 65	70	75
Glu Lys Pro Lys Pro Val Lys Glu Asn Pro Met Lys Lys Leu Lys Glu 85	90	95
Met Ser Gln Arg Pro Leu Pro Thr Gln His Gly Asp Gly Thr Tyr Pro 100	105	110
Thr Glu Lys Lys Leu Thr Gly Ile Gly Glu Asp Leu Lys His Ile Arg 115	120	125
Gly Tyr Asp Val Lys Thr Leu Leu Ala Met Val Lys Ser Lys Leu Lys 130	135	140
Gly Glu Lys Leu Lys Asp Asp Lys Thr Met Leu Met Glu Arg Val Met 145	150	155
Gln Leu Val Ala Arg Leu Pro Thr Glu Ser Lys Lys Arg Ala Glu Leu 165	170	175
Thr Asp Ser Leu Ile Asn Glu Leu Trp Glu Ser Leu Asp His Pro Pro 180	185	190
Leu Asn Tyr Leu Gly Pro Glu His Ser Tyr Arg Thr Pro Asp Gly Ser 195	200	205
Tyr Asn His Pro Phe Asn Pro Gln Leu Gly Ala Ala Gly Ser Arg Tyr 210	215	220
Ala Arg Ser Val Ile Pro Thr Val Thr Pro Pro Gly Ala Leu Pro Asp		

-continued

225					230					235					240
Pro	Gly	Leu	Ile	Phe	Asp	Ser	Ile	Met	Gly	Arg	Thr	Pro	Asn	Ser	Tyr
				245					250					255	
Arg	Lys	His	Pro	Asn	Asn	Val	Ser	Ser	Ile	Leu	Trp	Tyr	Trp	Ala	Thr
			260					265					270		
Ile	Ile	Ile	His	Asp	Ile	Phe	Trp	Thr	Asp	Pro	Arg	Asp	Ile	Asn	Thr
		275					280					285			
Asn	Lys	Ser	Ser	Ser	Tyr	Leu	Asp	Leu	Ala	Pro	Leu	Tyr	Gly	Asn	Ser
	290					295					300				
Gln	Glu	Met	Gln	Asp	Ser	Ile	Arg	Thr	Phe	Lys	Asp	Gly	Arg	Met	Lys
305					310					315					320
Pro	Asp	Cys	Tyr	Ala	Asp	Lys	Arg	Leu	Ala	Gly	Met	Pro	Pro	Gly	Val
				325					330					335	
Ser	Val	Leu	Leu	Ile	Met	Phe	Asn	Arg	Phe	His	Asn	His	Val	Ala	Glu
			340					345					350		
Asn	Leu	Ala	Leu	Ile	Asn	Glu	Gly	Gly	Arg	Phe	Asn	Lys	Pro	Ser	Asp
		355					360					365			
Leu	Leu	Glu	Gly	Glu	Ala	Arg	Glu	Ala	Ala	Trp	Lys	Lys	Tyr	Asp	Asn
370						375					380				
Asp	Leu	Phe	Gln	Val	Ala	Arg	Leu	Val	Thr	Ser	Gly	Leu	Tyr	Ile	Asn
385					390					395					400
Ile	Thr	Leu	Val	Asp	Tyr	Val	Arg	Asn	Ile	Val	Asn	Leu	Asn	Arg	Val
				405					410					415	
Asp	Thr	Thr	Trp	Thr	Leu	Asp	Pro	Arg	Gln	Asp	Ala	Gly	Ala	His	Val
			420					425					430		
Gly	Thr	Ala	Asp	Gly	Ala	Glu	Arg	Gly	Thr	Gly	Asn	Ala	Val	Ser	Ala
		435					440					445			
Glu	Phe	Asn	Leu	Cys	Tyr	Arg	Trp	His	Ser	Cys	Ile	Ser	Glu	Lys	Asp
450						455					460				
Ser	Lys	Phe	Val	Glu	Ala	Gln	Phe	Gln	Asn	Ile	Phe	Gly	Lys	Pro	Ala
465				470						475					480
Ser	Glu	Val	Arg	Pro	Asp	Glu	Met	Trp	Lys	Gly	Phe	Ala	Lys	Met	Glu
				485					490					495	
Gln	Asn	Thr	Pro	Ala	Asp	Pro	Gly	Gln	Arg	Thr	Phe	Gly	Gly	Phe	Lys
			500					505					510		
Arg	Gly	Pro	Asp	Gly	Lys	Phe	Asp	Asp	Asp	Asp	Leu	Val	Arg	Cys	Ile
		515					520					525			
Ser	Glu	Ala	Val	Glu	Asp	Val	Ala	Gly	Ala	Phe	Gly	Ala	Arg	Asn	Val
	530					535					540				
Pro	Gln	Ala	Met	Lys	Val	Val	Glu	Thr	Met	Gly	Ile	Ile	Gln	Gly	Arg
545					550					555					560
Lys	Trp	Asn	Val	Ala	Gly	Leu	Asn	Glu	Phe	Arg	Lys	His	Phe	His	Leu
				565					570					575	
Lys	Pro	Tyr	Ser	Thr	Phe	Glu	Asp	Ile	Asn	Ser	Asp	Pro	Gly	Val	Ala
			580					585					590		
Glu	Ala	Leu	Arg	Arg	Leu	Tyr	Asp	His	Pro	Asp	Asn	Val	Glu	Leu	Tyr
		595					600					605			
Pro	Gly	Leu	Val	Ala	Glu	Glu	Asp	Lys	Gln	Pro	Met	Val	Pro	Gly	Val
	610						615					620			
Gly	Ile	Ala	Pro	Thr	Tyr	Thr	Ile	Ser	Arg	Val	Val	Leu	Ser	Asp	Ala
625					630					635					640

-continued

---

Val Cys Leu Val Arg Gly Asp Arg Phe Tyr Thr Thr Asp Phe Thr Pro  
645 650 655

Arg Asn Leu Thr Asn Trp Gly Tyr Lys Glu Val Asp Tyr Asp Leu Ser  
660 665 670

Val Asn His Gly Cys Val Phe Tyr Lys Leu Phe Ile Arg Ala Phe Pro  
675 680 685

Asn His Phe Lys Gln Asn Ser Val Tyr Ala His Tyr Pro Met Val Val  
690 695 700

Pro Ser Glu Asn Lys Arg Ile Leu Glu Ala Leu Gly Arg Ala Asp Leu  
705 710 715 720

Phe Asp Phe Glu Ala Pro Lys Tyr Ile Pro Pro Arg Val Asn Ile Thr  
725 730 735

Ser Tyr Gly Gly Ala Glu Tyr Ile Leu Glu Thr Gln Glu Lys Tyr Lys  
740 745 750

Val Thr Trp His Glu Gly Leu Gly Phe Leu Met Gly Glu Gly Gly Leu  
755 760 765

Lys Phe Met Leu Ser Gly Asp Asp Pro Leu His Ala Gln Gln Arg Lys  
770 775 780

Cys Met Ala Ala Gln Leu Tyr Lys Asp Gly Trp Thr Glu Ala Val Lys  
785 790 795 800

Ala Phe Tyr Ala Gly Met Met Glu Glu Leu Leu Val Ser Lys Ser Tyr  
805 810 815

Phe Leu Gly Asn Asn Lys His Arg His Val Asp Ile Ile Arg Asp Val  
820 825 830

Gly Asn Met Val His Val His Phe Ala Ser Gln Val Phe Gly Leu Pro  
835 840 845

Leu Lys Thr Ala Lys Asn Pro Thr Gly Val Phe Thr Glu Gln Glu Met  
850 855 860

Tyr Gly Ile Leu Ala Ala Ile Phe Thr Thr Ile Phe Phe Asp Leu Asp  
865 870 875 880

Pro Ser Lys Ser Phe Pro Leu Arg Thr Lys Thr Arg Glu Val Cys Gln  
885 890 895

Lys Leu Ala Lys Leu Val Glu Ala Asn Val Lys Leu Ile Asn Lys Ile  
900 905 910

Pro Trp Ser Arg Gly Met Phe Val Gly Lys Pro Ala Lys Asp Glu Pro  
915 920 925

Leu Ser Ile Tyr Gly Lys Thr Met Ile Lys Gly Leu Lys Ala His Gly  
930 935 940

Leu Ser Asp Tyr Asp Ile Ala Trp Ser His Val Val Pro Thr Ser Gly  
945 950 955 960

Ala Met Val Pro Asn Gln Ala Gln Val Phe Ala Gln Ala Val Asp Tyr  
965 970 975

Tyr Leu Ser Pro Ala Gly Met His Tyr Ile Pro Glu Ile His Met Val  
980 985 990

Ala Leu Gln Pro Ser Thr Pro Glu Thr Asp Ala Leu Leu Leu Gly Tyr  
995 1000 1005

Ala Met Glu Gly Ile Arg Leu Ala Gly Thr Phe Gly Ser Tyr Arg Glu  
1010 1015 1020

Ala Ala Val Asp Asp Val Val Lys Glu Asp Asn Gly Arg Gln Val Pro  
1025 1030 1035 1040

-continued

---

Val Lys Ala Gly Asp Arg Val Phe Val Ser Phe Val Asp Ala Ala Arg  
 1045 1050 1055

Asp Pro Lys His Phe Pro Asp Pro Glu Val Val Asn Pro Arg Arg Pro  
 1060 1065 1070

Ala Lys Lys Tyr Ile His Tyr Gly Val Gly Pro His Ala Cys Leu Gly  
 1075 1080 1085

Arg Asp Ala Ser Gln Ile Ala Ile Thr Glu Met Phe Arg Cys Leu Phe  
 1090 1095 1100

Arg Arg Arg Asn Val Arg Arg Val Pro Gly Pro Gln Gly Glu Leu Lys  
 1105 1110 1115 1120

Lys Val Pro Arg Pro Gly Gly Phe Tyr Val Tyr Met Arg Glu Asp Trp  
 1125 1130 1135

Gly Gly Leu Phe Pro Phe Pro Val Thr Met Arg Val Met Trp Asp Asp  
 1140 1145 1150

Glu

<210> SEQ ID NO 7  
 <211> LENGTH: 530  
 <212> TYPE: PRT  
 <213> ORGANISM: Fusarium oxysporum

<400> SEQUENCE: 7

Met Lys Gly Ser Ala Thr Leu Ala Phe Ala Leu Val Gln Phe Ser Ala  
 1 5 10 15

Ala Ser Gln Leu Val Trp Pro Ser Lys Trp Asp Glu Val Glu Asp Leu  
 20 25 30

Leu Tyr Met Gln Gly Gly Phe Asn Lys Arg Gly Phe Ala Asp Ala Leu  
 35 40 45

Arg Thr Cys Glu Phe Gly Ser Asn Val Pro Gly Thr Gln Asn Thr Ala  
 50 55 60

Glu Trp Leu Arg Thr Ala Phe His Asp Ala Ile Thr His Asp Ala Lys  
 65 70 75 80

Ala Gly Thr Gly Gly Leu Asp Ala Ser Ile Tyr Trp Glu Ser Ser Arg  
 85 90 95

Pro Glu Asn Pro Gly Lys Ala Phe Asn Asn Thr Phe Gly Phe Phe Ser  
 100 105 110

Gly Phe His Asn Pro Arg Ala Thr Ala Ser Asp Leu Thr Ala Leu Gly  
 115 120 125

Thr Val Leu Ala Val Gly Ala Cys Asn Gly Pro Arg Ile Pro Phe Arg  
 130 135 140

Ala Gly Arg Ile Asp Ala Tyr Lys Ala Gly Pro Ala Gly Val Pro Glu  
 145 150 155 160

Pro Ser Thr Asn Leu Lys Asp Thr Phe Ala Ala Phe Thr Lys Ala Gly  
 165 170 175

Phe Thr Lys Glu Glu Met Thr Ala Met Val Ala Cys Gly His Ala Ile  
 180 185 190

Gly Gly Val His Ser Val Asp Phe Pro Glu Ile Val Gly Ile Lys Ala  
 195 200 205

Asp Pro Asn Asn Asp Thr Asn Val Pro Phe Gln Lys Asp Val Ser Ser  
 210 215 220

Phe His Asn Gly Ile Val Thr Glu Tyr Leu Ala Gly Thr Ser Lys Asn  
 225 230 235 240

-continued

---

Pro Leu Val Ala Ser Lys Asn Ala Thr Phe His Ser Asp Lys Arg Ile  
 245 250 255

Phe Asp Asn Asp Lys Ala Thr Met Lys Lys Leu Ser Thr Lys Ala Gly  
 260 265 270

Phe Asn Ser Met Cys Ala Asp Ile Leu Thr Arg Met Ile Asp Thr Val  
 275 280 285

Pro Lys Ser Val Gln Leu Thr Pro Val Leu Glu Ala Tyr Asp Val Arg  
 290 295 300

Pro Tyr Ile Thr Glu Leu Ser Leu Asn Asn Lys Asn Lys Ile His Phe  
 305 310 315 320

Thr Gly Ser Val Arg Val Arg Ile Thr Asn Asn Ile Arg Asp Asn Asn  
 325 330 335

Asp Leu Ala Ile Asn Leu Ile Tyr Val Gly Arg Asp Gly Lys Lys Val  
 340 345 350

Thr Val Pro Thr Gln Gln Val Thr Phe Gln Gly Gly Thr Ser Phe Gly  
 355 360 365

Ala Gly Glu Val Phe Ala Asn Phe Glu Phe Asp Thr Thr Met Asp Ala  
 370 375 380

Lys Asn Gly Ile Thr Lys Phe Phe Ile Gln Glu Val Lys Pro Ser Thr  
 385 390 395 400

Lys Ala Thr Val Thr His Asp Asn Gln Lys Thr Gly Gly Tyr Lys Val  
 405 410 415

Asp Asp Thr Val Leu Tyr Gln Leu Gln Gln Ser Cys Ala Val Leu Glu  
 420 425 430

Lys Leu Pro Asn Ala Pro Leu Val Val Thr Ala Met Val Arg Asp Ala  
 435 440 445

Arg Ala Lys Asp Ala Leu Thr Leu Arg Val Ala His Lys Lys Pro Val  
 450 455 460

Lys Gly Ser Ile Val Pro Arg Phe Gln Thr Ala Ile Thr Asn Phe Lys  
 465 470 475 480

Ala Thr Gly Lys Lys Ser Ser Gly Tyr Thr Gly Phe Gln Ala Lys Thr  
 485 490 495

Met Phe Glu Glu Gln Ser Thr Tyr Phe Asp Ile Val Leu Gly Gly Ser  
 500 505 510

Pro Ala Ser Gly Val Gln Phe Leu Thr Ser Gln Ala Met Pro Ser Gln  
 515 520 525

Cys Ser  
 530

<210> SEQ ID NO 8  
 <211> LENGTH: 358  
 <212> TYPE: PRT  
 <213> ORGANISM: Fusarium graminearum

<400> SEQUENCE: 8

Met Ala Ser Ala Thr Arg Gln Phe Ala Arg Ala Ala Thr Arg Ala Thr  
 1 5 10 15

Arg Asn Gly Phe Ala Ile Ala Pro Arg Gln Val Ile Arg Gln Gln Gly  
 20 25 30

Arg Arg Tyr Tyr Ser Ser Glu Pro Ala Gln Lys Ser Ser Ser Ala Trp  
 35 40 45

Ile Trp Leu Thr Gly Ala Ala Val Ala Gly Gly Ala Gly Tyr Tyr Phe  
 50 55 60

-continued

---

Tyr Gly Asn Ser Ala Ser Ser Ala Thr Ala Lys Val Phe Asn Pro Ser  
 65 70 75 80  
 Lys Glu Asp Tyr Gln Lys Val Tyr Asn Glu Ile Ala Ala Arg Leu Glu  
 85 90 95  
 Glu Lys Asp Asp Tyr Asp Asp Gly Ser Tyr Gly Pro Val Leu Val Arg  
 100 105 110  
 Leu Ala Trp His Ala Ser Gly Thr Tyr Asp Lys Glu Thr Gly Thr Gly  
 115 120 125  
 Gly Ser Asn Gly Ala Thr Met Arg Phe Ala Pro Glu Ser Asp His Gly  
 130 135 140  
 Ala Asn Ala Gly Leu Ala Ala Ala Arg Asp Phe Leu Gln Pro Val Lys  
 145 150 155 160  
 Glu Lys Phe Pro Trp Ile Thr Tyr Ser Asp Leu Trp Ile Leu Ala Gly  
 165 170 175  
 Val Cys Ala Ile Gln Glu Met Leu Gly Pro Ala Ile Pro Tyr Arg Pro  
 180 185 190  
 Gly Arg Ser Asp Arg Asp Val Ser Gly Cys Thr Pro Asp Gly Arg Leu  
 195 200 205  
 Pro Asp Ala Ser Lys Arg Gln Asp His Leu Arg Gly Ile Phe Gly Arg  
 210 215 220  
 Met Gly Phe Asn Asp Gln Glu Ile Val Ala Leu Ser Gly Ala His Ala  
 225 230 235 240  
 Leu Gly Arg Cys His Thr Asp Arg Ser Gly Tyr Ser Gly Pro Trp Thr  
 245 250 255  
 Phe Ser Pro Thr Val Leu Thr Asn Asp Tyr Phe Arg Leu Leu Val Glu  
 260 265 270  
 Glu Lys Trp Gln Trp Lys Lys Trp Asn Gly Pro Ala Gln Tyr Glu Asp  
 275 280 285  
 Lys Ser Thr Lys Ser Leu Met Met Leu Pro Ser Asp Ile Ala Leu Ile  
 290 295 300  
 Glu Asp Lys Lys Phe Lys Pro Trp Val Glu Lys Tyr Ala Lys Asp Asn  
 305 310 315 320  
 Asp Ala Phe Phe Lys Asp Phe Ser Asn Val Val Leu Arg Leu Phe Glu  
 325 330 335  
 Leu Gly Val Pro Phe Ala Gln Gly Thr Glu Asn Gln Arg Trp Thr Phe  
 340 345 350  
 Lys Pro Thr His Gln Glu  
 355

<210> SEQ ID NO 9  
 <211> LENGTH: 122  
 <212> TYPE: PRT  
 <213> ORGANISM: Chlamydomonas eugametos

<400> SEQUENCE: 9

Met Ser Leu Phe Ala Lys Leu Gly Gly Arg Glu Ala Val Glu Ala Ala  
 1 5 10 15  
 Val Asp Lys Phe Tyr Asn Lys Ile Val Ala Asp Pro Thr Val Ser Thr  
 20 25 30  
 Tyr Phe Ser Asn Thr Asp Met Lys Val Gln Arg Ser Lys Gln Phe Ala  
 35 40 45  
 Phe Leu Ala Tyr Ala Leu Gly Gly Ala Ser Glu Trp Lys Gly Lys Asp





-continued

115

<210> SEQ ID NO 12  
 <211> LENGTH: 147  
 <212> TYPE: PRT  
 <213> ORGANISM: *Aspergillus niger*

<400> SEQUENCE: 12

Met Pro Leu Thr Pro Glu Gln Ile Lys Ile Ile Lys Ala Thr Val Pro  
 1 5 10 15  
 Val Leu Gln Glu Tyr Gly Thr Lys Ile Thr Thr Ala Phe Tyr Met Asn  
 20 25 30  
 Met Ser Thr Val His Pro Glu Leu Asn Ala Val Phe Asn Thr Ala Asn  
 35 40 45  
 Gln Val Lys Gly His Gln Ala Arg Ala Leu Ala Gly Ala Leu Phe Ala  
 50 55 60  
 Tyr Ala Ser His Ile Asp Asp Leu Gly Ala Leu Gly Pro Ala Val Glu  
 65 70 75 80  
 Leu Ile Cys Asn Lys His Ala Ser Leu Tyr Ile Gln Ala Asp Glu Tyr  
 85 90 95  
 Lys Ile Val Gly Lys Tyr Leu Leu Glu Ala Met Lys Glu Val Leu Gly  
 100 105 110  
 Asp Ala Cys Thr Asp Asp Ile Leu Asp Ala Trp Gly Ala Ala Tyr Trp  
 115 120 125  
 Ala Leu Ala Asp Ile Met Ile Asn Arg Glu Ala Ala Leu Tyr Lys Gln  
 130 135 140  
 Ser Gln Gly  
 145

<210> SEQ ID NO 13  
 <211> LENGTH: 165  
 <212> TYPE: PRT  
 <213> ORGANISM: *Zea mays*

<400> SEQUENCE: 13

Met Ala Leu Ala Glu Ala Asp Asp Gly Ala Val Val Phe Gly Glu Glu  
 1 5 10 15  
 Gln Glu Ala Leu Val Leu Lys Ser Trp Ala Val Met Lys Lys Asp Ala  
 20 25 30  
 Ala Asn Leu Gly Leu Arg Phe Phe Leu Lys Val Phe Glu Ile Ala Pro  
 35 40 45  
 Ser Ala Glu Gln Met Phe Ser Phe Leu Arg Asp Ser Asp Val Pro Leu  
 50 55 60  
 Glu Lys Asn Pro Lys Leu Lys Thr His Ala Met Ser Val Phe Val Met  
 65 70 75 80  
 Thr Cys Glu Ala Ala Ala Gln Leu Arg Lys Ala Gly Lys Val Thr Val  
 85 90 95  
 Arg Glu Thr Thr Leu Lys Arg Leu Gly Ala Thr His Leu Arg Tyr Gly  
 100 105 110  
 Val Ala Asp Gly His Phe Glu Val Thr Gly Phe Ala Leu Leu Glu Thr  
 115 120 125  
 Ile Lys Glu Ala Leu Pro Ala Asp Met Trp Ser Leu Glu Met Lys Lys  
 130 135 140  
 Ala Trp Ala Glu Ala Tyr Ser Gln Leu Val Ala Ala Ile Lys Arg Glu

-continued

145 150 155 160

Met Lys Pro Asp Ala  
165

<210> SEQ ID NO 14

<211> LENGTH: 169

<212> TYPE: PRT

<213> ORGANISM: *Oryza sativa subsp. japonica*

<400> SEQUENCE: 14

Met Ala Leu Val Glu Gly Asn Asn Gly Val Ser Gly Gly Ala Val Ser  
1 5 10 15

Phe Ser Glu Glu Gln Glu Ala Leu Val Leu Lys Ser Trp Ala Ile Met  
20 25 30

Lys Lys Asp Ser Ala Asn Ile Gly Leu Arg Phe Phe Leu Lys Ile Phe  
35 40 45

Glu Val Ala Pro Ser Ala Ser Gln Met Phe Ser Phe Leu Arg Asn Ser  
50 55 60

Asp Val Pro Leu Glu Lys Asn Pro Lys Leu Lys Thr His Ala Met Ser  
65 70 75 80

Val Phe Val Met Thr Cys Glu Ala Ala Ala Gln Leu Arg Lys Ala Gly  
85 90 95

Lys Val Thr Val Arg Asp Thr Thr Leu Lys Arg Leu Gly Ala Thr His  
100 105 110

Phe Lys Tyr Gly Val Gly Asp Ala His Phe Glu Val Thr Arg Phe Ala  
115 120 125

Leu Leu Glu Thr Ile Lys Glu Ala Val Pro Val Asp Met Trp Ser Pro  
130 135 140

Ala Met Lys Ser Ala Trp Ser Glu Ala Tyr Asn Gln Leu Val Ala Ala  
145 150 155 160

Ile Lys Gln Glu Met Lys Pro Ala Glu  
165

<210> SEQ ID NO 15

<211> LENGTH: 160

<212> TYPE: PRT

<213> ORGANISM: *Arabidopsis thaliana*

<400> SEQUENCE: 15

Met Glu Ser Glu Gly Lys Ile Val Phe Thr Glu Glu Gln Glu Ala Leu  
1 5 10 15

Val Val Lys Ser Trp Ser Val Met Lys Lys Asn Ser Ala Glu Leu Gly  
20 25 30

Leu Lys Leu Phe Ile Lys Ile Phe Glu Ile Ala Pro Thr Thr Lys Lys  
35 40 45

Met Phe Ser Phe Leu Arg Asp Ser Pro Ile Pro Ala Glu Gln Asn Pro  
50 55 60

Lys Leu Lys Pro His Ala Met Ser Val Phe Val Met Cys Cys Glu Ser  
65 70 75 80

Ala Val Gln Leu Arg Lys Thr Gly Lys Val Thr Val Arg Glu Thr Thr  
85 90 95

Leu Lys Arg Leu Gly Ala Ser His Ser Lys Tyr Gly Val Val Asp Glu  
100 105 110

His Phe Glu Val Ala Lys Tyr Ala Leu Leu Glu Thr Ile Lys Glu Ala

-continued

---

	115					120						125			
Val	Pro	Glu	Met	Trp	Ser	Pro	Glu	Met	Lys	Val	Ala	Trp	Gly	Gln	Ala
	130					135						140			
Tyr	Asp	His	Leu	Val	Ala	Ala	Ile	Lys	Ala	Glu	Met	Asn	Leu	Ser	Asn
145					150					155					160

<210> SEQ ID NO 16  
 <211> LENGTH: 147  
 <212> TYPE: PRT  
 <213> ORGANISM: Pisum sativum

<400> SEQUENCE: 16

Met	Gly	Phe	Thr	Asp	Lys	Gln	Glu	Ala	Leu	Val	Asn	Ser	Ser	Trp	Glu
1				5					10					15	
Ser	Phe	Lys	Gln	Asn	Leu	Ser	Gly	Asn	Ser	Ile	Leu	Phe	Tyr	Thr	Ile
		20						25					30		
Ile	Leu	Glu	Lys	Ala	Pro	Ala	Ala	Lys	Gly	Leu	Phe	Ser	Phe	Leu	Lys
		35					40						45		
Asp	Thr	Ala	Gly	Val	Glu	Asp	Ser	Pro	Lys	Leu	Gln	Ala	His	Ala	Glu
	50					55					60				
Gln	Val	Phe	Gly	Leu	Val	Arg	Asp	Ser	Ala	Ala	Gln	Leu	Arg	Thr	Lys
65					70					75					80
Gly	Glu	Val	Val	Leu	Gly	Asn	Ala	Thr	Leu	Gly	Ala	Ile	His	Val	Gln
				85					90					95	
Arg	Gly	Val	Thr	Asp	Pro	His	Phe	Val	Val	Val	Lys	Glu	Ala	Leu	Leu
			100					105						110	
Gln	Thr	Ile	Lys	Lys	Ala	Ser	Gly	Asn	Asn	Trp	Ser	Glu	Glu	Leu	Asn
		115					120					125			
Thr	Ala	Trp	Glu	Val	Ala	Tyr	Asp	Gly	Leu	Ala	Thr	Ala	Ile	Lys	Lys
	130					135						140			
Ala	Met	Thr													
145															

<210> SEQ ID NO 17  
 <211> LENGTH: 145  
 <212> TYPE: PRT  
 <213> ORGANISM: Vigna unguiculata

<400> SEQUENCE: 17

Met	Val	Ala	Phe	Ser	Asp	Lys	Gln	Glu	Ala	Leu	Val	Asn	Gly	Ala	Tyr
1				5					10					15	
Glu	Ala	Phe	Lys	Ala	Asn	Ile	Pro	Lys	Tyr	Ser	Val	Val	Phe	Tyr	Thr
		20						25					30		
Thr	Ile	Leu	Glu	Lys	Ala	Pro	Ala	Ala	Lys	Asn	Leu	Phe	Ser	Phe	Leu
		35					40						45		
Ala	Asn	Gly	Val	Asp	Ala	Thr	Asn	Pro	Lys	Leu	Thr	Gly	His	Ala	Glu
	50					55					60				
Lys	Leu	Phe	Gly	Leu	Val	Arg	Asp	Ser	Ala	Ala	Gln	Leu	Arg	Ala	Ser
65					70					75					80
Gly	Gly	Val	Val	Ala	Asp	Ala	Ala	Leu	Gly	Ala	Val	His	Ser	Gln	Lys
				85				90						95	
Ala	Val	Asn	Asp	Ala	Gln	Phe	Val	Val	Val	Lys	Glu	Ala	Leu	Val	Lys
		100					105							110	
Thr	Leu	Lys	Glu	Ala	Val	Gly	Asp	Lys	Trp	Ser	Asp	Glu	Leu	Gly	Thr

-continued

---

115 120 125

Ala Val Glu Leu Ala Tyr Asp Glu Leu Ala Ala Ala Ile Lys Lys Ala  
 130 135 140

Tyr  
 145

<210> SEQ ID NO 18  
 <211> LENGTH: 154  
 <212> TYPE: PRT  
 <213> ORGANISM: Bos taurus

<400> SEQUENCE: 18

Met Gly Leu Ser Asp Gly Glu Trp Gln Leu Val Leu Asn Ala Trp Gly  
 1 5 10 15

Lys Val Glu Ala Asp Val Ala Gly His Gly Gln Glu Val Leu Ile Arg  
 20 25 30

Leu Phe Thr Gly His Pro Glu Thr Leu Glu Lys Phe Asp Lys Phe Lys  
 35 40 45

His Leu Lys Thr Glu Ala Glu Met Lys Ala Ser Glu Asp Leu Lys Lys  
 50 55 60

His Gly Asn Thr Val Leu Thr Ala Leu Gly Gly Ile Leu Lys Lys Lys  
 65 70 75 80

Gly His His Glu Ala Glu Val Lys His Leu Ala Glu Ser His Ala Asn  
 85 90 95

Lys His Lys Ile Pro Val Lys Tyr Leu Glu Phe Ile Ser Asp Ala Ile  
 100 105 110

Ile His Val Leu His Ala Lys His Pro Ser Asp Phe Gly Ala Asp Ala  
 115 120 125

Gln Ala Ala Met Ser Lys Ala Leu Glu Leu Phe Arg Asn Asp Met Ala  
 130 135 140

Ala Gln Tyr Lys Val Leu Gly Phe His Gly  
 145 150

<210> SEQ ID NO 19  
 <211> LENGTH: 154  
 <212> TYPE: PRT  
 <213> ORGANISM: Sus scrofa

<400> SEQUENCE: 19

Met Gly Leu Ser Asp Gly Glu Trp Gln Leu Val Leu Asn Val Trp Gly  
 1 5 10 15

Lys Val Glu Ala Asp Val Ala Gly His Gly Gln Glu Val Leu Ile Arg  
 20 25 30

Leu Phe Lys Gly His Pro Glu Thr Leu Glu Lys Phe Asp Lys Phe Lys  
 35 40 45

His Leu Lys Ser Glu Asp Glu Met Lys Ala Ser Glu Asp Leu Lys Lys  
 50 55 60

His Gly Asn Thr Val Leu Thr Ala Leu Gly Gly Ile Leu Lys Lys Lys  
 65 70 75 80

Gly His His Glu Ala Glu Leu Thr Pro Leu Ala Gln Ser His Ala Thr  
 85 90 95

Lys His Lys Ile Pro Val Lys Tyr Leu Glu Phe Ile Ser Glu Ala Ile  
 100 105 110

Ile Gln Val Leu Gln Ser Lys His Pro Gly Asp Phe Gly Ala Asp Ala

-continued

---

115	120	125
Gln Gly Ala Met Ser Lys	Ala Leu Glu Leu Phe	Arg Asn Asp Met Ala
130	135	140
Ala Lys Tyr Lys Glu Leu Gly Phe Gln Gly		
145	150	

<210> SEQ ID NO 20  
 <211> LENGTH: 154  
 <212> TYPE: PRT  
 <213> ORGANISM: Equus caballus

<400> SEQUENCE: 20

Met Gly Leu Ser Asp Gly Glu Trp Gln Gln Val Leu Asn Val Trp Gly		
1	5	10 15
Lys Val Glu Ala Asp Ile Ala Gly His Gly Gln Glu Val Leu Ile Arg		
	20	25 30
Leu Phe Thr Gly His Pro Glu Thr Leu Glu Lys Phe Asp Lys Phe Lys		
	35	40 45
His Leu Lys Thr Glu Ala Glu Met Lys Ala Ser Glu Asp Leu Lys Lys		
	50	55 60
His Gly Thr Val Val Leu Thr Ala Leu Gly Gly Ile Leu Lys Lys Lys		
	65	70 75 80
Gly His His Glu Ala Glu Leu Lys Pro Leu Ala Gln Ser His Ala Thr		
	85	90 95
Lys His Lys Ile Pro Ile Lys Tyr Leu Glu Phe Ile Ser Asp Ala Ile		
	100	105 110
Ile His Val Leu His Ser Lys His Pro Gly Asp Phe Gly Ala Asp Ala		
	115	120 125
Gln Gly Ala Met Thr Lys Ala Leu Glu Leu Phe Arg Asn Asp Ile Ala		
	130	135 140
Ala Lys Tyr Lys Glu Leu Gly Phe Gln Gly		
145	150	

<210> SEQ ID NO 21  
 <211> LENGTH: 124  
 <212> TYPE: PRT  
 <213> ORGANISM: Synechocystis PCC6803

<400> SEQUENCE: 21

Met Ser Thr Leu Tyr Glu Lys Leu Gly Gly Thr Thr Ala Val Asp Leu		
1	5	10 15
Ala Val Asp Lys Phe Tyr Glu Arg Val Leu Gln Asp Asp Arg Ile Lys		
	20	25 30
His Phe Phe Ala Asp Val Asp Met Ala Lys Gln Arg Ala His Gln Lys		
	35	40 45
Ala Phe Leu Thr Tyr Ala Phe Gly Gly Thr Asp Lys Tyr Asp Gly Arg		
	50	55 60
Tyr Met Arg Glu Ala His Lys Glu Leu Val Glu Asn His Gly Leu Asn		
	65	70 75 80
Gly Glu His Phe Asp Ala Val Ala Glu Asp Leu Leu Ala Thr Leu Lys		
	85	90 95
Glu Met Gly Val Pro Glu Asp Leu Ile Ala Glu Val Ala Ala Val Ala		
	100	105 110
Gly Ala Pro Ala His Lys Arg Asp Val Leu Asn Gln		

-continued

---

115                      120

<210> SEQ ID NO 22  
 <211> LENGTH: 183  
 <212> TYPE: PRT  
 <213> ORGANISM: *Synechococcus* sp. PCC 7335

<400> SEQUENCE: 22

Met Asp Val Ala Leu Leu Glu Lys Ser Phe Glu Gln Ile Ser Pro Arg  
 1                      5                      10                      15

Ala Ile Glu Phe Ser Ala Ser Phe Tyr Gln Asn Leu Phe His His His  
                     20                      25                      30

Pro Glu Leu Lys Pro Leu Phe Ala Glu Thr Ser Gln Thr Ile Gln Glu  
                     35                      40                      45

Lys Lys Leu Ile Phe Ser Leu Ala Ala Ile Ile Glu Asn Leu Arg Asn  
                     50                      55                      60

Pro Asp Ile Leu Gln Pro Ala Leu Lys Ser Leu Gly Ala Arg His Ala  
 65                      70                      75                      80

Glu Val Gly Thr Ile Lys Ser His Tyr Pro Leu Val Gly Gln Ala Leu  
                     85                      90                      95

Ile Glu Thr Phe Ala Glu Tyr Leu Ala Ala Asp Trp Thr Glu Gln Leu  
                     100                      105                      110

Ala Thr Ala Trp Val Glu Ala Tyr Asp Val Ile Ala Ser Thr Met Ile  
                     115                      120                      125

Glu Gly Ala Asp Asn Pro Ala Ala Tyr Leu Glu Pro Glu Leu Thr Phe  
                     130                      135                      140

Tyr Glu Trp Leu Asp Leu Tyr Gly Glu Glu Ser Pro Lys Val Arg Asn  
 145                      150                      155                      160

Ala Ile Ala Thr Leu Thr His Phe His Tyr Gly Glu Asp Pro Gln Asp  
                     165                      170                      175

Val Gln Arg Asp Ser Arg Gly  
                     180

<210> SEQ ID NO 23  
 <211> LENGTH: 118  
 <212> TYPE: PRT  
 <213> ORGANISM: *Nostoc commune*

<400> SEQUENCE: 23

Met Ser Thr Leu Tyr Asp Asn Ile Gly Gly Gln Pro Ala Ile Glu Gln  
 1                      5                      10                      15

Val Val Asp Glu Leu His Lys Arg Ile Ala Thr Asp Ser Leu Leu Ala  
                     20                      25                      30

Pro Val Phe Ala Gly Thr Asp Met Val Lys Gln Arg Asn His Leu Val  
                     35                      40                      45

Ala Phe Leu Ala Gln Ile Phe Glu Gly Pro Lys Gln Tyr Gly Gly Arg  
                     50                      55                      60

Pro Met Asp Lys Thr His Ala Gly Leu Asn Leu Gln Gln Pro His Phe  
 65                      70                      75                      80

Asp Ala Ile Ala Lys His Leu Gly Glu Arg Met Ala Val Arg Gly Val  
                     85                      90                      95

Ser Ala Glu Asn Thr Lys Ala Ala Leu Asp Arg Val Thr Asn Met Lys  
                     100                      105                      110

Gly Ala Ile Leu Asn Lys

-continued

115

&lt;210&gt; SEQ ID NO 24

&lt;211&gt; LENGTH: 146

&lt;212&gt; TYPE: PRT

<213> ORGANISM: *Vitreoscilla stercoraria*

&lt;400&gt; SEQUENCE: 24

```

Met Leu Asp Gln Gln Thr Ile Asn Ile Ile Lys Ala Thr Val Pro Val
1           5           10           15
Leu Lys Glu His Gly Val Thr Ile Thr Thr Thr Phe Tyr Lys Asn Leu
20           25           30
Phe Ala Lys His Pro Glu Val Arg Pro Leu Phe Asp Met Gly Arg Gln
35           40           45
Glu Ser Leu Glu Gln Pro Lys Ala Leu Ala Met Thr Val Leu Ala Ala
50           55           60
Ala Gln Asn Ile Glu Asn Leu Pro Ala Ile Leu Pro Ala Val Lys Lys
65           70           75           80
Ile Ala Val Lys His Cys Gln Ala Gly Val Ala Ala Ala His Tyr Pro
85           90           95
Ile Val Gly Gln Glu Leu Leu Gly Ala Ile Lys Glu Val Leu Gly Asp
100          105          110
Ala Ala Thr Asp Asp Ile Leu Asp Ala Trp Gly Lys Ala Tyr Gly Val
115          120          125
Ile Ala Asp Val Phe Ile Gln Val Glu Ala Asp Leu Tyr Ala Gln Ala
130          135          140
Val Glu
145

```

&lt;210&gt; SEQ ID NO 25

&lt;211&gt; LENGTH: 131

&lt;212&gt; TYPE: PRT

<213> ORGANISM: *Corynebacterium glutamicum*

&lt;400&gt; SEQUENCE: 25

```

Met Thr Thr Ser Glu Asn Phe Tyr Asp Ser Val Gly Gly Glu Glu Thr
1           5           10           15
Phe Ser Leu Ile Val His Arg Phe Tyr Glu Gln Val Pro Asn Asp Asp
20           25           30
Ile Leu Gly Pro Met Tyr Pro Pro Asp Asp Phe Glu Gly Ala Glu Gln
35           40           45
Arg Leu Lys Met Phe Leu Ser Gln Tyr Trp Gly Gly Pro Lys Asp Tyr
50           55           60
Gln Glu Gln Arg Gly His Pro Arg Leu Arg Met Arg His Val Asn Tyr
65           70           75           80
Pro Ile Gly Val Thr Ala Ala Glu Arg Trp Leu Gln Leu Met Ser Asn
85           90           95
Ala Leu Asp Gly Val Asp Leu Thr Ala Glu Gln Arg Glu Ala Ile Trp
100          105          110
Glu His Met Val Arg Ala Ala Asp Met Leu Ile Asn Ser Asn Pro Asp
115          120          125
Pro His Ala
130

```



-continued

---

<210> SEQ ID NO 26  
 <211> LENGTH: 132  
 <212> TYPE: PRT  
 <213> ORGANISM: *Bacillus subtilis*  
 <400> SEQUENCE: 26  
 Met Gly Gln Ser Phe Asn Ala Pro Tyr Glu Ala Ile Gly Glu Glu Leu  
 1 5 10 15  
 Leu Ser Gln Leu Val Asp Thr Phe Tyr Glu Arg Val Ala Ser His Pro  
 20 25 30  
 Leu Leu Lys Pro Ile Phe Pro Ser Asp Leu Thr Glu Thr Ala Arg Lys  
 35 40 45  
 Gln Lys Gln Phe Leu Thr Gln Tyr Leu Gly Gly Pro Pro Leu Tyr Thr  
 50 55 60  
 Glu Glu His Gly His Pro Met Leu Arg Ala Arg His Leu Pro Phe Pro  
 65 70 75 80  
 Ile Thr Asn Glu Arg Ala Asp Ala Trp Leu Ser Cys Met Lys Asp Ala  
 85 90 95  
 Met Asp His Val Gly Leu Glu Gly Glu Ile Arg Glu Phe Leu Phe Gly  
 100 105 110  
 Arg Leu Glu Leu Thr Ala Arg His Met Val Asn Gln Thr Glu Ala Glu  
 115 120 125  
 Asp Arg Ser Ser  
 130

<210> SEQ ID NO 27  
 <211> LENGTH: 136  
 <212> TYPE: PRT  
 <213> ORGANISM: *Bacillus megaterium*  
 <400> SEQUENCE: 27  
 Met Arg Glu Lys Ile His Ser Pro Tyr Glu Leu Leu Gly Gly Glu His  
 1 5 10 15  
 Thr Ile Ser Lys Leu Val Asp Ala Phe Tyr Thr Arg Val Gly Gln His  
 20 25 30  
 Pro Glu Leu Ala Pro Ile Phe Pro Asp Asn Leu Thr Glu Thr Ala Arg  
 35 40 45  
 Lys Gln Lys Gln Phe Leu Thr Gln Tyr Leu Gly Gly Pro Ser Leu Tyr  
 50 55 60  
 Thr Glu Glu His Gly His Pro Met Leu Arg Ala Arg His Leu Pro Phe  
 65 70 75 80  
 Glu Ile Thr Pro Ser Arg Ala Lys Ala Trp Leu Thr Cys Met His Glu  
 85 90 95  
 Ala Met Asp Glu Ile Asn Leu Glu Gly Pro Glu Arg Asp Glu Leu Tyr  
 100 105 110  
 His Arg Leu Ile Leu Thr Ala Gln His Met Ile Asn Ser Pro Glu Gln  
 115 120 125  
 Thr Asp Glu Lys Gly Phe Ser His  
 130 135

<210> SEQ ID NO 28  
 <211> LENGTH: 399  
 <212> TYPE: PRT  
 <213> ORGANISM: *Saccharomyces cerevisiae*  
 <400> SEQUENCE: 28

-continued

---

Met Leu Ala Glu Lys Thr Arg Ser Ile Ile Lys Ala Thr Val Pro Val  
 1 5 10 15  
 Leu Glu Gln Gln Gly Thr Val Ile Thr Arg Thr Phe Tyr Lys Asn Met  
 20 25 30  
 Leu Thr Glu His Thr Glu Leu Leu Asn Ile Phe Asn Arg Thr Asn Gln  
 35 40 45  
 Lys Val Gly Ala Gln Pro Asn Ala Leu Ala Thr Thr Val Leu Ala Ala  
 50 55 60  
 Ala Lys Asn Ile Asp Asp Leu Ser Val Leu Met Asp His Val Lys Gln  
 65 70 75 80  
 Ile Gly His Lys His Arg Ala Leu Gln Ile Lys Pro Glu His Tyr Pro  
 85 90 95  
 Ile Val Gly Glu Tyr Leu Leu Lys Ala Ile Lys Glu Val Leu Gly Asp  
 100 105 110  
 Ala Ala Thr Pro Glu Ile Ile Asn Ala Trp Gly Glu Ala Tyr Gln Ala  
 115 120 125  
 Ile Ala Asp Ile Phe Ile Thr Val Glu Lys Lys Met Tyr Glu Glu Ala  
 130 135 140  
 Leu Trp Pro Gly Trp Lys Pro Phe Asp Ile Thr Ala Lys Glu Tyr Val  
 145 150 155 160  
 Ala Ser Asp Ile Val Glu Phe Thr Val Lys Pro Lys Phe Gly Ser Gly  
 165 170 175  
 Ile Glu Leu Glu Ser Leu Pro Ile Thr Pro Gly Gln Tyr Ile Thr Val  
 180 185 190  
 Asn Thr His Pro Ile Arg Gln Glu Asn Gln Tyr Asp Ala Leu Arg His  
 195 200 205  
 Tyr Ser Leu Cys Ser Ala Ser Thr Lys Asn Gly Leu Arg Phe Ala Val  
 210 215 220  
 Lys Met Glu Ala Ala Arg Glu Asn Phe Pro Ala Gly Leu Val Ser Glu  
 225 230 235 240  
 Tyr Leu His Lys Asp Ala Lys Val Gly Asp Glu Ile Lys Leu Ser Ala  
 245 250 255  
 Pro Ala Gly Asp Phe Ala Ile Asn Lys Glu Leu Ile His Gln Asn Glu  
 260 265 270  
 Val Pro Leu Val Leu Leu Ser Ser Gly Val Gly Val Thr Pro Leu Leu  
 275 280 285  
 Ala Met Leu Glu Glu Gln Val Lys Cys Asn Pro Asn Arg Pro Ile Tyr  
 290 295 300  
 Trp Ile Gln Ser Ser Tyr Asp Glu Lys Thr Gln Ala Phe Lys Lys His  
 305 310 315 320  
 Val Asp Glu Leu Leu Ala Glu Cys Ala Asn Val Asp Lys Ile Ile Val  
 325 330 335  
 His Thr Asp Thr Glu Pro Leu Ile Asn Ala Ala Phe Leu Lys Glu Lys  
 340 345 350  
 Ser Pro Ala His Ala Asp Val Tyr Thr Cys Gly Ser Leu Ala Phe Met  
 355 360 365  
 Gln Ala Met Ile Gly His Leu Lys Glu Leu Glu His Arg Asp Asp Met  
 370 375 380  
 Ile His Tyr Glu Pro Phe Gly Pro Lys Met Ser Thr Val Gln Val  
 385 390 395

-continued

<210> SEQ ID NO 29  
 <211> LENGTH: 152  
 <212> TYPE: PRT  
 <213> ORGANISM: Nicotina tobaccum

<400> SEQUENCE: 29

Met Ser Ser Phe Ser Glu Glu Gln Glu Ala Leu Val Leu Lys Ser Trp  
 1 5 10 15  
 Asp Ser Met Lys Lys Asn Ala Gly Glu Trp Gly Leu Lys Leu Phe Leu  
 20 25 30  
 Lys Ile Phe Glu Ile Ala Pro Ser Ala Lys Lys Leu Phe Ser Phe Leu  
 35 40 45  
 Lys Asp Ser Asn Val Pro Leu Glu Gln Asn Ala Lys Leu Lys Pro His  
 50 55 60  
 Ala Lys Ser Val Phe Val Met Thr Cys Glu Ala Val Gln Leu Arg  
 65 70 75 80  
 Lys Ala Gly Lys Val Val Val Arg Asp Ser Thr Leu Lys Lys Leu Gly  
 85 90 95  
 Ala Ala His Phe Lys Tyr Gly Val Ala Asp Glu His Phe Glu Val Thr  
 100 105 110  
 Lys Phe Ala Leu Leu Glu Thr Ile Lys Glu Ala Val Pro Asp Met Trp  
 115 120 125  
 Ser Val Asp Met Lys Asn Ala Trp Gly Glu Ala Phe Asp Gln Leu Val  
 130 135 140  
 Asn Ala Ile Lys Thr Glu Met Lys  
 145 150

<210> SEQ ID NO 30  
 <211> LENGTH: 160  
 <212> TYPE: PRT  
 <213> ORGANISM: Medicago sativa

<400> SEQUENCE: 30

Met Gly Thr Leu Asp Thr Lys Gly Phe Thr Glu Glu Gln Glu Ala Leu  
 1 5 10 15  
 Val Val Lys Ser Trp Asn Ala Met Lys Lys Asn Ser Ala Glu Leu Gly  
 20 25 30  
 Leu Lys Leu Phe Leu Lys Ile Phe Glu Ile Ala Pro Ser Ala Gln Lys  
 35 40 45  
 Leu Phe Ser Phe Leu Lys Asp Ser Lys Val Pro Leu Glu Gln Asn Thr  
 50 55 60  
 Lys Leu Lys Pro His Ala Met Ser Val Phe Leu Met Thr Cys Glu Ser  
 65 70 75 80  
 Ala Val Gln Leu Arg Lys Ser Gly Lys Val Thr Val Arg Glu Ser Ser  
 85 90 95  
 Leu Lys Lys Leu Gly Ala Asn His Phe Lys Tyr Gly Val Val Asp Glu  
 100 105 110  
 His Phe Glu Val Thr Lys Phe Ala Leu Leu Glu Thr Ile Lys Glu Ala  
 115 120 125  
 Val Pro Glu Met Trp Ser Pro Ala Met Lys Asn Ala Trp Gly Glu Ala  
 130 135 140  
 Tyr Asp Gln Leu Val Asn Ala Ile Lys Ser Glu Met Lys Pro Ser Ser  
 145 150 155 160

-continued

<210> SEQ ID NO 31  
 <211> LENGTH: 161  
 <212> TYPE: PRT  
 <213> ORGANISM: Glycine max

<400> SEQUENCE: 31

```

Met Thr Thr Thr Leu Glu Arg Gly Phe Ser Glu Glu Gln Glu Ala Leu
1          5          10          15
Val Val Lys Ser Trp Asn Val Met Lys Lys Asn Ser Gly Glu Leu Gly
          20          25          30
Leu Lys Phe Phe Leu Lys Ile Phe Glu Ile Ala Pro Ser Ala Gln Lys
          35          40          45
Leu Phe Ser Phe Leu Arg Asp Ser Thr Val Pro Leu Glu Gln Asn Pro
          50          55          60
Lys Leu Lys Pro His Ala Val Ser Val Phe Val Met Thr Cys Asp Ser
          65          70          75          80
Ala Val Gln Leu Arg Lys Ala Gly Lys Val Thr Val Arg Glu Ser Asn
          85          90          95
Leu Lys Lys Leu Gly Ala Thr His Phe Arg Thr Gly Val Ala Asn Glu
          100         105         110
His Phe Glu Val Thr Lys Phe Ala Leu Leu Glu Thr Ile Lys Glu Ala
          115         120         125
Val Pro Glu Met Trp Ser Pro Ala Met Lys Asn Ala Trp Gly Glu Ala
          130         135         140
Tyr Asp Gln Leu Val Asp Ala Ile Lys Ser Glu Met Lys Pro Pro Ser
          145         150         155         160
Ser
    
```

<210> SEQ ID NO 32  
 <211> LENGTH: 97  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: encodes signal peptide

<400> SEQUENCE: 32

```

atgaaaaaga aaaaaacatg gaaacgcttc ttacactttt cgagtgcagc tctggetgca    60
ggtttgatat tcacttctgc tgetcccgca gaggcag                                97
    
```

<210> SEQ ID NO 33  
 <211> LENGTH: 33  
 <212> TYPE: PRT  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: signal peptide

<400> SEQUENCE: 33

```

Met Lys Lys Lys Lys Thr Trp Lys Arg Phe Leu His Phe Ser Ser Ala
1          5          10          15
Ala Leu Ala Ala Gly Leu Ile Phe Thr Ser Ala Ala Pro Ala Glu Ala
          20          25          30
Ala
    
```

<210> SEQ ID NO 34  
 <211> LENGTH: 131  
 <212> TYPE: DNA

-continued

---

<213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: encodes signal peptide

<400> SEQUENCE: 34

atgtcaccag cacaaagaag aattttactg tatatccttt catttatctt tgtcacggc 60  
 gcagtgctct attttgtaa aagcgattat ctgtttacgc tgattttcat tgccattgcc 120  
 attctggtcg g 131

<210> SEQ ID NO 35  
 <211> LENGTH: 43  
 <212> TYPE: PRT  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: signal peptide

<400> SEQUENCE: 35

Met Ser Pro Ala Gln Arg Arg Ile Leu Leu Tyr Ile Leu Ser Phe Ile  
 1 5 10 15  
 Phe Val Ile Gly Ala Val Val Tyr Phe Val Lys Ser Asp Tyr Leu Phe  
 20 25 30  
 Thr Leu Ile Phe Ile Ala Ile Ala Ile Leu Phe  
 35 40

<210> SEQ ID NO 36  
 <211> LENGTH: 90  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: encodes signal peptide

<400> SEQUENCE: 36

atggtcagca tccgccgag cttcgaagcg tatgtcgatg acatgaatat cattaactgtt 60  
 ctgattcctg ctgaacaaaa ggaaatcatg 90

<210> SEQ ID NO 37  
 <211> LENGTH: 30  
 <212> TYPE: PRT  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: signal peptide

<400> SEQUENCE: 37

Met Val Ser Ile Arg Arg Ser Phe Glu Ala Tyr Val Asp Asp Met Asn  
 1 5 10 15  
 Ile Ile Thr Val Leu Ile Pro Ala Glu Gln Lys Glu Ile Met  
 20 25 30

<210> SEQ ID NO 38  
 <211> LENGTH: 99  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: encodes signal peptide

<400> SEQUENCE: 38

atggctgcat atattatcag aagaacctta atgtetatcc cgattttatt gggaattacg 60  
 atttatcat ttgttatcat gaaagccgcg cccggagat 99

-continued

---

<210> SEQ ID NO 39  
 <211> LENGTH: 33  
 <212> TYPE: PRT  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: signal peptide  
  
 <400> SEQUENCE: 39  
  
 Met Ala Ala Tyr Ile Ile Arg Arg Thr Leu Met Ser Ile Pro Ile Leu  
 1 5 10 15  
  
 Leu Gly Ile Thr Ile Leu Ser Phe Val Ile Met Lys Ala Ala Pro Gly  
 20 25 30  
  
 Asp  
  
 <210> SEQ ID NO 40  
 <211> LENGTH: 87  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: encodes signal peptide  
  
 <400> SEQUENCE: 40  
  
 atgaaacggt caatctctat ttttattacg tgtttattga ttacgttatt gacaatgggc 60  
 ggcatgatag ctctgcggc atcagca 87  
  
 <210> SEQ ID NO 41  
 <211> LENGTH: 29  
 <212> TYPE: PRT  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: signal peptide  
  
 <400> SEQUENCE: 41  
  
 Met Lys Arg Ser Ile Ser Ile Phe Ile Thr Cys Leu Leu Ile Thr Leu  
 1 5 10 15  
  
 Leu Thr Met Gly Gly Met Ile Ala Ser Pro Ala Ser Ala  
 20 25  
  
 <210> SEQ ID NO 42  
 <211> LENGTH: 93  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: encodes signal peptide  
  
 <400> SEQUENCE: 42  
  
 atgccttata tgaaacgagt gttgctgctt cttgtcactg gattgtttat gagtttgttt 60  
 gcagtcactg ctactgcctc agctcaaaca ggt 93  
  
 <210> SEQ ID NO 43  
 <211> LENGTH: 31  
 <212> TYPE: PRT  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: signal peptide  
  
 <400> SEQUENCE: 43  
  
 Met Pro Tyr Leu Lys Arg Val Leu Leu Leu Leu Val Thr Gly Leu Phe  
 1 5 10 15  
  
 Met Ser Leu Phe Ala Val Thr Ala Thr Ala Ser Ala Gln Thr Gly  
 20 25 30



-continued

---

<210> SEQ ID NO 49  
<211> LENGTH: 24  
<212> TYPE: PRT  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: signal peptide

<400> SEQUENCE: 49

Met Lys Lys Arg Trp Ser Ile Val Thr Leu Met Leu Ile Phe Thr Leu  
1           5           10           15  
Val Leu Ser Ala Cys Gly Phe Gly  
          20

<210> SEQ ID NO 50  
<211> LENGTH: 99  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: encodes signal peptide

<400> SEQUENCE: 50

atgctaaaat atatcggaag acgcttagtc tatatgatta tcacactatt tgtgattgta   60  
actgtgacat tcttcttaat gcaagcagca cggggcggg                         99

<210> SEQ ID NO 51  
<211> LENGTH: 33  
<212> TYPE: PRT  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: signal peptide

<400> SEQUENCE: 51

Met Leu Lys Tyr Ile Gly Arg Arg Leu Val Tyr Met Ile Ile Thr Leu  
1           5           10           15  
Phe Val Ile Val Thr Val Thr Phe Phe Leu Met Gln Ala Ala Pro Gly  
          20           25           30

Gly

<210> SEQ ID NO 52  
<211> LENGTH: 126  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: encodes signal peptide

<400> SEQUENCE: 52

atgacaagcc caaccgcgag aagaactgcg aaacgcagac ggagaaaact aaataaaaga   60  
ggaaaactgt tgtttggctct ttttagcagtg atggtttgca ttacgatttg gaatgctctt   120  
catcga   126

<210> SEQ ID NO 53  
<211> LENGTH: 42  
<212> TYPE: PRT  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: signal peptide

<400> SEQUENCE: 53

Met Thr Ser Pro Thr Arg Arg Arg Thr Ala Lys Arg Arg Arg Arg Lys



-continued

---

```

1           5           10           15
Leu Asn Lys Arg Gly Lys Leu Leu Phe Gly Leu Leu Ala Val Met Val
      20           25           30

Cys Ile Thr Ile Trp Asn Ala Leu His Arg
      35           40

<210> SEQ ID NO 54
<211> LENGTH: 168
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: encodes signal peptide

<400> SEQUENCE: 54
atggcatacg acagtcgttt tgatgaatgg gtacagaaac tgaaagagga aagctttcaa    60
aacaatacgt ttgaccgccc caaatttatt caaggagcgg ggaagattgc aggactttct    120
cttgatttaa cgattgcccc gtcggttggg gcctttgaag taaatgct                168

```

```

<210> SEQ ID NO 55
<211> LENGTH: 56
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: signal peptide

<400> SEQUENCE: 55
Met Ala Tyr Asp Ser Arg Phe Asp Glu Trp Val Gln Lys Leu Lys Glu
1           5           10           15
Glu Ser Phe Gln Asn Asn Thr Phe Asp Arg Arg Lys Phe Ile Gln Gly
      20           25           30
Ala Gly Lys Ile Ala Gly Leu Ser Leu Gly Leu Thr Ile Ala Gln Ser
      35           40           45
Val Gly Ala Phe Glu Val Asn Ala
      50           55

```

```

<210> SEQ ID NO 56
<211> LENGTH: 117
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: encodes signal peptide

<400> SEQUENCE: 56
atggcgcgaa aacatgat atccagacgt caatttttga attatacgt cacagcgta    60
ggaggtttta tggcggetag tatgctcatg cctatggttc gcttcgcact cgaccgg    117

```

```

<210> SEQ ID NO 57
<211> LENGTH: 39
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: signal peptide

<400> SEQUENCE: 57
Met Gly Gly Lys His Asp Ile Ser Arg Arg Gln Phe Leu Asn Tyr Thr
1           5           10           15
Leu Thr Gly Val Gly Gly Phe Met Ala Ala Ser Met Leu Met Pro Met
      20           25           30

```

-continued

---

Val Arg Phe Ala Leu Asp Pro  
35

<210> SEQ ID NO 58  
 <211> LENGTH: 78  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: encodes signal peptide

<400> SEQUENCE: 58

atggttggtga aaaggagaat aggggttgcta ttaagtatgg ttggcgatt catgcttttg 60  
 gctggatgct cgagtgtg 78

<210> SEQ ID NO 59  
 <211> LENGTH: 26  
 <212> TYPE: PRT  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: signal peptide

<400> SEQUENCE: 59

Met Leu Leu Lys Arg Arg Ile Gly Leu Leu Leu Ser Met Val Gly Val  
 1 5 10 15

Phe Met Leu Leu Ala Gly Cys Ser Ser Val  
 20 25

<210> SEQ ID NO 60  
 <211> LENGTH: 129  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: encodes signal peptide

<400> SEQUENCE: 60

atgaaaaaaaa cactcaccac tattcgcaga tcatcaattg caaggagact tattatttct 60  
 ttctgctga tcttaattgt tccgataacc gccctttcgg ttagcgctta tcaatcagca 120  
 gttgcctca 129

<210> SEQ ID NO 61  
 <211> LENGTH: 43  
 <212> TYPE: PRT  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: signal peptide

<400> SEQUENCE: 61

Met Lys Lys Thr Leu Thr Thr Ile Arg Arg Ser Ser Ile Ala Arg Arg  
 1 5 10 15

Leu Ile Ile Ser Phe Leu Leu Ile Leu Ile Val Pro Ile Thr Ala Leu  
 20 25 30

Ser Val Ser Ala Tyr Gln Ser Ala Val Ala Ser  
 35 40

<210> SEQ ID NO 62  
 <211> LENGTH: 105  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: encodes signal peptide

-continued

&lt;400&gt; SEQUENCE: 62

```
atgaaaaaaa gaaagaggcg aaactttaa aggttcattg cagcattttt agtgttggct    60
ttaatgattt cattagtgcc agccgatgta ctagcaaaat ctaca                    105
```

&lt;210&gt; SEQ ID NO 63

&lt;211&gt; LENGTH: 35

&lt;212&gt; TYPE: PRT

&lt;213&gt; ORGANISM: Artificial Sequence

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: signal peptide

&lt;400&gt; SEQUENCE: 63

```
Met Lys Lys Arg Lys Arg Arg Asn Phe Lys Arg Phe Ile Ala Ala Phe
 1          5          10         15
Leu Val Leu Ala Leu Met Ile Ser Leu Val Pro Ala Asp Val Leu Ala
          20         25         30
Lys Ser Thr
          35
```

&lt;210&gt; SEQ ID NO 64

&lt;211&gt; LENGTH: 102

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Artificial Sequence

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: encodes signal peptide

&lt;400&gt; SEQUENCE: 64

```
atgaaaacgca gaaaattcag ctcggttggtg ggggcagtgct ttatttttgc actgattttc    60
agcctttttt ctccgggaac caaagctgca ggggcgggcy cg                        102
```

&lt;210&gt; SEQ ID NO 65

&lt;211&gt; LENGTH: 34

&lt;212&gt; TYPE: PRT

&lt;213&gt; ORGANISM: Artificial Sequence

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: signal peptide

&lt;400&gt; SEQUENCE: 65

```
Met Lys Arg Arg Lys Phe Ser Ser Val Val Ala Ala Val Leu Ile Phe
 1          5          10         15
Ala Leu Ile Phe Ser Leu Phe Ser Pro Gly Thr Lys Ala Ala Ala Ala
          20         25         30
Gly Ala
```

&lt;210&gt; SEQ ID NO 66

&lt;211&gt; LENGTH: 78

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Artificial Sequence

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: encodes signal peptide

&lt;400&gt; SEQUENCE: 66

```
atgaagaaac gcagaaagat atgttattgc aatactgccc tgctgcttat gatttttgct    60
gctggatgta cggacagt                                         78
```

&lt;210&gt; SEQ ID NO 67

&lt;211&gt; LENGTH: 26

&lt;212&gt; TYPE: PRT

&lt;213&gt; ORGANISM: Artificial Sequence

-continued

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: signal peptide

&lt;400&gt; SEQUENCE: 67

Met Lys Lys Arg Arg Lys Ile Cys Tyr Cys Asn Thr Ala Leu Leu Leu  
1 5 10 15Met Ile Leu Leu Ala Gly Cys Thr Asp Ser  
20 25

&lt;210&gt; SEQ ID NO 68

&lt;211&gt; LENGTH: 129

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Artificial Sequence

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: encodes signal peptide

&lt;400&gt; SEQUENCE: 68

atgaagaaaa gagttgctgg ctggtacagg cggatgaaga ttaaggataa gctgtttgtg 60

tttctatcgt tgattatggc cgtatccttt ctgtttgtat acagcggggg ccagtatgcc 120

tttcatgtg 129

&lt;210&gt; SEQ ID NO 69

&lt;211&gt; LENGTH: 43

&lt;212&gt; TYPE: PRT

&lt;213&gt; ORGANISM: Artificial Sequence

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: signal peptide

&lt;400&gt; SEQUENCE: 69

Met Lys Lys Arg Val Ala Gly Trp Tyr Arg Arg Met Lys Ile Lys Asp  
1 5 10 15Lys Leu Phe Val Phe Leu Ser Leu Ile Met Ala Val Ser Phe Leu Phe  
20 25 30Val Tyr Ser Gly Val Gln Tyr Ala Phe His Val  
35 40

&lt;210&gt; SEQ ID NO 70

&lt;211&gt; LENGTH: 114

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Artificial Sequence

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: encodes signal peptide

&lt;400&gt; SEQUENCE: 70

atgagaagga gctgtctgat gattagacga aggaaacgca tgttaccgc tgttacgttg 60

ctggtcttgt tggatgatggg aacctctgta tgtcctgtga aagetgaagg ggca 114

&lt;210&gt; SEQ ID NO 71

&lt;211&gt; LENGTH: 38

&lt;212&gt; TYPE: PRT

&lt;213&gt; ORGANISM: Artificial Sequence

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: signal peptide

&lt;400&gt; SEQUENCE: 71

Met Arg Arg Ser Cys Leu Met Ile Arg Arg Arg Lys Arg Met Phe Thr  
1 5 10 15Ala Val Thr Leu Leu Val Leu Leu Val Met Gly Thr Ser Val Cys Pro  
20 25 30

-continued

---

Val Lys Ala Glu Gly Ala  
35

<210> SEQ ID NO 72  
 <211> LENGTH: 114  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: encodes signal peptide

<400> SEQUENCE: 72

atgagaatac agaaaagacg aacacacgtc gaaaacattc tccgtattct ttgcccceca 60  
 attatgatac ttagcctaata cctcccaaca ccaccattc atgcagaaga aagc 114

<210> SEQ ID NO 73  
 <211> LENGTH: 38  
 <212> TYPE: PRT  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: signal peptide

<400> SEQUENCE: 73

Met Arg Ile Gln Lys Arg Arg Thr His Val Glu Asn Ile Leu Arg Ile  
 1 5 10 15  
 Leu Leu Pro Pro Ile Met Ile Leu Ser Leu Ile Leu Pro Thr Pro Pro  
 20 25 30  
 Ile His Ala Glu Glu Ser  
 35

<210> SEQ ID NO 74  
 <211> LENGTH: 147  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: encodes signal peptide

<400> SEQUENCE: 74

atgctgtctg tcgaaatgat aagcagacaa aatcgttgtc attatgtgta taagggagga 60  
 aatatgatga ggcgtattct gcatattgtg ttgatcacgg cattaatggt cttaaatgta 120  
 atgtacacgt tcgaagctgt aaaggca 147

<210> SEQ ID NO 75  
 <211> LENGTH: 49  
 <212> TYPE: PRT  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: signal peptide

<400> SEQUENCE: 75

Met Leu Ser Val Glu Met Ile Ser Arg Gln Asn Arg Cys His Tyr Val  
 1 5 10 15  
 Tyr Lys Gly Gly Asn Met Met Arg Arg Ile Leu His Ile Val Leu Ile  
 20 25 30  
 Thr Ala Leu Met Phe Leu Asn Val Met Tyr Thr Phe Glu Ala Val Lys  
 35 40 45

Ala

<210> SEQ ID NO 76  
 <211> LENGTH: 93

-continued

---

<212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: encodes signal peptide  
  
 <400> SEQUENCE: 76  
  
 atgctaagag atttaggaag aagagtagcg atcgcagcca ttttaagcgg aattattctt 60  
 ggaggcatga gcatttcttt ggcaaatatg ccc 93  
  
 <210> SEQ ID NO 77  
 <211> LENGTH: 31  
 <212> TYPE: PRT  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: signal peptide  
  
 <400> SEQUENCE: 77  
  
 Met Leu Arg Asp Leu Gly Arg Arg Val Ala Ile Ala Ala Ile Leu Ser  
 1 5 10 15  
  
 Gly Ile Ile Leu Gly Gly Met Ser Ile Ser Leu Ala Asn Met Pro  
 20 25 30  
  
 <210> SEQ ID NO 78  
 <211> LENGTH: 102  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: encodes signal peptide  
  
 <400> SEQUENCE: 78  
  
 atgaaaaaga tgtccagaag acaatttcta aaaggaatgt tcggcgctct tgetgcgggg 60  
 gctttaacgg cggcgggggg atatggctat gccaggtatc tc 102  
  
 <210> SEQ ID NO 79  
 <211> LENGTH: 34  
 <212> TYPE: PRT  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: signal peptide  
  
 <400> SEQUENCE: 79  
  
 Met Lys Lys Met Ser Arg Arg Gln Phe Leu Lys Gly Met Phe Gly Ala  
 1 5 10 15  
  
 Leu Ala Ala Gly Ala Leu Thr Ala Gly Gly Gly Tyr Gly Tyr Ala Arg  
 20 25 30  
  
 Tyr Leu  
  
 <210> SEQ ID NO 80  
 <211> LENGTH: 84  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: encodes signal peptide  
  
 <400> SEQUENCE: 80  
  
 atgagaagat ttttactaaa tgtcatatta gtcttagcca ttgtcttggc cttgagatat 60  
 gttcattact cattggaacc agaa 84  
  
 <210> SEQ ID NO 81  
 <211> LENGTH: 28

-continued

---

```

<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: signal peptide

<400> SEQUENCE: 81

Met Arg Arg Phe Leu Leu Asn Val Ile Leu Val Leu Ala Ile Val Leu
1           5           10           15

Phe Leu Arg Tyr Val His Tyr Ser Leu Glu Pro Glu
           20           25

<210> SEQ ID NO 82
<211> LENGTH: 87
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: encodes signal peptide

<400> SEQUENCE: 82

atgtttgaaa gtgaagcaga actgagacga atcaggattg cacttgtatg gatagctgtc      60
tttttactgt tcggggcggtg cggaat                                           87

<210> SEQ ID NO 83
<211> LENGTH: 29
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: signal peptide

<400> SEQUENCE: 83

Met Phe Glu Ser Glu Ala Glu Leu Arg Arg Ile Arg Ile Ala Leu Val
1           5           10           15

Trp Ile Ala Val Phe Leu Leu Phe Gly Ala Cys Gly Asn
           20           25

<210> SEQ ID NO 84
<211> LENGTH: 93
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: encodes signal peptide

<400> SEQUENCE: 84

atgaagaaga gaattacata ttcactgctt gctcttctag cagttgttgc tttcgctttc      60
actgattcat caaaagcaaa agcggcagaa gca                                       93

<210> SEQ ID NO 85
<211> LENGTH: 31
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: signal peptide

<400> SEQUENCE: 85

Met Lys Lys Arg Ile Thr Tyr Ser Leu Leu Ala Leu Leu Ala Val Val
1           5           10           15

Ala Phe Ala Phe Thr Asp Ser Ser Lys Ala Lys Ala Ala Glu Ala
           20           25           30

<210> SEQ ID NO 86
<211> LENGTH: 116

```

-continued

---

```

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: encodes signal peptide

<400> SEQUENCE: 86

atgcagaaat atagacgcag aaacacgggt gcctttacag tactagctta ttttactttt      60
tttgcgggag tatttttggt tagtatcgga ctctataatg ctgataatct ggaact      116

<210> SEQ ID NO 87
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: signal peptide

<400> SEQUENCE: 87

Met Gln Lys Tyr Arg Arg Arg Asn Thr Val Ala Phe Thr Val Leu Ala
1          5          10          15

Tyr Phe Thr Phe Phe Ala Gly Val Phe Leu Phe Ser Ile Gly Leu Tyr
          20          25          30

Asn Ala Asp Asn Leu Glu
          35

<210> SEQ ID NO 88
<211> LENGTH: 102
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: encodes signal peptide

<400> SEQUENCE: 88

atgatgttga atatgatcag acgtttgctg atgacctggt tattttctgct tgcatttggc      60
acgacatttt tatcagtgtc aggaattgaa gcgaaggact tg      102

<210> SEQ ID NO 89
<211> LENGTH: 34
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: signal peptide

<400> SEQUENCE: 89

Met Met Leu Asn Met Ile Arg Arg Leu Leu Met Thr Cys Leu Phe Leu
1          5          10          15

Leu Ala Phe Gly Thr Thr Phe Leu Ser Val Ser Gly Ile Glu Ala Lys
          20          25          30

Asp Leu

<210> SEQ ID NO 90
<211> LENGTH: 132
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: encodes signal peptide

<400> SEQUENCE: 90

atggctgaac gcgtagagt gcgtgtgcca aaaaagaaaa agagcaaacy taggaaaatt      60
ttaaaaagaa taatgttatt gttcgccctt gcactattgg tagttgtagg gcttgccggg      120

```



-continued

---

 tataaaacttt at 132

<210> SEQ ID NO 91  
 <211> LENGTH: 44  
 <212> TYPE: PRT  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: signal peptide

<400> SEQUENCE: 91

Met Ala Glu Arg Val Arg Val Arg Val Arg Lys Lys Lys Lys Ser Lys  
 1 5 10 15  
 Arg Arg Lys Ile Leu Lys Arg Ile Met Leu Leu Phe Ala Leu Ala Leu  
 20 25 30  
 Leu Val Val Val Gly Leu Gly Gly Tyr Lys Leu Tyr  
 35 40

<210> SEQ ID NO 92  
 <211> LENGTH: 141  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: encodes signal peptide

<400> SEQUENCE: 92

atgagcgatg aacagaaaaa gccagaacaa attcacagac gggacatttt aaaatgggga 60  
 gcgatggcgg gggcagccgt tgcgatcggg gccagcggtc tcggcggctt cgctccgctt 120  
 gttcagactg cggctaagcc a 141

<210> SEQ ID NO 93  
 <211> LENGTH: 47  
 <212> TYPE: PRT  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: signal peptide

<400> SEQUENCE: 93

Met Ser Asp Glu Gln Lys Lys Pro Glu Gln Ile His Arg Arg Asp Ile  
 1 5 10 15  
 Leu Lys Trp Gly Ala Met Ala Gly Ala Ala Val Ala Ile Gly Ala Ser  
 20 25 30  
 Gly Leu Gly Gly Leu Ala Pro Leu Val Gln Thr Ala Ala Lys Pro  
 35 40 45

<210> SEQ ID NO 94  
 <211> LENGTH: 204  
 <212> TYPE: PRT  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: fusion protein

<400> SEQUENCE: 94

Met Ala Tyr Asp Ser Arg Phe Asp Glu Trp Val Gln Lys Leu Lys Glu  
 1 5 10 15  
 Glu Ser Phe Gln Asn Asn Thr Phe Asp Arg Arg Lys Phe Ile Gln Gly  
 20 25 30  
 Ala Gly Lys Ile Ala Gly Leu Ser Leu Gly Leu Thr Ile Ala Gln Ser  
 35 40 45  
 Ala Ser Ala Ala Gly Ala His Met Glu Leu Gly Leu Ser Glu Glu Thr

-continued

50			55				60								
Ile	Arg	Val	Ile	Lys	Ser	Thr	Val	Pro	Leu	Leu	Lys	Glu	His	Gly	Thr
65				70					75						80
Glu	Ile	Thr	Ala	Arg	Met	Tyr	Glu	Leu	Leu	Phe	Ser	Lys	Tyr	Pro	Lys
			85					90						95	
Thr	Lys	Glu	Leu	Phe	Ala	Gly	Ala	Ser	Glu	Glu	Gln	Pro	Lys	Lys	Leu
		100					105						110		
Ala	Asn	Ala	Ile	Ile	Ala	Tyr	Ala	Thr	Tyr	Ile	Asp	Arg	Leu	Glu	Glu
		115					120					125			
Leu	Asp	Asn	Ala	Ile	Ser	Thr	Ile	Ala	Arg	Ser	His	Val	Arg	Arg	Asn
	130					135					140				
Val	Lys	Pro	Glu	His	Tyr	Pro	Leu	Val	Lys	Glu	Cys	Leu	Leu	Gln	Ala
145					150					155					160
Ile	Glu	Glu	Val	Leu	Asn	Pro	Gly	Glu	Glu	Val	Leu	Lys	Ala	Trp	Glu
			165						170						175
Glu	Ala	Tyr	Asp	Phe	Leu	Ala	Lys	Thr	Leu	Ile	Thr	Leu	Glu	Lys	Lys
			180					185						190	
Leu	Tyr	Ser	Gln	Pro	Arg	His	His	His	His	His	His				
		195					200								

&lt;210&gt; SEQ ID NO 95

&lt;211&gt; LENGTH: 156

&lt;212&gt; TYPE: PRT

&lt;213&gt; ORGANISM: Artificial Sequence

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: signal peptidase I recognition site

&lt;400&gt; SEQUENCE: 95

Ala	Ser	Ala	Ala	Gly	Ala	His	Met	Glu	Leu	Gly	Leu	Ser	Glu	Glu	Thr
1				5					10					15	
Ile	Arg	Val	Ile	Lys	Ser	Thr	Val	Pro	Leu	Leu	Lys	Glu	His	Gly	Thr
		20						25					30		
Glu	Ile	Thr	Ala	Arg	Met	Tyr	Glu	Leu	Leu	Phe	Ser	Lys	Tyr	Pro	Lys
		35				40						45			
Thr	Lys	Glu	Leu	Phe	Ala	Gly	Ala	Ser	Glu	Glu	Gln	Pro	Lys	Lys	Leu
	50					55					60				
Ala	Asn	Ala	Ile	Ile	Ala	Tyr	Ala	Thr	Tyr	Ile	Asp	Arg	Leu	Glu	Glu
65					70					75					80
Leu	Asp	Asn	Ala	Ile	Ser	Thr	Ile	Ala	Arg	Ser	His	Val	Arg	Arg	Asn
			85						90					95	
Val	Lys	Pro	Glu	His	Tyr	Pro	Leu	Val	Lys	Glu	Cys	Leu	Leu	Gln	Ala
		100						105						110	
Ile	Glu	Glu	Val	Leu	Asn	Pro	Gly	Glu	Glu	Val	Leu	Lys	Ala	Trp	Glu
		115					120						125		
Glu	Ala	Tyr	Asp	Phe	Leu	Ala	Lys	Thr	Leu	Ile	Thr	Leu	Glu	Lys	Lys
	130					135					140				
Leu	Tyr	Ser	Gln	Pro	Arg	His	His	His	His	His	His				
145					150						155				

&lt;210&gt; SEQ ID NO 96

&lt;211&gt; LENGTH: 10

&lt;212&gt; TYPE: PRT

&lt;213&gt; ORGANISM: Artificial Sequence

-continued

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: N-terminal peptide

&lt;400&gt; SEQUENCE: 96

Gly Ala His Met Glu Leu Gly Leu Ser Glu  
 1                    5                    10

What is claimed is:

1. A cell comprising an exogenous nucleic acid molecule comprising, in the 5' to 3' direction, a promoter sequence operably linked to a nucleic acid encoding a signal peptide operably linked to a nucleic acid encoding a heme-containing polypeptide.

2. The cell of claim 1, wherein the promoter sequence is a tissue-specific promoter.

3. The cell of claim 1, wherein the promoter sequence is a constitutive promoter.

4. The cell of claim 1, wherein the promoter sequence is an inducible promoter.

5. The cell of claim 1, wherein the signal peptide is a transit peptide.

6. The cell of claim 1, wherein the signal peptide is a secretion signal peptide.

7. The cell of claim 1, wherein the signal peptide has a sequence selected from the group consisting of 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, and 93.

8. The cell of claim 1, wherein the nucleic acid encoding a heme-containing polypeptide is selected from the group consisting of androglobin, cytoglobin, globin E, globin X, globin Y, hemoglobin, myoglobin, leghemoglobin, erythrocrucorin, beta hemoglobin, alpha hemoglobin, non-symbiotic

hemoglobin, flavohemoglobin, protoglobin, cyanoglobin, Hell's gate globin I, bacterial hemoglobin, ciliate myoglobin, histoglobin, neuroglobins, truncated 2/2 globin, HbN, HbO, and, G1b3.

9. The cell of claim 1, wherein the nucleic acid encoding a heme-containing polypeptide is a globin.

10. The cell of claim 1, wherein the nucleic acid encoding a heme-containing polypeptide is leghemoglobin.

11. The cell of claim 1, wherein the cell is selected from the group consisting of a bacterial cell, a yeast cell, an insect cell, a plant cell, and a mammalian cell.

12. The cell of claim 1, wherein the cell is a yeast cell.

13. The cell of claim 12, wherein the yeast cell is a *Pichia pastoris* yeast cell.

14. The cell of claim 1, wherein the cell is from a species other than *Nicotiana*.

15. The cell of claim 1, wherein the cell is from a species other than a filamentous fungi.

16. The cell of claim 1, further comprising a tag.

17. The cell of claim 1, wherein the tag is a detectable label.

18. The cell of claim 1, further comprising a transcription termination region.

\* \* \* \* \*