



US 20250263368A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2025/0263368 A1**

Zhang et al.

(43) **Pub. Date:** **Aug. 21, 2025**

(54) **CELLULOSE SYNTHASE INHIBITORS AS A NEW CLASS OF HERBICIDE AND NON-GMO CROPS THAT ARE RESISTANT TO THE HERBICIDE**

(71) Applicant: **Purdue Research Foundation**, West Lafayette, IN (US)

(72) Inventors: **Chunhua Zhang**, West Lafayette, IN (US); **Lei Huang**, West Lafayette, IN (US); **Xiaohui Li**, West Lafayette, IN (US)

(21) Appl. No.: **19/204,442**

(22) Filed: **May 9, 2025**

Related U.S. Application Data

(63) Continuation of application No. 16/765,788, filed on May 20, 2020, filed as application No. PCT/US18/61962 on Nov. 20, 2018.

(60) Provisional application No. 62/588,677, filed on Nov. 20, 2017.

Publication Classification

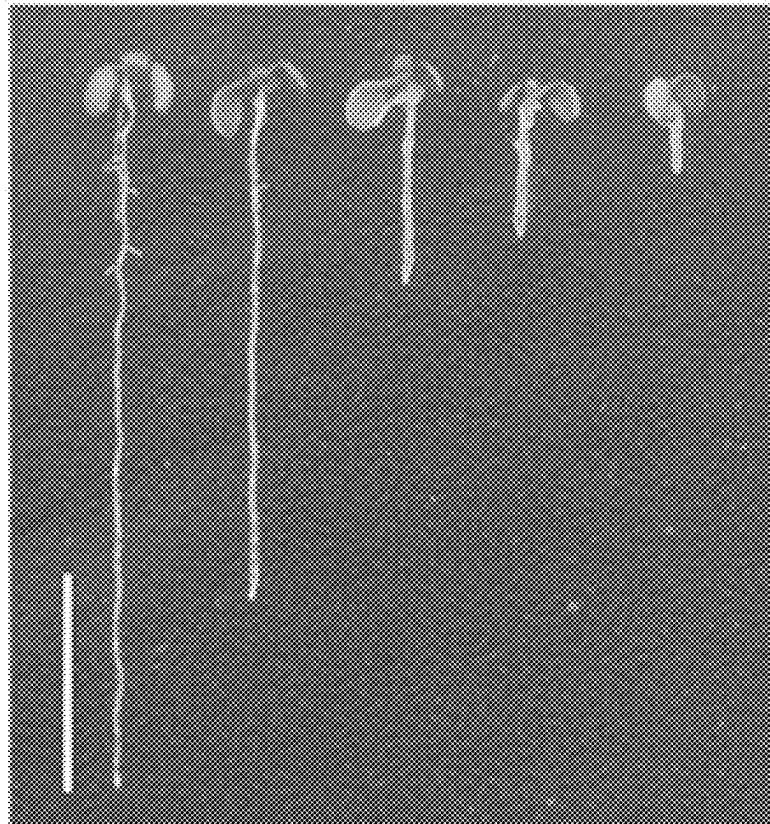
(51) **Int. Cl.**
C07C 335/40 (2006.01)
A01N 47/34 (2006.01)
CI2N 9/10 (2006.01)
(52) **U.S. Cl.**
CPC *C07C 335/40* (2013.01); *A01N 47/34* (2013.01); *CI2N 9/1059* (2013.01)

(57) **ABSTRACT**

Using chemical genetic screening, we discovered a small molecule Cellulosin (aka endosidin20 or ES20) that causes cell swollen and inhibits plant growth, but does not disrupt global vesicle trafficking. By doing mutant screening, we obtained multiple alleles of *Arabidopsis thaliana* that are resistant to Cellulosin inhibition in growth. Those mutated amino acid residues are conserved across plant species. Cellulosin targets a group of cellulose synthases (CesAs) of *Arabidopsis thaliana* by binding to a conserved domain essential for the catalytic activity of CesA. Cellulosin may target and inhibit all subtypes of CesAs in plants. The present invention relates to Cellulosin, a cellulose synthase inhibitor, its analogs or derivatives as a broad-spectrum herbicide. The mutated genes, their protein products and a cell or a plant having those mutated genes or expressing those protein products are within the scope of this disclosure.

Specification includes a Sequence Listing.

B



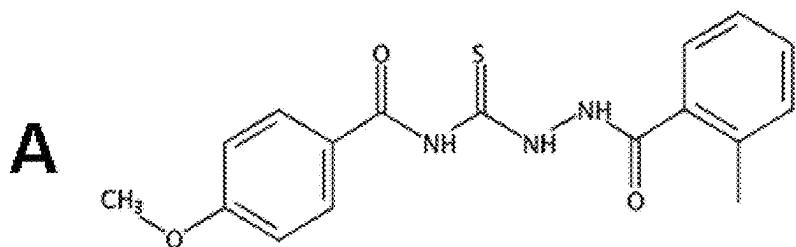


FIG. 1A

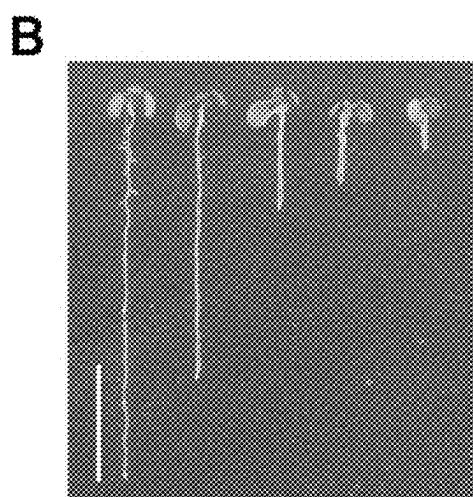


FIG. 1B

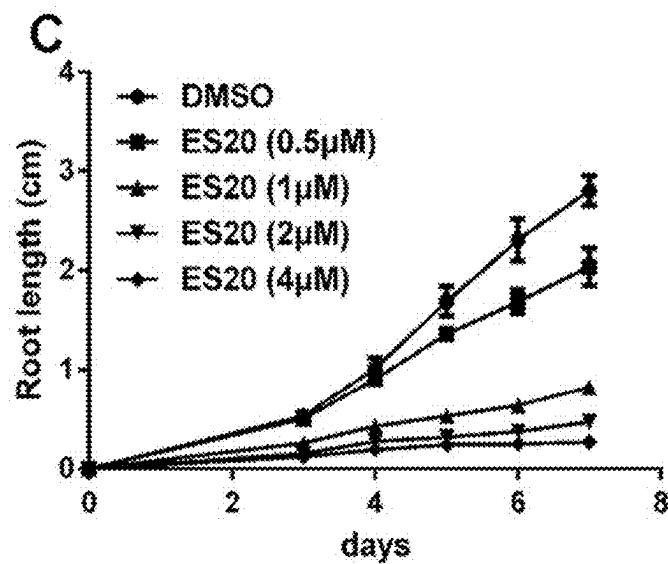


FIG. 1C

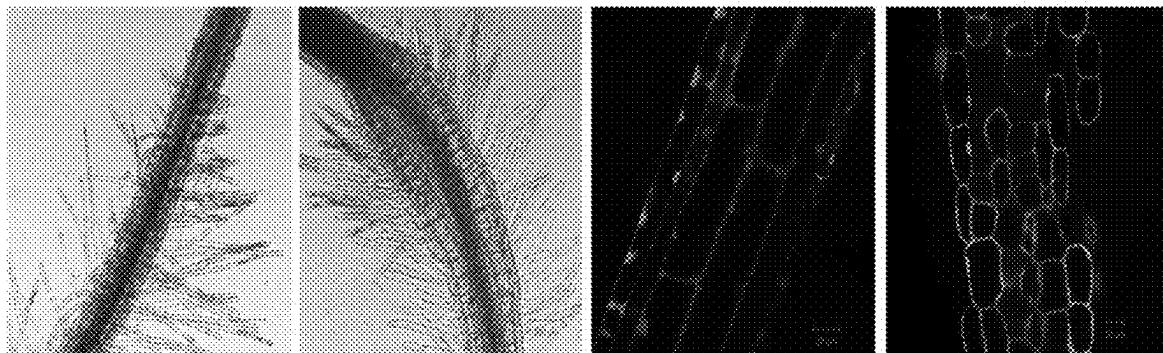


FIG. 1D

FIG. 1E

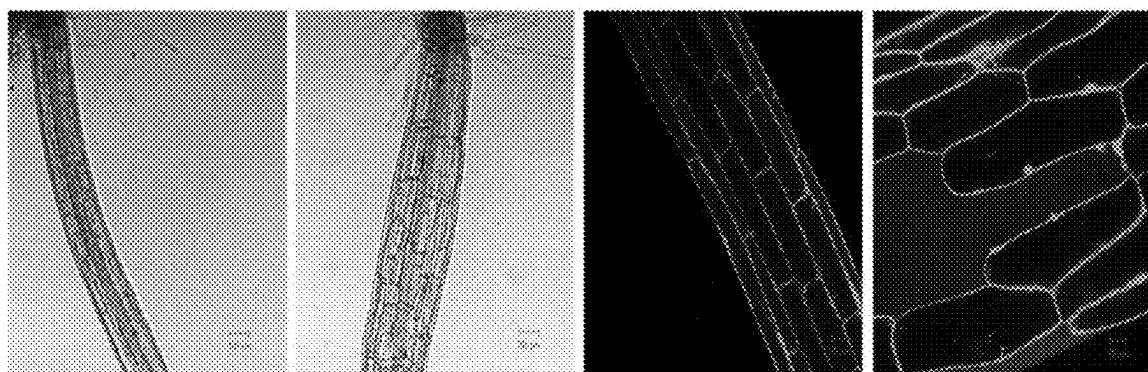


FIG. 1H

FIG. 1I

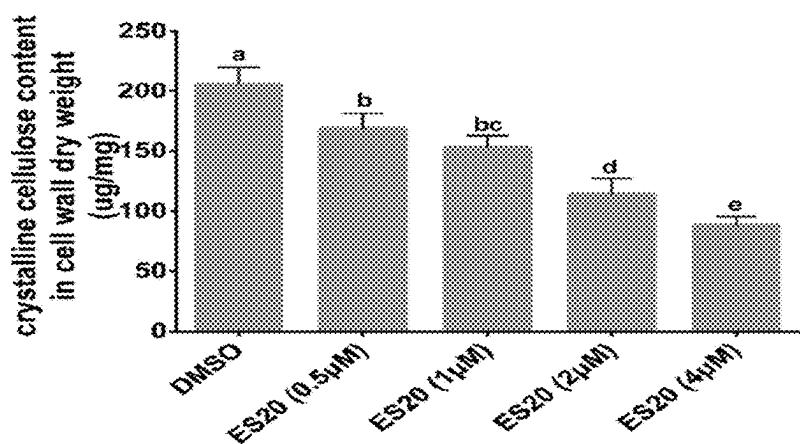


FIG. 1J

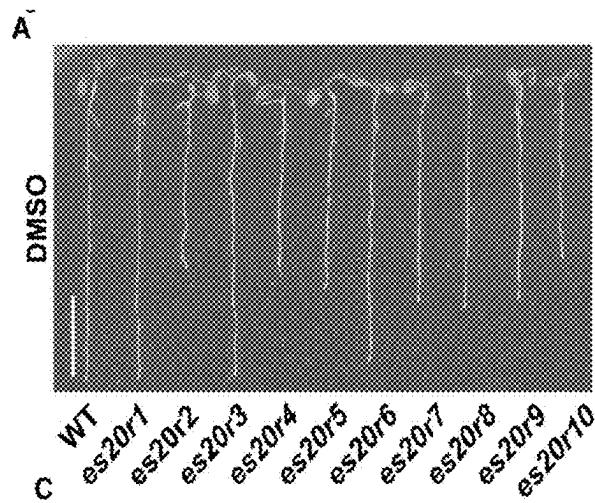


FIG. 2A

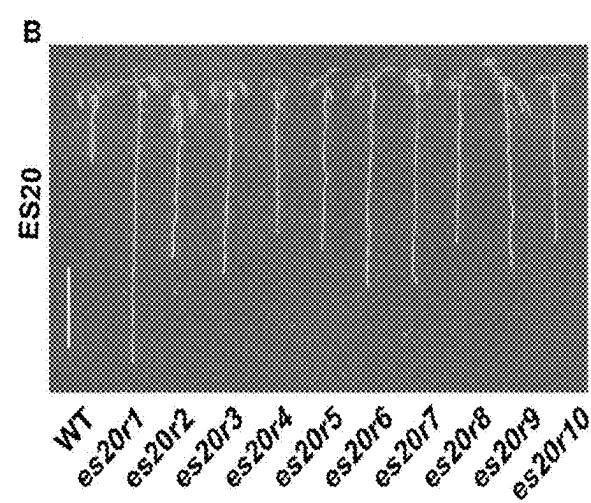


FIG. 2B

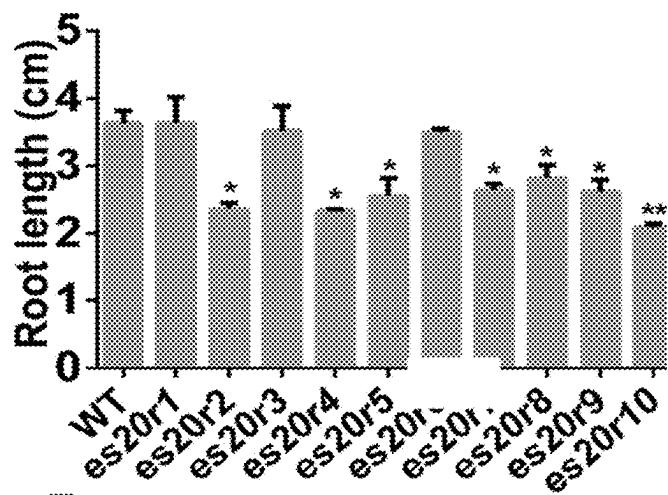


FIG. 2C

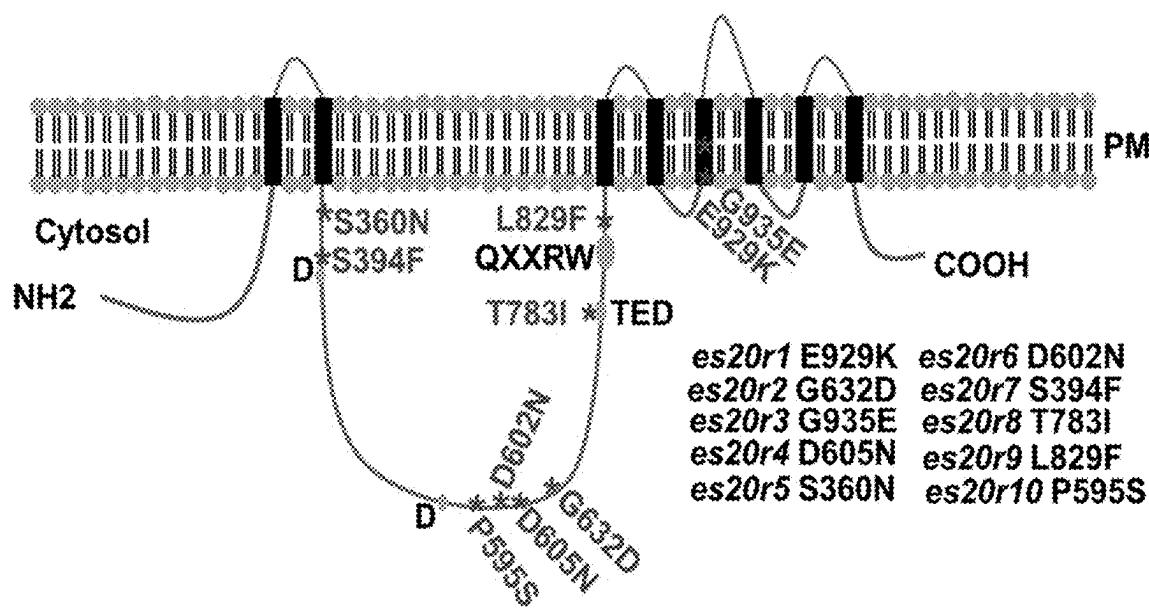
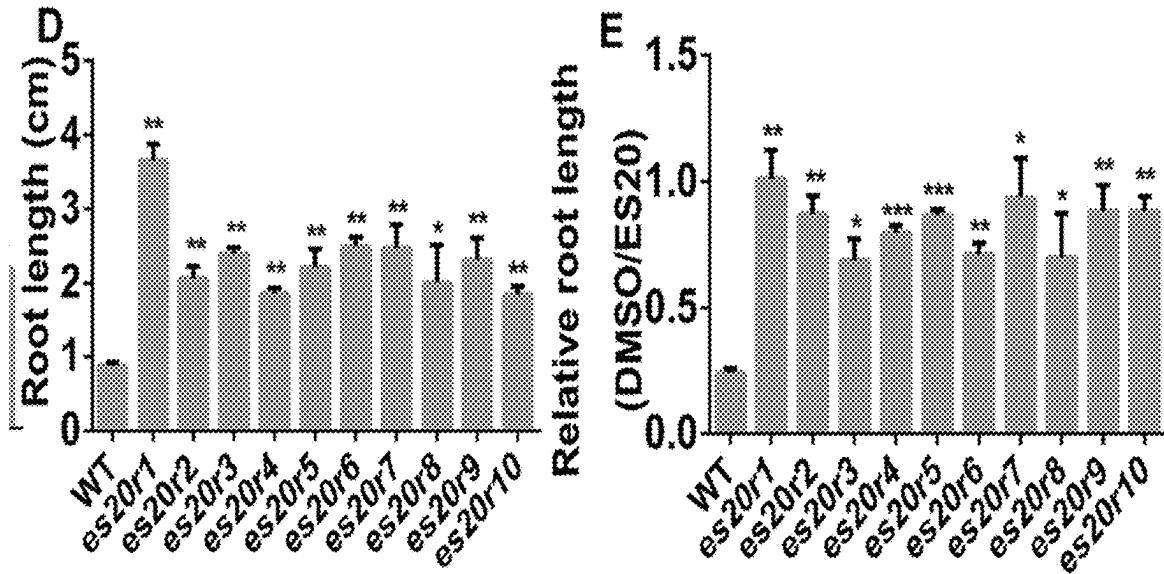


FIG. 2F

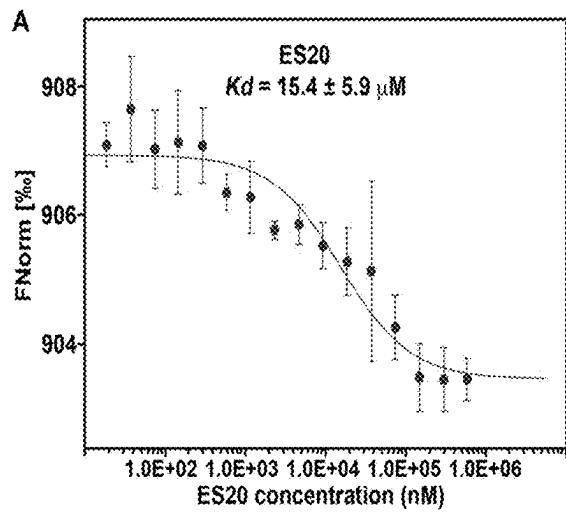


FIG. 3A

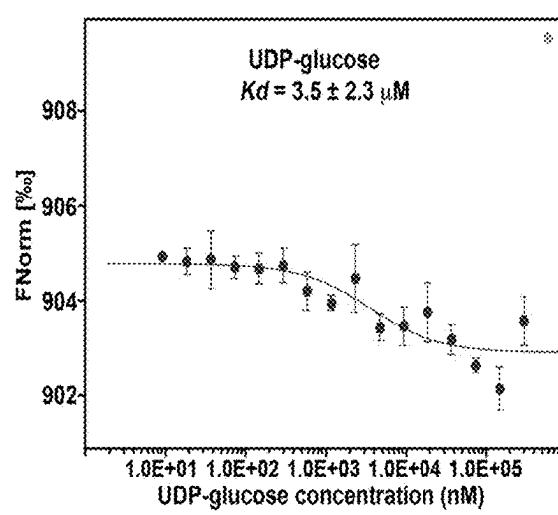


FIG. 3B

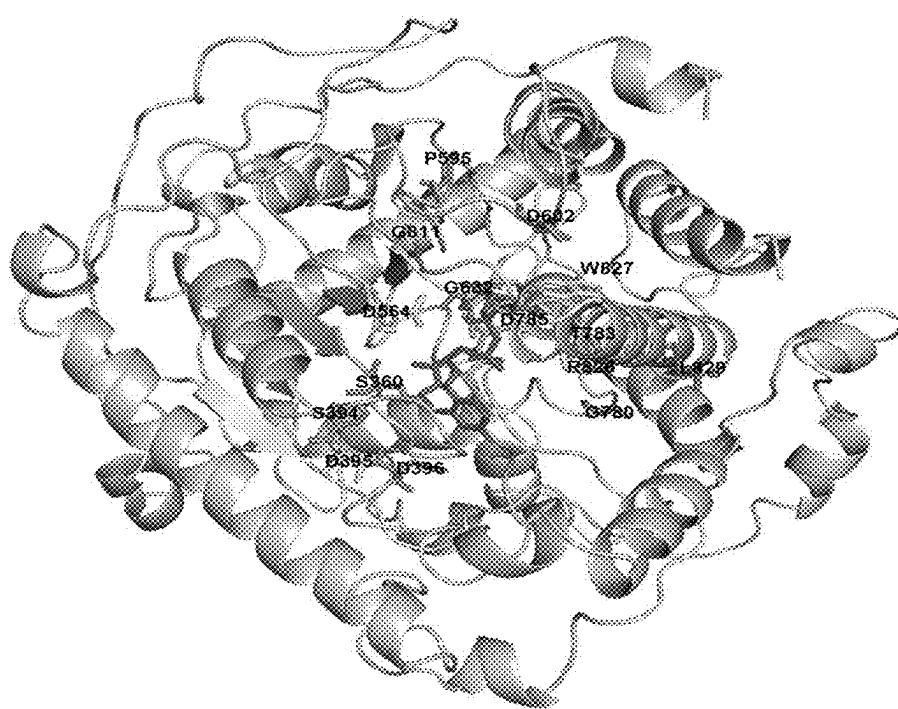


FIG. 3C

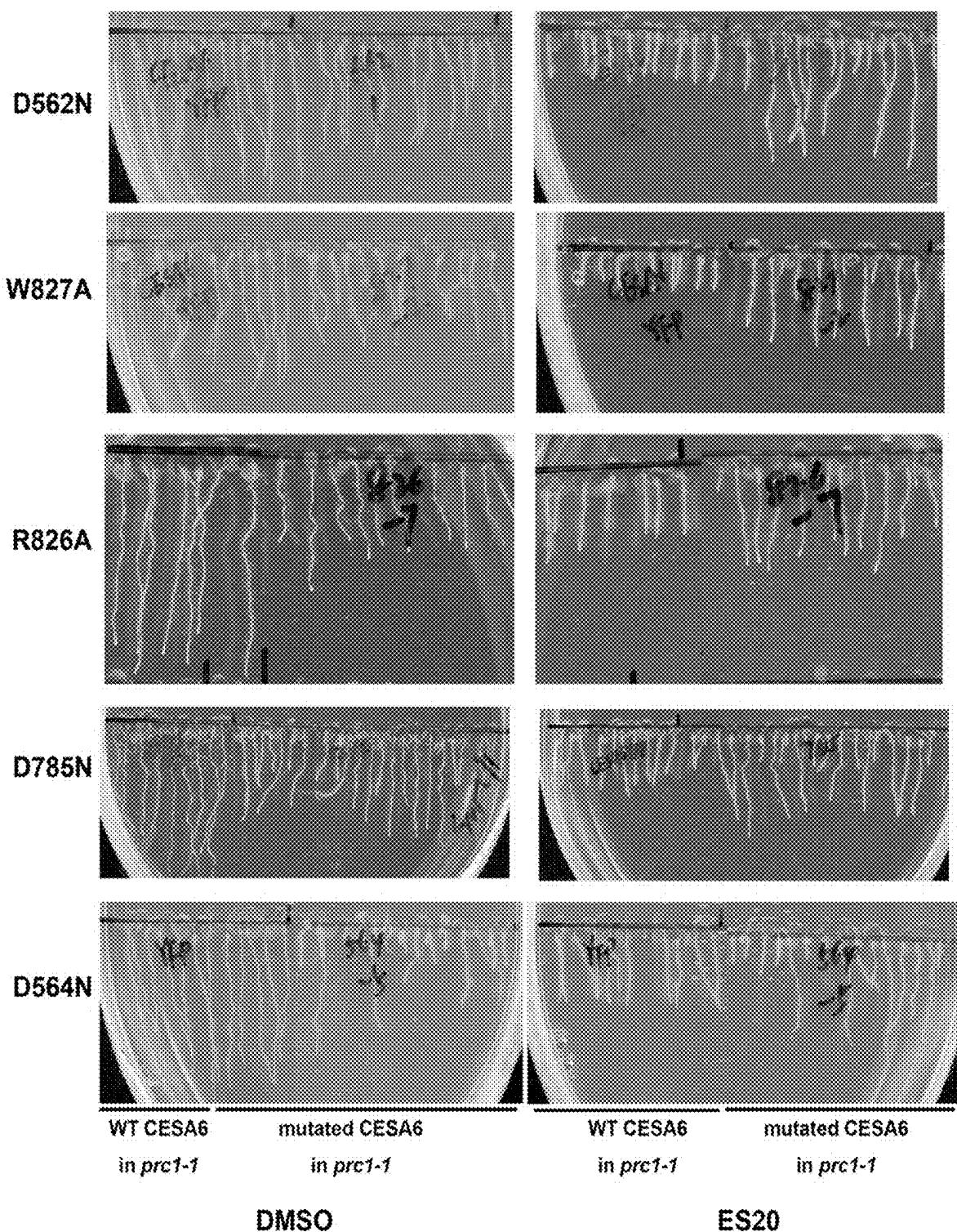


FIG. 4

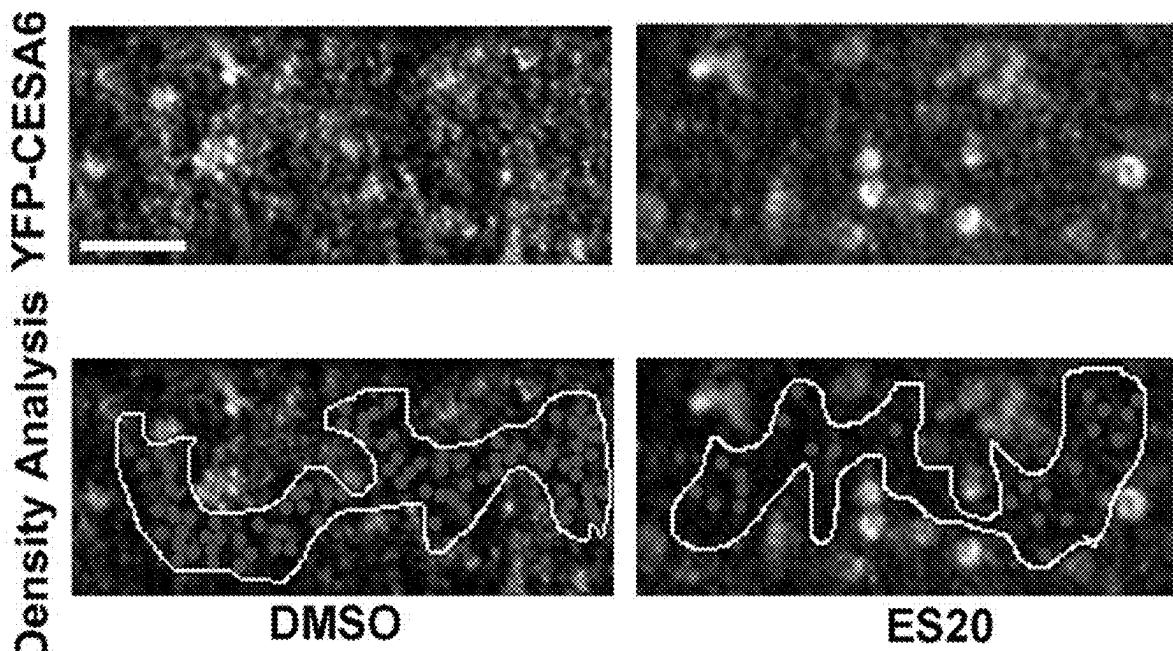


FIG. 5A

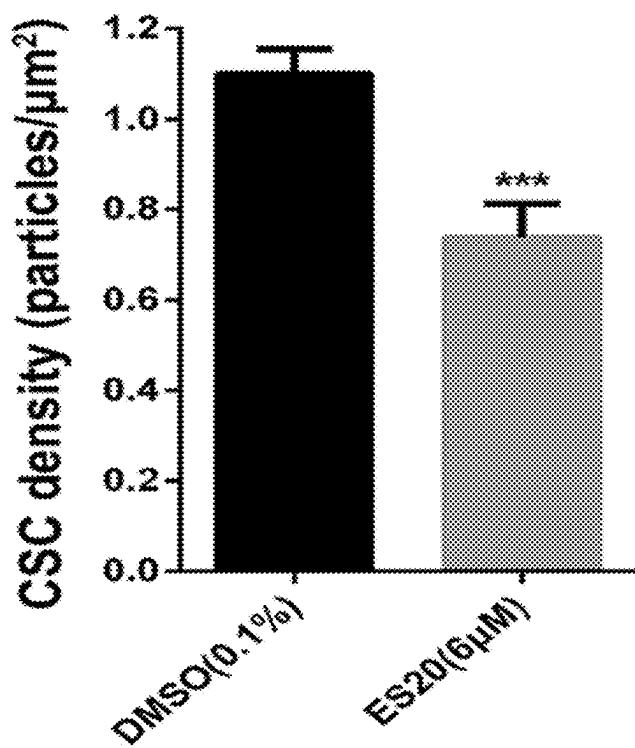


FIG. 5B

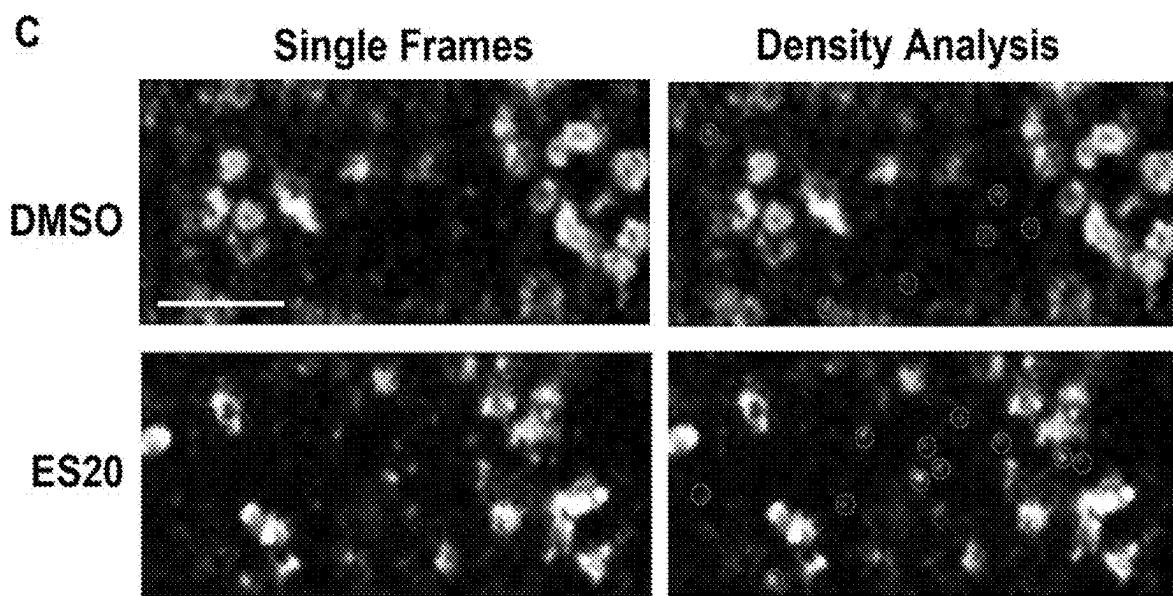


FIG. 5C

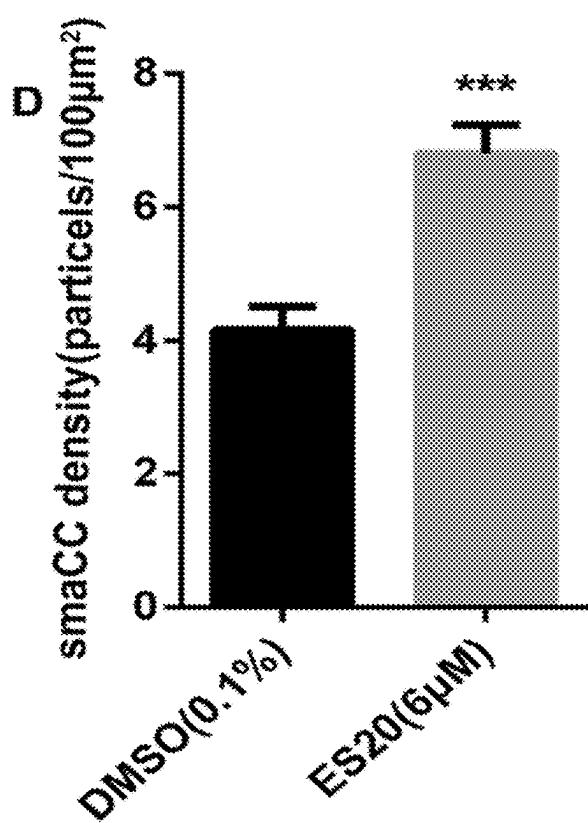


FIG. 5D

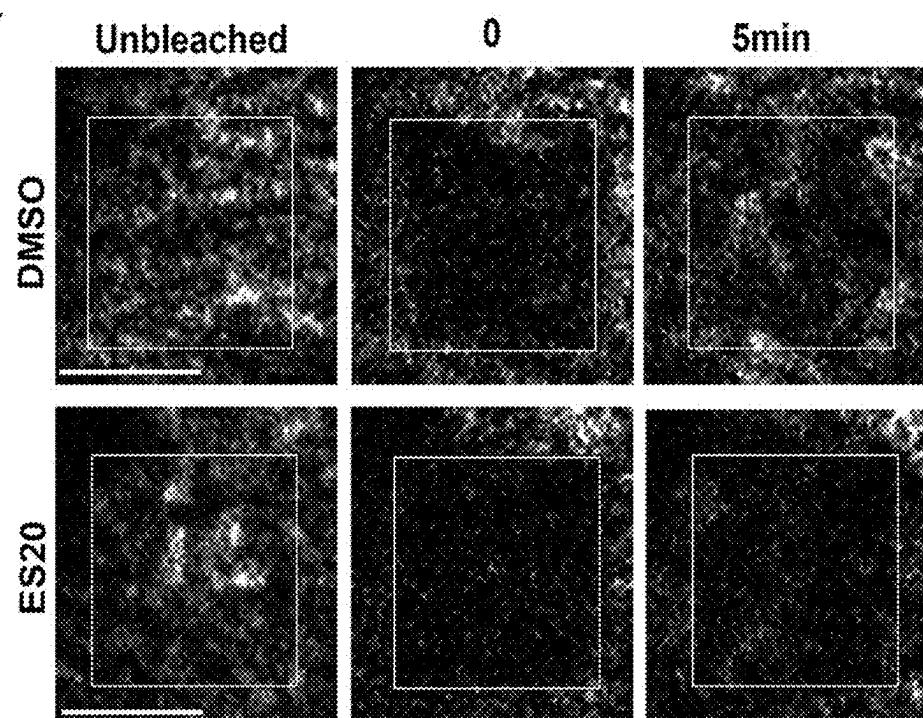


FIG. 5E

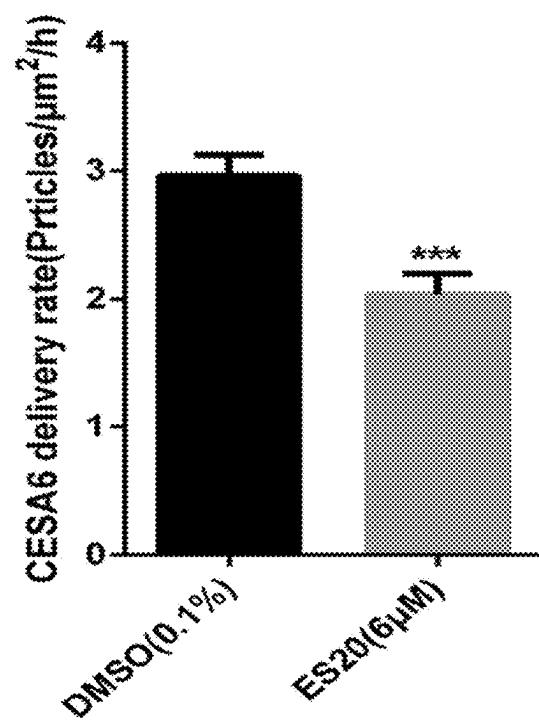


FIG. 5F

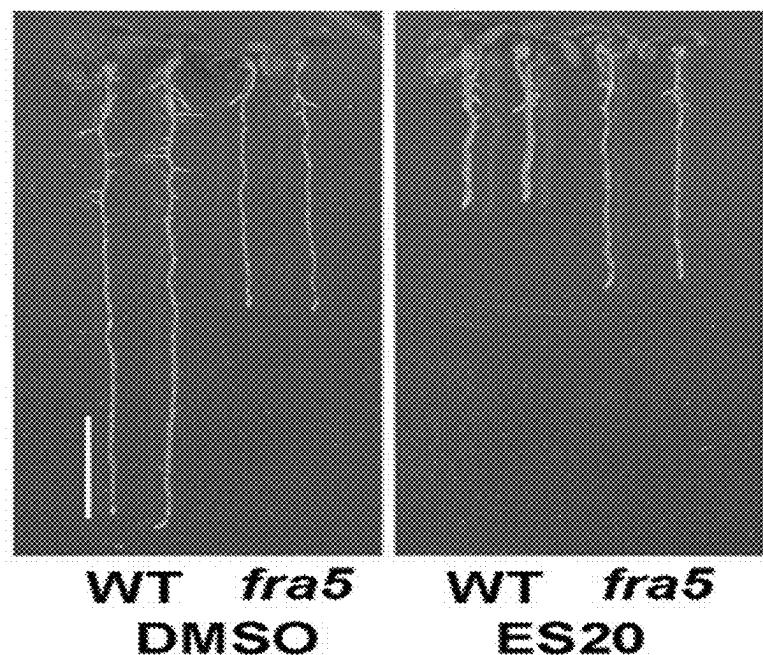


FIG. 6A

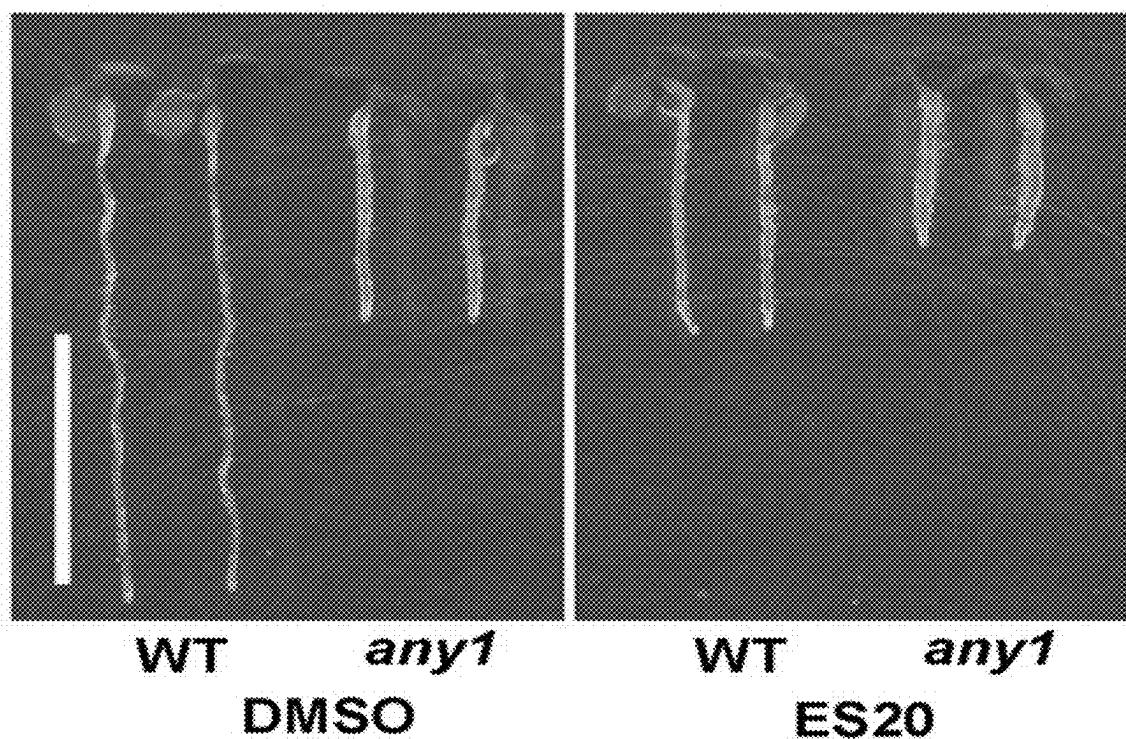


FIG. 6B

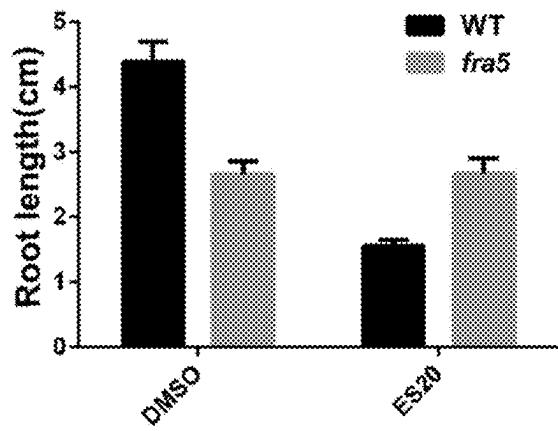


FIG. 6C

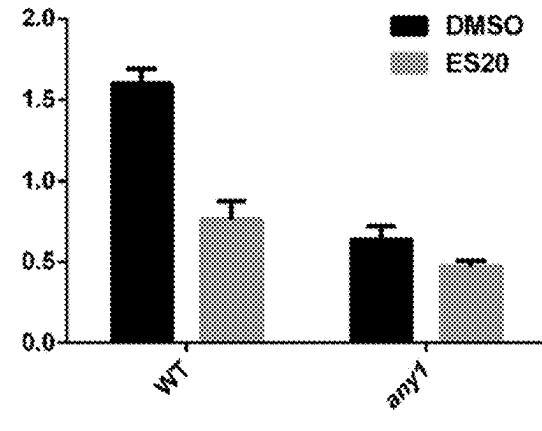


FIG. 6D

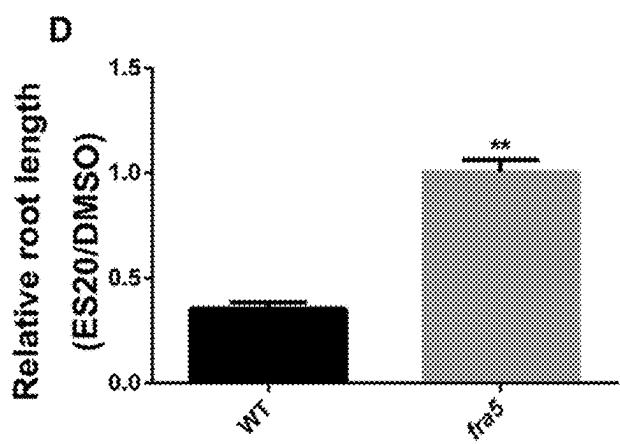


FIG. 6E

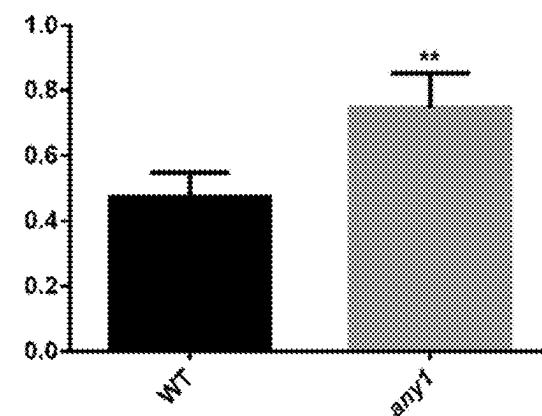


FIG. 6F

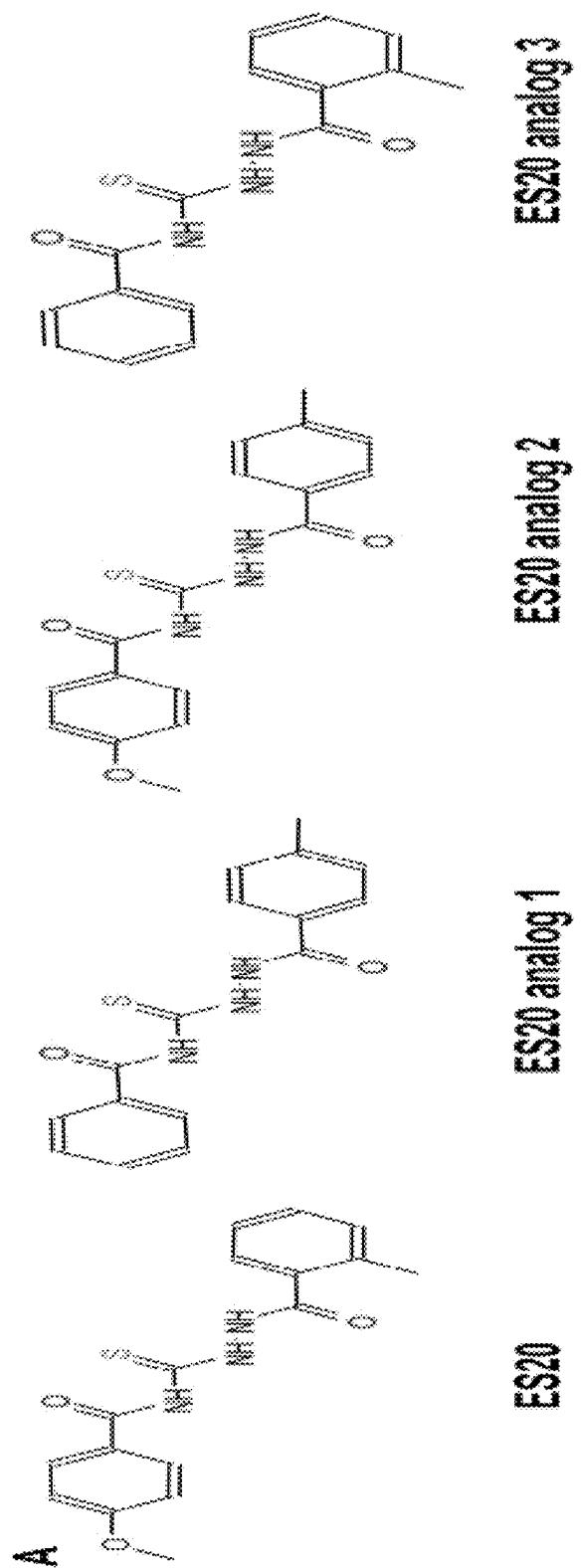


FIG. 7A

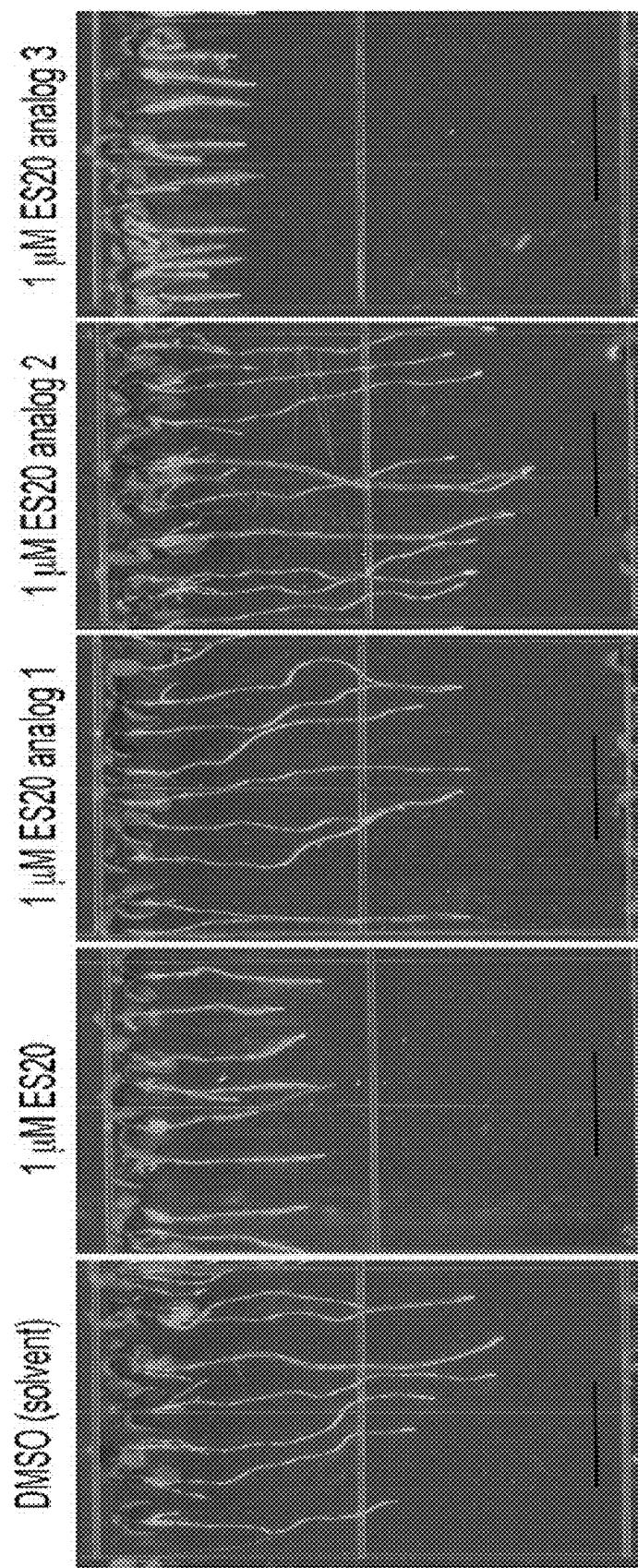


FIG. 7B

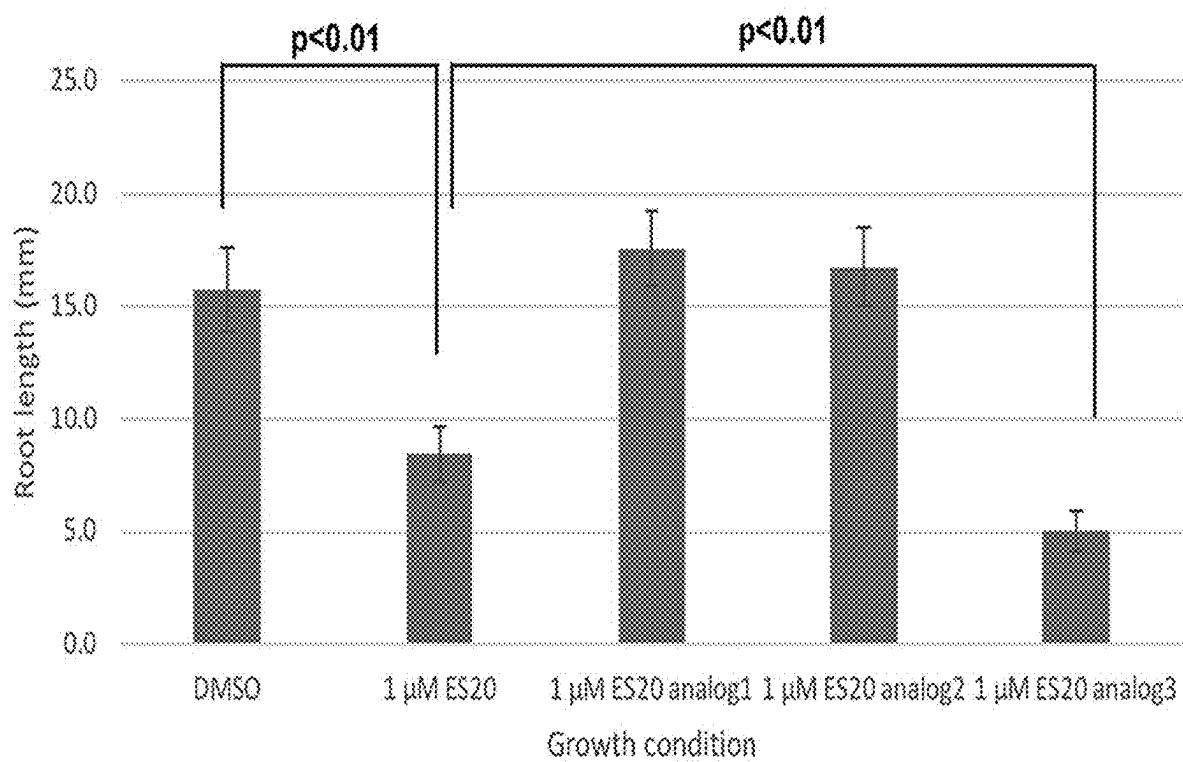


FIG. 7C

FIG. 8A

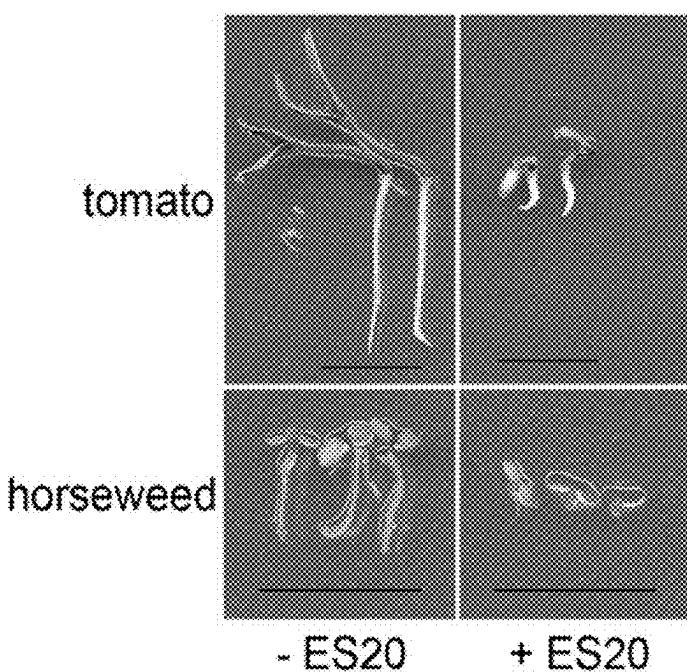
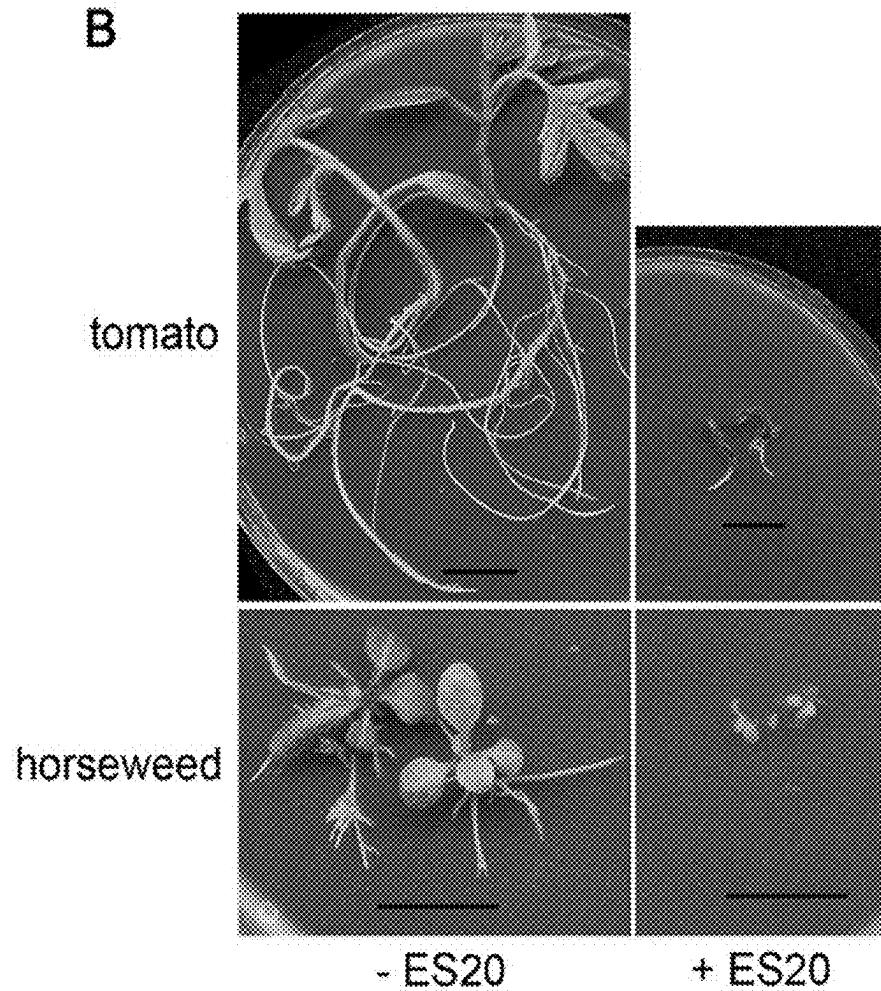


FIG. 8B

B



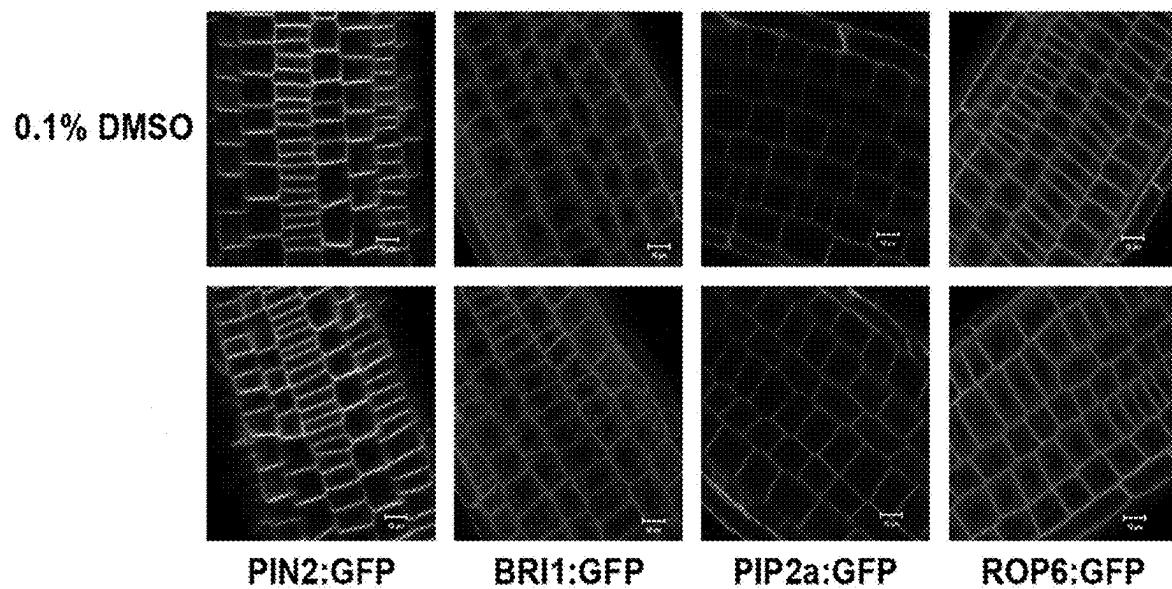


FIG. 9A

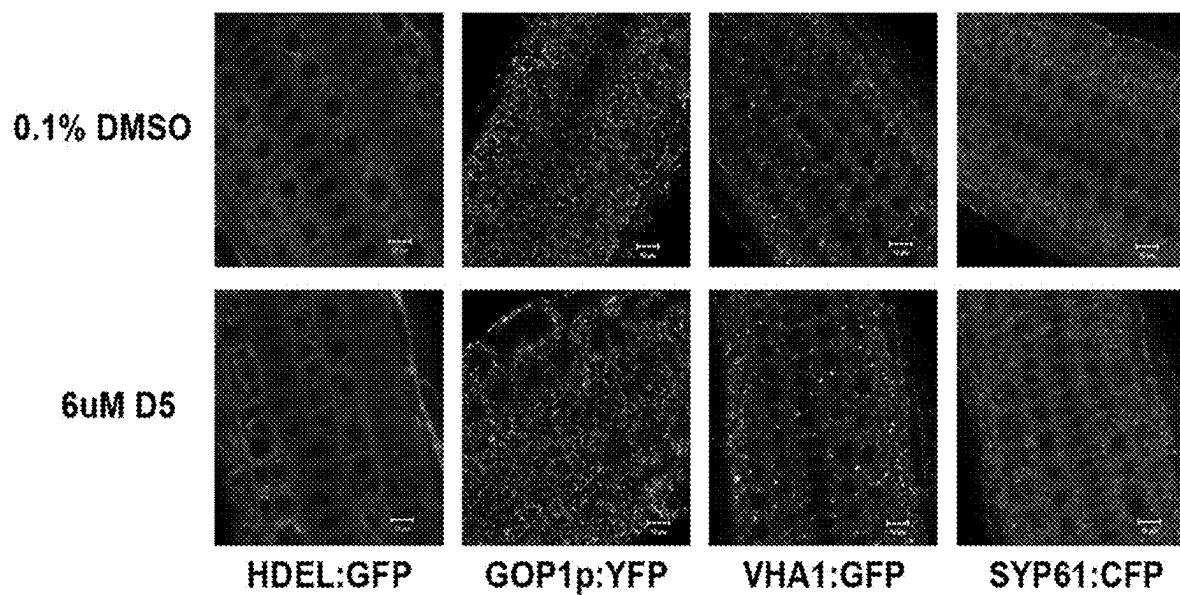


FIG. 9B

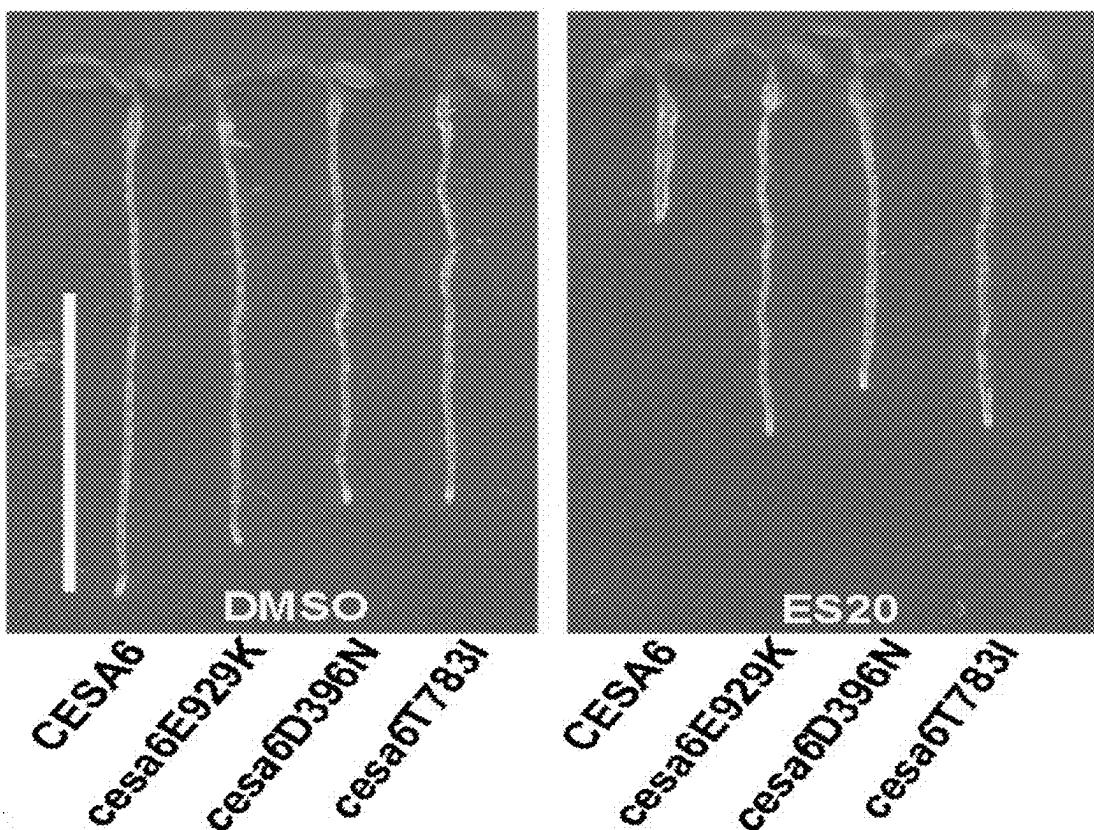


FIG. 10A

FIG. 10B

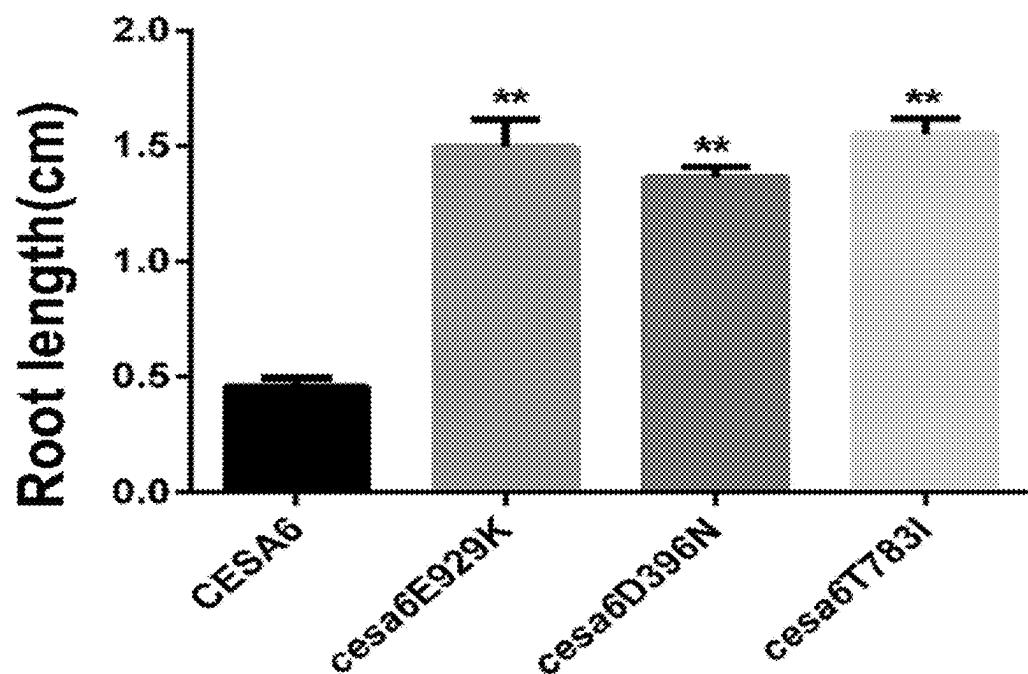


FIG. 10C

FIG. 11A

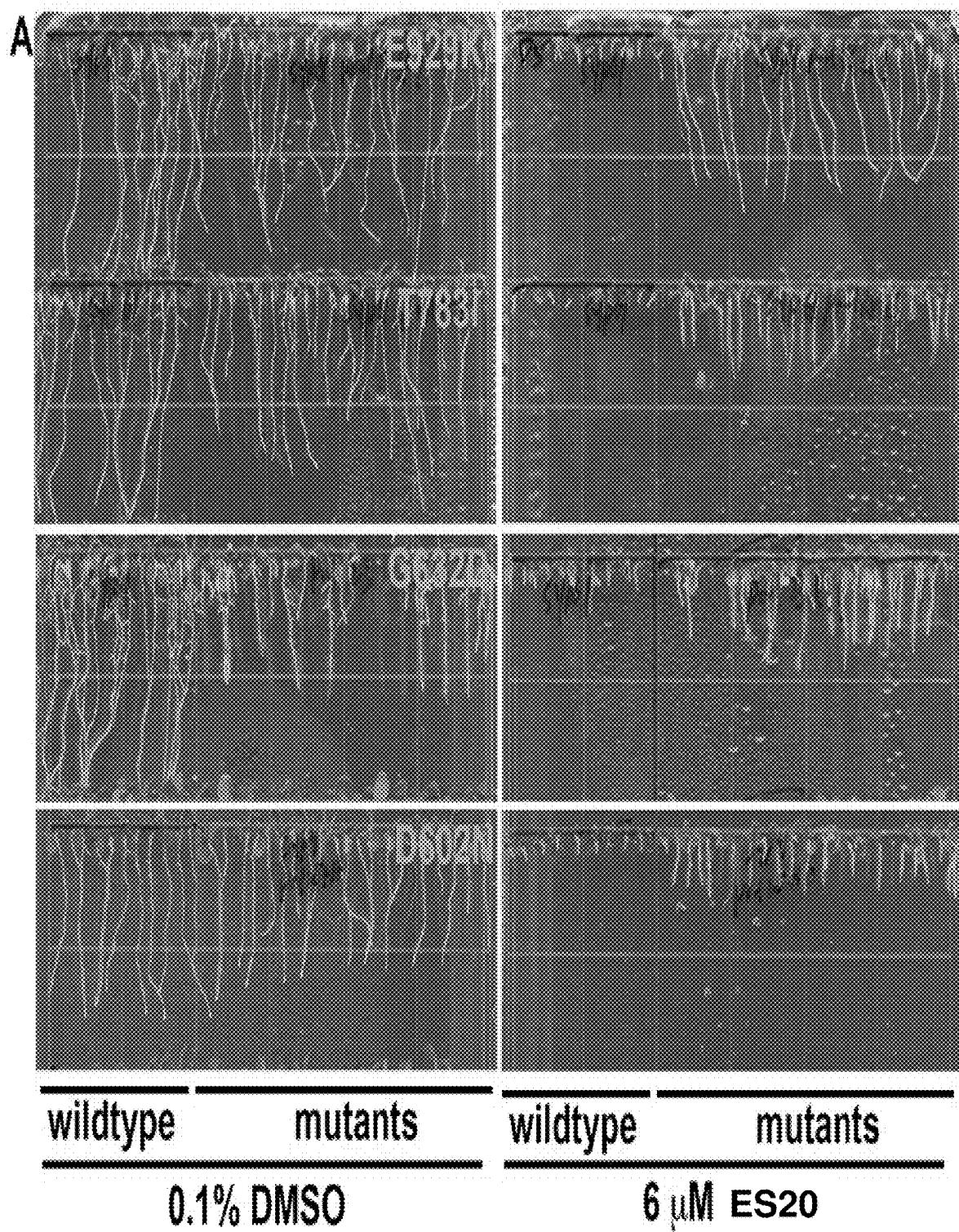


FIG. 11B

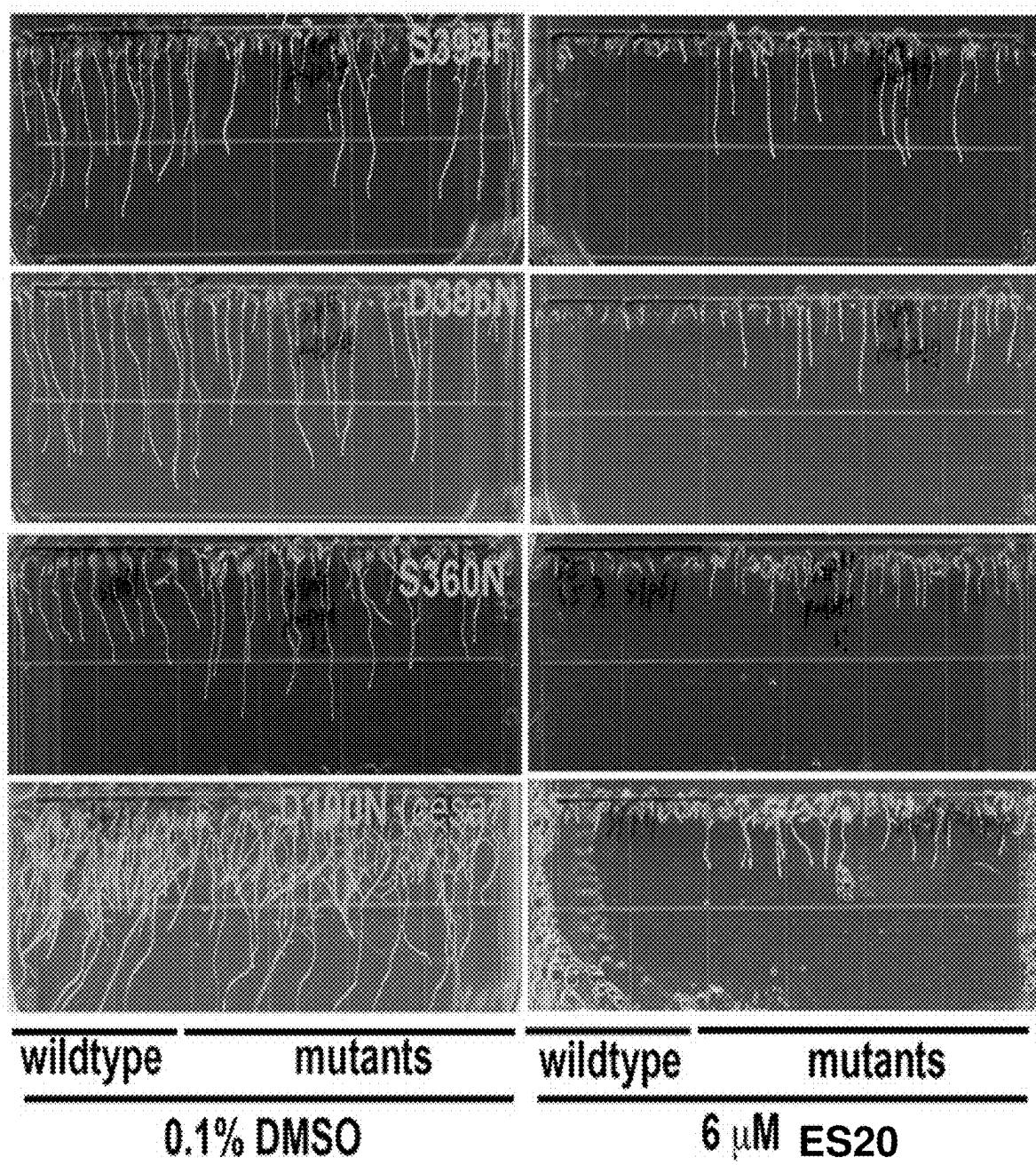


FIG. 11C

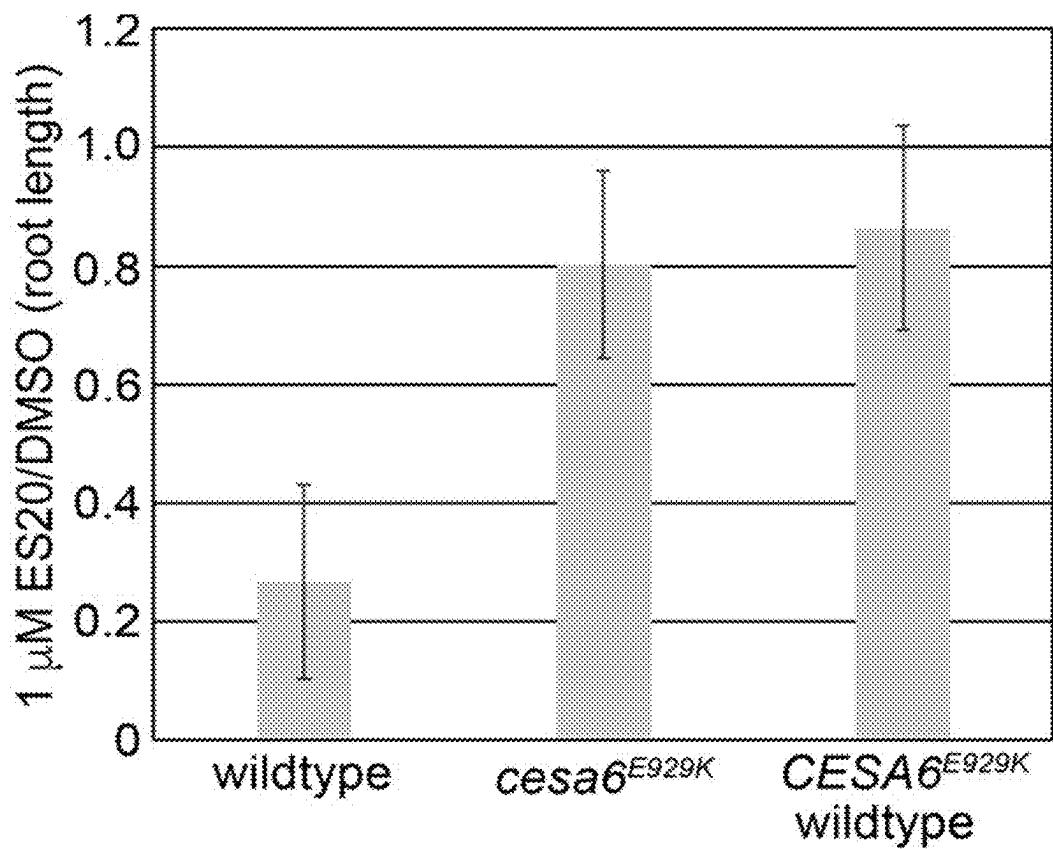


FIG. 11D

CELLULOSE SYNTHASE INHIBITORS AS A NEW CLASS OF HERBICIDE AND NON-GMO CROPS THAT ARE RESISTANT TO THE HERBICIDE

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present U.S. patent application relates to and claims the priority benefit of U.S. Provisional Patent Application Ser. No. 62/588,677, filed Nov. 20, 2017, the content of which is hereby incorporated by reference in its entirety.

STATEMENT OF SEQUENCE LISTING

[0002] A computer-readable form (CRF) of the Sequence Listing is submitted concurrently with this application. The file, generated on Nov. 19, 2018, is entitled Sequence_Listing_68083-02_ST25_txt. Applicant states that the content of the computer-readable form is the same and the information recorded in computer readable form is identical to the written sequence listing.

TECHNICAL FIELD

[0003] The present application relates to composition matters of cellulose synthase inhibitors as a herbicide and mutant genes that convey resistance to those herbicides to a plant that expresses those mutant genes.

BACKGROUNDS AND SUMMARY OF THE INVENTION

[0004] This section introduces aspects that may help facilitate a better understanding of the disclosure. Accordingly, these statements are to be read in this light and are not to be understood as admissions about what is or is not prior art.

[0005] Plant cell wall determines plant cell morphology and is the major composition of plant biomass. Cellulose is the polymer of $\beta(1,4)$ D-glucose and is an essential cell wall component that controls anisotropic plant cell growth. Cellulose is the most abundant biopolymers on earth and provides clothes, food, shelter and energy for human. Although the composition of cellulose is only the linear chain of $\beta(1,4)$ D-glucose units, its mechanisms of biosynthesis in plants is extremely complicated but poorly understood. Cellulose microfibrils are synthesized by cellulose synthase complexes (CSCs) that move along the plasma membrane through the direction of underlying cortical microtubules (Bashline, L. et al., *Plant Physiol.* 2013, 163, 150-160; Bashline, L., et al., *Proc. Natl Acad Sci. USA* 2015, 112, 12870-12875; Parredez, A. R. et al., *Science* 2006, 312, 1491-1495). CSCs are assembled in the Golgi and delivered to the plasma membrane via small secretory vesicles (Crowell, E. F. et al., *Plant Cell* 2009, 21, 1141-1154; Gutierrez, R., et al., *Nat. Cell Biol.* 2009, 11, 797-806). Although the rosette structured CSCs were first observed from freeze fractured maize cells almost forty years ago and plant CesA genes were first cloned from cotton fiber twenty years ago, there are still many knowledge gaps regarding the mechanisms of plant cellulose biosynthesis.

[0006] For example, the mechanisms of spatiotemporal control of CSC trafficking to the plasma membrane, the mechanisms of CSC catalyzing $\beta(1,4)$ D-glucose polymerization, the interaction between CSC and other proteins and lipids on the plasma membrane and the mechanisms of

integration between cellulose polymerization and anisotropic cell growth. CSC is composed of cellulose synthase (CesA) proteins (the number of cellulose synthase protein per complex is still under debate in the cell wall field) that have multiple transmembrane domains and are extremely challenging to manipulate *in vitro*. A good amount of knowledge about plant cellulose synthesis came from phenotypic analysis of different cesA mutants that carry missense mutations (Slabaugh, E. et al., *Trends Plant Sci.* 2014, 19, 99-106; Kumar, M. et al., *Phytochemistry*, 2015, 112, 91-99). Small molecules have also been helpful in understanding cellulose synthesis because some of the missense mutants were identified through chemical genetic screens, although the mechanisms of action of most of these small molecules are not known (Tateno, M. et al., *J Exp Bot* 2016, 67, 533-542). A collection of inhibitors that could target different domains of CesAs will greatly facilitate in understanding the function of these domains.

[0007] Using chemical genetic screen in *Arabidopsis thaliana*, we identified a small molecule Endosidin20 (ES20) that reduces plant growth and inhibits anisotropic cell expansion. Our preliminary data show that ES20 affects CSC trafficking at the plasma membrane and physically interacts with the CesA6 subunit of the CSC. Structure modeling and molecular docking experiments indicate that ES20 could bind to a conserved pocket in the catalytic domain of CesAs. In a ES20-suppression screen, we found seven mutant alleles of *Arabidopsis* CesA6 and one mutant allele of CesA7 that are less sensitive to ES20, but did not find mutants in other CesA genes expressed during primary wall formation. The seven mutant alleles of cesa6 contain missense mutations in amino acids located in the first central cytoplasmic catalytic domain and a poorly characterized smaller second cytoplasmic domain. The one cesa7 allele is located at the N-terminal cytoplasmic region. Our small molecule, live cell imaging tools, protein structure and molecular docking approaches and availability of a group of newly identified different mutant alleles place us in a unique position in understanding cellulose synthase trafficking, cellulose biosynthesis and their relationship to cell growth. In this proposal, we aim to understand the integration of cellulose synthesis and cell growth control by studying the function of seven amino acids that are essential for plants' sensitivity to ES20. We hypothesize that ES20 targets *Arabidopsis thaliana* CesAs, especially CesA6 and CesA7, at the core catalytic domain to affect cellulose synthesis. Amino acids at the core catalytic domain contribute different roles in cellulose biosynthesis and cell shape control. We test our hypothesis using the combination of chemical genetics, live cell imaging, cell wall analysis, biochemical analysis and plant phenotypic characterization.

[0008] Using chemical genetic screening, we discovered a small molecule Cellulosin (aka endosidin20 or ES20) that causes cell swollen and inhibits plant growth, but does not disrupt global vesicle trafficking. By doing mutant screening, we obtained multiple alleles of *Arabidopsis thaliana* that are resistant to Cellulosin inhibition in growth. Those mutated amino acid residues are conserved across plant species. Cellulosin targets a group of cellulose synthases (CesAs) of *Arabidopsis thaliana* by binding to a conserved domain essential for the catalytic activity of CesA. Cellulosin may target and inhibit all subtypes of CesAs in plants. The present invention relates to Cellulosin, a cellulose synthase inhibitor, its analogs or derivatives as a broad-

spectrum herbicide. The mutated genes, their protein products and a cell or a plant having those mutated genes or expressing those protein products are within the scope of this disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] A better understanding of the present invention will be obtained upon reference to the following description in conjunction with the accompanying drawings.

[0010] FIGS. 1A-1J depict ES20 inhibits anisotropic plant cell growth and reduces crystalline cellulose content in dark-grown seedlings. FIG. 1A shows the molecular structure of ES20. FIG. 1B shows ES20 inhibits *Arabidopsis* root growth and causes root swollen in a dosage-dependent manner. *Arabidopsis* wildtype seedlings grown on different concentrations of ES20 are shown. FIG. 1C. Growth curve of *Arabidopsis* wildtype plants grown on different concentrations of ES20. FIG. 1D. The tip region of 5 days old *Arabidopsis* roots treated with DMSO or 3 μ M ES20 for 12 hours. Seedlings treated with ES20 have swollen roots in compare with DMSO control. FIG. 1E. The epidermal cells of 5 days old *Arabidopsis* roots treated with 3 μ M ES20 for 12 hours are swollen in compare with DMSO control. FIG. 1F. ES20 inhibits *Arabidopsis* hypocotyl growth and cause hypocotyl swollen in a dosage dependent manner. *Arabidopsis* wildtype seedlings grown on different concentrations of ES20 under dark condition are shown. FIG. 1G. Quantification on the length of 5 days old dark grown hypocotyls in the presence of different concentrations of ES20. FIG. 1H. Hypocotyls from seedlings grown on ES20 media under dark condition are wider in compare with seedlings grown on DMSO control. FIG. 1I. The hypocotyl epidermal cells of plants grown in the presence of ES20 under dark condition are swollen in compare with DMSO control. J. ES20 reduces the content of crystalline cellulose in dark-grown *Arabidopsis* seedlings in a dosage dependent manner. If there is no common letter between two groups, it indicates they are significant different in crystalline cellulose content by paired t-test, $p<0.05$. Scale bars in FIGS. 1B and 1F are 1.0 cm; Bars in FIGS. 1D and 1H are 100 μ m. Scale bars in FIGS. 1E and 1I are 20 μ m.

[0011] FIGS. 2A-2F depict a group of cesa6 that have reduced sensitivity to ES20. FIG. 2A. 6 days old seedlings of wildtype (WT) and cesa6 mutants (es20r1 to es20r10) that were grown on growth media supplemented with DMSO solvent control. FIG. 2B. 6 days old seedlings of wildtype and cesa6 alleles grown on growth media supplemented with 1 μ M ES20. FIG. 2C. Quantification on the root length of 6 days old wildtype and mutant seedlings grown on growth media supplemented with DMSO solvent control. FIG. 2D. Quantification on the root length of wildtype and mutant seedlings grown on growth media supplemented with 1 μ M ES20. FIG. 2E. The ratio of root length from plants grown on 1 μ M ES20 and those grown on DMSO for each genotype. The root length of wildtype plants grown on 1 μ M ES20 was reduced to about 25% of those grown on DMSO. However, in some mutants, the root growth was not reduced at all (the ratio was close to 1) by 1 μ M ES20 or was slightly reduced but was not as significant as that of the wildtype plants. FIG. 2F. The diagram shows the topology of CesA6, the location, and the mutated amino acid in each mutant allele.

[0012] FIGS. 3A-3C demonstrate that ES20 directly interacts with CesA6 at the catalytic core. FIG. 3A. The ther-

mophoresis binding curve for the interaction of CesA6 central cytoplasmic domain with different concentrations of ES20. FIG. 3B. The thermophoresis binding curve for the interaction of CesA6 central cytoplasmic domain with different concentrations of UDP-glucose. FIG. 3C. Hypothesized binding site for ES20 (cyan) and UDP-glucose (magenta) on modeled structure of CesA6 central cytoplasmic domain. The amino acids highlighted by green indicate the ones found to have been mutated in our reduced sensitivity mutants. The amino acids highlighted by yellow indicate the ones that we mutated based on hypothesized binding site and have caused reduced sensitivity to ES20 in plants.

[0013] FIG. 4. Mutations potentially reduce sensitivity of a plant to ES20 and transgenic plants that express mutated CesA6 in prc1-1 genetic background have reduced sensitivity to ES20.

[0014] FIGS. 5A-5E show that ES20 treatment inhibits exocytic trafficking of CesA6. FIGS. 5A and 5B. The density of plasma membrane localized CSCs labeled by YFP-CesA6 was significantly reduced by ES20 treatment. FIGS. 5C and 5D. The numbers of SmaCCs were significantly increased by ES20 treatment. FIGS. 5E and 5F. The exocytic delivery of CSC to the plasma membrane was reduced after ES20 treatment. *** indicates $p<0.001$ by paired t-test.

[0015] FIGS. 6A-6F demonstrate that ES20 targets multiple CesAs in plants. Plants that carry a P557T mutation in CesA7 (fra5) and plants carry a D604N mutation in CesA1 (any1) are less sensitive to ES20. FIG. 6A. Seedlings of Wildtype and fra5 grown on growth media supplemented with DMSO or ES20. FIG. 6B. Seedlings of Wildtype and any1 grown on growth media supplemented with DMSO or ES20. FIGS. 6C, 6D, 6E and 6F depict quantification on the growth of wildtype, fra1 and any1 in the presence or absence of ES20.

[0016] FIGS. 7A-7C show that ES20 analog 3 is more potent than ES20 in inhibiting plant growth. FIG. 7A. Molecular structures of ES20 and analogs. FIG. 7B. The effect of ES20 and its analogs on *Arabidopsis* root growth. 7-day-old *Arabidopsis* seedlings grown endoc Cellulosin and analogs are shown. Bars, 1 cm. FIG. 7C. Quantification on the effects of Cellulosin and analogs on *Arabidopsis* root growth.

[0017] FIGS. 8A-8B demonstrate that ES20 is active in inhibiting tomato and horseweed growth. FIG. 8A shows 6-day-old tomato and horseweed grown on the media in the absence and presence of Cellulosin. FIG. 8B depicts 16-day-old tomato and horseweed grown on the media in the absence and presence of Cellulosin. Bars, 1 cm.

[0018] FIGS. 9A and 9B demonstrate that ES20 does not disrupt general endomembrane system of plants. Transgenic plants expressing different fluorescence-tagged proteins were examined for their cellular localization in the DMSO control treatment and after 2 hours of 5 μ M ES20 treatment.

[0019] FIGS. 10A, 10B and 10C depict expression of CesA6 with missense mutations in prc1-1 background is sufficient to create plants with reduced sensitivity to ES20.

[0020] FIG. 11A shows locations of point mutations in our cesa6 alleles. The sequence alignment was done by Slabaugh et al. (Slabaugh, D., et al., *Trends Plant Sci* 2014, 19, 99-106) and was adopted here to show the locations of mutations: Gray shade indicates predicted transmembrane helices, blue shade indicates BcsA interfacial helices, orange circles denote the locations of known missense mutations in AtCESAs. Orange squares are the mutations we discovered.

[0021] FIG. 11B shows comparison of seedling growth between wildtype and different mutants on DMSO and Cellulosin. The mutation in each mutant is labeled. All mutants are alleles of cesa6 except D100N that is a mutant of cesa7.

[0022] FIG. 11C shows demonstrates expression of mutant CesA construct in wildtype plants leads to resistance to ES20. The root length of wildtype plants grown on 1 μ M ES20 is about 20% of those that are grown on normal media.

[0023] FIG. 11D shows root growth of mutants CesA6 and the wildtype in the presence of 1 μ M of ES20 or DMSO as a control. The root length of cesa6E929k mutant and transgenic plants that express CesA6F929K in wildtype is about 80% of those grown on normal media.

DETAILED DESCRIPTION

[0024] For the purposes of promoting an understanding of the principles of the present disclosure, reference will now be made to the embodiments illustrated in the drawings, and specific language will be used to describe the same. It will nevertheless be understood that no limitation of the scope of this disclosure is thereby intended.

[0025] As used herein, the following terms and phrases shall have the meanings set forth below. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art.

[0026] In the present disclosure the term "about" can allow for a degree of variability in a value or range, for example, within 20%, within 10%, within 5%, or within 1% of a stated value or of a stated limit of a range. In the present disclosure, the term "substantially" can allow for a degree of variability in a value or range, for example, within 80%, within 90%, within 95%, 99%, 99.5%, 99.9%, 99.99%, or at least about 99.999% or more of a stated value or of a stated limit of a range.

[0027] In this document, the terms "a," "an," or "the" are used to include one or more than one unless the context clearly dictates otherwise. The term "or" is used to refer to a nonexclusive "or" unless otherwise indicated. In addition, it is to be understood that the phraseology or terminology employed herein, and not otherwise defined, is for the purpose of description only and not of limitation. Any use of section headings is intended to aid reading of the document and is not to be interpreted as limiting. Further, information that is relevant to a section heading may occur within or outside of that particular section. Furthermore, all publications, patents, and patent documents referred to in this document are incorporated by reference herein in their entirety, as though individually incorporated by reference. In the event of inconsistent usages between this document and those documents so incorporated by reference, the usage in the incorporated reference should be considered supplementary to that of this document; for irreconcilable inconsistencies, the usage in this document controls.

[0028] "Identity," as is well understood in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as determined by the match between strings of such sequences. Methods to determine "identity" are designed to give the largest match between the sequences tested. Moreover, methods to determine identity

are codified in publicly available programs. Computer programs can be used to determine "identity" between two sequences these programs include but are not limited to, GCG; suite of five BLAST programs, three designed for nucleotide sequences queries (BLASTN, BLASTX, and TBLASTX) and two designed for protein sequence queries (BLASTP and TBLASTN). The BLAST X program is publicly available from NCBI and other sources. The well-known Smith Waterman algorithm can also be used to determine identity.

[0029] The term "substituted" as used herein refers to a functional group in which one or more hydrogen atoms contained therein are replaced by one or more non-hydrogen atoms. The term "functional group" or "substituent" as used herein refers to a group that can be or is substituted onto a molecule. Examples of substituents or functional groups include, but are not limited to, a halogen (e.g., F, Cl, Br, and I); an oxygen atom in groups such as hydroxyl groups, alkoxy groups, aryloxy groups, aralkyloxy groups, oxo (carbonyl) groups, carboxyl groups including carboxylic acids, carboxylates, and carboxylate esters; a sulfur atom in groups such as thiol groups, alkyl and aryl sulfide groups, sulfoxide groups, sulfone groups, sulfonyl groups, and sulfonamide groups; a nitrogen atom in groups such as amines, azides, hydroxylamines, cyano, nitro groups, N-oxides, hydrazides, and enamines; and other heteroatoms in various other groups.

[0030] The term "alkyl" as used herein refers to substituted or unsubstituted straight chain and branched alkyl groups and cycloalkyl groups having from 1 to about 20 carbon atoms (C_1-C_{20}), 1 to 12 carbons (C_1-C_{12}), 1 to 8 carbon atoms (C_1-C_8), or, in some embodiments, from 1 to 6 carbon atoms (C_1-C_6). Examples of straight chain alkyl groups include those with from 1 to 8 carbon atoms such as methyl, ethyl, n-propyl, n-butyl, n-pentyl, n-hexyl, n-heptyl, and n-octyl groups. Examples of branched alkyl groups include, but are not limited to, isopropyl, iso-butyl, sec-butyl, t-butyl, neopentyl, isopentyl, and 2,2-dimethylpropyl groups. As used herein, the term "alkyl" encompasses n-alkyl, isoalkyl, and anteisoalkyl groups as well as other branched chain forms of alkyl. Representative substituted alkyl groups can be substituted one or more times with any of the groups listed herein, for example, amino, hydroxy, cyano, carboxy, nitro, thio, alkoxy, and halogen groups.

[0031] The term "alkenyl" as used herein refers to substituted or unsubstituted straight chain and branched divalent alkenyl and cycloalkenyl groups having from 2 to 20 carbon atoms (C_2-C_{20}), 2 to 12 carbons (C_2-C_{12}), 2 to 8 carbon atoms (C_2-C_8) or, in some embodiments, from 2 to 4 carbon atoms (C_2-C_4) and at least one carbon-carbon double bond. Examples of straight chain alkenyl groups include those with from 2 to 8 carbon atoms such as **13** $CH=CH-$, $-CH=CHCH_2-$, and the like. Examples of branched alkenyl groups include, but are not limited to, $-CH=C(CH_3)-$ and the like.

[0032] An alkynyl group is the fragment, containing an open point of attachment on a carbon atom that would form if a hydrogen atom bonded to a triply bonded carbon is removed from the molecule of an alkyne. The term "hydroxalkyl" as used herein refers to alkyl groups as defined herein substituted with at least one hydroxyl ($-OH$) group.

[0033] The term "cycloalkyl" as used herein refers to substituted or unsubstituted cyclic alkyl groups such as, but

not limited to, cyclopropyl, cyclobutyl, cyclopentyl, cyclohexyl, cycloheptyl, and cyclooctyl groups. In some embodiments, the cycloalkyl group can have 3 to about 8-12 ring members, whereas in other embodiments the number of ring carbon atoms range from 3 to 4, 5, 6, or 7. In some embodiments, cycloalkyl groups can have 3 to 6 carbon atoms (C_3-C_6). Cycloalkyl groups further include polycyclic cycloalkyl groups such as, but not limited to, norbornyl, adamantlyl, bornyl, camphenyl, isocamphenyl, and carenyl groups, and fused rings such as, but not limited to, decalinyl, and the like.

[0034] The term “acyl” as used herein refers to a group containing a carbonyl moiety wherein the group is bonded via the carbonyl carbon atom. The carbonyl carbon atom is also bonded to another carbon atom, which can be part of a substituted or unsubstituted alkyl, aryl, aralkyl cycloalkyl, cycloalkylalkyl, heterocyclyl, heterocyclylalkyl, heteroaryl, heteroarylalkyl group or the like. In the special case wherein the carbonyl carbon atom is bonded to a hydrogen, the group is a “formyl” group, an acyl group as the term is defined herein. An acyl group can include 0 to about 12-40, 6-10, 1-5 or 2-5 additional carbon atoms bonded to the carbonyl group. An acryloyl group is an example of an acyl group. An acyl group can also include heteroatoms within the meaning here. A nicotinoyl group (pyridyl-3-carbonyl) is an example of an acyl group within the meaning herein. Other examples include acetyl, benzoyl, phenylacetyl, pyridylacetyl, cinnamoyl, and acryloyl groups and the like. When the group containing the carbon atom that is bonded to the carbonyl carbon atom contains a halogen, the group is termed a “haloacyl” group. An example is a trifluoroacetyl group.

[0035] The term “aryl” as used herein refers to substituted or unsubstituted cyclic aromatic hydrocarbons that do not contain heteroatoms in the ring. Thus aryl groups include, but are not limited to, phenyl, azulenyl, heptalenyl, biphenyl, indacenyl, fluorenyl, phenanthrenyl, triphenylenyl, pyrenyl, naphthacenyl, chrysene, biphenylenyl, anthracenyl, and naphthyl groups. In some embodiments, aryl groups contain about 6 to about 14 carbons (C_6-C_{14}) or from 6 to 10 carbon atoms (C_6-C_{10}) in the ring portions of the groups. Aryl groups can be unsubstituted or substituted, as defined herein. Representative substituted aryl groups can be mono-substituted or substituted more than once, such as, but not limited to, 2-, 3-, 4-, 5-, or 6-substituted phenyl or 2-8 substituted naphthyl groups, which can be substituted with carbon or non-carbon groups such as those listed herein.

[0036] The term “aralkyl” and “arylalkyl” as used herein refers to alkyl groups as defined herein in which a hydrogen or carbon bond of an alkyl group is replaced with a bond to an aryl group as defined herein. Representative aralkyl groups include benzyl and phenylethyl groups and fused (cycloalkylaryl) alkyl groups such as 4-ethyl-indanyl. Aralkenyl groups are alkenyl groups as defined herein in which a hydrogen or carbon bond of an alkyl group is replaced with a bond to an aryl group as defined herein.

[0037] The term “alkoxy” as used herein refers to an oxygen atom connected to an alkyl group, including a cycloalkyl group, as are defined herein. Examples of linear alkoxy groups include but are not limited to methoxy, ethoxy, propoxy, butoxy, pentyloxy, hexyloxy, and the like. Examples of branched alkoxy include but are not limited to isopropoxy, sec-butoxy, tert-butoxy, isopentyloxy, isohexyloxy, and the like. Examples of cyclic alkoxy include but are not limited to cyclopropoxy, cyclobutyloxy, cyclopenty-

loxy, cyclohexyloxy, and the like. An alkoxy group can further include double or triple bonds, and can also include heteroatoms. For example, an allyloxy group is an alkoxy group within the meaning herein. A methoxyethoxy group is also an alkoxy group within the meaning herein, as is a methylenedioxy group in a context where two adjacent atoms of a structure are substituted therewith.

[0038] The term “amine” as used herein refers to primary, secondary, and tertiary amines having, e.g., the formula $N(\text{group})_3$ wherein each group can independently be H or non-H, such as alkyl, aryl, and the like. Amines include but are not limited to $R-\text{NH}_2$, for example, alkylamines, arylamines, alkylarylamines; $R_2\text{NH}$ wherein each R is independently selected, such as dialkylamines, diarylamines, aralkylamines, heterocyclylamines and the like; and $R_3\text{N}$ wherein each R is independently selected, such as trialkylamines, dialkylarylamines, alkyldiarylamines, triarylamines, and the like. The term “amine” also includes ammonium ions as used herein.

[0039] The term “amino group” as used herein refers to a substituent of the form $-\text{NH}_2$, $-\text{NHR}$, $-\text{NR}_2$, $-\text{NR}_3^+$, wherein each R is independently selected, and protonated forms of each, except for $-\text{NR}_3^+$, which cannot be protonated. Accordingly, any compound substituted with an amino group can be viewed as an amine. An “amino group” within the meaning herein can be a primary, secondary, tertiary, or quaternary amino group. An “alkylamino” group includes a monoalkylamino, dialkylamino, and trialkylamino group.

[0040] The terms “halo,” “halogen,” or “halide” group, as used herein, by themselves or as part of another substituent, mean, unless otherwise stated, a fluorine, chlorine, bromine, or iodine atom.

[0041] The term “haloalkyl” group, as used herein, includes mono-halo alkyl groups, poly-halo alkyl groups wherein all halo atoms can be the same or different, and per-halo alkyl groups, wherein all hydrogen atoms are replaced by halogen atoms, such as fluoro. Examples of haloalkyl include trifluoromethyl, 1,1-dichloroethyl, 1,2-dichloroethyl, 1,3-dibromo-3,3-difluoropropyl, perfluorobutyl, $-\text{CF}(\text{CH}_3)_2$ and the like.

[0042] The term “optionally substituted,” or “optional substituents,” as used herein, means that the groups in question are either unsubstituted or substituted with one or more of the substituents specified. When the groups in question are substituted with more than one substituent, the substituents may be the same or different. When using the terms “independently,” “independently are,” and “independently selected from” mean that the groups in question may be the same or different. Certain of the herein defined terms may occur more than once in the structure, and upon such occurrence each term shall be defined independently of the other.

[0043] The compounds described herein may contain one or more chiral centers, or may otherwise be capable of existing as multiple stereoisomers. It is to be understood that in one embodiment, the invention described herein is not limited to any particular stereochemical requirement, and that the compounds, and compositions, methods, uses, and medicaments that include them may be optically pure, or may be any of a variety of stereoisomeric mixtures, including racemic and other mixtures of enantiomers, other mixtures of diastereomers, and the like. It is also to be understood that such mixtures of stereoisomers may include a

single stereochemical configuration at one or more chiral centers, while including mixtures of stereochemical configuration at one or more other chiral centers.

[0044] Similarly, the compounds described herein may include geometric centers, such as cis, trans, E, and Z double bonds. It is to be understood that in another embodiment, the invention described herein is not limited to any particular geometric isomer requirement, and that the compounds, and compositions, methods, uses, and medicaments that include them may be pure, or may be any of a variety of geometric isomer mixtures. It is also to be understood that such mixtures of geometric isomers may include a single configuration at one or more double bonds, while including mixtures of geometry at one or more other double bonds.

[0045] As used herein, the term "salts" and "pharmaceutically acceptable salts" refer to derivatives of the disclosed compounds wherein the parent compound is modified by making acid or base salts thereof. Examples of pharmaceutically acceptable salts include, but are not limited to, mineral or organic acid salts of basic groups such as amines; and alkali or organic salts of acidic groups such as carboxylic acids. Pharmaceutically acceptable salts include the conventional non-toxic salts or the quaternary ammonium salts of the parent compound formed, for example, from non-toxic inorganic or organic acids. For example, such conventional non-toxic salts include those derived from inorganic acids such as hydrochloric, hydrobromic, sulfuric, sulfamic, phosphoric, and nitric; and the salts prepared from organic acids such as acetic, propionic, succinic, glycolic, stearic, lactic, malic, tartaric, citric, ascorbic, pamoic, maleic, hydroxymaleic, phenylacetic, glutamic, benzoic, salicylic, sulfanilic, 2-acetoxybenzoic, fumaric, toluenesulfonic, methanesulfonic, ethane disulfonic, oxalic, and isethionic, and the like.

[0046] Pharmaceutically acceptable salts can be synthesized from the parent compound which contains a basic or acidic moiety by conventional chemical methods. In some instances, such salts can be prepared by reacting the free acid or base forms of these compounds with a stoichiometric amount of the appropriate base or acid in water or in an organic solvent, or in a mixture of the two; generally, nonaqueous media like ether, ethyl acetate, ethanol, isopropanol, or acetonitrile are preferred. Lists of suitable salts are found in Remington's Pharmaceutical Sciences, 17th ed., Mack Publishing Company, Easton, Pa., 1985, the disclosure of which is hereby incorporated by reference.

[0047] The term "pharmaceutically acceptable carrier" is art-recognized and refers to a pharmaceutically-acceptable material, composition or vehicle, such as a liquid or solid filler, diluent, excipient, solvent or encapsulating material, involved in carrying or transporting any subject composition or component thereof. Each carrier must be "acceptable" in the sense of being compatible with the subject composition and its components and not injurious to the patient. Some examples of materials which may serve as pharmaceutically acceptable carriers include: (1) sugars, such as lactose, glucose and sucrose; (2) starches, such as corn starch and potato starch; (3) cellulose, and its derivatives, such as sodium carboxymethyl cellulose, ethyl cellulose and cellulose acetate; (4) powdered tragacanth; (5) malt; (6) gelatin; (7) talc; (8) excipients, such as cocoa butter and suppository waxes; (9) oils, such as peanut oil, cottonseed oil, safflower oil, sesame oil, olive oil, corn oil and soybean oil; (10) glycols, such as propylene glycol; (11) polyols, such as

glycerin, sorbitol, mannitol and polyethylene glycol; (12) esters, such as ethyl oleate and ethyl laurate; (13) agar; (14) buffering agents, such as magnesium hydroxide and aluminum hydroxide; (15) alginic acid; (16) pyrogen-free water; (17) isotonic saline; (18) Ringer's solution; (19) ethyl alcohol; (20) phosphate buffer solutions; and (21) other non-toxic compatible substances employed in pharmaceutical formulations.

[0048] As used herein, the term "administering" includes all means of introducing the compounds and compositions described herein to the patient, including, but are not limited to, oral (po), intravenous (iv), intramuscular (im), subcutaneous (sc), transdermal, inhalation, buccal, ocular, sublingual, vaginal, rectal, and the like. The compounds and compositions described herein may be administered in unit dosage forms and/or formulations containing conventional nontoxic pharmaceutically acceptable carriers, adjuvants, and vehicles.

[0049] Illustrative formats for oral administration include tablets, capsules, elixirs, syrups, and the like. Illustrative routes for parenteral administration include intravenous, intraarterial, intraperitoneal, epidural, intraurethral, intrasternal, intramuscular and subcutaneous, as well as any other art recognized route of parenteral administration.

[0050] Illustrative means of parenteral administration include needle (including microneedle) injectors, needle-free injectors and infusion techniques, as well as any other means of parenteral administration recognized in the art. Parenteral formulations are typically aqueous solutions which may contain excipients such as salts, carbohydrates and buffering agents (preferably at a pH in the range from about 3 to about 9), but, for some applications, they may be more suitably formulated as a sterile non-aqueous solution or as a dried form to be used in conjunction with a suitable vehicle such as sterile, pyrogen-free water. The preparation of parenteral formulations under sterile conditions, for example, by lyophilization, may readily be accomplished using standard pharmaceutical techniques well known to those skilled in the art. Parenteral administration of a compound is illustratively performed in the form of saline solutions or with the compound incorporated into liposomes. In cases where the compound in itself is not sufficiently soluble to be dissolved, a solubilizer such as ethanol can be applied.

[0051] The dosage of each compound of the claimed combinations depends on several factors, including: the administration method, the condition to be treated, the severity of the condition, whether the condition is to be treated or prevented, and the age, weight, and health of the person to be treated. Additionally, pharmacogenomic (the effect of genotype on the pharmacokinetic, pharmacodynamic or efficacy profile of a therapeutic) information about a particular patient may affect the dosage used.

[0052] It is to be understood that in the methods described herein, the individual components of a co-administration, or combination can be administered by any suitable means, contemporaneously, simultaneously, sequentially, separately or in a single pharmaceutical formulation. Where the co-administered compounds or compositions are administered in separate dosage forms, the number of dosages administered per day for each compound may be the same or different. The compounds or compositions may be administered via the same or different routes of administration. The compounds or compositions may be administered

according to simultaneous or alternating regimens, at the same or different times during the course of the therapy, concurrently in divided or single forms.

[0053] The term "therapeutically effective amount" as used herein, refers to that amount of active compound or pharmaceutical agent that elicits the biological or medicinal response in a tissue system, animal or human that is being sought by a researcher, veterinarian, medical doctor or other clinician, which includes alleviation of the symptoms of the disease or disorder being treated. In one aspect, the therapeutically effective amount is that which may treat or alleviate the disease or symptoms of the disease at a reasonable benefit/risk ratio applicable to any medical treatment. However, it is to be understood that the total daily usage of the compounds and compositions described herein may be decided by the attending physician within the scope of sound medical judgment. The specific therapeutically-effective dose level for any particular patient will depend upon a variety of factors, including the disorder being treated and the severity of the disorder; activity of the specific compound employed; the specific composition employed; the age, body weight, general health, gender and diet of the patient; the time of administration, route of administration, and rate of excretion of the specific compound employed; the duration of the treatment; drugs used in combination or coincidentally with the specific compound employed; and like factors well known to the researcher, veterinarian, medical doctor or other clinician of ordinary skill.

[0054] Depending upon the route of administration, a wide range of permissible dosages are contemplated herein, including doses falling in the range from about 1 g/kg to about 1 g/kg. The dosages may be single or divided, and may administered according to a wide variety of protocols, including q.d. (once a day), b.i.d. (twice a day), t.i.d. (three times a day), or even every other day, once a week, once a month, once a quarter, and the like. In each of these cases it is understood that the therapeutically effective amounts described herein correspond to the instance of administration, or alternatively to the total daily, weekly, month, or quarterly dose, as determined by the dosing protocol.

[0055] In addition to the illustrative dosages and dosing protocols described herein, it is to be understood that an effective amount of any one or a mixture of the compounds described herein can be determined by the attending diagnostician or physician by the use of known techniques and/or by observing results obtained under analogous circumstances. In determining the effective amount or dose, a number of factors are considered by the attending diagnostician or physician, including, but not limited to the species of mammal, including human, its size, age, and general health, the specific disease or disorder involved, the degree of or involvement or the severity of the disease or disorder, the response of the individual patient, the particular compound administered, the mode of administration, the bio-availability characteristics of the preparation administered, the dose regimen selected, the use of concomitant medication, and other relevant circumstances.

[0056] The term "patient" includes human and non-human animals such as companion animals (dogs and cats and the like) and livestock animals. Livestock animals are animals raised for food production. The patient to be treated is preferably a mammal, in particular a human being.

[0057] It is understood that, the herbicides disclosed herein can be applied to a field of a plant for weed control at the same time as a pre-formulated mixture, or applied individually as a separately pre-formulated product, consequentially or concurrently.

[0058] It is understood that, multiple application of said composition of herbicides may be needed in some cases in order achieve effective and efficient weed control for a field of a plant. As disclosed herein said plant is resistant to the herbicides applied.

[0059] In preparing a product for an end user, adjuvants, surfactants, anti-drifting agents, colorings, anti-freezing or other stabilizing chemicals may be included. An adjuvant is an additive (usually in relatively low amounts compared to the carrier) that improves or enhances application, performance, safety, storage, or handling of an active ingredient. Adjuvants include materials such as: Surfactants (spreaders, stickers, emulsifiers, wetting agents), which increase surface contact, reduce runoff, and increase penetration through leaf cuticle.

[0060] Pure herbicide molecules have to be properly formulated before it is delivered to the end user for various purposes of weed control and rarely used as the pure chemical. In addition, a given chemical may be formulated in a variety of differing formulations and sold under different trade names. The primary reason for formulating a herbicide is to allow the user to dispense it in a convenient carrier, such as water. The primary purpose of the carrier is to enable the uniform distribution of a relatively small amount of herbicide over a comparatively large area. In addition to providing the consumer with a form of herbicide that is easy to handle, formulating a herbicide can enhance the phytotoxicity of the herbicide, improve the shelf-life (storage) of the herbicide, and protect the herbicide from adverse environmental conditions while in storage or transit. Formulations vary according to the solubility of the herbicide active ingredient in water, oil and organic solvents, and the manner the formulation is applied (i.e., dispersed in a carrier such as water or applied as a dry formulation itself) (P. Miller and P. Westra, Colorado State University, www.colostate.edu).

[0061] Solution (S)—Solution formulations are designed for those active ingredients that dissolve readily in water. The formulation is a liquid and consists of the active ingredient and additives. When herbicides formulated as solutions are mixed with water, the active ingredient will not settle out of solution or separate.

[0062] Soluble Powder (SP)—Soluble powder formulations are similar to Solutions(S) in that, when mixed with water, these dry formulations dissolve readily and form a true solution. The formulation is dry and consists of the active ingredient and additives. When thoroughly mixed, no further agitation is necessary to keep the active ingredient dissolved in solution. Few formulations of this type are available because few active ingredients are highly soluble in water.

[0063] Emulsifiable Concentrate (E or EC)—Formulations of this type are liquids that contain the active ingredient, one or more solvents, and an emulsifier that allows mixing with water. Formulations of this type are highly concentrated and relatively inexpensive per pound of active ingredient; easy to handle, transport, and store; require little agitation (will not settle out or separate); and are not abrasive to machinery or spraying equipment. Formulations of this type may, however, have potentially greater phyto-

toxicity than other formulations; exhibit a potential for over-or under-dosing through mixing or calibration errors; are more easily absorbed through skin of humans or animals; and contain solvents that may cause deterioration of rubber or plastic hoses and pump parts.

[0064] **Wettable Powder (W or WP)**—Wettable powders are dry, finely ground formulations in which the active ingredient is combined with a finely ground carrier (usually mineral clay) along with other ingredients, to enhance the ability of the active ingredient plus carrier to suspend in water. The powder is mixed with water for application. Wettable powders are one of the most widely used herbicide formulations and offer low cost and ease of storage, transport, and handling; lower phytotoxicity potential than ECs and other liquid formulations; and less skin and eye absorption hazard than ECs and other liquid formulations. Some disadvantages are that they require constant and thorough agitation in the spray tank, are abrasive to pumps and nozzles (causing premature wear), may produce visible residues on plant and soil surfaces, and can create an inhalation hazard to the applicator while handling (pouring and mixing) the concentrated powder.

[0065] **Liquid Flowable (F or FL)**—Liquid flowable formulations consist of finely ground active ingredient suspended in a liquid. Flowables are mixed with water for application, are easily handled and applied, and seldom clog nozzles. Some of their disadvantages are that they may leave a visible residue on plant and soil surfaces, and typically require constant and thorough agitation to remain in suspension.

[0066] **Dry Flowables and Water-Dispersible Granules (DF, DG or WDG)**—Dry flowable and water-dispersible granule formulations are much like wettable powders except that the active ingredient is formulated on a large particle (granule) instead of onto a ground powder. This type of formulation offers essentially the same advantages and disadvantages as wettable powder formulations. However, these formulations generally are more easily mixed and measured than wettable powders. Because they create less dust when handling, they cause less inhalation hazard to the applicator during pouring and mixing.

[0067] **Granules and Pellets (G, P or PS)**—Used exclusively for soil applied herbicides, the active ingredient is formulated onto large particles (granules or pellets). The primary advantages of this type of formulation are that the formulation is ready to use with simple application equipment (seeders or spreaders), and the drift potential is low because the particles are large and settle quickly. The disadvantages of these formulations are that they do not adhere to foliage (not intended for foliar applications), and may require mixing into the soil in order to achieve adequate herbicidal activity.

[0068] As it is disclosed herein, a cellulosin refers to a class of compounds that acts as an inhibitor toward cellulose synthase (CesA), an enzyme that catalyzes the synthesis of cellulose. Examples of a cellulosin include Endosidin20 (ES20) (FIG. 1A) and analogues 1, 2 and 3 shown in FIG. 7A.

[0069] In some embodiments, the composition of a herbicide may be formulated in various dosage forms, including, but not limited to, dry formulation, liquid formulation, granular or pellet formulation. The practice and information are known in the arts. In some other embodiments, the final product of the composition disclosed herein may be formu-

lated as a suspension, a liquid spray, a powder, a nanoparticle, or an aerosol, together with one or more adjuvants, excipients or carriers.

[0070] In another aspect, the present invention also may include those polypeptides which exhibit at least 80%, more preferably at least 90% identity, and most preferably at least 95% identity to a polypeptide sequence selected from the group of sequences set forth above.

[0071] In some illustrative embodiments, the present invention relates to a mutant gene having the sequence of SEQ ID NOS: 3, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 49, 51, and 55, or a gene of 80% or more sequence identity having said mutated gene, respectively.

[0072] In some illustrative embodiments, the present invention relates to a cell or a plant having the mutant gene disclosed herein.

[0073] In some illustrative embodiments, the present invention relates to a protein having the sequence of SEQ ID NOS: 4, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 50, 52, and 56, or a polypeptide of 80% or more sequence identity having said mutated conserved amino acid residue, respectively.

[0074] In some illustrative embodiments, the present invention relates to a cell or a plant expressing the protein or polypeptide disclosed herein.

[0075] In some illustrative embodiments, the present invention relates to an isolated and cloned mutant gene encoding a cellulose synthase comprising a cellulose synthase gene with a specified nucleic acid sequence, wherein the nucleic acid sequence comprises the nucleotide sequence of SEQ ID NOS: 3, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 49, 51, and 55, or a gene of 80% or more sequence identity having said mutated gene, respectively. In some other embodiments, the present invention relates to a cell or a plant having at least a copy of those genes.

[0076] In some illustrative embodiments, the present invention relates to a cell or a plant having a mutant gene, wherein said mutant gene comprises the sequence of SEQ ID NOS: 3, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 49, 51, and 55, or a gene of 90% or more sequence identity having said mutated gene, respectively.

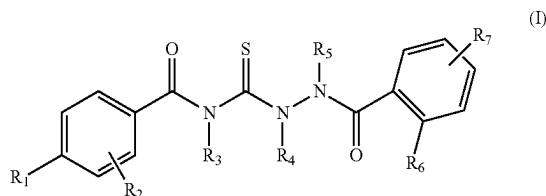
[0077] In some illustrative embodiments, the present invention relates to a cell or a plant expressing a protein, wherein said protein comprises the sequence of SEQ ID NOS: 4, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 50, 52, and 56, or a polypeptide of 90% or more sequence identity having said mutated conserved amino acid residue, respectively.

[0078] In some illustrative embodiments, the present invention relates to a method for weed control of a crop comprising cloning and expressing a mutant gene having the sequence of SEQ ID NOS: 3, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 49, 51, and 55, or a gene of 90% or more sequence identity having said mutated gene, respectively, in said crop, and applying an inhibitor of cellulose synthase to the field of said crop.

[0079] In some illustrative embodiments, the present invention relates to a method for weed control of a crop comprising cloning and expressing a protein having the sequence of SEQ ID NOS: 4, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 50, 52, and 56,

or a polypeptide of 90% or more sequence identity having said mutated conserved amino acid residue, respectively, in said crop, and applying an inhibitor of cellulose synthase to the field of said crop.

[0080] In some illustrative embodiments, the present invention relates to a compound of the following general formula (I)



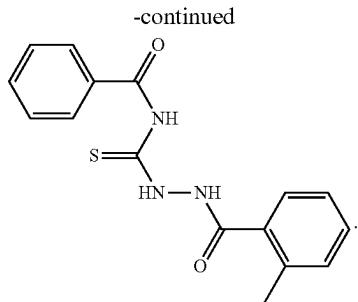
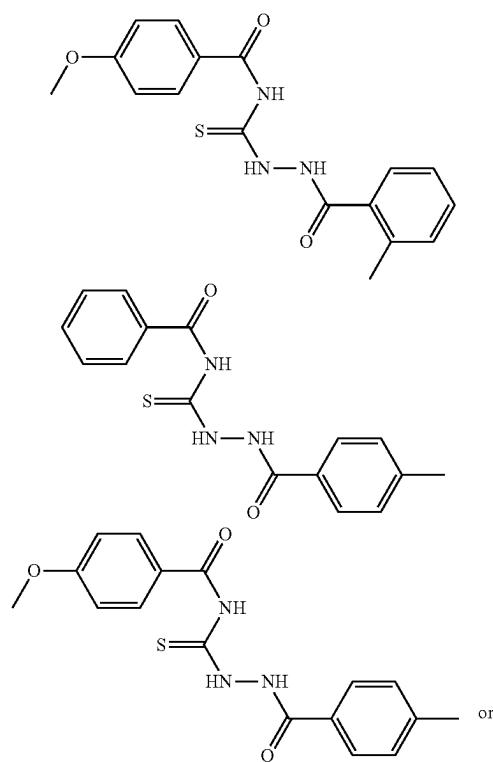
or an acceptable salt thereof, wherein

[0081] R₁ is hydrogen, a halo, an alkyl, alkenyl, or alkyloxy;

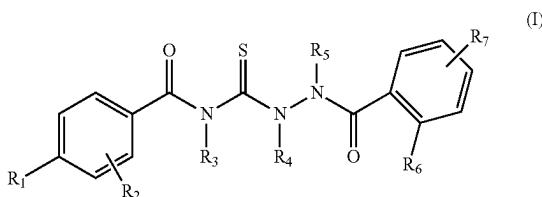
[0082] R₂ and R₇ represent four individual substituents independently selected from the group comprising hydrogen, a halo, an alkyl, alkenyl, and alkyloxy; and

[0083] R₃, R₄, R₅ and R₆ are independently hydrogen or an alkyl.

[0084] In some illustrative embodiments, the present invention relates to a compound of the following general formula (I), wherein the compound is



[0085] In some illustrative embodiments, the present invention relates to a cellulose synthase inhibitor having the following general formula (I)



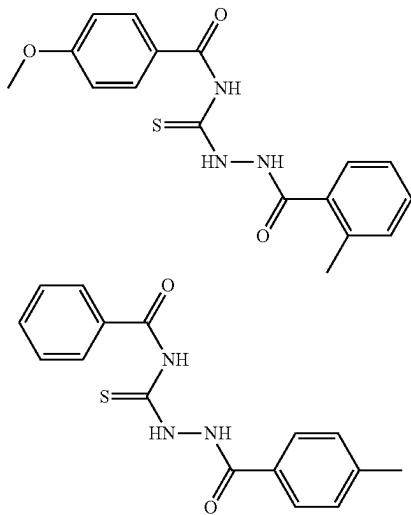
or an acceptable salt thereof, wherein

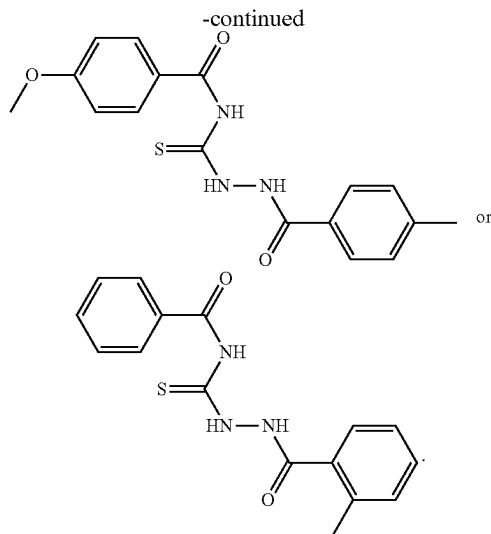
[0086] R₁ is hydrogen, a halo, an alkyl, alkenyl, or alkyloxy;

[0087] R₂ and R₇ represent four individual substituents independently selected from the group comprising hydrogen, a halo, an alkyl, alkenyl, and alkyloxy; and

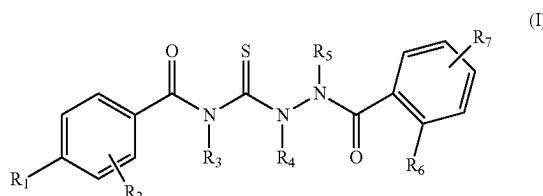
[0088] R₃, R₄, R₅ and R₆ are independently hydrogen or an alkyl.

[0089] In some illustrative embodiments, the present invention relates to a cellulose synthase inhibitor of the following general formula (I), wherein the inhibitor is





[0090] In some illustrative embodiments, the present invention relates to a herbicide having the following general formula (I)



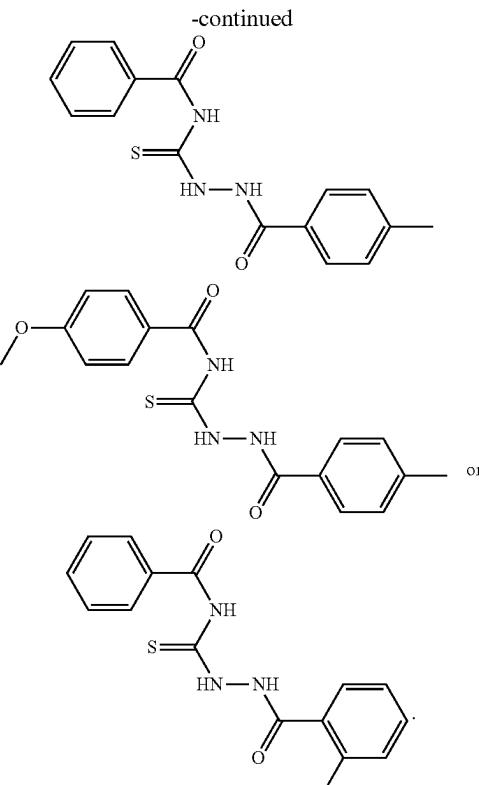
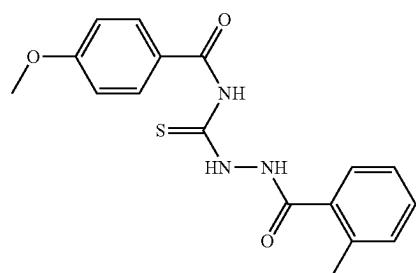
or an acceptable salt thereof, together with one or more diluents, excipients or carriers, wherein

[0091] R₁ is hydrogen, a halo, an alkyl, alkenyl, or alkyloxy;

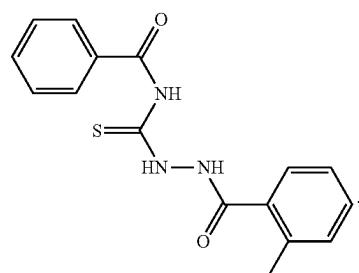
[0092] R₂ and R₇ represent four individual substituents independently selected from the group comprising hydrogen, a halo, an alkyl, alkenyl, and alkyloxy;

[0093] and R₃, R₄, R₅ and R₆ are independently hydrogen or an alkyl.

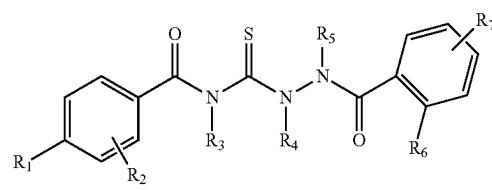
[0094] In some illustrative embodiments, the present invention relates to a herbicide of the following general formula (I), wherein the herbicide is



[0095] In some illustrative embodiments, the present invention relates to a herbicide of the following general formula (I), wherein the herbicide is



[0096] In some illustrative embodiments, the present invention relates to a method for weed control of a plant, comprising the step of applying an effective amount of an herbicide having the following general formula (I)



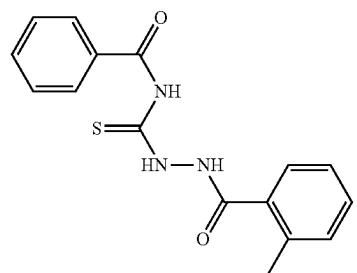
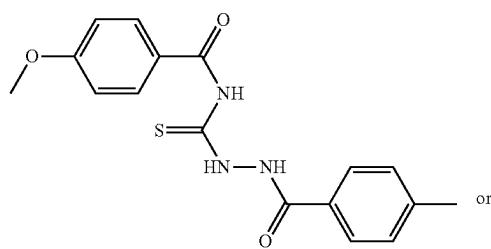
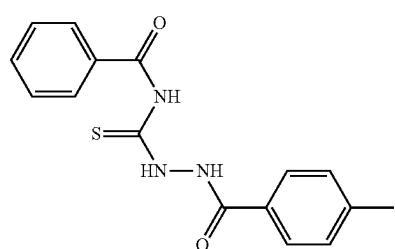
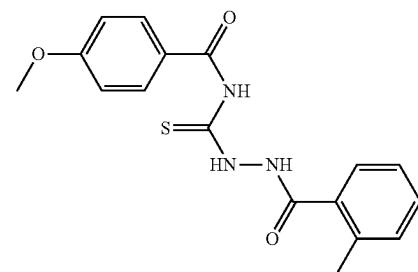
or an acceptable salt thereof, to the field of said plant, wherein

[0097] R_1 is hydrogen, a halo, an alkyl, alkenyl, or alkyloxy;

[0098] R_2 and R_7 represent four individual substituents independently selected from the group comprising hydrogen, a halo, an alkyl, alkenyl, and alkyloxy;

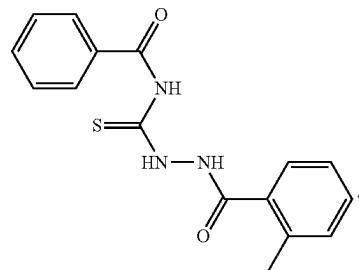
[0099] and R_3 , R_4 , R_5 and R_6 are independently hydrogen or an alkyl.

[0100] In some illustrative embodiments, the present invention relates to a method for weed control of a plant, comprising the step of applying an effective amount of an herbicide having the following general formula (I), wherein said herbicide is

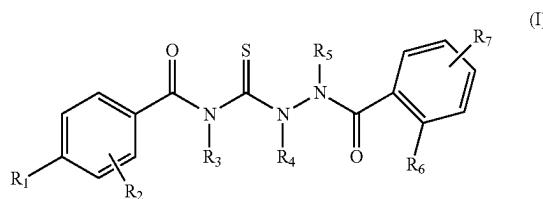


[0101] In some illustrative embodiments, the present invention relates to a method for weed control of a plant, comprising the step of applying an effective amount of an

herbicide having the following general formula (I), wherein said herbicide is



[0102] In some illustrative embodiments, the present invention relates to a method for treating a bacterial infection of a patient comprising administrating a therapeutically effective amount of a compound of formula (I), together with one or more diluents, excipients or carriers,



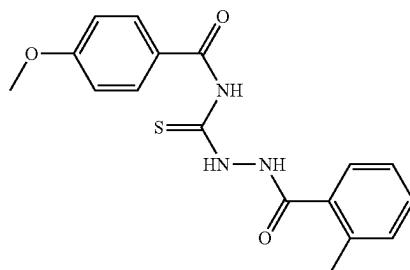
or a pharmaceutically acceptable salt thereof, to the patient in need of relief from said infection, wherein

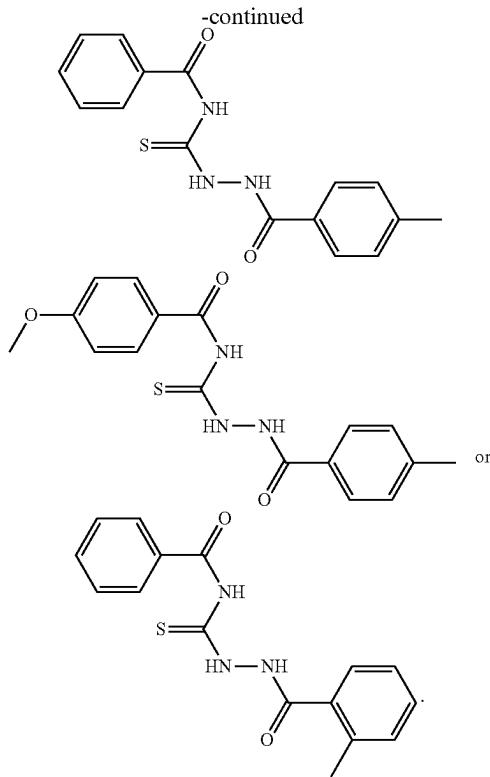
[0103] R_1 is hydrogen, a halo, an alkyl, alkenyl, or alkyloxy;

[0104] R_2 and R_7 represent four individual substituents independently selected from the group comprising hydrogen, a halo, an alkyl, alkenyl, and alkyloxy; and

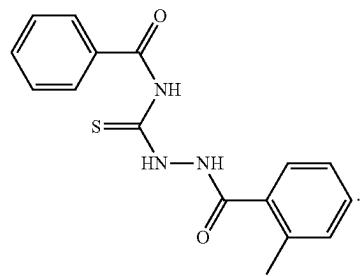
[0105] R_3 , R_4 , R_5 and R_6 are independently hydrogen or an alkyl.

[0106] In some illustrative embodiments, the present invention relates to a method for treating a bacterial infection of a patient comprising administrating a therapeutically effective amount of a compound of formula (I), together with one or more diluents, excipients or carriers, wherein the compound is

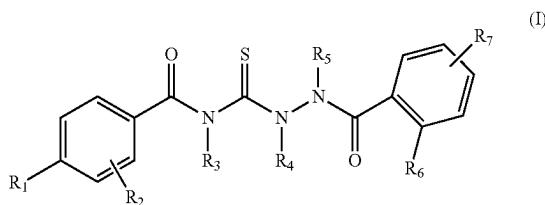




[0107] In some illustrative embodiments, the present invention relates to a method for treating a bacterial infection of a patient comprising administrating a therapeutically effective amount of a compound of formula (I), together with one or more diluents, excipients or carriers, wherein said compound is



[0108] In some illustrative embodiments, the present invention relates to a pharmaceutical composition for treating a bacterial infection comprising a compound of formula (I), together with one or more diluents, excipients or carriers,



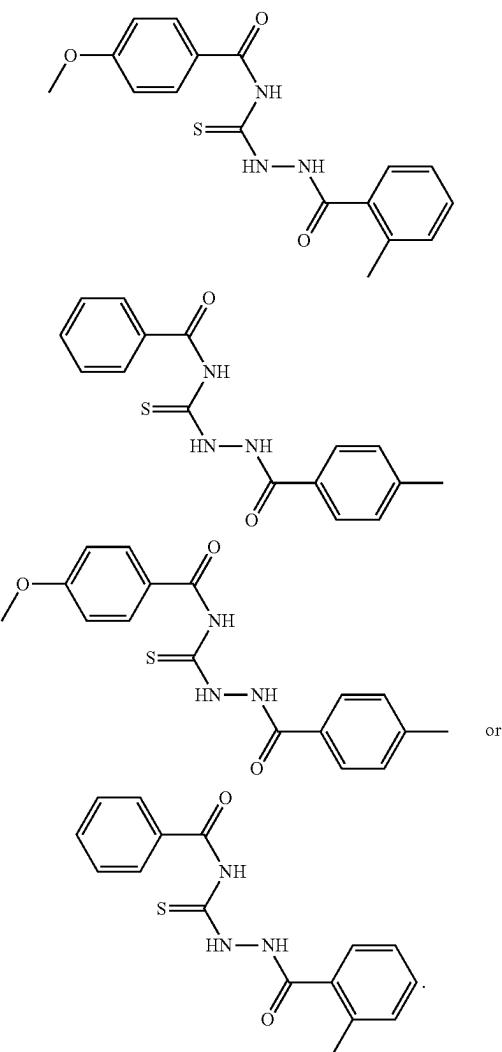
or a pharmaceutically acceptable salt thereof, wherein

[0109] R₁ is hydrogen, a halo, an alkyl, alkenyl, or alkyloxy;

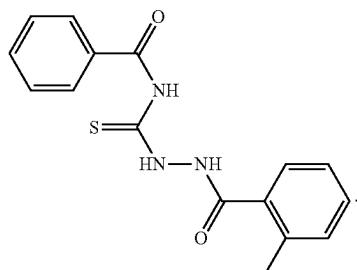
[0110] R₂ and R₇ represent four individual substituents independently selected from the group comprising hydrogen, a halo, an alkyl, alkenyl, and alkyloxy; and

[0111] R₃, R₄, R₅ and R₆ are independently hydrogen or an alkyl.

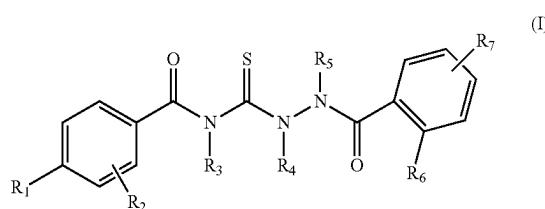
[0112] In some illustrative embodiments, the present invention relates to a pharmaceutical composition for treating a bacterial infection comprising a compound of formula (I), together with one or more diluents, excipients or carriers, wherein the compound is



[0113] In some illustrative embodiments, the present invention relates to a pharmaceutical composition for treating a bacterial infection comprising a compound of formula (I), together with one or more diluents, excipients or carriers, wherein the compound is



[0114] In some illustrative embodiments, the present invention relates to a medicament for treating a bacterial infection comprising a compound of formula (I), together with one or more diluents, excipients or carriers,



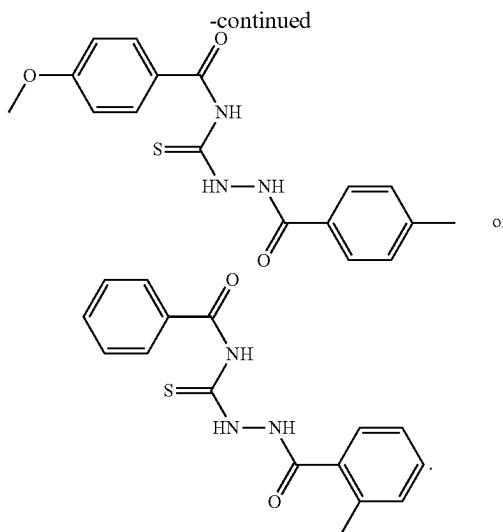
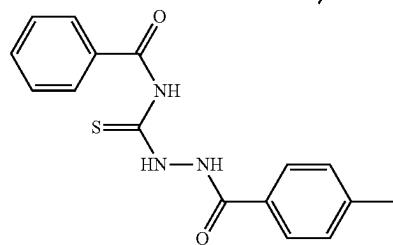
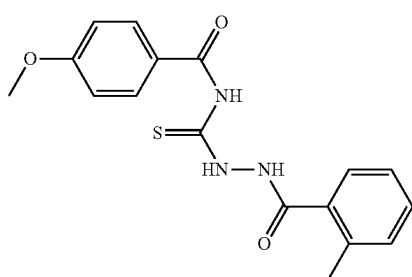
or a pharmaceutically acceptable salt thereof, wherein

[0115] R₁ is hydrogen, a halo, an alkyl, alkenyl, or alkyloxy;

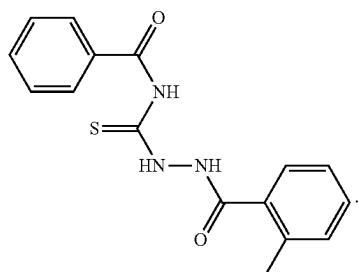
[0116] R₂ and R₇ represent four individual substituents independently selected from the group comprising hydrogen, a halo, an alkyl, alkenyl, and alkyloxy; and

[0117] R₃, R₄, R₅ and R₆ are independently hydrogen or an alkyl.

[0118] In some illustrative embodiments, the present invention relates to a pharmaceutical composition for treating a bacterial infection comprising a compound of formula (I), together with one or more diluents, excipients or carriers, wherein the compound is



[0119] In some illustrative embodiments, the present invention relates to a pharmaceutical composition for treating a bacterial infection comprising a compound of formula (I), together with one or more diluents, excipients or carriers, wherein the compound is



[0120] The following non-limiting exemplary embodiments are included herein to further illustrate the invention. These exemplary embodiments are not intended and should not be interpreted to limit the scope of the invention in any way. It is also to be understood that numerous variations of these exemplary embodiments are contemplated herein.

[0121] Like other eukaryotic organisms, plant cells contain compartmentalized organelles for controlled biochemical reactions. Membrane trafficking pathways transport molecules, such as proteins and lipids, among organelles and are essential for growth, development and environment adaptation et al. Cellulose is the polymer of $\beta(1,4)$ D-glucose and is an essential cell wall component that controls plant cell survival and growth. In plants, a large portion of solar energy and CO₂ fixed through photosynthesis is used to synthesize cellulose. Although the composition of cellulose is only the linear chain of repetitive $\beta(1,4)$ D-glucose units, its mechanisms of biosynthesis is poorly understood. Studies have shown that membrane trafficking pathways play essential roles in cellulose biosynthesis.

[0122] Plant cellulose is synthesized by the hexagonal rosette structured cellulose synthase complex (CSC) at the plasma membrane (Pear, J. R. et al., *Proc Natl Acad Sci USA* 1996, 93, 12637-12642; Arioli, T., et al., *Science* 1998, 279,

717-720; Giddings, T. H., *J Cell Biol* 1980, 84, 327-339). The exact molecular composition of CSC rosette is not clear. Each CSC is believed to contain at least 18 monomeric cellulose synthases (CesA) of three different isoforms at a 1:1:1 molar ratio (Gonneau, M. et al., *Plant Physiol* 2014, 166, 1709-1712; Newman, R. H. et al, *Plant Physiol* 2013, 163, 1558-1567). Each assembled CSC is about 25 nm in diameter with a molecular weight of at least 2 M Da. By examining cellular localization of a functional fluorescence-tagged CesA in plant cells, people have found that CSCs are not only localized to the plasma membrane, but also in the Golgi, and small CesA compartments. Plasma membrane localized CSCs co-align with the microtubules and move bidirectionally at a speed of about 300 nm/min during primary cell wall synthesis in etiolated hypocotyl epidermal cells of *Arabidopsis* (Paredez, A. R., *Science* 2006, 312, 1491-1495; Desprez, T, et al., *Proc Natl Acad Sci USA* 2007, 104, 15572-15577). It is understood that plasma membrane localized CSCs that move along the microtubules are performing cellulose polymerization. The conserved exocytic and endocytic trafficking pathways are responsible for the delivery and turnover of CSCs at the plasma membrane (Zhu, X. et al., *Proc Natl Acad Sci USA* 2018, 115, E3578-E3587; Bashline, L. et al., *Plant Physiol.* 2013, 163, 150-160; Bashline, L., et al., *Proc. Natl Acad Sci. USA* 2015, 112, 12870-12875), which indicates that membrane trafficking machineries are essential for the process of cellulose biosynthesis. However, the detailed mechanisms of CSC cellular trafficking have not been fully elucidated.

[0123] Plant CesA proteins are believed to have eight transmembrane domains and a large central cytoplasmic domain. The central cytoplasmic domain is located between two transmembrane domains are located at the N-terminal region of CesA and six transmembrane domains at the C-terminal of CesA. The central cytoplasmic domain of plant CesAs contain 492 to 536 amino acids, depending on isoforms and species (Kumar, M., et al., *Phytochemistry*, 2015, 112, 91-99). The central cytoplasmic domain of plant CesAs contains conserved amino acids that are responsible for the glycosyl transferase activities required for cellulose catalytic synthesis, as has been found in bacteria cellulose synthase (BcsA) (Morgan, J. L., et al., *Nat Struct Mol Biol* 2014, 21, 489-496; Morgan, J. L., et al., *Nature* 2016, 531, 329-334). Plant CesAs also contain two plant specific regions, plant conserved region (P-CR) and class specific region (CSR), in their central cytoplasmic domain. Based on structural studies from BcsA and sequence comparison between plant and bacteria cellulose synthases, it is believed that the central cytoplasmic domain of plant CesAs catalyze the synthesis of cellulose. However, it is poorly understood how the cellulose catalytic synthesis process is integrated to the trafficking of CesAs.

[0124] By screening chemical libraries using *Arabidopsis thaliana*, we discovered a small molecule Endosidin20 (ES20) that reduced plant growth, inhibited anisotropic cell expansion and reduced crystalline cellulose synthesis (see Table 1). In an ES20-suppression genetic screen, we discovered a number of mutant alleles of *Arabidopsis* CesA6 that are less sensitive to ES20. Ten of the those mutant alleles of cesA6 contain mis-sense mutations in amino acids located in the central cytoplasmic domain and one is located at the third transmembrane domain that is close to the central cytoplasmic domain. Structure modeling and molecular docking experiments indicate that ES20 could bind to a

pocket in the conserved catalytic domain of CesAs. Eight of the mutated amino acids in ES20-insensitive mutants are located in the catalytic domain where ES20 is believed to bind. We used biochemical assay to show that ES20 physically interacted with the central cytoplasmic domain of CesA, where UDP-glucose, the substrate for cellulose synthesis, also binds. Based on our modeled structure and hypothesized binding sites, we created eight new transgenic lines that carrying missense mutations in amino acids that are close to the hypothesized ES20 and UDP-glucose binding site. Interestingly, five of these eight mutations lead to reduced sensitivity to ES20. We also discovered that upon short-term ES20 treatment, plasma membrane localized YFP-CesA6 is mostly removed. We show that ES20 competes with UDP-glucose, the substrate for cellulose biosynthesis, to inhibit cellulose biosynthesis and the competition reduced CesA exocytic trafficking.

[0125] ES20 is a growth inhibitor that reduces crystalline cellulose content.

[0126] In order to identify small molecules that could be used in studying plant membrane trafficking, chemical libraries have been screened first using tobacco pollen growth assay and then *Arabidopsis* live cell imaging assay (Drakakaki G. et al., *Proc Natl Acad Sci USA* 2011, 108, 17850-17855). In primary screen, small molecules that inhibited pollen germination or pollen tube growth in in vitro culture condition were selected. For secondary screen, selected small molecules were tested in *Arabidopsis* plants expressing different fluorescence-tagged cellular markers. Small molecule ES20 (FIG. 1A) inhibited tobacco pollen tube growth but surprisingly it did not significantly affect the trafficking of two typical cargo proteins, PIN2 and BRI1. Locations of general organelle markers, such as HDEL-GFP (Endoplasmic Reticulum, ER), Gop1p-YFP (Golgi), SYP61-CFP (Trans-Golgi Network, TGN) and VHA1-GFP (TGN), are unaffected as well by ES20 treatment (FIGS. 9A and 9B). These results indicate that ES20 does not inhibit general membrane trafficking in plants. However, ES20 significantly affects cell growth.

[0127] When we grew wildtype *Arabidopsis* seedlings on the growth media supplemented with ES20, we observed shorter and swollen roots in a ES20 dosage dependent manner (FIGS. 1B and 1C). When we grew wildtype *Arabidopsis* seedlings on normal growth media and then treated them with 5 μ M ES20 for 12 hours in liquid growth media, the roots became swollen in the tip region, where growth occur (FIG. 1D). Staining on root epidermal cells with fluorescence dye showed that the individual cells in root elongation zone are swollen as well after 12 hours of 5 μ M ES20 treatment (FIG. 1E). *Arabidopsis* seedlings grown on media with ES20 under dark condition have shorter and swollen hypocotyls in a ES20 dosage dependent manner (FIGS. 1F, 1G and 1H). Individual hypocotyl epidermal cells from plants grown on ES20 media are swollen as well in compare with those grown on regular media (FIG. 1I). These growth assays indicate that ES20 inhibits cell growth and causes cells to grow in an isotropic manner instead of anisotropic manner.

[0128] The composition and organization of cell wall determines plant cell shapes. Swollen cells and organs in plants are often caused by direct or indirect disruption of cell wall biosynthesis or organization. For example, swollen cells and roots have been found in cesa mutants and in plants treated with inhibitors of cellulose synthesis and microtu-

bule organization (Arioli, T., et al., *Science* 1998, 279, 717-720; Burn, J. E., et al., *Plant Physiol.* 2002, 129, 797-807). The retarded growth and cell swollen phenotypes indicate that ES20 could affect cell growth through interfering with cell wall biosynthesis or organization. To test whether ES20 affects cell wall composition, we took the straightforward approach of analyzing crystalline cellulose content in seedlings grown in the presence of different concentrations of ES20 (Updegraff, D. M. *Anal. Biochem.* 1969, 32, 420-424). We found that in dark-grown *Arabidopsis* hypocotyls, the content of crystalline cellulose is reduced in a ES20 dosage dependent manner. The content of crystalline cellulose in seedlings grown on M ES20 was reduced to about 50% of the seedlings grown in the absence of ES20 (FIG. 1J). The reduction of crystalline cellulose in ES20 treated seedlings indicates that ES20 affects cell growth by interfering with cell wall biosynthesis.

[0129] A group of novel cesa6 alleles are less sensitive to ES20 inhibition.

[0130] In order to identify the cell growth pathways that are targeted by ES20, we performed a chemical genetic screen for mutants that have reduced sensitivity to this growth inhibitor. About 400,000 M2 seedlings from an EMS mutagenized population in Col-0 ecotype expressing SYP61-CFP were screened for their sensitivity to 5 μ M ES20, a dosage that completely arrests root growth in wild type plants. Majority of the seedlings in our screen showed chlorosis cotyledons/leaves and arrested root growths. However, we found plants that have normal green cotyledons and leaves and elongated roots. These plants became obviously healthier and stronger than other plants after two weeks of growth on ES20 and were transferred to soil to get the M3 population. After re-testing the seeds from M3, we confirmed a total of 30 individual lines that have reduced sensitivity to ES20 in growth. We generated the mapping population and used high-throughput whole genome sequencing to clone the mutant genes. After analysis on candidate SNPs in these mutants, we found that they belong to 12 mutant alleles carrying either C-T or G-A mutations. 10 out of the 12 mutants were in CesA6 and two in CesA7. After two times of backcrosses, we consistently detected less sensitivity in these mutants to ES20 (FIGS. 2A to 2E). After we further analyzed the location of the cesa6 mutations, we found that 8 of the 10 cesa6 alleles were new mutations that were located in the central cytoplasmic domain and the other two were located at the transmembrane domain at the C-terminal direction of the central cytoplasmic domain (FIG. 2F).

[0131] In order to confirm that the mutations at CesA6 caused the reduced sensitivity to ES20, we performed genetic complementation experiments. We cloned the genomic content of CesA6 and inserted a YFP tag to the 3' of the coding region so that we could generate CesA6 with a N-terminal YFP fusion. We then created one of the mutations that were found in our mutants using site-directed mutagenesis. We then transformed each CesA6 construct to a loss of function cesa6 mutant named prc1-1 (Fagard, M. et al., *Plant Cell* 2000, 12, 2409-2424). We found that the wild type CesA construct could rescue the growth phenotype of prc1-1 and the mutant CesA constructs could fully or partially rescue the growth phenotype of prc1-1, depending on the mutations (FIGS. 10A, 10B, and 10C). The transgenic plants expressing wild type CesA6 construct in prc1-1 background have the same sensitivity to ES20 treatment as

the wild type plants. However, the transgenic plants expressing mutant CesA6 construct in prc1-1 background have reduced sensitivity to ES20 treatment (FIGS. 10A, 10B, and 10C). These complementation experiments further confirmed that the mutations at CesA6 caused the insensitivity to ES20.

[0132] ES20 directly interacts with the central cytoplasmic domain of CesA6, likely to be at the catalytic core.

[0133] Often when a mutation occurs in the ligand binding site on a receptor protein, the ligand molecule may not be able bind to the receptor or may have reduced binding affinity to the receptor, so that the organisms could become insensitive to the ligand molecule. Multiple alleles of cesa6 that are less sensitive to ES20 made us hypothesize that ES20 could interact directly with CesA6 to inhibit its function for cellulose biosynthesis. This hypothesis also fits our observation that ES20 reduces the crystalline cellulose content in plants. In order to test for direct action between ES20 and CesA6, we expressed and purified the central cytoplasmic domain of CesA6 from *E. coli*. The purified central cytoplasmic domain protein contains a SUMO-His tag at the N-terminal region and a GFP tag at the C-terminal region. We then performed microscale thermophoresis (MST) assay to test for direct interaction between ES20 and the fusion protein. The MST binding curve showed that ES20 directly interacted with CesA6 with a dissociation constant (K_d) of $15.4 \pm 5.9 \mu\text{M}$ (FIG. 3A). We also tested the interaction between UDP-glucose and CesA6 central cytoplasmic domain using the same assay. We found that UDP-glucose interacts with CesA6 with a K_d of $3.5 \pm 2.3 \mu\text{M}$ (FIG. 3B).

[0134] In order to find possible binding site for ES20 on CesA6 central cytoplasmic domain, we used computational tools to model the structure of CesA6 central cytoplasmic domain using the solved structure of *Rhodobacter sphaeroides* cellulose synthase (RsBcsA) cytoplasmic domain. We then used molecular docking to explore the possible binding sites for ES20 and UDP-glucose. Surprisingly, we found both ES20 and UDP-glucose were docked to the same pocket in the center region of modeled CesA6 central cytoplasmic domain structure (FIG. 3C). When we examined the position of mutated amino acids in our reduced sensitivity mutants, we discovered that most of these amino acids are either very close to or directly located at the hypothesized binding pocket for ES20 and UDP-glucose (FIG. 3C, green colored). Two of the mutated amino acids D396 and T783, have been found to be located at the catalytic core of RsBcsA and participate in catalysis process (Morgan, J. L., et al., *Nature*, 2016, 531, 329-334). The bioinformatic data matched well with our genetic observation and indicated ES20 could inhibit cellulose biosynthesis by competing with UDP-glucose for the binding site on CesA6.

[0135] In order to further validate our structural modeling and molecular docking data and to further confirm that ES20 binds to the catalytic domain of CesA6, we hypothesized that if we mutated some other amino acids in the binding pocket, we should be able to reduce the binding of ES20 and the plants should have reduced ES20 sensitivity. We selected eight amino acids that are close to the hypothesized ES20 and UDP-glucose binding site and created eight CesA6 genomic constructs that each carries one missense mutation in one of these eight amino acids. We then transformed these constructs to loss of function mutant prc1-1. We obtained

eight new transgenic lines that each express one of the CesA6 genomic constructs containing a missense mutation. Then these transgenic lines were tested on their sensitivity to ES20. Five out of the eight transgenic lines expressing hypothesized mutations had reduced sensitivity to ES20 (FIG. 4). In total, 24 independent mutant alleles that have mutations around the hypothesized binding pocket discovered.

[0136] ES20 inhibits exocytic trafficking of CSCs.

[0137] A large knowledge gap in plant biology is that it is now known how the activities of large protein complexes, such as CSC, are integrated with their cellular transport. Golgi localized CSCs are delivered to the plasma membrane to catalyze cellulose synthesis. CSCs are also localized to some small vesicles named SmaCCs. It is not known what occurs to CSCs when the catalytic core fails to bind a functional substrate. In order to understand better on the integration of CSC catalytic activity and CSC trafficking, we took a straight approach and observed the localization of CesA6 after short-term ES20 treatment. We found that after ES20 treatment, the density of plasma membrane localized CesA6 in root epidermal cells was significantly reduced (FIGS. 5A and 5B). The numbers of SmaCCs that contain CesA6 were significantly increased (FIGS. 5C and 5D). Reduced CSC density at the plasma membrane and increased SmaCCs density could be the result of increased endocytosis or decreased exocytic delivery of CSC. We then performed fluorescence recovery after photobleaching experiment to test whether the exocytic delivery of CSC was affected by ES20 treatment. We bleached YFP-CesA6 fluorescence at the plasma membrane and then observed the delivery of new CSC to the plasma membrane (FIG. 5D). We discovered that in cells treated with ES20, CSC delivery was significantly lower than the DMSO control (FIG. 5E). These results indicate that ES20 treatment affects the exocytic delivery of CSCs to the plasma membrane.

[0138] ES20 targets multiple CesA in plants.

[0139] Due to the conservation of plant CesAs in their central cytoplasmic domains, we hypothesized that ES20 targeted multiple CesAs to inhibit growth. In order to test this hypothesis, we tested the response of fra5 and any1 to ES20. The mutant fra5 carries a mutation of P557T in *Arabidopsis thaliana* CesA7. CesA7 P557 aligns to CesA6 P595. Since the mutation of CesA6 P595S caused reduced sensitivity to ES20 in plants, we test the response of fra5 to ES20. Like CesA6 P595S mutation, fra5 plants have reduced sensitivity to ES20 too (FIGS. 6A, 6C, 6D, 6E and 6F). The mutant any1 carries a mutation of D604N in *Arabidopsis* CesA1. D604 in CesA1 aligns to D605 in CesA6. Since CesA6 D605N caused reduced sensitivity to ES20, we tested the response of any1 to ES20. We discovered that like CesA6 D605N, the mutation of CesA1 D604N made plants to be less sensitive to ES20 (FIGS. 6B, 6C, 6D, 6E and 6F). Reduced sensitivity to ES20 in mutants of CesA7 and CesA1 indicates that ES20 targets multiple CesAs in plants.

[0140] Analog3 of ES20 is more potent than ES20 in inhibiting growth.

[0141] In order to identify key structure in ES20 that causes its inhibition activity, we collected three close analogs of ES20 (FIG. 7A) and tested for their effect in inhibiting plant growth. We found that ES20 analog 1 and analog 2 did not inhibit *Arabidopsis* growth at 1 μM con-

centration. However, ES20 analog 3 is more potent than ES20 in inhibiting *Arabidopsis* growth (FIGS. 7B, and 7C).

[0142] ES20 is active in crops and weeds.

[0143] Because cellulose synthase is highly conserved in plants and ES20 targets the conserved catalytic domain, we hypothesize that ES20 is active in different plants, including common weeds. We tested this hypothesis by growing tomato and horseweed on growth media containing ES20. We found that ES20 could significantly inhibit the growth of both tomato and horseweed (FIG. 8). This indicates that ES20, and probably its more potent analogs, can serve as herbicide for weed control.

[0144] Use of cellulose synthase inhibitor ES20 or an analogue thereof for treatment of bacterial infections, especially for those bacteria that produce and form a biofilm at the site of infection.

[0145] Cellulose is also a component of the extracellular matrix of biofilm that is composed of groups of microorganism cells, including bacteria (Ziemba, C., et al., *NPJ Biofilms Microbiomes* 2016, 2, 1; Limoli, D. H., et al, *Microbiol Spectr*. 2015, 3; Mann, E. E., et al., *FEMS Microbiol Rev*. 2012, 36, 893-916). Biofilms cause human and animal tooth decay and many other types of infectious diseases such as urinary tract infection, cystic fibrosis and middle-ear infections (Bjarnsholt, T. *APMIS* 2013, Suppl., 1-51). The extracellular matrix of biofilms protects them from biotic and abiotic stresses and made biofilms difficult to manipulate. The application of general antibiotics often fails to cure biofilm infections and there is great need to develop alternative methods to control biofilms (Rybtko, M., et al., *J. Mol. Biol.* 2015, 427, 3628-3645). Inhibition of cellulose synthesis has the potential to inhibit the formation of biofilms extracellular matrix and makes it possible to overcome biofilm antibiotic resistance and control corresponding diseases caused by biofilms more efficiently.

[0146] Strategies for producing crops that are resistant to ES20.

[0147] We have discovered that single amino acid change in model plant *Arabidopsis thaliana* is sufficient to create plants that are resistant to ES20. A total of 24 amino acid mutations that lead to resistance to ES20 are disclosed in this present application. Those mutants are found in AtCesA4, AtCesA6, AtCesA7 and AtCesA8. These amino acids are highly conserved among different plant species. The same mutation in these conserved amino acids of any crops will have resistance to a cellulose synthase inhibitor, such as ES20. Creating a cell or a crop plant by incorporating those mutated genes and conveying cellulose synthase inhibitor resistance to the cell or a crop plant are known in the art. Furthermore, one can create a crop plant that is resistant to a cellulose synthase inhibitor, including ES20 and its analogues, using three different ways. (1). Clone the genomic construct of CesA6 homolog in a crop and create a single nucleotide mutation in the genomic construct. Then transform the construct that carries the single nucleotide mutation to the crop so that the construct expresses a CesA6 homolog protein carrying a mutated amino acid. The expression of CesA6 homologs with the mutated amino acid will allow the crop plant to be resistant to a cellulose synthase inhibitor. (2). Create a mutation in CesA6 homolog construct in a crop and express the construct under an inducible promoter. The mutated CesA6 homolog will only be expressed when a cellulose synthase inhibitor, like ES20, is applied. (3). Use CRISPR-cas9 to edit a crop genome to create a mutation at

the amino acid that we have disclosed herein. For specific examples, see D. K. Sohoo, et al., *Plant Biotechnology Journal* (2013) 11, pp. 362-372.; and J. L. Acebes, et al., Cellulose Biosynthesis Inhibitors. In: *Cellulose: Structure and Properties* . . . Editors: A. Lejeune, T. Deprez, pp. 39-73. Nova Science Publishers, Inc. 2010. ISBN: 978-1-60876-388-7. General Methods: NMR spectra were recorded on Bruker spectrometers (^1H at 500 MHz; ^{13}C at 125 MHz.

TABLE 1

Wild Type and Mutant Genes and Protein Products Listing						
Gene SEQ ID NO:	Mutant gene name	Gene	Location on genomic DNA	Location on coding sequence	Codon change(s)	Amino acid change(s) Protein SEQ ID NO:
1	wild type	CESA4				2
3	cesa4 ^{C53F}	CESA4	267G-T	158G-T	TGC-TTC	C53→F 4
5	wild type	CESA6				6
7	cesa6 ^{D396N}	CESA6	1916G-A	1186G-A	GAT-AAT	D396→N 8
9	cesa6 ^{D564N}	CESA6	2804G-A	1690G-A	GAT-AAT	D564→N 10
11	cesa6 ^{D562N}	CESA6	2798G-A	1684G-A	GAT-AAT	D562→N 12
13	cesa6 ^{D602N}	CESA6	2918G-A	1804G-A	GAT-AAT	D602→N 14
15	cesa6 ^{D605N}	CESA6	2927G-A	1813G-A	GAT-AAT	D605→N 16
17	cesa6 ^{D785N}	CESA6	3717G-A	2353G-A	GAT-AAT	D785→N 18
19	cesa6 ^{E929K}	CESA6	4309G-A	2785G-A	GAA-AAA	E929→K 20
21	cesa6 ^{G632D}	CESA6	3092G-A	1895G-A	GGT-GAT	G632→D 22
23	cesa6 ^{G780S}	CESA6	3702G-A	2338G-A	GGT-AGT	G780→S 24
25	cesa6 ^{G811E}	CESA6	3796G-A	2432G-A	GGA-GAA	G811→E 26
27	cesa6 ^{G935E}	CESA6	4328G-A	2804G-A	GGA-GAA	G935→E 28
29	cesa6 ^{L829F}	CESA6	3849C-T	2485C-T	CTT-TTT	L829→F 30
31	cesa6 ^{L286F}	CESA6	1503C-T	856C-T	CTT-TTT	L286→F 32
33	cesa6 ^{P595S}	CESA6	2897C-T	1783C-T	CCT-TCT	P595→S 34
35	cesa6 ^{Q823E}	CESA6	3831C-G	2467C-G	CAA-GAA	Q823→E 36
37	cesa6 ^{S360N}	CESA6	1809G-A	1079G-A	AGT-AAT	S360→N 38
39	cesa6 ^{S394F}	CESA6	2400C-T	1181C-T	TCT-TTT	S394→F 40
41	cesa6 ^{R826A}	CESA6	3840/3841	2476/2477	CGA-GCA	R826→A 42
			CG-GC	CG-GC		
43	cesa6 ^{W827A}	CESA6	3843/3844	2479/2480	TGG-GCG	W827→A 44
			TG-GC	TG-GC		
45	cesa6 ^{T783I}	CESA6	3712C-T	2348C-T	ACC-ATC	T783→I 46
47	wild type	CESA7				48
49	cesa7 ^{A243V}	CESA7	1607C-T	728C-T	GCT-GTT	A243→V 50
51	cesa7 ^{D100N}	CESA7	829G-A	298G-A	GAC-AAC	D100→N 52
53	wild type	CESA8				54
55	cesa8 ^{P174S}	CESA8	943C-T	520C-T	CCT-TCT	P174→S 56

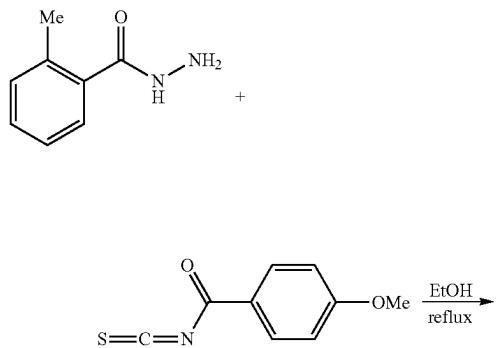
[0148] The mutations we found are located in *Arabidopsis thaliana* CesA6 (AT5G64740), CesA4 (AT5G44030), CesA7 (AT5G17420) and CesA8 (AT4G18780) genes. Single nucleotide change in these mutants causes single amino acid change in CesA proteins and lead to reduced sensitivity to cellulosin (endosidin20 or ES20). The mutations we have discovered are listed in Table 1 and FIG. 11A showing the conserved amino acid residues of those subtypes of cellulose synthases. Representative resistant phenotypes are shown in FIG. 11B. We also found that when we express CesA6 constructs that carry the mutations in wild-type *Arabidopsis*, the plants have reduced sensitivity to Cellulosin (FIG. 11C). This indicates that we can create Cellulosin resistant plants in two ways. One is to create mutations in CesA6 homology gene in crop. The second is to express the mutant CesA6 in crops without knocking out endogenous CesA genes.

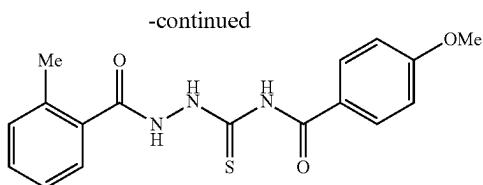
Synthesis of Compound ES20

[0149] Chemical shifts (δ) were given in ppm with reference to solvent signals [^1H NMR: CDCl_3 (7.26); ^{13}C NMR: CDCl_3 (77.2)]. Column chromatography was performed on

silica gel. All reactions were conducted under argon atmosphere and all solvents and reagents were used as obtained from commercial sources without further purification (Scheme 1).

Scheme 1. Chemical Synthesis of ES20





[0150] o-Toluidine hydrazide (1.50 g, 10.0 mmol) was added to a stirred solution of ethanol (40 mL) followed by 4-methoxybenzoyl isothiocyanate (1.93 g, 10.0 mmol) at room temperature. The solution was heated up to reflux under argon for 15 mins. Ethanol was removed under vacuum to give the crude product as a yellow solid. The crude product was recrystallized in ethanol to give 2.00 g purified product as a white crystal in 59% yield.

[0151] HRMS (ESI) [M+H⁺] calculated for C₁₇H₁₇N₃O₃S: 344.1063, found: 344.1064; FTIR (neat, cm⁻¹) ν_{max} 3216, 1667, 1603, 1525, 1498, 1428, 1258, 1174, 1028, 842, 758; ¹H NMR (500 MHz, CDCl₃) δ : 9.48 (s, 1H),

8.95 (s, 1H), 7.87 (d, J =8.8 Hz, 2H), 7.60 (d, J =7.6 Hz, 1H), 7.42 (t, J =7.5 Hz, 1H), 7.30 (d, J =7.8 Hz, 2H), 7.01 (d, J =8.9 Hz, 2H), 3.90 (s, 3H), 2.56 (s, 3H); ^{13}C NMR (126 MHz, CDCl_3) δ : 171.6, 166.0, 164.4, 164.1, 137.9, 131.6, 131.4, 129.8, 127.6, 126.0, 123.0, 114.5, 55.7, 20.2.

[0152] Those skilled in the art will recognize that numerous modifications can be made to the specific implementations described above. The implementations should not be limited to the particular limitations described. Other implementations may be possible.

[0153] While the inventions have been illustrated and described in detail in the drawings and foregoing description, the same is to be considered as illustrative and not restrictive in character, it being understood that only certain embodiments have been shown and described and that all changes and modifications that come within the spirit of the invention are desired to be protected. It should be understood by those skilled in the art that various alternatives to the embodiments described herein may be employed in practicing the claims without departing from the spirit and scope as defined in the following claims.

SEQUENCE LISTING

-continued

tatcatcaaca	caattgttta	tcctttcaact	tcttattccctc	tacttgcccta	ttgtactatt	2520
ccggccgtt	gtcttcac	cggcaaggtc	atccatccctc	cgcataacaa	tttgcgaagc	2580
atatggttt	tagctctgtt	cctatcaatc	attgcaacgg	ctatcttaga	actccgggtgg	2640
atgggtgtaa	gcataaata	tcgtccggct	aacgagcaat	tctgggtgat	ttggccgtgtc	2700
tcagctatc	tcttcgtgt	cttcaaggt	cttcttaaaag	ttctctttgg	tgttgacact	2760
aacttaccg	tcacgtcaaa	aggagccagt	gacgaacccg	atgagtttg	agatctttac	2820
ctcttttaat	ggacaaacct	tcttcatccc	ccgacaaactc	tcattatttc	caacatggtc	2880
gggttgttgg	ccggaggattc	tgacgttcat	aacaatggat	atgttcttgg	gggacccgtt	2940
tttggaaac	tgttttgc	gttttgggtc	atgttccatc	tttacccctt	cctcaaagggt	3000
ttgtatggta	gacaaaaccg	aacaccggacg	atgttgc	ttttgcgttat	cctttctcgat	3060
tgcatcttc	ctttgggtt	ggttccggat	gatccgttct	tgcggaaaca	gacttggccg	3120
ttaactgaacg	aatgtggcgt	cgactgttaa				3150

```
SEQ ID NO: 2          moltype = AA    length = 1049
FEATURE             Location/Qualifiers
source              1..1049
                     mol_type = protein
                     organism = Arabidopsis thaliana
```

SEQUENCE : 2	MEPNTTMASFD DEHRHSSFS A KICKVCGDEV KDDDNQGTFV ACHVCVYPVC KPCYEYERSN	60
GNKCCPQCNT LYKRHKGSPK I AGDEEINGP DDSDDDELNIK YRQDGSSIHQ NFAYGSENGD	120	
YNSKQQWRPN GRAFSSTFSV LGKDFFEAERD GYTDAEWKER VDKWKARQEK RGLVTKGQT	180	
NEDKKEDDEEE YLDAEARQPL WRKVPPISSK ISPYRIVIVL RLVLVLFVFR FRILTPAKDA	240	
YPLWLISIVC EIWFALSWL DQFPKWRPM RETYDRLSLM RFERGEKHN LAPDVVFVST	300	
DVPLKEPEPII TANTILSILA DVYPVNKVSC YVSDDGASML FDLTDLSETSE FARRWVPFCK	360	
KYNVEPRAPE FYFSEKIDYL KDKVQTTFVK DRRAMKREYE EFKVRINALV AKAQKKPPEG	420	
WVMQDGTWPW GNNTRDHPGM IQVYLGKEGA FDIDNGELPR LVVYVSREKRP GYAHKKKAGA	480	
MNAMQVRVSAW LTNAFPMLNL DCDHYINNSK AIRESMCFLM DPQLGKKLCY VQFPQRFDGI	540	
DLDNRXANRN 1VFFD1NMRQ LDGQIPGVWP GTGCVFNPRY LYGYEPVPSE KRKKMTDCW	600	
PSWIICCCGG GNRNHKSDDS KKKSGIKSLF SKLKKKKKK SDDKTMSSYS RKRSSTEAIF	660	
DLEDIEEGLE GYDELEKSSL MSQKNFEKRF GMSPVFIASL LMENGGLPEA TNTSSLIKEA	720	
IHVISCGYE KTEWGWKEIGW IYGSVTDL TGFRMHCRGW KSVYCMKPKR AFKGSPAPINL	780	
SDRHLQHRLW ALGSWEIFPS RHCPFLWYAWG KGLKILERLA YINTIVYPPFT SIPLLYACTI	840	
PAVCLLTGKF IIPTINNFA S IWFLALFLSI IATAILELRW SGVSINDLWR NEEFWVIGGV	900	
SAHLLFAVPGQ LLKVLFGVDT NFTVTSKGAS DEADEFGDLY LFKWTLLIP PTTLIIILAMV	960	
GVVAGVSDAI NNGYGSWGP FLGKLFPAFWV IVHLYPFLKG LMGRQNRTPT IVVLLWSILLA	1020	
SIFSFLWVMRI DPFLPKOTGE B S QCGVGDV	1049	

```
SEQ ID NO: 3          moltype = DNA  length = 3150
FEATURE              Location/Qualifiers
source               1..3150
                     mol_type = other DNA
                     organism = synthetic constru
```

-continued

agtgtatgata	aaacgatgag	cagctatagt	agaaaacgt	catcgacgg	ggcgatttt	1980
gatttggaa	atattggaga	agggottgaa	ggatatgtat	agcttgagaa	atcttcgtta	2040
atgtcgcaga	agaatttga	gaagagattt	ggaatgtctc	ctgtgttat	tgcgtcgacg	2100
ttgatggaga	acggaggatt	gccggaggcg	acqaacacga	gctcggtat	caaagaagct	2160
attcatgtta	ttatgtgtgg	ttatgaagag	aagactgaat	ggggcaaaa	gattggatgg	2220
atttatggat	cagtaacaga	agatatccct	actggattt	gaatgcattt	tagaggttgg	2280
aaatcagttt	actgtatgcc	taaaagacca	gcttc当地	gatcagctcc	gatcaattt	2340
tctgatcggt	tgcaccaagt	tcttagatgg	gctcttgggt	ccgtc当地	cttc当地	2400
cgtcattgtc	cttgggtta	tgcttgggg	ggcaagctta	agattctaga	acgtcttgc	2460
tacatcaaca	caattgttta	tccttactt	tctatcttct	tacttgcata	ttgtactatt	2520
ccggccgtt	gtcttcatac	ccgcaagttt	atcatttctt	cgatcaacaa	tttc当地	2580
atatggttt	tagctgttt	cctatcaatc	atgc当地	ctatcttgc	actccgggt	2640
agtgggttaa	gcatcaatga	cctctggcgt	aacgagcaat	tctgggtat	tggcgggtgc	2700
tcagctcatc	tcttcgtct	cttctttaag	ttcttcttgg	ttgttgcact	2760	
aacttcaccg	tcacgtcaaa	aggagccgt	gacgaagccg	atgagtttgg	agatcttac	2820
ctctttaat	ggacacccct	tctcatcccc	ccgacaacactt	tcattatcc	caacatggc	2880
ggtgttgg	ccggagttt	tgacgctatt	aacaatggat	atggttcttgc	gggaccgtt	2940
tttggaaac	tggttttgc	gttgggttgc	attgttcatc	tttatectt	cctcaaagg	3000
ttgatggta	gacaaaaccc	aacaccgcg	attgttgc	tttgggttat	cctctcgct	3060
tegatcttct	ctttgggttgc	gttgggttgc	tgccgttcc	tgccgaaaca	gactggtcc	3120
ttactgaagc	aatgtggcgt	cgactgttta				3150

SEQ ID NO: 4	moltype = AA	length = 1049				
FEATURE	Location/Qualifiers					
source	1..1049					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 4						
MEPNTMASFD	DEHRHSSFS	KICKVCDEV	KDDDNGQTFV	ACHVCVYPVC	KPFYEEYERSN	60
GNKCCPQCNT	LYKRHKGSPK	IAGDEENNGP	DDSDDELNIK	YRQDGSSIHQ	NFAYGSENGD	120
YNSKQQWRPN	GRAFSSTGSV	LGDKEFAERD	GYTDAEWKR	VDKWKRQEKG	RGLVTKEQET	180
NEDKEDEEE	YLDAEARQPL	WRKVPISSK	ISPYRIVIL	RLVLFVFR	FRIITPAKDA	240
YPLWLISVIC	EIWFALSWIL	DQFPKWFPI	RETYLDRLSM	RFERDGEKNK	LAPVDVFVST	300
VDPLKEPEII	TANTILS	VDYPVNKVSC	YVSDDGASM	LFDTLSETSE	FARRWPVFCK	360
KYNVEPRAPE	FYFSEKIDYL	KDKVQTTFVK	DRRAMKREYE	EFKVRINALV	AKAQKKPEEG	420
WVMQDGTPWP	CNNTRDHPGM	IQVYLGKEGA	FDIDGNELPR	LVVVSREKRP	GYAHHKAGA	480
MNAMRVSAV	LTNAPFMLNL	DCDHYINNSK	AIRESMCFLM	DPQLGKLLCY	VQFPQRFDGI	540
DLNDRYANRN	IVFFDINMRG	LDGIQGPVYV	GTGCVFNRP	LYGYEPVVSE	KRKMTCDCW	600
PSWICCCCGG	GNRNHKSDF	SKLKKKTSF	SDDKTMSSYS	RKRSSTEAIF		660
DLEDIEEGL	GYDELEKSSL	MSQNKEFPR	GMSPVFIAS	LMENGGLPEA	TNTSSLIKEA	720
IHVISCGYEE	KTEWGKEIGW	IYGSVTEDIL	TGFRMHCRGW	KSVYCMKP	AFKGSAPINL	780
SDRLHQVLRW	ALGSVEIFFS	RHCPLWYANG	GKLKILERLA	YINTIVYPFT	SIPLLAYCTI	840
PAVCLLTGKF	IIPPTINNAS	IWFILALFLSI	IATAILELRW	SGVSINDLWR	NEQFWVI	900
SAHLFAVFGQ	LLKVLFGVDT	NFTVTSKGS	DEADEFGDLY	LFKWTTLIP	PTTLLI	960
GUUVAGVSDAI	NNGYGSWGPL	FGKLFIAFWV	IVHLYPFLKG	LMGRQNRTPT	IVVLWSILLA	1020
SIFSLVWVRI	DPFLPKQTGP	LLKQCGVDC				1049

SEQ ID NO: 5	moltype = DNA	length = 3255				
FEATURE	Location/Qualifiers					
source	1..3255					
	mol_type = other DNA					
	organism = Arabidopsis thaliana					
SEQUENCE: 5						
atgaacacccg	gtggcgggtt	aatcgccgtt	tccacaaca	ggaatggat	tgtcttcatt	60
atggccatgt	agaatggcccg	aataagatca	gtccaagacg	tgagtggaca	gacatgtca	120
atctgcagat	atggatgtca	attgtatgtt	gatggagaa	cggttgc	atgttaacgaa	180
tgtgcattcc	ctgtgtgttag	accttgcata	gagtgacaa	gacgaaagg	caatcaagct	240
tgtccacagt	gaaaaaccccg	tttcaaacgt	ctttaaggaa	gtccaaaggt	tgaaggtat	300
gaagaggaa	atgacatttgc	tgatgttagtgc	aatgtatggaaa	taatgggat		360
ggatgttgc	aggtttctgt	aggtatgtca	atctctcg	gcaactccgg	tttcccacaa	420
tctgatgttt	attcagctcc	acctggctt	catatccat	tgcgtactt	ccggcagcag	480
gacgttgc	tttcttctgt	tagacatgt	cttattgttc	ctccctact	ttgtggtcat	540
ggcaatagag	ttcatctctgt	ttcttttctt	gaccgcacc	tggctgcaca	tccaggcct	600
atggatctt	agaaaatgtt	tgccgttata	gtgtatggaa	gtgtcgcttgc	gaaaatcg	660
atggagaaat	ggaagagaaaa	gcagaatgtat	aaacttcagg	ttgtttaggc	tgaaggagat	720
cctgatgttt	aatgtgtgtt	tgatgtctgt	tttccatgt	ttgtatgggg	aaggcagcc	780
ttgtcttaga	agataccat	caaatacgat	aagataaaatc	cttaccggat	gttattgt	840
ctacgttctt	tgattcttgc	tctttctt	cactaccgtt	ttcttcaccc	cgtcaagat	900
gcataatgtt	tgtggcttat	ttctgttata	tgttgatgtat	gttttgctgt	ttcatgggtt	960
cttgatgtt	tccctaaatgt	gtaccctatc	gagcggagaaa	cgtaacttgg	ccgactctca	1020
ttaaagatgt	agaaaagaa	gaaaccgtcg	ggactatccc	ctgtggatgt	attttgtat	1080
acagtggatc	cattgttgc	gcctccgtt	attactgc	atactgttt	gtcttatttt	1140
gctgttgc	atccgttgc	taagggttgc	tgttacgtat	ctgtatgttgc		1200
cttactttcg	aagcttttgc	tgagaccgtt	gaatttgc	ggaaatgggt	tcctttctgc	1260
aagaaatatt	gtattgttgc	tcgtgttcc	gaatgttgc	tctgtccata	aatggactac	1320
ttgaagaata	aagttcatcc	cgcattttgtt	agggagccgc	gagccatgaa	gagagat	1380

-continued

gaagaattca	aagtaaagat	caatgcttta	gtagcaacag	cacagaaaat	gcctgaggat	1440
ggtttggacta	tgcaagacgg	tacaccttgg	cccggtaaata	gtgtggaga	tcatccttgc	1500
atgattcagg	tcttcttgg	aagtgcgggt	gttcgtatg	tgcaaaaacaa	cgagttgcct	1560
cgatttagtt	acgtttctcg	tgagaagaga	cccggttttgc	atcaccat	aa gaaggcttgc	1620
gctatgaatt	ccctgatacg	agtctctggg	gttcttatcaa	atgctcttca	ccttctgaat	1680
gtcgatgtg	atcactacat	caacaatago	aaagcttca	gagaagcaat	gtgtttcatg	1740
atggatcctc	agtccaggaa	gaaaatcttct	tatgttca	tatgttca	aaaggttgcgg	1800
atttgatagc	acgatcgata	ctcaaatcg	aatgttgcgt	tctttgat	caatgttgc	1860
ggtttggatg	ggcttacaagg	gcctataac	gtcggttac	gttgtgttt	caggaggccaa	1920
gegccttacg	gatttgcgt	accgttacgg	aquaaggaaag	aaagaaaggcc	cacgttacgc	1980
tgcccaaaat	gtgttgcct	atgtttgtt	tcaagaaaga	atcgtaaagg	aaagacacgt	2040
gtctcgatata	agaagaaagaa	gaatggaa	gcgttcaaa	agatccacgc	attagaaaaat	2100
atcgaagagg	gcccgcgtac	taaagggttct	aacgttagac	agtcaaccg	ggcaatgc	2160
atgaagtgg	agaagaaatt	tggggcgtt	cctgttattt	ttgcatctgc	gcgtatgg	2220
aatgggttgg	tggcttgc	cgcaagcccg	gttgcgttgc	ttaaagaagg	catccaatgc	2280
attagtgcg	gatatgttgc	taaaatgttgc	tggggaaat	agatgttgcgt	gatctatgtt	2340
tctgttaccg	aaagatattct	tacgggtttt	aatgtcatt	ctcatgttgc	gagatcttgc	2400
tattgttacac	caaaggtagc	ggcttcaaa	ggatcgttgc	caatcaatct	ttcggatcgt	2460
ctccatcaag	tcttcgtat	ggcgttgcgt	tcgggttgc	ttttcttgc	taggcattgt	2520
cctatgttgc	atgttgcgtt	agggtgggttgc	aatatgttgcgt	agcgttgcgt	ctacattaaac	2580
tctgttgcgtt	accctgttgc	ctcttaccc	ctcatgttgc	actgttgcgt	ccctgcacatc	2640
tgtcttctca	ctggaaaatt	categttcc	gagatttagca	actatgcgg	tatctcttc	2700
atggcgttct	tcttcgtat	tgcataat	acgttgcgt	ggatcgttgc	ggcaaaatgt	2760
gggtatcgt	atgggttgc	aaacgttacgg	ttttgggtca	tttgggttgc	tttgcgtc	2820
ctgttgcgt	tcttcgttgc	tcttcgtat	gttcttgcgt	gttgcgttgc	taacttcaca	2880
gtcacatca	aacgttgcgt	tgtatggag	ttctcttgc	tttacccctt	caaattgttgc	2940
tcacttctca	tccttcgtat	gacttactt	atcataaaat	atcattgttgc	catatgttgc	3000
gtctctgtat	ccatcgttgc	tggatcgttgc	tcgttgcgttgc	cgcttttgc	aagactgttgc	3060
tttgactt	gggtatcgtat	tcatcttac	ccgttccat	aaagggttgc	tggaaacaa	3120
gatagaatgc	caaccattat	tgtatgttgc	tccatcttgc	tggcgttgc	tcttacactt	3180
cttgggttgc	gggttataatcc	gttgcgttgc	tttgcgttgc	ctatttgcgt	gatctgttgc	3240
tttagactgt	tgtatgttgc	tttgcgttgc	aaaggcggttgc	ctatttgcgt	gatctgttgc	3255

SEQ ID NO: 6	moltype = AA	length = 1084
FEATURE	Location/Qualifiers	
source	1..1084	
	mol_type = protein	
	organism = Arabidopsis thaliana	
SEQUENCE: 6		
MNTGGRLIAG SHNRNEFVLI NADENARIRS VQELSGQTCQ ICRDEIELTV DGEPEPVACNE	60	
CAFPVCRPCY EYERREGNQA CPQCKTRFKR LKGSPRVEGD EEEEDIDDD NEFEYGNNGI	120	
GFDQVSEGMS ISRRNSGFPQ SDLDSAPPGS QIPLLTYGDE DVEISSLRHA LIVPPSLGGH	180	
GNRVHPVSLS DPTVAHAPRP MVPQKDLAVY GYGSAWAKDR MEEWKRKQNE KLQVVRHEGD	240	
PDFEDGDDAD PDMDEGRQP LSRKLPKISS KINPYRMLIV LRLVILGLFF HYRILHPVKD	300	
AYALWLISVI CEIWFAVSVW LDQFPKWPYI ERBTYLDRLS LRYEKEGKPS GLSPVDFVFS	360	
TVDPKLEPPL ITANTVLSIL AVDYPVDKVA CYVSDDGAAM LTFEALSETA EFARKWVPFC	420	
KKYCIEPRAP EWYFCHKMDV RERRAMKRDV EEFKVKINAL VATAQKVPED	480	
GWTMQDGTGW PGNSVSRDHPG MIQVFLGSDG VRDVENNELP RLVYVSREKR PGFDHHKKAG	540	
AMNSLIRVSG VLSNAPYLLN VDCDHYINNS KALREAMCFM MDPQSGKKIC YVQFPQRFDG	600	
IDRHDRYSNR NVVFFDINMK GLDGLQGPYI VGTGCVFRQQ ALYGFDAPKK KKGPRKTCNC	660	
WPKWCLLCFG SRKNRKAFTV AADKKKKNRE ASKQIHALEN IEEGRVTKGS NVEQSTEAMQ	720	
MKLEKKFGQS PVFVASARME NGGMARNASP ACLLKEAIQV ISCGYEDKTE WGKEIGWIY	780	
SVTEFDILTGK KMHSHGWRSV YCTPKLAAFK GSAPINLSDR LHQVLRWALG SVEIFLRSRC	840	
PIWYGYGGGL KWLERLSYIN SVVYWPWTSLP LIVYCSLPAI CLLTGKFIIP EISNYASILF	900	
MALFSSIAIT GILEMQWGVK GIDDWWRNHQ FWIWIGGSAH LFALFQQLLK VLAVGDTNFT	960	
VTSKAADDGE PSDLYLFWKT SLLIPPMTLL IINVIGIVVG VSDAISNGYD SWGPLFGRFL	1020	
FALWVIIHLY PFLKGLLGKQ DRMPTIIVVV SILLASILTL LWVRVNPVVA KGGPILEICG	1080	
LDCL		1084

SEQ ID NO: 7	moltype = DNA	length = 3255
FEATURE	Location/Qualifiers	
source	1..3255	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 7		
atgaacacccg gtggcgggtt aatcgccgtt tctcacaaca ggaatggatgt tgccttcat	60	
aatgccatcg aagaatccccg aataagatca gtccaaagacg tgatggaca gacatgttca	120	
atctgcagat atgagatcga attgactgtt gatggacaacg cgtttgcgc atgtaacgaa	180	
tgtgcattcc ctgtgttgcgtt accttgcata gatgtacgaa gacgagaagg caatcaatgt	240	
tgtccacatgtt gcaaaaacccg ttccaaacgtt cttaaaggaa gtccaaaggat tgaagggtat	300	
gaagaggaaatg atgacatgttgcgtt aatgttgcgtt agatgttgcgtt taatgggttgcgtt	360	
ggatgttgcgtt aggtttctgttgcgtt aatgttgcgtt accttgcata gcaactccgg ttcccaatgttgcgtt	420	
tctgatttttttgcgtt accttgcata gatgttgcgtt cggcgacgcg atgatgttgcgtt	480	
gacgttgcgtt ttcttgcgtt accttgcata gatgttgcgtt tggcgttgcgtt tggcgttgcgtt	540	
ggcaatagatgc ttcatctgttgcgtt accttgcata gatgttgcgtt tggcgttgcgtt tggcgttgcgtt	600	
atgttgcgtt acaaatgttgcgtt accttgcata gatgttgcgtt tggcgttgcgtt tggcgttgcgtt	660	

-continued

```

atggagaaat ggaagagaaa gcagaatgag aaacttcagg ttgttaggca tgaaggagat 720
cctgattttg aagatggtg tgatgtctgat tttccaaatgaa tggatgaggg aaggcagcc 780
ttgtcttaga agataccaat ccaatcgago aagataatac cttaccggat gttaatttgt 840
ctacgtcttgc tgatcttgg tctcttctt cactaccgtt ttcttcaccc cgtcaaaatg 900
gcatatgttt tttgtggcttat ttctgttata tttgtgagat gttttgtgtt ttcatgggtt 960
cttgcattgtt tccctaaatg gtaccctatc gaggcagaaat cgtacttggc ccgactctca 1020
ttaagatgtt agaaaagaagg gaaaccgtcg ggactatccc ctgtggatgt atttgttagt 1080
acagtggatc catttggaaa gctcccggtt attactgtttt atactgtttt gtcttatttt 1140
gctgttggat atccgtcgaa taagggtgtt ttgttacgtat ctgataatgg tgctgtatg 1200
cttacttccg aagcttcgtt tgagacccgtt gaatccgcaaa gggaaatgggtt tccttcgtc 1260
aagaatattt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1320
tttggaaatca aagtttcatcc cgcattttttt tttttttttt tttttttttt tttttttttt 1380
gaagaattca aagtttcatcc cgcattttttt tttttttttt tttttttttt tttttttttt 1440
ggtttggacta tgcaagacgg tacacccgtt cccggtaatgg tttttttttt tttttttttt 1500
atgatttcgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1560
cgattttttt acgttttcgtt tgagaaatggaa cccggattttt atccatcaa gaaggctgg 1620
gctatgttattt ccctgtatcc agtctctggg gttcttatcaa atgcttctt ccttctt 1680
gtcgatttttgc atcaactatcc caaatatggaa aagatcttgc gttttttttt tttttttttt 1740
atggatccctc agtccggaaa gaaaatctgtt tatgttcttgc tttttttttt tttttttttt 1800
atttgcatttttcc agatccgtatcc caaatatggaa aagatcttgc gttttttttt tttttttttt 1860
ggtttggatg ggcttacaagg gccttataatcc gttttttttt tttttttttt tttttttttt 1920
gcgccttttccg gatttttttgc accggaaatggaa aagaaggggcc cacgttgc gttttttttt 1980
tggccaaat ggtgttcttgc atggttttttgc tttttttttt tttttttttt tttttttttt 2040
gtctgcggatc aagaaatggaa gaaatggggaa ggttgcggatc gttttttttt tttttttttt 2100
atcgaagagg ggcgcgttcc taaagggttcc aacgttgc gttttttttt tttttttttt 2160
atgaagtttttggaa aagaaatggaa tttttttttt tttttttttt tttttttttt tttttttttt 2220
atatgggttttggggatccggcc gttttttttt tttttttttt tttttttttt tttttttttt 2280
attatgggttttggggatccggcc gttttttttt tttttttttt tttttttttt tttttttttt 2340
tctgttccggcc aagatatttttgc tttttttttt tttttttttt tttttttttt tttttttttt 2400
tatttgcacac caaatatggaa aagatcttgc gttttttttt tttttttttt tttttttttt 2460
cttccatccatcc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2520
cttatttttttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2580
tctgttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2640
tgtcttctca ctggaaat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2700
atggcgttcc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2760
gggatcgtatcc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2820
ctgttttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2880
gtcacatcaa aagcagcttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2940
tcacttctca tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3000
gtctctgtatcc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3060
tttgcatttttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3120
gatagaatgc caaccattat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3180
cttttttttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3240
tttagactgtatcc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3255

```

```

SEQ ID NO: 8 moltype = AA length = 1084
FEATURE Location/Qualifiers
source 1..1084
mol_type = protein
organism = synthetic construct

```

```

SEQUENCE: 8
MNTGGRLIAG SHNRNEFVLI NADENARIRS VQELSGQTQ ICRDEIELTV DGEPPVACNE 60
CAFPVCRPCY EYERREGNQA CPQCKTRFKR LKGSPPRVEGD EEEEDDIDDD NEFEYGNNGI 120
GFDQVSEGMS ISRRNSGFPQ SDLDSAPPGS QIPLLTYGD DVEISSDRHA LIVPPSLGGH 180
GNRVHPSPS DPTVAAHPRP MVPQKDLAVY GYGSVAWKR MEEWKRKQNE KLQVVRHEGD 240
PDFEDGDDAD PMPPMDRGQP LSRKIPKINQ KINPYRMLIV LRLQVILGLFF HYRILHPVKD 300
AYALWLISVI CEIWFAVSVW LDQFPKWKYP ERETYLDRLS RLYKEKGKPS GLSPVDFVFS 360
TVDPLKEPPL ITANTVLSIL AVDYPVDKVA CYVSDNGAAM LTFEALSETA EFARKWPFC 420
KJYCIEPRAP EWYFCHKMDV LKNNVHPAFV RERRAMKRDY EEFKVKINAL VATAQKVPED 480
GWTMQDTPSW PGNNSRHDHPG MIQVFLGSDG VRDVENNELP RLVYVSRERK PGFDHHKKAG 540
AMNSLIRVPSG VLNSNAPYLLN VDCDHYINNS KALREAMCFM MDPQSGKKIC YVQFPQRFDG 600
IDRHDRYSNR NVVFFDINMK GLDGLQGPIY VGTGCVFRQQ ALYGFDAPKK KKGPRKTCNC 660
WPKWCLLCFG SRKNRKAATV AADKKKKNRE ASKOIHLEN IEEGRVTIKGS NVEQSTEAMQ 720
MKLEKKPGQS PVFVASARMV NGGMARNASP ACLLKEAIQV ISCGYEDKTT WGKEIGWIYG 780
SVTEDILTGK FKHSHGWRSV YCTPKLAASF GSAPINLSDR LHQVLRWALG SVEIFLSRHC 840
PIWYGYGGGL KWLERLSYIN SVVYWPSTSLP LIVYCSLPAI CLLTGKFIVP EISNYASILF 900
MALFSSIAIT GILEMQWGKV GIDDDWWRNEQ FWVIGGVSAA LFALFQGLLK VLAVGVDTNFT 960
VTSKAADDGE PSDLYLFKWT SLLIPPMTLL IIINVIGVVG VSDAISNGYD SWGPLFGRFL 1020
FALWVIIHLH PFLKGLLGKQ DRMPITIIVVV SILLASILTL LWVRVNPVVA KGGPILEICG 1080
LDCL 1084

```

```

SEQ ID NO: 9 moltype = DNA length = 3255
FEATURE Location/Qualifiers
source 1..3255
mol_type = other DNA
organism = synthetic construct

```

-continued

```
SEQ ID NO: 10          moltype = AA  length = 1084
FEATURE                Location/Qualifiers
source                 1..1084
                      mol_type = protein
                      organism = synthetic construct
```

-continued

SVTEDILITGF	KMHSHGWRSP	YCTPKLAAFK	GSAPINLSDR	LHQVLRWALG	SVEIFLSRHC	840
PWYGYGGGL	KWLRLSYIN	SVVYPWTSPL	LIVYCSLPAI	CLLTGKFIVP	EISNYASILF	900
MALFSSIAIT	GILEMQWGVK	GIDDWWRNEQ	FWVIGGVSAN	LFALFQGLLK	VLAGVDTNFT	960
VTSKAADDGE	FSDLYLFKWT	SLLIPPMTLL	IINVIGVIVG	VSDAISNGYD	SWGPILEICG	1020
FALWVIIHLY	PFLKGGLGKQ	DRMPTIIVVW	SILLASILTL	LWVRVNPFA	KGGPILEICG	1080
LDCL						1084

SEQ ID NO: 11	moltype = DNA length = 3255					
FEATURE	Location/Qualifiers					
source	1..3255					
	mol_type = other DNA					
	organism = synthetic construct					
SEQUENCE: 11						
atgaacaccg	gtggcggtt	aatcgccgtt	tctcacaaca	ggaatgagtt	tgtccctatt	60
aatgccatg	agaatgcccgg	ataaagatca	gtccaaaggc	tgagtgcaca	gacatgtcaa	120
atctgcacag	atgagatcga	attgactttt	gatggagaac	cgtttgtggc	atgtaacgaa	180
tgtgcattcc	ctgtgttag	accttgcata	gagtaaaaaa	gacgagaagg	caatcaagct	240
tgtccacagt	gaaaaaccgg	tttcaaaacgt	cattaaaggaa	gtccaaagg	tgaagggtat	300
gaagagggaa	atgacatgt	tgatgttag	aatggatgg	atgtatggaa	taatgggatt	360
ggatttgc	atgtttctga	atgtatgtca	atctctgc	gcaactccgg	tttccccacaa	420
tctgattttt	attcagctcc	acctggctct	cagattccat	tgctgactta	cggcgacgag	480
gacggttggaa	tttcttcgt	tagacatgt	cttattgttc	tccttcact	tgggtgtcat	540
gcoaatagag	ttcatectgt	ttcttcgttct	gacccgaccg	ttgtgcaca	tccaaaggcct	600
atggtaacctc	agaaatgtct	tgccgtttat	gtttagggaa	gtgtcgctt	gaaagatccg	660
atggggaaat	ggaagagaaa	gcagaatgt	aaacttcagg	ttgttaggca	tgaaggagat	720
cctgattttt	aagatgtgt	tgatgtctgt	tttccaatgt	tggatgagg	aaggcaggca	780
ttgtcttagga	agataaaat	caaatacgag	aaagataatc	tttacccgt	gttaatttgt	840
ctacgtcttg	tttattttt	tcttcgttct	cactaccgt	ttcttcaccc	cgtcaaaat	900
gcatatgttt	tgtggcttat	tttctgtata	tgtgagat	gttttgctgt	ttcatgggtt	960
cttgcattatgt	ttcccaatgt	gtaccctata	gagcggaaa	ctgacttgg	ccgactctca	1020
ttaagatgtat	aaaaaaagg	gaaacgtcg	ggactatccc	ctgtggatgt	atgttttagt	1080
acagtggatc	cattgaaaga	gcctcgttct	attactgc	atactgtctt	gtctattttt	1140
gctgttggatt	atccgtcg	taagggtgt	tgttacgtat	ctgtatgtt	tgctgctatg	1200
cttactttcg	aaagtctttc	tgagaccgt	gaattcgc	ggaaatgggt	tccttcgtc	1260
aagaaatattt	gtatttgag	tcgtgtctt	gaatgttatt	tctgcattaa	aatggactac	1320
tttgcattatcc	aaagtcatcc	cgccattttt	aggggagcgg	gagccatgaa	gagagattat	1380
gaagaatttca	aatgtttat	caatgttta	gtaccaac	cacagaaat	gcctcgaggat	1440
ggttggacta	tgcaagacgg	tacacccgtt	cccggtataa	gtgtgcgaga	tcatccctgg	1500
atgattcagg	tcttcgttgg	aaatggcgat	tttcgtgtat	tcgaaaaacaa	cgagttgcct	1560
cgatttagttt	acgtttctcg	tggatggat	cccggtattt	atccattaa	gaaggcttgg	1620
gctatgtttt	ccctgtatcc	agtttgcatt	tttcttatca	atgtcttca	cctctgtat	1680
gtcaatttgt	atcaactat	caacaatgt	aaagcttta	gagaagcaat	gtgtttcat	1740
atggatcttc	agtcaagaa	gaaaatctgt	tatgttgc	tccctcaaa	gttcgtatgg	1800
attgtatggc	aaatgttgcata	cttcaatgc	aatgttgc	tctttgtat	catatgtaaa	1860
ggtttggatg	ggctcaagg	gcctatata	gtcggat	gtttgtt	caggaggca	1920
gcgttttacg	gatttgatgc	accgaagaag	aagaaggcc	cacgtaaag	atgcaattgc	1980
tggccaaat	gtgttgcctt	atgttttgc	tcaaaaaaa	atcgtaaag	aaagacatgt	2040
gctgcggata	agaagaaagaa	gaatagggg	gggtcaaa	agatccacg	attagaaaaat	2100
atcgaaggagg	gcccgtc	taagggttct	aaatgggtt	agtcaacca	ggcaatgca	2160
atgaagttgg	agaagaaattt	tggcagtt	cctgtattt	tttgcatt	tgctatggag	2220
atatggggaa	tggcttggaa	cgcaacccgg	gtttgtctc	ttaaaaag	catccaatgc	2280
attatgttgc	gatatgtt	taaaaactt	ttggggaaa	agatgggtt	gttctatgt	2340
tctgttaccg	aaatgttct	tacgggttt	aaatgttgcatt	cttgcatttgc	gagatctgt	2400
tattgttac	caaagttagc	ggcttcaaa	ggatcgttgc	caatcaat	ttcggatgt	2460
ctccatcaag	tttccatcgat	ggcgttggg	tcgggttgc	ttttcttgc	taggcattgt	2520
ccttattttgt	atgggttgc	agggttgggtt	aaatgggttgc	agcgttgc	ctacattaa	2580
tctgttgc	accgttggac	cttcttgc	cttgcatttgc	actgttctt	ccctgcattc	2640
tgtcttgc	ctggaaaattt	catgttccc	gagatttgc	actatgc	gatgttgc	2700
atggcgctct	tctgttgc	tgcaataac	ggatttgc	agatgttgc	ggccaaatgt	2760
gggatcgtat	atgggttgg	aaacaaac	tcgttgc	tttggggat	ttctgcgt	2820
ctgtttgc	tcttccaa	tttgcatttgc	tttgcatttgc	gtgtcgat	taacttca	2880
gtcacatca	aagcgttgc	tgatgttgc	tttgcatttgc	tttgcatttgc	caaatgttgc	2940
tcaatttgc	tccctcaat	gacttactt	atcataaaac	tcatttgc	catatgttgc	3000
gtctctgtat	ccatcgac	tggatcgt	tcgttgc	tttgcatttgc	catatgttgc	3060
tttgcatttgc	gggttcat	tcatatttgc	tcgttgc	tttgcatttgc	catatgttgc	3120
gatagaatgtc	caaccattat	tgtgttgc	tcatttgc	tcgttgc	tttgcatttgc	3180
cttgggttcc	gggttatacc	gtttgttgc	tttgcatttgc	tcatttgc	catatgttgc	3240
tttagactgt	tgttgc					3255

SEQ ID NO: 12	moltype = AA length = 1084					
FEATURE	Location/Qualifiers					
source	1..1084					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 12						
MNTGGRLIAG	SHNRNEFVLI	NADENARIRS	VQELSGQTCQ	ICRDEIELTV	DGEPPVACNE	60

-continued

CAFPVCRPCY	EYERREGNQA	CPQCKTRFKR	LKGSPRVEGD	EEEDDDDDLD	NEFEYGNNGI	120
GFDQVSEGMS	ISRRNSGFPQ	SDLDSAPPGS	QIPLLTYGDE	DVEISSDRHA	LIVPPSLLGHF	180
GNRHPVPSL	DPTVAAHPRP	MVPQKDLAVY	GYGSVAKDR	MEEWKRKQNE	KLQVVRHEGD	240
PFDENGDDAD	FPMMDEGRQP	SLRKLPIKPKS	KINPYRMLIV	LRLVILGLFF	HYRLHVPVKD	300
AYALWLISVI	CEIWFAVSWU	LDQFPKWYPI	ERETYLDRLS	LRYEKEGKPS	GLSPDVVFVS	360
TVDPKKEPPL	ITANTVLSIL	AVDYPVDKVA	CYVSSDDGAAM	LTFEALSETA	EFARKWVPF	420
KKYCIEPRAP	EWYFCHKMDY	LKNKVHPAFV	RERRAMKRDY	EEFKVTKINAL	VATAQKVPED	480
GWTMQDGTWP	PGNSVRDHGP	MIQVFLGSDG	VRDVENNELP	RLVYVSREREK	PGFDHHKKAG	540
AMNSLRIVSG	VLSNAPAYLN	VNCDIHYINNS	KALREAMCFM	MDPOSGGKIC	YVQFPQRFDG	600
IDRHDRYSNR	NVVFFDINMK	GLDGLQGPI	VGTGCVFRRQ	ALYGFADPKK	KKGPRKTNC	660
WPWKWCLLCFG	SRKNRKAHTV	AADKKKKRNRE	ASKQIHALEN	IEEGRVTKGS	NVEQSTEAMQ	720
MKLEKKFGQS	PVFVASARME	NGGMARNASP	ACLLKIAIQV	ISCGYEDKTE	WGKEIGWIGY	780
SVTEDLTGTF	KMHSHGWRSV	YCTPKLAIFG	GSAPINLSDR	HLQVRLWALG	SVEIFLRSHC	840
PIWYGYGGGL	KWLRLERSYIN	SVVPWTSL	LIVYCSLPAI	CLLTGKFIVP	EISNYASILF	900
MALFSSIAIT	GILEMOWGKV	GIDDWWRNEQ	FWWIGVGSAH	LFALFQGLLK	VLAGVDTNTFT	960
VTSKAADDGE	FSDLYLFWT	SLLIPPMTL	IINVIGVIVG	VSDAISNGYD	SWGPLFGRLF	1020
FALWVIIHLY	PFLKGLLGKQ	DRMPTIIVVW	SILLASILTL	LWVRVNPFPVA	KGGPILEICG	1080
DLCL						1084

SEQ ID NO: 13 moltype = DNA length = 3255

FEATURE	Location/Qualifiers
---------	---------------------

source 1..3255

mol_type = other DNA

atgaacaccg gtggcggtt aatcgcccg ttcacaaca ggaatgagg ttgcctcatt 60
 aatccgcgtt agaatggcc aataagatca gtcggaaagc tgaggggaca gacatgtcaa 120
 atctgcagag attagatcgta attgactgtt gatggagaac cgtttgtggc atgttaacgaa 180
 tgtgcattcc ctgtgtgt tag accttgctt gaggacaaa gacgagaagg caatccaagt 240
 tgtccacagt gcaaaaacccg tttcaaacgt cttaaaggaa gtcggaaaggt tgaagggtat 300
 gaaggaggaa atgacatttg tgatttagac aatggatgtt agtagtggaa taatggggaa 360
 ggatgttgcgat aggttctgtt aggtatgtca atcttcgtt gcaactccgg tttccacaa 420
 tctgatttgc attcagctcc acctggctt cagatccat tgcgtactt cggcgcacg 480
 gacgttggaa ttcttctgtt tagacatgtt ctattttgtt ctcttcact tgggtgttgc 540
 ggcataatagg ttcatctgtt ttcttctt gaccggaccc tggtgcaca tccaggcct 600
 atggatcttc agaaagatctt tgccgtttt ggtttagggaa tggtgcctt gaaagatcg 660
 atggaggaaat ggaagagaaa gcagaatgg aaacttcagg ttgttaggc tgaaggagat 720
 ctgtatgttgc aagatggta tgatgtgtt ttccaaatg tggtaggggg aaggcagcc 780
 ttgttcttaga agataccat caatcgac aagataatac ttaccggat gttattgttgc 840
 ctacgttgc tgattttttt ctcttctt cactacgtt tttttccccccgttcaaggat 900
 geatatgtttt tttctgttata tttctgttata tttctgttata tttctgttata tttctgttata 960
 ctgtatcgtt tttctgttata tttctgttata tttctgttata tttctgttata tttctgttata 1020
 ttaatgtatgtt aagaaagaaat gaaaccgtcg gggacttccctt ctgttgcgtt atttttgtatgtt 1080
 acagtggatcc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1140
 getgttgcattt atccctgttca taagggttgc tttttttttt tttttttttt tttttttttt 1200
 cttaatcttcg aagcttccctt tgagaacccgtt gttttttttt tttttttttt tttttttttt 1260
 aagaatataat ttatgttgc tttttttttt tttttttttt tttttttttt tttttttttt 1320
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1380
 gaagaattca aagttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1440
 ggttggacta tgcaagacccg tacacccctt tttttttttt tttttttttt tttttttttt 1500
 atgttgcatttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1560
 cgatgttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1620
 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1680
 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1740
 atggatccctt aatgttgc tttttttttt tttttttttt tttttttttt tttttttttt 1800
 attaatggcc acatgttgc tttttttttt tttttttttt tttttttttt tttttttttt 1860
 ggtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1920
 ggcgttttacg gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1980
 tggccaaaaat ggttgcatttgc tttttttttt tttttttttt tttttttttt tttttttttt 2040
 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2100
 atcgaaaggcc ggcgttttacg tttttttttt tttttttttt tttttttttt tttttttttt 2160
 atgaaggatgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2220
 aatgggtggaa tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2280
 tggcttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2340
 attatgttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2400
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2460
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2520
 ctccatcaag tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2580
 cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2640
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2700
 atggcgctt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2760
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2820
 gggatcgatg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2880
 atgggtggaa tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2940
 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3000
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3060

-continued

tttgcactt	gggtcatcat	tcatctttac	cgttcctta	aaggtttgt	tggaaacaa	3120
gatagaatgc	caaccattatc	tgtcgctctgg	tccatctcc	tggcctcgat	tcttacactt	3180
cttgggtcc	gggttaatcc	gtttgtggcg	aaaggcggtc	ctattctcga	gatctgtgg	3240
tttagactgct	tgtga					3255
 SEQ ID NO: 14		moltype = AA	length = 1084			
FEATURE		Location/Qualifiers				
source		1..1084				
		mol_type = protein				
		organism = synthetic construct				
SEQUENCE: 14						
MTGGRLLIAG SHNRNEFVLI NADENARIRS VQELSGQTQ ICRDEIELTV DGEPPVACNE				60		
CAFPVCRPCY EYERREGNQA CPQCKTRFKR LKGSPRVEGD EEEIDDLD NEFEYGNNGI				120		
GFDQVSEGMS ISRRRNQGSD SSDLSDAPPGS QIPLLTYGDE DVEISSLRHA LIVPPSLGGH				180		
GNRVHPVLSL DPTVAAPRPM MVPQKDLAVY GYGSAWKRDR MEEWKRKQNE KLQVVRHEGD				240		
PDPEDEGDDAD FPMMDEGRQP LSRKIPIKSS KINPYRMLIV LRLVILGLFF HYRILHPVKD				300		
AYALWLISVI CEIWFAVSWV LDQFPKWPI ERETYLDRRLS LRYEKEGKPS GLSPVDFVFS				360		
TVDPKKEPPL ITANTVLSIL AVDYPVDKVA CYVSDDGAAM LTFEALSETA EFARKWVPFC				420		
KKYCIEPRAP EWYFCHKMDY EEFKVKINAL VATAQKVPED				480		
GWTMQDGTWP PGNSVRDHDPG MIQVFLGSDG VRDVENNELP RLVYVSREREK PGFDHHKKAG				540		
AMNSLIRVSG VLSNAPYLLN VDCDHYINNS KALREAMCFM MDPQSGKKIC YVQFPQRFDG				600		
INRHDRYSNR NVVFFDINMK GLDGLQGPYI VGTGCVFRQQ ALYGFADPKK KKGRPKTCNC				660		
WPWKWCLLCFG SRKNRKAKTV AADKKKKNNRE ASKQIHALEN IEEGRVTKGS NVEQSTEAMQ				720		
MKLEKKFGQS PVFVASARM NGGMARNASP ACLLKEIAQV ISCGYEDKTE WGKEIGWIYG				780		
SVTEDILTGK KMHSHGWRSV YCTPKLAAPK GSAPINLSDR LHQVLRWALG SVEIFLSRHC				840		
PWYGYGGGL KWLERLRSYIN SVVYPTWTSLP LIVYCSLPAI CLLTGKPIV P EISNYASILF				900		
MAFLSSIAIT GILEMQWGKV GIIDWWRNRE FWIIGGVSAH LFALFQGLLK VLAVGVDNTFT				960		
VTSKAADDGE PSDLYLFKWT SLLIPPMTLL IINVIGVING VSDAISNGYD SWGPLFGRLF				1020		
FALWVIIHLY PFLKGGLGKQ DRMPTIIVVVW SILLASILTL LWVRVNPFA KGGPILEICG				1080		
LDCL				1084		
 SEQ ID NO: 15		moltype = DNA	length = 3255			
FEATURE		Location/Qualifiers				
source		1..3255				
		mol_type = other DNA				
		organism = synthetic construct				
SEQUENCE: 15						
atgaacaccc gttggcgggt aatcgccgt ttcacacaaca ggaatggat tgcctcatt				60		
aatggccatg agaatggccc aataagatca gtccaaagac tgagtggaca gacatgtcaa				120		
atctgcagag attagatcg attgtatggat gatggagac ctgttgccg atgtaaacgaa				180		
tgtgcattcc ctgtgtgtat accttgcata gatgtacaaa gacggaaagg caatcaagct				240		
tgtccacatg gcaaaaacccg tttcaaacgt cttaaaggaa gtccaaagat tgaaggatgt				300		
gaagaggaaat atgacatttgc tgatgtatggaaat aatgtatggatggatggatggat				360		
ggatgtatggatggatggatggatggatggatggatggatggatggatggatggatggatggat				420		
tctcgatggatggatggatggatggatggatggatggatggatggatggatggatggatggat				480		
gacgttggata ttctttctgtatggatggatggatggatggatggatggatggatggatggat				540		
ggcaatagatg ttccatctgtatggatggatggatggatggatggatggatggatggatggat				600		
atggatccatc aaaaatgtatggatggatggatggatggatggatggatggatggatggatggat				660		
atggaggaaat ggaagagaaa gcaaaatgtatggatggatggatggatggatggatggatggat				720		
cctgatggatggatggatggatggatggatggatggatggatggatggatggatggatggatggat				780		
ttgtcttagaa agatataatcaatgtatggatggatggatggatggatggatggatggatggat				840		
ctacgttcttgc tgatttcttgc tgatgtatggatggatggatggatggatggatggatggat				900		
gcataatgtatggatggatggatggatggatggatggatggatggatggatggatggatggatggat				960		
cttgatcatgtatggatggatggatggatggatggatggatggatggatggatggatggatggat				1020		
tttaagatgtatggatggatggatggatggatggatggatggatggatggatggatggatggat				1080		
acatgttgcattgtatggatggatggatggatggatggatggatggatggatggatggatggat				1140		
gtctgttgcattgtatggatggatggatggatggatggatggatggatggatggatggatggat				1200		
cttactttcg aagcttttc tgagaccgtatggatggatggatggatggatggatggatggat				1260		
aagaaatattgtatggatggatggatggatggatggatggatggatggatggatggatggat				1320		
tttgaagatataatgtatggatggatggatggatggatggatggatggatggatggatggat				1380		
gaagaatgtatggatggatggatggatggatggatggatggatggatggatggatggatggat				1440		
ggttggactatgtatggatggatggatggatggatggatggatggatggatggatggatggat				1500		
atgattcagg ttcttcttgc aagtgtatggatggatggatggatggatggatggatggat				1560		
cgatgtatggatggatggatggatggatggatggatggatggatggatggatggatggatggat				1620		
gtctatgtatggatggatggatggatggatggatggatggatggatggatggatggatggat				1680		
gtcgatgtatggatggatggatggatggatggatggatggatggatggatggatggatggat				1740		
atggatccatc agtcaggaaa gaaaatctgtatggatggatggatggatggatggatggat				1800		
attgtatggatggatggatggatggatggatggatggatggatggatggatggatggatggat				1860		
ggttggatggatggatggatggatggatggatggatggatggatggatggatggatggatggat				1920		
gctatgtatggatggatggatggatggatggatggatggatggatggatggatggatggat				1980		
tggccaaatgtatggatggatggatggatggatggatggatggatggatggatggatggat				2040		
gctgatgtatggatggatggatggatggatggatggatggatggatggatggatggatggat				2100		
atcgaagagg ggcgcgtcac taaaggatgtatggatggatggatggatggatggatggat				2160		
atgaagtggatggatggatggatggatggatggatggatggatggatggatggatggatggat				2220		
aatggatggatggatggatggatggatggatggatggatggatggatggatggatggatggat				2280		
attagttcgatgtatggatggatggatggatggatggatggatggatggatggatggatggat				2340		

-continued

tctgttaccg	aagatattct	tacgggttt	aagatgcatt	ctcatgggt	gagatctgtt	2400
tattgtacac	caaagttagc	ggcttccaaa	ggatcagtc	caatacatt	ttcgccatgt	2460
ctccatcaag	ttcttcgtat	ggcgggttgg	tcgggttgg	ttttttttgg	taggcatttg	2520
cttatttgtt	atggttatgg	agggtgggtt	aaatggcttg	agegggttc	ctatattaac	2580
tctgtggttt	accctggac	ctctcttacc	ctcatcgtt	actgttctct	ccctggccatc	2640
tgtttcttcata	ctggaaaatt	categttccc	gagatttagca	actatgcgag	tatcccttcc	2700
atggcgctet	tctcgtcgat	tcgaaaaatc	ggtaataacg	ggtatttctcg	agatgcgaat	2760
gggatcgtat	atggttggaa	aaacgaaacat	ttttttttgg	ttttgggggt	ttttggggat	2820
ctgtttgtctc	tcttccaagg	tctcttcaag	gttcttgcgt	gtgtcgacac	taacttcaca	2880
gtcacatcaa	aaggcagctg	tgatggagag	tttttttttttt	tttttttttttt	caaatggat	2940
tcaacttcata	tccctcccaat	gactctactt	atcataaaacg	tcatttttttt	catatgggtt	3000
gtctctgtat	ccatcagcaa	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	3060
tttgcacttt	gggtcatcat	tcatctttac	ccgtttcccta	aaagggttgc	tttttttttttt	3120
gatagaatgc	caaccattat	tgtcgctcgg	tccatcttcc	ttggccctcgat	tcttacactt	3180
ctttgggttc	gggttaatcc	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	3240
tttagactgt	tgtgtt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	3255

```
SEQ ID NO: 16          moltype = AA  length = 1084
FEATURE              Location/Qualifiers
source               1..1084
                     mol_type = protein
                     organism = synthetic constru
```

SEQUENCE : 16	MNTGGRLIAQ SHNRNEFVLLI NADENARIRS VQELSGQTCQ ICRDEIELTV DGEFPVACNE	60
CAFPVCPRCY	EYERREGNQA CPQCKTRFKR LKGSPRVEGD EEEDDIDDLD NEFEYGNNGNI	120
GFDQVSSEGMS	ISRRNSGFPQ SDLSDAPPS QIPILLTYGDE DVEISSDRHA LIVPPSLGHH	180
GNRVHPVSLS	DPTVAAHPRP MVPQKDLAVY GYGSVAWKDR MEEWKRKQNE KLQVRHEDG	240
PDFEDGDDAD	FPMMDEGRQP LSRKIPKISS KINPYRMLIV LRLVLIGLFY HYRLHPVKD	300
AYALWLISVCE	CEIWFAVWSW LDQFPKWYPI ERETYLDRLS RLYKEGKPS GLSPVDVFWS	360
TVDPVLEBPPPL	ITANTVLSL AVDYPVDKWA CYVSSDGAAM LTFLAESETA EFARKWVPC	420
KKYCIEPRAP	EWYFCHKMDY LKNKVHPAFV RERRAMKRDY EEFVKVINAL VATAQKVPED	480
GWTMODOGTWPG	PGNNSVRDHGP MIQVFLGSDC VRDVENNELP RLVYVSREKRG PGDFDHKKAG	540
AMNSLLIRVSG	VLSNAPYLVD CTDCHYINNS KALCREAMCFM MDPQSGKKIC YVQFPQRFDG	600
IDRHNNRYSNRN	NNVVFDINMK GLDGLQGPPI VGTGCVFRRO ALYGFDAKPX KQPKRKTNC	660
WPWKCLLCFG	SRKNRKAHTV AADKKKKNRE ASKQIHALEN IEEGRTVKGS NVEQSTEAMQ	720
MKLEKKPGQS	PVFSVASARME NGGMARNASN ACLLKEAIQV ISCGYEDKTE WGKEIGWIYG	780
SVTEDILTGFT	KMHSHGWRSV YCTPKLAFK GSAPINLSDR LHQVLRWALG SVEIFLRRHC	840
PIWYGGGL	KWLERLSYIN SVVYPTWLSL LIVYCSLPAI CLTGTGFIIVP EISNYASILF	900
MALEFSSIAIT	GILEMQWGKV GIDDWWRNEQ FWVIGGSAAH LFALFQGLLK VLAGVDTNTFT	960
VTSKAADDGE	FSDLYLFWKT SLLIPPMTLI IINVIGIVG VSDAISNGYD SWGPFGLRF	1020
FALWVIIHLHY	PFLKGLLGKQ DRMPTIIVVV SILLASILTL LWVRVNPFVA KGGPILEICG	1080
DLCL		1084

```
SEQ ID NO: 17          moltype = DNA    length = 3255
FEATURE                Location/Qualifiers
source                 1..3255
mol_type = other DNA
organism = synthetic construct
```

SEQUENCE: 17

atgaacacccg	gtggcgggtt	aatcgccggt	tctcacaaaca	ggaatggagtt	tgtccctcatt	60
aatggccatgg	agaatggcccg	ataaataatca	gtccaaagagc	tgatggccaa	gacatgtcgg	120
atctgcacagg	atagatcgatca	attgactgttc	gatggagaac	cgttggggcc	atgtacaaacgg	180
tgtgcattcc	ctgtgtgttag	accttgcata	gagtagacaaa	gacgagaagg	caatcaagct	240
tgtccacagt	gaaaaaccgg	tttcaaacgt	cttaaaggaa	gtccaaagagt	tgaagggtat	300
gaagaggaggaa	atgacatgttga	tgatgttgcac	aatggatgttgc	tgatgttggaa	taatggggat	360
ggatggatgttgc	agggttctgttgc	aggtatgttgc	atctctcgcc	gcaactccgg	ttttccacaaa	420
tctgttgcgttgc	attcaatgttgc	accttgcgttgc	cagatccat	tgtgtacttgc	cgccgacacgg	480
gacgttggata	tttcttctgttgc	tagacatgttgc	cttattgttgc	ctccctcaact	ttgggtgtcat	540
ggcacatagatgttgc	ttcacatctgttgc	tttcttctgttgc	gccccggccgt	ttggccgtcaca	tccaaggccgt	600
atggtatccatgttgc	agaaagatgttgc	tttcttctgttgc	ttttatggaa	ttgttcgttgc	gaaagatcggt	660
atggggatgttgc	ggaaagagaaaa	tttcttctgttgc	ttttatggaa	ttgttcgttgc	ttttatggaa	720
cctgttgcgttgc	ttttatggaa	tttcttctgttgc	ttttatggaa	ttgttcgttgc	ttttatggaa	780
tttgtcttaggttgc	ttttatggaa	tttcttctgttgc	ttttatggaa	ttgttcgttgc	ttttatggaa	840
tttgtcttaggttgc	ttttatggaa	tttcttctgttgc	ttttatggaa	ttgttcgttgc	ttttatggaa	900
tttgtcttaggttgc	ttttatggaa	tttcttctgttgc	ttttatggaa	ttgttcgttgc	ttttatggaa	960
tttgtcttaggttgc	ttttatggaa	tttcttctgttgc	ttttatggaa	ttgttcgttgc	ttttatggaa	1020
tttgtcttaggttgc	ttttatggaa	tttcttctgttgc	ttttatggaa	ttgttcgttgc	ttttatggaa	1080
tttgtcttaggttgc	ttttatggaa	tttcttctgttgc	ttttatggaa	ttgttcgttgc	ttttatggaa	1140
tttgtcttaggttgc	ttttatggaa	tttcttctgttgc	ttttatggaa	ttgttcgttgc	ttttatggaa	1200
tttgtcttaggttgc	ttttatggaa	tttcttctgttgc	ttttatggaa	ttgttcgttgc	ttttatggaa	1260
tttgtcttaggttgc	ttttatggaa	tttcttctgttgc	ttttatggaa	ttgttcgttgc	ttttatggaa	1320
tttgtcttaggttgc	ttttatggaa	tttcttctgttgc	ttttatggaa	ttgttcgttgc	ttttatggaa	1380
tttgtcttaggttgc	ttttatggaa	tttcttctgttgc	ttttatggaa	ttgttcgttgc	ttttatggaa	1440
tttgtcttaggttgc	ttttatggaa	tttcttctgttgc	ttttatggaa	ttgttcgttgc	ttttatggaa	1500
tttgtcttaggttgc	ttttatggaa	tttcttctgttgc	ttttatggaa	ttgttcgttgc	ttttatggaa	1560
tttgtcttaggttgc	ttttatggaa	tttcttctgttgc	ttttatggaa	ttgttcgttgc	ttttatggaa	1620

-continued

gtatgaatt	ccctgataacg	agtctctggg	gttctatcaa	atgctcccta	ccttcgtaat	1680
gtcgattgt	atcaactacat	caacaatage	aaagcttta	gagaagcaat	gttttcatcg	1740
atggatcctc	agtcaggaaaa	aaaaatctgt	tatgttccatg	tccctccaag	gttcatggg	1800
atcgataggc	acgcgcata	ctcaaaatcc	aatgttgtgt	tctttgatata	caataatgaaa	1860
ggtttggatg	ggctacaagg	gcctatatac	gtcggtacag	gttgggtttt	caggaggcaa	1920
gcccgttacg	gatttgatgc	accgaaaggag	aagaagggcc	cacgttacaaag	atcgaaatgc	1980
tggccaaaat	gggtgtctct	atgttttggt	tcaagaaaaa	atcgtaaagc	aaagacagtg	2040
gtgcggatata	agaagaaaggaa	gaataggggaa	ggctcaaaagc	atgcattacgc	attagaaaaat	2100
atcgaagagg	gcccgcgtcac	taaagggtct	aacgttagaac	agtcaaccga	ggcaatgc当地	2160
atgaaggatgg	agaagaaatgg	ttggccatgt	cctgtatattg	ttgcatactgc	gctgtatggag	2220
aatggggggg	tggctagaaaa	cgaaaggccc	gttgggttgc	ttaaaagac	cattcaacgt	2280
attatgtcgc	gatataaaga	taaaactgaa	tggggaaaaag	agatgggtgt	gatctatgtt	2340
tctgttaccc	aaaatattct	tacggtttt	aaagatgcatt	ctcatgggtt	gagatctgtt	2400
tattgtacac	caaaggtagc	ggcttccaaa	ggatcagctc	caataatctt	ttccgatctgt	2460
ctccatcaag	ttcttcgtatg	ggcggttggg	tcgggttggaa	tttttttgag	taggcattgt	2520
ccttattgtt	atgggtatgg	agggtgggtt	aaatggcttg	agcgggttgc	ctataatcaa	2580
tctgtggttt	accctggac	ctcttaccc	ctcatcggtt	actgttctct	ccctgcacatc	2640
tgtcttctca	ctggaaaatt	catcgcccc	gagattatgc	actatgcgtt	tatctcttcgt	2700
atggcgctct	tctctgtatg	tgcataatacg	ggtattttctcg	agatgcattgt	gggcaaaatgt	2760
gggatcgtatg	atgggtggag	aaacacaaacg	tttttgggtca	ttggggatgt	ttctcgatcat	2820
ctgtttgtctc	tcttccaaagg	tctccatcaag	gttcttgctg	gtgtcgacac	taacttcacaa	2880
gtcacatcatca	aaggcgtatg	tgatggagag	ttctctggacc	tttactctt	caatggact	2940
tcacttctca	tccctccat	gactctactatc	atcataaaacg	tcattggatgt	catatcggtt	3000
gtctctgtatg	ccatcgccaa	ttggatacac	tcgttggggac	cgcttttcgg	aaagactgttcc	3060
tttgcacttt	gggtcatcat	tcatctttac	ccgttcccta	aagggttgcgt	tggggaaacaa	3120
gatagaatgc	caaccattat	tgtctgttgc	tccatctcc	ttggcttcgtat	tcttacatct	3180
ctttgggttc	gggttaatcc	gttgggtggcg	aaaggcggtc	ctatcttcga	gatctgtgtt	3240
tttagactgt	tgtga					3255

```
SEQ ID NO: 18          moltype = AA  length = 1084
FEATURE                Location/Qualifiers
source                 1..1084
                      mol_type = protein
                      organism = synthetic construct
```

SEQ ID NO: 19 moltype = DNA length = 3255
FEATURE Location/Qualifiers
source 1..3255
mol_type = other DNA
organism = synthetic construct

organism = synthetic construct

SEQUENCE:	19					
atgaacacgg	gtggtcgggt	aatcgccggt	tctcacaaca	ggaatgagggt	tgtccatt	60
aatgccatg	agaatgccc	aataagatca	gtccaaagagc	tgagtggaca	gacatgtcaa	120
atctgcagatg	atgagatcga	attgactgtt	gatggagaaac	cgttttgtgc	atgtaaacaa	180
tgtgcatttc	ctgtgttagt	accttgcata	gatgtacgaaa	gacgagaaagg	caatacaact	240
tgtcacagt	gcggaaaa	tttccaaacgt	ctttaaaaggaa	gtcacaaggat	tgaagggttat	300
gaagaggaaag	atgacatttg	tgatttttagac	aatgaggat	atgtatggaaa	taatgggatt	360
ggatgttgc	agggtttcgt	aggtatgtca	atctcttcgtc	gcaactccgg	ttttccaccaa	420
tctgttgg	attcagctcc	acctggcctt	cagattccat	tgtctgttccat	ccggccgacgac	480
gacgttggaga	ttttttctgt	tagacatgtt	cttattttgtt	ctcttcact	tggtgttcat	540
ggcaatagag	tttcatccgt	tttttttttt	gaccggaccc	tggctgcaca	tccaaaggcct	600
atggtacctc	agaaaagatct	tgccggttat	ggttatggaa	gtgtcgctt	gaaagatccgg	660
atggggat	ggaaagagaaa	gcagaatgag	aaaccttcagg	ttgtttagcca	tgaaggagat	720
cctgttgg	aaatgtgttgc	tgatgtctgt	tttccaaatgt	tggatgggg	aaggccggcc	780
ttgttgcgg	agatataat	caaatcgacg	aaatataatcc	cttacccggat	gttattttgt	840
ctacgttcttgc	ttgttcttttgc	tctttttttt	cactacggta	ttttttccccc	cgtccaaatgt	900

-continued

```

gcatatgtt ttttgttata tgtgagatat ggtttgcgt ttcatgggtt 960
cttgatcgt tccataatg gtacccatc gagcggaaa cgtacttga ccgactctca 1020
ttaagatgt agaaaagg gaaaccgtcg ggactatccc ctgtggatgt atttggtagt 1080
acagtggatc cattgaaaaga gcctccgtt attactgcaa atactgttt gtctatttt 1140
gctgttggatt atccgtcga taagggttgc ttacgtatc ctgtatgtt tgctgtatgt 1200
cttactttcg aagcttttc tgagaccgtc gaatttgcataa ggaaatgggt tccttctgc 1260
aagaatattt gtattggcc tctgttccg gaatgggtt ctgcataaa aatggactac 1320
tttgaagaata aagttatcc cgcatttttggggagccg gaggcatgaa gagagattat 1380
gaagaattca aagtaaagat caatgttta gtagcaacg cacagaaatg gcctgaggat 1440
ggttggacta tgcaacacgg tacacccgtt cccggtaataa gtgtggagaa tcatctggc 1500
atgattcagg ttttccgtt aagtggatgtt gttgtgtatc tggaaaacaa cgagttgcet 1560
cgattgtttt acgtttctcg tgagaagaga cccggattttt atcaccataa gaaggctgg 1620
gctatgaattt ccctgtatcc agtctctggg gttcttatca atgtcttca ctttctgaat 1680
gtcgatgttgc atactacat caacaatago aaagcttca gagaagcaat gtgtttcatg 1740
atggatccgc atcaggaaaaa gaaaatgttgc ttttgcgttccatcaaaatgatgggg 1800
attgtatggc acgatcgata ctcaaatcg aatgttgcgtt ttttgcgttcaatgttgc 1860
ggtttggatg ggctacaagg gcctatatac gtcggtaacg gttgtttt caggaggca 1920
gcgccttacg gatttgcgttcc accgaagaagg aagaaggccg caccgtaaagc atgcaatgtc 1980
tggccaaatgttgcgttccat atgtttccat tcaagaaaggc atcgtaaaggc aaagacactg 2040
gctgcggatc aagaagaaaga aatggggaa gctgtcaacgc agatccacgc attagaaat 2100
atcgaagagg gccgcgtcac taaagggttca aacgttagaaac gtcacccgcg ggcacatgca 2160
atgaagtgg aagaagaaatggggatgtt cctgttgcgtt ttgcacatgcg cgcgtatggag 2220
atattgtggg tggcttagaa cgcacggccg gttgttgcgttcc ttaaagaagc catccaatgtc 2280
attagttgcg gatatggaa taaaactgaa tggggaaaaggc agatgggttgc gatctatgtt 2340
tctgttaccgc aagatattct tacgggtttt aagatgcattt ctatgttgcgtt gatctgtt 2400
tattgttacac caaaggtagc ggctttcaaa ggtacgtacgc caatcaatct ttcggatcgt 2460
ctcatacaag ttttgcgttggg ggcgttgcgtt ttttgcgttgcgtt gatctgttgcgtt 2520
cttattttgtt atgtttatggg aggtgggttgcg aatgggttgcg cttatgttgcgttccat 2580
tctgttgcgtt acccggtggac ctcttgcgtt ctatgttgcgtt actgttctt ccctgcgttccat 2640
tgcgttgcgttccat atcgatgttccat gatgttgcgttccat atcgatgttccat 2700
atgatgttgcgttccat atcgatgttccat atcgatgttccat atcgatgttccat 2760
gggatgtatgttgcgttccat atcgatgttccat atcgatgttccat atcgatgttccat 2820
ctgttgcgttccat atcgatgttccat atcgatgttccat atcgatgttccat atcgatgttccat 2880
gtcacatcaaa aagcagctga tgatggagag ttctctgacc ttacatctt ccaatggact 2940
tcaacttcaaa tccctcaatc gacttctactt atcataaaacg tcattggactt cattatgttccat 3000
gtctctgttgcgttccat atcgatgttccat atcgatgttccat atcgatgttccat 3060
tttgcgttccat atcgatgttccat atcgatgttccat atcgatgttccat atcgatgttccat 3120
gatagaatgc caaccattat tgcgttgcgttccatccatc tggcctcgat tcttacactt 3180
cttgggttccat atcgatgttccat atcgatgttccat atcgatgttccat atcgatgttccat 3240
tttagactgtgtga 3255

```

```

SEQ ID NO: 20      moltype = AA length = 1084
FEATURE          Location/Qualifiers
source           1..1084
mol_type = protein
organism = synthetic construct

```

```

SEQUENCE: 20
MNTGGRLIAG SHNRNEEFVLI NADENARIRS VQELSGQTQ ICRDEIELTV DGEPPVACNE 60
CAPPVCRPCY EYERREGNQA CPQCKTRFKR LKGSPPRVEGD EEEDDIDLD NEFEYGNNGI 120
GFDQVSEGMIS RRRNSGFPQ SDLDSAPPGS QIPLLTYGD DVEISSDRHA LIVPPSLGGH 180
GNRVPHVPSL DPTVAAPHRP MVPQKDLAVY GYGSVAWKR MEEWKRKQDNE KLQVVRHEGD 240
PDFEDGDDAD PFMMDREGQP LSRKIPMLIV KINPYRMLIV LRLVILGLFF HYRILHPVKD 300
AYALWLISVI CEIWFAVSWL LDQFPKWKPI ERETYLDRLLS RLYEKEKGKPS GLSPDVVFVS 360
TVDPLKEPPL ITANTVLSIL AVDYPVDKVA CYVSDDGAAM LTFEALSETA EFARKWVPFC 420
KKYCIEPRAP EWYFCHKMDV LKNKWHVPAFW RERRAMKRDY EEFKVVKINAL VATAQKVPED 480
GWTMQDGTWP PGNSVRDHPG MIQVFLGSDG VRDVENNELLP RLVYVSRERK PGFDHHKKAG 540
AMNSLIRVSG VLSNAPYLLN VDCDHYINNS KALREAMCFM MDPQSGKKIC YVQFPQRFDG 600
IDRHDRYSNR NVVFFDINMK GLDGLQGPIY VGTGCVFRQQ ALYGFDAPKK KKGPRKTCNC 660
WPWKWCLLCFG SRKNRRAKTV AADKKKKRNRE ASKQIHALEN IEEGRVTKGS NVEQSTEAMQ 720
MKLEKKFQGS PVFVASARME NGGMARNASP ACLLKEAIQV ISCGYEDKTE WGKEIGWIYG 780
SVTFEDILITGF KMHSHGWRSV YCTPKLAAFK GSAPINLSDR LHQVLRWALG SVEIFLSRHC 840
PIWYGYGGGL KWLERLRSYIN SVVYPWTSLP LIVYCSLPAI CLLTGKFIIP EISNYASILF 900
MALEFSSTAIT GILEMQWGKV GIDDWWRNKQ FWVIGGVSAH LFALFQQLK VLAVGDTNFT 960
VTSKAADDGE FSDLYLFWKWT SLLIPPMTLL IINVIGVIVG VSDAISNGYD SWGPLFGRFLF 1020
FALWVIIHLH PFLKGLLGKQ DRMPTIIVVV SILLASILTL LWVRVNPFVA KGGPILEICG 1080
LDCL 1084

```

```

SEQ ID NO: 21      moltype = DNA length = 3255
FEATURE          Location/Qualifiers
source           1..3255
mol_type = other DNA
organism = synthetic construct

```

```

SEQUENCE: 21
atgaacaccg gtgggtcggtt aatcgccgtt tctcacaaca ggaatggatgttgcgtt 60
atgcccgtt agaatggccg aataagatca gtccaaaggc tgagtgccatc gacatgttgcgtt 120
atctcgatgttgcgttccat atcgatgttccat atcgatgttccat atcgatgttccat 180

```

-continued

tgtgcattcc	ctgtgttag	accttgcata	gagtaacgaa	gacgagaagg	caatcaagct	240
tgtccacagt	gaaaaacccgt	ttcaaaacgt	cttaaaggaa	gtccaagagt	tgaaggtgat	300
gaagaggaaag	atgacatgt	tgatggat	aatgagttt	atgtggaaa	taatggatt	360
ggatttgc	aggtttctg	aggtatgtca	atctctcg	gcaactccgg	tttcccacaa	420
tctgatttgg	attcagctcc	acctggctct	cagattccat	tgctgactt	cggcgacgag	480
gacgttgaga	tttcttctg	tagacatgt	cttattgtt	ctccctca	tggttgctcat	540
ggcaatagag	ttcatctgt	ttcttctt	gaccgaccc	tggtgcaca	tccaaaggcc	600
atggatcc	agaaaatgt	tcgggtttat	ggttatggaa	gtgtcgott	gaaagatcg	660
atggagaaat	ggaagagaaa	gcagaatgag	aaacttcagg	ttgttaggca	tgaaggagat	720
cctgattttg	aatgtgtt	tgatgtctat	tttccaatg	tggatgaggg	aaggcagcca	780
ttgtcttaga	agataacaaat	caaatcgac	aagataaatc	tttacccggat	gttaatttgt	840
ctacgttctg	tgatttctgg	tcttctt	cactaccgt	tttccaccc	cgtcaaagat	900
gcatacgctt	tgtggcttat	ttctgttata	tgtgagat	ggtttgcgt	ttcatgggtt	960
cttgcatacg	tccctaaatg	gtccctat	gacgcgaaa	cgtaacttgg	ccgactctca	1020
ttaagatgt	agaaaagaagg	gaaaccgtcg	ggactatccc	ctgtggatgt	atttggatgt	1080
acagtggatc	cattaaagaa	gcctccgtt	attactgc	atactgtctt	gtcttatttt	1140
gctgttggatt	atccgtcg	taagggtgt	tgttacgt	ctgtatgtt	tgctgtctatg	1200
cttactttcg	aaagtcttc	tgagaccgt	gaattcgaa	ggaaatgggt	tccttctgc	1260
aagaaatatt	tgattggcc	tcgtgtcc	gaatggat	tctgcataaa	aatggactac	1320
ttgaagaata	aaagtcatcc	cgcatttgg	agggagccgc	gagccatgaa	gagagattat	1380
gaagaattca	aaagtaaagat	caatgctta	gtagcaacag	cacagaaagt	gcctgaggat	1440
ggtttggacta	tgcaagacgg	tacaccttgg	cccggttaata	gtgtgcgaga	tcatcttggc	1500
atgattcagg	tcttcttgg	aagtgcgtt	tttcgtatgt	teggaaaacaa	cgaggcttgc	1560
cgatttttcg	aaagtcccg	tgagaagaga	cccggttgc	atcaccat	aaaggcttgc	1620
gctatgaatt	ccctgtatcg	agtccttgg	tttctatcaa	atgtcttca	ccttgcataat	1680
gtcgatttgc	atactacat	caacaatgc	aaagcttca	gagaagcaat	gtgtttcatg	1740
atggatctc	agtcaaggaaa	gaaaatctgt	tatgttca	tccctcaaaag	tttcgtatggg	1800
attgtatggc	acgatccgata	ctcaaatcg	aatgttgcgt	tctttgat	caatatgaaa	1860
ggtttggat	ggctacaagg	gcctatatac	gtcgatcac	gttgcgtt	caggaggca	1920
gcccgttacg	gatttgcgt	accgaagaag	aagaaggccc	cacgtaaagac	atgcaatgtc	1980
tggccaaat	ggtgttcc	atgttttgcgt	tcaagaaaggaa	atcgtaaaagc	aaagacatgt	2040
gtcgcggata	agaagaagaa	gaatggggaa	ggctcaaaac	agatccacgc	attagaaaaat	2100
atcgaagagg	gcccgtc	taaagggtt	aaactgttgc	agtcaaccga	ggcaatgca	2160
atgaagtgg	agaagaaatt	tgggcagt	cctgtatgtt	ttgcatctgc	gcgtatggag	2220
aatgggtggg	ttggtagaaa	cgcaacccgg	gcttgcgtc	ttaaagaagc	catccaatgc	2280
attagttgcg	gatatgaga	taaaactgaa	tggggaaaag	agatgggttgc	gatctatgtt	2340
tctgttaccg	aaagatattct	tacgggttt	aagatgcatt	ctcatgttgc	gagatctgtt	2400
tattgtacac	caaaggtagc	ggcttcaaa	ggatcagtc	caatcaatct	ttcggatcgt	2460
ctccatcaag	tttcttgcgt	ggcgttggg	tcgggttgc	ttttcttgc	taggcattgt	2520
ccttattttgt	atgtttatgg	aggttgcgtt	aatgggttgc	agcgggttgc	ctacattaaac	2580
tctgttgcgtt	accctgttgc	ctcttacgg	ctatcgat	actgttctt	ccctgcacat	2640
tgtcttctca	ctggaaatt	catcgcccc	gagatttgc	actatgcgag	tatcttcttc	2700
atggcgctc	tctcgat	tgcataatgc	gggttctc	agatgcatttgc	gggcaaaagt	2760
gggatcgat	atggggat	aaacaaacg	tttgcgttgc	tttgcgttgc	tttgcgttgc	2820
ctgtttgcgt	tcttcaagg	ttcttc	tttgcgttgc	tttgcgttgc	tttgcgttgc	2880
gtcacatcaa	aaagcagctg	tgatggag	ttctctgc	tttacatc	caaatggact	2940
tcacttctca	tcctccat	gacttctact	atcataaaacg	tcattttgcgtt	catatgcgga	3000
gtctctgtat	tcatcgacaa	ttggatacg	tcgtggggac	tcgtttccg	aaagacttttc	3060
tttgcattt	gggtcatcat	tcatcttac	ccgttccat	aaagtttgc	ttggaaacaa	3120
gatagaatgc	caaccattat	tgtcgat	ttccatcttcc	ttccatcttcc	ttccatcttcc	3180
cttgggtcc	gggttaatcc	ttttgttgcgt	aaaggcggtc	ctatttcg	ttccatcttcc	3240
tttagactgt	tgtga					3255

SEQ ID NO: 22 moltype = AA length = 1084

FEATURE Location/Qualifiers

source 1..1084

mol_type = protein

organism = synthetic construct

SEQUENCE: 22

MNTGGRLIAG	SHNRNEFVLI	NADENARIRS	VQELSGQTQ	ICRDEIELTV	DGEPFVACNE	60
CAPPVCRPCY	EYERREGNQA	CPQCKTRFKR	LKGSPRVERGD	EEEDDIDLD	NEFEYGNNGI	120
GFDQVSEGMIS	ISRRNSGFPQ	SDLDSAPPGS	QIPLLTYGD	DVEISSDRHA	LIVPPSLGGH	180
GNRVHPVSLSDPTVAHPRP	MVPQKDLAVY	GYGSVAWKRDR	MEEWKRKQNE	KLQVVRHEGD	240	
PDFEDGDDA	FPMMDEGRQP	LSRKIPIKSS	KINPYRMLIV	LRLVILGLFF	HYRILHPVKD	300
AYALWLISVI	CEIWFAVSVW	LDQFKWYPI	ERETYLDRILS	LYREKEGPSP	GLSPVDFVFS	360
TVDPPLKEPPL	ITANTVLSIL	AVDYPVDKVA	CYVSDDGAAM	LTFEALSETA	EFARKWPVFC	420
KKYCIERPRAP	EWYFCHKMDY	LKNKVHPAFV	RERRAMKRDY	EEFKVKINAL	VATAQKVPED	480
GWTMQDGT	PGNSVRDH	MIQVFLGS	VRDVENN	RLVYVSR	PGFDHHKKAG	540
AMNSLIRVSG	VLSNAPYLLN	VDCDHYINNS	KALREAMCFM	MDPQSGKKIC	YVQFQRFDG	600
IDRHDYRSNR	NVVFFDINMK	GLDGLQGPYI	VDTGCVFRQ	ALYGFDAPKK	KKGPRKTCNC	660
WPKWCLLCFG	SRKNRKA	AADKKKKNRE	ASKQIHALEN	IEEGRTVKGS	NVEQSTEAMQ	720
MKLEKKFGQS	PVFVASARME	NGGMARNASP	ACLLKEAQV	ISCGYEDKTE	WGKEIGWIY	780
SVTFEDILTG	KMHSHGWRSV	YCTPKLA	GSAPINLSDR	LHQVLRWALG	SVEIPLSRHC	840
PIWYGYGGGL	KWLERLSYIN	SVVYPWTS	LIVYCSLP	CLLTGKFIVP	EISNYASILF	900
MALFSSIAIT	GILEMQWGKV	GIDDWWRNEQ	FWWIGGVSAH	LFALFQGLLK	VLAGVDTNFT	960
VTSKAADDGE	PSDLYLFKWT	SLLIPPM	TLL IINVIGIVG	VSDAISNGYD	SWGPLFGR	1020

-continued

FALWVIIHLY PFLKGGLGKQ DRMPTIIVWW SILLASILTL LWVRVNPVVA KGGPILEICG 1080
LDCL 1084

SEQ ID NO: 23 moltype = DNA length = 3255
FEATURE Location/Qualifiers
source 1..3255
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 23
atgaacacccg gtggcggtt aatgcgcgtt tctcacaaca ggaatgagtt tgccctatt 60
aatgcgcgtg aagaatggcccg aataagatca gtccaagagc tgagtggaca gacatgtcaa 120
atgtcgacag atgagatcgat attgactgtt gatggagaac cgtttgccc atgttaacgaa 180
tgtgcattcc ctgtgtgttag accttgcata gactacgaaa gacgagaagg caatcaagct 240
tgtccacagt gcaaaacccg ttccaaacgt cttaaaggaa gtccaaaggt tgaagggtat 300
gaagaggaaag atgacattga tgatttagac aatgagttt agtatggaaa taatggatt 360
ggatttgcgatc aggtttctgtt aggtatgtca atcttcgtc gcaactccgg tttcccccaa 420
tctgatttgg attcagctcc acctggctct cagattccat tgctgactt cggcgacag 480
gacgttggata ttcttcgtt tagacatgtt ctattgttgc ctcccttact ttgtggctat 540
ggcaatagag ttcatctgtt ttcttcgtt gacccgaccg tggctgcaca tccaaaggct 600
atggatcc tcgatccca gaaagatct tgccgtttat ggttatggaa gtgtcgctt gaaagatccg 660
atggaggaaat ggaagagaaa gcagaatggaa aaacttcagg ttgttaggca tgaaggagat 720
ctcgattttg aagatgggtt tgatgtctt ttcccaatgtt tggtatgggg aaggcaggca 780
ttgtcttagga agataccat caaatcgatc aagataatc cttaccggat gttattttgt 840
ctacgttcttgg tttttttttt cttttttttt cactaccgtt ttcttcaccc cgtcaaaatg 900
gcatatgtttt tttttttttt ttctgttata tttttttttt gttttgtt gttttttttt 960
cttgcattttt tttttttttt ttctgttata tttttttttt gttttgtt gttttttttt 1020
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1080
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1140
gtttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1200
cttactttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1260
aagaaatattt ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1320
ttttaaaaaat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1380
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1440
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1500
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1560
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1620
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1680
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1740
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1800
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1860
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1920
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1980
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2040
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2100
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2160
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2220
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2280
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2340
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2400
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2460
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2520
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2580
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2640
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2700
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2760
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2820
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2880
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2940
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3000
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3060
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3120
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3180
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3240
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3255

SEQ ID NO: 24 moltype = AA length = 1084
FEATURE Location/Qualifiers
source 1..1084
mol_type = protein
organism = synthetic construct
SEQUENCE: 24
MNTGGRLIAG SHNRNEFVLI NADENARIRS VQELSGQTCQ ICRDEIELTV DGEPFVACNE 60
CAPVPCRCPY EYERREGNQA CPQCKTRFKR LKGSPRVERGD EEEEDIDDDNEFEYGNNGI 120
GFDQVSEGMS ISRRNSGFPQ SDLDSAPPGS QIPLLTYGDE DVEISSDRHA LIVPPSLGGH 180
GNRVHPVSLSDPTVAAHPRP MVPQKDLAVY GYGSVAWKDR MEEWKRKQNE KLQVVRHEGD 240
PDFPEDGDDAD FPMMDEGRQP LSRKIPIKSS KINPYRMLIV LRLVILGLFF HYRILHPVKD 300

-continued

AYALWLISVI	CEIWFVAWSWV	LDQFPKWYPI	ERETYLDRLS	LRYEKEGKPS	GLSPVDFVFS	360
TVDPPLKEPPL	ITANTVLSIL	AVDYPVDKVA	CYVSDDGAAM	LTFEALSETA	EFARKWVPFC	420
KKYCIEPRAP	EWYFCHKMDY	LKNVHPAFV	RERRAMKRDY	EEFKVKINAL	VATAQKVPED	480
GWTMQDGTWP	PGNSVRDHPG	MIQVFLGSIDG	VRDVENNELP	RLVYVSRKRR	PGFDHHHKAG	540
AMNSLIRVSG	VLSNAPYLLN	VDCDHYINNS	KALREAMCFM	MDPQSGKKIC	YVQFPQRFDG	600
IDRHDRYSNR	MNVFFDINMK	GLDGLQGPIY	VGTGCVFRQQ	ALYGFDAPKK	KKGPRTKCNC	660
WPKWCLLCFG	SRKNRKAHTV	AADKKKKNRE	ASKQIHALEN	IEEGRVTKGS	NVEQSTEAMQ	720
MKLEKKFQGS	PVFVASARME	NGGMARNASP	ACLLKEIAQV	ISCGYEDKTE	WGKEIGWIYS	780
SVTEDILTG	KMHSHGWRSV	YCTPKLAALK	GSAPINLSDR	LHQVLRWALG	SVEIFLSRHC	840
PIWYGYGGGL	KWLRLSYIN	SVVYWPSTSPL	LIVYCSLPAI	CLLTGKFIVP	EISNYASILF	900
MALFSSIAIT	GILEMQWKV	GIDDWRNEQ	FVWVIGVSAH	LFALFQQLLK	VLAGVDTNFT	960
VTSKAADDGE	FSDLYLFWKT	SLLIPPMTLL	IINVIGVIVG	VSDAISNGYD	SWGPLFGRLF	1020
FALWVIIHLY	PFLKGLLGKQ	DRMPTIIVVV	SILLASILTL	LWVRVNPFVA	KGGPILEICG	1080
LDCL						1084

SEQ_ID NO: 25	moltype = DNA	length = 3255
FEATURE	Location/Qualifiers	
source	1..3255	
	mol_type = other DNA	
	organism = synthetic construct	

SEQUENCE: 25						
atgaacacccg	gtggcggtt	aatccgcgtt	tctcacaaaca	ggaataggatt	tgtcctcatt	60
aatgccatcg	agaatccccg	aataagatca	gtccaaagacg	tgatgtggaca	gacatgtcaa	120
atctgcacatcg	atggatcgatcg	atctactgtt	gatggagaaac	cgttttgcc	atgttaacgaa	180
tgtgcattcc	ctgtgtgttag	accttgcata	gagtaacgaa	gacgagaagg	caatcaagct	240
tgtccacagt	gaaaaaccgg	tttcaaacgt	ctttaaaggaa	gtccaagagt	tgaagggtat	300
gaagagggaaag	atgacatcg	tgatgttagt	aatgaggat	tgatgtggaa	taatgggat	360
ggatgttgc	agggttctga	aggatgtca	atctctcgtc	gcacactccc	tttcccacaa	420
tctgattttg	attcagctcc	acctggctct	cagattccat	tgcgtactt	cggcgacag	480
gacgttggaga	tttcttctga	tagacatgt	cttattgttgc	ctcccttact	tgggtgtcat	540
ggcaatagag	ttcatactgt	ttctcttttgc	gaccggaccc	tgcgtgcaca	tccaaaggcc	600
atggatccatcg	atggatccatcg	ttctatgttgc	ggtttatgttgc	gtgcgttgc	gaaagatccg	660
atggaggaaat	ggaagagaaa	gcagaatgag	aaacttcagg	ttgttaggca	tgaaggagat	720
cctgattttg	aaatgttgta	tgatgtctgt	tttccaaatga	ttgtatgggg	aaggcagcca	780
ttgtcttagga	agataccat	caaatacg	aaatgttgttgc	tttccatgttgc	tttttgtat	840
ctacgttcttgc	tttcttttgc	tctcttttgc	cactacgttgc	ttcttcaccc	cgtcaaaat	900
gcataatgtt	tgtggatatt	ttctgttata	tgtgagatat	tttttgtgttgc	tttcatgttgc	960
cttgatcgtt	tcccttaatgc	gtaccctatc	gagcgagaaa	cgtacttgc	ccgactctca	1020
ttaaagatgtt	agaaaaaagg	gaaacgcgt	ggactatccc	ctgtggatgt	atttttgtat	1080
acatgttgc	cattgttgc	gccttcgttgc	attactgttgc	atactgttgc	gtcttattttc	1140
gtctgttgc	atccgttgc	taagggttgc	tgttacgtat	ctgtatgttgc	ttgtgtatgt	1200
cttacttttcg	aaatgttttc	tgagaccgt	gaattcgca	ggaaatgggt	ttcttttgc	1260
aaatgttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	1320
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	1380
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	1440
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	1500
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	1560
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	1620
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	1680
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	1740
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	1800
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	1860
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	1920
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	1980
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	2040
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	2100
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	2160
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	2220
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	2280
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	2340
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	2400
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	2460
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	2520
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	2580
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	2640
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	2700
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	2760
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	2820
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	2880
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	2940
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	3000
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	3060
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	3120
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	3180
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	3240
tttgcataatcg	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	tttttgttgc	3255

-continued

```

SEQ ID NO: 26          moltype = AA  length = 1084
FEATURE
source
1..1084
mol_type = protein
organism = synthetic construct
SEQUENCE: 26
MTTGGRLIAG SHNRNEFVLI NADENARIRS VQELSGQTCA ICRDEIELTV DGEPPVACNE 60
CAFVPCRKY EYERREGNQA CPQCKTRFKR LKGSPRVEGD EEEDDIDDLD NEFEYGNNGI 120
GFDQVSECGMS ISRRNSGFPQ SDLDSAPPGS QIPLLTYGDE DVEISSLRHA LIVPPSLGGH 180
GNRVHPVSL S DPTVAAPRP MVPKDLAGVY GYGSAWKR MEEWKRKQNE KLQVVRHEGD 240
PFDGEDGDDA FPMMDEGRQP LSRKIPIKSS KINPYRMLIV LRLVILGLFL HYRILHPVKD 300
AYALWLISVI CEIWFAVSWV LDQFPKWYPI ERETYLDRLS LRYEKEGKPS GLSPVDFVFS 360
TVDPKKEPPL ITANTVLSIL AVDYPVDKVA CYVSDDGAAM LTFEALSETA EFARKWVPC 420
KKYCIEPRAP EWYFCHKMDY LKNKVFPAFV RERRAMKRDY EEFKVKINAL VATAQKVPED 480
GWTMMDGTPW PGNSVRDHPG MIQVFLGSDG VRDVENNELP RLVYVSREREK PGFDHHKKAG 540
AMNSLIRVSG VLSNAPYLLN VDCDHYINNS KALREAMCFM MDPQSGKKIC YVQFPQRFDG 600
IDRHDRYSNR NVVFFDINMG GLDGLQGPYI VGTGCVFRQQ ALYGFADPKK KKGRPKTCNC 660
WPWKCLLCFG SRKNRAKTM AADKCKKRNRE ASKQIHALEN IEEGRVTKGS NVEQSTEAMQ 720
MKLEKKFCQG P VPFVASARMV NGGMARNASP ISCGYEDKTE WGKEIGWIYG 780
SVTEDILTGK KMHSHGWRSV YCTPKLAALK ESAPINLSDR LHQVLRWALG SVEIFLSRHC 840
PIWYGYGGGL KWLERLSYIN SVVYWPWTSLP LIVYCSLPAI CLLTGKFIVP EISNYASILF 900
MALFSSIAIT GILEMQWGKV GIDDWRRNEQ FWVIGGVSA LFALFQGLLK VLAGVDTNFT 960
VTSKAADDGE PSDLYLFKWT SLLIPMPLL IINVIGVIVG VSDAISNGYD SWGPLFGRFL 1020
FALWVIIHLH PFLKGGLGKQ DRMPTIIVVV SILLASILTL LWVRVNPVVA KGGPILEICG 1080
LDCL                                         1084

SEQ ID NO: 27          moltype = DNA  length = 3255
FEATURE
source
1..3255
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 27
atgaacacccg gtggtcgggtt aatgcgcgtt tctcacaaca ggaatgaggt tgccttcatt 60
aatgcgcgtt agaatgccccg aataagatca gtccaagagc tgagtggaca gacatgtcaa 120
atctgcagatc agatagatca attgactttt gatggagaca cgttggcgc atgtaacgaa 180
tgtgcattcc ctgtgtgttag accttgcata gtagtacaaa gacgagaagg caatcaagct 240
tgtccacagt gaaaaaccgg tttcaaacgt cttaaaggaa gtccaaaggt tgaagggtat 300
gaagaggaaatg atgacatgtc tgatgttaga aatgagtttgg agtatggaaa taatgggat 360
ggatgttgc acgttttgcg aggtatgtca atctctcgctc gcaactccgg tttcccacaa 420
tctgatgtttt attcagctcc acctggctctc caatggccat tgcgtactt cggcgacag 480
gacgttggaa tttcttctgtc tagacatgtc ttatgttcc tccttcact tgggttcat 540
ggcaatagatc ttcatcttcgtt ttccatccat gacccgaccc tggctgcaca tccaaaggcc 600
atggtaacctc agaaagatct tcgcgttattt ggttatggaa gtgtcgcttg gaaagatccg 660
atggaggaaat ggaagagaaa cgcaaatggaa aaacttcagg ttgttaggca tgaaggagat 720
cctgatgtttt aagatgggtc tgatgtctat tttccatgtc tggatgggg aaggcagcca 780
ttgtcttaga agatacaatc caaatcgacg aaatggatctt ctatccggat gttaatttgc 840
ctacgttcttgc tgatgttttgc ttttttttcaacttgcgtt ttccatccc cgtcaaaatg 900
gcataatgtt ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 960
cttgatcaatgc tcccttaatgc gtaccctatc gagcgagaaa cgtacttggc ccgactctca 1020
ttaaatgtt agaaaggaaagg gaaaccgtcg ggactatccc ctgtggatgt attttttatgt 1080
acatggatc ctttgcgttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 1140
gtctgttattt atccgttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 1200
cttacttttcg aagcttttc tgagaccgtt gaaatcgca gaaatgggt tcctttctgc 1260
aaagaaatattt gtatttgcgtt tcgtgtctcc gaatggatatt ttttttttgc ttttttttgc 1320
tttgcgttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 1380
gaagaaatca aatggatgttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 1440
ggttggacta tgcaagacgg tacacccgtt cccggataa gtgtgcgaga tcatacttcgtt 1500
atgattcagg ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 1560
cgatttagttt acgttttgcgtt tgagaagaga cccggatattt atccatcaa gaaggcttgc 1620
gtctatgtttt ccctgtatcc acgttttgcgtt ttttttttgc ttttttttgc ttttttttgc 1680
gtctgttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 1740
atggatcttc agtcaggaaa gaaaatctgt tatgttcaatgc ttttttttgc ttttttttgc 1800
attgtatggc acgtatgttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 1860
ggtttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 1920
ggcgttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 1980
tggccaaat ggtgttccat ttttttttgc ttttttttgc ttttttttgc ttttttttgc 2040
gctgcggata agaagaagaa gaataggggaa gcgtcaaaatgc agatccacgc attagaaat 2100
atcgaagagg ggcgcgttgc taaagtttgc aatgttgcgtt ttttttttgc ttttttttgc 2160
atgaagtggg aagaaatattt tggttgcgtt ctttttttgc ttttttttgc ttttttttgc 2220
aattggggaa tggctcaagg gcctatataatgc ttttttttgc ttttttttgc ttttttttgc 2280
attagttgcg gatgtatggaa taaaactgtttt gtttttttgc ttttttttgc ttttttttgc 2340
tctgttaccg aagatatttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 2400
tattgttacac caaagtttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 2460
ctccatcaag ttcttcgttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 2520
ccttatttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc ttttttttgc 2580

```

-continued

tctgtggtt	acccgtggac	ctctctaccg	ctcatcgtt	actgttctct	ccctgccatc	2640
tgtcttcga	ctggaaaatt	categttccc	gagatttagca	actatgcggag	tatcctcttc	2700
atggcgctct	tctcgatc	tgcataaacg	ggtattctcg	agatgcataatg	gggcaaaagt	2760
gggatcgatg	atgggtggag	aaacgaaacag	ttttgggtca	ttgaagggtgt	ttctcgccat	2820
ctgtttgtct	tcttccaagg	tctctcaag	gttcttgctg	gtgtcgacac	taacttcaca	2880
gtcacatcaa	aaggcactga	tgtatggagag	tctctgacc	tttaccttt	caaatggact	2940
tcacttctca	tccctccat	gactctactc	atcataaacg	tcatgtgggt	catatgcgga	3000
gttctgtatg	ccatcagcaa	tggatacga	tcgtggggac	cgctttccgg	aagactgtt	3060
tttgcacttt	gggtcatcat	tcatcttac	cegttccta	aagggttgc	tggaaacaa	3120
gatagaatgc	caaccattat	tgtctgtctgg	tcctatctcc	tggcctcgat	tcttacactt	3180
cttgggtatcc	gggttaatcc	gtttgtggcg	aaaggcggtc	ctatttcga	gatctgttgt	3240
tttagactgt	tgtga					3255

SEQ ID NO: 28 moltype = AA length = 1084
 FEATURE Location/Qualifiers
 source 1..1084
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 28
 MTGGRRIAG SHNRNEFVLI NADENARIRS VQELSGQTQ ICRDEIELTV DGEPPVACNE 60
 CAFPVCRPCY EYERREGNQA CPQCKTRFKR LKGSPRVEGD EEEIDDLD NEFEYGNNGI 120
 GFQVSEGMS ISRRNSGFPQ SDLDSAPPGS QIPLLTYGDE DVEISSLRHA LIVPPSLGGH 180
 GNRVHPVSL S DPTVAAHPRP MVPQKDLAVY GYGSVAWKDR MEEWKRKQNE KLQVVRHEGD 240
 PDPEGDGDDAD FPMMDEGRQP LSRKIPIKSS KINPYRMLIV LRLVILGLFF HYRILHPVKD 300
 AYALWLISVI CEIWFVAWSW LDQFPKWYPI ERETYLDRLS LRYEKEGKPS GLSPVDFVFS 360
 TVDPLKEPPL ITANTVLSIL AVDYPVDKVA CYVSDDGAAM LTFEALSETA EFARKWVPFC 420
 KKCYCIEPAF EWYFCHKMDV LKNKVPAP RERRAMKRDY EEFKVKNAL VATAQKVPED 480
 GWFMQDGTWP PGNSVRDHDP MIQVFLGSDG VRDVENNELP RLVYVSRERK PGFDHHKKAG 540
 AMNSLIRVSG VLSNAPYLLN VDCDHYINNS KALREAMCFM MDPQSGKKIC YVQFPQRFDG 600
 IDRHDYRSNR NVVFFDINMK GLDGLQGPYI VGTGCVFRQQ ALYGFADPKK KKGPRKTCNC 660
 WPKWCLLCFG SRKNRKAFTV AADKKKKRN ASKQIHALEN IEEGRVTKGS NVEQSTEAMQ 720
 MKLEKKFGQS PVFVASARM NGGMARNASP ACILKKEAIQV ISCGYEDKTE WGKEIGWIYG 780
 SVTEDILTGK KMHSHGWRSV YCTPKLAAPK GSAPINLSDR LHQVLRWALG SVEIFLSRHC 840
 PIWYGYGGGL KWLERLSYIN SVVYWPWTSLP LIVYCSLPAI CLLTGKPIVP EISNYASILF 900
 MALFSSIAIT GILEMQWGVK GIDDWWRNEAQ FWVIEGVSAH LFALFQGLLK VLAVGDTNFT 960
 VTSKAADDGE PSDLYLFKWT SLLIPMTLL IINVIGIVVG VSDAISNGYD SWGPLFGRLF 1020
 FALWVIIHLH PFLKGLLGKQ DRMPTIIVVV SILLASILTL LWVRVNPFVA KGGPILEICG 1080
 LDCL 1084

SEQ ID NO: 29 moltype = DNA length = 3255
 FEATURE Location/Qualifiers
 source 1..3255
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 29
 atgaacaccg gtggctggtt aatcgccgt ttcacacaaca ggaatggatt tgccttcatt 60
 aatcccgatg agaatccccg aataagatca gtccaaagac tgatgtggaca gacatgtcaa 120
 atctgcagag attagatcgaa attgactgtt gatggagaaac cggttggc atgtaaacgaa 180
 tttgtcattcc ctgtgtgtat accttgcata ggttacgaaa gacgagaagg caatcaagct 240
 tttgtccacagt gcaaaaacctg tttcaaaacgt cttaaaggaa gtccaaaggt tgaagggtat 300
 gaaaggaggaaat atgacattga tgatgttaga aatgatggttt agtatggaaa taatggatt 360
 ggatgttgcgat aggtttctga aggtatgtca atctctcgctc gcaactccgg ttccacaaa 420
 tttgtatggg attcaagtc accttgcgtt ccattttcatc tgcttgcattt cggcgacgag 480
 gacgttggata tttcttctga tagacatgtt cttattgttc ctcccttact tgggtgtcat 540
 ggcataatagat ttcatctgtt ttctttttt gacccgaccc tggctgcaca tccaaaggct 600
 atggtaatctc agaaagatct tcgggtttat ggttatggaa gtgtcgctt gaaagatcg 660
 atggaggaaat ggaagagaaa gcaaatatgaa aaaaatccgg ttgttaggca tgaaggagat 720
 cttgtatggg attcaagtc accttgcgtt ccattttcatc tgcttgcattt cggcgacgag 780
 ttgtcttagga agataatcaa caaatcgacg aagataaaatc ttatccggat gttatgttg 840
 ctacgttcttgc tgatgtttgg tttttttttt cactatcgta ttcttcaccc cgtcaaaatg 900
 gcatatgtt gttgtgttat ttctgttata tgatgtat ggttgcgtt ttcatgggtt 960
 cttgtatgtt tcccttaatg gtatccatc gacgttgcata cgtacttggaa ccgactctca 1020
 ttaagatatg agaaagaagg gaaaccgtcg ggactatccc ctgtggatgtt atttgttagt 1080
 acagtggatc cttttttttt gtcctccgtt attactgtca aatctgtttt gtcttattttt 1140
 gctgttgcattt atccctgtcga taagggtgtt tgatgtatcgtt tgctgtatgtt 1200
 cttaatccgtt aagcttttc tgatgtatcgtt gatgtatcgtt ttcttcaccc cgtcaaaatg 1260
 aagaaatattt gtatgtatcgtt tgatgtatcgtt gatgtatcgtt ttcttcaccc cgtcaaaatg 1320
 ttgtatgtt aatgttgcata cgcattttgtt agggagccgc gagccatgaa gagagatatt 1380
 gaagaatgttca aatgtatcgtt gatgtatcgtt gatgtatcgtt ttcttcaccc cgtcaaaatg 1440
 ggttggacta tgcaagacgg tacacccgtt cccggtaataa gtgtcgacgg tcatccggc 1500
 atgttgcattt tttttttttt aatgtatcgtt gatgtatcgtt ttcttcaccc cgtcaaaatg 1560
 cgatgtatgtt acgtttctcg tgatgtatcgtt gatgtatcgtt ttcttcaccc cgtcaaaatg 1620
 gctatgtatgtt aatgtatcgtt gatgtatcgtt ttcttcaccc cgtcaaaatg 1680
 gtcgttgcattt atccctgtcga aaacatgttca gatgtatcgtt ttcttcaccc cgtcaaaatg 1740
 atggatccctc atgtatgttca gatgtatcgtt ttcttcaccc cgtcaaaatg 1800
 atgtatgttca aatgtatcgtt gatgtatcgtt ttcttcaccc cgtcaaaatg 1860

-continued

ggtttggatg	ggctacaagg	gcctataac	gtcggtacag	gttgttgc	caggaggca	1920
ggccttacg	gatttgatc	accgaagaag	aagaaggccc	cacgtaaagac	atgcaattgc	1980
tggccaaat	ggtgtctct	atgtttgtt	tcaagaaa	atcgtaaagc	aaagacagtg	2040
gctgcggata	agaagaagaa	gaataggaa	gctgtcaaagc	agatccacgc	attagaaaaat	2100
atcgaaagagg	gcccgtcac	taaagggtct	aacgtagaac	agtcaaccga	ggcaatgca	2160
atgaagttgg	agaagaaattt	tggcoagtct	cctgtatttgc	ttgcatactgc	gcgtatggag	2220
aatgggtggg	ttggctagaaa	cgcaagcccc	gcttgcgtc	ttaaaaagaa	catccaagtc	2280
attagtgcg	gatatgaaga	taaaactgaa	tggggaaaag	agattgggtg	gatctatgg	2340
tctgttaccg	aatatattct	tacgggttt	aatatgcatt	ctcatgggt	gagatctgtt	2400
tatgttacaa	caaaatgtac	ggcttcaaa	ggatcagctc	caatcaact	ttcgatcg	2460
ctccatcaag	ttcttcgtat	ggcgttgggg	tcgggtgaga	tttttttgag	taggatgtt	2520
cttattttgt	atggatatgg	agggtgggtt	aaatggcttgc	tcacattaaac	ctacattaaac	2580
tctgtggttt	accctgtggac	ctctctaccg	ctcatcgttt	actgttctt	ccctgcacatc	2640
tgtcttctca	ctggaaaaat	categttcc	gagatttagca	actatgcgg	tatctcttcc	2700
atggcgcet	tctcgatcgat	tgcataaacg	ggtattctcg	agatgcata	ggccaaagtt	2760
gggatcgtat	atggatggag	aaacgaaacag	ttttgggtca	tttgggggtt	ttctcgcat	2820
ctgtttgtc	tcttccaaagg	tctctcaag	gttcttgcgt	gtgtcgacac	taaattcaca	2880
gtcacatcaa	aacgagctat	tgtatggag	ttctctgacc	tttacccctt	caaattggact	2940
tcacttctca	tcccttccat	gactctactc	atcataaacg	tcattggagt	catatgcgga	3000
gtctctgtat	ccatcagcaa	tgcatacgc	tcgtggggac	cgcttccgg	aagactgtt	3060
tttgactt	gggtcatcat	tcatcttac	ccgttcccta	aaggttgt	tggggaaacaa	3120
gatagaatgc	caaccattat	tgtctgtcg	tccatccctc	tggcctcgat	tcttacactt	3180
cttgggttcc	gggttaatcc	gtttgtgggg	aaaggcggtt	ctattctcg	gatctgtgt	3240
tttagactgt	tgtga					3255

SEQ ID NO: 30 moltype = AA length = 1084
 FEATURE Location/Qualifiers
 source 1..1084
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 30
 MNIGGRILAG SHNRNEFVLI NADENARIRS VQELSGQTCQ ICRDEIELTV DGEPPVACNE 60
 CAFPVCRPCY EYERREGNQA CPQCKTRFKR LKGSPRVEGD EEEIDDIDLD NEFEYGNNGI 120
 GFDQVSEMGMS ISRRNSGFPQ SDLDSAPPGS QIPLLTYGDE DVEISSLRHA LIVPPSLGGH 180
 GNRVHPVSLSD PVTVAAHPRP MVPQKDLAVY GYGSAWKDR MEEWKRKQNE KLQVVRHEGD 240
 PDPEGDGDDAD FPMMDEGRQP LSRKIPIKSS KINPYRMLIV LRLVIFGLFF HYRILHPVKD 300
 AYALWLISVI CEIWFVAWSWV LDQFPKWKYPI ERETYLDRLS LRYEKEGKPS GLSPDVDFVS 360
 TVDPLKEEPL ITANTVLSIL AVDYPVDKVA CYVSDDGAAM LTFEALSETA EFARKWVPFC 420
 KKCYIEPRP EWYFCHKMDY RERRAMKRDY EEFVKVINAL VATAQKVPED 480
 GWTMMDGTPW PGNSVRDHPG MIQVFLGSDG VRDVENNELP RLVYVSRERK PGFDHHKKAG 540
 AMNSLIRVSG VLSNAPYLLN VDCDHYINNS KALREAMCFM MDPQSGKKIC YVQFPQRFDG 600
 IDRHDYRSNR NVVFFDINMK GLDGLQGPYI VGTGCVFRQQ ALYGFDAPKK KKGPRKTCNC 660
 WPKWCLLCFG SRKNRRAKTV AADKKKKRNE ASKOIHALEN IEEGRVTKGS NVEQSTEAMQ 720
 MKLEKKFGQS PVFVASARMV NGGMARNASP ACLLKEAIQV ISCGYEDKTE WGKEIGWIYG 780
 SVTEDILITGF KMHSHGWRSV YCTPKLAAFK GSAPINLSDR LHQVLRWALG SVEIFLRSRHC 840
 PIWGYGGGL KWLRLRSYIN SVVYPWTSLP LIVYCSLPAI CLLTGKFIVP EISNYASILF 900
 MALFSSIAIT GILEMQWGKV GIDDDWRNBQ FWIWGGSAH LFALFQGLLK VLAVGDTNFT 960
 VTSKAADDGE PSDLILFKWT SLLIPPMTLL IINVIGIVG VSDAISNGYD SWGPLFGRLF 1020
 FALWVIIHLY PFLKGLLGKQ DRMPTIIVVV SILLASILTL LWVRVNPFLVA KGGPILEICG 1080
 LDCL 1084

SEQ ID NO: 31 moltype = DNA length = 3255
 FEATURE Location/Qualifiers
 source 1..3255
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 31
 atgaacacgg gtggtegggtt aatcgccgt ttcacacaaca ggaatggatt tgccttcatt 60
 atatgcgtatc agaatggccc aataatgcata gtcacaaatc tgactggaca gacatgtcaa 120
 atatgcgtatc agatggatcgatc attcgactttt gatggggaaac cggttgcgc atgtatggaa 180
 tttgtcattcc ctgtgtgtat accttgcata ggttgcataa gacggaaaatggaaatc 240
 tttgtccacatg gcaaaaatcccg tttcaaacgt cttaaaggaa gtttgcataa gtttgcataa 300
 gtttgcataa gtttgcataa gtttgcataa gtttgcataa gtttgcataa gtttgcataa 360
 gtttgcataa gtttgcataa gtttgcataa gtttgcataa gtttgcataa gtttgcataa 420
 tttgttcattcc attcgttcc accttgcata cttttttttt tttttttttt tttttttttt 480
 gtttgcataa gtttgcataa gtttgcataa gtttgcataa gtttgcataa gtttgcataa 540
 gtttgcataa gtttgcataa gtttgcataa gtttgcataa gtttgcataa gtttgcataa 600
 atatgcgtatc agaaatgcata tgcgttttat gtttgcataa gtttgcataa gtttgcataa 660
 atatgcgtatc agaaatgcata tgcgttttat gtttgcataa gtttgcataa gtttgcataa 720
 cttgtatggg attcgttcc accttgcata cttttttttt tttttttttt tttttttttt 780
 ttgttctatggg agatggatcgatc ttgtatggg tttttttttt tttttttttt tttttttttt 840
 ctatcgatcgatc ttgtatggg tttttttttt tttttttttt tttttttttt tttttttttt 900
 gtttgcataa gtttgcataa gtttgcataa gtttgcataa gtttgcataa gtttgcataa 960
 gtttgcataa gtttgcataa gtttgcataa gtttgcataa gtttgcataa gtttgcataa 1020
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1080
 acatgtatc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1140

-continued

gctgttgcatt	atccctgtcga	taagggttgc	tgttacgtat	ctgatgtatgg	tgctgtctatg	1200
cttacttcg	aagtttttc	tgaggaccgt	gaatttcgaa	ggaaatgggt	tcctttctgc	1260
aagaataattt	gtatttgc	tcgtgtctcc	gaatggtat	tctgcataaa	aatggactac	1320
ttagaagaata	aagtcatcc	cgcattttt	agggagccgc	gagccatgaa	gagagattt	1380
gaagaattca	aagtaaagat	caatgttta	gtagcaacag	cacagaaatg	gcctgaggat	1440
ggtgttgcata	tgcaagacgg	tacaccttgg	ccccgttaat	gtgtcgagaa	tcatcttgc	1500
atgatttcagg	tcttcttgg	aagtgcgggt	gttcgtatgt	tgcggaaacaa	cgagttgcct	1560
cgatttagttt	acgtttctcg	tgagaagaga	ccccggattt	atcacatcaa	gaaggcttgg	1620
gctatgaattt	ccctgtatcg	agtctctggg	gttctatcaa	atgtcttca	ccttctgaat	1680
gtcgatgttg	atcacatcg	caacaatagc	aaagcttca	gagaagcaat	gtgttctatg	1740
atggatccctt	agtcaggaaa	gaaaatccat	tatgttcaat	tccctcaaa	gttcgtatgg	1800
atttgataggc	acgatcgata	ctcaaatcg	aatgttgc	tctttgatat	caatatgaaa	1860
ggtttggatg	ggctacaagg	gcctatatac	gtcggtacag	gttgcgtttt	caggaggca	1920
gcgccttacg	gatttgcgtc	accgaagaagg	aagaaggccc	cacgtaaagac	atgcaatgtc	1980
tggccaaat	ggtgttctct	atgttttgc	tcaagaaaggaa	atcgtaaaggc	aaagacagt	2040
gctgcggataa	agaagaagaa	gaataggaa	gcgtcaaaac	agatccacgc	attaaaaat	2100
atcgaagagg	gcgcgtcac	taaagggtct	aacgttagaac	agtcaaccga	ggcaatgca	2160
atgaagtgg	agaagaattt	tggggactgt	cctgtatttt	ttgcatctgc	gcgtatggag	2220
aatgggttgg	tgggttggaa	cgcaagccc	gcttgcgtc	ttaaagaaggc	catccaatgc	2280
attagtttgcg	gatatgaaa	taaaactgaa	tggggaaaag	agatgggtt	gatctatgtt	2340
tctgttaccg	agatatttct	tacgggtttt	aagatgcatt	ctcatgttgc	gagatcttgc	2400
tattgttacac	caaaggtagc	gggttcaaa	ggatcagctc	caatcaatct	ttcggatcgt	2460
ctccatcaag	tcttcgtat	ggcgttgggg	tcgggttgc	ttttcttgc	tagcatgt	2520
cctattttgtt	atgggtatgg	agggtgggtt	aatatggcttgc	agcgggttgc	ctacattaaac	2580
tctgttgcgtt	accctgtggac	ctcttacgg	ctcatcgttt	actgttcttct	ccctgcacatc	2640
tgtcttctca	ctggaaaattt	categttcc	gagattagca	actatgcgag	tatcttcttc	2700
atggcgctct	tcttcgtat	tgcaatcgat	gttatttctcg	agatgcataatg	gggcaaaatgt	2760
gggatcgtat	atgggttgg	aaacgaacag	ttttgggtca	tttgggggtt	ttctgcgcac	2820
ctgtttgc	tcttcgaagg	tcttcgtat	gttcttgc	gtgtcgacac	taatctcaca	2880
gtcacatcaa	aagcagctga	tgatggagag	tctcttgc	tttacccctt	caaatggact	2940
tcacttctca	tcccttccat	gactcttactc	atcataaacg	tcatttgc	catatgcgga	3000
gttcttgcgtat	ccatcgaa	tggatacgac	tcgttggggac	cgcttttccg	aaagcttgc	3060
tttgcactttt	gggtcatcat	tcatcttac	ccgttccat	aagggttgc	tggaaacaa	3120
gatagaatgc	caaccattat	tgtcgcttgc	tccatccctc	tggcctcgat	tcttacactt	3180
cttgggttcc	gggttaatcc	gttgggttgc	aaaggcggtc	ctatttcg	gatctgttgc	3240
tttagactgt	tgtga					3255

SEQ ID NO: 32	moltype = AA	length = 1084
FEATURE	Location/Qualifiers	
source	1..1084	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 32		
MNTGGRLIAG SHNRNEFVLI NADENARIRS VQELSGQTCQ ICRDEIELTV DGEFPVACNE	60	
CAPPVCRPCY EYERREGNQA CPQCKTRFKR LKGSPRVEGD EEEIDDIDLD NEFEYGNNGI	120	
GFDQVSEGMIS ISRRNSGFPQ SDLDSAPPGS QIPLLTYGDE DVEISSDRHA LIVPPSLGGH	180	
GNRVHPVSLSDPTVAAHPRP MVPQKDLAVY GYGSVAWKDR MEEWKRKQNE KLQVVRHEGD	240	
PDPEDGDDAD PFMMDGRQP LSRKIPRIKINPYRMLIV LRLVILGLFF HYRILHPVKD	300	
AYALWLISVI CEIWFAVSVW LDQFPKWYPI ERBTYLDRLS LRYEKEGKPS GLSPDVVFV	360	
TVDPPLKEPPL ITANTVLSIL AVDYPVDKVA CYVSDDGAAM LTFEALSETA EFARKWV	420	
KKYCIEPRAP EWYFCHKMDY LKNKVHPAFV RERRAMKRDY EEFKVVKINAL VATAQKVPED	480	
GWTMQDGTWP PGNSVRDHPG MIQVFLGSDG VRDVENNELP RLVYVSREKR PGFDHHKKAG	540	
AMNSLIRVSG VLSNAPYLLN VDCDHYINNS KALRQAMCFM MDPQSGKKIC YVQFPQRFDG	600	
IDRHDRYSNR NVVFFDINMK GLDGLQGPIY VGTGCVFRQQ ALYGFDAPKK KKGPRKTCNC	660	
WPKWCLLCFG SRKNRKAKTV AADKKKKNRE ASKQIHALEN IEEGRVTKGS NVEQSTEAMQ	720	
MKLEKKFGQS PVFVASARME NGGMARNASP ACLLKEAIQV ISCGYEDKTE WGKEIGWIYG	780	
SVTEPDLITGF KMHSHGWRSV YCTPKLAAPK GSAPINLSDR LHQVLRWFAG SVEIFLRSRC	840	
PIWYGYGGGL KWLERLSYIN SVVYPWTSLP LIVYCSLPAI CLLTGKFI	900	
MALFSSIAIT GILEMOWGKV GIDDDWWRNEQ FWVIGGVSAH LFALFOGLLK VLAVGDTNF	960	
VTSKAADDGE PSDLYLFKWT SLLIPPMTLL IIINVIGIVVG VSDAISNGYD SWGPLFGR	1020	
FALWVIIHLY PFLKGLLGKQ DRMPTIIVVV SILLASILTL LWVRVNPVFA KGGPILEICG	1080	
LDCL		1084

SEQ ID NO: 33	moltype = DNA	length = 3255
FEATURE	Location/Qualifiers	
source	1..3255	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 33		
atgaacacccg gtggtcgggtt aatcgccgt ttcacacaaca ggaatggatgt tgccttcatt	60	
aatgccatcg aagaatccccg aataaagatca gtccaaagac tgagttggaca gacatgtcaa	120	
atctgcacatcg atgatgtca attgtactgtt gatggagaac cgtttgcgc atgtacaa	180	
tgtgcattcc ctgtgtgttag accttgcata gatgtacaa gacgagaagg caatcaagct	240	
tgtccacatcg gcaaaaacccg ttccaaacgt cttaaaggaa gtccaaagatgt tgaagggtat	300	
gaagaggaag atgacattga tgatgttgc aatgtatggatgtt agatggaaa taatggatt	360	
ggattttgcgtatcgatgtca aggtatgtca atctctcgatc gcaactccgg tttccacaa	420	

-continued

tctgattttgg	attcagctcc	acctggctct	cagattccat	tgctgactta	cgggacgag	480
gacgtttagaa	tttccttctga	tagacatgt	cttattttgc	ctcttcact	tggttgtcat	540
ggcaataatgg	ttcatactgt	ttctctttct	gaccgacccg	tggtgcaca	tccaaggct	600
atggtaacctc	agaagatct	tgccgtttat	ggttatggaa	tggtgccttgc	aaaagatccg	660
atggaggaaat	ggaagagaaaa	gcagaatgt	aaacttcagg	ttgtttaggc	tgaaggatgg	720
cctgattttg	aagatggta	tgatgctgt	tttccaatga	tggatgaggg	aaggcagcc	780
ttgtcttagaa	agataccaaat	caaatcgac	aagataaaatc	cttacccgttgc	gttaatttgc	840
ctacgtcttgc	tgatttcttgc	tctcttttct	caactcgta	ttcttcaccc	cgtcaaaatgt	900
gcatatgtt	tgtggcttat	ttctgttata	tgtgagat	ggtttgcgttgc	tcatgggtt	960
cttgatcagt	tccctaaatgt	gtaccctatc	gagcggaaaa	cgtacttgg	ccgactctca	1020
ttaagatgtt	agaaaqaagg	gaaacccgtcg	ggactatcccc	ctgtggatgtt	attingttatgt	1080
acagttggatc	cattgaaaga	gcctccgtt	attactgc	atactgttctt	gtcttatttt	1140
gtgttgtgtt	atcttcgtca	taagggttgttgc	tggtacgtat	ctgtatgttgc	tgctgtctatgt	1200
cttactttcg	aagcttttc	tgagaccgt	gaattcgttgc	ggaaatgggtt	tcctttctgc	1260
aagaaatattt	gtatttgcgg	tcgtgtccccc	gaatgttatt	ttctgcatttt	aatggactac	1320
ttagaaagat	aagtccatcc	cgccattttgtt	aggggacgggc	gagcataatgg	gagaggat	1380
gaagaatccaa	aagtaaataat	caatgttca	tgtagcaac	cacagaaatgt	gcctggatgtt	1440
ggttggactt	tgcaagacgg	tacacccgttgc	cccggttata	tggtgcggat	tcatcttcgg	1500
atgatttcagg	tcttccttgc	aagtgcgttgc	gttcgtatgt	tcgaaacaaat	cgagggtgc	1560
cgtatgttgc	acgttttcgt	tgagaagaga	cccggttgc	atccatttttt	gaaggcttgc	1620
gtctatgttgc	ccctgtatc	agtcgttgc	ttcttcatca	atgtcccttgc	cctttgttgc	1680
gtcgattttgt	atcactacat	caacaatag	aaagcttta	gagaagcaat	gtgtttcatgt	1740
atggatcctc	agtccaggaaa	gaaaatctgt	tatgtttcgt	tctcttcaatgt	gttgcgttgc	1800
attgtatggc	acgtatcgata	ctccaaatcg	aatgttgcgt	tctttgtat	caatatgtaaa	1860
ggttttggat	ggctatcaagg	gccttatatac	tcgtgtatgc	gttgcgttgc	caggaggat	1920
gcgttttacgc	gatttgcgt	accgttgc	aaagggcc	cgtatgttgc	atgtcaatgc	1980
tggccaaat	ggtgttctct	atgttttgc	tcaagaaaa	atcgtaaagg	aaagacatgt	2040
gtgtcgatgtt	agaagaagaa	gaataggggaa	gcgttcaatgc	agatccacgc	attaaaaatgt	2100
atcgaagagg	ggccgtgttac	taaagggttct	aacgttgc	agtcaacccg	ggcaatgc	2160
taaagtttgc	agaagaaatgtt	tgggcaatgt	cctgtatgttgc	tgatgttgc	gtgtatgttgc	2220
aatgggtggaa	tggctagaaaa	cgcaaggcc	gcttgcgttgc	ttaaaaaggc	catccatgtc	2280
attagttgc	gatgttgc	taaaaactgaa	tggggaaaatgt	agatgggttgc	gatctatgttgc	2340
tctgttaccc	aaagatatttt	tcacgggtttt	aaagatgttgc	ctctatgttgc	tgatgttgc	2400
tattgtatcc	caaagttagc	ggcttttcaaa	ggatgttgc	caatataatc	ttccgtatgttgc	2460
ctccatcaat	ttcttcgtatc	ggcgcttggaa	tcgggttgc	tttttttgc	ttggccatgttgc	2520
ccttattttgtt	atggttatgg	aggtgggttgc	aaatgggttgc	agccgggttgc	ctatattttatc	2580
tctgttgc	accctgtggac	ctctcttaccc	ctcatcggttgc	actgttgc	ccctggccat	2640
tgttttctca	ctggaaaatt	catgttcc	gagatttagca	actatgtc	tatcttcttc	2700
atggcgcttc	tcttcgtatc	tgcaataatcg	ggatatttcgt	agatgttgc	ggggaaatgt	2760
gggatcgatg	atgggtggag	aaacgttgc	ttttgggtca	ttggggatgt	ttctgtcgat	2820
ctgtttgttc	tcttccttgc	tctcttgc	gttcttcgttgc	gtgtgcacatgc	taatttcaca	2880
gtcataatcg	aaggatgttgc	tgatgttgc	ttcttcgttgc	ttatcttgc	caatgttgc	2940
tctacttcc	ccctccatgttgc	gacttctactc	atcataatcg	tcattttgc	catatgttgc	3000
gtctctgtat	ccatcgacaa	tggatagcc	tcgtggggatc	cgcttttcgg	aaagactgttgc	3060
tttgcacttgc	gggtcatatc	tcatctttac	ccgttccatc	aagggttgc	tggggaaacaa	3120
gatagaatgc	caaccatatttgc	tgtgttgc	ttcatcttgc	ttggccatgttgc	tcttactatc	3180
ctttgggttgc	gggttaatcc	gttgcgttgc	aaaggccgttgc	ctatatttgc	gatctgttgc	3240
tttagactgt	tgttgc					3255

SEQ ID NO: 34 moltype = AA length = 1084
FEATURE Location/Qualifiers
source 1..1084
mol_type = protein

SEQUENCE :	34	organism = synthetic construct		
MNTGGRLIAG	SHNRNEFVLI	NADENARIRS	VQELSGQTQ ICRDEIELTV DGEFPVACNE	60
CAFPVCRPCY	EYERREGNQA	CPQCKTRFKR	LKGSPRVEGD EEEDDIDDLD NEFEYGNNGI	120
GFDQVSSEGMS	ISRRNSGFPQ	SLDLSQAPPGS	QIPLLTYGDE DVEISSDRHA LIVPPSLGHH	180
GNVRHPVSLIS	DPTVAHPRP	MVPQDADLPA	GYGSVAKDR MEEWKRKQNE KLVQVRHEGD	240
PDFEDGDDAD	PFMMDEGRQP	LSRKPIKSS	KINPYRMLIV RLRLVILGLFF HYRLPHVKD	300
AYALWLISVI	CEIWFAVSWV	LDQFPKWPYPI	ERETYLDRLS LRYKEKGKPS GLSPVDVFVS	360
TVDPLKEEPL	ITANTWLSIL	AVDYPDVKVA	CYVSDGAAM LTFAESETA EFARKWVFC	420
KKYCIEPRAP	EWYFCHKMDY	LKNKVHPAFV	RERRAMKRDY EEFKVKINAL VATAQKVPED	480
GWTMGTDPW	PGNSVRDHPG	MIQVFLGSDQ	VRDVENNELP RLUVVSEREK PGPDHWHKKAG	540
AMMSLIRVSG	VLSNAPYLLN	VDCDHYINNS	KALREAMCFM MDPQSGKKIC YVQFSQRFDG	600
IDRHDRYSNR	NVVFFDINMK	GLDGLQGPI	VGTGCVFRRQ ALYGFADPKK KKGPRTKTC	660
WPWKCLLCFG	SRKNRKAATV	AADKKKKNNRE	ASKQIHALEN IEEGRVTKGS NVBQSTEAMQ	720
MKLEKEFFGQS	PVFVASARME	NGGMARNASP	ACLLKEAIQV ISCGYEDKTE WGKEIGWVY	780
SVTDELLTGF	KMHSHGWRSL	YCTPLKLAIF	GSAPINLSDR HLQVLRLWALG SWEIFLSRHC	840
PIIWGYGGGL	KWLRLRSYIN	SVVYPTWSL	LIVYCSLPAI CLLTGKFIVP EISNYASILF	900
MALFSSIAIT	GILEMQWGKV	GIDDWWRNQE	FWVIGGSAAH LFALFQGLLK VLAGVDTNFT	960
VTSKAADDGE	FSDLYLFWKT	SLLIPPMTLL	IINVIGVIVG VSDAISNGYD SWGPLFGRLF	1020
FALWVIIHLH	PFLKGLLGKQ	DRMPTIIVVW	SILLASILTL LWVRVNPFTA KGGPILEICG	1080
LDCL,				1084

SEQ ID NO: 35 moltype = DNA length = 3255

-continued

FEATURE	Location/Qualifiers
source	1..3255
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 35	
atgaacacccg gtggcggtt aatcgccggt tcttcacaaca ggaatgagg ttgcctcatt 60	
aatgccgatg agaatggccg aataagatca gtc当地 gacatgtcaa 120	
atctgcagatc atgatgtca attactgttt gatggagaac cgtttgcgc atgtacgaa 180	
tgtgcattcc ctgtgttag accttgcatt gatgtacaaa gacgagaagg caatcaagct 240	
tgtccacagt gcaaaacccg ttccaaacgt cttaaaggaa gtccaaaggt tgaagggtat 300	
gaagaggaaag atgatgtca tgatgttag aatgaggttt agtatggaaa taatggatt 360	
ggatgttgc atgggttctg aggtatgtca atcttcgtc gcaactccg tttccacaa 420	
tctgatttgg attcagctcc acctggctc cagattccat tgctgactt cggcgacgag 480	
gacgttggaa ttcttctgt tagacatgtt cttatgttgc ctcccttact tgggtgtcat 540	
ggcaatagag ttccatctgt ttcttctgt gaccggaccg tggctgcaca tccaaaggcct 600	
atgttaccc tc aaaaatgtc tgggtttat gggttgcata gttgtcgtt gaaatgtccg 660	
atggaggaaat ggaagagaaa gcagaatgag aaacttcagg ttgttaggca tgaaggagat 720	
cctgattttt aagatgtgt tgc当地 gggatccatgt tttccatgtt gggatgggg aaggcaggca 780	
ttgtcttagga agataccat caaatcgac agataaaatc cttaccggat gttatttttg 840	
ctacgttctg tgatcttgc ttcttctt cactaccgtt ctccatccc cgtcaaaatgt 900	
gcataatgtt tgc当地 tttttttat ttctgttata tgc当地 gatgttgcattt gttttgttgc ttcatgggtt 960	
cttgcatttgtt cc当地 tttttttat gggatccatgtt gggatccatgtt gggatccatgtt gggatccatgtt 1020	
ttaagatgtt agaaaaggaa gaaaccgtcg ggactatccc ctgtggatgtt atttttttgtt 1080	
acagtggatc cttccgtt cttccgtt attactgtca atactgttgc ttcttattttt 1140	
gtctgttgcattt atccgttgc taagggttgc ttcttgcattt gtttgcattt gtttgcattt 1200	
cttacttcg aagcttttc tggaccgtt gatgtccaa gggaaatgggt ttcccttgc 1260	
aagaatattt gtatttgcggc tggcttgc当地 gaatgttgcattt tttttttat aatggactac 1320	
tttgc当地 aagtttcc cggatgttgc当地 gggggccggc gagatgttgc当地 gagatgttgc当地 1380	
gaagaatttca aagtaaagat caatgttgc当地 gtatgttgc当地 gagatgttgc当地 gagatgttgc当地 1440	
gggttggacta tgcaagacgg tacacccgtt cccggtaataa gtgtgc当地 tcatcttgc当地 1500	
atgatttgcggg tttttttat aatgttgc当地 gtttgc当地 tggaaaacccg cggatgttgc当地 1560	
cgtatgttgc当地 aatgttgc当地 tggatgttgc当地 cccggatgttgc当地 atcaccataa gaaggcttgc当地 1620	
gtctatgttgc当地 ccttgc当地 atgttgc当地 atgttgc当地 cttttttat 1680	
gtctgatttgc当地 atcactatcat caacaatgttgc当地 aaagcttgc当地 gagaagcaat gtgttgc当地 1740	
atggatgttgc当地 atgatgttgc当地 gaaaatgttgc当地 tatgttgc当地 tcccttcaaaatgttgc当地 1800	
attgtatgttgc当地 acgatgttgc当地 ctcaatgttgc当地 aatgttgc当地 tttttttat caatgttgc当地 1860	
gggttggatgttgc当地 ggcttacaagg gccttatataatgttgc当地 gtctgttgc当地 gtttgc当地 cttttttat 1920	
gcgttgc当地 gtttgc当地 accgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 1980	
tggcccaatgttgc当地 gtttgc当地 atgttgc当地 tttttttat tttttttat 2040	
gtctgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 2100	
atcgaagagg ggc当地 gtc当地 tttttttat aatgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 2160	
atgaagtggg aaaaaatgttgc当地 tggatgttgc当地 tttttttat tttttttat 2220	
aatggatgttgc当地 tggatgttgc当地 tttttttat aatgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 2280	
attgttgc当地 gatgttgc当地 taaaatgttgc当地 tttttttat aatgttgc当地 aacgttgc当地 aacgttgc当地 2340	
tctgttgc当地 aagatgttgc当地 tttttttat aatgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 2400	
tattgttgc当地 caaaggatgttgc当地 ggttgc当地 aatgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 2460	
cttccatgttgc当地 ttccatgttgc当地 ggttgc当地 aatgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 2520	
cattttttgtt atgttgc当地 tttttttat aatgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 2580	
tctgttgc当地 accgttgc当地 ctccatgttgc当地 atgttgc当地 tttttttat 2640	
tgttgc当地 ctggaaatgttgc当地 catgttgc当地 gagatgttgc当地 aatgttgc当地 aacgttgc当地 aacgttgc当地 2700	
atggatgttgc当地 ttccatgttgc当地 ggttgc当地 aatgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 2760	
gggatgttgc当地 atgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 2820	
cttccatgttgc当地 ttccatgttgc当地 ggttgc当地 aatgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 2880	
gtcacatgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 2940	
tcacttgc当地 tcccttcaatgttgc当地 gacttgc当地 atcataatgttgc当地 tttttttat 3000	
gtctgttgc当地 ctccatgttgc当地 tttttttat aatgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 3060	
tttgc当地 ggttgc当地 ctccatgttgc当地 tttttttat aatgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 3120	
gatagaatgttgc当地 caaccatgttgc当地 tttttttat aatgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 3180	
cttgggttgc当地 ggttgc当地 tttttttat aatgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 aacgttgc当地 3240	
tttagactgttgc当地 tttttttat 3255	
SEQ ID NO: 36	moltype = AA length = 1084
FEATURE	Location/Qualifiers
source	1..1084
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 36	
MNIGGRILAG SHNRNEFVLI NADENARIRS VQELSGQTCQ ICRDEIELTV DGEPPVACNE 60	
CAFPVCRPCY EYERREGNQA CPQCKTRFKR LKGSPRVEGD EEEDDIDLD NEFEYGNNGI 120	
GFDQVSEGMS ISRRNSGFPQ SDLDSAPPGS QIPLLTYGDE DVEISSDRHA LIVPPSLGGH 180	
GNRVHPVSLIS DPTVAAHPRP MVPQKDLAVY GYGSVAWKRDR MEEWKRKQNE KLQVVRHEGD 240	
PDFEDGDDAD FPMMDEGRQP LSRKIPIKSS KINPYRMLIV LRLVILGLFF HYRILHPVKD 300	
AYALWLISVI CEIWFVAWSW LDQFPKWKYPI ERETYLDRLS LRYEKEKPS GLSPVDFVFS 360	
TVDPPLKEPPL ITANTVLSIL AVDYPVDKVA CYVSDDGAAAT LTFEALSETA EFARKWVPFC 420	
KKYCIEPRAP EWYFCHKMDY LKNKVHPAFV RERRAMKRKY EEFKVKINAL VATAQKVPED 480	
GWQMJDGTWP PGNSVRDHPG MIQVFLGSDG VRDVENNELP RLVYVSREREKR PGFDHHKKAG 540	

-continued

AMNSLIRVSG	VLSNAPYLLN	VDCDHINNS	KALREAMCFM	MDPQSGKKIC	YVQFPQRFDG	600
IDRHDYRSNR	NVVFFDINMK	GLDGLQGPYI	VGTGCVFRQ	ALYGFDAPKK	KKGPRKTCNC	660
WPKWCLLCFG	SRKNRKAKTV	AADKKKNRE	ASKQIHALEN	IEEGRVTKGS	NVEQSTEAMQ	720
MKLEKKFQGS	PVFVASARMV	NGGMARNASP	ACLLKEAIQV	ISCGYEDKTE	WGKEIGWIYG	780
SVTEDILTG	KMHSHGWRSV	YCTPKLAAFK	GSAPINLSDR	LHEVLRWALG	SVEIFLSRHC	840
PIWYGYGGGL	KWLERLSYIN	SVVYWPWTSLP	LIVYCSLPAI	CLLTGKFIVP	EISNYASILF	900
MALFSSIAIT	GILEMQWGV	GIDDWWRNEQ	FWVIGGVSAG	LFALFQGLLK	VLAGVDTNFT	960
VTSKAADDGE	FSDLYLFKWT	SLLIPPMTLL	IINVIGIVG	VSDAISNGYD	SWGPLEGRLF	1020
FALWVIIHLY	PFLKGLLGKQ	DRMPTIIVVW	SILLASILTL	LWVRVNPVVA	KGGPILEICG	1080
LDCL						1084

SEQ ID NO: 37	moltype = DNA	length = 3255				
FEATURE	Location/Qualifiers					
source	1..3255					
	mol_type = other DNA					
	organism = synthetic construct					
SEQUENCE: 37						
atgaacaccc	gtggcggtt	aatgcgggt	tctcacaaca	ggaatgagtt	tgtccctatt	60
aatgccatg	agaatcccc	ataaagatca	gtccaaagagg	tgagtggaca	gacatgtcaa	120
atctgcagag	atagatcgta	attgactgtt	gatggagaac	cgtttgtggc	atgttaacgaa	180
tgtgcattcc	ctgtgtgttag	accttgcata	gagtacgaaa	gacgagaagg	caatcaagct	240
tgtccacagt	gaaaaaccgg	tttcaaacgt	cattaaaggaa	gtccaaagagt	tgaagggtat	300
gaagagggaa	atgacatgt	tgatgttaga	aatgagtttg	agtatggaaa	taatgggatt	360
ggatttgc	atgtttctga	acgtatgtca	atctctgc	gcaactccgg	tttcccacaa	420
tctgatttgg	attcagctcc	acctggctct	cagatccat	tgcgtactta	cggcgacgag	480
gacgttggaa	tttcttcgt	tagacatgt	cttattgttgc	tccttcaact	tggttgtcat	540
ggcaatagag	tttcttcgt	tttcttcgt	gaccgcacgg	tggctgoaca	tccaaaggcc	600
atggatccctc	atccatgt	tgccgttata	gggttatggaa	gtgtcgctt	gaaagatcgg	660
atggaggaaat	ggaagagaaa	gcagaatgt	aaacttcagg	ttgttaggca	tgaaggagat	720
cctgattttg	aaatgtgt	tgatgtctgt	tttccaatgt	tggatgaggg	aaggcagcca	780
ttgtcttagga	agataccat	caaataatgt	aaatccatgt	tttccatgtt	tttccatgtt	840
ctacgttttgc	tgatgtttgg	tcttttttgc	cactaccgt	tttccatccc	cgtcaaaat	900
gcataatgtt	tgtgggttat	tttctgtata	tgtgagatat	ggtttgcgtt	tttcatgggtt	960
cttgatcgt	tcccttaatgt	gtacccttc	gacgcgagaa	cgtaacttgc	ccgactctca	1020
ttaaagatgt	agaaaaaagg	gaaacgcgt	ggactatccc	ctgtggatgt	atttgttaat	1080
acagttggatc	cattgaaaga	gcctccgtt	attactgtca	atactgttt	gtctattttt	1140
gctgttggatt	atccgtcgt	taagggtgt	tgttacgtat	tttccatgtt	tttccatgtt	1200
cttactttcg	aaatgttttc	tgagaccgt	gaattcgca	ggaaatgggt	tttccatgtt	1260
aaagaaatatt	gtatttgag	tcgtgtcttc	ggatgttttt	tctgcataaa	aatggactac	1320
tttgaagaatgt	aaatgttcatcc	cgccattttgtt	ggggagcggc	gagccatgaa	gagatgttt	1380
gaagaatttca	aaatgtgt	caatgttta	gtacgttca	cacagaaatgt	gccttgcggat	1440
ggtttggacta	tgcaagacgg	tacacccgtt	cccggtata	gtgtgcgaga	tcatccgttgc	1500
atgattcagg	tcttccgttgg	aaatgtgttgc	tttccatgtt	tttccatgtt	tttccatgtt	1560
cgatttagtt	acgttttctcg	tgagaagagg	cccggttgc	atccatgtt	aaaggcttgc	1620
gctatgttgc	ccctgtatcc	agtttgcgttgc	tttccatgtt	atgttgcata	cccttgcgtt	1680
gtcgatttgc	atcaatcat	caacaatgt	aaatgttgc	gagaagcaat	gtgttgcgtt	1740
atggatccctc	atgtgttgc	aaatgttgc	tttccatgtt	tttccatgtt	tttccatgtt	1800
attgtatggc	acgtatgttgc	aaatgttgc	tttccatgtt	tttccatgtt	tttccatgtt	1860
ggtttggatgt	ggcttacagg	gttgcataat	gttgcataat	gttgcataat	gttgcataat	1920
gcgcgtttacg	gatttgcgt	accgttgcgt	aaatgttgc	atgttgcgtt	atgttgcgtt	1980
tggcccaat	gtgtgttgc	atgtttgttgc	tttccatgtt	tttccatgtt	tttccatgtt	2040
gctgttggatgt	aaatgttgc	aaatgttgc	tttccatgtt	tttccatgtt	tttccatgtt	2100
atcgaagagg	gcgcgttac	taaaatgttgc	aaatgttgc	aaatgttgc	aaatgttgc	2160
atgaagtgg	aaatgttgc	aaatgttgc	tttccatgtt	tttccatgtt	tttccatgtt	2220
atatgttgc	tttccatgtt	aaatgttgc	tttccatgtt	tttccatgtt	tttccatgtt	2280
attatgttgc	tttccatgtt	aaatgttgc	tttccatgtt	tttccatgtt	tttccatgtt	2340
tctgttgcgt	aaatgttgc	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	2400
tattgttacac	aaatgttgc	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	2460
ctccatcat	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	2520
cctatgttgc	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	2580
tctgttgcgtt	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	2640
tgttgcgttca	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	2700
atggcgctt	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	2760
gggtatgttgc	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	2820
ctgttgcgtt	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	2880
gtcacatca	aaatgttgc	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	2940
tcacttgc	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	3000
gtctctgtat	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	3060
tttgcgttgc	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	3120
gatagaatgc	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	3180
ctttgggttcc	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	3240
tttagactgt	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	tttccatgtt	3255

SEQ ID NO: 38	moltype = AA	length = 1084
FEATURE	Location/Qualifiers	
source	1..1084	

-continued

```

mol_type = protein
organism = synthetic construct

SEQUENCE: 38
MTTGGRLIAG SHNRNEFVLI NADENARIRS VOELSGQTQ ICRDEIELTV DGEPPVACNE 60
CAFPVCRPCY EYERREGNQA CPQCKTRFKR LKGSPRVEGD EEEIDDLD NEFEYGNNGI 120
GFDQVSEGMS ISRRNSGFPQ SDLDSAPPGS QIPLLTYGDE DVEISSLRHA LIVPPSLGGH 180
GNRHPVPSLS DPTVAAHPRP MVPQKDLAVY YGGSVAWKR MEEWKRKQNE KLQVVRHEGD 240
PDFEDGDDAD FPMMDEGRQF LSRKIPIKSS KINPYRMLIV LRLVILGLFF HYRILHPVKD 300
AYALWLISVI CEIWFAVSWV LDQFPKWYPI ERETYLDRLS LRYEKEGKPS GLSPVDVFVN 360
TVDPKKEPPL ITANTVLSIL AVDYPVDKVA CYVSDDGAAM LTFEALSETA EFARKWVFC 420
KKYCIEPRAPI WYFCHKMDY LKNKVCHPAP RERRAMKRDY EEFKVKINAL VATAQKVPED 480
GWTMMDGTPW PGNSVRDHDPG MIQVFLGSDG VRDVENNELP RLIVYVSRERK PGFDHHKKAG 540
AMNSLIRVSG VLSNAPYLLN VDCDHYINNS KALREAMCFM MDPQSGKKIC YVQFPQRFDG 600
IDRHDRYSNR NVVFFDINMG GLDGLQGPYI VGTGCVFRQQ ALYGFDAPKK KKGPRKTCNC 660
WPWKCLLCFG SRKNRKAVT AADKKKKRN ASKQIHALEN IEEGRVTKGS NVEQSTEAMQ 720
MKLEKKFQGS PVFVASARME NGGMARNASP ACLLKEIAQV ISCGYEDKTE WGKEIGWIYG 780
SVTEDILTGK KMHSHGWRSP YCTPKLAALK GSAPINLSDR LHQVLRWALG SVEIFLRSRH 840
PIWYGYGGGL KWLERLSYIN SVVYWPSTSPL LIVYCSLPAI CLLTGKFIVP EISNYASILF 900
MALFSSIAIT GILEMQWGVK GIDDWWRNEQ FWVIGGVSAA LFALFQGLLK VLAGVDTNFT 960
VTSKAADDGE FSDLYLFKWT SLLIPMPLL IIINVIGIVG VSDAISNGYD SWGPLFGRRLF 1020
FALWVIIHLH PFLKGGLGKQ DRMPHIIVVV SILLASILTL LWVRVNPFA KGGPILEICG 1080
LDCL                                         1084

```

```

SEQ ID NO: 39      moltype = DNA length = 3255
FEATURE           Location/Qualifiers
source            1..3255
mol_type = other DNA
organism = synthetic construct

```

```

SEQUENCE: 39
atgaacacccg gtggcgggtt aatcgccgtt tctccacaaca ggaatggatt tgtccctatt 60
aatgccatgtt agaatccccg aataagatca gtccaaagagc tgagtggaca gacatgtcaa 120
atttcgcaggat attagatcgta attactgtt gatggagaaac cggttgtggc atgttaacggaa 180
tgtgcattttt ctgtgtgtat accttgctat gagtacgaaa gacgagaagg caatcaagct 240
tgtccacagt gcaaaaacccg tttcaaacgt cttaaaggaa gtccaagagt tgaagggtat 300
gaagaggaaatc atgacattgtt tgatggaaatc atatggaaa taatggattt 360
ggatttgcgtt aggttgcgtt acgtatgtca atctctcgcc gcaactccggg ttcccaccaa 420
tctcgatttt attcaagctcc accttcgcgtt cagattccat tgctgactt cggcgacag 480
gacgttggata tttcttctgtt tagacatgtt ctattgttcc tccttcactt tggtggcat 540
ggcaatagatc ttcatctgtt ttctttttt gacccgaccgg tggctgcaca tccaaggcc 600
atggtaacctc agaaatgtt tgcgggttat ggttatggaa gtgtcgctt gaaagatccg 660
atggaggaaat ggaagagaaa gcagaatgtt aaatccagggttggtagca tgaaggatgtt 720
cctcgatttt aagatgggtt tgatgtcttgc tttccaaatgttggatgggg aaggcagcc 780
ttgtcttggaa agatataatc caaatcgacgaaatccatgttggatgggg 840
ctacgttcttgc tgatttttttgc tttttttttt cactaccgtt ttcttcaccc cgtcaaaatgtt 900
gcatacgatgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 960
cttgcgttgcgtt tccctttaatgtt gtttttttttgc ttttttttttgc ttttttttttgc 1020
ttaaatgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1080
acatggatccatgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1140
gtctgttgcgtt atccctgttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1200
cttacttttcgtt aagtttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1260
aagaaatgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1320
tttgcgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1380
gaagaaatccatgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1440
gggttggactt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1500
atggatccatgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1560
cgatgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1620
gtctatgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1680
gtcgattgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1740
atggatccatgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1800
atggatccatgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1860
gggttggatgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1920
gcgttttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1980
tggccaaatgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2040
gtctgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2100
atcgaaatgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2160
atcgaaatgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2220
aatgggtggaaatgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2280
attagttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2340
tctgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2400
tattgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2460
ctccatgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2520
cctatgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2580
tctgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2640
tgcgtttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2700
atggatccatgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2760
gggttggatgttgcgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2820

```

-continued

ctgtttgctc	tcttccaagg	tctccctaag	gttcttgctg	gtgtcgacac	taacttcaca	2880
gtcacatcaa	aaggcagctga	tgtatggagag	ttctcttgacc	tttacatctt	caaatgggt	2940
tcacttttca	tcccttccaat	gactctactc	atccataaaacg	tccatggagt	catatgcgg	3000
gtctctgtat	ccatcagcaga	tgatcagacat	tcgttggggac	cgttttcggd	aaagactgttc	3060
tttgcactt	gggtcatcat	tcatctttac	ccgttcccta	aagggttgc	tggaaacaaa	3120
gatagaatgc	caaccattat	tgtctgtctgg	tccatctcc	tggcttcgat	tcttacactt	3180
ctttgggttc	gggtaatcc	gttttgtggcg	aaaggccggtc	ctatattctcg	gatctgtggt	3240
tttagactgt	tgtga					3255

```
SEQ ID NO: 40          moltype = AA  length = 1084
FEATURE                Location/Qualifiers
source                 1..1084
                       mol_type = protein
                       organism = synthetic construct
```

SEQUENCE : 40		
MNTGGRLIAQ SHNRNEFVLI NADENARIRS VQEELSGQTQ ICRDEIELTV DGEFPVACNE	60	
CAFPCVRCY EYEREGNQA CPQCKTRFKR LKGSPRVEGD EEEIDDLDL NEFEYGVNNGI	120	
GFDQVSSEGMS ISRRNSGFPQ SDLDSAPPGS QIPLLLTYGDE DVEISSDRHA LIVPPSLGGH	180	
GNRVHPVSLs DPTVAAHPRP MVPQKDLAVY GYGSVAWKDR MEEWKRKQNE KLQVVRHEGD	240	
PDFEDGDDAD FPMMDEGRQP LSRKIPKISS KINPYRMLIV LRLVILGLFF HYRLHPVKD	300	
AYALWLISVI CEIWFAVSWV LDQDFPKWYPI ERETYLDRRLS LRYEKEGPKS GLSPVDVFVs	360	
TVTDPKLBPPL ITANTVLSL VPYDPPVDKA CYYVFDGAAM LTFLEALSETA EFARKWVPPC	420	
KKYCIEPRAP EWYFCHKM DY LKNKVHPAFV RERRAMKR DY EEFKVKINAL VATAQKPED	480	
GWTMODOGTW PGNSVRDH PG MIQVFLGSSD VRDVENNEL RLVYVSREKR PGDFDHKKAG	540	
AMNSLLIRVSG VLSNAPYL VNDCDHYINN KALREACMFMD PQSGKKIC YVQFPQRFDG	600	
IDRHDHRSYRN NVFVFDINM GLDGLQGPPI VGTGCFVRQRQ ALYGFADPKK KKGRPKTCNC	660	
WPWKCLLCFG SRKNRKAKTV AADKKKKNRE ASKQIHALEN IEEGRVTKGS NVEQSTEAMQ	720	
MKLEKKPGQS PVFVASARME NGGMARNASP ACLLKEAIQV ISCCGYEDKTE WGKEIGWIYG	780	
SVTEDILTGT KMHSHGWRSV YCTPKLAFAK GSAPINLSDR LHQVLRLWALG SVEIFLRSRH	840	
PWIYWGCGGL KWLERLSYIN PWVSYPTSL LIVYCSLPAI CLTGGKFIVE EISNYASILF	900	
MALFSSIAIT GILEMQWKV GIDDWWRNEQ FWVIGGSVAH LFALFQGLLK VLAGVDTNFT	960	
VTSKAADDGE FSDLYLFWKT SLLIPPMTLL IINVIGVIVG VSDAISNGYD SWGFLGRLF	1020	
FALWVIIHLH PFLKGLLGKQ DRMPTIIVVV SILLASILTL LWVRVNPVFA KGGFILEICG	1080	
DLCL	1084	

SEQ ID NO: 41 moltype = DNA length = 3255
FEATURE Location/Qualifiers
source 1..3255
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 41

	organism	synthetic construct	60			
atgaacacgg	gtggcgggtt	aatcgccggt	tctcacaaaca	ggaatggagtt	tgccttcatt	120
aatggccgat	agaatggccg	aataaatgca	gtccaaagacg	tgatggccaa	gacatgtcaa	180
atctgcagag	atagagatcga	attgactgtt	gatggagaac	cgttttgtgc	atgtaaacgaa	240
tgtgcattcc	ctgtgtgttag	accttgctat	gagtagacaa	gacgagaagg	caatccaagt	300
tgtccacagt	gcaaaaaccgg	tttcaaacgt	cttaaaggaa	gtccaaagagt	tgaagggtat	360
gaagaggaga	atgacattgtt	tgatgttagac	aatgaggatgtt	agtagtggaaa	taatggggat	420
ggatgttgc	agggttctgtt	aggatgttgc	atctctcg	gcaactccgg	ttttccccaaa	480
tctgttttgtt	attcagotcc	acctggctct	cagatccat	tgcgtactta	cgggcagcag	540
gacgttggaa	tttcttctgt	tagacatgtt	cttattgtt	ctccctca	tgggttgtcat	600
ggcacataaggg	ttcacatctgtt	tttcttcttct	gaccggacccg	tggtcgcaca	tccaaaggccc	660
atggtatcc	agaaagatgt	tgccgtttat	ggtttagggaa	gtgtgccttgc	gaaaatgtcg	720
atggggat	ggaagagaaaa	gcagaatgttag	aaacttcagg	ttgtttaggc	tgaaggagat	780
cctgatttt	aaagatgggtt	tgatgtgtat	tttccaatgt	tggatgaggg	aaggcagccaa	840
tttgttctggaa	agataccataa	caatccggac	aagataaaatc	cttacccgtt	gtttaattgttgc	900
ctacgttgc	tgatttcttttgc	tctcttttttgc	cactacccgt	tcttcaccc	cgttcaaagat	960
gcataatgttt	tgtggcttat	ttctgttata	tgtgagatata	gggttgcgttgc	tctatgggtt	1020
cttgatcagt	tcccttaatgt	gtaccctatc	gagcgagaaa	cgtacttgg	ccgactctca	1080
ttaagatgtt	agaaaagaaagg	gaaaccctcg	ggactatccc	ctgtggatgtt	atttggatgtt	1140
acatggatc	cattggaaaga	gcctccgtt	attactgc	atactgttctt	gtctatgtt	1200
gtgtgttgc	atccctgtcga	taagggttgc	tgttacgtat	ctgtatgttgc	tgcgtctatgt	1260
cttactttcg	aaagtctttc	tgagaccgt	gaattccaa	ggaaatgggt	tcctttctgc	1320
aagaatataat	gttattggggc	tcgtgttccc	gaatgttatt	tgcgcataaa	aatggactac	1380
ttgaagaata	aagtgttcc	ccgcattttgtt	agggggccgc	gagccatgaa	gagagattat	1440
aaagaatcca	aagtagttaat	caatgttta	gtaccaacag	cacagaaatgt	gccttggatgt	1500
ggttggactt	tgcaagacgg	tacaccccttgc	cccggtata	tgtgtcgaga	tcatcttgc	1560
atgattcagg	tcttccttgg	aaagtgcgggt	gttcgtgtat	tcgaaaacaa	cgagttgcct	1620
cgatttagttt	acgtttctcg	tgagaagaga	cccggttgc	atccacataa	gaaggcttgc	1680
gtctatgtt	ccctgtatacg	agtctcttgg	gttctatcaa	atgtcttca	cctttgtatgt	1740
gtcgatgttgc	atctacatcat	caacaatgc	aaagcttta	gagaacaaat	gtgtttcatgt	1800
atggatcctc	agtccggaaa	gaaaatctgt	tatgttca	tcccttca	atgttgcgttgg	1860
attgatggc	acgtatcgata	ctcaaatcg	aatgttgcgtt	tctttgtat	caatatgaaa	1920
gggttggatgt	ggcttacaaagg	gccttatatac	gtcggtacag	gttgcgtttt	caggaggccaa	1980
gcgttttac	gatttgcgttgc	acggaaagag	aagaaggccc	cacgttgc	atgtcaattgc	2040
tgcccaaat	gggttgcgtt	atgttttgttgc	tcaagaaaga	atgtcaatgc	aaagacatgttgc	2100
gtgtgcgttgc	agaagagaaa	gaataggggaa	gcgttca	atgttgcgttgc	attagaaaaat	

-continued

atcgaagagg	gcccgtcac	taaagggtct	aacgtagaac	agtcaaccg	ggcaatgcaa	2160
atgaagtgg	agaagaatt	tggcagtt	cctgtatttgc	ttgcacatcg	gcgtatggag	2220
aatgggtggg	tggttagaaa	cgcaagccg	gcttgtctgc	ttaaagaagc	catccaagtc	2280
attagttcg	gatatgaaga	taaaactaa	tggggaaaag	agattgggt	gatctatgg	2340
tctgttacc	aatgatattct	tacgggttt	aagatgcatt	ctcatgggt	gagatctgtt	2400
tattgtacac	caaagttagc	ggcttcaaa	ggatcagctc	caatcaactt	ttcggatcgt	2460
ctccatcaag	ttcttgcatg	ggcgcttggg	tcgggttgg	tttcttgag	taggtatgt	2520
cctatgggt	atgggtatgg	agggtgggt	aaatgggttgc	agcgggttgc	ctacattaaac	2580
tctgtggtt	accctggac	ctctctacc	ctcatcgtt	actgttctt	ccctgccatc	2640
tgcttctca	ctggaaaatt	categttc	gagattaaacg	actatcgag	tatctcttc	2700
atggcgct	tctctgtcgat	tgcataaacc	ggatcataac	atgtgcata	gggcaaaagt	2760
gggatcgtat	atgggtggag	aaacgaaacag	tttgggtca	ttggagggtt	ttctgcgtat	2820
ctgtttgtc	tcttcaagg	tctctcaag	gttcttgct	gtgtcgac	taacttcaca	2880
gtcacatcaa	aacgagct	tgatggagag	ttctctgacc	tttacccctt	caaatggact	2940
tcacttccat	tccctccat	gactctact	atcataaacg	tcattgggt	catatcgga	3000
gtctctgtat	ccatcagaa	tggatac	tcgtggggac	cgcttccgg	aagactgtt	3060
tttgcactt	gggtcatcat	tcatcttac	cggttctta	aaaggttgt	ttggaaacaa	3120
gatagaatgc	caaccaattat	tgtctgtgg	tcacatctcc	ttggcctcgat	tcttacactt	3180
cttgggtcc	gggttaatcc	gtttgtgg	aaaggcggtc	ctattctcg	gatctgtgt	3240
ttagactgt	tgtga					3255

SEQ ID NO: 42 moltype = AA length = 1084
 FEATURE Location/Qualifiers
 source 1..1084
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 42
 MTNGGRILAG SHNRNEFVLI NADENARIRS VQELSGQTCQ ICRDEIELTV DGEPPVACNE 60
 CAFPVCRPCY EYERREGNQA CPQCKTRFKR LKGSPRVEGD EEEIDDIDLD NEFEYGNNGI 120
 GFDQVSEGMS ISRRNSGFPQ SDLDAPPGS QIPLLTYGDE DVEISSLRHA LIVPPSLGGH 180
 GNRVHPVSLIS PTDVAHKPRM MVPQKDLAVY GYGSAWKR MEEWKRKQNE KLQVVRHEGD 240
 PDPEGDGDDA FPMMDEGRQP LSRKIPIKSS KINPYRMLIV LRLVILGLFF HYRILHPVKD 300
 AYALWLISVI CEIWFVAWSVW LDQFPKWYPI ERETYLDRS LRYEKEGKPS GLSPVDVFVS 360
 TVDPLKEPPL ITANTVLSIL AVDYPVDKVA CYVSDDGAAM LTFEALSETA EFARKWPFC 420
 KKCYCIEPRAP EWYFCHKMDY LKNKVHPAFV RERRAMKRDY EEFKVKINAL VATAQKVPED 480
 GWTMQDGTPE PGNSVRDHDP MIQVFLGSDG VRDVENNELP RLVYVSRERK PGFDHHKKAG 540
 AMNSLIRVSG VLSNAPYLLN VDCDHYINNS KALREAMCFM MDPQSGKKIC YVQFPQRFDG 600
 IDRHDYRSNR NVVFFDINMK GLDGLQGPIY VGTGCVFRQ ALYGFDAPKK KKGPRKTCNC 660
 WPWKWCLLCFG SRKNRKAKTV AADKKKKRN ASKQIHLEN IEEGRVTKGS NVEQSTEAMQ 720
 MKLEKKFGQS PVFVASARME NGGMARNASP ACLLKEAIQV ISCGYEDKTE WGKEIGWIYG 780
 SVTEDILTGK KMHSHGWRSV YCTPKLAALK GSAPINLSDR LHQVLAWALG SVEIFLSRHC 840
 PIWYGYGGGL KWLERLRSYIN SVVYPTWTSLP LIVYCSLPAI CLLTGKFIIVP EISNYASILF 900
 MALFSSTAIT GILEMQWGKV GIDDDWWRNEQ FWVIGGVSAH LFALFQGLLK VLAGVDTNFT 960
 VTSKAADDGE PSDLYLFKWT SLLIPPMTLL IINVIGIVG VSDAISNGYD SWGPLFGRFL 1020
 FALWVIIHLY PFLKGLLGKQ DRMPTIIVVV SILLASILTL LWVRVNPVVA KGGPILEICG 1080
 LDCL 1084

SEQ ID NO: 43 moltype = DNA length = 3255
 FEATURE Location/Qualifiers
 source 1..3255
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 43
 atgaacaccg gtggcgggt aatccgcgt ttcacacaaca ggaatggatt tgccttcatt 60
 atgcccgtat agaatggccg aataagatca gtccaaagac tgagtggaca gacatgtcaa 120
 atctgcagag atgatgtca attgtatgtt gatggagaa cgtttgtggc atgttaacgaa 180
 tggatccctt ctgtgtgtat acctgtctat gaggatcggaa gacgagaaat caatcaagct 240
 tgcacacatgt gcaaaaaccgg tttcaaacgt cttaaaggaa gtccaaaggt tgaagggtat 300
 gaagggaaat atgacatttgc tgatgttaga aatggatgtt agtatggaaa taatgggatt 360
 ggatgttgcgat aggtttctga aggtatgtca atctctcgat gcaactccgg ttcccacaa 420
 tctgatgtttt attcagtc acctggctt ctgtgtactt tgctgactt ccggcagcgg 480
 gacgttggaa ttcttcgtat tagacatgtt ctattgttc ctccctact tgggtgtcat 540
 ggcacatagat ttcatctgtt ttcttttgc gacccgaccg tggctgcaca tccaaaggct 600
 atggatccctt agaaaatgtt tgccgttataat ggtttatggaa gtgtcgctt gaaaatcg 660
 atggaggaaat ggaagagaaaa gcaagatggaa aaaaattcgg tggtttaggg tgaaggagat 720
 cttcgatgtttt aagatgtgtat tgatgtctat ttccatgtat tggtatgggg aaggcaggca 780
 ttgtcttagaa agataccat caaatcgago aagataaaatc ttaccggat gttattgtt 840
 ctacgttctt tgattcttgc tctttctt cactaccgtt ttcttcaccc cgtcaagat 900
 gcatatgtttt ttctgtttat tggtgatgtt ggtttgtgtt ttcatgggtt 960
 cttcgatgtttt cccttaatgtt tgacccatc gacccgtt gacccgtt ccgactctca 1020
 ttaaagatgtt agaaaatggaa gaaaccgtcg ggactatccc ctgtggatgtt atttggatgt 1080
 acagtggatc ctttgcggat gcctccgtt attactgtca atactgtttt gtcttattttt 1140
 gctgttggat atccgtcgat taagggttgc tggatgtat ctgtatgtt tgctgtat 1200
 ctactttcg aagcttttc tgagaccgtt gaaattcgaa ggaaatgggt tcctttctgc 1260
 aagaaatattt gtattggcc tctgtctcc gatgtgttattt tctggccatcaa aatggactac 1320
 ttgaagaata aagttcatcc cgccatgtt agggagccgc gacccatgaa gagagatatt 1380

-continued

gaagaattca	aagtaaagat	caatgcttta	gttagcaacag	cacagaaaatg	gcctgaggat	1440
ggtttggacta	tgcaagacgg	tacaccttgg	cccggtaaata	gtgtggaga	tcatccttgc	1500
atgattcagg	tcttcttgg	aagtgcgggt	gttcgtatg	tgcaaaaacaa	cgagttgcct	1560
cgatttagtt	acgtttctcg	tgagaagaga	cccggttttgc	atcaccat	aa gaaggcttgc	1620
gctatgaatt	ccctgatacg	agtctctggg	gttcttatcaa	atgtccttta	ccttctgaat	1680
gtcgatgtg	atcactacat	caacaatago	aaagcttta	gagaagcaat	gtgtttcatg	1740
atggatcctc	agtccaggaa	gaaaatcttct	tatgttcaat	tccctcaaaag	gttcgtatgg	1800
atttgatagc	acgatcgata	ctcaaatcg	aatgttgcgt	tctttgatata	caatgtaaaa	1860
ggtttggatg	ggcttacaagg	gcctataac	gtcggtacag	gttgtgttt	caggaggccaa	1920
gegccttacg	gatttgcgtc	accgtacaa	aquaaggccg	cacgtaaagac	atgtcaatgc	1980
tggccaaaat	gtgttgcct	atgtttgtt	tcaagaaaga	atcgtaaagc	aaagacatgt	2040
gtctcgccata	agaagaagaa	gaatggaa	gcgtccaaac	agatccacgc	attagaaaaat	2100
atcgaagagg	gccgcgtcac	taaagggtct	aacgttagac	agtcaaccg	ggcaatgc	2160
atgaagtgg	agaagaaatt	tggggcgtt	cctgtat	ttgcatctgc	gcgtatgg	2220
aatgggttgg	tggctgat	aaaaccccg	gttgcgttgc	ttaaagaagc	catccaatgc	2280
attagtgcg	gatatgtaa	aaaatgtaa	tggggaaatg	agatgggttgc	gatctatgtt	2340
tctgttaccg	agatattct	tacgggttt	aatgtcatt	ctcatgggt	gagatcttgc	2400
tattgttacac	caaaggtagc	ggcttcaaa	ggatcgctc	caatcaatct	ttcggatcgt	2460
ctccatcaag	ttcttcgagc	ggcgttgc	tcgggttgc	ttttcttgc	taggcattgt	2520
cctatgttgc	atgttgcgtt	agggtgggttgc	aatatgggttgc	agcgttgc	ctacattaaac	2580
tctgttgcgtt	accctgttgc	ctcttaccc	ctcatcgttt	actgttctt	ccctgcacatc	2640
tgtcttctca	ctggaaaatt	categttcc	gagatttagca	actatgcgg	tatctcttc	2700
atggcgcgtct	tcttcgtat	tcaataac	ggtatttctc	agatgcata	gggcaaaatgt	2760
gggatcgat	atgggttgg	aaaaccccg	ttttgggtca	tttgggggttgc	ttctgcgtc	2820
ctgttgcgtc	tcttcaagg	tcttgcgtt	gttgcgttgc	gttgcgttgc	taacttcaca	2880
gtcacatcaa	aaccaatgt	tgtatggag	ttctcttgc	tttacccctt	caaattgttgc	2940
tcacttctca	tccttcata	gacttactt	atcataaact	tcattttgcgt	catatgcgttgc	3000
gtctctgtat	ccatcagca	tggatcgc	tcgttggggac	cgctttccg	aagactgttgc	3060
tttgactt	gggtatcat	tcatcttac	ccgttccat	aagggttgc	tggggaaacaa	3120
gatagaatgc	caaccattat	tgtcgcttgc	tccatccccc	tggcctcgat	tcttacactt	3180
cttgggttcc	gggttataatcc	gtttgttgc	aaaggccgttgc	ctatttgc	gatctgttgc	3240
tttagactgt	tgtga					3255

SEQ ID NO: 44	moltype = AA	length = 1084
FEATURE	Location/Qualifiers	
source	1..1084	
	mol_type = protein	
	organism = synthetic construct	

SEQUENCE: 44						
MNTGGRLIAG	SHNRNEFVLI	NADENARIRS	VQELSGQTCQ	ICRDEIELTV	DGEPEPVACNE	60
CAFPVCRPCY	EYERREGNQA	CPQCKTRFKR	LKGSPRVEGD	EEEDDIDLD	NEFEYGNNGI	120
GFDQVSEGMS	ISRRNSGFPQ	SDLDSAPPGS	QIPLLTYGDE	DVEISSLRHA	LIVPPSLGGH	180
GNRVHPVSL	DPTVAAHPRP	MVPQKDLAVY	GYGSVANKDR	MEEWKRKQNE	KLQVVRHEGD	240
PDFEDGDDAD	PDFMDEGRQP	LSRKLPPIKSS	KINPYRMLIV	LRLVILGLFF	TYRILHPVKD	300
AYALWLISVI	CEIWFAVSVW	LDQFPKWPYI	ERBTYLDRLS	LYREKEGKPS	GLSPVDFVFS	360
TVDPKLEPPL	ITANTVLSIL	AVDYPVDKVA	CYVSDDGAAM	LTFEALSETA	EFARKWVPF	420
KKYCIEPRAP	EWYFCHKMDY	LKNKVHPAFV	RERRAMKRDV	EEFKVKINAL	VATAQKVPED	480
GWTMQDGT	PGNSVDRHPG	MIQVFLGSDG	VRDVENNELP	RLVYVSREKR	PGFDHHKKAG	540
AMNSLIRVSG	VLSNAPYLLN	VDCDHYINNS	KALREAMCFM	MDPQSCKKIC	YVQFPQRFDG	600
IDRHDRYSNR	NVVFFDINMK	GLDGLQGPYI	VGTGCVFRQQ	ALYGFDAPKK	KKGPRKTCNC	660
WPKWCLLCFG	SRKNRKAFTV	AADKKKKRNRE	ASKQIHALEN	IEEGRVTKGS	NVEQSTEAMQ	720
MKLEKKFGQS	PVFVASARME	NGGMARNASP	ACLLKEAIQV	ISCGYEDKTE	WGKEIGWIYG	780
SVTEPDLTGF	KMHSHGWRSV	YCTPKLAAPK	GSAPINLSDR	LHQVLRALG	SVEIFLRSRC	840
PIWYGYGGGL	KWLERLSYIN	SVYVPTWSLP	LIVYCSLPAI	CLLTGKIVP	EISNYASILF	900
MALFSSIAIT	GILEMQWGV	GIDDDWRNRBQ	FWVIGGVS	LFALFQQLK	VLAGVDTNFT	960
VTSKAADDGE	FSDLYLFWKT	SLLIPPMTLL	IINVIGIVG	VSDAISNGYD	SWGPFGRFL	1020
FALWVIIHLY	PFLKGLLGKQ	DRMPTIIVVW	SILLASILTL	LWVRVNPVVA	KGGPILEICG	1080
LDCL						1084

SEQ ID NO: 45	moltype = DNA	length = 3255
FEATURE	Location/Qualifiers	
source	1..3255	
	mol_type = other DNA	
	organism = synthetic construct	

SEQUENCE: 45						
atgaacacccg	gtggtcgggt	aatcgccgtt	tctcacaaca	ggaatggat	tgtcttcat	60
aatggccatg	agaatgcccc	aataagatca	gtccaagacg	ttagtggaca	gacatgtcaa	120
atctgcagag	atgagatcga	attgactgtt	gatggagaac	cgtttgc	atgttaacgaa	180
tgtgcattcc	ctgtgtgtat	accttgcata	gagtacgaaa	gacgagaagg	caatcaatgt	240
tgtccacatgt	gcaaaaacccg	tttcaaacgt	ctttaaggaa	gtccaaatgt	tgaagggtat	300
gaagaggaaat	atgacatgt	tgatgttgc	aatggatgttgc	atgtatggaa	taatgggttgc	360
ggatgttgc	agggttctgt	aggtatgtca	atctctcg	gcaactccgg	tttcccaatgt	420
tctgatgttgc	attcagcttcc	acctggcttgc	cagattccat	tgctgactt	cggcgacatgt	480
gacgttgc	tttcttctgt	tagacatgt	cttattgttgc	ctcccttact	tgggttgc	540
ggcaatagat	ttcataatgt	ttcttcttgc	gacccgaccc	tggctgcaca	tccaaaggcct	600
atggtacatc	agaaatgt	tgcgggttat	ggttatggaa	gtgtcgcttgc	gaaatgt	660

-continued

SEQUENCE: 47

```

atggaagota ggcgggtct tgcggcgtt tctcataacc gtaatgaact agtcgtcatt 60
cacaaccatg aagagccaaa gcctctgaag aatcttagatg gacaattctg tgagatatgt 120
ggagatcaga ttggtttaac agtagaaaggaa gacctctcg tagcttgcaa tgagtgggt 180
tttccggcgt ttagacccctg ctatgagtag gagagaagag aaggaacaca aaactgtcct 240
cagtgtaaaga ctcgttacaa gctgttcaga ggaagccaa gagtggaggg agatgaagac 300
gaagaagata ttgtatgatg tgatgttgcgaa tttatatcgc aacatgaaca agatgaagac 360
aagcattctg ctgaggctat gctttatggg aaaatgacgt atgaaaggagg tcctgaggat 420
gatgagaatg ggagatccc acctgttata gctgggtgc atagtgaga atttccagtt 480
ggaggagggtt atggtaatgg agaagatggg cttcataaagc gtgtgcaccc atatccatca 540
tctgaaggctg ggagtgggg aggatgggg gaaagaatgg atgactggaa gctccagcat 600
ggaatcttgc ggcggaaacc agatgtatg cttgagatgg gactgtatcga cgaggcacgg 660
cagccactct cgcggaaagt tccccattgc tcaagcaaga tcaatccata tcggatggtc 720
atcgctcgata ggcttgcgtat tctatcgatg tttctggcgtt ataggcttt gaatccagtg 780
catgtatcgc tgggattatg gtgcacccgt tgatgttgc aaatctgggtt cgcgtctct 840
tggattcttgc atcgttccc caatgttgc cttatgtac tctatgtatcgg 900
cttccctca ggtacgagag agaagggtaa ccaaataatgc ttggccctgtt agatgtcttt 960
gtcagtaacgg tggaccattt gaaggagctt cccctctgtca catccaaacac tggctgtca 1020
atcttggcca tggactacc agttggagaa atcttctgtt atgtctgtca cgcacgggtc 1080
tcaatgttgc catgtatcgc tctctggcaaa atcgttgcgtt atgggttccc 1140
ttctgtaaaga aattctccat agagccacgg gcaccggaga tggacttgc ttttcaatgg 1200
gattatctcc accagaaatgg ccaccaaca ttgtttaagg aacgtcgagc catgaagaga 1260
gaatatgagg agttcaagggtt caggatgtatcgtt gctcaatgttgc cgaaggccctt 1320
ctagaagggtt ggatcatgc agatggaaaca cccgtggccca ggaacaaacac caaggaccac 1380
cccggtatgaa tccaaatgtt cctctggccca agcggaggat ttgtatgtca agggcatgag 1440
cttcctcgcc ttgtgtacgtt gtcggcgttgc aacgttgcgtt gtttcaaca ccacaaggaaa 1500
gttggcgcoca tggatgttgcgtt gtcggcgttgc tcaacaaatgc tcccttcatcg 1560
ctgaaacttggg atctgtatcgc tctatgttgc aacacgttgc cccgtggggaa agcaatgtgt 1620
tttttgcgtt atcttctgtt tggaaaggaaat ttttgcgttgc ttcagttccc tcaaaagggtt 1680
gatggcatttgc acacaaatggc tggatgttgc aacagaaaca cagtcttctt tgatgttcaat 1740
atgaaagggttgc tagatggatc ccaagggttgc tggatgttgc ttttgcgttgc 1800
cgacaaatggc tggatgttgc tggatgttgc aacgggttca aacgttgcgtt gatgtatcgc 1860
tggatgttgc tggatgttgc tggatgttgc aacgggttca aacgttgcgtt gatgtatcgc 1920
atgaaatgttgc acgttgcgttgc cccgtggggaa gcaaggatgttgc ttttgcgttgc 1980
gaaatgttgc tggatgttgc tggatgttgc tggatgttgc ttttgcgttgc 2040
gaagaagggttgc tggatgttgc tggatgttgc tggatgttgc ttttgcgttgc 2100
gtcataatggc tggatgttgc tggatgttgc tggatgttgc ttttgcgttgc 2160
ggatgttgc tggatgttgc tggatgttgc tggatgttgc ttttgcgttgc 2220
atttactgttgc tggatgttgc tggatgttgc tggatgttgc ttttgcgttgc 2280
aggtttaacc aggttttgcgttgc tggatgttgc tggatgttgc ttttgcgttgc 2340
agtcctcttgc tggatgttgc tggatgttgc tggatgttgc ttttgcgttgc 2400
gccaacacaaatggc tggatgttgc tggatgttgc tggatgttgc ttttgcgttgc 2460
gcccattgttgc tggatgttgc tggatgttgc tggatgttgc ttttgcgttgc 2520
tttttgcgttgc tggatgttgc tggatgttgc tggatgttgc ttttgcgttgc 2580
ggatgttgc tggatgttgc tggatgttgc tggatgttgc ttttgcgttgc 2640
gtcataatggc tggatgttgc tggatgttgc tggatgttgc ttttgcgttgc 2700
tttccatgttgc tggatgttgc tggatgttgc tggatgttgc ttttgcgttgc 2760
tggatgttgc tggatgttgc tggatgttgc tggatgttgc ttttgcgttgc 2820
gcaatgttgc tggatgttgc tggatgttgc tggatgttgc ttttgcgttgc 2880
cttttgcgttgc tggatgttgc tggatgttgc tggatgttgc ttttgcgttgc 2940
agacaaatggc tggatgttgc tggatgttgc tggatgttgc ttttgcgttgc 3000
tctttgcgttgc tggatgttgc tggatgttgc tggatgttgc ttttgcgttgc 3060
aagtgttgc tggatgttgc tggatgttgc tggatgttgc ttttgcgttgc 3081

```

SEQ ID NO: 48 moltype = AA length = 1026
 FEATURE Location/Qualifiers
 source 1..1026
 mol_type = protein
 organism = Arabidopsis thaliana

SEQUENCE: 48

```

MEASAGLVAG SHNRNRELVVI HNHEEPKPLK NLDGQFCEIC GDQIGLTVEG DLFVACNECG 60
FPACRPCYEE ERREGTQNCP QCKTRYKRLR GSPRVEGDED EEDIDDIYE FNIEHEQDKH 120
KHSAEAMLYG KMSYGRGPED DENGFRPPVI AGGHSGEFPV GGGYGNCEHG LHKRVHPYPS 180
SEAGSEGGWR ERMDDWKLQH GNGLGEPDDE PEMGLIDEAR QPLSRKVPIA SSKINPYRMV 240
IVARLVLAV PLRYRLLNPV HDALGLWLTS VICEIWFAVS WILDQFPKWF PIERTYLDR 300
LSLRYEREGE PNMLAPVDVF VSTVDPKLEP PLVTSNTVLS ILAMDYPVEK ISCVYSDG 360
SMLTFESLSE TAEFARKWVP FCKKFSIEPR APEMYFTLK VDLQDKVHPT FVKERRAMKR 420
EYEEFKVRIN AQVAKASKVP LEGWIMQDGTL PWPGNNNTKDH PGMIQVFLGH SGGFDVEGHE 480
LPLRVYVSRE KRPGFQHHKK AGAMNALVRV AGVLTNAPFM LNLDCCDHVNV NSKAVERAMC 540
FLMDPQIGKK VCYVQFPQRDF DGIDTNDRYA NRNTVFFDIN MKGLDGIQGP VYVGTCVFK 600
RQALYGYEPP KGPKRPKMS CGCCPCFGRR RKNKKFSKND MNQGDAVALLGG AEGDKEHLM 660
EMNFETKFGQ SSIFVTSTLM EEEGVPPSSS PAVLLKEAIH VISCGYEDKT EWGTELGIYI 720
GSITTEDILTG FKMHCRGWRS IYCMKPRAAF KGSAPINSLD RLNVQLRWAL GSVEIFFSRH 780
SPLWYGYKGG KLKWLERFAY ANTTIYPFTS IPLLAYCILP AICLLTDKFI MPPISTFASL 840
FFISLFBMSII VTGILELRWS GVSIEEWWRN EQFWVIGGIIS AHLFAVVQGL LKILAGIDTN 900
FTVTSKATDD DDFGELYAFK WTTLLIPPTT VLIINIVGVV AGISDAINNG YQSWGPLFGK 960

```

-continued

LFFFSFWVIH LYPFLKGLMG RQNRTPTIVV IWSVLLASIF SLLWVRIDPF VLKTKGPDTS 1020
KCGINC 1026

SEQ ID NO: 49 moltype = DNA length = 3081
 FEATURE Location/Qualifiers
 source 1..3081
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 49
 atggaaactg gcgcgggtc tgcgcgggt ttcataacc gtaatgaact agtcgttac 60
 cacaaccatg aagagccaaa gctctgaag aatctatcg gacaatttc tgagatgt 120
 ggagatcaga ttggtttaac agtagaaagg gacctcttc tagcttgc aa 180
 ttccggcgt gtagacccctg ctatgatc gagagaagag aaggaacaca aaactgtcc 240
 cagtgttca ctcgttacaa ggttccaga ggaagccaa gatgggggg agatggaaac 300
 gaagaagata ttgtatgtatc tgatgtatc gttatatacg aacatgaaca agataagcat 360
 aacgttctg ctgaggctat gttttatggg aaaatgtatcg atggaaaggat tccttgaggat 420
 gatgagaatg ggagattccc acctgttata gctgttgc atatgtggaa atttccagg 480
 ggaggaggat atggatgtatc agaacatgtt cccttatacg gtgtgcacc atatccatca 540
 ttcgttgcgtt gggatggggg agatggccg gaaagatgtt atactggaa gtcggccat 600
 gggaaatcttgc ggccagaacc agatgtatc cctgtatgg gactgtatc cgaggcacgg 660
 cagccacttc cgccggaaatg tccatttgc tcaagcaaga tcaatccata tcggatgttc 720
 atcgttgcgtt gggttgcgtt cttagcgtt ttcttgcgtt ataggcttca gatccatgt 780
 catgtatgtc tgggattatg gtgcattttt gtgtatgtt aaatctgtt cgctgttct 840
 tggatcttgc atcgttccca caatgttgc cttatgttgc gtgagacccat tctatgttgc 900
 ctttccctca ggtacgagag agaagggtt ccaaataatgc ttggccctgtt agatgttctt 960
 gtcagtacgg tggaccattt gggaggactt cccctcgta catccaacac tggctgtca 1020
 atcttgcgtt tggacttcccc atgttgcgtt ttcttgcgtt atgttcttgc cgacgggtgt 1080
 tcaatgttca catttgcgtt cttcttgcgtt actgttgcgtt ttgcgttcaaaa atgggttccc 1140
 ttctgttcaaga aattcttccat agagccacgg gcaccggaga tggacttccat tctatgttgc 1200
 gattatcttgc aggacaaatg ccacccaaaca ttgttgcgtt aacgtcgacg catgaagaga 1260
 gaatatgagg agttcaatgg caggatcaac gtcgttgcgtt ccaggcccttca aagggttctt 1320
 cttaggttgcgtt ggtatgtca agtggaaaca ccgttgcgtt ccaggccatc aacgggacc 1380
 cccggatgtatc tccaaatgtt cttcttgcgtt acggggaggat ttgtatgtca agggcatgt 1440
 ctcttgcgtt ttgttgcgtt gtccctgcgtt aacgttgcgtt ttgttgcgtt ccacaagaaaa 1500
 gttggccgtt tgaatgttgcgtt ggttgcgtt ccaggccatc tccaaatgtt ctcttgcgtt 1560
 ctgtatgttgcgtt atcgttgcgtt ctatgttgcgtt aacaggccatc ccaggccatc tccaaatgtt 1620
 tttttgtatgg atcgttgcgtt tggaaagaag gtcgttgcgtt ttgttgcgtt ccacaaggttt 1680
 gatggatgttgcgtt acacaaacgcg tccgttgcgtt aacaggccatc cgttgcgtt ttgttgcgtt 1740
 atggaaagggttgcgtt tagatgttgcgtt ccaaggccatc gtttgcgtt gtacttgcgtt ttgttgcgtt 1800
 ccacaaggccatc tggatgttgcgtt tggatgttgcgtt aacaggccatc aacgttgcgtt ttgttgcgtt 1860
 tggatgttgcgtt gtcgttgcgtt ttggccgggg agaaaagaata agaaatttc caagaatgtac 1920
 atgaatgttgcgtt acgttgcgtt ctttgcgtt gcaaggccatc aacaggccatc ttgttgcgtt 1980
 gaaatgttgcgtt ttggaaacatc ttgttgcgtt tccatgttgcgtt ttgttgcgtt aac 2040
 gaaagggttgcgtt ttgttgcgtt gtcgttgcgtt ttgttgcgtt aac 2100
 gtcataatgttgcgtt gccgttgcgtt agacaagact ggttgcgtt ccaggccatc aacgttgcgtt ttgttgcgtt 2160
 ggttgcgttcatc caggatgttgcgtt ttgttgcgtt aac 2220
 atttactgttgcgtt tggatgttgcgtt ttgttgcgtt aac 2280
 tgcgttgcgtt gtcgttgcgtt ttgttgcgtt aac 2340
 aggttgcgtt aac 2400
 agtcttgcgtt ttgttgcgtt gtcgttgcgtt ttgttgcgtt aac 2460
 gtcataatgttgcgtt gccgttgcgtt agacaagact ggttgcgtt ccaggccatc aacgttgcgtt ttgttgcgtt 2520
 gtcgttgcgttcatc caggatgttgcgtt ttgttgcgtt aac 2580
 ttcttgcgttcatc cactgttgcgtt gtcgttgcgtt ttgttgcgtt aac 2640
 ggttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 2700
 ttcacccgttgcgttcatc caggatgttgcgtt ttgttgcgtt aac 2760
 tggacaaacatc tggatgttgcgtt ttgttgcgtt aac 2820
 gtcgttgcgttcatc cactgttgcgtt ttgttgcgtt aac 2880
 ctcttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 2940
 agacaaacatc tggatgttgcgtt ttgttgcgtt aac 3000
 tttttgtatgg atcgttgcgtt ttgttgcgtt aac 3060
 aatgttgcgtt ggttgcgtt ttgttgcgtt aac 3120
 aatgttgcgttcatc cactgttgcgtt ttgttgcgtt aac 3180
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 3240
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 3300
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 3360
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 3420
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 3480
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 3540
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 3600
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 3660
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 3720
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 3780
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 3840
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 3900
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 3960
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 4020
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 4080
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 4140
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 4200
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 4260
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 4320
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 4380
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 4440
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 4500
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 4560
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 4620
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 4680
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 4740
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 4800
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 4860
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 4920
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 4980
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 5040
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 5100
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 5160
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 5220
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 5280
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 5340
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 5400
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 5460
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 5520
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 5580
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 5640
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 5700
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 5760
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 5820
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 5880
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 5940
 aatgttgcgttcatc ctttgcgttcatc cactgttgcgtt ttgttgcgtt aac 6000

```

SEQ ID NO: 50          moltype = AA  length = 1026
FEATURE                  Location/Qualifiers
source                   1..1026
                        mol_type = protein
                        organism = synthetic construct

SEQUENCE: 50
MEASAGLVAG SHNRNELVVI HNHEEPKPLK NLDGQFCCEIC GDQIGLTVEG DLFVACNECG 60
FPACRPCYVE ERREGTQNCP QCKTRYKRLR GSPRVEGDED EEDIDDIEYE FNIEHEQDKH 120
KHSAAEAMLYG KMSYGRGPED DENGRFPVVI AGGHSGEFPV GGGYGNGEHGH LHKRVHPYPS 180
SEAGSEGGWR ERMDDWKLQH GNLGPEPDDD PEMGLIDEAR QPLSRKVPIA SSKINPYRMV 240
IVVRLVLILAV FLRYRQLLNH HDALGLWLTS VICEIWFAVS WILDQFPKWF PIERETYTLDR 300
LSLYEREGERE PNMLAPVDVF VSTVDPKLKEP PLVTSNTVLS ILAMDQPKWV IISCYVSDDGK 360
SMLTFESLSE TAEFKRWPV FCKKFSIESPR APEMYFTLKVH DYLQDKVPHET FVKERRAMRK 420
EYEEFKVRIN AQAKASKWPV LEGWIMQDGT PGMQIVQFLGH SGGFDFVEGHE 480

```

-continued

LPRLVYVSRE	KRPGFQHHKK	AGAMNALVRV	AGVLTNAPFM	LNLDCDHVNV	NSKAVREAMC	540
FLMDPQIYKK	VCYVQFPQRF	DGIDTNDRYA	NRNTVFFDI	MKGLDGIQGP	VYVTGCVFK	600
RQALYGYEPP	KGPKRPMIS	CGCCPCFGRR	RKNKKFSKND	MNGDVAALGG	AEGDKEHLM	660
EMNFEKTFQQ	SSIFVITSLM	EEGGVPPSSS	PAVLLKEIAH	VISCGYEDKT	EWGTELGIWY	720
GSITEDILTG	PKMHCRGWRS	IYCMPPRPAF	KGSAPINLSD	RLNQVLRWAL	GSVEIFFSRH	780
SPLWYGYKGG	KLKWLRLRFAY	ANTTIYPFITS	IPLLAYCILP	AICLLTDKFI	MPPPISTFAFL	840
FFISLFBMSII	VTGILELRWS	GVSIEEWWRN	EQFWVIGGIS	AHLFAVVQGL	LKILAGIDTN	900
FTVTSKATDD	DDFGELYAFK	WTLLLIPPT	VLIINIVGVV	AGISDAINNG	YQSWGPLFGK	960
LFFSFVWIVH	LYPFLKGLMG	RQNRTPTIVV	IWSVLLASIF	SLLWVRIDPF	VLKTKGPDT	1020
KCGINC						1026

SEQ ID NO: 51	moltype = DNA length = 3081					
FEATURE	Location/Qualifiers					
source	1..3081					
	mol_type = other DNA					
	organism = synthetic construct					
SEQUENCE: 51						
atggaagcta	gogccggctc	tgtcgccgtt	tctcataacc	gtaatgaact	agtctgtcatt	60
cacaaccatg	aagaggccaa	gcctctgaag	aatcttagatg	gacaatttcg	tgagatatgt	120
ggagatcaga	ttgggttaac	atgagaagg	gacctctcg	tagcttgcac	tgagtgttgt	180
tttccggcgt	gtagacctt	ctatgagta	gagagaagag	aaggaaacaca	aaactgtcct	240
caagtgtaa	ctcgatcacat	gctgttcaga	ggaaggccaa	gagtggaggg	agatgaaaac	300
gaagaagat	ttgtatgtat	tgatgtat	ttaatatatcg	aacatgaaaca	agataagcat	360
aaggatctt	ctgaggatct	gtctttatggg	aaaatggact	atggaagg	tcctgaggat	420
gatgagaatg	ggagattccc	acctgttata	gctgggtgtc	atagtggaga	atttccagtt	480
ggaggagggt	atggatatgg	agaacatggg	cttcataacg	gtgtgcaccc	atatccatca	540
tctgaagct	ggagtggagg	aggatggcg	gaaagaatgg	atgactggaa	gtctccagcat	600
ggaaatctt	ggccgaaacc	agatgtat	ccttggatgt	gactgtatcg	cgaggcacgg	660
cagccactt	cgccggaaagt	tccccattgc	tcaagcaaga	tcaatccata	tcggatgtc	720
atcgatcgat	ggcttgcgtat	tctagcgtat	tttctgcgtt	ataggctt	gaatccagt	780
catgtatgc	ttggattgtat	gctgtatgtc	gtgatctgt	aaatctggc	cgctgtct	840
tggattctt	atcagtcccc	caagtggtc	ccttataac	gtgagaccta	tctagatcgg	900
cttccctca	ggtaqagag	agaaggtaa	ccaaatatgc	ttgccccctgt	agatgtctt	960
gtcagtaacgg	ttggaccatt	gaaggagct	ccccctgtca	catccaacac	tgtgtgtca	1020
atcttggcc	ttggactaccc	agttgagaa	atcttgcgtat	atgtctgt	cgacggctgt	1080
tcaatgcctt	cattcgaatc	tctctggaa	actgtcgat	ttgcaagaaa	atgggttccc	1140
ttctgtaa	aatttccat	agacccacgg	gcacccggaa	tgtacttac	gttgaagatt	1200
gattatctcc	aggacaaagt	ccacccaaca	tttggtaagg	aacgtcgagc	catgaagaga	1260
gaatatgagg	agttcaaggt	caggatgtat	gtctcaagt	cgaggcctc	aaagggttcc	1320
ctagaaggat	ggatcatgc	agatggaa	ccgtggccaa	ggaacaaacac	caaggaccac	1380
cccggtatga	tccaaqtctt	ccttggccac	agcggaggat	ttgatgtcg	agggcatgag	1440
cttcctcggc	ttgtgtacgt	gtccccgtg	aagcgtctt	gttttcaaca	ccacaagaaa	1500
gtctggcoca	taatgtccct	gggtcgatgt	gcaggcgatc	tcacaaatgc	tcctttcatg	1560
ctgaacttgg	actgtgtatca	ctatgtaa	aacagcaagg	ccgtggagg	agcaatgtgt	1620
tttttgcgtt	atccctcagat	tggaaaaag	gtctgtat	tccatgtccc	tcaaaggtt	1680
gatggcattt	acacaaacga	tcgttacgc	aacagaaaca	caagtctt	tgatataat	1740
atgaaaggctc	tagatgtat	ccaaatgtc	gtttacgtt	gtactgtt	tgttttcaaa	1800
cgacaaggctc	tgtatgtt	tgaaccacca	aagggttca	aacgtccaa	gtgataagc	1860
tgtgggttt	gtccctgtt	tggggccgg	agaaaaata	agaaattttc	caagaatgtac	1920
atgaatgtt	acgttagcgc	cettggagga	gcagaagggt	ataaggagca	tttgcgtct	1980
gaaaatgcct	ttggaaaaacc	atttggggca	tcatccatct	ttgttaaccc	aactttgtat	2040
gaagaaggct	gttttccctcc	gtcatcaat	cttcgtatgc	tccttaaaa	ggcaatccat	2100
gtcataact	ggggatgtat	agacaaagact	qagtggggaa	ctgagctgg	ttgatgtat	2160
ggctctatca	cagaggat	tttgacggg	ttcaagatgc	attgcgtgg	atggaggct	2220
atttactgc	tcgcctaaag	gcctgcattc	aaagggttca	ctccatattaa	tctatcagac	2280
aggttaacc	agggtttgc	ttggggactt	ggatcggtt	agatatttt	cagccggcac	2340
agtccctctt	ggatgtatca	caaaggaggc	aaactcaat	ggcttgcgg	tttgcttat	2400
gccaacacaa	caatctaccc	cttcacatct	ataccactt	ttgcctact	tatcttcca	2460
gcccattgtc	tccttactc	caaatttc	atgccaccc	taagcacatt	tgctgtatc	2520
ttcttcatt	cactgtttat	gtcgtat	gtaacgggaa	tcttggatt	gatggggc	2580
ggagttacga	ttggaggat	gtggagaaac	gcaatattc	gggttattgg	aggaatctca	2640
gtctcatctt	ttgcgggtt	ccaaatgtc	ctcaaaatct	tagcaggcat	tgacacaaac	2700
ttcacccgtca	catcaaaggc	aacagatgt	gtgactt	gagaactt	cgcattcaaa	2760
tggacaacac	tgctgtatcc	tccaaact	gtcttaatca	taaaccatgt	ttggcggtt	2820
gcaggccat	cagatgtat	taacaatgt	tatcgtat	ggggacctt	atttggtaaa	2880
cttcttctt	ctttttgggt	cattgttcat	cttcacatct	tcctcaagg	tctgtatgg	2940
agacagaaca	gaacaccaac	cattgtgggt	atttggtca	tgttattggc	atctatctt	3000
tctttgtttt	gggttaagaat	tgatccctt	gtgtcaaga	ccaaaggacc	tgacacttcc	3060
aagtgtggca	tcaactgtct	a				3081

SEQ ID NO: 52	moltype = AA length = 1026
FEATURE	Location/Qualifiers
source	1..1026
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 52	

-continued

MEASAGLVAG	SHNRNELVVI	HNHEEPKPLK	NLDGQFCEIC	GDQIGLTVEG	DLFVACNECG	60
FPAACRCPYEY	ERREGTONCP	QCKTRYKRRLR	GSPRVEGDEN	EEDIDIEYE	FNIEHQDKH	120
KHSAEAMLYG	KMSYGRGPED	DENGRFPPVI	AGGHSGEFPV	GGGYGNGEHG	LHKRVHPYPS	180
SEAGSEGGWR	ERMDDWKLQH	GNLGPEDPDDD	PEMGLIDEAR	QPLSRKPPIA	SSKINPYRMV	240
IVARLVLILAV	FLRYRLLNPV	HDALGLWLTS	VICEIWFAVS	WILDQFPKWF	PIERE TYLDR	300
LSLRYEREGE	PNMLAPVDV	VSTVDPKKEP	PLVTSNTVLS	ILAMDYPVEK	ISCVYSDG	360
SMLTFESLSE	TAEFARKWVP	FCKKSIEPR	APEMYFTLKV	DYLQDKVHPT	FVKERRAMKR	420
EYEEFKVRIN	AQVAKASKVP	LEGWIMQDGT	PWPGNNTKDH	PGMIQVFLGH	SGGFDEVGHE	480
LPRLVYVSRE	KRPQFQHHKK	AGAMNALVRV	AGVLTNAPFM	LNLDCDHVYN	NSKAVREAMC	540
FLMDPQIGKK	VCYVQFPQRF	DGIDTNDRYA	NRNTVFFDI	MKGLDGIQGQ	VYVTGCVFK	600
RQALYGYEPP	KGPKRPMIS	CCCPCKGRR	RKNKKFSKND	MNGDVAALGG	AEGDKEHLM	660
EMNFEKTFQQ	SSI	FVITSTLM	EEGGVPPSSS	PAVLLKBAIH	VISCGYEDK	720
GSITEDILTG	FKMHCRGWR	IYCMKPRAF	KGSAPINLSD	RLNQVRLWAL	GSVEIFFSRH	780
SPLWYGYKGG	KLKWLERFAV	ANNTTYPFM	IPLLAYCILP	AICLLTDKF1	MPPPISTFA	840
FFISLFMSII	VTGILELRWS	GVSIEEWWRN	EQFWVIGGIS	AHLFAVVQGL	LKLAGIDTN	900
FTVTSKATDD	DDFGELEYAFF	WTLLIPPTT	VLI	INIIVGVV	AGISDAING	YQSWGPLFGK
LFFSFWVIVH	LYPFLKGLMG	RQNRTPTIVV	IWSVLLASIF	SLLWVRIDPF	VLTKGPDT	1020
KGINC						1026

SEQ ID NO: 53	moltype = DNA	length = 2958
FEATURE	Location/Qualifiers	
source	1..2958	
	mol_type = other DNA	
	organism = <i>Arabidopsis thaliana</i>	
SEQUENCE: 53		
atgatggagt ctaggtctcc catctgcac ac ttgtggtg aagagatttg tggaaaatca	60	
aacggaggt ttttgtggc ttgtcatgag tttgtttcc cgatctgc aa gcttgc tt	120	
gagttatgaaat tc aaaaaggc tcgaagaatt tgcttcgcgt gcggcaatcc ttacgatgag	180	
aatgtgtttt atgatgttgc gacaagaca tc taaaactc aatccatgtt tc caacacag	240	
accaataaca ct ttcaggc ttcaagggtt catgcatac atataagtac agtctcaaca	300	
atagacatgt aactgtatgc tgaatatggc aatccaattt ggaagaacag atggagagc	360	
tggaaagaca aaaaaggc aagagc aagaaaga aatgtccaaa agcaacaaaa	420	
gctgaacaac atgaggctca gattcctacc caacagcaca tggaaagatc gcccacccaa	480	
acagaatctg gtgctcaga tttgtttcg gttgtgttcc ctatccaa gacaaaatc	540	
acttcatata gttatgttcat catcgcgg ttgatcatc tggctctgtt tttaactatc	600	
cgtatcacgc atcctgtcga tagcgttac ggtttatggc taacatgtt gatatgtgag	660	
atttggttt ctgtttctt ggtgttggat cagttccctt aatggctcc tattaacccg	720	
gaaaacttaca tccgaccgtt atccgcaaga ttcgaaagag aaggcgcgca atcacagtt	780	
gcagctgtt tagttttttt tagtgcgtt gatccattaa aggaggccacc ttgtataact	840	
gcaaacacgg ttctttcgat cctcgcgtt gattatccgg tggataaagt ctcttgcet	900	
gtatctgtt atgggtctgc aatgtttcg tttgagttttt tgggttgcgac agcaatgtt	960	
gcttaggaat gggtaacctt ctgcacaaag tactccatcg agccacgc tcccgagttt	1020	
tacttctccg taaaatcga ttacttctt gataaagtcc aacccctttt ttgtaaaagaa	1080	
cgttagagcca tggaaaagaga ttatggat tttaaaataataa gaatgtatgc tttagtcgc	1140	
aaggctcaa agacaccaga agaaggatgg acaatgc aatggaaatc ttggccccgg	1200	
aacaacactc gtgaccatcc cgggatgtt cagggttttc ttggatatacg cggtgtcgc	1260	
gacatttggat gaaatgtact tccaaatgtt gtttacgttct cttagagaaa gagacttgg	1320	
tatcagcatc aaaaaggc cggggcagag aacgcattgg tgagggtgtc tgccgttta	1380	
acgaatgtc cattcatttc taacccgtt tttgtatcaact acgtcaacaa tagcaagcc	1440	
gtgcgtttagt caatgtgtt tttatggat cctgttgcgtt gtcaagacgt ttgcgttgc	1500	
cagtccccc agagatttg tggatcgtac aagagtgtc gatatgtaa ccgcacactt	1560	
gttttcttcg atgttataat gagagggtt gatgttccat ttatgttgc	1620	
acaggtactt ctttttagaaatc acaacactt ttcgttacata gtcacccctt aaaacccagg	1680	
atttttacccg aatcttcatc atcgtctgt tttgttgcata ccaagaagaa acaacactaa	1740	
gtcccttcg agatttacaa agatgtcaag cgagaagaa ttgtatgtc aatcttttaat	1800	
cttggggacc tagacatcc cgtatgtatc gacatgtca tttgttgcgtt acaaaacagc	1860	
tttggataaa ctgtttgtt ctatcgttgcgtt tttatgttgcgtt ctactcttat ggagaatggc	1920	
gggtgttccg actctgtaaa cccgtcaacg ctatcaaaag aagcttataa tttgttgcgtt	1980	
tgtggatcgtt aagaaaaac tttatggat tttatggat tttatggat tttatggat	2040	
accgaagaca ttttgcgtt tttatggat tttatggat tttatggat tttatggat	2100	
atggccatccaa gaccgttcc tttatggat tttatggat tttatggat tttatggat	2160	
cagggttccat tttatggat tttatggat tttatggat tttatggat tttatggat	2220	
tggtaatgtt tttatggat tttatggat tttatggat tttatggat tttatggat	2280	
tttgcgttacc tttatggat tttatggat tttatggat tttatggat tttatggat	2340	
tttgcgttacc tttatggat tttatggat tttatggat tttatggat tttatggat	2400	
gggtgttccat tttatggat tttatggat tttatggat tttatggat tttatggat	2460	
atgtttatggat tttatggat tttatggat tttatggat tttatggat tttatggat	2520	
tttgcgttacc tttatggat tttatggat tttatggat tttatggat tttatggat	2580	
acatcgaaaa cccgtcaacg tttatggat tttatggat tttatggat tttatggat	2640	
cttgcgttacc tttatggat tttatggat tttatggat tttatggat tttatggat	2700	
tttgcgttacc tttatggat tttatggat tttatggat tttatggat tttatggat	2760	
ggcttttggg tgatttttca tttatggat tttatggat tttatggat tttatggat	2820	
agaacacacta tttatggat tttatggat tttatggat tttatggat tttatggat	2880	
tgggttgcgtt tttatggat tttatggat tttatggat tttatggat tttatggat	2940	
cttttgcgttacc tttatggat tttatggat tttatggat tttatggat tttatggat	2958	

-continued

-continued

gccttttggg tgattttca tctttatcca ttccctcaaag gtcttatggg aagacaaaac	2820
agaacacacta ctattgttat tctctggct atcttgctt cttctgtgtt ttcaactt	2880
tgggttcgta tcaatccctt cgctccaaa accgatacga ctcccttctc tctgaactgt	2940
cttttgcata attgctaa	2958
 SEQ ID NO: 56	
FEATURE	moltype = AA length = 985
source	Location/Qualifiers 1..985
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 56	
MMESRSPICN TCGEEIVGKS NGEFFVACHE CSFPICKACL EYEFKEGRRI CLRCGNPYDE	60
NVFDVETKTKT SKTQSIVPTQ TNNTSQDSGI HARHISTVST IDSELNDEYG NPIWKNRVES	120
WKDKKKDKSK KKKKDPKATK AEQHEAQIPT QOHMEDTPPN TESGATDVL VVISIPRTKI	180
TSYRIVIIMR LIILALFFNY RITHPVDSAY GLWLTSVICE IWFAVSWLQ QFPKWSPINR	240
ETYIDRLSAR FEREGBSQL AAVDFVFVSTV DPLKEPPLIT ANTVLSILAL DYPVDFKVSCY	300
VSDDGAAMLS FESLVETADF ARKWFVFCKK YSIEPRPEAF YFSLKIDYL DKVQPSFVKE	360
RRAKMDRKYEE KAQKTMNALVA TMQDGTSWPNG NNTRDHPGMI QVFLGYSGAR	420
DIEGNELPLR VVVSREKRPG YQHHKKAGAE NALVRVSAVL TNAPFILNLD CDHYVNNNSKA	480
VREAMCFLMD PVVGQDVCFV QFPQRFDGID KSDRYANRNI VFFDVNMRLG DGIQGPVYVG	540
TGTVFRRQAL YGYSPPSKPR ILPQSSSSC CCLTKKKQPO DPSEIYKDAK REELDAIFN	600
LGDLNDNDEY DRSM LISQTS FEKTFGLSTV FIESTLMENG GVPDSVNPNST LIKEAIHVIS	660
CGYEKTEWG KEIGWIYGSI TEDILTGFKM HCRGWRSSIYC MPLRPAFKGS APINLSDRLH	720
QVLRWALGSV EIFLRSRHCPL WYGCSSGRKL LLQRLAYINT IVYPFTSLPL VAYCTLPAIC	780
LITGKFIIPT LSNLASMLFL GLFISIILTS VLELRWSGVS IEDLWRNEQF WVIGGVS AHL	840
FAVFOGFLKM LAGLDTNFTV TSKTADDLEF GELYIVKWTI LLIPPTSLLI INLGVVAGF	900
SDALNKGYEA WGPLFGKVFF AFWVILHLYP FLKGLMGRCQN RTPTIVILWS ILLASVFSLV	960
WVRINPFVSK TDTTSLSLNC LLIDC	985

What is claimed is:

1. A mutant gene having the sequence of SEQ ID NOS: 3, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 49, 51, and 55, or a gene of 80% or more sequence identity having said mutated gene, respectively, wherein said gene confers resistance to a cellulose synthase inhibitor.

2. A cell or a plant having the mutant gene of claim 1.

3. The mutant gene according to claim 1, wherein said gene expresses a protein having the sequence of SEQ ID NOS: 4, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 50, 52, and 56, or a polypeptide of 80% or more sequence identity having said mutated conserved amino acid residue, respectively, wherein said protein confers resistance to a cellulose synthase inhibitor.

4. (canceled)

5. An isolated and cloned mutant gene encoding a cellulose synthase comprising a cellulose synthase gene with a specified nucleic acid sequence, wherein the nucleic acid sequence comprises the nucleotide sequence of SEQ ID NOS: 3, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 49, 51, and 55, or a gene of 80% or more sequence identity having said mutated gene, respectively.

6. A cell or a plant having the gene of claim 5.

7. (canceled)

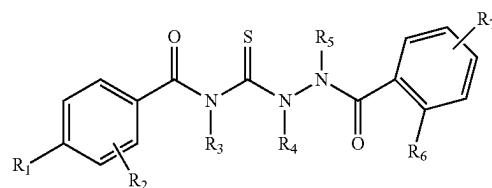
8. The cell or a plant according to claim 6, wherein said mutant gene expresses a protein having the sequence of SEQ ID NOS: 4, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 50, 52, and 56, or a polypeptide of 90% or more sequence identity having said mutated conserved amino acid residue, respectively.

9. A method for weed control of a crop comprising cloning and expressing a mutant gene having the sequence

of SEQ ID NOS: 3, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 49, 51, and 55, or a gene of 90% or more sequence identity having said mutated gene, respectively, in said crop, and applying an inhibitor of cellulose synthase to the field of said crop.

10. The method for weed control of a crop according to claim 9, wherein said mutant gene expresses a protein having the sequence of SEQ ID NOS: 4, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 50, 52, and 56, or a polypeptide of 90% or more sequence identity having said mutated conserved amino acid residue, respectively, in said crop, and applying an inhibitor of cellulose synthase to the field of said crop.

11. The method for weed control of a crop according to claim 9, wherein said inhibitor of cellulose synthase is a compound of the following general formula:



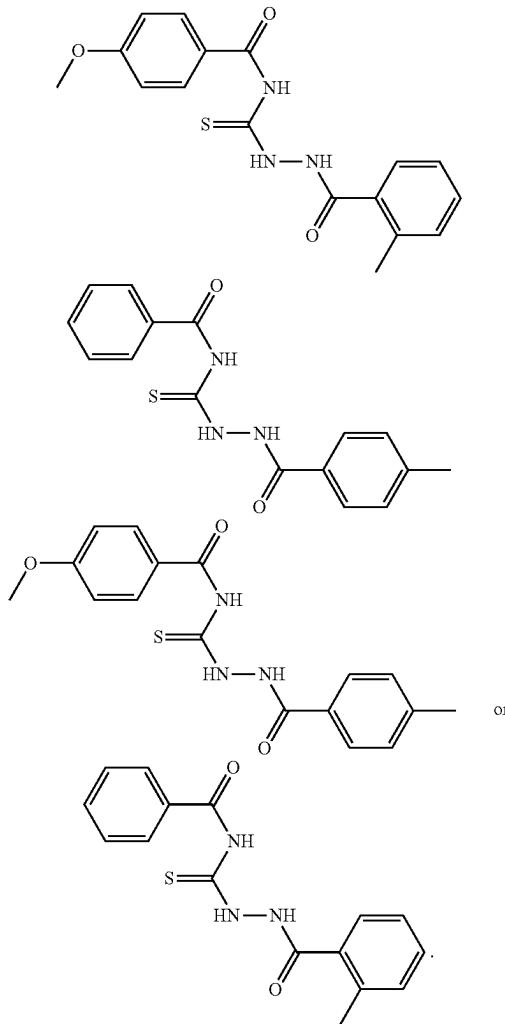
or an acceptable salt thereof, wherein

R₁ is hydrogen, a halo, an alkyl, alkenyl, or alkyloxy;

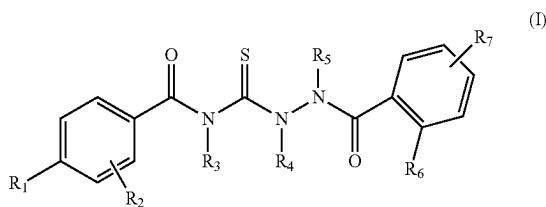
R₂ and R₇ represent four individual substituents independently selected from the group comprising hydrogen, a halo, an alkyl, alkenyl, and alkyloxy; and

R₃, R₄, R₅ and R₆ are independently hydrogen or an alkyl.

12. The method for weed control of a crop according to claim **11**, wherein the compound is



13. A cellulose synthase inhibitor having the following general formula (I)

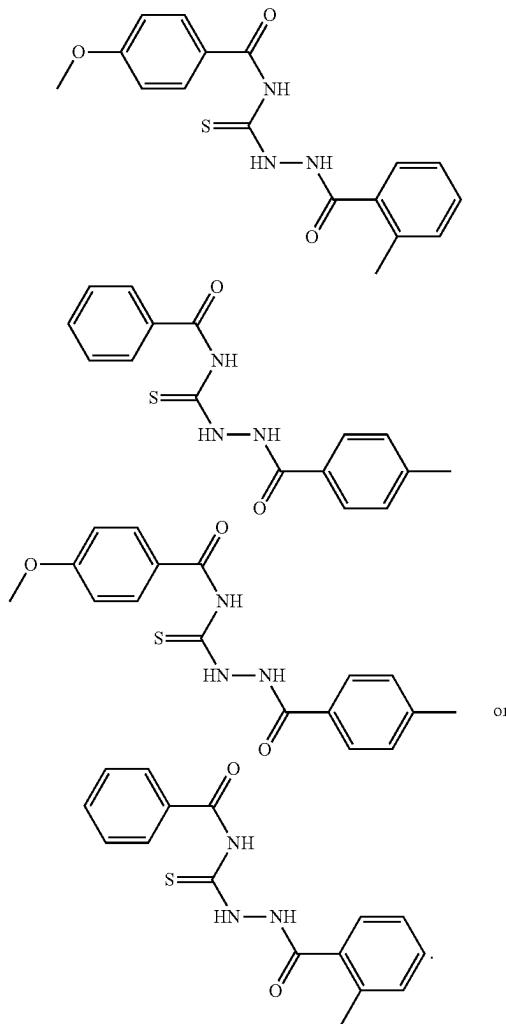


or an acceptable salt thereof, wherein

R₁ is hydrogen, a halo, an alkyl, alkenyl, or alkoxy; R₂ and R₇ represent four individual substituents independently selected from the group comprising hydrogen, a halo, an alkyl, alkenyl, and alkoxy; and

R₃, R₄, R₅ and R₆ are independently hydrogen or an alkyl.

14. The cellulose synthase inhibitor of claim **13**, wherein the inhibitor is

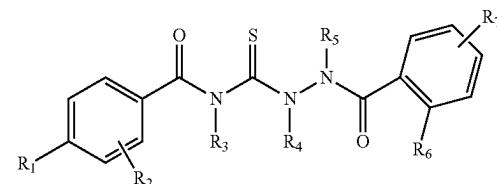


15. The cellulose synthase inhibitor of claim **13**, wherein said inhibitor is a herbicide.

16. (canceled)

17. (canceled)

18. A method for weed control of a plant, comprising the step of applying an effective amount of an herbicide having the following general formula (I)



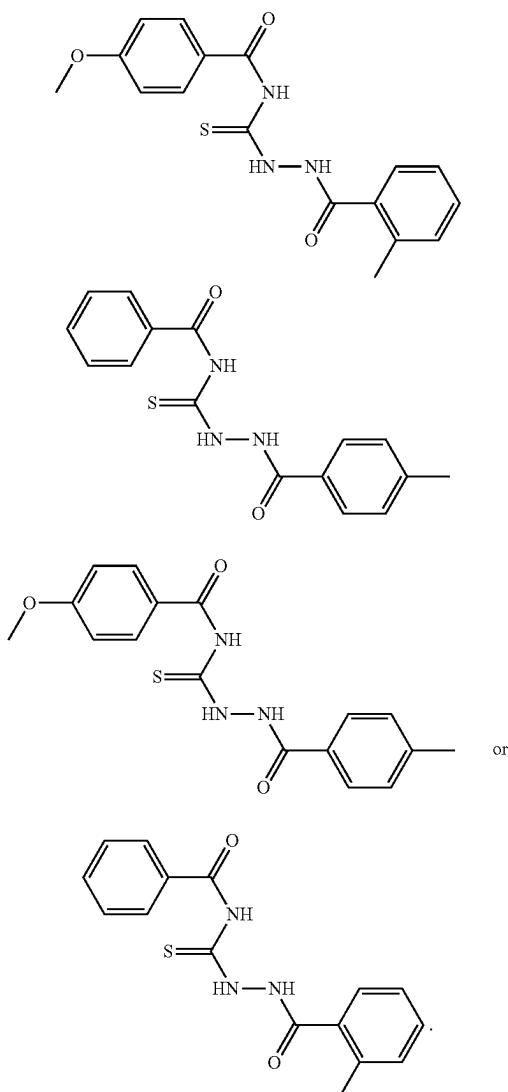
or an acceptable salt thereof, to the field of said plant, wherein

R₁ is hydrogen, a halo, an alkyl, alkenyl, or alkyloxy;

R₂ and R₇ represent four individual substituents independently selected from the group comprising hydrogen, a halo, an alkyl, alkenyl, and alkyloxy;

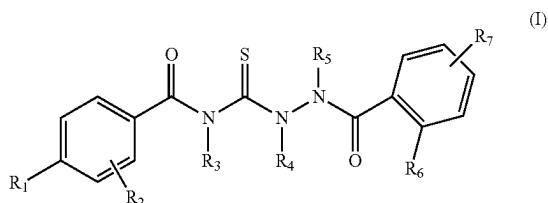
and R₃, R₄, R₅ and R₆ are independently hydrogen or an alkyl.

19. The method of claim 18, wherein said herbicide is



20. (canceled)

21. A method for treating a bacterial infection of a patient comprising administrating a therapeutically effective amount of a compound of formula (I), together with one or more diluents, excipients or carriers,



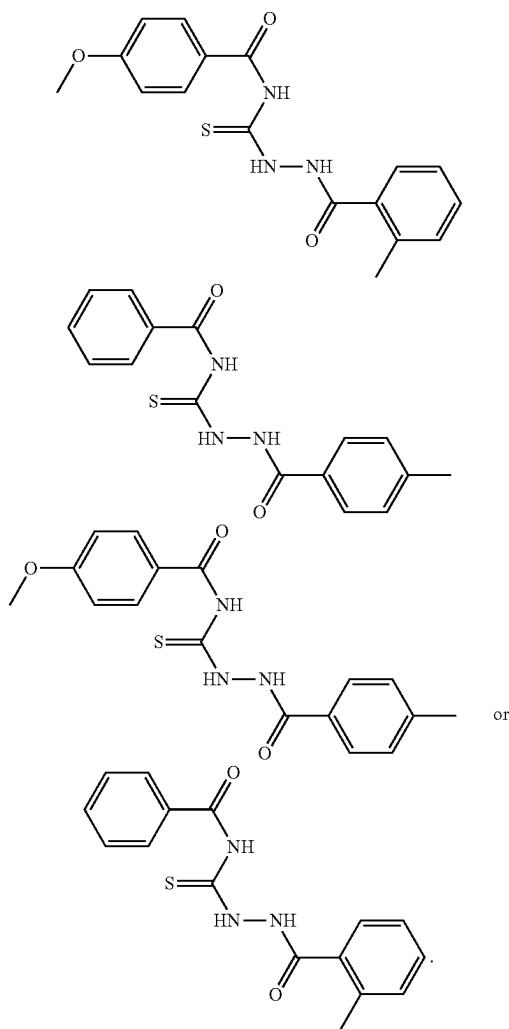
or a pharmaceutically acceptable salt thereof, to the patient in need of relief from said infection, wherein

R₁ is hydrogen, a halo, an alkyl, alkenyl, or alkyloxy;

R₂ and R₇ represent four individual substituents independently selected from the group comprising hydrogen, a halo, an alkyl, alkenyl, and alkyloxy; and

R₃, R₄, R₅ and R₆ are independently hydrogen or an alkyl.

22. The method of claim 21, wherein the compound is



23.-29. (canceled)

* * * * *