



US012391948B2

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

(10) **Patent No.:** US 12,391,948 B2
(45) **Date of Patent:** Aug. 19, 2025

(54) **RNA-REGULATED FUSION PROTEINS AND METHODS OF THEIR USE**(71) Applicant: **CORNELL UNIVERSITY**, Ithaca, NY (US)(72) Inventors: **Samie R. Jaffrey**, New York, NY (US); **Jiahui Wu**, New York, NY (US)(73) Assignee: **CORNELL UNIVERSITY**, Ithaca, NY (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 849 days.

(21) Appl. No.: **17/637,940**(22) PCT Filed: **Aug. 31, 2020**(86) PCT No.: **PCT/US2020/048781**

§ 371 (c)(1),

(2) Date: **Feb. 24, 2022**(87) PCT Pub. No.: **WO2021/042050**PCT Pub. Date: **Mar. 4, 2021**(65) **Prior Publication Data**

US 2022/0290161 A1 Sep. 15, 2022

Related U.S. Application Data

(60) Provisional application No. 62/894,651, filed on Aug. 30, 2019.

(51) **Int. Cl.**

C12N 15/62 (2006.01)
A61K 31/7105 (2006.01)
C07K 19/00 (2006.01)
C12N 5/071 (2010.01)
C12N 9/02 (2006.01)
C12N 15/115 (2010.01)
C12N 15/52 (2006.01)
C12N 15/85 (2006.01)
C12Q 1/6816 (2018.01)

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

CPC **C12N 15/62** (2013.01); **C12N 5/0602** (2013.01); **C12N 15/115** (2013.01); **C12N 15/52** (2013.01); **C12N 15/85** (2013.01); **C12Q 1/6816** (2013.01); **C12Y 111/01011** (2013.01); **C12Y 113/12013** (2013.01); **C12Y 201/01043** (2013.01); **C12Y 603/0401** (2013.01); **C07K 2319/60** (2013.01); **C07K 2319/61** (2013.01); **C07K 2319/85** (2013.01); **C12N 2310/16** (2013.01); **C12N 2800/107** (2013.01)

(58) **Field of Classification Search**

None

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

9,388,418 B2 2008/0300187 A1*	7/2016 12/2008	Rossi et al. Sauvageau	C07K 14/47 435/375
2012/0213737 A1 2014/0255361 A1 2019/0185934 A1	8/2012 9/2014 6/2019	Zhu et al. Wandless et al. Jaffrey et al.	

OTHER PUBLICATIONS

International Search Report and Written Opinion for corresponding International Application No. PCT/US2020/048781 (mailed Feb. 8, 2021).

Wu et al., "Live Imaging of mRNA using RNA-Stabilized Fluorogenic Proteins," *Nat. Method.* 16(9):862-865 (2019).

Song et al., "Imaging RNA Polymerase III Transcription Using a Photostable RNA-Fluorophore Complex," *Nat. Chem. Biol.* 13(11):1187-1194 (2017).

Extended European Search Report for European Application No. 20950485.1 (mailed Mar. 4, 2025).

Tok et al., "A Comparative Binding Study of Modified Bovine Immunodeficiency Virus TAR RNA Against its TAT Peptide," *Bioorg. Med. Chem. Lett.* 14(24):6101-6105 (2004).

Bonger et al., "Small-Molecule Displacement of a Cryptic Degron Causes Conditional Protein Degradation," *Nat. Chem. Biol.* 7(8):531-537 (2011).

Battiste et al., "α Helix-RNA Major Groove Recognition in an HIV-1 Rev Peptide-RRE RNA Complex," *Science* 273:1547-1551 (1996).

Legault et al., "NMR Structure of the Bacteriophage λ N Peptide/boxB RNA Complex: Recognition of a GNRA Fold by an Arginine-Rich Motif," *Cell* 93(2):289-299 (1998).

Yamamoto et al., "A Novel RNA Motif that Binds Efficiently and Specifically to the Tat Protein of HIV and Inhibits the Trans-Activation by Tat of Transcription in vitro and in vivo," *Genes to Cells* 5:371-388 (2000).

Rakhit et al., "Evaluation of FKBP and DHFR based Destabilizing Domains in *Saccharomyces cerevisiae*," *Bioorg. Med. Chem. Lett.* 21(17):4965-4968 (2011) [Author manuscript].

Iwamoto et al., "A General Chemical Method to Regulate Protein Stability in the Mammalian Central Nervous System," *Chem. Biol.* 17(9):981-988 (2010) [Author Manuscript].

Athanassiou et al., "Structural Mimicry of Retroviral Tat Proteins by Constrained β-Hairpin Peptidomimetics: Ligands with High Affinity and Selectivity for Viral TAR RNA Regulatory Elements," *J. Am. Chem. Soc.* 126(22):6906-6913 (2004).

(Continued)

Primary Examiner — Manjunath N Rao

Assistant Examiner — Jae W Lee

(74) *Attorney, Agent, or Firm* — Troutman Pepper Locke LLP (Rochester)

ABSTRACT

The present disclosure is directed to RNA-regulated fusion proteins comprising a protein of interest and an RNA-regulated destabilization domain. Also disclosed are RNA aptamers that bind specifically to a RNA-regulated destabilization domain. Nucleic acid molecules encoding the RNA-regulated fusion proteins and RNA aptamers and methods of use thereof are also disclosed.

10 Claims, 49 Drawing Sheets

Specification includes a Sequence Listing.

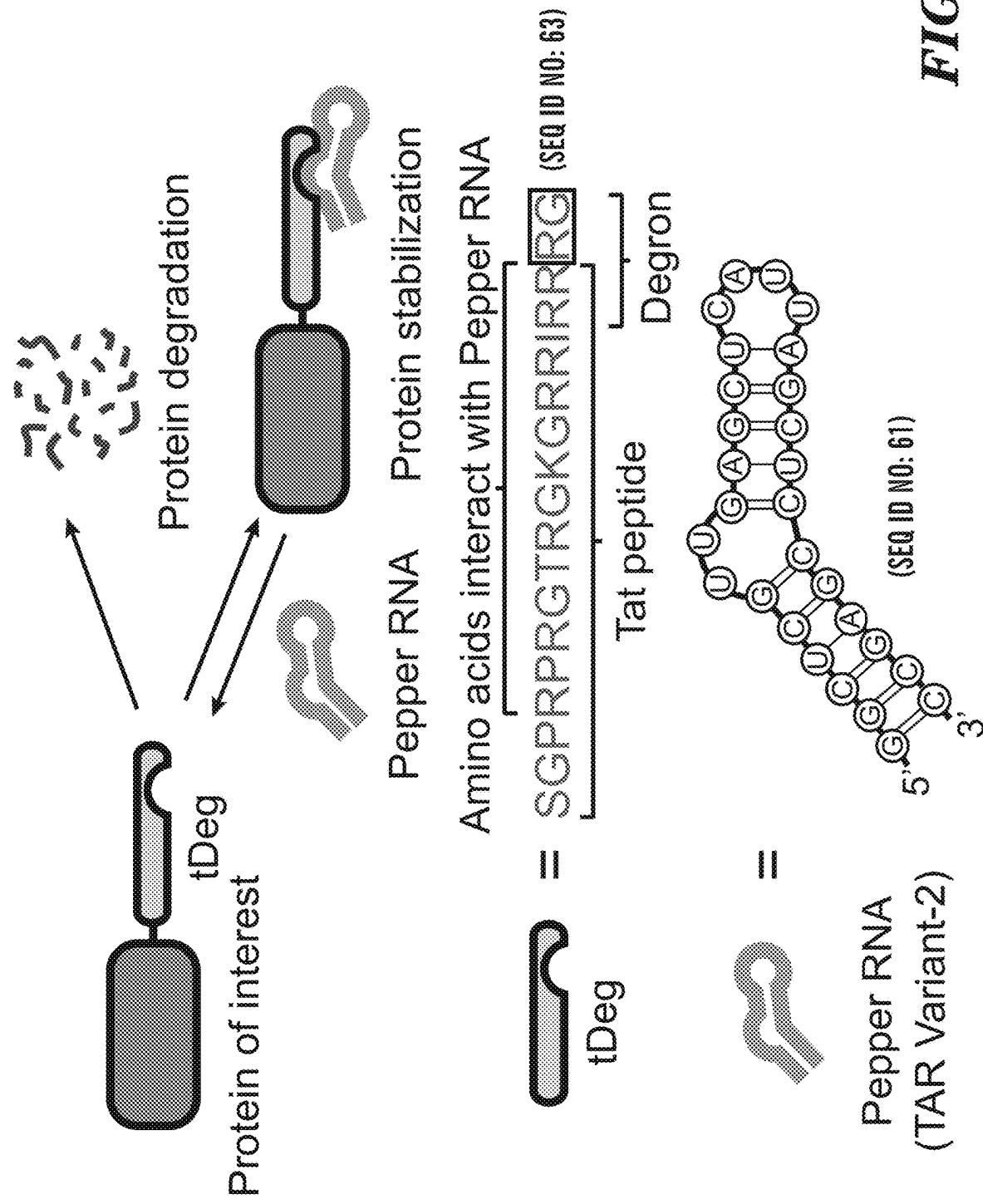
(56)

References Cited

OTHER PUBLICATIONS

- Chen et al., "A Peptide Interaction in the Major Groove of RNA Resembles Protein Interactions in the Minor Groove of DNA," Proc. Natl. Acad. Sci. USA 92:5077-5081 (1995).
- Chen and Frankel, "An RNA-Binding Peptide from Bovine Immunodeficiency Virus Tat Protein Recognizes an Unusual RNA Structure," Biochemistry 33:2708-2715 (1994).
- Koren et al., "The Eukaryotic Proteome is Shaped by E3 Ubiquitin Ligases Targeting C-Terminal Degrons," Cell 173:1622-1635 (2018).
- Smith et al., "Altering the Context of an RNA Bulge Switches the Binding Specificities of Two Viral Tat Proteins," Biochemistry 37:10808-10814 (1998).

* cited by examiner



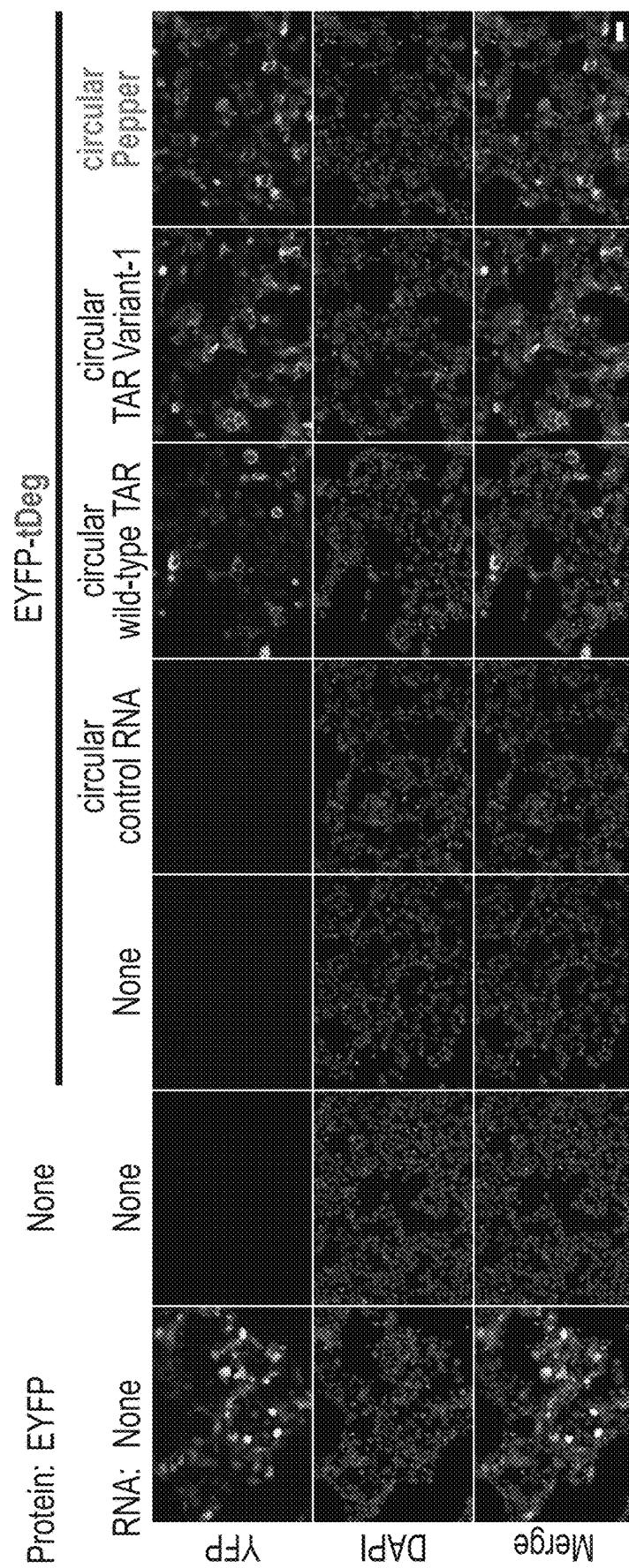
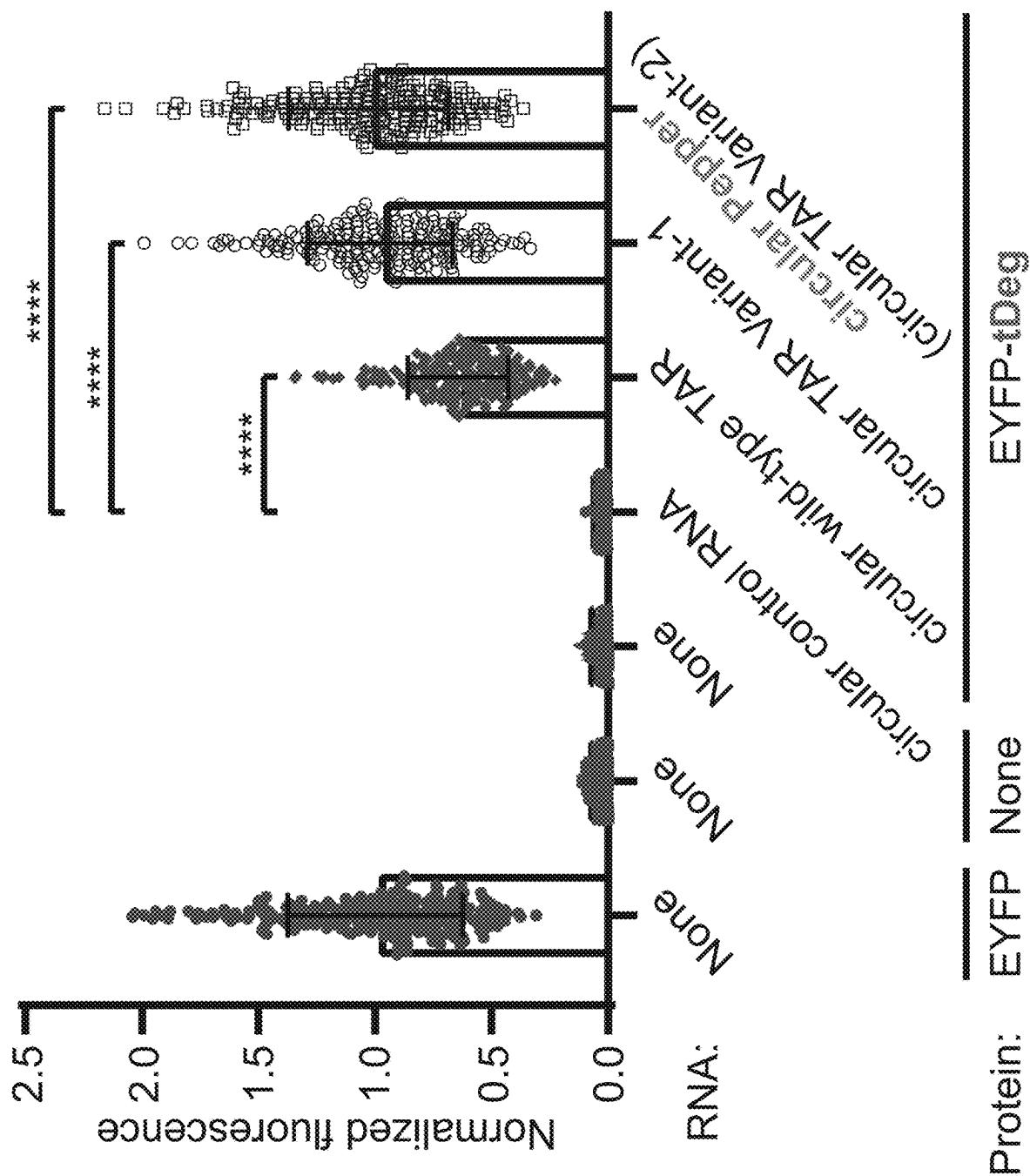
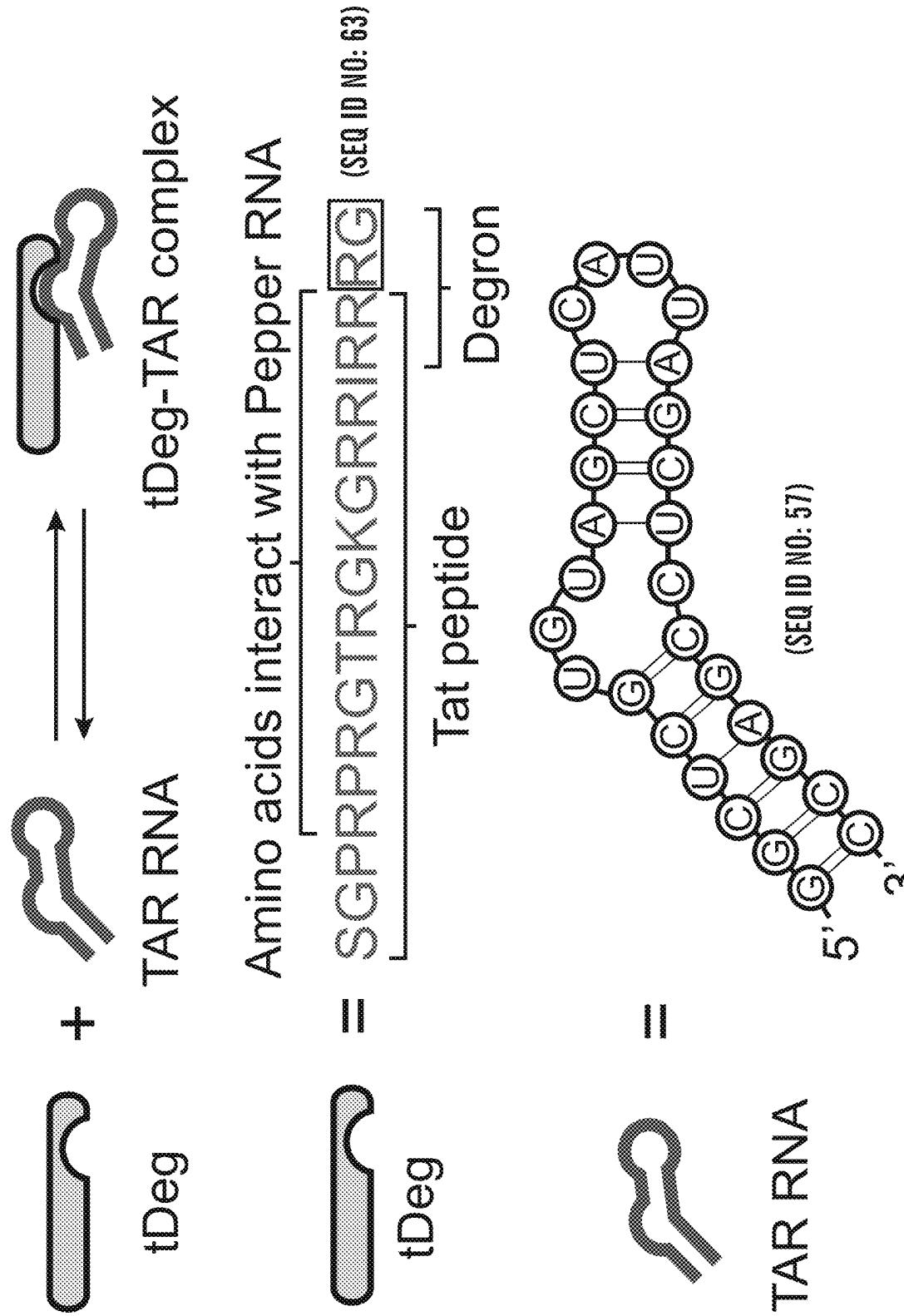


FIG. 1B

FIG. 1C





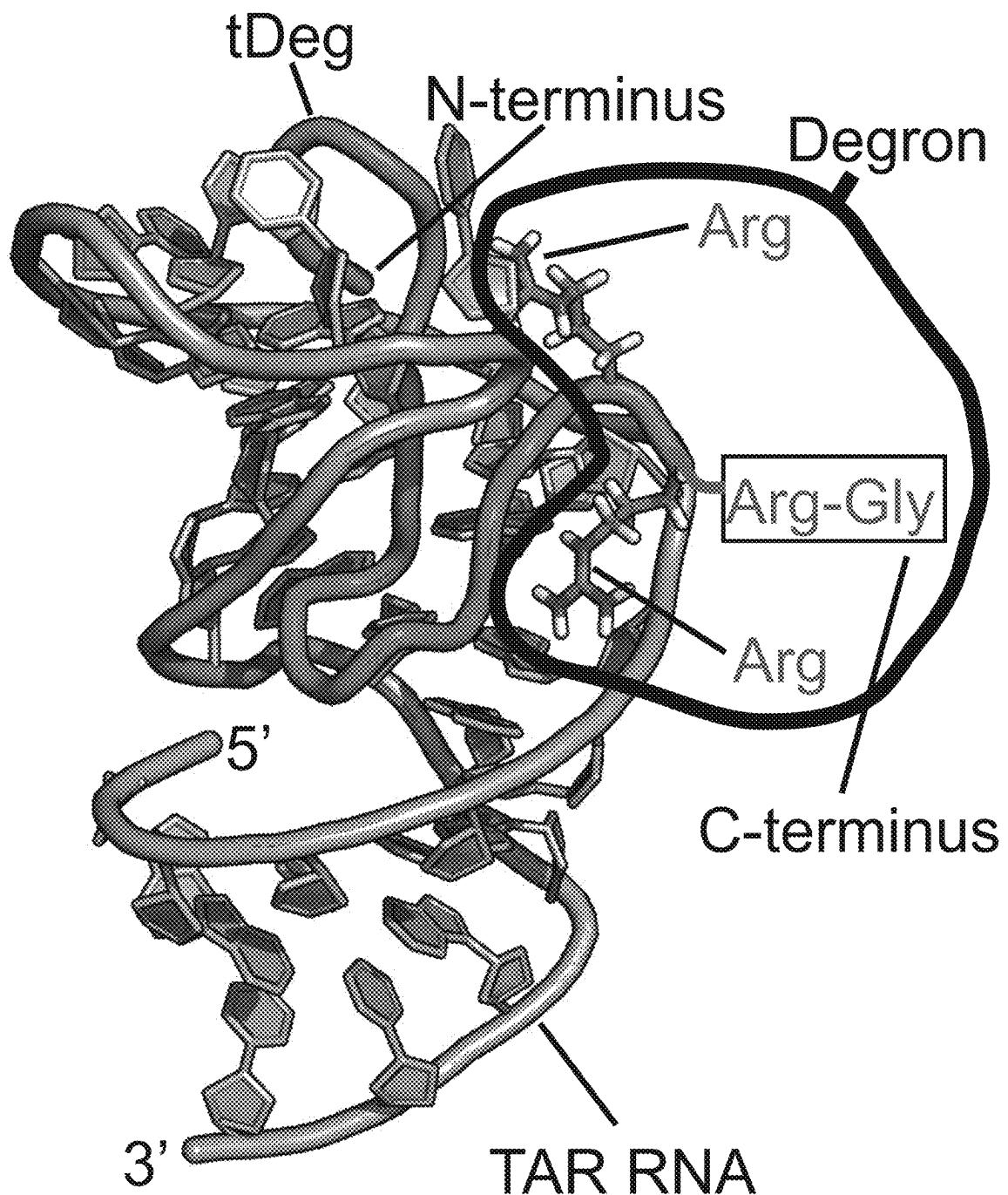
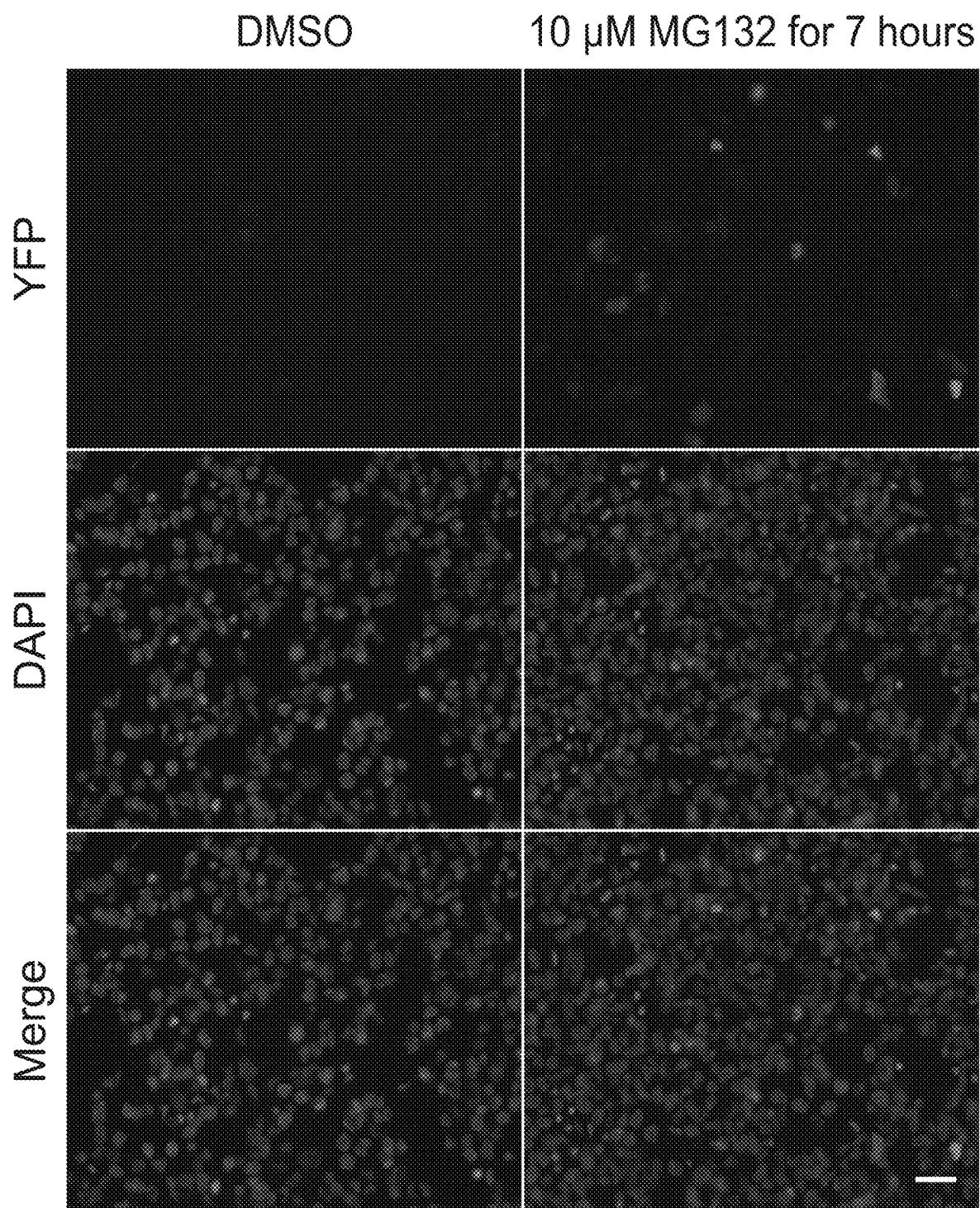


FIG. 2B

EYFP-tDeg

**FIG. 3A**

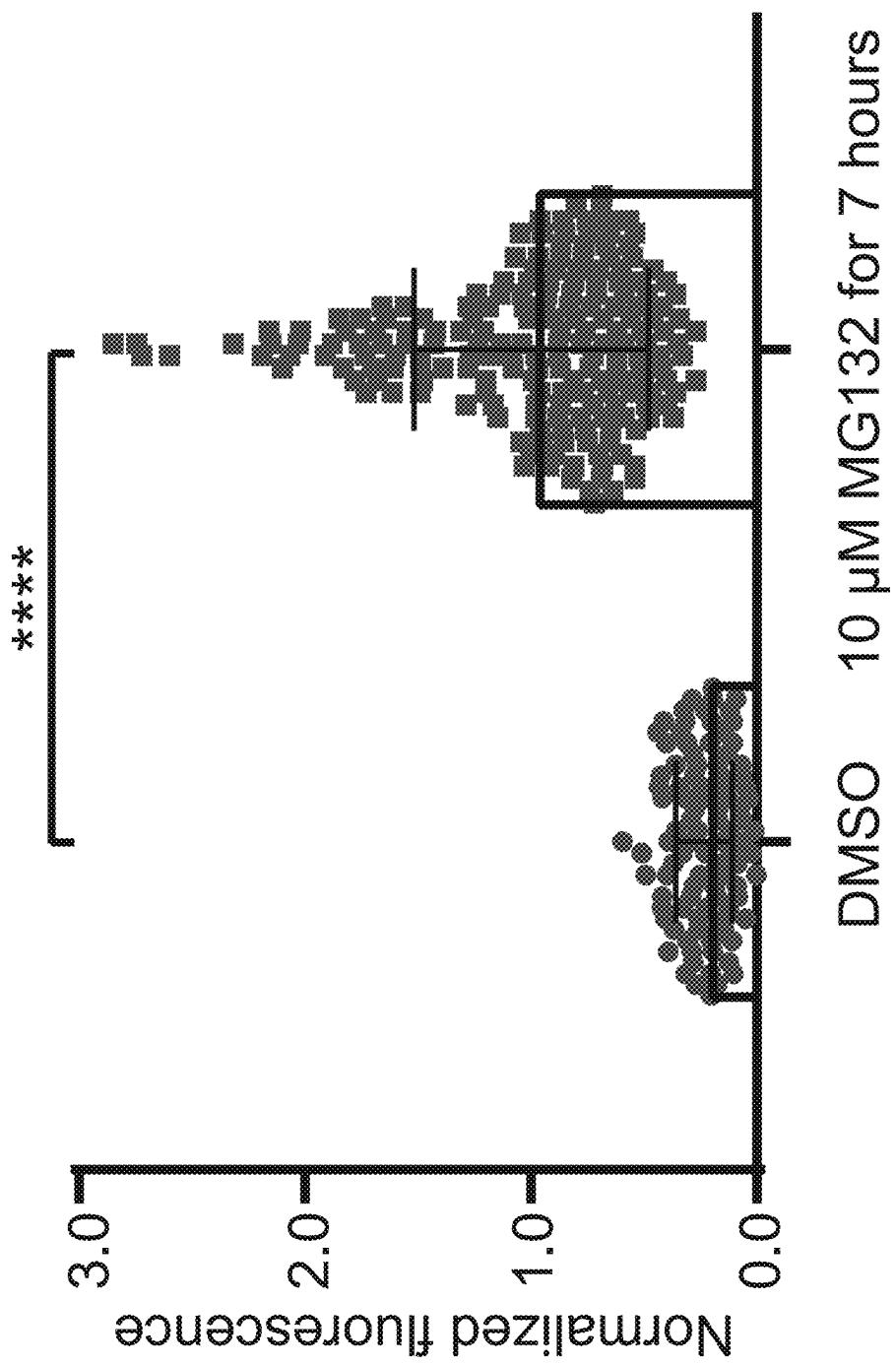


FIG. 3B

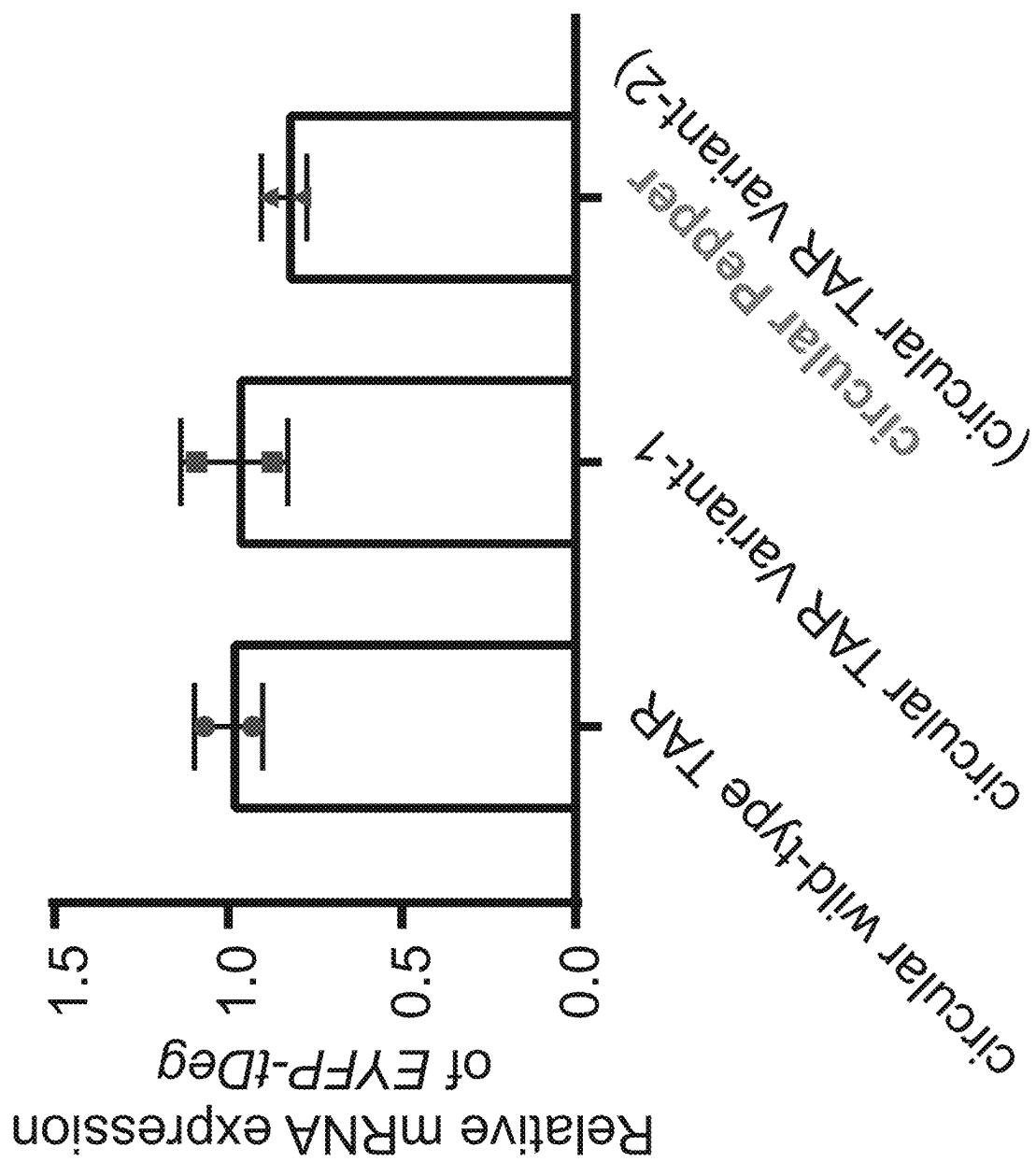
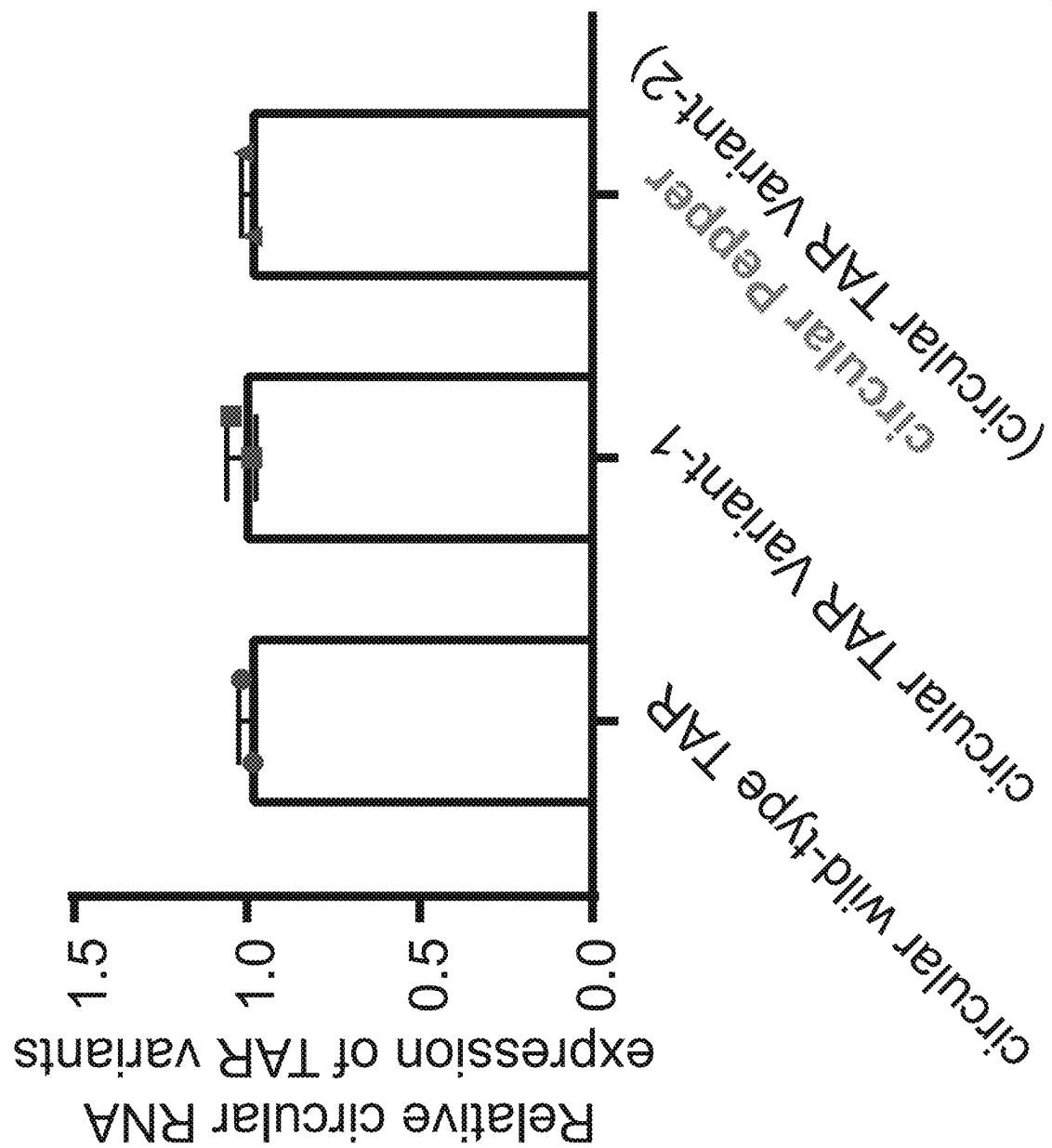
FIG. 4A

FIG. 4B



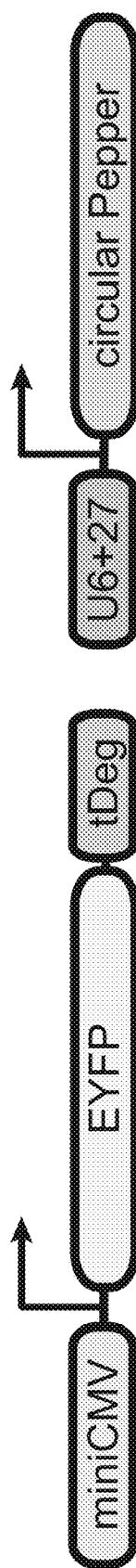


FIG. 5A

U2OS
EYFP-tDeg
circular + Pepper
- Pepper

COS-7
EYFP-tDeg
circular + Pepper
- Pepper

HeLa
EYFP-tDeg
circular + Pepper
- Pepper

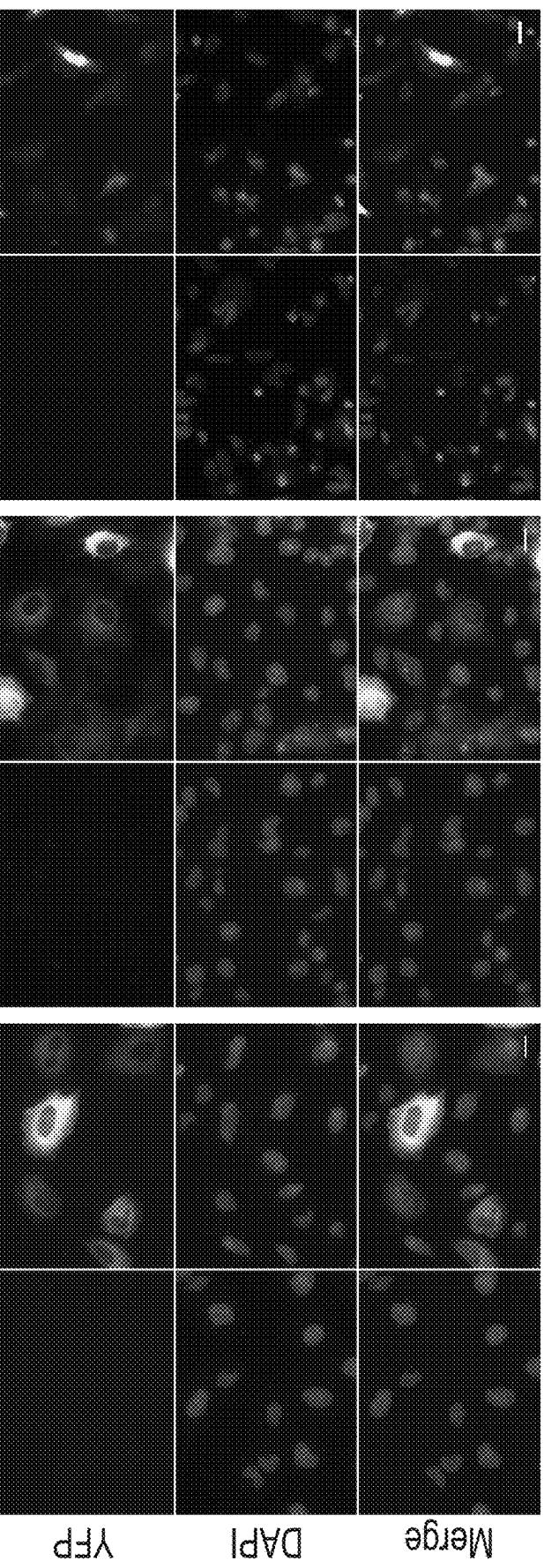


FIG. 5B

FIG. 5C

FIG. 5D

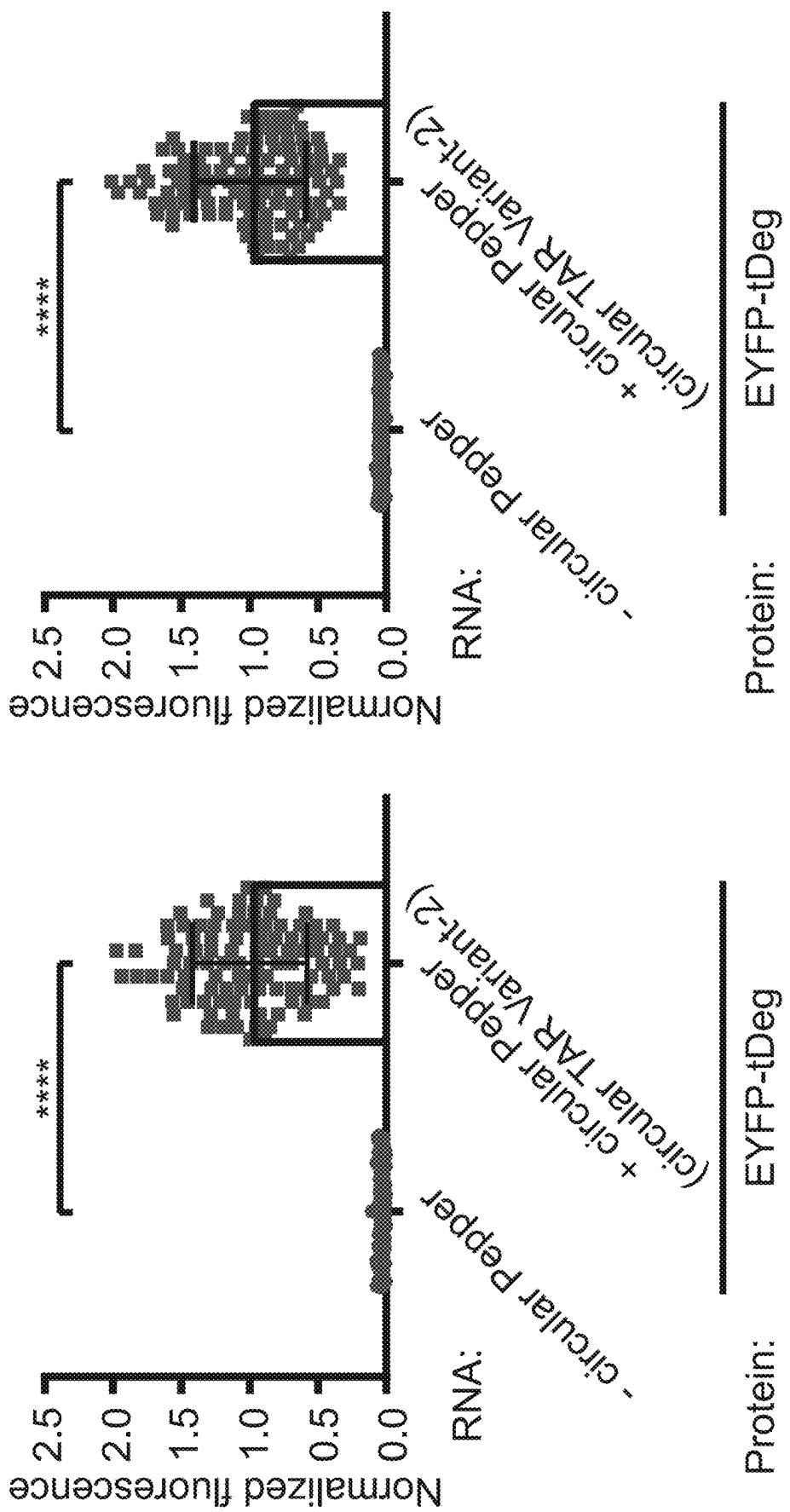


FIG. 5E

FIG. 5F

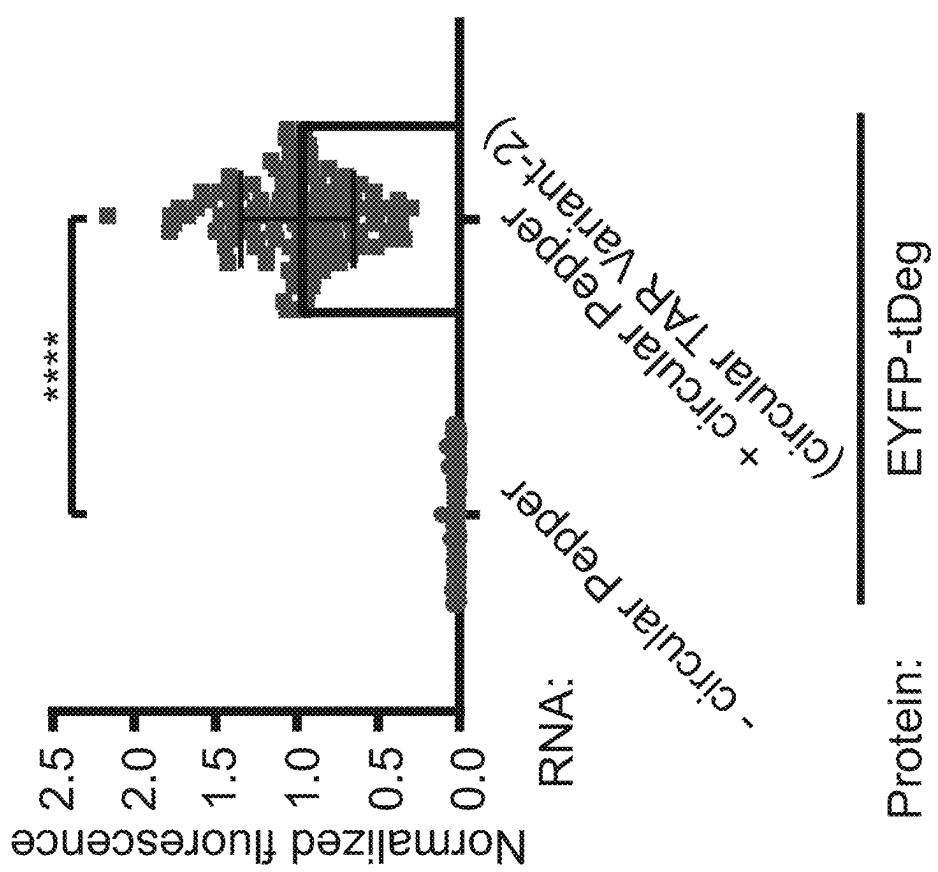


FIG. 5G

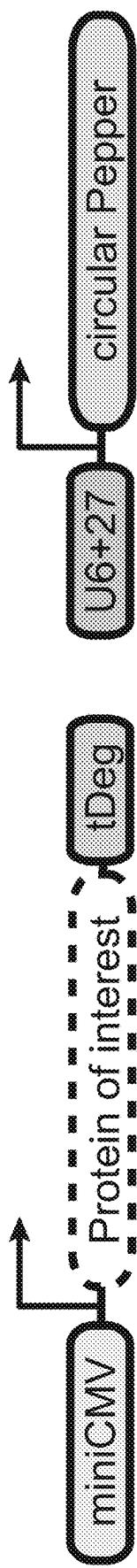


FIG. 6A
Protein of interest = mNeonGreen or mCherry or Nanoluc or Nanoluc

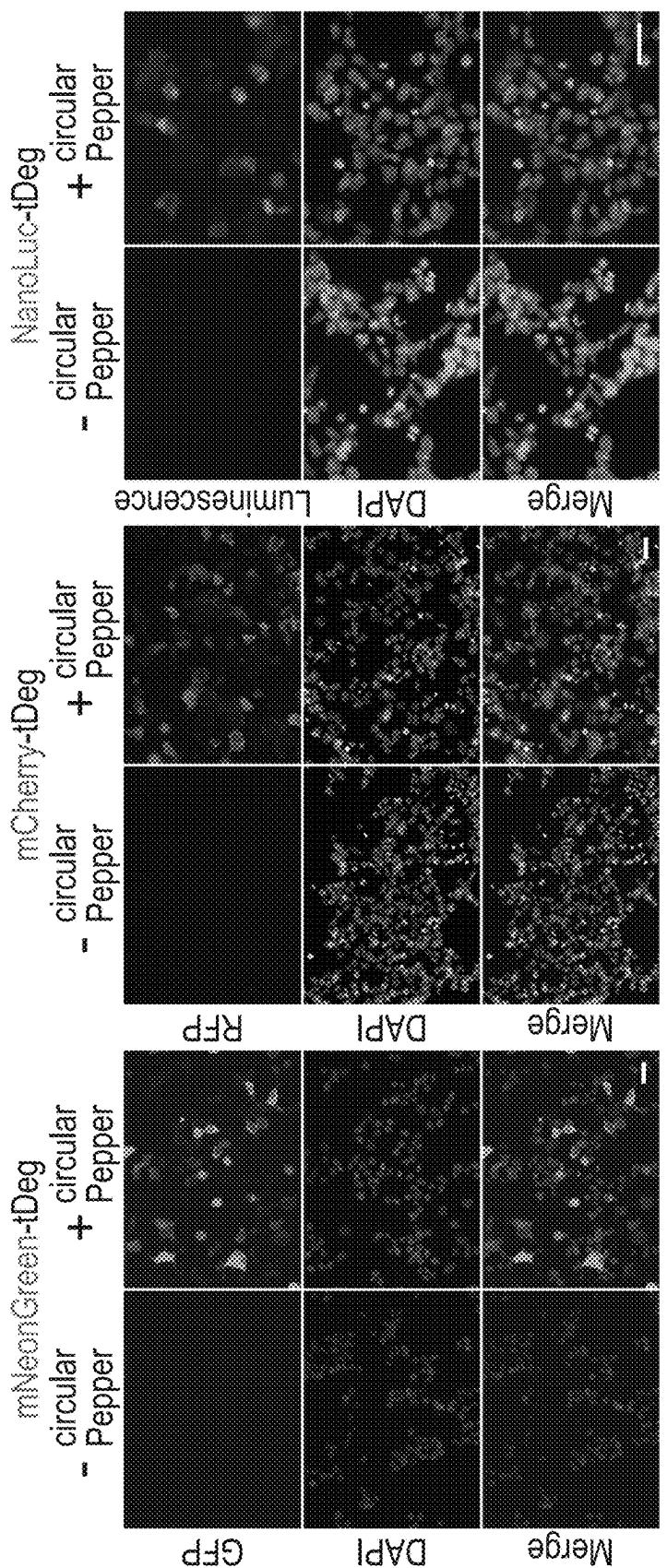


FIG. 6B
FIG. 6C
FIG. 6D

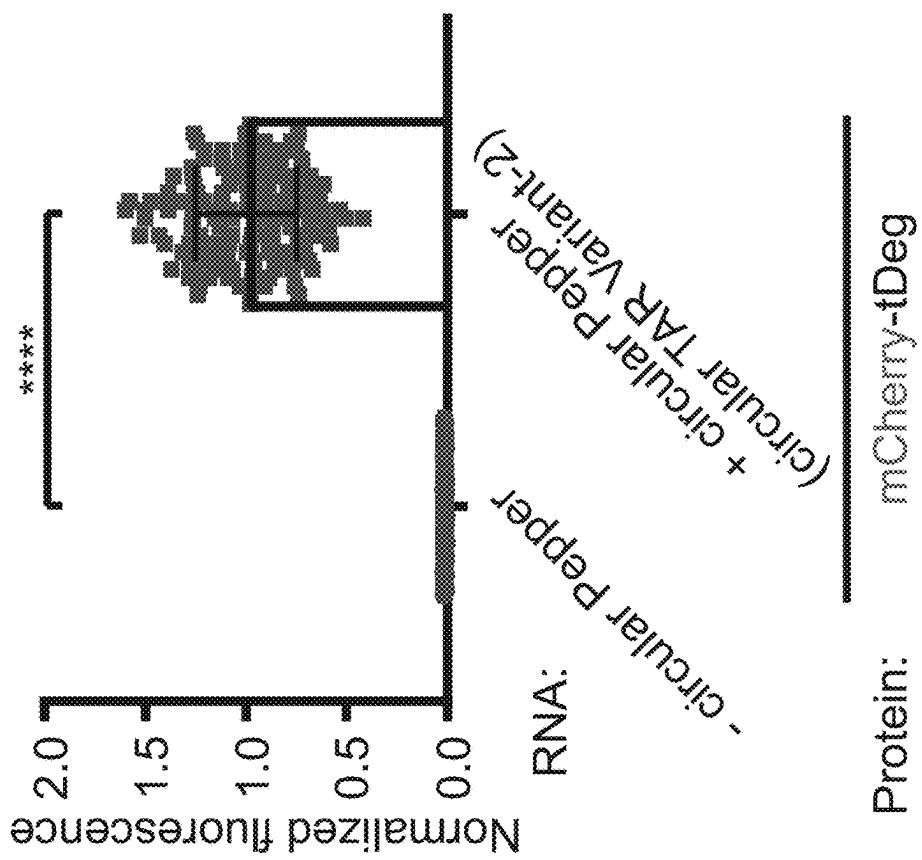


FIG. 6F

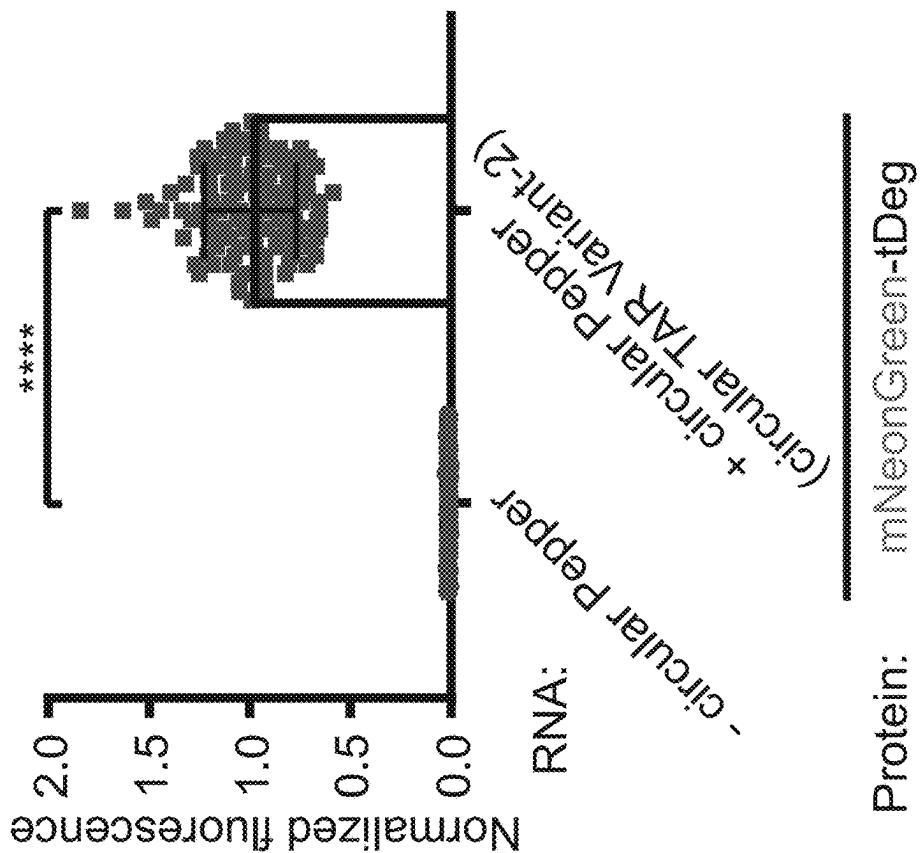


FIG. 6E

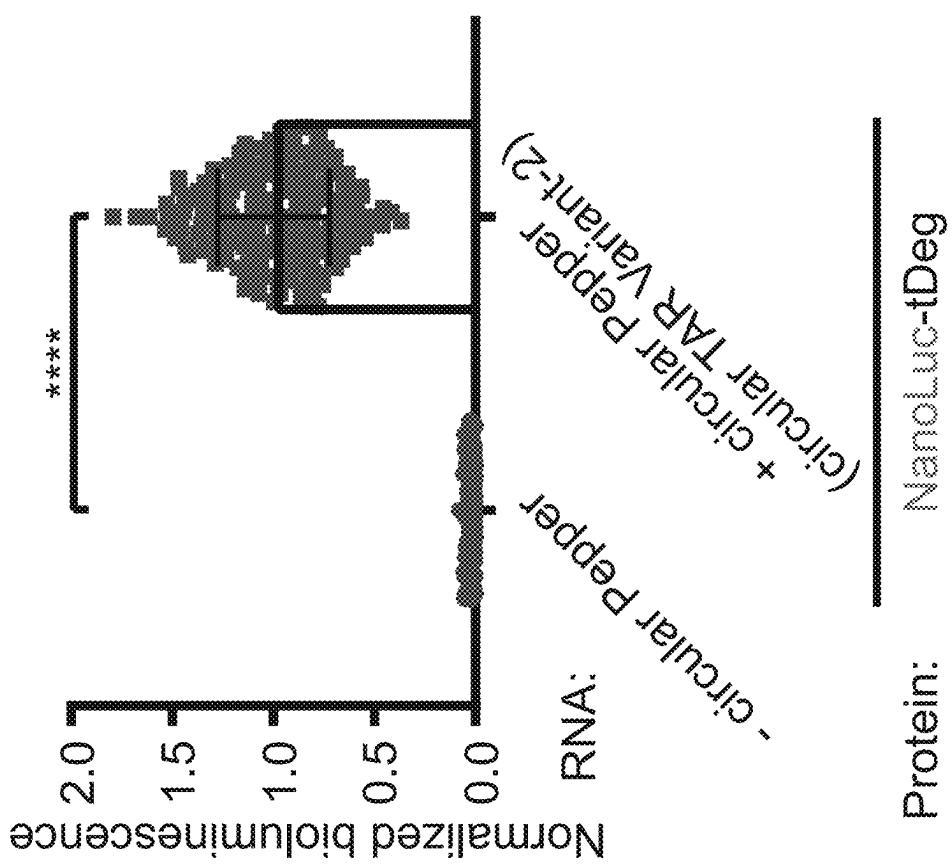


FIG. 6G

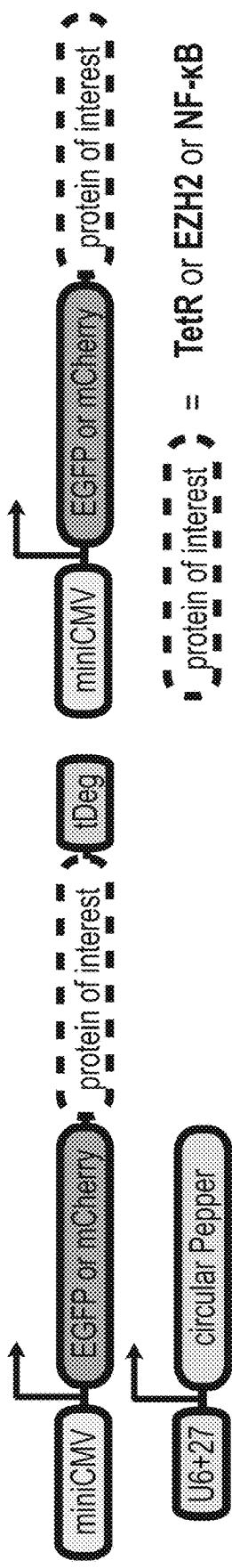


FIG. 7A

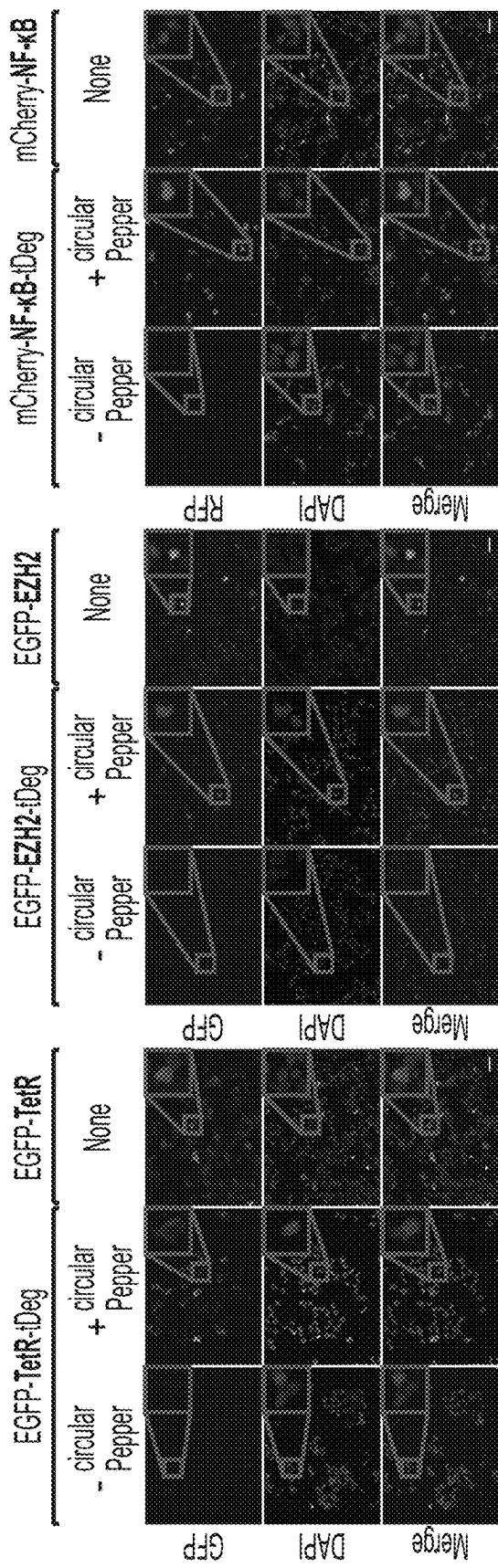


FIG. 7B

FIG. 7D

FIG. 7C

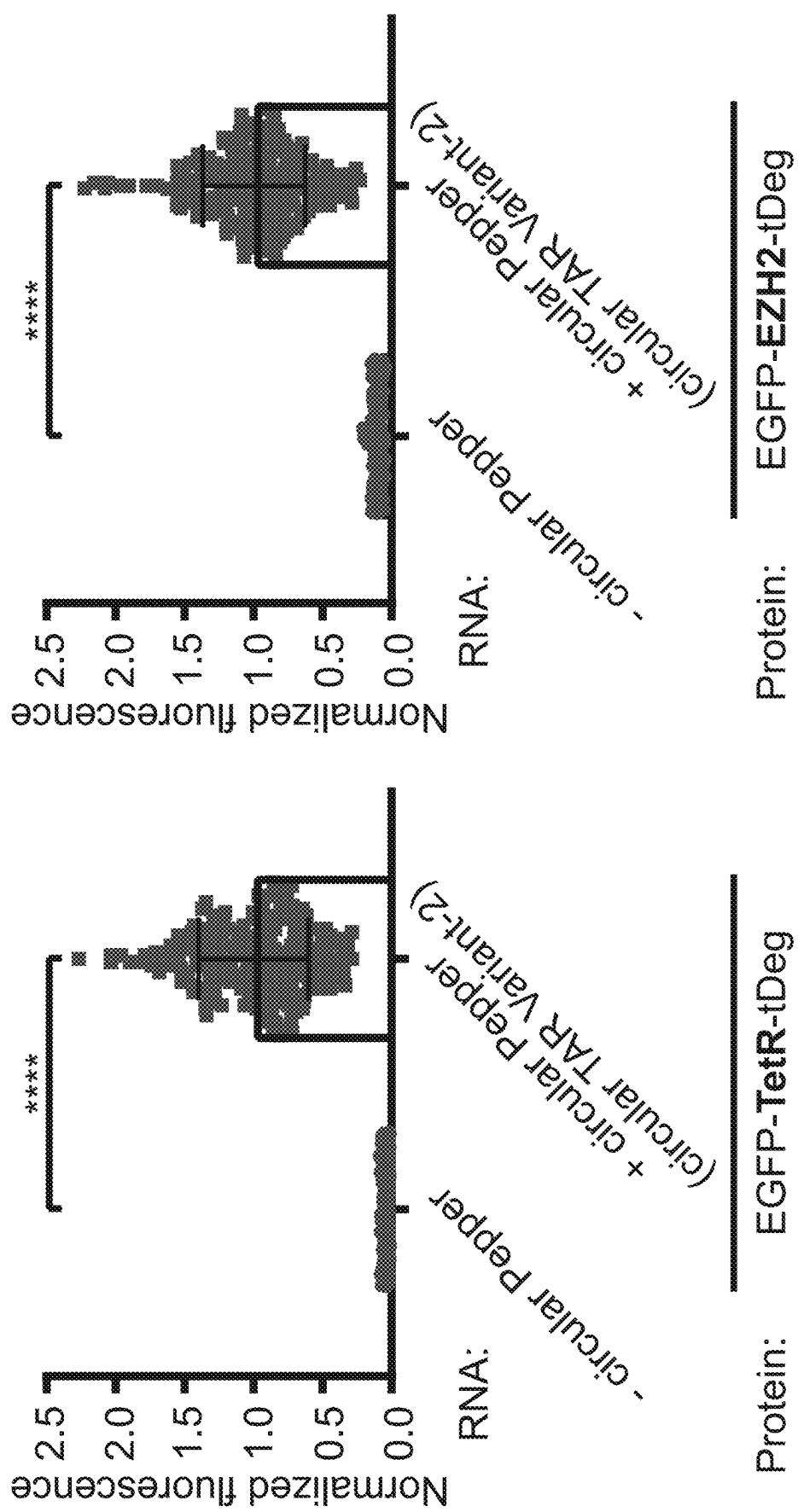


FIG. 7F

FIG. 7E

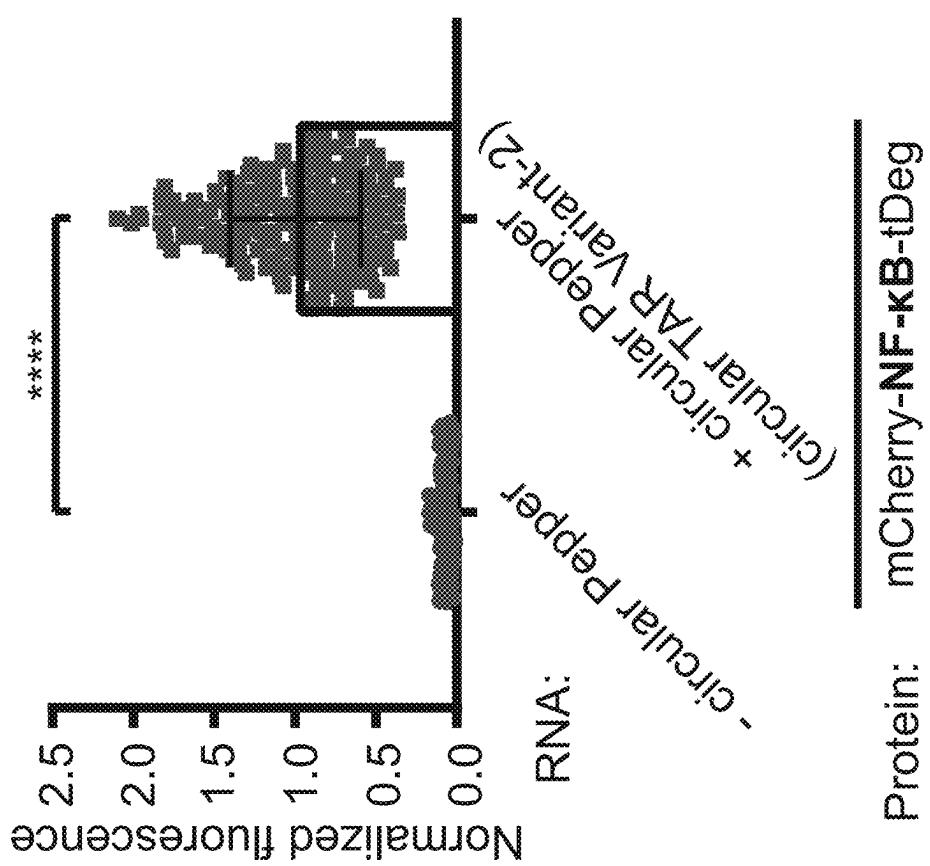


FIG. 7G



FIG. 8A

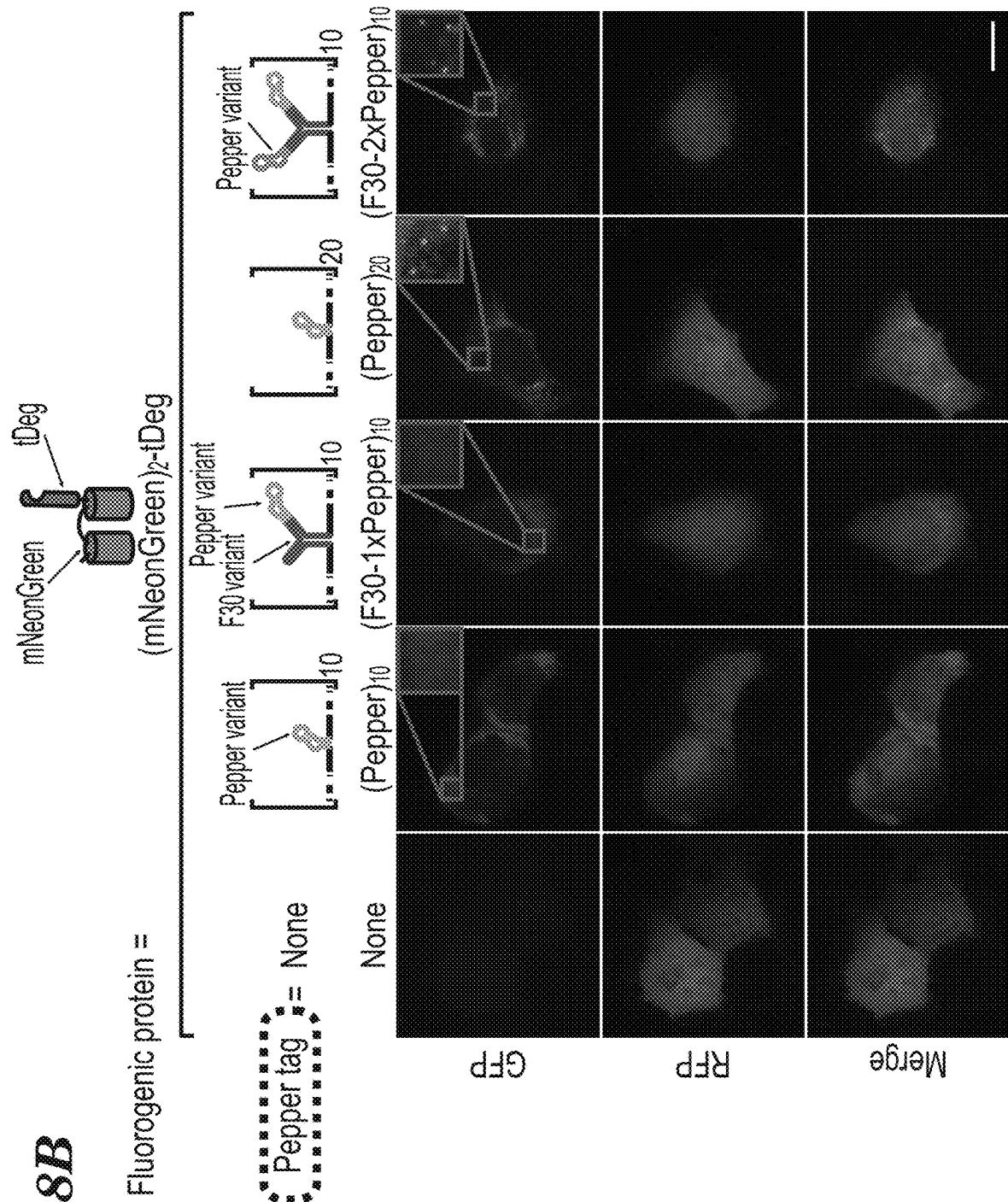
FIG. 8B

FIG. 9A

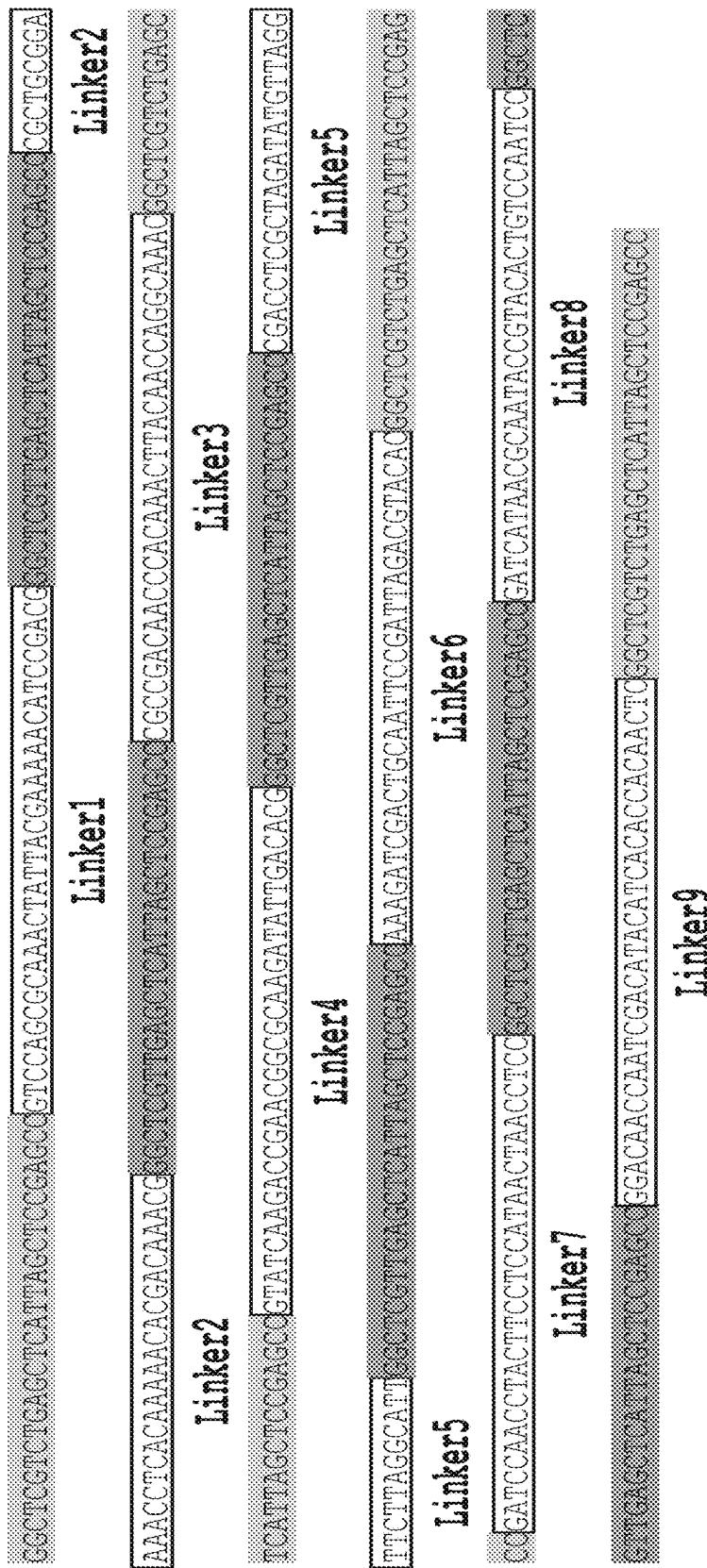
(Pepper)₁₀ tag (586 bp):

卷之三

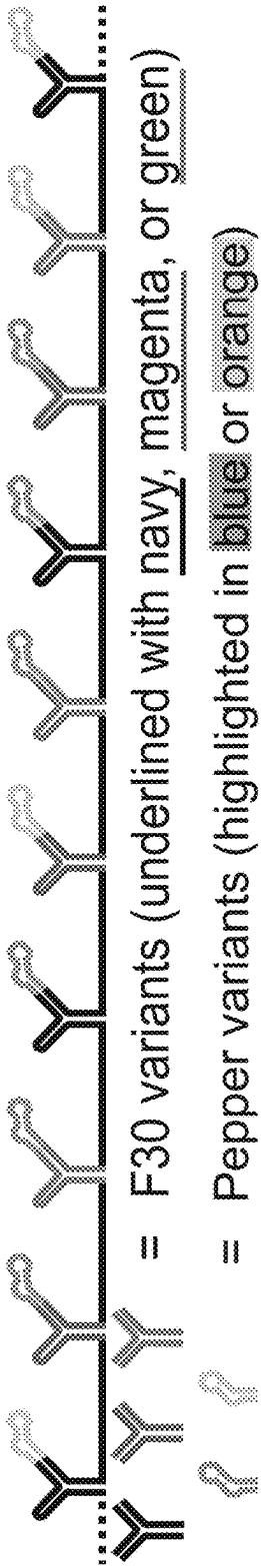
= Pepper variants (highlighted in blue or orange)

= Different linkers (highlighted in black boxes)

(Pepper)₁₀ tag sequence:



(F30-1xPepper)₁₀ tag (1466 bp):



F30 variants (underlined with navy, magenta, or green)

= Pepper variants (highlighted in blue or orange)

= Different linkers (highlighted in black boxes)

(F30-1xPepper)₁₀ tag sequence:

卷之三

10

卷之三

卷之三

四百一十九

E30_VdElitz

CCACGGTTCCCTACAA

卷之三

卷之三

CATACCTACAA

三

卷之三

四百一十一

LIBRARY

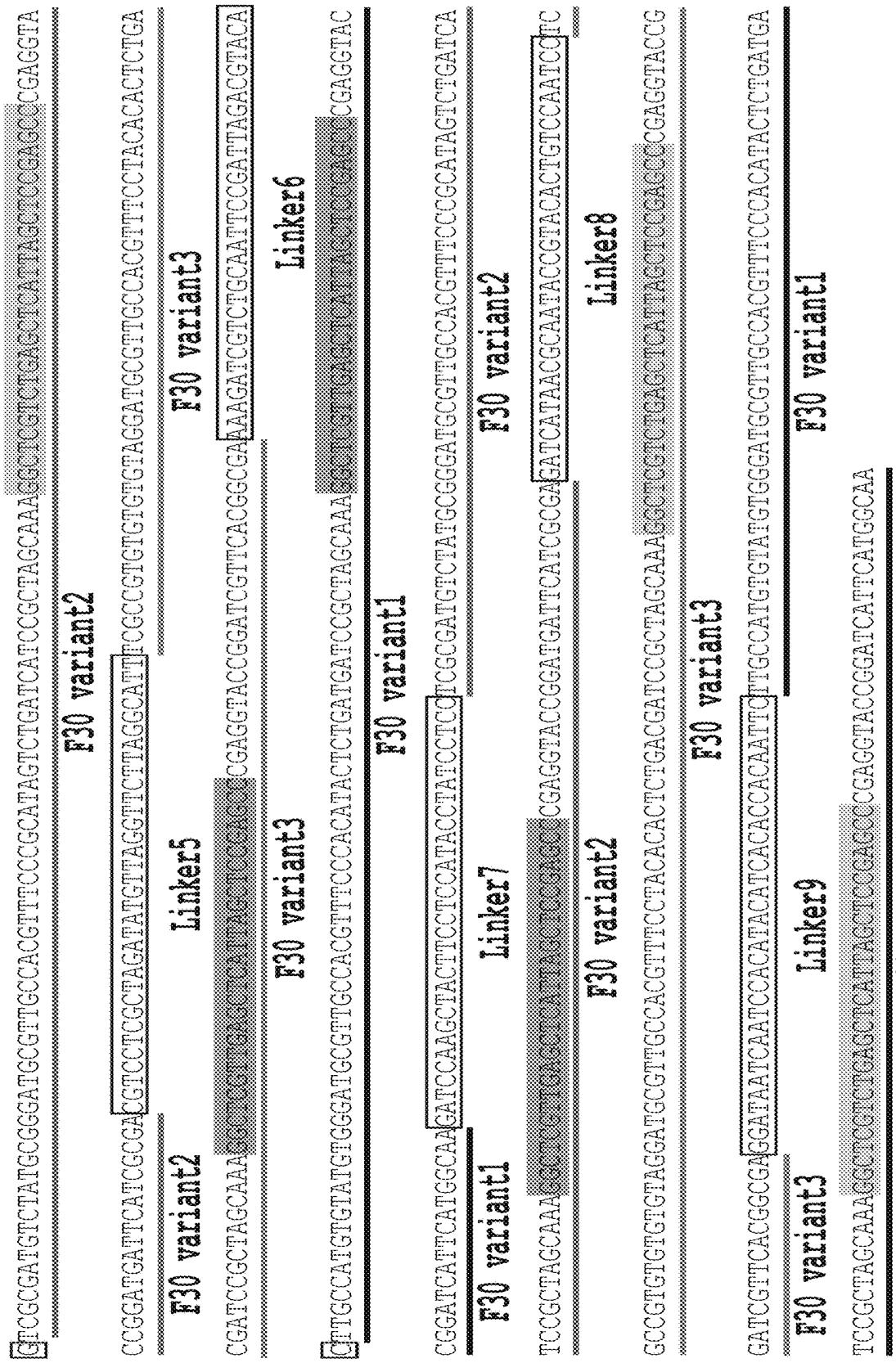


FIG. 9B (cont.)

(Pepper)₂₀ tag (1228 bp):

(Pepper)₂₀ tag (1228 bp):



= Pepper variants (highlighted in blue or orange)

— = Different linkers (highlighted in black boxes)

(Pepper)20 tag sequence:

Linker1	GGCTTGTGACTCTATTAACCTTCGGAAACTTACCGAAAACATCCGACC	CGCTTGCGGA
Linker2	AACCTCACAAAACACGGACAACCG	CGGGGACAAACCCACAAACTTACAAACCAGGAAACGCCGTTCTCAGC
Linker3	TGATTAACCTTCGGAAACGGGGCAAGATAATTGACACG	CGAACCTCGCTAGATATGTTAGG
Linker4	TCTTGTAGGCATTT	AAAGATGGACTGAAATTCCGATTAGACGGTACACGTTGCTTCCTGGC
Linker5	GATCCAACCTACTTCTCCATAACTAACCTCC	GATCUATAACCGAATAACCGTACACTGCCAAATCG
Linker6		
Linker7		
Linker8		
Linker9		
Linker10		

FIG. 9C (*cont.*)

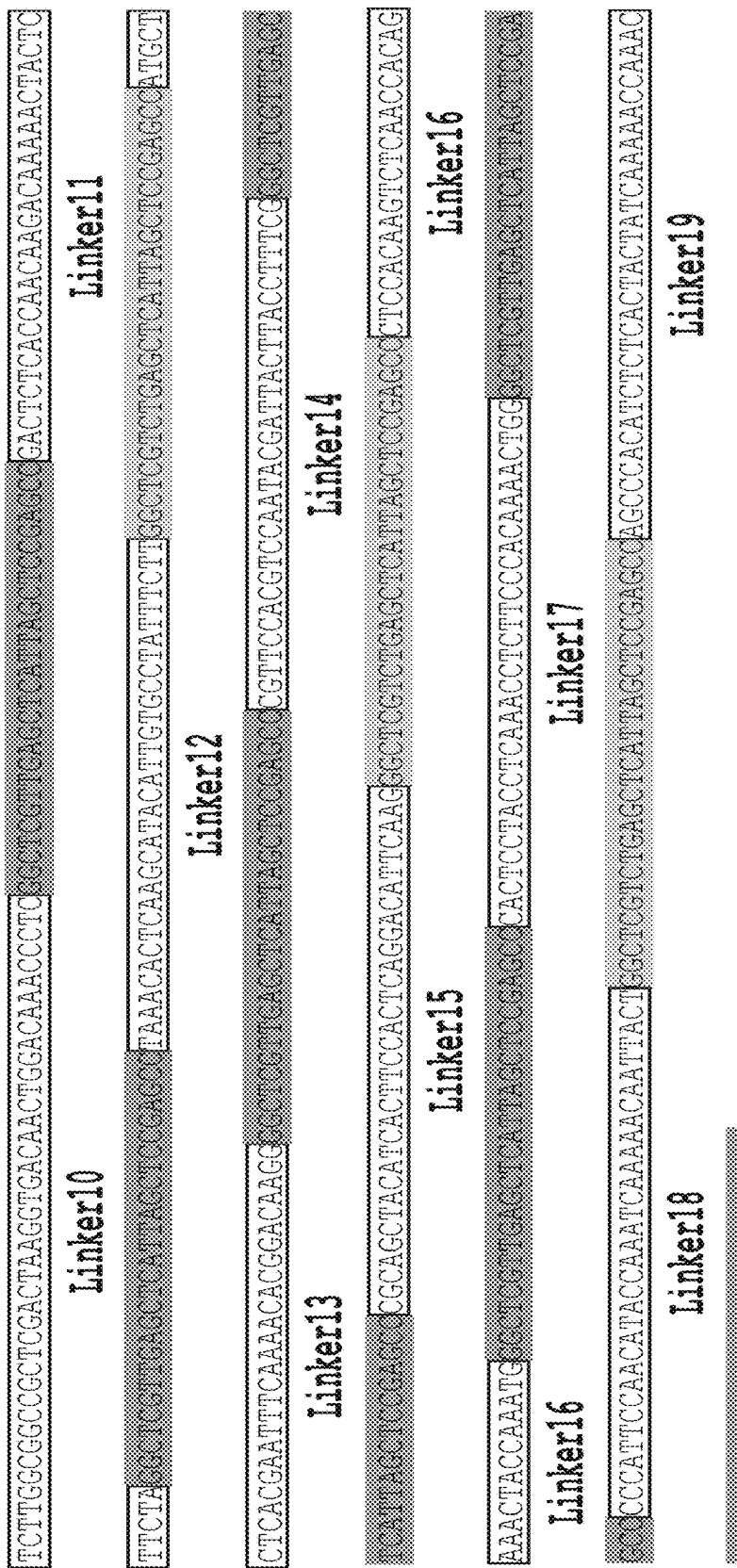
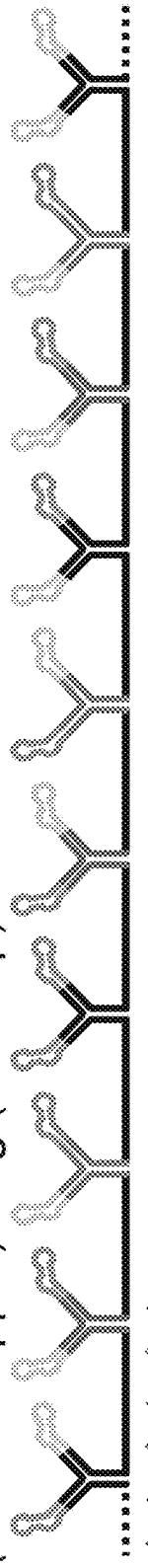


FIG. 9D

(F30-2xPepper)₁₀ tag (1812 bp):



= F30 variants (underlined with navy, magenta, or green)

20

A vertical column of black dots, likely representing a sequence or a list item.

(F30-2xPepper)₁₀ tag sequence:

TGGCCATGTGTATGTCGAAGCTAGAACTCGACTACGGTTCCACATACTCTGATGATCGCTAACGCC

112

TCCTGTCAGCCATTACCTCCGAGGTTACCGGATCATTCAAGGAATCTTACGAAATCATCCGACCTGATGGCTATGC

E30 Variant

Tolka
G

E30 Variante 2

GGGAAGCGTAGAAACGCCGCTGCTGATTTACCTTGGACCGATACTGGATCAGTCTGATCATCCGCTAGCAAA

30 variant2

CGAGGTACCGGATGATTCACTGGGAAATCTACAAAATACGTCAAACGTGCGCGTGTGTAGGAAGCGTAGAAAC

30 Vattanez

Linker

varianze

FIG. 9D (*cont.*)

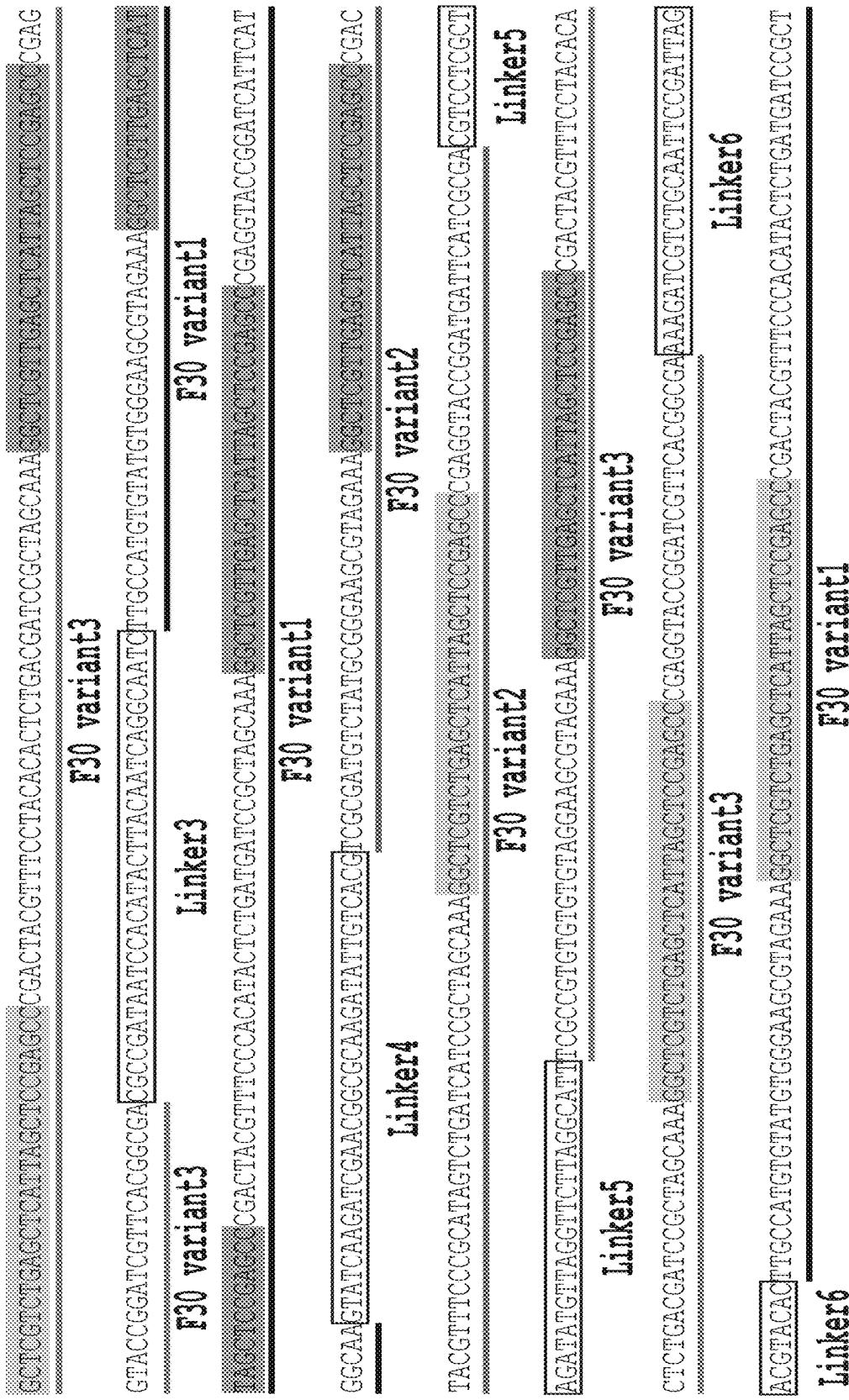
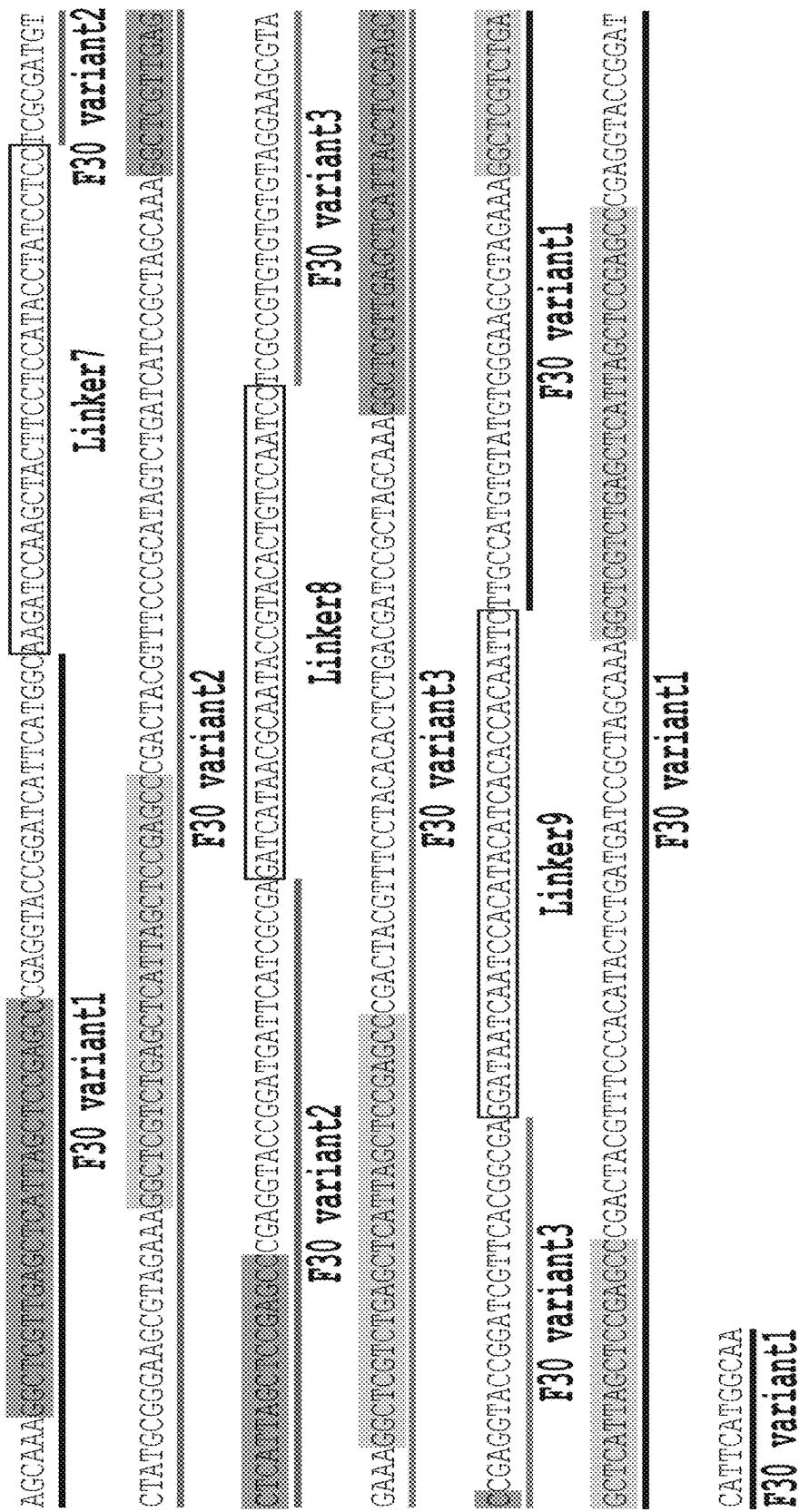


FIG. 9D (*cont.*)



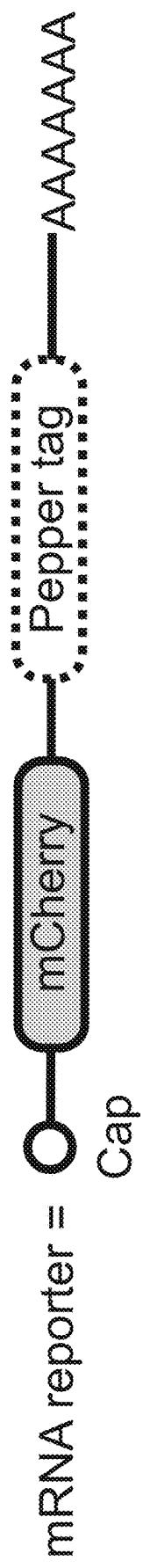


FIG. 10A

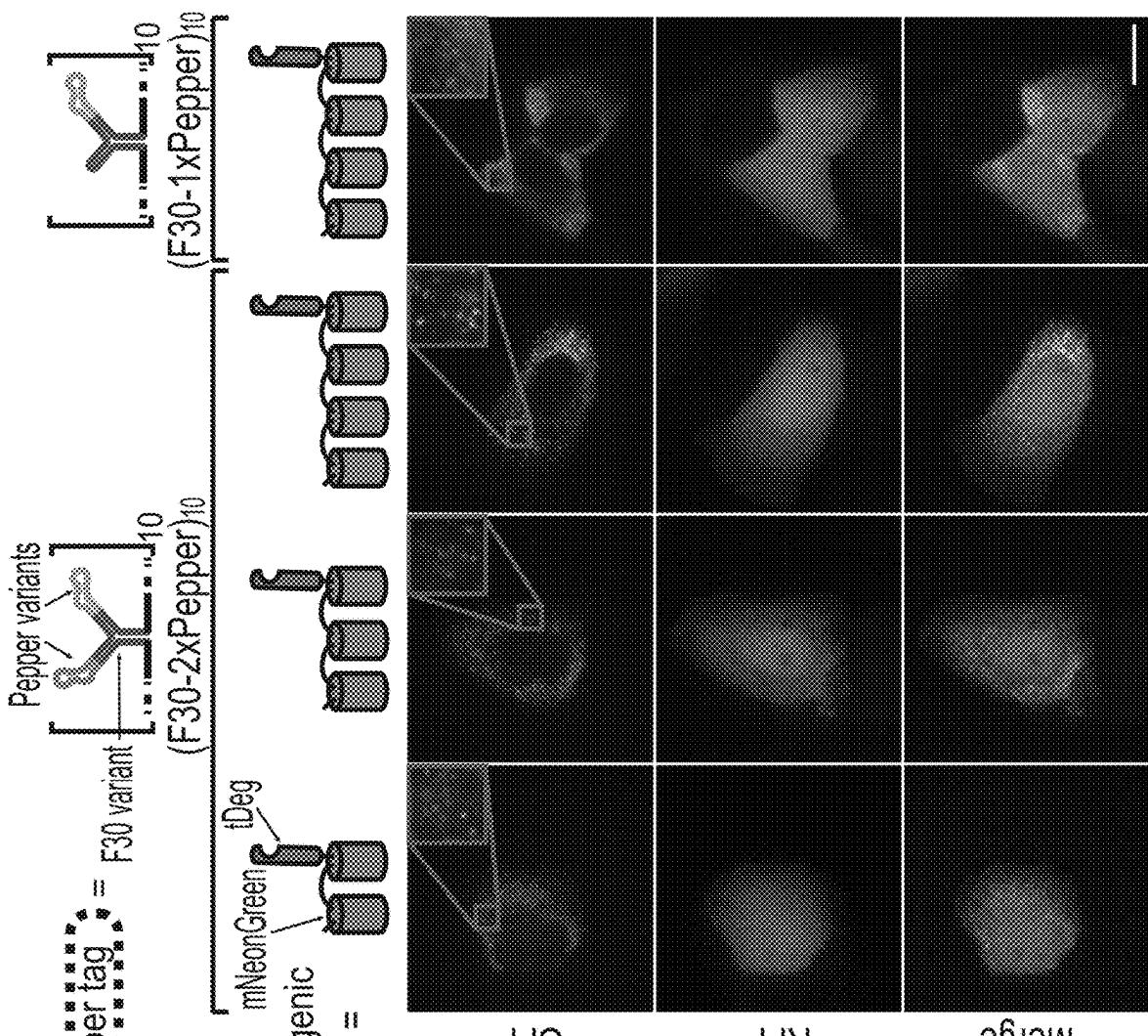


FIG. 10B

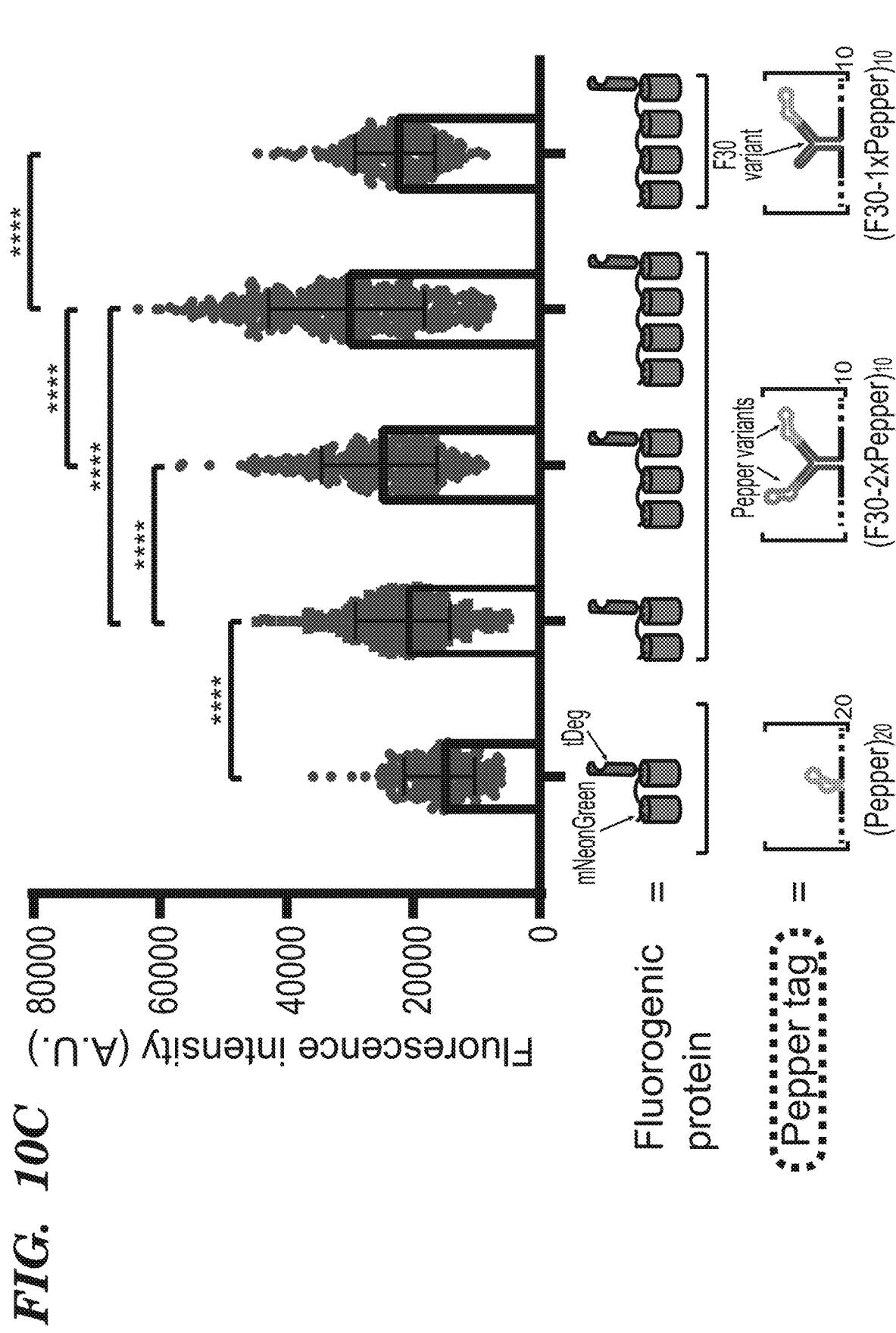




FIG. 11A

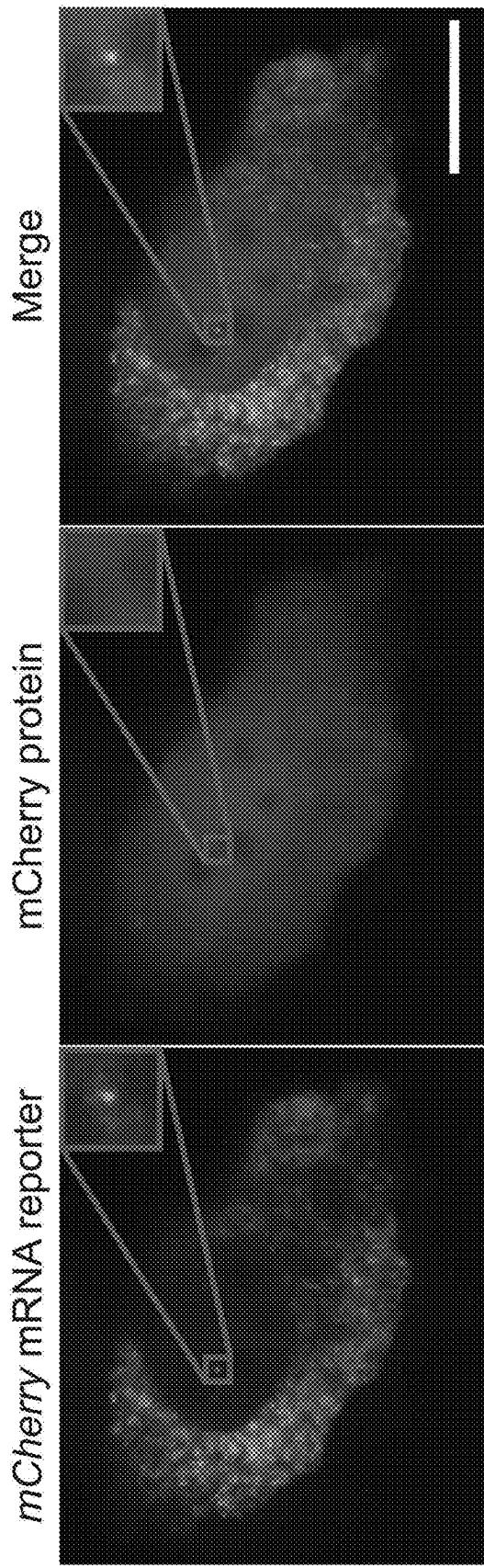


FIG. 11B

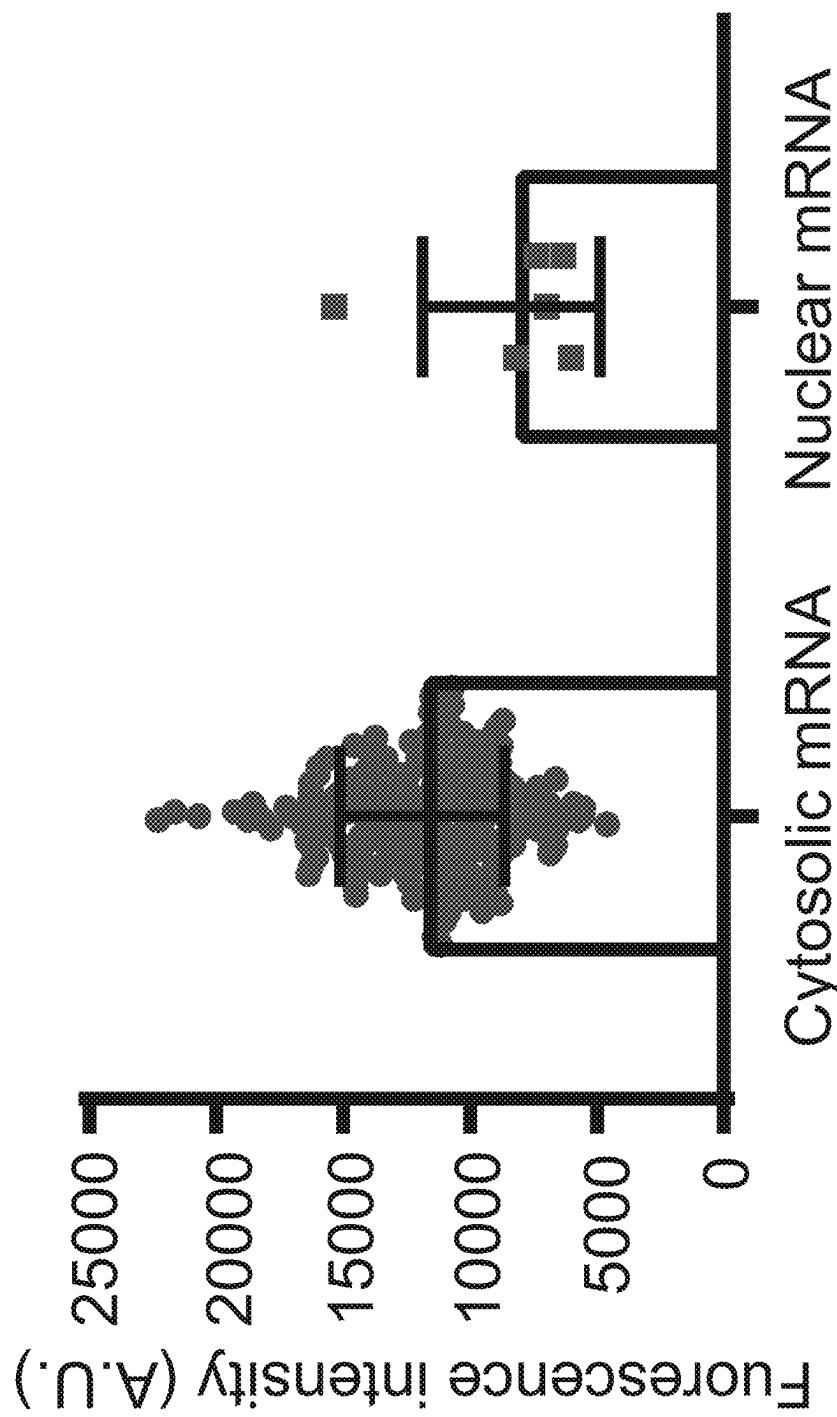


FIG. 11C

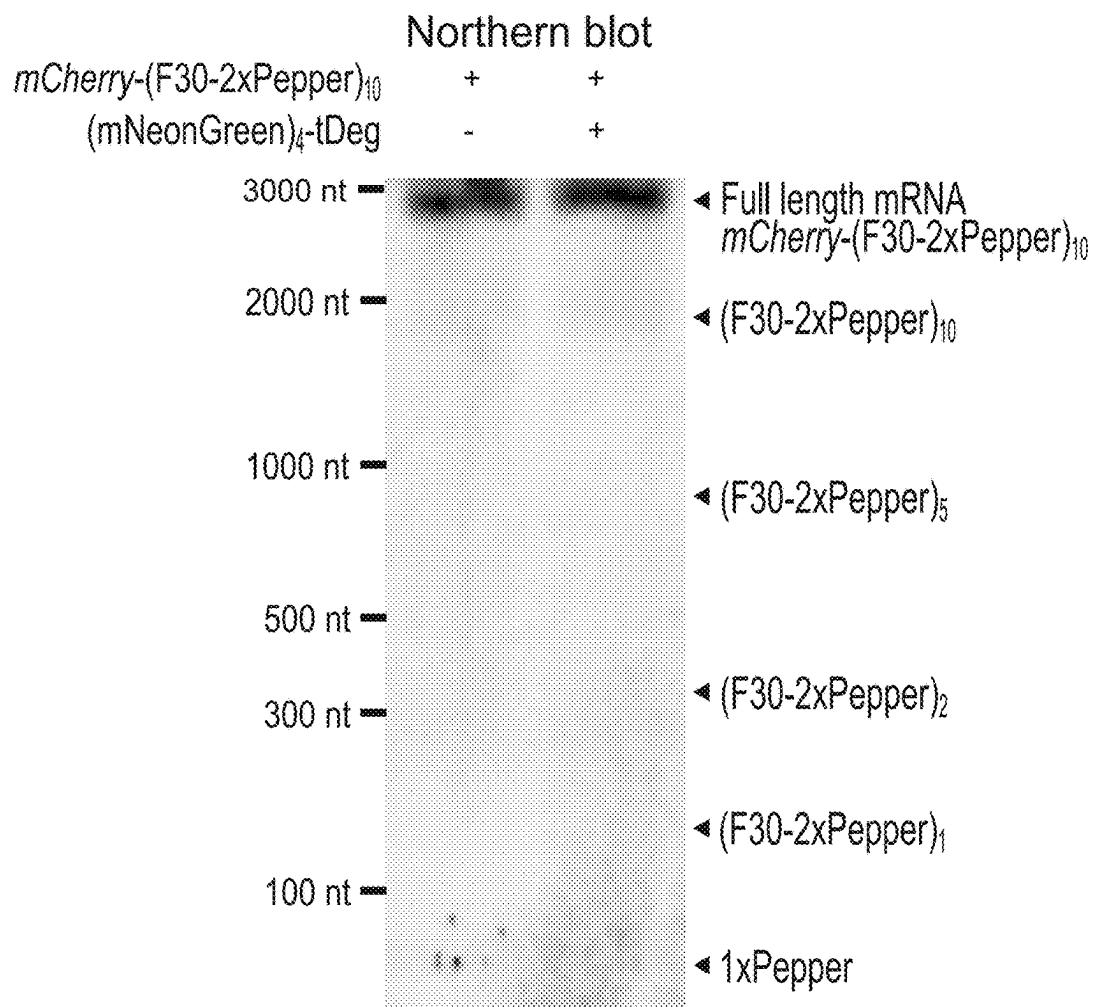


FIG. 12A

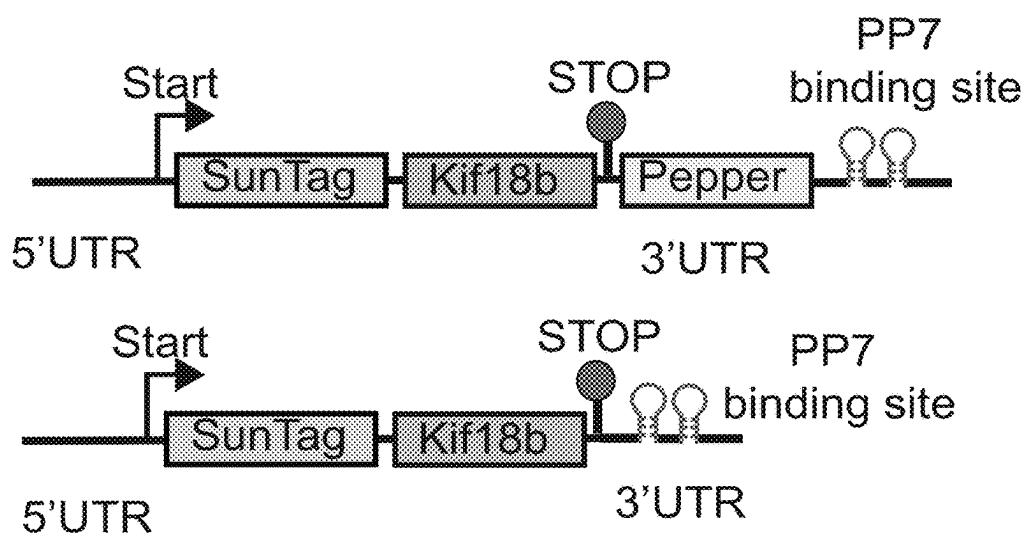
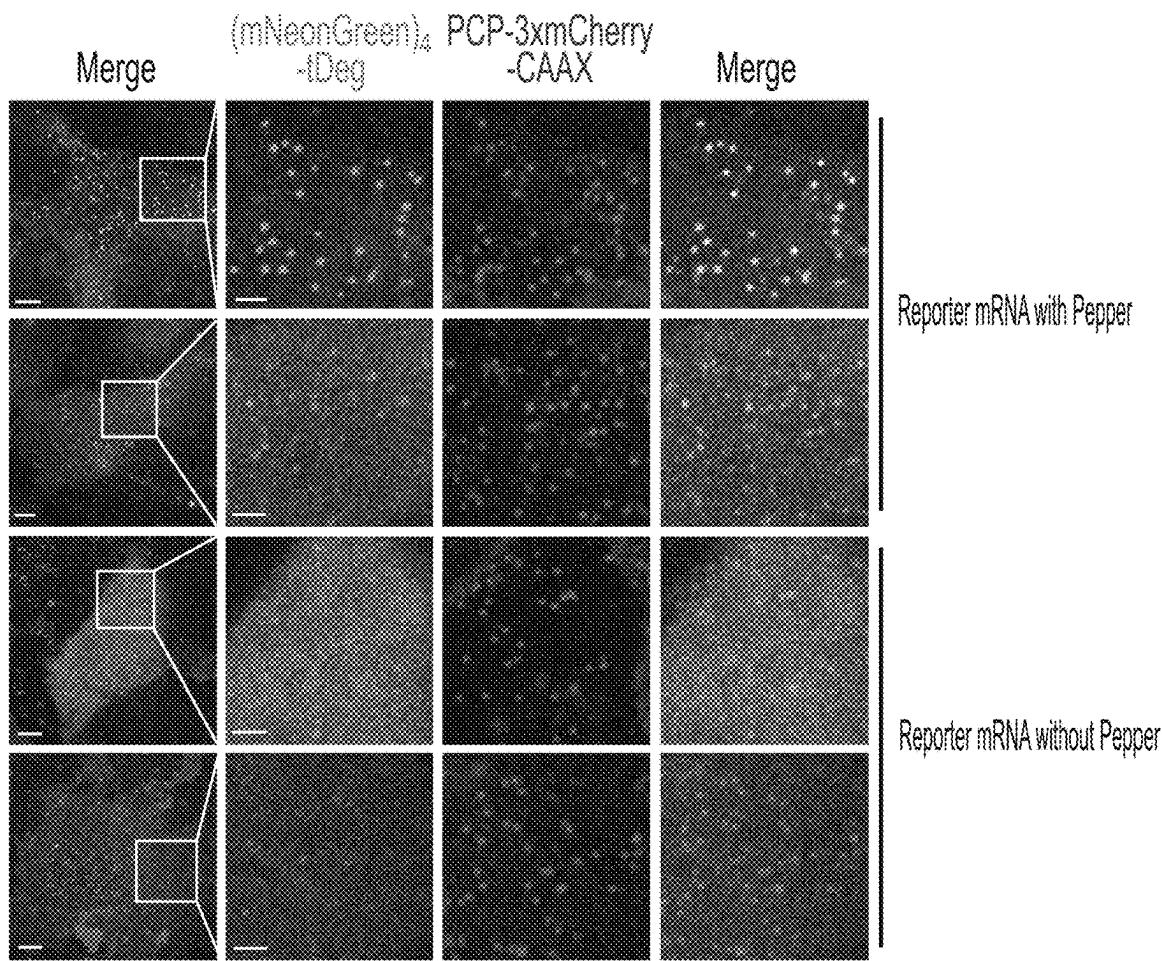
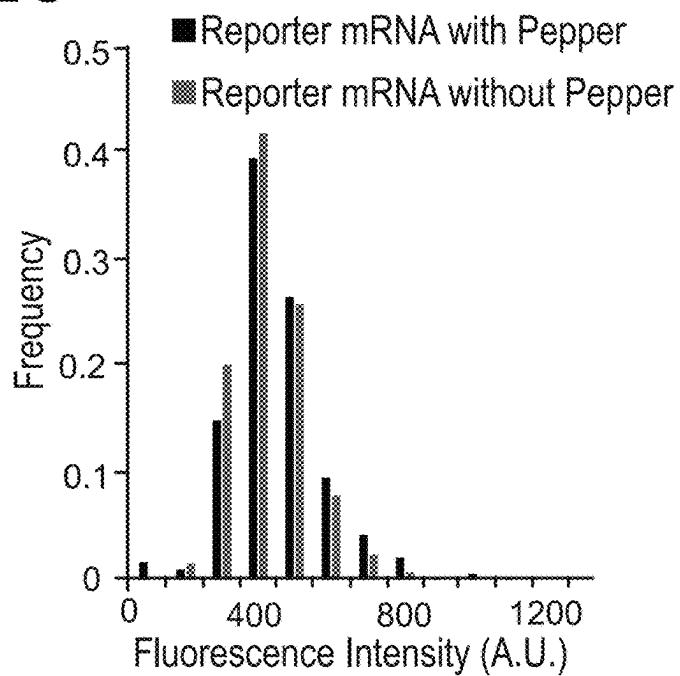


FIG. 12B

FIG. 12C**FIG. 12D**

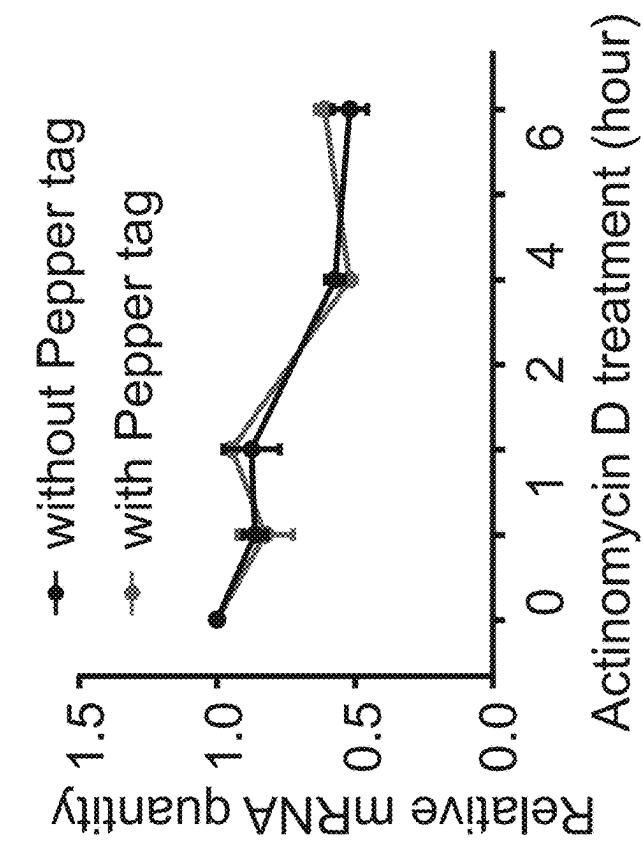


FIG. 13A

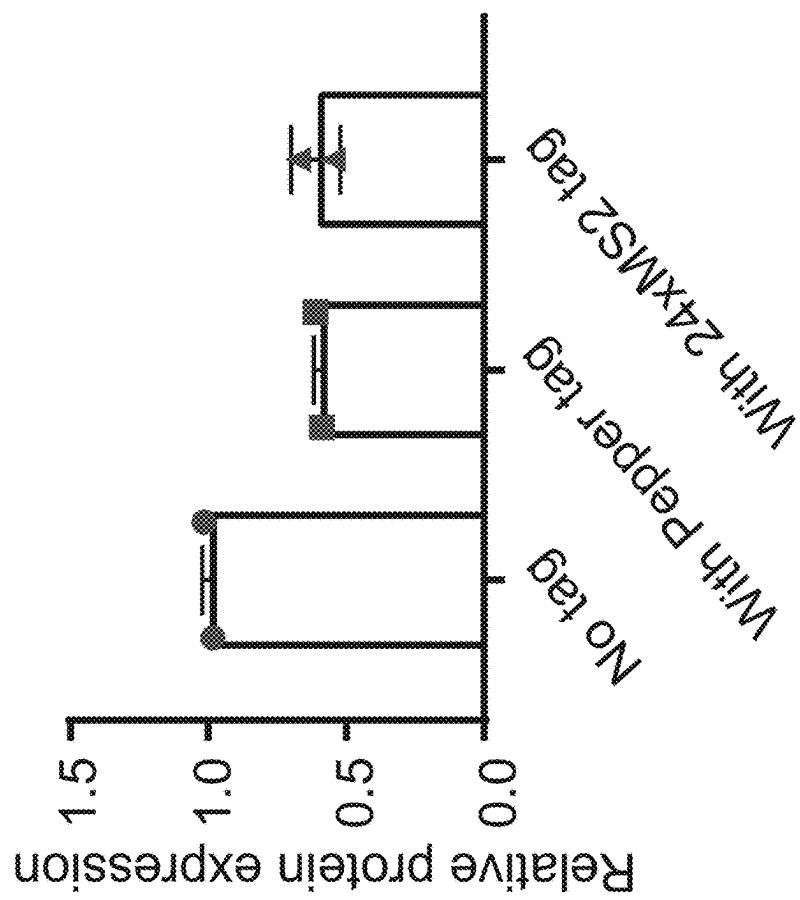


FIG. 13B

FIG. 13D

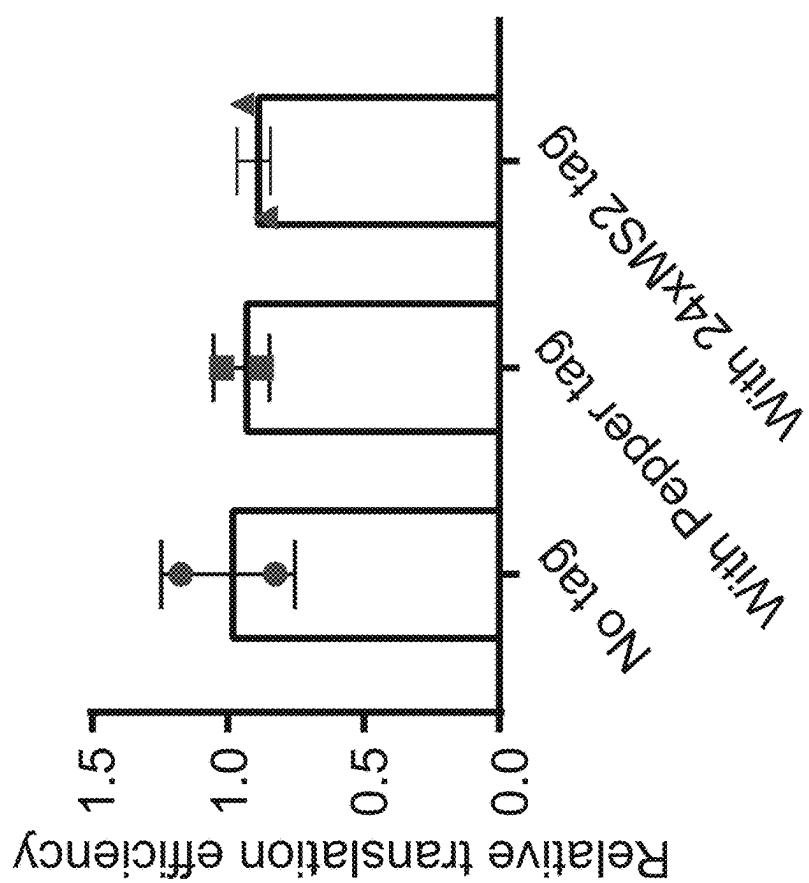
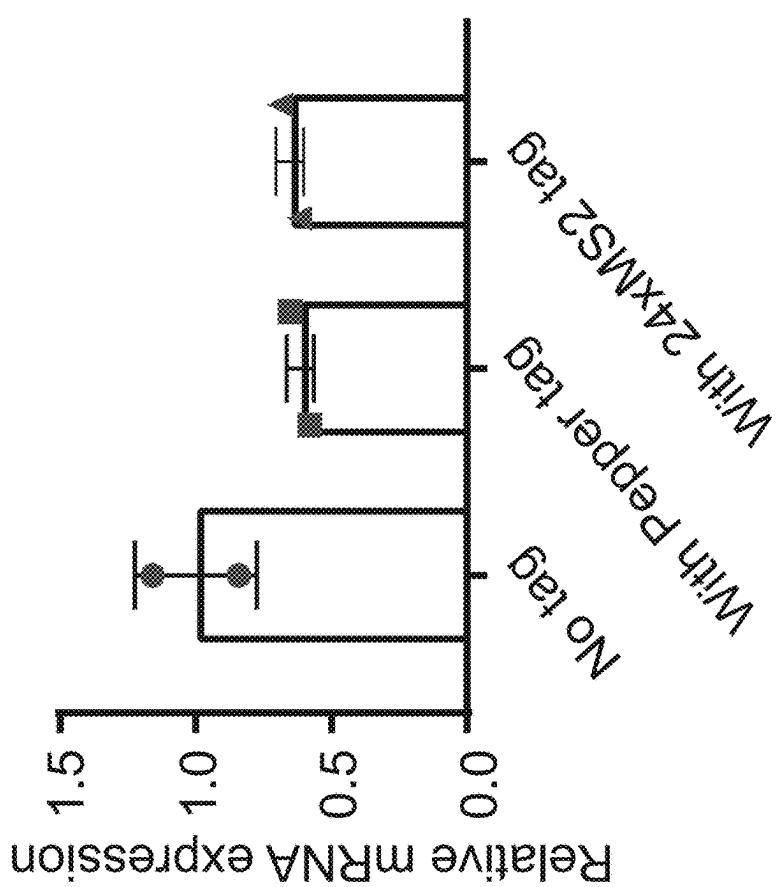
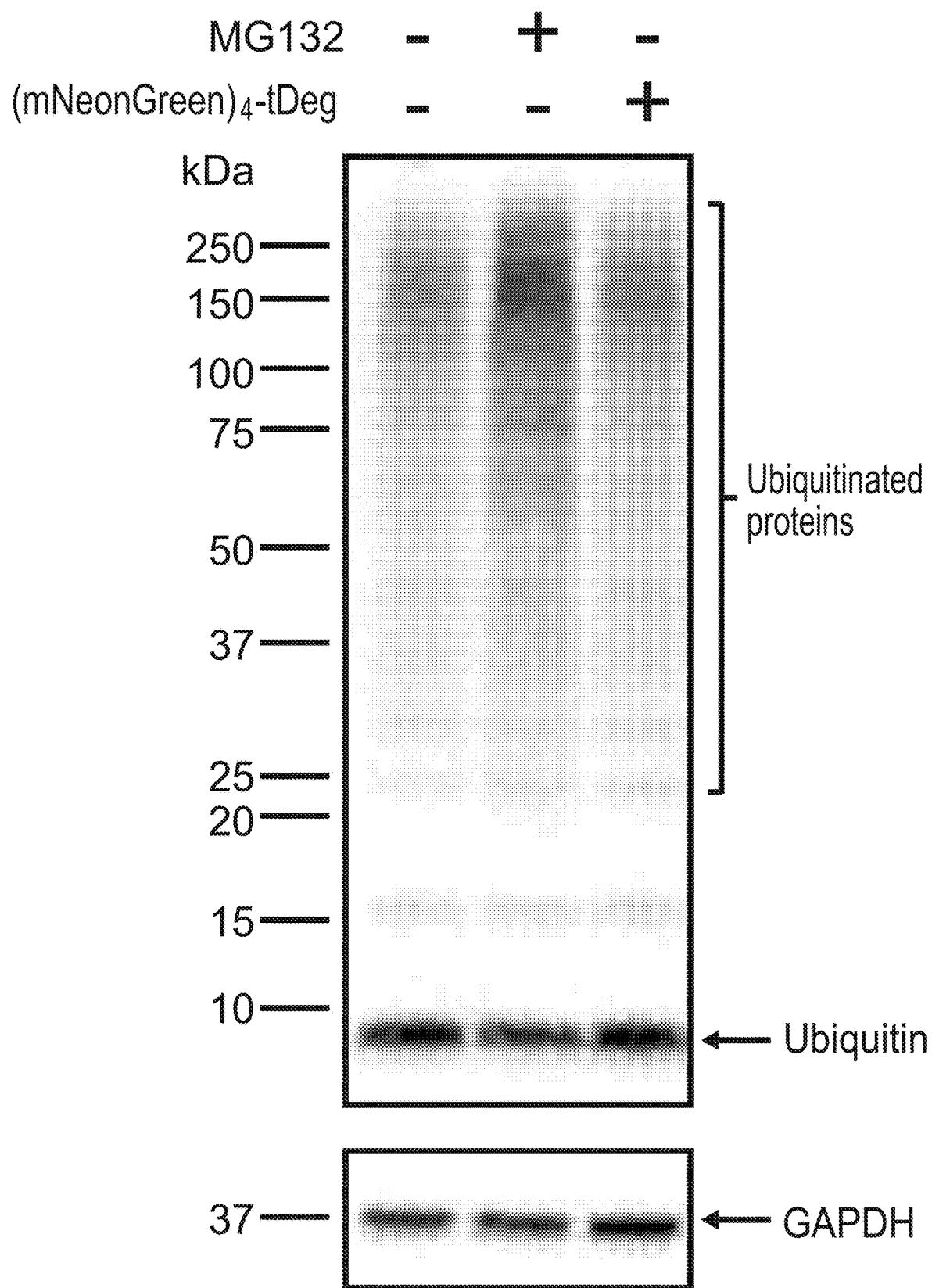


FIG. 13C



**FIG. 13E**

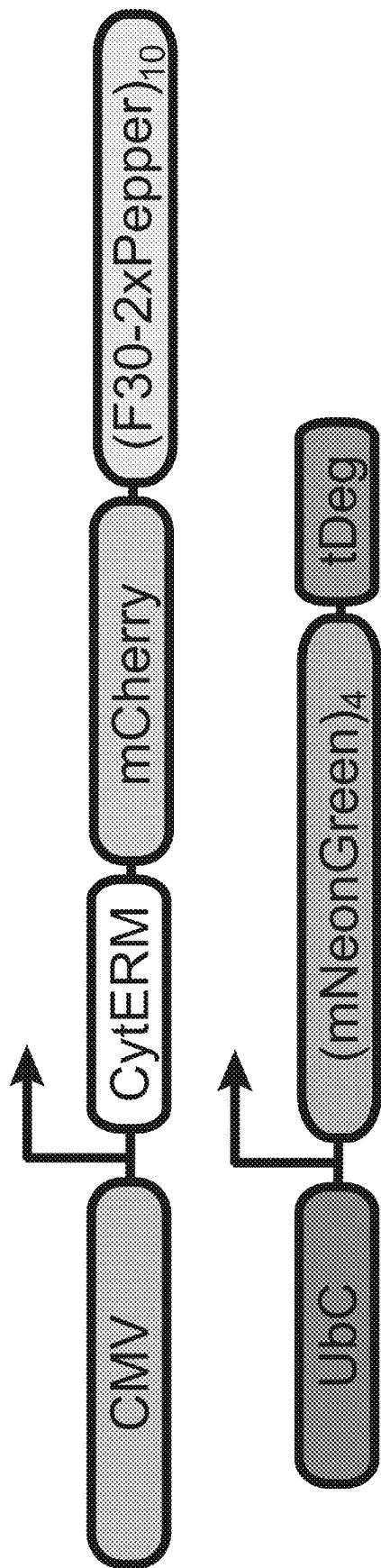


FIG. 14A

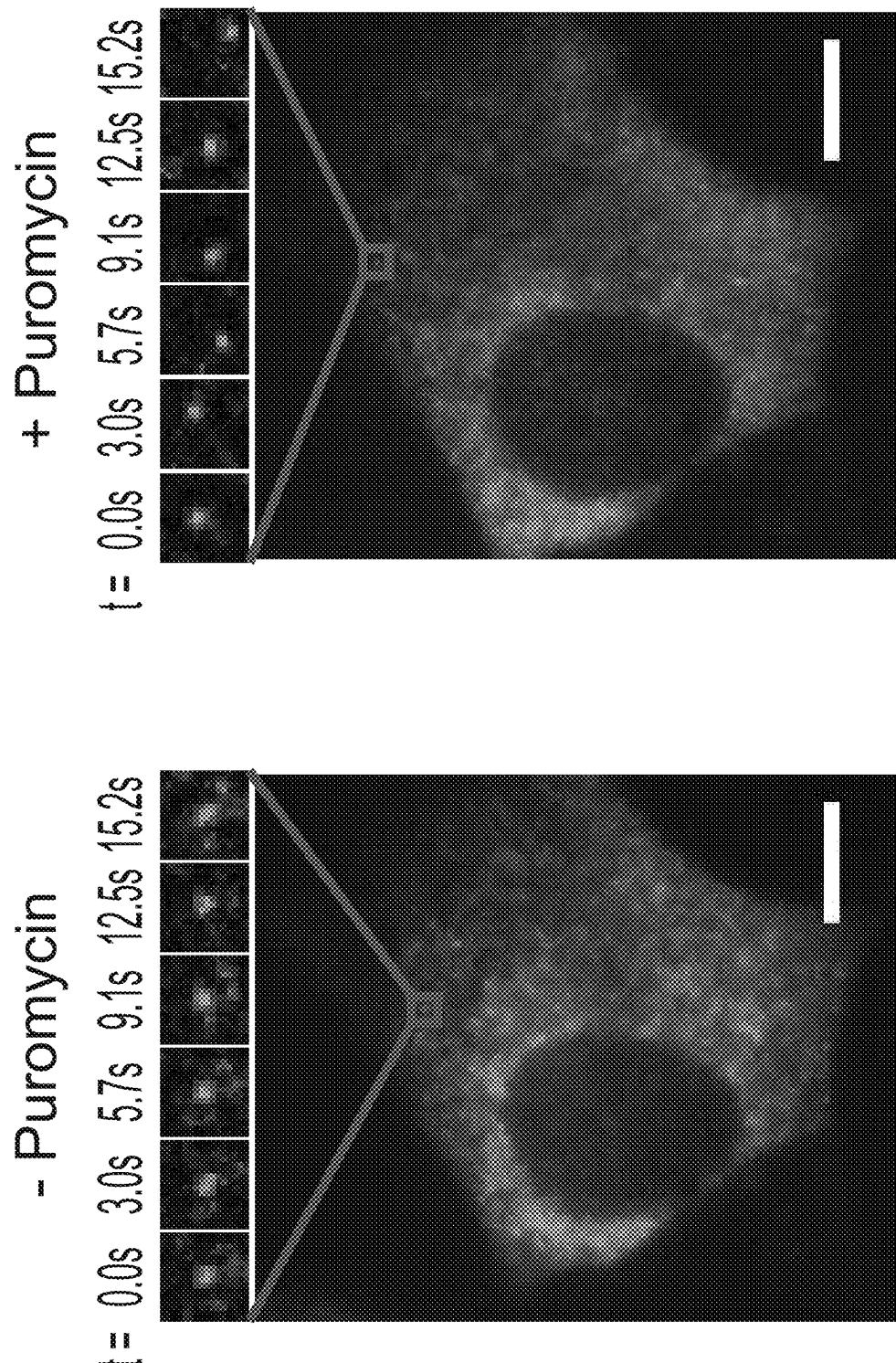


FIG. 14B
FIG. 14C

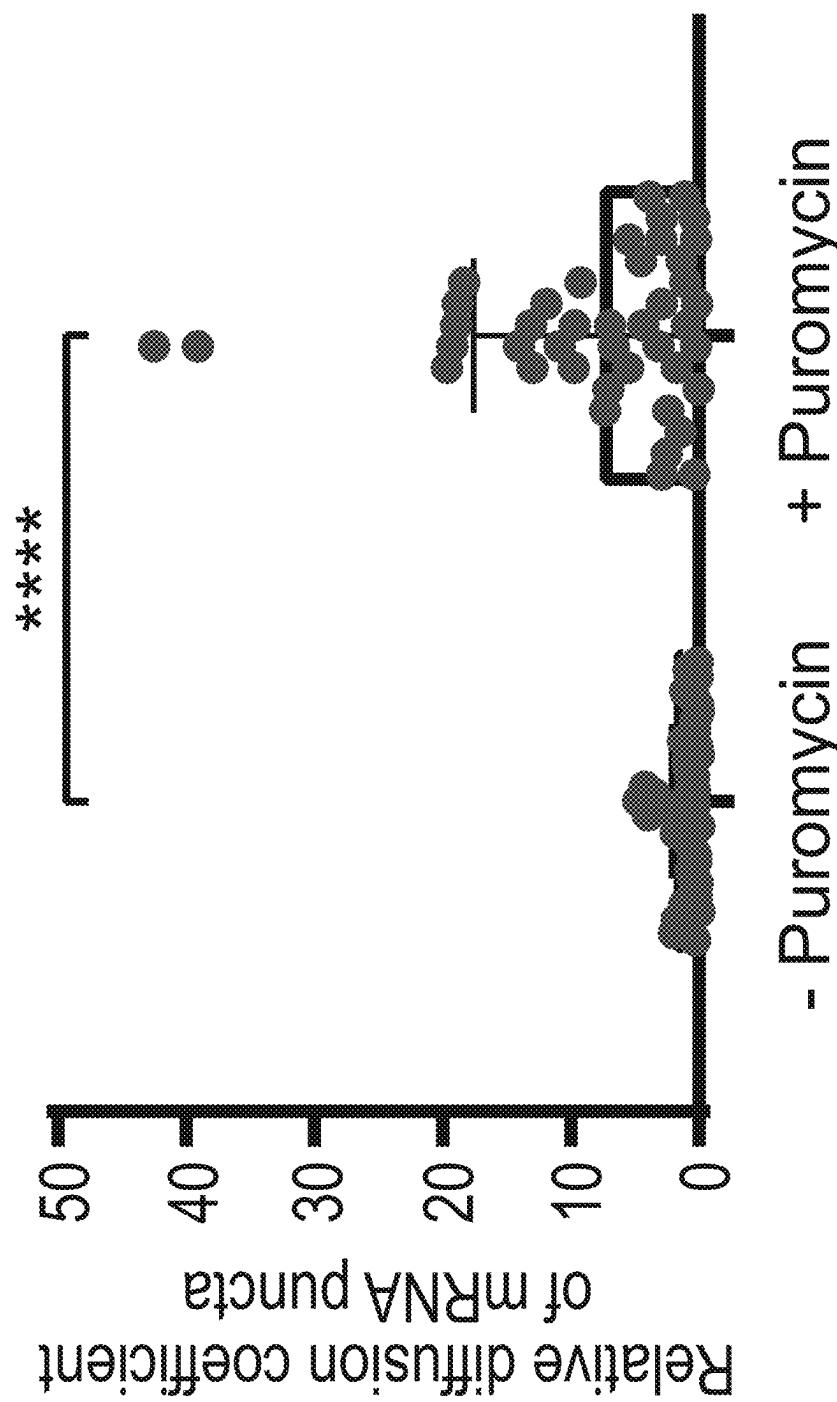


FIG. 14D

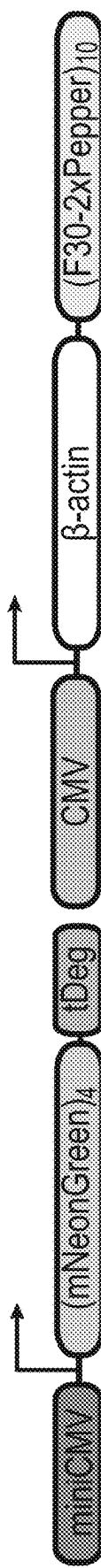


FIG. 15A

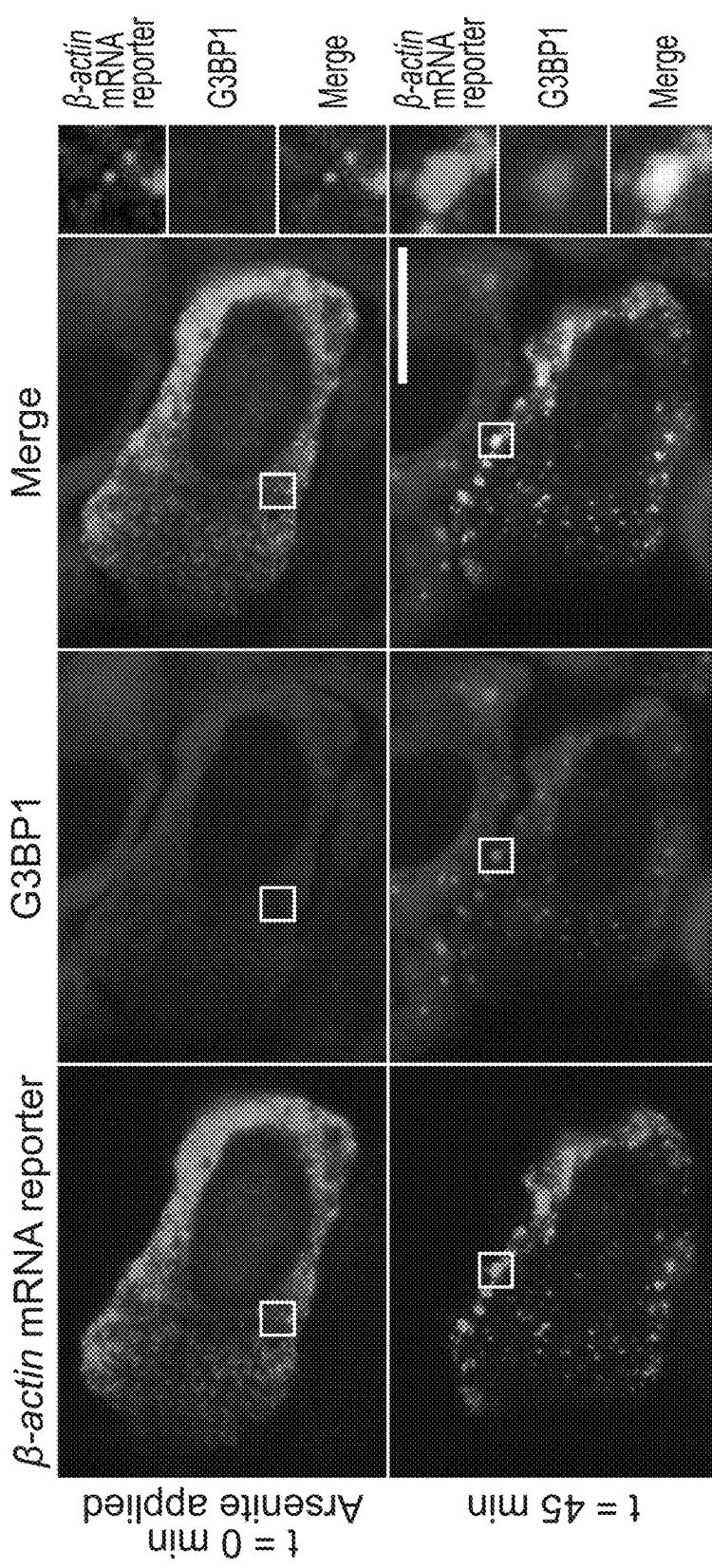


FIG. 15B

G3BP1

Merge

$\beta\text{-actin}$
mRNA
reporter

G3BP1

Merge

$\beta\text{-actin}$
mRNA
reporter

G3BP1

Merge

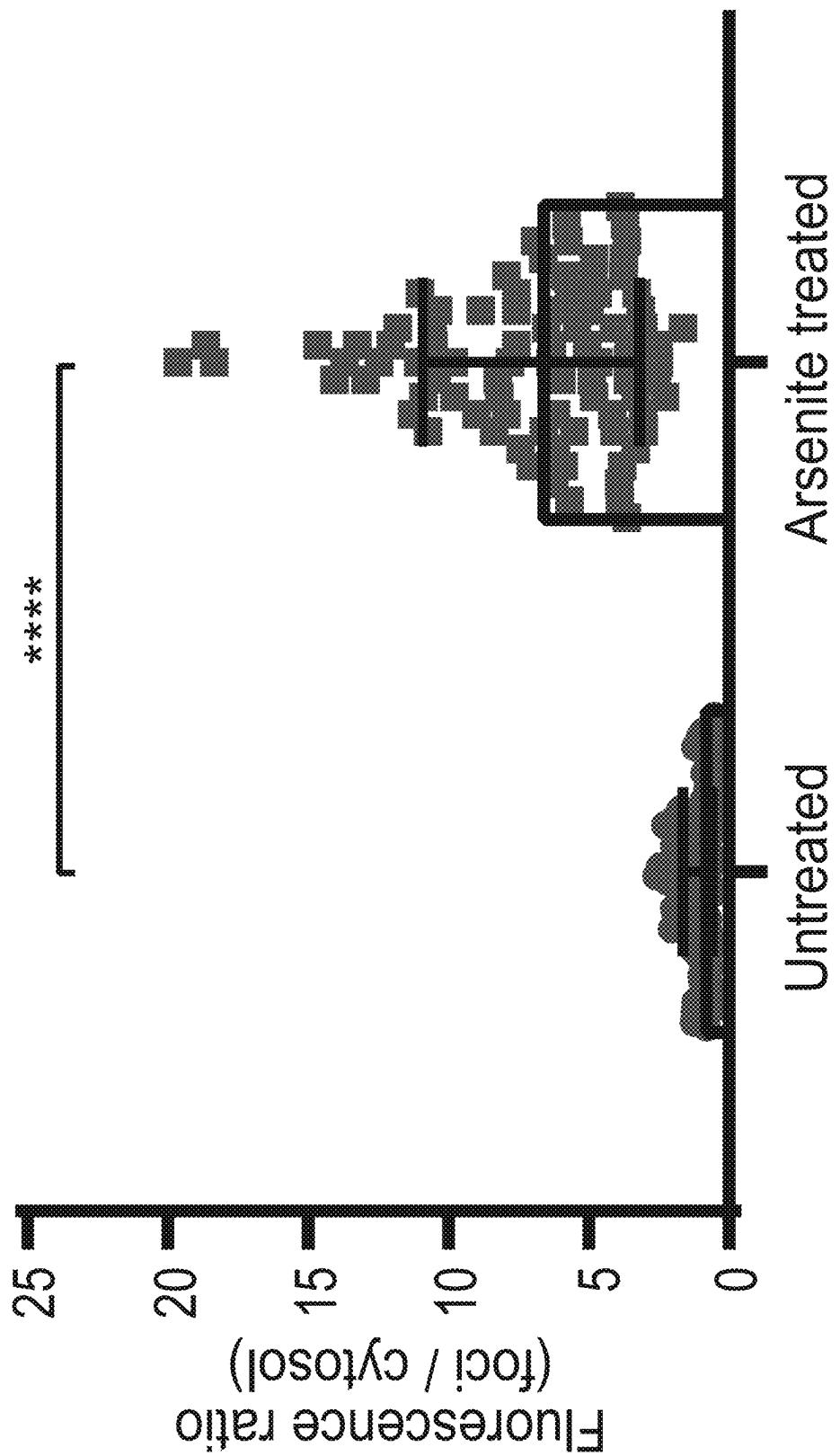


FIG. 15C

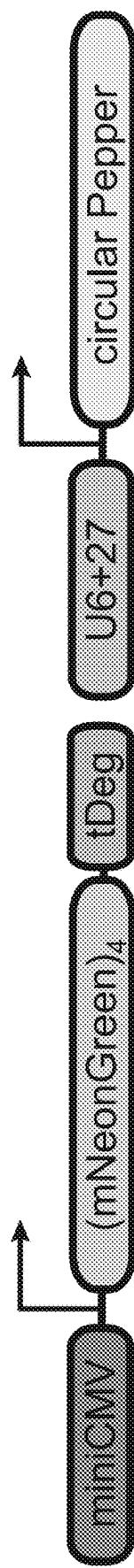


FIG. 16A

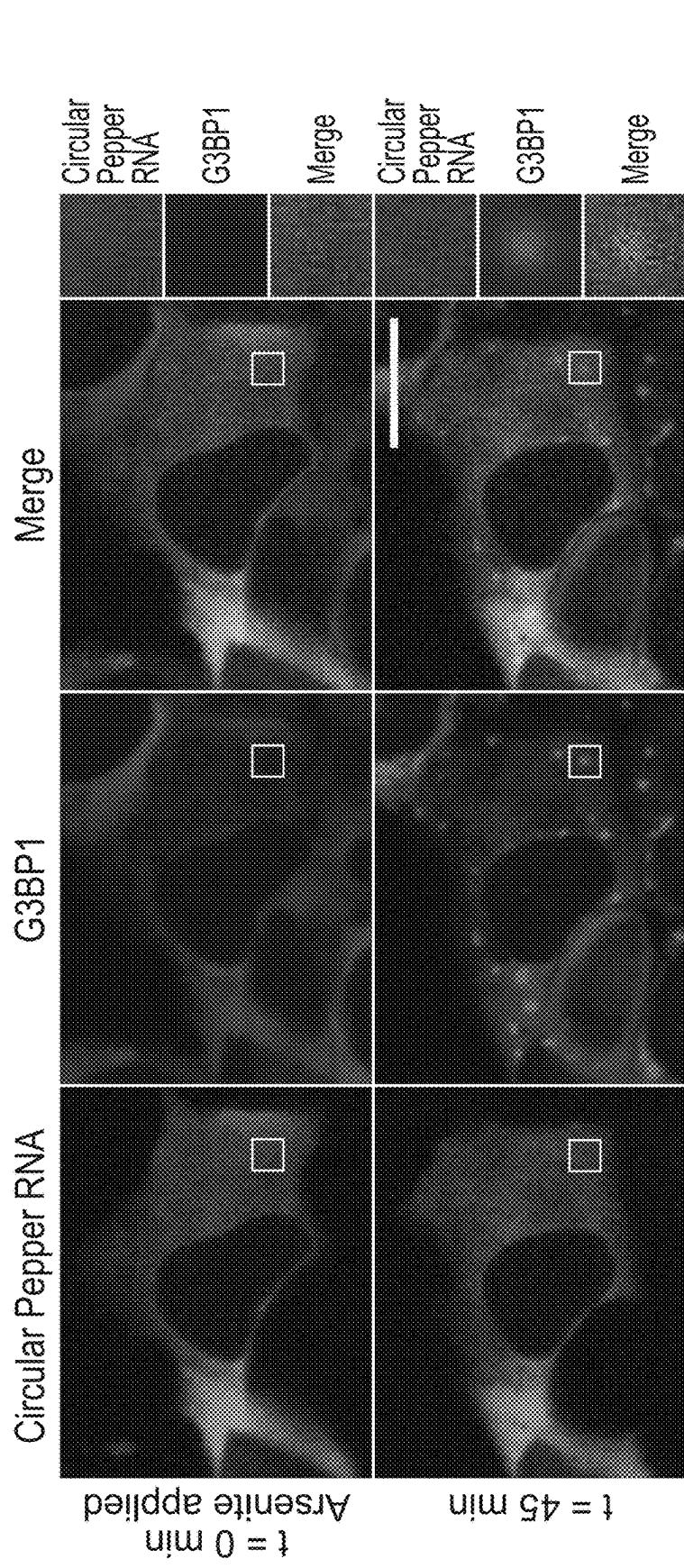


FIG. 16B

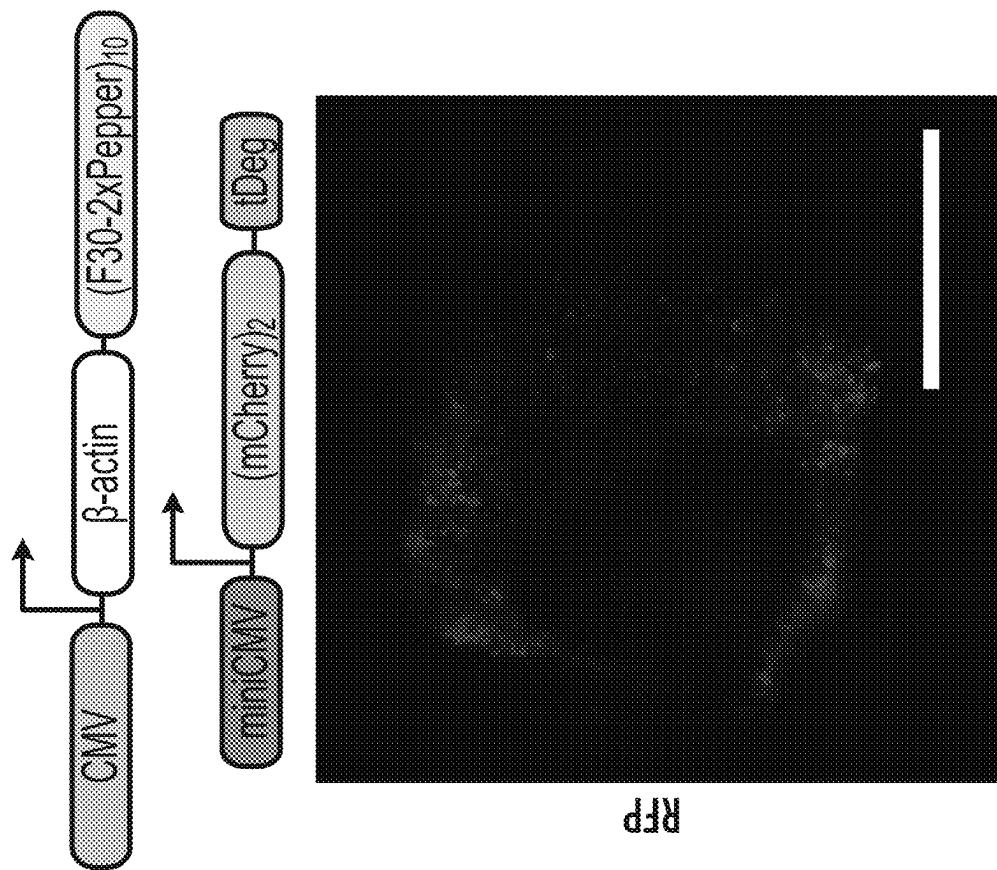


FIG. 17B

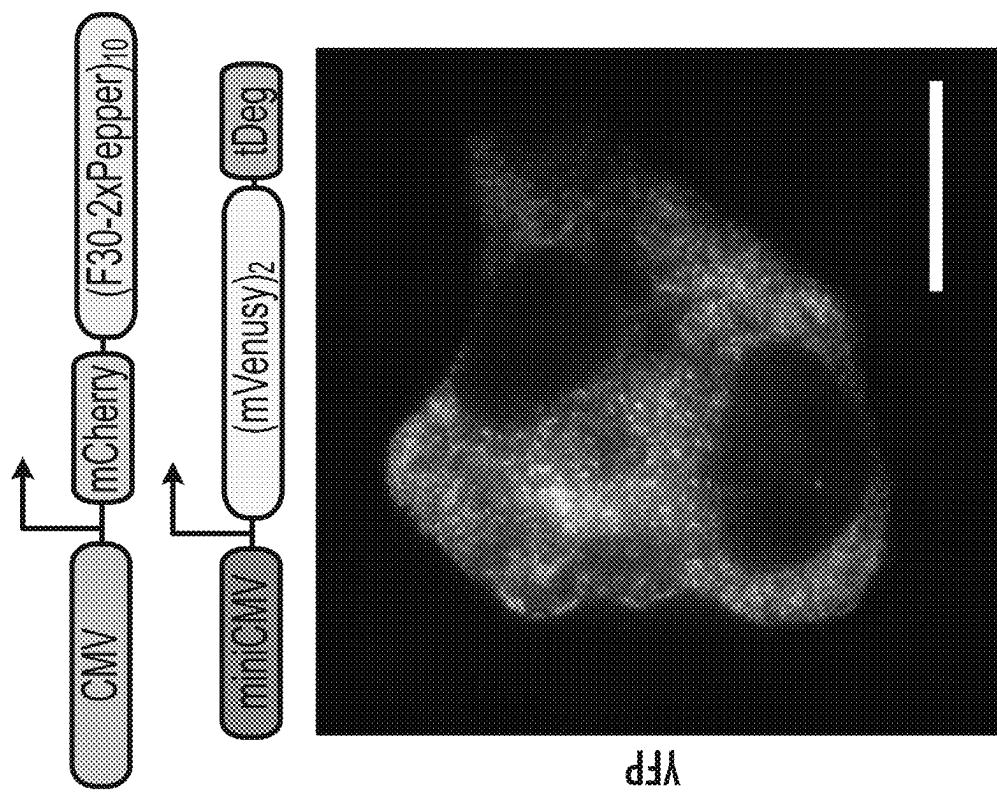


FIG. 17A

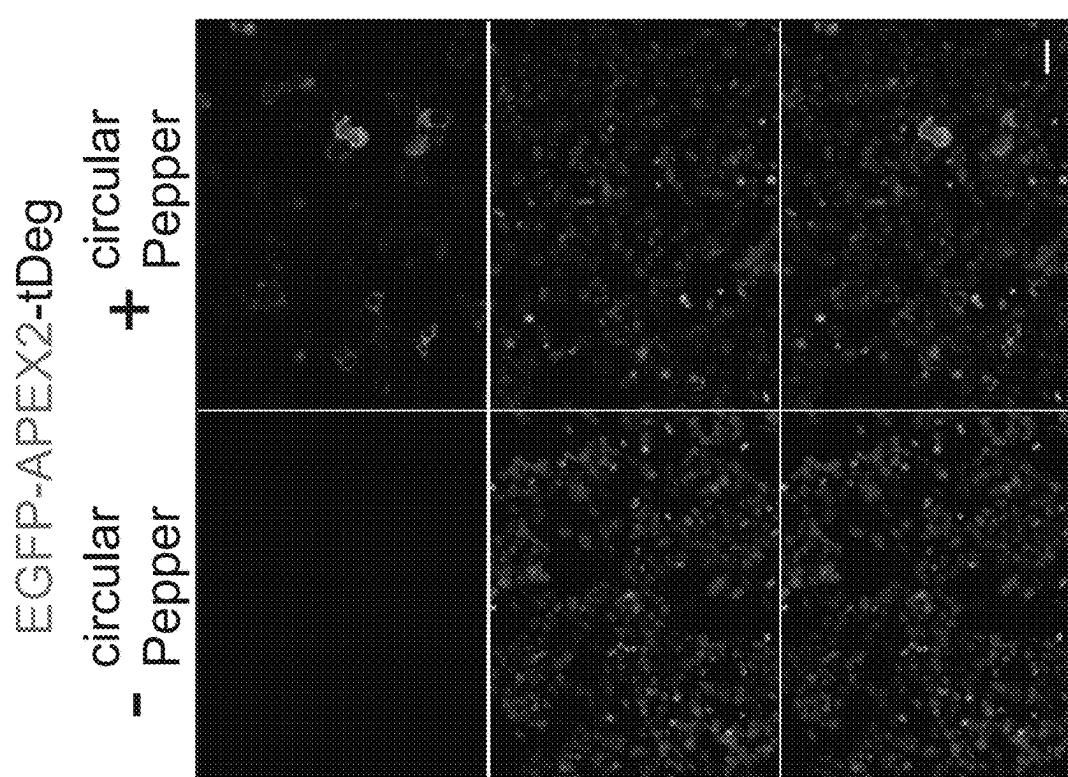


FIG. 18B

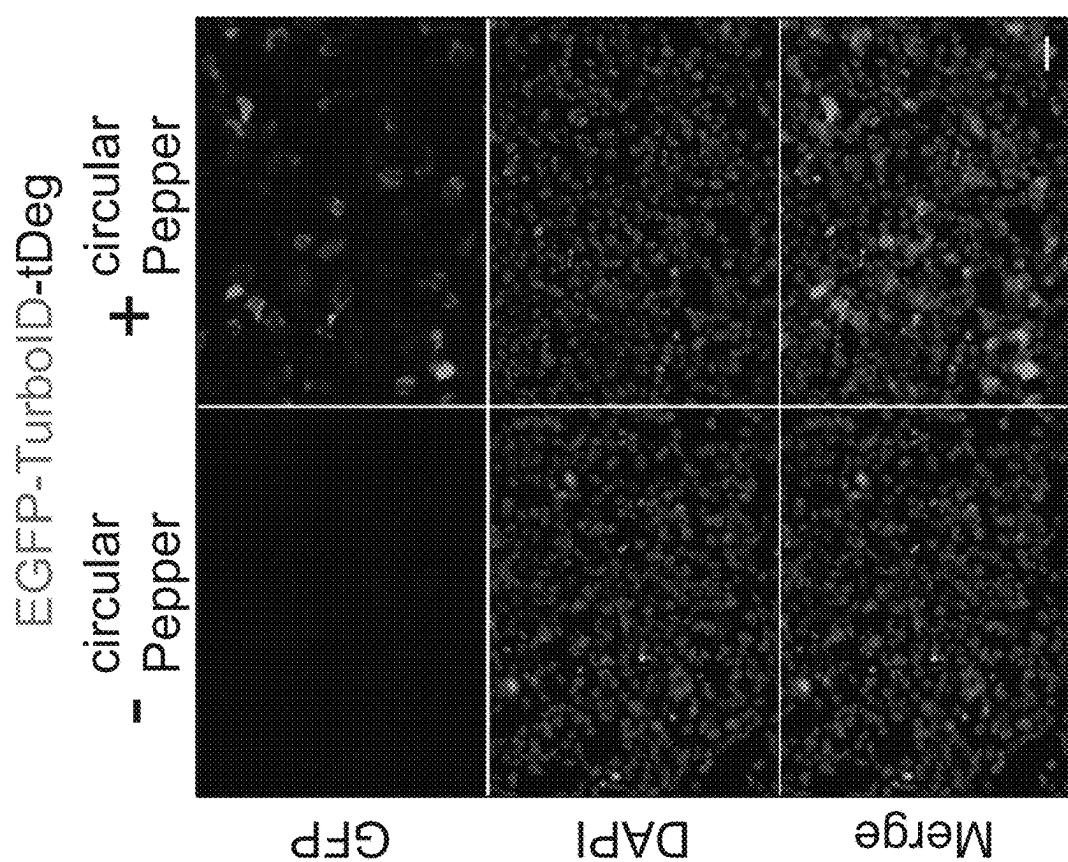


FIG. 18A

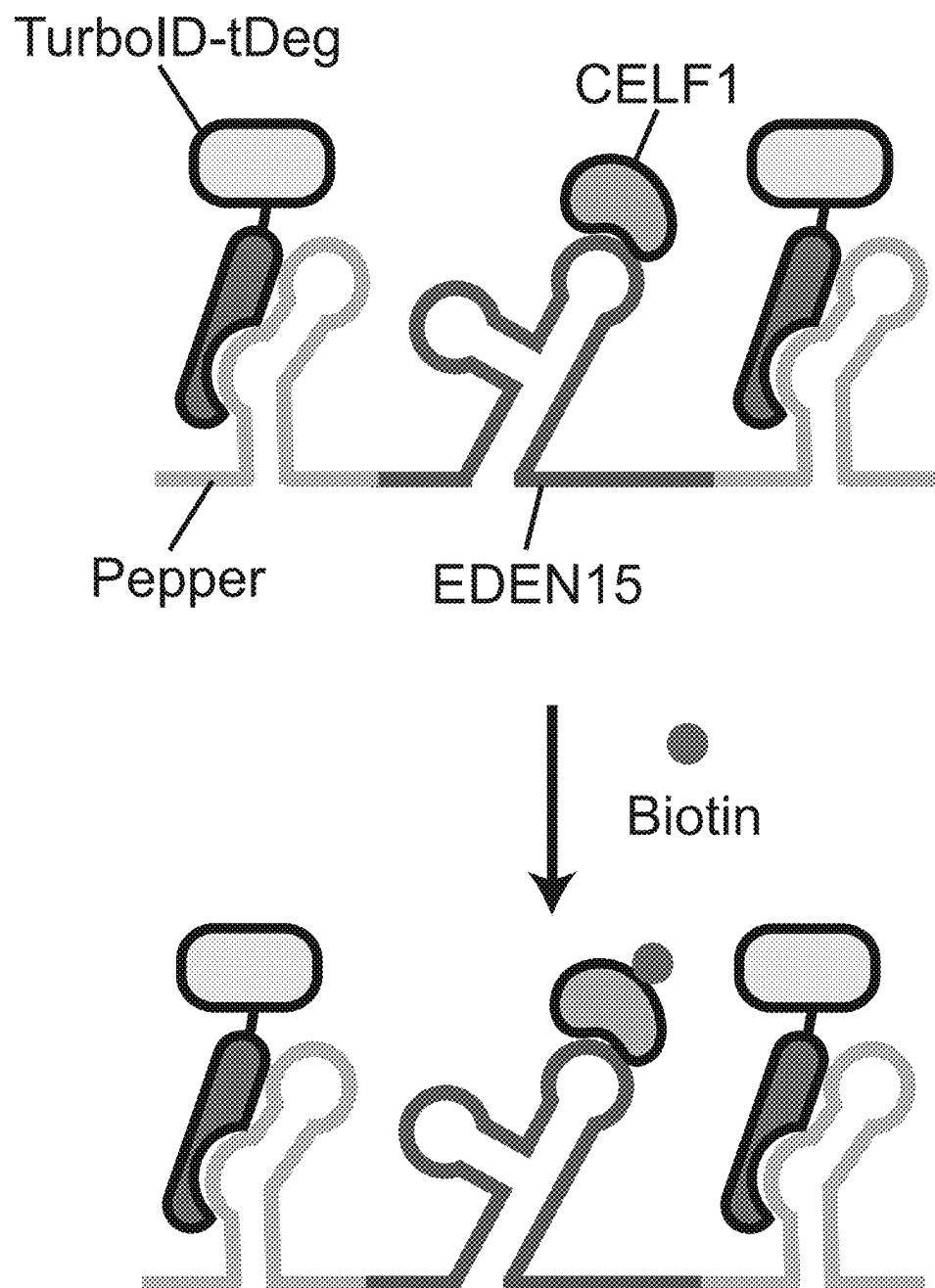


FIG. 18C

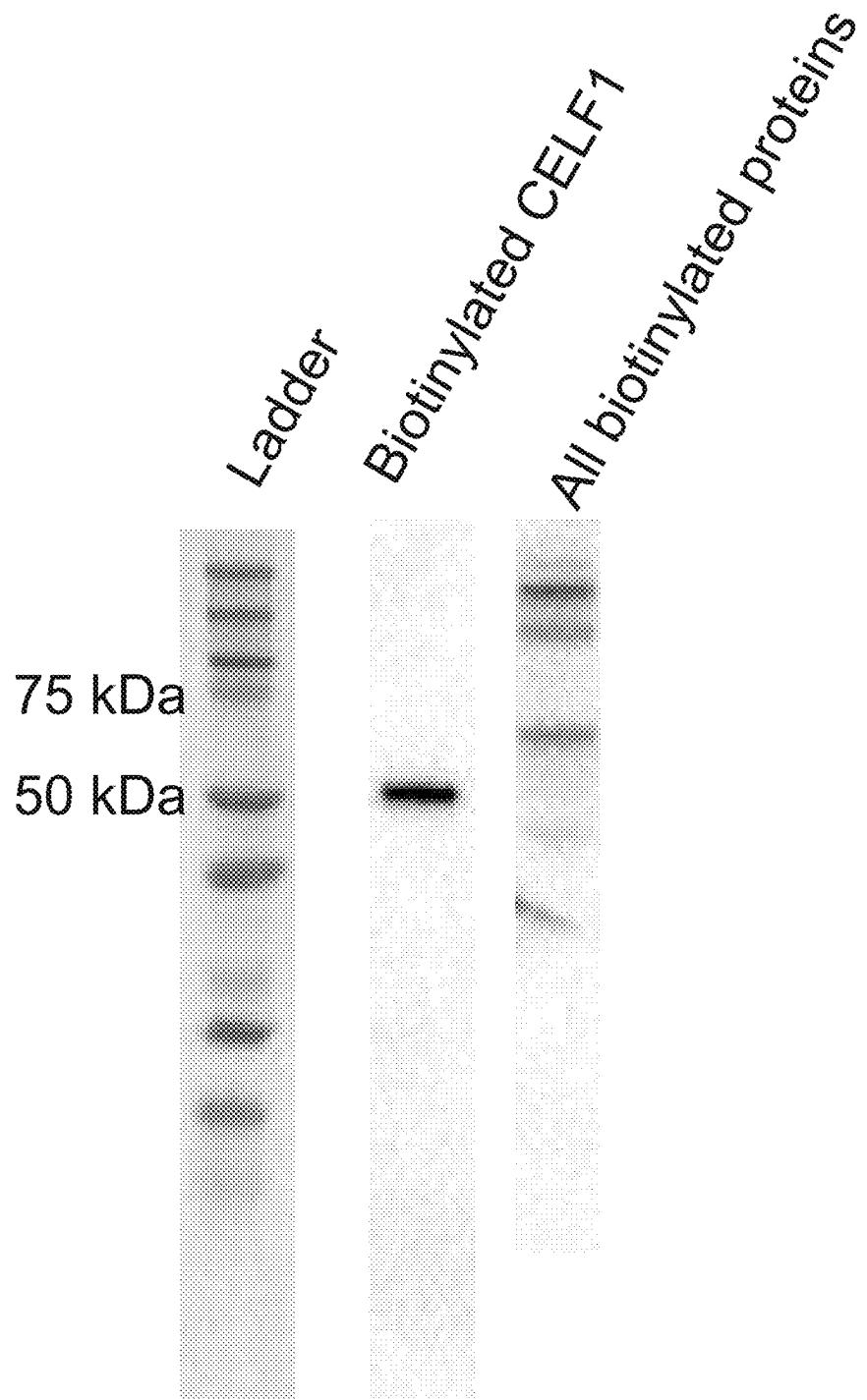


FIG. 18D

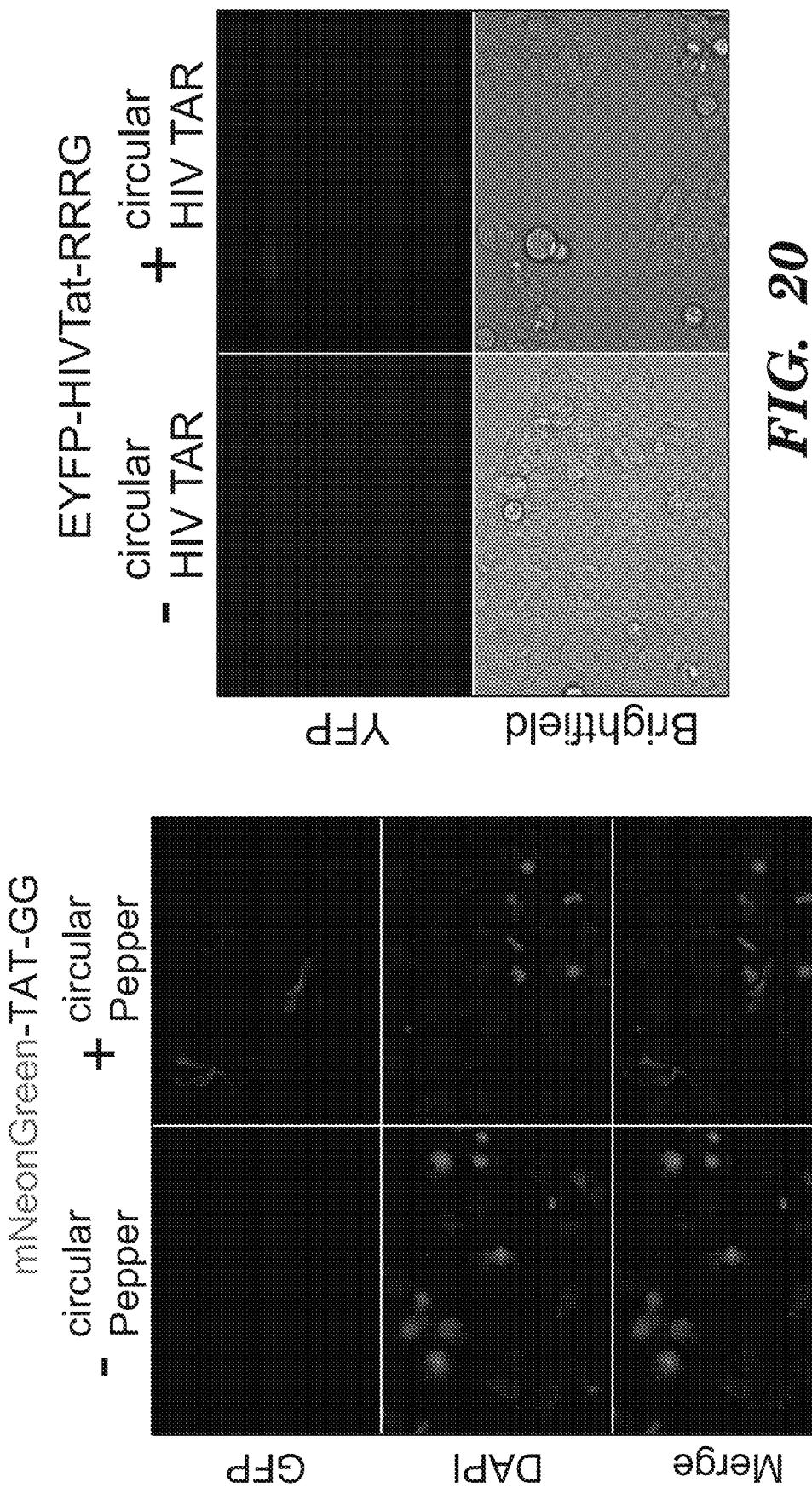


FIG. 19

FIG. 20

1

RNA-REGULATED FUSION PROTEINS AND METHODS OF THEIR USE

This application is a national stage application under 35 U.S.C. § 371 of International Application No. PCT/US2020/048781, filed Aug. 31, 2020, which claims priority benefit of U.S. Provisional Patent Application Ser. No. 62/894,651 filed Aug. 30, 2019, which is hereby incorporated by reference in its entirety.

This invention was made with government support under Grant Number MH109087 awarded by the National Institutes of Health. The government has certain rights in the invention.

FIELD OF THE DISCLOSURE

This present disclosure relates to RNA-regulated fusion proteins and methods of their use.

BACKGROUND

Fluorogenic RNA aptamers are RNA aptamers that bind otherwise nonfluorescent molecules and switch them to a fluorescent form. These fluorogenic dyes can be applied to cells, enabling RNAs tagged with these fluorogenic aptamers to be imaged using fluorescence microscopy (Paige et al., “RNA Mimics of Green Fluorescent Protein,” *Science* 333: 642-646 (2011) and Braselmann et al., “A Multicolor Riboswitch-Based Platform for Imaging of RNA in Live Mammalian Cells,” *Nat. Chem. Biol.* 14:964-971 (2018)). However, few fluorogenic aptamers have been developed since there are not many fluorogenic dyes that meet the criteria required for use in live cells. For example, most dyes show nonspecific fluorescence activation by cellular lipids or DNA (Löber, G., “The Fluorescence of Dye—Nucleic Acid Complexes,” *Journal of Luminescence* 22:221-265 (1981) and Fam et al., “Recent Advances in Fluorescent Probes for Lipid Droplets,” *Materials (Basel)* 11 (2018)). This nonspecific binding leads to background fluorescence that obscures the fluorescence of the RNA-dye complexes. Another problem is that the fluorogenic dyes are not genetically encoded and therefore need to be added exogenously for RNA imaging. A genetically encoded conditionally fluorescent dye would provide a simple alternative to the use of fluorogenic RNA aptamers.

The present disclosure is directed to overcoming deficiencies in the art.

SUMMARY

A first aspect of the disclosure relates to a nucleic acid molecule encoding an RNA-regulated fusion protein. The nucleic acid molecule includes: a first nucleic acid sequence encoding a protein of interest and a second nucleic acid sequence encoding an RNA-regulated destabilization domain, where the second nucleic acid sequence is operably coupled to the first nucleic acid sequence.

Another aspect of the disclosure relates to a nucleic acid molecule encoding a lentiviral transactivator of transcription (Tar) RNA aptamer sequence.

A further aspect of the disclosure relates to an RNA-regulated fusion protein comprising a protein of interest and an RNA-regulated destabilization domain.

Yet another aspect of the disclosure relates to a molecular complex comprising: an RNA-regulated fusion protein comprising (i) a protein of interest and (ii) an RNA-regulated

2

destabilization domain; and an RNA aptamer bound specifically to the RNA-regulated destabilization domain.

Another aspect of the invention relates to a method of imaging RNA in a cell. This method involves providing a first vector encoding an RNA-regulated fusion protein, where the RNA-regulated fusion protein comprises a fluorescent protein, a bioluminescent protein, or an enzyme fused to an RNA-regulated destabilization domain; providing a second vector encoding an RNA molecule comprising (i) an RNA sequence of interest and (ii) an RNA aptamer sequence, where the RNA-regulated destabilization domain specifically binds to the RNA aptamer sequence; transfecting a host cell with the first vector and the second vector; and imaging said transfected cells.

Yet another aspect of the invention relates to a method of imaging RNA in a cell. This method involves providing a vector encoding an RNA-regulated fusion protein, where the RNA-regulated fusion protein comprises a fluorescent protein, a bioluminescent protein, or an enzyme fused to an RNA-regulated destabilization domain; transfecting a host cell with the first vector; contacting said transfected cell with an RNA molecule comprising (i) an RNA sequence of interest and (ii) an RNA aptamer sequence, where the RNA-regulated destabilization domain specifically binds to the RNA aptamer sequence; and imaging said contacted cells.

A further aspect of the invention relates to a method of selectively modifying an RNA-binding protein. This method involves providing a first expression vector encoding an RNA-regulated fusion protein, where the RNA-regulated fusion protein comprises an enzyme fused to an RNA-regulated destabilization domain; providing a second expression vector encoding (i) an RNA sequence of interest and (ii) an RNA aptamer sequence, where the RNA-regulated destabilization domain specifically binds to the RNA aptamer sequence; transfecting a host cell with the first and second expression vectors; and allowing the enzyme to be expressed, where the expressed enzyme selectively modifies a protein that binds to the RNA sequence of interest.

Another aspect of the invention relates to a method of regulating expression of an RNA-stabilized protein of interest. This method involves providing a first vector encoding an RNA-regulated fusion protein, where the RNA-regulated fusion protein comprises a protein of interest fused to an RNA-regulated destabilization domain; providing a second vector encoding an RNA aptamer sequence, where the RNA-regulated destabilization domain specifically binds to the RNA aptamer sequence; providing a host cell comprising a functional ubiquitination system; transfecting the host cell with the first and second expression vectors; and expressing the first and second expression vectors within the host cell, where said expressing the first and second expression vectors regulates proteomic stability of the RNA-regulated fusion protein; and where, in the absence of any expressed

RNA aptamer sequence in the host cell, the RNA-regulated destabilization domain promotes degradation of the RNA-regulated fusion protein by the ubiquitination system; and where the RNA-regulated fusion protein is stabilized by the expressed RNA aptamer sequence.

Another aspect of the invention relates to a method of regulating expression of an RNA-stabilized protein of interest. This method involves providing a first vector encoding an RNA-regulated fusion protein, where the RNA-regulated fusion protein comprises a protein of interest fused to an RNA-regulated destabilization domain; providing a second vector encoding an RNA aptamer sequence, where the RNA-regulated destabilization domain specifically binds to

the RNA aptamer sequence; providing a mammalian cell lysate or solution comprising (i) a ubiquitin ligase, (ii) proteosomal degradation machinery, (iii) transcriptional machinery, and (iv) translational machinery; contacting the mammalian cell lysate or solution with the first and second expression vectors; and expressing the first and second expression vectors, where said expressing the first and second expression vectors regulates proteomic stability of the RNA-regulated fusion protein; and where, in the absence of any expressed RNA aptamer sequence in the cell lysate or solution, the RNA-regulated destabilization domain promotes degradation of the RNA-regulated fusion protein by the proteosomal degradation system; and where the RNA-regulated fusion protein is stabilized by the expressed RNA aptamer sequence.

Another aspect of the present application relates to a treatment method. This method involves contacting a cell with an RNA aptamer, where upon said contacting, the aptamer interacts with an RNA-regulated destabilization domain fused to a protein of interest in the cell to stabilize the protein of interest in the cell.

Another aspect of the present invention relates to a treatment method. This method involves contacting a cell with a vector according to the present application under conditions effective to express an RNA molecule as described herein to treat the cell.

The examples described herein below demonstrate the use of RNA-regulated fluorescent fusion proteins whose fluorescence is stabilized by RNA aptamers. In some embodiments, the RNA-regulated fluorescent fusion proteins are highly unstable until they bind RNA aptamers inserted in mRNAs, resulting in fluorescent RNA-protein complexes that enable live imaging of mRNA in living cells. In some embodiments, the technology described herein is an imaging system that bypasses the limitations of using fluorogenic RNA aptamers and conditionally fluorescent small molecule dyes for imaging. In some embodiments, this is achieved by engineering a peptide degron sequence whose activity can be regulated by an RNA aptamer. When fused to a fluorescent protein, this peptide degron sequence can send the fluorescent protein to degradation. However, this degradation function of the peptide degron is impeded when bound to a specific RNA aptamer sequence. In some embodiments, a peptide degron sequence causes rapid degradation of the unbound fluorescent proteins when expressed in mammalian cells. This is different from previous methods. In some embodiments, methods described herein utilize an RNA aptamer sequence that can effectively abrogate the degradation function of the peptide degron once they are bound. This is also different from previous methods. Methods described herein enable fluorescent proteins and other proteins to carry out their native function only when they are bound to a specific RNA sequence. In the case of enhanced yellow fluorescent protein (EYFP), a 38 fold fluorescent enhancement was observed when bound to the engineered RNA aptamer described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGS. 1A-1C show the design and optimization of an RNA-regulated protein destabilization domain. FIG. 1A is a schematic drawing of a Pepper RNA-regulated protein destabilization domain, tDeg. tDeg is a bifunctional peptide that includes the Tat peptide, which is capable of binding to the Pepper RNA aptamer, and the previously described C-terminal Arg-Arg-Arg-Gly degron (Bonger et al., "Small-Molecule Displacement of a Cryptic Degrone Causes Con-

ditional Protein Degradation," *Nat. Chem. Biol.* 7:531-7 (2011), which is hereby incorporated by reference in its entirety). When fused to a protein of interest, tDeg causes protein degradation. However, the protein destabilization function of tDeg is impeded when it binds to the Pepper RNA aptamer. Amino acids Arg-Gly, highlighted in a black box, are appended to the C-terminus of Tat to make the full Arg-Arg-Arg-Gly degron. FIG. 1B demonstrates that Pepper RNA stabilizes EYFP fused to tDeg in cells. To test whether tDeg functions as an RNA-regulated destabilization domain, EYFP-tDeg was coexpressed with different circular RNAs, and the yellow fluorescence in HEK293T cells was imaged. Without circular wild-type TAR RNA or its variants, cells coexpressing EYFP-tDeg and the circular control RNA only showed minimal fluorescence above background fluorescence. Cells exhibit yellow fluorescence only when circular wild-type TAR RNA, TAR Variant-1, or TAR Vairnat-2 (named Pepper) was coexpressed. Notably, higher yellow fluorescence signals were observed in the cytosol compared to the nucleus when EYFP-tDeg was coexpressed with the circular wild-type TAR RNA or its variants. This is consistent with the cytosolic expression of small circular RNAs using the Tornado expression system (Litke & Jaffrey, "Highly Efficient Expression of Circular RNA Aptamers in Cells Using Autocatalytic Transcripts," *Nat. Biotechnol.* 37:667-675 (2019), which is hereby incorporated by reference in its entirety). All cells were stained with Hoechst dye. Scale bar, 40 μ m. FIG. 1C shows the summary data of normalized fluorescence of untransfected HEK293T cells, or HEK293T cells expressing EYFP or EYFP-tDeg with different RNAs as in (FIG. 1B). Total cellular yellow fluorescence of individual cells is plotted ($n=4$ independent cell cultures). Values are means \pm s.d.

****P_{circular} *wild-type* $TAR = 7.9 \times 10^{-113}$;
 ****P_{circular} *TAR* $TAR_{Variant-1} = 2.1 \times 10^{-117}$;
 ****P_{circular} *TAR Variant-2* $= 1.7 \times 10^{-115}$ by one-way ANOVA.

FIGS. 2A-2B are schematic illustrations showing the design of tDeg, an RNA-regulated destabilization domain. Shown is a structural representation of how TAR binds to the tDeg, and may therefore obstruct recognition of the Arg-Arg-Arg-Gly degradation-inducing signal. RNA is depicted in grey, and peptide sequence is shown letters of the polypeptide chain. A schematic representation of RNA binding to the tDeg sequence is shown in FIG. 2A. Here, a bifunctional peptide sequence, called tDeg, that functions both as a destabilization domain and as a binding site for the bovine immunodeficiency virus TAR RNA (in grey) was designed. Knowing that the TAR RNA binds to specific amino acids in the Tat peptide including the two C-terminal arginines, an Arg-Gly (highlighted in a black box) was added to the C-terminus of the Tat peptide to make the full Arg-Arg-Arg-Gly degron. When the TAR RNA binds to this bifunctional domain, it impedes the function of the destabilization domain by sterically blocking recognition of the Arg-Arg-Arg-Gly degron by proteasomal machinery. The structure model (FIG. 2B) of the Tat-TAR complex shows that the first two arginines of the Arg-Arg-Arg-Gly degron would be inaccessible to any Arg-Arg-Arg-Gly-binding protein that mediates its degradation. The additional Arg-Gly residues are modeled into the C-terminus of Tat in a black box in FIG. 2B. The structure representation in FIG. 2B is based on the NMR structure of the bovine immunodeficiency virus Tat-TAR complex (PDB entry: 1BIV) (Puglisi et al., "Solution Structure of a Bovine Immunodeficiency Virus Tat-TAR Peptide-RNA Complex," *Science* 270:1200-3 (1995), which is hereby incorporated by reference in its entirety).

FIGS. 3A-3B demonstrate that tDeg confers protein instability to EYFP by proteasomal degradation. In FIG. 1B, it

was shown that tDeg confers protein instability to EYFP. However, the lack of yellow fluorescence of EYFP-tDeg in FIG. 1B could be due to protein misfolding or aggregation. In FIG. 3A, whether the lack of yellow fluorescence of EYFP-tDeg is due to proteasomal degradation was examined. In these experiments, HEK293T cells were transiently transfected with a plasmid expressing EYFP-tDeg. These cells were then treated with vehicle (DMSO) or a proteasome inhibitor (10 μ M MG132) for 7 hours, respectively. When treated with vehicle (DMSO), minimal yellow fluorescence was detected. This result is consistent with the result from FIG. 1B. However, when proteasome activity was inhibited by treatment of 10 μ M MG132 for 7 hours, the yellow fluorescence of EYFP-tDeg was restored. Thus, this confirmed that the tDeg tag markedly reduces the stability of EYFP by inducing its proteasomal degradation. All cells were stained with Hoechst dye. Scale bar, 40 μ m. In FIG. 3B, normalized total cellular yellow fluorescence of individual cells is plotted (n=3 independent cell cultures). Values are means \pm s.d. ***P=5.6 \times 10 $^{-36}$ by unpaired two-tailed Student's t-test.

FIGS. 4A-4B demonstrate that engineered TAR variants' higher efficiency in stabilizing EYFP-tDeg proteins is not due to expression differences in EYFP-tDeg mRNA or the circular TAR RNAs. In FIGS. 1B and 1C, it was shown that circular wild-type TAR, Variant-1, and Variant-2 showed 24-fold, 36-fold, and 38-fold fluorescence increases, respectively. However, the improved efficiency in stabilizing EYFP-tDeg protein could be due to uneven expression levels of the EYFP-tDeg mRNA, or the uneven expression levels of the circular TAR RNA variants. Here, the relative expression of EYFP-tDeg mRNA (FIG. 4A) and the relative expression of circular TAR RNA variants (FIG. 4B) was compared. In these experiments, HEK293T cells were transiently transfected with a plasmid expressing EYFP-tDeg and the corresponding circular TAR RNA variant as shown in FIGS. 1B and 1C. Total RNA was extracted by TRIzol® extraction. EYFP-tDeg mRNA expression level was quantified using RT-qPCR. Each circular TAR RNA variant's expression level was quantified by running the extracted total RNA on a TBE-Urea gel followed by SYBR™ Gold nucleic acid gel staining. These results show that there is no significant expression difference in the EYFP-tDeg mRNA or the circular TAR RNA variants. Thus, this confirms that the engineered circular TAR RNA variants indeed show higher efficiency in stabilizing tDeg-tagged EYFP. Data were collected from two independent cell cultures. Values are means \pm s.d.

FIGS. 5A-5G demonstrate that tDeg can be regulated by the Pepper RNA aptamer in diverse mammalian cell types. In FIGS. 1A-1C, it was shown that EYFP-tDeg can be regulated by the Pepper RNA aptamer in HEK293T cells. Here, whether tDeg can be regulated by the Pepper RNA aptamer in various mammalian cell types was examined (FIG. 5A). In these experiments, U2OS cells (FIG. 5B, FIG. 5E), COS-7 cells (FIG. 5C, FIG. 5F), or HeLa cells (FIG. 5D, FIG. 5G) were transiently expressed EYFP-tDeg with and without the circular Pepper RNA aptamer, respectively. In each case, cells showed low or undetectable levels of yellow fluorescence without the circular Pepper RNA aptamer. The yellow fluorescence of EYFP-tDeg was only restored when the circular Pepper RNA aptamer was coexpressed. Thus, tDeg can be regulated by the Pepper RNA aptamer in diverse mammalian cell types. All cells were stained with Hoechst dye. Scale bar, 20 μ m. Normalized total cellular fluorescence (FIGS. 5E, 5F, and 5G) of individual cells is plotted (n=3 independent cell cultures). Values

are means \pm s.d. ***P_{U2OS}=5.7 \times 10 $^{-59}$; ***P_{COS-7}=1.6 \times 10 $^{-46}$; ***P_{HeLa}=2.0 \times 10 $^{-139}$ by unpaired two-tailed Student's t-test.

FIGS. 6A-6G demonstrate that tDeg confers Pepper RNA-dependent regulation to diverse proteins. To test whether Pepper RNA stabilizes different proteins fused to tDeg, HEK293T cells expressing mNeonGreen (FIG. 6B, FIG. 6E), mCherry (FIG. 6C, FIG. 6F), and the luciferase NanoLuc (FIG. 6D, FIG. 6G) fused to a C-terminal tDeg tag with and without circular Pepper RNA (FIG. 6A) were imaged, respectively. In each case, there was a considerable increase of fluorescence (FIG. 6E, FIG. 6F) or bioluminescence (FIG. 6G) of the tDeg-tagged protein only when circular Pepper RNA was coexpressed in cells. For detecting bioluminescence, cells were incubated in media with fumimazine (from Promega Nano-Glo® Luciferase Assay System, diluted 100 \times) and imaged using a 460 \pm 25 nm emission filter cube. All cells were stained with Hoechst dye. Scale bar, 40 μ m. Normalized total cellular fluorescence (FIG. 6E and FIG. 6F) or bioluminescence (FIG. 6G) of individual cells is plotted (n=3 independent cell cultures). Values are means \pm s.d.

****P_{mNeonGreen-tDeg}=1.1 \times 10 $^{-123}$; ****P_{mCherry-tDeg}=3.0 \times 10 $^{-131}$; ****P_{NanoLuc-tDeg}=1.7 \times 10 $^{-120}$ by unpaired two-tailed Student's t-test.

FIGS. 7A-7G demonstrate that tDeg confers Pepper RNA-dependent regulation to diverse proteins. In FIGS. 6A-6G, it was shown that tDeg confers Pepper RNA-dependent regulation of different fluorescent proteins and the luciferase, NanoLuc (Hall et al., "Engineered Luciferase Reporter from a Deep Sea Shrimp Utilizing a Novel Imidazopyrazinone Substrate," *ACS Chem. Biol.* 7:1848-57 (2012), which is hereby incorporated in its entirety). Whether tDeg confers Pepper-dependent regulation to proteins with different functions and localizations in cells was tested here (FIG. 7A). In these experiments, HEK293T cells transiently expressed EGFP-TetR-tDeg (FIG. 7B, FIG. 7E), EGFP-EZH2-tDeg (FIG. 7C, FIG. 7F), or mCherry-NF- κ B-tDeg (FIG. 7D, FIG. 7G), with and without the circular Pepper RNA aptamer, respectively. In each case, proteins were nearly undetectable unless coexpressed with the circular Pepper RNA. Furthermore, protein localization of these proteins without tDeg and the circular Pepper RNA was compared to their stabilized counterparts by tDeg and circular Pepper RNA. It was observed that EGFP-TetR-tDeg with circular Pepper RNA showed more green fluorescent signals in the cytosol compared to EGFP-TetR. Significant change of protein localization in the case of EGFP-EZH2-tDeg or mCherry-NF- κ B-tDeg with the circular Pepper RNA was not observed. It was concluded that tDeg is a versatile tag for RNA-dependent protein stabilization. All cells were stained with Hoechst dye. Scale bar, 40 μ m. Normalized total cellular fluorescence (FIGS. 7E, 7F, and 7G) of individual cells is plotted (n=3 independent cell cultures). Values are means \pm s.d. ****P_{EGFP-TetR4Deg}=2.9 \times 10 $^{-136}$; ****P_{EGFP-EZH2-tDeg}=1.1 \times 10 $^{-120}$; ****P_{mCherry-NF- κ B-tDeg}=3.5 \times 10 $^{-119}$ by unpaired two-tailed Student's t-test.

FIGS. 8A-8B demonstrate the optimization of a concatenated Pepper tag to image mRNAs in live cells. Pepper RNA-regulated fluorescent proteins were used to fluorescently tag mRNAs in live cells. As a first step, the best way to incorporate the Pepper aptamers in the 3'UTR of a transcript of interest was determined. In these experiments, a fluorescent protein (mNeonGreen)₂-tDeg and an mCherry mRNA reporter (FIG. 8A) containing 3'UTR tags comprising 10 or 20 concatenated Pepper aptamers with and without a folding scaffold, F30, were expressed respectively. In the case of the (Pepper)₂₀ and (F30-2 \times Pepper)₁₀ tags, mobile

green fluorescent puncta in the cytosol were observed (FIG. 8B). A signal to noise ratio was evident when the (F30-2×Pepper)₁₀ tag (signal to noise ratio=1.8) was used, compared to the (Pepper)₂₀ tag (signal to noise ratio=1.5). However, puncta were not readily detectable with either the (Pepper)₁₀ tag or the (F30-1×Pepper)₁₀ tag. Therefore, the (F30-2×Pepper)₁₀ tag was used to image mRNAs in the subsequent experiments. Scale bar, 20 μm. This experiment was performed three times with similar results.

FIGS. 9A-9D show the design of Pepper tags for imaging mRNA. Design and sequences of four Pepper tags used in FIG. 8B: (Pepper)₁₀ (FIG. 9A; SEQ ID NO: 119), (F30-1×Pepper)₁₀ (FIG. 9B; SEQ ID NO: 120), (Pepper)₂₀ (FIG. 9C; SEQ ID NO: 121), and (F30-1×Pepper)₁₀ (FIG. 9D; SEQ ID NO: 122).

FIGS. 10A-10C demonstrate the optimization of the number of fluorescent mNeonGreen monomers in the fluorescent protein for imaging mRNA in live cells. In FIG. 8B, it was observed that (F30-2×Pepper)₁₀ is the optimal tag for imaging mRNAs in live cells. To further optimize the system of using Pepper RNA-regulated fluorogenic protein to image mRNAs, it was determined whether increasing the number of fluorescent mNeonGreen could increase the fluorescence signal to background noise ratio of the mobile green fluorescent puncta. In these experiments, an mCherry mRNA reporter tagged with (F30-2×Pepper)₁₀ and tandem fluorescent mNeonGreen with 2, 3, or 4 copies were transiently expressed, respectively, in cells. Here, an increase of fluorescence intensity of the green fluorescent puncta as the number of tandem mNeonGreen increased from 2, 3, to 4 copies, respectively (FIG. 10B) and (FIG. 10C) was observed. mRNAs tagged with (F30-1×Pepper)₁₀ using the (mNeonGreen)₄-tDeg fluorescent fusion protein were also re-tested. It was shown that puncta were detectable, but not as pronounced as when the (F30-2×Pepper)₁₀ tag was used. Thus, it was concluded that (mNeonGreen)₄-tDeg provides a high signal to noise ratio for imaging mRNAs. Scale bar, 20 μm. FIG. 10C is a graph showing the fluorescence intensity of green fluorescent puncta of individual cells is plotted (n=3 independent cell cultures). Values are means±s.d.

$$\text{****P}_{(\text{Pepper})20:(\text{F30-2}\times\text{Pepper})10}=4.6\times10^{19};$$

****P_{(mNeonGreen)2-tDeg:(mNeonGreen)3-tDeg}=7.7×10⁻⁹; ****P_{(mNeonGreen)2-tDeg:(mNeonGreen)4-tDeg}=2.5×10⁻²⁹; ****P_{(mNeonGreen)3-tDeg:(mNeonGreen)4-tDeg}=2.0×10⁻⁹; ****P_{(F30-2×Pepper)10:(F30-1×Pepper)10}=5.6×10⁻¹⁷ by one-way ANOVA.

FIGS. 11A-11C demonstrate that Pepper tag enables visualization of both nuclear and cytosolic mRNAs. FIG. 11A is a schematic representation of the DNA plasmid constructs used for imaging mRNAs in the nucleus and cytosol. To image nascent transcription of mRNA, cells coexpressing an mCherry mRNA reporter containing a 3'UTR green Pepper mRNA tag, (F30-2×Pepper)₁₀, and a green fluorescent fusion protein, (mNeonGreen)₄-tDeg were imaged (FIG. 11B). Cytosolic green fluorescent puncta reflecting mCherry mRNA transcripts and nuclear green fluorescent puncta, potentially reflecting mCherry mRNA transcripts were observed. Less green fluorescent puncta in the nucleus were observed as compared to the cytosol. This potentially reflects that most of the nuclear mCherry mRNA transcripts were exported out of the nucleus. Scale bar, 20 μm. FIG. 11C is a graph providing summary data of cytosolic and nuclear mRNA fluorescence intensity in FIG. 11B (n=201 fluorescent puncta). Values are means±s.d. This experiment was performed three times with similar results.

FIGS. 12A-12D demonstrate that Pepper tag and fluorescent fusion protein enable visualization of individual

mRNAs. To examine whether the puncta observed when imaging Pepper-tagged mRNAs might be stable degradation intermediates, northern blot was performed on total RNA extracted from cells expressing (F30-2×Pepper)₁₀-tagged mCherry RNA transcripts with and without coexpressing the fluorescent fusion protein, (mNeonGreen)₄-tDeg. In these experiments, only full-length mRNA transcript was detected (FIG. 12A). Therefore, it was concluded that the fluorescent puncta in cells largely reflects the full-length transcript, and that degraded or liberated Pepper aptamers do not accumulate in cells. To assess whether the mobile green fluorescent puncta seen in cells expressing Pepper-tagged mRNA represent single mRNAs, a previously described mRNA imaging method in which the resulting puncta were validated to represent single mRNA was used (Yan et al., "Dynamics of Translation of Single mRNA Molecules In Vivo," *Cell* 165:976-89 (2016), which is hereby incorporated by reference in its entirety). This system uses 24 PP7 RNA hairpins in the 3'UTR of a reporter mRNA, and a 3×mCherry-CAAX protein fused to PCP (PP7 coat protein), the PP7-binding protein. The PCP-3×mCherry-CAAX fusion protein is anchored to the membrane via the CAAX sequence, which reduces puncta motility and facilitates quantitative fluorescence measurements. A PP7-containing reporter mRNA was imaged with and without the (F30-2×Pepper)₁₀ tag (FIG. 12B). The (mNeonGreen)₄-tDeg fluorescent fusion protein was used to image the Pepper-tagged mRNAs. If the Pepper tag or the green fluorescent fusion protein caused mRNA to aggregate, the Pepper-tagged reporter mRNA puncta would have been expected to have higher red fluorescence (from PCP-3×mCherry-CAAX) compared to the reporter mRNA puncta without the Pepper tag. The results of these experiments showed that the red fluorescence intensity distribution of the reporter mRNA is not significantly different with and without the Pepper tag (FIG. 12C) (Black bars, 19 cells, 485 mRNAs; Shaded bars, 13 cells, 384 mRNAs). This suggests that the Pepper tag and the green fluorescent fusion protein do not cause mRNA aggregation. Furthermore, colocalization between the green and magenta fluorescent puncta was observed only when the reporter mRNA contained the Pepper tag (FIG. 12D). These results suggest that the green fluorescent puncta observed using the Pepper tag and green fluorescent fusion protein are indeed individual mRNAs. Scale bar, 5 μm (left panel in FIG. 12D), 1 μm (right panel in FIG. 12D). In FIG. 12D, the experiment of reporter mRNA with Pepper was performed three times with similar results, the experiment of reporter mRNA without Pepper was performed twice with similar results.

FIGS. 13A-13E demonstrate that Pepper tag and fluorescent fusion protein do not have observable effects on mRNA turnover kinetics, mRNA translation efficiency, or proteasome activity in cells. To test whether adding the Pepper tag to an mRNA transcript affects its stability, reporter plasmids expressing mCherry transcripts with and without the (F30-2×Pepper)₁₀ tag were constructed. HEK293T cells were transfected with these two reporter plasmids, respectively. In each case, the same cells were cotransfected with the (mNeonGreen)₄-tDeg fluorescent fusion protein. The cells were treated with 5 μg/mL actinomycin D to inhibit new transcription. The amount of reporter mRNA transcripts remaining at each time point was quantified by RT-qPCR at t=0, 1, 2, 4, and 6 hours of actinomycin D treatment. The results showed that fusing the Pepper tag to the reporter mRNA (half-life=5.9 hours) does not significantly affect its turnover rate compared to its untagged counterpart (half-life=6.0 hours) (FIG. 13A). Thus, these data suggest that Pepper-tagged mRNA transcripts have similar turnover kinetics as mRNAs without the Pepper tag. Data were

collected from 2 independent cell cultures. Values are means \pm s.d. To test whether adding the Pepper tag to an mRNA transcript affects its protein translation efficiency, the protein translation efficiency of an mCherry mRNA was compared with and without the (F30-2 \times Pepper)₁₀ Pepper tag. HEK293T cells expressing mCherry mRNA or mCherry-(F30-2 \times Pepper)₁₀ mRNA were harvested. The amount of mCherry protein and mCherry mRNA was quantified by western blotting and RT-qPCR, respectively. A slight decrease of mRNA levels in the Pepper-tagged mCherry mRNA was observed compared to its untagged counterpart (FIG. 13C). The same phenomenon was also observed in the mCherry mRNA tagged with the 24 \times MS2 hairpins (Wu et al., "Synonymous Modification results in High-Fidelity Gene Expression of Repetitive Protein and Nucleotide Sequences," *Genes Dev.* 29:876-86 (2015), which is hereby incorporated by reference in its entirety). This may due to the longer transcript length associate with 3'UTR-tagged mRNAs. Protein translation efficiency was calculated by normalizing the amount of mCherry protein to the amount of mCherry mRNA (FIGS. 13B-13D). No significant difference in protein translation efficiency was found between the untagged mCherry mRNA transcript and the Pepper-tagged mCherry mRNA transcript (FIG. 13D). These results suggest that Pepper tag does not significantly affect protein translation of these mRNA reporter transcripts. Data were collected from 2 independent cell cultures. Values are means \pm s.d. Since the degradation mechanism of the fluorescent RNA-regulated fusion proteins described herein relies on ubiquitination and subsequent proteasomal degradation, expression of fluorescent RNA-regulated fusion proteins could lead to the overload of proteasome activity in cells. To test whether the expression of fluorescent RNA-regulated fusion proteins overloads proteasome activity, a RNA-regulated fluorescent fusion protein, (mNeonGreen)₄-tDeg was expressed in HEK293T cells. If the expression of (mNeonGreen)₄-tDeg overloads the activity of the proteasome, an accumulation of the ubiquitinated protein in cells would be expected. FIG. 13E shows western blotting results using an anti-ubiquitin antibody of untransfected cells and cells expressing (mNeonGreen)₄-tDeg. Significant difference in the ubiquitinated proteins were not observed. As a control, untransfected cells treated with a proteasome inhibitor (10 μ M MG132) for 5 hours showed a significant increase of the ubiquitinated proteins (FIG. 13E). Thus, these results suggest that expression of fluorescent RNA-regulated fusion proteins does not overload proteasome activity in cells. Data shown here is a representative image from 2 independent cell cultures.

FIGS. 14A-14D demonstrate that Pepper tag does not disrupt the localization of mRNAs. To determine whether the Pepper tag disrupts an mRNA's proper cellular localization, an ER-targeting reporter mRNA was chosen, and its localization in cells was imaged using the (F30-2 \times Pepper)₁₀ Pepper tag and the (mNeonGreen)₄-tDeg fluorescent fusion protein (FIG. 14A). This ER-targeting reporter mRNA encodes the first 29 amino acids of cytochrome p450, CytTERM, and the encoding sequence of mCherry followed by (F30-2 \times Pepper)₁₀ in the 3'UTR (FIG. 14A). During protein translation, the CytTERM peptide will direct this reporter mRNA to the outer ER membrane, and confine the mRNA's mobility. Indeed, green fluorescent puncta with low mobility were observed (FIGS. 14B, 14D), suggesting that the reporter mRNA is localized to the outer ER membrane. To further validate the localization of the ER-targeting reporter mRNA, the cells were treated with a translation inhibitor (100 μ g/mL, puromycin) to liberate the reporter

mRNA from the ER into the cytosol. A significant mobility increase of the green fluorescent puncta was observed (FIG. 14C, FIG. 14D), reflecting the dissociation of the reporter mRNA from the ER. Together, these results confirmed that the Pepper tag does not disrupt the localization of mRNAs. Scale bar in (FIG. 14B, FIG. 14C), 10 μ m. Relative diffusion coefficient of mRNA puncta is plotted (n=2 independent cell cultures). Values are means \pm s.d. ***P=2.7 \times 10⁻⁶ by unpaired two-tailed Student's t-test.

FIGS. 15A-15C demonstrate the imaging of green Pepper-tagged β -actin mRNA in live cells. FIG. 15A shows DNA plasmid constructs used for imaging β -actin mRNA in live cells. To image β -actin mRNA localization in response to arsenite stress, a β -actin mRNA reporter containing a 3'UTR green Pepper mRNA tag, (F30-2 \times Pepper)₁₀ was constructed (FIG. 15B). Cells coexpressing this β -actin mRNA reporter and a green fluorescent RNA-regulated fusion protein, (mNeonGreen)₄-tDeg were imaged before and 45 minutes after arsenite (500 μ M) treatment to induce stress granules. Individual mRNA transcripts were observed to rapidly accumulated to form stress granules as evidenced by coexpression of tetramethylrhodamine-labeled HaloTag-G3BP1 to label stress granules. Scale bar, 20 μ m. FIG. 15C shows the fluorescence ratio of foci/cytosol in untreated cells vs. arsenite treated cells is plotted (n=3 independent cell cultures). Values are means \pm s.d. ***P=2.5 \times 10⁻³¹ by unpaired two-tailed Student's t-test.

FIGS. 16A-16B demonstrate that (mNeonGreen)₄-tDeg without the Pepper-tagged β -actin mRNA does not accumulate in stress granules upon arsenite treatment. In FIGS. 15A-15C, cytosolic green fluorescent puncta were shown to accumulate in stress granules to form foci upon application of 500 μ M arsenite. However, the formation of green fluorescent foci in stress granules could be due to aggregation of the fluorescent RNA-regulated fusion protein, (mNeonGreen)₄-tDeg, regardless of the present of the β -actin mRNA. To test whether this is the case, (mNeonGreen)₄-tDeg was coexpressed with circular Pepper RNA in U2OS cells (FIG. 16A). Before arsenite treatment, cytosolic green fluorescent was observed without any puncta, which is consistent with the results in FIGS. 5A-5G. Upon application of 500 μ M arsenite, green fluorescent foci formation was not observed (FIG. 15B). These results confirmed that the formation of green fluorescent foci in FIGS. 15A-15C were indeed due to the β -actin mRNA. This experiment was performed twice with similar results. Scale bar, 20 μ m.

FIGS. 17A-17B demonstrate imaging of mRNAs using Pepper RNA-regulated fluorescent fusion proteins with different hues. So far, mRNA imaging using the green Pepper RNA tag, comprising the Pepper aptamer and a Pepper-regulated fluorescent mNeonGreen fusion protein has been described herein. To further expand the color palette for mRNA imaging, (mVenus)₂-tDeg and (mCherry)₂-tDeg were expressed to generate yellow Pepper and red Pepper complexes on mRNA. In these experiments, (mVenus)₂-tDeg was used to image an mCherry mRNA reporter tagged with (F30-2 \times Pepper)₁₀ (FIG. 17A), and (mCherry)₂-tDeg was used to image a β -actin mRNA reporter tagged with (F30-2 \times Pepper)₁₀ (FIG. 17B), respectively. In both cases, mobile fluorescent puncta were observed in cells. This experiment was performed twice with similar results. Scale bar, 20 μ m.

FIGS. 18A-18D demonstrate the use of the tDeg-Pepper system to selectively biotinylate RNA-binding protein. tDeg was first shown to confer Pepper RNA-dependent regulation of a biotin ligase, TurboID, and a peroxidase, APEX2. HEK293T cells transiently expressed EGFP-TurboID-tDeg

11

(FIG. 18A), and EGFP-APEX2-tDeg (FIG. 18B), with and without the Pepper RNA aptamer, respectively. In each case, proteins were nearly undetectable unless coexpressed with the Pepper RNA. FIG. 18C is a schematic showing that a selectively activated biotin ligase (TurboID-tDeg) specifically biotinylates an RNA-binding protein (CELF1) that bind to the RNA sequence of interest (EDEN15). FIG. 18 D shows that TurboID-tDeg enables selective biotinylation of CELF1, while minimizing nonspecific biotinylation of proteins that do not bind to the RNA of interest (EDEN15).

FIG. 19 demonstrates that Tat-GG confers Pepper RNA-dependent Regulation. In these experiments, U2OS cells transiently expressed mNeonGreen-Tat-GG fusion protein with and without the circular Pepper RNA aptamer, respectively. mNeonGreen was nearly undetectable (left panels) unless coexpressed with circular Pepper RNA (right panels). All cells were stained with Hoechst dye. Scale bar, 20 μ m.

FIG. 20 demonstrate that HIV Tat-RRRG (SEQ ID NO: 127) confers HIV TAR RNA-dependent regulation. In these experiments, cells transiently expressed YFP-HIV Tat-RRRG fusion protein with and without the circular HIV TAR RNA aptamer, respectively. YFP was nearly undetectable (top left panel) unless coexpressed with circular HIV TAR RNA aptamer (right panel). Bottom panels show brightfield microscopy of cells transfected with EYFP-HIV Tat-RRRG in the absence (left panel) or presence (right panel) of circular HIV TAR RNA (SEQ ID NO: 128).

DETAILED DESCRIPTION

A first aspect of the disclosure relates to a nucleic acid molecule encoding an RNA-regulated fusion protein. The nucleic acid molecule includes: a first nucleic acid sequence encoding a protein of interest and a second nucleic acid sequence encoding an RNA-regulated destabilization domain, where the second nucleic acid sequence is operably coupled to the first nucleic acid sequence.

12

The terms protein and polypeptide are generally used interchangeably and refer to a single polypeptide chain. It will be appreciated that such polypeptide chains may bind to other polypeptides or proteins, or other molecules such as cofactors. The terms protein and polypeptide also refer to variants, mutants, biologically active fragments, modifications, analogs and/or derivatives of the polypeptides described herein. The term fusion protein refers to a protein that is comprised of two or more amino acid sequences, from two or more proteins or polypeptide sequences that are not found linked in nature and that are physically linked by a peptide bond.

A protein of interest refers to a protein/polypeptide that is desired and/or being assessed. In other words, a protein of interest may be any protein. In some embodiments, the protein of interest is a protein that is the subject of research. In some embodiments, the protein of interest is known to be involved in a disease state, and is specifically targeted in treatment of the disease state.

In some embodiments, the protein of interest is a fluorescent protein, a bioluminescent protein, an enzyme, or a transcriptional regulator.

In some embodiments, the protein of interest is a fluorescent protein. As used herein, the term “fluorescent protein” refers to a protein or polypeptide which fluoresces, or emits light, when excited with appropriate electromagnetic radiation.

Suitable fluorescent proteins include, without limitation, Green Fluorescent Protein, Enhanced Green Fluorescent Protein (EGFP), Enhanced Yellow Fluorescent Protein (EYFP), Venus, mVenus, Citrine, mCitrine, Cerulean, mCerulean, Orange Fluorescent Protein (OFP), mNeonGreen, mNeonGreen, mCherry, mTagBFP, Venus, mVenus, mTurquoise, mScarlet, mWasabi, mOrange, and dTomato. Suitable fluorescent protein amino acid sequences are shown in Table 1 below.

TABLE 1

Exemplary Fluorescent Protein Amino Acid Sequences			
Fluorescent Protein	Amino Acid Sequence	SEQ ID NO:	
Green Fluorescent Protein (GFP)	MSKGEEELFTGVVPILVELGDVNGHKFSVSGEGEGLDATYGKLTLKF CTIGKLPVPWPPTLVITFSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQ ERTIFFKDDGNYKTRAEVKFEGLDTLVNRIELKGIDFKEDGNILGHKL EYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPI GDGPVLLPDNHYLSTSQALSKDNEKRDHMVLEFVTAAGITHGMDE LYK	1	
Enhanced Green Fluorescent Protein (EGFP)	MVSKGEEELFTGVVPILVELGDVNGHKFSVSGEGEGLDATYGKLTLKF ICTIGKLPVPWPPTLVTTILTYGVQCFSRYPDHMKQHDFFKSAMPEGYV QERTIFFKDDGNYKTRAEVKFEGLDTLVNRIELKGIDFKEDGNILGHK LEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTP IGDGPVLLPDNHYLSTSQALSKDNEKRDHMVLEFVTAAGITLGMD ELYK	2	
Enhanced Yellow Fluorescent Protein (EYFP)	MVSKGEEELFTGVVPILVELGDVNGHKFSVSGEGEGLDATYGKLTLKF ICTIGKLPVPWPPTLVTTFGYGLQCFARYPDHMKQHDFFKSAMPEGYV QERTIFFKDDGNYKTRAEVKFEGLDTLVNRIELKGIDFKEDGNILGHK LEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTP IGDGPVLLPDNHYLSTSQALSKDNEKRDHMVLEFVTAAGITLGMD ELYK	3	
Venus	MVSKGEEELFTGVVPILVELGDVNGHKFSVSGEGEGLDATYGKLTLKF ICTIGKLPVPWPPTLVTTGLYGLQCFARYPDHMKQHDFFKSAMPEGYV QERTIFFKDDGNYKTRAEVKFEGLDTLVNRIELKGIDFKEDGNILGHK LEYNYNSHNVYITADKQKNGIKANFKIRHNIEDGGVQLADHYQQNTP IGDGPVLLPDNHYLSTSQALSKDNEKRDHMVLEFVTAAGITLGMD ELYK	4	

TABLE 1-continued

Exemplary Fluorescent Protein Amino Acid Sequences		
Fluorescent Protein	Amino Acid Sequence	SEQ ID NO:
mVenus	MVKGEELFTGVVPILVELGDGVNGHKFSVSGEGEGDATYGKLTTLK ICTIGKLPVPWPWTLVTTLGYGLQCFARYPDHMKQHDFFKSAMPEGYV QERTIFFKDDGNYKTRAEVKPEGDTLVNRIELKGIDFKEDGNILGHK LEYNNYNSHNVYITADKQKNGIKANFKIRHNIEDGGVQLADHYQQNTP IGDGPVLLPDNHYLSSYQSKLSKDPNEKRDHMVLLFVTAAGITLGMD ELYK	5
Citrine	MVKGEELFTGVVPILVELGDGVNGHKFSVSGEGEGDATYGKLTLF ICTIGKLPVPWPWTLVTTFGYGLMCFARYPDHMKQHDFFKSAMPEGYV QERTIFFKDDGNYKTRAEVKPEGDTLVNRIELKGIDFKEDGNILGHK LEYNNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTP IGDGPVLLPDNHYLSSYQSKLSKDPNEKRDHMVLLFVTAAGITLGMD ELYK	6
mCitrine	MVKGEELFTGVVPILVELGDGVNGHKFSVSGEGEGDATYGKLTLF ICTIGKLPVPWPWTLVTTFGYGLMCFARYPDHMKQHDFFKSAMPEGYV QERTIFFKDDGNYKTRAEVKPEGDTLVNRIELKGIDFKEDGNILGHK LEYNNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTP IGDGPVLLPDNHYLSSYQSKLSKDPNEKRDHMVLLFVTAAGITLGMD ELYK	7
Cerulean	MVKGEELFTGVVPILVELGDGVNGHKFSVSGEGEGDATYGKLTLF ICTIGKLPVPWPWTLVTTLWGQCFARYPDHMKQHDFFKSAMPEGYV QERTIFFKDDGNYKTRAEVKPEGDTLVNRIELKGIDFKEDGNILGHK LEYNAISDNVYITADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTP IGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLFVTAAGITLGMD ELYK	8
mCerulean	MVKGEELFTGVVPILVELGDGVNGHKFSVSGEGEGDATYGKLTLF ICTIGKLPVPWPWTLVTTLWGQCFARYPDHMKQHDFFKSAMPEGYV QERTIFFKDDGNYKTRAEVKPEGDTLVNRIELKGIDFKEDGNILGHK LEYNAISDNVYITADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTP IGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLFVTAAGITLGMD ELYK	9
Orange Fluorescent Protein (OFP)	MNL SKNVS VSY MKGN VN NH FEY D GEGG DP Y TG KYS M KMTL RGQ N CLP FS YDI ITTA F QY G R VF T K Y P E G I V D Y F K D S L P D A F Q W N R R I V F EDGG VL NM SS D I TY K D N V L H G D V W A V G V N F P P N G P V M K N E I V M E E P T E E T F T P K N G V L V G F C P K A Y L L K D G S Y Y G N M T T F Y R S K K S G Q A P P G Y HF V K H R L V K I N V G H G F K T V E Q T E Y A T A H V S D L P K	10
mNeon Green	MVKGEEDNMASLPATHELHIFGSINGVDFDMVGQGTGNPNPDGYEEL NLKSTKGDLQFSPWILVPHIGYGFHQYLPYPDGMSPFQAAAMDGSGY QVHRTMQFEDGASLTVNRYRTYEGSHIKGEAQVKGTGFPADGPVMTN SLTAADWSRSKKTYPNDKTIIISTFKWSTYTGNGKRYRSTARTTYFA KPM A ANY LKNQPMYVFRKTELKHSKTELNFKEWQKAFTDVMGMDELY K	11
moxNeon Green	MVKGEEDNMASLPATHELHIFGSINGVDFDMVGQGTGNPNPDGYEEL NLKSTKGDLQFSPWILVPHIGYGFHQYLPYPDGMSPFQAAAMDGSGY QVHRTMQFEDGASLTVNRYRTYEGSHIKGEAQVKGTGFPADGPVMTN SLTAADWSRSKKTYPNDKTIIISTFKWSTYTGNGKRYRSTARTTYFA KPM A ANY LKNQPMYVFRKTELKHSKTELNFKEWQKAFTDVMGMDELY K	12
mCherry	MVKGEEDNMMAI I KEPMRFKVHMEGSVNGHEFEI EGE G EGR PY EGT Q TAKLK VTKGGPLPFAW DILSPQFMYGSKAVVKHPADIPDYLKLSFPE GPKWERVMNFEDGGVVTVTQDSSLQDGEFLYKVKLRGTNPSDGPM QKKTMGWEASSERMPY PEDGALKGEIKQRLKLKDGGHYDAEVKTTYKA KKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELY K (GenBank Accession No. QEM23462.1, which is hereby incorporated by reference in its entirety)	13
mTagBFP	MVKGEELIKENMHMKLYMEGTVDNHHFKCTSEGE GKP YEGT QTMRI KVEGGPLPFAFDILATSFYLSKTFINHTQGIPDFKQS PEGPTW ERVTTYEDGGVLTATQDTSLQDGCLIIYVNKIRGVNFTSNGPVMQKKT LGWEAFTETLYPADGGL EGRNDMALKLVGGSHLIANAKTTYRSKKPA KNLKMPGVYYDYRLERIKEANNETYVEQHEVAVARYCDLPSKLGHK LN	14
Venus	MVKGEELFTGVVPILVELGDGVNGHKFSVSGEGEGDATYGKLTKL ICTIGKLPVPWPWTLVTTLGYGLQCFARYPDHMKQHDFFKSAMPEGYV QERTIFFKDDGNYKTRAEVKPEGDTLVNRIELKGIDFKEDGNILGHK LEYNNYN SHNVYITADKQKNGIKANFKIRHNIEDGGVQLADHYQQNTP	15

TABLE 1-continued

Exemplary Fluorescent Protein Amino Acid Sequences		
Fluorescent Protein	Amino Acid Sequence	SEQ ID NO:
	IGDGPVLLPDNHYLSYQSALS KDPNEKRDHMVLLEFVTAAGITLGMD ELYK	
mVenus	MVKGEELFTGVVPILVELDGDVNNGHKFVS GEGEGDATYGKLT LKL ICTIGKLPVPWP TLVTTLGYGLQCFARYPDHMKQHDFFKSAMPEGYV QERTIFFKDDGNYKTRAEVFKFEGDTLVNRI ELKGIDFKEDGNILGHK LEYNYNSHNVVI TADKQKNGI KANFKIRHNIEDGGVQLADHYQQNTP IGDGPVLLPDNHYLSYQSCLS KDPNEKRDHMVLLEFVTAAGITLGMD ELYK	16
mTurquoise	MVKGEELFTGVVPILVELDGDVNNGHKFsys GEGEGDATyGKLT LKF ICTIGKLPVPWP TLVTL SWGVQCFCARYPDHMKQHDFFKSAMPEGYV QERTIFFKDDGNYKTRAEVFKFEGDTLVNRI ELKGIDFKEDGNILGHK LEYNYISDNVYI TADKQKNGI KANFKIRHNIEDGGVQLADHYQQNTP IGDGPVLLPDNHYLSQSKLS KDPNEKRDHMVLLEFVTAAGITLGMD ELYK	17
mScarlet	MVKGEAVIKEFMRFKVHMEGSMNGHEFEIEGEGEGRPYEGTQTAKL KVKGGLPLFWDILSPQFMYGSRAFTKHPADIPDYKQSFPEGFKW ERVMNFEDGGAVIVTQDTSLEDGTLIYKVKLRGINFPPDGPMQKKT MGWEASTERLYPEDGVLKGDIKMALLRKDGGRYLA DFKITYKAKKPV QMPGAYNVDRKLDITSHNEDYT VVQEYERSEGRHSTGMD ELYK	18
mWasabi	MVKGEETTMGVIKPDMKIKLKMEGNVNGHAFVIEGEGEGRPYEGTQ TINLEVKEGAPLPFSYDILTTAFSYCNRAFTKYPDDIPNYFKQSPPE GYSWERTMTFEDKGIVVKVKS DISMEEDSFYIEIHLGENFPNGPVM QKETTGWDASTERMYVRDGVLKGDVKMALLLEGGGHHRVDFKTIYRA KKAVKL PDYHFDVDRHIEILNHDKDYNKIVVYETAVARNSTDGMDELYK	19
mOrange	MVKGEENNMAIKEFMRFKVRM EGSMNGHEFEIEGEGEGRPYEGFQ TAKLKVTGGPLPFAW DILSPQFTYGSKAVVKHPADIPDYKQLSPPE GFKWERVMNFEDGGVVIVTQDSSLQDGFIYKVKLRGINFPPSDGPVM QKKTMGWEASSERMPEDGALKGEIKMRLKLKDGGHHTSEVKITYKA KKPVQLPGAYIVGIKL DITSHNEDYTIVEQYERAEGRHSTGMD ELYK	20
dTomato	MVKGEAVIKEFMRFKVRM EGSMNGHEFEIEGEGEGRPYEGTQ KVTKGGPLPFAW DILSPQFMYGSKAVVKHPADIPDYKQLSPPE ERVMNFEDGGLVTVTQDSSLQDGFIYKVKM RGINFPPDGPMQKKT MGWEASTERLYPRDGVLKGEIHQALKLKDGGH YLVEFKTIYMAKKPV QLPGYYYVDTKLDITSHNEDYTIVEQYERSEGRHHLFLYGMDELYK	21

In other embodiments, the protein of interest is a bioluminescent protein. As used herein, the term “bioluminescent protein” refers to any protein capable of acting on a suitable substrate and producing luminescence. As used herein, the term “substrate” refers to any molecule capable of producing or absorbing luminescence with a bioluminescent protein.

Suitable bioluminescent proteins include, without limitation, luciferase, β -galactosidase, β -lactamase, peroxidase, alkaline phosphatase, β -glucuronidase, and β -glucosidase. Exemplary bioluminescent amino acid sequences are shown in Table 2 below.

TABLE 2

Exemplary Bioluminescent Protein Amino Acid Sequences		
Bioluminescent Protein	Amino Acid Sequence	SEQ ID NO:
Nanoluc luciferase (Nluc)	MVFTLEDFVGDWQTAGYNLDQVLEQGGVSSLFQNLGVSVTP IQRIVLSENGLKIDIHVIIPYEGLSGDQMCGQIEKIFKVVYP VDDHHFKVILHYGTLVIDGTPNMIDYFGRPYEGIAVFDGKK ITVTGTLWNGNKIIDERLINPDGSLLFRVTINGVTGWRLCER ILA (GenBank Accession No. AFI79290.1, which is hereby incorporated by reference in its entirety)	22
Firefly luciferase	MEDA KNIKKGPAPFYPLEDGTAGEQLH KAMKRYALVPGTIAF TDAHIEVNITYA EYFEMSVRLAEAMKRYGLNTNHRIVVCSEN S1QFFMPVL GALFIGVAVAPANDIYNRELLNSMNISQPTVV FVSKKGLQKILNVQKLP I IQKIIIMDSKTDYQGFQSMYTFV TSHLPPGFNEYDFV PESFDRDKTIALIMNSSG STGLPKGV AL PHRTACVRF SHARDPIFGNQIIPDTAILS VVFPFH GFG MFTT	23

TABLE 2-continued

Exemplary Bioluminescent Protein Amino Acid Sequences		
Bioluminescent Protein	Amino Acid Sequence	SEQ ID NO:
	<p>LYGLICGFRVVLMYRFEELFLRSLQDYKIQSALLVPTLFSF FAKSTLIDKYDLSNLHEIASGGAPLSKEVGEAVAKRFLPGI RQGYLTETTSAILITPEGDDKPGAVGVVPFFEAKVVDLDT GKTILGVNQRGELCVRGPMIMSGYVNNEATNALIDKDGLHLS GDIAYWDEDEHFFIVDRLKSLIKYGQVAPAAELESILLQHP NIFDAGVAGLPDDDAELPAAVVLEHGKTMTEKEIVDYVAS QVTAAKKLRGGVVFVDEVPKGLTGLDARKIREILIKAKKGG KSKL (GenBank Accession No. CAB91857.1, which is hereby incorporated by reference in its entirety)</p>	
Renilla luciferase (Rluc)	<p>MASKVYDPERRQRKRMITGPQWWARCKQMNVLDSSFINYYDSEKH AENAVIFLHGNAASSYLWRHVPHIEPVARCIIPDLIGMGKS GKSGNGSYRLLDHYKYLTAWFELLNLPKKIIIFVGHDWGACLA PHYSYEHQDKIKIAIVHAESVVDVIESWDEWPDIEDIALIKS EEGEKMKVLENNFFVETMLPSKIMRKLEPEEFAAYLEPFKEKG EVRRPTLSWPREIPLVKGGKPDVVQIVRNYYNAYLRASDLLPK MFIESDPGFFSNAIIVEGAKKFPNTEFVKVKGLHFQSEDAPDE MGKYIKSFVERVLNEQ (GenBank Accession No. ABA41680.1, which is hereby incorporated by reference in its entirety)</p>	24
Gaussia luciferase	<p>MGVKVLFALICIAVAEAKPTENNEDFNIVAVASNFAATTDLDA DRGKLPKGKLLPLEVLKEMEANARKAGCTRGCLICLSHIKCTP KMKKP1PGRCHTYFGDKESAQGGIGEAIVDIPETPGFKDLEP MEQFIAQVQLCVDCCTTGCLKGLANVQCS DLLKKWLPQRCATF ASKIQGQVVDKIKGAGGD (GenBank Accession No. BAR71165.1, which is hereby incorporated by reference in its entirety)</p>	25
β -galactosidase	<p>VVLQRRDWENPGVTQLNRLAAHPPFASWRNSEEARTDRPSQQ LRSINGEWRFAWFPFAPEAVPESWLECDLPEADTVVPSNWQM HGYDAPITYTNVTPYITVNPPFPVTENPTGCYSLTFNVDESWL QEGQTRIIIFDGVNSAFHLWCNGRWWGYGODSRLPSEFDLSAF LRAGENRLAVMVLRWSDGSYLEDQDMWRMSGIFRDVSLLHKP TTQISDFHVATRFNNDFSRAVLEAEVQMCGELRDYLRLTVSL WQGETQVASGTAPPGEIIDERGGYADRVTLRLNVENPKLWS AEIIPNLYRAVVELHTADGTLIEAAEACDVGFRREVRIENGLLL NGKPLLIRGVNRHEHHPLHGQVMDEQTMVQDILLMKQNNFNA VRCSHYPNHPWLWYTLCDRGLYVVDANEIETHGMVPMNRLTD DPRWLPMAMSERVTRMVQRDRNHPSVIWIISLGNESGHGANHDA LYRNIKSVDPSRPVQYEGGGADTTATDIIICPMYARVDEDQPF PAVPKWSIKKWLSPGETRPLILCEYAHAMGNLSGGFAKYWQ AFROYPRLQCGFWWDVWDQSLIKVDENGNPWSAYGGDFGDTP NDRQFCMNGLGFADRTPHPALIEAKHQQQFFQFRLSGQTIEV TSEYLFRHSDELLHWMVALDGKPLASGEVPLDVAPQGKQLI ELPELPQPESSAGQLWLTVRVVQPQNATAWSEAGHIISAWQQWRL AENLISVTLPAASHIAIPHITTSEMDFCIELGNKRWQFNROSDF LSQMWIGDKKQLLTPLRDQFTRAPLDNDIGVSEATRIDPNAW VERWKAAGHYQAEAAALLQCTADTLADAVLTTAHAWHQGKT LFISRKTYRIDGSGQMAITVDVEVASDTPHPARIQLNCOLAQ VAERVNWLGLGPQENYPDRLTAACFDRNWLPLSDMYTPYVFP SENGLRCGTRELNYGPHQWRGDFQFNISRYSQQQLMETSHRH LLHABEGTWNIDGFMIGGDDSWSPSVSAEFQLSAGRYHY QLWCQK (GenBank Accession No. CAB90353.1, which is hereby incorporated by reference in its entirety)</p>	26
β -lactamase (HaloTag)	<p>MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARV GYIELDLSNGKILEFSRPEERFPMMSTFKVLLCGAVLSRIDA GQEOLGRRRIHYSQNDLVEYSPVTKEHLDGMTVRELCSSAIT MSDNTAANLLTTIGGPKELTAFLHNMGDHVTRLDRWEPELN EAIPNDERDTTMRPAMATTLRKLLTGEELLTLASRQQLIDWME ADKVAGPILLRSALPAGWFIADKSGAGERGSRGIIAALGPDGK PSRIVVIYTTGSQATMDERNRQIAEIGASLIKHW (GenBank Accession No. AEQ28652.1, which is hereby incorporated by reference in its entirety)</p>	27
Ascorbate peroxidase 1, cytosolic (<i>Glycine max</i>)	<p>MGKSYPTVSADYQKAVEKAKKKLRGFIAEKRCAPLMLRLAWH SAGTFDKGKTGTPGFGTIKHPAELAHSANGLDIAVRLLEPL KAEFPILSYADFYQLAGVVAVEVTGPEVPFHPGREDKPEPP PEGLRPDATKGSDHLDVFGKAMGLTDQDIVALSGGHTIGAA HKERSGFEGPWTNSPLIFDNSYFTELLSGEKEGLLQLPSDKA LLSDPVFRPLVDKYAADEDAFFADYAEAHQKLSELGFADA</p>	28

TABLE 2-continued

Exemplary Bioluminescent Protein Amino Acid Sequences		
Bioluminescent Protein	Amino Acid Sequence	SEQ ID NO:
(GenBank Accession No. NP_001237785.1, which is hereby incorporated by reference in its entirety)		
Ascorbate peroxidase 1 (<i>Arabidopsis thaliana</i>)	MTKNYPPTVSEDYKKAVERKCRRLGLIAEKNCAPIMVRLAWH SAGTFDCQSRGTGGPGTMRFDAEQAHGANSGIHIALRLLDPI REQFPFTISFADFHQLAGVVAVEVTGGPDIPFHPGREDKPQP PEGRLPDAATKGCDHLRDVFAKQMLSDLKDIVALSGAHTLGRC HKDRSGFEGAWTSNPLIFDNSYFKELLSGEKEGLLQLVSDKA LLDDPVFRPLVEKYAADEDAFFADYAEAHMKLSELGFADA (GenBank Accession No. NP_172267.1, which is hereby incorporated by reference in its entirety)	29
Ascorbate peroxidase 2 (<i>Arabidopsis thaliana</i>)	MVKKSYPEVKEEYKKAVQRCKRKLRGLIAEKHCAPIVRLAWH HSAGTFDVTKTGTGGPGTIRHPQELAHANDNGLDIAVRLLDP IKELPILSYADFYQLAGVVAVEITGGPEIPFHGRLDKVEP PPEGRLPQATKGVDHLDVFGMGLNDKDIVALSGGHTLGRC HKERSGFEAWTPNPLIFDNSYFKEILSGEKEGLLQLPTDKA LLDDPLFLPVEKYAADEDAFFEDYTEAHKLKLSLGFADE (GenBank Accession No. AEE74792.1, which is hereby incorporated by reference in its entirety)	30
Ascorbate peroxidase (<i>Pisum sativum</i>)	MGKSYPTVSPDYQKAIEKAKRKLRGFIAEKKCAPLILRLAWH SAGTFDSKTKTGGPGTIKHQAELAHGANGLDIAVRLEPI KEQFPIVSYADFYQLAGVVAVEITGGPEVPFHPGREDKPEPP PEGRLPDAATKGSDHLDVFGKAMGLSDQDIVALSGGHTIGAA HKERSGFEAWTPNPLIFDNSYFTELLTGKDGLLQLPSDKA LLTDHSVFRPLVEKYAADEDVFFADYAEAHKLKLSLGFAEA (GenBank Accession No. AAA33645. 1, which is hereby incorporated by reference in its entirety)	31
APEX2 (soybean ascorbate peroxidase)	MGKSYPTVSADYQDAVEKAKKLRLGFIAEKRCAPLMLRLAFH SAGTFDKGTTGTGGPGTIKHPAELAHSANGLDIAVRLEPL KAEPFILSYADFYQLAGVVAVEITGGPKVPFHPGREDKPEPP PEGRLPDPTKGSDHLDVFGKAMGLSDQDIVALSGGHTIGAA HKERSGFEAWTPNPLIFDNSYFTELLSGEKEGLLQLPSDKA LLSDPVFRPLVDKYAADEDAFFADYAEAHQKLSELGFADA (see, e.g., Ganapathy et al., "Compartment-Specific Labeling of Bacterial Periplasmic Proteins by Peroxidase-Mediated Biotinylation," ACS Infect. Dis. 4(6): 918-925 (2018) and Lam et al., "Directed Evolution of APEX2 for Electron Microscopy and Proximity Labeling," Nature Methods 12:51-54 (2014), which are hereby incorporated by reference in their entirety)	32
Horseradish peroxidase (<i>Armoracia rusticana</i>)	MQLTPTFYDNSCPNVSNIVRDTIVNELRSDPRIIASILRLHF HDCFVNCGDASILLDNTTANSARGFPVIDRMKAIVESACPR TVSCADLLTIAAQSVTLAGGPSWRVPLGRRDSLQAFDLAN ANLPAFFFTLPQLKDSFRNVGLNRSSDLVALSGGHTFGKNQC RFIMDRLYNFNSTGLPDPILNITYLQTLRGLCPNGNLNALV DFDLRPTTIFDNKYYVNLEEQKGILIQSDQELFSSPNATDTIP LVRSFANSTQTFFNAFVEAMDRMGNIPTLTGTQQQIRLNCRV VNSNS (GenBank Accession No. CAA00083.1, which is hereby incorporated by reference in its entirety)	33
Alkaline phosphatase	MKQSTIALALLPLLFTPVTKARTPEMPLQGTAVDGGGSMHA SLEVLENRAAQGDITAPGGARRLTGDQTAALRDSLSDKPAKN ILLIGDGMDSETTAARNYAEAGAGGFFKGIDALPLTGQYTH YALNKKTGTKPDYVTDAAAATANSTGVKTYNGALGVDIHEKD HPTILEMAAGLATGNVSTAELQDATPAALVAHVTSRKCYG PSATSEKCPGNALEKGKGGSITEBQLLNARADVTLLGGGAKTFA ETATAGEWQGKTLRBRQAQARGYQIVLSDAASLNSVTEANQQP LLGLFADGNMPVRWLGPKATYHNIDKPAVTCTPNPQRNDSV PTLAQMTPDKAIELLSKNEKGFFLQVEGASIDKQDHAANPCGQ IGETVLDLDEAVQRALEFAKKEGNTLVIVTADHHAHASQIVAPD TKAPGLTQALNTKDGAVMVMSYGNSEEDSQEHTGSQLRIAAY GPHAAANVVGLTDQTDLYTMTKAALGLK (GenBank Accession No. AAK73766.1, which is hereby incorporated by reference in its entirety)	34
Alkaline phosphatase (<i>Escherichia coli</i>)	MKQSTIALALLPLLFTPVTKARTPEMPVLENRAAQGDITAPG GARRLTGDQTAALRDSLSDKPAKNILLIGDGMDSETTAAR NYAEGAGGFFKGIDALPLTGQYTHYALNKKTGTKPDYVTDAA SATAWSTGVKTYNGALGVDIHEKDHTILEMAAGLATGNV STAELQDATPAALVAHVTSRKCYGAPSATSEKCPGNALEKGKG	35

TABLE 2-continued

Exemplary Bioluminescent Protein Amino Acid Sequences		
Bioluminescent Protein	Amino Acid Sequence	SEQ ID NO:
	GSITEQQLLNARADVTLLGGGAKTFAETATAGEWQGKTLREQAQ ARGYQLVSDAASLMSVTEANQQKPLLGLFADGNPVRWLGPK ATYHGNIDKPAVTCTPNPQRNDSPVTLAQMTDKAIELLSKNE KGFLFLQVEGASIDKQDHAANPCGQIGETVLDDEAVQRALEFA KKEGNTLVIVTADHAAHASQVAPDTKAPGLTQALNTKDGAQM VMSYGNSEEDSQEHTGSQLRIAAYGPHAANVVLTDQTDLFY TMKAALGLK (GenBank Accession No. WP_001364609.1, which is hereby incorporated by reference in its entirety)	
β -glucuronidase (<i>Escherichia coli</i>)	MLRPVETPTREIKKLDGLWAFSLDRNCIDQRWWESALQES RAIAVPGSFNDQFADADIRNYAGNVWYQREVFIPIKGWAGQRI VLRFDATVHYGKVWVNQNQEVMEHQGGYTFPEADVTPYVIAGK SVRITVCVNNELNWQTIPIPPGMVITDENGGKKQSYFHDFNNYA GIHRSVVMLYTPNTWDDITVVTHVAQDCNHSASVDWQVVANG DVSVELRDAQQVATGGTSGTLQVNVNPHLWQPGEGYLYEL CVTAKSQTECDIYPLRVGIRSVAVKGQQPLINHKPFYFTGFG RHEDADLRKGKGFDNVLVMVHDHALMDWIGANSYRTSHYPYAAE MLDWADEHGIVVVIDETAAVGFNLNSLGIGFEAGNPKPELYSEE AVNGETQQAHLQAIKELIARDKNHPSVMMWSIANEPDTRPQV HGNISPLAEATRKLDPTPITCVNMFCDAHTDTISDLFDVL CLNRYYGWWVQSGLDETAEKVLEKELLAWQEKLHQPIIIITEY GVDTLAGLHSMYTDMWSEEEYQCAWLDMYHRVFDRSAVVGEO VWNFADFATSQGILRVGGNKKGIPTRDRPKSAAFLLQKRWT GMNFGEKPQOGGKQ (GenBank Accession No. AAC53703.1, which is hereby incorporated by reference in its entirety)	36
β -glucosidase (<i>Francisella tularensis</i>)	MSTNSNIROQKLGQLIMMDFRYWGDSNNQRIPIFTKINDIVNK IFKDYNLGGFILFRENIQNNEQVISLLRDQLQANTNTPIFFAT DQEGRVRNLQQGTSGCGNMALAATDNPNAUTMAKIIGDEL YSLGININFAPAVDVNSNKNNPILGVRYSNDNPDIVIDYAKN AINGYHDAKIIDCIKHFPGHDTATDSHLGNVNLDTLKELO TTELLPFSKLARDCSMIMTAHISVPALDDTQYQSVSTSENIV VPATLSYKIIITKLLKQQMFKFDGLVVS DAMDMHAIAKHFGTIE ASKLAILAGIDILLMPVRVWEENDLYKLEELFCLEKCYNQN SNFANAVDNVYTNTIDFKAKHKLDES LFKLSQDEQLKYANQ IVNSNKHQOIALDIAKQSTTVVKNSGIIIPCDLNKLKNILIVD SDNQRLADFHSELQKIVLDDNSNVINCENINHHNIKTIIEN ADLILLISANREYNQTYSYITSIKPEQTINIAALTPYDINY IDNIINYVCITYGATSMQTNYTKTSLKINIQTTLLENIFGNKE IKGVLVPSL (GenBank Accession No. AAC53703.1, which is hereby incorporated by reference in its entirety)	37

The protein of interest may be an enzyme. In some embodiments, the enzyme is selected from the group consisting of a ligase and a methyltransferase.

As described herein, the term "ligase" refers to an enzyme that catalyzes the joining of two large molecules by forming a new chemical bond, usually with accompanying hydrolysis of a small pendant chemical group on one of the larger molecules or the enzyme catalyzing the linking together of two compounds. Suitable ligases include, without limitation, DNA ligases, RNA ligases, amino acid-tRNA ligases (e.g., tyrosine-tRNA ligase, tryptophan-tRNA ligase, threonine-tRNA ligase, leucine-tRNA ligase, isoleucine-tRNA ligase, lysine-tRNA ligase, alanine-tRNA ligase, valine-tRNA ligase, methionine-tRNA ligase, serine-tRNA ligase, aspartate-tRNA ligase, D-alanine-tRNA ligase, glycine-tRNA ligase, proline-tRNA ligase, cysteine-tRNA ligase, glutamate-tRNA ligase, glutamine-tRNA ligase, arginine-tRNA ligase, phenylalanine-tRNA ligase, histidine-tRNA ligase, asparagine-tRNA ligase, aspartate-tRNA ligase, glutamate-tRNA ligase), acetate-CoA ligase, succinate-CoA ligase, biotin-CoA ligase (i.e., biotin ligase), carboxylic acid-CoA ligase, acetate-CoA ligase, and aspartate-ammonia ligase (see, e.g., McDonald, Andrew, "The Enzyme List Class 6—Li-

45 gases," *ExplorEnz Database* (2019), which is hereby incorporated by reference in its entirety).

In some embodiments, the ligase is a biotin ligase. As described herein, biotin ligases catalyze the formation of biotin-5'-AMP anhydride, which diffuses out of the active site to biotinylate proximal endogenous proteins on nucleophilic residues such as lysine. In some embodiments, the biotin ligase is selected from TurboID, miniTurbo, and *E. coli* BirA (see, e.g., Branion et al., "Efficient Proximity Labeling in Living Cells and Organisms with TurboID," *Nat. Biotechnol.* 36(9):880-887 (2018), which is hereby incorporated by reference in its entirety).

50 The methyltransferase may be a histone methyltransferase, an N-terminal methyltransferase, a DNA/RNA methyltransferase, a natural product methyltransferase, a non-SAM dependent methyltransferase, or a radical SAM methyltransferase. As described herein, histone methyl transferases catalyze the transfer of one, two, or three methyl groups to lysine and arginine residues of histone proteins. In some embodiments, the histone methyltransferase is a histone-lysine N-methyltransferase selected from the group consisting of enhancer of zeste homolog 1 (EZH1), enhancer of zeste homolog 2 (EZH2), disruptor of telomeric silencing 55 60 65 1-like (DOT1-like), ASH1L, euchromatic histone-lysine

N-methyltransferase 1 (EHMT1), euchromatic histone-lysine N-methyltransferase 2 (EHMT2), histone-lysine N-methyltransferase 2A, histone-lysine N-methyltransferase 2D (KMT2D), lysine N-methyltransferase 2C (KMT2C), myeloid/lymphoid or mixed-lineage leukemia 4 (MLL4), lysine methyltransferase 2E, and nuclear receptor binding SET domain protein 1 (NSD1). In other embodiments, the histone methyltransferase is a histone-arginine

N-methyltransferases selected from the group consisting of protein arginine N-methyltransferase 1, protein arginine N-methyltransferase 3, protein arginine N-methyltransferase 4, protein arginine N-methyltransferase 5, and protein arginine N-methyltransferase 7.

Non-limiting examples of suitable enzymes are identified in Table 3 below.

TABLE 3

Exemplary Enzyme Amino Acid Sequences		
Enzyme	Amino Acid Sequence	SEQ ID NO:
E. coli BirA (Biotin-CoA ligase)	MKDNTVPLKLIAALLANGEFHSGEQLGETLGMSRAAINKHIQTLR DWGVDFVTVPKGKVSLPPEPIQLLNAAQQLGQLDGGSSAVLPVID STNQYLLDRIGELKSGDACAIEYQQAGRGRGRKFWSFGANLY LSMPWRLLEGPAAAIGLSSLVIGIVMAEVLRKLGADKVRVKWPND LYLQDRKLAGILVELTGKTDAAQIVIGAGINMAMRRVEESVNN QGWITLQEAGINLDRNTLAATLIRELRAALELFQEGLAPYLSR WEKLDNFINRPVKKLIIGDKEIFGISRGIDKQGALLLEQDGIIKP WMGGEISLRSAEK (GenBank Accession No. NP_418404.1, which is hereby incorporated by reference in its entirety)	38
miniTurbo biotin ligase	MIPLLNAKQILGQLDGGSSAVLPVVDDSTNQYLLDRIGELKSGDA CIAEYQQAGRGSRGKWFSPFGANLYLSMFWRKLKGPAAGLGP VIGIVMAEALRKLGADKVRVKWPNDLYLQDRKLAGILVELAGIT GDAAQIVIGAGINVAMRRVEESVNNQGWITLQEAGINLDRNTLA AMLIRELRAALELFQEGLAPYLSRWEKLDNFINRPVKKLIIGD EIFGISRGIDKQGALLLEQDGVIKPWMGGEISLRSAEK (see, e.g., Branen et al., "Efficient Proximity Labeling in Living Cells and Organisms with TurboID," Nat. Biotechnol. 36(9):880-887 (2018), which is hereby incorporated by reference in its entirety)	39
Turbo ID biotin ligase	MKDNTVPLKLIAALLANGEFHSGEQLGETLGMSRAAINKHIQTLR DWGVDFVTVPKGKVSLPPEPIQLLNAAQQLGQLDGGSSAVLPVVD STNQYLLDRIGELKSGDACAIEYQQAGRGRGRKFWSFGANLY LSMPWRLKRGPAAIGLGPVIGIVMAEALRKLGADKVRVKWPNDL YLQDRKLAGILVELAGITGDAAQIVIGAGINVAMRRVEESVNNQ GWITLQEAGINLDRNTLAATLIRELRAALELFQEGLAPYLPW EKLDNFNIRPVKKLIIGDKEIFGISRGIDKQGALLLEQDGVIKP MGGEISLRSAEK (see, e.g., Branen et al., "Efficient Proximity Labeling in Living Cells and Organisms with TurboID," Nat. Biotechnol. 36(9):880-887 (2018), which is hereby incorporated by reference in its entirety)	40
Biotin ligase (Mammalian expression vector pCBio)	MDYKDDDKSPRSMKDNTVPLKLIAALLANGEFHSGEQLGETLGM SRAAINKHIQTLRDWGDFVTVPKGKVSLPPEPIQLLNAAQQLGQ LDGGSSAVLPVIDSTNQYLLDRIGELKSGDACAIEYQQAGRGR GRKFWSFGANLYLSMFWRLEQGPAAAGLSSLVIGIVMAEVLRK LGADKVRVKWPNDLYLQDRKLAGILVELTGKTDAAQIVIGAGI NMAMRRVEESVNNQGWITLQEAGINLDRNTLAAMLIRELRAALE LFQEGLAPYLSRWEKLDNFINRPVKKLIIGDKEIFGISRGIDKQ GALLLEQDGIIKPWMGGEISLRSAEK (GenBank Accession No. ABF74577.1, which is hereby incorporated by reference in its entirety)	41
Enhancer of Zeste Homolog 2 (Homo sapiens) methyl-transferase	MGQTGKKSEKGPVCWRKRVKSEYMRRLQLKRFRRADEVKSMFSS NRQKILERTEILNQEWKQRRIQPVHILTSVSSLRGTRCSVSD LDFPTQVPLKTLNAVASVPMYWSPLQQNFMVEDETVLHNIP YMGDEVLDQDGTFIELIKNYDGKVHGDRECQFINDEIFVELVN ALGQYNDDDDDDDDGDPPEEREEQKDLEDHRDDEKSPRKRFFS DKIFEAISMFPDKGTAEELKEKYKELTEQQLPGALPPECTPNI DGPNAKSQVREQSLHSFHTLCRRCFKYDCFLHPFATPNTYKR KNTETALDNKPCGPQCYQHLEGAKEFAAALTAEIRKTTPKRPGG RRRGRLPNSSSRPSTPTINVLESKDTDSREAGTETGGENNNDKE EEEKKDETSSSEANSRCQTPIMKMPNIEPHENVEWSGAEASMF RVLIGTYYDNFCIAARLIGTKTCRQVYEFRVKESSIIAPAPABD VDTPPRKKKRKHRLWAACRKRQIQLKKGSSNHVNYQPCDHPRQ PCDSSCPVIAQNFCEKFCQCSSECQNRFPGCRCKAQCNTKQCP CYLAVRECDPDLCIICGAADHWDKNSVCKNCISIQRGSKHHLL APSIVAGWGFIFKDPVQKNEFISEYCGEIISQDEADRRGVYD YMCFSFLFNLNNDFVVDATRKGKNGKIRFANHSVNPNCYAKVMMVNG DHRIGIFAKRAIQTGEELFFDYRYSQADALKYVGIEREMEIP (GenBank Accession No. AAC51520.1, which is hereby incorporated by reference in its entirety)	42

Additional suitable proteins of interest include, but are not limited to, a G-protein coupled receptor (GPCR), a nuclear receptor, a voltage gated ion channel, a ligand gated channel, a receptor tyrosine kinase, a growth factor, a phosphatase, a protein kinase, a viral regulator, a bacterial cell division protein, a scaffold protein, a DNA repair protein, a cytoskeletal protein, a ribosome, a histone deacetylase, an apoptosis regulator, a chaperone protein, a kinase, a phosphorylase, a phosphatase, deacetylase, a cytoskeletal protein (e.g., myosin, actin, dynein, kinesin, and tubulin).

As described herein, a G-protein coupled receptor (GPCR) refers to a membrane protein which binds to a signaling molecule. Upon binding, a conformational change occurs, which allows binding of the GPCR to, and activation of, a G-protein. The activated G-protein then interacts with an effector molecule, which is typically involved in a second messenger pathway. Suitable G-protein coupled receptors may be selected from the group consisting of a luteinizing hormone receptor, a follicle stimulating hormone receptor, a thyroid stimulating hormone receptor, a calcitonin receptor, a glucagon receptor, a glucagon-like peptide 1 receptor (GLP-1), a metabotropic glutamate receptor, a parathyroid hormone receptor, a vasoactive intestinal peptide receptor, a secretin receptor, a growth hormone releasing factor (GRF)

receptor, protease-activated receptors (PARs), cholecystokinin receptors, somatostatin receptors, melanocortin receptors, nucleotide receptors (e.g., ADP receptors), adenosine receptors, thromboxane receptors, platelet activating factor receptor, adrenergic receptors, 5-hydroxytryptamine (5-HT) receptors, a chemokine receptor (e.g., CXCR4, CCR5), chemokine receptors, neuropeptide receptors, opioid receptors, erythropoietin receptor, von Willebrand receptor, parathyroid hormone (PTH) receptor, vasoactive intestinal peptide (VIP) receptor, and collagen receptors. Exemplary protease-activated receptors include, without limitation, PAR1, PAR2, PAR3, or PAR4 receptors.

In some embodiments, the protein of interest is a transcription factor. Transcription factors include proteins that are involved in gene regulation in prokaryotic and/or eukaryotic organisms. In one embodiment, transcription factors have a positive effect on gene expression and, thus, may be referred to as an activator or a transcriptional activation factor. In another embodiment, a transcription factor negatively regulates gene expression and, thus, may be referred to as a repressor or a transcription repression factor. Suitable transcription factors include, without limitation, c-Myc, c-Fos, c-Jun, CREB, GATA-2, GAL4, GAL4Np16, c-Myb, MyoD, and NFkB, and tetR. Exemplary transcription factors are identified in Table 4 below.

TABLE 4

Exemplary Transcription Factor Amino Acid Sequences		
Transcription Factor	Amino Acid Sequence	SEQ ID NO:
c-Myc (<i>Homo sapiens</i>)	MPLNVNSFTNRNYDLDYDSVQPYFYCDEEENFYQQQQQSELQPPAP SEDIWKKFELLPTPPLSPSRSGLCSPSVAVTPFSLRGDNNDGG GSFSTADQLEMVTELLGGDMVNQSFCIDPDETFIGKIIIQDCMW SGFSAAKLVLSEKLASYQAARKDGSSPNPARGHHSVCTSSLYLQD LSAAASECIDPSVVFPPYPLNDSSSPKSCASQDSSAFSPSSDSLSS STESSPQGSPEPLVHLHEETPPPTTSDSEEQEDEEEIDVVSVKEKR QAPGKRSESGSPSAGGHSKPHSPVPLVKRCHVSTHQHNYAAPPST RKDYPAAKRVKLDSVRVLRQISNNRKCTS PRSSDTEENVKRRTHN VLERQRNRNELKRSFFALRDQIPELENNEKAPKVILKKATAYILS VQAEEQKLISEEDLLRKRRQLKHKEQLRNSCA (GenBank Accession No. AAA36340.1, which is hereby incorporated by reference in its entirety)	43
c-Fos (<i>Homo sapiens</i>)	MMFSGFNADYEASSSRCSSASPAGDSLSSYYHSPADSFSMMGSPVN AQDFCTDLAVSSANFIPTVTAISTSPDLQWLQPALVSSVAPSQT RAPHPFGVPAPSAGAYSAGVSKMTGGRQASIGRRGKVEQLSPE EEEKRRIRRNKMAAKCRNRRRELTDTLQAETDQLEDEKSALQ TEIANLLKEKEKLEFILAHHRPACKIPPDLGFPBEEMSVASLDLTG GLPEVATPESEEAFTPLLLNDEPKPKPSVEPVKSISSSMELKTEPFD DFLFPASSRPGSETARSVPDMDSLGSFYAADWEPLHSGSLGMGP MATELEPLCIPVVICTPSCTAYTSSFVFTYPEADSFPSCAAHRK GSSSNEPSSDSLSSPTLLAL (GenBank Accession No. AAA52471.1, which is hereby incorporated by reference in its entirety)	44
c-Jun (<i>Homo sapiens</i>)	MTAKMETTFYDDALNASFLPSESAGPYGYSNPKILKQSMTLNLADP VGSLKPHLRAKNSDLTSPLDVGILLKLASPELERLIIQSSNGHITT TPTPTQFLCPKNUVTDEQEGFAEGFVRALAEHLHSQNTLPSVTSAAQ PVNGAGMVAVASVAGGGGGFSASLHSPEPPVYANLSNFNPGA LSSGGGAPSYGAAGLAFFPAQPOQQQQQQPPPHLPQQMPVQHPRQLF KEEPQTVPMPGETPPLSPIDMESQERIKAERKMRNRIAASKCR KRKLERIARLEEKVKTLLKAQNSELASTANMLREQVAQLKQKVMNH VNSGCQLMLTQQLOTF (GenBank Accession No. NP_002219.1, which is hereby incorporated by reference in its entirety)	45
CREB (<i>Homo sapiens</i>)	MIMESGAENQQSGDAAVTEAENQQMTVQAQPQIATLAQVSMPAAH ATSSAPTVLVLQLEPNQGTVQVHGVIQAAQPSVIQSPQVQTVQIST IAEEDSQESVDSVTDSQKRREILSRRPSYRKILNLDLSSDAPGVP RIEEKSEETSAPAITIVTVPTPIYQTSSGQYIAITQGGAIQLA NNGTGVQGLQTLTMNAATQPGTTILOYAQTTDGQQQILVPSNQ VVVQAASGDVQTYQIRTAFTSTIAPGVVMASSPALPTQPAAEAR KREVRMLKNREAARECRKKKEYVKCLENRAVLENQNKTLLIEEL	46

TABLE 4-continued

Exemplary Transcription Factor Amino Acid Sequences		
Transcription Factor	Amino Acid Sequence	SEQ ID NO:
	KALKDLYCHKSD (GenBank Accession No. AAA35715. 1, which is hereby incorporated by reference in its entirety)	
GATA-2 (<i>Homo sapiens</i>)	MEVAPEQPGWMAHPAVLNAQHPSHHPGLAHNHYMEPAHVLPPIPDEV DVFFNHLDQSQQNPYYANPAQRGVSYSPAHHARLTGGMCRPHLLHS PGLPWLDDGKAALSAHHHKWTWVSPFSKTPHLPSAAGGPGGHSLC TQGLGVGGSSGSSVASLPTAAHSGSHLFGFPPRHPKELSPDPS TTGAASPASSAGGSARGEDKGVYQASLTESMKMESGRPLRP GLATMGTQPATHPIPYPSYPAAAHDYSSGLFHPGSFLGGPAS SFTPQRSKTRSCSEGRCVNCGATATPLWRDGTHYLNCACGF YHMKGQNRPPLIKPKRRLSAARRAGTCCANCQTTLWRRNANG DPVCNAACGLYKKLHNVRPLTMKKEGIQTRNRKMSNKSKKSKGA ECFEELSCKCMQEKSPPFSAALAGHMAPMGHLLPPFSHSGHILPTP TPIHPSSSLSGFHPHSSMTAMG (GenBank Accession No. AAA35869. 1, which is hereby incorporated by reference in its entirety)	47
GAL4 (<i>Saccharomyces revisiae</i>)	MKLSSIEQACDICRLLKLCSEKPKCAKCLKNNECRYSPKT RSPLTRAHLTEVESRLERLEQLFLIFPREDLDMILKMDSLQDIK ALLTGLFVQDNVNKAUTDRLASVETDMLPLTRQHRISATSSSEE SSNKGQRQLTVSIDSAAHHDNSTIPLDFMPRDLAHGFDWSEEDDM SDGLPFLKTDPNNGFFGDGSLCILRSIGFKPENYTNSNVNRLP TMI TDRYTLASRSTTSRLLQSYLNNFHPYCPIVHSPTLMMLYNNQ IEIASKDQWQILFNCILAIGAWCIEGESTIDIVFYYQNAKSHLTS KVFESEGSIIILVTALHLLSRYTQWRQTKNTSYNFHFSIIMAIISLG LNRLDLPSSFSDSILEQRRRIWWSVYSWEIQLSLLYGRSIQLSQN TISFPSSVDDVQRTTGTPTIYHGIIETARLLQVFTKIYELDKTVT AEKSPICAKKCLMICNEIEEVSRQAPKFLQMDISTTALTNLKEH PWLSFTRFELWKQQLSLIIVLRFDTNFQKQSLEQDQNDHQ YEVKRCSIMLSDAARQTVMSVSYMDNHNVTPTYFAWNCSYLFNA VLVPITKLLSNKSNAENNNTAQQLQQINTVLMLLKKLATFKIQT CEKYIQVLEEVCAFPFLLSQCAIPLPHISYNNSNGSAIKNIVGSAT IAQYPTLPEENVNNISVKYVSPGSVPVPLKGASFSDLVKKL SNRPPSRNSPVTPRSTPSHRSTPFLGQQQQLQSLVPLTPSALF GGANFQSGNIADSSLSTFTTNSSNGPNLITQTNSQALSQPIAS SNVHDNFMNNEITASKIDDGNNSKPLSPGWTDQTAAYNAFGITTGM FNTTTMDDVNYNLFDDEDTTPNPKKE (GenBank Accession No. AAA34626. 1, which is hereby incorporated by reference in its entirety)	48
GAL4Np16 (<i>Saccharomyces revisiae</i>)	MKLSSIEQACDICRLLKLCSEKPKCAKCLKNNECRYSPKT RSPLTRAHLTEVESRLERLEQLFLIFPREDLDMILKMDSLQDIK ALLTGLFVQDNVNKAUTDRLASVETDMLPLTRQHRISATSSSEE SSNKGQRQLTVSIEFSRGRTRNNGSTIEGLLDPDDDAPEAG LVAPRMSFLSAGQRPRRLSTAPITDVSVDLRLGEEVDMFTA DALDDFDLEMGLDVESPSPGMTHDPVSYGALDVDDFEQMFDTA LGIDDFCG (GenBank Accession No. AAN86074.1, which is hereby incorporated by reference in its entirety)	49
c-Myb (<i>Homo sapiens</i>)	MARRPRHSIYSSDEDDEDDEMCDHDYDGLLPKGKRHLGKTRWTR EE (GenBank Accession No. AAA72118. 1, which is hereby incorporated by reference in its entirety)	50
MyoD (<i>Mus musculus</i>)	MELLSPPLRIDLTGPDGSLCSFETADDYDDPCFDSPDLRFED LDPRLVHVGALLKPEEEAHFSTAVHPGPAREDEHVRAPS GHQ GRCLLWACKACKRKTNNADRRAATMRERRRLSKVNEAFETLKRC ISSNPNQRLPKVEILRNARYIEGLQALLRDQDAAPPGAAAFYAP GFLPPGRGSEHSGDASSPRSNCSDGMMDSGPPSGPQRQNGY DTAYYSEAVRESRPGKSAAVSSLDCLSSIVERISIDSPAAPALL ADAPPESPPGPPEGASLSDTEQGTQTPSPDAAPQCPAGSNPNAIY QVL (GenBank Accession No. AAA39798.1, which is hereby incorporated by reference in its entirety)	51
NF-KB (<i>Homo sapiens</i>)	MDELFPFLIPPAEQPKQRGMFRYKCEGRSAGSIPGERSTDTTKTH PTIKINGYTGP GTVIRISLVTKDPHPRPHPHELVKGKDCRDGFYEAE LCPDRCIHSFQNLGIQCVKKRDLQEQAISQRIQTNNNPFQVPIEEQ RGDYDLNAVRCLFCQVTVRDPSGRPLRLPPVLSHPIFDNRAPNTAE LKI CRVNRNSGSCLGGDEI FLLCDKVQKEDI E VYPTGPGWEARGS FSQADVHRQVAIVVRTTPYADPSLQAPVRVSMQLRRPSDRELSEP MEFQYLPDTDHRRIEKRKRTYETFKSIMKSPFSGPTDPRPPP RRIAVPSSRSSASVPKAPQPYPTSSLSTINYDEFPTMVPPSGQI	52

TABLE 4 -continued

Exemplary Transcription Factor Amino Acid Sequences		
Transcription Factor	Amino Acid Sequence	SEQ ID NO:
	SQASALAPAPPQVLPQAPAPAPAPAMVSALAQQAPAPVVLAPGPP QAVAPPAPKPTQAGEGTLSSEALLQLQFDDEDLGGALLGNSTDPAVF TDLASVDNSEPQQLLNQGIPVAPHTTEPMILMEYPEAITRLVTAQR PPDPAPAPLGAPGLPNGLSGDEDFSSIADMDFSALLSQISS (GenBank Accession No. 2006293A, which is hereby incorporated by reference in its entirety)	
TetR (Proteobacteria)	MFISDKVSSMTKLQPNTVIRAALDLLNEVGVDGLTRRKLAERLGV QQPALYWHRNKRALLDALAEAMLAENHHTHSVPRAADDWRSFLIG NARSFRQALLAYRDGARIHAGTRPGAPQMESTADAQLRFLCEAGFS AGDAVNALMTISYFTVGAVLLEEQAGDSDAGERGGTVEQAPLSPLL RAAIADFDEAGPDAAFEQGLAVIVDGLAKRRLVVVRNVEGPRKGDD (GenBank Accession No. WP_000470728.1, which is hereby incorporated by reference in its entirety)	53

Additional exemplary transcription factors are identified ²⁰ in Table 5 below.

TABLE 5

Additional Exemplary Transcription Factors	
Transcription Factor Family	Transcriptions Factors
Basic Helix-Loop-Helix (bHLH) Family	AHR, ARNT/HIF-1 beta , ASCL1/Mash1, ASCL2/Mash2, CLOCK, DEC2, HAND1, HAND2, HES-1, HES-4, HIF-1 alpha/HIF1A, HIF-2, alpha/EPAS1, c-Maf, Max, MESP1, MITF, MLX, Mxi1, c-Myc, MYCL1/L-Myc, MYF-5, MyoD, Myogenin, NeuroD1, NeuroD2, Neurogenin-1, Neurogenin-2, Neurogenin-3, Olig1, Olig2, Olig3, SCL/Tal1, SREBP2, TCF-12/HTF4, TFEB, Twist-1
Basic Leucine Zipper (bZIP) Family	ATF1, ATF2, ATF4, BACH1, BATF, BATF3, c-Fos, CEBP alpha, CEBP epsilon, CREB, FosB/G0S3, FRA-1, GADD153, HSF1, HSF2, HSF4, c-Jun, JunB, JunD, c-Maf, MafB, MafF, MafG, MafK, Max, MITF, MLX, Mxi1, MYB, c-Myc, MYCL1/L-Myc, NFL3/E4BP4, Nrf1, Nrf2, NRL, OASIS/CREB3L1, SREBP2, TSC22, XBP1
ETS (E-twenty six) Family	ELF3, Ets-1, ETV1, ETV2/ER71, ETV5, ETV6, FLI1, PU.1/Spi-1, Spi-B
Forkhead Domain Family	FoxC1, FoxC2, FoxD3, FoxF1, FoxF2, FoxH1, FoxJ1, FoxJ3, FoxK1, FOXL2, FoxM1, FoxN1, FoxO1/FKHR, FoxO3, FoxP1, FoxP2, FoxP3, FoxP4, HNF-3 alpha/FoxA1, HNF-3 beta/FoxA2
GATA Family Hypoxia Inducible Factors (HIFs) Family	GATA-1, GATA-2, GATA-3, GATA-4, GATA-5, GATA-6, TRPS1 HIF-1, HIF-2, HIF-3, ARNT/HIF-1 beta
High Mobility Group (HMG) Family	HMGAlB, HMGAl2, HMGB1/HMG-1, HMGB3, HMGNI, LEF1, SOX1, SOX2, SOX3, SOX5, SOX6, SOX7, SOX9, SOX10, SOX11, SOX15, SOX17, SOX18, SOX21, TCF7/TCF1, TCF7L1/TCF3
Homeodomain (Hox) Family	ADNP, ARX, ATBF1/ZFHX3, CDX2, CDX4, CRX, DLX5, DUX4, DUX4/DUX4c, DUX4c, EMX2, GBX2, Gooseco, HHEX, HNF-6/ONECUT1, HOXA1, HOXB1, HOXB7, HOXB13, HOXD10, Islet-1, Islet-2, LHX5, LIM1, MSX1, MSX2, Nanog, NKX2.2, NKX2.5, NKX3.1, NKX6.1, Oct-1, Oct-3/4, Oct-4A, Oct-4B, ONECUT2/OC-2, Oct2, PDX-1/IPF1, PHOX2B, PITX2, POU3F2, Prox1, SATB1, TCF-2/HNF-1 beta, TCF-3/E2A, TGIF1, TTF-1/NKX2-1, VSTM2L, ZEB1
Immunoglobulin-Like Domain Family	CSL, NFkB, p50 (NFkB1), p52 (NFkB2), p53, p63/TP73L, NFkBp65/RelA, RelB, c-Rel, STAT (STAT1, STAT2, STAT3, STAT4, STAT5a/b, STAT5a, STAT5b, STAT6)
Interferon-Regulatory Factor (IRF) Family	IRF1, IRF2, IRF3, IRF4, IRF5, IRF6, IRF8
Kruppel-like Family	KLF2, KLF4, KLF5, KLF6, KLF10, KLF12, KLF17
Paired Box (Pax) Family	Pax2, Pax3, Pax4, Pax5/BSAP, Pax6, Pax7
Mothers against decapentaplegic homolog (Smad) Family	FOXL2, Smad1, Smad2, Smad2/3, Smad3, Smad4, Smad5, Smad7, Smad8, Smad9
Additional Transcription	AP-2 beta, AP-2 gamma, AP-2 epsilon, Autoimmune Regulator/AIRE, BLIMP1/PRDM1, C1D, DACH2, DC-SCRIPT/ZNF366, DIDO1, E2F-

TABLE 5-continued

Additional Exemplary Transcription Factors	
Transcription Factor Family	Transcription Factors
Factors	1, E2F-2, E2F-4, EGR1, GLI-1, GLI-2, GLI-3, HNF-4 alpha/NR2A1, HNF-4 gamma/NR2A2, LMO2, LMO4, LPP, MEF2C, PREB, RFX6, Teneurin-1, Teneurin-2, Teneurin-4, TFCP2L1, ZSCAN21

RNA-regulated destabilization domains are amino acid sequences that, when functionally coupled to a protein of interest, modulate the stability of the protein of interest in a RNA-dependent manner. In some embodiments, when the RNA-regulated destabilization domain is fused to a protein of interest, the RNA-regulated destabilization domain mediates protein degradation. In accordance with such embodiments, the protein destabilization function of the RNA-regulated destabilization domain is impeded when it binds to a specific RNA molecule (e.g., an aptamer).

In some embodiments, the RNA-regulated destabilization domain comprises a bifunctional peptide comprising an RNA-binding domain and a degron peptide. The RNA-binding domain may be any peptide to which an RNA molecule can bind, where such binding sterically inhibits the interaction of the degron peptide with a proteosomal pathway component (e.g., an E3 ubiquitin ligase). Thus, in some embodiments, the RNA-binding domain is MDARTRR-RERRAEKQAAQWKAAN (lambdaN; SEQ ID NO: 123), which is derived from the lambda bacteriophage antiterminator protein N. In accordance with such embodiments, the RNA-binding domain is specific for BoxB (SEQ ID NO: 124): GGGCCCUGAAGAAGGGCCC (see, e.g., “NMR Structure of the Bacteriophage Lambda N Peptide/boxB RNA Complex: Recognition of a GNRA Fold by an Arginine-Rich Motif,” *Cell* 93(2):289-299 (1998), which is hereby incorporated by reference in its entirety).

In other embodiments, the RNA-binding domain is DTRQARRNRRRRWRERQRAAAAR (HIV-1 Rev; SEQ ID NO: 125), which is derived from HIV-1 Rev peptide. In accordance with such embodiments, the RNA-binding domain is specific for RRE RNA (SEQ ID NO: 126): GGUCUGGGCGCAGCGCAAGCUGCGGACAGGCC (see, e.g., Battiste et al., “Alpha Helix—RNA Major Groove Recognition in an HIV-1 Rev Peptide—RRE RNA Complex,” *Science* 273:1547-1551 (1996), which is hereby incorporated by reference in its entirety).

The RNA-regulated destabilization domain may comprise a bifunctional peptide comprising a lentiviral transactivator of transcription (Tat) peptide and a degron peptide.

In some embodiments, the lentiviral Tat peptide is a bovine immunodeficiency virus Tat peptide. In other embodiments, the lentiviral Tat peptide is a human immunodeficiency virus Tat peptide.

According to some embodiments, the Tat peptide has the sequence of RKKRRQRRR (SEQ ID NO: 129). See, e.g., Yamamoto et al., “A Novel RNA Motif that Binds Efficiently and Specifically to the Tat Protein of HIV and Inhibits the Trans-Activation by Tat of Transcription In Vitro and In Vivo,” *Genes Cells* 5:371-388 (2000), which is hereby incorporated by reference in its entirety.

According to some embodiments, the Tat peptide has the consensus sequence of SEQ ID NO: 54 as follows: XXXXXXXXXXXXXXXXX, where X at position 1 can be S or A; X at position 2 can be G or A; X at position 3 can be P or A; X at position 4 can be R or K; X at position 5 can

be P, A, I, Y, K, or R; X at position 6 can be R, K, V, or Y; X at position 7 can be G, A, or R; X at position 8 can be T or A; X at position 9 can be R or K; X at position 10 can be G or A; X at position 11 can be K or A; X at position 12 can be G or A; X at position 13 can be R or K; X at position 14 can be I or A; X at position 15 can be R, K, Y, or G; and X at position 16 can be R, K, V, T, or Y. See, e.g., Athanassiou et al., “Structural Mimicry of Retroviral Tat Proteins by Constrained β-Hairpin Peptidomimetics: Ligands with High Affinity and Selectivity for Viral TAR RNA Regulatory Elements,” *J. Am. Chem. Soc.* 126:6906-6913 (2004); Chen & Frankel, “A Peptide Interaction in the Major Groove of RNA Resembles Protein Interactions in the Minor Groove of DNA,” *Proc. Natl. Acad. Sci. USA* 92:5077-5081 (1995); and Koren et al., “The Eukaryotic Proteome is Shaped by E3 Ubiquitin Ligases Targeting C-Terminal Degrons,” *Cell* 173:1622-1635 (2018), which are hereby incorporated by reference in their entirety). For example, the Tat peptide may have the amino acid sequence of SEQ ID NO: 55 as follows: SGPRPRGTRGKGIRR.

In some embodiments, the lentiviral Tat peptide comprises an RNA binding site. The RNA binding site may correspond to amino acid residues 4-17 of SEQ ID NO: 54 or amino acid residues 4-17 of SEQ ID NO: 55.

In some embodiments, the RNA binding site is specific for an RNA aptamer. An aptamers is a nucleic acid molecule that binds with high affinity and specificity to a target. Nucleic acid aptamers may be single-stranded, partially single-stranded, partially double-stranded, or double-stranded nucleotide sequences. Aptamers include, without limitation, defined sequence segments and sequences comprising nucleotides (e.g., ribonucleotides, nucleotide analogs, modified nucleotides, and nucleotides comprising backbone modifications, branchpoints, and non-nucleotide residues, groups, or bridges). Nucleic acid aptamers include partially and fully single-stranded and double-stranded nucleotide molecules and sequences; synthetic RNA, DNA, and chimeric nucleotides; hybrids; duplexes; heteroduplexes; and any ribonucleotide, deoxyribonucleotide, or chimeric counterpart thereof and/or corresponding complementary sequence, promoter, or primer-annealing sequence needed to amplify, transcribe, or replicate all or part of the aptamer molecule or sequence.

As described herein, the RNA binding site is specific for an RNA aptamer having the consensus sequence of SEQ ID NO: 56 as follows: NNNNN-

SHSYWSBMNNNNDSBHSNNNNN, where N can be A, C, G, or U; S can be C or G; H can be A, C, or U; Y can be C or U; W can be A or U; B can be C, G, or U; M can be A or C; and D can be A, G, or U. Thus, in some embodiments, the RNA aptamer has the sequence of wild-type TAR RNA (SEQ ID NO: 57) as follows: GGCUCUGUAGCUCAUUAGCUCCGAGCC.

According to some embodiments, the RNA binding site is specific for an RNA aptamer having the consensus sequence of SEQ ID NO: 58 as follows: NNNNN-

SHCYSWSBMNNNNDHSBHSNNNN, where N can be A, C, G, or U; S can be C or G; H can be A, C, or U; Y can be C or U; W can be A or U; B can be C, G, or U; M can be A or C; and D can be A, G, or U. Thus, in some embodiments, the RNA aptamer has the sequence of TAR Variant-1 (SEQ ID NO: 59) as follows: GGCUCGU-CUGAGCUAUAGCUCCGAGCC.

In other embodiments, the RNA binding site is specific for an RNA aptamer having the consensus sequence of SEQ ID NO: 60 as follows: NNNNNNSI-TYSWSBMNNNNDHSBHSNNNN, where N can be A, C, G, or U; S can be C or G; H can be A, C, or U; Y can be C or U; W can be A or U; B can be C, G, or U; M can be A or C; and D can be A, G, or U. Thus, in some embodiments, the RNA aptamer has the sequence of TAR Variant-2 (Pepper; SEQ ID NO: 61) as follows: GGCUCGUUGAG-CUCAUUAGCUCCGAGCC.

In further embodiments, the RNA binding site is specific for an RNA aptamer having the sequence of HIV TAR (SEQ ID NO: 128) as follows: ACGAACG-UUGAUCCCCGUUUGCCGGUCGAUCGCUUCGA.

As used herein, the term “degron” or “degradation signal” or “degron peptide” refers to an amino acid element within a protein that is sufficient for recognition and degradation by a proteolytic system. In some embodiments, the degron is a ubiquitin-pathway degron. In accordance with such embodiments, the degron comprises a region specific for E3 binding (see, e.g., Ravid & Hochstrasser, “Diversity of Degradation Signals in the Ubiquitin-Proteasome System,” *Nat. Rev. Mol. Cell Biol.* 9:679-689 (2008), which is hereby incorporated by reference in its entirety).

The degron peptide may be selected from a monopeptide, a dipeptide, a tripeptide, a tetrapeptide, a pentapeptide, a hexapeptide, a heptapeptide, or an octapeptide. Exemplary degron peptides are well known in the art and are listed in Table 6 below.

TABLE 6

Exemplary Degrone Peptides

Degrone Peptide	Amino Acid Sequences
Monopeptide	P, E
Dipeptide	RG, GG, EE, AP, RP, NP, DP, CP, EP, QP, GP, HP, IP, LP, KP, MP, FP, PP, SP, TP, WP, YP, VP, SA, SR, SN, SD, SC, SE, SQ, SG, SH, SI, SL, SK, SM, SF, SP, SS, ST, SW, SY, SV, AN, RN, NN, DN, CN, EN, QN, GN, HN, IN, LN, KN, MN, FN, PN, SN, TN, WN, YN, VN, AD, RD, ND, DD, CD, ED, QD, GD, HD, ID, LD, KD, MD, FD, PD, SD, TD, WD, YD, VD, CA, CR, CN, CD, CC, CE, CQ, CG, CH, CI, CL, CK, CM, CF, CP, CS, CT, CW, CY, CV, AE, RE, NE, DE, CE, EE, QE, GE, HE, IE, LE, KE, ME, FE, PE, SE, TE, WE, YE, VE

In some embodiments, the degron peptide is SEQ ID NO: 130 as follows: RRRG. In accordance with such embodiments, the destabilization domain has the sequence of HIV Tat-RRRG (SEQ ID NO: 127) as follows: RKKRQRQRRRG.

In other embodiments, the degron peptide is selected from the group consisting of FKBP12, dihydrofolate reductase, and derivatives thereof. See, e.g., Rakshit et al., “Evaluation of FKBP and DHFR Based Destabilizing Domains in *Saccharomyces Cerevisiae*,” *Bioorg. Med. Chem. Lett.* 21:4965-4968 (2011) and Iwamoto et al., “A General Chemical Method to Regulate Protein Stability in the Mammalian Central Nervous System,” *Chem. Biol.* 17:981-988 (2010), which are hereby incorporated by reference in their entirety). In some embodiments, the FKBP12 is a human FKBP12. In some embodiments, the dihydrofolate reductase is an *E. coli* dehydrogenase (ecDHFR). As described herein, aptam-

ers that selectively bind to FKBP12, DHFR, or derivatives thereof may be used to confer stability to a protein of interest comprising FKBP12, ecDHFR, or a derivative thereof as a fusion partner.

In some embodiments, the destabilization domain has the consensus sequence of SEQ ID NO: 62 as follows: XXXXXXXXXXXXXXXXXX, where X at position 1 can be S or A; X at position 2 can be G or A; X at position 3 can be P or A; X at position 4 can be R or K; X at position 5 can be P, A, I, Y, K, or R; X at position 6 can be R, K, V, or Y; X at position 7 can be G, A, or R; X at position 8 can be T or A; X at position 9 can be R or K; X at position 10 can be G or A; X at position 11 can be K or A; X at position 12 can be G or A; X at position 13 can be R or K; X at position 14 can be I or A; X at position 15 can be R, K, Y, or G; X at position 16 can be R, K, V, T, or Y; X at position 17 can be any amino acid but preferably R, G, E, S, or C; and x at position 18 is optional and can be any amino acid, but preferably G, E, O, N, D, or E.

In some embodiments the destabilization domain has the sequence of tDeg (SEQ ID NO: 63) as follows: SGPR-PRGTRGKGRRIRRRG.

The nucleic acid molecule described herein may further comprise a third nucleic acid sequence encoding a second protein of interest, wherein the third nucleic acid sequence is located between the first nucleic acid sequence and second nucleic acid sequence. Suitable proteins of interest are described in more detail above and include, without limitation, a fluorescent protein, a bioluminescent protein, an enzyme, or a transcriptional regulator.

Another aspect of the invention relates to a nucleic acid molecule encoding a lentiviral transactivator of transcription (Tar) RNA aptamer sequence.

In some embodiments, the lentiviral transactivator of transcription (Tar) RNA aptamer sequence is a bovine immunodeficiency virus (BIV) Tar sequence. In other

embodiments, the lentiviral transactivator of transcription (Tar) RNA sequence is a human immunodeficiency virus (HIV) Tar sequence.

According to some embodiments, the nucleic acid molecule encoding the lentiviral Tar RNA sequence is a DNA molecule according to the consensus sequence of SEQ ID NO: 64 as follows: NNNNN-SHSYWSBMNNNNDHSBHSNNNN, where N can be A, C, G, or T; S can be C or G; H can be A, C, or T; Y can be C or T; W can be A or T; B can be C, G, or T; M can be A or C; and D can be A, G, or T. For example, the nucleic acid molecule encoding the lentiviral Tar RNA sequence may be a DNA molecule encoding wild-type TAR RNA as follows: GGCTCGTAGCTCATTAGCTCCGAGCC (SEQ ID NO: 65).

According to some embodiments, the nucleic acid molecule encoding the lentiviral TAR RNA sequence is a DNA molecule according to the consensus sequence of SEQ ID NO: 66 as follows: NNNNN-SHCYWSBMNNNNDSBHBSNNNNN, where N can be A, C, G, or T; S can be C or G; H can be A, C, or T; Y can be C or T; W can be A or T; B can be C, G, or T; M can be A or C; and D can be A, G, or T. For example, the nucleic acid molecule encoding the lentiviral Tar RNA sequence may be a DNA molecule encoding TAR Variant-1 as follows: GGCTCGTCTGAGCTCATTAGCTCCGAGCC (SEQ ID NO: 67).

According to some embodiments, the nucleic acid molecule encoding the lentiviral TAR RNA sequence is a DNA

molecule according to the consensus sequence of SEQ ID NO: 68 as follows: NNNNNSI-TYSWSBMNNNNDSBHBSNNNNN, where N can be A, C, G, or T; S can be C or G; H can be A, C, or T; Y can be C or T; W can be A or T; B can be C, G, or T; M can be A or C; and D can be A, G, or T. For example, the nucleic acid molecule encoding the lentiviral Tar RNA sequence may be a DNA molecule encoding TAR Variant-2 (Pepper) as follows: GGCTCGTTGAGCTCATTAGCTCCGAGCC (SEQ ID NO: 69).

Suitable additional lentiviral transactivator of transcription (Tar) RNA aptamer sequences of the present application are shown in Table 7 below.

TABLE 7

TAR RNA	Sequence	SEQ ID NO:
(Pepper) ₁₀ tag	GCGUCGUGAGCUAUUAGCUCGGAGCCGUCCAGCGCAAACUAU UACGAAAAAACUCCGACGGGCUCGUUGAGCUAUAGCUCCGAGC CCGUGCGGAAACCUCACAAAACAGCAAAACGGGUCGUUGA GCUCAUUAGCUCCGAGCCGCGACAACCCACAACUUAACCCA GCAAACGGCUCGGUCAGCUAUAGCUCGGAGCCGUUAACAGA CCGAACGGCGCAAGAUAUUGACACGGGUCUGUUGAGCUAUAGC UCCGAGCCGACCCUCGUAGAUAGUAGGUUCUUAGGCAUUGGC UCGUUGAGCUAUAGCUCGGAGCCAAAGAUUCGACUGCAAUUCCG AUUAGACGUACACGGCUCGUUGAGCUAUAGCUCGGAGCCGAU CCAACCUACUUCUCCAUAAACUACCCGGUCGUUGAGCUCAU UAGCUCGGAGCCGAUAAACCAAACUCCGAGCCGAGAACCAAUC GGCUCGUUGAGCUCAUAGCUCGGUCUGAGCUAUAGCUCGGAGC ACAUACACACCACACUCCGUCGUAGCUCAUAGCUCGGAGC C	70
(F30- 1xPepper) ₁₀ tag	UUGCCAUGUGUAUGUGGGAUUGCUCGUUGCCACGUUUCACAUAC UAGAUGACCGUAGCAAGGCUUGAGCUAUAGCUCGGAG CCCGAGGUACCGGAUCAUCAUGGCAAGGUCCAGGCAAUCUAAA CGAAAUAUCAUCGGACGUUCGCGGAUUGCUAUGCAGGGAU CGUUCUCCGCAUAGCUCGAUCAUCGGUAGCAGGCUUGAG CUCAUAGCUCGGAGCCGAGGUACCGGAUAGUUAUCGGCAGC UGCGGAAAACUCACAAAACUGCUAAACGUCGCCGUGUGUG UAGGAUGCGGUUCCGACGUUCCUACACACUGACGAUCCGCUAG CAAAGGCUCGUUGAGCUAUAGCUCGGAGCCGAGGUACCGGA CGUUCACGGCGACGCCGUAUAUCCACAUACUACUACAGGCAAU CUUUGCCAUGUGUAUGUGGGAUUGCUCGUUGCCACGUUUC CUGAUGAUCCGCUAGCAAAGGCUCGUUGAGCUAUAGCUCGGAG CCCAGGGUACCGGAUCAUCAUGGCAAGAUCAAGAUCAACGGC GCAAGGAAUUGUCACGGCAGUUGCUAUAGCGGUUAGCG CGUUCUCCGCAUAGCUCGAUCAUCGGUAGCAGGCUUGCC GUCAUUAUAGCUCGGAGCCGAGGUACCGGAUAGUUAUCGG UCCUCGCUAGAUUAUGUAGGUUCUAGGCAUUCGGCUGUGUG GUAGGAUGCGUJGCCACGUUCCUACACACUUCUGACGAUCCG GCAAAAGGCUCGUUGAGCUAUAGCUCGGAGCCGAGGUACCGGA UCGUUCACGGCGAAAGAACUGCUUGCGUAAUUCGAGCUACA CUUUGCCAUGUGUAUGUGGGAUUGCUCGUUGCCACGUUUC CUGAUGAUCCGCUAGCAAAGGCUCGUUGAGCUAUAGCUCGG CCCAGGGUACCGGAUCAUCAUGGCAAGAUCAAGCUACUCC CAUACAUAAUCCUCGGCAUUCGCUAGCUAGGCUUGCCAC UUUCCCGCAUAGCUCGUCAUCCGCUAGCAAGGCUCGUUGAG CAUAGCUCGGCGAGGUACCGGAUAGCUAUUCGCGAGA UAACGCAAUACCGUACACUGCUAAUCCUCGGCUGUGUGUG GAUGCGUJGCCACGUUCCUACACACUUCUGACGAUCCG AGGCUCGUUGAGCUAUAGCUCGGAGCCGAGGUACCGGA UUACCGCGAGGAUUAACUACAUACACACCACAAUC UCCCAUGUGUAUGUGGGAUUGCUCGUUGCCACGUUUC GAUGAUCCGCUAGCAAAGGCUCGUUGAGCUAUAGCUCGG CCGAGGUACCGGAUCAUCAUGGCA	71
(Pepper) ₂₀ tag	GCCUCGUAGCUCAUUAGCUCGGAGCCGUCCAGCGCAAACUAU UACGAAAAAACUCCGACGGGCUCGUUGAGCUAUAGCUCGGAGC CCGUGCGGAAACCUCACAAAACAGCAAAACGGGUCGUUGA GCUCAUUAGCUCCGAGCCGCGACAACCCACAACUUAACCCA GCAAACGGCUCGUUGAGCUAUAGCUCGGAGCCGUUA CCGAACGGCGCAAGAUUAGACGGGUCUGUUGAGCUAUAGC UCCGAGCCGACCCUCGUAGAUAGUAGGUUCUAGGCAU UCGUUGAGCUAUUAGCUCGGAGCCAAAGAUAGCUCGG C	72

TABLE 7 -continued

TAR RNA Sequences		SEQ ID NO:
TAR RNA	Sequence	
	AUUAGACGUACACGGCUCGUCUGAGCUAAUAGCUCCGAGCCGAU CCAACCUACUUCUCUCAUAACUAACCUCCGGCUCGUUGAGCUAU UAGCUCCGAGCCGAUCUAACGCAAAUACCGUACACUGUCCAAUCC GGCUCGUUGAGCUAUAGCUCGGCAGCCGAACACAUUGACAU ACAUACACACCAACUCGCUCGUAGCUCAUUAACGUCCGAGC CGAAUUGGUUCGUUCUUCUUGGGCGCGCUCGUACUAAGGUGACA UGGACCAAACCCUCCGGCUCGUAGCUCAUUAAGCUCGGAGCCAC CUCACCAACAAGACAAAACUACUUCUAGGUUCGUUGAGCUA UUAGCUCCGAGCCUAACACUAGCAUACAUUGUGGCCAUUU UGGCUCGUUGAGCUAUAGCUCGGAGCCAUUGCUCACGAAU UCAAACACAGGAAACUACAAAAGGGGUUCGUUGAGCUAUAGCUC CGUUCACGUCCAAUACGAUUAACUACUUCUUGGGCGUUGAGC UCAUAGCUCGGCAGCCCCAGCUACAUACUUCACUAGGACAU UCAAGGGCUCGUUGAGCUAUAGCUCGGAGCCCUCCACAAGUC UCAACCCACAGAAACUACAAAAGGGGUUCGUUGAGCUAUAGCUC CGAGCCCACUCCUACCUAACCCUUCUCCCAACAAAAGGGGUUC GUUGAGCUAUAGCUCGGAGCCCCCAUCCAACAUACCAA AAAACAAUUAUCGGCUCGUUGAGCUAUAGCUCGGAGCCAGCC CACAUUCUCACUACUAACAAAACAAACGGCUCGUUGAGCUA UUAGCUCCGAGCC	
(F30- 2xPepper) _{10tag}	UUGCCAUGUGUAUGUGGGAAAGCGUAGAAAGGCUCGUUGAGCUAU UAGCUCCGAGCCGAUCAGGUUCCACAUACUUGAGCUAUAGCUC GGACAAAGGCUCGUUGAGCUAUAGCUCGGAGCCGAGGUACC GGGUACAUUAGCAGGCAAGUCCAGCGCAACUUAACGAAAAUAC CGACGUCCGGAUGCUUAUGCGGGAAAGGUAGAAAGGUCUGUGA GUCAUUAAGCUCGGAGCCGAUACGUUUCUCCGCAUAGCUCGA AUCCGCUAGCAGGCAAGGCUUCGUUGAGCUAUAGCUCGGAG GUACGGGAUGAUUACUCCGAGCUCGUUGAGCUAUAGCUCGGAG ACGUACAAAGCUCGCCGUGUGUGUAGGAAGCGUAGAAAGGUCG UCUGAGCUAUUAGCUCGGAGCCGACUACGUUUCUACACACUC UGACGAUCCGCUAGCAAAGGCUCGUUGAGCUAUAGCUCGGAG CCGAGGUACCCGAUCGUUCACGGCAGCCGAUAAUCCACAUACU UACAAUAGCAGGAAUUCUUGCCAUGUGUAUGUGGGAAAGCGUAG GCUCGUUGAGCUAUAGCUCGGAGCCGACUACGUUUCUCCACAU ACUCCGAGGUACCCGAUCAGGCUAGGUACUACGGCAGCCGAC GAGCCGAGGUACCCGAUCGUACUUCUAGGCAAGUAUACAGAU GGCCGAAAGAUUUGUCACGUCCGAGGUACUAGCGGGAAAGCGUAG AAAGGCUCGUUGAGCUAUAGCUCGGAGCCGACUACGUUUC GCAUAGCUCGUAGCAGGCUACACUAGCUCGGAGGUACU GGCUCGGAGCCGAGGUACCCGAUCGUACUACGGCAGCC AGAUUAGGUAGGUUUCUAGGCAUUCUCCGUGUGUGUAGGAAG CGUAGAAAGGCUCGUUGAGCUAUAGCUCGGAGCCGACUACGU UUCCUACACACUUCUGAGCUACCCGUAGCAGGCUACGGCUG CAUAGCUCGGAGCCGAGGUACCCGAUCGUACGGCAGGAAAGA UCGUUCGAAUUCUCCGAGGUACGGCUACUACUUGCCAUGUGU GAAGCGUAGAAAGGCUCGUUGAGCUAUAGCUCGGAGCCGAC UACGUUUCUCCACAUACUUCUGAGCUACCCGUAGCAGGCUUG AGCUCAUAGCUCGGAGCCGAGGUACCCGAUCGUACGGCAAG AUCCAGUACUUCUCCACAUACUACUCCUCCUCCGAGGUAC CGGGAGCGUAGAAAGGCUCGUUGAGCUAUAGCUCGGAGCC GACUACGUUUCUCCGAGGUACCCGUAGCAGGCUACGGCUG UUGAGCUAUAGCAGGCUACACUACUACGGGUACCCGAUC GAGAUCAUACGCAAACGGUACACUACGGGUACCCGUUGUG UGUGUAGGAAGCGUAGAAAGGCUCGUUGAGCUAUAGCUCGG GCCGGACUACGUUUCUCCACACUACUACGGGUACCCGAUC CUCGUUGAGCUAUAGCUCGGAGCCGAGGUACCCGAUC CGGGAGGUAAAUCUACUACACACACAAUUCUUGCC AUGUGUAUGUGGGAAAGCGUAGAAAGGCUCGUUGAGCUAUAGC UCCGAGCCGGACUACGUUUCUCCACAUACUUCUGAGCUAC AAAGGCUCGUUGAGGUACUCCGAGGUACCCGA CAUUCUAGGCAA	73

In some embodiments, the nucleic acid molecule further encodes at least one additional RNA aptamer. Thus, in some embodiments, the nucleic acid molecule may encode a lentiviral transactivator of transcription (Tar) RNA aptamer operably coupled to at least one additional RNA aptamer. The at least one additional aptamer may be a S-adenosyl-methionine (SAM)-binding aptamer. For example, the nucleic acid molecule may encode a SAM-binding aptamer

operably linked to the lentiviral transactivator of transcription (Tar) RNA aptamer. As described herein, binding of SAM to its aptamer promotes folding of other linked aptamers, such as Pepper. In this way, the expressed RNA is a "sensor" which couples SAM levels to Pepper folding.

Also contemplated are nucleic acid molecules encoding a protein-binding RNA sequence. Thus, in some embodiments, the nucleic acid molecule encodes a non-lentiviral

transactivator of transcription (Tar) RNA sequence. In accordance with such embodiments, the protein-binding RNA sequence is BoxB or RRE.

Some embodiments of the present application relate to a vector comprising a nucleic acid molecule described herein (i.e., a nucleic acid molecule encoding an RNA-regulated fusion protein and/or a lentiviral transactivator of transcription (Tar) RNA sequence). As used herein, the term vector means any genetic element, such as a plasmid, phage, transposon, cosmid, chromosome, virus, virion, etc., which is capable of replication when associated with the proper control elements and which is capable of transferring gene sequences between cells. Thus, the term includes cloning and expression vectors, as well as viral vectors. The heterologous nucleic acid molecule is inserted into the expression system or vector in proper sense (5' to 3') orientation and correct reading frame. The vector contains the necessary elements for the transcription and/or translation of the inserted protein and/or RNA coding sequences of the present application.

In one embodiment, the vector is a plasmid. Numerous vectors suitable for use in the compositions of the present application are known to those of skill in the art, and many are commercially available. The following vectors are provided by way of example; for eukaryotic cells: pcDNA3.1 (+), Tornado (Litke & Jaffrey, "Highly Efficient Expression of Circular RNA Aptamers in Cells Using Autocatalytic Transcripts," *Nat. Biotechnol.* 37(6):667-675(2019), which is hereby incorporated by reference in its entirety), pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG, and pSVLSV40 (Pharmacia). However, any other vector may be used so long as it is compatible with the cell.

In another embodiment, the vector is a viral vector. Suitable viral expression vectors include, but are not limited to, viral vectors based on vaccinia virus; poliovirus; adenovirus (see, e.g., PCT Patent Application Publication Nos. WO 94/12649 to Gregory et al., WO 93/03769 to Crystal et al., WO 93/19191 to Haddada et al., WO 94/28938 to Wilson et al., WO 95/11984 to Gregory, and WO 95/00655 to Graham, which are hereby incorporated by reference in their entirety); adeno-associated virus (see, e.g., Flannery et al., "Efficient Photoreceptor-Targeted Gene Expression In Vivo by Recombinant Adeno-Associated Virus," *PNAS* 94:6916-6921 (1997); Bennett et al., "Real-Time, Noninvasive In Vivo Assessment of Adeno-Associated Virus-Mediated Retinal Transduction," *Invest. Ophthalmol. Vis. Sci.* 38:2857-2863 (1997); Jomary et al., "Nonviral Ocular Gene Transfer," *Gene Ther.* 4:683-690 (1997); Rolling et al., "Evaluation of Adeno-Associated Virus-Mediated Gene Transfer into the Rat Retina by Clinical Fluorescence Photography," *Hum. Gene. Ther.* 10:641-648 (1999); Ali et al., "Gene Transfer Into the Mouse Retina Mediated by an Adeno-Associated Viral Vector," *Hum. Mol. Genet.* 5:591-594 (1996); Samulski et al., "Helper-Free Stocks of Recombinant Adeno-Associated Viruses: Normal Integration Does not Require Viral Gene Expression," *J. Vir.* 63:3822-3828 (1989); Mendelson et al., "Expression and Rescue of a Nonselected Marker from an Integrated AAV Vector," *Virol.* 166:154-165 (1988); and Flotte et al., "Stable In Vivo Expression of the Cystic Fibrosis Transmembrane Conductance Regulator With an Adeno-Associated Virus Vector," *PNAS* 90:10613-10617 (1993), which are hereby incorporated by reference in their entirety); SV40; herpes simplex virus; human immunodeficiency virus (see, e.g., Miyoshi et al., "Stable and Efficient Gene Transfer into the Retina Using an HIV-Based Lentiviral Vector," *PNAS* 94:10319-10323 (1997), which is hereby incorporated by reference in its

entirety); a retroviral vector, e.g., Murine Leukemia Virus, spleen necrosis virus, and vectors derived from retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, a lentivirus, human immunodeficiency virus, myeloproliferative sarcoma virus, and mammary tumor virus and the like.

As described herein supra, the nucleic acid molecules encoding a protein of interest described herein may be inserted into a vector in the sense (i.e., 5' to 3') direction, such that the nucleic acid sequence encoding an RNA-regulated fusion protein is properly oriented for the expression of the encoded protein under the control of a promoter of choice. In some embodiments, the nucleic acid molecules encoding a RNA aptamer are inserted into the vector in the sense direction, such that the nucleic acid molecule encoding the RNA aptamer is properly oriented for the expression of a desired RNA aptamer. Single or multiple nucleic acid molecules may be ligated into an appropriate vector in this 10 way, under the control of a suitable promoter, to prepare a nucleic acid construct. A promoter is a DNA sequence which contains the binding site for RNA polymerase and initiates transcription of a downstream nucleic acid sequence. In one embodiment, the vector comprises a promoter. Thus, in 15 some embodiments, the vector comprises a nucleic acid molecule encoding a lentiviral transactivator of transcription (Tar) aptamer (e.g., Pepper) operably coupled to a promoter. In other embodiments, the vector comprises a nucleic acid molecule encoding a lentiviral transactivator of transcription 20 (Tar) aptamer (e.g., Pepper) and at least one additional aptamer sequence (e.g., a S-adenosylmethionine (SAM)-binding aptamer) operably coupled to a promoter.

The promoter may be a constitutively active promoter (i.e., a promoter that is constitutively in an active or "on" 25 state), an inducible promoter (i.e., a promoter whose state, active or inactive state, is controlled by an external stimulus, e.g., the presence of a particular temperature, compound, or protein), a spatially restricted promoter (i.e., transcriptional 30 control element, enhancer, etc.) (e.g., tissue specific promoter, cell type specific promoter, etc.), or a temporally restricted promoter (i.e., the promoter is in the "on" state or "off" state during specific stages of a biological process).

Suitable promoters can be derived from viruses and can therefore be referred to as viral promoters, or they can be 45 derived from any organism, including prokaryotic or eukaryotic organisms. Suitable promoters can be used to drive expression by any RNA polymerase (e.g., RNA Polymerase I, RNA Polymerase II, RNA Polymerase III). The promoter may be a viral promoter. Exemplary promoters 50 include, but are not limited to the SV40 early promoter, mouse mammary tumor virus long terminal repeat (LTR) promoter; adenovirus major late promoter (Ad MLP); a herpes simplex virus (HSV) promoter, a cytomegalovirus (CMV) promoter such as the CMV immediate early promoter region (CMVIE), a rous sarcoma virus (RSV) promoter, a human U6 small nuclear promoter (U6) (Miyagishi et al., "U6 Promoter-Driven siRNAs with Four Uridine 3' Overhangs Efficiently Suppress Targeted Gene Expression in Mammalian Cells," *Nat. Biotechnol.* 20:497-500 (2002), 55 which is hereby incorporated by reference in its entirety), an enhanced U6 promoter (e.g., Xia et al., "An Enhanced U6 Promoter for Synthesis of Short Hairpin RNA," *Nucleic Acids Res.* 31(17):e100 (2003), which is hereby incorporated by reference in its entirety), a human H1 promoter ("H1"), 60 and the like. In some embodiments the promoter is a phage promoter, e.g., a T7 promoter that has been engineered to be expressed in a mammalian cell.

41

Examples of inducible promoters include, but are not limited to T7 RNA polymerase promoter, T3 RNA polymerase promoter, isopropyl-beta-D-thiogalactopyranoside (IPTG)-regulated promoter, lactose induced promoter, heat shock promoter, tetracycline-regulated promoter, steroid-regulated promoter, metal-regulated promoter, estrogen receptor-regulated promoter, etc. Inducible promoters can therefore be regulated by molecules including, but not limited to, doxycycline, RNA polymerase, e.g., T7 RNA polymerase, an estrogen receptor, an estrogen receptor fusion, etc.

In some embodiments, the promoter is a eukaryotic RNA polymerase promoter or a derivative thereof. Exemplary RNA polymerase II promoters include, without limitation,

42

cytomegalovirus ("CMV"), phosphoglycerate kinase-1 ("PGK-1"), and elongation factor 1 α ("EF1 α ") promoters. In yet another embodiment, the promoter is a eukaryotic RNA polymerase III promoter selected from the group consisting of U6, H1, 56, 7SK, and derivatives thereof.

The RNA Polymerase promoter may be mammalian. Suitable mammalian promoters include, without limitation, human, murine, bovine, canine, feline, ovine, porcine, 5 ursine, and simian promoters. In one embodiment, the RNA polymerase promoter sequence is a human promoter.

According to one embodiment, the vector is a plasmid and has the sequence of pCMV-mCherry-(F30-2 \times Pepper)₁₀ (SEQ ID NO: 74; GenBank Accession No. MN052904.1, which is hereby incorporated by reference) as follows:

```

1   GACGGATCGG GAGATCTCCC GATCCCCTAT GGTGCACTCT CAGTACAATC
51  TGCTCTGATG CCGCATAGTT AAGCCAGTAT CTGCTCCCTG CTTGTTGTT
101 GGAGGTCGCT GAGTAGTGCG CGAGCAAAAT TTAAGCTACA ACAAGGCAAG
151 GCTTGACCGA CAATTGATG AAGAACATGC TTAGGGTTAG GCGTTTGCG
201 CTGCTTCGCG ATGTACGGGC CAGATATAACG CGTTGACATT GATTATTGAC
251 TAGTTATTAA TAGTAATCAA TTACGGGTC ATTAGTTCAT AGCCCATATA
301 TGGAGTTCCG CGTTACATAA CTTACGGTAA ATGGCCCCC TGGCTGACCG
351 CCCAACGACC CCCGCCATT GACGTCAATA ATGACGTATG TTCCCATAGT
401 AACGCCAATA GGGACTTTCC ATTGACGTCA ATGGGTGGAG TATTTACGGT
451 AAACTGCCA CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTACGCC
501 CCTATTGACG TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCAGTA
551 CATGACCTTA TGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA
601 TCGCTATTAC CATGGTGATG CGGTTTGGC AGTACATCAA TGGCGTGGA
651 TAGCGGTTG ACTCACGGGG ATTCCAAGT CTCCACCCCA TTGACGTCAA
701 TGGGAGTTTG TTTTGGCACC AAAATCAACG GGACTTTCCA AAATGTGCGTA
751 ACAACTCCGC CCCCATGGAC GCAAATGGGC GGTAGGGCGT TACGGTGGGA
801 GGTCTATATA AGCAGAGCTC TCTGGCTAAC TAGAGAACCC ACTGCTTACT
851 GGCTTATCGA AATTAATACG ACTCACTATA GGAGAACCCA AGCTGGCTAG
901 CGTTTAAACT TAAGCTTGCC ACCATGGTGA GCAAGGGCGA GGAGGATAAC
951 ATGGCCATCA TCAAGGAGTT CATGCGCTTC AAGGTGCACA TGGAGGGCTC
1001 CGTGAACGGC CACGAGTTCG AGATCGAGGG CGAGGGCGAG GGCGCCCCCT
1051 ACAGGGGCAC CCAGACCGCC AAGCTGAAGG TGACCAAGGG TGGCCCCCTG
1101 CCCTTCGCCT GGGACATCCT GTCCCCTCAG TTCAATGTACG GCTCCAAGGC
1151 CTACGTGAAG CACCCCGCCG ACATCCCCGA CTACTTGGAG CTGTCCTTCC
1201 CCGAGGGCTT CAAGTGGGAG CGCGTGATGA ACTTCGAGGA CGGCGCGTGT
1251 GTGACCGTGA CCCAGGACTC CTCCCTGCAG GACGGCGAGT TCATCTACAA
1301 GGTGAAGCTG CGCGGCACCA ACTTCCCCTC CGACGGCCCC GTAATGCAGA
1351 AGAAGACCAT GGGCTGGGAG GCCTCCTCCG AGCGGATGTA CCCCAGGAC
1401 GGCAGCCCTGA AGGGCGAGAT CAAGCAGAGG CTGAAGCTGA AGGACGGCGG
1451 CCACTACGAC GCTGAGGTCA AGACCACCTA CAAGGCCAAG AAGCCCGTGC
1501 AGCTGCCGG CGCCTACAAAC GTCAACATCA AGTTGGACAT CACCTCCAC
1551 AACGAGGACT ACACCATCGT GGAACAGTAC GAACGCGCCG AGGGCCGCCA

```

- continued

1601 CTCCACCGGC GGCATGGACG AGCTGTACAA GTAACTCGAG ATCCGTTACG
 1651 GCCGGAATCA ATCGCTAATC ACTCAACTTG CCATGTGTAT GTGGGAAGCG
 1701 TAGAAAGGCT CGTTGAGCTC ATTAGCTCCG AGCCCAGCTA CGTTTCCCAC
 1751 ATACTCTGAT GATCCGCTAG CAAAGGCTCG TCTGAGCTCA TTAGCTCCGA
 1801 GCCCGAGGTA CCGGATCATT CATGGCAAGT CCAGCGCAAT CTATTACGAA
 1851 AATCATCCGA CGTCGCGATG TCTATGCCGG AAGCGTAGAA AGGCTCGTCT
 1901 GAGCTCATTA GCTCCGAGCC CGACTACGTT TCCCGCATAG TCTGATCATC
 1951 CGCTAGAAA GGCTCGTTGA GCTCATTAGC TCCGAGCCCG AGGTACCGGA
 2001 TGATTCTATCG CGACGCTGCG GAAAATCTCA CAAAATCACG TCAAACGTCG
 2051 CCGTGTGTGT GTAGGAAGCG TAGAAAGGCT CGTCTGAGCT CATTAGCTCC
 2101 GAGCCGACT ACGTTTCCTA CACACTCTGA CGATCCGCTA GCAAAGGCTC
 2151 GTTGAGCTCA TTAGCTCCGA GCCCGAGGTA CCGGATCGTT CACGGCGACG
 2201 CCGATAATCC ACATACTTAC AATCAGGAA TCTTGCCATG TGTATGTGGG
 2251 AAGCGTAGAA AGGCTCGTT AGCTCATTAG CTCCGAGCCC GACTACGTTT
 2301 CCCACATACT CTGATGATCC GCTAGCAAAG GCTCGTTGAG CTCATTAGCT
 2351 CCGAGCCCGA GGTACCGGAT CATTGATGGC AAGTATCAAG ATCGAACGGC
 2401 GCAAGATATT GTCACGTCGC GATGTCTATG CGGGAAAGCGT AGAAAGGCTC
 2451 GTTGAGCTCA TTAGCTCCGA GCCCGACTAC GTTTCCCGCA TAGTCTGATC
 2501 ATCCGCTAGC AAAGGCTCGT CTGAGCTCAT TAGCTCCGAG CCCGAGGTAC
 2551 CGGATGATTC ATCGCGACGT CCTCGCTAGA TATGTTAGGT TCTTAGGCAT
 2601 TTCGCCGTGT GTGTGTAGGA AGCGTAGAAA GGCTCGTTGA GCTCATTAGC
 2651 TCCGAGCCCG ACTACGTTTC CTACACACTC TGACGATCCG CTAGCAAAGG
 2701 CTCGTCTGAG CTCATTAGCT CCGAGCCCGA GGTACCGGAT CGTTCACGGC
 2751 GAAAAGATCG TCTGCAATTG CGATTAGACG TACACTTGCC ATGTGTATGT
 2801 GGGAAAGCGTA GAAAGGCTCG TCTGAGCTCA TTAGCTCCGA GCCCGACTAC
 2851 GTTTCCCAACA TACTCTGATG ATCCGCTAGC AAAGGCTCGT TGAGCTCATT
 2901 AGCTCCGAGC CCGAGGTACC GGATCATTCA TGGCAAGATC CAAGCTACTT
 2951 CCTCCATACC TATCCTCCTC GCGATGTCTA TGCGGGAAAGC GTAGAAAGGC
 3001 TCGTCTGAGC TCATTAGCTC CGAGCCCGAC TACGTTCCCG GCATAGTCTG
 3051 ATCATCCGCT AGCAAAGGCT CGTTGAGCTC ATTAGCTCCG AGCCCGAGGT
 3101 ACCGGATGAT TCATCGCGAG ATCATAACGC AATACCGTAC ACTGTTCAAT
 3151 CCTCGCCGTG TGTGTGTAGG AAGCGTAGAA AGGCTCGTCT GAGCTCATTA
 3201 GCTCCGAGCC CGACTACGTT TCCTACACAC TCTGACGATC CGCTAGCAA
 3251 GGCTCGTTGA GCTCATTAGC TCCGAGCCCG AGGTACCGGA TCGTTCACGG
 3301 CGAGGATAAT CAATCCACAT ACATCACACCC ACAATTCTTG CCATGTGTAT
 3351 GTGGGAAGCG TAGAAAGGCT CGTCTGAGCT CATTAGCTCC GAGCCGACT
 3401 ACGTTTCCCA CATACTCTGA TGATCCGCTA GCAAAGGCTC GTCTGAGCTC
 3451 ATTAGCTCCG AGCCCGAGGT ACCGGATCAT TCATGGCAAG AATTGGTCGT
 3501 TCTTCTTGGC GGCGCTCGA CTAATCACC GGTAATCTTC TTGTCATCT
 3551 AGACCTTATA AAGATCTTG TACAAGGGCC CGTTAAACC CGCTGATCAG

-continued

3601 CCTCGACTGT GCCTTCTAGT TGCCAGCCAT CTGTTGTTG CCCCTCCCC
 3651 GTGCCTCCT TGACCCCTGGA AAGGTGCCAC TCCCCTGTC CTTTCTTAAT
 3701 AAAATGAGGA AATTGCATCG CATTGTCCTGA TAGGGTGTCA TTCTTATTCTG
 3751 GGGGGTGGGG GTGGGGGCAG GACAGCAAGG GGGAGGATTG GGAAGACAAT
 3801 AGCAGGCATG CTGGGGATGC GGTGGGCTCT ATGGCTTCTG AGGCGGAAAG
 3851 AACAGCTGG GGCTCTAGGG GGTATCCCCA CGCGCCCTGT AGCGGCGCAT
 3901 TAAGCGCGGC GGGTGTGGTG GTTACGCGCA CGGTGACCGC TACACTTGCC
 3951 AGCGCCCTAG CGCCCGCTCC TTTCGCTTTC TTCCCTTCCT TTCTCGCCAC
 4001 GTTCGCCGGC TTTCCTCGTC AAGCTCTAAA TCAGGGGCTC CCTTTAGGGT
 4051 TCCGATTTAG TGCTTACGG CACCTCGACC CCAAAAAACT TGATTAGGGT
 4101 GATGGTCAC GTAGTGGGC ATCGCCCTGA TAGACGGTTT TTGCCCCCTT
 4151 GACGTTGGAG TCCACGTTCT TTAATAGTGG ACTCTTGTTC CAAACTGGAA
 4201 CAACACTCAA CCCTATCTCG GTCTATTCTT TTGATTATA AGGGATTTG
 4251 CCGATTTCGG CCTATTGGTT AAAAAATGAG CTGATTTAAC AAAAAATTAA
 4301 CGCGAATTAA TTCTGTGGAA TGTGTGTAG TTAGGGTGTG GAAAGTCCCC
 4351 AGGCTCCCCA GCAGGCAGAA GTATGCAAAG CATGCATCTC AATTAGTCAG
 4401 CAACCAGGTG TGAAAGTCC CCAGGCTCCC CAGCAGGCAG AAGTATGCAA
 4451 AGCATGCATC TCAATTAGTC AGCAACCATA GTCCCGCCCC TAACTCCGCC
 4501 CATCCCGCCC CTAACCTCCGC CCAGTTCCGC CCATTCTCCG CCCATGGCT
 4551 GACTAATTCTT TTTTATTAT GCAGAGGCCG AGGCCGCCTC TGCCCTTGAG
 4601 CTATTCCAGA AGTAGTGAGG AGGCTTTTT GGAGGCCAG GCTTTGCAA
 4651 AAAGCTCCCG GGAGCTTGTA TATCCATTCTT CGGATCTGAT CAAGAGACAG
 4701 GATGAGGATC GTTTCGCATG ATTGAACAAG ATGGATTGCA CGCAGGTTCT
 4751 CCGCCCGCTT GGGTGGAGAG GCTATTCCGC TATGACTGGG CACACAGAC
 4801 AATCGGCTGC TCTGATGCCG CCGTGTCCG GCTGTCAGCG CAGGGCGCC
 4851 CGGTTCTTT TGTCAAGACC GACCTGTCCG GTGCCCTGAA TGAACGTGAG
 4901 GACGAGGCAG CGCGGCTATC GTGGCTGCC ACACGGCGG TTCCTGCC
 4951 AGCTGTGCTC GACGTTGTCA CTGAAGCGGG AAGGGACTGG CTGCTATTGG
 5001 GCGAAGTGCC GGGCAGGAT CTCCGTGTCAT CTCACCTTGC TCCCTGCCAG
 5051 AAAGTATCCA TCATGGCTGA TGCAATGCCG CGGCTGCATA CGCTTGATCC
 5101 GGCTACCTGC CCATTGACCA ACCAAGCGAA ACATCGCATC GAGCGAGCAC
 5151 GTACTCGGAT GGAAGCCGGT CTTGTGATC AGGATGATCT GGACGAAGAG
 5201 CATCAGGGGC TCGCGCCAGC CGAAACTGTC GCCAGGCTCA AGGCGCGCAT
 5251 GCCCGACGGC GAGGATCTCG TCGTGACCCA TGGCGATGCC TGCTTGCGA
 5301 ATATCATGGT GAAAATGGC CGCTTTCTG GATTGATCGA CTGTTGCCGG
 5351 CTGGGTGTGG CGGACCGCTA TCAGGACATA GCGTTGGCTA CCCGTGATAT
 5401 TGCTGAAGAG CTTGGCGGCCG AATGGGCTGA CCGCTTCCTC GTGCTTTACG
 5451 GTATCGCCGC TCCCGATTG CAGCGCATCG CCTTCTATCG CCTTCTTGAC
 5501 GAGTTCTCT GAGCGGGACT CTGGGGTTCG AAATGACCGA CCAAGCGACG
 5551 CCCAACCTGC CATCACGAGA TTTCGATTCC ACCGCCGCCT TCTATGAAAG
 5601 GTTGGCTTC GGAATCGTT TCCGGGACGC CGGCTGGATG ATCCTCCAGC

- continued

5651 GCGGGGATCT CATGCTGGAG TTCTTCGCC ACCCAAAC TT GTTTATTGCA
 5701 GCTTATAATG GTTACAAATA AAGCAATAGC ATCACAAATT TCACAAATAA
 5751 AGCATTTTT TCACTGCATT CTAGTTGTGG TTTGTCCAAA CTCATCAATG
 5801 TATCTTATCA TGTCTGTATA CCGTCGACCT CTAGCTAGAG CTTGGCGTAA
 5851 TCATGGTCAT AGCTGTTCC TGTGTGAAT TGTATCCGC TCACAATTCC
 5901 ACACAAACATA CGAGCCGGAA GCATAAAAGTG TAAAGCCTGG GGTGCCTAA
 5951 GAGTGAGCTA ACTCACATTA ATTGCGTTGC GCTCACTGCC CGCTTCCAG
 6001 TCGGGAAACC TGCGTGCCA GCTGCATTAA TGAATCGGCC AACCGCGGG
 6051 GAGAGGCGGT TTGCGTATTG GCGCCTCTTC CGCTTCCTCG CTCACTGACT
 6101 CGCTCGCTC GGTGCGTCGG CTGCGGCGAG CGGTATCAGC TCACTCAAAG
 6151 GCGGTAATAC GGTTATCCAC AGAACATCAGGG GATAACGCAG GAAAGAACAT
 6201 GTGAGAAAAA GGCCAGCAAA AGGCCAGGAA CCGTAAAAAG GCCGCGTTGC
 6251 TGGCGTTTT CCATAGGCTC CGCCCCCTG ACGAGCATCA CAAAAATCGA
 6301 CGCTCAAGTC AGAGGTGGCG AAACCCGACA GGACTATAAA GATACCAGGC
 6351 GTTCCCCCT GGAAGCTCCC TCGTGCCTC TCCTGTTCCG ACCCTGCCGC
 6401 TTACCGGATA CCTGTCCGCC TTTCTCCCTT CGGGAAAGCGT GGCGCTTTCT
 6451 CATA GCTCAC GCTGTAGGTA TCTCAGTTCG GTGTAGGTCG TTCGCTCAA
 6501 GCTGGGCTGT GTGCACGAAC CCCCCGTTCA GCCCGACCGC TGCGCCTTAT
 6551 CCGGTAACTA TCGTCTTGAG TCCAACCCGG TAAGACACGA CTTATGCCA
 6601 CTGGCAGCAG CCACTGGTAA CAGGATTAGC AGAGCGAGGT ATGTTAGGCGG
 6651 TGCTACAGAG TTCTTGAAGT GGTGGCTAA CTACGGCTAC ACTAGAACAGAA
 6701 CAGTATTGAG TATCTGCGCT CTGCTGAAGC CAGTTACCTT CGGAAAAAGA
 6751 GTTGGTAGCT CTTGATCCGG CAAACAAACC ACCGCTGGTA GCGGTTTTT
 6801 TGTTTGCAAG CAGCAGATTAA CGCGCAGAAA AAAAGGATCT CAAGAACATC
 6851 CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA AAAACTCACGT
 6901 TAAGGGATTT TGTCATGAG ATTATCAAAA AGGATTTCA CCTAGATCCT
 6951 TTTAAATTAA AAATGAAGTT TAAATCAAT CTAAAGTATA TATGAGTAAA
 7001 CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC TATCTCAGCG
 7051 ATCTGTCTAT TTGCTTCATC CATAGTTGCC TGACTCCCCG TCGTGTAGAT
 7101 AACTACGATA CGGGAGGGCT TACCATCTGG CCCCAGTGCT GCAATGATAC
 7151 CGCGAGACCC ACGCTCACCG GCTCCAGATT TATCAGCAAT AAACCAGCCA
 7201 GCCGGAAGGG CGGAGCGCAG AAGTGGTCT GCAACTTTAT CGGCCTCCAT
 7251 CCAGTCTATT AATTGTTGCC GGGAGCTAG AGTAAGTAGT TCGCCAGTTA
 7301 ATAGTTGCG CAACTTGTTT GCCATTGCTA CAGGCATCGT GGTGTCACGC
 7351 TCGTCGTTG GTATGGCTTC ATTCAAGCTCC GGTTCCCAAC GATCAAGGCG
 7401 AGTTACATGA TCCCCATGT TGTGAAAAA AGCGGTTAGC TCCTTCGGTC
 7451 CTCCGATCGT TGTCAAGAGT AAGTTGGCG CAGTGTATC ACTCATGGTT
 7501 ATGGCAGCAC TGCATAATTC TCTTACTGTC ATGCCATCCG TAAGATGCTT
 7551 TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC
 7601 GGCAGCGAG TTGCTCTTGC CCGCGTCAA TACGGGATAA TACCGCGCCA

-continued

7651 CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG
 7701 AAAACTCTCA AGGATCTTAC CGCTGTTGAG ATCCAGTTCG ATGTAACCCA
 7751 CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTCAC CAGCGTTCT
 7801 GGGTGAGCAA AAACAGGAAG GCAAAATGCC GCAAAAAAGG GAATAAGGCC
 7851 GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTCAA TATTATTGAA
 7901 GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT
 7951 TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC
 8001 ACCTGACGTC

According to one embodiment, the vector is a plasmid and has the sequence of pminiCMV-(mNeonGreen)₄-tDeg (SEQ ID NO: 75; GenBank Accession No. MN052905.1, which is hereby incorporated by reference) as follows:

1 GACGGATCGG GAGATCTCCC GATCCCCTAT GGTGCACTCT CAGTACAATC
 51 TGCTCTGATG CCGCATAGTT AAGCCAGTAT CTGCTCCCTG CTTGTGTGTT
 101 GGAGGTCGCT GAGTAGTGCG CGAGCAAAAT TTAAGCTACA ACAAGGCAAG
 151 GCTTGACCGA CAATTGCATG AAGAATCTGC TTAGGGTTAG GCGTTTGC
 201 CTGCTTCGCG ATGTACGGGC CAGATATAAG CGTTGGTAGG CGTGTACGGT
 251 GGGAGGCCTA TATAAGCAGA GCTAACGTTG CCACCATGGT GAGCAAGGGC
 301 GAGGAGGATA ACATGGCCTC TCTCCCAGCG ACACATGAGT TACACATCTT
 351 TGGCTCCATC AACGGTGTGG ACTTTGACAT GGTGGGTAG GGCACCGGCA
 401 ATCCAAATGA TGGTTATGAG GAGTTAAACC TGAAGTCCAC CAAGGGTGAC
 451 CTCCAGTTCT CCCCTGGAT TCTGGTCCCT CATATCGGGT ATGGCTTCCA
 501 TCAGTACCTG CCCTACCCCTG ACGGGATGTC GCCTTTCCAG GCCGCCATGG
 551 TAGATGGCTC CGGATACCAA GTCCATCGCA CAATGCAGTT TGAAGATGGT
 601 GCCTCCCTTA CTGTTAACTA CCGCTACACC TACGAGGGAA GCCACATCAA
 651 AGGAGAGGCC CAGGTGAAGG GGACTGGTTT CCCTGCTGAC GGTCTGTGA
 701 TGACCAAATC GCTGACCGCT CGCGACTGGT GCAGGTCGAA GAAGACTTAC
 751 CCCAACGACA AAACCACAT CAGTACCTTT AAGTGGAGTT ACACCACTGG
 801 AAATGGCAAG CGCTACCGGA GCACTGCGCG GACCACCTAC ACCTTTGCCA
 851 AGCCAATGGC GGCTAACTAT CTGAAGAACC AGCCGATGTA CGTGTCCGT
 901 AAGACGGAGC TCAAGCACTC CAAGACCGAG CTCAACTTCA AGGAGTGGCA
 951 AAAGGCCTTT ACCGATGTGA TGGGCATGGA CGAGCTGTAC AAGGGTGGAC
 1001 ATATGGGCAC AGGGTCCACA GGCGGTACCG GCGGAGTTTC CAAAGGAGAA
 1051 GAAGACAATA TGGCATCACT CCCCGCAACC CACGAGTTGC ATATTTCCG
 1101 TTCAATTAAT GGAGTAGATT TCGATATGGT TGGCCAGGGAA ACAGGAAACC
 1151 CAAACGACGG ATATGAAGAG CTTAATCTCA AAAGTACCAA AGGCGATCTG
 1201 CAATTTCTC CGTGGATACT CGTGCACAC ATTGGATACG GATTCACCA
 1251 ATATCTCCG TATCCGGATG GAATGTCCCC CTTCAAGCA GCAATGGTGG
 1301 ACGGGAGTGG TTATCAGGTA CACAGAACCA TGCAGTTCGA GGACGGGGCT
 1351 TCTCTGACCG TAAATTATAG GTTACATTAT GAAGGCTCAC ATATTAAGGG
 1401 CGAAGCACAG GTTAAAGGAA CGGGGTTTCC TGCGGATGGC CCCGTATGA
 1451 CTAATTCTCT GACAGCCGCA GATTGGTGTC GCTCCAAAAA GACATACCCG

- continued

1501 AATGATAAAGA CTATAATCTC AACATTCAAA TGGTCCTATA CGACAGGCCA
 1551 CGGGAAACGA TATAGATCCA CGGCTCGAAC AACTTACACA TTCGCTAAC
 1601 CTATGGCCGC CAATTACCTC AAAAATCAGC CCATGTATGT GTTTAGGAAA
 1651 ACCGAATTGA AGCATTCTAA AACCGAACCTT AATTTTAAGG AATGGCAGAA
 1701 GGCTTTCACA GACGTAATGG GGATGGATGA ACTCTATAAA TCAGGTCTCG
 1751 AGTCCTCAGG GGGAACGGGT GGGTCCGGAG GAGTTAGTAA AGGTGAAGAG
 1801 GACAATATGG CAAGTTGCC TGCGACTCAC GAGCTTCATA TCTTTGGTC
 1851 TATAATGGC GTTGACTTCG ATATGGTGG CCAAGGTACT GGCAACCCCA
 1901 ATGACGGTTA CGAGGAGTTG AATCTCAAGT CCACAAAAGG TGATCTTCAG
 1951 TTCAGCCCTT GGATTCTCGT ACCTCATATT GGATATGGCT TTCACCAGTA
 2001 CCTTCCATAC CCAGACGGTA TGTCACCCTT TCAAGCTGCG ATGGTGGATG
 2051 GTTCCGGCTA TCAGGTCCAC CGAACGATGC AATTCGAGGA CGGGGCCAGC
 2101 CTCACCGTTA ATTATAGGTA CACCTATGAG GGAAGTCACA TAAAGGGAGA
 2151 AGCCAAGTG AAAGGAACAG GATTCCCAGC TGATGGTCCA GTAATGACGA
 2201 ACTCCTTGAC AGCGGCTGAC TGGTGTAGAA GCAAAAAGAC GTATCCTAAT
 2251 GACAAGACCA TCATTAGCAC TTCAAAATGG AGTTATACCA CAGGAAACGG
 2301 CAAACGGTAC AGAACGCACTG CTAGAACTAC CTACACTTTC GCAAAGCCGA
 2351 TGGCTGCAAA CTATTGAGG AATCAGGCCA TGTACGTGTT TCGAAAAACG
 2401 GAACTTAAGC ACAGTAAGAC TGAACCTTAAT TTCAAGGAGT GGCAGAAGGC
 2451 GTTCACGGAT GTCATGGGTA TGGATGAAC GTATAAGGGG GGGTCTGGCA
 2501 CTGGGGGCAC TGCCAGCAGC GGATCCGGTG CGGGTGTGAG CAAGGGCGAG
 2551 GAGGATAACA TGCGCTCTCT CCCAGCGACA CATGAGTTAC ACATTTGG
 2601 CTCCATCAAC GGTGTGGACT TTGACATGGT GGGTCAGGGC ACCGGCAATC
 2651 CAAATGATGG TTATGAGGAG TAAACCTGA AGTCCACCAA GGGTGACCTC
 2701 CAGTTCTCCC CCTGGATTCT GGTCCCTCAT ATCGGGTATG GCTCCCATCA
 2751 GTACCTGCCA TACCTTGACG GGATGTCGCC TTTCCAGGCC GCCATGGTAG
 2801 ATGGCTCCGG ATACCAAGTC CATCGCACAA TGCAGTTGA AGATGGTGCC
 2851 TCCCTTACTG TTAACTACCG CTACACCTAC GAGGGAAAGCC ACATCAAAGG
 2901 AGAGGCCAG GTGAAGGGGA CTGGTTTCCC TGCTGACGGT CCTGTGATGA
 2951 CCAACTCGCT GACCGCTGCG GACTGGTCCA GGTCGAAGAA GACTTACCCC
 3001 AACGACAAAA CCATCATCAG TACCTTTAAG TGGAGTTACA CCACTGGAAA
 3051 TGGCAAGCGC TACCGGAGCA CTGCGCGGC CACCTACACC TTTGCCAAGC
 3101 CAATGGCGGC TAACTATCTG AAGAACCGAC CGATGTACGT GTTCCGTAAG
 3151 ACGGAGCTCA AGCACTCCAA GACCGAGCTC AACTTCAAGG AGTGGCAAAA
 3201 GGCCTTTACC GATGTGATGG GCATGGACGA GCTGTACAAG GGCGGAAGAT
 3251 CCGGTGGTGG TTCTGGTCCT CGTCCCCGTG GTACTCGTGG TAAAGGTGCG
 3301 CGTATTGTC GCGCGGGTTA ATCTAGAGGG CCCGTTAAA CCCGCTGATC
 3351 AGCCTCGACT GTGCCCTCTA GTTGCCAGCC ATCTGTTGTT TGCCCTCCC
 3401 CCGTGCCCTC CTTGACCCCTG GAAAGGTGCC ACTCCCACGT TCCCTTCCTA
 3451 ATAAAATGAG GAAATTGCAT CGCATTGTCT GAGTAGGTGT CATTCTATTG

-continued

3501 TGGGGGGTGG GGGTGGGGC AGGACAGCAA GGGGGAGGAT TGGGAAGACA
 3551 ATAGCAGGCA TGCTGGGAT GCGGTGGCT CTATGGCTTC TGAGGCAGAA
 3601 AGAACCAAGCT GGGGCTCTAG GGGGTATCCC CACCGCCCT GTAGCGGC
 3651 ATTAAGCGCG GCGGGTGTGG TGGTTACGCG CAGCGTGACC GCTACACTTG
 3701 CCAGCGCCCT AGCGCCCGCT CCTTTCGCTT TCTTCCCTTC CTTTCTCGCC
 3751 ACGTTCGCCG GCTTCCCGC TCAAGCTCTA AATGGGGC TCCCTTAGG
 3801 GTTCCGATTT AGTGCTTTAC GGCAACCTCGA CCCCCAAAAAA CTTGATTAGG
 3851 GTGATGGTTC ACCTAGTGGG CCATCGCCCT GATAGACGGT TTTTCGCC
 3901 TTGACGTTGG AGTCCACGTT CTTTAATAGT GGACTCTTGT TCCAAACTGG
 3951 AACAAACACTC AACCTATCT CGGTCTATTTC TTTTGATTAA TAAGGGATT
 4001 TGCCGATTTG GGCCTATTGG TAAAAAAATG AGCTGATTAA ACAAAAATTT
 4051 AACCGCAATT AATTCTGTGG AATGTGTGTC AGTTAGGGTG TGGAAAGTCC
 4101 CCAGGCTCCC CAGCAGGCAG AAGTATGCAA AGCATGCATC TCAATTAGTC
 4151 AGCAACCAGG TGTGGAAAGT CCCCAGGCTC CCCAGCAGGC AGAAGTATGC
 4201 AAAGCATGCA TCTCAATTAG TCAGCAACCA TAGTCCCGCC CCTAACTCCG
 4251 CCCATCCCGC CCCTAACTCC GCCCAGTTCC GCCCATTCTC CGCCCCATGG
 4301 CTGACTAATT TTTTTTATTG ATGCAGAGGC CGAGGCGGCC TCTGCCTCTG
 4351 AGCTATTCCA GAAGTAGTGA GGAGGCTTT TTGGAGGCCT AGGCTTTGC
 4401 AAAAAGCTCC CGGGAGCTTG TATATCCATT TTCGGATCTG ATCAAGAGAC
 4451 AGGATGAGGA TCGTTCGCA TGATTGAACA AGATGGATTG CACGCAGGTT
 4501 CTCCGGCCGC TTGGGTGGAG AGGCTATTCTG GCTATGACTG GGCACAACAG
 4551 ACAATCGGCT GCTCTGATGC CGCCGTGTTG CGGCTGTCAG CGCAGGGCG
 4601 CCCGGTTCTT TTTGTCAAGA CGCACCTGTC CGGTGCCCTG AATGAACTGC
 4651 AGGACGAGGC AGCCGGCTA TCGTGGCTGG CCACGACGGG CGTTCC
 4701 GCAGCTGTGC TCGACGTTGT CACTGAAGCG GGAAGGGACT GGCTGCTATT
 4751 GGGCGAAGTG CCGGGGCAGG ATCTCCTGTC ATCTCACCTT GCTCCTGCC
 4801 AGAAAGTATC CATCATGGCT GATGCAATGC GGCGGCTGCA TACGCTTGAT
 4851 CCGGCTACCT GCCCATTGCA CCACCAAGCG AAACATCGCA TCGAGCGAGC
 4901 ACGTACTCGG ATGGAAGCCG GTCTGTGCA TCAGGATGAT CTGGACGAAG
 4951 AGCATCAGGG GCTCGCGCCA GCCGAAGTGT TCGCCAGGCT CAAGGCGCG
 5001 ATGCCCGACG GCGAGGATCT CGTCGTGACC CATGGCGATG CCTGCTTGCC
 5051 GAATATCATG GTGGAAAATG GCCGCTTTTC TGGATTGATC GACTGTGGCC
 5101 GGCTGGGTGT GGCGGACCGC TATCAGGACA TAGCGTGGC TACCCGTGAT
 5151 ATTGCTGAAG AGCTTGGCGG CGAATGGGCT GACCGCTTCC TCGTGT
 5201 CGGTATCGCC GCTCCCGATT CGCAGCGCAT CGCCTTCTAT CGCCTTCTG
 5251 ACGAGTTCTT CTGAGCGGGC CTCTGGGTT CGAAATGACC GACCAAGCGA
 5301 CGCCCAACCT GCCATCACGA GATTCGATT CCACCGCCGC CTTCTATGAA
 5351 AGGTTGGCT TCGGAATCGT TTTCCGGAC GCCGGCTGGA TGATCCTCCA
 5401 GCGCGGGGAT CTCATGCTGG AGTTCTTCGCG CCACCCCAAC TTGTTTATTG
 5451 CAGCTTATAA TGTTTACAAA TAAAGCAATA GCATCACAAA TTTCACAAAT
 5501 AAAGCATTAA TTTCAGTGCAT TTCTAGTTGT GGTTTGTCCA AACTCATCAA

- continued

5551 TGTATCTTAT CATGCTGTA TACCGTCGAC CTCTAGCTAG AGCTTGGCGT
 5601 AACATGGTC ATAGCTGTT CCTGTGTGAA ATTGTTATCC GCTCACATT
 5651 CCACACAACA TACGAGCCGG AAGCATAAAG TGAAAGCCT GGGGTGCCTA
 5701 ATGAGTGAGC TAACTCACAT TAATTGCGTT GCGCTCACTG CCCGCTTCC
 5751 AGTCGGAAA CCTGTCGTGC CAGCTGCATT AATGAATCGG CCAACGGCG
 5801 GGGAGAGCG GTTGTGCTAT TGGGCGCTCT TCCGCTTCCT CGCTCACTGA
 5851 CTCGCTGC GC TCGTCGTTC GGCTGCGCG AGCGGTATCA GCTCACTCAA
 5901 AGGCGTAAT ACAGTTATCC ACAGAACATAG GGGATAACGC AGGAAAGAAC
 5951 ATGTGAGCAA AAGGCCAGCA AAAGGCCAGG AACCGTAAAA AGGCCGCGTT
 6001 GCTGGCGTTT TTCCATAGGC TCCGCCCGG TGACGAGCAT CACAAAATC
 6051 GACGCTCAAG TCAGAGGTGG CGAAACCGA CAGGACTATA AAGATACCAG
 6101 GCGTTTCCCC CTGGAAGCTC CCTCGTGC CG TCTCCTGTT CGACCCCTGCC
 6151 GCTTACCGGA TACCTGTCGG CCTTTCTCCC TTCGGGAAGC GTGGCGCTTT
 6201 CTCATAGCTC ACCTGTAGG TATCTCAGTT CGGTGTAGGT CGTTCGCTCC
 6251 AAGCTGGGCT GTGTGCACGA ACCCCCCGTT CAGCCCGACC GCTGCGCCTT
 6301 ATCCGGTAAC TATCGTCTTG AGTCCAACCC GGTAAGACAC GACTTATCGC
 6351 CACTGGCAGC AGCCACTGGT AACAGGATTA GCAGAGCGAG GTATGTAGGC
 6401 GGTGCTACAG AGTCTTGAA GTGGTGGCCT AACTACGGCT AACTAGAAC
 6451 AACAGTATTG GGTATCTGCG CTCTGCTGAA GCCAGTTACC TTGGAAAAAA
 6501 GAGTTGGTAG CTCTTGATCC GGAAACAAAA CCACCGCTGG TAGCGTTTT
 6551 TTTGTTGCA AGCAGCAGAT TACGCGCAGA AAAAAGGAT CTCAGAAGA
 6601 TCCTTGATC TTTTCTACGG GGTCTGACGC TCAGTGGAAC GAAAACCTCAC
 6651 GTTAAGGGAT TTTGGTCATG AGATTATCAA AAAGGATCTT CACCTAGATC
 6701 CTTTTAAATT AAAATGAAG TTTAAATCA ATCTAAAGTA TATATGAGTA
 6751 AACCTGGTCT GACAGTTACC AATGCTTAAT CAGTGAGGCA CCTATCTCAG
 6801 CGATCTGTCT ATTTCGTTCA TCCATAGTTG CCTGACTCCC CGTCTGTAG
 6851 ATAACACGA TACGGGAGGG CTTACCATCT GGCCCCAGTG CTGCAATGAT
 6901 ACCGGAGAC CCACGCTCAC CGGCTCCAGA TTTATCAGCA ATAAACCAGC
 6951 CAGCCGGAAG GGCGAGCGC AGAAGTGGTC CTGCAACTTT ATCCGCTCC
 7001 ATCCAGTCTA TTAATTGTTG CCGGGAAGCT AGAGTAAGTA GTTCGCCAGT
 7051 TAATAGTTG CGAACAGTTG TTGCCATTGC TACAGGCATC GTGGTGTAC
 7101 GCTCGTCGTT TGGTATGGCT TCATTCTAGCT CCGGTTCCCA ACGATCAAGG
 7151 CGAGTTACAT GATCCCCCAT GTTGTGCAA AAAGCGGTTA GCTCCTTCGG
 7201 TCCCTCGATC GTTGTGAGAA GTAAGTTGGC CGCAGTGTAA TCACTCATGG
 7251 TTATGGCAGC ACTGCATAAT TCTCTTACTG TCATGCCATC CGTAAGATGC
 7301 TTTCTGTGA CTGGTGAGTA CTCACACCAAG TCATTCTGAG AATAGTGTAT
 7351 GCGCGACCG AGTTGCTCTT GCCCGGCGTC AATACGGGAT AATACCGCGC
 7401 CACATAGCAG AACTTTAAA GTGCTCATCA TTGGAAAACG TTCTTCGGGG
 7451 CGAAAACCTCT CAAGGATCTT ACCGCTGTTG AGATCCAGTT CGATGTAACC
 7501 CACTCGTGCA CCCAACTGAT CTTCAAGCATC TTTTACTTTC ACCAGCGTTT

- continued

```

7551 CTGGGTGAGC AAAAACAGGA AGGCAAAATG CCGCAAAAAA GGGATAAGG
7601 GCGACACGGA AATGTTGAAT ACTCATACTC TTCCCTTTTC AATATTATTG
7651 AAGCATTAT CAGGGTTATT GTCTCATGAG CGGATACATA TTTGAATGTA
7701 TTTAGAAAAA TAAACAAATA GGGGTTCCGC GCACATTCC CCGAAAAGTG
7751 CCACCTGACG TC

```

10

According to one embodiment, the vector is a plasmid and has the sequence of pCMV-CytERM-mCherry-(F30-2×Pep-per)₁₀ (SEQ ID NO: 76; GenBank Accession No. MN052906.1, which is hereby incorporated by reference) as follows:

```

1 GACGGATCGG GAGATCTCCC GATCCCCTAT GGTGCACTCT CAGTACAATC
51 TGCTCTGATG CCCCATAGTT AAGCCAGTAT CTGCTCCCTG CTTGTGTGTT
101 GGAGGTCGCT GAGTAGTGC CGAGCAAAT TTAAGCTACA ACAAGGCAAG
151 GCTTGACCGA CAATTGCATG AAGAATCTGC TTAGGGTTAG GCGTTTGCG
201 CTGCTTCGCG ATGTACGGGC CAGATATACTG CGTTGACATT GATTATTGAC
251 TAGTTATTAA TAGTAATCAA TTACGGGTC ATTAGTTCAT AGCCCATATA
301 TGGAGTCCG CGTTACATAA CTTACGGTAATGGCCGCC TGGCTGACCG
351 CCCAACGACC CCCGCCATT GACGTCAATA ATGACGTATG TTCCCATAGT
401 AACGCCAATA GGGACTTTCC ATTGACGTCA ATGGGTGGAG TATTTACGGT
451 AAACCTGCCA CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTACGCC
501 CCTATTGACG TCAATGACGG TAAATGGCC GCCTGGCATT ATGCCAGTA
551 CATGACCTTA TGGGACTTTTC CTACTTGCCA CTACATCTAC GTATTAGTCA
601 TCGCTATTAC CATGGTGATG CGGTTTGGC AGTACATCAA TGGCGTGGAA
651 TAGCGTTTG ACTCACGGGG ATTTCCAAGT CTCCACCCCA TTGACGTCAA
701 TGGGAGTTTG TTTGGCACC AAAATCAACG GGACTTCCA AAATGTCGA
751 ACAACTCCGC CCCCATTGAC GCAAATGGC GGTAGGCGTG TACGGTGGGA
801 GGGTCTATATA AGCAGAGCTC TCTGGCTAAC TAGAGAACCC ACTGTTACT
851 GGCTTATCGA ATTAATACG ACTCACTATA GGGAGACCC AGCTGGCTAG
901 CGTTTAAACT TGCCACCATG GACCCTGTGG TGGTGCTGGG GCTCTGTCTC
951 TCCCTGTTGC TTCTCCTTTC ACTCTGGAAA CAGAGCTATG GGGGAGGGAA
1001 ACTGGGCGGA AGCGGAGGGGA CGGGGGGTTA AGGAACCTCA GGGGGTGTGA
1051 GCAAGGGCGA GGAGGATAAC ATGGCCATCA TCAAGGAGTT CATGCGCTTC
1101 AAGGTGCACA TGGAGGGCTC CGTGAACGGC CACGAGTTCG AGATCGAGGG
1151 CGAGGGCGAG GGCGCCCCCT ACAGAGGGCAC CCAGACCGCC AAGCTGAAGG
1201 TGACCAAGGG TGGCCCCCTG CCCTTCGCTT GGGACATCCT GTCCCTCAG
1251 TTCATGTACG GCTCCAAGGC CTACGTGAAG CACCCCGCCG ACATCCCCGA
1301 CTACTTGAAG CTGTCTTCC CCGAGGGCTT CAAGTGGGAG CGCGTGATGA
1351 ACTTCGAGGA CGCGGGCGTG GTGACCGTGA CCCAGGACTC CTCCCTGCAG
1401 GACGGCGAGT TCATCTACAA GGTGAAGCTG CGCGGCACCA ACTTCCCCTC
1451 CGACGGCCCG GTAATGCAGA AGAAGACCAT GGGCTGGGAG GCCTCCTCCG
1501 AGCGGATGTA CCCCAGGAC GGCGCCCTGA AGGGCGAGAT CAAGCAGAGG

```

-continued

1551 CTGAAGCTGA AGGACGGCGG CCACTACGAC GCTGAGGTCA AGACCACCTA
 1601 CAAGGCCAAG AAGCCCGTGC AGCTGCCCGG CGCCTACAAC GTCAACATCA
 1651 AGTTGGACAT CACCTCCCCAC AACGAGGACT ACACCATCGT GGAACAGTAC
 1701 GAACCGGCCG AGGGCCGCCA CTCCACCGGC GGCATGGACG AGCTGTACAA
 1751 GTAACTCGAG ATCCGTTACG GCCGGAATCA ATCGCTAATC ACTCAACTTG
 1801 CCATGTGTAT GTGGGAAGCG TAGAAAGGCT CGTTGAGCTC ATTAGCTCG
 1851 AGCCCGACTA CGTTCCAC ATACTCTGAT GATCCGCTAG CAAAGGCTCG
 1901 TCTGAGCTCA TTAGCTCCGA GCCCGAGGT CGGGATCATT CATGGCAAGT
 1951 CCAGCGCAAT CTATTACGAA AATCATCCGA CGTCGCGATG TCTATGCGGG
 2001 AAGCGTAGAA AGGCTCGTCT GAGCTCATTA GCTCCGAGCC CGACTACGTT
 2051 TCCCGCATAG TCTGATCATC CGCTAGCAA GGCTCGTTGA GCTCATTAGC
 2101 TCCGAGCCCG AGGTACCGGA TGATTCATCG CGACGCTGCG GAAAATCTCA
 2151 CAAAATCACG TCAAACGTG CGGTGTGTGT GTAGGAAGCG TAGAAAGGCT
 2201 CGTCTGAGCT CATTAGCTCC GAGCCCGACT ACGTTTCTA CACACTCTGA
 2251 CGATCCGCTA GCAAAGGCTC GTTGAGCTCA TTAGCTCCGA GCCCGAGGT
 2301 CCGGATCGTT CACGGCGACG CCGATAATCC ACATACTTAC AATCAGGC
 2351 TCTTGCCATG TGTATGTGGG AAGCGTAGAA AGGCTCGTTG AGCTCATTAG
 2401 CTCCGAGCCC GACTACGTTT CCCACATACT CTGATGATCC GCTAGCAAAG
 2451 GCTCGTTGAG CTCATTAGCT CCGAGCCCGA GGTACCGGAT CATTGATGGC
 2501 AAGTATCAAG ATCGAACGGC GCAAGATATT GTCACGTCGC GATGCTATG
 2551 CGGGAAAGCGT AGAAAGGCTC GTTGAGCTCA TTAGCTCCGA GCCCGACTAC
 2601 GTTCCCGCA TAGTCTGATC ATCCGCTAGC AAAGGCTCGT CTGAGCTCAT
 2651 TAGCTCCGAG CCCGAGGTAC CGGATGATTC ATCGCGACGT CCTCGCTAGA
 2701 TATGTTAGGT TCTTAGGCAT TTCGGCGTGT GTGTGTAGGA AGCGTAGAAA
 2751 GGCTCGTTGA GCTCATTAGC TCCGAGCCCG ACTACGTTTC CTACACACTC
 2801 TGACGATCCG CTAGCAAAGG CTCGTCTGAG CTCATTAGCT CCGAGCCCGA
 2851 GGTACCGGAT CGTTCACGGC GAAAAGATCG TCTGCAATTG CGATTAGACG
 2901 TACACTTGCC ATGTGTATGT GGGAAAGCGTA GAAAGGCTCG TCTGAGCTCA
 2951 TTAGCTCCGA GCCCGACTAC GTTCCCAACA TACTCTGATG ATCCGCTAGC
 3001 AAAGGCTCGT TGAGCTCATT AGCTCCGAGC CCGAGGTACC GGATCATTCA
 3051 TGGCAAGATC CAAGCTACTT CCTCCATACC TATCCTCCTC GCGATGTCTA
 3101 TCGGGGAAGC GTAGAAAGGC TCGTCTGAGC TCATTAGCTC CGAGCCCGAC
 3151 TACGTTCCC GCATAGTCTG ATCATCCGCT AGCAAAGGCT CGTTGAGCTC
 3201 ATTAGCTCCG AGCCCGAGGT ACCGGATGAT TCATCGCGAG ATCATAACGC
 3251 AATACCGTAC ACTGTCCAAT CCTCGCCGTG TGTGTGTAGG AAGCGTAGAA
 3301 AGGCTCGTCT GAGCTCATTA GCTCCGAGCC CGACTACGTT TCCTACACAC
 3351 TCTGACGATC CGCTAGCAA GGCTCGTTGA GCTCATTAGC TCCGAGCCCG
 3401 AGGTACCGGA TCGTTCACGG CGAGGATAAT CAATCCACAT ACATCACACC
 3451 ACAATTCTTG CCATGTGTAT GTGGGAAGCG TAGAAAGGCT CGTCTGAGCT
 3501 CATTAGCTCC GAGCCCGACT ACGTTTCCCA CATACTCTGA TGATCCGCTA
 3551 GCAAAGGCTC GTCTGAGCTC ATTAGCTCCG AGCCCGAGGT ACCGGATCAT

- continued

3601 TCATGGCAAG AATTGGTCGT TCTTCTTGGC GGCGGCTCGA CTAAATCACC
 3651 GGTAATCTTC TTGTCATCT AGACCTTATA AAGATTTTG TACAAGGGCC
 3701 CGTTTAAACC CGCTGATCAG CCTCGACTGT GCCTTCTAGT TGCCAGCCAT
 3751 CTGTTGTTG CCCCTCCCC GTGCCTTCT TGACCCCTGGA AAGGTGCCAC
 3801 TCCCCTGTG CTTTCTAAT AAAATGAGGA AATTGCATCG CATTGTCTGA
 3851 GTAGGTGTCA TTCTATTCTG GGGGGTGGGG GTGGGGGCAG GACAGCAAGG
 3901 GGGAGGATTG GGAAGACAAT AGCAGGCATG CTGGGGATGC GGTGGCTCT
 3951 ATGGCTCTG AGCGGAAAG ACCAGCTGG GGCTCTAGGG GGTATCCCCA
 4001 CGCGCCCTGT AGCGCGCAT TAAGCGCGC GGGTGTGGTG GTTACCGC
 4051 GCGTGACCGC TACACTTGCC AGCGCCCTAG CGCCCGCTCC TTTCGCTTTC
 4101 TTCCCTTCCT TTCTGCCAC GTTCGCCGGC TTTCCCGTC AAGCTCTAAA
 4151 TCGGGGCTC CCTTTAGGGT TCCGATTTAG TGCTTTACGG CACCTCGACC
 4201 CCAAAAAACT TGATTAGGGT GATGGTTCAC GTAGTGGGCC ATCGCCCTGA
 4251 TAGACGGTTT TTCGCCCTT GACGTTGGAG TCCACGTTCT TTAATAGTGG
 4301 ACTCTTGTTC CAAACTGGAA CAACACTCAA CCCTATCTCG GTCTATTCTT
 4351 TTGATTATA AGGGATTTTG CCGATTTCGG CCTATTGGTT AAAAATGAG
 4401 CTGATTTAAC AAAAATTTAA CGCGAATTAA TTCTGTGGAA TGTGTGTCAG
 4451 TTAGGGTGTG GAAAGTCCCC AGGCTCCCCA GCAGGCAGAA GTATGCAAAG
 4501 CATGCATCTC ATTAGTCAG CAACCAGGTG TGGAAAGTCC CCAGGCTCCC
 4551 CAGCAGGCAG AAGTATGCAA AGCATGCATC TCAATTAGTC AGCAACCATA
 4601 GTCCCGCCCC TAATCCGCC CATCCCGCC CTAACTCCGC CCAGTTCCGC
 4651 CCATTCTCCG CCCCATGGCT GACTAATTTT TTTTATTAT GCAGAGGCCG
 4701 AGGCCGCCTC TGCCCTGAG CTATTCCAGA AGTAGTGAGG AGGTTTTT
 4751 GGAGGCCTAG GCTTTGCAA AAAGCTCCC GGAGCTTGTA TATCCATT
 4801 CGGATCTGAT CAAGAGACAG GATGAGGATC GTTTCGCATG ATTGAACAAG
 4851 ATGGATTGCA CGCAGGTTCT CGGGCCGCTT GGGTGGAGAG GCTATTCCGC
 4901 TATGACTGGG CACACAGAC AATCGGCTGC TCTGATGCCG CCGTGTCCG
 4951 GCTGTCAGCG CAGGGCGGCC CGGTTCTTT TGTCAAGACC GACCTGTCCG
 5001 GTGCCCTGAA TGAACCTGCAG GACGAGGAG CGCGGCTATC GTGGCTGGCC
 5051 ACGACGGCGC TTCCCTGCCAG AGCTGTGCTC GACGTTGTCA CTGAAGCGGG
 5101 AAGGGACTGG CTGCTATTGG GCGAAGTGC CGGGCAGGAT CTCCGTGTCAT
 5151 CTCACCTTGC TCCGCCAG AAAGTATCCA TCATGGCTGA TGCAATGCCG
 5201 CGGCTGCATA CGCTTGATCC GGCTACCTGC CCATTCGACC ACCAAGCGAA
 5251 ACATCGCATC GAGCGAGCAC GTACTCGGAT GGAAGCCGGT CTTGTCGATC
 5301 AGGATGATCT GGACGAAGAG CATCAGGGC TCGCGCCAGC CGAAACTGTT
 5351 GCCAGGCTCA AGGCGCGCAT GCCCGACGGC GAGGATCTCG TCGTGACCCA
 5401 TGGCGATGCC TGCTTGCCGA ATATCATGGT GGAAAATGGC CGCTTTCTG
 5451 GATTCACTGA CTGTGGCCGG CTGGGTGTGG CGGACCGCTA TCAGGACATA
 5501 GCGTTGGCTA CCCGTGATAT TGCTGAAGAG CTTGGCGGCG AATGGGCTGA
 5551 CCGCTTCCTC GTGCTTTACG GTATCGCCGC TCCCGATTG CAGCGCATCG

-continued

5601 CCTTCTATCG CCTTCTTGAC GAGTTCTCT GAGCGGGACT CTGGGGTTCG
 5651 AAATGACCGA CCAAGCGACG CCCAACCTGC CATCACGAGA TTTCGATTCC
 5701 ACCGCCGCCT TCTATGAAAG GTTGGGCTTC GGAATCGTTT TCCGGGACGC
 5751 CGGCTGGATG ATCCTCCAGC GCGGGGATCT CATGCTGGAG TTCTCGCCC
 5801 ACCCCAACCTT GTTTATTGCA GCTTATAATG GTTACAAATA AAGCAATAGC
 5851 ATCACAAATT TCACAAATAA AGCATTTTT TCACTCATT CTAGTTGTGG
 5901 TTTGTCCAAA CTCATCAATG TATCTTATCA TGTCTGTATA CCGTCGACCT
 5951 CTAGCTAGAG CTTGGCGTAA TCATGGTCAT AGCTGTTCC TGTGTGAAAT
 6001 TGTATCCGC TCACAATTCC ACACAACATA CGAGCCGAA GCATAAAGTG
 6051 TAAAGCCTGG GGTGCCTAAT GAGTGAGCTA ACTCACATTA ATTGCCTTGC
 6101 GCTCACTGCC CGCTTCCAG TCGGGAAACC TGTCGTGCCA GCTGCATTAA
 6151 TGAATCGGCC AACCGCGGGG GAGAGGCGGT TTGCGTATTG GGCGCTCTTC
 6201 CGCTTCCTCG CTCACTGACT CGCTGCGCTC GGTCGTTCGG CTGCGCGAG
 6251 CGGTATCAGC TCACCAAAG GCGGTAATAC GGTTATCCAC AGAACATCAGGG
 6301 GATAACGCAG GAAAGAACAT GTGAGCAAAA GGCCAGCAAA AGGCCAGGAA
 6351 CCGTAAAAAG GCCCGTTCGC TGGCGTTTT CCATAGGCTC CGCCCCCTG
 6401 ACGAGCATCA CAAAATCGA CGCTCAAGTC AGAGGTGGCG AAACCCGACA
 6451 GGACTATAAA GATACCAGGC GTTTCCCCCT GGAAGCTCCC TCGTGCCTC
 6501 TCCCTGTCG ACCCTGCCGC TTACCGGATA CCTGTCGCC TTTCTCCCTT
 6551 CGGGAAGCGT GGCGCTTCT CATAGCTCAC GCTGTAGGTA TCTCAGTTCG
 6601 GTGTAGGTCG TTCGCTCAA GCTGGGCTGT GTGCACGAAC CCCCCGTTCA
 6651 GCCCGACCGC TCGCCTTAT CCGGTAACTA TCGTCTTGAG TCCAACCCGG
 6701 TAAGACACGA CTTATCGCCA CTGGCAGCAG CCACCTGGTAA CAGGATTAGC
 6751 AGAGCGAGGT ATGTAGGCCG TGCTACAGAG TTCTTGAAGT GGTGGCTAA
 6801 CTACGGCTAC ACTAGAAGAA CAGTATTGG TATCTGCGCT CTGCTGAAGC
 6851 CAGTTACCTT CGGAAAAAGA GTTGGTAGCT CTTGATCCGG CAAACAAACC
 6901 ACCGCTGGTA CGGGTTTTTG TGTTGCAAG CAGCAGATT CGCGCAGAAA
 6951 AAAAGGATCT CAAGAAGATC CTTGATCTT TTCTACGGGG TCTGACGCTC
 7001 AGTGGAACGA AAACTCACGT TAAGGGATT TGGTCATGAG ATTATCAAA
 7051 AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT
 7101 CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA TGCTTAATCA
 7151 GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTGCTTCATC CATACTTGCC
 7201 TGACTCCCCG TCGTGTAGAT AACTACGATA CGGGAGGGCT TACCATCTGG
 7251 CCCCAGTGCT GCAATGATAC CGCGAGACCC ACGCTCACCG GCTCCAGATT
 7301 TATCAGCAAT AAACCAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT
 7351 GCAACTTTAT CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGGAAAGCTAG
 7401 AGTAAGTAGT TCGCCAGTTA ATAGTTGCG CAACGTTGTT GCCATTGCTA
 7451 CAGGCATCGT GGTGTCACGC TCGTCGTTG GTATGGCTTC ATTCACTC
 7501 GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCCATGT TGTGCAAAAA
 7551 AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG
 7601 CAGTGTATC ACTCATGGTT ATGGCAGCAC TGCATAATT CTCCTACTGTC

US 12,391,948 B2

65**66**

- continued

```

7651 ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC
7701 ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC CGGGCGTCAA
7751 TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT
7801 GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC CGCTGTTGAG
7851 ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGGCATCTT
7901 TTACTTTCAC CAGCGTTCTC GGGTGAGCAA AAACAGGAAG GCAAAATGCC
7951 GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT
8001 CCTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG
8051 GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC
8101 ACATTTCCCC GAAAAGTGCC ACCTGACGTC

```

According to one embodiment, the vector is a plasmid and has the sequence of pUbC-(mNeonGreen)₄-tDeg (SEQ ID

NO: 77; GenBank Accession No. MN052907.1, which is hereby incorporated by reference) as follows:

```

1 GACGGATCGG GAGATCTCCC GATCCCTAT GGTGCACTCT CAGTACAATC
51 TGCTCTGATG CCGCATAGTT AAGCCAGTAT CTGCTCCCTG CTTGTGTGTT
101 GGAGGTCGCT GAGTAGTGCG CGAGCAAAAT TTAAGCTACA ACAAGGCAAG
151 GCTTGACCGA CAATTGCGATG AAGAATCTGC TTAGGGTTAG GCGTTTGCG
201 CTGCTTCGCG ATGTACGGGC CAGATATACG CGTTGGCCTC CGCGCCGGT
251 TTTGGCGCCT CCCGCGGGCG CCCCCCTCCT CACGGCGAGC GCTGCCACGT
301 CAGACGAAGG CGCGCAGCGAG CGTCCTGATC CTTCCGCCCG GACGCTCAGG
351 ACAGCGGCC GCTGCTCATA AGACTCGGCC TTAGAACCCC AGTATCAGCA
401 GAAGGACATT TTAGGACGGG ACTTGGGTGA CTCTAGGGCA CTGGTTTTCT
451 TTCCAGAGAG CGGAACAGGC GAGGAAAAGT AGTCCCTCT CGCGGATTCT
501 GCGGAGGGAT CTCCGTGGG CGTGAACAGC CGATGATTAT ATAAGGACGC
551 GCCGGGTGTG GCACAGCTAG TTCCGTCGCA GCCGGGATT GGGTCGCGGT
601 TCTTGTGTTGT GGATCGCTGT GATCGTCACT TGGAAGCTTG CCACCATGGT
651 GAGCAAGGGC GAGGAGGATA ACATGGCCTC TCTCCCAGCG ACACATGAGT
701 TACACATCTT TGGCTCCATC AACGGTGTGG ACTTGACAT GGTGGGTCAG
751 GGCACCGCA ATCCAAATGA TGGTTATGAG GAGTTAAACC TGAAGTCCAC
801 CAAGGGTGAC CTCCAGTTCT CCCCCGGAT TCTGGTCCCT CATATCGGGT
851 ATGGCTTCCA TCAGTACCTG CCCTACCGCT ACGGGATGTC GCCTTTCCAG
901 GCCCACATCAA AGGAGAGGCC CAGGTGAAGG GGACTGGTT CCCTGCTGAC
951 TGAAGATGGT GCCTCCCTTA CTGTTAACTA CCGCTACACC TACGAGGGAA
1001 GCCCACATCAA AGGAGAGGCC CAGGTGAAGG GGACTGGTT CCCTGCTGAC
1051 GGTCCCTGTGA TGACCAAATC GCTGACCGCT GCGGACTGGT GCAGGTCGAA
1101 GAAGACTTAC CCCAACGACA AAACCATCAT CAGTACCTT AAGTGGAGTT
1151 ACACCACTGG AAATGGCAAG CGCTACCGGA GCACTGGCG GACCACCTAC
1201 ACCTTTGCCA AGCCAATGGC GGCTAACTAT CTGAAGAACC AGCCGATGTA
1251 CGTGTCCGT AAGACGGAGC TCAAGCACTC CAAGACCGAG CTCAACTTCA
1301 AGGAGTGGCA AAAGGCCTTT ACCGATGTGA TGGGCATGGA CGAGCTGTAC

```

-continued

1351 AAGGGTGGAC ATATGGGCAC AGGTCACACA GGCAGTACCG GCGGAGTTTC
 1401 CAAAGGAGAA GAAGACAATA TGGCATCACT CCCCACCAACC CACGAGTTGC
 1451 ATATTTTCGG TTCAATTAAT GGACTAGATT TCGATATGGT TGGCCAGGGAA
 1501 ACAGGAAACCC CAAACGACGG ATATGAAGAG CTTAATCTCA AAAGTACCAA
 1551 AGGCGATCTG CAATTTCTC CGTGGATACT CGTGCCACAC ATTGGATACG
 1601 GATTCACCA ATATCTCCG TATCCGGATG GAATGTCCCC CTTCAAGCA
 1651 GCAATGGTGG ACGGGAGTGG TTATCAGGTA CACAGAACCA TGCAAGTCGA
 1701 GGACGGGGCT TCTCTGACCG TAAATTATAG GTATACTTAT GAAGGCTCAC
 1751 ATATTAAGGG CGAACGACAG GTTAAAGGAA CCGGGTTCC TGCGGATGGC
 1801 CCCGTATGA CTAATTCTCT GACAGCCGCA GATTGGTGTC GCTCCAAAAA
 1851 GACATACCCG AATGATAAGA CTATAATCTC AACATTCAAA TGGTCCATA
 1901 CGACAGGCAA CGGGAAACGA TATAGATCCA CGGCTCGAAC AACATTACACA
 1951 TTCGCTAACAC CTATGCCGC CAATTACCTC AAAATCAGC CCATGTATGT
 2001 GTTTAGAAA ACCGAATTGA AGCATTCTAA AACGGAACCTT AATTTAAGG
 2051 AATGGCAGAA GGCTTCACA GACGTAATGG GGATGGATGA ACTCTATAAA
 2101 TCAGGTCTCG AGTCCTCAGG GGGAACGGGT GGGTCCGGAG GAGTTAGTAA
 2151 AGGTGAAGAG GACAATATGG CAAGTTGCC TGCGACTCAC GAGCTTCATA
 2201 TCTTTGGGTC TATAATGGC GTTGACTTCG ATATGGTTGG CCAAGGTACT
 2251 GGCAACCCCA ATGACGGTTA CGAGGAGTTG AATCTCAAGT CCACAAAAGG
 2301 TGATCTTCAG TTCAGCCCTT GGATTCTCGT ACCTCATATT GGATATGGCT
 2351 TTCACCGATA CCTTCCATAC CCAGACGGTA TGTACCCCTT TCAAGCTGCG
 2401 ATGGTGGATG GTTCCGGCTA TCAGGTCCAC CGAACGATGC AATTGAGGA
 2451 CGGGGCCAGC CTCACCGTTA ATTATAGGTA CACCTATGAG GGAAGTCACA
 2501 TAAAGGAGA AGCCCAGTG AAAGGAACAG GATTCAGGAG TGATGGTCCA
 2551 GTAATGACGA ACTCCTGAC AGCGGCTGAC TGGTAGAA GCAAAAGAC
 2601 GTATCCTAAT GACAAGACCA TCATTAGCAC TTTCAAATGG AGTTATACCA
 2651 CAGGAAACGG CAAACGGTAC AGAACGACTG CTAGAACTAC CTACACTTTC
 2701 GCAAAGCCGA TGGCTGAAA CTATTGAAG AATCAGCCCA TGACGTGTT
 2751 TCGAAAAACG GAACTTAAGC ACAGTAAGAC TGAACCTAAT TTCAAGGAGT
 2801 GGCAGAAGGC GTTCACGGAT GTCATGGTA TGGATGAACT GTATAAGGGAA
 2851 GGGTCTGGCA CTGGGGGCAC TGCCAGCAGC GGATCCGGTG GCGGTGTGAG
 2901 CAAGGGCGAG GAGGATAACA TGGCTCTCT CCCAGCGACA CATGAGTTAC
 2951 ACATCTTGG CTCCATCAAC GGTGTGGACT TTGACATGGT GGGTCAGGGC
 3001 ACCGGCAATC CAAATGATGG TTATGAGGAG TTAAACCTGA AGTCCACCAA
 3051 GGGTGACCTC CAGTTCTCCC CCTGGATTCT GGTCCCTCAT ATCGGGTATG
 3101 GCTTCCATCA GTACCTGCC TACCGTGACG GGATGTCGCC TTTCCAGGCC
 3151 GCCATGGTAG ATGGCTCCGG ATACCAAGTC CATCGCACAA TGCAAGTTGA
 3201 AGATGGTGCC TCCCTTACTG TTAACCTACCG CTACACCTAC GAGGGAAAGCC
 3251 ACATCAAAGG AGAGGCCAG GTGAAGGGGA CTGGTTCCC TGCTGACGGT
 3301 CCTGTGATGA CCAACTCGCT GACCGCTGCG GACTGGTGCA GGTGAAAGAA
 3351 GACTTACCCC AACGACAAAAA CCATCATCAG TACCTTAAG TGGAGTTACA

-continued

3401 CCACTGGAAA TGGCAAGCGC TACCGGAGCA CTGCGCGGAC CACCTACACC
 3451 TTTGCCAACG CAATGGCGGC TAACTATCTG AAGAACCGAGC CGATGTACGT
 3501 GTTCCGTAAG ACGGAGCTCA AGCACTCCAA GACCGAGCTC AACTTCAAGG
 3551 AGTGGCAAAA GGCCTTACCG ATGTGATGG GCATGGACGA GCTGTACAAG
 3601 GGCGGAAGAT CCGGTTGGG TTCTGGTCCT CGTCCCCGTG GTACTCGTGG
 3651 TAAAGGTGCG CCGTATTGTC GCCGCGGTTA ATCTAGAGGG CCCGTTAAA
 3701 CCCGCTGATC AGCCTCGACT GTGCCCTCTA GTGCCAGCC ATCTGTTGTT
 3751 TGCCCCCTCCC CCGTGCCCTTC CTTGACCCCTG GAAAGGTGCC ACTCCCAC TG
 3801 TCCTTTCTTA ATAATGAG GAAATTGCAT CGCATTGTCT GAGTAGGTGT
 3851 CATTCTATTG TGGGGGGTGG GGGTGGGGC AGGACAGCAA GGGGGAGGAT
 3901 TGGGAAGACA ATAGCAGGCA TGCTGGGAT GCGGTGGCT CTATGGCTTC
 3951 TGAGGCGGAA AGAACACAGCT GGGGCTCTAG GGGGTATCCC CACCGCCCT
 4001 GTAGCGCGC ATTAAGCGCG GCGGGTGTGG TGGTACGCG CAGCGTGACC
 4051 GCTACACTTG CCAGCGCCCT AGCGCCCGCT CCTTCGCTT TCTCCCTTC
 4101 CTTTCTGCC ACGTTGCCG GCTTCCCCG TCAAGCTCTA AATCGGGGGC
 4151 TCCCTTTAGG GTTCCGATTT AGTGTGTTAC GGACACCTCGA CCCCCAAAAAA
 4201 CTTGATTAGG GTGATGGTTC ACGTAGTGGG CCATCGCCCT GATAGACGGT
 4251 TTTTCGCCCT TTGACGTTGG AGTCCACGTT CTTTAATAGT GGACTCTTGT
 4301 TCCAAACTGG AACAAACACTC AACCTATCT CGGTCTATTG TTTGATTAA
 4351 TAAGGGATT TGCCGATTTC GGCCTATTGG TTAAAAAAATG AGCTGATTAA
 4401 ACAAAAATTT AACCGGAATT AATTCTGTGG AATGTGTGTC AGTTAGGGTG
 4451 TGGAAAGTC CCAGGCTCCC CAGCAGGAG AAGTATGCAA AGCATGCATC
 4501 TCAATTAGTC AGCAACCAGG TGTGGAAAGT CCCCAGGCTC CCCAGCAGGC
 4551 AGAAGTATGC AAAGCATGCA TCTCAATTAG TCAGCAACCA TAGTCCCGCC
 4601 CCTAACTCCG CCCATCCCGC CCCTAACTCC GCCCAGTTCC GCCCATTCTC
 4651 CGCCCCATGG CTGACTAATT TTTTTTATTT ATGCAGAGGG CGAGGCGGCC
 4701 TCTGCCTCTG AGCTATTCCA GAAGTAGTGA GGAGGCTTT TTGGAGGCCT
 4751 AGGCTTTGC AAAAGCTCC CGGGAGCTTG TATATCCATT TTGGATCTG
 4801 ATCAAGAGAC AGGATGAGGA TCGTTCGCA TGATTGAACA AGATGGATTG
 4851 CACGCAGGTT CTCCGGCCGC TTGGGTGGAG AGGCTATTG GCTATGACTG
 4901 GGCACAAACAG ACAATCGGCT GCTCTGATGC CGCCGTGTT CGGCTGTCA
 4951 CGCAGGGCG CCGGGTCTT TTTGTCAGA CCGACCTGTC CGGTGCCCTG
 5001 AATGAACTGC AGGACGAGGC AGCGCGGCTA TCGTGGCTGG CCACGACGGG
 5051 CGTCCCTGTC GCAGCTGTGC TCGACGTTGT CACTGAAGCG GGAAGGGACT
 5101 GGCTGCTATT GGGCGAAGTG CGGGGCAGG ATCTCCTGTC ATCTCACCTT
 5151 GCTCCTGCCG AGAAAGTATC CATCATGGCT GATGCAATGC GGCGCTGCA
 5201 TACGCTTGAT CCGGCTACCT GCCCATCGA CCACCAAGCG AAACATCGCA
 5251 TCGAGCGAGC ACGTACTCGG ATGGAAGCCG GTCTTGTGCA TCAGGATGAT
 5301 CTGGACGAAG AGCATCAGGG GCTCGCGCA GCGAACTGT TCGCCAGGCT
 5351 CAAGGCGCGC ATGCCCGACG GCGAGGATCT CGTCGTGACC CATGGCGATG

-continued

5401 CCTGCTTGC CAAATATCATG GTGGAAAATG GCCGCTTTTC TGGATTCACT
 5451 GACTGTGGCC GGCTGGGTGT GGCGGACCGC TATCAGGACA TAGCGTTGGC
 5501 TACCCGTGAT ATTGCTGAAG AGCTTGGCGG CGAATGGGCT GACCGCTTCC
 5551 TCGTGCTTTA CGGTATCGCC GCTCCCGATT CGCAGCGCAT CGCCTTCTAT
 5601 CGCCTTCTTG ACGAGTTCTT CTGAGCGGGA CTCTGGGTT CGAAATGACC
 5651 GACCAAGCGA CGCCCAACCT GCCATCACGA GATTTCGATT CCACCGCCGC
 5701 CTTCTATGAA AGGTTGGCT TCGGAATCGT TTTCCGGAC GCCGGCTGGA
 5751 TGATCCTCCA CGCGGGGAT CTCATGCTGG AGTTCTTCGC CCACCCCCAAC
 5801 TTGTTTATTG CAGCTTATAA TGGTTACAAA TAAAGCAATA GCATCACAAA
 5851 TTCACAAAT AAAGCATTTC TTTCACTGCA TTCTAGTTGT GGTTTGTCCA
 5901 AACTCATCAA TGTATCTTAT CATGCTGTAA TACCGTCGAC CTCTAGCTAG
 5951 AGCTTGGCGT AATCATGGTC ATAGCTGTTT CCTGTGTGAA ATTGTTATCC
 6001 GCTCACAATT CCACACAAACA TACGAGCCGG AAGCATAAAG TGAAAGCCT
 6051 GGGGTGCCATA ATGAGTGAGC TAACTCACAT TAATTGCGTT GCGCTCACTG
 6101 CCCGCTTCC AGTCGGAAA CCTGTCGTGC CAGCTGCATT AATGAATCGG
 6151 CCAACGCGC GGGAGAGGCG GTTTCGTAT TGGCGCTCT TCCGCTTCCT
 6201 CGCTCACTGA CTCGCTGC CTCGGTGTTC GGCTGCGCG AGCGGTATCA
 6251 GCTCACTCAA AGGCGGTAAT ACGGTTATCC ACAGAATCAG GGGATAACGC
 6301 AGGAAAGAAC ATGTGAGCAA AAGGCCAGCA AAAGGCCAGG AACCGTAAAAA
 6351 AGGCCGCGTT GCTGGCGTT TTCCATAGGC TCCGCCCCC TGACGAGCAT
 6401 CACAAAAATC GACGCTCAAG TCAGAGGTGG CGAAACCCGA CAGGACTATA
 6451 AAGATACCAG GCGTTCCCC CTGGAAGCTC CCTCGTGC CTCCTGTTC
 6501 CGACCCCTGCC GCTTACCGGA TACCTGTCCG CCTTCTCCC TTGGGAAGC
 6551 GTGGCGCTTT CTCATAGCTC ACGCTGTAGG TATCTCAGTT CGGTGTAGGT
 6601 CGTCGCTCC AAGCTGGCT GTGTGCACGA ACCCCCCGTT CAGCCCCGACC
 6651 GCTGCGCCTT ATCCGTAAC TATCGTCTTG AGTCCAACCC GGTAAGACAC
 6701 GACTTATCGC CACTGGCAGC AGCCACTGGT AACAGGATTA GCAGAGCGAG
 6751 GTATGTAGGC GGTGCTACAG AGTTCTGAA GTGGTGGCCT AACTACGGCT
 6801 ACACATAGAAC AACAGTATTT GGTATCTGC CTCGTGTAA GCCAGTTACC
 6851 TTCGGAAAAA GAGTTGGTAG CTCTTGATCC GGCAAACAAA CCACCGCTGG
 6901 TAGCGGTTTT TTGTTTGCA AGCAGCAGAT TACCGCGAGA AAAAAAGGAT
 6951 CTCAGGAAAGA TCCCTTGATC TTTCTACGG GGTCTGACGC TCAGTGGAAC
 7001 GAAAACCTAC GTTAAGGGAT TTTGGTCATG AGATTATCAA AAAGGATCTT
 7051 CACCTAGATC CTTTAAATT AAAAATGAAG TTTAAATCA ATCTAAAGTA
 7101 TATATGAGTA AACTTGGTCT GACAGTTACC AATGCTTAAT CAGTGAGGCA
 7151 CCTATCTCAG CGATCTGTCT ATTCGTTCA TCCATAGTTG CCTGACTCCC
 7201 CGTCGTGTAG ATAACCTACGA TACGGGGAGGG CTTACCATCT GGCCCCAGTG
 7251 CTGCAATGAT ACCCGAGAC CCACGCTCAC CGGCTCCAGA TTTATCAGCA
 7301 ATAAACCAAGC CAGCGGAAAG GGCGGAGCGC AGAAGTGGTC CTGCAACTTT
 7351 ATCCGCTCC ATCCAGTCTA TTAATTGTTG CGGGAAAGCT AGAGTAAGTA
 7401 GTTCGCCAGT TAATAGTTG CGAACGTTG TTGCCATTGC TACAGGCATC

-continued

7451 GTGGTGTAC GCTCGTCGTT TGGTATGGCT TCATTCAGCT CCGGTTCCCA
 7501 ACGATCAAGG CGAGTTACAT GATCCCCAT GTTGTGAAA AAAGCGTTA
 7551 GCTCCTTCGG TCCTCCGATC GTTGTAGAA GTAAGTTGGC CGCAGTGTAA
 7601 TCACTCATGG TTATGGCAGC ACTGCATAAT TCTCTTACTG TCATGCCATC
 7651 CGTAAGATGC TTTTCTGTGA CTGGTGAGTA CTCAACCAAG TCATTCTGAG
 7701 AATAGTGTAT GCGGCGACCG AGTTGCTCTT GCGGGCGTC AATAACGGGAT
 7751 AATAACCGCGC CACATAGCAG AACTTTAAAA GTGCTCATCA TTGGAAAACG
 7801 TTCTTCGGGG CGAAAACTCT CAAGGATCTT ACCGCTGTTG AGATCCAGTT
 7851 CGATGTAACC CACTCGTGCA CCCAACTGAT CTTCAGCATC TTTTACTTTC
 7901 ACCAGCGTTT CTGGGTGAGC AAAAACAGGA AGGCAAATG CCGCAAAAAAA
 7951 GGGATAAAGG GCGACACGGA AATGTTGAAT ACTCATACTC TTCTTTTC
 8001 AATATTATTG AAGCATTAT CAGGGTTATT GTCTCATGAG CGGATAACATA
 8051 TTGAAATGTA TTTAGAAAAAA TAAACAAATA GGGGTTCCGC GCACATTCC
 8101 CCGAAAAGTG CCACCTGACG TC

According to one embodiment, the vector is a plasmid and has the sequence of pAV-U6+27-Tornado-F30-Pepper(TAR Variant-2) (SEQ ID NO: 78; GenBank Accession No. MN052908.1, which is hereby incorporated by reference in its entirety) as follows:

1 GCCGGATCCA AGGTGGGCA GGAAGAGGGC CTATTCCTCA TGATTCCCTC
 51 ATATTTGCAT ATACGATACA AGGCTGTTAG AGAGATAATT AGAATTAATT
 101 TGACTGTAAA CACAAAGATA TTAGTACAAA ATACGTGACG TAGAAAGTAA
 151 TAATTTCTTG GGTAGTTGC AGTTTAAAAA TTATGTTTTA AAATGGACTA
 201 TCATATGCTT ACCGTAACCTT GAAAGTATTT CGATTTCTTG GCTTTATATA
 251 TCTTGTGGAA AGGACGAAAC ACCGTGCTCG CTTCGGCAGC ACATATACTA
 301 GTCGACGGGC CGCACTCGCC GGTCCAAGC CCGGATAAAA TGGGAGGGGG
 351 CGGGAAACCG CCTAACCATG CCGAGTGCAG CGCCTTGCCA TGTGTATGTG
 401 GGACGCGTTG CCACGTTCC CACATACTCT GATGATCCGC TAGCAAAGGC
 451 TCGTTGAGCT CATTAGCTCC GAGCCCGAGG TACCGGATCA TTGATGGCAA
 501 GCGGCCGCGG TCGGCGTGG A CTGTAGAAC A CTGCCAATGC CGGTCCAAG
 551 CCCGGATAAA AGTGGAGGGT ACAGTCCACG CTCTAGAGCG GACTTCGGTC
 601 CGCTTTTAC TAGGACCTGC AGGCATGCAA GCTTGACGTC GGTTACCGAT
 651 ATCCATATGG CGACCGCATC GATCTCGAGC CGAGGACTAG TAACTTGTGTT
 701 ATTGCAAGCTT ATAATGGTTA CAAATAAAGC AATAGCATCA CAAATTCAC
 751 AAATAAAGCA TTTTTTCAC TGCAATTCTAG TTGTGGTTG TCCAAACTCA
 801 TCAATGTATC TTATCATGTC TTACGTAGAT AAGTAGCATG GCGGGTTAAT
 851 CATTAACTAC AAGGAACCCC TAGTGATGGA GTTGGCCACT CCCCTCTGCG
 901 GCGCTCGCTC GCTCACTGAG GCCGGCGAC CAAAGGTGCG CCGACGCCCG
 951 GGCTTTGCC GGGCGGCCCTC AGTGAGCGAG CGAGCGCGA GAGAGGGAGT
 1001 GGCCAAAGAT CTCTGGCGTA ATAGCGAAGA GGCCCGCACC GATCGCCCTT
 1051 CCCAACAGTT GCGCAGCCTG AATGGCTAAT GGGAAATTGT AACGTTAAT

-continued

1101 ATTTTGTAA TATTTGTTA AAATTCGCGT TAAATTTTG TTAAATCAGC
 1151 TCATTTTTA ACCAATAGGC CGAAATCGGC AAAATCCCTT ATAAATCAAA
 1201 AGAATAGACC GAGATAGGGT TGAGTGTGT TCCAGTTGG AACAAAGAGTC
 1251 CACTATTAAA GAACGTGGAC TCCAACGTCA AAGGGCGAAA AACCGTCTAT
 1301 CAGGGCGATG GCCCCACTACG TGAAACCATCA CCCTAATCAA GTTTTTGGG
 1351 GTCGAGGTGC CGTAAGCAGC TAAATCGGA CCCTAAAGGG ATGCCCGAT
 1401 TTAGAGCTTG ACGGGAAAG CCGGCGAACG TGGCGAGAAA GGAAGGGAAAG
 1451 AAAGCGAAG GAGCGGGCGC TAGGGCGCTG GCAAGTGTAG CGGTACGCT
 1501 GCGCGTAACC ACCACACCCG CCGCGCTTAA TGCGCCGCTA CAGGGCGCGT
 1551 CAGGTGGCAC TTTTCGGGAA AATGTGCGCG GAACCCCTAT TTGTTTATTT
 1601 TTCTAAATAC ATTCAAATAT GTATCCGCTC ATGAGACAAT AACCCGTATA
 1651 AATGCTTCAA TAATATTGAA AAAGGAAGAG TATGAGTATT CAACATTCC
 1701 GTGTGCCCT TATTCCCTTT TTTGGGGCAT TTTGCCTTCC TGTTTTGCT
 1751 CACCCAGAAA CGCTGGTGAA AGTAAAAGAT GCTGAAGATC AGTTGGGTGC
 1801 ACGAGTGGGT TACATCGAAC TGGATCTCAA CAGCGGTAAAG ATCCTTGAGA
 1851 GTTTTCCGCC CGAAGAACGT TTTCCAATGA TGAGCACTTT TAAAGTTCTG
 1901 CTATGTGGCG CGGTATTATC CCGTATTGAC GCCGGGCAAG AGCAACTCGG
 1951 TCGCCGCATA CACTATTCTC AGAATGACTT GGTTGAGTAC TCACCGAGTC
 2001 CAGAAAAGCA TCTTACGGAT GGCAATGACAG TAAGAGAATT ATGCAGTGCT
 2051 GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT
 2101 CGGAGGACCG AAGGAGCTAA CCGCTTTTT GCACAAACATG GGGGATCATG
 2151 TAACTGCCCT TGATCGTTGG GAACCGGAGC TGAATGAAGC CATAACAAAC
 2201 GACGAGCGTG ACACCAACGAT GCCTGTAGCA ATGGCAACAA CGTTGCGCAA
 2251 ACTATTAACG TGCAGACTAC TTACTCTAGC TTCCCGGCAA CAATTAATAG
 2301 ACTGGATGGA GGCGGATAAA GTTGCAGGAC CACTTCTGCG CTCGGCCCTT
 2351 CGGGCTGGCT GTTTTATTGC TGATAAATCT GGAGCCGGTG AGCGTGGGTG
 2401 TCGCGGTATC ATTGCAGCAC TGGGCCAGA TGGTAAGCCC TCCCGTATCG
 2451 TAGTTATCTA CACGACGGGG AGTCAGGCAA CTATGGATGA ACGAAATAGA
 2501 CAGATCGCTG AGATAGGTGC CTCACTGATT AAGCATTGGT AACTGTCAGA
 2551 CCAAGTTAC TCATATATAC TTTAGATTGA TTTAAAACCTT CATTTTTAAT
 2601 TTAAAAGGAT CTAGGTGAAG ATCCTTTTG ATAATCTCAT GACCAAAATC
 2651 CCTTAACGTG AGTTTCGTT CCACTGAGCG TCAGACCCCG TAGAAAAGAT
 2701 CAAAGGATCT TCTTGAGATC CTTTTTTCT GCGCGTAATC TGCTGCTTGC
 2751 AAACAAAAAA ACCACCGCTA CCAGCGGTGG TTTGTTGCC GGATCAAGAG
 2801 CTACCAACTC TTTTCGAA GGTAACTGGC TTCAGCAGAG CGCAGATACC
 2851 AAATACTGTC CTTCTAGTGT AGCCGTAGTT AGGCCACCAC TTCAAGAACT
 2901 CTGTAGCACC GCCTACATAC CTCGCTCTGC TAATCCTGTT ACCAGTGGCT
 2951 GCTGCCAGTG GCGATAAGTC GTGTCTTACC GGGTTGGACT CAAGACGATA
 3001 GTTACCGGAT AAGGCGCAGC GGTCGGGCTG AACGGGGGGT TCGTGCAACA
 3051 CAGCCAGCTT GGAGCGAACG ACCTACACCG AACTGAGATA CCTACAGCGT

-continued

3101 GAGCATTGAG AAAGCGCCAC GCTTCCCGAA GGGAGAAAGG CGGACAGGTA
 3151 TCCGGTAAGC GGCAGGGTCG GAACAGGAGA GCGCACGAGG GAGCTTCCAG
 3201 GGGGAAACGC CTGGTATCTT TATAGTCCTG TCGGGTTTCG CCACCTCTGA
 3251 CTTGAGCGTC GATTTTGAG ATGCTCGTCA GGGGGCGGA GCCTATGGAA
 3301 AAACGCCAGC AACGCCGCCT TTTTACGGTT CCTGGCCTT TGCTGGCCTT
 3351 TTGCTCACAT GTTCTTCCT GCGTTATCCC CTGATTCTGT GGATAACCGT
 3401 ATTACCGCCT TTGAGTGAGC TGATACCGCT CGCCGCAGCC AACGACCGA
 3451 GCGCAGCGAG TCAGTGAGCG AGGAAGCGGA AGAGCGCCCA ATACGCAAAC
 3501 CGCCTCTCCC CGCGCGTTGG CCGATTCAATT AATGCAGAGA TCTTTGCCA
 3551 CTCCCTCTCT GCGCGCTCGC TCGCTCACTG AGGCCGGCG ACCAAAGGTC
 3601 GCCCGACGCC CGGGCTTGC CCAGCGGGCG TCAGTGAGCG AGCGAGCGCG
 3651 CAGAGAGGGA GTGGCCAAGT CCATCACTAG GGGTCCTGG AGGGGTGGAG
 3701 TCGTGACGTG AATTACGTCA TAGGTTAGG GAGGTCTGG ATCGATCCAG
 3751 ACATGATAAG ATACATTGAT GAGTTGGAC AAACACAAAC TAGAATGCAG
 3801 TGAAAAAAAT GCTTTATTG TGAAATTGT GATGCTATTG CTTTATTGTT
 3851 AACCATATA AGCTGCAATA AACAAAGTTAA CAACAACAAT TGCATTCAATT
 3901 TTATGTTCA GGTCAGGGG GAGGTGTGGG AGGTTTTTA AAGCAAGTAA
 3951 AACCTCTACA AATGTGGTAT GGCTGATTAT GATCTCTAGT CAAGGCACTA
 4001 TACATCAAAT ATTCCATTAA AACCCCTTTA CAAATTAAAA AGCTAAAGGT
 4051 ACACAATTT TGAGCATAGT TATTAATAGC AGACACTCTA TGCTGTGTG
 4101 GAGTAAGAAA AAACAGTATG TTATGATTAT AACTGTTATG CCTACTTATA
 4151 AAGGTTACAG AATATTTC CATAATTTC TTGTATAGCA GTGCAGCTT
 4201 TTCCTTTGTG GTGTAATAG CAAAGCAAGC AAGAGTTCTA TTACTAAACA
 4251 CAGCATGACT CAAAAAAACTT AGCAATTCTG AAGGAAAGTC CTTGGGTCT
 4301 TCTACCTTTC TCTTCTTTT TGGAGGAGTA GAATGTTGAG AGTCAGCAGT
 4351 AGCCTCATCA TCACTAGATG GCATTCTTC TGAGCAAAAC AGGTTTCCT
 4401 CATTAAGGC ATTCCACCAC TGCTCCCATT CATCAGTTCC ATAGGTTGGA
 4451 ATCTAAAATA CACAAACAAT TAGAACAGT AGTTAACAC ATTATACACT
 4501 TAAAAATTAAAT ATATTCACCT TAGAGCTTAA AATCTCTGTA GGTAGTTGTT
 4551 CCAATTATGT CACACCACAG AAGTAAGGTT CCTTCACAAA GATCCGGGAC
 4601 CAAAGCGGCC ATCGTGCCTC CCCACTCCTG CAGTCGGGG GCATGGATGC
 4651 GCGGATAGCC GCTGCTGGTT TCCCTGGATGC CGACGGATTG GCACTGCCGG
 4701 TAGAACTCCG CGAGGTGTC CAGCCTCAGG CAGCAGCTGA ACCAAACTCGC
 4751 GAGGGGATCG AGCCCCGGGT GGGCGAAGAA CTCCAGCATG AGATCCCCGC
 4801 GCTGGAGGAT CATCCAGCCG GCGTCCCGGA AAACGATTCC GAAGCCCAAC
 4851 CTTTCATAGA AGGCGCGGGT GGAATCGAAA TCTCGTGATG GCAGGTTGGG
 4901 CGTCGCTTGG TCGGTATTT CGAACCCAG AGTCCCGCTC AGAAGAACTC
 4951 GTCAAGAAGG CGATAGAAGG CGATGCGCTG CGAACATGGGA CGGGCGATAAC
 5001 CGTAAAGCAC GAGGAAGCGG TCAGCCCATT CGCCGCCAAG CTCTTCAGCA
 5051 ATATCACGGG TAGCCAACGC TATGCTCTGA TAGCGGTCCG CCACACCCAG
 5101 CGGGCCACAG TCGATGAATC CAGAAAAGCG GCCATTTC ACCATGATAT

-continued

5151 TCGGCAAGCA GGCATGCCA TGGGTACAGA CGAGATCCTC GCCGTCGGC
 5201 ATGCGCCCT TGAGCCTGGC GAACAGTTCG GCTGGCGCGA GCCCCGTATG
 5251 CTCTTGCTCA GATCATCCTG ATCGACAAGA CCGGCTTCCA TCCGAGTACG
 5301 TGCTCGCTCG ATGCGATGTT CGCTTGGTGG TCGAATGGGC AGGTAGCCGG
 5351 ATCAAGCGTA TGCAGCCGCC GCATTGCATC AGCCATGATG GATACTTTCT
 5401 CGGCAGGAGC AAGGTGAGAT GACAGGAGAT CCTGCCCCGG CACTTCGCC
 5451 AATAGCAGCC AGTCCTTCC CGCTTCAGTG ACAACGTCGA GCACAGCTGC
 5501 GCAAGGAACG CCCGTCGTGG CCAGCACAGA TAGCCGCGCT GCCTCGTCC
 5551 GCAGTTCAATT CAGGGCACCG GACAGGTCGG TCTTGACAAA AAGAACCGGG
 5601 CGCCCCCTGCG CTGACAGCCG GAACACGGCG GCATCAGAGC AGCCGATTGT
 5651 CTGTTGTGCC CAGTCATAGC CGAATAGCCT CTCCACCCAA GCGGCCGGAG
 5701 AACCTGCGTG CAATCCATCT TGTTCAATCA TGCAGAACAGA TCCTCATCCT
 5751 GTCTCTTGAT CAGATCTTGA TCCCCGCGC CATCAGATCC TTGGCGGCAA
 5801 GAAAGCCATC CAGTTTACTT TGCAAGGGCTT CCCAACCTTA CCAGAGGGCG
 5851 CCCCAGCTGG CAATTCCGGT TCGCTTGCTG TCCATAAAAC CGCCCAGTCT
 5901 AGCTATCGGC ATGTAAGCCC ACTGCAAGCT ACCTGCTTTC TCTTTGCGCT
 5951 TGCCTTTCC CTTGTCCAGA TAGCCCAGTA GCTGACATTC ATCCGGGGTC
 6001 AGCACCGTTT CTGCGGACTG GCTTTCTACG TGTTCCGCTT CCTTTAGCAG
 6051 CCCTTGCGCC CTGAGTGCTT GCGGCAGCGT GAAGCTTTT GCAAAAGCCT
 6101 AGGCCTCCAA AAAAGCCTCC TCAACTACTTC TGGAATAGCT CAGAGGCCGA
 6151 GCGGGCTCG GCCTCTGCAT AAATAAAAAA ATTAGTCAG CCATGGGGCG
 6201 GAGAATGGGC GGAACGGGC GGAGTTAGGG GCGGGATGGG CGGAGTTAGG
 6251 GGCAGGGACTA TGGTTGCTGA CTAATTGAGA TGCATGCTT GCATACTTCT
 6301 GCCTGCTGGG GAGCCTGGGG ACTTTCCACA CCTGGTTGCT GACTAATTGA
 6351 GATGCATGCT TTGCATACTT CTGCTTGCTG GGGAGCCTGG GGACTTTCCA
 6401 CACCTAACT GACACACATT CCACA

According to one embodiment, the vector is a plasmid and has the sequence of pAV-U6+27-Tornado-F30-TAR Variant-1 (SEQ ID NO: 79; GenBank Accession No. MN052909.1, which is hereby incorporated by reference in its entirety) as follows:

1 GCCGGATCCA AGGTGGGCA GGAAGAGGGC CTATTCCTCA TGATTCCCTC
 51 ATATTTGCAT ATACGATACA AGGCTGTTAG AGAGATAATT AGAATTAATT
 101 TGACTGTAAA CACAAAGATA TTAGTACAAA ATACGTGACG TAGAAAGTAA
 151 TAATTTCTTG GGTAGTTGC AGTTTTAAA TTATGTTTA AAATGGACTA
 201 TCATATGCTT ACCGTAACCT GAAAGTATT CGATTTCTTG GCTTTATATA
 251 TCTTGTGGAA AGGACGAAAC ACCGTGCTCG CTTCGGCAGC ACATATACTA
 301 GTCGACGGGC CGCACTCGCC GGTCCCAAGC CCGGATAAAA TGGGAGGGGG
 351 CGGGAAACCG CCTAACCATG CCGAGTGCAGG CCGCTTGCCA TGTGTATGTG
 401 GGACGCCTG CCACGTTCC CACACTCT GATGATCCGC TAGCAAAGGC
 451 TCGTCTGAGC TCATTAGCTC CGAGCCGAG GTACCGGATC ATTCAATGGCA

-continued

501 AGCGGCCGCG GTCGGCGTGG ACTGTAGAAC ACTGCCAATG CCGGTCCCAA
 551 GCCCCGATAA AAGTGGAGGG TACAGTCCAC GCTCTAGAGC GGACTTCGGT
 601 CCGCTTTTA CTAGGACCTG CAGGCATGCA AGCTTGACGT CGGTTACCGA
 651 TATCCATATG GCGACCGCAT CGATCTCGAG CCGAGGACTA GTAACATTGTT
 701 TATTGCAGCT TATAATGGTT ACAAAATAAG CAATAGCATC ACAAAATTCA
 751 CAAATAAACG ATTTTTTCA CTGCATTCTA GTTGTGGTT GTCCAAACTC
 801 ATCAATGTAT CTTATCATGT CTTACGTAGA TAAGTAGCAT GGC GGTTAA
 851 TCATTAACTA CAAGGAACCC CTAGTGATGG AGTTGGCCAC TCCCTCTCTG
 901 CGCGCTCGCT CGCTCACTGA GGCGGGCGA CCAAAGGTCG CCCGACGCC
 951 GGGCTTTGCC CGGGCGGCCT CAGTGAGCGA GCGAGCGCGC AGAGAGGGAG
 1001 TGGCCAAAGA TCTCTGGCGT AATAGCGAAG AGGCCCGCAC CGATGCCCT
 1051 TCCCCAACAGT TGCGCAGCCT GAATGGCTAA TGGGAAATTG TAAACGTTAA
 1101 TATTTTGTAA ATATTTGTT AAAATTCGCG TTAAATTTT GTAAATCAG
 1151 CTCATTTTT AACCAATAGG CCGAAATCGG CAAAATCCCT TATAATCAA
 1201 AAGAATAGAC CGAGATAGGG TTGAGTGTG TTCCAGTTG GAACAAGAGT
 1251 CCACTATTAA AGAACGTGGA CTCCAACGTC AAAGGGCGAA AAACCGTCTA
 1301 TCAGGGCGAT GGCCCACACTAC GTGAACCATC ACCCTAATCA AGTTTTTGG
 1351 GGTCGAGGTG CCGTAAAGCA CTAAATCGGA ACCCTAAAGG GATGCCCGA
 1401 TTTAGAGCTT GACGGGAAA GCCGGCGAAC GTGGCGAGAA AGGAAGGGAA
 1451 GAAAGCGAAA GGAGCGGGCG CTAGGGCGCT GGCAAGTGTAA CGGGTCACGC
 1501 TCGCGTAAAC CACCACACCC GCCCGCGCTTA ATCGCGCGCT ACAGGGCGCG
 1551 TCAGGGTGCCTA CTTTTCGGGG AAATGTGCGC GGAACCCCTA TTTGTTTATT
 1601 TTTCTAAATA CATTCAAATA TGTATCCGCT CATGAGACAA TAACCTGAT
 1651 AAATGCTTCA ATAATATTGA AAAAGGAAGA GTATGAGTAT TCAACATTTC
 1701 CGTGTGCCCT TTATTCCTT TTTGCGGCA TTTGCCTTC CTGTTTTGC
 1751 TCACCCGAAAC CGCTGGTGA AAGTAAAAGA TGCTGAAGAT CAGTTGGTG
 1801 CACGAGTGGG TTACATCGAA CTGGATCTCA ACAGCGGTAA GATCCTTGAG
 1851 AGTTTCGCC CGGAAGAACG TTTTCCAATG ATGAGCACTT TTAAAGTTCT
 1901 GCTATGTGGC GCGGTATTAT CCCGTATTGA CGCCGGCAA GAGCAACTCG
 1951 GTCGCCCAT ACACATTCT CAGAATGACT TGGTTGAGTA CTCACCAGTC
 2001 ACAGAAAAGC ATCTTACGGA TGGCATGACA GTAAGAGAAAT TATGCAAGTGC
 2051 TGCCATAACC ATGAGTGATA ACACGTGGC CAACTTACTT CTGACAACGAA
 2101 TCGGAGGACC GAAGGGAGCTA ACCGCTTTT TGCAACACAT GGGGGATCAT
 2151 GTAACTCGCC TTGATCGTT GGAAACCGGAG CTGAATGAAG CCATACAAAA
 2201 CGACGAGCGT GACACCACGA TGCCTGTAGC AATGGCAACA ACGTTGCGCA
 2251 AACTATTAAC TGGCGAACTA CTTACTCTAG CTTCCCGGCA ACAATTAAATA
 2301 GACTGGATGG AGGCGGATAA AGTTGCAGGA CCACCTCTGC GCTCGGCCCT
 2351 TCCGGCTGGC TGGTTATTG CTGATAAAATC TGGAGCCGGT GAGCGTGGGT
 2401 CTCGCGGTAT CATTGAGCA CTGGGGCCAG ATGGTAAGGCC CTCCCGTATC
 2451 GTAGTTATCT ACACGACGGG GAGTCAGGCA ACTATGGATG AACGAAATAG

-continued

2501 ACAGATCGCT GAGATAGGTG CCTCACTGAT TAAGCATTGG TAACTGTCAG
 2551 ACCAAGTTA CTCATATATA CTTTAGATTG ATTTAAAAC TCACTTTTAA
 2601 TTAAAGAGA TCTAGGTGAA GATCCTTTT GATAATCTCA TGACCAAAAT
 2651 CCCTTAACGT GAGTTTCGT TCCACTGAGC GTCAGACCCC GTAGAAAAGA
 2701 TCAAAGGATC TTCTTGAGAT CCTTTTTTC TGCGCGTAAT CTGCTGCTTG
 2751 CAAACAAAAA AACCACCGCT ACCAGCGGTG GTTTGTTGC CGGATCAAGA
 2801 GCTACCAACT CTTTTCCGA AGGTAACTGG CTTCAGCAGA GCGCAGATAC
 2851 CAAATACTGT CTTCTAGTG TAGCCGTAGT TAGGCCACCA CTTCAAGAAC
 2901 TCTGTAGCAC CGCCTACATA CCTCGCTCTG CTAATCCTGT TACCAAGTGGC
 2951 TGCTGCCAGT GGCGATAAGT CGTGTCTTAC CGGGTTGGAC TCAAGACGAT
 3001 AGTTACCGGA TAAGGCGCAG CGGTGGGCT GAACGGGGGG TTCTGTGCAAC
 3051 ACAGCCAGCT TGGAGCGAAC GACCTACACC GAACTGAGAT ACCTACAGCG
 3101 TGAGCATTGA GAAAGCGCCA CGCTCCCCGA AGGGAGAAAG CGGGACAGGT
 3151 ATCCGGTAAG CGGCAGGGTC GGAACAGGGAG AGCGCACGAG GGAGCTTCCA
 3201 GGGGAAACG CCTGGTATCT TTATAGTCCT GTCGGGTTTC GCCACCTCTG
 3251 ACTTGAGCGT CGATTTTGT GATGTCGTC AGGGGGCGG AGCTATGGA
 3301 AAAACGCCAG CAACCGGGCC TTTTACGGT TCCTGGCCTT TTGCTGGCCT
 3351 TTTGCTCACA TGTTCTTCC TGCGTTATCC CCTGATTCTG TGGATAACCG
 3401 TATTACCGCC TTGAGTGAG CTGATAACCGC TCGCCGCAGC CGAACGACCG
 3451 AGCGCAGCGA GTCAGTGAGC GAGGAAGCGG AAGAGCGCCC AATACGCAA
 3501 CGCCTCTCC CGCGCGCTTG GCCGATTCA TAATGCAGAG ATCTTGGCC
 3551 ACTCCCTCTC TGCGCGCTCG CTCGCTCACT GAGGCCGGGC GACCAAAGGT
 3601 CGCCCGACGC CGGGGCTTG CCCGGCGGC CTCAGTGAGC GAGCGAGCGC
 3651 GCAGAGAGGG AGTGGCCAAC TCCATCACTA GGGGTTCTG GAGGGTGG
 3701 GTCGTGACGT GAATTACGTC ATAGGGTTAG GGAGGTCTG GATCGATCCA
 3751 GACATGATAA GATACATTGA TGAGTTGGA CAAACCACAA CTAGAATGCA
 3801 GTGAAAAAAA TGCTTATTG TGAAATTG TGATGCTATT GCTTTATTG
 3851 TAACCATTAT AAGCTGCAAT AAACAAGTT ACAACAACAA TTGCATTCA
 3901 TTTATGTTTC AGGTCAGGG GGAGGTGTGG GAGGTTTTT AAAGCAAGTA
 3951 AACCTCTAC AAATGTGGTA TGGCTGATTA TGATCTCTAG TCAAGGCACT
 4001 ATACATCAA TATTCCTTAT AAACCCCTT ACAATTAAA AAGCTAAAGG
 4051 TACACATTG TTGAGCATAG TTATTAATAG CAGACACTCT ATGCCCTGTG
 4101 GGAGTAAGAA AAAACAGTAT GTTATGATTA TAACTGTTAT GCCTACTTAT
 4151 AAAGGTTACA GAATTTTT CCATAATTG CTTGTATAGC AGTCAGCAG
 4201 TTCCCTTGT GGTGTTAAATA GCAAAGCAAG CAAGAGTTCT ATTACTAAC
 4251 ACAGCATGAC TCAAAAAACT TAGCAATTCT GAAGGAAAGT CCTGGGGTC
 4301 TTCTACCTT CTCTTCTT TTGGAGGAGT AGAATGTTGA GAGTCAGCAG
 4351 TAGCCTCATC ATCACTAGAT GGCATTCTT CTGAGCAAA CAGGTTTCC
 4401 TCATTAAGG CATTCCACCA CTGCTCCCAT TCATCAGTTC CATAGGTTGG
 4451 AATCTAAAT ACACAAACAA TTAGAATCAG TAGTTAACCA CATTATACAC
 4501 TAAAAAATT TATATTACG TTAGAGCTTT AAATCTCTGT AGGTAGTTG

-continued

4551 TCCAATTATG TCACACCACA GAAGTAAGGT TCCTTCACAA AGATCCGGGA
 4601 CCAAAGCGGC CATCGTGCCT CCCCACTCCT GCAGTTGGGG GGCATGGATG
 4651 CGCGGATAGC CGCTGCTGGT TTCTGGATG CCGACGGATT TGCACGCCG
 4701 GTAGAACTCC GCGAGGTCGT CCAGCCTAG GCAGCAGCTG AACCAACTCG
 4751 CGAGGGGATC GAGCCCCGGG TGGCGAAGA ACTCCAGCAT GAGATCCCCG
 4801 CGCTGGAGGA TCATCCAGCC GGCGTCCCGG AAAACGATTC CGAAGCCAA
 4851 CCTTCATAG AAGGCGGC GGAAATCGAA ATCTCGTGT GGCAGGTTGG
 4901 GCGTCGCTTG GTCGGTCATT TCGAACCCCA GAGTCCCCT CAGAAGAACT
 4951 CGTCAAGAAG GCGATAGAAG GCGATGCGCT GCGAATCGGG AGCGGCGATA
 5001 CGCTAAAGCA CGAGGAAGCG GTCAGCCCAT TCGCCGCAA GCTCTTCAGC
 5051 AATATCACGG GTAGCCAACG CTATGCTCTG ATAGCGGTCC GCCACACCCA
 5101 GCGGCCACA GTCGATGAAT CCAGAAAAGC GGCCATTTC CACCATGATA
 5151 TTCGGCAAGC AGGCATCGCC ATGGGTCAAG ACAGAGATCCT CGCCGTCGGG
 5201 CATGCGCGC TTGAGCCTGG CGAACAGTTC GGCTGGCGCG AGCCCCTGAT
 5251 GCTCTTGCC AGATCATCCT GATCGACAAG ACCGGCTTCC ATCCGAGTAC
 5301 GTGCTCGCTC GATGCGATGT TCGCTTGGTG GTCGAATGGG CAGGTAGCCG
 5351 GATCAAGCGT ATGCAGCCGC CGCATTGCAT CAGCCATGAT GGATACTTTC
 5401 TCGGCAGGAG CAAGGTGAGA TGACAGGAGA TCCTGCCCG GCACTTCGCC
 5451 CAATAGCAGC CAGTCCTTC CCGCTTCAGT GACAACGTCG AGCACAGCTG
 5501 CGCAAGGAAC GCGCGTCGTG GCCAGCCACG ATAGCCGCGC TGCTCGTCC
 5551 TGCAGTTCAT TCAGGGCACC GGACAGGTG GTCTTGACAA AAAGAACCGG
 5601 GCGCCCTGC GCTGACAGCC GGAACACGGC GGCATCAGAG CAGCCGATTG
 5651 TCTGTTGTGC CCAGTCATAG CCGAATAGCC TCTCCACCCA AGCGGCCGG
 5701 GAAACCTGCGT GCAATCCATC TTGTTCAATC ATGCGAAACG ATCCTCATCC
 5751 TGTCTCTTGA TCAGATCTTG ATCCCTGCG CCATCAGATC CTTGGCGGCA
 5801 AGAAAGCCAT CCAGTTTACT TTGCAGGGCT TCCCAACCTT ACCAGAGGGC
 5851 GCGCCAGCTG GCAATTCGG TTCGCTTGCT GTCCATAAAA CGGCCAGTC
 5901 TAGCTATCGG CATGTAAGCC CACTGCAAGC TACCTGCTT CTCTTGCGC
 5951 TTGCGTTTTC CCTTGCCAG ATAGCCCAGT AGCTGACATT CATCCGGGT
 6001 CAGCACCGTT TCTGCGGACT GGCTTCTAC GTGTTCCGCT TCCTTTAGCA
 6051 GCGCTTGCAGC CCTGAGTGCT TGCGGCAGCG TGAAGCTTT TGCAAAAGCC
 6101 TAGGCCTCCA AAAAAGCTC CTCACTACTT CTGGAATAGC TCAGAGGCCG
 6151 AGGCAGGCTC GGCCTCTGCA TAAATAAAA AAATTAGTCA GCCATGGGGC
 6201 GGAGAATGGG CGGAACATGGG CGGAGTTAGG GGCGGGATGG GCGGAGTTAG
 6251 GGGCGGGACT ATGGTTGCTG ACTAATTGAG ATGCATGCTT TGCATACTTC
 6301 TGCCTGCTGG GGAGCCTGGG GACTTCCAC ACCTGGTTGC TGACTAATTG
 6351 AGATGCGATGC TTGCGATACT TCTGCTGCT GGGGAGCCTG GGGACTTTCC
 6401 ACACCCCTAAC TGACACACAT TCCACA

As described herein, the vector may comprise two, three, four, five, or more nucleic acid sequences according to the present application. In some embodiments, the vector comprises a first nucleic acid sequences encoding a first RNA-regulated fusion protein and a second nucleic acid sequence encoding a second RNA-regulated fusion protein. In other embodiments, the vector may further comprise a third nucleic acid molecule encoding a third RNA-regulated fusion protein, etc. For example, the vector may comprise 3-10 or more nucleic acid molecules, each encoding an independently selected RNA fusion protein according to the present application.

In some embodiments, where the vector encodes multiple RNA-regulated fusion proteins, each independent fusion protein may comprise a component of a metabolic pathway. In some embodiments, the metabolic pathway is glucose metabolism and the independent fusion proteins comprise insulin, glucagon, and/or protein kinase C epsilon. In other embodiments, the metabolic pathway is a GPCR signaling pathway and the independent fusion proteins are selected from the group consisting of α , β , and γ subunits of G-proteins.

In other embodiments, where the vector encodes multiple RNA-regulated fusion proteins, each RNA-regulated fusion protein comprises a distinct protein of interest. Suitable proteins of interest are described in detail above. In some embodiments, the proteins of interest comprise fluorescent proteins. In accordance with such embodiments, the fluorescent proteins have fluorescent emission spectra that do not substantially overlap with one another.

In some embodiments, the present application relates to an expression system comprising an expression vector into which is inserted a nucleic acid molecule described herein. In one embodiment, the expression system comprises a first vector encoding an RNA-regulated fusion protein and a second vector encoding a lentiviral transactivator of transcription (Tar) RNA aptamer.

Some embodiments of the present application relate to a host cell comprising a nucleic acid molecule (i.e., a nucleic acid molecule encoding an RNA-regulated fusion protein and/or a lentiviral transactivator of transcription (Tar) RNA sequence) or a vector (i.e., a vector comprising a nucleic acid molecule encoding an RNA-regulated fusion protein and/or a lentiviral transactivator of transcription (Tar) RNA sequence) described herein.

In some embodiments, the host cell is a mammalian cell. Suitable mammalian cells include, without limitation, rodent cells (i.e., mouse or rat cells), rabbit cells, guinea pig cells, feline cells, canine cells, porcine cells, equine cells, bovine cell, ovine cells, monkey cells, non-human primate, or human cells. In some embodiments, the host cell is a human

cell. Suitable cells comprising the nucleic acid molecule or vector as described herein include primary or immortalized embryonic cells, fetal cells, or adult cells, at any stage of their lineage, e.g., totipotent, pluripotent, multipotent, or differentiated cells.

The nucleic acid molecules and/or vectors described herein may be introduced into cells via transformation, particularly transduction, conjugation, lipofection, protoplast fusion, mobilization, particle bombardment, microinjection, transfection, or electroporation. In some embodiments, the nucleic acid molecules described herein are incorporated into the host cell using standard cloning procedures known in the art, as described by Sambrook et al., Molecular Cloning: A Laboratory Manual, Second Edition, Cold Springs Laboratory, Cold Springs Harbor, N.Y. (1989), which is hereby incorporated by reference in its entirety.

In some embodiments, the host cell may comprise an endogenous RNA ligase. As described herein, the endogenous RNA ligase has the ability to catalyze the circularization of a ribonucleic acid molecule having a 5'-OH and a 2'-3'-cyclic phosphate. In accordance with this embodiment, the endogenous RNA ligase is RtcB.

Another aspect of the present application relates to an RNA-regulated fusion protein comprising a protein of interest and an RNA-regulated destabilization domain. Suitable proteins of interest and RNA-regulated destabilization domains are described in more detail supra.

In some embodiments, the protein of interest is a fluorescent protein, a bioluminescent protein, an enzyme, or a transcription factor. Suitable fluorescent proteins, bioluminescent proteins, enzymes, or transcription factors are described in more detail supra.

In some embodiments, the RNA-regulated destabilization domain has the consensus sequence of SEQ ID NO: 62 as follows: XXXXXXXXXXXXXXXXX, where X at position 1 can be S or A; X at position 2 can be G or A; X at position 3 can be P or A; X at position 4 can be R or K; X at position 5 can be P, A, I, Y, K, or R; X at position 6 can be R, K, V, or Y; X at position 7 can be G, A, or R; X at position 8 can be T or A; X at position 9 can be R or K; X at position 10 can be G or A; X at position 11 can be K or A; X at position 12 can be G or A; X at position 13 can be R or K; X at position 14 can be I or A; X at position 15 can be R, K, Y, or G; X at position 16 can be R, K, V, T, or Y; X at position 17 can be any amino acid but preferably R, G, E, S, or C; and x at position 18 is optional and can be any amino acid, but preferably G, E, O, N, D, or E.

In some embodiments the RNA-regulated destabilization domain has the sequence of tDeg (SEQ ID NO: 63) as follows: SGPRPRGTRGKGRIRRRG.

Exemplary RNA-regulated fusion proteins are identified in Table 8 below.

TABLE 8

Exemplary RNA-Regulated Fusion Proteins		
Vector	Sequence	SEQ ID NO:
(mNeonGreen) ₄ -tDeg	MVKGEEDNMASLPATHELHIFGSINGVDFDMVGQGTGNPNDGYE ELNLKSTKGDLQFSPWILVPHIGYGFHQYLPPDGMSPFQAAMVD GSGYQVHRTMQFEDGASLTVNRYTYEGSHI KGEAQVKGTGFPAD GPVMTNSLTAADWCRSKTYPNDKTIISTFKWSYTTGNGKRYRST ARTTYTFAKPMAANYLKNQPMYFRKTELKHSKTELNKFKEWKAF TDVGMGDELYKGHHMGTGSTGGVSKGEEDNMASLPATHELHI FGSINGVDFDMVGQGTGNPNDGYBEFLNLKSTKGDLQFSPWLVLPH IGYGFHQYLPPDGMSPFQAAMVDGSGYQVHRTMQFEDGASLTVN YRYTYEGSHIKGEAQVKGTGFADGPVMTNSLTAADWCRSKKTYP NDKTIISTFKWSYTTGNGKRYRSTARTTYTFAKPMAANYLKNQPM	80

TABLE 8-continued

Exemplary RNA-Regulated Fusion Proteins		
Vector	Sequence	SEQ ID NO:
	YVFRKTELHSKTELNFKEWQKAFTDVMGMDELYKSGLESSGGTG GSGGVSKGEEDNMASLPATHELHIFGSINGVDFDMVGQGTGNPND GYEELNLKSTKGLLQFSPWILVPHIGYGFHOYLPPDGMSPFQAA MVDGSGYQVHRTMQFEDGASLTVNRYTYEGSHIKGEAQVKGTGF PADGPVMTNSLTAADWCRSKKTYPNDKTIISTFKWSYTTGNGKRY RSTARTTTFAKPMAANYLKNQPMVFRKTELHSKTELNFKEWQ KAPTDVMGMDELYKGGSCTGGTASSGSGGGVSKEGEDNMASLPAT HELHIFGSINGVDFDMVGQGTGNPNDGYEEELNKSTKGLLQFSPW IILVPHIGYGFHOYLPPDGMSPFQAAAMDGSYQVHRTMQFEDGA SLTVNRYTYEGSHIKGEAQVKGTGFADGPVMTNSLTAADWCRS KKTPNDKTIISTFKWSYTTGNGKRYRSTARTTTFAKPMAANYL KNQPMVFRKTELHSKTELNFKEWQKAFTDVMGMDELYKGGRSG GGSGPRPRGTRGKGRRIRRG (GenBank Accession No. QEM23463.1 and GenBank Accession No. QEM23465.1, which are hereby incorporated by reference in their entirety)	
mNeonGreen-tDeg	MVSKEEDNMASLPATHELHIFGSINGVDFDMVGQGTGNPNDGYE EELNLKSTKGLLQFSPWILVPHIGYGFHOYLPPDGMSPFQAAAMD GSGYQVHRTMQFEDGASLTVNRYTYEGSHIKGEAQVKGTGFAD GPVMTNSLTAADWCRSKKTPNDKTIISTFKWSYTTGNGKRYRST ARTTTTFAKPMAANYLKNQPMVFRKTELHSKTELNFKEWQKAFT TDVMGMDELYKGGMGGGGGGSPRPRGTRGKGRRIRRG	81
mCherry-tDeg	MVSKEEDNMAMIEKFMRKVHMEGSVNGHEFEIEEGEGEGRPYEG TQATAKLVTKGGPLPFAWDLSPQFMYGSKAYVKHPADIPDYLKL SFPEGFKWERVMNMFEDGGVVTVTQDSSLQDGFEIYKVKLRGTNFP SDGPVMMQKKTMGWEASSERMYPEDGALKGEIKQLLKDGHHYDA EVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGR HSTGGMDELYKGSSGGGGSPRPRGTRGKGRRIRRG	82
NanoLuc-tDeg	MVFTLEDFTVGDWRQTAGYNLDQVLEQGGVSSLFQNLGVSVTPIQR IVLSEGENGLKIDIHVIIPYEGLSQDMQIEKIFKVVYPVDDHHF KVILHYGLVIDGVTTPNMIDYFGRPYEGIAVPDGKKITVTGILWN GNKIIDERLINPDGSLLFRVTINGVTGWLRCERILAGGSHMGGSG GGSGPRPRGTRGKGRRIRRG	83
EYFP-tDeg	MVSKEEELFTGVVPILVELGDVNGHKFSVSGECEGEGDATYGKLT KFICTIGKLPVPWPWTLLTIVYQVQCFSRYPDHMKQHDFFKSAMP EGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDG NILGHKLEYNNYNSHNVYIMADQKQNGIKVNFKIRHNIEDGSVOLA DHYQONTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLFV TAAGITLGMDELYKGTCACGTSGGRLDKSKVINSALELLNEVGIE GLTTRKLAQKLGVEQPTLYWHVKNKRALLDALAIEMLDRHHTFC PLEGESWQDFLRNNAKSFRCALLSHRDGAKVHLGTRPTEKQYETL ENQLAFLCQQGFSLENALYALSAVGHFTLGCVLEDQEHQVAKEER ETPTTDSMPPLLRAQIELFDHQGAEPAPLFGLELIICGLEKQLKC ESGSGSGTGGIGGSGGSPRPRGTRGKGRRIRRG	84
EGFP-TetR-tDeg	MVSKEEELFTGVVPILVELGDVNGHKFSVSGECEGEGDATYGKLT KFICTGKLPVPWPWTLLTIVYQVQCFSRYPDHMKQHDFFKSAMP EGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDG NILGHKLEYNNYNSHNVYIMADQKQNGIKVNFKIRHNIEDGSVOLA DHYQONTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLFV TAAGITLGMDELYKGTCACGTSGGRLDKSKVINSALELLNEVGIE GLTTRKLAQKLGVEQPTLYWHVKNKRALLDALAIEMLDRHHTFC PLEGESWQDFLRNNAKSFRCALLSHRDGAKVHLGTRPTEKQYETL ENQLAFLCQQGFSLENALYALSAVGHFTLGCVLEDQEHQVAKEER ETPTTDSMPPLLRAQIELFDHQGAEPAPLFGLELIICGLEKQLKC ESGSGSGTGGIGGSGGSPRPRGTRGKGRRIRRG	85
mCherry-TetR-tDeg	MVSKEEELFTGVVPILVELGDVNGHKFSVSGECEGEGDATYGKLT KFICTGKLPVPWPWTLLTIVYQVQCFSRYPDHMKQHDFFKSAMP EGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDG NILGHKLEYNNYNSHNVYIMADQKQNGIKVNFKIRHNIEDGSVOLA DHYQONTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLFV TAAGITLGMDELYKGTCACGTSGGRLDKSKVINSALELLNEVGIE GLTTRKLAQKLGVEQPTLYWHVKNKRALLDALAIEMLDRHHTFC PLEGESWQDFLRNNAKSFRCALLSHRDGAKVHLGTRPTEKQYETL ENQLAFLCQQGFSLENALYALSAVGHFTLGCVLEDQEHQVAKEER ETPTTDSMPPLLRAQIELFDHQGAEPAPLFGLELIICGLEKQLKC ESGSGSGTGGIGGSGGSPRPRGTRGKGRRIRRG	86
EGFP-EZH2-tDeg	MVSKEEELFTGVVPILVELGDVNGHKFSVSGECEGEGDATYGKLT KFICTGKLPVPWPWTLLTIVYQVQCFSRYPDHMKQHDFFKSAMP EGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDG NILGHKLEYNNYNSHNVYIMADQKQNGIKVNFKIRHNIEDGSVOLA DHYQONTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLFV TAAGITLGMDELYKGTCACGTSGGRLDKSKVINSALELLNEVGIE EYMLRQLKRFRRADEVKSMFSSNRQKILERTEILNQEWKQRRIQ	87

TABLE 8-continued

Exemplary RNA-Regulated Fusion Proteins		
Vector	Sequence	SEQ ID NO:
	<p>PVHILTSVSSLRGTRCECSVTSDLDFPTQVIPLKTLNAVAVSPIMY SWSPLQQNFMVEDETVLHNIPYMGDEVLDQDGTIEELIKNYDGK VHGDRRECGFINDEIFVFLVNLGQYNDDDDDGGDPPEEREEKQK DLEDHRDDKESRPPRKPSDKIFEAFIAISMFPDKGTAELKEKYKE LTECQLPGALPPCTPNIDGPNAKSQREQSLHSFHTLFCRRCFK YDCFLHPFHATPNTYKRKNTEALDNKPCGPQCYQHLEGAKEFAA ALTAERIKTPPKRPGRRGRRLPNNSRPSTPTINVLESKDTDSR REAGTETGGENNDKEEEEKKDETSSSEANSRCQTPIKMKPNIEP PENVEWSGAEASMPRVLIGTYYDNFCAIARLIGTKTCRQVYFRV KESSIAPAPAEDVDTPRKKRKHRLWAACRKLQKKGSSNH VYNQPCDHPQCDSSCPVCIAQNFCEKFCQCSSECQNRFPGCR CKAQCNTPQCPYLAVERCDPDLCLTCGAADHWDISKVSNCKNSI QRGSKHHLLAPSDVAGWGFIDPVQNEFISEYCGEIISQDEA DRRGKVYDKYMCFLFNLLNDFVVDACTRKGKIRFANHSVPNCY AKVMVNGDHRIGIFAKRAIQTGEELFFDYRYSQADALKYVGIER EMEIPGSGTGGIGGSGPRPRGTRKGRRIRRGG</p>	
mCherry-EZH2-tDeg	<p>MVSKEEEDNMAIKEFMRKVHMEGSVNGHEFEIEGEGERPYEG TQTAKLKVTKGGPLPFAWDLSPQFMYGSKAYVKHPADIPDYLKL SPEGFKWERVMNFEDGGVTVTQDSSLQDGEFIYVKVLRGTNFP SDGPMQKTMGWEASSERMPEDGALKGEIKQRLKLKDGGHYDA EVKTTYKAKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGR HSTGGMDELYKGTGACGTSGGMGTGKSEKGPGVCRKRVKSEYM RLRQLKFRRADEVKSMFSSNRQKILERTEILNQEWKQRRRIQPVH ILTSVSSLRGTRCESVTSDDLDFPTQVILPLKTLNAVAVSPIMYSWS PLQONFMEDETVLHNIPYMGDEVLDQDGTIEELIKNYDGKVHG DRECGFINDEIFVFLVNLGQYNDDDDDGGDPPEEREEKQKDL DHRDDKESRPPRKPSDKIFEAFIAISMFPDKGTAEEELKEKYKELTE QQLPGALPPECTPNIDGPNAKSQREQSLHSFHTLFCRRCFKYDC FLHPFHATPNTYKRKNTEALDNKPCGPQCYQHLEGAKEFAAALT AERIKTPPKRPGRRGRRLPNNSRPSTPTINVLESKDTDSREA GTBTGENNDKEEEEKKDETSSSEANSRCQTPIKMKPNIEPEN VEWSGAEASMPRVLIGTYYDNFCAIARLIGTKTCRQVYFRV SIIAPAPAEDVDTPRKKRKHRLWAACRKLQKKGSSNHVYN YQPCDHPQCDSSCPVCIAQNFCEKFCQCSSECQNRFPGCRKA QCNTPQCPYLAVERCDPDLCLTCGAADHWDISKVSNCKNSIQRG SKKHLLLAPSDVAGWGFIDPVQNEFISEYCGEIISQDEADRR GKVIDKYMCFLFNLLNDFVVDACTRKGKIRFANHSVPNCYAKV MMVNGDHRIGIFAKRAIQTGEELFFDYRYSQADALKYVGIEREME IPGSGTGGIGGSGPRPRGTRKGRRIRRGG</p>	88
EGFP-NFKB-tDeg	<p>MVSKEEELFTGVVPILVLDGDVNGHKFSVSGEGERPYEG KFICTIGKLPVPWPWTILTGYVQCFSPDHMKQHDFFKSAMP EGVYQERTIFFKDKDGNYKTRAEVKPEGDTLVRIEKGIDFKEDG NILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVOLA DHYQONTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLFV TAAGITLGMDELYKGSSGSSGGSSGGSGTGAEDDPYLGRPEQMFH LDPSTLTHITFNPEVFQPFQMLPTADGPYLQILEQPKQRGFRFRYV CEGPHGGGLPGASSEKNKKSYPQVKICNYVGPVAKVIVQLVTNGKN IHLHAHSLVGKHCEDGICCTVTAGPKDMVVGFAVLGILHVTKKVF ETLEARTEACIRGYNPGLLWHPDLAYLQAEGGGDRQLGDREKEL IRQALQQTKEMDLSVVRMLFTAFLPDTSTGSFTRRLEPVVSDAIY DSKAPNASNLKIVRMDRTAGCTVGGEEIYLLCDKVQKDDIQIRFY EEEEENGWEGFGDFSPTDVHQFAIVPKTPKYKDINITKPKASVF VQLRKSDELTESEPKPFLYYPEIKDKEEVQRKRQKLMNPNSDSFG GGSGAGAGGGGMPGSSGGGGTGSTGPVYSPPHYGFTYGGITFH PGTTSNAGMKHGTMDTESKKDPEGCDKSDDKNTVNLFGKDPRGS LSGGTGGSGPGRGTRKGRRIRRGG</p>	89
mCherry-NFKB-tDeg	<p>MVSKEEEDNMAIKEFMRKVHMEGSVNGHEFEIEGEGERPYEG TQTAKLKVTKGGPLPFAWDLSPQFMYGSKAYVKHPADIPDYLKL SPEGFKWERVMNFEDGGVTVTQDSSLQDGEFIYVKVLRGTNFP SDGPMQKTMGWEASSERMPEDGALKGEIKQRLKLKDGGHYDA EVKTTYKAKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGR HSTGGMDELYKGGSSGSSGGSGTGAEDDPYLGRPEQMFHLD SLHTTIFNPEVFQPFQMLPTADGPYLQILEQPKQRGFRFRYCEG PSHGGGLPGASSEKNKKSYPQVKICNYVGPVAKVIVQLVTNGKNIHL HAHSLVGKHCEDGICCTVAGPKDMVVGFAVLGILHVTKKVFETL EARTEACIRGYNPGLLWHPDLAYLQAEGGGDRQLGDREKELIRQ AALQQTKEMDLSVVRMLFTAFLPDTSTGSFTRRLEPVVSDAIYDSK APNASNLKIVRMDRTAGCTVGGEEIYLLCDKVQKDDIQIRFYEEE ENGGWEGFGDFSPTDVHQFAIVPKTPKYKDINITKPKASVFQV RRKSDELTESEPKPFLYYPEIKDKEEVQRKRQKLMNPNSDSFGGG GAGAGGGGMPGSSGGGGTGSTGPVYSPPHYGFTYGGITFH TKSNAGMKHGTMDTESKKDPEGCDKSDDKNTVNLFGKDPRGSLSG</p>	90

TABLE 8-continued

Exemplary RNA-Regulated Fusion Proteins		
Vector	Sequence	SEQ ID NO:
EGFP-TurboID-tDeg	MVSKEEELFTGVPILVELGDGVNGHKFSVSGECEGEGDATYGKLTL KFICTIGKLPVPWPTLVTTLTGYVQCFSRYPDHMKQHDFFKSAMP EGVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDG NILGHKLEYNNYNSHNVYIMADQKQNGIKVNFKIRHNIEDGSVOLA DHYQQNTPIGDGPVLLPDNHYLSTQSLSKDPNEKRDHMVILLEFV TAAGITLGMDELYKGTGACGTSGGMKDNTVPLKLIAALLANGEFHS GEQLGETLGMMSRAINKHIQTLRDWGVDVFTVPGKGYSLPEPIPL LNAKQILQGLDGGSVAVLPPVDS TNQYLLDRIGELKSGDACIAEY QQAGRSRGRKWFSPFGANLYLSMFWRLKRGPAAILGPVIGIVM AEALRKLGADKVRVKWPNDLYLQDRKLAGILVELAGITGDAAQIV IGAGINVAMRRVEESVVNQGWITLQEAGINLDRNTLAATLIRELR AALELFEQEGLAPYLPRWEKLDNFINRPVKLIIGDKEIFGSRGI DKQGALLLEQDGVIKPWMGGEISLRAEKGSGTGGTGGSGPRPRG TRGKGRRIRRRG	91
EGFP-APEX-tDeg	MVSKEEELFTGVPILVELGDGVNGHKFSVSGECEGEGDATYGKLTL KFICTGKLPVPWPTLVTTLTGYVQCFSRYPDHMKQHDFFKSAMP EGVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDG NILGHKLEYNNYNSHNVYIMADQKQNGIKVNFKIRHNIEDGSVOLA DHYQQNTPIGDGPVLLPDNHYLSTQSLSKDPNEKRDHMVILLEFV TAAGITLGMDELYKGTGACGTSGKSPTVSADYQDAVEKAKKKLR GFIAEKRCAPLMLRLAFLPHSAGTFDKGKTGTGGPFGTIKHPAELAHS ANNGLDIAVRLLEPLKAEPILSYADFYQLAGVVAEVVTGGPKVP FHPGREDKPEPPPPEGRLDPPTKGSDHLDVFGKAMGLTDQDIVAL SGGHTIGAAHKERSGFEGPWTSNPLIFDNNSYFTELLSGEKEGLLQ LPSDKALLSDPVFRPLVDKYAADEDAFFADYAEAHQKLSELGFAD AGSGTGGTGGSGPRPRGTRGKGRRIRRRG	92

Yet another aspect of the disclosure relates to a molecular complex comprising an RNA-regulated fusion protein comprising (i) a protein of interest and (ii) an RNA-regulated destabilization domain and an RNA aptamer bound specifically to the RNA-regulated destabilization domain.

In some embodiments, the protein of interest is a fluorescent protein, a bioluminescent protein, an enzyme, or a transcription factor. Suitable fluorescent proteins, bioluminescent proteins, enzymes, and transcription factors are described in detail *supra*.

In some embodiments, the RNA-regulated destabilization domain has the sequence of SEQ ID NO: 62, where X at position 1 is S or A; X at position 2 is G or A; X at position 3 is P or A; X at position 4 is R or K; X at position 5 is P, A, I, Y, K, or R; X at position 6 is R, K, V, or Y; X at position 7 is G, A, or R; X at position 8 is T or A; X at position 9 is R or K; X at position 10 is G or A; X at position 11 is K or A; X at position 12 is G or A; X at position 13 is R or K; X at position 14 is I or A; X at position 15 is R, K, Y, or G; X at position 16 is R, K, V, T, or Y; X at position 17 is any amino acid; and x at position 18 is optional and can be any amino acid. For example, the RNA-regulated destabilization domain may be tDeg (SEQ ID NO: 63).

Suitable RNA aptamer sequences are described in detail *supra*. In some embodiments, the RNA aptamer comprises the consensus sequence of SEQ ID NO: 56, SEQ ID NO: 58, or SEQ ID NO: 60, wherein N can be A, C, G, or U; S can be C or G; H can be A, C, or U; Y can be C or U; W can be A or U; B can be C, G, or U; M can be A or C; and D can be A, G, or U. For example, the RNA aptamer may comprise the sequence of wild-type TAR RNA (SEQ ID NO: 57), TAR Variant-1 (SEQ ID NO: 59), or TAR Variant-2 (Pepper; SEQ ID NO: 61).

Additional exemplary RNA aptamers may be selected from the group consisting of SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 72, and SEQ ID NO: 73.

Some embodiments of the present application relate to a host cell comprising a molecular complex described herein (i.e., a molecular complex comprising an RNA-regulated fusion protein and an RNA aptamer bound specifically to the RNA-regulated destabilization domain). Suitable host cells are described in detail *supra*.

In some embodiments, the host cell is a mammalian cell. As described herein above, suitable mammalian cells include, without limitation, rodent cells (i.e., mouse or rat cells), rabbit cells, guinea pig cells, feline cells, canine cells, porcine cells, equine cells, bovine cell, ovine cells, monkey cells, non-human primate, or human cells. In some embodiments, the host cell is a human cell.

Another aspect of the invention relates to a method of imaging RNA in a cell. This method involves providing a first vector encoding an RNA-regulated fusion protein, wherein the RNA-regulated fusion protein comprises a fluorescent protein, a bioluminescent protein, or an enzyme fused to an RNA-regulated destabilization domain; providing second vector encoding an RNA molecule comprising (i) an RNA sequence of interest and (ii) an RNA aptamer sequence, where the RNA-regulated destabilization domain specifically binds to the RNA aptamer sequence; transfecting a host cell with the first vector and the second vector; and imaging said contacted cells.

Suitable vectors for carrying out the methods of imaging RNA in a cell are described in more detail *supra* and include, e.g., a plasmid (e.g., an expression vector) and a viral vector (e.g., a lentiviral or adenoviral vector).

Suitable RNA-regulated fusion proteins for carrying out the methods of the present application are described in more detail *supra*. In some embodiments of the methods described herein, the RNA-regulated fusion protein is a fluorescent protein selected from the group consisting of Green Fluorescent Protein, Enhanced Green Fluorescent Protein

(EGFP), Enhanced Yellow Fluorescent Protein (EYFP), Venus, mVenus, Citrine, mCitrine, Cerulean, mCerulean, Orange Fluorescent Protein (OFP), mNeonGreen, mox-NeonGreen, mCherry, mTagBFP, Venus, mVenus, mTurquoise, mScarlet, mWasabi, mOrange, and dTomato.

In other embodiments of the methods described herein, the RNA-regulated fusion protein is a bioluminescent protein selected from the group consisting of luciferase, β -galactosidase, β -lactamase, peroxidase, alkaline phosphatase, β -glucuronidase, and β -glucosidase. In some embodiments, the bioluminescent protein is a luciferase selected from the group consisting of Nanoluc luciferase (Nluc), Firefly luciferase, and *Renilla* luciferase (Rluc).

In further embodiments of the methods described herein, the RNA-regulated fusion protein is an enzyme, wherein the enzyme is a biotin ligase. Suitable biotin ligases are described in detail supra and include, e.g., TurboID, miniTurbo, or *E. coli* BirA.

As described in more detail supra, the RNA-regulated destabilization domain may comprise a bifunctional peptide having a lentiviral transactivator of transcription (Tat) peptide and a degron peptide. Lentiviral transactivator of transcription (Tat) peptides and a degron peptides are described in more detail supra.

In some embodiments of the methods described herein, the RNA-regulated destabilization domain comprises the consensus sequence of SEQ ID NO: 62, where X at position 1 is S or A; X at position 2 is G or A; X at position 3 is P or A; X at position 4 is R or K; X at position 5 is P, A, I, Y, K, or R; X at position 6 is R, K, V, or Y; X at position 7 is G, A, or R; X at position 8 is T or A; X at position 9 is R or K; X at position 10 is G or A; X at position 11 is K or A; X at position 12 is G or A; X at position 13 is R or K; X at position 14 is I or A; X at position 15 is R, K, Y, or G; X at position 16 is R, K, V, T, or Y; X at position 17 is any amino acid; and x at position 18 is optional and can be any amino acid. Thus, in some embodiments, the RNA-regulated destabilization domain is tDeg (SEQ ID NO: 63).

As used herein, an RNA of interest is an RNA molecule that is desired and/or is being assessed. The RNA of interest may be a messenger RNA (mRNA) or a noncoding RNA (ncRNA). A messenger RNA or "mRNA" refers to a single-stranded RNA molecule that specifies the amino acid sequence of a protein. The mRNA molecule may comprise a 5' untranslated region (5' UTR), a coding region, and a 3' untranslated region (3' UTR). A 5' UTR is an untranslated nucleotide segment in an RNA molecule immediately preceding the AUG start codon. A 3' UTR is an untranslated nucleotide segment in an RNA molecule immediately following the translation termination codon.

In some embodiments, the RNA of interest is an mRNA and the RNA aptamer is located within a coding region of the mRNA. In other embodiment, the RNA of interest is a mRNA and the RNA aptamer is located upstream of the 5' UTR, within the 5' UTR, within the 3' UTR, or downstream of the 3' UTR.

In other embodiments, the RNA of interest is a noncoding RNA (ncRNA). As described herein, a noncoding RNA refers to a functional RNA molecule that is not translated into a protein. The RNA of interest may be a noncoding RNA selected from the group consisting of ribosomal RNA (rRNA), transfer RNA (tRNA), heterogeneous nuclear RNA (hnRNA), small cytoplasmic RNA (scRNA), small nuclear (snRNA), small nucleolar (snoRNA), ribozymes, and regulatory RNA (e.g., siRNA, miRNA, microRNA, etc.).

In some embodiments, the RNA of interest is an artificial, engineered synthetic RNA.

Suitable RNA aptamers are described in detail supra. In some embodiments of the methods described herein, the RNA aptamer comprises the consensus sequence of SEQ ID NO: 56, SEQ ID NO: 58, or SEQ ID NO: 60, where N can be A, C, G, or U; S can be C or G; H can be A, C, or U; Y can be C or U; W can be A or U; B can be C, G, or U; M can be A or C; and D can be A, G, or U. For example, the RNA aptamer may comprise the sequence of wild-type TAR RNA (SEQ ID NO: 57), TAR Variant-1 (SEQ ID NO: 59), or TAR Variant-2 (Pepper; SEQ ID NO: 61). In some embodiments of the methods described herein, the RNA aptamer comprises the sequence of SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 72, or SEQ ID NO: 73.

Methods of transfecting a host cell are well known in the art and described in more detail supra. According to some embodiments of the methods described herein, transfecting the host cell with the first vector and the second vector is carried out simultaneously. In other embodiments, transfecting the host cell with the first vector and the second vector is carried out sequentially.

Methods of imaging cells are well known in the art. In some embodiments, imaging said transfected cells is carried out by fluorescence microscopy or imaging flow cytometry (see, e.g., Wu et al., "Live Imaging of mRNA Using RNA-Stabilized Fluorogenic Proteins," *Nature Methods* 16:862-565 (2019) and Wu & Jaffrey, Live Imaging of mRNA Using Pepper RNA-Stabilized Fluorogenic Proteins," *Nature Methods*, DOI: 10.1101/rs.2.11494/v1 (2019), which are hereby incorporated by reference in their entirety).

Yet another aspect of the invention relates to a method of imaging RNA in a cell. This method involves providing a vector encoding an RNA-regulated fusion protein, where the RNA-regulated fusion protein comprises a fluorescent protein, a bioluminescent protein, or an enzyme fused to an RNA-regulated destabilization domain; transfecting a host cell with the first vector; contacting said transfected cell with an RNA molecule comprising (i) an RNA sequence of interest and (ii) an RNA aptamer sequence, where the RNA-regulated destabilization domain specifically binds to the RNA aptamer sequence; and imaging said contacted cells.

Suitable vectors for carrying out the methods of imaging RNA in a cell are described in more detail supra and include, e.g., a plasmid (e.g., an expression vector) and a viral vector (e.g., a lentiviral or adenoviral vector).

Suitable RNA-regulated fusion proteins for carrying out the methods of the present application are described in more detail supra. In some embodiments of the methods described herein, the RNA-regulated fusion protein is a fluorescent protein selected from the group consisting of Green Fluorescent Protein, Enhanced Green Fluorescent Protein (EGFP), Enhanced Yellow Fluorescent Protein (EYFP), Venus, mVenus, Citrine, mCitrine, Cerulean, mCerulean, Orange Fluorescent Protein (OFP), mNeonGreen, mox-NeonGreen, mCherry, mTagBFP, Venus, mVenus, mTurquoise, mScarlet, mWasabi, mOrange, and dTomato.

In other embodiments of the methods described herein, the RNA-regulated fusion protein is a bioluminescent protein selected from the group consisting of luciferase, β -galactosidase, β -lactamase, peroxidase, alkaline phosphatase, β -glucuronidase, and β -glucosidase. In some embodiments, the bioluminescent protein is a luciferase selected from the group consisting of Nanoluc luciferase (Nluc), Firefly luciferase, and *Renilla* luciferase (Rluc).

In further embodiments of the methods described herein, the RNA-regulated fusion protein is an enzyme, wherein the

enzyme is a biotin ligase. Suitable biotin ligases are described in detail supra and include, e.g., TurboID, miniTurbo, or *E. coli* BirA.

As described in more detail supra, the RNA-regulated destabilization domain may comprise a bifunctional peptide having a lentiviral transactivator of transcription (Tat) peptide and a degron peptide. Lentiviral transactivator of transcription (Tat) peptides and a degron peptides are described in more detail supra.

In some embodiments of the methods described herein, the RNA-regulated destabilization domain comprises the consensus sequence of SEQ ID NO: 62, where X at position 1 is S or A; X at position 2 is G or A; X at position 3 is P or A; X at position 4 is R or K; X at position 5 is P, A, I, Y, K, or R; X at position 6 is R, K, V, or Y; X at position 7 is G, A, or R; X at position 8 is T or A; X at position 9 is R or K; X at position 10 is G or A; X at position 11 is K or A; X at position 12 is G or A; X at position 13 is R or K; X at position 14 is I or A; X at position 15 is R, K, Y, or G; X at position 16 is R, K, V, T, or Y; X at position 17 is any amino acid; and x at position 18 is optional and can be any amino acid. Thus, in some embodiments, the RNA-regulated destabilization domain is tDeg (SEQ ID NO: 63).

In some embodiments, the RNA of interest is a mRNA and the RNA aptamer is located within a coding region of the mRNA. In other embodiment, the RNA of interest is a mRNA and the RNA aptamer is located upstream of the 5' UTR, within the 5' UTR, within the 3' UTR, or downstream of the 3' UTR.

In other embodiments, the RNA of interest is a noncoding RNA (ncRNA). As described herein, the term "noncoding RNA" refers to a functional RNA molecule that is not translated into a protein. The RNA of interest may be a noncoding RNA selected from the group consisting of ribosomal RNA (rRNA), transfer RNA (tRNA), heterogeneous nuclear RNA (hnRNA), small cytoplasmic RNA (scRNA), small nuclear (snRNA), small nucleolar (snoRNA), ribozymes, and regulatory RNA (e.g., siRNA, miRNA, microRNA, etc.).

Suitable RNA aptamers are described in detail supra. In some embodiments of the methods described herein, the RNA aptamer comprises the consensus sequence of SEQ ID NO: 56, SEQ ID NO: 58, or SEQ ID NO: 60, wherein N can be A, C, G, or U; S can be C or G; H can be A, C, or U; Y can be C or U; W can be A or U; B can be C, G, or U; M can be A or C; and D can be A, G, or U. For example, the RNA aptamer may comprise the sequence of wild-type TAR RNA (SEQ ID NO: 57), TAR Variant-1 (SEQ ID NO: 59), or TAR Variant-2 (Pepper; SEQ ID NO: 61). In some embodiments of the methods described herein, the RNA aptamer comprises the sequence of SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 72, or SEQ ID NO: 73.

The RNA molecule comprising the (i) RNA sequence of interest and (ii) the RNA aptamer sequence may be a circular RNA molecule or a linear RNA molecule.

Methods of transfecting a host cell are well known in the art and described in more detail supra.

Contacting the transfected cell may be carried out by allowing the RNA molecule comprising the (i) RNA sequence of interest and (ii) the RNA aptamer sequence may be a circular RNA molecule or a linear RNA molecule to diffuse into the cell.

Methods of imaging cells are well known in the art. In some embodiments, imaging said contacted cells is carried out by fluorescence microscopy or imaging flow cytometry.

A further aspect of the invention relates to a method of selectively modifying an RNA-binding protein. This method

involves providing a first expression vector encoding a RNA-regulated fusion protein, where the RNA-regulated fusion protein comprises an enzyme fused to an RNA-regulated destabilization domain; providing a second expression vector encoding (i) an RNA sequence of interest and (ii) an RNA aptamer sequence, where the RNA-regulated destabilization domain specifically binds to the RNA aptamer sequences; transfecting a host cell with the first and second expression vectors; and allowing the enzyme to be expressed, wherein the expressed enzyme selectively modifies a protein that binds to the RNA sequence of interest.

Suitable enzymes are described in more detail supra. In some embodiments, the enzyme is selected from the group consisting of a ligase, a peroxidase, and a methyltransferase.

In some embodiments of the methods described herein, the enzyme is a biotin ligase selected from the group consisting of TurboID, miniTurbo, and *E. coli* BirA.

In some embodiments of the methods described herein, the enzyme is a peroxidase selected from the group consisting of an ascorbate peroxidase and a horseradish peroxidase. The ascorbate peroxidase may be APEX2.

As described in more detail supra, the RNA-regulated destabilization domain may comprise a bifunctional peptide having a lentiviral transactivator of transcription (Tat) peptide and a degron peptide. Lentiviral transactivator of transcription (Tat) peptides and a degron peptides are described in more detail supra.

In some embodiments of the methods described herein, the RNA-regulated destabilization domain comprises the consensus sequence of SEQ ID NO: 62, where X at position 1 is S or A; X at position 2 is G or A; X at position 3 is P or A; X at position 4 is R or K; X at position 5 is P, A, I, Y, K, or R; X at position 6 is R, K, V, or Y; X at position 7 is G, A, or R; X at position 8 is T or A; X at position 9 is R or K; X at position 10 is G or A; X at position 11 is K or A; X at position 12 is G or A; X at position 13 is R or K; X at position 14 is I or A; X at position 15 is R, K, Y, or G; X at position 16 is R, K, V, T, or Y; X at position 17 is any amino acid; and x at position 18 is optional and can be any amino acid. Thus, in some embodiments, the RNA-regulated destabilization domain is tDeg (SEQ ID NO: 63).

In some embodiments, the RNA of interest is a mRNA and the RNA aptamer is located within a coding region of the mRNA. In other embodiment, the RNA of interest is a mRNA and the RNA aptamer is located upstream of the 5' UTR, within the 5' UTR, within the 3' UTR, or downstream of the 3' UTR.

In other embodiments, the RNA of interest is a noncoding RNA (ncRNA). As described herein, the term "noncoding RNA" refers to a functional RNA molecule that is not translated into a protein. The RNA of interest may be a noncoding RNA selected from the group consisting of ribosomal RNA (rRNA), transfer RNA (tRNA), heterogeneous nuclear RNA (hnRNA), small cytoplasmic RNA (scRNA), small nuclear (snRNA), small nucleolar (snoRNA), ribozymes, and regulatory RNA (e.g., siRNA, miRNA, microRNA, etc.).

Suitable RNA aptamers are described in detail supra. In some embodiments of the methods described herein, the RNA aptamer comprises the consensus sequence of SEQ ID NO: 56, SEQ ID NO: 58, or SEQ ID NO: 60, wherein N can be A, C, G, or U; S can be C or G; H can be A, C, or U; Y can be C or U; W can be A or U; B can be C, G, or U; M can be A or C; and D can be A, G, or U. For example, the RNA aptamer may comprise the sequence of wild-type TAR RNA (SEQ ID NO: 57), TAR Variant-1 (SEQ ID NO: 59), or TAR Variant-2 (Pepper; SEQ ID NO: 61). In some

99

embodiments of the methods described herein, the RNA aptamer comprises the sequence of SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 72, or SEQ ID NO: 73.

In some embodiments of the methods of selectively modifying an RNA-binding protein described herein, the method further involves identifying a protein that is selectively modified by the enzyme within the transfected cells. See, e.g., Ramanathan et al., "RNA-Protein Interaction Detection in Living Cells," *Nature Methods* 15:207-212 (2018), which is hereby incorporated by reference in its entirety.

Another aspect relates to a method of regulating expression of an RNA-stabilized protein of interest. This method involves providing a first vector encoding an RNA-regulated fusion protein, where the RNA-regulated fusion protein comprises a protein of interest fused to an RNA-regulated destabilization domain; providing a second vector encoding an RNA aptamer sequence, where the RNA-regulated destabilization domain specifically binds to the RNA aptamer sequence; providing a host cell comprising a functional ubiquitination system; transfecting the host cell with the first and second expression vectors; and expressing the first and second expression vectors within the host cell, where said expressing the first and second expression vectors regulates proteomic stability of the RNA-regulated fusion protein; and where, in the absence of any expressed RNA aptamer sequence in the host cell, the RNA-regulated destabilization domain promotes degradation of the RNA-regulated fusion protein by the ubiquitination system; and where the RNA-regulated fusion protein is stabilized by the expressed RNA aptamer sequence.

Another aspect of the invention relates to a method of regulating expression of an RNA-stabilized protein of interest. This method involves providing a first vector encoding an RNA-regulated fusion protein, where the RNA-regulated fusion protein comprises a protein of interest fused to an RNA-regulated destabilization domain; providing a second vector encoding an RNA aptamer sequence, where the RNA-regulated destabilization domain specifically binds to the RNA aptamer sequence; providing a mammalian cell lysate or solution comprising (i) a ubiquitin ligase, (ii) proteosomal degradation machinery, (iii) transcriptional machinery, and (iv) translational machinery; contacting the mammalian cell lysate or solution with the first and second expression vectors; and expressing the first and second expression vectors, where said expressing the first and second expression vectors regulates proteomic stability of the RNA-regulated fusion protein; and where, in the absence of any expressed RNA aptamer sequence in the cell lysate or solution, the RNA-regulated destabilization domain promotes degradation of the RNA-regulated fusion protein by the proteosomal degradation system; and where the RNA-regulated fusion protein is stabilized by the expressed RNA aptamer sequence.

Suitable proteins of interest for use in the methods described herein are described in more detail supra. In some embodiments, the protein of interest is a fluorescent protein, a bioluminescent protein, an enzyme, or a transcription factor. In other embodiments, the protein of interest is selected from the group consisting of a G-protein coupled receptor (GPCR), a nuclear receptor, a voltage gated ion channel, a ligand gated channel, a receptor tyrosine kinase, a growth factor, a phosphatase, a protein kinase, a viral regulator, a bacterial cell division protein, a scaffold protein, a DNA repair protein, a cytoskeletal protein, a ribosome, a histone deacetylase, an apoptosis regulator, a chaperone

100

protein, a kinase, a phosphorylase, a phosphatase, deacetylase, a cytoskeletal protein (e.g., myosin, actin, dynein, kinesin, and tubulin).

Suitable expression vectors encoding RNA-regulated fusion proteins and vectors encoding an RNA aptamer sequence for use in the methods described herein are described in detail supra and include, e.g., a plasmid (e.g., an expression vector) and a viral vector (e.g., a lentiviral or adenoviral vector).

As described in more detail supra, the RNA-regulated destabilization domain may comprise a bifunctional peptide having a lentiviral transactivator of transcription (Tat) peptide and a degron peptide. Lentiviral transactivator of transcription (Tat) peptides and a degron peptides are described in more detail supra.

In some embodiments of the methods described herein, the RNA-regulated destabilization domain comprises the consensus sequence of SEQ ID NO: 62, where X at position 1 is S or A; X at position 2 is G or A; X at position 3 is P or A; X at position 4 is R or K; X at position 5 is P, A, I, Y, K, or R; X at position 6 is R, K, V, or Y; X at position 7 is G, A, or R; X at position 8 is T or A; X at position 9 is R or K; X at position 10 is G or A; X at position 11 is K or A; X at position 12 is G or A; X at position 13 is R or K; X at position 14 is I or A; X at position 15 is R, K, Y, or G; X at position 16 is R, K, V, T, or Y; X at position 17 is any amino acid; and x at position 18 is optional and can be any amino acid. Thus, in some embodiments, the RNA-regulated destabilization domain is tDeg (SEQ ID NO: 63).

Suitable RNA aptamer sequences for use in the methods described herein are described in more detail supra. In some embodiments, the RNA aptamer comprises the consensus sequence of SEQ ID NO: 56, SEQ ID NO: 58, or SEQ ID NO: 60, wherein N can be A, C, G, or U; S can be C or G; H can be A, C, or U; Y can be C or U; W can be A or U; B can be C, G, or U; M can be A or C; and D can be A, G, or U. For example, the RNA aptamer may comprises the sequence of wild-type TAR RNA (SEQ ID NO: 57), TAR Variant-1 (SEQ ID NO: 59), or TAR Variant-2 (Pepper; SEQ ID NO: 61). In other embodiments, the RNA aptamer comprises the sequence of SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 72, or SEQ ID NO: 73.

Suitable host cells for use in the methods described herein are described in more detail supra. In some embodiments, the host cell is a mammalian cell.

Suitable mammalian cell lysates include, for example and without limitation, human cell lysates, non-human primate cell lysates, feline cell lysates, canine cell lysates, ovine cell lysates, hircine cell lysates, bovine cell lysates, equine cell lysates, porcine cell lysates, leporine cell lysates, and murine cell lysates.

Suitable solutions comprising (i) a ubiquitin ligase, (ii) proteosomal degradation machinery, (iii) transcriptional machinery, and (iv) translational machinery are well known in the art.

Exemplary ubiquitin ligases include, without limitation, ubiquitin E3 ligases (Li et al., "Genome-Wide and Functional Annotation of Human E3 Ubiquitin Ligases Identifies MULAN, A Mitochondrial E3 that Regulates the Organelle's Dynamics and Signaling," *PLoS One* 3(1):e1487 (2008); Berndsen & Wolberger, "New Insights into Ubiquitin E3 Ligase Mechanism," *Nat. Struct. Mol. Biol.* 21(4): 301-307 (2014), which are hereby incorporated by reference in their entirety). In some embodiments, the ubiquitin E3 ligase is selected form the group consisting of Really Interesting New Gene/U-box (RING) E3 ligase, Homologous to E6AP C-Terminus (HECT) E3 ligase, and RING

101

between RING (RBR) E3 ligase (see, e.g., Metzger et al., "RING-Type E3 Ligases: Master Manipulators of E2 Ubiquitin-Conjugating Enzymes and Ubiquitination," *Biochim. Biophys. Acta.* 1843(1):47-60 (2014); Rotin & Kumar, "Physiological Functions of the HECT Family of Ubiquitin Ligases," *Nat. Rev. Mol. Cell. Biol.* 10(6):398-409 (2009); Sluimer & Distel, "Regulating the Human HECT E3 Ligases," *Cell Mol. Life Sci.* 75(17):3121-3141 (2018); Reiter & Klevit, "Characterization of RING-Between-RING E3 Ubiquitin Transfer Mechanisms," *Methods. Mol. Biol.* 1844:3-17 (2018); and Dove & Klevit, "RING-Between-RING E3 Ligases: Emerging Themes Amid the Variations," *J. Mol. Biol.* 429(22):3363-3375 (2017), which are hereby incorporated by reference in their entirety).

Methods of transfecting cells are well known in the art and described in more detail supra.

Another aspect of the present application relates to a treatment method. This method involves contacting a cell with an RNA aptamer, where upon said contacting, the aptamer interacts with an RNA-regulated destabilization domain fused to a protein of interest in the cell to stabilize the protein of interest in the cell.

According to one embodiment, this and other treatment methods described herein are effective to treat a cell, e.g., a cell under a stress or disease condition. Exemplary cell stress conditions may include, without limitation, exposure to a toxin; exposure to chemotherapeutic agents, irradiation, or environmental genotoxic agents such as polycyclic hydrocarbons or ultraviolet (UV) light; exposure of cells to conditions such as glucose starvation, inhibition of protein glycosylation, disturbance of Ca²⁺ homeostasis and oxygen; exposure to elevated temperatures, oxidative stress, or heavy metals; and exposures to a pathological disease state (e.g., diabetes, Parkinson's disease, cardiovascular disease (e.g., myocardial infarction, end-stage heart failure, arrhythmogenic right ventricular dysplasia, and Adriamycin-induced cardiomyopathy), and various cancers (Fulda et al., "Cellular Stress Responses: Cell Survival and Cell Death," *Int. J. Cell Biol.* (2010), which is hereby incorporated by reference in its entirety).

In some embodiments, contacting a cell with an RNA molecule (aptamer) of the present application involves introducing an RNA molecule into a cell. Suitable methods of introducing RNA molecules into cells are well known in the art and include, but are not limited to, the use of transfection reagents, electroporation, microinjection, or via viruses.

The cell may be a eukaryotic cell. Exemplary eukaryotic cells include a yeast cell, an insect cell, a fungal cell, a plant cell, and an animal cell (e.g., a mammalian cell). Suitable mammalian cells include, for example without limitation, human, non-human primate, cat, dog, sheep, goat, cow, horse, pig, rabbit, and rodent cells.

The RNA molecule of the present invention may be isolated or present in in vitro conditions for extracellular expression and/or processing. According to this embodiment, the RNA molecule is contacted by an RNAligase (e.g., RtcB) in vitro, purified, circularized, and then the circularized RNA molecule is administered to a cell or subject for treatment.

Treating cells also includes treating the organism in which the cells reside. Thus, by this and the other treatment methods of the present invention, it is contemplated that treatment of a cell includes treatment of a subject in which the cell resides.

In some embodiments, the treatment method further comprises introducing the protein of interest into the cell prior to said contacting.

102

In some embodiments, the cell is in a patient.

In some embodiments, introducing is carried out by any one or more of injecting mRNA encoding for the protein of interest into the patient, injecting a plasmid encoding for the protein of interest into the patient, injecting the protein of interest into the patient, or systemically delivering the protein of interest into the patient.

In some embodiments, the patient is a human.

Another aspect of the present application relates to a treatment method. This method involves contacting a cell with a vector according to the present application under conditions effective to express an RNA molecule as described herein to treat the cell.

A further aspect of the present application relates to a kit comprising a vector encoding an RNA-regulated destabilization domain and a vector encoding an RNA aptamer that specifically binds to said RNA-regulated destabilization domain. Suitable RNA-regulated destabilization domains and RNA aptamers are described in detail supra.

In some embodiments, the kit comprises a vector encoding tDeg and vector encoding a Pepper aptamer.

EXAMPLES

The following examples are provided to illustrate embodiments of the present invention but they are by no means intended to limit its scope.

Materials and Methods for Examples 1-5

General methods and materials. Single stranded synthetic DNA oligonucleotides for PCR were purchased from Integrated DNA Technologies. Phusion® High-Fidelity DNA Polymerase (NEB M0530) was used for routine PCR amplifications. PCR products were run on 1% TAE agarose gels. PCR products with correct size were then excised and purified with the Qiaquick Gel Extraction kit (Qiagen 28704). Restriction endonucleases used for restriction digest were purchased from New England Biolabs, and used according to the manufacturer's recommended protocol. DNA ligation reactions were carried out using the Quick 40 Ligation™ Kit (NEB M2200L). DNA plasmids were propagated using chemically competent *E. coli* (Agilent 200314). The QIAprep Spin Plasmid Miniprep Kit (Qiagen 27106) was used for DNA plasmid extraction and purification from *E. coli*. DNA sequencing (GENEWIZ) was used to verify 45 the inserted gene sequences.

Cell culture and transfection. HEK293T/17 (ATCC CRL-11268), U2OS (ATCC HTB-96), COS-7 (ATCC CRL-1651), and HeLa (ATCC CCL-2) cells were cultured in DMEM (Thermo Fisher Scientific 11995-065) supplemented with 10% fetal bovine serum (Corning 35-010-CV), 100 U ml⁻¹ penicillin and 100 µg ml⁻¹ of streptomycin (Thermo Fisher Scientific 15140122) under 37° C. with 5% CO₂. TrypLE Express (Thermo Fisher Scientific 12604013) was used for detaching cells from culture flasks during cell 50 passage. All cell lines used in this study were transfected using FuGENE HD (Promega 2311) according to the manufacturer's instructions. Prior to live-cell imaging, cells were changed to imaging media: phenol red-free DMEM (Thermo Fisher Scientific 31053-028) supplemented with 10% fetal bovine serum (Corning 35-010-CV), 100 U ml⁻¹ penicillin and 100 µg ml⁻¹ of streptomycin (Thermo Fisher Scientific 15140122), 1x GlutaMAX™ (Thermo Fisher Scientific 35050-061), and 1 mM sodium pyruvate (Thermo Fisher Scientific 11360-070).

Fluorescence and bioluminescence imaging of tDeg-tagged proteins. To construct an expression vector for EYFP, EYFP-tDeg, mNeonGreen-tDeg, mCherry-tDeg, NanoLuc-

103

tDeg, EGFP-TetR-tDeg, EGFP-EZH2-tDeg, or mCherry-NF- κ B-tDeg, a pcDNA3.1(+) vector was digested by MluI and XbaI and ligated to an insert comprising a miniCMV promoter (5'-GGTAGGCGTGTACGGTGGGAGGCC-TATATAAGCAG AGCT-3' (SEQ ID NO: 93), a HindIII restriction site, a Kozak sequence (5'-GCCACC-3'), and the gene encoding EYFP, EYFP, mNeonGreen, mCherry, NanoLuc, EGFP-TetR, EGFP-EZH2, or mCherry-NF- κ B, respectively, fused with tDeg. These expression vectors were called miniCMV-EYFP, miniCMV-EYFP-tDeg, miniCMV-mNeonGreen-tDeg, miniCMV-mCherry-tDeg, miniCMV-NanoLuc-tDeg, miniCMV-EGFP-TetR-tDeg, miniCMV-EGFP-EZH2-tDeg, and miniCMV-mCherry-NF- κ B-tDeg respectively. For control constructs of miniCMV-EGFP-TetR, miniCMV-EGFP-EZH2, and miniCMV-mCherry-NF- κ B, a stop codon was inserted on the immediate upstream of the coding sequence of tDeg using QuikChange Site-Directed Mutagenesis Kits (Agilent).

To construct an expression vector for different circular RNAs, the Tornado expression plasmid (Litke et al., Highly Efficient Expression of Circular RNA Aptamers in Cells using Autocatalytic Transcripts," *Nat. Biotechnol.* 37:667-675 (2019), which is hereby incorporated by reference in its entirety) containing an F30 scaffold was digested, then ligated to inserts encoding the following sequences, respectively: wild-type TAR RNA (5'-GGCTCGTAGCTCATT-AGCTCCGAGCC-3' (SEQ ID NO: 65)), TAR Variant-1 (5'-GGCTCGTAGCTCATTAGCTCCGAGCC-3'(SEQ ID NO: 67)), Pepper (TAR Variant-2) (5'-GGCTCGTT-GAGCTCATTAGCTCCGAGCC-3'(SEQ ID NO: 69), or a control RNA, the MS2 hairpin (5'-ACATGAGGATCACCATGT-3'(SEQ ID NO: 94)). These vectors were called: U6+27-tnd-wildtype TAR, TAR Variant-1, Pepper (TAR Variant-2), control RNA, respectively.

For live-cell imaging experiments with HEK293T cells, HEK293T cells were seeded into 12-well flat bottom cell culture plates (Corning™ 3513) with 2×10^5 cells per well, and were cultured overnight. On the next day, cells were transfected using FuGENE HD (Promega 2311) according to the manufacturer's instructions. Specifically, for imaging experiments in FIGS. 1A-C, 550 ng of miniCMV-EYFP-tDeg were cotransfected with 550 ng of U6+27-tnd-wildtype TAR, TAR Variant-1, Pepper (TAR Variant-2), or control RNA, respectively. In the case of EYFP, 550 ng of miniCMV-EYFP was transfected with 550 ng of diluent DNA (pUC19 plasmid) to maintain 1.1 μ g of total plasmid DNA per well. For imaging experiments in FIGS. 6A-6G and FIGS. 7A-7G, 550 ng of miniCMV-protein X-tDeg (protein X=mNeonGreen, mCherry, NanoLuc, EGFP-TetR, EGFP-EZH2, or mCherry-NF- κ B) was cotransfected with 550 ng of circular Pepper (TAR Variant-2) or with 550 ng of diluent DNA (pUC19 plasmid). At 24 hours after transfection, cells were subcultured into 35 mm imaging dishes precoated with poly-D-lysine (Mattek Corporation P35GC-1.5-14C) and mouse laminin I (Cultrex® 3401-010-02) in culture media. Cells were then cultured overnight. Cell culture media was changed imaging media prior to fluorescence or bioluminescence live-cell imaging.

For live-cell imaging experiments in FIGS. 4A-4B, U2OS cells, COS-7 cells, or HeLa cells were seeded into 35 mm imaging dishes precoated with poly-D-lysine (Mattek Corporation P35GC-1.5-14C) with 2×10^5 cells per dish, respectively. On the next day, cells were transfected using FuGENE HD (Promega 2311) according to the manufacturer's instructions. Specifically, 1.4 μ g of miniCMV-EYFP-tDeg was cotransfected with 1.4 μ g of circular Pepper (TAR Variant-2) or 1.4 μ g of diluent DNA (pUC19 plasmid). At 48

104

hours after transfection, cell culture media was changed imaging media prior to fluorescence live-cell imaging.

Prior to live-cell fluorescence or bioluminescence imaging, 1 μ L of Hoechst 33342 (Thermo Fisher Scientific H3570) per 2 ml of imaging media was added to the cells. In the case of proteasome inhibitor treatment, cells were treated with either DMSO or 10 μ M (final concentration in the media) MG132 for 7 hours prior to live-cell imaging. In the case of bioluminescence imaging of NanoLuc, 20 μ L of furimazine (Promega Nano-Glo® Luciferase Assay System) per 2 ml of imaging media was added to the cells prior to bioluminescence imaging.

For live-cell fluorescence or bioluminescence imaging, an epifluorescence inverted microscope (Nikon Eclipse TE2000-E) equipped with a CoolSnap HQ2 CCD camera and a 130-W Nikon mercury lamp was used. The NIS-Elements Advanced Research software (Nikon) was used to control the microscope and camera. Cells were imaged with a 20 \times /0.75-NA (numerical aperture) or a 40 \times /0.75-NA air objective (Nikon) at 37° C. A FITC filter cube (with excitation filter 470 \pm 20 nm, dichroic mirror 495 nm (long pass), and emission filter 525 \pm 25 nm) was used for detecting EGFP-TetR-tDeg or EGFP-EZH2-tDeg with an exposure time of 500 msec. A YFP filter cube (with excitation filter 500 \pm 12 nm, dichroic mirror 520 nm (long pass), and emission filter 542 \pm 13.5 nm) was used for detecting EYFP, EYFP-tDeg, or mNeonGreen-tDeg with an exposure time of 500 msec. A TRITC filter cube (with excitation filter 560 \pm 20 nm, dichroic mirror 585 nm (long pass), and emission filter 630 \pm 37.5 nm) was used for detecting mCherry-tDeg, or mCherry-NF- κ B-tDeg with an exposure time of 500 msec. A filter cube (with emission filter 460 \pm 25 nm) was used for detecting the bioluminescence of NanoLuc with an exposure time of 3 minutes. A DAPI filter cube (with 350 \pm 25 nm excitation filter, 400 nm (long pass) dichroic mirror, and 460 \pm 25 nm emission filter) was used for detecting the Hoechst-stained nuclei in cells with an exposure time of 100-500 msec. All filters used in these filter cubes are purchased from Chroma Technology. Cell fluorescence/bioluminescence was calculated using ImageJ by measuring the mean fluorescence/bioluminescence signal in a cell's area and subtracting background based on average signal of culture media. Normalized fluorescence/bioluminescence was calculated by dividing the cell fluorescence/bioluminescence intensity of each cell to the averaged cell fluorescence/bioluminescence of the whole cell population.

RT-qPCR. Total RNA was isolated from cells using Trizol according to the manufacturer's instruction. To remove residual DNA contaminations, the purified RNA was treated with DNaseI (Thermo-Fisher) according to the manufacturer's instructions. The same amount of DNaseI-treated RNA was reverse transcribed to cDNA using SuperScript IV First-Strand kit (Invitrogen) with random hexamers according to the manufacturer's instructions. To measure relative expression levels of the RNAs of interest, qPCR measurements were performed using the iQ SYBR Green Supermix with 0.250 ng of cDNA in the final reaction mix. For the amplification, the following protocol was used: 98° C. for 2 minutes, 40 cycles of 95° C. for 10 seconds, 60° C. for 40 seconds. Primer sets for amplifying the cDNA of EYFP and mCherry are listed in Table 9. Every primer set was tested for its efficiency. To test primer specificity, melting curves were performed at the end of the 40 cycles of amplification. In the case of mCherry quantification, an untransfected sample was added as additional negative control. Relative

105

measurements ($2^{-\Delta Cq}$) of mCherry, EYFP were performed using GAPDH and RPS18 as housekeeping genes. Biological replicates were tested.

TABLE 9

ssDNA oligo probes used in RT-qPCR		
EYFP fw	ACGTAAACGCCACAAGTTC	SEQ ID NO: 95
EYFP rv	CTTCATGTTGGTCGGGGTAGC	SEQ ID NO: 96
mCherry fw	CACGAGTTGAGATCGAGGG	SEQ ID NO: 97
mCherry rv	CAAGTAGTCGGGGATGTCGG	SEQ ID NO: 98

Gel staining. Total RNA was isolated from cells using TRIzol® according to the manufacturer's instruction. Then, 2.5 µg of isolated total RNA was separated using a precast 6% TBE-Urea Gel (Life Technologies EC68655). This gel was run at 200 V in TBE buffer until completion, and stained with SYBR Gold (ThermoFisher S11494) diluted 1:10,000 in TBE buffer for 15 minutes. After SYBR Gold staining, RNA bands were imaged on a ChemiDoc XRS+ system (Bio-Rad).

mRNA imaging using tDeg and Pepper. To construct an expression vector for RNA-regulated fluorescent fusion proteins used in mRNA imaging, a pcDNA3.1(+) vector was digested by MluI and XbaI and ligated to an insert comprising a miniCMV promoter (5'-GGTAGGCCTGTACGGTGGGAGGCC-TATATAAGCAGAG CT-3' (SEQ ID NO: 118)), a HindIII restriction site, a Kozak sequence (5'-GCCACC-3'), and the gene encoding tandem copies of mNeonGreen, mVenus, or mCherry, respectively. To construct an expression vector for an mCherry mRNA reporter containing different 3'UTR tags comprising 10 or 20 concatenated Pepper aptamers, a pcDNA3.1(+) vector was first digested by HindIII and XbaI and ligated to an insert encoding the gene of mCherry followed by XhoI after its stop codon. This vector was called CMV-mCherry. CMV-mCherry was then digested XhoI and XbaI, and ligated to different Pepper tags, respectively. All the Pepper tags were synthesized by GenScript.

U2OS cells were seeded into 35 mm imaging dishes precoated with poly-D-lysine (Mattek Corporation P35GC-1.5-14C) with 2×10^5 cells per dish. On the next day, cells were transfected using FuGENE HD (Promega 2311) according to the manufacturer's instructions. Specifically, 1.4 µg of RNA-regulated fluorescent fusion protein plasmids were cotransfected with 1.4 µg of mRNA reporter plasmids. At 48 hours after transfection, cell culture media was changed to imaging media prior to imaging experiments.

For mRNA imaging experiments, an epifluorescence inverted microscope (Olympus IX-70) equipped with a Evolve® 512 EMCCD OEM camera (Photometrics) and an Insight SSI 7 color solid state illumination system (Applied Precision) was used. The Resolve3D softWoRx-Acquire Version: 6.5.2 was used to control the microscope and camera. Cells were imaged with a 100×/1.4-NA oil objective at 37° C., with N=1.520 immersion oil (Applied Precision). A FITC filter cube (with excitation filter 475±14 nm, dichroic mirror with a reflection band of 481-502 nm, and a transmission band of 506-543 nm), and emission filter 525±25 nm) was used for detecting mNeonGreen with an exposure time of 50 msec. A YFP filter cube (with excitation filter 513±8.5 nm, dichroic mirror with a reflection band of 496-528 nm, and a transmission band of 537-550 nm, and emission filter 559±19 nm) was used for detecting mVenus

106

with an exposure time of 100 msec. A TRITC filter cube (with excitation filter 542±13.5 nm, dichroic mirror with a reflection band of 547-565 nm, and a transmission band of 576-630 nm, and emission filter 594±22.5 nm) was used for detecting reporter plasmids encoding mCherry with an exposure time of 10-100 msec. Signal-to-noise ratio of the fluorescent puncta was calculated by the mean fluorescence intensity of each mRNA puncta divided by the mean fluorescence intensity of the adjacent cytosolic background fluorescence.

Northern blot. HEK293T cells were seeded into 10 cm culture dish with 3×10^6 cells per dish. On the next day, cells were cotransfected with CMV-mCherry-(F30-2×Pepper)₁₀ and miniCMV-(mNeonGreen)₄-tDeg or pUC19, respectively. A total amount of 19 µg plasmid DNA was used for each culture dish, and pUC19 vector was used here as a diluent DNA to ensure the same amount of plasmid DNA transfected to the cells. All transfections were performed using FuGENE HD (Promega 2311) according to the manufacturer's instructions. Cells were harvested after 48 hours of transfection. Total RNA was extracted with TRIzol® (Thermo Fisher Scientific 15596026) followed by isopropanol precipitation. The purified total RNA was then subjected to RNase-free DNase I (Thermo Fisher Scientific AM2224) digestion at 37° C. for 1 hour. After digestion, the RNA was subjected to phenol-chloroform (Thermo Fisher Scientific AM9720) extraction and ethanol purification.

For gel electrophoresis, a 1.5% agarose/formaldehyde gel (20 mM MOPS, 5 mM sodium acetate, 1 mM EDTA, 1.5% w/v agarose, 2% formaldehyde) was used. 20 µg of total RNA was loaded in each lane. The RNA was resuspended in 20 µL of RNA sample buffer (20 mM MOPS, 5 mM sodium acetate, 1 mM EDTA, 50% v/v formamide, 3.7% formaldehyde). The RNA samples were heated at 70° C. for 10 minutes, and then chilled on ice for more than 1 minute. Before loading the RNA samples into the gel, the RNA samples were mixed with 2 µL of loading buffer (50% glycerol, 5 mM EDTA, 0.4% bromophenol blue, 0.4% xylene cyanol). The gel was run at 70 V for 2 hours. After electrophoresis, the gel was stained with 1×SYBR™ Gold Nucleic Acid Gel Stain (Thermo Fisher Scientific S11494) to assess the quality of the RNA and check for separation. All solutions mentioned above were made in diethylpyrocarbonate (DEPC)-treated water.

After electrophoresis, the RNA was transferred to Amersham Hybond-N+ nylon membrane (GE Healthcare Life Sciences RPN203B) using the VacuGene XL Vacuum Blotting System (GE Healthcare Life Sciences) according to the manufacturer's instructions. The RNA was then UV cross-linked to the nylon membrane. The membrane was washed with NorthernMax® Prehybridization/Hybridization Buffer (Thermo Fisher Scientific AM8677) at 42° C. for at least 30 minutes. Biotinylated (at 5') single-stranded DNA probes (Integrated DNA Technologies) as shown in Table 10 were mixed with NorthernMax® Prehybridization/Hybridization Buffer and incubated with the membrane at 42° C. overnight. On the following day, the membrane was washed in 50 mL of wash buffer 1 (2×SSC, 0.1% SDS) twice at 42° C. for 10 minutes each time, and then washed with wash buffer 2 (0.1×SSC, 0.1% SDS) twice at 42° C. for 15 minutes. The membrane was visualized by Chemiluminescent Nucleic Acid Detection Module Kit (Thermo Fisher Scientific 89880).

TABLE 10

ssDNA oligo probes used in FIG. 12A		
Probe-1	GTTGAGTGATTAGCGATTGA TTCCGGCC	SEQ ID NO: 99
Probe-2	GTCGGATGATTTCTGTAATA GATTGCGCTG	SEQ ID NO: 100
Probe-3	TTGACGTGATTTCTGAGAT TTTCCGCA	SEQ ID NO: 101
Probe-4	TGCCTGATTGTAAGTATGTG GATTATCGG	SEQ ID NO: 102
Probe-5	GGATAGGTATGGAGGAAGTA GCTTGGAA	SEQ ID NO: 103
Probe-6	ACAATATCTTGCGCCGTTCG ATCTTG	SEQ ID NO: 104
Probe-7	GGCCGCCAAGAACGACC AA	SEQ ID NO: 105
Probe-8	CCTAAGAACCTAACATATCT AGCGAGG	SEQ ID NO: 106
Probe-9	TGTGCACCTTGAAAGCGCATGAA	SEQ ID NO: 107
Probe-10	CCTGGGTACGGTCACCACG	SEQ ID NO: 108
Probe-11	GCCCCATGGTCTTCTCTGC	SEQ ID NO: 109
Probe-12	GGGTGCTTCACGTAGGCCTT	SEQ ID NO: 110
Probe-13	GTCACCTTCAGCTTGGCGTC	SEQ ID NO: 111
Probe-14	GCCTCTGCTTGATCTGCCCTC	SEQ ID NO: 112
Probe-15	GTCTTGACCTCAGCGTCGTAGTG	SEQ ID NO: 113
Probe-16	CGGCGCGTTCGTACTGTTCC	SEQ ID NO: 114
Probe-17	GCCGATAATCCACATACTTACAA TCAGG	SEQ ID NO: 115

Imaging membrane-tethered mRNA. U2OS cells were seeded 72 hours before imaging in 96-well glass bottom dishes (Matrplates, Brooks Life Science Systems) at 40% confluence. Cells were transfected with DNA plasmids that encode miniCMV-(mNeonGreen)₄-tDeg, PCP-3×mCherry-CAAX and the mRNA reporter 48 hours before imaging using 0.5 µl FuGENE 6 (Promega) and 200-300 ng DNA per well. The transfection mix was prepared in OptiMEM (Sigma-Aldrich) and added to the cells in a total volume 150-200 µl of medium.

Twenty-four hours prior to imaging, transcription of the reporters was induced by addition of doxycycline (1 ng/ml) (Sigma-Aldrich). Thirty minutes before imaging, the cell culture medium was replaced with pre-warmed CO₂-independent Leibovitz's-15 medium (Gibco) with doxycycline. Images were acquired using a Nikon TI inverted microscope with perfect focus system equipped with a Yokagawa CSU-X1 spinning disc, a 100× 1.49 NA objective and an iXon Ultra 897 EMCCD camera (Andor) and was controlled by NIS software (Nikon). During the experiment, cells were maintained at a constant temperature of 37° C. Single Z-plane images were acquired, with the bottom plasma membrane of the cell in the focal plane. Camera exposure times of 500 ms were used for both mNeonGreen and mCherry.

To determine the fluorescence intensity of mRNA foci, mean spot intensities were measured in Image J in a region of interest (ROI) 0.53×0.53 µm in size. For each spot, local

background fluorescence intensity was measured in a ROI (0.53×0.53 µm in size) directly next to the spot of interest, and mean background fluorescence intensities were subtracted from the mean spot intensity. Cells with very high number of mRNAs (more than ~50) were excluded from the analysis.

Western Blotting. Cells were lysed in whole cell lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM EDTA, 50 mM NaCl, 1% Triton X-100, 0.1% SDS) containing 1× protease and phosphatase inhibitor (Pierce, 78440). Lysates were cleared by centrifugation (12,000 g for 10 minutes). Protein quantification was performed using the Pierce BCA protein assay kit according to the manufacturer's instruction (Thermo Fisher Scientific, 23227). Equal quantities of proteins were mixed with loading dye, and incubated at 95° C. for 5 minutes before they were separated on 4-12% Bis-Tris gels (Invitrogen) and transferred onto a PVDF membrane at constant 350 mA at 4° C. for 1 hour. Membranes were blocked by incubation in 5% milk for 1 hour at room temperature under agitation and then incubated with the following primary antibodies: mouse anti-GAPDH (Santa Cruz) with a 1:5000 dilution in 1% milk overnight, or rabbit anti-mCherry (Abcam, ab167453) with a 1:1000 dilution in 1% milk overnight, or rabbit anti-ubiquitin (Abcam, ab19247) with a 1:1000 dilution in 1% milk overnight. After incubation with the appropriate secondary antibodies conjugated to HRP and extensive washing, blots were imaged on a ChemiDoc XRS+ system (Bio-Rad).

Imaging ER-targeting mRNA. To construct an expression vector for an ER-targeting mRNA reporter, DNA sequence that encodes the first 29 amino acids of cytochrome p450, CytER, and a linker sequence (MDPVVVLGLCLSLLLSLWKQSYGGKLGGSGG TGGSGTSGG (SEQ ID NO: 116) was cloned into the upstream of the mCherry sequence of the CMV-mCherry-(F30-2×Pepper)₁₀ plasmid to make CMV-CytER-mCherry-(F30-2×Pepper)₁₀. To construct the plasmid that encodes the RNA-regulated fluorescent fusion protein used in this experiment, the miniCMV promoter sequence in miniCMV-(mNeonGreen)₄-tDeg was replaced with the human ubiquitin C promoter sequence to make UbC-(mNeonGreen)₄-tDeg.

U2OS cells were seeded into 35 mm imaging dishes precoated with poly-D-lysine (Mattek Corporation P35GC-1.5-14C) with 2×10⁵ cells per dish. On the following day, cells were cotransfected with 1.4 µg of CMV-CytER-mCherry-(F30-2×Pepper)₁₀, 0.28 µg of UbC-(mNeonGreen)₄-tDeg, and 1.12 µg of pUC19 (as a diluent DNA) using FuGENE HD (Promega 2311) according to the manufacturer's instructions. At 48 hours after transfection, cell culture media was changed to imaging media prior to imaging experiments. This imaging setup for these experiments are the same as the one used for mRNA imaging using tDeg and Pepper.

Imaging β-actin mRNA after arsenite stress. To construct an expression vector for a β-actin mRNA reporter containing a (F30-2×Pepper)₁₀ tag, the full length β-actin gene (from Addgene Plasmid #27123) was amplified by PCR and digested by XhoI and HindIII, and then ligated to a vector from CMV-mcherry-(F30-2×Pepper)₁₀ digested by the same restriction endonucleases to cut out the gene encoding mCherry. This expression vector was called CMV-O-actin-(F30-2×Pepper)₁₀.

U2OS cells stably expresses Halo-G3BP1 were seeded into 35 mm imaging dishes precoated with poly-D-lysine (Mattek Corporation P35GC-1.5-14C) with 2×10⁵ cells per dish. On the following day, cells were cotransfected with 1.4 µg of miniCMV-(mNeonGreen)₄-tDeg with 1.4 µg of CMV-

109

O-actin-(F30-2×Pepper)₁₀ using FUGENE HD (Promega 2311) according to the manufacturer's instructions. For control experiments, 1.4 µg of miniCMV-(mNeonGreen)₄-tDeg with 1.4 µg of U6+27-tnd-Pepper was used following the same transfection protocol. At ~40 hours after transfection, cell culture media was changed to imaging media with the HaloTag® TMRDirect™ Ligand (Promega G2991) for 5 hours. Cells were then rinsed with 1×PBS (Thermo Fisher Scientific 10010049) and incubated in imaging media prior to imaging experiments. The same microscope setup as in the above mRNA imaging experiments was used. To induce stress granule formation, 1 mL of imaging media supplemented with 1000 µM of sodium arsenite was added to the cells cultured in 1 mL of imaging media to reach a final concentration of 500 µM of sodium arsenite.

Statistical analysis. All data were expressed as means±s.d. with sample sizes (n) listed for each experiment. Statistical analyses were performed using Excel (Microsoft) and Prism (Graphpad). For different circular TAR variants' inhibition of tDeg's destabilizing effect, and optimization of the number of fluorescent mNeonGreen monomers in the RNA-regulated fluorescent fusion protein for imaging mRNA in live cells, one-way ANOVA was used to analyze significant differences between group means. For Pepper RNA-dependent regulation of protein stability, imaging green Pepper-tagged β-actin mRNA, proteasomal inhibition, imaging membrane-tethered mRNA, two tailed Student's t-tests were used to analyze significant differences between group means. P values were reported for each experiment.

Example 1—tDeg Reduces Protein Stability by Inducing Proteasomal Degradation

In order to expand fluorescent aptamer-based imaging, Applicant sought to create a new class of RNA-regulated fluorescent dyes that are genetically encoded. Fluorescent proteins are particularly useful since a diverse array of spectrally distinct proteins have been described (Rodriguez et al., "The Growing and Glowing Toolbox of Fluorescent and Photoactive Proteins," *Trends Biochem. Sci.* 42:111-129 (2017), which is hereby incorporated by reference in its entirety). However, these proteins are constitutively fluorescent. To make them dependent on RNA, Applicant considered making them rapidly degraded in cells except when bound by a specific RNA aptamer. In this way, fluorescence would be selectively associated with RNA-protein complexes, and not with unbound fluorescent protein. This would be functionally equivalent to RNA-induced fluorescence of small molecule dyes.

First, a "destabilization domain" that can be inhibited by an RNA aptamer was developed. Previously, the Arg-Arg-Arg-Gly (SEQ ID NO: 117) was described as a degron sequence when appended to the C-terminus of proteins (Bonger et al., "Small-Molecule Displacement of a Cryptic Degron Causes Conditional Protein Degradation," *Nat. Chem. Biol.* 7:531-537 (2011), which is hereby incorporated by reference in its entirety). This sequence is similar to the arginine-rich RNA-binding domain of the Tat protein, which contains Arg-Arg as its last two amino acids. Therefore, Arg-Gly was appended to extend this Arg-Arg sequence so that the full Arg-Arg-Arg-Gly (SEQ ID NO: 117) degron is at the C-terminus of this peptide (FIGS. 1A-1B and FIGS. 2A-2B). This 19-amino acid-long bifunctional peptide was termed "tDeg." Tat binds a 28 nt-long RNA hairpin termed TAR (Ye et al., "Molecular Recognition in the Bovine Immunodeficiency Virus Tat Peptide-TAR RNA Complex," *Chem. Biol.* 2:827-40 (1995) and Puglisi et al., "Solution

110

Structure of a Bovine Immunodeficiency Virus Tat-TAR Peptide-RNA Complex," *Science* 270:1200-1203 (1995), which are hereby incorporated by reference in their entirety), which may shield the degron and thus prevent recruitment of the proteasomal machinery needed for proteolysis (FIG. 1A and FIGS. 2A-2B).

Whether tDeg confers instability to proteins was first investigated. To do so, tDeg was fused to the C-terminus of enhanced yellow fluorescent protein (EYFP), and the resulting fusion protein (EYFP-tDeg) was expressed in HEK293T cells. While EYFP was readily detectable, EYFP-tDeg was nearly undetectable (FIGS. 1B-1C). EYFP-tDeg was restored by proteasome inhibition (FIGS. 3A-3B) indicated that tDeg reduces protein stability by inducing proteasomal degradation.

Example 2—tDeg is Regulated by TAR RNA and TAR RNA Variants

Whether the tDeg can be regulated by the TAR RNA was next investigated. The TAR RNA was expressed as a circular RNA using the Tornado ribozyme-assisted circularization approach to achieve high expression in mammalian cells (Litke & Jaffrey, "Highly Efficient Expression of Circular RNA Aptamers in Cells Using Autocatalytic Transcripts," *Nat. Biotechnol.* 37:667-675 (2019), which is hereby incorporated by reference in its entirety). When TAR was expressed, EYFP-tDeg-expressing cells exhibited a 24-fold increase of fluorescence relative to control RNA (FIGS. 1B-1C). TAR variants that bind Tat with higher affinity, Variant-1 and Variant-2 (Smith et al., "Altering the Context of an RNA Bulge Switches the Binding Specificities of Two Viral Tat Proteins," *Biochemistry* 37:10808-10814 (1998), which is hereby incorporated by reference in its entirety), were even more efficient at inducing EYFP-tDeg, with Variant-2 exhibiting a 38-fold increase in cellular fluorescence (FIGS. 1B-1C; FIGS. 4A-4B). Expression of Variant-2 induced EYFP-tDeg cellular fluorescence levels similar to levels in cells expressing EYFP without the tDeg (FIG. 1C). Furthermore, Variant-2 induced EYFP-tDeg fluorescence in diverse cell types (FIGS. 5A-5G). Thus, the EYFP-tDeg is a RNA-regulated fluorescent fusion protein that is regulated by TAR.

Because the TAR Variant-2 aptamer can control the expression of different colored fluorescent proteins, as described infra, this aptamer was named after the multicolored vegetable Pepper, in keeping with the vegetable nomenclature system used previously for fluorogenic RNA aptamers.

Example 3—tDeg Tag is a Versatile Tag for Pepper-Dependent Protein Stabilization

Whether the expression level of other proteins could be controlled by the Pepper RNA was next investigated. Addition of tDeg to the C-terminus of mNeonGreen, mCherry, NanoLuc, tetracycline repressor protein (TetR), EZH2, and NF-κB, resulted in minimal or undetectable protein levels in control cells and clear induction in circular Pepper-expressing cells (FIGS. 6A-6G and FIGS. 7A-7G). Taken together, these data indicate that the tDeg tag is a versatile tag for RNA-dependent protein stabilization.

Example 4—Intracellular Imaging Using Pepper-Modified mRNA

mRNAs are commonly imaged using tethered fluorescent proteins. For example, a GFP fusion with MS2 phage coat

111

protein (MCP) can be recruited to mRNAs containing 24-48 consecutive MS2 RNA hairpins in their 3'UTRs (Bertrand et al., "Localization of ASH1 mRNA Particles in Living Yeast," *Mol. Cell* 2:437-45 (1998), which is hereby incorporated by reference in its entirety). In this way, many GFPs are recruited to single mRNAs resulting in an aggregate fluorescence that can be detected by fluorescence microscopy. Typically nuclear localization elements are added to the GFP-MCP fusion to remove the unbound fluorescent protein from the cytoplasm into the nucleus (Bertrand et al., "Localization of ASH1 mRNA Particles in Living Yeast," *Mol. Cell* 2:437-45 (1998), which is hereby incorporated by reference in its entirety). This can reduce the fluorescence background in the cytosol, facilitating mRNA detection. However, this may introduce a potential artifact since the MS2-tagged mRNAs will contain dozens of nuclear localization sequences due to the recruited fluorescent proteins (Tyagi, S., "Imaging Intracellular RNA Distribution and Dynamics in Living Cells," *Nat. Methods* 6:331-338 (2009), which is hereby incorporated by reference in its entirety). The RNA aptamers described herein do not introduce a cellular trafficking element and may therefore bypass this concern.

To investigate the use of RNA aptamers in intracellular imaging, a tag for mRNA imaging consisting of consecutive Pepper aptamers was next generated. In optimization experiments, an mCherry mRNA reporter containing different 3'UTR tags comprising 10 or 20 concatenated Pepper aptamers and Pepper aptamers that were inserted into an RNA three-way junction sequence termed F30 were imaged. Aptamers inserted within the F30 show improved folding (Filonov et al., "In-Gel Imaging of RNA Processing Using Broccoli Reveals Optimal Aptamer Expression Strategies," *Chem. Biol.* 22:649-60 (2015), which is hereby incorporated by reference in its entirety). mCherry mRNA was readily detectable as mobile fluorescent puncta in the cytoplasm when the tag contained 20 Pepper aptamers. The brightest puncta were seen when using the (F30-2×Pepper)₁₀ tag, which comprises 10 consecutive F30 sequences, with each of the two arms of F30 containing one Pepper aptamer (FIGS. 8A-B; FIGS. 9A-9D; and FIGS. 10A-10C).

mRNA imaging using RNA-regulated fluorescent fusion proteins of different brightness was also investigated. These proteins comprised 2, 3, or 4 tandem mNeonGreen monomers with a C-terminal tDeg. In these experiments, a RNA-regulated fluorescent fusion protein comprising four mNeonGreens provided the highest signal-to-noise ratio for imaging mRNAs (FIGS. 10A-10C). Although most fluorescent puncta were detected in the cytoplasm, occasional puncta were detected in the nucleus, potentially reflecting mRNAs prior to nuclear export (FIGS. 11A-11C).

Cellular puncta likely reflect single mRNA molecules rather than Pepper-containing mRNA fragments since northern blotting of total cellular RNA derived from cells expressing (F30-2×Pepper)₁₀-tagged mRNA, either with or without coexpression of the (mNeonGreen)₄-tDeg showed mostly full-length transcripts (FIG. 12A). Furthermore, puncta derived from mRNAs tagged with (F30-2×Pepper)₁₀ were the same size and intensity as mRNAs tagged using the PP7 fluorescent protein recruitment system, which was previously shown to reflect single mRNA molecules (Yan et al., "Dynamics of Translation of Single mRNA Molecules In Vivo," *Cell* 165:976-989 (2016), which is hereby incorporated by reference in its entirety) (FIGS. 12B-12D).

Adding the Pepper tag to an mRNA could adversely affect mRNA fate. However, the (F30-2×Pepper)₁₀ Pepper tag was not found to substantially alter the stability of the mCherry

112

transcript (FIG. 13A). Similarly, a significant difference in protein translation between the untagged and Pepper-tagged mCherry mRNA transcript was not observed (FIGS. 13B-13D). Lastly, expression of RNA-regulated fluorescent fusion proteins did not significantly affect total cellular proteasome activity (FIG. 13E).

mRNAs that exhibit specific subcellular localizations were next imaged. mRNA localization to the endoplasmic reticulum (ER) was imaged using an ER-targeted reporter mRNA that encodes the first 29 amino acids of cytochrome P450, CytERM (cytoplasmic end of an endoplasmic reticulum signal-anchor membrane protein) (Costantini et al., "Assessing the Tendency of Fluorescent Proteins to Oligomerize Under Physiologic Conditions," *Traffic* 13:643-649 (2012), which is hereby incorporated by reference in its entirety). This sequence tethers the mRNA to the outer ER membrane during protein translation, and restricts the mRNA's mobility. Indeed, fluorescent puncta with low mobility were observed when this mRNA was expressed with a 3'UTR (F30-2×Pepper)₁₀ Pepper tag (FIGS. 14A-14D). Treatment with puromycin, which disrupts the ribosome and dissociates the mRNA from the nascent peptide, significantly increased puncta mobility, consistent with dissociation of the reporter mRNA from the ER (FIGS. 14A-14D).

Next, β-actin mRNA containing a 3'UTR (F30-2×Pepper)₁₀ tag was expressed and its localization was imaged in response to arsenite treatment, which induces stress granule formation (Tourrière et al., "The RasGAP-Associated Endoribonuclease G3BP Assembles Stress Granules," *J. Cell Biol.* 160:823-831 (2003), which is hereby incorporated by reference in its entirety). Upon application of 500 μM arsenite, the individual fluorescent puncta rapidly accumulated to form stress granules as evidenced by coexpression of Halo-tagged G3BP1 to label stress granules (FIGS. 15A-C and FIGS. 16A-B).

Example 5—Imaging of Pepper-Regulated mVenus and Pepper-Regulated mCherry

To expand the color palette of RNA-regulated fluorescent fusion proteins, two tandem copies of mVenus and two tandem copies of mCherry were fused with a C-terminal tDeg tag to convert them into RNA-regulated fluorescent fusion proteins, respectively, for imaging mRNAs. In both cases, fluorescent puncta were detected in the yellow and red fluorescence channels, respectively (FIGS. 17A-17B). Together, these data show that Pepper-tagged mRNAs can be imaged in different colors using different fluorogenic proteins.

Discussion of Examples 1-5

The studies described infra demonstrate how constitutively fluorescent proteins can be converted to fluorescent proteins that are regulated by RNA aptamers. RNA-regulation was conferred to a protein by making its proteomic stability controlled by an RNA aptamer, Pepper. In this way, unbound RNA-regulated fluorescent fusion protein is rapidly degraded, but the RNA-regulated fluorescent fusion protein bound to an specific RNA aptamer (e.g., Pepper) remains stable. Thus, these Pepper-regulated fluorescent fusion proteins are functionally analogous to RNA-regulated fluorogenic dyes. This system has the advantage of being able to use diverse fluorescent proteins with diverse spectral properties. Additionally, unlike the Spinach system (Paige et al., "RNA Mimics of Green Fluorescent Protein," *Science* 333:642-646 (2011), which is hereby incorporated by ref-

erence in its entirety), the fluorescent system described herein is fully genetically encoded.

Fluorophore maturation kinetics may also contribute to the low fluorescence of the Pepper system. Since the tDeg tag is highly efficient, it is possible that newly synthesized mNeonGreen is degraded prior to chromophore maturation. mNeonGreen that is bound to the RNA may persist for a sufficiently long time to mature to a fluorescent form while bound to RNA. This may further contribute to the low background fluorescence in cells.

Unlike previous mRNA imaging systems, no nuclear localization elements are added to fluorescent proteins to lower cytosolic background fluorescence. Instead, low background fluorescence is achieved by the highly efficient degradation of the unbound RNA-regulated fluorescent fusion protein. The simplicity of this system should simplify mRNA imaging.

An important question is whether the tagged mRNA faithfully recapitulates behavior of the endogenous mRNA. The Pepper tag did not substantially affect the stability, translation, and localization of the specific mRNAs described herein. Nevertheless, imaging tags are best used when comparing two mRNAs that differ by a single sequence alteration, or the same mRNA compared in two different conditions. In this way the role of a putative functional RNA element or RNA-regulatory pathway can be inferred and then validated with the endogenous mRNA.

Although the RNA-regulated destabilization domains were used to create fluorescent fusion proteins for RNA imaging, the ability to control protein expression levels through the Pepper aptamer can potentially enable novel synthetic biology applications. For these applications, Pepper can be expressed on its own, rather than part of an mRNA. By expressing tDeg-tagged proteins, diverse types of protein functions can be regulated by RNA aptamer expression levels.

Example 6—the tDeg-Pepper System can be Used to Selectively Modify RNA-Binding Proteins

RNA-binding proteins (RBPs) bind to RNA molecules to orchestrate most biological functions in the cell. A major way to uncover previously unknown biological functions is to discover the RBPs involved in these processes. Current methods for discovering RBPs have low sensitivity. This is because current methods rely on recruiting a biotin ligase or a peroxidase to an RNA of interest to biotinylate any RBPs that are bound to this RNA. The major problem of these methods is the promiscuous activity of the biotin ligase or peroxidase would also nonspecifically biotinylate irrelevant proteins in the cytosol.

To address this problem, new method for identifying RBPs with high sensitivity was developed. In this method, a biotin ligase and a peroxidase, whose activity is only turned on when it binds to the RNA target, was engineered.

To achieve this, tDeg was fused to a biotin ligase, called TurboID, and an engineered peroxidase, called APEX2, respectively. The stability of these two proteins can be regulated by the Pepper RNA. This method drastically decreases the nonspecific biotinylation due to the promiscuous activity of this biotin ligase and peroxidase, thereby enabling the discovery of RBPs in living cells with high sensitivity.

tDeg confers Pepper RNA-dependent regulation of a biotin ligase, TurboID, and a peroxidase, APEX2. FIG. 18A-18B show that HEK293T cells transiently express EGFP-TurboID-tDeg (FIG. 18A), and EGFP-APEX2-tDeg (FIG. 18B), with and without the Pepper RNA aptamer, respectively. In each case, proteins were nearly undetectable unless coexpressed with the Pepper RNA. FIG. 18C provides a schematic showing that a selectively activated biotin ligase (TurboID-tDeg) specifically biotinylates an RNA-binding protein (CELF1) that bind to the RNA sequence of interest (EDEN15). FIG. 18 D shows that TurboID-tDeg enables selective biotinylation of CELF1, while minimizing nonspecific biotinylation of proteins that do not bind to the RNA of interest (EDEN15). These results demonstrate that the tDeg-Pepper system can be used to selectively modify RNA-binding proteins.

Example 7—Tat-GG Confers Pepper RNA-Dependent Regulation

Next, whether a variant of tDeg, Tat-GG, can be regulated by the Pepper RNA aptamer was examined. In these experiments, U2OS cells transiently expressed mNeonGreen-Tat-GG fusion protein with and without the circular Pepper RNA aptamer, respectively. Cells showed undetectable levels of green fluorescence without the circular Pepper RNA aptamer (FIG. 19). The green fluorescence of mNeonGreen-Tat-GG was only restored when the circular Pepper RNA aptamer was coexpressed (FIG. 19). Thus, these results confirm that the tDeg variant Tat-GG can be regulated by the Pepper RNA aptamer.

Example 8—HIV Tat-RRRG Confers HIV TAR-Dependent Regulation

Next, whether HIV Tat-RRRG (RKKRQRRRG; SEQ ID NO: 127) can be regulated by the HIV TAR sequence (ACGAAGCUUGAUCCCCGUUUGCCGGUCGAU CGC-UUCGA; SEQ ID NO: 128) was examined. In these experiments, cells transiently expressed YFP-HIV Tat-RRRG fusion protein with and without the circular HIV TAR RNA aptamer, respectively. Cells showed undetectable levels of yellow fluorescence without the circular HIV TAR RNA aptamer (FIG. 20). The yellow fluorescence of YFP-HIV Tat-RRRG was restored when the circular HIV TAR RNA aptamer was coexpressed (FIG. 20). Thus, these results confirm that HIV Tat-RRRG can be regulated by the HIV TAR RNA aptamer.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 130

```

<210> SEQ ID NO 1
<211> LENGTH: 238
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Green Fluorescent Protein

```

-continued

<400> SEQUENCE: 1

Met Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val	5	10	15
---	---	----	----

Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu	20	25	30
---	----	----	----

Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys	35	40	45
---	----	----	----

Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe	50	55	60
---	----	----	----

Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Gln	65	70	75	80
---	----	----	----	----

His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg	85	90	95
---	----	----	----

Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val	100	105	110
---	-----	-----	-----

Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile	115	120	125
---	-----	-----	-----

Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn	130	135	140
---	-----	-----	-----

Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly	145	150	155	160
---	-----	-----	-----	-----

Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val	165	170	175
---	-----	-----	-----

Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro	180	185	190
---	-----	-----	-----

Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser	195	200	205
---	-----	-----	-----

Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val	210	215	220
---	-----	-----	-----

Thr Ala Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys	225	230	235
---	-----	-----	-----

<210> SEQ ID NO 2

<211> LENGTH: 239

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: Enhanced Green Fluorescent Protein

<400> SEQUENCE: 2

Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu	1	5	10	15
---	---	---	----	----

Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly	20	25	30
---	----	----	----

Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile	35	40	45
---	----	----	----

Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr	50	55	60
---	----	----	----

Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys	65	70	75	80
---	----	----	----	----

Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu	85	90	95
---	----	----	----

Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu	100	105	110
---	-----	-----	-----

US 12,391,948 B2

117

118

-continued

Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
115 120 125

Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
130 135 140

Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
145 150 155 160

Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
165 170 175

Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
180 185 190

Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu
195 200 205

Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
210 215 220

Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys
225 230 235

<210> SEQ ID NO 3

<211> LENGTH: 239

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: Enhanced Yellow Fluorescent Protein

<400> SEQUENCE: 3

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

Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
20 25 30

Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
35 40 45

Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
50 55 60

Phe Gly Tyr Gly Leu Gln Cys Phe Ala Arg Tyr Pro Asp His Met Lys
65 70 75 80

Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
85 90 95

Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
100 105 110

Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
115 120 125

Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
130 135 140

Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
145 150 155 160

Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
165 170 175

Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
180 185 190

Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Tyr Gln Ser Ala Leu
195 200 205

Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
210 215 220

Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys
225 230 235

US 12,391,948 B2

119**120**

-continued

<210> SEQ ID NO 4
<211> LENGTH: 239
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Venus

<400> SEQUENCE: 4

```

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

Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
20          25          30

Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Leu Ile
35          40          45

Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
50          55          60

Leu Gly Tyr Gly Leu Gln Cys Phe Ala Arg Tyr Pro Asp His Met Lys
65          70          75          80

Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
85          90          95

Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
100         105         110

Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
115         120         125

Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
130         135         140

Asn Tyr Asn Ser His Asn Val Tyr Ile Thr Ala Asp Lys Gln Lys Asn
145         150         155         160

Gly Ile Lys Ala Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Gly
165         170         175

Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
180         185         190

Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Tyr Gln Ser Ala Leu
195         200         205

Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
210         215         220

Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys
225         230         235

<210> SEQ ID NO 5  

<211> LENGTH: 239  

<212> TYPE: PRT  

<213> ORGANISM: Artificial  

<220> FEATURE:  

<223> OTHER INFORMATION: mVenus


<400> SEQUENCE: 5



```

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

Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
20 25 30

Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Leu Ile
35 40 45

Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
50 55 60

Leu Gly Tyr Gly Leu Gln Cys Phe Ala Arg Tyr Pro Asp His Met Lys
65 70 75 80

```


```

-continued

Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
 85 90 95

Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
 100 105 110

Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
 115 120 125

Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
 130 135 140

Asn Tyr Asn Ser His Asn Val Tyr Ile Thr Ala Asp Lys Gln Lys Asn
 145 150 155 160

Gly Ile Lys Ala Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Gly
 165 170 175

Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
 180 185 190

Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Tyr Gln Ser Lys Leu
 195 200 205

Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
 210 215 220

Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys
 225 230 235

<210> SEQ ID NO 6
<211> LENGTH: 239
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Citrine

<400> SEQUENCE: 6

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

Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
 20 25 30

Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
 35 40 45

Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
 50 55 60

Phe Gly Tyr Gly Leu Met Cys Phe Ala Arg Tyr Pro Asp His Met Lys
 65 70 75 80

Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
 85 90 95

Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
 100 105 110

Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
 115 120 125

Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
 130 135 140

Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
 145 150 155 160

Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
 165 170 175

Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
 180 185 190

Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Tyr Gln Ser Ala Leu
 195 200 205

-continued

Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
 210 215 220

Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys
 225 230 235

<210> SEQ_ID NO 7
 <211> LENGTH: 239
 <212> TYPE: PRT
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: mCitrine

<400> SEQUENCE: 7

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

Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
 20 25 30

Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
 35 40 45

Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
 50 55 60

Phe Gly Tyr Gly Leu Met Cys Phe Ala Arg Tyr Pro Asp His Met Lys
 65 70 75 80

Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
 85 90 95

Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
 100 105 110

Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
 115 120 125

Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
 130 135 140

Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
 145 150 155 160

Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
 165 170 175

Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
 180 185 190

Pro Val Leu Pro Asp Asn His Tyr Leu Ser Tyr Gln Ser Lys Leu
 195 200 205

Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
 210 215 220

Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys
 225 230 235

<210> SEQ_ID NO 8
 <211> LENGTH: 239
 <212> TYPE: PRT
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: Cerulean

<400> SEQUENCE: 8

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

Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
 20 25 30

Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile

US 12,391,948 B2

125**126**

-continued

35	40	45
Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr		
50	55	60
Leu Thr Trp Gly Val Gln Cys Phe Ala Arg Tyr Pro Asp His Met Lys		
65	70	75
Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu		
85	90	95
Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu		
100	105	110
Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly		
115	120	125
Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr		
130	135	140
Asn Ala Ile Ser Asp Asn Val Tyr Ile Thr Ala Asp Lys Gln Lys Asn		
145	150	155
Gly Ile Lys Ala Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser		
165	170	175
Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly		
180	185	190
Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu		
195	200	205
Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe		
210	215	220
Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys		
225	230	235

<210> SEQ ID NO 9
<211> LENGTH: 239
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: mCerulean

<400> SEQUENCE: 9

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

US 12,391,948 B2

127

128

-continued

165	170	175
-----	-----	-----

Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly 180	185	190
--	-----	-----

Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Lys Leu 195	200	205
--	-----	-----

Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe 210	215	220
--	-----	-----

Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys 225	230	235
--	-----	-----

<210> SEQ ID NO 10

<211> LENGTH: 222

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: Orange Fluorescent Protein

<400> SEQUENCE: 10

Met Asn Leu Ser Lys Asn Val Ser Val Tyr Met Lys Gly Asn 1	5	10	15
--	---	----	----

Val Asn Asn His Glu Phe Glu Tyr Asp Gly Glu Gly Gly Asp Pro 20	25	30
---	----	----

Tyr Thr Gly Lys Tyr Ser Met Lys Met Thr Leu Arg Gly Gln Asn Cys 35	40	45
---	----	----

Leu Pro Phe Ser Tyr Asp Ile Ile Thr Thr Ala Phe Gln Tyr Gly Phe 50	55	60
---	----	----

Arg Val Phe Thr Lys Tyr Pro Glu Gly Ile Val Asp Tyr Phe Lys Asp 65	70	75	80
---	----	----	----

Ser Leu Pro Asp Ala Phe Gln Trp Asn Arg Arg Ile Val Phe Glu Asp 85	90	95
---	----	----

Gly Gly Val Leu Asn Met Ser Ser Asp Ile Thr Tyr Lys Asp Asn Val 100	105	110
--	-----	-----

Leu His Gly Asp Val Trp Ala Val Gly Val Asn Phe Pro Pro Asn Gly 115	120	125
--	-----	-----

Pro Val Met Lys Asn Glu Ile Val Met Glu Glu Pro Thr Glu Glu Thr 130	135	140
--	-----	-----

Phe Thr Pro Lys Asn Gly Val Leu Val Gly Phe Cys Pro Lys Ala Tyr 145	150	155	160
--	-----	-----	-----

Leu Leu Lys Asp Gly Ser Tyr Tyr Gly Asn Met Thr Thr Phe Tyr 165	170	175
--	-----	-----

Arg Ser Lys Lys Ser Gly Gln Ala Pro Pro Gly Tyr His Phe Val Lys 180	185	190
--	-----	-----

His Arg Leu Val Lys Thr Asn Val Gly His Gly Phe Lys Thr Val Glu 195	200	205
--	-----	-----

Gln Thr Glu Tyr Ala Thr Ala His Val Ser Asp Leu Pro Lys 210	215	220
--	-----	-----

<210> SEQ ID NO 11

<211> LENGTH: 236

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: mNeon Green

<400> SEQUENCE: 11

Met Val Ser Lys Gly Glu Glu Asp Asn Met Ala Ser Leu Pro Ala Thr 1	5	10	15
--	---	----	----

US 12,391,948 B2

129**130**

-continued

His Glu Leu His Ile Phe Gly Ser Ile Asn Gly Val Asp Phe Asp Met
 20 25 30
 Val Gly Gln Gly Thr Gly Asn Pro Asn Asp Gly Tyr Glu Glu Leu Asn
 35 40 45
 Leu Lys Ser Thr Lys Gly Asp Leu Gln Phe Ser Pro Trp Ile Leu Val
 50 55 60
 Pro His Ile Gly Tyr Gly Phe His Gln Tyr Leu Pro Tyr Pro Asp Gly
 65 70 75 80
 Met Ser Pro Phe Gln Ala Ala Met Val Asp Gly Ser Gly Tyr Gln Val
 85 90 95
 His Arg Thr Met Gln Phe Glu Asp Gly Ala Ser Leu Thr Val Asn Tyr
 100 105 110
 Arg Tyr Thr Tyr Glu Gly Ser His Ile Lys Gly Glu Ala Gln Val Lys
 115 120 125
 Gly Thr Gly Phe Pro Ala Asp Gly Pro Val Met Thr Asn Ser Leu Thr
 130 135 140
 Ala Ala Asp Trp Cys Arg Ser Lys Lys Thr Tyr Pro Asn Asp Lys Thr
 145 150 155 160
 Ile Ile Ser Thr Phe Lys Trp Ser Tyr Thr Thr Gly Asn Gly Lys Arg
 165 170 175
 Tyr Arg Ser Thr Ala Arg Thr Thr Tyr Thr Phe Ala Lys Pro Met Ala
 180 185 190
 Ala Asn Tyr Leu Lys Asn Gln Pro Met Tyr Val Phe Arg Lys Thr Glu
 195 200 205
 Leu Lys His Ser Lys Thr Glu Leu Asn Phe Lys Glu Trp Gln Lys Ala
 210 215 220
 Phe Thr Asp Val Met Gly Met Asp Glu Leu Tyr Lys
 225 230 235

<210> SEQ ID NO 12
 <211> LENGTH: 236
 <212> TYPE: PRT
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: moxNeon Green
 <400> SEQUENCE: 12

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

-continued

Ala Ala Asp Trp Ser Arg Ser Lys Lys Thr Tyr Pro Asn Asp Lys Thr
145 150 155 160

Ile Ile Ser Thr Phe Lys Trp Ser Tyr Thr Thr Gly Asn Gly Lys Arg
165 170 175

Tyr Arg Ser Thr Ala Arg Thr Thr Tyr Thr Phe Ala Lys Pro Met Ala
180 185 190

Ala Asn Tyr Leu Lys Asn Gln Pro Met Tyr Val Phe Arg Lys Thr Glu
195 200 205

Leu Lys His Ser Lys Thr Glu Leu Asn Phe Lys Glu Trp Gln Lys Ala
210 215 220

Phe Thr Asp Val Met Gly Met Asp Glu Leu Tyr Lys
225 230 235

<210> SEQ ID NO 13

<211> LENGTH: 236

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: mCherry

<400> SEQUENCE: 13

Met Val Ser Lys Gly Glu Glu Asp Asn Met Ala Ile Ile Lys Glu Phe
1 5 10 15

Met Arg Phe Lys Val His Met Glu Gly Ser Val Asn Gly His Glu Phe
20 25 30

Glu Ile Glu Gly Glu Gly Arg Pro Tyr Glu Gly Thr Gln Thr
35 40 45

Ala Lys Leu Lys Val Thr Lys Gly Gly Pro Leu Pro Phe Ala Trp Asp
50 55 60

Ile Leu Ser Pro Gln Phe Met Tyr Gly Ser Lys Ala Tyr Val Lys His
65 70 75 80

Pro Ala Asp Ile Pro Asp Tyr Leu Lys Leu Ser Phe Pro Glu Gly Phe
85 90 95

Lys Trp Glu Arg Val Met Asn Phe Glu Asp Gly Gly Val Val Thr Val
100 105 110

Thr Gln Asp Ser Ser Leu Gln Asp Gly Glu Phe Ile Tyr Lys Val Lys
115 120 125

Leu Arg Gly Thr Asn Phe Pro Ser Asp Gly Pro Val Met Gln Lys Lys
130 135 140

Thr Met Gly Trp Glu Ala Ser Ser Glu Arg Met Tyr Pro Glu Asp Gly
145 150 155 160

Ala Leu Lys Gly Glu Ile Lys Gln Arg Leu Lys Leu Lys Asp Gly Gly
165 170 175

His Tyr Asp Ala Glu Val Lys Thr Thr Tyr Lys Ala Lys Lys Pro Val
180 185 190

Gln Leu Pro Gly Ala Tyr Asn Val Asn Ile Lys Leu Asp Ile Thr Ser
195 200 205

His Asn Glu Asp Tyr Thr Ile Val Glu Gln Tyr Glu Arg Ala Glu Gly
210 215 220

Arg His Ser Thr Gly Gly Met Asp Glu Leu Tyr Lys
225 230 235

<210> SEQ ID NO 14

<211> LENGTH: 237

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

-continued

<223> OTHER INFORMATION: mTagBFP

<400> SEQUENCE: 14

Met Val Ser Lys Gly Glu Glu Leu Ile Lys Glu Asn Met His Met Lys
 1 5 10 15

Leu Tyr Met Glu Gly Thr Val Asp Asn His His Phe Lys Cys Thr Ser
 20 25 30

Glu Gly Glu Gly Lys Pro Tyr Glu Gly Thr Gln Thr Met Arg Ile Lys
 35 40 45

Val Val Glu Gly Pro Leu Pro Phe Ala Phe Asp Ile Leu Ala Thr
 50 55 60

Ser Phe Leu Tyr Gly Ser Lys Thr Phe Ile Asn His Thr Gln Gly Ile
 65 70 75 80

Pro Asp Phe Phe Lys Gln Ser Phe Pro Glu Gly Phe Thr Trp Glu Arg
 85 90 95

Val Thr Thr Tyr Glu Asp Gly Gly Val Leu Thr Ala Thr Gln Asp Thr
 100 105 110

Ser Leu Gln Asp Gly Cys Leu Ile Tyr Asn Val Lys Ile Arg Gly Val
 115 120 125

Asn Phe Thr Ser Asn Gly Pro Val Met Gln Lys Lys Thr Leu Gly Trp
 130 135 140

Glu Ala Phe Thr Glu Thr Leu Tyr Pro Ala Asp Gly Gly Leu Glu Gly
 145 150 155 160

Arg Asn Asp Met Ala Leu Lys Leu Val Gly Gly Ser His Leu Ile Ala
 165 170 175

Asn Ala Lys Thr Thr Tyr Arg Ser Lys Lys Pro Ala Lys Asn Leu Lys
 180 185 190

Met Pro Gly Val Tyr Tyr Val Asp Tyr Arg Leu Glu Arg Ile Lys Glu
 195 200 205

Ala Asn Asn Glu Thr Tyr Val Glu Gln His Glu Val Ala Val Ala Arg
 210 215 220

Tyr Cys Asp Leu Pro Ser Lys Leu Gly His Lys Leu Asn
 225 230 235

<210> SEQ ID NO 15

<211> LENGTH: 239

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: Venus

<400> SEQUENCE: 15

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

Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
 20 25 30

Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Leu Ile
 35 40 45

Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
 50 55 60

Leu Gly Tyr Gly Leu Gln Cys Phe Ala Arg Tyr Pro Asp His Met Lys
 65 70 75 80

Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
 85 90 95

Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
 100 105 110

-continued

Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
115 120 125

Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
130 135 140

Asn Tyr Asn Ser His Asn Val Tyr Ile Thr Ala Asp Lys Gln Lys Asn
145 150 155 160

Gly Ile Lys Ala Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Gly
165 170 175

Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
180 185 190

Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Tyr Gln Ser Ala Leu
195 200 205

Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
210 215 220

Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys
225 230 235

<210> SEQ ID NO 16
<211> LENGTH: 239
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: mVenus

<400> SEQUENCE: 16

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

Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
20 25 30

Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Leu Ile
35 40 45

Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
50 55 60

Leu Gly Tyr Gly Leu Gln Cys Phe Ala Arg Tyr Pro Asp His Met Lys
65 70 75 80

Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
85 90 95

Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
100 105 110

Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
115 120 125

Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
130 135 140

Asn Tyr Asn Ser His Asn Val Tyr Ile Thr Ala Asp Lys Gln Lys Asn
145 150 155 160

Gly Ile Lys Ala Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Gly
165 170 175

Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
180 185 190

Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Tyr Gln Ser Lys Leu
195 200 205

Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
210 215 220

Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys
225 230 235

-continued

<210> SEQ ID NO 17
<211> LENGTH: 239
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: mTurquoise

<400> SEQUENCE: 17

```

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

Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
20          25          30

Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
35          40          45

Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
50          55          60

Leu Ser Trp Gly Val Gln Cys Phe Ala Arg Tyr Pro Asp His Met Lys
65          70          75          80

Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
85          90          95

Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
100         105         110

Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
115         120         125

Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
130         135         140

Asn Tyr Ile Ser Asp Asn Val Tyr Ile Thr Ala Asp Lys Gln Lys Asn
145         150         155         160

Gly Ile Lys Ala Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Gly
165         170         175

Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
180         185         190

Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Lys Leu
195         200         205

Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Glu Phe
210         215         220

Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys
225         230         235

```

<210> SEQ ID NO 18
<211> LENGTH: 232
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: mScarlet

<400> SEQUENCE: 18

```

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

Val His Met Glu Gly Ser Met Asn Gly His Glu Phe Glu Ile Glu Gly
20          25          30

Glu Gly Glu Gly Arg Pro Tyr Glu Gly Thr Gln Thr Ala Lys Leu Lys
35          40          45

Val Thr Lys Gly Gly Pro Leu Pro Phe Ser Trp Asp Ile Leu Ser Pro
50          55          60

Gln Phe Met Tyr Gly Ser Arg Ala Phe Thr Lys His Pro Ala Asp Ile

```

US 12,391,948 B2

139**140**

-continued

65	70	75	80
Pro Asp Tyr Tyr Lys Gln Ser Phe Pro Glu Gly Phe Lys Trp Glu Arg			
85		90	95
Val Met Asn Phe Glu Asp Gly Gly Ala Val Thr Val Thr Gln Asp Thr			
100		105	110
Ser Leu Glu Asp Gly Thr Leu Ile Tyr Lys Val Lys Leu Arg Gly Thr			
115		120	125
Asn Phe Pro Pro Asp Gly Pro Val Met Gln Lys Lys Thr Met Gly Trp			
130		135	140
Glu Ala Ser Thr Glu Arg Leu Tyr Pro Glu Asp Gly Val Leu Lys Gly			
145		150	155
Asp Ile Lys Met Ala Leu Arg Leu Lys Asp Gly Gly Arg Tyr Leu Ala			
165		170	175
Asp Phe Lys Thr Thr Tyr Lys Ala Lys Lys Pro Val Gln Met Pro Gly			
180		185	190
Ala Tyr Asn Val Asp Arg Lys Leu Asp Ile Thr Ser His Asn Glu Asp			
195		200	205
Tyr Thr Val Val Glu Gln Tyr Glu Arg Ser Glu Gly Arg His Ser Thr			
210		215	220
Gly Gly Met Asp Glu Leu Tyr Lys			
225		230	

<210> SEQ ID NO 19
<211> LENGTH: 236
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: mWasabi

<400> SEQUENCE: 19

Met Val Ser Lys Gly Glu Glu Thr Thr Met Gly Val Ile Lys Pro Asp			
1	5	10	15
Met Lys Ile Lys Leu Lys Met Glu Gly Asn Val Asn Gly His Ala Phe			
20	25	30	
Val Ile Glu Gly Glu Gly Lys Pro Tyr Asp Gly Thr Asn Thr			
35	40	45	
Ile Asn Leu Glu Val Lys Glu Gly Ala Pro Leu Pro Phe Ser Tyr Asp			
50	55	60	
Ile Leu Thr Thr Ala Phe Ser Tyr Gly Asn Arg Ala Phe Thr Lys Tyr			
65	70	75	80
Pro Asp Asp Ile Pro Asn Tyr Phe Lys Gln Ser Phe Pro Glu Gly Tyr			
85	90	95	
Ser Trp Glu Arg Thr Met Thr Phe Glu Asp Lys Gly Ile Val Lys Val			
100	105	110	
Lys Ser Asp Ile Ser Met Glu Glu Asp Ser Phe Ile Tyr Glu Ile His			
115	120	125	
Leu Lys Gly Glu Asn Phe Pro Pro Asn Gly Pro Val Met Gln Lys Glu			
130	135	140	
Thr Thr Gly Trp Asp Ala Ser Thr Glu Arg Met Tyr Val Arg Asp Gly			
145	150	155	160
Val Leu Lys Gly Asp Val Lys Met Lys Leu Leu Glu Gly Gly			
165	170	175	
His His Arg Val Asp Phe Lys Thr Ile Tyr Arg Ala Lys Lys Ala Val			
180	185	190	
Lys Leu Pro Asp Tyr His Phe Val Asp His Arg Ile Glu Ile Leu Asn			

US 12,391,948 B2

141

-continued

195	200	205
-----	-----	-----

His Asp Lys Asp Tyr Asn Lys Val Thr Val Tyr Glu Ile Ala Val Ala		
210	215	220

Arg Asn Ser Thr Asp Gly Met Asp Glu Leu Tyr Lys		
225	230	235

<210> SEQ ID NO 20
<211> LENGTH: 236
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: mOrange

<400> SEQUENCE: 20

Met Val Ser Lys Gly Glu Glu Asn Asn Met Ala Ile Ile Lys Glu Phe			
1	5	10	15

Met Arg Phe Lys Val Arg Met Glu Gly Ser Val Asn Gly His Glu Phe		
20	25	30

Glu Ile Glu Gly Glu Gly Glu Arg Pro Tyr Glu Gly Phe Gln Thr		
35	40	45

Ala Lys Leu Lys Val Thr Lys Gly Gly Pro Leu Pro Phe Ala Trp Asp		
50	55	60

Ile Leu Ser Pro Gln Phe Thr Tyr Gly Ser Lys Ala Tyr Val Lys His			
65	70	75	80

Pro Ala Asp Ile Pro Asp Tyr Phe Lys Leu Ser Phe Pro Glu Gly Phe		
85	90	95

Lys Trp Glu Arg Val Met Asn Phe Glu Asp Gly Gly Val Val Thr Val		
100	105	110

Thr Gln Asp Ser Ser Leu Gln Asp Gly Glu Phe Ile Tyr Lys Val Lys		
115	120	125

Leu Arg Gly Thr Asn Phe Pro Ser Asp Gly Pro Val Met Gln Lys Lys		
130	135	140

Thr Met Gly Trp Glu Ala Ser Ser Glu Arg Met Tyr Pro Glu Asp Gly			
145	150	155	160

Ala Leu Lys Gly Glu Ile Lys Met Arg Leu Lys Leu Lys Asp Gly Gly		
165	170	175

His Tyr Thr Ser Glu Val Lys Thr Thr Tyr Lys Ala Lys Lys Pro Val		
180	185	190

Gln Leu Pro Gly Ala Tyr Ile Val Gly Ile Lys Leu Asp Ile Thr Ser		
195	200	205

His Asn Glu Asp Tyr Thr Ile Val Glu Gln Tyr Glu Arg Ala Glu Gly		
210	215	220

Arg His Ser Thr Gly Gly Met Asp Glu Leu Tyr Lys		
225	230	235

<210> SEQ ID NO 21
<211> LENGTH: 234
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: dTomato

<400> SEQUENCE: 21

Met Val Ser Lys Gly Glu Glu Val Ile Lys Glu Phe Met Arg Phe Lys			
1	5	10	15

Val Arg Met Glu Gly Ser Met Asn Gly His Glu Phe Glu Ile Glu Gly		
20	25	30

142

-continued

Glu Gly Glu Gly Arg Pro Tyr Glu Gly Thr Gln Thr Ala Lys Leu Lys
35 40 45

Val Thr Lys Gly Gly Pro Leu Pro Phe Ala Trp Asp Ile Leu Ser Pro
50 55 60

Gln Phe Met Tyr Gly Ser Lys Ala Tyr Val Lys His Pro Ala Asp Ile
65 70 75 80

Pro Asp Tyr Lys Lys Leu Ser Phe Pro Glu Gly Phe Lys Trp Glu Arg
85 90 95

Val Met Asn Phe Glu Asp Gly Gly Leu Val Thr Val Thr Gln Asp Ser
100 105 110

Ser Leu Gln Asp Gly Thr Leu Ile Tyr Lys Val Lys Met Arg Gly Thr
115 120 125

Asn Phe Pro Pro Asp Gly Pro Val Met Gln Lys Lys Thr Met Gly Trp
130 135 140

Glu Ala Ser Thr Glu Arg Leu Tyr Pro Arg Asp Gly Val Leu Lys Gly
145 150 155 160

Glu Ile His Gln Ala Leu Lys Leu Lys Asp Gly Gly His Tyr Leu Val
165 170 175

Glu Phe Lys Thr Ile Tyr Met Ala Lys Lys Pro Val Gln Leu Pro Gly
180 185 190

Tyr Tyr Tyr Val Asp Thr Lys Leu Asp Ile Thr Ser His Asn Glu Asp
195 200 205

Tyr Thr Ile Val Glu Gln Tyr Glu Arg Ser Glu Gly Arg His His Leu
210 215 220

Phe Leu Tyr Gly Met Asp Glu Leu Tyr Lys
225 230

<210> SEQ ID NO 22
<211> LENGTH: 171
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Nanoluc luciferase

<400> SEQUENCE: 22

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

Gly Tyr Asn Leu Asp Gln Val Leu Glu Gln Gly Gly Val Ser Ser Leu
20 25 30

Phe Gln Asn Leu Gly Val Ser Val Thr Pro Ile Gln Arg Ile Val Leu
35 40 45

Ser Gly Glu Asn Gly Leu Lys Ile Asp Ile His Val Ile Ile Pro Tyr
50 55 60

Glu Gly Leu Ser Gly Asp Gln Met Gly Gln Ile Glu Lys Ile Phe Lys
65 70 75 80

Val Val Tyr Pro Val Asp Asp His His Phe Lys Val Ile Leu His Tyr
85 90 95

Gly Thr Leu Val Ile Asp Gly Val Thr Pro Asn Met Ile Asp Tyr Phe
100 105 110

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

Val Thr Gly Thr Leu Trp Asn Gly Asn Lys Ile Ile Asp Glu Arg Leu
130 135 140

Ile Asn Pro Asp Gly Ser Leu Leu Phe Arg Val Thr Ile Asn Gly Val
145 150 155 160

US 12,391,948 B2

145

-continued

Thr Gly Trp Arg Leu Cys Glu Arg Ile Leu Ala
165 170

146

<210> SEQ ID NO 23
<211> LENGTH: 550
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Firefly luciferase

<400> SEQUENCE: 23

Met Glu Asp Ala Lys Asn Ile Lys Lys Gly Pro Ala Pro Phe Tyr Pro
1 5 10 15

Leu Glu Asp Gly Thr Ala Gly Glu Gln Leu His Lys Ala Met Lys Arg
20 25 30

Tyr Ala Leu Val Pro Gly Thr Ile Ala Phe Thr Asp Ala His Ile Glu
35 40 45

Val Asn Ile Thr Tyr Ala Glu Tyr Phe Glu Met Ser Val Arg Leu Ala
50 55 60

Glu Ala Met Lys Arg Tyr Gly Leu Asn Thr Asn His Arg Ile Val Val
65 70 75 80

Cys Ser Glu Asn Ser Leu Gln Phe Phe Met Pro Val Leu Gly Ala Leu
85 90 95

Phe Ile Gly Val Ala Val Ala Pro Ala Asn Asp Ile Tyr Asn Glu Arg
100 105 110

Glu Leu Leu Asn Ser Met Asn Ile Ser Gln Pro Thr Val Val Phe Val
115 120 125

Ser Lys Lys Gly Leu Gln Lys Ile Leu Asn Val Gln Lys Lys Leu Pro
130 135 140

Ile Ile Gln Lys Ile Ile Ile Met Asp Ser Lys Thr Asp Tyr Gln Gly
145 150 155 160

Phe Gln Ser Met Tyr Thr Phe Val Thr Ser His Leu Pro Pro Gly Phe
165 170 175

Asn Glu Tyr Asp Phe Val Pro Glu Ser Phe Asp Arg Asp Lys Thr Ile
180 185 190

Ala Leu Ile Met Asn Ser Ser Gly Ser Thr Gly Leu Pro Lys Gly Val
195 200 205

Ala Leu Pro His Arg Thr Ala Cys Val Arg Phe Ser His Ala Arg Asp
210 215 220

Pro Ile Phe Gly Asn Gln Ile Ile Pro Asp Thr Ala Ile Leu Ser Val
225 230 235 240

Val Pro Phe His His Gly Phe Gly Met Phe Thr Thr Leu Gly Tyr Leu
245 250 255

Ile Cys Gly Phe Arg Val Val Leu Met Tyr Arg Phe Glu Glu Leu
260 265 270

Phe Leu Arg Ser Leu Gln Asp Tyr Lys Ile Gln Ser Ala Leu Leu Val
275 280 285

Pro Thr Leu Phe Ser Phe Phe Ala Lys Ser Thr Leu Ile Asp Lys Tyr
290 295 300

Asp Leu Ser Asn Leu His Glu Ile Ala Ser Gly Gly Ala Pro Leu Ser
305 310 315 320

Lys Glu Val Gly Glu Ala Val Ala Lys Arg Phe His Leu Pro Gly Ile
325 330 335

Arg Gln Gly Tyr Gly Leu Thr Glu Thr Thr Ser Ala Ile Leu Ile Thr
340 345 350

US 12,391,948 B2

147**148**

-continued

Pro Glu Gly Asp Asp Lys Pro Gly Ala Val Gly Lys Val Val Pro Phe
355 360 365

Phe Glu Ala Lys Val Val Asp Leu Asp Thr Gly Lys Thr Leu Gly Val
370 375 380

Asn Gln Arg Gly Glu Leu Cys Val Arg Gly Pro Met Ile Met Ser Gly
385 390 395 400

Tyr Val Asn Asn Pro Glu Ala Thr Asn Ala Leu Ile Asp Lys Asp Gly
405 410 415

Trp Leu His Ser Gly Asp Ile Ala Tyr Trp Asp Glu Asp Glu His Phe
420 425 430

Phe Ile Val Asp Arg Leu Lys Ser Leu Ile Lys Tyr Lys Gly Tyr Gln
435 440 445

Val Ala Pro Ala Glu Leu Glu Ser Ile Leu Leu Gln His Pro Asn Ile
450 455 460

Phe Asp Ala Gly Val Ala Gly Leu Pro Asp Asp Asp Ala Gly Glu Leu
465 470 475 480

Pro Ala Ala Val Val Val Leu Glu His Gly Lys Thr Met Thr Glu Lys
485 490 495

Glu Ile Val Asp Tyr Val Ala Ser Gln Val Thr Thr Ala Lys Lys Leu
500 505 510

Arg Gly Val Val Phe Val Asp Glu Val Pro Lys Gly Leu Thr Gly
515 520 525

Lys Leu Asp Ala Arg Lys Ile Arg Glu Ile Leu Ile Lys Ala Lys Lys
530 535 540

Gly Gly Lys Ser Lys Leu
545 550

<210> SEQ ID NO 24
<211> LENGTH: 311
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Renilla luciferase

<400> SEQUENCE: 24

Met Ala Ser Lys Val Tyr Asp Pro Glu Gln Arg Lys Arg Met Ile Thr
1 5 10 15

Gly Pro Gln Trp Trp Ala Arg Cys Lys Gln Met Asn Val Leu Asp Ser
20 25 30

Phe Ile Asn Tyr Tyr Asp Ser Glu Lys His Ala Glu Asn Ala Val Ile
35 40 45

Phe Leu His Gly Asn Ala Ala Ser Ser Tyr Leu Trp Arg His Val Val
50 55 60

Pro His Ile Glu Pro Val Ala Arg Cys Ile Ile Pro Asp Leu Ile Gly
65 70 75 80

Met Gly Lys Ser Gly Lys Ser Gly Asn Gly Ser Tyr Arg Leu Leu Asp
85 90 95

His Tyr Lys Tyr Leu Thr Ala Trp Phe Glu Leu Leu Asn Leu Pro Lys
100 105 110

Lys Ile Ile Phe Val Gly His Asp Trp Gly Ala Cys Leu Ala Phe His
115 120 125

Tyr Ser Tyr Glu His Gln Asp Lys Ile Lys Ala Ile Val His Ala Glu
130 135 140

Ser Val Val Asp Val Ile Glu Ser Trp Asp Glu Trp Pro Asp Ile Glu
145 150 155 160

US 12,391,948 B2

149**150**

-continued

Glu Asp Ile Ala Leu Ile Lys Ser Glu Glu Gly Glu Lys Met Val Leu
165 170 175

Glu Asn Asn Phe Phe Val Glu Thr Met Leu Pro Ser Lys Ile Met Arg
180 185 190

Lys Leu Glu Pro Glu Glu Phe Ala Ala Tyr Leu Glu Pro Phe Lys Glu
195 200 205

Lys Gly Glu Val Arg Arg Pro Thr Leu Ser Trp Pro Arg Glu Ile Pro
210 215 220

Leu Val Lys Gly Gly Lys Pro Asp Val Val Gln Ile Val Arg Asn Tyr
225 230 235 240

Asn Ala Tyr Leu Arg Ala Ser Asp Asp Leu Pro Lys Met Phe Ile Glu
245 250 255

Ser Asp Pro Gly Phe Phe Ser Asn Ala Ile Val Glu Gly Ala Lys Lys
260 265 270

Phe Pro Asn Thr Glu Phe Val Lys Val Lys Gly Leu His Phe Ser Gln
275 280 285

Glu Asp Ala Pro Asp Glu Met Gly Lys Tyr Ile Lys Ser Phe Val Glu
290 295 300

Arg Val Leu Lys Asn Glu Gln
305 310

<210> SEQ ID NO 25
<211> LENGTH: 185
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: *Gaussia luciferase*

<400> SEQUENCE: 25

Met Gly Val Lys Val Leu Phe Ala Leu Ile Cys Ile Ala Val Ala Glu
1 5 10 15

Ala Lys Pro Thr Glu Asn Asn Glu Asp Phe Asn Ile Val Ala Val Ala
20 25 30

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

Gly Lys Lys Leu Pro Leu Glu Val Leu Lys Glu Met Glu Ala Asn Ala
50 55 60

Arg Lys Ala Gly Cys Thr Arg Gly Cys Leu Ile Cys Leu Ser His Ile
65 70 75 80

Lys Cys Thr Pro Lys Met Lys Phe Ile Pro Gly Arg Cys His Thr
85 90 95

Tyr Glu Gly Asp Lys Glu Ser Ala Gln Gly Gly Ile Gly Glu Ala Ile
100 105 110

Val Asp Ile Pro Glu Ile Pro Gly Phe Lys Asp Leu Glu Pro Met Glu
115 120 125

Gln Phe Ile Ala Gln Val Asp Leu Cys Val Asp Cys Thr Thr Gly Cys
130 135 140

Leu Lys Gly Leu Ala Asn Val Gln Cys Ser Asp Leu Leu Lys Lys Trp
145 150 155 160

Leu Pro Gln Arg Cys Ala Thr Phe Ala Ser Lys Ile Gln Gly Gln Val
165 170 175

Asp Lys Ile Lys Gly Ala Gly Gly Asp
180 185

<210> SEQ ID NO 26
<211> LENGTH: 1015

US 12,391,948 B2

151

-continued

<212> TYPE: PRT
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: Beta-galactosidase
 <400> SEQUENCE: 26

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

152

-continued

Pro Asn His Pro Leu Trp Tyr Thr Leu Cys Asp Arg Tyr Gly Leu Tyr
 385 390 395 400

 Val Val Asp Glu Ala Asn Ile Glu Thr His Gly Met Val Pro Met Asn
 405 410 415

 Arg Leu Thr Asp Asp Pro Arg Trp Leu Pro Ala Met Ser Glu Arg Val
 420 425 430

 Thr Arg Met Val Gln Arg Asp Arg Asn His Pro Ser Val Ile Ile Trp
 435 440 445

 Ser Leu Gly Asn Glu Ser Gly His Gly Ala Asn His Asp Ala Leu Tyr
 450 455 460

 Arg Trp Ile Lys Ser Val Asp Pro Ser Arg Pro Val Gln Tyr Glu Gly
 465 470 475 480

 Gly Gly Ala Asp Thr Thr Ala Thr Asp Ile Ile Cys Pro Met Tyr Ala
 485 490 495

 Arg Val Asp Glu Asp Gln Pro Phe Pro Ala Val Pro Lys Trp Ser Ile
 500 505 510

 Lys Lys Trp Leu Ser Leu Pro Gly Glu Thr Arg Pro Leu Ile Leu Cys
 515 520 525

 Glu Tyr Ala His Ala Met Gly Asn Ser Leu Gly Gly Phe Ala Lys Tyr
 530 535 540

 Trp Gln Ala Phe Arg Gln Tyr Pro Arg Leu Gln Gly Gly Phe Val Trp
 545 550 555 560

 Asp Trp Val Asp Gln Ser Leu Ile Lys Tyr Asp Glu Asn Gly Asn Pro
 565 570 575

 Trp Ser Ala Tyr Gly Gly Asp Phe Gly Asp Thr Pro Asn Asp Arg Gln
 580 585 590

 Phe Cys Met Asn Gly Leu Val Phe Ala Asp Arg Thr Pro His Pro Ala
 595 600 605

 Leu Thr Glu Ala Lys His Gln Gln Gln Phe Phe Gln Phe Arg Leu Ser
 610 615 620

 Gly Gln Thr Ile Glu Val Thr Ser Glu Tyr Leu Phe Arg His Ser Asp
 625 630 635 640

 Asn Glu Leu Leu His Trp Met Val Ala Leu Asp Gly Lys Pro Leu Ala
 645 650 655

 Ser Gly Glu Val Pro Leu Asp Val Ala Pro Gln Gly Lys Gln Leu Ile
 660 665 670

 Glu Leu Pro Glu Leu Pro Gln Pro Glu Ser Ala Gly Gln Leu Trp Leu
 675 680 685

 Thr Val Arg Val Val Gln Pro Asn Ala Thr Ala Trp Ser Glu Ala Gly
 690 695 700

 His Ile Ser Ala Trp Gln Gln Trp Arg Leu Ala Glu Asn Leu Ser Val
 705 710 715 720

 Thr Leu Pro Ala Ala Ser His Ala Ile Pro His Leu Thr Thr Ser Glu
 725 730 735

 Met Asp Phe Cys Ile Glu Leu Gly Asn Lys Arg Trp Gln Phe Asn Arg
 740 745 750

 Gln Ser Gly Phe Leu Ser Gln Met Trp Ile Gly Asp Lys Lys Gln Leu
 755 760 765

 Leu Thr Pro Leu Arg Asp Gln Phe Thr Arg Ala Pro Leu Asp Asn Asp
 770 775 780

 Ile Gly Val Ser Glu Ala Thr Arg Ile Asp Pro Asn Ala Trp Val Glu
 785 790 795 800

 Arg Trp Lys Ala Ala Gly His Tyr Gln Ala Glu Ala Ala Leu Leu Gln

US 12,391,948 B2

155**156**

-continued

805	810	815	
Cys Thr Ala Asp Thr Leu Ala Asp Ala Val	Leu Ile Thr Thr Ala His		
820	825	830	
Ala Trp Gln His Gln Gly Lys Thr Leu Phe Ile Ser Arg	Lys Thr Tyr		
835	840	845	
Arg Ile Asp Gly Ser Gly Gln Met Ala Ile Thr Val Asp Val	Glu Val		
850	855	860	
Ala Ser Asp Thr Pro His Pro Ala Arg Ile Gly Leu Asn Cys	Gln Leu		
865	870	875	880
Ala Gln Val Ala Glu Arg Val Asn Trp Leu Gly Leu Gly	Pro Gln Glu		
885	890	895	
Asn Tyr Pro Asp Arg Leu Thr Ala Ala Cys Phe Asp Arg	Trp Asp Leu		
900	905	910	
Pro Leu Ser Asp Met Tyr Thr Pro Tyr Val Phe Pro Ser	Glu Asn Gly		
915	920	925	
Leu Arg Cys Gly Thr Arg Glu Leu Asn Tyr Gly Pro His	Gln Trp Arg		
930	935	940	
Gly Asp Phe Gln Phe Asn Ile Ser Arg Tyr Ser Gln Gln	Gln Leu Met		
945	950	955	960
Glu Thr Ser His Arg His Leu Leu His Ala Glu Glu Gly	Thr Trp Leu		
965	970	975	
Asn Ile Asp Gly Phe His Met Gly Ile Gly Gly Asp Asp	Ser Trp Ser		
980	985	990	
Pro Ser Val Ser Ala Glu Phe Gln	Leu Ser Ala Gly Arg	Tyr His Tyr	
995	1000	1005	
Gln Leu Val Trp Cys Gln Lys			
1010	1015		

<210> SEQ ID NO 27

<211> LENGTH: 286

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: Beta-lactamase

<400> SEQUENCE: 27

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

US 12,391,948 B2

157

-continued

158

145	150	155	160
Asp Arg Trp Glu Pro Glu Leu Asn Glu Ala Ile Pro Asn Asp Glu Arg			
165	170	175	
Asp Thr Thr Met Pro Val Ala Met Ala Thr Thr Leu Arg Lys Leu Leu			
180	185	190	
Thr Gly Glu Leu Leu Thr Leu Ala Ser Arg Gln Gln Leu Ile Asp Trp			
195	200	205	
Met Glu Ala Asp Lys Val Ala Gly Pro Leu Leu Arg Ser Ala Leu Pro			
210	215	220	
Ala Gly Trp Phe Ile Ala Asp Lys Ser Gly Ala Gly Glu Arg Gly Ser			
225	230	235	240
Arg Gly Ile Ile Ala Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg Ile			
245	250	255	
Val Val Ile Tyr Thr Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn			
260	265	270	
Arg Gln Ile Ala Glu Ile Gly Ala Ser Leu Ile Lys His Trp			
275	280	285	

<210> SEQ ID NO: 28
<211> LENGTH: 250
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Ascorbate peroxidase 1, cytosolic

<400> SEQUENCE: 28

Met Gly Lys Ser Tyr Pro Thr Val Ser Ala Asp Tyr Gln Lys Ala Val			
1	5	10	15
Glu Lys Ala Lys Lys Lys Leu Arg Gly Phe Ile Ala Glu Lys Arg Cys			
20	25	30	
Ala Pro Leu Met Leu Arg Leu Ala Trp His Ser Ala Gly Thr Phe Asp			
35	40	45	
Lys Gly Thr Lys Thr Gly Gly Pro Phe Gly Thr Ile Lys His Pro Ala			
50	55	60	
Glu Leu Ala His Ser Ala Asn Asn Gly Leu Asp Ile Ala Val Arg Leu			
65	70	75	80
Leu Glu Pro Leu Lys Ala Glu Phe Pro Ile Leu Ser Tyr Ala Asp Phe			
85	90	95	
Tyr Gln Leu Ala Gly Val Val Ala Val Glu Val Thr Gly Gly Pro Glu			
100	105	110	
Val Pro Phe His Pro Gly Arg Glu Asp Lys Pro Glu Pro Pro Pro Glu			
115	120	125	
Gly Arg Leu Pro Asp Ala Thr Lys Gly Ser Asp His Leu Arg Asp Val			
130	135	140	
Phe Gly Lys Ala Met Gly Leu Thr Asp Gln Asp Ile Val Ala Leu Ser			
145	150	155	160
Gly Gly His Thr Ile Gly Ala Ala His Lys Glu Arg Ser Gly Phe Glu			
165	170	175	
Gly Pro Trp Thr Ser Asn Pro Leu Ile Phe Asp Asn Ser Tyr Phe Thr			
180	185	190	
Glu Leu Leu Ser Gly Glu Lys Glu Gly Leu Leu Gln Leu Pro Ser Asp			
195	200	205	
Lys Ala Leu Leu Ser Asp Pro Val Phe Arg Pro Leu Val Asp Lys Tyr			
210	215	220	
Ala Ala Asp Glu Asp Ala Phe Phe Ala Asp Tyr Ala Glu Ala His Gln			

US 12,391,948 B2

159**160**

-continued

225	230	235	240
-----	-----	-----	-----

Lys Leu Ser Glu Leu Gly Phe Ala Asp Ala
245 250

<210> SEQ ID NO 29

<211> LENGTH: 250

<212> TYPE: PRT

<213> ORGANISM: Arabidopsis thaliana

<400> SEQUENCE: 29

Met Thr Lys Asn Tyr Pro Thr Val Ser Glu Asp Tyr Lys Ala Val	1	15
5 10		

Glu Lys Cys Arg Arg Lys Leu Arg Gly Leu Ile Ala Glu Lys Asn Cys	20	25	30

Ala Pro Ile Met Val Arg Leu Ala Trp His Ser Ala Gly Thr Phe Asp	35	40	45

Cys Gln Ser Arg Thr Gly Gly Pro Phe Gly Thr Met Arg Phe Asp Ala	50	55	60

Glu Gln Ala His Gly Ala Asn Ser Gly Ile His Ile Ala Leu Arg Leu	65	70	75	80

Leu Asp Pro Ile Arg Glu Gln Phe Pro Thr Ile Ser Phe Ala Asp Phe	85	90	95

His Gln Leu Ala Gly Val Val Ala Val Glu Val Thr Gly Pro Asp	100	105	110

Ile Pro Phe His Pro Gly Arg Glu Asp Lys Pro Gln Pro Pro Glu	115	120	125

Gly Arg Leu Pro Asp Ala Thr Lys Gly Cys Asp His Leu Arg Asp Val	130	135	140

Phe Ala Lys Gln Met Gly Leu Ser Asp Lys Asp Ile Val Ala Leu Ser	145	150	155	160

Gly Ala His Thr Leu Gly Arg Cys His Lys Asp Arg Ser Gly Phe Glu	165	170	175

Gly Ala Trp Thr Ser Asn Pro Leu Ile Phe Asp Asn Ser Tyr Phe Lys	180	185	190

Glu Leu Leu Ser Gly Glu Lys Glu Gly Leu Leu Gln Leu Val Ser Asp	195	200	205

Lys Ala Leu Leu Asp Asp Pro Val Phe Arg Pro Leu Val Glu Lys Tyr	210	215	220

Ala Ala Asp Glu Asp Ala Phe Phe Ala Asp Tyr Ala Glu Ala His Met	225	230	235	240

Lys Leu Ser Glu Leu Gly Phe Ala Asp Ala
245 250

<210> SEQ ID NO 30

<211> LENGTH: 251

<212> TYPE: PRT

<213> ORGANISM: Arabidopsis thaliana

<400> SEQUENCE: 30

Met Val Lys Lys Ser Tyr Pro Glu Val Lys Glu Glu Tyr Lys Lys Ala	1	5	10	15

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

Cys Ala Pro Ile Val Leu Arg Leu Ala Trp His Ser Ala Gly Thr Phe	35	40	45

Asp Val Lys Thr Lys Thr Gly Gly Pro Phe Gly Thr Ile Arg His Pro

US 12,391,948 B2

161**162**

-continued

50 55 60

Gln Glu Leu Ala His Asp Ala Asn Asn Gly Leu Asp Ile Ala Val Arg
 65 70 75 80

Leu Leu Asp Pro Ile Lys Glu Leu Phe Pro Ile Leu Ser Tyr Ala Asp
 85 90 95

Phe Tyr Gln Leu Ala Gly Val Val Ala Val Glu Ile Thr Gly Gly Pro
 100 105 110

Glu Ile Pro Phe His Pro Gly Arg Leu Asp Lys Val Glu Pro Pro Pro
 115 120 125

Glu Gly Arg Leu Pro Gln Ala Thr Lys Gly Val Asp His Leu Arg Asp
 130 135 140

Val Phe Gly Arg Met Gly Leu Asn Asp Lys Asp Ile Val Ala Leu Ser
 145 150 155 160

Gly Gly His Thr Leu Gly Arg Cys His Lys Glu Arg Ser Gly Phe Glu
 165 170 175

Gly Ala Trp Thr Pro Asn Pro Ile Phe Asp Asn Ser Tyr Phe Lys
 180 185 190

Glu Ile Leu Ser Gly Glu Lys Glu Gly Leu Leu Gln Leu Pro Thr Asp
 195 200 205

Lys Ala Leu Leu Asp Asp Pro Leu Phe Leu Pro Phe Val Glu Lys Tyr
 210 215 220

Ala Ala Asp Glu Asp Ala Phe Phe Glu Asp Tyr Thr Glu Ala His Leu
 225 230 235 240

Lys Leu Ser Glu Leu Gly Phe Ala Asp Lys Glu
 245 250

<210> SEQ ID NO 31

<211> LENGTH: 250

<212> TYPE: PRT

<213> ORGANISM: Pisum sativum

<400> SEQUENCE: 31

Met Gly Lys Ser Tyr Pro Thr Val Ser Pro Asp Tyr Gln Lys Ala Ile
 1 5 10 15

Glu Lys Ala Lys Arg Lys Leu Arg Gly Phe Ile Ala Glu Lys Lys Cys
 20 25 30

Ala Pro Leu Ile Leu Arg Leu Ala Trp His Ser Ala Gly Thr Phe Asp
 35 40 45

Ser Lys Thr Lys Thr Gly Gly Pro Phe Gly Thr Ile Lys His Gln Ala
 50 55 60

Glu Leu Ala His Gly Ala Asn Asn Gly Leu Asp Ile Ala Val Arg Leu
 65 70 75 80

Leu Glu Pro Ile Lys Glu Gln Phe Pro Ile Val Ser Tyr Ala Asp Phe
 85 90 95

Tyr Gln Leu Ala Gly Val Val Ala Val Glu Ile Thr Gly Pro Glu
 100 105 110

Val Pro Phe His Pro Gly Arg Glu Asp Lys Pro Glu Pro Pro Pro Glu
 115 120 125

Gly Arg Leu Pro Asp Ala Thr Lys Gly Ser Asp His Leu Arg Asp Val
 130 135 140

Phe Gly Lys Ala Met Gly Leu Ser Asp Gln Asp Ile Val Ala Leu Ser
 145 150 155 160

Gly Gly His Thr Ile Gly Ala Ala His Lys Glu Arg Ser Gly Phe Glu
 165 170 175

US 12,391,948 B2

163**164**

-continued

Gly	Pro	Trp	Thr	Ser	Asn	Pro	Leu	Ile	Phe	Asp	Asn	Ser	Tyr	Phe	Thr
180															
							185						190		

Glu	Leu	Leu	Thr	Gly	Glu	Lys	Asp	Gly	Leu	Leu	Gln	Leu	Leu	Pro	Ser	Asp
195							200						205			

Lys	Ala	Leu	Leu	Thr	Asp	Ser	Val	Phe	Arg	Pro	Leu	Val	Glu	Lys	Tyr	
210							215						220			

Ala	Ala	Asp	Glu	Asp	Val	Phe	Phe	Ala	Asp	Tyr	Ala	Glu	Ala	His	Leu
225								230				235			240

Lys	Leu	Ser	Glu	Leu	Gly	Phe	Ala	Glu	Ala						
							245						250		

<210> SEQ ID NO 32

<211> LENGTH: 250

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: soybean ascorbate peroxidase

<400> SEQUENCE: 32

Met	Gly	Lys	Ser	Tyr	Pro	Thr	Val	Ser	Ala	Asp	Tyr	Gln	Asp	Ala	Val
1							5				10			15	

Glu	Lys	Ala	Lys	Lys	Lys	Leu	Arg	Gly	Phe	Ile	Ala	Glu	Lys	Arg	Cys
						20				25			30		

Ala	Pro	Leu	Met	Leu	Arg	Leu	Ala	Phe	His	Ser	Ala	Gly	Thr	Phe	Asp
						35			40			45			

Lys	Gly	Thr	Lys	Thr	Gly	Gly	Pro	Phe	Gly	Thr	Ile	Lys	His	Pro	Ala
						50			55		60				

Glu	Leu	Ala	His	Ser	Ala	Asn	Asn	Gly	Leu	Asp	Ile	Ala	Val	Arg	Leu
65						70			75			80			

Leu	Glu	Pro	Leu	Lys	Ala	Glu	Phe	Pro	Ile	Leu	Ser	Tyr	Ala	Asp	Phe
						85			90			95			

Tyr	Gln	Leu	Ala	Gly	Val	Val	Ala	Val	Glu	Val	Thr	Gly	Pro	Lys	
						100			105			110			

Val	Pro	Phe	His	Pro	Gly	Arg	Glu	Asp	Lys	Pro	Glu	Pro	Pro	Glu	
						115			120			125			

Gly	Arg	Leu	Pro	Asp	Pro	Thr	Lys	Gly	Ser	Asp	His	Leu	Arg	Asp	Val
						130			135			140			

Phe	Gly	Lys	Ala	Met	Gly	Leu	Thr	Asp	Gln	Asp	Ile	Val	Ala	Leu	Ser
145							150			155			160		

Gly	Gly	His	Thr	Ile	Gly	Ala	Ala	His	Lys	Glu	Arg	Ser	Gly	Phe	Glu
						165			170			175			

Gly	Pro	Trp	Thr	Ser	Asn	Pro	Leu	Ile	Phe	Asp	Asn	Ser	Tyr	Phe	Thr
							180			185			190		

Glu	Leu	Leu	Ser	Gly	Glu	Lys	Glu	Gly	Leu	Leu	Gln	Leu	Pro	Ser	Asp
						195			200			205			

Lys	Ala	Leu	Leu	Ser	Asp	Pro	Val	Phe	Arg	Pro	Leu	Val	Asp	Lys	Tyr
							210			215			220		

Ala	Ala	Asp	Glu	Asp	Ala	Phe	Phe	Ala	Asp	Tyr	Ala	Glu	Ala	His	Gln
							225			230			235		240

Lys	Leu	Ser	Glu	Leu	Gly	Phe	Ala	Asp	Ala						
							245			250					

<210> SEQ ID NO 33

<211> LENGTH: 299

<212> TYPE: PRT

<213> ORGANISM: Armoracia rusticana

US 12,391,948 B2

165

-continued

<400> SEQUENCE: 33

```

Met Gln Leu Thr Pro Thr Phe Tyr Asp Asn Ser Cys Pro Asn Val Ser
1           5          10          15

Asn Ile Val Arg Asp Thr Ile Val Asn Glu Leu Arg Ser Asp Pro Arg
20          25          30

Ile Ala Ala Ser Ile Leu Arg Leu His Phe His Asp Cys Phe Val Asn
35          40          45

Gly Cys Asp Ala Ser Ile Leu Leu Asp Asn Thr Thr Asn Ala Asn Ser
50          55          60

Ala Arg Gly Phe Pro Val Ile Asp Arg Met Lys Ala Ala Val Glu Ser
65          70          75          80

Ala Cys Pro Arg Thr Val Ser Cys Ala Asp Leu Leu Thr Ile Ala Ala
85          90          95

Gln Gln Ser Val Thr Leu Ala Gly Gly Pro Ser Trp Arg Val Pro Leu
100         105         110

Gly Arg Arg Asp Ser Leu Gln Ala Phe Leu Asp Leu Ala Asn Ala Asn
115         120         125

Leu Pro Ala Pro Phe Phe Thr Leu Pro Gln Leu Lys Asp Ser Phe Arg
130         135         140

Asn Val Gly Leu Asn Arg Ser Ser Asp Leu Val Ala Leu Ser Gly Gly
145         150         155         160

His Thr Phe Gly Lys Asn Gln Cys Arg Phe Ile Met Asp Arg Leu Tyr
165         170         175

Asn Phe Ser Asn Thr Gly Leu Pro Asp Pro Thr Leu Asn Thr Thr Tyr
180         185         190

Leu Gln Thr Leu Arg Gly Leu Cys Pro Leu Asn Gly Asn Leu Ser Ala
195         200         205

Leu Val Asp Phe Asp Leu Arg Thr Pro Thr Ile Phe Asp Asn Lys Tyr
210         215         220

Tyr Val Asn Leu Glu Glu Gln Lys Gly Leu Ile Gln Ser Asp Gln Glu
225         230         235         240

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

Phe Ala Asn Ser Thr Gln Thr Phe Asn Ala Phe Val Glu Ala Met
260         265         270

Asp Arg Met Gly Asn Ile Thr Pro Leu Thr Gly Thr Gln Gly Gln Ile
275         280         285

Arg Leu Asn Cys Arg Val Val Asn Ser Asn Ser
290         295

```

<210> SEQ ID NO 34

<211> LENGTH: 489

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: Alkaline phosphatase

<400> SEQUENCE: 34

```

Met Lys Gln Ser Thr Ile Ala Leu Ala Leu Pro Leu Leu Phe Thr
1           5          10          15

Pro Val Thr Lys Ala Arg Thr Pro Glu Met Pro Leu Gln Gly Thr Ala
20          25          30

Val Asp Gly Gly Gly Ser Met His Ala Ser Leu Glu Val Leu Glu
35          40          45

Asn Arg Ala Ala Gln Gly Asp Ile Thr Ala Pro Gly Gly Ala Arg Arg

```

166

US 12,391,948 B2

167

-continued

50	55	60
Leu Thr Gly Asp Gln Thr Ala Ala	Leu Arg Asp Ser Leu Ser Asp Lys	
65	70	75
Pro Ala Lys Asn Ile Ile Leu Leu Ile	Gly Asp Gly Met Gly Asp Ser	
85	90	95
Glu Ile Thr Ala Ala Arg Asn Tyr Ala Glu Gly Ala Gly Phe Phe		
100	105	110
Lys Gly Ile Asp Ala Leu Pro Leu Thr Gly Gln Tyr Thr His Tyr Ala		
115	120	125
Leu Asn Lys Lys Thr Gly Lys Pro Asp Tyr Val Thr Asp Ser Ala Ala		
130	135	140
Ser Ala Thr Ala Trp Ser Thr Gly Val Lys Thr Tyr Asn Gly Ala Leu		
145	150	155
Gly Val Asp Ile His Glu Lys Asp His Pro Thr Ile Leu Glu Met Ala		
165	170	175
Lys Ala Ala Gly Leu Ala Thr Gly Asn Val Ser Thr Ala Glu Leu Gln		
180	185	190
Asp Ala Thr Pro Ala Ala Leu Val Ala His Val Thr Ser Arg Lys Cys		
195	200	205
Tyr Gly Pro Ser Ala Thr Ser Glu Lys Cys Pro Gly Asn Ala Leu Glu		
210	215	220
Lys Gly Gly Lys Gly Ser Ile Thr Glu Gln Leu Leu Asn Ala Arg Ala		
225	230	235
Asp Val Thr Leu Gly Gly Ala Lys Thr Phe Ala Glu Thr Ala Thr		
245	250	255
Ala Gly Glu Trp Gln Gly Lys Thr Leu Arg Glu Gln Ala Gln Ala Arg		
260	265	270
Gly Tyr Gln Leu Val Ser Asp Ala Ala Ser Leu Asn Ser Val Thr Glu		
275	280	285
Ala Asn Gln Gln Lys Pro Leu Leu Gly Leu Phe Ala Asp Gly Asn Met		
290	295	300
Pro Val Arg Trp Leu Gly Pro Lys Ala Thr Tyr His Gly Asn Ile Asp		
305	310	315
Lys Pro Ala Val Thr Cys Thr Pro Asn Pro Gln Arg Asn Asp Ser Val		
325	330	335
Pro Thr Leu Ala Gln Met Thr Asp Lys Ala Ile Glu Leu Ser Lys		
340	345	350
Asn Glu Lys Gly Phe Phe Leu Gln Val Glu Gly Ala Ser Ile Asp Lys		
355	360	365
Gln Asp His Ala Ala Asn Pro Cys Gly Gln Ile Gly Glu Thr Val Asp		
370	375	380
Leu Asp Glu Ala Val Gln Arg Ala Leu Glu Phe Ala Lys Lys Glu Gly		
385	390	395
Asn Thr Leu Val Ile Val Thr Ala Asp His Ala His Ala Ser Gln Ile		
405	410	415
Val Ala Pro Asp Thr Lys Ala Pro Gly Leu Thr Gln Ala Leu Asn Thr		
420	425	430
Lys Asp Gly Ala Val Met Val Met Ser Tyr Gly Asn Ser Glu Glu Asp		
435	440	445
Ser Gln Glu His Thr Gly Ser Gln Leu Arg Ile Ala Ala Tyr Gly Pro		
450	455	460
His Ala Ala Asn Val Val Gly Leu Thr Asp Gln Thr Asp Leu Phe Tyr		
465	470	475
		480

168

US 12,391,948 B2

169

-continued

Thr Met Lys Ala Ala Leu Gly Leu Lys
485

<210> SEQ ID NO 35
<211> LENGTH: 471
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 35

Met Lys Gln Ser Thr Ile Ala Leu Ala	Leu Leu Pro Leu Leu Phe Thr
1 5	10 15

Pro Val Thr Lys Ala Arg Thr Pro Glu Met Pro Val Leu Glu Asn Arg	
20 25	30

Ala Ala Gln Gly Asp Ile Thr Ala Pro Gly Gly Ala Arg Arg Leu Thr	
35 40	45

Gly Asp Gln Thr Ala Ala Leu Arg Asp Ser Leu Ser Asp Lys Pro Ala	
50 55	60

Lys Asn Ile Ile Leu Leu Ile Gly Asp Gly Met Gly Asp Ser Glu Ile	
65 70	75 80

Thr Ala Ala Arg Asn Tyr Ala Glu Gly Ala Gly Gly Phe Phe Lys Gly	
85 90	95

Ile Asp Ala Leu Pro Leu Thr Gly Gln Tyr Thr His Tyr Ala Leu Asn	
100 105	110

Lys Lys Thr Gly Lys Pro Asp Tyr Val Thr Asp Ser Ala Ala Ser Ala	
115 120	125

Thr Ala Trp Ser Thr Gly Val Lys Thr Tyr Asn Gly Ala Leu Gly Val	
130 135	140

Asp Ile His Glu Lys Asp His Pro Thr Ile Leu Glu Met Ala Lys Ala	
145 150	155 160

Ala Gly Leu Ala Thr Gly Asn Val Ser Thr Ala Glu Leu Gln Asp Ala	
165 170	175

Thr Pro Ala Ala Leu Val Ala His Val Thr Ser Arg Lys Cys Tyr Gly	
180 185	190

Pro Ser Ala Thr Ser Glu Lys Cys Pro Gly Asn Ala Leu Glu Lys Gly	
195 200	205

Gly Lys Gly Ser Ile Thr Glu Gln Leu Leu Asn Ala Arg Ala Asp Val	
210 215	220

Thr Leu Gly Gly Ala Lys Thr Phe Ala Glu Thr Ala Thr Ala Gly	
225 230	235 240

Glu Trp Gln Gly Lys Thr Leu Arg Glu Gln Ala Gln Ala Arg Gly Tyr	
245 250	255

Gln Leu Val Ser Asp Ala Ala Ser Leu Asn Ser Val Thr Glu Ala Asn	
260 265	270

Gln Gln Lys Pro Leu Leu Gly Leu Phe Ala Asp Gly Asn Met Pro Val	
275 280	285

Arg Trp Leu Gly Pro Lys Ala Thr Tyr His Gly Asn Ile Asp Lys Pro	
290 295	300

Ala Val Thr Cys Thr Pro Asn Pro Gln Arg Asn Asp Ser Val Pro Thr	
305 310	315 320

Leu Ala Gln Met Thr Asp Lys Ala Ile Glu Leu Leu Ser Lys Asn Glu	
325 330	335

Lys Gly Phe Phe Leu Gln Val Glu Gly Ala Ser Ile Asp Lys Gln Asp	
340 345	350

His Ala Ala Asn Pro Cys Gly Gln Ile Gly Glu Thr Val Asp Leu Asp

170

US 12,391,948 B2

171

172

-continued

355 360 365

Glu Ala Val Gln Arg Ala Leu Glu Phe Ala Lys Lys Glu Gly Asn Thr
 370 375 380

Leu Val Ile Val Thr Ala Asp His Ala His Ala Ser Gln Val Val Ala
 385 390 395 400

Pro Asp Thr Lys Ala Pro Gly Leu Thr Gln Ala Leu Asn Thr Lys Asp
 405 410 415

Gly Ala Val Met Val Met Ser Tyr Gly Asn Ser Glu Glu Asp Ser Gln
 420 425 430

Glu His Thr Gly Ser Gln Leu Arg Ile Ala Ala Tyr Gly Pro His Ala
 435 440 445

Ala Asn Val Val Gly Leu Thr Asp Gln Thr Asp Leu Phe Tyr Thr Met
 450 455 460

Lys Ala Ala Leu Gly Leu Lys
 465 470

<210> SEQ ID NO 36

<211> LENGTH: 602

<212> TYPE: PRT

<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 36

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

Gly Leu Trp Ala Phe Ser Leu Asp Arg Glu Asn Cys Gly Ile Asp Gln
 20 25 30

Arg Trp Trp Glu Ser Ala Leu Gln Glu Ser Arg Ala Ile Ala Val Pro
 35 40 45

Gly Ser Phe Asn Asp Gln Phe Ala Asp Ala Asp Ile Arg Asn Tyr Ala
 50 55 60

Gly Asn Val Trp Tyr Gln Arg Glu Val Phe Ile Pro Lys Gly Trp Ala
 65 70 75 80

Gly Gln Arg Ile Val Leu Arg Phe Asp Ala Val Thr His Tyr Gly Lys
 85 90 95

Val Trp Val Asn Asn Gln Glu Val Met Glu His Gln Gly Gly Tyr Thr
 100 105 110

Pro Phe Glu Ala Asp Val Thr Pro Tyr Val Ile Ala Gly Lys Ser Val
 115 120 125

Arg Ile Thr Val Cys Val Asn Asn Glu Leu Asn Trp Gln Thr Ile Pro
 130 135 140

Pro Gly Met Val Ile Thr Asp Glu Asn Gly Lys Lys Lys Gln Ser Tyr
 145 150 155 160

Phe His Asp Phe Phe Asn Tyr Ala Gly Ile His Arg Ser Val Met Leu
 165 170 175

Tyr Thr Thr Pro Asn Thr Trp Val Asp Asp Ile Thr Val Val Thr His
 180 185 190

Val Ala Gln Asp Cys Asn His Ala Ser Val Asp Trp Gln Val Val Ala
 195 200 205

Asn Gly Asp Val Ser Val Glu Leu Arg Asp Ala Asp Gln Gln Val Val
 210 215 220

Ala Thr Gly Gln Gly Thr Ser Gly Thr Leu Gln Val Val Asn Pro His
 225 230 235 240

Leu Trp Gln Pro Gly Glu Gly Tyr Leu Tyr Glu Leu Cys Val Thr Ala
 245 250 255

-continued

Lys Ser Gln Thr Glu Cys Asp Ile Tyr Pro Leu Arg Val Gly Ile Arg
260 265 270

Ser Val Ala Val Lys Gly Gln Gln Phe Leu Ile Asn His Lys Pro Phe
275 280 285

Tyr Phe Thr Gly Phe Gly Arg His Glu Asp Ala Asp Leu Arg Gly Lys
290 295 300

Gly Phe Asp Asn Val Leu Met Val His Asp His Ala Leu Met Asp Trp
305 310 315 320

Ile Gly Ala Asn Ser Tyr Arg Thr Ser His Tyr Pro Tyr Ala Glu Glu
325 330 335

Met Leu Asp Trp Ala Asp Glu His Gly Ile Val Val Ile Asp Glu Thr
340 345 350

Ala Ala Val Gly Phe Asn Leu Ser Leu Gly Ile Gly Phe Glu Ala Gly
355 360 365

Asn Lys Pro Lys Glu Leu Tyr Ser Glu Glu Ala Val Asn Gly Glu Thr
370 375 380

Gln Gln Ala His Leu Gln Ala Ile Lys Glu Leu Ile Ala Arg Asp Lys
385 390 395 400

Asn His Pro Ser Val Val Met Trp Ser Ile Ala Asn Glu Pro Asp Thr
405 410 415

Arg Pro Gln Val His Gly Asn Ile Ser Pro Leu Ala Glu Ala Thr Arg
420 425 430

Lys Leu Asp Pro Thr Arg Pro Ile Thr Cys Val Asn Val Met Phe Cys
435 440 445

Asp Ala His Thr Asp Thr Ile Ser Asp Leu Phe Asp Val Leu Cys Leu
450 455 460

Asn Arg Tyr Tyr Gly Trp Tyr Val Gln Ser Gly Asp Leu Glu Thr Ala
465 470 475 480

Glu Lys Val Leu Glu Lys Glu Leu Leu Ala Trp Gln Glu Lys Leu His
485 490 495

Gln Pro Ile Ile Ile Thr Glu Tyr Gly Val Asp Thr Leu Ala Gly Leu
500 505 510

His Ser Met Tyr Thr Asp Met Trp Ser Glu Glu Tyr Gln Cys Ala Trp
515 520 525

Leu Asp Met Tyr His Arg Val Phe Asp Arg Val Ser Ala Val Val Gly
530 535 540

Glu Gln Val Trp Asn Phe Ala Asp Phe Ala Thr Ser Gln Gly Ile Leu
545 550 555 560

Arg Val Gly Gly Asn Lys Lys Gly Ile Phe Thr Arg Asp Arg Lys Pro
565 570 575

Lys Ser Ala Ala Phe Leu Leu Gln Lys Arg Trp Thr Gly Met Asn Phe
580 585 590

Gly Glu Lys Pro Gln Gln Gly Gly Lys Gln
595 600

<210> SEQ ID NO 37
<211> LENGTH: 555
<212> TYPE: PRT
<213> ORGANISM: Francisella tularensis

<400> SEQUENCE: 37

Met Ser Thr Asn Ser Asn Ile Arg Gln Lys Leu Gly Gln Leu Ile Met
1 5 10 15

Met Asp Phe Arg Tyr Trp Gly Glu Asp Ser Asn Asn Gln Arg Ile Pro
20 25 30

US 12,391,948 B2

175**176**

-continued

Phe Thr Lys Thr Asn Asp Ile Val Asn Lys Ile Phe Lys Asp Tyr Asn
 35 40 45

Leu Gly Gly Phe Ile Leu Phe Arg Glu Asn Ile Gln Asn Asn Glu Gln
 50 55 60

Val Ile Ser Leu Leu Arg Asp Leu Gln Ala Asn Thr Asn Thr Pro Ile
 65 70 75 80

Phe Phe Ala Thr Asp Gln Glu Gly Arg Val Asn Arg Leu Gln Gln
 85 90 95

Gly Thr Ser Gly Cys Gly Asn Met Ala Leu Ala Ala Thr Asp Asn Pro
 100 105 110

His Asn Ala Tyr Thr Met Ala Lys Ile Ile Gly Asp Glu Leu Tyr Ser
 115 120 125

Leu Gly Ile Asn Ile Asn Phe Ala Pro Ala Val Asp Val Asn Ser Asn
 130 135 140

Lys Asn Asn Pro Ile Ile Gly Val Arg Ser Tyr Ser Asp Asn Pro Asp
 145 150 155 160

Ile Val Ile Asp Tyr Ala Lys Asn Ala Ile Asn Gly Tyr His Asp Ala
 165 170 175

Lys Ile Ile Asp Cys Ile Lys His Phe Pro Gly His Gly Asp Thr Ala
 180 185 190

Thr Asp Ser His Leu Gly Asn Val Asn Leu Asp Lys Thr Leu Lys Glu
 195 200 205

Leu Gln Thr Thr Glu Leu Leu Pro Phe Ser Lys Leu Ala Arg Asp Cys
 210 215 220

Ser Met Ile Met Thr Ala His Ile Ser Val Pro Ala Leu Asp Asp Thr
 225 230 235 240

Gln Tyr Gln Ser Val Ser Thr Ser Glu Asn Ile Tyr Val Pro Ala Thr
 245 250 255

Leu Ser Tyr Lys Ile Ile Thr Lys Leu Leu Lys Gln Gln Met Lys Phe
 260 265 270

Asp Gly Leu Val Val Ser Asp Ala Met Asp Met His Ala Ile Ala Lys
 275 280 285

His Phe Gly Thr Ile Glu Ala Ser Lys Leu Ala Ile Leu Ala Gly Ile
 290 295 300

Asp Ile Leu Leu Met Pro Val Arg Val Trp Ser Glu Asn Asp Leu Tyr
 305 310 315 320

Lys Leu Glu Glu Leu Phe Cys Glu Leu Glu Lys Gly Tyr Asn Gln Asn
 325 330 335

Ser Asn Phe Ala Asn Ala Val Asp Asn Val Tyr Thr Asn Ile Thr Asp
 340 345 350

Phe Lys Ala Lys His Lys Leu Asp Glu Ser Leu Ile Phe Lys Leu Ser
 355 360 365

Gln Asp Glu Gln Leu Lys Tyr Ala Asn Gln Ile Val Asn Ser Asn Lys
 370 375 380

His Gln Gln Ile Ala Leu Asp Ile Ala Lys Gln Ser Thr Thr Val Val
 385 390 395 400

Lys Asn Ser Gly Ile Ile Pro Cys Asp Leu Asn Lys Leu Lys Asn Ile
 405 410 415

Leu Ile Val Asp Ser Asp Asn Gln Arg Leu Ala Asp Phe His Ser Glu
 420 425 430

Leu Gln Lys Ile Val Leu Asp Asn Asn Ser Asn Val Ile Ile Asn Cys
 435 440 445

US 12,391,948 B2

177

178

-continued

Glu Asn Ile Asn Asn His Asn Ile Lys Thr Ile Ile Glu Asn Ala Asp			
450	455	460	
Leu Ile Leu Leu Ile Ser Ala Asn Leu Arg Glu Tyr Asn Gln Thr Tyr			
465	470	475	480
Ser Tyr Ile Thr Ser Ile Lys Pro Glu Gln Thr Ile Asn Ile Ala Ala			
485	490	495	
Leu Thr Pro Tyr Asp Ile Asn Tyr Ile Asp Asn Ile Ile Asn Tyr Val			
500	505	510	
Cys Ile Tyr Gly Ala Thr Ser Met Asp Gln Thr Asn Tyr Thr Lys Thr			
515	520	525	
Ser Leu Lys Ile Asn Ile Gln Thr Thr Leu Glu Asn Ile Phe Gly Asn			
530	535	540	
Lys Glu Ile Lys Gly Val Leu Pro Val Ser Leu			
545	550	555	

<210> SEO ID NO 38

<211> LENGTH: 321

<212> TYPE: PRT

<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 38

Met Lys Asp Asn Thr Val Pro Leu Lys Leu Ile Ala Leu Leu Ala Asn
1 5 10 15

Gly Glu Phe His Ser Gly Glu Gln Leu Gly Glu Thr Leu Gly Met Ser
20 25 30

Arg Ala Ala Ile Asn Lys His Ile Gln Thr Leu Arg Asp Trp Gly Val
35 40 45

Asp Val Phe Thr Val Pro Gly Lys Gly Tyr Ser Leu Pro Glu Pro Ile
50 55 60

Gln Leu Leu Asn Ala Lys Gln Ile Leu Gly Gln Leu Asp Gly Gly Ser
65 70 75 80

Val Ala Val Leu Pro Val Ile Asp Ser Thr Asn Gln Tyr Leu Leu Asp

Arg Ile Gly Glu Leu Lys Ser Gly Asp Ala Cys Ile Ala Glu Tyr Gln

Gln Ala Gly Arg Gly Arg Arg Gly Arg Lys Trp Phe Ser Pro Phe Gly

Ala Asn Leu Tyr Leu Ser Met Phe Trp Arg Leu Glu Gln Gly Pro Ala

Ala Ala Ile Gly Leu Ser Leu Val Ile Gly Ile Val Met Ala Glu Val

115 130 135 135

180 185 190

Met Ala Met Arg Arg Val Glu Glu Ser Val Val Asn Gln Gly Trp Ile
210 215 220

Thr	Leu	Gln	Glu	Ala	Gly	Ile	Asn	Leu	Asp	Arg	Asn	Thr	Leu	Ala	Ala
225					230					235					240

Met Leu Ile Arg Glu Leu Arg Ala Ala Leu Glu Leu Phe Glu Gln Glu
245 250 255

Gly Leu Ala Pro Tyr Leu Ser Arg Trp Glu Lys Leu Asp Asn Phe Ile
260 265 270

-continued

<400> SEQUENCE: 40

Met	Lys	Asp	Asn	Thr	Val	Pro	Leu	Lys	Leu	Ile	Ala	Leu	Leu	Ala	Asn
1					5				10					15	

Gly Glu Phe His Ser Gly Glu Gln Leu Gly Glu Thr Leu Gly Met Ser
 20 25 30

Arg Ala Ala Ile Asn Lys His Ile Gln Thr Leu Arg Asp Trp Gly Val
35 40 45

Asp Val Phe Thr Val Pro Gly Lys Gly Tyr Ser Leu Pro Glu Pro Ile
50 55 60

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

Val Ala Val Leu Pro Val Val Asp Ser Thr Asn Gln Tyr Leu Leu Asp
85 90 95

Arg Ile Gly Glu Leu Lys Ser Gly Asp Ala Cys Ile Ala Glu Tyr Gln
100 105 110

Gln Ala Gly Arg Gly Ser Arg Gly Arg Lys Trp Phe Ser Pro Phe Gly
 115 120 125

Ala Asn Leu Tyr Leu Ser Met Phe Trp Arg Leu Lys Arg Gly Pro Ala
130 135 140

Ala Ile Gly Leu Gly Pro Val Ile Gly Ile Val Met Ala Glu Ala Leu
145 150 155 160

Arg Lys Leu Gly Ala Asp Lys Val Arg Val Lys Trp Pro Asn Asp Leu
165 170 175

Tyr Leu Gln Asp Arg Lys Leu Ala Gly Ile Leu Val Glu Leu Ala Gly
180 185 190

Ile Thr Gly Asp Ala Ala Gln Ile Val Ile Gly Ala Gly Ile Asn Val
195 200 205

Ala Met Arg Arg Val Glu Glu Ser Val Val Asn Gln Gly Trp Ile Thr
210 215 220

Leu Gln Glu Ala Gly Ile Asn Leu Asp Arg Asn Thr Leu Ala Ala Thr
325 326 327 328 329 330 331 332 333 334 335 336 337 338

Leu Ile Arg Glu Leu Arg Ala Ala Leu Glu Leu Phe Glu Gln Glu Gly
215 250 285

Leu Ala Pro Tyr Leu Pro Arg Trp Glu Lys Leu Asp Asn Phe Ile Asn

Arg Pro Val Lys Leu Ile Ile Gly Asp Lys Glu Ile Phe Gly Ile Ser

Arg Gly Ile Asp Lys Gln Gly Ala Leu Leu Leu Glu Gln Asp Gly Val

Ile Lys Pro Trp Met Gly Gly Glu Ile Ser Leu Arg Ser Ala Glu Lys

585 590 595 600

<210> SEQ ID NO: 41
<211> LENGTH: 334
<212> TYPE: PRT

<213> ORGANISM: Artificial
<220> FEATURE:

<400> SEQUENCE: 41

Met	Asp	Tyr	Lys	Asp	Asp	Asp	Asp	Lys	Ser	Pro	Arg	Ser	Met	Lys	Asp
1				5				10					15		

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

US 12,391,948 B2

183**184**

-continued

His	Ser	Gly	Glu	Gln	Leu	Gly	Glu	Thr	Leu	Gly	Met	Ser	Arg	Ala	Ala
35					40						45				
Ile	Asn	Lys	His	Ile	Gln	Thr	Leu	Arg	Asp	Trp	Gly	Val	Asp	Val	Phe
50					55					60					
Thr	Val	Pro	Gly	Lys	Gly	Tyr	Ser	Leu	Pro	Glu	Pro	Ile	Gln	Leu	Leu
65					70					75					80
Asn	Ala	Lys	Gln	Ile	Leu	Gly	Gln	Leu	Asp	Gly	Gly	Ser	Val	Ala	Val
					85				90					95	
Leu	Pro	Val	Ile	Asp	Ser	Thr	Asn	Gln	Tyr	Leu	Leu	Asp	Arg	Ile	Gly
									105					110	
Glu	Leu	Lys	Ser	Gly	Asp	Ala	Cys	Ile	Ala	Glu	Tyr	Gln	Gln	Ala	Gly
					115				120					125	
Arg	Gly	Arg	Arg	Gly	Arg	Lys	Trp	Phe	Ser	Pro	Phe	Gly	Ala	Asn	Leu
									135					140	
Tyr	Leu	Ser	Met	Phe	Trp	Arg	Leu	Glu	Gln	Gly	Pro	Ala	Ala	Ala	Ile
145					150				155					160	
Gly	Leu	Ser	Leu	Val	Ile	Gly	Ile	Val	Met	Ala	Glu	Val	Leu	Arg	Lys
					165				170					175	
Leu	Gly	Ala	Asp	Lys	Val	Arg	Val	Lys	Trp	Pro	Asn	Asp	Leu	Tyr	Leu
					180				185					190	
Gln	Asp	Arg	Lys	Leu	Ala	Gly	Ile	Leu	Val	Glu	Leu	Thr	Gly	Lys	Thr
					195				200					205	
Gly	Asp	Ala	Ala	Gln	Ile	Val	Ile	Gly	Ala	Gly	Ile	Asn	Met	Ala	Met
					210				215					220	
Arg	Arg	Val	Glu	Glu	Ser	Val	Val	Asn	Gln	Gly	Trp	Ile	Thr	Leu	Gln
225					230				235					240	
Glu	Ala	Gly	Ile	Asn	Leu	Asp	Arg	Asn	Thr	Leu	Ala	Ala	Met	Leu	Ile
					245				250					255	
Arg	Glu	Leu	Arg	Ala	Ala	Leu	Glu	Leu	Phe	Glu	Gln	Glu	Gly	Leu	Ala
					260				265					270	
Pro	Tyr	Leu	Ser	Arg	Trp	Glu	Lys	Leu	Asp	Asn	Phe	Ile	Asn	Arg	Pro
					275				280					285	
Val	Lys	Leu	Ile	Ile	Gly	Asp	Lys	Glu	Ile	Phe	Gly	Ile	Ser	Arg	Gly
					290				295					300	
Ile	Asp	Lys	Gln	Gly	Ala	Leu	Leu	Glu	Gln	Asp	Gly	Ile	Ile	Lys	
					305				310					320	
Pro	Trp	Met	Gly	Gly	Glu	Ile	Ser	Leu	Arg	Ser	Ala	Glu	Lys		
					325				330						

<210> SEQ ID NO 42

<211> LENGTH: 746

<212> TYPE: PRT

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 42

Met	Gly	Gln	Thr	Gly	Lys	Ser	Glu	Lys	Gly	Pro	Val	Cys	Trp	Arg
1					5				10					15

Lys	Arg	Val	Lys	Ser	Glu	Tyr	Met	Arg	Leu	Arg	Gln	Leu	Lys	Arg	Phe
					20				25					30	

Arg	Arg	Ala	Asp	Glu	Val	Lys	Ser	Met	Phe	Ser	Ser	Asn	Arg	Gln	Lys
					35				40					45	

Ile	Leu	Glu	Arg	Thr	Glu	Ile	Leu	Asn	Gln	Glu	Trp	Lys	Gln	Arg	Arg
					50				55					60	

Ile	Gln	Pro	Val	His	Ile	Leu	Thr	Ser	Val	Ser	Ser	Leu	Arg	Gly	Thr
					65				70					80	

US 12,391,948 B2

185**186**

-continued

Arg Glu Cys Ser Val Thr Ser Asp Leu Asp Phe Pro Thr Gln Val Ile
 85 90 95
 Pro Leu Lys Thr Leu Asn Ala Val Ala Ser Val Pro Ile Met Tyr Ser
 100 105 110
 Trp Ser Pro Leu Gln Gln Asn Phe Met Val Glu Asp Glu Thr Val Leu
 115 120 125
 His Asn Ile Pro Tyr Met Gly Asp Glu Val Leu Asp Gln Asp Gly Thr
 130 135 140
 Phe Ile Glu Glu Leu Ile Lys Asn Tyr Asp Gly Lys Val His Gly Asp
 145 150 155 160
 Arg Glu Cys Gly Phe Ile Asn Asp Glu Ile Phe Val Glu Leu Val Asn
 165 170 175
 Ala Leu Gly Gln Tyr Asn Asp Asp Asp Asp Asp Asp Gly Asp Asp
 180 185 190
 Pro Glu Glu Arg Glu Glu Lys Gln Lys Asp Leu Glu Asp His Arg Asp
 195 200 205
 Asp Lys Glu Ser Arg Pro Pro Arg Lys Phe Pro Ser Asp Lys Ile Phe
 210 215 220
 Glu Ala Ile Ser Ser Met Phe Pro Asp Lys Gly Thr Ala Glu Glu Leu
 225 230 235 240
 Lys Glu Lys Tyr Lys Glu Leu Thr Glu Gln Gln Leu Pro Gly Ala Leu
 245 250 255
 Pro Pro Glu Cys Thr Pro Asn Ile Asp Gly Pro Asn Ala Lys Ser Val
 260 265 270
 Gln Arg Glu Gln Ser Leu His Ser His Thr Leu Phe Cys Arg Arg
 275 280 285
 Cys Phe Lys Tyr Asp Cys Phe Leu His Pro Phe His Ala Thr Pro Asn
 290 295 300
 Thr Tyr Lys Arg Lys Asn Thr Glu Thr Ala Leu Asp Asn Lys Pro Cys
 305 310 315 320
 Gly Pro Gln Cys Tyr Gln His Leu Glu Gly Ala Lys Glu Phe Ala Ala
 325 330 335
 Ala Leu Thr Ala Glu Arg Ile Lys Thr Pro Pro Lys Arg Pro Gly Gly
 340 345 350
 Arg Arg Arg Gly Leu Pro Asn Asn Ser Ser Arg Pro Ser Thr Pro
 355 360 365
 Thr Ile Asn Val Leu Glu Ser Lys Asp Thr Asp Ser Asp Arg Glu Ala
 370 375 380
 Gly Thr Glu Thr Gly Gly Glu Asn Asn Asp Lys Glu Glu Glu Lys
 385 390 395 400
 Lys Asp Glu Thr Ser Ser Ser Glu Ala Asn Ser Arg Cys Gln Thr
 405 410 415
 Pro Ile Lys Met Lys Pro Asn Ile Glu Pro Pro Glu Asn Val Glu Trp
 420 425 430
 Ser Gly Ala Glu Ala Ser Met Phe Arg Val Leu Ile Gly Thr Tyr Tyr
 435 440 445
 Asp Asn Phe Cys Ala Ile Ala Arg Leu Ile Gly Thr Lys Thr Cys Arg
 450 455 460
 Gln Val Tyr Glu Phe Arg Val Lys Glu Ser Ser Ile Ile Ala Pro Ala
 465 470 475 480
 Pro Ala Glu Asp Val Asp Thr Pro Pro Arg Lys Lys Lys Arg Lys His
 485 490 495

US 12,391,948 B2

187

-continued

Arg Leu Trp Ala Ala His Cys Arg Lys Ile Gln Leu Lys Lys Asp Gly
500 505 510

Ser Ser Asn His Val Tyr Asn Tyr Gln Pro Cys Asp His Pro Arg Gln
515 520 525

Pro Cys Asp Ser Ser Cys Pro Cys Val Ile Ala Gln Asn Phe Cys Glu
530 535 540

Lys Phe Cys Gln Cys Ser Ser Glu Cys Gln Asn Arg Phe Pro Gly Cys
545 550 555 560

Arg Cys Lys Ala Gln Cys Asn Thr Lys Gln Cys Pro Cys Tyr Leu Ala
565 570 575

Val Arg Glu Cys Asp Pro Asp Leu Cys Leu Thr Cys Gly Ala Ala Asp
580 585 590

His Trp Asp Ser Lys Asn Val Ser Cys Lys Asn Cys Ser Ile Gln Arg
595 600 605

Gly Ser Lys Lys His Leu Leu Leu Ala Pro Ser Asp Val Ala Gly Trp
610 615 620

Gly Ile Phe Ile Lys Asp Pro Val Gln Lys Asn Glu Phe Ile Ser Glu
625 630 635 640

Tyr Cys Gly Glu Ile Ile Ser Gln Asp Glu Ala Asp Arg Arg Gly Lys
645 650 655

Val Tyr Asp Lys Tyr Met Cys Ser Phe Leu Phe Asn Leu Asn Asn Asp
660 665 670

Phe Val Val Asp Ala Thr Arg Lys Gly Asn Lys Ile Arg Phe Ala Asn
675 680 685

His Ser Val Asn Pro Asn Cys Tyr Ala Lys Val Met Met Val Asn Gly
690 695 700

Asp His Arg Ile Gly Ile Phe Ala Lys Arg Ala Ile Gln Thr Gly Glu
705 710 715 720

Glu Leu Phe Phe Asp Tyr Arg Tyr Ser Gln Ala Asp Ala Leu Lys Tyr
725 730 735

Val Gly Ile Glu Arg Glu Met Glu Ile Pro
740 745

<210> SEQ ID NO 43
<211> LENGTH: 439
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 43

Met Pro Leu Asn Val Ser Phe Thr Asn Arg Asn Tyr Asp Leu Asp Tyr
1 5 10 15

Asp Ser Val Gln Pro Tyr Phe Tyr Cys Asp Glu Glu Glu Asn Phe Tyr
20 25 30

Gln Gln Gln Gln Ser Glu Leu Gln Pro Pro Ala Pro Ser Glu Asp
35 40 45

Ile Trp Lys Phe Glu Leu Leu Pro Thr Pro Pro Leu Ser Pro Ser
50 55 60

Arg Arg Ser Gly Leu Cys Ser Pro Ser Tyr Val Ala Val Thr Pro Phe
65 70 75 80

Ser Leu Arg Gly Asp Asn Asp Gly Gly Gly Ser Phe Ser Thr Ala
85 90 95

Asp Gln Leu Glu Met Val Thr Glu Leu Leu Gly Gly Asp Met Val Asn
100 105 110

Gln Ser Phe Ile Cys Asp Pro Asp Asp Glu Thr Phe Ile Lys Asn Ile
115 120 125

188

US 12,391,948 B2

189**190**

-continued

Ile Ile Gln Asp Cys Met Trp Ser Gly Phe Ser Ala Ala Ala Lys Leu
 130 135 140

Val Ser Glu Lys Leu Ala Ser Tyr Gln Ala Ala Arg Lys Asp Ser Gly
 145 150 155 160

Ser Pro Asn Pro Ala Arg Gly His Ser Val Cys Ser Thr Ser Ser Leu
 165 170 175

Tyr Leu Gln Asp Leu Ser Ala Ala Ser Glu Cys Ile Asp Pro Ser
 180 185 190

Val Val Phe Pro Tyr Pro Leu Asn Asp Ser Ser Ser Pro Lys Ser Cys
 195 200 205

Ala Ser Gln Asp Ser Ser Ala Phe Ser Pro Ser Ser Asp Ser Leu Leu
 210 215 220

Ser Ser Thr Glu Ser Ser Pro Gln Gly Ser Pro Glu Pro Leu Val Leu
 225 230 235 240

His Glu Glu Thr Pro Pro Thr Thr Ser Ser Asp Ser Glu Glu Glu Gln
 245 250 255

Glu Asp Glu Glu Ile Asp Val Val Ser Val Glu Lys Arg Gln Ala
 260 265 270

Pro Gly Lys Arg Ser Glu Ser Gly Ser Pro Ser Ala Gly Gly His Ser
 275 280 285

Lys Pro Pro His Ser Pro Leu Val Leu Lys Arg Cys His Val Ser Thr
 290 295 300

His Gln His Asn Tyr Ala Ala Pro Pro Ser Thr Arg Lys Asp Tyr Pro
 305 310 315 320

Ala Ala Lys Arg Val Lys Leu Asp Ser Val Arg Val Leu Arg Gln Ile
 325 330 335

Ser Asn Asn Arg Lys Cys Thr Ser Pro Arg Ser Ser Asp Thr Glu Glu
 340 345 350

Asn Val Lys Arg Arg Thr His Asn Val Leu Glu Arg Gln Arg Arg Asn
 355 360 365

Glu Leu Lys Arg Ser Phe Phe Ala Leu Arg Asp Gln Ile Pro Glu Leu
 370 375 380

Glu Asn Asn Glu Lys Ala Pro Lys Val Val Ile Leu Lys Lys Ala Thr
 385 390 395 400

Ala Tyr Ile Leu Ser Val Gln Ala Glu Glu Gln Lys Leu Ile Ser Glu
 405 410 415

Glu Asp Leu Leu Arg Lys Arg Arg Glu Gln Leu Lys His Lys Leu Glu
 420 425 430

Gln Leu Arg Asn Ser Cys Ala
 435

<210> SEQ ID NO 44
 <211> LENGTH: 380
 <212> TYPE: PRT
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 44

Met Met Phe Ser Gly Phe Asn Ala Asp Tyr Glu Ala Ser Ser Ser Arg
 1 5 10 15

Cys Ser Ser Ala Ser Pro Ala Gly Asp Ser Leu Ser Tyr Tyr His Ser
 20 25 30

Pro Ala Asp Ser Phe Ser Ser Met Gly Ser Pro Val Asn Ala Gln Asp
 35 40 45

Phe Cys Thr Asp Leu Ala Val Ser Ser Ala Asn Phe Ile Pro Thr Val

US 12,391,948 B2

191**192**

-continued

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

<210> SEQ ID NO 45

<211> LENGTH: 331

<212> TYPE: PRT

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 45

Met Thr Ala Lys Met Glu Thr Thr Phe Tyr Asp Asp Ala		
1	5	10
15		

Ser Phe Leu Pro Ser Glu Ser Gly Pro Tyr Gly Tyr Ser Asn Pro Lys		
20	25	30

Ile Leu Lys Gln Ser Met Thr Leu Asn Leu Ala Asp Pro Val Gly Ser		
35	40	45

US 12,391,948 B2

193**194**

-continued

Leu Lys Pro His Leu Arg Ala Lys Asn Ser Asp Leu Leu Thr Ser Pro
 50 55 60

Asp Val Gly Leu Leu Lys Leu Ala Ser Pro Glu Leu Glu Arg Leu Ile
 65 70 75 80

Ile Gln Ser Ser Asn Gly His Ile Thr Thr Pro Thr Pro Thr Gln
 85 90 95

Phe Leu Cys Pro Lys Asn Val Thr Asp Glu Gln Glu Gly Phe Ala Glu
 100 105 110

Gly Phe Val Arg Ala Leu Ala Glu Leu His Ser Gln Asn Thr Leu Pro
 115 120 125

Ser Val Thr Ser Ala Ala Gln Pro Val Asn Gly Ala Gly Met Val Ala
 130 135 140

Pro Ala Val Ala Ser Val Ala Gly Gly Ser Gly Ser Gly Phe Ser
 145 150 155 160

Ala Ser Leu His Ser Glu Pro Pro Val Tyr Ala Asn Leu Ser Asn Phe
 165 170 175

Asn Pro Gly Ala Leu Ser Ser Gly Gly Ala Pro Ser Tyr Gly Ala
 180 185 190

Ala Gly Leu Ala Phe Pro Ala Gln Pro Gln Gln Gln Gln Pro Pro
 195 200 205

His His Leu Pro Gln Gln Met Pro Val Gln His Pro Arg Leu Gln Ala
 210 215 220

Leu Lys Glu Glu Pro Gln Thr Val Pro Glu Met Pro Gly Glu Thr Pro
 225 230 235 240

Pro Leu Ser Pro Ile Asp Met Glu Ser Gln Glu Arg Ile Lys Ala Glu
 245 250 255

Arg Lys Arg Met Arg Asn Arg Ile Ala Ala Ser Lys Cys Arg Lys Arg
 260 265 270

Lys Leu Glu Arg Ile Ala Arg Leu Glu Glu Lys Val Lys Thr Leu Lys
 275 280 285

Ala Gln Asn Ser Glu Leu Ala Ser Thr Ala Asn Met Leu Arg Glu Gln
 290 295 300

Val Ala Gln Leu Lys Gln Lys Val Met Asn His Val Asn Ser Gly Cys
 305 310 315 320

Gln Leu Met Leu Thr Gln Gln Leu Gln Thr Phe
 325 330

<210> SEQ ID NO 46

<211> LENGTH: 327

<212> TYPE: PRT

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 46

Met Thr Met Glu Ser Gly Ala Glu Asn Gln Gln Ser Gly Asp Ala Ala
 1 5 10 15

Val Thr Glu Ala Glu Asn Gln Gln Met Thr Val Gln Ala Gln Pro Gln
 20 25 30

Ile Ala Thr Leu Ala Gln Val Ser Met Pro Ala Ala His Ala Thr Ser
 35 40 45

Ser Ala Pro Thr Val Thr Leu Val Gln Leu Pro Asn Gly Gln Thr Val
 50 55 60

Gln Val His Gly Val Ile Gln Ala Ala Gln Pro Ser Val Ile Gln Ser
 65 70 75 80

Pro Gln Val Gln Thr Val Gln Ile Ser Thr Ile Ala Glu Ser Glu Asp
 85 90 95

-continued

Ser Gln Glu Ser Val Asp Ser Val Thr Asp Ser Gln Lys Arg Arg Glu
 100 105 110

Ile Leu Ser Arg Arg Pro Ser Tyr Arg Lys Ile Leu Asn Asp Leu Ser
 115 120 125

Ser Asp Ala Pro Gly Val Pro Arg Ile Glu Glu Glu Lys Ser Glu Glu
 130 135 140

Glu Thr Ser Ala Pro Ala Ile Thr Thr Val Thr Val Pro Thr Pro Ile
 145 150 155 160

Tyr Gln Thr Ser Ser Gly Gln Tyr Ile Ala Ile Thr Gln Gly Ala
 165 170 175

Ile Gln Leu Ala Asn Asn Gly Thr Asp Gly Val Gln Gly Leu Gln Thr
 180 185 190

Leu Thr Met Thr Asn Ala Ala Ala Thr Gln Pro Gly Thr Thr Ile Leu
 195 200 205

Gln Tyr Ala Gln Thr Thr Asp Gly Gln Gln Ile Leu Val Pro Ser Asn
 210 215 220

Gln Val Val Val Gln Ala Ala Ser Gly Asp Val Gln Thr Tyr Gln Ile
 225 230 235 240

Arg Thr Ala Pro Thr Ser Thr Ile Ala Pro Gly Val Val Met Ala Ser
 245 250 255

Ser Pro Ala Leu Pro Thr Gln Pro Ala Glu Glu Ala Ala Arg Lys Arg
 260 265 270

Glu Val Arg Leu Met Lys Asn Arg Glu Ala Ala Arg Glu Cys Arg Arg
 275 280 285

Lys Lys Lys Glu Tyr Val Lys Cys Leu Glu Asn Arg Val Ala Val Leu
 290 295 300

Glu Asn Gln Asn Lys Thr Leu Ile Glu Glu Leu Lys Ala Leu Lys Asp
 305 310 315 320

Leu Tyr Cys His Lys Ser Asp
 325

<210> SEQ ID NO 47
 <211> LENGTH: 474
 <212> TYPE: PRT
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 47

Met Glu Val Ala Pro Glu Gln Pro Gly Trp Met Ala His Pro Ala Val
 1 5 10 15

Leu Asn Ala Gln His Pro Asp Ser His His Pro Gly Leu Ala His Asn
 20 25 30

Tyr Met Glu Pro Ala His Val Leu Pro Pro Asp Glu Val Asp Val Phe
 35 40 45

Phe Asn His Leu Asp Ser Gln Gly Asn Pro Tyr Tyr Ala Asn Pro Ala
 50 55 60

Gln Arg Gly Val Ser Tyr Ser Pro Ala His Ala Arg Leu Thr Gly Gly
 65 70 75 80

Gln Met Cys Arg Pro His Leu Leu His Ser Pro Gly Leu Pro Trp Leu
 85 90 95

Asp Gly Gly Lys Ala Ala Leu Ser Ala Ala His His Lys Thr Trp Thr
 100 105 110

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

Pro Gly Gly His Ser Leu Cys Thr Gln Gly Leu Gly Val Gly Gly

US 12,391,948 B2

197**198**

-continued

130	135	140
Ser Ser Gly Ser Ser Val Ala Ser Leu Thr Pro Thr Ala Ala His Ser		
145	150	155
Gly Ser His Leu Phe Gly Phe Pro Pro Arg His Pro Lys Glu Leu Ser		
165	170	175
Pro Asp Pro Ser Thr Thr Gly Ala Ala Ser Pro Ala Ser Ser Ala		
180	185	190
Gly Gly Ser Ser Ala Arg Gly Glu Asp Lys Asp Gly Val Lys Tyr Gln		
195	200	205
Ala Ser Leu Thr Glu Ser Met Lys Met Glu Ser Gly Arg Pro Leu Arg		
210	215	220
Pro Gly Leu Ala Thr Met Gly Thr Gln Pro Ala Thr His His Pro Ile		
225	230	235
Pro Thr Tyr Pro Ser Tyr Val Pro Ala Ala Ala His Asp Tyr Ser Ser		
245	250	255
Gly Leu Phe His Pro Gly Ser Phe Leu Gly Gly Pro Ala Ser Ser Phe		
260	265	270
Thr Pro Lys Gln Arg Ser Lys Thr Arg Ser Cys Ser Glu Gly Arg Glu		
275	280	285
Cys Val Asn Cys Gly Ala Thr Ala Thr Pro Leu Trp Arg Arg Asp Gly		
290	295	300
Thr Gly His Tyr Leu Cys Asn Ala Cys Gly Phe Tyr His Lys Met Lys		
305	310	315
Gly Gln Asn Arg Pro Leu Ile Lys Pro Lys Arg Arg Leu Ser Ala Ala		
325	330	335
Arg Arg Ala Gly Thr Cys Cys Ala Asn Cys Gln Thr Thr Thr Thr		
340	345	350
Leu Trp Arg Arg Asn Ala Asn Gly Asp Pro Val Cys Asn Ala Cys Gly		
355	360	365
Leu Tyr Tyr Lys Leu His Asn Val Asn Arg Pro Leu Thr Met Lys Lys		
370	375	380
Glu Gly Ile Gln Thr Arg Asn Arg Lys Met Ser Asn Lys Ser Lys Lys		
385	390	395
Ser Lys Lys Gly Ala Glu Cys Phe Glu Glu Leu Ser Lys Cys Met Gln		
405	410	415
Glu Lys Ser Ser Pro Phe Ser Ala Ala Ala Leu Ala Gly His Met Ala		
420	425	430
Pro Met Gly His Leu Pro Pro Phe Ser His Ser Gly His Ile Leu Pro		
435	440	445
Thr Pro Thr Pro Ile His Pro Ser Ser Ser Leu Ser Phe Gly His Pro		
450	455	460
His Pro Ser Ser Met Val Thr Ala Met Gly		
465	470	

<210> SEQ ID NO 48

<211> LENGTH: 881

<212> TYPE: PRT

<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 48

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

Lys Lys Leu Lys Cys Ser Lys Glu Lys Pro Lys Cys Ala Lys Cys Leu		
20	25	30

US 12,391,948 B2

199**200**

-continued

Lys Asn Asn Trp Glu Cys Arg Tyr Ser Pro Lys Thr Lys Arg Ser Pro
35 40 45

Leu Thr Arg Ala His Leu Thr Glu Val Glu Ser Arg Leu Glu Arg Leu
50 55 60

Glu Gln Leu Phe Leu Leu Ile Phe Pro Arg Glu Asp Leu Asp Met Ile
65 70 75 80

Leu Lys Met Asp Ser Leu Gln Asp Ile Lys Ala Leu Leu Thr Gly Leu
85 90 95

Phe Val Gln Asp Asn Val Asn Lys Asp Ala Val Thr Asp Arg Leu Ala
100 105 110

Ser Val Glu Thr Asp Met Pro Leu Thr Leu Arg Gln His Arg Ile Ser
115 120 125

Ala Thr Ser Ser Ser Glu Glu Ser Ser Asn Lys Gly Gln Arg Gln Leu
130 135 140

Thr Val Ser Ile Asp Ser Ala Ala His His Asp Asn Ser Thr Ile Pro
145 150 155 160

Leu Asp Phe Met Pro Arg Asp Ala Leu His Gly Phe Asp Trp Ser Glu
165 170 175

Glu Asp Asp Met Ser Asp Gly Leu Pro Phe Leu Lys Thr Asp Pro Asn
180 185 190

Asn Asn Gly Phe Phe Gly Asp Gly Ser Leu Leu Cys Ile Leu Arg Ser
195 200 205

Ile Gly Phe Lys Pro Glu Asn Tyr Thr Asn Ser Asn Val Asn Arg Leu
210 215 220

Pro Thr Met Ile Thr Asp Arg Tyr Thr Leu Ala Ser Arg Ser Thr Thr
225 230 235 240

Ser Arg Leu Leu Gln Ser Tyr Leu Asn Asn Phe His Pro Tyr Cys Pro
245 250 255

Ile Val His Ser Pro Thr Leu Met Met Leu Tyr Asn Asn Gln Ile Glu
260 265 270

Ile Ala Ser Lys Asp Gln Trp Gln Ile Leu Phe Asn Cys Ile Leu Ala
275 280 285

Ile Gly Ala Trp Cys Ile Glu Gly Glu Ser Thr Asp Ile Asp Val Phe
290 295 300

Tyr Tyr Gln Asn Ala Lys Ser His Leu Thr Ser Lys Val Phe Glu Ser
305 310 315 320

Gly Ser Ile Ile Leu Val Thr Ala Leu His Leu Leu Ser Arg Tyr Thr
325 330 335

Gln Trp Arg Gln Lys Thr Asn Thr Ser Tyr Asn Phe His Ser Phe Ser
340 345 350

Ile Arg Met Ala Ile Ser Leu Gly Leu Asn Arg Asp Leu Pro Ser Ser
355 360 365

Phe Ser Asp Ser Ser Ile Leu Glu Gln Arg Arg Arg Ile Trp Trp Ser
370 375 380

Val Tyr Ser Trp Glu Ile Gln Leu Ser Leu Leu Tyr Gly Arg Ser Ile
385 390 395 400

Gln Leu Ser Gln Asn Thr Ile Ser Phe Pro Ser Ser Val Asp Asp Val
405 410 415

Gln Arg Thr Thr Thr Gly Pro Thr Ile Tyr His Gly Ile Ile Glu Thr
420 425 430

Ala Arg Leu Leu Gln Val Phe Thr Lys Ile Tyr Glu Leu Asp Lys Thr
435 440 445

Val Thr Ala Glu Lys Ser Pro Ile Cys Ala Lys Lys Cys Leu Met Ile
450

US 12,391,948 B2

201

-continued

202

450	455	460
Cys Asn Glu Ile Glu Glu Val Ser Arg Gln Ala Pro Lys Phe Leu Gln		
465	470	475
Met Asp Ile Ser Thr Thr Ala Leu Thr Asn Leu Leu Lys Glu His Pro		
485	490	495
Trp Leu Ser Phe Thr Arg Phe Glu Leu Lys Trp Lys Gln Leu Ser Leu		
500	505	510
Ile Ile Tyr Val Leu Arg Asp Phe Thr Asn Phe Thr Gln Lys Lys		
515	520	525
Ser Gln Leu Glu Gln Asp Gln Asn Asp His Gln Ser Tyr Glu Val Lys		
530	535	540
Arg Cys Ser Ile Met Leu Ser Asp Ala Ala Gln Arg Thr Val Met Ser		
545	550	555
Val Ser Ser Tyr Met Asp Asn His Asn Val Thr Pro Tyr Phe Ala Trp		
565	570	575
Asn Cys Ser Tyr Tyr Leu Phe Asn Ala Val Leu Val Pro Ile Lys Thr		
580	585	590
Leu Leu Ser Asn Ser Lys Ser Asn Ala Glu Asn Asn Glu Thr Ala Gln		
595	600	605
Leu Leu Gln Gln Ile Asn Thr Val Leu Met Leu Leu Lys Lys Leu Ala		
610	615	620
Thr Phe Lys Ile Gln Thr Cys Glu Lys Tyr Ile Gln Val Leu Glu Glu		
625	630	635
640		
Val Cys Ala Pro Phe Leu Leu Ser Gln Cys Ala Ile Pro Leu Pro His		
645	650	655
Ile Ser Tyr Asn Asn Ser Asn Gly Ser Ala Ile Lys Asn Ile Val Gly		
660	665	670
Ser Ala Thr Ile Ala Gln Tyr Pro Thr Leu Pro Glu Glu Asn Val Asn		
675	680	685
Asn Ile Ser Val Lys Tyr Val Ser Pro Gly Ser Val Gly Pro Ser Pro		
690	695	700
Val Pro Leu Lys Ser Gly Ala Ser Phe Ser Asp Leu Val Lys Leu Leu		
705	710	715
720		
Ser Asn Arg Pro Pro Ser Arg Asn Ser Pro Val Thr Ile Pro Arg Ser		
725	730	735
Thr Pro Ser His Arg Ser Val Thr Pro Phe Leu Gly Gln Gln Gln		
740	745	750
Leu Gln Ser Leu Val Pro Leu Thr Pro Ser Ala Leu Phe Gly Gly Ala		
755	760	765
Asn Phe Asn Gln Ser Gly Asn Ile Ala Asp Ser Ser Leu Ser Phe Thr		
770	775	780
Phe Thr Asn Ser Ser Asn Gly Pro Asn Leu Ile Thr Thr Gln Thr Asn		
785	790	795
800		
Ser Gln Ala Leu Ser Gln Pro Ile Ala Ser Ser Asn Val His Asp Asn		
805	810	815
Phe Met Asn Asn Glu Ile Thr Ala Ser Lys Ile Asp Asp Gly Asn Asn		
820	825	830
Ser Lys Pro Leu Ser Pro Gly Trp Thr Asp Gln Thr Ala Tyr Asn Ala		
835	840	845
Phe Gly Ile Thr Thr Gly Met Phe Asn Thr Thr Met Asp Asp Val		
850	855	860
Tyr Asn Tyr Leu Phe Asp Asp Glu Asp Thr Pro Pro Asn Pro Lys Lys		
865	870	875
880		

-continued

Glu

<210> SEQ ID NO 49

<211> LENGTH: 278

<212> TYPE: PRT

<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 49

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

Lys	Lys	Leu	Lys	Cys	Ser	Lys	Glu	Lys	Pro	Lys	Cys	Ala	Lys	Cys	Leu
						20		25			30				

Lys	Asn	Asn	Trp	Glu	Cys	Arg	Tyr	Ser	Pro	Lys	Thr	Lys	Arg	Ser	Pro
						35		40			45				

Leu	Thr	Arg	Ala	His	Leu	Thr	Glu	Val	Glu	Ser	Arg	Leu	Glu	Arg	Leu
						50		55			60				

Glu	Gln	Leu	Phe	Leu	Leu	Ile	Phe	Pro	Arg	Glu	Asp	Leu	Asp	Met	Ile
65						70		75			80				

Leu	Lys	Met	Asp	Ser	Leu	Gln	Asp	Ile	Lys	Ala	Leu	Leu	Thr	Gly	Leu
						85		90			95				

Phe	Val	Gln	Asp	Asn	Val	Asn	Lys	Asp	Ala	Val	Thr	Asp	Arg	Leu	Ala
						100		105			110				

Ser	Val	Glu	Thr	Asp	Met	Pro	Leu	Thr	Leu	Arg	Gln	His	Arg	Ile	Ser
						115		120			125				

Ala	Thr	Ser	Ser	Glu	Ser	Ser	Asn	Lys	Gly	Gln	Arg	Gln	Leu		
						130		135			140				

Thr	Val	Ser	Ile	Glu	Phe	Ser	Arg	Gly	Arg	Thr	Arg	Asn	Asn	Tyr	Gly
145						150		155			160				

Ser	Thr	Ile	Glu	Gly	Leu	Leu	Asp	Leu	Pro	Asp	Asp	Asp	Asp	Ala	Pro
						165		170			175				

Ala	Glu	Ala	Gly	Leu	Val	Ala	Pro	Arg	Met	Ser	Phe	Leu	Ser	Ala	Gly
						180		185			190				

Gln	Arg	Pro	Arg	Arg	Leu	Ser	Thr	Thr	Ala	Pro	Ile	Thr	Asp	Val	Ser
						195		200			205				

Leu	Val	Asp	Glu	Leu	Arg	Leu	Asp	Gly	Glu	Glu	Val	Asp	Met	Thr	Pro
						210		215			220				

Ala	Asp	Ala	Leu	Asp	Asp	Phe	Asp	Leu	Glu	Met	Leu	Gly	Asp	Val	Glu
						225		230			235			240	

Ser	Pro	Ser	Pro	Gly	Met	Thr	His	Asp	Pro	Val	Ser	Tyr	Gly	Ala	Leu
						245		250			255				

Asp	Val	Asp	Asp	Phe	Glu	Gln	Met	Phe	Thr	Asp	Ala	Leu	Gly		
						260		265			270				

Ile	Asp	Asp	Phe	Gly	Gly										
					275										

<210> SEQ ID NO 50

<211> LENGTH: 47

<212> TYPE: PRT

<213> ORGANISM: *Homo sapiens*

<400> SEQUENCE: 50

Met	Ala	Arg	Arg	Pro	Arg	His	Ser	Ile	Tyr	Ser	Ser	Asp	Glu	Asp	Asp
1						5		10			15				

Glu	Asp	Phe	Glu	Met	Cys	Asp	His	Asp	Tyr	Asp	Gly	Leu	Leu	Pro	Lys
						20		25			30				

US 12,391,948 B2

205**206**

-continued

Ser Gly Lys Arg His Leu Gly Lys Thr Arg Trp Thr Arg Glu Glu
 35 40 45

<210> SEQ ID NO 51

<211> LENGTH: 318

<212> TYPE: PRT

<213> ORGANISM: Mus musculus

<400> SEQUENCE: 51

Met Glu Leu Leu Ser Pro Pro Leu Arg Asp Ile Asp Leu Thr Gly Pro
 1 5 10 15

Asp Gly Ser Leu Cys Ser Phe Glu Thr Ala Asp Asp Phe Tyr Asp Asp
 20 25 30

Pro Cys Phe Asp Ser Pro Asp Leu Arg Phe Phe Glu Asp Leu Asp Pro
 35 40 45

Arg Leu Val His Val Gly Ala Leu Leu Lys Pro Glu Glu His Ala His
 50 55 60

Phe Ser Thr Ala Val His Pro Gly Pro Gly Ala Arg Glu Asp Glu His
 65 70 75 80

Val Arg Ala Pro Ser Gly His His Gln Ala Gly Arg Cys Leu Leu Trp
 85 90 95

Ala Cys Lys Ala Cys Lys Arg Lys Thr Thr Asn Ala Asp Arg Arg Lys
 100 105 110

Ala Ala Thr Met Arg Glu Arg Arg Leu Ser Lys Val Asn Glu Ala
 115 120 125

Phe Glu Thr Leu Lys Arg Cys Thr Ser Ser Asn Pro Asn Gln Arg Leu
 130 135 140

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

Gln Ala Leu Leu Arg Asp Gln Asp Ala Ala Pro Pro Gly Ala Ala Ala
 165 170 175

Phe Tyr Ala Pro Gly Pro Leu Pro Pro Gly Arg Gly Ser Glu His Tyr
 180 185 190

Ser Gly Asp Ser Asp Ala Ser Ser Pro Arg Ser Asn Cys Ser Asp Gly
 195 200 205

Met Met Asp Tyr Ser Gly Pro Pro Ser Gly Pro Arg Arg Gln Asn Gly
 210 215 220

Tyr Asp Thr Ala Tyr Tyr Ser Glu Ala Val Arg Glu Ser Arg Pro Gly
 225 230 235 240

Lys Ser Ala Ala Val Ser Ser Leu Asp Cys Leu Ser Ser Ile Val Glu
 245 250 255

Arg Ile Ser Thr Asp Ser Pro Ala Ala Pro Ala Leu Leu Ala Asp
 260 265 270

Ala Pro Pro Glu Ser Pro Pro Gly Pro Pro Glu Gly Ala Ser Leu Ser
 275 280 285

Asp Thr Glu Gln Gly Thr Gln Thr Pro Ser Pro Asp Ala Ala Pro Gln
 290 295 300

Cys Pro Ala Gly Ser Asn Pro Asn Ala Ile Tyr Gln Val Leu
 305 310 315

<210> SEQ ID NO 52

<211> LENGTH: 537

<212> TYPE: PRT

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 52

US 12,391,948 B2

207

208

-continued

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

Arg Gly Met Arg Phe Arg Tyr Lys Cys Glu Gly Arg Ser Ala Gly Ser
 20 25 30

Ile Pro Gly Glu Arg Ser Thr Asp Thr Thr Lys Thr His Pro Thr Ile
 35 40 45

Lys Ile Asn Gly Tyr Thr Gly Pro Gly Thr Val Arg Ile Ser Leu Val
 50 55 60

Thr Lys Asp Pro Pro His Arg Pro His Glu Leu Val Gly Lys
 65 70 75 80

Asp Cys Arg Asp Gly Phe Tyr Glu Ala Glu Leu Cys Pro Asp Arg Cys
 85 90 95

Ile His Ser Phe Gln Asn Leu Gly Ile Gln Cys Val Lys Lys Arg Asp
 100 105 110

Leu Glu Gln Ala Ile Ser Gln Arg Ile Gln Thr Asn Asn Asn Pro Phe
 115 120 125

Gln Val Pro Ile Glu Glu Gln Arg Gly Asp Tyr Asp Leu Asn Ala Val
 130 135 140

Arg Leu Cys Phe Gln Val Thr Val Arg Asp Pro Ser Gly Arg Pro Leu
 145 150 155 160

Arg Leu Pro Pro Val Leu Ser His Pro Ile Phe Asp Asn Arg Ala Pro
 165 170 175

Asn Thr Ala Glu Leu Lys Ile Cys Arg Val Asn Arg Asn Ser Gly Ser
 180 185 190

Cys Leu Gly Gly Asp Glu Ile Phe Leu Leu Cys Asp Lys Val Gln Lys
 195 200 205

Glu Asp Ile Glu Val Tyr Phe Thr Gly Pro Gly Trp Glu Ala Arg Gly
 210 215 220

Ser Phe Ser Gln Ala Asp Val His Arg Gln Val Ala Ile Val Phe Arg
 225 230 235 240

Thr Pro Pro Tyr Ala Asp Pro Ser Leu Gln Ala Pro Val Arg Val Ser
 245 250 255

Met Gln Leu Arg Arg Pro Ser Asp Arg Glu Leu Ser Glu Pro Met Glu
 260 265 270

Phe Gln Tyr Leu Pro Asp Thr Asp Asp Arg His Arg Ile Glu Glu Lys
 275 280 285

Arg Lys Arg Thr Tyr Glu Thr Phe Lys Ser Ile Met Lys Lys Ser Pro
 290 295 300

Phe Ser Gly Pro Thr Asp Pro Arg Pro Pro Arg Arg Ile Ala Val
 305 310 315 320

Pro Ser Arg Ser Ser Ala Ser Val Pro Lys Pro Ala Pro Gln Pro Tyr
 325 330 335

Pro Phe Thr Ser Ser Leu Ser Thr Ile Asn Tyr Asp Glu Phe Pro Thr
 340 345 350

Met Val Phe Pro Ser Gly Gln Ile Ser Gln Ala Ser Ala Leu Ala Pro
 355 360 365

Ala Pro Pro Gln Val Leu Pro Gln Ala Pro Ala Pro Ala Pro Ala Pro
 370 375 380

Ala Met Val Ser Ala Leu Ala Gln Ala Pro Ala Pro Val Pro Val Leu
 385 390 395 400

Ala Pro Gly Pro Pro Gln Ala Val Ala Pro Pro Ala Pro Lys Pro Thr
 405 410 415

Gln Ala Gly Glu Gly Thr Leu Ser Glu Ala Leu Leu Gln Leu Gln Phe

US 12,391,948 B2

209**210**

-continued

420 425 430

Asp Asp Glu Asp Leu Gly Ala Leu Leu Gly Asn Ser Thr Asp Pro Ala
435 440 445

Val Phe Thr Asp Leu Ala Ser Val Asp Asn Ser Glu Phe Gln Gln Leu
450 455 460

Leu Asn Gln Gly Ile Pro Val Ala Pro His Thr Thr Glu Pro Met Leu
465 470 475 480

Met Glu Tyr Pro Glu Ala Ile Thr Arg Leu Val Thr Ala Gln Arg Pro
485 490 495

Pro Asp Pro Ala Pro Ala Pro Leu Gly Ala Pro Gly Leu Pro Asn Gly
500 505 510

Leu Leu Ser Gly Asp Glu Asp Phe Ser Ser Ile Ala Asp Met Asp Phe
515 520 525

Ser Ala Leu Leu Ser Gln Ile Ser Ser
530 535

<210> SEQ ID NO 53

<211> LENGTH: 225

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: TetR

<400> SEQUENCE: 53

Met Phe Ile Ser Asp Lys Val Ser Ser Met Thr Lys Leu Gln Pro Asn
1 5 10 15

Thr Val Ile Arg Ala Ala Leu Asp Leu Leu Asn Glu Val Gly Val Asp
20 25 30

Gly Leu Thr Thr Arg Lys Leu Ala Glu Arg Leu Gly Val Gln Gln Pro
35 40 45

Ala Leu Tyr Trp His Phe Arg Asn Lys Arg Ala Leu Leu Asp Ala Leu
50 55 60

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

Asp Asp Asp Trp Arg Ser Phe Leu Ile Gly Asn Ala Arg Ser Phe Arg
85 90 95

Gln Ala Leu Leu Ala Tyr Arg Asp Gly Ala Arg Ile His Ala Gly Thr
100 105 110

Arg Pro Gly Ala Pro Gln Met Glu Thr Ala Asp Ala Gln Leu Arg Phe
115 120 125

Leu Cys Glu Ala Gly Phe Ser Ala Gly Asp Ala Val Asn Ala Leu Met
130 135 140

Thr Ile Ser Tyr Phe Thr Val Gly Ala Val Leu Glu Glu Gln Ala Gly
145 150 155 160

Asp Ser Asp Ala Gly Glu Arg Gly Thr Val Glu Gln Ala Pro Leu
165 170 175

Ser Pro Leu Leu Arg Ala Ala Ile Asp Ala Phe Asp Glu Ala Gly Pro
180 185 190

Asp Ala Ala Phe Glu Gln Gly Leu Ala Val Ile Val Asp Gly Leu Ala
195 200 205

Lys Arg Arg Leu Val Val Arg Asn Val Glu Gly Pro Arg Lys Gly Asp
210 215 220

Asp
225

-continued

```

<210> SEQ ID NO 54
<211> LENGTH: 16
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Tat peptide consensus sequence
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(1)
<223> OTHER INFORMATION: Xaa at position 1 can be S or A
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (2)..(2)
<223> OTHER INFORMATION: Xaa at position 2 can be G or A
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Xaa at position 3 can be P or A
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (4)..(4)
<223> OTHER INFORMATION: Xaa at position 4 can be R or K
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (5)..(5)
<223> OTHER INFORMATION: Xaa at position 5 can be P, A, I, Y, K, or R
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (6)..(6)
<223> OTHER INFORMATION: Xaa at position 6 can be R, K, V, or Y
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (7)..(7)
<223> OTHER INFORMATION: Xaa at position 7 can be G, A, or R
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (8)..(8)
<223> OTHER INFORMATION: Xaa at position 8 can be T or A
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (9)..(9)
<223> OTHER INFORMATION: Xaa at position 9 can be R or K
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (10)..(10)
<223> OTHER INFORMATION: Xaa at position 10 can be G or A
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (11)..(11)
<223> OTHER INFORMATION: Xaa at position 11 can be K or A
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (12)..(12)
<223> OTHER INFORMATION: Xaa at position 12 can be G or A
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (13)..(13)
<223> OTHER INFORMATION: Xaa at position 13 can be R or K
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (14)..(14)
<223> OTHER INFORMATION: Xaa at position 14 can be I or A
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (15)..(15)
<223> OTHER INFORMATION: Xaa at position 15 can be R, K, Y, or G
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (16)..(16)
<223> OTHER INFORMATION: X at position 16 can be R, K, V, T, or Y

<400> SEQUENCE: 54
Xaa Xaa
1 5 10 15

```

```

<210> SEQ ID NO 55
<211> LENGTH: 16

```

-continued

<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Tat peptide
<400> SEQUENCE: 55

Ser Gly Pro Arg Pro Arg Gly Thr Arg Gly Lys Gly Arg Ile Arg Arg
1 5 10 15

<210> SEQ ID NO 56
<211> LENGTH: 28
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA binding site consensus sequence
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(5)
<223> OTHER INFORMATION: n is a, c, g, or u
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (14)..(17)
<223> OTHER INFORMATION: n is a, c, g, or u
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (24)..(28)
<223> OTHER INFORMATION: n is a, c, g, or u

<400> SEQUENCE: 56

nnnnnnshsyw sbmnnnnndsb hbsnnnnn

28

<210> SEQ ID NO 57
<211> LENGTH: 28
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: TAR RNA

<400> SEQUENCE: 57

ggcucgugua gcucaauagc uccgagcc

28

<210> SEQ ID NO 58
<211> LENGTH: 29
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA aptamer consensus sequence
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(5)
<223> OTHER INFORMATION: n is a, c, g, or u
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (15)..(18)
<223> OTHER INFORMATION: n is a, c, g, or u
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (25)..(29)
<223> OTHER INFORMATION: n is a, c, g, or u

<400> SEQUENCE: 58

nnnnnnshcys wsbnmnnndsb bhbsnnnnn

29

<210> SEQ ID NO 59
<211> LENGTH: 29
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: TAR Variant-1

<400> SEQUENCE: 59

-continued

ggcucgucug agcuauuag cuccgagcc

29

```

<210> SEQ ID NO 60
<211> LENGTH: 28
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: RNA aptamer consensus sequence
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(5)
<223> OTHER INFORMATION: n is a, c, g, or u
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (14)..(17)
<223> OTHER INFORMATION: n is a, c, g, or u
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (24)..(28)
<223> OTHER INFORMATION: n is a, c, g, or u

<400> SEQUENCE: 60

```

nnnnnshysw sbmnnnndsb hbsnnnnn

28

```

<210> SEQ ID NO 61
<211> LENGTH: 28
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: TAR Variant-2

<400> SEQUENCE: 61

```

ggcucguuga gcuauuagc uccgagcc

28

```

<210> SEQ ID NO 62
<211> LENGTH: 18
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: destabilization domain consensus sequence
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(1)
<223> OTHER INFORMATION: Xaa at position 1 can be S or A
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (2)..(2)
<223> OTHER INFORMATION: Xaa at position 2 can be G or A
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Xaa at position 3 can be P or A
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (4)..(4)
<223> OTHER INFORMATION: Xaa at position 4 can be R or K
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (5)..(5)
<223> OTHER INFORMATION: Xaa at position 5 can be P, A, I, Y, K, or R
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (6)..(6)
<223> OTHER INFORMATION: Xaa at position 6 can be R, K, V, or Y
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (7)..(7)
<223> OTHER INFORMATION: Xaa at position 7 can be G, A, or R
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (8)..(8)
<223> OTHER INFORMATION: Xaa at position 8 can be T or A
<220> FEATURE:

```

-continued

```

<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (9)..(9)
<223> OTHER INFORMATION: Xaa at position 9 can be R or K
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (10)..(10)
<223> OTHER INFORMATION: Xaa at position 10 can be G or A
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (11)..(11)
<223> OTHER INFORMATION: Xaa at position 11 can be K or A
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (12)..(12)
<223> OTHER INFORMATION: Xaa at position 12 can be G or A
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (13)..(13)
<223> OTHER INFORMATION: Xaa at position 13 can be R or K
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (14)..(14)
<223> OTHER INFORMATION: Xaa at position 14 can be I or A
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (15)..(15)
<223> OTHER INFORMATION: Xaa at position 15 can be R, K, Y, or G
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (16)..(16)
<223> OTHER INFORMATION: Xaa at position 16 can be R, K, V, T, or Y
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (17)..(17)
<223> OTHER INFORMATION: Xaa at position 17 can be any amino acid but
    preferably R, G, E, S, or C
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (18)..(18)
<223> OTHER INFORMATION: Xaa at position 18 is optional and can be any
    amino acid, but preferably G, E, O, N, D, or E

```

<400> SEQUENCE: 62

Xaa													
1													15

Xaa Xaa

```

<210> SEQ ID NO 63
<211> LENGTH: 19
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: tDeg

```

<400> SEQUENCE: 63

Ser	Gly	Pro	Arg	Pro	Arg	Gly	Thr	Arg	Gly	Lys	Gly	Arg	Arg	Ile	Arg
1														15	

Arg Arg Gly

```

<210> SEQ ID NO 64
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: lentiviral Tar RNA consensus sequence
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(5)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (14)..(17)
<223> OTHER INFORMATION: n is a, c, g, or t

```

-continued

<220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (24)..(28)
 <223> OTHER INFORMATION: n is a, c, g, or t
 <400> SEQUENCE: 64

nnnnnshsyw sbmnnnndsb hbsnnnnn

28

<210> SEQ ID NO 65
 <211> LENGTH: 28
 <212> TYPE: DNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: TAR RNA

<400> SEQUENCE: 65

ggctcgtgt gtcattagc tccgagcc

28

<210> SEQ ID NO 66
 <211> LENGTH: 29
 <212> TYPE: DNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: lentiviral TAR RNA consensus sequence
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (1)..(5)
 <223> OTHER INFORMATION: n is a, c, g, or t
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (15) ..(18)
 <223> OTHER INFORMATION: n is a, c, g, or t
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (25)..(29)
 <223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 66

nnnnnshcys wsbnnnnnds bhbsnnnnn

29

<210> SEQ ID NO 67
 <211> LENGTH: 29
 <212> TYPE: DNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: TAR Variant-1

<400> SEQUENCE: 67

ggctcgtctg agctcattag ctccgagcc

29

<210> SEQ ID NO 68
 <211> LENGTH: 28
 <212> TYPE: DNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: lentiviral TAR RNA consensus sequence
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (1)..(5)
 <223> OTHER INFORMATION: n is a, c, g, or t
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (14) ..(17)
 <223> OTHER INFORMATION: n is a, c, g, or t
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (24)..(28)
 <223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 68

-continued

nnnnnshysw sbmnnnndsb hbsnnnnn

28

<210> SEQ ID NO 69
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: TAR Variant-2 (Pepper)

<400> SEQUENCE: 69

ggctcggtga gtcatttgc tccgagcc

28

<210> SEQ ID NO 70
<211> LENGTH: 586
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: (Pepper)10 tag

<400> SEQUENCE: 70

ggcucgug	agcuauuag	cuccgagccg	uccagcgcaa	acuauuacga	aaaacaucgg	60
acgggcucgu	ugagcuauu	agcuccgago	ccgcugcgga	aaaccucaca	aaaacacgac	120
aaacgggcuc	guugagcuca	uuagcuccga	gcccggccac	aacccacaaa	cuuacaacca	180
ggcaaacggc	ucgucugagc	ucaauuagcuc	cgagccguau	caagaccgaa	cgccgcaga	240
uaauugacacg	ggcucguuga	gcuauuago	uccgagcccg	accucgcuag	auauguuagg	300
uucuuaggca	uiuggcucguu	gagcuauua	gcuccgagcc	aaagaucgac	ugcaauuuccg	360
auuagacqua	cacggcucgu	cugagcuau	uagcuccgag	ccgauccaac	cuacuuccuc	420
cauaacuaac	cuccggcucg	uugagcuau	uagcuccgag	ccgaucaua	cgcaauuaccg	480
uacacugucc	aaucggcuc	guugagcuca	uuagcuccga	gccggacaac	caaucgacau	540
acaucacacc	acaacucggc	ucgucugagc	ucaauagcuc	cgagcc		586

<210> SEQ ID NO 71
<211> LENGTH: 1466
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: (P30-1xPepper)10 tag

<400> SEQUENCE: 71

uugccaugug	uaugugggau	gcguugccac	guuuccccaca	uacucugaug	auccgcuagc	60
aaaggcucgu	cugagcuau	uagcuccgag	cccgaggguac	cggaucauuc	auggcaaguc	120
cagcgcaauc	uaauacgaaa	aucauccgac	gucgcgaugu	cuaugeggga	ugcguuggca	180
cguuuuccgc	auagucugau	cauccgcua	caaaggcucg	uugagcucau	uagcuccgag	240
cccgagguac	cggaugauuc	aucgcgcacgc	ugcgaaaaau	cucacaaaaa	cacgucaa	300
gucgcccugu	guguguagga	ugcguugcca	cgguuuccuac	acacucugac	gauccgcuag	360
caaaggcucg	uugagecuau	uagcuccgag	cccgagguac	cggaucguuc	acggcgacgc	420
cguauaaucca	cauacuuaca	aucaggcaau	cuugccaugu	guauguggga	ugcguuggca	480
cguuuuccac	auacucugau	gauccgcua	caaaggcucg	uugagcucau	uagcuccgag	540
cccgagguac	cggaucauuc	auggcaagua	ucaagaucga	acggcgcaag	auauugucac	600
gucgcgaugu	uaugcgggga	ugcguugcca	cgguuuccgc	auagucugau	cauccgcua	660
caaaggcucg	ucugagcuca	uuagcuccga	gcccgaggua	ccggauauu	caucgcgacg	720
uccucgcua	auaughuagg	uucuuaggca	uuucgcccug	uguguguagg	augeguugcc	780

-continued

acguuuuccua cacacucuga cgauccgcua gcaaaggcuc guugagcuca uuagcuccga	840
gcccggaggua ccgggaucgau cacggcgaaa agaucgugcug caauuuccgau uagacguaca	900
cuuggccaugu guauggggga ugcuugugcca cguuuucccac auacucugau gauccgcuag	960
caaaggcucg uugagcucau uagcucccgag cccgaggguac cggaucauuc auggcaagau	1020
ccaagcuacu uccuccauac cuauccuccu cgcgaugucu augcggaug cguugccacg	1080
uuuucccgcau agucugauca uccgcuagca aaggcucgguu gagcucauua gcuccgagcc	1140
cgggguaccg gaugauucau cgcgagauca uaacgcaaua cgcguacacug uccaaucuc	1200
gcccggugug uguaggugugc guggccacgu uuccuacaca cucugacgau cgcguagcaa	1260
aggcucgucu gagcucauua gcucccgagcc cgaggguaccg gaucguucac ggcgaggaua	1320
aucaauccac auacaucaca ccacaauuuc uggcaugugu augugggaug cguugccacg	1380
uuuucccacau acucugauga uccgcuagca aaggcucguc ugagcucauu agcuccgagc	1440
ccgaggguacc ggaucauuca uggcaa	1466

<210> SEQ ID NO 72

<211> LENGTH: 1228

<212> TYPE: RNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: (Pepper)20 tag

<400> SEQUENCE: 72

ggcucgcugc agcucauuag cucccgagccg uccagcgcaaa acuauuacgaaa aacauccg	60
acggggcucgu ugagcuauu agcucccgagcc cgcugcgaa aaaccucaca aaaacacgac	120
aaacggggcuc guugagcuca uuagcuccgaa gccccggcgc aacccacaaa cuuacaacca	180
ggcaaaacggc ucgucugagc ucauuagcuc cgagccguau caagaccgaa cggcgcaaga	240
uaauugacacg ggcucguuga gcucuuuagc uccegagcccc accucgcuag auauuguagg	300
uuuuuaggca uuggccgauu gagcucauua gcucccgagcc aaagauucgac ugcaauuccg	360
auuuagacgaa cacggcucgu cugagcuau uagcucccgag cgcgaucac cuacuuccuc	420
cuaaacuaac cuccggcucg uugagcuau uagcucccgag cgcgaucuaa cgcaauaccg	480
uacacugucc aaucgggcuc guugagcuca uuagcuccgaa gccccgacaac caaucgacau	540
acaucacacc acaacucggc ucgucugagc ucauuagcuc cgagccgaaau ugguucguuc	600
ucuuuggggc cgcucgacua aggugacaac ugacaaacc cgcucguugc ugagcucauu	660
agcucccgagc cgacucucac caacaagaca aaaacuacuc uucuaggcuc guugagcuca	720
uuuagcuccga gccuaaacac ucaagcauac auugugccua uuuuuggcuc ggcucgagc	780
cuaauagcucc gggccaaugc cucacgaaau ucaaaacacg gacaaggggc ucgugagc	840
cuaauagcucc gagcccguauc cacguccaaau acgauuacuc accuuucggc cgcugagc	900
ucaauagcuc cgcggccgca gcuacauac uuccacucag gacaauucaag ggcucgucug	960
agcucauuag cucccgagccc uccacaaguc ucaaccacag aaacuaccaa augggcucgu	1020
ugagcuauu agcucccgagc ccacuccuac cucaaaccuc uucccacaaa acuggggcuc	1080
guugagcuca uuagcuccga gcccccauuc caacauacca aauaaaaac auuuacuggc	1140
ucgucugagc ucaauagcuc cgagccgagcc cacaucucuc acuacuauc aaaaaccaac	1200
ggcucguuga gcuauuagc uccgagcc	1228

<210> SEQ ID NO 73

-continued

<211> LENGTH: 1812
 <212> TYPE: RNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: (F30-2xPepper)10 tag

 <400> SEQUENCE: 73

uuugccauug	uaugugggaa	gcguagaaag	gcucguugag	cucauuagcu	ccgagcccg	60
cuacguuuuc	cacauacucu	gaugauccgc	uagcaaaggc	ucgucugagc	ucauuagcuc	120
cgagccccag	guacgggauc	auucauggca	aguuccagcgc	aaucuauuac	gaaaaucauc	180
cgacgucg	augucuaugc	ggaaagcgu	gaaaggcuc	ucugagcu	uuagcuccg	240
gcccga	acuac	guuuucccg	uagucuga	auccgcu	aaaggcuc	300
acuccg	gagc	ggaga	ggaugau	ucgcgac	gcccggaa	360
acgucaa	acg	uguguagg	gcuagaa	gcucg	ucauuagc	420
uccgagcc	acuacg	uuuc	cuacacac	ugacgau	aggcaagg	480
ucauuag	uc	guac	ggggaa	cg	ucguugag	540
uacaau	cagg	gg	ggggaa	u	ucguu	600
uagcucc	cccg	gg	ggggaa	ac	uac	660
gagcuau	uu	gg	ggggaa	u	uac	720
ggcgcaag	au	gg	ggggaa	cg	uac	780
ucauuag	uac	gg	ggggaa	u	uac	840
cgucug	ggcc	gg	ggggaa	cg	uac	900
agauau	gg	gg	ggggaa	g	uac	960
ugagcuau	ag	gg	ggggaa	gg	uac	1020
aggcucgu	u	gg	ggggaa	gg	uac	1080
ucgucu	gg	gg	ggggaa	gg	uac	1140
uacuuag	gg	gg	ggggaa	gg	uac	1200
cgagcc	gg	gg	ggggaa	gg	uac	1260
auccaag	gg	gg	ggggaa	gg	uac	1320
ggcucgu	gg	gg	ggggaa	gg	uac	1380
gcuagca	gg	gg	ggggaa	gg	uac	1440
gagaau	gg	gg	ggggaa	gg	uac	1500
gaaaggc	gg	gg	ggggaa	gg	uac	1560
auccgc	gg	gg	ggggaa	gg	uac	1620
cggcgagg	aa	gg	ggggaa	gg	uac	1680
gcuuagg	gg	gg	ggggaa	gg	uac	1740
uugaug	gg	gg	ggggaa	gg	uac	1800
cauucaugg	aa					1812

<210> SEQ ID NO 74
 <211> LENGTH: 8010
 <212> TYPE: DNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: pCMV-mCherry- (F30-2xPepper)10

<400> SEQUENCE: 74

US 12,391,948 B2

227

228

-continued

gacggatcgg gagatctccc gatccctat ggtgcactt cagataatc tgctctgatg
ccgcatatgtt aagccagtat ctgctccctg cttgtgtgtt ggaggcgct gagtagtgcg
cgagcaaat ttaagtcaca acaaggcaag gcttgaccga caatttgcatt aagaatctgc
ttagggttag gcttttgcg ctgcttcgcg atgtacgggc cagatatacg cggtgacatt
gattattgac tagttattaa tagtaatcaa ttacgggtc attagttcat agccatata
tggagtccg cgttacataa cttacggtaa atggccgcgc tggctgaccg cccaaacgacc
ccggccatt gacgtcaata atgacgtatg ttcccatagt aacgcctaata gggacttcc
attgacgtca atgggtggag tatttacgtt aaactgccc cttggcagta catcaagtgt
atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt
atgcccagta catgaccta tggacttgc ctacttggca gtacatctac gtattagtca
tcgttattac catggtgatg cgggttggc agtacatcaa tggggctgga tagcgggttgc
actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttgc ttttggcacc
aaaatcaacg ggacttcca aatgtcgta acaactccgc cccattgac gcaaattggc
ggtagggcgtg tacgggtggga ggtctatata agcagagctc tctggtaac tagagaaccc
actgcttact ggcttatcga aattaatacg actcactata gggagacccca agctggctag
cgtttaact taagcttgc accatggta gcaagggcga ggaggataac atggccatca
tcaaggagt catgcgttc aaggtgcaca tggagggtc cgtgaacggc cacgagttcg
agatcgaggg cgagggcgag ggccgcctt acgagggcac ccagaccgccc aagctgaagg
tgaccaaggg tggccccctg cccttcgcctt gggacatctt gtcccttcag ttcatgtacg
gctccaaggc ctacgtgaag caccggcccg acatccccga ctacttgaag ctgtcccttc
ccgagggtt caagtggag cgcgtatga acttcgagga cggcggcgtg gtgaccgtga
cccaggactc ctccctgcag gacggcgagt tcacatcaca ggtgaagctg cggccacca
acttccctc cgacggcccc gtaatgcaga agaagaccat gggctggag gcttccctccg
agcggatgtt ccccgaggac ggcgcctga agggcgagat caagcagagg ctgaagctga
aggacggcgcc ccactacgac gctgaggta agaccaccta caaggccaaag aagccgtgc
agctgccccgg cgcctacaac gtcacatca agttggacat cacctccac aacgaggact
acaccatcg ggaacagtac gaacgcggc agggccgcctt ctccacccggc ggcattggacg
agctgtacaa gtaactcgag atccgttacg gccggaatca atcgctaatac actcaacttg
ccatgtgtat gtggaaagcg tagaaaggct cgttggatctt attagctccg agcccgacta
cgtttccac atactctgtat gatccgttagt caaaggctcg tctggatctt ttatgtcccg
gcccggatgtt ccggatcatt catggcaagt ccagcgcaat ctattacgaa aatcatccga
cgtcgccatgtt tctatgcggg aagcgtagaa aggctcgatc gagctcattt gctccgagcc
cgactacgtt tcccgcatag tctgtatcatc cgctagcaaa ggctcgatc gctcatttgc
tccgagcccg aggtaccgga tgattcattc cgacgcgtcg gaaaatctca caaaatcact
tcaaacgtcg ccgtgtgtgtt gtaggaagcg tagaaaggct cgtctggatctt cattagctcc
gagccccact acgtttccata cacactctga cgatccgcata gcaaaggctc gttgagctca
ttagctccga gcccggatgtt ccggatcattt cacggcgacg ccgataatcc acatacttac
aatcaggcaat tcttgcctatc tgtagtgcggg aagcgtagaa aggctcgatc agctcatttgc
ctccgagcccg gactacgtt cccacataact ctgtatcatc gctagcaaaag gctcgatc
2100
2160
2220
2280
2340

US 12,391,948 B2

229

230

-continued

ctcattagct ccgagccgaa ggtaccggat cattcatggc aagtatcaag atcgaacggc	2400
gcaagatatt gtcacgtcgc gatgtctatg cggaaagegt agaaaggctc gttgagctca	2460
ttagctccga gcccgactac gtttcccga tagtctgatc atccgctagc aaaggctcgt	2520
ctgagctcat tagctccgag cccgaggtac cggatgattc atcgcgacgt cctcgctaga	2580
tatgttaggt tcttaggcat ttgcgcgtgt gtgtgttagga agcgttagaaa ggctcggtga	2640
gctcattagc tccgagcccg actacgttgc ctacacactc tgacgatccg ctagcaaagg	2700
ctcgctctgag ctcattagct ccgagccgaa ggtaccggat cgttcacggc gaaaagatcg	2760
tctgcaattc cgattagacg tacacttgcg atgtgtatgt gggaaagcgta gaaaggctcg	2820
tctgagctca ttagctccga gcccgactac gtttcccaca tactctgatg atccgctagc	2880
aaaggctcgt tgagctcatt agtcccgagc cccgaggtacc ggatcattca tggcaagatc	2940
caagctactt cctccataacc tatectccctc gcatgtctta tgcggaaagc gtagaaaggc	3000
tcgtctgagc tcattagctc cgagcccgac tacgtttccc gcatagtctg atcatccgct	3060
agcaaaaggct cggttagcgc attagctccg agcccggagg accggatgtat tcatcgccag	3120
atcataacgc aataccgtac actgtccaaat cctcgccgtg tgggtgttagg aagcgttagaa	3180
aggctcgtct gagctcatttta gctccgagcc cgactacgtt tcctacacac tctgacgatc	3240
cgctagcaaa ggctcggtga gctcattago tccgagcccg aggtaccggta tcgttcacgg	3300
cgaggataat caatccacat acatcacacc acaatttctt ccatgtgtat gtggaaagcg	3360
tagaaaggct cgtctgagct cattagctcc gagcccgact acgtttccca catactctga	3420
tgatccgcta gcaaaaggctc gtctgagctc attagctccg agcccggagg accggatcat	3480
tcatggcaag aattggtcgt tcttcttggc ggccgctcgta ctaaatcacc ggtaatcttc	3540
ttgtccatct agaccttata aagatctttg tacaaggccc cgtttaaacc cgctgatcat	3600
cctcgactgt gccttctagt tgccagccat ctgttgtttg cccctccccc gtgccttcct	3660
tgacccttggaa aagggtccac tcccactgtc ctttcctaaat aaaatgagga aattgcatcg	3720
cattgtctga gtaggtgtca ttcttattctg ggggggtgggg gtggggcag gacagcaagg	3780
ggggaggatgg ggaagacaat agcaggcatg ctggggatgc ggtggggctct atggcttctg	3840
aggcggaaag aaccagctgg ggctcttaggg ggtatccccca cgcccccgtgt agcggccat	3900
taagcgcggc ggggtgtggg gttacgcgcga gcgtgaccgc tacacttgcc agcgccttag	3960
cgcccgctcc ttccgttttc ttcccttcct ttctcgccac gtccgcggc tttccccgtc	4020
aagctctaaa tcgggggctc cttttaggt tccgatttag tgctttacgg cacctcgacc	4080
ccaaaaaaact tgattagggt gatgggtcac gtagtgggcc atcgccctga tagacggttt	4140
ttccgcctttt gacgttggag tccacgttct ttaatagtgg actcttggtc caaactggaa	4200
caacactcaa ccctatctcg gtcttattctt ttgatttata agggattttg ccgatttcgg	4260
cctattgggt aaaaatgag ctgatttaac aaaaatttaa cgcgattaa ttctgtggaa	4320
tgtgtgtcag ttaggggtgt gaaagtcccc aggctccccca gcaggcagaa gtatgaaag	4380
catgcacatctc aatttagtcag caaccaggtg tggaaagtc ccaggtccc cagcaggcag	4440
aagtatgcaa agcatgcac tcaatttagtc agcaaccata gtcccgcccc taactccgccc	4500
catcccgccc ctaactccgc ccagttccgc ccattctccg ccccatggct gactaatttt	4560
tttttattttt gcaaggcccg aggcggccctc tgccctgttag ctattccaga agtagtgagg	4620
aggctttttt ggaggccctag gcttttgcaa aaagctcccc ggagcttgta tatccatttt	4680
cggatctgat caagagacag gatgaggatc gttcgcatg attgaacaag atggattgca	4740

-continued

cgcagggtct cggccgctt gggggagag gctattccgc tatgactggg cacaacagac	4800
aatcggtgc tctgatgccg ccgtgttccg gctgtcagcg cagggggccc cggttcttt	4860
tgtcaagacc gacctgtccg gtgcctgaa tgaactgcag gacgaggcag cgccgctatc	4920
gtggctggcc acgacggcgc ttccctgcgc agctgtgctc gacggttgtca ctgaagcggg	4980
aaggggactgg ctgctattgg gcgaagtgcc ggggcaggat ctccctgtcat ctcaccttgc	5040
tcctgcccag aaagtatcca tcatggctga tgcaatgcgg cggctgcata cgcttgatcc	5100
ggctacactgc ccatcgacc accaagcgaa acatcgatc gagcgagcac gtactcgat	5160
ggaagccggt ctgtcgatc agggatgtct ggacgaagag catcaggggc tcgcgcacgc	5220
cgaactgttc gccaggctca aggccgcgcat gccccacggc gaggatctcg tcgtgaccca	5280
tggcgatgcc tgcttgcga atatcatggt gggaaatggc cgcttttctg gattcatcga	5340
ctgtggccgg ctgggtgtgg cggaccgcta tcaggacata gctgtggcta cccgtgatata	5400
tgctgaagag ctggggggcg aatgggctga ccgcgttcctc gtgcgttacg gtatcgccgc	5460
tcccgattcg cagcgcatcg ccttctatcg cttcttgcac gagttttctt gaggcggact	5520
ctgggggttcg aaatgaccga ccaagcgacg cccaaacctgc catcagcaga tttcgattcc	5580
acogccgcct tctatgaaag gttgggcttc ggaatcgttt tccgggacgc cggtggatg	5640
atccctccagc gcggggatct catgtggag ttcttcgcacc accccaaactt gtttttgca	5700
gcttataatg gttacaaaata aagcaatago atcacaaattt tcaaaaaataa agcattttt	5760
tcactgcattt ctagttgtgg tttgtccaaa ctcattcaatg tatcttatca tgtctgtata	5820
ccgtcgacct ctagctagag cttggcgtaa tcatggtcat agctgtttcc tttgtgaaat	5880
tgttatccgc tcaacattcc acacaacata cgagccggaa gcataaaatg taaaggctgg	5940
ggtgtccataat gagttagctt actcacatta attgcgttgc gctcaactgcc cgctttccag	6000
tcggggaaacc tgcgtgttca gctgcattaa tgaatcgcc aacgcgcggg gagaggcgg	6060
tttgcgttgc ggcgccttcc cgcttcctcg ctcaactgact cgctgcgcgc ggtcggttgc	6120
ctgcggcggag cggtatcagc tcactcaaag gcggttaatacg gtttatccac agaatcagg	6180
gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag	6240
ggccgcgtgc tggcggtttt ccataggctc cgcggccctgc acgagcatca caaaaatcga	6300
cgctcaagtc agagggtggcg aaacccgaca ggactataaa gataccaggc gttttccct	6360
ggaagctccc tcgtgtcgctc tcctgttccg accctgcgcgc ttacccgata cctgtccgc	6420
tttctccctt cgggaagcgtt ggcgttttct catagtcac gctgttaggtt tctcagtcg	6480
gtgttaggtcg ttcgttccaa gctgggtgtt gtgcacgaac ccccggttca gccccaccgc	6540
tgcgcccttat ccggtaacta tcgttccat gccaacccgg taagacacga cttatcgcc	6600
ctggcgcgcag ccactggtaa caggatttgc agagcgaggat atgttaggcgg tgctacagag	6660
ttcttgaagt ggtggcctaa ctacggctac actagaagaa cagtttttttgc tatctgcgc	6720
ctgtgtaaac cagtttaccc cggaaaaaga gttggtagctt ctgtatccgg caaacaaacc	6780
accgctggta cgggtttttt tgggttgcag cagcagatca cgcgcagaaa aaaaggatct	6840
caagaagatc ctttgcgttcc ttcgttccgg tctgtacgcgc agtggaaacga aaactcacgt	6900
taagggttgc tgggtcatgatc attatcaaaa aggttccatca ccttagatcctt tttaaatata	6960
aaatgaagtt ttaaatcaat ctaaaatgtata tatgagtaaa cttgggtctga cagttacca	7020
tgcttaatca gtgaggcacc tatctcagcg atctgtctat ttcgttcatc catagttgcc	7080

-continued

tgactccccg tcgttagat aactacgata cgggaggggct taccatctgg ccccagtgtc	7140
gcaatgatac cgcgagaccc acgctcaccc gctccagatt tatcagcaat aaaccagcca	7200
gccggaaggg ccgagcgac agtggctt gcaactttat ccgcctccat ccagtctatt	7260
aattgttgcg gggaaagctag agtaagtagt tcgcagttt atagttgcg caacgttgc	7320
gccattgcta caggcatcgt ggtgtcacgc tcgtcggtt gtagtggctc attcagctcc	7380
ggttccaaac gatcaaggcg agttacatga tccccatgt tggcaaaaa agcggttagc	7440
tccctcggtc ctccgatcgt tgtcagaatg aagttggccg cagtgttatac actcatgtt	7500
atggcagcac tgcataattc tcttactgtc atgcacatccg taagatgtt ttctgtgact	7560
ggtgagtaact caaccaagtc attctgagaa tagtgtatgc ggcgaccggag ttgctctgc	7620
ccggcgtcaa tacggataa taccgcgcca catagcgaaaa cttaaaaatg gctcatcatt	7680
ggaaaacgtt ctccggggcg aaaactctca aggtatcttac cgctgttgcg atccagttcg	7740
atgttaaccca ctcgtgcacc caactgtatct tcagcatctt ttactttcac cagcgttct	7800
gggtgagcaa aaacaggaag gcaaaatgcc gcaaaaaagg gaataaggc gacacggaaa	7860
tgttgaatac tcataactctt ccttttcaa tattattgaa gcatttatca gggttattgt	7920
ctcatgagcg gatacatatt tgaatgtatt tagaaaaata aacaatagg gggtccgcgc	7980
acatttcccc gaaaagtgc acctgacgtc	8010

<210> SEQ ID NO 75

<211> LENGTH: 7762

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: pminiCMV- (mNeonGreen) 4-tDeg

<400> SEQUENCE: 75

gacggatcg gagatctccc gatccctat ggtgcactct cagttacaatc tgctctgtat	60
ccgcataatgg aagccagttt ctgcctccgtt cttgtgttgg gggggcgctt gagtagtgc	120
cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcattt aagaatctgc	180
ttagggtagt gggttttgcg ctgttgcgtt atgtacggccg cagatatacg cgttggtagg	240
cgtgtacggt gggaggccata tataaggcaga gctaagcttgc ccaccatgtt gagcaaggc	300
gaggaggata acatggccctc tctccctatgc acacatgttgc tacatctt tggctccatc	360
aacgggtgtgg actttgacat ggtgggtcag ggcacccggca atccaaatga tggttatgag	420
gagttaaacc tgaagtccac caagggttgc ctccagtttcc cccctggat tctggccct	480
catatcggtt atggcttcca tcagttacccgc ccctaccctgc acgggtatgc gcctttccat	540
gccgcatgg tagatggctc cggatccaa gtccatcgca caatgcgtt tgaagatgg	600
gcctccctta ctgttaacta ccgttacacc tacggggaa gccacatcaa aggagaggcc	660
cagggtgttgg ggactgggtt ccctgttgcg ggttctgtgttgc tgaccaactc gctgttccgt	720
gcggacttgtt gcagggtcgaa gaagacttac cccaaacgaca aaaccatcat cgttccgtt	780
aagtggagtt acaccactgg aatggcaag cgctaccggc gcactgcgcg gaccacccatc	840
acctttggca agccaaatggc ggcttactat ctgttgcgtt agccgtatgtt cgttccgtt	900
aagacggagc tcaaggactc caagaccggat ctcacttca aggagtggca aaaggccccc	960
acccatgttgc tgggtatggc cgatgttgc aagggtggac atatggccac agggccacaca	1020
ggcggttccg gggggatccc caaaggagaa gaagacaata tggcatact ccccgcaacc	1080
cacgagttgc atatccggg ttcaattaat ggatgttgc tgcataatgg tggccaggaa	1140

-continued

acaggaaacc caaacgacgg atatgaagag cttaatctca aaagtaccaa aggcgatctg	1200
caattttctc cgtggatact cgtgccacac attggatacg gatttcacca atatctcccg	1260
tatccggatg gaatgtcccc ctttcaagca gcaatggtgg acgggagtg ttatcaggta	1320
cacagaacca tgcagttcga ggacggggct tctctgaccg taaattatag gtatacttat	1380
gaaggctcac atattaaggg cgaagcacag gttaaaggaa ccgggttcc tgccgatggc	1440
cccgcatga ctaattctct gacagccgca gattgggtgc gtcacaaaaa gacatacccg	1500
aatgataaga ctataatctc aacattcaaa tggtcctata cgacaggcaa cgggaaacga	1560
tatagatcca cggctcgaac aacttacaca ttgcctaaac ctatggccgc caattacctc	1620
aaaaatcagc ccatgtatgt gtttaggaaa accgaattga agcattctaa aacggaactt	1680
aattnaagg aatggcagaa ggcttcaca gacgtaatgg ggatggatga actctataaa	1740
tcaaggctcg agtccctcagg gggAACGGGT gggccggag gagttagtaa aggtgaagag	1800
gacaatatgg caagtttgc tgccgactcac gagttcatac tctttgggtc tataaatggc	1860
gttgacttcg atatggttgg ccaaggtact ggcaacccca atgacggtaa cgaggagttg	1920
aatctcaagt ccacaaaagg tgatottcag ttcaaggccc ttgattctcgt acctcatatt	1980
ggatatggct ttcaaggcta cttccatacc ccagacggta tgtaacccct tcaagctgcg	2040
atgggtggatg gttccggcta tcagggtccac cgaacgatgc aattcgagga cggggccagc	2100
ctcaccgtta attataggtt cacctatgag ggaagtcaca taaaggaga agcccaagtg	2160
aaaggaacag gattcccgac tgatggtcca gtaatgacgaa actccctgac agccgctgac	2220
tgggttagaa gaaaaagac gtatcctaattt gacaagacca tcattagcac tttcaatgg	2280
agttataccca caggaaacgg caaacggtaa agaaggactc cttagaactac ctacacttc	2340
gcaaaaggca tggctgcaaa ctatttgaag aatcagccca tgtacgttt tcgaaaaacg	2400
gaacttaagc acagtaagac tgaacttaat ttcaaggagt ggcagaaggc gttcacggat	2460
gtcatggta tggatgact gtataaggga gggctggca ctgggggcac tgccagcagc	2520
ggatccgggt ggggtgttag caaggccgag gaggataaca tggccctctt cccagcgaca	2580
catgagttac acatctttgg ctccatcaac ggtgtggact ttgacatggt gggcagggc	2640
accggcaatc caaatgatgg ttatgaggag ttaaacctga agtccaccaa gggtgacctc	2700
cagttctccc cctggattct ggtccctcat atcgggtatg gttccatca gtacctgccc	2760
taccctgacg ggtgtcgcc ttccaggcc gccatggtag atggctccgg ataccaagtc	2820
catcgacaa tgcagttga agatggtccc tcccttactg ttaactaccc ctacacccat	2880
gagggaaagcc acatcaaagg agaggcccgat gtgaagggaa ctggttccc tgctgacgg	2940
cctgtatgca ccaactcgct gaccgctcg gactgggtca ggtcgaagaa gacttacccc	3000
aacgacaaaaa ccatcatcag tacctttaag tggagttaca ccactggaaa tggcaagcgc	3060
taccggagca ctgcgcggac cacctacacc tttgccaacg caatggccgc taactatctg	3120
aagaaccagc cgatgtacgt gttccgttaag acggagctca agcactccaa gaccgagctc	3180
aacttcaagg agtggcaaaa ggccttaccg gatgtgttagt gcatggacgaa gctgtacaag	3240
ggcggaaat cccgggtgg ttctggctt cgtccccctg gtactcggtt taaagggtcgc	3300
cgtatcgatc gccgcggta atcttagaggg cccgtttaaa cccgctgatc agcctcgact	3360
gtgccttcata gttgccagcc atctgttggt tgccctccc ccgtgccttc cttgaccctg	3420
gaaaggtgcc actcccactg tcctttcata ataaaatgag gaaattgcat cgcatgtct	3480

-continued

gaggtagtgt cattcttatc tgggggttgg ggggtggggc aggacagcaa gggggaggat 3540
tggaaagaca atagcaggca tgctggggat ggggtgggc ctatggttc tgaggcgaa 3600
agaaccacgt ggggctctag ggggtatccc cacgcgcctt gtagccgcg attaagcgcg 3660
gggggtgtgg tggtaacgca cagcgtgacc gctacacttg ccagccctt agcccccgt 3720
ccttcgtttt tttcccttc ctttcgtcc acgttcgcgg gtttccccg tcaagctcta 3780
aatcggggc tccctttagg gttccgattt agtgctttac ggacccctga ccccaaaaaa 3840
cttgatttagg gtgtatggttc acgttagtggg ccacgcctt gatagacggg tttccctt 3900
ttgacgttgg agtccacgtt cttaatagt ggactcttgc tccaaactgg aacaacactc 3960
aacccatatct cggcttatc ttttgattt taagggattt tgccgattt ggcctattgg 4020
ttaaaaaatg agctgattt aaaaaattt aacgcgaatt aattctgtgg aatgtgttc 4080
agttagggtg tggaaagtcc ccaggctccc cagcaggcag aagtatgca agcatgcattc 4140
tcaattatgc agcaaccagg tggaaagtcc cccagggctt cccagcaggc agaagtatgc 4200
aaagcatgca tctcaattatgc tcaagcaacca tagtccgcctt cctaaactccg cccatcccgc 4260
ccctaaactcc gcccagtcc gcccatttc cgcggatgg ctgacttaatt ttttttattt 4320
atgcagaggc cgaggccgcc tctgcctctg agtattcca gaagtagtga ggaggctttt 4380
ttggaggccctt aggttttgc aaaaagctcc cgggagcttgc tatattccatt ttcggatctg 4440
atcaagagac aggttgcgc tggatttgc tggatttgc tggatttgc agatggattt caccgggtt 4500
ctccggccgc ttgggtggag aggctattcg gctatgactt ggcacaaacag acaatcggt 4560
gtctgtatgc cggctgttc cggctgtca cggcaggggcg cccgggtttt tttgtcaaga 4620
ccgacctgtc cgggtcccttgc aatgaacttgc aggacggaggc agcgcggctt tggctgtgg 4680
ccacgacggg cgttcccttgc cggactgttc tggacgttgc cactgaagcg ggaaggact 4740
ggctgtattt gggcgaagtg cggggccagg atcttcgttc atctcacctt gctctggcc 4800
agaaaagtatc catcatggct gatgcaatgc gggggcttca tacgttgc tggcttgc 4860
gcccatttgc caaccaagcg aaacatgc tggaggcaggc agtacttgc atggaaaggcc 4920
gtcttgcga tcaggatgtt cggacggatc agcatcaggc gtcggccca gccaactgt 4980
tcggccaggctt caaggcgccgc atggccgcac gggaggatctt cgtcgccatc catggcgatc 5040
cctgttgcga aatatcatg tggaaaatg gcccattttc tggatttgc gacttggcc 5100
ggctgggtgtt gggggccgc tttttttttt tttttttttt tttttttttt tttttttttt 5160
agcttggccggc cgaatgggtt gaccgcttgc tggacttttgc cggatcgcc gctcccgatt 5220
cgccggccat cggccattttt cggccattttt acggatgtttt cggatcgcc gctcccgatt 5280
cgaaatgacc gaccaagcg cggccaaactt gccatcagca gatttcgattt ccacccggcc 5340
cttctatggaa aggttgggtt tggaaatgtt tttccggcc gcccggccg tggatcttcc 5400
ggccggggat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 5460
tggttacaaa taaagcaata gcatcacaaa tttcacaaat aaagcattttt ttttacttgc 5520
ttcttagtgc ggtttgttca aactcatcaat tttttttttt tttttttttt tttttttttt 5580
ctcttagcttgc agcttggccgtt aatcatggcc tttttttttt tttttttttt tttttttttt 5640
gctcacaattt ccacacaaca tacggccgg aaggatcattttt tttttttttt tttttttttt 5700
atggatggcc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 5760
cttgcgtgc cagctgcattt aatgaatcgcc ccaacggccg gggagaggcc gtttgcgtat 5820
tggggcgctt tccgcttcc cggctacttgc cttggctggcc tgggtcggtt gggccggcc 5880

US 12,391,948 B2

239

240

-continued

agcggtatca gtcactcaa aggccgtaat acggtatcc acagaatcag gggataaacgc	5940
aggaaagaac atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa aggccgcgtt	6000
gctggcgaaa ttccataggc tccgcggccc tgacgagcat cacaaaaatc gacgctcaag	6060
tcaagagggtgg cgaaacccga caggactata aagataccag gcgtttcccc ctggaagctc	6120
cctcggtgcgc tctcctgttc cgaccctgccc gcttaccgga tacctgtccg ccttctccc	6180
ttcgggaagc gtggcgctt ctcatalogt acgctgttagg tatctcagtt cggtgttaggt	6240
cgttcgatcc aagctgggtgtgtcaega accccccgtt cagcccgacc gctggcgctt	6300
atcccgtaac tatacgcttg agtccaaccc ggtaagacac gacttatcgc cactggcagc	6360
agccactggtaacaggatttgcagagcgag gtatgttaggc ggtgctacag agttcttcaa	6420
gtgggtggcactaactacggtaacactagaag aacagtattt ggtatctgcgtctgtcaaa	6480
gccagttacc ttccggaaaaa gagttggtag ctcttgatcc ggcaaaacaaa ccaccgcgtt	6540
tagcgggtttt tttgtttgtca agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga	6600
tcctttgatc ttttctacgg ggtctgacgc tcagtgaaac gaaaactcac gttttagggat	6660
tttggtcatg agattatcaa aaaggatctt caccttagatc cttttaattt aaaaatgaag	6720
ttttaaatca atctaaagta tatatgagta aacttggctt gacagttacc aatgtttaat	6780
cagtgaggc cctatctcag cgatctgtctt attcgttca tcctatgttgcctactccc	6840
cgtcgtgttag ataactacga tacgggaggg cttaccatct ggccccagtg ctgcaatgt	6900
accgcgcgac ccacgcac cggctccaga ttatcagca ataaaccagc cagccggaaag	6960
ggcccgagcgc agaagtggc ctgcaacttt atccgcctcc atccagtcata ttaatttttgc	7020
ccggggaaagctt agagtaagttt gttcgccagt taatagttt cgcaacgttgc ttgcattgc	7080
tacaggcatacgttgggtgtcac gctcgatcgatc tggatggctt tcattcagtttccca	7140
acgatcaagg egagttacat gatccccat gtttgcaaa aaagcggtaa gtccttcgg	7200
tcctccgatc gttgtcagaa gtaagttggc cgcagtgatc tcactcatgg ttagggcagc	7260
actgcataat tctcttactg tcatgccatc cgtaagatgc ttttctgtga ctgggtgatc	7320
ctcaaccaag tcattctgag aatagtgtat gcggcgaccg agttgttctt gccccggcgtc	7380
aatacggatc aataccgcgc cacatagcag aactttaaa gtcgtatca ttggaaaacg	7440
ttcttcgggg cgaaaactctt caaggatctt accgctgtt agatccagtt cgatgttacc	7500
cactcgtgc cccaaactgtat cttcagatc ttttactttc accagegtt ctgggtgagc	7560
aaaaacagga aggcaaaaatg ccgcggaaaaa gggataagg ggcacacggaaatgttgaat	7620
actcataactc ttccttttc aatattattt aagcattttt cagggttattt gtctcatgag	7680
cgatcataata tttgtatgtt ttttagaaaaa taaacaaata ggggttccgc gcacatttcc	7740
ccgaaaaatgc ccacctgacgc tc	7762

<210> SEQ ID NO 76
<211> LENGTH: 8130
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: pCMV-CyTERM-mCherry- (F30-2xPepper)10

<400> SEQUENCE: 76

gacggatcgagatctccc gatccccat ggtgcactct cagtcataatc tgctctgtat	60
ccgcataatc aagccagttat ctgtccctg cttgtgttgg ggggttccgc gagtagtgcg	120

-continued

cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc	180
ttagggtagt gcgtttgcg ctgcttcgat atgtacgggc cagataacg cggtgacatt	240
gattattgac tagttattaa tagtaatcaa ttacgggtc attagttcat agccatata	300
tggagttccg cggttacataa cttacggtaa atggcccgcc tggctgaccc cccaacgacc	360
ccccccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggacttcc	420
attgacgtca atgggtggag tatttacgtt aaactgccc cttggcagta catcaagtgt	480
atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt	540
atgcccagta catgaccta tgggacttcc ctacttggca gtacatctac gtattagtca	600
tcgctattac catggtgatg cggtttggc agtacatcaa tgggcgtgaa tagcggttg	660
actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttggc accacc	720
aaaatcaacg ggactttcca aaatgtcgta acaactccgc ccccattgac gcaaatggc	780
ggtagggcgtg tacggtggga ggtctatata agcagagctc tctggtaac tagagaaccc	840
actgcttact ggcttatcga aattaatacg actcaactata gggagaccga agctggctag	900
cgtttaact tgccaccatg gaccctgtgg tggtgctggg gctctgtctc tcctgttgc	960
ttctcccttc actctggaaa cagagctatg ggggaggggaa actggggcggaa agcggaggga	1020
cgggggggttc aggaacttca ggggggtgtga gcaaggcgca ggaggataac atggccatca	1080
tcaaggagtt catgcgttcc aaggtgcaca tggagggctc cgtgaacggc cacgagttcg	1140
agatcgaggg cgagggcgag ggcccccct acgagggc acagaccggc aagctgaagg	1200
tgaccaaggg tggcccccctg cccttcgcctt gggacatctt gtccttcag ttcatgtacg	1260
gctccaaggc ctacgtgaag caccggccg acatccccca ctacttgaag ctgtccttc	1320
ccgagggctt caagtgggag cgctgtatga acttcggagga cggcggcgtg gtgaccgtga	1380
cccaggactc ctccctgcag gacggcgagt tcatctacaa ggtgaagctg cgccggacca	1440
acttcccttc cgacggcccc gtaatgcaga agaagaccat gggctggag gcctccctcg	1500
agcggatgta ccccgaggac ggcgcctga agggcgagat caagcagagg ctgaagctga	1560
aggacggcgccg ccactacgac gctgaggatca agaccaccta caaggccaag aagcccgatc	1620
agctgccccg cgccctacaac gtcaacatca agttggacat cacctccac aacgaggact	1680
acaccatcgat ggaacagtac gaacgcgcg aggccgccta ctccaccggc ggcacatggacg	1740
agctgtacaa gtaactcgag atccgttacg gccgaaatca atcgctaatc actcaacttg	1800
ccatgtgtat gtggaaagcg tagaaaggct cgttgagctc attagctcg agcccgacta	1860
cgtttccac atactctgat gatccgctcg caaaggctcg tctgagctca ttagctccga	1920
gccccgaggtt ccggatcatt catggcaagt ccagcgcaat ctattacgaa aatcatccga	1980
cgtcgcgatg tctatgcggg aagcgttagaa aggctcgat gagctcatta gctccgagcc	2040
cgactacgtt tcccgatcgatc tctgatcatac cgctagcaaa ggctcggtga gctcattac	2100
tccgagcccg aggtaccggta tgattcatcg cgacgctcg gaaaatctca caaaatcag	2160
tcaaacgtcg ccgtgtgtgt gttaggaagcg tagaaaggct cgtctgagct cattagctcc	2220
gagccccact acgtttccata cacactctga cgatccgcta gcaaaggctc gttgagctca	2280
ttagctccga gccccgaggtt ccggatcgatc cacggcgacg ccgataatcc acataacttac	2340
aatcaggcaat tcttgcctatg tttatgtggg aagcgttagaa aggctcgatc agtctattag	2400
ctcccgagccg gactacgtttt cccacatact ctgatgtatcc gctagcaag gctcggttag	2460
ctcattagct ccgagccccgaa ggtaccggat cattcatggc aagtatcaag atcgaacggc	2520

-continued

gcaagatatt gtcacgtcgc gatgtctatg cgggaaggcgt agaaaggctc gttgagctca	2580
ttagctccga gcccgactac gtttcccga tagtctgatc atccgctagc aaaggctcgt	2640
ctgagctcat tagctccgag cccgaggtac cggatgatcc atcgcgacgt cctcgctaga	2700
tatgttaggt tcttaggcat ttgcgggtgt gtgtgttagga agcgttagaaa ggctcggtga	2760
gctcatttgc tccgagcccc actacgttcc ctacacactc tgacgatccg ctagcaaagg	2820
ctcgctcgag ctcattagct ccgagccccga ggttccggat cgttcacggc gaaaaggatcg	2880
tctgcaattc cgatttagacg tacacttgcc atgtgtatgt gggaaaggcgt aaaaaggctcg	2940
tctgagctca ttagctccga gcccgactac gtttcccaca tactctgtatc atccgctagc	3000
aaaggctcgat tgagctcatt agctccgagc ccgaggtacc ggatcattca tggcaagatc	3060
caagctactt cctccatacc tatccctcgc gcatgtctca tgcgggaagc gttagaaaggc	3120
tctgctcgagc tcattagctc cgagccccac tacgtttccc gcatagtctg atcatccgct	3180
agccaaaggct cgtttagatcc attagctccg agccccggat accggatgtat tcatcgccag	3240
atcataacgc aataccgtac actgtccaat cctcgccgtg tttgtgttagg aagcgttagaa	3300
aggctcgctc gagctcatta gctccgagcc cgactacgtt tcttacacac tctgacgatc	3360
cgttagccaaa ggctcggtga gctcattago tccgagcccc aggttccggta tcggttacgg	3420
cgaggataat caatccacat acatcacacc acaattcttgc ccatgtgtat gtgggaagcg	3480
tagaaaggct cgtctgagct cattagctcc gagccccact acgttccca catactctga	3540
tgtatccgcta gcaaaggctc gtctgagctc attagctccg agccccggat accggatcat	3600
tcatggcaag aattggtcgt tttttttggc ggccgctcga ctaaatcacc ggttatcttc	3660
ttgtccatct agacccatata aagatctttg tacaaggccg cgtttaaacc cgctgatctg	3720
cctcgactgt gccttctagt tgccagccat ctgttgggtt cccctcccccc gtgccttct	3780
tgaccctgga aagggtccac tccactgtc ctttccataa aaaatggatgg aattgcacgt	3840
cattgtctga gttagtgcgtt ttcttattctg ggggggtgggg gtggggggcag gacagcaagg	3900
ggggaggattt ggaagacaat agcaggcatg ctggggatgc ggtggggctt atgggttctg	3960
aggccggaaag aaccagctgg ggctcttaggg ggtatccccca cgcccccgtt agccggccat	4020
taagcgeggc ggggtgtggc gttacgcgc gctgtacccg tacacttgcg agcgeccctag	4080
cgcggccgtcc ttctcgatcc ttcccttctt ttctcgccac gttcggccgc ttcccccgct	4140
aagctctaaa tcgggggctc cctttaggtt tccgatttag tgctttacgg cacctcgacc	4200
ccaaaaaaact tgattagggt gatgggtcac gtagtgggcc atcgccctga tagacggttt	4260
ttcgcccttt gacgttggag tccacgttct ttaatagtgg actcttgcgtc caaactggaa	4320
caacactcaa ccctatctcg gtcttattctt ttgatttata agggattttt ccgatttccgg	4380
ccttattgggtt aaaaaatggat ctgatttaac aaaaatttaa cgcgaattaa ttctgtggaa	4440
tgtgtgtcg ttaggggtgt gaaaaggcccc aggttccca gcaggccagaa gtatgcggaa	4500
catgcacatc aatttagtcgtt caaccagggtg tggaaaggccc ccaggccccc cagcaggccag	4560
aagtatgcac agcacatc tcaatttagtc agcaaccata gttccggcccc taactccgccc	4620
catccccggcc ctaactccgc ccagttccgc ccattctccg ccccatggct gactaatttt	4680
tttttattttt gcaaggccgcg aggccgcctc tgccctgtatc ctatccaga agtagtggagg	4740
aggctttttt ggaggccctag gcttttgcaaa aagctcccc ggagcttgcataatccat	4800
cggatctgtat caagagacag gatgaggatc gtttcgcgtatc attgaacaag atggattgc	4860

-continued

cgcagggtct cggcccgctt gggtgagag gctattcggc tatgactggg cacaacagac	4920
aatcggtgc tctgatgccg ccgtgttccg gctgtcagcg cagggggccg cggttcttt	4980
tgtcaagacc gacctgtccg gtgcctgaa tgaactgcag gacgaggcag cgccgtatc	5040
gtggctggcc acgacggcg ttccttgcgc agctgtctc gacgttgtca ctgaagcggg	5100
aaggggactgg ctgttattgg gcgaagtgcg ggggcaggat ctccctgtcat ctcaccttgc	5160
tcctgcccag aaagtatcca tcatggctga tgcaatgcgg cggctgcata cgcttgcattc	5220
ggctacctgc ccattcgacc accaagcgaa acatcgcatc gagcggcac gtactcgat	5280
ggaaggccgtt cttgtcgatc aggtatgtct ggacgaagag catcaggggc tcgcgcacgc	5340
cgaactgttc gccaggctca aggccgcgcat gccccacggc gaggatctcg tcgtgaccca	5400
tggcgatgcc tgcttgcga atatcatgtt ggaaaatggc cgctttctg gattcatcga	5460
ctgtggccgg ctgggtgtgg cggaccgcta tcaggacata gcgttgcata cccgtatgtat	5520
tgctgaagag cttggcgccg aatgggctga ccgttgcctc gtgtttaacg gtatcgccgc	5580
tcccgcattcg cagcgcattcg ccttctatcg ccttcttgcag gagttttctt gaggccgact	5640
ctggggttcg aaatgaccga ccaagcgacg cccaacctgc catcagaga ttgcatttcc	5700
accggccgcct tctatgaaag gttgggcattc ggaatcgattt tccgggacgc cggctggatg	5760
atcctccagc gccccggatct catgctggag ttcttcgccc accccaaactt gtttattgca	5820
gcttataatg gttacaataa aagcaatgc atcacaaattt tcacaaataa agcatttttt	5880
tcactgcatt ctatgttggg tttgtccaaa ctcatcaatg tatcttatca tgcgtgtata	5940
ccgtcgacct ctagctagag cttggcgtaa tcatggctcat agctgttcc tgcgtgaaat	6000
tgttatccgc tcacaattcc acacaacata cgagccggaa gcataaagtg taaaggctgg	6060
gggtgcctaat gagtgagctta actcacatta attgcgttgc gctcaactgccc cgctttccag	6120
tcgggaaacc tgctgtgcctt gctgcattaa tgaatcgccc aacgcgcggg gagaggccgt	6180
ttgcgtatttggcgttcccg cgttccctcg ctcaactgact cgctgcgcgc ggtcggttccgg	6240
ctgcggcgag cggatcgc tcaactcaaag gcggtaatac gtttccac agaattcagg	6300
gataacgcac gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag	6360
ccgcgcgttgc tggcggtttt ccataggctc cgccccctcg acgagcatca caaaaatcga	6420
cgctcaagtc agagggtggcg aaacccgaca ggactataaa gataccaggc gtttccct	6480
ggaagctccc tcgtgcgttc tctgttcccg accctgcgcgc ttaccggata cctgtccgc	6540
tttctccctt cggggaaacgtt ggcgtttctt catacgctac gctgttaggt tctcgatgtcg	6600
gtgttaggtcg ttgcgtccaa gctgggtgtgt gtgcacgaaac ccccggttca gcccgcacgc	6660
tgcgccttat ccggtaacta tcgttgcgttgc tccaaacccgg taagacacga cttatcgcca	6720
ctggcagcagc ccactggtaa caggattgc agagcgaggt atgtaggccg tgctacagag	6780
ttcttgcgttgc ggtggcttaa ctacggctac actagaagaa cagtatttttgc tgcgtgcgt	6840
ctgcgtgaagc cagttacccctt cgaaaaaga gttggtagctt cttgtatccgg caaacaacc	6900
accgctggta gcgggttttt tggttgcag cagcagatta cgccgcagaaa aaaaggatct	6960
caagaagatc ctttgcgttgc tctgcgttc tgcgtgcgtc agtggaaacga aaactcacgt	7020
taaggggattt tggtcatgag attatcaaaa aggtatccca cctagatccct tttaaattaa	7080
aaatgaagtt taaaatcaat ctaaaatgtata tatgagtaaa ctgggtotga cagttaccaa	7140
tgcttaatca gtgaggcacc tatctcgatcg atctgtctat ttcgttgcata catatgtcc	7200
tgactcccccg tcgtgttagat aactacgata cggggaggcgt taccatctgg ccccaatgtct	7260

-continued

gcaatgatac cgcgagaccc acgctcacgg gctccagatt tatcagcaat aaaccagcca	7320
gccggagggg ccgagcgcag aagtggctt gcaactttat ccgcctccat ccagtctatt	7380
aattgttgcc gggaaagctag agtaagtagt tcgcccagtta atagtttgcg caacgttgtt	7440
gcattgtcta caggcatcgt ggtgtcacgc tcgtcggtt gtagtggcttc attcagctcc	7500
ggttcccaac gatcaaggcg agttacatga tccccatgt tgtgaaaaa agcggttagc	7560
tccttcggtc ctccgatcgt tgtcagaatg aagttggccg cagtgttatac actcatggtt	7620
atggcagcac tgcataattc tcttactgtc atgcacatccg taagatgctt ttctgtgact	7680
ggtgagacta caaccaagtc attctgagaa tagtgtatgc ggccggccgat ttgctctgc	7740
ccggcgtcaa tacggataa taccgcgcacatagcagaa cttaaaaagt gctcatcatt	7800
ggaaaacgtt ctccggggcg aaaactctca aggtatcttac cgctgttgag atccagttcg	7860
atgtaaacca ctcgtgcacc caactgatct tcagcatctt ttactttcac cagcgttct	7920
gggtgagcaa aaacaggaag gcaaaatgcg gcaaaaaagg gaataaggc gacacggaaa	7980
tgttgaatac tcataactctt ccttttcaa tattattgaa gcatttatca gggttattgt	8040
ctcatgagcg gatacatatt tgaatgtatt tagaaaaata aacaaatagg ggttccgcgc	8100
acatttcccc gaaaaagtgcc acctgacgto	8130

<210> SEQ ID NO 77

<211> LENGTH: 8122

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: pUBC- (mNeonGreen)4-tDeg

<400> SEQUENCE: 77

gacggatcg gagatctccc gatcccattt ggtgcactct cagtaaatc tgctctgtat	60
ccgcatagtt aagccagttt ctgtccctgt ctgtgtgtt ggaggctcgat gatgtgtcg	120
cggccaaat ttaagtcaca acaaggcaag gcttgaccga caattgtatc aagaatctgc	180
ttagggtagt gctttttcgat ctgtttcgat atgtacgggc cagatatacg cgtggcctc	240
cggccgggtt ttggggcctt cccggggcgcccccttcac cggccggccgat gctggccacgt	300
cagacgaagg ggcggccggat cgttctgtat cttccggccgat gacgctcagg acageggccc	360
gctgctata agactcgccat ttagaaccat agtacgtatc gatggacatt ttaggacggg	420
acttgggtatc ctcttagggca ctgtttttctt ttccagagat cggaaacaggc gaggaaaat	480
atgtcccttc cggcgattttt gctggggat ctccgtgggg cggtaacgc cgtatgtttat	540
ataaggacgc gcccgggtgt gcaacagctat ttccgtcgat gcccggatatt ggttccgggt	600
tcttgggtt ggttccgtgtt gatgttcaat tggaaagtttccatccatgtt gagcaaggc	660
gaggaggata acatggccctc tctcccgatc acacatgtat tacacatctt tggctccatc	720
aacgggtgtgg actttgacat ggtgggtcgat ggcacccggca atccaaatgatgatgtt	780
gagttaaacc tgaagttccac caagggtgac ctccagtttccatccatgtt gagcaaggc	840
catatcggtt atggcttcca tcagttacgtt ccctaccctgtt acgggtatgc gctttccatc	900
ggccggccatgg tagatggctc cggatccaa gtccatcgatc caatgcgtt tgaagatgg	960
gcctccctta ctgttaacta ccgttacacc tacggggaa gccacatcaa aggagaggcc	1020
caggtgaagg ggactgggtttt ccctgtgtat ggttccgtgtat gatccaaactc gctgaccgt	1080
gcggactggat gcaagggtcgaa gaagacttac cccaaacgaca aaaccatcat cagttttt	1140

US 12,391,948 B2

249**250**

-continued

aagtggagtt acaccactgg aaatggcaag cgctaccgga gcactgcgcg gaccacctac	1200
acctttgccaa agccaaatggc ggcttaactat ctgaagaacc agccgatgtt cgttgtccgt	1260
aagacggagc tcaaggactc caagaccgag ctcaacttca aggagtggca aaaggccccc	1320
accgatgttga tggggatggc cgagctgtac aagggtggac atatgggcac agggccaca	1380
ggcggttccg gggggattttc caaaggagaa gaagacaata tggcatcaact ccccgcaacc	1440
cacgagttgc atattttccg ttcaattaat ggagtagatt tcgatatgg tggccaggaa	1500
acaggaaacc caaacgacgg atatgaagag cttaatctca aaagtaccaa aggccatctg	1560
caatttctc cgtggatact cgtccacac attggatacg gatttcacca atatctccg	1620
tatccggatg gaatgtcccc ctttcaagca gcaatgggtt acggggatgg ttatcaggtt	1680
cacagaacca tgcagttcgaa ggacggggct tctctgaccc taaattatag gtatacttat	1740
gaaggctcac atattaaggc cgaaggcacag gttaaaggaa cggggttcc tgccggatggc	1800
cccgcatgaa ctaattctct gacagccgca gattgggttc gtcacaaaaa gacataccg	1860
aatgataaga ctataatctc aacattcaaa tggtcctata cgacaggccaa cgggaaacga	1920
tatagatcca cggctcgaaac aacttacaca ttgcctaaac ctatggccgc caattacctc	1980
aaaaatcagc ccattgtatgt gtttagggaa accgaatttgaa agcattctaa aacggaaactt	2040
aattttaagg aatggcagaa ggctttcaca gacgtaatgg ggatggatgtt actctataaa	2100
tcaggtctcg agtcctcagg gggAACGGGT gggtccggag gagtttagtta aggtgttgc	2160
gacaatatgg caagtttgcg tgcactcac gacccatca tctttgggtc tataaatggc	2220
gttgacttcg atatgggtt ccaaggtaact ggcaacccca atgacggttt cggggatgg	2280
aatctcaagt ccacaaaagg tgcatttcgat ttcagccctt ggattctcgat acctcatatt	2340
ggatataggc ttcaccagta ccttcatacc ccaagcggtt tgcacccctt tcaagctgcg	2400
atgggtggatg gttccggcta tcagggtccac cgaacgtgc aattcgagga cggggccagc	2460
ctcaccgtta attataggta cacctatgg ggaagtcaca taaaggggaa agcccaagtg	2520
aaaggaacag gattcccgac tgcattggca gtaatgcacg actccttgcg acgggttgc	2580
tgggttagaa gaaaaaagac gtatcctaat gacaagaccat tcattagcac tttcaatgg	2640
agttataccca caggaaacgg caaacgggtt agaaggactg ctgcacccatc ctacacttc	2700
gcaaaaggcgaa tggctcgaaa ctatggtaag aatcggccca tgcacgtttt tcgaaaaacg	2760
gaacttaagc acagtaagac tgaacttaat ttcaaggatgg ggcagaaggc gttcacggat	2820
gtcatgggtt tggatggact gtataaggga gggcttggca ctggggggccac tgcccgac	2880
ggatccgggtt ggggtgttgcgaa caagggcgag gaggataaca tggctctctt cccagcgaca	2940
catgagttac acatcttgg ctccatcaac ggtgtggact ttgcacatggt gggtcaggc	3000
accggcaatc caaatgtatgg ttatgaggag ttaaacctgaa agtccaccaa gggtgaccc	3060
cagttctccc cttggattctt ggtccctcat atcgggtatgtt gttccatca gtacccccc	3120
taccctgacg ggtatgtcgcc tttccaggcc gccatggtag atggctccgg ataccaagtc	3180
catcgccacaa tgcagtttgcgaa agatgggtcc tcccttactgtttaactaccc ctacaccc	3240
gagggaagcc acatcaaagg agaggcccgatgttggggaa ctgggttcccg tgctgacgg	3300
cctgtatgttgc ccaactcgctt gaccgctgcg gactgggttca ggtcgaagaa gacttaccc	3360
aacgacaaaaa ccatcatcgat tacctttaag tggagttaca ccactggaaa tggcaaggc	3420
taccggagca ctgcggggac cacctacacc tttgccaagc caatggccgc taactatctg	3480
aagaaccaggc cgtatgtacgt gttccgttgcg acggagctca agcactccaa gaccgagctc	3540

-continued

aacttcaagg agtgccaaaa ggccttacc gatgtatgg gcatggacga gctgtacaag 3600
ggcggaaagat ccgggtgggg ttctggtct cgtccccgtg gtactcggt taaaggtcg 3660
cgtagtcgat gcccgggta atcttagaggg cccgttaaaa cccgcgtatc agcctcgact 3720
gtgccttcta gttgccagcc atctgttgg tgeccctccc cegtgccctc cttgaccctg 3780
gaaagggtgc actcccaactg tccttcctata ataaaatgtg gaaattgtat cgcatgtct 3840
gagtaggtgt cattctatc tgggggggtgg ggggtggggc aggacaccaa gggggaggat 3900
tggaaagaca atagcaggca tgctggggat ggggtgggt ctatgggttc tgaggcgaa 3960
agaaccagct ggggtcttag ggggtatccc caegcgccct gtageggcgc attaagcg 4020
gggggtgtgg tgggtacgca cagcgatggc gctacacttg ccagcgccct agcgccccgt 4080
ccttcgttctt ctttcgttcc acgttcgtcc gttttcccg tcaagctcta 4140
aatcgggggc tccctttagg gttccgattt agtgcttac ggcacctcgaa ccccaaaaaaa 4200
cttgattagg gtgtatgggtc acgttagtggg ccacgcggct gatagacggt ttttcgcct 4260
ttgacgttgg agtccacgtt cttaatagt ggactcttgt tccaaacttgg aacaacactc 4320
aacccatctt cggcttatc ttttgattta taagggatt tgccgatttc ggcctattgg 4380
ttaaaaaatg agctgattta acaaaaatatt aacgcgaaatt aattctgtgg aatgtgtgtc 4440
agttagggttgg tggaaagtcc ccaggctccc cagcaggcag aagtatgcaaa agcatgcata 4500
tcaatttagtc agcaaccagg tggaaagt ccccgatgtc cccagcaggc agaagtatgc 4560
aaagcatgca tctcaatttag tcagcaacca tagtcccgcc cctaaactccg cccatcccg 4620
ccctaaactcc gcccaggttcc gcccatttc cggccatgg ctgactaatt ttttttattt 4680
atgcagaggc cgaggccgccc tctgcctctg agtattcca gaagtagtga ggaggctttt 4740
ttggaggccat aggcttttgc aaaaagctcc cgggagcttgc tatatccatt ttcggatctg 4800
atcaagagac aggttaggaa tcgtttcgtca tgattgaaca agatggattt caccgagg 4860
ctccggccgc ttgggtggag aggctatttcg gctatgactg ggcacaacag acaatcggt 4920
gctctgtatc cggcggttcc cggctgtcc cgcaggggcg cccgggtttt tttgtcaaga 4980
ccgacactgtc cgggtccctg aatgaacttc aggacgaggc agcgcggctt tcgtggctgg 5040
ccacgacggg cgttccctgc gcagctgtgc tcgcacgttgc cactgaagcg ggaaggggact 5100
ggctgtatc gggcgaagtgc cgggggcagg atctcgttc atctcacctt gctctggccg 5160
agaaagtatc catcatggct gatgaatgc ggccgtcgatc tacgcttgat cccgttacct 5220
gcccatcgat ccaccaagcg aaacatcgca tcgagcgagc acgtactcgat atggaaaggccg 5280
gttttgcgtca tcaggatgtat ctggacgaaag agcatcaggc gtcgcggccaa gccgaactgt 5340
tcgcccaggat caaggcgccgc atgcccgtacg gcggaggatct cgtcgatccatcgtcgat 5400
cctgttgcgtca gaatatcatg gtggaaaatg gccgttttc tggatttcgtac gactgtggcc 5460
ggctgggtgtt ggccggaccgc tatcaggaca tagcgttggc taccctgtat attgtgtaa 5520
agtttggccgc cgaatgggttgc gaccgttcc tcgtgttttgc cggatcgcc gctcccgatt 5580
cgacgacgtat cggcttctat cggcttcttgc acgagttttt ctgagcgggc ctctgggtt 5640
cgaaatgacc gaccaaggcga cggccaaactt gccatcacga gatttcgttcc ccacccggc 5700
cttctatgaa aggttgggttgc tcggtatcgat ttccggccat gccggctggat tgatcttca 5760
gcccggggat ctcgtatcgat agtttccgtcc ccaacccaaatc ttgttttattt cagcttataa 5820
tggtttacaaa taaagcaata gcatcacaaa tttcacaaaatc aaagcattttt tttactgtca 5880

-continued

ttcttagtgtt ggtttgtcca aactcatcaa tgtatcttat catgtctgta taccgtcgac	5940
ctcttagtag agcttggcgt aatcatggtc atagctgtt cctgtgtgaa attgttatcc	6000
gctcacaatt ccacacaaca tacgagccgg aagcataaag tgtaaagcct ggggtgccta	6060
atgagtgagc taactcacat taattgcgtt gcgcactg cccgccttcc agtcgggaaa	6120
cctgtcgctc cagctgcatt aatgaatcg ccaacgcgcg gggagaggcg gtttgcgtat	6180
tggcgctct tcgcgttcct cgctactga ctgcgtgcg tcggtegttc ggctgcgcg	6240
agcggtatca gtcactcaa aggccgtaat acggttatcc acagaatcag gggataacgc	6300
aggaaagaac atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa aggccgcgtt	6360
gctggcggtt ttccataggc tccgcctccc tgacgagcat cacaaaaatc gacgtcaag	6420
tcaaggggtgg cgaaacccga caggactata aagataccag gcgttcccc ctggaaagtc	6480
cctcggtgcgc tctccgttcc cgaccctgcg gcttaccgga tacctgtccg ccttctccc	6540
ttcgggaaaggc gtggcgctt ctcatacgct acgctgttagg tatctcagtt cggtgttagt	6600
cgttcgtcc aagctgggct gtgtgcacga accccccgtt cagccccgacc gctgcgcctt	6660
atccggtaac tatcgcttt agtccaaaccg ggtaaagacac gacttacgcg cactggcagc	6720
agccactggt aacaggatta gcagagcgg gtagttaggc ggtgctacag agttttgaa	6780
gtgggtggcct aactacggct acactagaag aacagtattt ggtatctgcg ctctgctgaa	6840
gccagttacc ttccggaaaaa gagttggtag ctcttgatcc ggcaaacaaaa ccaccgctgg	6900
tagcgggtttt tttgtttgca agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga	6960
tcctttgatc ttttctacgg ggtctgacgc tcagtggAAC gaaaactcac gttaaaggat	7020
tttggtcatg agattatcaa aaaggatctt cacctagatc cttttaaatt aaaaatgaaag	7080
ttttaaatca atctaaagta tatatgagta aacttggctc gacagttaacc aatgttaat	7140
cagtggggca cctatctcag cgatctgtct atttcgttca tccatagttt cctgactccc	7200
cgtcggttag ataactacga tacggggaggc cttaccatct ggccccagtg ctgcaatgat	7260
accggcggac ccacgctcac cggctccaga tttatcagca ataaaccacg cagccggaaag	7320
ggccggagcgc agaagtggtc ctgcaacttt atccgcctcc atccaggctta ttaatttttg	7380
ccggggaaagct agagtaagta gttcgccagt taatagtttgc gcaacgttg ttgccattgc	7440
tacaggcatacgtt gttgtgtcac gtcgtcggtt tggatggct tcattcagct ccgggtccca	7500
acgatcaagg ctagttacat gatccccat gttgtcAAA aaagcggtaa gtccttcgg	7560
tcctccgatc gttgtcagaa gtaagttggc cgcagtgttca tcactcatgg ttatggcagc	7620
actgcataat tctcttactg tcataccatc cgtaagatgc ttttctgtga ctggtagta	7680
ctcaaccaag tcattctgag aatagtgtat gccggcggacc agttgtctt gccccggcgc	7740
aatacgggat aataccggcgc cacatagcag aactttaaaaa gtgctcatca ttggaaaacg	7800
tttttcgggg cggaaactct caaggatctt accgctgtttt agatccagg ctatgttaacc	7860
cactcggtca cccaaactgtat cttcagcatac ttttactttc accagcggtt ctgggtgagc	7920
aaaaacagga aggccaaaatg ccgcAAAAAAA gggataaagg ggcacacgg aatgttaat	7980
actcatactc ttcccttttc aatattattt aagcatttat cagggttatt gtctcatgag	8040
cggatacata tttagaatgta tttagaaaaa taaacaaataa ggggttccgc gcacattcc	8100
ccgaaaaatgc ccacactgacg tc	8122

-continued

<212> TYPE: DNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: pAV-U6+27-Tornado-F30-Pepper (TAR Variant-2)

<400> SEQUENCE: 78

gccggatcca	aggctggca	ggaagagggc	ctatttccca	tgattccttc	atatttgcata	60
atacgataca	aggctgttag	agagataatt	agaattaatt	tgactgtaaa	cacaaagata	120
ttagtacaaa	atacgtgacg	tagaaagtaa	taatttcttgc	ggtagttgc	agttttaaaa	180
ttatgtttta	aatggacta	tcatatgctt	accgttaactt	gaaagtattt	cgatttcttgc	240
gttttatata	tcttgtggaa	aggacgaaac	accgtgctgc	cttcggcagc	acatataacta	300
gtcgacgggc	cgcactcgcc	ggtcccaagc	ccggataaaaa	tgggaggggg	cgggaaaccg	360
cctaaccatg	cggagtgcgg	ccgcttgcaca	tgtgtatgttgc	ggacgcgttgc	ccacgtttcc	420
cacatactct	gatgatccgc	tagcaaaggc	tcgtttagtgc	cattagctcc	gagcccgagg	480
taccggatca	ttcatggcaa	gccccgcgg	tcggcgtggaa	ctgttagaaca	ctgccaatgc	540
cggtcccaag	cccgataaaa	agtggagggt	acagtccacg	ctctagagcg	gacttcggtc	600
cgttttttac	taggacctgc	aggcatgca	gcttgacgtc	ggttaccgtat	atccatatgg	660
cgcaccgcata	gatctcgagc	cgaggactag	taacttgcgtt	attgcagctt	ataatggta	720
caaataaaagc	aatagcatca	caaatttcac	aaataaaagc	ttttttcac	tgcattcttag	780
ttgtggtttg	tccaaactca	tcaatgtatc	ttatcatgtc	ttacgttagat	aagtagcatg	840
gggggttaat	cattaaactac	aaggaacccc	tagtgtatggaa	gttggccact	ccctctctgc	900
gctcgctc	gctcaactgag	gccggggcgcac	caaagggtgc	ccgacgcoccg	ggctttgccc	960
ggggggcctc	agtgagcgcag	cgagcgcgcac	gagagggaggt	ggccaaagat	ctctggcgta	1020
atagcgaaga	ggcccgacc	gategcccctt	ccaaacagtt	gctcgacgtc	aatggctaat	1080
ggaaatttgt	aaacgttaat	attttgcgtt	atatttgcgtt	aaattcgcgt	taaattttgcgtt	1140
ttaaatcgc	tcatttttta	accaataggc	cgaatcggc	aaaatccctt	ataaaatcaaa	1200
agaatagacc	gagatagggt	tgagtgttgt	tccagtttgcgtt	aacaagagtc	cactattaaa	1260
gaacgtggac	tccaaacgtca	aaggcgaaa	aaccgtctat	caggcgcgtat	gcccaactacg	1320
tgaaccatca	ccctaatacaa	gtttttgggg	gtcgaggatgc	cgtaaagcac	taaatcggaa	1380
ccctaaagggt	atgccccgat	ttagagcttgc	acggggaaag	ccggcgaacgc	tggcgagaaa	1440
ggaaaggaaag	aaagcgaaag	gagccccgcgc	tagggcgctgc	gcaagtgttag	cggtcagcgt	1500
gcgcgttaacc	accacaccccg	ccgcgcctaa	tgcgcgcata	cagggcgcgt	caggtggcac	1560
tttcggggaa	aatgtgcgcg	gaacccttat	ttgtttattt	ttctaaatac	attcaaataat	1620
gtatccgctc	atgagacaat	aaccctgata	aatgcttca	taatattgaa	aaaggaagag	1680
ttatgtgttatt	caacatttcc	gtgtgcgcctt	tattcccttt	tttgcggcat	tttgccttcc	1740
ttttttgtct	cacccagaaa	cgctgggtgaa	agtaaaagat	gctgaagatc	agttgggtgc	1800
acgagtggtt	tacatcgaaac	tggatctcaa	cagcggtaaag	atcccttgcata	gttttcggcc	1860
cgaagaacgt	tttccaaatga	tgagoacttt	taaagttctgc	ctatgtggcg	cggttattatc	1920
ccgtatttgc	gccggggcaag	agcaactcg	tcgcccata	cactattctc	agaatgactt	1980
ggttggttgc	tcaccaggta	cagaaaagca	tcttacggat	ggcatgcacag	taagagaatt	2040
atgcagtgtc	gccataacca	tgagtgtataa	cactgcggcc	aacttacttc	tgacaacgt	2100
cggaggaccg	aaggagctaa	ccgtttttt	gcacaacatg	ggggatcatg	taactcgcc	2160

-continued

tgatcggtgg gaaccggagc tgaatgaago cataccaaac gacgagcggtg acaccacat	2220
gcctgttagca atggcaacaa cggtgcgcaa actattaact ggccaactac ttactctagc	2280
ttccccggcaa caattaatag actggatgga ggccgataaa gttgcaggac cacttctcg	2340
ctcgccctt ccggctggct ggtttattgc tgataaatct ggagccggtg agcgtgggtc	2400
tccgcgtatc attgcagcac tggggccaga tggtaagccc tcccgtatcg tagttatcta	2460
cacgacgggg agtcaggcaa ctatggatga acgaaataga cagategctg agataggtgc	2520
ctcactgatt aagcatttgt aactgtcaga ccaagttac tcataatatac ttttagatga	2580
tttaaaaactt catttttaat ttAAAAGGAT cttagtgaag atccttttg ataatctcat	2640
gacccaaatc ccttaacgtg agtttcgtt ccactgagcg tcagaccccg tagaaaagat	2700
caaaggatct tcttgagatc cttttttct gcgcgtaaatc tgctgtttgc aaacaaaaaa	2760
accaccgcta ccagegggtgg tttgtttgcc ggatcaagag ctaccaactc ttttccgaa	2820
ggtaactggc ttcagcagag cgcaagatacc aaatactgtc cttctagtgt agccgtagtt	2880
aggccaccac ttcaagaact ctgttagcacc gcctacatac ctgcgtctgc taatcctgtt	2940
accagtggtc gctgccagtg gcgataagtc gtgtcttacc gggttggact caagacgata	3000
gttaccggat aaggcgcagc ggtcgggctg aacgggggggt tcgtgcaaca cagccagtt	3060
ggagcgaacg acctacaccg aactgagata cctacagcggt gaggattgag aaagcgcac	3120
gcttcccggaa gggagaaagg cggacaggtt tccggtaagc ggcagggtcg gaacaggaga	3180
gcgcacgagg gagttccag gggaaacgc ctggtatctt tatagtccctg tcgggtttcg	3240
ccacctctga cttgagcgtc gattttgcgt atgctgtca gggggggcgga gcctatggaa	3300
aaacgcccagc aacgcggcct ttttacgggtt cctggccttt tgctggcctt ttgctcacat	3360
gttctttctt gcgttatccc ctgattctgt ggataaccgtt attaccgcct ttgagtggac	3420
tgataaccgtc cgccgcagcc gaacgaccga gcgcagcgag tcagtggact aggaagcgga	3480
agagcgcggca atacgcaaac cgcctctccc cgccgttgg ccgattcatt aatgcagaga	3540
tctttggcca ctccctctct gcgcgtcg tcgtcactg aggccggggcg accaaaggtc	3600
ggccgacgcc cgggcttgc cggggggggcc tcagtggact agcgagcgcc cagagaggga	3660
gtggccaaact ccatcactag gggttctgg aggggtggag tcgtgacgtg aattacgtca	3720
taggggtttagg gaggtcctgg atcgatccag acatgataag atacattgtat gagttggac	3780
aaaccacaac tagaatgcag tgaaaaaaat gctttatgg tgaatattgt gatgtattt	3840
ctttatgttta aaccattata agtgcataa aacaaggtaa caacaacaat tgcattcatt	3900
ttatgtttca ggttcagggg gaggtgtggg aggtttttta aagcaagtaa aacctctaca	3960
aatgtggatggtt ggcgtgattat gatctctagt caaggcacta tacatcaaattt attccttatt	4020
aaccccttta caaattaaaa agctaaaggta acacaatttt tgagcatagt tattaatagc	4080
agacactcta tgcctgtgtg gagtaagaaa aaacagtatg ttatgttatt aactgtttatg	4140
cctacttata aagggtacag aatattttc cataattttc ttgtatagca gtgcagctt	4200
ttcccttgcgt gttaaatag caaaagcaago aaggttctta ttactaaaca cagcatgact	4260
caaaaaactt agcaattctg aaggaaaggcc cttggggtct tctaccttcc tcttttttt	4320
tggaggagta gaatgttgag agtcagcgat agcctcatca tcactagatg gcatttttc	4380
tgagcaaaac aggtttccctt cattaaaggcc attccaccac tgctccatt catcgatcc	4440
ataggttggaa atctaaaata cacaacaat tagaatcagt agtttaacac attatacact	4500
taaaaaattttt atatccatct tagagcttta aatctctgtt ggtgtttgtt ccaattatgt	4560

US 12,391,948 B2

259

-continued

260

cacaccacag aagttaagggtt ccttcacaaa gatccgggac caaagcggcc atcgtgcctc 4620
 cccactcctg cagttcgaaaa gcatggatgc gcggatagcc gctgctgggt tcctggatgc 4680
 cgacggattt gcactgccgg tagaactccg cgaggtcgctc cagcctcagg cagcagctga 4740
 accaactcgc gaggggatcg agccccgggt gggcgaagaa ctccagcatg agatccccgc 4800
 gctggaggat catccagccg gcgtcccgga aaacgattcc gaagcccaac ctttcataga 4860
 aggccgggggt ggaatcgaaa tctcgatgc gtaggttggg cgtcgcttgg tcggatcttt 4920
 cgaaccccaag agtcccgtc agaagaactc gtcaagaagg cgatagaagg cgatgcgctg 4980
 cgaatcggga gcggcgatac cgtaaagcac gaggaaagcg tcagccatt cgccgccaag 5040
 ctcttcagca atatcacggg tagccaacgc tatgtcctga tagcggtccg ccacacccag 5100
 cccggccacag tcgatgaatc cagaaaageg gccatttcc accatgatata tcggcaagca 5160
 ggcatcgcca tgggtcacga cgagatccctc gccgtcgaaa atgcgcgcct tgagcgtggc 5220
 gaacagttcg gtcggcgca gcccctgtat ctcttgccca gatcatccgt atcgacaaga 5280
 cccggcttcca tccgagttacg tgctcgatcg atgcgatgtt cgcttgggtt tcgaatgggc 5340
 aggttageccgg atcaagcgta tgcagccgc gcattgcattt agccatgtat gatactttct 5400
 cggcaggagc aaggtgagat gacaggagat cctgccccgg cacttcgccc aatagcagcc 5460
 agtcccttcc cgcttcaatgc acaacgtcgca gcacagctgc gcaaggaaacg cccgtcggtt 5520
 ccagccacga tagcccgctt gcctcgatc gcagttcattt cagggcaccg gacaggtcggtt 5580
 tcttgacaaa aagaaccggg cggccctcgat ctgacagccg gaacacggcg gcatcagagc 5640
 agccgattgt ctgttgtgcc cagtcatatgc cgaatagccct ctccacccaa gcccggag 5700
 aacctgcgtt caatccatct tggtaatca tgcgaaacga ttctcatctt gtctttgtat 5760
 cagatcttgc tccctgcgc catcagatcc ttggcgaa gaaagccatc cagttactt 5820
 tgcagggtt cccaaaccta ccagggcgcc cccagctgg caatccggc tcgcttgcgtt 5880
 tccataaaac cggccagtc agctatcgcc atgttaagccc actgcaagct acctgttttc 5940
 tctttgegt tgcgtttcc cttgtccaga tagcccaatgc gtcgatcatc atccgggttc 6000
 agcaccgtt ctgcggactg gcttctaaatg tggccgtt cctttagcag cccttgcgc 6060
 ctgagtgott gggcggcgatgtt gaaaaagectt agggccatccaa aaaaggctcc 6120
 tcactacttc tggaaatagct cagaggccga ggccggctcg gcctctgcattt aaataaaaa 6180
 aattagttagt ccatggggcg gagaatgggc ggaactgggg ggagtttaggg gcccggatggg 6240
 cggagttagg ggccggacta tgggtgctga ctaattgaga tgcgtgtttt gcataacttct 6300
 gcctgctggg gaggctgggg actttccaca cctgggttgc gactaatttgc gatgcgtgtt 6360
 ttgcatactt ctgcctgtt gggagccctgg ggactttccaa caccctaactt gacacacattt 6420
 ccaca 6425

<210> SEQ ID NO 79
 <211> LENGTH: 6426
 <212> TYPE: DNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: pAV-U6+27-Tornado-F30-TAR Variant-1

<400> SEQUENCE: 79

gccggatcca aggtcgccca ggaagagggc ctatccatccatc tgattccttc atatttgcattt 60
 atacgataca aggctgttagt agagataattt agaattaattt tgactgtaaa cacaaagata 120

-continued

ttatgtttta aaatggacta tcatacgctt accgtaactt gaaagtattt cgattcttg
gcctttata tcttgtggaa aggacgaaac accgtgcgtc ctccggcagc acatatacta
gtcgacgggc cgcaactcgcc ggtcccaago ccggataaaa tgggggggg cggggaaaccg
cctaaccatg ccgagtgccg cccgttgcctatgtg ggacgcgttg ccacgttcc
cacatactct gatgatccgc tagcaaaggc tcgtctgagc tcattagctc cgagcccgag
gtaccggatc attcatggca agcggccgcg gtccggcgtgg actgtagaac actgccaatg
ccggtcccaa gccccgataa aagtgggggg tacagttcac gctctagagc ggacttcgg
ccgctttta ctaggacctg caggcatgca agcttgacgt cggttaccga tatccatatg
gcgaccgcattt cgatctcgag ccggacta gtaacttgg tattgcagct tataatgggt
acaataaaag caatagcatc acaaatttca caaataaaagc attttttca ctgcattctat
gttgtggttt gtccaaactc atcaatgtat cttatcatgt cttacgtaga taatgtat
ggcgggttaa tcattaacta caaggaaccc ctatgtatgg agttggccac tccctctctg
cgcgctcgct cgctactga ggccggcga ccaaaggctg cccgacgccc gggcttgc
cgccggccct cagttagcga gcgagcgcgc agagagggag tggccaaaga tctctggcgt
aatagcgaag agggccgcac cgatgcctt tcccaacagt tgcgacgctt gaatggctaa
tggaaattt taaacgttaa tattttgtta atattttgtt aaaattcgcg ttaaattttt
gttaaatcag ctcatatccaa aaccaatagg ccgaaatcgg caaaatccct tataaatcaa
aagaatagac cgagataggg tttagtgg tttccagtttgg gaacaagagt ccactattaa
agaacgtgaa ctccaaacgtc aaaggccgaa aaaccgtctt tcaaggccgat ggccactac
gtgaaccatc accctaataatca agtttttgg ggtcgaggtg ccgtaaagca ctaaatcgga
accctaaagg gatgccccga ttttagagctt gacggggaaa gcccggcggac gttggcgagaa
aggaagggaa gaaagcggaaa ggagcggcggctt ccggccgtt ggcaagtgtt ggggtcacgc
tgcgcgttaac caccacaccc gcccggctt atgcgcgtt acaggccgcg tcagggtggca
cttttccgggg aaatgtgcgc ggaaccccta tttgtttatt tttctaaata cattcaaata
tgtatccgct catgagacaa taaccctgtt aatgttttca ataatattgtt aaaagggaaaa
gtatgtat tcaacatttc cgtgtcgcccc ttatccctt ttttgcggca ttttgccttc
ctgttttgc tcacccagaa acgctggtga aagtaaaaaga tgctgaagat cagttgggt
cacgagtggg ttacatcgaa ctggatctca acagcggtaa gatccttgg agtttgcct
ccgaagaacg ttttccaaatg atgagactt ttaaaggctt gctatgtggc gccgtattat
cccgatgttga cggccggccaa gagcaactcg gtccggccat acactattct cagaatgact
tgggtggatc ctcaccagtc acagaaaaaggc atcttacggta tggcatgaca gtaagagaat
tatgcagtgc tgccataacc atgagtgata acactgcggc caacttactt ctgacaacgaa
tcggaggacc gaaggagctt accgctttt tgcacaacat gggggatcat gtaactcgcc
tttgcgttgc ggaacccggag ctgatgttgc ccataccaaa cgacgacgtt gacaccacgaa
tgcgttgc aatggcaaca acgttgcgcgca aactattaaac tggcgacta cttactcttag
cttccggccaa acaattaata gactggatgg agggggataa agttgcgggaa ccacttctgc
gctcgccct tccggctggc tgggttattt ctgataaaatc tggagccgggt gaggcgtgggt
ctcgccgtat cattgcagca ctggggccag atggtaagcc ctcccgatc gtatgtatct
acacgacgggg gaggcggccaa actatggatg aacgaaaatg acagatcgct gagatggatg
2400
2380
2360
2340
2320
2300
2280
2260
2240
2220
2200
2180
2160
2140
2120
2100
2080
2060
2040
2020
2000
1980
1960
1940
1920
1900
1880
1860
1840
1820
1800
1780
1760
1740
1720
1700
1680
1660
1640
1620
1600
1580
1560
1540
1520
1500
1480
1460
1440
1420
1400
1380
1360
1340
1320
1300
1280
1260
1240
1220
1200
1180
1160
1140
1120
1100
1080
1060
1040
1020
1000
980
960
940
920
900
880
860
840
820
800
780
760
740
720
700
680
660
640
620
600
580
560
540
520
500
480
460
440
420
400
380
360
340
320
300
280
260
240
220
200
180
160
140
120
100
80
60
40
20
0

-continued

cctcactgat taagcattgg taactgtcg accaagttt ctcataata ctttagattg	2580
attnaaaact tcattttaa tttaaaagga tctaggtgaa gatcctttt gataatctca	2640
tgacccaaat cccttaacgt gagtttcgt tccactgagc gtcagacccc gtagaaaaga	2700
tcaaaggatc ttcttgagat ccttttttc tgcgcgtaat ctgctgcttg caaacaaaaa	2760
aaccaccgct accagcggtg gtttgggtc cggtcaaga gctaccaact cttttccga	2820
aggttaactgg cttcagcaga ggcgacatac caaatactgt ctttctagtg tagccgtgt	2880
taggccacca cttcaagaac tctgttagcac cgcttacata cttcgctctg ctaatccgt	2940
taccagtggc tgctgcagt ggcgataagt cgtgtcttac cgggttggac tcaagacgat	3000
agttaccgga taaggcgcag cggtcggtt gAACGGGGGGG ttctgtcaac acagccagct	3060
tggagcgaac gacctacacc gaactgagat acctacagcg tgagcattga gaaagcgcca	3120
cgcttccoga agggagaaag gcggacaggt atccggtaag cggcagggtc ggaacaggag	3180
agcgcacgag ggagcttcca gggggaaacg cctggtatct ttatagtctt gtctgggttc	3240
gccacccctcg acttgagcgt cgattttgt gatgctcgcc aggggggggg agcctatgga	3300
aaaacgcggc caacgcggcc tttttacggt tcttgcctt ttgtggctt tttgtcaca	3360
tgttcttcc tgcgttatcc cctgattctg tggataaccg tattacogcc tttgagttag	3420
ctgataccgc tcgcccgcgc cgaacgacccg agegcagegca gtcagtggc gaggaagcgg	3480
aagagcggcc aatacgcaaa ccgcctctcc ccgcgcgttg gccgattcat taatgcagag	3540
atctttggcc actccctctc tgcgcgctcg ctgcgtact gaggcggggc gaccaaggt	3600
cgccccgacgc cggggcttg cccggggggc ctcaagtggc gagcggcgcgc gcagagaggg	3660
agtggccaaac tccatacta ggggttcctg gagggttggg gtcgtgacgt gaattacgtc	3720
atagggttag ggagggtcctg gategatcca gacatgataa gatacattga tgagtttgg	3780
caaaccacaa ctagaatgca gtgaaaaaaaaa tgctttattt gtgaaatttg tgatgtatt	3840
gtcttatttgc taaccattat aagctgcaat aaacaagttt acaacaacaa ttgcattcat	3900
tttatgtttc aggttcaggg ggagggttgg gagggtttt aaagcaagta aaacctctac	3960
aaatgtggta tggctgatta tgatctctag tcaaggcact atacatcaaa tattccttat	4020
taacccctt acaaattaaa aagctaaagg tacacaattt ttgagcatag ttattatag	4080
cagacactct atgcctgtgt ggagtaagaa aaaacagtat gttatgatata taactgttat	4140
gcctacttat aaaggttaca gaatattttt ccataattttt cttgtatagc agtgcagctt	4200
tttccttgc ggtgtaaata gcaaagcaag caagagttct attactaaac acagcatgac	4260
tcaaaaaact tagcaattct gaaggaaagt cttggggtc ttctacctt ctcttcttt	4320
ttggaggagt agaardttga gagtccggc tagcctcatc atcaactatg ggcatttttt	4380
ctgagcaaaa cagggttcc tcattaaagg cattccacca ctgctcccat tcatcaggcc	4440
cataggttgg aatctaaaat acacaaacaa tttagaatcag tagtttaca cattatacac	4500
ttaaaaaattt tatatttacc tttagagcttt aaatctctgt aggttagtttgc tccaattatg	4560
tcacaccaca gaagtaaggt tccttcacaa agatccggga ccaaagcggc catcgccct	4620
ccccactcct gcagttcggg ggcattggatg cgccggatagc cgctgttgtt ttcctggatg	4680
ccgacggatt tgcactgccc gtagaactcc gcgaggtcg ccagcctcag gcagcagctg	4740
aaccaactcg cgaggggatc gagccgggg tgggcgaaga actccagcat gagatccccg	4800
cgctggagga tcatccagcc ggcgtcccgaaaacgattc cgaagccaa ctttcata	4860

-continued

aaggcggcg	tggaatcgaa	atctcgtat	ggcaggttgg	gcgtcgcttg	gtcggtcatt	4920	
tcgaacccca	gagtcccgct	cagaagaact	cgtcaagaag	gcatagaag	gcgtatgcgct	4980	
gccaatcg	ggcggegata	ccgtaaagca	cgaggaagcg	gtcageccat	tcgcggccaa	5040	
gctttcagc	aatatcacgg	gtagccaacg	ctatgtcctg	atagcggtcc	gccacaccca	5100	
gcggccaca	gtcgatgaat	ccagaaaago	ggccattttc	caccatgata	ttcggcaagc	5160	
aggcategc	atgggtcactg	acgagatct	cgecgteggg	catgcgcgc	ttgagcctgg	5220	
cgaacagttc	ggctggcgcg	agccccctgat	gctttgtcc	agatcatcct	gatcgacaag	5280	
accggcttc	atccgagta	gtgctcgctc	gatgcgtat	tcgcttggtg	gtcgaatggg	5340	
caggttagcc	gatcaagcgt	atgcagccgc	cgcattgcat	cagccatgat	ggatactttc	5400	
tcggcaggag	caaggtgaga	tgacaggaga	tcctgcccc	gcacttcgccc	caatagcagc	5460	
cagtcccttc	ccgcttcagt	gacaacgtcg	agcacagctg	cgcaaggaac	gcccgtcg	5520	
gccagccacg	atagccgcgc	tgccctcgcc	tgcaagttcat	tcagggcacc	ggacaggctg	5580	
gtcttgacaa	aaagaaccgg	gcccctgc	gctgacagcc	ggaacacggc	ggcatcagag	5640	
cagccgattg	tctgttgtgc	ccagtcata	ccgaatagcc	tctccacc	agcggccgga	5700	
gaacctgcgt	gcaatccatc	ttgttcaatc	atgcgaaacg	atccctatcc	tgtctcttga	5760	
tcagatctt	atccccctgc	ccatcagatc	cttggcggca	agaaagccat	ccagttact	5820	
ttgcagggct	tcccaacctt	accagaggc	gccccagctg	gcaattccgg	ttcgcttgct	5880	
gtccataaaa	ccgcccagtc	tagtatcg	catgtaa	gccc	actgc	5940	
ctctttgcgc	ttgcgtttc	ccttgccag	atagcccagt	agctgacatt	catccgggt	6000	
cagcacgtt	tctgcccact	ggcttctac	gtgttccgct	tcctttagca	gccc	ttgcgc	6060
cctgagtgt	tgccgcagcg	tgaagcttt	tgcaaaagcc	taggcctcca	aaaaagcctc	6120	
ctcaactt	ctggaatagc	tcagaggccg	aggccgcctc	ggcctctgca	taaataaaaa	6180	
aaattagtca	ccatggggc	ggagaatggg	cggaaactggg	cgagatagg	ggcgggatgg	6240	
gcggagttag	ggccgggact	atggttgtc	actaattgag	atgcgtat	tgcataactc	6300	
tgccctgtgg	ggagccctgg	gacttccac	acctgggtgc	tgactaattg	agatgcgtc	6360	
tttgcatact	tctgcctgt	ggggagcc	gggactttc	acacccta	tgacacacat	6420	
tccaca						6426	

<210> SEQ ID NO 80

<211> LENGTH: 1011

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: (mNeonGreen) 4-tDeg

<400> SEQUENCE: 80

Met	Val	Ser	Lys	Gly	Glu	Glu	Asp	Asn	Met	Ala	Ser	Leu	Pro	Ala	Thr	
1																

5 10 15

His	Glu	Leu	His	Ile	Phe	Gly	Ser	Ile	Asn	Gly	Val	Asp	Phe	Asp	Met
20															

25 30

Val	Gly	Gln	Gly	Thr	Gly	Asn	Pro	Asn	Asp	Gly	Tyr	Glu	Leu	Asn
35														

40 45

Leu	Lys	Ser	Thr	Lys	Gly	Asp	Leu	Gln	Phe	Ser	Pro	Trp	Ile	L	Val
50															

55 60

Pro	His	Ile	Gly	Tyr	Gly	Phe	His	Gln	Tyr	L	Pro	Tyr	Pro	Asp	Gly
65															

70 75 80

-continued

Met Ser Pro Phe Gln Ala Ala Met Val Asp Gly Ser Gly Tyr Gln Val
 85 90 95

 His Arg Thr Met Gln Phe Glu Asp Gly Ala Ser Leu Thr Val Asn Tyr
 100 105 110

 Arg Tyr Thr Tyr Glu Gly Ser His Ile Lys Gly Glu Ala Gln Val Lys
 115 120 125

 Gly Thr Gly Phe Pro Ala Asp Gly Pro Val Met Thr Asn Ser Leu Thr
 130 135 140

 Ala Ala Asp Trp Cys Arg Ser Lys Lys Thr Tyr Pro Asn Asp Lys Thr
 145 150 155 160

 Ile Ile Ser Thr Phe Lys Trp Ser Tyr Thr Thr Gly Asn Gly Lys Arg
 165 170 175

 Tyr Arg Ser Thr Ala Arg Thr Thr Tyr Thr Phe Ala Lys Pro Met Ala
 180 185 190

 Ala Asn Tyr Leu Lys Asn Gln Pro Met Tyr Val Phe Arg Lys Thr Glu
 195 200 205

 Leu Lys His Ser Lys Thr Glu Leu Asn Phe Lys Glu Trp Gln Lys Ala
 210 215 220

 Phe Thr Asp Val Met Gly Met Asp Glu Leu Tyr Lys Gly Gly His Met
 225 230 235 240

 Gly Thr Gly Ser Thr Gly Gly Thr Gly Val Ser Lys Gly Glu Glu
 245 250 255

 Asp Asn Met Ala Ser Leu Pro Ala Thr His Glu Leu His Ile Phe Gly
 260 265 270

 Ser Ile Asn Gly Val Asp Phe Asp Met Val Gly Gln Gly Thr Gly Asn
 275 280 285

 Pro Asn Asp Gly Tyr Glu Glu Leu Asn Leu Lys Ser Thr Lys Gly Asp
 290 295 300

 Leu Gln Phe Ser Pro Trp Ile Leu Val Pro His Ile Gly Tyr Gly Phe
 305 310 315 320

 His Gln Tyr Leu Pro Tyr Pro Asp Gly Met Ser Pro Phe Gln Ala Ala
 325 330 335

 Met Val Asp Gly Ser Gly Tyr Gln Val His Arg Thr Met Gln Phe Glu
 340 345 350

 Asp Gly Ala Ser Leu Thr Val Asn Tyr Arg Tyr Thr Tyr Glu Gly Ser
 355 360 365

 His Ile Lys Gly Glu Ala Gln Val Lys Gly Thr Gly Phe Pro Ala Asp
 370 375 380

 Gly Pro Val Met Thr Asn Ser Leu Thr Ala Ala Asp Trp Cys Arg Ser
 385 390 395 400

 Lys Lys Thr Tyr Pro Asn Asp Lys Thr Ile Ile Ser Thr Phe Lys Trp
 405 410 415

 Ser Tyr Thr Thr Gly Asn Gly Lys Arg Tyr Arg Ser Thr Ala Arg Thr
 420 425 430

 Thr Tyr Thr Phe Ala Lys Pro Met Ala Ala Asn Tyr Leu Lys Asn Gln
 435 440 445

 Pro Met Tyr Val Phe Arg Lys Thr Glu Leu Lys His Ser Lys Thr Glu
 450 455 460

 Leu Asn Phe Lys Glu Trp Gln Lys Ala Phe Thr Asp Val Met Gly Met
 465 470 475 480

 Asp Glu Leu Tyr Lys Ser Gly Leu Glu Ser Ser Gly Gly Thr Gly Gly
 485 490 495

 Ser Gly Gly Val Ser Lys Gly Glu Glu Asp Asn Met Ala Ser Leu Pro

US 12,391,948 B2

269**270**

-continued

500	505	510
Ala Thr His Glu Leu His Ile Phe Gly Ser Ile Asn Gly Val Asp Phe		
515	520	525
Asp Met Val Gly Gln Gly Thr Gly Asn Pro Asn Asp Gly Tyr Glu Glu		
530	535	540
Leu Asn Leu Lys Ser Thr Lys Gly Asp Leu Gln Phe Ser Pro Trp Ile		
545	550	555
Leu Val Pro His Ile Gly Tyr Gly Phe His Gln Tyr Leu Pro Tyr Pro		
565	570	575
Asp Gly Met Ser Pro Phe Gln Ala Ala Met Val Asp Gly Ser Gly Tyr		
580	585	590
Gln Val His Arg Thr Met Gln Phe Glu Asp Gly Ala Ser Leu Thr Val		
595	600	605
Asn Tyr Arg Tyr Thr Tyr Glu Gly Ser His Ile Lys Gly Glu Ala Gln		
610	615	620
Val Lys Gly Thr Phe Pro Ala Asp Gly Pro Val Met Thr Asn Ser		
625	630	635
Leu Thr Ala Ala Asp Trp Cys Arg Ser Lys Lys Thr Tyr Pro Asn Asp		
645	650	655
Lys Thr Ile Ile Ser Thr Phe Lys Trp Ser Tyr Thr Thr Gly Asn Gly		
660	665	670
Lys Arg Tyr Arg Ser Thr Ala Arg Thr Thr Tyr Thr Phe Ala Lys Pro		
675	680	685
Met Ala Ala Asn Tyr Leu Lys Asn Gln Pro Met Tyr Val Phe Arg Lys		
690	695	700
Thr Glu Leu Lys His Ser Lys Thr Glu Leu Asn Phe Lys Glu Trp Gln		
705	710	715
720		
Lys Ala Phe Thr Asp Val Met Gly Met Asp Glu Leu Tyr Lys Gly Gly		
725	730	735
Ser Gly Thr Gly Gly Thr Ala Ser Ser Gly Ser Gly Gly Val Ser		
740	745	750
Lys Gly Glu Glu Asp Asn Met Ala Ser Leu Pro Ala Thr His Glu Leu		
755	760	765
His Ile Phe Gly Ser Ile Asn Gly Val Asp Phe Asp Met Val Gly Gln		
770	775	780
Gly Thr Gly Asn Pro Asn Asp Gly Tyr Glu Glu Leu Asn Leu Lys Ser		
785	790	795
800		
Thr Lys Gly Asp Leu Gln Phe Ser Pro Trp Ile Leu Val Pro His Ile		
805	810	815
Gly Tyr Gly Phe His Gln Tyr Leu Pro Tyr Pro Asp Gly Met Ser Pro		
820	825	830
Phe Gln Ala Ala Met Val Asp Gly Ser Gly Tyr Gln Val His Arg Thr		
835	840	845
Met Gln Phe Glu Asp Gly Ala Ser Leu Thr Val Asn Tyr Arg Tyr Thr		
850	855	860
Tyr Glu Gly Ser His Ile Lys Gly Glu Ala Gln Val Lys Gly Thr Gly		
865	870	875
880		
Phe Pro Ala Asp Gly Pro Val Met Thr Asn Ser Leu Thr Ala Ala Asp		
885	890	895
Trp Cys Arg Ser Lys Lys Thr Tyr Pro Asn Asp Lys Thr Ile Ile Ser		
900	905	910
Thr Phe Lys Trp Ser Tyr Thr Thr Gly Asn Gly Lys Arg Tyr Arg Ser		
915	920	925

-continued

Thr Ala Arg Thr Thr Tyr Thr Phe Ala Lys Pro Met Ala Ala Asn Tyr
 930 935 940
 Leu Lys Asn Gln Pro Met Tyr Val Phe Arg Lys Thr Glu Leu Lys His
 945 950 955 960
 Ser Lys Thr Glu Leu Asn Phe Lys Glu Trp Gln Lys Ala Phe Thr Asp
 965 970 975
 Val Met Gly Met Asp Glu Leu Tyr Lys Gly Gly Arg Ser Gly Gly
 980 985 990
 Ser Gly Pro Arg Pro Arg Gly Thr Arg Gly Lys Gly Arg Arg Ile Arg
 995 1000 1005
 Arg Arg Gly
 1010

<210> SEQ ID NO 81
 <211> LENGTH: 266
 <212> TYPE: PRT
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: mNeonGreen-tDeg

<400> SEQUENCE: 81

Met Val Ser Lys Gly Glu Glu Asp Asn Met Ala Ser Leu Pro Ala Thr
 1 5 10 15
 His Glu Leu His Ile Phe Gly Ser Ile Asn Gly Val Asp Phe Asp Met
 20 25 30
 Val Gly Gln Gly Thr Gly Asn Pro Asn Asp Gly Tyr Glu Glu Leu Asn
 35 40 45
 Leu Lys Ser Thr Lys Gly Asp Leu Gln Phe Ser Pro Trp Ile Leu Val
 50 55 60
 Pro His Ile Gly Tyr Gly Phe His Gln Tyr Leu Pro Tyr Pro Asp Gly
 65 70 75 80
 Met Ser Pro Phe Gln Ala Ala Met Val Asp Gly Ser Gly Tyr Gln Val
 85 90 95
 His Arg Thr Met Gln Phe Glu Asp Gly Ala Ser Leu Thr Val Asn Tyr
 100 105 110
 Arg Tyr Thr Tyr Glu Gly Ser His Ile Lys Gly Glu Ala Gln Val Lys
 115 120 125
 Gly Thr Gly Phe Pro Ala Asp Gly Pro Val Met Thr Asn Ser Leu Thr
 130 135 140
 Ala Ala Asp Trp Cys Arg Ser Lys Lys Thr Tyr Pro Asn Asp Lys Thr
 145 150 155 160
 Ile Ile Ser Thr Phe Lys Trp Ser Tyr Thr Thr Gly Asn Gly Lys Arg
 165 170 175
 Tyr Arg Ser Thr Ala Arg Thr Tyr Thr Phe Ala Lys Pro Met Ala
 180 185 190
 Ala Asn Tyr Leu Lys Asn Gln Pro Met Tyr Val Phe Arg Lys Thr Glu
 195 200 205
 Leu Lys His Ser Lys Thr Glu Leu Asn Phe Lys Glu Trp Gln Lys Ala
 210 215 220
 Phe Thr Asp Val Met Gly Met Asp Glu Leu Tyr Lys Gly Gly His Met
 225 230 235 240
 Gly Gly Gly Ser Gly Gly Ser Gly Pro Arg Pro Arg Gly Thr Arg
 245 250 255
 Gly Lys Gly Arg Arg Ile Arg Arg Gly
 260 265

-continued

<210> SEQ ID NO 82
<211> LENGTH: 261
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: mCherry-tDeg

<400> SEQUENCE: 82

```

Met Val Ser Lys Gly Glu Glu Asp Asn Met Ala Ile Ile Lys Glu Phe
1          5           10          15

Met Arg Phe Lys Val His Met Glu Gly Ser Val Asn Gly His Glu Phe
20         25           30

Glu Ile Glu Gly Glu Gly Arg Pro Tyr Glu Gly Thr Gln Thr
35         40           45

Ala Lys Leu Lys Val Thr Lys Gly Gly Pro Leu Pro Phe Ala Trp Asp
50         55           60

Ile Leu Ser Pro Gln Phe Met Tyr Gly Ser Lys Ala Tyr Val Lys His
65         70           75           80

Pro Ala Asp Ile Pro Asp Tyr Leu Lys Leu Ser Phe Pro Glu Gly Phe
85         90           95

Lys Trp Glu Arg Val Met Asn Phe Glu Asp Gly Gly Val Val Thr Val
100        105          110

Thr Gln Asp Ser Ser Leu Gln Asp Gly Glu Phe Ile Tyr Lys Val Lys
115        120          125

Leu Arg Gly Thr Asn Phe Pro Ser Asp Gly Pro Val Met Gln Lys Lys
130        135          140

Thr Met Gly Trp Glu Ala Ser Ser Glu Arg Met Tyr Pro Glu Asp Gly
145        150          155          160

Ala Leu Lys Gly Glu Ile Lys Gln Arg Leu Lys Leu Lys Asp Gly Gly
165        170          175

His Tyr Asp Ala Glu Val Lys Thr Thr Tyr Lys Ala Lys Lys Pro Val
180        185          190

Gln Leu Pro Gly Ala Tyr Asn Val Asn Ile Lys Leu Asp Ile Thr Ser
195        200          205

His Asn Glu Asp Tyr Thr Ile Val Glu Gln Tyr Glu Arg Ala Glu Gly
210        215          220

Arg His Ser Thr Gly Gly Met Asp Glu Leu Tyr Lys Gly Ser Gly
225        230          235          240

Gly Gly Ser Gly Pro Arg Pro Arg Gly Thr Arg Gly Lys Gly Arg Arg
245        250          255

Ile Arg Arg Arg Gly
260

```

<210> SEQ ID NO 83
<211> LENGTH: 201
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: NanoLuc-tDeg

<400> SEQUENCE: 83

```

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

Gly Tyr Asn Leu Asp Gln Val Leu Glu Gln Gly Gly Val Ser Ser Leu
20         25           30

Phe Gln Asn Leu Gly Val Ser Val Thr Pro Ile Gln Arg Ile Val Leu

```

US 12,391,948 B2

275**276**

-continued

35	40	45
Ser	Gly	Glu
50	55	60
Gly	Gly	
65	70	75
Leu	Ser	Gly
		Asp
		Gln
		Met
		Gly
		Gln
		Ile
		Glu
		Lys
		Ile
		Asp
		His
		Ile
		Val
		Ile
		Ile
		Pro
		Tyr
Val	Val	Tyr
		Pro
		Val
		Asp
		Asp
		His
		His
		Phe
		Lys
		Ile
		Phe
		Lys
		Val
		Ile
		Leu
		His
		Tyr
Gly	Thr	Leu
		Val
		Ile
		Asp
		Gly
		Val
		Thr
		Pro
		Asn
		Met
		Ile
		Asp
		Tyr
		Phe
Gly	Arg	Pro
		Tyr
		Glu
		Gly
		Ile
		Ala
		Val
		Phe
		Asp
		Gly
		Lys
		Lys
		Ile
		Thr
		Phe
Val	Thr	Gly
		Thr
		Leu
		Trp
		Asn
		Gly
		Asn
		Lys
		Ile
		Ile
		Asp
		Glu
		Arg
		Leu
Ile	Asn	Pro
		Asp
		Gly
		Ser
		Leu
		Leu
		Phe
		Arg
		Val
		Thr
		Ile
		Asn
		Gly
145	150	155
100	105	110
115	120	125
130	135	140
180	185	190
195	200	

<210> SEQ ID NO 84
<211> LENGTH: 264
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: EYFP-tDeg

<400> SEQUENCE: 84

Met	Val	Ser	Lys	Gly	Glu	Glu	Leu	Phe	Thr	Gly	Val	Val	Pro	Ile	Leu
1							5		10				15		
Val	Glu	Leu	Asp	Gly	Asp	Val	Asn	Gly	His	Lys	Phe	Ser	Val	Ser	Gly
							20		25				30		
Glu	Gly	Glu	Gly	Asp	Ala	Thr	Tyr	Gly	Lys	Leu	Thr	Leu	Lys	Phe	Ile
						35		40			45				
Cys	Thr	Thr	Gly	Lys	Leu	Pro	Val	Pro	Trp	Pro	Thr	Leu	Val	Thr	Thr
						50		55			60				
Phe	Gly	Tyr	Gly	Leu	Gln	Cys	Phe	Ala	Arg	Tyr	Pro	Asp	His	Met	Lys
						65		70			75			80	
Gln	His	Asp	Phe	Phe	Lys	Ser	Ala	Met	Pro	Glu	Gly	Tyr	Val	Gln	Glu
						85		90			95				
Arg	Thr	Ile	Phe	Phe	Lys	Asp	Asp	Gly	Asn	Tyr	Lys	Thr	Arg	Ala	Glu
						100		105			110				
Val	Lys	Phe	Glu	Gly	Asp	Thr	Leu	Val	Asn	Arg	Ile	Glu	Leu	Lys	Gly
						115		120			125				
Ile	Asp	Phe	Lys	Glu	Asp	Gly	Asn	Ile	Leu	Gly	His	Lys	Leu	Glu	Tyr
						130		135			140				
Asn	Tyr	Asn	Ser	His	Asn	Val	Tyr	Ile	Met	Ala	Asp	Lys	Gln	Lys	Asn
						145		150			155			160	
Gly	Ile	Lys	Val	Asn	Phe	Lys	Ile	Arg	His	Asn	Ile	Glu	Asp	Gly	Ser
						165		170			175				
Val	Gln	Leu	Ala	Asp	His	Tyr	Gln	Gln	Asn	Thr	Pro	Ile	Gly	Asp	Gly
						180		185			190				
Pro	Val	Leu	Leu	Pro	Asp	Asn	His	Tyr	Leu	Ser	Tyr	Gln	Ser	Ala	Leu

US 12,391,948 B2

277

-continued

278

195	200	205
Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe		
210	215	220
Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys Gly		
225	230	235
Gly Ser Gly Gly Ser Gly Pro Arg Pro Arg Gly Thr Arg Gly Lys		
245	250	255
Gly Arg Arg Ile Arg Arg Arg Gly		
260		

<210> SEQ ID NO 85
<211> LENGTH: 482
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: EGFP-TetR-tDeg

<400> SEQUENCE: 85

Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu		
1	5	10
15		
Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly		
20	25	30
Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile		
35	40	45
Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr		
50	55	60
Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys		
65	70	75
80		
Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu		
85	90	95
Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu		
100	105	110
Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly		
115	120	125
Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr		
130	135	140
Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn		
145	150	155
160		
Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser		
165	170	175
Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Glu Asp Gly		
180	185	190
Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Lys Leu		
195	200	205
Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe		
210	215	220
Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys Gly		
225	230	235
240		
Thr Gly Ala Cys Gly Thr Ser Gly Gly Arg Leu Asp Lys Ser Lys Val		
245	250	255
Ile Asn Ser Ala Leu Glu Leu Leu Asn Glu Val Gly Ile Glu Gly Leu		
260	265	270
Thr Thr Arg Lys Leu Ala Gln Lys Leu Gly Val Glu Gln Pro Thr Leu		
275	280	285
Tyr Trp His Val Lys Asn Lys Arg Ala Leu Leu Asp Ala Leu Ala Ile		

US 12,391,948 B2

279**280**

-continued

290

295

300

Glu Met Leu Asp Arg His His Thr His Phe Cys Pro Leu Glu Gly Glu
 305 310 315 320

Ser Trp Gln Asp Phe Leu Arg Asn Asn Ala Lys Ser Phe Arg Cys Ala
 325 330 335

Leu Leu Ser His Arg Asp Gly Ala Lys Val His Leu Gly Thr Arg Pro
 340 345 350

Thr Glu Lys Gln Tyr Glu Thr Leu Glu Asn Gln Leu Ala Phe Leu Cys
 355 360 365

Gln Gln Gly Phe Ser Leu Glu Asn Ala Leu Tyr Ala Leu Ser Ala Val
 370 375 380

Gly His Phe Thr Leu Gly Cys Val Leu Glu Asp Gln Glu His Gln Val
 385 390 395 400

Ala Lys Glu Glu Arg Glu Thr Pro Thr Asp Ser Met Pro Pro Leu
 405 410 415

Leu Arg Gln Ala Ile Glu Leu Phe Asp His Gln Gly Ala Glu Pro Ala
 420 425 430

Phe Leu Phe Gly Leu Glu Leu Ile Ile Cys Gly Leu Glu Lys Gln Leu
 435 440 445

Lys Cys Glu Ser Gly Ser Gly Ser Gly Thr Gly Gly Thr Gly Gly Ser
 450 455 460

Gly Pro Arg Pro Arg Gly Thr Arg Gly Lys Gly Arg Arg Ile Arg Arg
 465 470 475 480

Arg Gly

<210> SEQ ID NO 86

<211> LENGTH: 479

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: mCherry-TetR-tDeg

<400> SEQUENCE: 86

Met Val Ser Lys Gly Glu Glu Asp Asn Met Ala Ile Ile Lys Glu Phe
 1 5 10 15

Met Arg Phe Lys Val His Met Glu Gly Ser Val Asn Gly His Glu Phe
 20 25 30

Glu Ile Glu Gly Glu Gly Glu Arg Pro Tyr Glu Gly Thr Gln Thr
 35 40 45

Ala Lys Leu Lys Val Thr Lys Gly Gly Pro Leu Pro Phe Ala Trp Asp
 50 55 60

Ile Leu Ser Pro Gln Phe Met Tyr Gly Ser Lys Ala Tyr Val Lys His
 65 70 75 80

Pro Ala Asp Ile Pro Asp Tyr Leu Lys Leu Ser Phe Pro Glu Gly Phe
 85 90 95

Lys Trp Glu Arg Val Met Asn Phe Glu Asp Gly Gly Val Val Thr Val
 100 105 110

Thr Gln Asp Ser Ser Leu Gln Asp Gly Glu Phe Ile Tyr Lys Val Lys
 115 120 125

Leu Arg Gly Thr Asn Phe Pro Ser Asp Gly Pro Val Met Gln Lys Lys
 130 135 140

Thr Met Gly Trp Glu Ala Ser Ser Glu Arg Met Tyr Pro Glu Asp Gly
 145 150 155 160

Ala Leu Lys Gly Glu Ile Lys Gln Arg Leu Lys Leu Lys Asp Gly Gly
 165 170 175

US 12,391,948 B2

281

-continued

His Tyr Asp Ala Glu Val Lys Thr Thr Tyr Lys Ala Lys Lys Pro Val
 180 185 190
 Gln Leu Pro Gly Ala Tyr Asn Val Asn Ile Lys Leu Asp Ile Thr Ser
 195 200 205
 His Asn Glu Asp Tyr Thr Ile Val Glu Gln Tyr Glu Arg Ala Glu Gly
 210 215 220
 Arg His Ser Thr Gly Gly Met Asp Glu Leu Tyr Lys Gly Thr Gly Ala
 225 230 235 240
 Cys Gly Thr Ser Gly Gly Arg Leu Asp Lys Ser Lys Val Ile Asn Ser
 245 250 255
 Ala Leu Glu Leu Leu Asn Glu Val Gly Ile Glu Gly Leu Thr Thr Arg
 260 265 270
 Lys Leu Ala Gln Lys Leu Gly Val Glu Gln Pro Thr Leu Tyr Trp His
 275 280 285
 Val Lys Asn Lys Arg Ala Leu Leu Asp Ala Leu Ala Ile Glu Met Leu
 290 295 300
 Asp Arg His His Thr His Phe Cys Pro Leu Glu Gly Glu Ser Trp Gln
 305 310 315 320
 Asp Phe Leu Arg Asn Asn Ala Lys Ser Phe Arg Cys Ala Leu Leu Ser
 325 330 335
 His Arg Asp Gly Ala Lys Val His Leu Gly Thr Arg Pro Thr Glu Lys
 340 345 350
 Gln Tyr Glu Thr Leu Glu Asn Gln Leu Ala Phe Leu Cys Gln Gln Gly
 355 360 365
 Phe Ser Leu Glu Asn Ala Leu Tyr Ala Leu Ser Ala Val Gly His Phe
 370 375 380
 Thr Leu Gly Cys Val Leu Glu Asp Gln Glu His Gln Val Ala Lys Glu
 385 390 395 400
 Glu Arg Glu Thr Pro Thr Thr Asp Ser Met Pro Pro Leu Leu Arg Gln
 405 410 415
 Ala Ile Glu Leu Phe Asp His Gln Gly Ala Glu Pro Ala Phe Leu Phe
 420 425 430
 Gly Leu Glu Leu Ile Ile Cys Gly Leu Glu Lys Gln Leu Lys Cys Glu
 435 440 445
 Ser Gly Ser Gly Ser Gly Thr Gly Gly Thr Gly Ser Gly Pro Arg
 450 455 460
 Pro Arg Gly Thr Arg Gly Lys Gly Arg Arg Ile Arg Arg Arg Gly
 465 470 475

282

<210> SEQ ID NO 87
 <211> LENGTH: 1023
 <212> TYPE: PRT
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: EGFP-EZH2-tDeg

<400> SEQUENCE: 87

Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu
 1 5 10 15
 Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
 20 25 30
 Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
 35 40 45
 Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
 50 55 60

-continued

Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys
 65 70 75 80
 Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
 85 90 95
 Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
 100 105 110
 Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
 115 120 125
 Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
 130 135 140
 Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
 145 150 155 160
 Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
 165 170 175
 Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
 180 185 190
 Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Lys Leu
 195 200 205
 Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
 210 215 220
 Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys Gly
 225 230 235 240
 Thr Gly Ala Cys Gly Thr Ser Gly Gly Met Gly Gln Thr Gly Lys Lys
 245 250 255
 Ser Glu Lys Gly Pro Val Cys Trp Arg Lys Arg Val Lys Ser Glu Tyr
 260 265 270
 Met Arg Leu Arg Gln Leu Lys Arg Phe Arg Arg Ala Asp Glu Val Lys
 275 280 285
 Ser Met Phe Ser Ser Asn Arg Gln Lys Ile Leu Glu Arg Thr Glu Ile
 290 295 300
 Leu Asn Gln Glu Trp Lys Gln Arg Arg Ile Gln Pro Val His Ile Leu
 305 310 315 320
 Thr Ser Val Ser Ser Leu Arg Gly Thr Arg Glu Cys Ser Val Thr Ser
 325 330 335
 Asp Leu Asp Phe Pro Thr Gln Val Ile Pro Leu Lys Thr Leu Asn Ala
 340 345 350
 Val Ala Ser Val Pro Ile Met Tyr Ser Trp Ser Pro Leu Gln Gln Asn
 355 360 365
 Phe Met Val Glu Asp Glu Thr Val Leu His Asn Ile Pro Tyr Met Gly
 370 375 380
 Asp Glu Val Leu Asp Gln Asp Gly Thr Phe Ile Glu Glu Leu Ile Lys
 385 390 395 400
 Asn Tyr Asp Gly Lys Val His Gly Asp Arg Glu Cys Gly Phe Ile Asn
 405 410 415
 Asp Glu Ile Phe Val Glu Leu Val Asn Ala Leu Gly Gln Tyr Asn Asp
 420 425 430
 Asp Asp Asp Asp Asp Gly Asp Asp Pro Glu Glu Arg Glu Glu Lys
 435 440 445
 Gln Lys Asp Leu Glu Asp His Arg Asp Asp Lys Glu Ser Arg Pro Pro
 450 455 460
 Arg Lys Phe Pro Ser Asp Lys Ile Phe Glu Ala Ile Ser Ser Met Phe
 465 470 475 480

-continued

Pro Asp Lys Gly Thr Ala Glu Glu Leu Lys Glu Lys Tyr Lys Glu Leu
 485 490 495

Thr Glu Gln Gln Leu Pro Gly Ala Leu Pro Pro Glu Cys Thr Pro Asn
 500 505 510

Ile Asp Gly Pro Asn Ala Lys Ser Val Gln Arg Glu Gln Ser Leu His
 515 520 525

Ser Phe His Thr Leu Phe Cys Arg Arg Cys Phe Lys Tyr Asp Cys Phe
 530 535 540

Leu His Pro Phe His Ala Thr Pro Asn Thr Tyr Lys Arg Lys Asn Thr
 545 550 555 560

Glu Thr Ala Leu Asp Asn Lys Pro Cys Gly Pro Gln Cys Tyr Gln His
 565 570 575

Leu Glu Gly Ala Lys Glu Phe Ala Ala Ala Leu Thr Ala Glu Arg Ile
 580 585 590

Lys Thr Pro Pro Lys Arg Pro Gly Gly Arg Arg Gly Arg Leu Pro
 595 600 605

Asn Asn Ser Ser Arg Pro Ser Thr Pro Thr Ile Asn Val Leu Glu Ser
 610 615 620

Lys Asp Thr Asp Ser Asp Arg Glu Ala Gly Thr Glu Thr Gly Gly Glu
 625 630 635 640

Asn Asn Asp Lys Glu Glu Glu Lys Lys Asp Glu Thr Ser Ser Ser
 645 650 655

Ser Glu Ala Asn Ser Arg Cys Gln Thr Pro Ile Lys Met Lys Pro Asn
 660 665 670

Ile Glu Pro Pro Glu Asn Val Glu Trp Ser Gly Ala Glu Ala Ser Met
 675 680 685

Phe Arg Val Leu Ile Gly Thr Tyr Tyr Asp Asn Phe Cys Ala Ile Ala
 690 695 700

Arg Leu Ile Gly Thr Lys Thr Cys Arg Gln Val Tyr Glu Phe Arg Val
 705 710 715 720

Lys Glu Ser Ser Ile Ile Ala Pro Ala Pro Ala Glu Asp Val Asp Thr
 725 730 735

Pro Pro Arg Lys Lys Arg Lys His Arg Leu Trp Ala Ala His Cys
 740 745 750

Arg Lys Ile Gln Leu Lys Lys Asp Gly Ser Ser Asn His Val Tyr Asn
 755 760 765

Tyr Gln Pro Cys Asp His Pro Arg Gln Pro Cys Asp Ser Ser Cys Pro
 770 775 780

Cys Val Ile Ala Gln Asn Phe Cys Glu Lys Phe Cys Gln Cys Ser Ser
 785 790 795 800

Glu Cys Gln Asn Arg Phe Pro Gly Cys Arg Cys Lys Ala Gln Cys Asn
 805 810 815

Thr Lys Gln Cys Pro Cys Tyr Leu Ala Val Arg Glu Cys Asp Pro Asp
 820 825 830

Leu Cys Leu Thr Cys Gly Ala Ala Asp His Trp Asp Ser Lys Asn Val
 835 840 845

Ser Cys Lys Asn Cys Ser Ile Gln Arg Gly Ser Lys Lys His Leu Leu
 850 855 860

Leu Ala Pro Ser Asp Val Ala Gly Trp Gly Ile Phe Ile Lys Asp Pro
 865 870 875 880

Val Gln Lys Asn Glu Phe Ile Ser Glu Tyr Cys Gly Glu Ile Ile Ser
 885 890 895

Gln Asp Glu Ala Asp Arg Arg Gly Lys Val Tyr Asp Lys Tyr Met Cys

US 12,391,948 B2

287

-continued

900	905	910
Ser Phe Leu Phe Asn Leu Asn Asn Asp Phe Val Val Asp Ala Thr Arg		
915	920	925
Lys Gly Asn Lys Ile Arg Phe Ala Asn His Ser Val Asn Pro Asn Cys		
930	935	940
Tyr Ala Lys Val Met Met Val Asn Gly Asp His Arg Ile Gly Ile Phe		
945	950	955
Ala Lys Arg Ala Ile Gln Thr Gly Glu Glu Leu Phe Phe Asp Tyr Arg		
965	970	975
Tyr Ser Gln Ala Asp Ala Leu Lys Tyr Val Gly Ile Glu Arg Glu Met		
980	985	990
Glu Ile Pro Gly Ser Gly Thr Gly Gly Thr Gly Ser Gly Pro Arg		
995	1000	1005
Pro Arg Gly Thr Arg Gly Lys Gly Arg Arg Ile Arg Arg Arg Gly		
1010	1015	1020

<210> SEQ ID NO: 88

<211> LENGTH: 1020

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: mCherry-EZH2-tDeg

<400> SEQUENCE: 88

Met Val Ser Lys Gly Glu Glu Asp Asn Met Ala Ile Ile Lys Glu Phe		
1	5	10
Met Arg Phe Lys Val His Met Glu Gly Ser Val Asn Gly His Glu Phe		
20	25	30
Glu Ile Glu Gly Glu Gly Arg Pro Tyr Glu Gly Thr Gln Thr		
35	40	45
Ala Lys Leu Lys Val Thr Lys Gly Gly Pro Leu Pro Phe Ala Trp Asp		
50	55	60
Ile Leu Ser Pro Gln Phe Met Tyr Gly Ser Lys Ala Tyr Val Lys His		
65	70	75
80		
Pro Ala Asp Ile Pro Asp Tyr Leu Lys Leu Ser Phe Pro Glu Gly Phe		
85	90	95
Lys Trp Glu Arg Val Met Asn Phe Glu Asp Gly Gly Val Val Thr Val		
100	105	110
Thr Gln Asp Ser Ser Leu Gln Asp Gly Glu Phe Ile Tyr Lys Val Lys		
115	120	125
Leu Arg Gly Thr Asn Phe Pro Ser Asp Gly Pro Val Met Gln Lys Lys		
130	135	140
Thr Met Gly Trp Glu Ala Ser Ser Glu Arg Met Tyr Pro Glu Asp Gly		
145	150	155
160		
Ala Leu Lys Gly Glu Ile Lys Gln Arg Leu Lys Leu Lys Asp Gly Gly		
165	170	175
His Tyr Asp Ala Glu Val Lys Thr Thr Tyr Lys Ala Lys Lys Pro Val		
180	185	190
Gln Leu Pro Gly Ala Tyr Asn Val Asn Ile Lys Leu Asp Ile Thr Ser		
195	200	205
His Asn Glu Asp Tyr Thr Ile Val Glu Gln Tyr Glu Arg Ala Glu Gly		
210	215	220
Arg His Ser Thr Gly Gly Met Asp Glu Leu Tyr Lys Gly Thr Gly Ala		
225	230	235
240		
Cys Gly Thr Ser Gly Gly Met Gly Gln Thr Gly Lys Lys Ser Glu Lys		

US 12,391,948 B2

289**290**

-continued

245	250	255
Gly Pro Val Cys Trp Arg Lys Arg Val Lys Ser Glu Tyr Met Arg Leu		
260	265	270
Arg Gln Leu Lys Arg Phe Arg Arg Ala Asp Glu Val Lys Ser Met Phe		
275	280	285
Ser Ser Asn Arg Gln Lys Ile Leu Glu Arg Thr Glu Ile Leu Asn Gln		
290	295	300
Glu Trp Lys Gln Arg Arg Ile Gln Pro Val His Ile Leu Thr Ser Val		
305	310	315
320		
Ser Ser Leu Arg Gly Thr Arg Glu Cys Ser Val Thr Ser Asp Leu Asp		
325	330	335
Phe Pro Thr Gln Val Ile Pro Leu Lys Thr Leu Asn Ala Val Ala Ser		
340	345	350
Val Pro Ile Met Tyr Ser Trp Ser Pro Leu Gln Gln Asn Phe Met Val		
355	360	365
Glu Asp Glu Thr Val Leu His Asn Ile Pro Tyr Met Gly Asp Glu Val		
370	375	380
Leu Asp Gln Asp Gly Thr Phe Ile Glu Glu Leu Ile Lys Asn Tyr Asp		
385	390	395
400		
Gly Lys Val His Gly Asp Arg Glu Cys Gly Phe Ile Asn Asp Glu Ile		
405	410	415
Phe Val Glu Leu Val Asn Ala Leu Gly Gln Tyr Asn Asp Asp Asp		
420	425	430
Asp Asp Asp Gly Asp Asp Pro Glu Glu Arg Glu Glu Lys Gln Lys Asp		
435	440	445
Leu Glu Asp His Arg Asp Asp Lys Glu Ser Arg Pro Pro Arg Lys Phe		
450	455	460
Pro Ser Asp Lys Ile Phe Glu Ala Ile Ser Ser Met Phe Pro Asp Lys		
465	470	475
480		
Gly Thr Ala Glu Glu Leu Lys Glu Lys Tyr Lys Glu Leu Thr Glu Gln		
485	490	495
Gln Leu Pro Gly Ala Leu Pro Pro Glu Cys Thr Pro Asn Ile Asp Gly		
500	505	510
Pro Asn Ala Lys Ser Val Gln Arg Glu Gln Ser Leu His Ser Phe His		
515	520	525
Thr Leu Phe Cys Arg Arg Cys Phe Lys Tyr Asp Cys Phe Leu His Pro		
530	535	540
Phe His Ala Thr Pro Asn Thr Tyr Lys Arg Lys Asn Thr Glu Thr Ala		
545	550	555
560		
Leu Asp Asn Lys Pro Cys Gly Pro Gln Cys Tyr Gln His Leu Glu Gly		
565	570	575
Ala Lys Glu Phe Ala Ala Ala Leu Thr Ala Glu Arg Ile Lys Thr Pro		
580	585	590
Pro Lys Arg Pro Gly Gly Arg Arg Gly Arg Leu Pro Asn Asn Ser		
595	600	605
Ser Arg Pro Ser Thr Pro Thr Ile Asn Val Leu Glu Ser Lys Asp Thr		
610	615	620
Asp Ser Asp Arg Glu Ala Gly Thr Glu Thr Gly Gly Glu Asn Asn Asp		
625	630	635
640		
Lys Glu Glu Glu Lys Lys Asp Glu Thr Ser Ser Ser Ser Glu Ala		
645	650	655
Asn Ser Arg Cys Gln Thr Pro Ile Lys Met Lys Pro Asn Ile Glu Pro		
660	665	670

US 12,391,948 B2

291

-continued

Pro Glu Asn Val Glu Trp Ser Gly Ala Glu Ala Ser Met Phe Arg Val
 675 680 685
 Leu Ile Gly Thr Tyr Tyr Asp Asn Phe Cys Ala Ile Ala Arg Leu Ile
 690 695 700
 Gly Thr Lys Thr Cys Arg Gln Val Tyr Glu Phe Arg Val Lys Glu Ser
 705 710 715 720
 Ser Ile Ile Ala Pro Ala Pro Ala Glu Asp Val Asp Thr Pro Pro Arg
 725 730 735
 Lys Lys Lys Arg Lys His Arg Leu Trp Ala Ala His Cys Arg Lys Ile
 740 745 750
 Gln Leu Lys Lys Asp Gly Ser Ser Asn His Val Tyr Asn Tyr Gln Pro
 755 760 765
 Cys Asp His Pro Arg Gln Pro Cys Asp Ser Ser Cys Pro Cys Val Ile
 770 775 780
 Ala Gln Asn Phe Cys Glu Lys Phe Cys Gln Cys Ser Ser Glu Cys Gln
 785 790 795 800
 Asn Arg Phe Pro Gly Cys Arg Cys Lys Ala Gln Cys Asn Thr Lys Gln
 805 810 815
 Cys Pro Cys Tyr Leu Ala Val Arg Glu Cys Asp Pro Asp Leu Cys Leu
 820 825 830
 Thr Cys Gly Ala Ala Asp His Trp Asp Ser Lys Asn Val Ser Cys Lys
 835 840 845
 Asn Cys Ser Ile Gln Arg Gly Ser Lys Lys His Leu Leu Leu Ala Pro
 850 855 860
 Ser Asp Val Ala Gly Trp Gly Ile Phe Ile Lys Asp Pro Val Gln Lys
 865 870 875 880
 Asn Glu Phe Ile Ser Glu Tyr Cys Gly Glu Ile Ile Ser Gln Asp Glu
 885 890 895
 Ala Asp Arg Arg Gly Lys Val Tyr Asp Lys Tyr Met Cys Ser Phe Leu
 900 905 910
 Phe Asn Leu Asn Asn Asp Phe Val Val Asp Ala Thr Arg Lys Gly Asn
 915 920 925
 Lys Ile Arg Phe Ala Asn His Ser Val Asn Pro Asn Cys Tyr Ala Lys
 930 935 940
 Val Met Met Val Asn Gly Asp His Arg Ile Gly Ile Phe Ala Lys Arg
 945 950 955 960
 Ala Ile Gln Thr Gly Glu Leu Phe Phe Asp Tyr Arg Tyr Ser Gln
 965 970 975
 Ala Asp Ala Leu Lys Tyr Val Gly Ile Glu Arg Glu Met Glu Ile Pro
 980 985 990
 Gly Ser Gly Thr Gly Gly Thr Gly Gly Ser Gly Pro Arg Pro Arg Gly
 995 1000 1005
 Thr Arg Gly Lys Gly Arg Arg Ile Arg Arg Arg Gly
 1010 1015 1020

292

<210> SEQ ID NO 89
 <211> LENGTH: 746
 <212> TYPE: PRT
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: EGFP-NFkB-tDeg

 <400> SEQUENCE: 89

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

-continued

Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
 20 25 30
 Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
 35 40 45
 Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
 50 55 60
 Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys
 65 70 75 80
 Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
 85 90 95
 Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
 100 105 110
 Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
 115 120 125
 Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
 130 135 140
 Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
 145 150 155 160
 Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
 165 170 175
 Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
 180 185 190
 Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Lys Leu
 195 200 205
 Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
 210 215 220
 Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys Gly
 225 230 235 240
 Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Thr Gly Ala
 245 250 255
 Glu Asp Asp Pro Tyr Leu Gly Arg Pro Glu Gln Met Phe His Leu Asp
 260 265 270
 Pro Ser Leu Thr His Thr Ile Phe Asn Pro Glu Val Phe Gln Pro Gln
 275 280 285
 Met Ala Leu Pro Thr Ala Asp Gly Pro Tyr Leu Gln Ile Leu Glu Gln
 290 295 300
 Pro Lys Gln Arg Gly Phe Arg Phe Arg Tyr Val Cys Glu Gly Pro Ser
 305 310 315 320
 His Gly Gly Leu Pro Gly Ala Ser Ser Glu Lys Asn Lys Lys Ser Tyr
 325 330 335
 Pro Gln Val Lys Ile Cys Asn Tyr Val Gly Pro Ala Lys Val Ile Val
 340 345 350
 Gln Leu Val Thr Asn Gly Lys Asn Ile His Leu His Ala His Ser Leu
 355 360 365
 Val Gly Lys His Cys Glu Asp Gly Ile Cys Thr Val Thr Ala Gly Pro
 370 375 380
 Lys Asp Met Val Val Gly Phe Ala Asn Leu Gly Ile Leu His Val Thr
 385 390 395 400
 Lys Lys Lys Val Phe Glu Thr Leu Glu Ala Arg Met Thr Glu Ala Cys
 405 410 415
 Ile Arg Gly Tyr Asn Pro Gly Leu Leu Val His Pro Asp Leu Ala Tyr
 420 425 430

US 12,391,948 B2

295**296**

-continued

Leu Gln Ala Glu Gly Gly Asp Arg Gln Leu Gly Asp Arg Glu Lys
 435 440 445
 Glu Leu Ile Arg Gln Ala Ala Leu Gln Gln Thr Lys Glu Met Asp Leu
 450 455 460
 Ser Val Val Arg Leu Met Phe Thr Ala Phe Leu Pro Asp Ser Thr Gly
 465 470 475 480
 Ser Phe Thr Arg Arg Leu Glu Pro Val Val Ser Asp Ala Ile Tyr Asp
 485 490 495
 Ser Lys Ala Pro Asn Ala Ser Asn Leu Lys Ile Val Arg Met Asp Arg
 500 505 510
 Thr Ala Gly Cys Val Thr Gly Gly Glu Glu Ile Tyr Leu Leu Cys Asp
 515 520 525
 Lys Val Gln Lys Asp Asp Ile Gln Ile Arg Phe Tyr Glu Glu Glu
 530 535 540
 Asn Gly Gly Val Trp Glu Gly Phe Gly Asp Phe Ser Pro Thr Asp Val
 545 550 555 560
 His Arg Gln Phe Ala Ile Val Phe Lys Thr Pro Lys Tyr Lys Asp Ile
 565 570 575
 Asn Ile Thr Lys Pro Ala Ser Val Phe Val Gln Leu Arg Arg Lys Ser
 580 585 590
 Asp Leu Glu Thr Ser Glu Pro Lys Pro Phe Leu Tyr Tyr Pro Glu Ile
 595 600 605
 Lys Asp Lys Glu Glu Val Gln Arg Lys Arg Gln Lys Leu Met Pro Asn
 610 615 620
 Phe Ser Asp Ser Phe Gly Gly Ser Gly Ala Gly Ala Gly Gly
 625 630 635 640
 Gly Met Phe Gly Ser Gly Gly Gly Gly Thr Gly Ser Thr Gly
 645 650 655
 Pro Gly Tyr Ser Phe Pro His Tyr Gly Phe Pro Thr Tyr Gly Ile
 660 665 670
 Thr Phe His Pro Gly Thr Thr Lys Ser Asn Ala Gly Met Lys His Gly
 675 680 685
 Thr Met Asp Thr Glu Ser Lys Lys Asp Pro Glu Gly Cys Asp Lys Ser
 690 695 700
 Asp Asp Lys Asn Thr Val Asn Leu Phe Gly Lys Asp Pro Arg Gly Ser
 705 710 715 720
 Leu Ser Gly Gly Thr Gly Ser Gly Pro Arg Pro Arg Gly Thr Arg
 725 730 735
 Gly Lys Gly Arg Arg Ile Arg Arg Arg Gly
 740 745

<210> SEQ ID NO 90
 <211> LENGTH: 743
 <212> TYPE: PRT
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: mCherry- NFkB-tDeg

<400> SEQUENCE: 90

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

US 12,391,948 B2

297**298**

-continued

Ala Lys Leu Lys Val Thr Lys Gly Gly Pro Leu Pro Phe Ala Trp Asp
 50 55 60

Ile Leu Ser Pro Gln Phe Met Tyr Gly Ser Lys Ala Tyr Val Lys His
 65 70 75 80

Pro Ala Asp Ile Pro Asp Tyr Leu Lys Leu Ser Phe Pro Glu Gly Phe
 85 90 95

Lys Trp Glu Arg Val Met Asn Phe Glu Asp Gly Gly Val Val Thr Val
 100 105 110

Thr Gln Asp Ser Ser Leu Gln Asp Gly Glu Phe Ile Tyr Lys Val Lys
 115 120 125

Leu Arg Gly Thr Asn Phe Pro Ser Asp Gly Pro Val Met Gln Lys Lys
 130 135 140

Thr Met Gly Trp Glu Ala Ser Ser Glu Arg Met Tyr Pro Glu Asp Gly
 145 150 155 160

Ala Leu Lys Gly Glu Ile Lys Gln Arg Leu Lys Leu Lys Asp Gly Gly
 165 170 175

His Tyr Asp Ala Glu Val Lys Thr Thr Tyr Lys Ala Lys Lys Pro Val
 180 185 190

Gln Leu Pro Gly Ala Tyr Asn Val Asn Ile Lys Leu Asp Ile Thr Ser
 195 200 205

His Asn Glu Asp Tyr Thr Ile Val Glu Gln Tyr Glu Arg Ala Glu Gly
 210 215 220

Arg His Ser Thr Gly Gly Met Asp Glu Leu Tyr Lys Gly Gly Ser Gly
 225 230 235 240

Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Thr Gly Ala Glu Asp Asp
 245 250 255

Pro Tyr Leu Gly Arg Pro Glu Gln Met Phe His Leu Asp Pro Ser Leu
 260 265 270

Thr His Thr Ile Phe Asn Pro Glu Val Phe Gln Pro Gln Met Ala Leu
 275 280 285

Pro Thr Ala Asp Gly Pro Tyr Leu Gln Ile Leu Glu Gln Pro Lys Gln
 290 295 300

Arg Gly Phe Arg Phe Arg Tyr Val Cys Glu Gly Pro Ser His Gly Gly
 305 310 315 320

Leu Pro Gly Ala Ser Ser Glu Lys Asn Lys Lys Ser Tyr Pro Gln Val
 325 330 335

Lys Ile Cys Asn Tyr Val Gly Pro Ala Lys Val Ile Val Gln Leu Val
 340 345 350

Thr Asn Gly Lys Asn Ile His Leu His Ser Leu Val Gly Lys
 355 360 365

His Cys Glu Asp Gly Ile Cys Thr Val Thr Ala Gly Pro Lys Asp Met
 370 375 380

Val Val Gly Phe Ala Asn Leu Gly Ile Leu His Val Thr Lys Lys Lys
 385 390 395 400

Val Phe Glu Thr Leu Glu Ala Arg Met Thr Glu Ala Cys Ile Arg Gly
 405 410 415

Tyr Asn Pro Gly Leu Leu Val His Pro Asp Leu Ala Tyr Leu Gln Ala
 420 425 430

Glu Gly Gly Asp Arg Gln Leu Gly Asp Arg Glu Lys Glu Leu Ile
 435 440 445

Arg Gln Ala Ala Leu Gln Gln Thr Lys Glu Met Asp Leu Ser Val Val
 450 455 460

Arg Leu Met Phe Thr Ala Phe Leu Pro Asp Ser Thr Gly Ser Phe Thr

US 12,391,948 B2

299**300**

-continued

465	470	475	480
Arg Arg Leu Glu Pro Val Val Ser Asp Ala Ile Tyr Asp Ser Lys Ala			
485	490	495	
Pro Asn Ala Ser Asn Leu Lys Ile Val Arg Met Asp Arg Thr Ala Gly			
500	505	510	
Cys Val Thr Gly Gly Glu Glu Ile Tyr Leu Leu Cys Asp Lys Val Gln			
515	520	525	
Lys Asp Asp Ile Gln Ile Arg Phe Tyr Glu Glu Glu Asn Gly Gly			
530	535	540	
Val Trp Glu Gly Phe Gly Asp Phe Ser Pro Thr Asp Val His Arg Gln			
545	550	555	560
Phe Ala Ile Val Phe Lys Thr Pro Lys Tyr Lys Asp Ile Asn Ile Thr			
565	570	575	
Lys Pro Ala Ser Val Phe Val Gln Leu Arg Arg Lys Ser Asp Leu Glu			
580	585	590	
Thr Ser Glu Pro Lys Pro Phe Leu Tyr Tyr Pro Glu Ile Lys Asp Lys			
595	600	605	
Glu Glu Val Gln Arg Lys Arg Gln Lys Leu Met Pro Asn Phe Ser Asp			
610	615	620	
Ser Phe Gly Gly Ser Gly Ala Gly Ala Gly Gly Gly Met Phe			
625	630	635	640
Gly Ser Gly Gly Gly Gly Thr Gly Ser Thr Gly Pro Gly Tyr			
645	650	655	
Ser Phe Pro His Tyr Gly Phe Pro Thr Tyr Gly Gly Ile Thr Phe His			
660	665	670	
Pro Gly Thr Thr Lys Ser Asn Ala Gly Met Lys His Gly Thr Met Asp			
675	680	685	
Thr Glu Ser Lys Lys Asp Pro Glu Gly Cys Asp Lys Ser Asp Asp Lys			
690	695	700	
Asn Thr Val Asn Leu Phe Gly Lys Asp Pro Arg Gly Ser Leu Ser Gly			
705	710	715	720
Gly Thr Gly Gly Ser Gly Pro Arg Pro Arg Gly Thr Arg Gly Lys Gly			
725	730	735	
Arg Arg Ile Arg Arg Arg Gly			
740			

<210> SEQ ID NO 91

<211> LENGTH: 597

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: EGFP-TurboID-tDeg

<400> SEQUENCE: 91

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

US 12,391,948 B2

301**302**

-continued

85	90	95
Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu		
100	105	110
Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly		
115	120	125
Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr		
130	135	140
Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn		
145	150	155
Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser		
165	170	175
Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly		
180	185	190
Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Lys Leu		
195	200	205
Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe		
210	215	220
Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys Gly		
225	230	235
Thr Gly Ala Cys Gly Thr Ser Gly Gly Met Lys Asp Asn Thr Val Pro		
245	250	255
Leu Lys Leu Ile Ala Leu Leu Ala Asn Gly Glu Phe His Ser Gly Glu		
260	265	270
Gln Leu Gly Glu Thr Leu Gly Met Ser Arg Ala Ala Ile Asn Lys His		
275	280	285
Ile Gln Thr Leu Arg Asp Trp Gly Val Asp Val Phe Thr Val Pro Gly		
290	295	300
Lys Gly Tyr Ser Leu Pro Glu Pro Ile Pro Leu Leu Asn Ala Lys Gln		
305	310	315
Ile Leu Gly Gln Leu Asp Gly Gly Ser Val Ala Val Leu Pro Val Val		
325	330	335
Asp Ser Thr Asn Gln Tyr Leu Leu Asp Arg Ile Gly Glu Leu Lys Ser		
340	345	350
Gly Asp Ala Cys Ile Ala Glu Tyr Gln Gln Ala Gly Arg Gly Ser Arg		
355	360	365
Gly Arg Lys Trp Phe Ser Pro Phe Gly Ala Asn Leu Tyr Leu Ser Met		
370	375	380
Phe Trp Arg Leu Lys Arg Gly Pro Ala Ala Ile Gly Leu Gly Pro Val		
385	390	395
Ile Gly Ile Val Met Ala Glu Ala Leu Arg Lys Leu Gly Ala Asp Lys		
405	410	415
Val Arg Val Lys Trp Pro Asn Asp Leu Tyr Leu Gln Asp Arg Lys Leu		
420	425	430
Ala Gly Ile Leu Val Glu Leu Ala Gly Ile Thr Gly Asp Ala Ala Gln		
435	440	445
Ile Val Ile Gly Ala Gly Ile Asn Val Ala Met Arg Arg Val Glu Glu		
450	455	460
Ser Val Val Asn Gln Gly Trp Ile Thr Leu Gln Glu Ala Gly Ile Asn		
465	470	475
Leu Asp Arg Asn Thr Leu Ala Ala Thr Leu Ile Arg Glu Leu Arg Ala		
485	490	495
Ala Leu Glu Leu Phe Glu Gln Glu Gly Leu Ala Pro Tyr Leu Pro Arg		
500	505	510

-continued

Trp Glu Lys Leu Asp Asn Phe Ile Asn Arg Pro Val Lys Leu Ile Ile
515 520 525

Gly Asp Lys Glu Ile Phe Gly Ile Ser Arg Gly Ile Asp Lys Gln Gly
530 535 540

Ala Leu Leu Leu Glu Gln Asp Gly Val Ile Lys Pro Trp Met Gly Gly
545 550 555 560

Glu Ile Ser Leu Arg Ser Ala Glu Lys Gly Ser Gly Thr Gly Gly Thr
565 570 575

Gly Gly Ser Gly Pro Arg Pro Arg Gly Thr Arg Gly Lys Gly Arg Arg
580 585 590

Ile Arg Arg Arg Gly
595

<210> SEQ ID NO 92
<211> LENGTH: 524
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: EGFP-APEX-tDeg

<400> SEQUENCE: 92

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

Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
20 25 30

Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
35 40 45

Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
50 55 60

Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys
65 70 75 80

Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
85 90 95

Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
100 105 110

Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
115 120 125

Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
130 135 140

Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
145 150 155 160

Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
165 170 175

Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
180 185 190

Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Lys Leu
195 200 205

Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
210 215 220

Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys Gly
225 230 235 240

Thr Gly Ala Cys Gly Thr Ser Gly Lys Ser Tyr Pro Thr Val Ser Ala
245 250 255

Asp Tyr Gln Asp Ala Val Glu Lys Ala Lys Lys Leu Arg Gly Phe
260 265 270

-continued

Ile Ala Glu Lys Arg Cys Ala Pro Leu Met Leu Arg Leu Ala Phe His
275 280 285

Ser Ala Gly Thr Phe Asp Lys Gly Thr Lys Thr Gly Gly Pro Phe Gly
290 295 300

Thr Ile Lys His Pro Ala Glu Leu Ala His Ser Ala Asn Asn Gly Leu
305 310 315 320

Asp Ile Ala Val Arg Leu Leu Glu Pro Leu Lys Ala Glu Phe Pro Ile
325 330 335

Leu Ser Tyr Ala Asp Phe Tyr Gln Leu Ala Gly Val Val Ala Val Glu
340 345 350

Val Thr Gly Gly Pro Lys Val Pro Phe His Pro Gly Arg Glu Asp Lys
355 360 365

Pro Glu Pro Pro Glu Gly Arg Leu Pro Asp Pro Thr Lys Gly Ser
370 375 380

Asp His Leu Arg Asp Val Phe Gly Lys Ala Met Gly Leu Thr Asp Gln
385 390 395 400

Asp Ile Val Ala Leu Ser Gly Gly His Thr Ile Gly Ala Ala His Lys
405 410 415

Glu Arg Ser Gly Phe Glu Gly Pro Trp Thr Ser Asn Pro Leu Ile Phe
420 425 430

Asp Asn Ser Tyr Phe Thr Glu Leu Leu Ser Gly Glu Lys Glu Gly Leu
435 440 445

Leu Gln Leu Pro Ser Asp Lys Ala Leu Leu Ser Asp Pro Val Phe Arg
450 455 460

Pro Leu Val Asp Lys Tyr Ala Ala Asp Glu Asp Ala Phe Phe Ala Asp
465 470 475 480

Tyr Ala Glu Ala His Gln Lys Leu Ser Glu Leu Gly Phe Ala Asp Ala
485 490 495

Gly Ser Gly Thr Gly Gly Ser Gly Pro Arg Pro Arg Gly
500 505 510

Thr Arg Gly Lys Gly Arg Arg Ile Arg Arg Arg Gly
515 520

<210> SEQ ID NO 93
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: miniCMV promoter

<400> SEQUENCE: 93

ggtaggcgtg tacgggtggga ggcctatata agcagagct

39

<210> SEQ ID NO 94
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: MS2 hairpin

<400> SEQUENCE: 94

acatgaggat cacccatgt

19

<210> SEQ ID NO 95
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:

-continued

<223> OTHER INFORMATION: EYFP fw probe

<400> SEQUENCE: 95

acgtaaacgg ccacaaggttc

20

<210> SEQ ID NO 96

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: EYFP rv probe

<400> SEQUENCE: 96

cttcatgtgg tcggggtagc

20

<210> SEQ ID NO 97

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: mCherry fw probe

<400> SEQUENCE: 97

cacgagttcg agatcgaggg

20

<210> SEQ ID NO 98

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: mCherry rv probe

<400> SEQUENCE: 98

caagtatgtcg gggatgtcgg

20

<210> SEQ ID NO 99

<211> LENGTH: 28

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: Probe-1

<400> SEQUENCE: 99

gttgagtgat tagcgattga ttccggcc

28

<210> SEQ ID NO 100

<211> LENGTH: 30

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: Probe-2

<400> SEQUENCE: 100

gtcggatgat tttcgtaata gattgcgcgt

30

<210> SEQ ID NO 101

<211> LENGTH: 29

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: Probe-3

<400> SEQUENCE: 101

ttgacgtgat tttgtgagat ttccgcag

29

-continued

<210> SEQ ID NO 102
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Probe-4

<400> SEQUENCE: 102

tgccctgattt taagtatgtg gattatcg

29

<210> SEQ ID NO 103
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Probe-5

<400> SEQUENCE: 103

ggataggtat ggaggaagta gcttgaa

27

<210> SEQ ID NO 104
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Probe-6

<400> SEQUENCE: 104

acaatatatctt gcgcgcgttcg atcttg

26

<210> SEQ ID NO 105
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Probe-7

<400> SEQUENCE: 105

ggccgcctaag aagaacgacc aa

22

<210> SEQ ID NO 106
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Probe-8

<400> SEQUENCE: 106

cctaagaacc taacatatct agcgagg

27

<210> SEQ ID NO 107
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Probe-9

<400> SEQUENCE: 107

tgtgcacattt gaagcgcatg aa

22

<210> SEQ ID NO 108
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Probe-10

-continued

<400> SEQUENCE: 108

cctgggtcac ggtcaccacg

20

<210> SEQ ID NO 109
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Probe-11

<400> SEQUENCE: 109

gccccatggtc ttcttctgc

19

<210> SEQ ID NO 110
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Probe-12

<400> SEQUENCE: 110

gggtgcttca cgtaggcctt

20

<210> SEQ ID NO 111
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Probe-13

<400> SEQUENCE: 111

gtcaccttca gtttggcggt c

21

<210> SEQ ID NO 112
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Probe-14

<400> SEQUENCE: 112

gcctctgtttt gatctcgccc ttc

23

<210> SEQ ID NO 113
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Probe-15

<400> SEQUENCE: 113

gtcttgacct cagcgctcgta gtg

23

<210> SEQ ID NO 114
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Probe-16

<400> SEQUENCE: 114

cggcgcggttc gtactgttcc

20

<210> SEQ ID NO 115
<211> LENGTH: 28

-continued

<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Probe-17

<400> SEQUENCE: 115

gccgataatc cacatactta caatcagg

28

<210> SEQ ID NO 116
<211> LENGTH: 43
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Linker sequence

<400> SEQUENCE: 116

Met Asp Pro Val Val Val Leu Gly Leu Cys Leu Ser Cys Leu Leu
1 5 10 15

Leu Ser Leu Trp Lys Gln Ser Tyr Gly Gly Lys Leu Gly Gly Ser
20 25 30

Gly Gly Thr Gly Ser Gly Thr Ser Gly Gly
35 40

<210> SEQ ID NO 117
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Destabilization domain

<400> SEQUENCE: 117

Arg Arg Arg Gly
1

<210> SEQ ID NO 118
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: miniCMV promoter

<400> SEQUENCE: 118

ggtaggctgt tacggtggga ggcctatata agcagagct

39

<210> SEQ ID NO 119
<211> LENGTH: 586
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: (Pepper)10 tag sequence

<400> SEQUENCE: 119

ggctcgctg	agctcattag	ctccgagccg	tccagcgcaa	actattacga	aaaacatccg	60
acggggctcgt	tgagctcatt	agctccgago	ccgctgcca	aaacacctaca	aaaacacgac	120
aaacacggctc	gttgagctca	ttagctccga	gcccccccac	aacccacaaa	cttacaacca	180
ggcaaacggc	tcgtctgagc	tcattagctc	cgagccgtat	caagacogaa	cggcgcaaga	240
tattgacacg	ggctcggtga	gctcattago	tccgagcccg	acctcgctag	atatgttagg	300
ttctttagca	ttggctcggt	gagctcatta	gctccgagcc	aaagatcgac	tgcaattccg	360
attagacgta	cacggctcgt	ctgagctcat	tagctccgag	ccgatccaaac	ctacttcctc	420
cataactaac	ctccggctcg	ttgagctcat	tagctccgag	ccgatcataa	cgcaataaccg	480

-continued

tacactgtcc aatccggctc gttgagctca ttagctccga gccggacaac caatcgacat	540
acatcacacc acaactcgcc tcgtctgago tcattagctc cgagcc	586

<210> SEQ ID NO 120
<211> LENGTH: 1466
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: (F30-1xPepper)10 tag sequence

<400> SEQUENCE: 120

ttgccccatgtt tatgtggat gcggttgcac gtttcccaca tactctgtat atccgctatc	60
aaaggctcgat ctgagctcat tagctccgag cccgaggatc cgatcatc atggcaatgc	120
cagcgcaatc tattacgaaa atcatccgac gtgcgtatc ctatgcggga tgccgttgcac	180
cgttcccgat atagtctgtat catccgctcg caaaggctcg ttgagctcat tagctccgag	240
cccgaggatc cgatgttgc acgcgttgc tgcggaaaat ctcacaaaaat cacgtcaaac	300
gtcgccgtgt gtgtgttaga tgccgttgcac cgttccctac acactctgtat gatccgctatc	360
caaaggctcg ttgagctcat tagctccgag cccgaggatc cgatgttgc acggcgacgc	420
cgataatcca catacttaca atcaggcaat ctgcgtatgt gtatgtggga tgccgttgcac	480
cgttccac atactctgtat gatccgctatc caaaggctcg ttgagctcat tagctccgag	540
cccgaggatc cgatcatc atggcaatgc tcaagatgc acggcgcaag atattgtcac	600
gtcggtatc ctatgcggga tgccgttgcac cgttcccgat atagtctgtat catccgctatc	660
caaaggctcg tctgagctca ttatgcgtatc gccccggatc ccggatgttgc acgtcgacgc	720
tcctcgctatc atatgtttagg ttctttaggc tttcgccgtg tggtgttagg atgcgttgcac	780
acgtttccatc cacactctgtatc cgatccgatc gcaaggctcg gttgagctca ttatgcgtatc	840
gccccggatc ccggatgttgc ttcacggatc agatcgatgc caattccgtat tagacgtaca	900
cttgccatgtt gtatgtggga tgccgttgcac cgttccac atactctgtat gatccgctatc	960
caaaggctcg ttgagctcat tagctccgag cccgaggatc cgatcatc atggcaatgc	1020
ccaaaggctact tcctccatc ctatccctatc cgatgttgc acgtcgatgtatc cggttgcac	1080
tttcccgat atgttgcgtatc tccgtatgc aaggctcgatc gagctatgc gtcggagcc	1140
cgaggatccatc gatgttgcgtatc cgatgttgc acgtcgatgtatc cggttgcac	1200
gccccggatc tgtaggtatc gttgccacgt ttatgcgtatc ctatgcgtatc ccgtatgc	1260
aggctcgatc gagctatgc gtcggagcc cgaggatccatc gatgttgc acgtcgatgtatc	1320
atcaatccatc atacatcaca ccacaattct tgcgtatgc atgtggatc cggttgcac	1380
tttcccgatcatc atgttgcgtatc tccgtatgc aaggctcgatc tgtagcttgc acgtccgac	1440
cccgaggatc ggatcattca tggcaatgc	1466

<210> SEQ ID NO 121
<211> LENGTH: 1228
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: (Pepper)20 tag sequence

<400> SEQUENCE: 121

ggatcgatgttgc acgtcgatgtatc tccgtatgc gtcggagcc cgaggatccatc	60
acggggctcgat tgtagcttgc acgtcgatgtatc tccgtatgc gtcggagcc	120
aaacccggatc gttgagctca ttatgcgtatc gccccggatc ccgtatgc acgtcgatgtatc	180

-continued

ggcaaacggc tcgtctgagc tcattagctc cgagccgtat caagaccgaa cggcgcaaga	240
tattgacacg ggctcggtga gtcattago tccgagcccg acctcgctag atatgttagg	300
ttctttaggca ttggctcggt gagctcatta gctccgagcc aaagatcgac tgcaattccg	360
attagacgta cacggctcggt ctgagctcat tagtccgag ccgatccaac ctacttcctc	420
cataactaac ctccggctcg ttgagctcat tagtccgag ccgatcataa cgcaataccg	480
taacactgtcc aatccggctc gttgagctca ttagtccgag gccggacaac caatcgacat	540
acatcacacc acaactcgcc tcgtctgagc tcattagctc cgagccgaat tggcggtct	600
tcttggggcc cgctcgacta aggtgacaac tggacaaacc ctccggctcg tgagctcatt	660
agctccgagc cgactctcac caacaagaca aaaactactc ttctaggctc gttgagctca	720
ttagctcoga gcctaaacac tcaagcatac attgtgccta tttcttgct cgtctgagct	780
cattagctcc gagccatgct ctcacgaatt tcaaaaacacg gacaaggggc tcggttagct	840
cattagctcc gagcccggttc cacgtccaat acgattactt acctttcggt ctcgttgagc	900
tcattagctc cgagcccgca gctacatcac ttccactcag gacattcaag ggctcgctg	960
agotcattag ctcccgagcc tccacaagtc tcaaccacag aaactaccaa atgggctcg	1020
ttagctcatt agctccgagc ccactctcac ctcaaaacctc ttcccacaaa actggggctc	1080
gtttagctca ttagctccga gccccatcc caacataccca aataaaaac aattactggc	1140
tcgtctgagc tcattagctc cgagccagec cacatctctc actactatca aaaaccaaac	1200
ggctcggtga gtcattagc tccgagcc	1228

<210> SEQ ID NO 122
 <211> LENGTH: 1812
 <212> TYPE: DNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: (F30-1xPepper)10 tag sequence

<400> SEQUENCE: 122	
ttgccatgttatgttggaa gcttagaaag gctcggttag ctcattagct ccgagccga	60
ctacgtttcc cacatactct gatgatccgc tagcaaaggc tcgtctgagc tcattagctc	120
cgagcccgag gtacccggatc attcatggca agtccagegc aatctattac gaaaatcatc	180
cgacgtcgcg atgtctatgc gggaaagcgta gaaaggctcg tctgagctca ttagctccga	240
gccccactac gttcccgca tagtctgatc atccgctagc aaaggctcg tgagctcatt	300
agctccgagc ccgagggtacc ggatgattca tcgcgacgct gcgaaaatc tcacaaaatc	360
acgtcaaacg tcgcccgtgt tttgttaggaa gcttagaaag gctcgctgatc gtcattagc	420
tccgagcccg actacgttcc ctacacactc tgacgatccg ctagcaagg ctcgttgagc	480
tcattagctc cgagcccgag gtacccggatc gttcacggcc acgcccataa tccacatact	540
tacaatcagg caatcttgcg atgtgtatgt gggaaagcgta gaaaggctcg ttgagctcat	600
tagctccgag cccgactacg tttccacat actctgtatc tccgctagca aaggctcg	660
gagctcatttta gtcggccatc cgaggatccg gatcattcat ggcaagtatc aagatcgaaac	720
ggcgcaagat attgtcacgt cgcgatgtatc atgcggaaag cgtagaaagg ctcgttgagc	780
tcattagctc cgagcccgac tacgtttccgc gcatagtctg atcatccgct agcaaaggct	840
cgtctgagct cattagctcc gagcccgagg taccggatca ttcatcgca cgtccctcgct	900
agatatgtta gggtttagg catttcggccg tttgtgtgtaa ggaagcgtag aaaggctcg	960

US 12,391,948 B2

319

320

-continued

ttagctcatt agctccgagc ccgactacgt ttccctacaca ctctgacgat ccgcgtacaa	1020
aggctcgct gagctcatta gctccgagcc cgaggtaccc gatcgttcac ggccaaaaaga	1080
tcgtctgaa ttccgattag acgtacactt gccatgtgta tgtggaaagc gtagaaaggc	1140
tcgtctgagc tcattagctc cgagccgac tacgttccc acatactctg atgatccgct	1200
agcaaaggct cgtttagctc attagctcg agcccgaggt accggatcat tcatggcaag	1260
atccaagcta cttccatccat acctatcctc ctcgcgtatg ctatgeggga agcgtagaaa	1320
ggctcgctg agctcattag ctccgagcc gactacgtt cccgcatagt ctgatcatcc	1380
gctagcaaag gctcggttag ctcattagct ccgagcccg ggtaccggat gattcatcgc	1440
gagatcataa cgcaataccg tacactgtcc aatcctcgcc gtgtgtgtt aggaagcgtta	1500
gaaaaggctcg tctgagctca tttagtccga gccccactac gtttcttaca cactctgacg	1560
atcccgtagc aaaggctcg ttagctcatt agctccgagc ccgaggtacc ggatcggtca	1620
cggcgaggat aatcaatcca catacatcac accacaattc ttgccatgtg tatgtggaa	1680
gcgtagaaag gctcgctga gctcattago tccgagcccg actacgttcc ccacatactc	1740
tgatgatccg cttagcaaagg ctcgtctgag ctcattagct ccgagcccg ggtaccggat	1800
cattcatggc aa	1812

<210> SEQ ID NO 123

<211> LENGTH: 22

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: LambdaN RNA-binding domain

<400> SEQUENCE: 123

Met Asp Ala Arg Thr Arg Arg Arg Glu Arg Arg Ala Glu Lys Gln Ala			
1	5	10	15

Gln Trp Lys Ala Ala Asn	
20	

<210> SEQ ID NO 124

<211> LENGTH: 19

<212> TYPE: RNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: BoxB RNA-binding domain

<400> SEQUENCE: 124

gggcccugaa gaagggccc 19

<210> SEQ ID NO 125

<211> LENGTH: 23

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: HIV-1 Rev RNA-binding domain

<400> SEQUENCE: 125

Asp Thr Arg Gln Ala Arg Arg Asn Arg Arg Arg Arg Trp Arg Glu Arg			
1	5	10	15

Gln Arg Ala Ala Ala Ala Arg	
20	

<210> SEQ ID NO 126

<211> LENGTH: 32

<212> TYPE: RNA

<213> ORGANISM: Artificial

-continued

<220> FEATURE:
<223> OTHER INFORMATION: RRE RNA-binding domain

<400> SEQUENCE: 126

ggucuggggcg cagcgcaagg ugcggacagg cc

32

<210> SEQ ID NO 127
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: HIV Tat -RRRG

<400> SEQUENCE: 127

Arg Lys Lys Arg Arg Gln Arg Arg Arg Gly
1 5 10

<210> SEQ ID NO 128
<211> LENGTH: 37
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: HIV TAR RNA

<400> SEQUENCE: 128

acgaagcuug aucccguaug ccggucgauc gcuucga

37

<210> SEQ ID NO 129
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Tat peptide

<400> SEQUENCE: 129

Arg Lys Lys Arg Arg Gln Arg Arg Arg
1 5

<210> SEQ ID NO 130
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Degron peptide

<400> SEQUENCE: 130

Arg Arg Arg Gly
1

50

What is claimed:

1. A nucleic acid molecule encoding an RNA-regulated fusion protein, said nucleic acid molecule comprising:
a first nucleic acid sequence encoding a protein of interest and a second nucleic acid sequence encoding an RNA-regulated destabilization domain,
wherein the second nucleic acid sequence is operably coupled to the first nucleic acid sequence,
wherein the RNA-regulated destabilization domain is a bifunctional peptide comprising:
a lentiviral transactivator of transcription (Tat) peptide and
a degron peptide,
wherein an RNA aptamer interacts with the RNA-regulated destabilization domain to stabilize the protein of interest, and

55 wherein the RNA-regulated destabilization domain is tDeg as set forth in SEQ ID NO: 63.

2. The nucleic acid molecule according to claim 1, wherein the protein of interest is a fluorescent protein, a bioluminescent protein, an enzyme, or a transcription factor.

3. The nucleic acid molecule according to claim 1, wherein the lentiviral transactivator of transcription (Tat) peptide comprises an RNA binding site corresponding to or amino acid residues 4-17 of SEQ ID NO: 55.

60 4. The nucleic acid molecule according to claim 1 further comprising:

a third nucleic acid sequence encoding a second protein of interest, wherein the third nucleic acid sequence is located between the first nucleic acid sequence and second nucleic acid sequence.

65 5. A vector comprising the nucleic acid molecule according to claim 1.

323

6. An expression system comprising an expression vector into which is inserted the nucleic acid molecule according to claim 1.

7. A host cell comprising the nucleic acid molecule of according to claim 1. 5

8. An RNA-regulated fusion protein encoded by the nucleic acid molecule according to claim 1.

9. A molecular complex comprising:
an RNA-regulated fusion protein encoded by the nucleic acid molecule according to claim 1 comprising 10
(i) a protein of interest and
(ii) an RNA-regulated destabilization domain; and
an RNA aptamer bound specifically to the RNA-regulated destabilization domain.

10. A host cell containing the molecular complex accord- 15
ing to claim 9.

324

* * * * *