(12) STANDARD PATENT (11) Application No. **AU 2020207840 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
RNA-GUIDED TARGETING OF GENETIC AND EPIGENOMIC REGULATORY PROTEINS TO SPECIFIC GENOMIC LOCI

(51) International Patent Classification(s)
C12N 15/62 (2006.01)

(21) Application No: **2020207840** (22) Date of Filing: **2020.07.23**

(43) Publication Date: **2020.08.13**
(43) Publication Journal Date: **2020.08.13**
(44) Accepted Journal Date: **2022.07.21**

(62) Divisional of:
2014239665

(71) Applicant(s)
The General Hospital Corporation

(72) Inventor(s)
Joung, J., Keith;Maeder, Morgan

(74) Agent / Attorney
RnB IP Pty Ltd, PO Box 9530, DEAKIN, ACT, 2600, AU

(56) Related Art
WO 2013176772 A1
WO 2014093655 A2

ABSTRACT

Methods and constructs for RNA-guided targeting of heterologous functional domains such as transcriptional activators to specific genomic loci. This invention relates to methods and constructs for RNA-guided targeting of genetic and epigenomic regulatory proteins, e.g., transcriptional activators, histone modification enzymes, DNA methylation modifiers, to specific genomic loci. At least in part, the present invention is based on the development of a fusion protein including a heterologous functional domain (e.g., a transcriptional activation domain) fused to a Cas9 nuclease that has had its nuclease activity inactivated by mutations (also known as "dCas9"). See Fig. 1C.

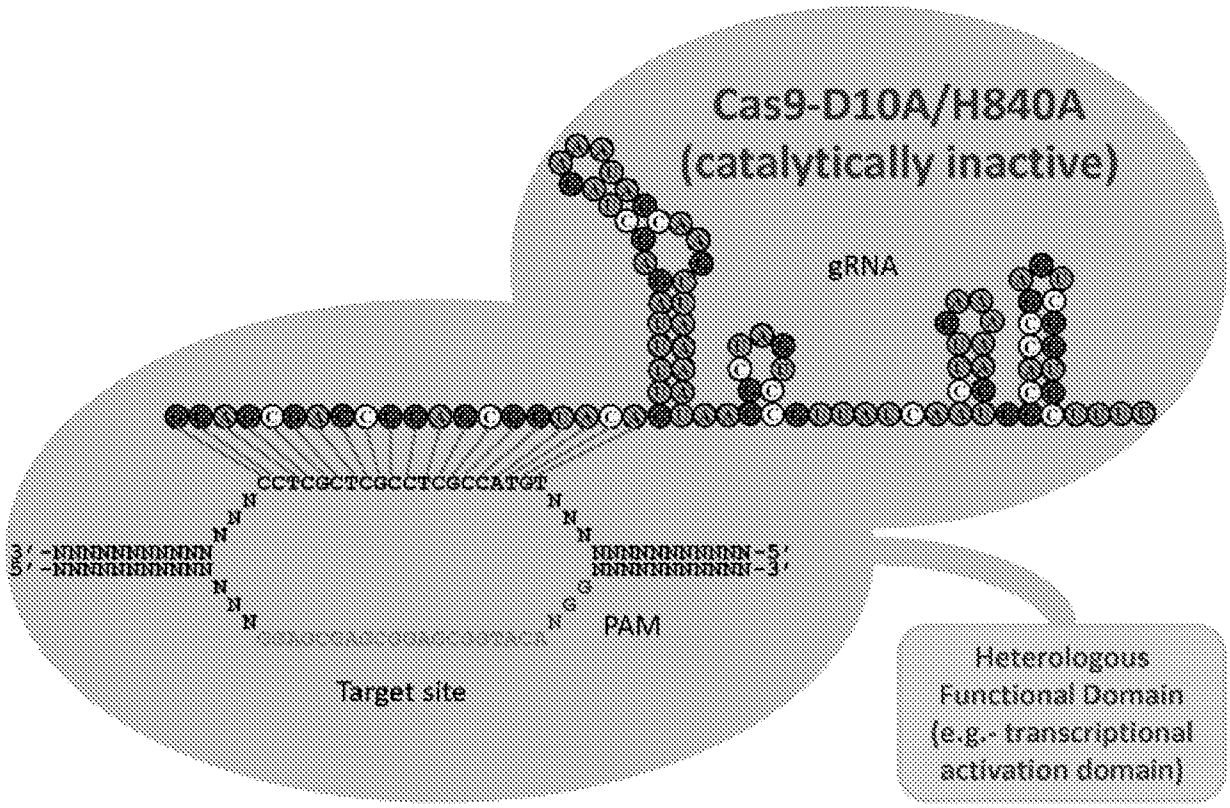


FIG. 1C

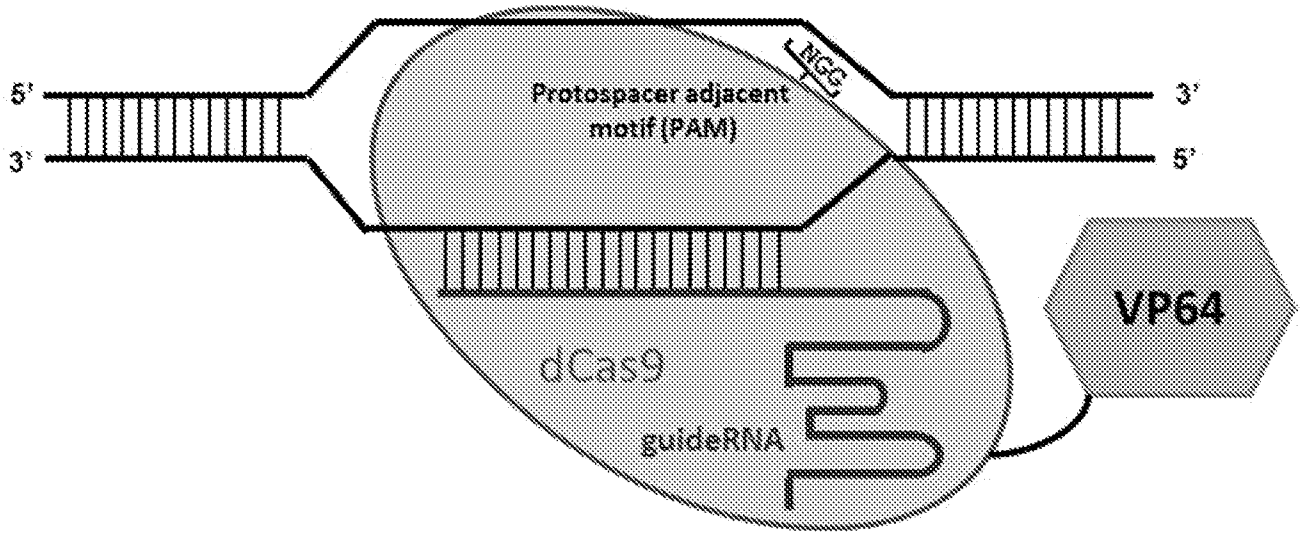


FIG. 1D

RNA-GUIDED TARGETING OF GENETIC AND EPIGENOMIC REGULATORY PROTEINS TO SPECIFIC GENOMIC LOCI

CLAIM OF PRIORITY

This application claims the benefit of U.S. Patent Application Serial Nos. 61/799,647, filed on March 15, 2013; 61/838,178, filed on June 21, 2013; 61/838,148, filed on June 21, 2013; and 61/921,007, filed on December 26, 2013. The entire contents of the foregoing are hereby incorporated by reference.

FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

This invention was made with Government support under Grant No. DP1GM105378 awarded by the National Institutes of Health and W911NF-11-2-0056 awarded by the Defense Advanced Research Projects Agency (DARPA) of the Department of Defense. The Government has certain rights in the invention.

TECHNICAL FIELD

This invention relates to methods and constructs for RNA-guided targeting of genetic and epigenomic regulatory proteins, e.g., transcriptional activators, histone modification enzymes, DNA methylation modifiers, to specific genomic loci.

BACKGROUND

Clustered Regulatory Interspaced Short Palindromic Repeats (CRISPR), and CRISPR-associated (cas) genes, referred to as CRISPR/Cas systems, are used by various bacteria and archaea to mediate defense against viruses and other foreign nucleic acid. These systems use small RNAs to detect and silence foreign nucleic acids in a sequence-specific manner.

Three types of CRISPR/Cas systems have been described (Makarova et al., Nat. Rev. Microbiol. 9, 467 (2011); Makarova et al., Biol. Direct 1, 7 (2006); Makarova et al., Biol. Direct 6, 38 (2011)). Recent work has shown that Type II CRISPR/Cas systems can be engineered to direct targeted double-stranded DNA breaks in vitro to specific sequences by using a single "guide RNA" with complementarity to the DNA target site and a Cas9 nuclease (Jinek et al., Science 2012; 337:816–821). This targetable Cas9-based system also works in cultured human cells (Mali et al., Science. 2013 Feb 15;339(6121):823-6; Cong et al., Science. 2013 Feb 15;339(6121):819-23) and in vivo in zebrafish (Hwang and Fu et al.,

Nat Biotechnol. 2013 Mar;31(3):227-9) for inducing targeted alterations into endogenous genes.

SUMMARY

At least in part, the present invention is based on the development of a fusion protein including a heterologous functional domain (e.g., a transcriptional activation domain) fused to a Cas9 nuclease that has had its nuclease activity inactivated by mutations (also known as "dCas9"). While published studies have used guide RNAs to target catalytically active and inactive Cas9 nuclease proteins to specific genomic loci, no work has yet adapted the use of this system to recruit additional effector domains. This work also provides the first demonstration of an RNA-guided process that results in an increase (rather than a decrease) in the level of expression of a target gene.

In addition, the present disclosure provides the first demonstration that multiplex gRNAs can be used to bring multiple dCas9-VP64 fusions to a single promoter, thereby resulting in synergistic activation of transcription.

Thus, in a first aspect, the invention provides fusion proteins comprising a catalytically inactive CRISPR associated 9 (dCas9) protein linked to a heterologous functional domain (HFD) that modifies gene expression, histones, or DNA, e.g., transcriptional activation domain, transcriptional repressors (e.g., silencers such as Heterochromatin Protein 1 (HP1), e.g., HP1α or HP1β, or a transcriptional repression domain, e.g., Krueppel-associated box (KRAB) domain, ERF repressor domain (ERD), or mSin3A interaction domain (SID)), enzymes that modify the methylation state of DNA (e.g., DNA methyltransferase (DNMT) or Ten-Eleven Translocation (TET) proteins, e.g., TET1, also known as Tet Methylcytosine Dioxygenase 1), or enzymes that modify histone subunit (e.g., histone acetyltransferases (HAT), histone deacetylases (HDAC), or histone demethylases). In some embodiments, the heterologous functional domain is a transcriptional activation domain, e.g., a transcriptional activation domain from VP64 or NF-κB p65; an enzyme that catalyzes DNA demethylation, e.g., a TET; or histone modification (e.g., LSD1, histone methyltransferase, HDACs, or HATs) or a transcription silencing domain, e.g., from Heterochromatin Protein 1 (HP1), e.g., HP1α or HP1β; or a biological tether, e.g., CRISPR/Cas Subtype Ypest protein 4 (Csy4), MS2,or lambda N protein.

In some embodiments, the catalytically inactive Cas9 protein is from *S. pyogenes*.

In some embodiments, the catalytically inactive Cas9 protein comprises mutations at comprises mutations at D10, E762, H983, or D986; and at H840 or N863, e.g., at D10 and H840, e.g., D10A or D10N and H840A or H840N or H840Y.

In some embodiments, the heterologous functional domain is linked to the N terminus or C terminus of the catalytically inactive Cas9 protein, with an optional intervening linker, wherein the linker does not interfere with activity of the fusion protein.

In some embodiments, the fusion protein includes one or both of a nuclear localization sequence and one or more epitope tags, e.g., c-myc, 6His, or FLAG tags, on the N-terminus, C-terminus, or in between the catalytically inactive CRISPR associated 9 (Cas9) protein and the heterologous functional domain, optionally with one or more intervening linkers.

In further aspect, the invention provides nucleic acids encoding the fusion proteins described herein, as well as expression vectors including the nucleic acids, and host cells expressing the fusion proteins.

In an additional aspect, the invention provides methods for increasing expression of a target gene in a cell. The methods include expressing a Cas9-HFD fusion protein as described herein in the cell, e.g., by contacting the cell with an expression vector including a sequence encoding the fusion protein, and also expressing in the cell one or more guide RNAs with complementarity directed to the target gene, e.g., by contacting the cell with one or more expression vectors comprising nucleic acid sequences encoding one or more guide RNAs.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Methods and materials are described herein for use in the present invention; other, suitable methods and materials known in the art can also be used. The materials, methods, and examples are illustrative only and not intended to be limiting. All publications, patent applications, patents, sequences, database entries, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control.

Other features and advantages of the invention will be apparent from the following detailed description and figures, and from the claims.

DESCRIPTION OF DRAWINGS

The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

FIG. 1A is a schematic illustration showing a single guide RNA (sgRNA) recruiting Cas9 nuclease to a specific DNA sequence and thereby introducing targeted alterations. The sequence of the guide RNA shown is GGAGCGAGCGGAGCGGUACAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUA AGGCUAGUCCG (SEQ ID NO:9)

FIG. 1B is a schematic illustration showing a longer version of the sgRNA used to recruit Cas9 nuclease to a specific DNA sequence and to thereby introduce targeted alterations. The sequence of the guide RNA shown is GGAGCGAGCGGAGCGGUACAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUA AGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU (SEQ ID NO:10).

FIG. 1C is a schematic illustration showing a Cas9 protein containing D10A and H840A mutations to render the nuclease portion of the protein catalytically inactive, fused to a transcriptional activation domain and recruited to a specific DNA sequence by a sgRNA. The sequence of the guide RNA shown is GGAGCGAGCGGAGCGGUACAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUA AGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU (SEQ ID NO:10).

FIG. 1D is a schematic depicting recruitment of dCas9-VP64 fusion protein to a specific genomic target sequence by a chimeric sgRNA.

FIG. 1E is a diagram illustrating the positions and orientations of 16 sgRNAs targeted to the endogenous human VEGFA gene promoter. Small horizontal arrows represent the first 20 nts of the gRNA complementary to the genomic DNA sequence with the arrow pointing 5' to 3'. Grey bars indicate DNaseI hypersensitive sites previously defined in human 293 cells (Liu et al., J Biol Chem. 2001 Apr 6;276(14):11323-34), numbered relative to the transcription start site (right-angle arrow).

FIG. 2A is a bar graph showing activation of VEGFA protein expression in 293 cells by various sgRNAs, each expressed with (grey bars) or without (black bars) dCas9-VP64. Fold-activation of VEGFA was calculated relative to the off-target sgRNA control as described in Methods. Each experiment was performed in triplicate and error bars

represent standard errors of the mean. Asterisks indicate samples that are significantly elevated above the off-target control as determined by a paired, one-sided t-test ($p<0.05$).

FIG. 2B is a bar graph showing multiplex sgRNA expression induces synergistic activation of VEGFA protein expression by dCas9-VP64 protein. Fold-activation of VEGFA protein in 293 cells in which the indicated combinations of sgRNAs were co-expressed with dCas9-VP64 is shown. Note that in all of these experiments the amount of each individual sgRNA expression plasmid used for transfection was the same. Fold-activation values were calculated as described in 2A and shown as grey bars. The calculated sum of mean fold-activation values induced by individual sgRNAs is shown for each combination as black bars. Asterisks indicate all combinations that were found to be significantly greater than the expected sum as determined by an analysis of variance (ANOVA) ($p<0.05$).

FIG. 3A is a diagram illustrating the positions and orientations of six sgRNAs targeted to the endogenous human *NTF3* gene promoter. Horizontal arrows represent the first 20 nts of the sgRNA complementary to the genomic DNA sequence with the arrow pointing 5' to 3'. Grey line indicates region of potential open chromatin identified from the ENCODE DNaseI hypersensitivity track on the UCSC genome browser with the thicker part of the bar indicating the first transcribed exon. Numbering shown is relative to the transcription start site (+1, right-angle arrow).

FIG. 3B is a bar graph showing activation of *NTF3* gene expression by sgRNA-guided dCas9-VP64 in 293 cells. Relative expression of *NTF3* mRNA, detected by quantitative RT-PCR and normalized to a *GAPDH* control (deltaCt x 10^4), is shown for 293 cells co-transfected with the indicated amounts of dCas9-VP64 and *NTF3*-targeted sgRNA expression plasmids. All experiments were performed in triplicate with error bars representing standard errors of the mean. Asterisks indicate samples that are significantly greater than the off-target gRNA control as determined by a paired, one-sided T-test ($P<0.05$).

FIG. 3C is a bar graph showing multiplex gRNA expression induces synergistic activation of NTF3 mRNA expression by dCas9-VP64 protein. Relative expression of NTF3 mRNA, detected by quantitative RT-PCR and normalized to a GAPDH control (deltaCt x 104), is shown for 293 cells co-transfected with dCas9-VP64 and the indicated combinations of NTF3-targeted gRNA expression plasmids. Note that in all of these experiments the amount of each individual gRNA expression plasmid used for transfection was the same. All experiments were performed in triplicate with error bars representing

standard errors of the mean. The calculated sum of mean fold-activation values induced by individual gRNAs is shown for each combination.

FIG. 4 is an exemplary sequence of an sgRNA expression vector.

FIG. 5 is an exemplary sequence of CMV-T7-Cas9 D10A/H840A-3XFLAG-VP64 expression vector.

FIG. 6 is an exemplary sequence of CMV-T7-Cas9 recoded D10A/H840A-3XFLAG-VP64 expression vector.

FIG. 7 is an exemplary sequence of a Cas9-HFD, i.e., a Cas9-activator. An optional 3xFLAG sequence is underlined; the nuclear localization signal PKKKRKVS (SEQ ID NO:11) is in lower case; two linkers are in bold; and the VP64 transcriptional activator sequence, DALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDML (SEQ ID NO:12), is boxed.

FIGs. 8A-8B are exemplary sequences of (8A) dCas9-NLS-3XFLAG-HP1alpha and (8B) dCas9-NLS-3XFLAG-HP1beta. Box = nuclear localization signal; underline = triple flag tag; double underline = HP1alpha hinge and chromoshadow domains.

FIG. 9 is an exemplary sequence of dCas9-TET1.

FIG. 10 is a bar graph showing results obtained with various dCas9-VP64 fusion constructs. Of those tested, the optimized dCas9-VP64 architecture included an N-terminal NLS (NFN) and an additional NLS (N) or FLAG tag/NLS (NF) placed between dCas9 and VP64. Expression of the VEGFA gene in human HEK293 cells was activated by transcriptional activation mediated by RNA-guided dCas9-VP64 fusions. Expression plasmids encoding variants of dCas9-VP64 were co-transfected with a plasmid that expressed three gRNAs that targeted sites in a region upstream of the VEGFA start codon (in this experiment, the gRNAs were expressed from a single gRNA and processed out by the Csy4 endoribonuclease). VEGFA protein expression is measured by ELISA, and the mean of two replicates is shown with error bars indicating standard errors of the mean.

FIGs. 11A-B are bar graphs showing the activities of dCas9-VP64 activators bearing alternative substitution mutations to catalytically inactivate Cas9 function. (11A) Plasmids expressing dCas9-VP64 proteins bearing various Cas9 inactivating substitutions to residues D10 and H840 were each co-transfected into HEK293 cells with either a single gRNA or three distinctly targeted gRNAs targeting the VEGFA upstream region (blue and red bars, respectively). (11B) Plasmids expressing these dCas9-VP64 variants were also transfected into a HEK293 cell-line that stably expresses a single VEGFA-targeted gRNA.

VEGFA protein levels were determined by ELISA with mean of two replicates and standard errors of the mean (error bars) shown.

DETAILED DESCRIPTION

Described herein are fusion proteins of a heterologous functional domain (e.g., a transcriptional activation domain) fused to a catalytically inactivated version of the Cas9 protein for the purpose of enabling RNA-guided targeting of these functional domains to specific genomic locations in cells and living organisms.

The CRISPR/Cas system has evolved in bacteria as a defense mechanism to protect against invading plasmids and viruses. Short protospacers, derived from foreign nucleic acid, are incorporated into CRISPR loci and subsequently transcribed and processed into short CRISPR RNAs (crRNAs). These crRNAs, complexed with a second tracrRNA, then use their sequence complementarity to the invading nucleic acid to guide Cas9-mediated cleavage, and consequent destruction of the foreign nucleic acid. In 2012, Doudna and colleagues demonstrated that a single guide RNA (sgRNA) composed of a fusion of a crRNA with tracrRNA can mediate recruitment of Cas9 nuclease to specific DNA sequences in vitro (Fig. 1C; Jinek et al., Science 2012).

More recently, a longer version of the sgRNA has been used to introduce targeted alterations in human cells and zebrafish (Fig. 1B; Mali et al. Science 2013, Hwang and Fu et al., Nat Biotechnol. 2013 Mar;31(3):227-9). Qi et al. demonstrated that gRNA-mediated recruitment of a catalytically inactive mutant form of Cas9 (referred to as dCas9) could lead to repression of specific endogenous genes in *E. coli* as well as of an EGFP reporter gene in human cells (Qi et al., Cell 152, 1173–1183 (2013)). Although this study demonstrated the potential to adapt RNA-guided Cas9 technology for regulation of gene expression, it did not test or demonstrate whether heterologous functional domains (e.g.—transcriptional activation domains) could be fused to dCas9 without disrupting its ability to be recruited to specific genomic sites by programmable sgRNAs or dual gRNAs (dgRNAs – i.e.- a customized crRNA and a tracrRNA).

As described herein, in addition to guiding Cas9-mediated nuclease activity, it is possible to use CRISPR-derived RNAs to target heterologous functional domains fused to Cas9 (Cas9-HFD) to specific sites in the genome (Figure 1C). For example, as described herein, it is possible to use single guide RNAs (sgRNAs) to target Cas9-HFD, e.g., Cas9-transcriptional activators (hereafter referred to as Cas9-activators) to the promoters of specific genes and thereby increase expression of the target gene. Thus Cas9-HFD can be

localized to sites in the genome, with target specificity defined by sequence complementarity of the guide RNA. The target sequence also includes a PAM sequence (a 2-5 nucleotide sequence specified by the Cas9 protein which is adjacent to the sequence specified by the RNA).

The Cas9-HFD are created by fusing a heterologous functional domain (e.g., a transcriptional activation domain, e.g., from VP64 or NF-κB p65), to the N-terminus or C-terminus of a catalytically inactive Cas9 protein.

Cas9

A number of bacteria express Cas9 protein variants. The Cas9 from *Streptococcus pyogenes* is presently the most commonly used; some of the other Cas9 proteins have high levels of sequence identity with the *S. pyogenes* Cas9 and use the same guide RNAs. Others are more diverse, use different gRNAs, and recognize different PAM sequences as well (the 2-5 nucleotide sequence specified by the protein which is adjacent to the sequence specified by the RNA). Chylinski et al. classified Cas9 proteins from a large group of bacteria (RNA Biology 10:5, 1–12; 2013), and a large number of Cas9 proteins are listed in supplementary figure 1 and supplementary table 1 thereof, which are incorporated by reference herein. Additional Cas9 proteins are described in Esvelt et al., Nat Methods. 2013 Nov; 10(11):1116-21 and Fonfara et al., "Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems." Nucleic Acids Res. 2013 Nov 22. [Epub ahead of print] doi:10.1093/nar/gkt1074.

Cas9 molecules of a variety of species can be used in the methods and compositions described herein. While the *S. pyogenes* and *S. thermophilus* Cas9 molecules are the subject of much of the disclosure herein, Cas9 molecules of, derived from, or based on the Cas9 proteins of other species listed herein can be used as well. In other words, while the much of the description herein uses *S. pyogenes* and *S. thermophilus* Cas9 molecules, Cas9 molecules from the other species can replace them. Such species include those set forth in the following table, which was created based on supplementary figure 1 of Chylinski et al., 2013.

Alternative Cas9 proteins	
GenBank Acc No.	**Bacterium**
303229466	*Veillonella atypica ACS-134-V-Col7a*
34762592	*Fusobacterium nucleatum subsp. vincentii*
374307738	*Filifactor alocis ATCC 35896*
320528778	*Solobacterium moorei F0204*
291520705	*Coprococcus catus GD-7*

Alternative Cas9 proteins	
GenBank Acc No.	**Bacterium**
42525843	*Treponema denticola ATCC 35405*
304438954	*Peptoniphilus duerdenii ATCC BAA-1640*
224543312	*Catenibacterium mitsuokai DSM 15897*
24379809	*Streptococcus mutans UA159*
15675041	*Streptococcus pyogenes SF370*
16801805	*Listeria innocua Clip11262*
116628213	*Streptococcus thermophilus LMD-9*
323463801	*Staphylococcus pseudintermedius ED99*
352684361	*Acidaminococcus intestini RyC-MR95*
302336020	*Olsenella uli DSM 7084*
366983953	*Oenococcus kitaharae DSM 17330*
310286728	*Bifidobacterium bifidum S17*
258509199	*Lactobacillus rhamnosus GG*
300361537	*Lactobacillus gasseri JV-V03*
169823755	*Finegoldia magna ATCC 29328*
47458868	*Mycoplasma mobile 163K*
284931710	*Mycoplasma gallisepticum str. F*
363542550	*Mycoplasma ovipneumoniae SC01*
384393286	*Mycoplasma canis PG 14*
71894592	*Mycoplasma synoviae 53*
238924075	*Eubacterium rectale ATCC 33656*
116627542	*Streptococcus thermophilus LMD-9*
315149830	*Enterococcus faecalis TX0012*
315659848	*Staphylococcus lugdunensis M23590*
160915782	*Eubacterium dolichum DSM 3991*
336393381	*Lactobacillus coryniformis subsp. torquens*
310780384	*Ilyobacter polytropus DSM 2926*
325677756	*Ruminococcus albus 8*
187736489	*Akkermansia muciniphila ATCC BAA-835*
117929158	*Acidothermus cellulolyticus 11B*
189440764	*Bifidobacterium longum DJO10A*
283456135	*Bifidobacterium dentium Bd1*
38232678	*Corynebacterium diphtheriae NCTC 13129*
187250660	*Elusimicrobium minutum Pei191*
319957206	*Nitratifractor salsuginis DSM 16511*
325972003	*Sphaerochaeta globus str. Buddy*
261414553	*Fibrobacter succinogenes subsp. succinogenes*
60683389	*Bacteroides fragilis NCTC 9343*
256819408	*Capnocytophaga ochracea DSM 7271*
90425961	*Rhodopseudomonas palustris BisB18*
373501184	*Prevotella micans F0438*
294674019	*Prevotella ruminicola 23*
365959402	*Flavobacterium columnare ATCC 49512*
312879015	*Aminomonas paucivorans DSM 12260*
83591793	*Rhodospirillum rubrum ATCC 11170*
294086111	*Candidatus Puniceispirillum marinum IMCC1322*

Alternative Cas9 proteins	
GenBank Acc No.	**Bacterium**
121608211	*Verminephrobacter eiseniae EF01-2*
344171927	*Ralstonia syzygii R24*
159042956	*Dinoroseobacter shibae DFL 12*
288957741	*Azospirillum sp- B510*
92109262	*Nitrobacter hamburgensis X14*
148255343	*Bradyrhizobium sp- BTAi1*
34557790	*Wolinella succinogenes DSM 1740*
218563121	*Campylobacter jejuni subsp. jejuni*
291276265	*Helicobacter mustelae 12198*
229113166	*Bacillus cereus Rock1-15*
222109285	*Acidovorax ebreus TPSY*
189485225	*uncultured Termite group 1*
182624245	*Clostridium perfringens D str.*
220930482	*Clostridium cellulolyticum H10*
154250555	*Parvibaculum lavamentivorans DS-1*
257413184	*Roseburia intestinalis L1-82*
218767588	*Neisseria meningitidis Z2491*
15602992	*Pasteurella multocida subsp. multocida*
319941583	*Sutterella wadsworthensis 3 1*
254447899	*gamma proteobacterium HTCC5015*
54296138	*Legionella pneumophila str. Paris*
331001027	*Parasutterella excrementihominis YIT 11859*
34557932	*Wolinella succinogenes DSM 1740*
118497352	*Francisella novicida U112*

The constructs and methods described herein can include the use of any of those Cas9 proteins, and their corresponding guide RNAs or other guide RNAs that are compatible. The Cas9 from *Streptococcus thermophilus* LMD-9 CRISPR1 system has been shown to function in human cells in Cong et al (Science 339, 819 (2013)). Additionally, Jinek et al. showed *in vitro* that Cas9 orthologs from S. *thermophilus* and *L. innocua*, (but not from *N. meningitidis* or *C. jejuni,* which likely use a different guide RNA), can be guided by a dual *S. pyogenes* gRNA to cleave target plasmid DNA, albeit with slightly decreased efficiency.

In some embodiments, the present system utilizes the Cas9 protein from *S. pyogenes*, either as encoded in bacteria or codon-optimized for expression in mammalian cells, containing mutations at D10, E762, H983, or D986 and H840 or N863, e.g., D10A/D10N and H840A/H840N/H840Y, to render the nuclease portion of the protein catalytically inactive; substitutions at these positions could be alanine (as they are in Nishimasu al., Cell 156, 935–949 (2014)) or they could be other residues, e.g., glutamine, asparagine, tyrosine, serine, or aspartate, e.g., E762Q, H983N, H983Y, D986N, N863D, N863S, or N863H (Figure 1C). The sequence of the catalytically inactive *S. pyogenes* Cas9

that can be used in the methods and compositions described herein is as follows; the exemplary mutations of D10A and H840A are in bold and underlined.

```
        10         20         30         40         50         60
MDKKYSIGL**A** IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE

        70         80         90        100        110        120
ATRLKRTARR RYTRRKNRIC YLQEIFSNEM AKVDDSFFHR LEESFLVEED KKHERHPIFG

       130        140        150        160        170        180
NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRLIYLALAH MIKFRGHFLI EGDLNPDNSD

       190        200        210        220        230        240
VDKLFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN

       250        260        270        280        290        300
LIALSLGLTP NFKSNFDLAE DAKLQLSKDT YDDDLDNLLA QIGDQYADLF LAAKNLSDAI

       310        320        330        340        350        360
LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI FFDQSKNGYA

       370        380        390        400        410        420
GYIDGGASQE EFYKFIKPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH

       430        440        450        460        470        480
AILRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE

       490        500        510        520        530        540
VVDKGASAQS FIERMTNFDK NLPNEKVLPK HSLLYEYFTV YNELTKVKYV TEGMRKPAFL

       550        560        570        580        590        600
SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLLKI

       610        620        630        640        650        660
IKDKDFLDNE ENEDILEDIV LTLTLFEDRE MIEERLKTYA HLFDDKVMKQ LKRRRYTGWG

       670        680        690        700        710        720
RLSRKLINGI RDKQSGKTIL DFLKSDGFAN RNFMQLIHDD SLTFKEDIQK AQVSGQGDSL

       730        740        750        760        770        780
HEHIANLAGS PAIKKGILQT VKVVDELVKV MGRHKPENIV IEMARENQTT QKGQKNSRER

       790        800        810        820        830        840
MKRIEEGIKE LGSQILKEHP VENTQLQNEK LYLYYLQNGR DMYVDQELDI NRLSDYDVD**A**

       850        860        870        880        890        900
IVPQSFLKDD SIDNKVLTRS DKNRGKSDNV PSEEVVKKMK NYWRQLLNAK LITQRKFDNL

       910        920        930        940        950        960
TKAERGGLSE LDKAGFIKRQ LVETRQITKH VAQILDSRMN TKYDENDKLI REVKVITLKS

       970        980        990       1000       1010       1020
KLVSDFRKDF QFYKVREINN YHHAHDAYLN AVVGTALIKK YPKLESEFVY GDYKVYDVRK

      1030       1040       1050       1060       1070       1080
MIAKSEQEIG KATAKYFFYS NIMNFFKTEI TLANGEIRKR PLIETNGETG EIVWDKGRDF

      1090       1100       1110       1120       1130       1140
ATVRKVLSMP QVNIVKKTEV QTGGFSKESI LPKRNSDKLI ARKKDWDPKK YGGFDSPTVA
```

```
      1150       1160       1170       1180       1190       1200
YSVLVVAKVE KGKSKKLKSV KELLGITIME RSSFEKNPID FLEAKGYKEV KKDLIIKLPK

      1210       1220       1230       1240       1250       1260
YSLFELENGR KRMLASAGEL QKGNELALPS KYVNFLYLAS HYEKLKGSPE DNEQKQLFVE

      1270       1280       1290       1300       1310       1320
QHKHYLDEII EQISEFSKRV ILADANLDKV LSAYNKHRDK PIREQAENII HLFTLTNLGA

      1330       1340       1350       1360
PAAFKYFDTT IDRKRYTSTK EVLDATLIHQ SITGLYETRI DLSQLGGD   (SEQ ID NO:13)
```

In some embodiments, the Cas9 nuclease used herein is at least about 50% identical to the sequence of *S. pyogenes* Cas9, i.e., at least 50% identical to SEQ ID NO:13. In some embodiments, the nucleotide sequences are about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% identical to SEQ ID NO:13.

In some embodiments, the catalytically inactive Cas9 used herein is at least about 50% identical to the sequence of the catalytically inactive *S. pyogenes* Cas9, i.e., at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% identical to SEQ ID NO:13, wherein the mutations at D10 and H840, e.g., D10A/D10N and H840A/H840N/H840Y are maintained.

In some embodiments, any differences from SEQ ID NO:13 are in non-conserved regions, as identified by sequence alignment of sequences set forth in Chylinski et al., RNA Biology 10:5, 1–12; 2013 (e.g., in supplementary figure 1 and supplementary table 1 thereof); Esvelt et al., Nat Methods. 2013 Nov;10(11):1116-21 and Fonfara et al., Nucl. Acids Res. (2014) 42 (4): 2577-2590. [Epub ahead of print 2013 Nov 22] doi:10.1093/nar/gkt1074, and wherein the mutations at D10 and H840, e.g., D10A/D10N and H840A/H840N/H840Y are maintained.

To determine the percent identity of two sequences, the sequences are aligned for optimal comparison purposes (gaps are introduced in one or both of a first and a second amino acid or nucleic acid sequence as required for optimal alignment, and non-homologous sequences can be disregarded for comparison purposes). The length of a reference sequence aligned for comparison purposes is at least 50% (in some embodiments, about 50%, 55%, 60%, 65%, 70%, 75%, 85%, 90%, 95%, or 100% of the length of the reference sequence) is aligned. The nucleotides or residues at corresponding positions are then compared. When a position in the first sequence is occupied by the same nucleotide or residue as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number

of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. For purposes of the present application, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch ((1970) J. Mol. Biol. 48:444-453) algorithm which has been incorporated into the GAP program in the GCG software package, using a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

Heterologous Functional Domains

The transcriptional activation domains can be fused on the N or C terminus of the Cas9. In addition, although the present description exemplifies transcriptional activation domains, other heterologous functional domains (e.g., transcriptional repressors (e.g., KRAB, ERD, SID, and others, e.g., amino acids 473–530 of the *ets2* repressor factor (ERF) repressor domain (ERD), amino acids 1–97 of the KRAB domain of KOX1, or amino acids 1–36 of the Mad mSIN3 interaction domain (SID); see Beerli et al., PNAS USA 95:14628-14633 (1998)) or silencers such as Heterochromatin Protein 1 (HP1, also known as swi6), e.g., HP1α or HP1β; proteins or peptides that could recruit long non-coding RNAs (lncRNAs) fused to a fixed RNA binding sequence such as those bound by the MS2 coat protein, endoribonuclease Csy4, or the lambda N protein; enzymes that modify the methylation state of DNA (e.g., DNA methyltransferase (DNMT) or TET proteins); or enzymes that modify histone subunits (e.g., histone acetyltransferases (HAT), histone deacetylases (HDAC), histone methyltransferases (e.g., for methylation of lysine or arginine residues) or histone demethylases (e.g., for demethylation of lysine or arginine residues)) as are known in the art can also be used. A number of sequences for such domains are known in the art, e.g., a domain that catalyzes hydroxylation of methylated cytosines in DNA. Exemplary proteins include the Ten-Eleven-Translocation (TET)1-3 family, enzymes that converts 5-methylcytosine (5-mC) to 5-hydroxymethylcytosine (5-hmC) in DNA.

Sequences for human TET1-3 are known in the art and are shown in the following table:

	GenBank Accession Nos.	
Gene	Amino Acid	Nucleic Acid
TET1	NP_085128.2	NM_030625.2
TET2*	NP_001120680.1 (var 1) NP_060098.3 (var 2)	NM_001127208.2 NM_017628.4
TET3	NP_659430.1	NM_144993.1

* Variant (1) represents the longer transcript and encodes the longer isoform (a). Variant (2) differs in the 5' UTR and in the 3' UTR and coding sequence compared to variant 1. The resulting isoform (b) is shorter and has a distinct C-terminus compared to isoform a.

In some embodiments, all or part of the full-length sequence of the catalytic domain can be included, e.g., a catalytic module comprising the cysteine-rich extension and the 2OGFeDO domain encoded by 7 highly conserved exons, e.g., the Tet1 catalytic domain comprising amino acids 1580-2052, Tet2 comprising amino acids 1290-1905 and Tet3 comprising amino acids 966-1678. See, e.g., Fig. 1 of Iyer et al., Cell Cycle. 2009 Jun 1;8(11):1698-710. Epub 2009 Jun 27, for an alignment illustrating the key catalytic residues in all three Tet proteins, and the supplementary materials thereof (available at ftp site ftp.ncbi.nih.gov/pub/aravind/DONS/supplementary_material_DONS.html) for full length sequences (see, e.g., seq 2c); in some embodiments, the sequence includes amino acids 1418-2136 of Tet1 or the corresponding region in Tet2/3.

Other catalytic modules can be from the proteins identified in Iyer et al., 2009.

In some embodiments, the heterologous functional domain is a biological tether, and comprises all or part of (e.g., DNA binding domain from) the MS2 coat protein, endoribonuclease Csy4, or the lambda N protein. These proteins can be used to recruit RNA molecules containing a specific stem-loop structure to a locale specified by the dCas9 gRNA targeting sequences. For example, a dCas9 fused to MS2 coat protein, endoribonuclease Csy4, or lambda N can be used to recruit a long non-coding RNA (lncRNA) such as XIST or HOTAIR; see, e.g., Keryer-Bibens et al., Biol. Cell 100:125–138 (2008), that is linked to the Csy4, MS2 or lambda N binding sequence. Alternatively, the Csy4, MS2 or lambda N protein binding sequence can be linked to another protein, e.g., as described in Keryer-Bibens et al., supra, and the protein can be targeted to the dCas9

binding site using the methods and compositions described herein. In some embodiments, the Csy4 is catalytically inactive.

In some embodiments, the fusion proteins include a linker between the dCas9 and the heterologous functional domains. Linkers that can be used in these fusion proteins (or between fusion proteins in a concatenated structure) can include any sequence that does not interfere with the function of the fusion proteins. In preferred embodiments, the linkers are short, e.g., 2-20 amino acids, and are typically flexible (i.e., comprising amino acids with a high degree of freedom such as glycine, alanine, and serine). In some embodiments, the linker comprises one or more units consisting of GGGS (SEQ ID NO:14) or GGGGS (SEQ ID NO:15), e.g., two, three, four, or more repeats of the GGGS (SEQ ID NO:14) or GGGGS (SEQ ID NO:15) unit. Other linker sequences can also be used.

Methods of Use

The described Cas9-HFD system is a useful and versatile tool for modifying the expression of endogenous genes. Current methods for achieving this require the generation of novel engineered DNA-binding proteins (such as engineered zinc finger or transcription activator-like effector DNA binding domains) for each site to be targeted. Because these methods demand expression of a large protein specifically engineered to bind each target site, they are limited in their capacity for multiplexing. Cas9-HFD, however, require expression of only a single Cas9-HFD protein, which can be targeted to multiple sites in the genome by expression of multiple short gRNAs. This system could therefore easily be used to simultaneously induce expression of a large number of genes or to recruit multiple Cas9-HFDs to a single gene, promoter, or enhancer. This capability will have broad utility, e.g., for basic biological research, where it can be used to study gene function and to manipulate the expression of multiple genes in a single pathway, and in synthetic biology, where it will enable researchers to create circuits in cell that are responsive to multiple input signals. The relative ease with which this technology can be implemented and adapted to multiplexing will make it a broadly useful technology with many wide-ranging applications.

The methods described herein include contacting cells with a nucleic acid encoding the Cas9-HFD described herein, and nucleic acids encoding one or more guide RNAs directed to a selected gene, to thereby modulate expression of that gene.

Guide RNAs (gRNAs)

Guide RNAs generally speaking come in two different systems: System 1, which uses separate crRNA and tracrRNAs that function together to guide cleavage by Cas9, and System 2, which uses a chimeric crRNA-tracrRNA hybrid that combines the two separate guide RNAs in a single system (referred to as a single guide RNA or sgRNA, see also Jinek et al., Science 2012; 337:816–821). The tracrRNA can be variably truncated and a range of lengths has been shown to function in both the separate system (system 1) and the chimeric gRNA system (system 2). For example, in some embodiments, tracrRNA may be truncated from its 3' end by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 or 40 nts. In some embodiments, the tracrRNA molecule may be truncated from its 5' end by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 or 40 nts. Alternatively, the tracrRNA molecule may be truncated from both the 5' and 3' end, e.g., by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 or 20 nts on the 5' end and at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 or 40 nts on the 3' end. See, e.g., Jinek et al., Science 2012; 337:816–821; Mali et al., Science. 2013 Feb 15;339(6121):823-6; Cong et al., Science. 2013 Feb 15;339(6121):819-23; and Hwang and Fu et al., Nat Biotechnol. 2013 Mar;31(3):227-9; Jinek et al., Elife 2, e00471 (2013)). For System 2, generally the longer length chimeric gRNAs have shown greater on-target activity but the relative specificities of the various length gRNAs currently remain undefined and therefore it may be desirable in certain instances to use shorter gRNAs. In some embodiments, the gRNAs are complementary to a region that is within about 100-800 bp upstream of the transcription start site, e.g., is within about 500 bp upstream of the transcription start site, includes the transcription start site, or within about 100-800 bp, e.g., within about 500 bp, downstream of the transcription start site. In some embodiments, vectors (e.g., plasmids) encoding more than one gRNA are used, e.g., plasmids encoding, 2, 3, 4, 5, or more gRNAs directed to different sites in the same region of the target gene.

Cas9 nuclease can be guided to specific 17-20 nt genomic targets bearing an additional proximal protospacer adjacent motif (PAM), e.g., of sequence NGG, using a guide RNA, e.g., a single gRNA or a tracrRNA/crRNA, bearing 17-20 nts at its 5' end that are complementary to the complementary strand of the genomic DNA target site. Thus, the present methods can include the use of a single guide RNA comprising a crRNA fused to a normally trans-encoded tracrRNA, e.g., a single Cas9 guide RNA as described in Mali et al., Science 2013 Feb 15; 339(6121):823-6, with a sequence at the 5' end that is complementary to the target sequence, e.g., of 25-17, optionally 20 or fewer nucleotides (nts), e.g., 20, 19, 18, or 17 nts, preferably 17 or 18 nts, of the complementary strand to a

target sequence immediately 5' of a protospacer adjacent motif (PAM), e.g., NGG, NAG, or NNGG. In some embodiments, the single Cas9 guide RNA consists of the sequence: (X_{17-20})GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCG(X_N) (SEQ ID NO:1);
(X_{17-20})GUUUUAGAGCUAUGCUGAAAAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUC(X_N) (SEQ ID NO:2);
(X_{17-20})GUUUUAGAGCUAUGCUGUUUUGGAAACAAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUC(X_N) (SEQ ID NO:3);
(X_{17-20})GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC(X_N) (SEQ ID NO:4),
(X_{17-20})GUUUAAGAGCUAGAAAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC(SEQ ID NO:5);
(X_{17-20})GUUUUAGAGCUAUGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:6); or
(X_{17-20})GUUUAAGAGCUAUGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:7);
wherein X_{17-20} is the nucleotide sequence complementary to 17-20 consecutive nucleotides of the target sequence. DNAs encoding the single guide RNAs have been described previously in the literature (Jinek et al., Science. 337(6096):816-21 (2012) and Jinek et al., Elife. 2:e00471 (2013)).

The guide RNAs can include X_N which can be any sequence, wherein N (in the RNA) can be 0-200, e.g., 0-100, 0-50, or 0-20, that does not interfere with the binding of the ribonucleic acid to Cas9.

In some embodiments, the guide RNA includes one or more Adenine (A) or Uracil (U) nucleotides on the 3' end. In some embodiments the RNA includes one or more U, e.g., 1 to 8 or more Us (e.g., U, UU, UUU, UUUU, UUUUU, UUUUUU, UUUUUUU, UUUUUUUU) at the 3' end of the molecule, as a result of the optional presence of one or more Ts used as a termination signal to terminate RNA PolIII transcription.

Although some of the examples described herein utilize a single gRNA, the methods can also be used with dual gRNAs (e.g., the crRNA and tracrRNA found in naturally occurring systems). In this case, a single tracrRNA would be used in conjunction with multiple different crRNAs expressed using the present system, e.g., the following:
(X_{17-20})GUUUUAGAGCUA (SEQ ID NO:102);
(X_{17-20}) GUUUUAGAGCUAUGCUGUUUUG (SEQ ID NO:103); or

(X_{17-20})GUUUUAGAGCUAUGCU (SEQ ID NO:104); and a tracrRNA sequence. In this case, the crRNA is used as the guide RNA in the methods and molecules described herein, and the tracrRNA can be expressed from the same or a different DNA molecule. In some embodiments, the methods include contacting the cell with a tracrRNA comprising or consisting of the sequence GGAACCAUUCAAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:8) or an active portion thereof (an active portion is one that retains the ability to form complexes with Cas9 or dCas9). In some embodiments, the tracrRNA molecule may be truncated from its 3' end by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 or 40 nts. In another embodiment, the tracrRNA molecule may be truncated from its 5' end by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 or 40 nts. Alternatively, the tracrRNA molecule may be truncated from both the 5' and 3' end, e.g., by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 or 20 nts on the 5' end and at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 or 40 nts on the 3' end. Exemplary tracrRNA sequences in addition to SEQ ID NO:8 include the following: UAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCG AGUCGGUGC (SEQ ID NO:105) or an active portion thereof; or AGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGC ACCGAGUCGGUGC (SEQ ID NO:106) or an active portion thereof.

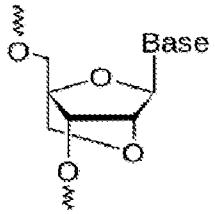
In some embodiments when (X_{17-20})GUUUUAGAGCUAUGCUGUUUUG (SEQ ID NO:102) is used as a crRNA, the following tracrRNA is used: GGAACCAUUCAAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:8) or an active portion thereof.

In some embodiments when (X_{17-20})GUUUUAGAGCUA (SEQ ID NO:102) is used as a crRNA, the following tracrRNA is used: UAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCG AGUCGGUGC (SEQ ID NO:105) or an active portion thereof.

In some embodiments when (X_{17-20}) GUUUUAGAGCUAUGCU (SEQ ID NO:104) is used as a crRNA, the following tracrRNA is used: AGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGC ACCGAGUCGGUGC (SEQ ID NO:106) or an active portion thereof.

In some embodiments, the gRNA is targeted to a site that is at least three or more mismatches different from any sequence in the rest of the genome in order to minimize off-target effects.

Modified RNA oligonucleotides such as locked nucleic acids (LNAs) have been demonstrated to increase the specificity of RNA-DNA hybridization by locking the modified oligonucleotides in a more favorable (stable) conformation. For example, 2'-O-methyl RNA is a modified base where there is an additional covalent linkage between the 2' oxygen and 4' carbon which when incorporated into oligonucleotides can improve overall thermal stability and selectivity (**Formula I**).



Formula I – Locked Nucleic Acid

Thus in some embodiments, the tru-gRNAs disclosed herein may comprise one or more modified RNA oligonucleotides. For example, the truncated guide RNAs molecules described herein can have one, some or all of the region of the guideRNA complementary to the target sequence are modified, e.g., locked (2'-O-4'-C methylene bridge), 5'-methylcytidine, 2'-O-methyl-pseudouridine, or in which the ribose phosphate backbone has been replaced by a polyamide chain (peptide nucleic acid), e.g., a synthetic ribonucleic acid.

In other embodiments, one, some or all of the nucleotides of the tru-gRNA sequence may be modified, e.g., locked (2'-O-4'-C methylene bridge), 5'-methylcytidine, 2'-O-methyl-pseudouridine, or in which the ribose phosphate backbone has been replaced by a polyamide chain (peptide nucleic acid), e.g., a synthetic ribonucleic acid.

In some embodiments, the single guide RNAs and/or crRNAs and/or tracrRNAs can include one or more Adenine (A) or Uracil (U) nucleotides on the 3' end.

Existing Cas9-based RGNs use gRNA-DNA heteroduplex formation to guide targeting to genomic sites of interest. However, RNA-DNA heteroduplexes can form a more promiscuous range of structures than their DNA-DNA counterparts. In effect, DNA-DNA duplexes are more sensitive to mismatches, suggesting that a DNA-guided nuclease may not bind as readily to off-target sequences, making them comparatively more specific than RNA-guided nucleases. Thus, the guide RNAs usable in the methods described herein

can be hybrids, i.e., wherein one or more deoxyribonucleotides, e.g., a short DNA oligonucleotide, replaces all or part of the gRNA, e.g., all or part of the complementarity region of a gRNA. This DNA-based molecule could replace either all or part of the gRNA in a single gRNA system or alternatively might replace all of part of the crRNA and/or tracrRNA in a dual crRNA/tracrRNA system. Such a system that incorporates DNA into the complementarity region should more reliably target the intended genomic DNA sequences due to the general intolerance of DNA-DNA duplexes to mismatching compared to RNA-DNA duplexes. Methods for making such duplexes are known in the art, See, e.g., Barker et al., BMC Genomics. 2005 Apr 22;6:57; and Sugimoto et al., Biochemistry. 2000 Sep 19;39(37):11270-81.

In addition, in a system that uses separate crRNA and tracrRNA, one or both can be synthetic and include one or more modified (e.g., locked) nucleotides or deoxyribonucleotides.

In a cellular context, complexes of Cas9 with these synthetic gRNAs could be used to improve the genome-wide specificity of the CRISPR/Cas9 nuclease system.

The methods described can include expressing in a cell, or contacting the cell with, a Cas9 gRNA plus a fusion protein as described herein.

Expression Systems

In order to use the fusion proteins and guide RNAs described herein, it may be desirable to express them from a nucleic acid that encodes them. This can be performed in a variety of ways. For example, a nucleic acid encoding a guide RNA or fusion protein can be cloned into an intermediate vector for transformation into prokaryotic or eukaryotic cells for replication and/or expression. Intermediate vectors are typically prokaryote vectors, e.g., plasmids, or shuttle vectors, or insect vectors, for storage or manipulation of the nucleic acid encoding the fusion protein or for production of the fusion protein. The nucleic acid encoding the guide RNA or fusion protein can also be cloned into an expression vector, for administration to a plant cell, animal cell, preferably a mammalian cell or a human cell, fungal cell, bacterial cell, or protozoan cell.

To obtain expression, a sequence encoding a guide RNA or fusion protein is typically subcloned into an expression vector that contains a promoter to direct transcription. Suitable bacterial and eukaryotic promoters are well known in the art and described, e.g., in Sambrook et al., Molecular Cloning, A Laboratory Manual (3d ed. 2001); Kriegler, Gene Transfer and Expression: A Laboratory Manual (1990); and Current

Protocols in Molecular Biology (Ausubel et al., eds., 2010). Bacterial expression systems for expressing the engineered protein are available in, e.g., *E. coli*, *Bacillus* sp., and *Salmonella* (Palva et al., 1983, Gene 22:229-235). Kits for such expression systems are commercially available. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known in the art and are also commercially available.

The promoter used to direct expression of the nucleic acid depends on the particular application. For example, a strong constitutive promoter is typically used for expression and purification of fusion proteins. In contrast, when the fusion protein is to be administered in vivo for gene regulation, either a constitutive or an inducible promoter can be used, depending on the particular use of the fusion protein. In addition, a preferred promoter for administration of the fusion protein can be a weak promoter, such as HSV TK or a promoter having similar activity. The promoter can also include elements that are responsive to transactivation, e.g., hypoxia response elements, Gal4 response elements, lac repressor response element, and small molecule control systems such as tetracycline-regulated systems and the RU-486 system (see, e.g., Gossen & Bujard, 1992, Proc. Natl. Acad. Sci. USA, 89:5547; Oligino et al., 1998, Gene Ther., 5:491-496; Wang et al., 1997, Gene Ther., 4:432-441; Neering et al., 1996, Blood, 88:1147-55; and Rendahl et al., 1998, Nat. Biotechnol., 16:757-761).

In addition to the promoter, the expression vector typically contains a transcription unit or expression cassette that contains all the additional elements required for the expression of the nucleic acid in host cells, either prokaryotic or eukaryotic. A typical expression cassette thus contains a promoter operably linked, e.g., to the nucleic acid sequence encoding the fusion protein, and any signals required, e.g., for efficient polyadenylation of the transcript, transcriptional termination, ribosome binding sites, or translation termination. Additional elements of the cassette may include, e.g., enhancers, and heterologous spliced intronic signals.

The particular expression vector used to transport the genetic information into the cell is selected with regard to the intended use of the fusion protein, e.g., expression in plants, animals, bacteria, fungus, protozoa, etc. Standard bacterial expression vectors include plasmids such as pBR322 based plasmids, pSKF, pET23D, and commercially available tag-fusion expression systems such as GST and LacZ. A preferred tag-fusion protein is the maltose binding protein (MBP). Such tag-fusion proteins can be used for purification of the engineered TALE repeat protein. Epitope tags can also be added to

recombinant proteins to provide convenient methods of isolation, for monitoring expression, and for monitoring cellular and subcellular localization, e.g., c-myc or FLAG.

Expression vectors containing regulatory elements from eukaryotic viruses are often used in eukaryotic expression vectors, e.g., SV40 vectors, papilloma virus vectors, and vectors derived from Epstein-Barr virus. Other exemplary eukaryotic vectors include pMSG, pAV009/A+, pMTO10/A+, pMAMneo-5, baculovirus pDSVE, and any other vector allowing expression of proteins under the direction of the SV40 early promoter, SV40 late promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, or other promoters shown effective for expression in eukaryotic cells.

The vectors for expressing the guide RNAs can include RNA Pol III promoters to drive expression of the guide RNAs, e.g., the H1, U6 or 7SK promoters. These human promoters allow for expression of gRNAs in mammalian cells following plasmid transfection. Alternatively, a T7 promoter may be used, e.g., for in vitro transcription, and the RNA can be transcribed in vitro and purified. Vectors suitable for the expression of short RNAs, e.g., siRNAs, shRNAs, or other small RNAs, can be used.
Some expression systems have markers for selection of stably transfected cell lines such as thymidine kinase, hygromycin B phosphotransferase, and dihydrofolate reductase. High yield expression systems are also suitable, such as using a baculovirus vector in insect cells, with the fusion protein encoding sequence under the direction of the polyhedrin promoter or other strong baculovirus promoters.

The elements that are typically included in expression vectors also include a replicon that functions in *E. coli*, a gene encoding antibiotic resistance to permit selection of bacteria that harbor recombinant plasmids, and unique restriction sites in nonessential regions of the plasmid to allow insertion of recombinant sequences.

Standard transfection methods are used to produce bacterial, mammalian, yeast or insect cell lines that express large quantities of protein, which are then purified using standard techniques (see, e.g., Colley et al., 1989, J. Biol. Chem., 264:17619-22; Guide to Protein Purification, in Methods in Enzymology, vol. 182 (Deutscher, ed., 1990)). Transformation of eukaryotic and prokaryotic cells are performed according to standard techniques (see, e.g., Morrison, 1977, J. Bacteriol. 132:349-351; Clark-Curtiss & Curtiss, Methods in Enzymology 101:347-362 (Wu et al., eds, 1983).

Any of the known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, polybrene,

protoplast fusion, electroporation, nucleofection, liposomes, microinjection, naked DNA, plasmid vectors, viral vectors, both episomal and integrative, and any of the other well-known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (see, e.g., Sambrook et al., supra). It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host cell capable of expressing the protein of choice.

In some embodiments, the fusion protein includes a nuclear localization domain which provides for the protein to be translocated to the nucleus. Several nuclear localization sequences (NLS) are known, and any suitable NLS can be used. For example, many NLSs have a plurality of basic amino acids, referred to as a bipartite basic repeats (reviewed in Garcia-Bustos et al, 1991, Biochim. Biophys. Acta, 1071:83-101). An NLS containing bipartite basic repeats can be placed in any portion of chimeric protein and results in the chimeric protein being localized inside the nucleus. In preferred embodiments a nuclear localization domain is incorporated into the final fusion protein, as the ultimate functions of the fusion proteins described herein will typically require the proteins to be localized in the nucleus. However, it may not be necessary to add a separate nuclear localization domain in cases where the DBD domain itself, or another functional domain within the final chimeric protein, has intrinsic nuclear translocation function.

The present invention includes the vectors and cells comprising the vectors.

EXAMPLES

The invention is further described in the following examples, which do not limit the scope of the invention described in the claims.

Example 1. Engineering CRISPR/Cas Activator System:

It was hypothesized that RNA-guided transcriptional activators could be created by fusing the strong synthetic VP64 activation domain (Beerli et al., Proc Natl Acad Sci USA 95, 14628–14633 (1998)) to the carboxy-terminus of the catalytically inactivated dCas9 protein (Fig. 1D).

To express guide RNAs (gRNAs) in human cells, a vector was engineered that would express the full length chimeric gRNA (a fusion of crRNA and tracrRNA originally described by Jinek et al. (Science 2012)) driven by a U6 promoter. Construction of the gRNA expression plasmids was performed as follows. Pairs of DNA oligonucleotides

encoding the variable 20 nt gRNA targeting sequences were annealed together to generate short double-strand DNA fragments with 4bp overhangs (Table 1).

Table 1. VEGFA and NTF3 gene target sites and associated oligonucleotides used to construct gRNA expression plasmids.

gRNA	Target Site (including PAM)	SEQ ID NO:
V1	GTGTGCAGACGGCAGTCACTAGG	16.
V2	GAGCAGCGTCTTCGAGAGTGAGG	17.
V3	GGTGAGTGAGTGTGTGCGTGTGG	18.
V4	GTTGGAGCGGGGAGAAGGCCAGG	19.
V5	GGGTGGGGGGAGTTTGCTCCTGG	20.
V6	GGCTTTGGAAAGGGGGTGGGGGG	21.
V7	GGGGCGGGGTCCCGGCGGGGCGG	22.
V8	GCTCGGAGGTCGTGGCGCTGGGG	23.
V9	GACTCACCGGCCAGGGCGCTCGG	24.
V10	GGCGCAGCGGTTAGGTGGACCGG	25.
V11	GGCGCATGGCTCCGCCCCGCCGG	26.
V12	GCCACGACCTCCGAGCTACCCGG	27.
V13	GCGGCGTGAGCCCTCCCCCTTGG	28.
V14	GGAGGCGGGGTGGAGGGGGTCGG	29.
V15	GGGCTCACGCCGCGCTCCGGCGG	30.
V16	GACCCCCTCCACCCCGCCTCCGG	31.
N1	GAGCGCGGAGCCATCTGGCCGGG	32.
N2	GCGCGGCGCGGAAGGGGTTAAGG	33.
N3	GCGGCGCGGCGCGGGCCGGCGGG	34.
N4	GCCGCGCCGCCCTCCCCCGCCGG	35.
N5	GCGGTTATAACCAGCCAACCCGG	36.
N6	GTGCGCGGAGCTGTTCGGAAGGG	37.
gRNA	top oligo	SEQ ID NO:
V1	ACACCGTGTGCAGACGGCAGTCACTG	38.
V2	ACACCGAGCAGCGTCTTCGAGAGTGG	39.
V3	ACACCGGTGAGTGAGTGTGTGCGTGG	40.
V4	ACACCGTTGGAGCGGGGAGAAGGCCG	41.
V5	ACACCGGGTGGGGGGAGTTTGCTCCG	42.
V6	ACACCGGCTTTGGAAAGGGGGTGGGG	43.
V7	ACACCGGGGCGGGGTCCCGGCGGGGG	44.
V8	ACACCGCTCGGAGGTCGTGGCGCTGG	45.
V9	ACACCGACTCACCGGCCAGGGCGCTG	46.
V10	ACACCGGCGCAGCGGTTAGGTGGACG	47.
V11	ACACCGGCGCATGGCTCCGCCCCGCG	48.
V12	ACACCGCCACGACCTCCGAGCTACCG	49.
V13	ACACCGCGGCGTGAGCCCTCCCCCTG	50.
V14	ACACCGGAGGCGGGGTGGAGGGGGTG	51.
V15	ACACCGGGCTCACGCCGCGCTCCGGG	52.
V16	ACACCGACCCCCTCCACCCCGCCTCG	53.
N1	ACACCGAGCGCGGAGCCATCTGGCCG	54.
N2	ACACCGCGCGGCGCGGAAGGGGTTAG	55.
N3	ACACCGCGGCGCGGCGCGGGCCGGCG	56.
N4	ACACCGCCGCGCCGCCCTCCCCCGCG	57.
N5	ACACCGCGGTTATAACCAGCCAACCG	58.
N6	ACACCGTGCGCGGAGCTGTTCGGAAG	59.
gRNA	bottom oligo	SEQ ID NO:
V1	AAAACAGTGACTGCCGTCTGCACACG	60.
V2	AAAACCACTCTCGAAGACGCTGCTCG	61.
V3	AAAACCACGCACACACTCACTCACCG	62.

V4	AAAACGGCCTTCTCCCCGCTCCAACG	63.
V5	AAAACGGAGCAAACTCCCCCCACCCG	64.
V6	AAAACCCCACCCCCTTTCCAAAGCCG	65.
V7	AAAACCCCCGCCGGGACCCCGCCCCG	66.
V8	AAAACCAGCGCCACGACCTCCGAGCG	67.
V9	AAAACAGCGCCCTGGCCGGTGAGTCG	68.
V10	AAAACGTCCACCTAACCGCTGCGCCG	69.
V11	AAAACGCGGGGCGGAGCCATGCGCCG	70.
V12	AAAACGGTAGCTCGGAGGTCGTGGCG	71.
V13	AAAACAGGGGGAGGGCTCACGCCGCG	72.
V14	AAAACACCCCCTCCACCCCGCCTCCG	73.
V15	AAAACCCGGAGCGCGGCGTGAGCCCG	74.
V16	AAAACGAGGCGGGGTGGAGGGGGTCG	75.
N1	AAAACGGCCAGATGGCTCCGCGCTCG	76.
N2	AAAACTAACCCCTTCCGCGCCGCGCG	77.
N3	AAAACGCCGGCCCGCGCCGCGCCGCG	78.
N4	AAAACGCGGGGGAGGGCGGCGCGGCG	79.
N5	AAAACGGTTGGCTGGTTATAACCGCG	80.
N6	AAAACTTCCGAACAGCTCCGCGCACG	81.

These fragments were ligated into BsmBI-digested plasmid pMLM3636 to yield DNA encoding a chimeric ~102 nt single-chain guide RNA (Mali et al., Science. 2013 Feb 15;339(6121):823-6; Hwang et al., Nat Biotechnol. 2013 Mar;31(3):227-9) expressed by a human U6 promoter. The pMLM3636 plasmid and its full DNA sequence are available from Addgene. See Fig. 4.

To engineer a Cas9-activator the D10A, H840A catalytic mutations (previously described in Jinek et al., 2012; and Qi et al., 2013) were introduced into either the wild-type or a codon-optimized Cas9 sequence (Fig. 5). These mutations render the Cas9 catalytically inactive so that it will no longer induce double-strand breaks. In one construct, a triple flag tag, nuclear localization signal and the VP64 activation domain were fused to the C-terminus of the inactive Cas9 (Fig. 6). Expression of this fusion protein was driven by the CMV promoter.

Construction of dCas-VP64 expression plasmids was performed as follows. DNA encoding the Cas9 nuclease harboring inactivating D10A/H840A mutations (dCas9) was amplified by PCR from plasmid pMJ841 (Addgene plasmid #39318) using primers that add a T7 promoter site 5' to the start codon and a nuclear localization signal at the carboxy-terminal end of the Cas9 coding sequences and cloned into a plasmid containing a CMV promoter as previously described (Hwang et al., Nat Biotechnol 31, 227–229 (2013)) to yield plasmid pMLM3629. Oligonucleotides encoding a triple FLAG epitope were annealed and cloned into XhoI and PstI sites in plasmid pMLM3629 to generate plasmid pMLM3647 expressing dCas9 with a C-terminal flag FLAG tag. DNA sequence encoding a Gly_4Ser linker followed by the synthetic VP64 activation domain was introduced

downstream of the FLAG-tagged dCas9 in plasmid pMLM3647 to yield plasmid pSL690. The D10A/H840A mutations were also introduced by QuikChange site-directed mutagenesis (Agilent) into plasmid pJDS247, which encodes a FLAG-tagged Cas9 sequence that has been codon optimized for expression in human cells, to yield plasmid pMLM3668. DNA sequence encoding the Gly_4Ser linker and the VP64 activation domain were then cloned into pMLM3668 to yield a codon-optimized dCas9-VP64 expression vector named pMLM3705.

Cell Culture, Transfection and ELISA Assays were performed as follows. Flp-In T-Rex 293 cells were maintained in Advanced DMEM supplemented with 10% FBS, 1% penstrep and 1% Glutamax (Invitrogen). Cells were transfected by Lipofectamine LTX (Invitrogen) according to manufacturer's instructions. Briefly, 160,000 293 cells were seeded in 24-well plates and transfected the following day with 250ng gRNA plasmid, 250ng Cas9-VP64 plasmid, 30ng pmaxGFP plasmid (Lonza), 0.5ul Plus Reagent and 1.65ul Lipofectamine LTX. Tissue culture media from transfected 293 cells was harvested 40 hours after transfection, and secreted VEGF-A protein assayed using R&D System's Human VEGF-A ELISA kit "Human VEGF Immunoassay."

16 sgRNAs were constructed for target sequences within three DNase I hyper-sensitive sites (**HSSs**) located upstream, downstream or at the transcription start site of the human *VEGFA* gene in 293 cells (**Fig. 1E**).

Before testing the abilities of the 16 *VEGFA*-targeted gRNAs to recruit a novel dCas9-VP64 fusion protein, each of these gRNAs was first assessed for its ability to direct Cas9 nuclease to its intended target site in human 293 cells. For this purpose, gRNA and Cas9 expression vectors were transfected in a 1:3 ratio because previous optimization experiments demonstrated a high level of Cas9-induced DNA cleavage in U2OS cells using this ratio of plasmids.

Transfections of 293 cells were performed as described above for the dCas9-VP16 *VEGFA* experiments except that cells were transfected with 125 ng of plasmid encoding *VEGFA*-targeted gRNAs and 375 ng of plasmid encoding active Cas9 nuclease (pMLM3639). 40 hours post-transfection, genomic DNA was isolated using the QIAamp DNA Blood Mini kit (Qiagen) according to manufacturer's instructions. PCR amplification of the three different targeted regions in the *VEGFA* promoter was performed using Phusion Hot Start II high-fidelity DNA polymerase (NEB) with 3% DMSO and the following touchdown PCR cycle: 10 cycles of 98 °C, 10 s; 72–62 °C, −1 °C/cycle, 15 s; 72 °C, 30 s, followed by 25 cycles of 98 °C, 10 s; 62 °C, 15 s; 72 °C, 30 s. The -500 region was

amplified using primers oFYF434 (5'- TCCAGATGGCACATTGTCAG-3' (SEQ ID NO:82)) and oFYF435 (5'- AGGGAGCAGGAAAGTGAGGT-3' (SEQ ID NO:83)). The region around the transcription start site was amplified using primers oFYF438 (5'- GCACGTAACCTCACTTTCCT-3' (SEQ ID NO:84)) and oFYF439 (5'- CTTGCTACCTCTTTCCTCTTTCT-3' (SEQ ID NO:85)). The +500 region was amplified using primers oFYF444 (5'- AGAGAAGTCGAGGAAGAGAGAG-3' (SEQ ID NO:86)) and oFYF445 (5'- CAGCAGAAAGTTCATGGTTTCG-3' (SEQ ID NO:87)). PCR products were purified using Ampure XP beads (Agencourt) and T7 Endonuclease I assays were performed and analyzed on a QIAXCEL capillary electrophoresis system as previously described (Reyon et al., Nat Biotech 30, 460-465 (2012)).

All 16 gRNAs were able to mediate the efficient introduction of Cas9 nuclease-induced indel mutations at their respective target sites as assessed using a previously described T7E1 genotyping assay (Table 2). Thus all 16 gRNAs can complex with Cas9 nuclease and direct its activity to specific target genomic sites in human cells.

Table 2. Frequencies of indel mutations induced by *VEGFA*-targeted gRNAs and Cas9 nuclease

gRNA	Mean Indel Mutation Frequency (%) ± SEM
V1	18.05 ± 0.47
V2	41.48 ±0.62
V3	33.22 ± 1.05
V4	16.97 ± 0.06
V5	7.46 ± 0.50
V6	16.99 ± 0.51
V7	1.42 ± 0.11
V8	34.07 ± 0.90
V9	24.53 ± 1.40
V10	35.65 ± 1.35
V11	4.45 ± 0.22
V12	23.95 ± 0.41
V13	9.45 ± 0.74
V14	12.17 ± 0.36
V15	14.28 ± 0.54
V16	18.82 ± 1.48

To test whether dCas9-VP64 protein could also be targeted to specific genomic sites in human cells by these same gRNAs, Enzyme-Linked Immunoblot Assays of VEGFA protein were performed as follows. Culture medium of Flp-In T-Rex HEK293 cells transfected with plasmids encoding VEGFA-targeted sgRNA and dCas9-VP64 was harvested 40 hours post-transfection and VEGFA protein expression was measured by

ELISA as previously described (Maeder et al., Nat Methods 10, 243–245 (2013)). Fold-activation of VEGFA expression was calculated by dividing the concentration of VEGFA protein in media from cells in which both a sgRNA and dCas9-VP64 were expressed by the concentration of VEGFA protein in media from cells in which an off-target sgRNA (targeted to a sequence in the *EGFP* reporter gene) and dCas9-VP64 were expressed.

15 of the 16 gRNAs tested induced significant increases in VEGFA protein expression when co-expressed with dCas9-VP64 in human 293 cells (Fig. 2A). The magnitude of VEGFA induction observed ranged from two- to 18.7-fold-activation with a mean of five-fold-activation. Control experiments revealed that expression of each of the 16 gRNAs alone, dCas9-VP64 alone, and dCas9-VP64 together with an "off-target" gRNA designed to bind an EGFP reporter gene sequence all failed to induce elevated VEGFA expression (Fig. 2A), demonstrating that co-expression of a specific gRNA and the dCas9-VP64 protein are both required for promoter activation. Thus dCas9-VP64 is stably expressed and can be directed by gRNAs to activate transcription of specific genomic loci in human cells. The greatest increase in VEGFA was observed in cells transfected with gRNA3, which induced protein expression by 18.7-fold. Interestingly, the three best gRNAs, and 6 of the 9 gRNAs capable of inducing expression by 3-fold or more, target the -500 region (~500bp upstream of the transcription start site).

Because in one aspect the system described herein uses variable gRNAs to recruit a common dCas9-VP64 activator fusion, one can envision that the expression of multiple guide RNAs in a single cell might enable multiplex or combinatorial activation of endogenous gene targets. To test this possibility, 293 cells were transfected with dCas9-VP64 expression plasmid together with expression plasmids for four gRNAs (V1, V2, V3, and V4) that each individually induced expression from the VEGFA promoter. Co-expression of all four gRNAs with dCas9-VP64 induced synergistic activation of VEGFA protein expression (i.e., a fold-activation greater than the expected additive effects of each individual activator) (Fig. 2B). In addition, various combinations of three of these four activators also activated the VEGFA promoter synergistically (Fig. 2B). Because synergistic activation of transcription is believed to result from the recruitment of multiple activator domains to a single promoter, multiple gRNA/dCas9-VP64 complexes are likely to be simultaneously binding to the VEGFA promoter in these experiments.

These experiments demonstrate that co-expression of a Cas9-HFD, e.g., a Cas9-activator protein (harboring the VP64 transcriptional activation domain) and a sgRNA with 20nt of sequence complementarity to sites in the human VEGF-A promoter in human

HEK293 cells can result in upregulation of VEGF-A expression. Increases in VEGF-A protein were measured by ELISA assay and it was found that individual gRNAs can function together with a Cas9-activator fusion protein to increase VEGF-A protein levels by up to ~18-fold (Fig. 2A). Additionally, it was possible to achieve even greater increases in activation through transcriptional synergy by introducing multiple gRNAs targeting various sites in the same promoter together with Cas9-activator fusion proteins (Fig. 2B).

Example 2. Engineering CRISPR/Cas Activator System targeting the endogenous human *NTF3* gene

To extend the generality of the present findings, we tested whether the RNA-guided activator platform could be used to induce the expression of the human NTF3 gene. To do this, six sgRNAs were designed to a predicted DNase I hypersensitive site (HSS) in the human *NTF3* promoter and plasmids expressing each of these gRNAs were co-transfected with a plasmid encoding dCas9-VP64 protein that had been codon optimized for human cell expression (Fig. 3A).

All six gRNAs tested induced significant increases in NTF3 transcript levels as detected by quantitative RT-PCR (Fig. 3B). Although fold-activation values for these six RNA-guided activators could not be accurately calculated (because basal levels of transcript were essentially undetectable), the mean levels of activated NTF3 mRNA expression varied over a four-fold range. Decreasing the amounts of gRNA and dCas9-VP64 expression plasmids transfected resulted in less activation of the NTF3 gene (Fig. 3B), demonstrating a clear dose-dependent effect.

In addition, 293 cells were co-transfected with dCas9-VP64 and NTF3-targeted gRNA expression plasmids alone and in single and double combinations. Relative expression of NTF3 mRNA was detected by quantitative RT-PCR and normalized to a GAPDH control (deltaCt x 10^4). In all of these experiments the amount of each individual gRNA expression plasmid used for transfection was the same. FIG. 3B shows that this multiplex gRNA expression induced synergistic activation of NTF3 mRNA expression by dCas9-VP64 protein.

Example 3. Engineering CRISPR/Cas-MS2, -Csy4 and –Lambda N Fusion Systems – Creating Biological Tethers

Fusion proteins are made in which an MS2 coat protein, Csy4 nuclease (preferably catalytically inactive Csy4, e.g., the H29A mutant described in Haurwitz et al.

329(5997):1355-8 (2010)), or the lambda N are fused to the N- or C-terminus of the inactivated dCas9. MS2 and lambda N are bacteriophage proteins that bind to a specific RNA sequence, and thus can be used as adapters to tether to the dCas9 protein a heterologous RNA sequence tagged with the specific MS2 or lambda N RNA binding sequence. dCas9-MS2 fusions or dCas9-lambda N fusions are co-expressed with chimeric long non-coding RNAs (lncRNAs) fused to the MS2 or lambda N stem loop recognition sequence on either their 5' or 3' end. Chimeric Xist or chimeric RepA lncRNAs will be specifically recruited by the dCas9 fusions and the ability of this strategy to induce targeted silencing will be assayed by measuring target gene expression. The system will be optimized by testing various alterations to the coat proteins and chimeric RNAs. The N55K and deltaFG mutations to the MS2 coat protein have been previously demonstrated to prevent protein aggregation and increase affinity for the stem-loop RNA. Additionally, we will test the high-affinity C-loop RNA mutant reported to increase affinity for the MS2 coat protein. Exemplary sequences for the MS2 and lambda N proteins are given below; the MS2 functions as a dimer, therefore the MS2 protein can include a fused single chain dimer sequence.

1. Exemplary sequences for Fusions of single MS2 coat protein (wt, N55K or deltaFG) to the N-terminus or C-terminus of the dCas9.

MS2 coat protein amino acid sequence:
MASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQ
NRKYTIKVEVPKVATQTVGGVELPVAAWRSYLNMELTIPIFATNSDCELIVKAMQG
LLKDGNPIPSAIAANSGIY (SEQ ID NO:88)

MS2 N55K:
MASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQ
KRKYTIKVEVPKVATQTVGGVELPVAAWRSYLNMELTIPIFATNSDCELIVKAMQG
LLKDGNPIPSAIAANSGIY (SEQ ID NO:89)

MS2deltaFG:
MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQ
NRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAA
NSGIY (SEQ ID NO:90)

2. Exemplary sequences for Fusions of fused dimeric MS2 coat protein (wt, N55K or deltaFG) to the N-terminus or C-terminus of dCas9.

Dimeric MS2 coat protein:
MASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQ
NRKYTIKVEVPKVATQTVGGVELPVAAWRSYLNMELTIPIFATNSDCELIVKAMQG
LLKDGNPIPSAIAANSGLYGAMASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWI
SSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKVATQTVGGVELPVAAWRSYLN
MELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSLIN (SEQ ID NO91

MASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQ
KRKYTIKVEVPKVATQTVGGVELPVAAWRSYLNMELTIPIFATNSDCELIVKAMQG
LLKDGNPIPSAIAANSGLYGAMASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWI
SSNSRSQAYKVTCSVRQSSAQKRKYTIKVEVPKVATQTVGGVELPVAAWRSYLN
MELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSLIN (SEQ ID NO:92)

Dimeric MS2deltaFG:
MASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQ
KRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAA
NSGLYGAMASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCS
VRQSSAQKRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGN
PIPSAIAANSLIN (SEQ ID NO:93)

3. Exemplary sequences for Fusions of Lambda N to N-terminus or C-terminus of dCas9.

Lambda N amino acid sequence:
MDAQTRRRERRAEKQAQWKAAN (SEQ ID NO:94) or
MDAQTRRRERRAEKQAQWKAANPLLVGVSAKPVNRPILSLNRKPKSRVESALNPI
DLTVLAEYHKQIESNLQRIERKNQRTWYSKPGERGITCSGRQKIKGKSIPLI (SEQ
ID NO:95)

4. Exemplary sequence for Fusions of Csy4 to N-terminus or C-terminus of dCas9

Exemplary sequences for Cys4 are given in Haurwitz et al. 329(5997):1355-8 (2010), e.g., the inactivated form.

The constructs are expressed in cells also expressing a regulatory RNA, e.g., a long non-coding RNA (lncRNA) such as HOTAIR, HOTTIP, XIST or XIST RepA, that has been fused with the cognate stem-loop recognition sequence for the lambda N or MS2 on either its 5' or 3' end. The wild type and high-affinity sequences for MS2 are AAACAUGAGGAUUACCCAUGUCG (SEQ ID NO:96) and AAACAUGAGGAUCACCCAUGUCG (SEQ ID NO:97), respectively (see Keryer-Bibens et al., supra, FIG. 2); the nutL and nutR BoxB sequences to which lambda N binds are GCCCUGAAGAAGGGC (SEQ ID NO:98) and GCCCUGAAAAAGGGC (SEQ ID NO:99), respectively. The sequence to which Csy4 binds is GTTCACTGCCGTATAGGCAG (truncated 20 nt) (SEQ ID NO:100) or GUUCACUGCCGUAUAGGCAGCUAAGAAA (SEQ ID NO:101).

The binding of the dCas9/MS2 to a target site in a cell expressing an MS2-binding sequence tagged lncRNA recruits that lncRNA to the dCas9 binding site; where the lncRNA is a repressor, e.g., XIST, genes near the dCas9 binding site are repressed. Similarly, binding of the dCas9/lambdaN to a target site in a cell expressing an lambdaN-binding sequence tagged lncRNA recruits that lncRNA to the dCas9 binding site.

Example 4. Engineering CRISPR/Cas-HP1 Fusion Systems –Sequence-Specific Silencing

The dCas9 fusion proteins described herein can also be used to target silencing domains, e.g., Heterochromatin Protein 1 (HP1, also known as swi6), e.g., HP1α or HP1β. Truncated versions of HP1α or HP1β in which the chromodomain has been removed can be targeted to specific loci to induce heterochromatin formation and gene silencing. Exemplary sequences of truncated HP1 fused to dCas9 are shown in Figs. 8A-8B. The HP1 sequences can be fused to the N- or C-terminus of the inactivated dCas9 as described above.

Example 5. Engineering CRISPR/Cas-TET Fusion Systems –Sequence-Specific Demethylation

The dCas9 fusion proteins described herein can also be used to target enzymes that modify the methylation state of DNA (e.g., DNA methyltransferase (DNMT) or TET proteins). Truncated versions of TET1 can be targeted to specific loci to catalyze DNA demethylation. Exemplary sequences of truncated TET1 fused to dCas9 are shown in Fig. 9. The TET1 sequence can be fused to the N- or C-terminus of the inactivated dCas9 as described above.

Example 6. Engineering Optimized CRISPR/Cas-VP64 Fusions

The activities of dCas9-based transcription activators harboring the VP64 activation domain were optimized by varying the number and position of the nuclear localization signal(s) (NLS) and 3xFLAG-tags within these fusions (Figure 10). dCas9-VP64 fusions that contain both an N-terminal NLS and an NLS that lies between the dCas9 and VP64 sequences consistently induce higher levels of target gene activation, perhaps resulting from enhanced nuclear localization of the activator (Figure 10). Furthermore, even greater levels of activation were observed when a 3xFLAG tag was placed between the C-terminal end of dCas9 and the N-terminal end of VP64. The 3xFLAG tag may act as an artificial linker, providing necessary spacing between dCas9 and VP64 and perhaps allowing for better folding of the VP64 domain (that may not be possible when constrained near dCas9) or better recognition of VP64 by transcriptional mediator complexes that recruit RNA polymerase II. Alternatively, the negatively charged 3xFLAG tag might also function as a fortuitous transcriptional activation domain, enhancing the effects of the VP64 domain.

Example 7. Optimized CatalyticallyCatlytically Inactive Cas9 Proteins (dCas9)

Additional optimization of the activities of dCas9-VP64 activators was performed by changing the nature of the inactivating mutations that abolish the nuclease activity of Cas9 in the dCas9 domain (Figure 11A-B). In published studies to date, the catalytic residues D10 and H840 were mutated to alanine (D10A and H840A) to disrupt the active site networks that mediate the hydrolysis of DNA. It was hypothesized that alanine substitutions at these positions might result in destabilization of dCas9 and therefore suboptimal activity. Therefore, more structurally conservative substitutions at D10 or H840 (for example, to asparagine or tyrosine residues: D10N, H840N, and H840Y) were tested to see if they might lead to greater gene activation by dCas9-VP64 fusions bearing these different mutations. When dCas9-VP64 variants bearing these variant substitutions were co-transfected into HEK293 cells with three gRNAs targeting upstream regions of the endogenous human VEGFA gene, greater VEGFA protein expression was observed for all but one of these variants (Figure 11A). However, this effect was not as significant when the dCas9-VP64 variants were co-transfected with only one of these gRNAs (Figure 11A), or when transfected into a HEK293 derivative cell-line that expresses a single VEGFA-targeted gRNA (Figure 11B).

OTHER EMBODIMENTS

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

WHAT IS CLAIMED IS:

1. A method of increasing expression of an endogenous target gene in a mammalian cell *in vitro* or *ex vivo*, the method comprising expressing in the mammalian cell a fusion protein comprising a catalytically inactive *S. pyogenes* Cas9 protein linked to a heterologous functional domain, and expressing one or more guide RNAs directed to the target gene,

wherein the catalytically inactive Cas9 protein comprises mutations at D10A and H840A, and the heterologous functional domain is a transcriptional activation domain,

wherein the expressed Cas9 fusion protein is directed to the target gene by the one or more guide RNAs, and wherein the transcriptional activation domain mediates the increased expression, wherein the transcriptional activation domain comprises NF-κB p65.

2. The method of any one of the preceding claims, wherein the heterologous functional domain is linked to the N terminus or the C terminus of the catalytically inactive Cas9 protein, with an optional intervening linker, wherein the linker does not interfere with activity of the fusion protein.

3. The method of any one of the preceding claims, further comprising one or both of a nuclear localization sequence and one or more epitope tags on the N-terminus, the C-terminus, or in between the catalytically inactive Cas9 protein and the heterologous functional domain, optionally with one or more intervening linkers.

4. The method of claim 3, wherein the one or more epitope tags is selected from the group consisting of c-myc, 6His, and FLAG tags.

5. The method of any one of the preceding claims, wherein expressing the fusion protein in the mammalian cell comprises contacting the mammalian cell with an expression vector comprising a promoter operably linked to a nucleic acid encoding the fusion protein.

6. The method of any one of the preceding claims, wherein expressing the one or more guide RNAs directed to the target gene in the mammalian cell comprises contacting the

mammalian cell with one or more expression vectors comprising a promoter operably linked to nucleic acid sequences encoding the one or more guide RNAs directed to the target gene.

7. The method of any one of the preceding claims, wherein the one or more gRNAs are single guide RNAs directed to the promoter of the target gene.

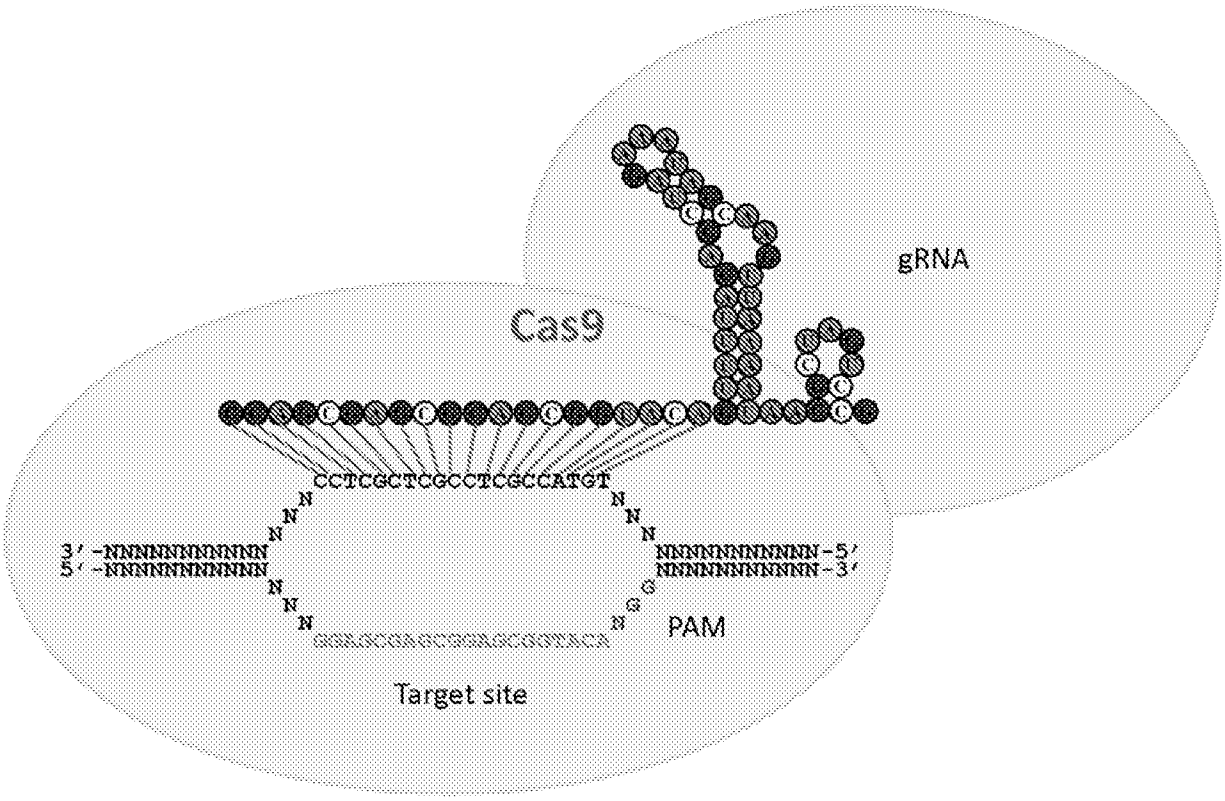


FIG. 1A

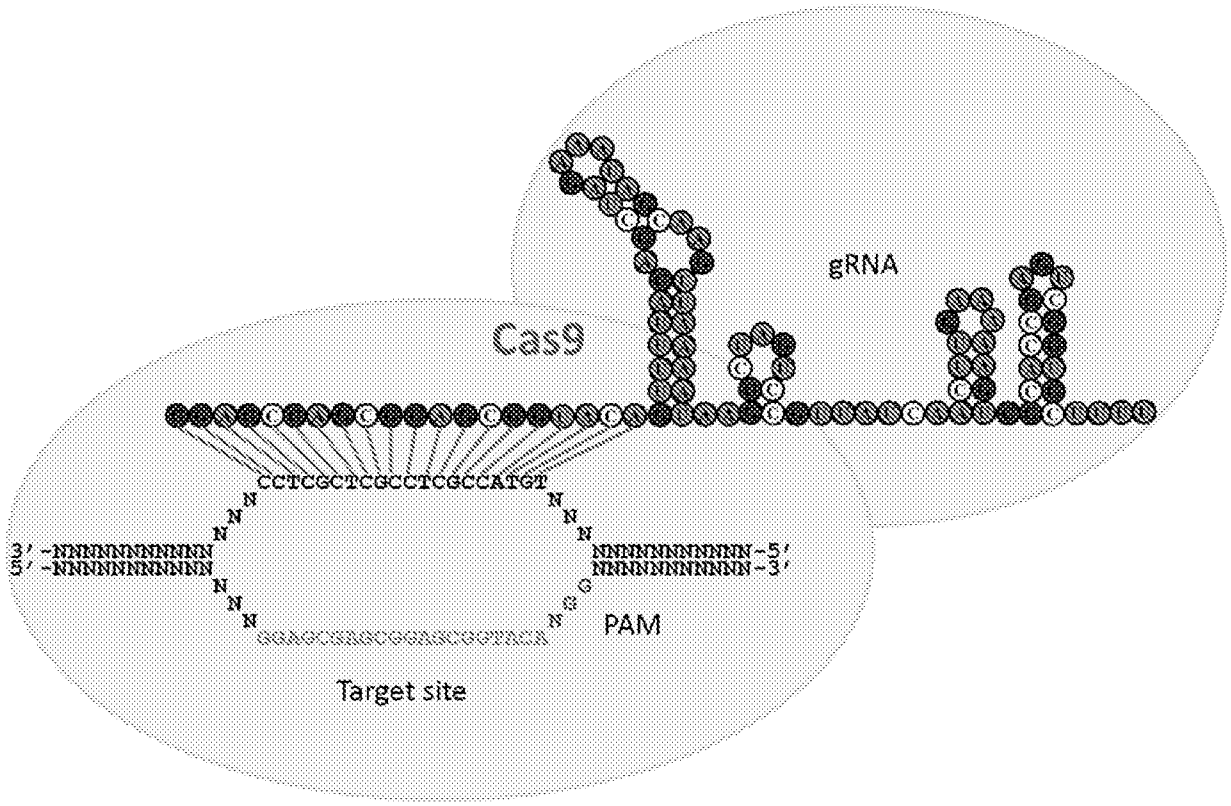


FIG. 1B

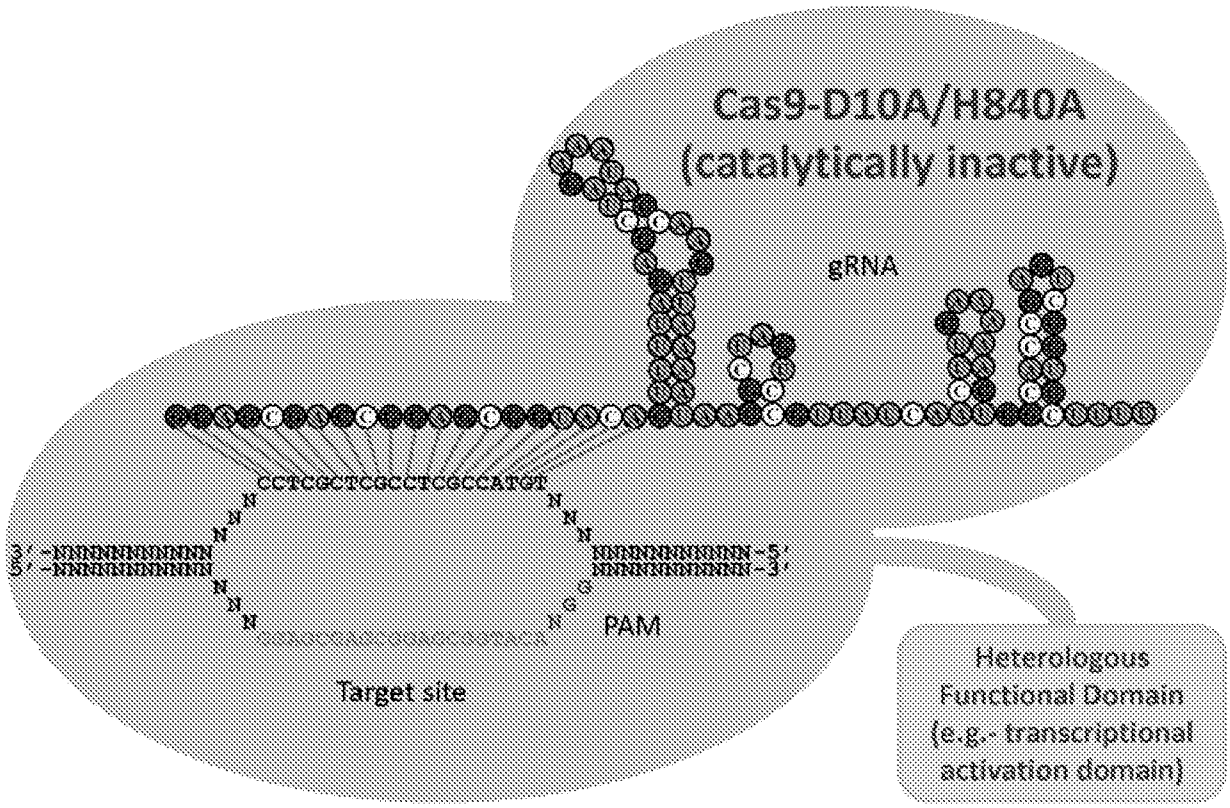


FIG. 1C

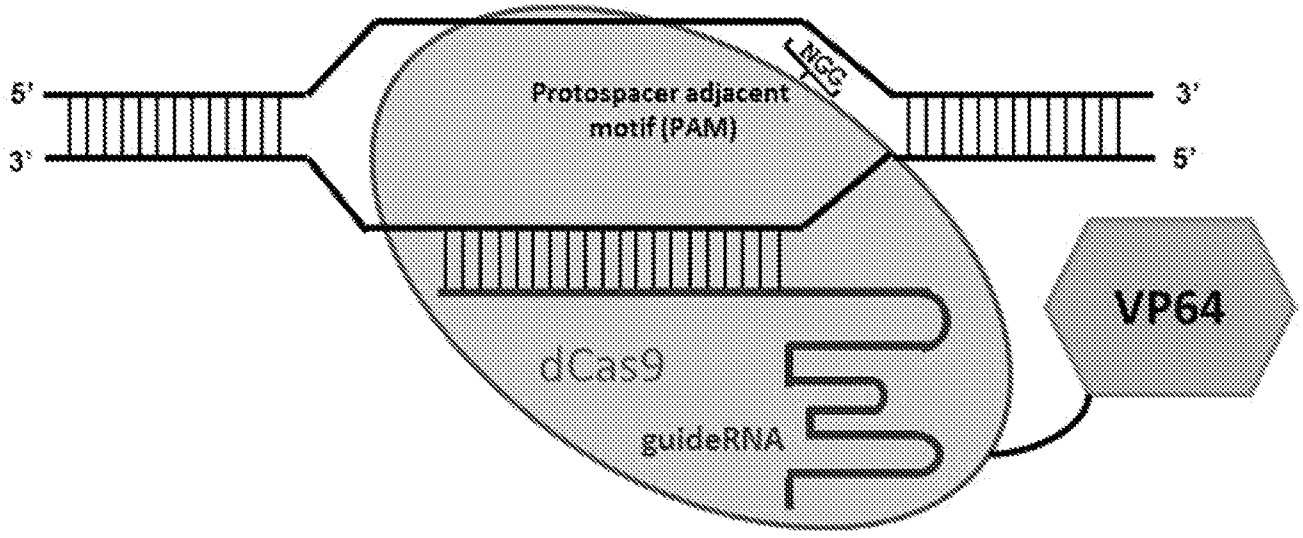


FIG. 1D

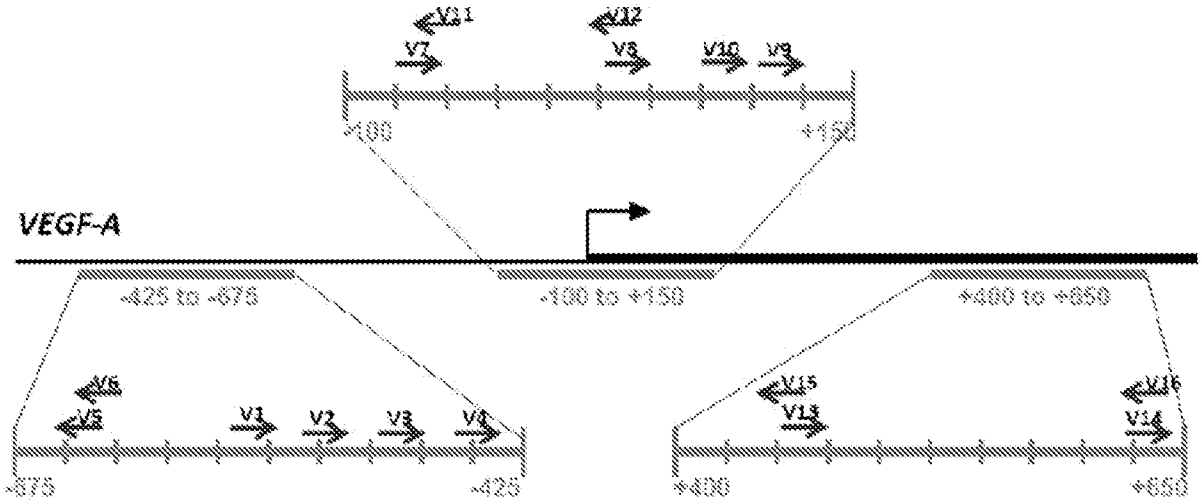


FIG. 1E

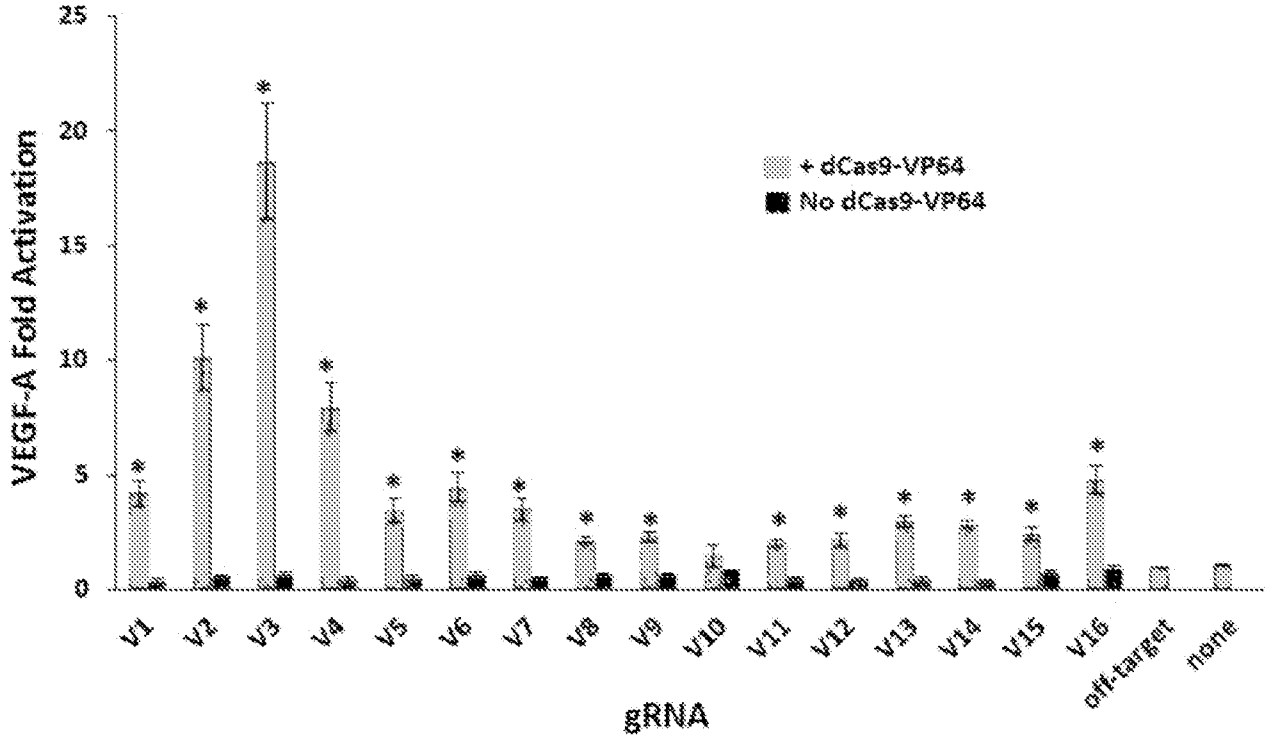


FIG. 2A

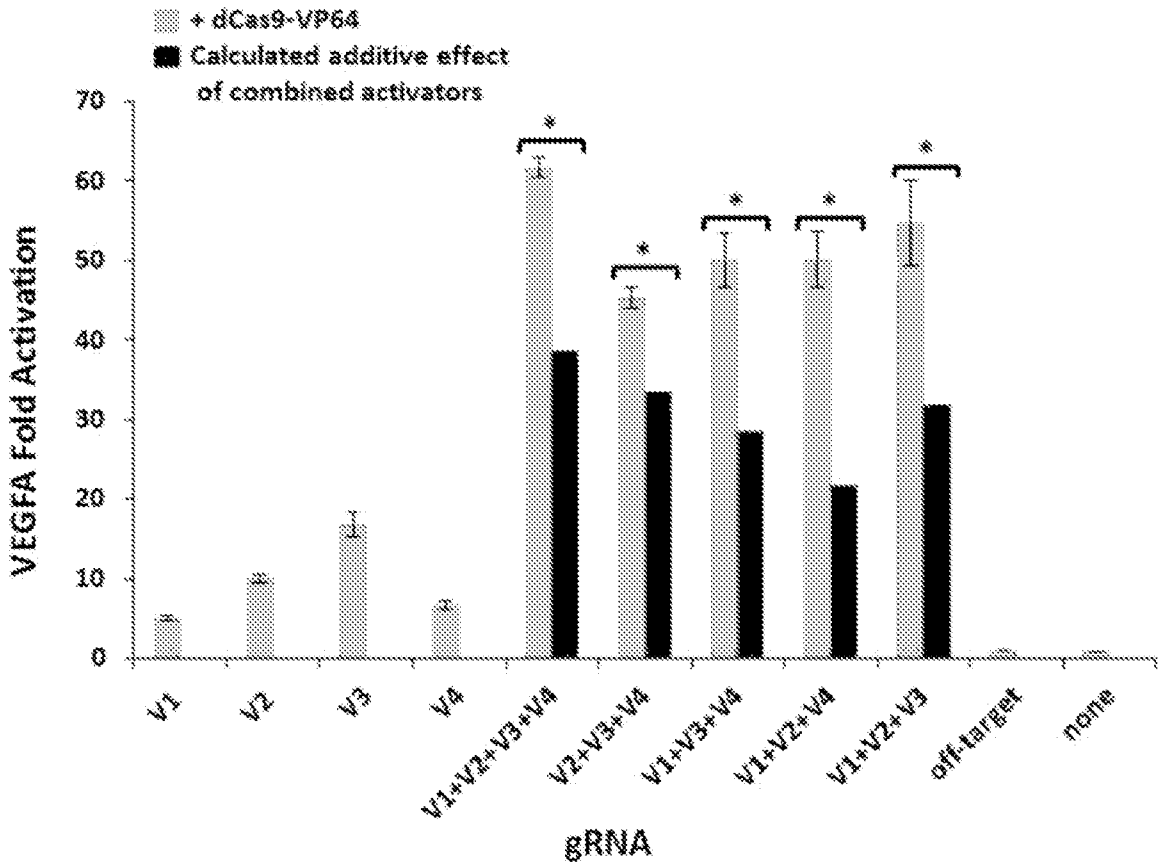


FIG. 2B

FIG. 3A

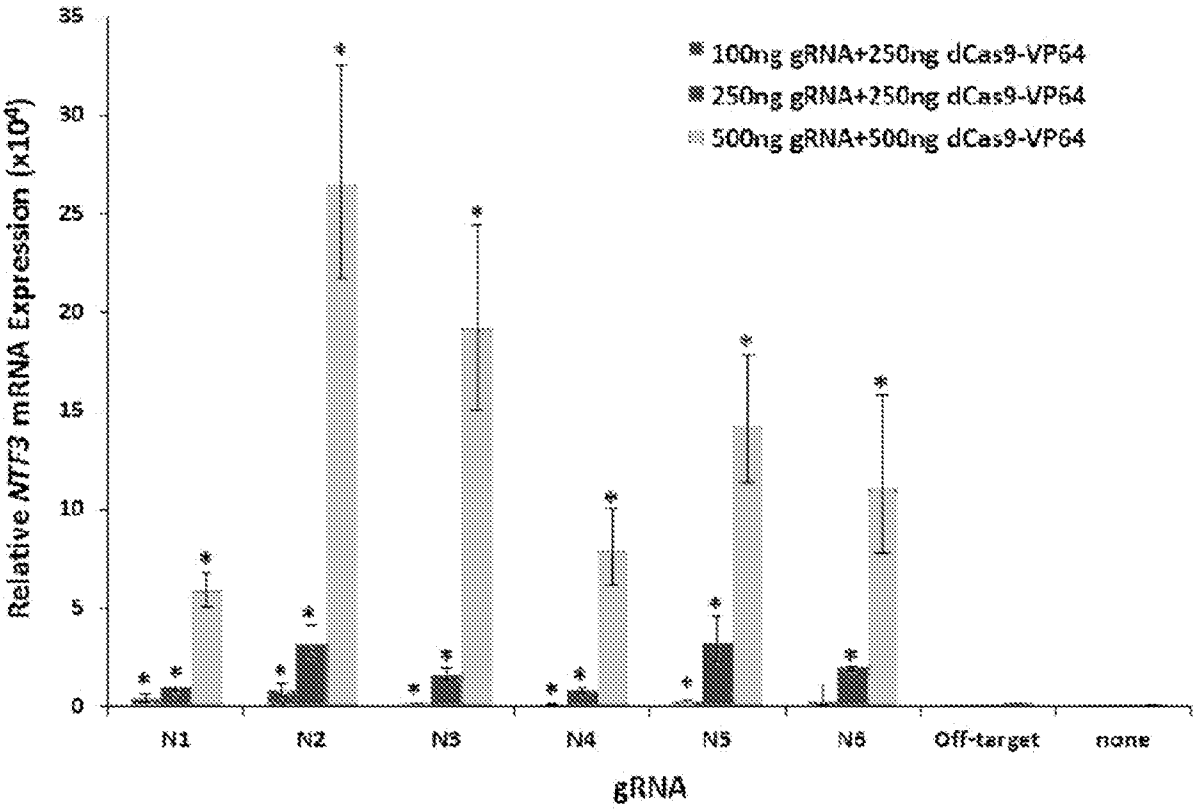


FIG. 3B

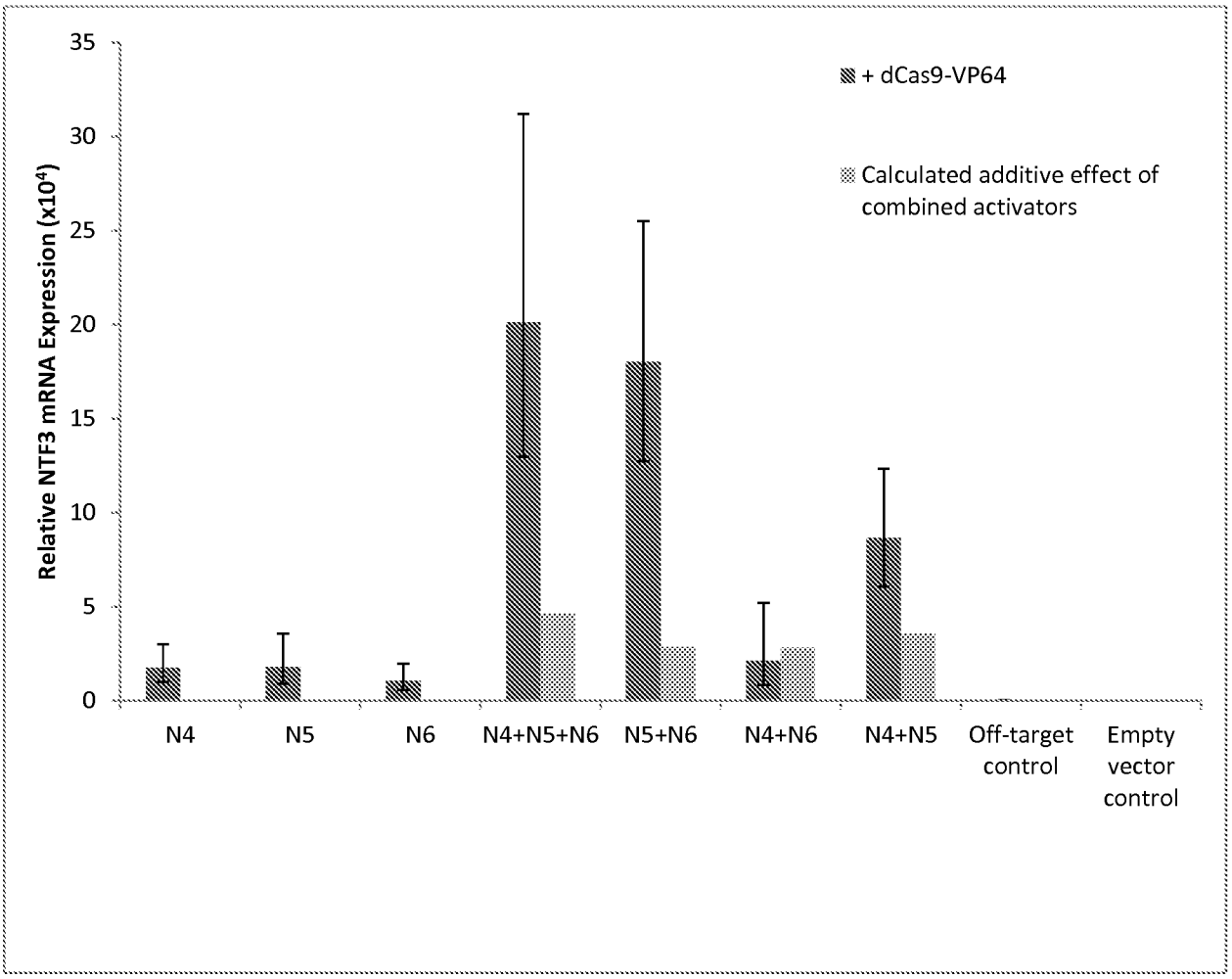


FIG. 3C

FIG. 4 - Guide RNA expression vector sequence

GACGTCGCTAGCTGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGGTACCAA
GGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTA
GAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAG
TAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGT
AACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCNNNNNNNNNN
NNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAA
GTGGCACCGAGTCGGTGCTTTTTTTAAGCTTGGGCCGCTCGAGGTACCTCTCTACATATGACATGTGA
GCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCG
CCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA
GATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGA
TACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAG
TTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCG
CCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCC
ACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAA
CTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAA
GAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAG
CAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCA
GTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCC
TTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTAC
CAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACT
CCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGC
GAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGA
AGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAG
TTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGT
TTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGC
AAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACT
CATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTG
GTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCA
ATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGG
GCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACT
GATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCA
AAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAG
CATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAG
GGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCT (SEQ ID NO:107)

FIG. 5 - CMV-T7-Cas9 D10A/H840A-3XFLAG-VP64:

ATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGAC
CTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGC
AGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGG
GAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATG
GGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGAG
ATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACCATGGATAAGAAATACTCAATAGGCTTAGc
TATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTT
CTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAG
CGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACA
GGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGG
AAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATA
TCCAACTATCTATCATCTGCGAAAAAAATTGGTAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGG
CCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTG
GACAAACTATTTATCCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCCTATTAACGCAAGTGGAGT
AGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCG
GTGAGAAGAAAAATGGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCAAAT
TTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATTTATTGGC
GCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGATAT
CCTAAGAGTAAATACTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATC
AAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAAT
CAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAAT
TTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACG
GACCTTTGACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAA
GACTTTTATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTT
GGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGA
ATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAAAAT
CTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAA
GGTCAAATATGTTACTGAAGGAATGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGAT
TTACTCTTCAAAACAAATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTT
TGATAGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGTACCTACCATGATTTGCTAAAAA
TTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACC
TTATTTGAAGATAGGGAGATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGA
AACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAA
GCAATCTGGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCC
ATGATGATAGTTTGACATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATGA
ACATATTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAGTTGTTGATGAAT
TGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAGACAACTC
AAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAG
ATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGG
AAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATgcCATTGTTCCAC
AAAGTTTCCTTAAAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAATCGTGGTAAATCGGA
TAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATC
ACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACTTGATAAAGCTGGTTTTA
TCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATAC
TAAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACT
TCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAAT
GCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGT

FIG. 5 - CMV-T7-Cas9 D10A/H840A-3XFLAG-VP64:

TTATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACT
CTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAA
ACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCC
ATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAA
AAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCC
AACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAA
AGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTAAA
GGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTC
GTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGA
ATTTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTT
GTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAGC
AGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAA
AATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGAT
CGTAAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGA
AACACGCATTGATTTGAGTCAGCTAGGAGGTGACGGTTCTCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTA
CAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGctgcaggaggcgg
tggaagcGGGCGCGCCGACGCGCTGGACGATTTCGATCTCGACATGCTGGGTTCTGATGCCCTCGATGACTTTG
ACCTGGATATGTTGGGAAGCGACGCATTGGATGACTTTGATCTGGACATGCTCGGCTCCGATGCTCTGGACGA
TTTCGATCTCGATATGTTATAAccggtCATCATCACCATCACCATTGAGTTTAAACCCGCTGATCAGCCTCGACTG
TGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCA
CTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGG
GGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCT
ATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCGATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCA
TGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAA
GTGTAAAGCCTAGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGT
CGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGC
GCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCA
AAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCA
AAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCAC
AAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGA
AGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGC
GTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGT
GCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGA
CACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACA
GAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGC
CAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTT
TGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCT
GACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGA
TCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAAT
GCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGT
AGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACC
GGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATC
CGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAAC
GTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCA
ACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTT
GTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCC
ATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCG
AGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTG
GAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCG
TGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAAT

FIG. 5 - CMV-T7-Cas9 D10A/H840A-3XFLAG-VP64:

GCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAA
GCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGT
TCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCGACGGATCGGGAGATCGATCTCCCGATCCCCTAGGG
TCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGT
CGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCT
GCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACT
AGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTAC
GGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATA
GTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTAC
ATCAAGTGTATC**C (SEQ ID NO:108)**

FIG. 6 - MV-T7-Cas9 recoded D10A/H840A-3XFLAG-VP64

ATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGAC
CTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGC
AGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGG
GAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATG
GGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGAG
ATCCGCGGCCGCTAATACGACTCACTATAGGGAGAGCCGCCACCATGGATAAAAAGTATTCTATTGGTTTAGc
CATCGGCACTAATTCCGTTGGATGGGCTGTCATAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTG
TTGGGGAACACAGACCGTCATTCGATTAAAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGG
CAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTACTTAC
AAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGTTTGGAAGAGTCCTTCCTTGTC
GAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAGGTGGCATATCATGAAAAG
TACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGAGGTTAATCTACTT
GGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAATCCGGACAACTCGGAT
GTCGACAAACTGTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGCAAGTG
GCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCACAATT
ACCCGGAGAGAAGAAAAATGGGTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAG
TCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATC
TACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTA
TCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATG
AACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATT
CTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTT
TATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACTCAATCGCGAAGATCTACT
GCGAAAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTT
AGAAGGCAGGAGGATTTTTATCCGTTCCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCA
TACCTTACTATGTGGGACCCCTGGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAA
CGATTACTCCATGGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGAC
CAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTCACAGTGT
ACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCGCCTTTCTAAGCGGAGAACAGA
AGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTT
TAAGAAAATTGAATGCTTCGATTCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACG
TATCATGACCTCCTAAAGATAATTAAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAAG
ATATAGTGTTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCACCT
GTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTCGCGGAAACT
TATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAAT
AGGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCG
GACAAGGGGACTCATTGCACGAACATATTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCA
GACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGAT
GGCACGCGAAAATCAAACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAG
GGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAACGAGAAA
CTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGTTTATCTGA
TTACGACGTCGATgcCATTGTACCCCAATCCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTC
GGATAAGAACCGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGC
GGCAGCTCCTAAATGCGAAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCT
TGTCTGAACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGCATGTTGC
ACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCGGGAAGTCAAAGTAAT
CACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGATAAATAACT
ACCACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAGAAATACCCGAAGCTAGA
AAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAGAT

FIG. 6 - MV-T7-Cas9 recoded D10A/H840A-3XFLAG-VP64

AGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAA
ACGGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGC
CGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAG
ACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACT
GGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGA
GAAGGGAAAATCCAAGAAACTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTT
TGAAAAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAAACTA
CCAAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGCTTCAAAAG
GGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCCTGTATTTAGCGTCCCATTACGAGAAGTTGAAAG
GTTCACCTGAAGATAACGAACAGAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTCGACGAAATCATAGA
GCAAATTTCGGAATTCAGTAAGAGAGTCATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAAC
AAGCACAGGGATAAACCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCG
CTCCAGCCGCATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAGA
CGCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGTCACAGCTTGGGGGTGAC
GGATCCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCATGAC
ATCGATTACAAGGATGACGATGACAAGGctgcaggaggcggtggaagcGGGCGCGCCGACGCGCTGGACGATTTCG
ATCTCGACATGCTGGGTTCTGATGCCCTCGATGACTTTGACCTGGATATGTTGGGAAGCGACGCATTGGATGA
CTTTGATCTGGACATGCTCGGCTCCGATGCTCTGGACGATTTCGATCTCGATATGTTATAAccggtCATCATCACC
ATCACCATTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCT
CCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCG
CATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGA
AGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCT
CGATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCG
CTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTAGGGTGCCTAATGAGTGAGCTAA
CTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAAT
CGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGC
TCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGG
GATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCT
GGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAAC
CCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCC
GCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATC
TCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGC
CTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGT
AACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTAC
ACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTT
GATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAA
AAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGG
GATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAA
TCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATC
TGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATC
TGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCA
GCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGG
AAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTC
ACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGT
TGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACT
CATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGT
ACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAA
TACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGG
ATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTC
ACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAA

FIG. 6 - MV-T7-Cas9 recoded D10A/H840A-3XFLAG-VP64

ATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATA
CATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGAC
GTCGACGGATCGGGAGATCGATCTCCCGATCCCCTAGGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCAT
AGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTAC
AACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGAT
GTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGT
TCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGAC
CCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATG
GGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATC (SEQ ID NO:109)

FIG. 7 - Cas9-activator protein

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD**GS**pkkkrkvssDYKDHDGDYKDHDIDYKDDDDKAA**GGGGS**GRADALDDFDLDMLGSDA

LDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDML (SEQ ID NO:110)

FIG. 8A - dCas9-NLS-3XFLAG-HP1alpha

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGDGSPKKKRKVSSDYKDHDGDYKDHDIDYKDDDDKAAGGGGSMKEGENNKPREKSESNKR
KSNFSNSADDIKSKKKREQSNDIARGFERGLEPEKIIGATDSCGDLMFLMKWKDTDEADLVL
AKEANVKCPQIVIAFYEERLTWHAYPEDAENKEKETAKS (SEQ ID NO:111)

FIG. 8B- dCas9-NLS-3XFLAG-HP1beta

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGDGS PKKKRKV SSDYKDHDGDYKDHDIDYKDDDDKAAGGGGSTAHETDKSEGGKRKADSD
SEDKGEESKPKKKKEESEKPRGFARGLEPERIIGATDSSGELMFLMKWKNSDEADLVPAKEA
NVKCPQVVISFYEERLTWHSYPSEDDDKKDDKN. (SEQ ID NO:112)

box = nuclear localization signal

underline = triple flag tag

double underline = HP1alpha hinge and chromoshadow domains

FIG. 9- dCas9-3XFLAG-TET1CD

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGDGSPKKKRKVSSDYKDHDGDYKDHDIDYKDDDDKAAGGGGSLPTCSCLDRVIQKDKGPY
YTHLGAGPSVAAVREIMENRYGQKGNAIRIEIVVYTGKEGKSSHGCPIAKWVLRRSSDEEKV
LCLVRQRTGHHCPTAVMVVLIMVWDGIPLPMADRLYTELTENLKSYNGHPTDRRCTLNENRT
CTCQGIDPETCGASFSFGCSWSMYFNGCKFGRSPSPRRFRIDPSSPLHEKNLEDNLQSLATR
LAPIYKQYAPVAYQNQVEYENVARECRLGSKEGRPFSGVTACLDFCAHPHRDIHNMNNGSTV
VCTLTREDNRSLGVIPQDEQLHVLPLYKLSDTDEFGSKEGMEAKIKSGAIEVLAPRRKKRTC
FTQPVPRSGKKRAAMMTEVLAHKIRAVEKKPIPRIKRKNNSTTTNNSKPSSLPTLGSNTETV
QPEVKSETEPHFILKSSDNTKTYSLMPSAPHPVKEASPGFSWSPKTASATPAPLKNDATASC
GFSERSSTPHCTMPSGRLSGANAAAADGPGISQLGEVAPLPTLSAPVMEPLINSEPSTGVTE
PLTPHQPNHQPSFLTSPQDLASSPMEEDEQHSEADEPPSDEPLSDDPLSPAEEKLPHIDEYW
SDSEHIFLDANIGGVAIAPAHGSVLIECARRELHATTPVEHPNRNHPTRLSLVFYQHKNLNK
PQHGFELNKIKFEAKEAKNKKMKASEQKDQAANEGPEQSSEVNELNQIPSHKALTLTHDNVV
TVSPYALTHVAGPYNHWV

(SEQ ID NO:113)

box = nuclear localization signal
underline = triple flag tag
double underline = TET1CD

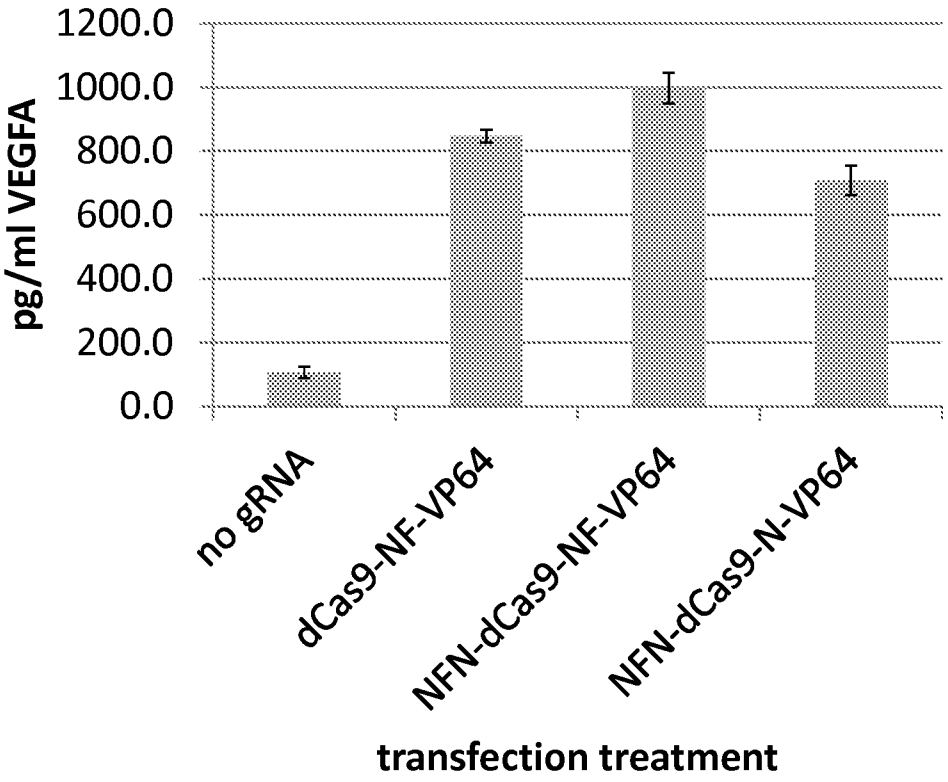


FIG. 10

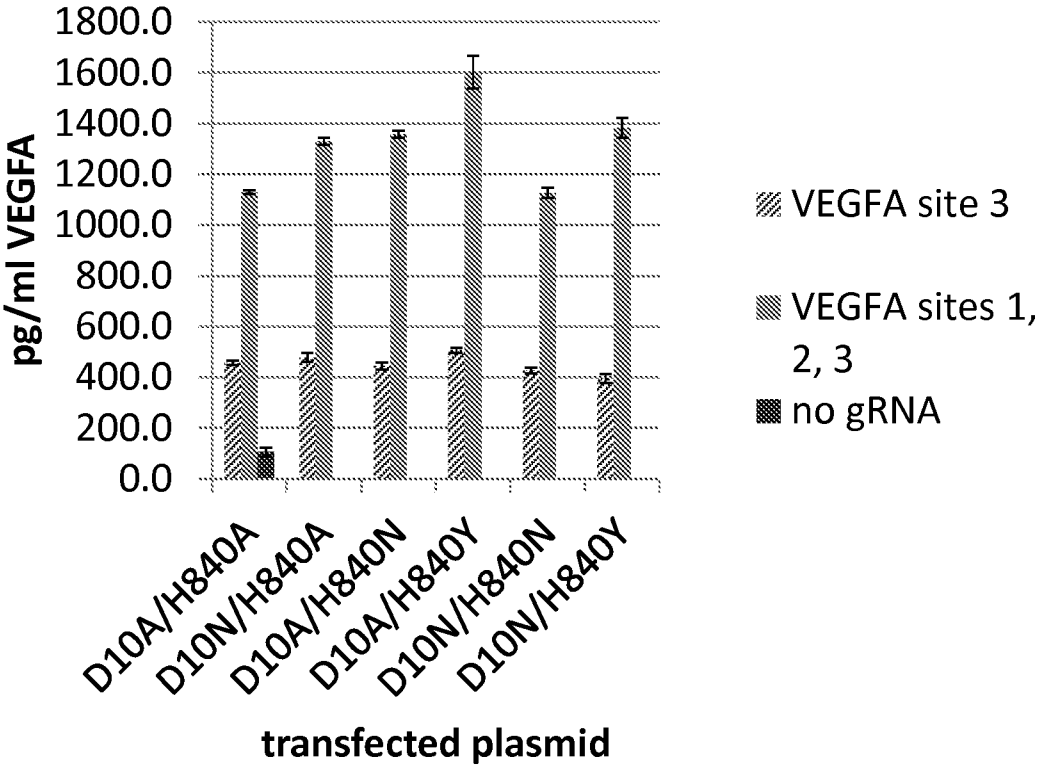


FIG. 11A

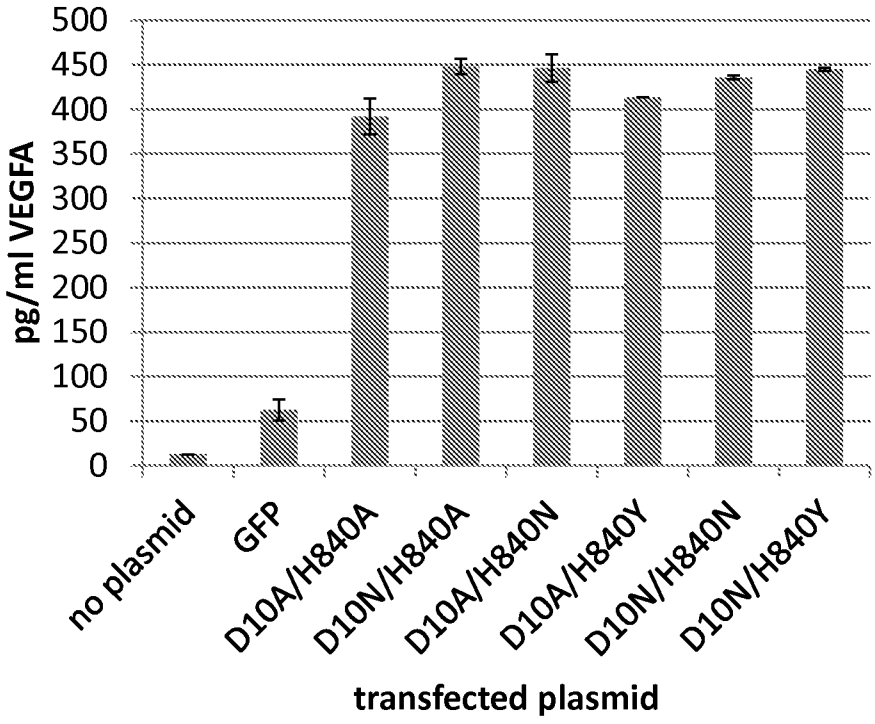


FIG. 11B

SEQUENCE LISTING

<110> THE GENERAL HOSPITAL CORPORATION

<120> RNA-GUIDED TARGETING OF GENETIC AND EPIGENOMIC REGULATORY PROTEINS TO SPECIFIC GENOMIC LOCI

<130> 00786-0882WO1

<140> PCT/US2014/027335
<141> 2014-03-14

<150> 61/921,007
<151> 2013-12-26

<150> 61/838,178
<151> 2013-06-21

<150> 61/838,148
<151> 2013-06-21

<150> 61/799,647
<151> 2013-03-15

<160> 113

<170> PatentIn version 3.5

<210> 1
<211> 262
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic guide polynucleotide

<220>
<221> modified_base
<222> (1)..(20)
<223> a, c, u, g, unknown or other and this region may encompass 17-20 nucleotides, wherein some positions may be absent

<220>
<221> modified_base
<222> (63)..(262)
<223> a, c, u, g, unknown or other and this region may encompass 0-200 nucleotides, wherein some positions may be absent

<400> 1
nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cgnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240

nnnnnnnnnn nnnnnnnnnn nn 262

<210> 2
<211> 275
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
guide polynucleotide

<220>
<221> modified_base
<222> (1)..(20)
<223> a, c, u, g, unknown or other and this region may encompass
17-20 nucleotides, wherein some positions may be absent

<220>
<221> modified_base
<222> (76)..(275)
<223> a, c, u, g, unknown or other and this region may encompass
0-200 nucleotides, wherein some positions may be absent

<400> 2
nnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcugaaa agcauagcaa guuaaaauaa 60

ggcuaguccg uuaucnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnn 275

<210> 3
<211> 287
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
guide polynucleotide

<220>
<221> modified_base
<222> (1)..(20)
<223> a, c, u, g, unknown or other and this region may encompass
17-20 nucleotides, wherein some positions may be absent

<220>
<221> modified_base
<222> (88)..(287)
<223> a, c, u, g, unknown or other and this region may encompass
0-200 nucleotides, wherein some positions may be absent

<400> 3
nnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcuguuu uggaaacaaa acagcauagc 60

aaguuaaaau aaggcuaguc cguuaucnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnn 287

<210> 4
<211> 296
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
guide polynucleotide

<220>
<221> modified_base
<222> (1)..(20)
<223> a, c, u, g, unknown or other and this region may encompass
17-20 nucleotides, wherein some positions may be absent

<220>
<221> modified_base
<222> (97)..(296)
<223> a, c, u, g, unknown or other and this region may encompass
0-200 nucleotides, wherein some positions may be absent

<400> 4
nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu ggcaccgagu cggugcnnnn nnnnnnnnnn nnnnnnnnnn 120

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnn 296

<210> 5
<211> 96
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
guide oligonucleotide

<220>
<221> modified_base
<222> (1)..(20)
<223> a, c, u, g, unknown or other and this region may encompass
17-20 nucleotides, wherein some positions may be absent

<400> 5
nnnnnnnnnn nnnnnnnnnn guuuaagagc uagaaauagc aaguuuaaau aaggcuaguc 60

cguuaucaac uugaaaaagu ggcaccgagu cggugc 96

<210> 6
<211> 106
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
guide polynucleotide

<220>
<221> modified_base
<222> (1)..(20)
<223> a, c, u, g, unknown or other and this region may encompass
17-20 nucleotides, wherein some positions may be absent

<400> 6
nnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcuggaa acagcauagc aaguuuaaau 60

aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugc 106

<210> 7
<211> 106
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
guide polynucleotide

<220>
<221> modified_base
<222> (1)..(20)
<223> a, c, u, g, unknown or other and this region may encompass
17-20 nucleotides, wherein some positions may be absent

<400> 7
nnnnnnnnnn nnnnnnnnnn guuuaagagc uaugcuggaa acagcauagc aaguuuaaau 60

aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugc 106

<210> 8
<211> 79
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
guide oligonucleotide

<400> 8
ggaaccauuc aaaacagcau agcaaguuaa aauaaggcua guccguuauc aacuugaaaa 60

aguggcaccg agucggugc 79

<210> 9
<211> 62
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
guide oligonucleotide

<400> 9
ggagcgagcg gagcgguaca guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cg 62

<210> 10
<211> 100
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
guide polynucleotide

<400> 10
ggagcgagcg gagcgguaca guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100

<210> 11
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
nuclear localization signal peptide

<400> 11
Pro Lys Lys Lys Arg Lys Val Ser
1 5

<210> 12
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
VP64 domain polypeptide

<400> 12
Asp Ala Leu Asp Asp Phe Asp Leu Asp Met Leu Gly Ser Asp Ala Leu
1 5 10 15

Asp Asp Phe Asp Leu Asp Met Leu Gly Ser Asp Ala Leu Asp Asp Phe
20 25 30

Asp Leu Asp Met Leu Gly Ser Asp Ala Leu Asp Asp Phe Asp Leu Asp
35 40 45

Met Leu

50

<210> 13
<211> 1368
<212> PRT
<213> Streptococcus pyogenes

<400> 13

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                  10                  15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

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

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175
```

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

```
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590
```

```
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800
```

```
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005
```

```
Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065

Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125

Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140

Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170

Phe Glu  Lys Asn Pro Ile Asp  Phe Leu Glu Ala Lys  Gly Tyr Lys
    1175                1180                1185

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200
```

```
Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                1240                1245

Pro Glu  Asp Asn Glu Gln Lys  Gln Leu Phe Val Glu  Gln His Lys
    1250                1255                1260

His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile Ser Glu  Phe Ser Lys
    1265                1270                1275

Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp Lys Val  Leu Ser Ala
    1280                1285                1290

Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg Glu Gln  Ala Glu Asn
    1295                1300                1305

Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu Gly Ala  Pro Ala Ala
    1310                1315                1320

Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg Lys Arg  Tyr Thr Ser
    1325                1330                1335

Thr Lys  Glu Val Leu Asp Ala  Thr Leu Ile His Gln  Ser Ile Thr
    1340                1345                1350

Gly Leu  Tyr Glu Thr Arg Ile  Asp Leu Ser Gln Leu  Gly Gly Asp
    1355                1360                1365


<210> 14
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
```

linker peptide

<400> 14
Gly Gly Gly Ser
1

<210> 15
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
linker peptide

<400> 15
Gly Gly Gly Gly Ser
1 5

<210> 16
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 16
gtgtgcagac ggcagtcact agg 23

<210> 17
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 17
gagcagcgtc ttcgagagtg agg 23

<210> 18
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 18
ggtgagtgag tgtgtgcgtg tgg 23

<210> 19
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 19
gttggagcgg ggagaaggcc agg 23

<210> 20
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 20
gggtgggggg agtttgctcc tgg 23

<210> 21
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 21
ggctttggaa agggggtggg ggg 23

<210> 22
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 22
ggggcggggt cccggcgggg cgg 23

<210> 23
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 23
gctcggaggt cgtggcgctg ggg 23

<210> 24
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 24
gactcaccgg ccagggcgct cgg 23

<210> 25
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 25
ggcgcagcgg ttaggtggac cgg 23

<210> 26
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 26
ggcgcatggc tccgccccgc cgg 23

<210> 27
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 27
gccacgacct ccgagctacc cgg 23

<210> 28
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 28
gcggcgtgag ccctcccccc tgg 23

<210> 29
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 29
ggaggcgggg tggaggggtg cgg 23

<210> 30
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 30
gggctcacgc cgcgctccgg cgg 23

<210> 31
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 31
gacccctcc accccgcctc cgg 23

<210> 32
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 32
gagcgcggag ccatctggcc ggg 23

<210> 33
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 33
gcgcggcgcg gaaggggtta agg 23

<210> 34
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 34
gcggcgcggc gcgggccggc ggg 23

<210> 35
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 35
gccgcgccgc cctcccccgc cgg 23

<210> 36
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 36
gcggttataa ccagccaacc cgg 23

<210> 37
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 37
gtgcgcggag ctgttcggaa ggg 23

<210> 38
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 38
acaccgtgtg cagacggcag tcactg 26

<210> 39
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 39
acaccgagca gcgtcttcga gagtgg 26

<210> 40
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 40
acaccggtga gtgagtgtgt gcgtgg 26

<210> 41
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 41
acaccgttgg agcggggaga aggccg 26

<210> 42
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 42
acaccgggtg gggggagttt gctccg 26

<210> 43
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 43
acaccggctt tggaaagggg gtgggg 26

<210> 44
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 44
acaccggggc ggggtcccgg cggggg 26

<210> 45
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 45
acaccgctcg gaggtcgtgg cgctgg 26

<210> 46
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic target binding site oligonucleotide

<400> 46
acaccgactc accggccagg gcgctg 26

<210> 47
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic target binding site oligonucleotide

<400> 47
acaccggcgc agcggttagg tggacg 26

<210> 48
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic target binding site oligonucleotide

<400> 48
acaccggcgc atggctccgc cccgcg 26

<210> 49
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic target binding site oligonucleotide

<400> 49
acaccgccac gacctccgag ctaccg 26

<210> 50
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 50
acaccgcggc gtgagccctc cccctg 26

<210> 51
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 51
acaccggagg cggggtggag ggggtg 26

<210> 52
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 52
acaccgggct cacgccgcgc tccggg 26

<210> 53
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 53
acaccgaccc cctccacccc gcctcg 26

<210> 54
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 54
acaccgagcg cggagccatc tggccg 26

<210> 55
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 55
acaccgcgcg gcgcggaagg ggttag 26

<210> 56
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 56
acaccgcggc gcggcgcggg ccggcg 26

<210> 57
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 57
acaccgccgc gccgccctcc cccgcg 26

<210> 58
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 58
acaccgcggt tataaccagc caaccg 26

<210> 59
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 59
acaccgtgcg cggagctgtt cggaag 26

<210> 60
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 60
aaaacagtga ctgccgtctg cacacg 26

<210> 61
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 61
aaaaccactc tcgaagacgc tgctcg 26

<210> 62
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 62
aaaaccacgc acacactcac tcaccg 26

<210> 63
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 63
aaaacggcct tctccccgct ccaacg 26

<210> 64
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 64
aaaacggagc aaactccccc cacccg 26

<210> 65
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 65
aaaaccccac ccccttttcca aagccg 26

<210> 66
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 66
aaaacccccg ccgggacccc gccccg 26

<210> 67
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 67
aaaaccagcg ccacgacctc cgagcg 26

<210> 68
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 68
aaaacagcgc cctggccggt gagtcg 26

<210> 69
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 69
aaaacgtcca cctaaccgct gcgccg 26

<210> 70
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 70
aaaacgcggg gcggagccat gcgccg 26

<210> 71
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 71
aaaacggtag ctcggaggtc gtggcg 26

<210> 72
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 72
aaaacagggg gagggctcac gccgcg 26

<210> 73
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 73
aaaacacccc ctccaccccg cctccg 26

<210> 74
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 74
aaaacccgga gcgcggcgtg agcccg 26

<210> 75
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 75
aaaacgaggc ggggtggagg gggtcg 26

<210> 76
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 76
aaaacggcca gatggctccg cgctcg 26

<210> 77
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 77
aaaactaacc ccttccgcgc cgcgcg 26

<210> 78
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 78
aaaacgccgg cccgcgccgc gccgcg 26

<210> 79
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 79
aaaacgcggg ggagggcggc gcggcg 26

<210> 80
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 80
aaaacggttg gctggttata accgcg 26

<210> 81
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
target binding site oligonucleotide

<400> 81
aaaacttccg aacagctccg cgcacg 26

<210> 82
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 82
tccagatggc acattgtcag 20

<210> 83
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 83
agggagcagg aaagtgaggt 20

<210> 84
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 84
gcacgtaacc tcactttcct 20

<210> 85
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 85
cttgctacct ctttcctctt tct 23

<210> 86
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 86
agagaagtcg aggaagagag ag 22

<210> 87
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 87
cagcagaaag ttcatggttt cg 22

<210> 88
<211> 130
<212> PRT
<213> Enterobacteria phage lambda

<400> 88
Met Ala Ser Asn Phe Thr Gln Phe Val Leu Val Asp Asn Gly Gly Thr
1 5 10 15

Gly Asp Val Thr Val Ala Pro Ser Asn Phe Ala Asn Gly Val Ala Glu
20 25 30

Trp Ile Ser Ser Asn Ser Arg Ser Gln Ala Tyr Lys Val Thr Cys Ser
35 40 45

Val Arg Gln Ser Ser Ala Gln Asn Arg Lys Tyr Thr Ile Lys Val Glu
50 55 60

Val Pro Lys Val Ala Thr Gln Thr Val Gly Gly Val Glu Leu Pro Val
65 70 75 80

Ala Ala Trp Arg Ser Tyr Leu Asn Met Glu Leu Thr Ile Pro Ile Phe
85 90 95

Ala Thr Asn Ser Asp Cys Glu Leu Ile Val Lys Ala Met Gln Gly Leu

100 105 110

Leu Lys Asp Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala Asn Ser Gly
115 120 125

Ile Tyr
130

<210> 89
<211> 130
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic lambda bacteriophage MS2 N55K mutant polypeptide

<400> 89
Met Ala Ser Asn Phe Thr Gln Phe Val Leu Val Asp Asn Gly Gly Thr
1 5 10 15

Gly Asp Val Thr Val Ala Pro Ser Asn Phe Ala Asn Gly Val Ala Glu
20 25 30

Trp Ile Ser Ser Asn Ser Arg Ser Gln Ala Tyr Lys Val Thr Cys Ser
35 40 45

Val Arg Gln Ser Ser Ala Gln Lys Arg Lys Tyr Thr Ile Lys Val Glu
50 55 60

Val Pro Lys Val Ala Thr Gln Thr Val Gly Gly Val Glu Leu Pro Val
65 70 75 80

Ala Ala Trp Arg Ser Tyr Leu Asn Met Glu Leu Thr Ile Pro Ile Phe
85 90 95

Ala Thr Asn Ser Asp Cys Glu Leu Ile Val Lys Ala Met Gln Gly Leu
100 105 110

Leu Lys Asp Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala Asn Ser Gly
115 120 125

Ile Tyr
130

<210> 90
<211> 117
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
lambda bacteriophage MS2 deltaFG mutant polypeptide

<400> 90
Met Ala Ser Asn Phe Thr Gln Phe Val Leu Val Asp Asn Gly Gly Thr
1 5 10 15

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

Trp Ile Ser Ser Asn Ser Arg Ser Gln Ala Tyr Lys Val Thr Cys Ser
35 40 45

Val Arg Gln Ser Ser Ala Gln Asn Arg Lys Tyr Thr Ile Lys Val Glu
50 55 60

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

Pro Ile Phe Ala Thr Asn Ser Asp Cys Glu Leu Ile Val Lys Ala Met
85 90 95

Gln Gly Leu Leu Lys Asp Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala
100 105 110

Asn Ser Gly Ile Tyr
115

<210> 91
<211> 262
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic dimeric MS2 coat polypeptide

<400> 91

```
Met Ala Ser Asn Phe Thr Gln Phe Val Leu Val Asp Asn Gly Gly Thr
1               5                  10                  15

Gly Asp Val Thr Val Ala Pro Ser Asn Phe Ala Asn Gly Val Ala Glu
            20                  25                  30

Trp Ile Ser Ser Asn Ser Arg Ser Gln Ala Tyr Lys Val Thr Cys Ser
        35                  40                  45

Val Arg Gln Ser Ser Ala Gln Asn Arg Lys Tyr Thr Ile Lys Val Glu
    50                  55                  60

Val Pro Lys Val Ala Thr Gln Thr Val Gly Gly Val Glu Leu Pro Val
65                  70                  75                  80

Ala Ala Trp Arg Ser Tyr Leu Asn Met Glu Leu Thr Ile Pro Ile Phe
                85                  90                  95

Ala Thr Asn Ser Asp Cys Glu Leu Ile Val Lys Ala Met Gln Gly Leu
            100                 105                 110

Leu Lys Asp Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala Asn Ser Gly
        115                 120                 125

Leu Tyr Gly Ala Met Ala Ser Asn Phe Thr Gln Phe Val Leu Val Asp
    130                 135                 140

Asn Gly Gly Thr Gly Asp Val Thr Val Ala Pro Ser Asn Phe Ala Asn
145                 150                 155                 160

Gly Val Ala Glu Trp Ile Ser Ser Asn Ser Arg Ser Gln Ala Tyr Lys
                165                 170                 175

Val Thr Cys Ser Val Arg Gln Ser Ser Ala Gln Asn Arg Lys Tyr Thr
            180                 185                 190
```

Ile Lys Val Glu Val Pro Lys Val Ala Thr Gln Thr Val Gly Gly Val
195 200 205

Glu Leu Pro Val Ala Ala Trp Arg Ser Tyr Leu Asn Met Glu Leu Thr
210 215 220

Ile Pro Ile Phe Ala Thr Asn Ser Asp Cys Glu Leu Ile Val Lys Ala
225 230 235 240

Met Gln Gly Leu Leu Lys Asp Gly Asn Pro Ile Pro Ser Ala Ile Ala
245 250 255

Ala Asn Ser Leu Ile Asn
260

<210> 92
<211> 262
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic dimeric MS2 N55K mutant coat polypeptide

<400> 92
Met Ala Ser Asn Phe Thr Gln Phe Val Leu Val Asp Asn Gly Gly Thr
1 5 10 15

Gly Asp Val Thr Val Ala Pro Ser Asn Phe Ala Asn Gly Val Ala Glu
20 25 30

Trp Ile Ser Ser Asn Ser Arg Ser Gln Ala Tyr Lys Val Thr Cys Ser
35 40 45

Val Arg Gln Ser Ser Ala Gln Lys Arg Lys Tyr Thr Ile Lys Val Glu
50 55 60

Val Pro Lys Val Ala Thr Gln Thr Val Gly Gly Val Glu Leu Pro Val
65 70 75 80

Ala Ala Trp Arg Ser Tyr Leu Asn Met Glu Leu Thr Ile Pro Ile Phe

85 90 95

Ala Thr Asn Ser Asp Cys Glu Leu Ile Val Lys Ala Met Gln Gly Leu
100 105 110

Leu Lys Asp Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala Asn Ser Gly
115 120 125

Leu Tyr Gly Ala Met Ala Ser Asn Phe Thr Gln Phe Val Leu Val Asp
130 135 140

Asn Gly Gly Thr Gly Asp Val Thr Val Ala Pro Ser Asn Phe Ala Asn
145 150 155 160

Gly Val Ala Glu Trp Ile Ser Ser Asn Ser Arg Ser Gln Ala Tyr Lys
165 170 175

Val Thr Cys Ser Val Arg Gln Ser Ser Ala Gln Lys Arg Lys Tyr Thr
180 185 190

Ile Lys Val Glu Val Pro Lys Val Ala Thr Gln Thr Val Gly Gly Val
195 200 205

Glu Leu Pro Val Ala Ala Trp Arg Ser Tyr Leu Asn Met Glu Leu Thr
210 215 220

Ile Pro Ile Phe Ala Thr Asn Ser Asp Cys Glu Leu Ile Val Lys Ala
225 230 235 240

Met Gln Gly Leu Leu Lys Asp Gly Asn Pro Ile Pro Ser Ala Ile Ala
245 250 255

Ala Asn Ser Leu Ile Asn
260

<210> 93
<211> 236
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic dimeric MS2 deltaFG mutant coat polypeptide

<400> 93

```
Met Ala Ser Asn Phe Thr Gln Phe Val Leu Val Asp Asn Gly Gly Thr
1               5                  10                  15

Gly Asp Val Thr Val Ala Pro Ser Asn Phe Ala Asn Gly Val Ala Glu
            20                  25                  30

Trp Ile Ser Ser Asn Ser Arg Ser Gln Ala Tyr Lys Val Thr Cys Ser
        35                  40                  45

Val Arg Gln Ser Ser Ala Gln Lys Arg Lys Tyr Thr Ile Lys Val Glu
    50                  55                  60

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

Pro Ile Phe Ala Thr Asn Ser Asp Cys Glu Leu Ile Val Lys Ala Met
                85                  90                  95

Gln Gly Leu Leu Lys Asp Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala
            100                 105                 110

Asn Ser Gly Leu Tyr Gly Ala Met Ala Ser Asn Phe Thr Gln Phe Val
        115                 120                 125

Leu Val Asp Asn Gly Gly Thr Gly Asp Val Thr Val Ala Pro Ser Asn
    130                 135                 140

Phe Ala Asn Gly Val Ala Glu Trp Ile Ser Ser Asn Ser Arg Ser Gln
145                 150                 155                 160

Ala Tyr Lys Val Thr Cys Ser Val Arg Gln Ser Ser Ala Gln Lys Arg
                165                 170                 175

Lys Tyr Thr Ile Lys Val Glu Val Pro Lys Gly Ala Trp Arg Ser Tyr
            180                 185                 190
```

Leu Asn Met Glu Leu Thr Ile Pro Ile Phe Ala Thr Asn Ser Asp Cys
195 200 205

Glu Leu Ile Val Lys Ala Met Gln Gly Leu Leu Lys Asp Gly Asn Pro
210 215 220

Ile Pro Ser Ala Ile Ala Ala Asn Ser Leu Ile Asn
225 230 235

<210> 94
<211> 22
<212> PRT
<213> Enterobacteria phage lambda

<400> 94
Met Asp Ala Gln Thr Arg Arg Arg Glu Arg Arg Ala Glu Lys Gln Ala
1 5 10 15

Gln Trp Lys Ala Ala Asn
20

<210> 95
<211> 107
<212> PRT
<213> Enterobacteria phage lambda

<400> 95
Met Asp Ala Gln Thr Arg Arg Arg Glu Arg Arg Ala Glu Lys Gln Ala
1 5 10 15

Gln Trp Lys Ala Ala Asn Pro Leu Leu Val Gly Val Ser Ala Lys Pro
20 25 30

Val Asn Arg Pro Ile Leu Ser Leu Asn Arg Lys Pro Lys Ser Arg Val
35 40 45

Glu Ser Ala Leu Asn Pro Ile Asp Leu Thr Val Leu Ala Glu Tyr His
50 55 60

Lys Gln Ile Glu Ser Asn Leu Gln Arg Ile Glu Arg Lys Asn Gln Arg
65 70 75 80

Thr Trp Tyr Ser Lys Pro Gly Glu Arg Gly Ile Thr Cys Ser Gly Arg
85 90 95

Gln Lys Ile Lys Gly Lys Ser Ile Pro Leu Ile
100 105

<210> 96
<211> 23
<212> RNA
<213> Enterobacteria phage lambda

<400> 96
aaacaugagg auuacccaug ucg 23

<210> 97
<211> 23
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic high affinity MS2 binding oligonucleotide

<400> 97
aaacaugagg aucacccaug ucg 23

<210> 98
<211> 15
<212> RNA
<213> Enterobacteria phage lambda

<400> 98
gcccugaaga agggc 15

<210> 99
<211> 15
<212> RNA
<213> Enterobacteria phage lambda

<400> 99
gcccugaaaa agggc 15

<210> 100
<211> 20
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
truncated Csy4 binding site oligonucleotide

<220>
<223> Description of combined DNA/RNA molecule: Synthetic
truncated Csy4 binding site oligonucleotide

<400> 100
gttcactgcc gtataggcag 20

<210> 101
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
Csy4 binding site oligonucleotide

<400> 101
guucacugcc guauaggcag cuaagaaa 28

<210> 102
<211> 32
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
crRNA oligonucleotide

<220>
<221> modified_base
<222> (1)..(20)
<223> a, c, u, g, unknown or other and this region may encompass
17-20 nucleotides, wherein some positions may be absent

<400> 102
nnnnnnnnnn nnnnnnnnnn guuuuagagc ua 32

<210> 103
<211> 42
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
crRNA oligonucleotide

<220>
<221> modified_base
<222> (1)..(20)
<223> a, c, u, g, unknown or other and this region may encompass
17-20 nucleotides, wherein some positions may be absent

<400> 103
nnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcuguuu ug 42

<210> 104
<211> 36
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
crRNA oligonucleotide

<220>
<221> modified_base
<222> (1)..(20)
<223> a, c, u, g, unknown or other and this region may encompass
17-20 nucleotides, wherein some positions may be absent

<400> 104
nnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcu 36

<210> 105
<211> 60
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
tracrRNA oligonucleotide

<400> 105
uagcaaguua aaauaaggcu aguccguuau caacuugaaa aaguggcacc gagucggugc 60

<210> 106
<211> 64
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
tracrRNA oligonucleotide

<400> 106
agcauagcaa guuaaaauaa ggcuaguccg uuaucaacuu gaaaaagugg caccgagucg 60

gugc 64

<210> 107
<211> 2279
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
RNA expression vector polynucleotide

<220>
<221> modified_base
<222> (331)..(350)
<223> a, c, t, g, unknown or other

<400> 107
gacgtcgcta gctgtacaaa aaagcaggct ttaaaggaac caattcagtc gactggatcc 60

ggtaccaagg tcgggcagga agagggccta tttcccatga ttccttcata tttgcatata 120

cgatacaagg ctgttagaga gataattaga attaatttga ctgtaaacac aaagatatta 180

gtacaaaata cgtgacgtag aaagtaataa tttcttgggt agtttgcagt tttaaaatta 240

tgttttaaaa tggactatca tatgcttacc gtaacttgaa agtatttcga tttcttggct 300

ttatatatct tgtggaaagg acgaaacacc nnnnnnnnnn nnnnnnnnnn gttttagagc 360

tagaaatagc aagttaaaat aaggctagtc cgttatcaac ttgaaaaagt ggcaccgagt 420

cggtgctttt tttaagcttg ggccgctcga ggtacctctc tacatatgac atgtgagcaa 480

aaggccagca aaaggccagg aaccgtaaaa aggccgcgtt gctggcgttt ttccataggc 540

tccgcccccc tgacgagcat cacaaaaatc gacgctcaag tcagaggtgg cgaaacccga 600

caggactata aagataccag gcgtttcccc ctggaagctc cctcgtgcgc tctcctgttc 660

cgaccctgcc gcttaccgga tacctgtccg cctttctccc ttcgggaagc gtggcgcttt 720

ctcatagctc acgctgtagg tatctcagtt cggtgtaggt cgttcgctcc aagctgggct 780

gtgtgcacga acccccccgtt cagcccgacc gctgcgcctt atccggtaac tatcgtcttg 840

agtccaaccc ggtaagacac gacttatcgc cactggcagc agccactggt aacaggatta 900

gcagagcgag gtatgtaggc ggtgctacag agttcttgaa gtggtggcct aactacggct 960

acactagaag aacagtattt ggtatctgcg ctctgctgaa gccagttacc ttcggaaaaa 1020

gagttggtag ctcttgatcc ggcaaacaaa ccaccgctgg tagcggtggt ttttttgttt 1080

gcaagcagca gattacgcgc agaaaaaaag gatctcaaga agatcctttg atcttttcta 1140

cggggtctga cgctcagtgg aacgaaaact cacgttaagg gattttggtc atgagattat 1200

caaaaaggat cttcacctag atccttttaa attaaaaatg aagttttaaa tcaatctaaa 1260

gtatatatga gtaaacttgg tctgacagtt accaatgctt aatcagtgag gcacctatct 1320

cagcgatctg tctatttcgt tcatccatag ttgcctgact ccccgtcgtg tagataacta 1380

cgatacggga gggcttacca tctggcccca gtgctgcaat gataccgcga gacccacgct 1440

caccggctcc agatttatca gcaataaacc agccagccgg aagggccgag cgcagaagtg 1500

gtcctgcaac tttatccgcc tccatccagt ctattaattg ttgccgggaa gctagagtaa 1560

gtagttcgcc agttaatagt ttgcgcaacg ttgttgccat tgctacaggc atcgtggtgt 1620

cacgctcgtc gtttggtatg gcttcattca gctccggttc ccaacgatca aggcgagtta 1680

catgatcccc catgttgtgc aaaaaagcgg ttagctcctt cggtcctccg atcgttgtca 1740

gaagtaagtt ggccgcagtg ttatcactca tggttatggc agcactgcat aattctctta 1800

ctgtcatgcc atccgtaaga tgcttttctg tgactggtga gtactcaacc aagtcattct 1860

gagaatagtg tatgcggcga ccgagttgct cttgcccggc gtcaatacgg gataataccg 1920

cgccacatag cagaacttta aaagtgctca tcattggaaa acgttcttcg gggcgaaaac 1980

tctcaaggat cttaccgctg ttgagatcca gttcgatgta acccactcgt gcacccaact 2040

gatcttcagc atcttttact ttcaccagcg tttctgggtg agcaaaaaca ggaaggcaaa 2100

atgccgcaaa aaagggaata agggcgacac ggaaatgttg aatactcata ctcttccttt 2160

ttcaatatta ttgaagcatt tatcagggtt attgtctcat gagcggatac atatttgaat 2220

gtatttagaa aaataaacaa atagggggttc cgcgcacatt tccccgaaaa gtgccacct 2279

<210> 108
<211> 7786
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
CMV-T7-Cas9 D10A/H840A-3xFlag-VP64 polynucleotide

<400> 108
atatgccaag tacgcccccct attgacgtca atgacggtaa atggcccgcc tggcattatg 60

cccagtacat gaccttatgg gactttccta cttggcagta catctacgta ttagtcatcg 120

ctattaccat ggtgatgcgg ttttggcagt acatcaatgg gcgtggatag cggtttgact 180

cacggggatt tccaagtctc caccccattg acgtcaatgg gagtttgttt tggcaccaaa 240

atcaacggga ctttccaaaa tgtcgtaaca actccgcccc attgacgcaa atgggcggta 300

ggcgtgtacg gtgggaggtc tatataagca gagctggttt agtgaaccgt cagatccgct 360

agagatccgc ggccgctaat acgactcact atagggagag ccgccaccat ggataagaaa 420

tactcaatag gcttagctat cggcacaaat agcgtcggat gggcggtgat cactgatgaa 480

tataaggttc cgtctaaaaa gttcaaggtt ctgggaaata cagaccgcca cagtatcaaa 540

aaaaatctta taggggctct tttatttgac agtggagaga cagcggaagc gactcgtctc 600

aaacggacag ctcgtagaag gtatacacgt cggaagaatc gtatttgtta tctacaggag 660

atttttttcaa atgagatggc gaaagtagat gatagtttct ttcatcgact tgaagagtct 720

tttttggtgg aagaagacaa gaagcatgaa cgtcatccta tttttggaaa tatagtagat 780

gaagttgctt atcatgagaa atatccaact atctatcatc tgcgaaaaaa attggtagat 840

tctactgata aagcggattt gcgcttaatc tatttggcct tagcgcatat gattaagttt 900

cgtggtcatt ttttgattga gggagattta aatcctgata atagtgatgt ggacaaacta 960

tttatccagt tggtacaaac ctacaatcaa ttatttgaag aaaaccctat taacgcaagt 1020

ggagtagatg ctaaagcgat tctttctgca cgattgagta aatcaagacg attagaaaat 1080

ctcattgctc agctccccgg tgagaagaaa aatggcttat ttgggaatct cattgctttg 1140

tcattgggtt tgacccctaa ttttaaatca aattttgatt tggcagaaga tgctaaatta 1200

cagctttcaa aagatactta cgatgatgat ttagataatt tattggcgca aattggagat 1260

caatatgctg atttgttttt ggcagctaag aatttatcag atgctatttt actttcagat 1320

atcctaagag taaatactga aataactaag gctcccctat cagcttcaat gattaaacgc 1380

tacgatgaac atcatcaaga cttgactctt ttaaaagctt tagttcgaca acaacttcca 1440

gaaaagtata aagaaatctt ttttgatcaa tcaaaaaacg gatatgcagg ttatattgat 1500

gggggagcta gccaagaaga attttataaa tttatcaaac caattttaga aaaaatggat 1560

ggtactgagg aattattggt gaaactaaat cgtgaagatt tgctgcgcaa gcaacggacc 1620

tttgacaacg gctctattcc ccatcaaatt cacttgggtg agctgcatgc tattttgaga 1680

agacaagaag acttttatcc atttttaaaa gacaatcgtg agaagattga aaaaatcttg 1740

acttttcgaa ttccttatta tgttggtcca ttggcgcgtg gcaatagtcg ttttgcatgg 1800

atgactcgga agtctgaaga aacaattacc ccatggaatt ttgaagaagt tgtcgataaa 1860

ggtgcttcag ctcaatcatt tattgaacgc atgacaaact ttgataaaaa tcttccaaat 1920

gaaaaagtac taccaaaaca tagtttgctt tatgagtatt ttacggttta taacgaattg 1980

acaaaggtca aatatgttac tgaaggaatg cgaaaaccag catttctttc aggtgaacag 2040

aagaaagcca ttgttgattt actcttcaaa acaaatcgaa aagtaaccgt taagcaatta 2100

aaagaagatt atttcaaaaa aatagaatgt tttgatagtg ttgaaatttc aggagttgaa 2160

gatagattta atgcttcatt aggtacctac catgatttgc taaaaattat taaagataaa 2220

gatttttgg ataatgaaga aaatgaagat atcttagagg atattgtttt aacattgacc 2280

ttatttgaag atagggagat gattgaggaa agacttaaaa catatgctca cctctttgat 2340

gataaggtga tgaaacagct taaacgtcgc cgttatactg gttggggacg tttgtctcga 2400

aaattgatta atggtattag ggataagcaa tctggcaaaa caatattaga ttttttgaaa 2460

tcagatggtt ttgccaatcg caattttatg cagctgatcc atgatgatag tttgacattt 2520

aaagaagaca ttcaaaaagc acaagtgtct ggacaaggcg atagtttaca tgaacatatt 2580

gcaaatttag ctggtagccc tgctattaaa aaaggtattt tacagactgt aaaagttgtt 2640

gatgaattgg tcaaagtaat ggggcggcat aagccagaaa atatcgttat tgaaatggca 2700

cgtgaaaatc agacaactca aaagggccag aaaaattcgc gagagcgtat gaaacgaatc 2760

gaagaaggta tcaaagaatt aggaagtcag attcttaaag agcatcctgt tgaaaatact 2820

caattgcaaa atgaaaagct ctatctctat tatctccaaa atggaagaga catgtatgtg 2880

gaccaagaat tagatattaa tcgtttaagt gattatgatg tcgatgccat tgttccacaa 2940

agtttcctta aagacgattc aatagacaat aaggtcttaa cgcgttctga taaaaatcgt 3000

ggtaaatcgg ataacgttcc aagtgaagaa gtagtcaaaa agatgaaaaa ctattggaga 3060

caacttctaa acgccaagtt aatcactcaa cgtaagtttg ataatttaac gaaagctgaa 3120

cgtggaggtt tgagtgaact tgataaagct ggttttatca aacgccaatt ggttgaaact 3180

cgccaaatca ctaagcatgt ggcacaaatt ttggatagtc gcatgaatac taaatacgat 3240

gaaaatgata aacttattcg agaggttaaa gtgattacct taaaatctaa attagtttct 3300

gacttccgaa aagatttcca attctataaa gtacgtgaga ttaacaatta ccatcatgcc 3360

catgatgcgt atctaaatgc cgtcgttgga actgctttga ttaagaaata tccaaaactt 3420

gaatcggagt ttgtctatgg tgattataaa gtttatgatg ttcgtaaaat gattgctaag 3480

tctgagcaag aaataggcaa agcaaccgca aaatatttct tttactctaa tatcatgaac 3540

ttcttcaaaa cagaaattac acttgcaaat ggagagattc gcaaacgccc tctaatcgaa 3600

actaatgggg aaactggaga aattgtctgg gataaagggc gagattttgc cacagtgcgc 3660

aaagtattgt ccatgcccca agtcaatatt gtcaagaaaa cagaagtaca gacaggcgga 3720

ttctccaagg agtcaatttt accaaaaaga aattcggaca agcttattgc tcgtaaaaaa 3780

gactgggatc caaaaaaata tggtggtttt gatagtccaa cggtagctta ttcagtccta 3840

gtggttgcta aggtggaaaa agggaaatcg aagaagttaa aatccgttaa agagttacta 3900

gggatcacaa ttatggaaag aagttccttt gaaaaaaatc cgattgactt tttagaagct 3960

aaaggatata aggaagttaa aaaagactta atcattaaac tacctaaata tagtcttttt 4020

gagttagaaa acggtcgtaa acggatgctg gctagtgccg gagaattaca aaaaggaaat 4080

gagctggctc tgccaagcaa atatgtgaat tttttatatt tagctagtca ttatgaaaag 4140

ttgaagggta gtccagaaga taacgaacaa aaacaattgt ttgtggagca gcataagcat 4200

tatttagatg agattattga gcaaatcagt gaattttcta agcgtgttat tttagcagat 4260

gccaatttag ataaagttct tagtgcatat aacaaacata gagacaaacc aatacgtgaa 4320

caagcagaaa atattattca tttatttacg ttgacgaatc ttggagctcc cgctgctttt 4380

aaatattttg atacaacaat tgatcgtaaa cgatatacgt ctacaaaaga agttttagat 4440

gccactctta tccatcaatc catcactggt ctttatgaaa cacgcattga tttgagtcag 4500

ctaggaggtg acggttctcc caagaagaag aggaaagtct cgagcgacta caaagaccat 4560

gacggtgatt ataaagatca tgacatcgat tacaaggatg acgatgacaa ggctgcagga 4620

ggcggtggaa gcgggcgcgc cgacgcgctg gacgatttcg atctcgacat gctgggttct 4680

gatgccctcg atgactttga cctggatatg ttgggaagcg acgcattgga tgactttgat 4740

ctggacatgc tcggctccga tgctctggac gatttcgatc tcgatatgtt ataaccggtc 4800

atcatcacca tcaccattga gtttaaaccc gctgatcagc ctcgactgtg ccttctagtt 4860

gccagccatc tgttgtttgc ccctcccccg tgccttcctt gaccctggaa ggtgccactc 4920

ccactgtcct ttcctaataa aatgaggaaa ttgcatcgca ttgtctgagt aggtgtcatt 4980

ctattctggg gggtggggtg gggcaggaca gcaaggggga ggattgggaa gacaatagca 5040

ggcatgctgg ggatgcggtg ggctctatgg cttctgaggc ggaaagaacc agctggggct 5100

cgataccgtc gacctctagc tagagcttgg cgtaatcatg gtcatagctg tttcctgtgt 5160

gaaattgtta tccgctcaca attccacaca acatacgagc cggaagcata aagtgtaaag 5220

cctagggtgc ctaatgagtg agctaactca cattaattgc gttgcgctca ctgcccgctt 5280

tccagtcggg aaacctgtcg tgccagctgc attaatgaat cggccaacgc gcggggagag 5340

gcggtttgcg tattgggcgc tcttccgctt cctcgctcac tgactcgctg cgctcggtcg 5400

ttcggctgcg gcgagcggta tcagctcact caaaggcggt aatacggtta tccacagaat 5460

caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta 5520

aaaaggccgc gttgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa 5580

atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc 5640

cccctggaag ctccctcgtg cgctctcctg ttccgaccct gccgcttacc ggatacctgt 5700

ccgcctttct cccttcggga agcgtggcgc tttctcaatg ctcacgctgt aggtatctca 5760

gttcggtgta ggtcgttcgc tccaagctgg gctgtgtgca cgaacccccc gttcagcccg 5820

accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggtaaga cacgacttat 5880

cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta 5940

cagagttctt gaagtggtgg cctaactacg gctacactag aaggacagta tttggtatct 6000

gcgctctgct gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac 6060

aaaccaccgc tggtagcggt ggtttttttg tttgcaagca gcagattacg cgcagaaaaa 6120

aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaacgaaa 6180

actcacgtta agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt 6240

taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact tggtctgaca 6300

gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcatcca 6360

tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggctta ccatctggcc 6420

ccagtgctgc aatgataccg cgagacccac gctcaccggc tccagattta tcagcaataa 6480

accagccagc cggaagggcc gagcgcagaa gtggtcctgc aactttatcc gcctccatcc 6540

agtctattaa ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca 6600

acgttgttgc cattgctaca ggcatcgtgg tgtcacgctc gtcgtttggt atggcttcat 6660

tcagctccgg ttcccaacga tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag 6720

cggttagctc cttcggtcct ccgatcgttg tcagaagtaa gttggccgca gtgttatcac 6780

tcatggttat ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt 6840

ctgtgactgg tgagtactca accaagtcat tctgagaata gtgtatgcgg cgaccgagtt 6900

gctcttgccc ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc 6960

tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat 7020

ccagttcgat gtaacccact cgtgcaccca actgatcttc agcatctttt actttcacca 7080

gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataagggcga 7140

cacggaaatg ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg 7200

gttattgtct catgagcgga tacatatttg aatgtattta gaaaaataaa caaatagggg 7260

ttccgcgcac atttccccga aaagtgccac ctgacgtcga cggatcggga gatcgatctc 7320

ccgatcccct agggtcgact ctcagtacaa tctgctctga tgccgcatag ttaagccagt 7380

atctgctccc tgcttgtgtg ttggaggtcg ctgagtagtg cgcgagcaaa atttaagcta 7440

caacaaggca aggcttgacc gacaattgca tgaagaatct gcttagggtt aggcgttttg 7500

cgctgcttcg cgatgtacgg gccagatata cgcgttgaca ttgattattg actagttatt 7560

aatagtaatc aattacgggg tcattagttc atagcccata tatggagttc cgcgttacat 7620

aacttacggt aaatggcccg cctggctgac cgcccaacga cccccgccca ttgacgtcaa 7680

taatgacgta tgttcccata gtaacgccaa tagggacttt ccattgacgt caatgggtgg 7740

actatttacg gtaaactgcc cacttggcag tacatcaagt gtatcc 7786

<210> 109
<211> 7785
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
MV-T7-Cas9 recoded D10A/H840A-3xFLAG-VP64 polynucleotide

<400> 109
atatgccaag tacgcccccc attgacgtca atgacggtaa atggcccgcc tggcattatg 60

cccagtacat gaccttatgg gactttccta cttggcagta catctacgta ttagtcatcg 120

ctattaccat ggtgatgcgg ttttggcagt acatcaatgg gcgtggatag cggtttgact 180

cacggggatt tccaagtctc caccccattg acgtcaatgg gagtttgttt tggcaccaaa 240

atcaacggga ctttccaaaa tgtcgtaaca actccgcccc attgacgcaa atgggcggta 300

ggcgtgtacg gtgggaggtc tatataagca gagctggttt agtgaaccgt cagatccgct 360

agagatccgc ggccgctaat acgactcact atagggagag ccgccaccat ggataaaaag 420

tattctattg gtttagccat cggcactaat tccgttggat gggctgtcat aaccgatgaa 480

tacaaagtac cttcaaagaa atttaaggtg ttggggaaca cagaccgtca ttcgattaaa 540

aagaatctta tcggtgccct cctattcgat agtggcgaaa cggcagaggc gactcgcctg 600

aaacgaaccg ctcggagaag gtatacacgt cgcaagaacc gaatatgtta cttacaagaa 660

atttttagca atgagatggc caaagttgac gattctttct ttcaccgttt ggaagagtcc 720

ttccttgtcg aagaggacaa gaaacatgaa cggcacccca tctttggaaa catagtagat 780

gaggtggcat atcatgaaaa gtacccaacg atttatcacc tcagaaaaaa gctagttgac 840

tcaactgata aagcggacct gaggttaatc tacttggctc ttgcccatat gataaagttc 900

cgtgggcact ttctcattga gggtgatcta aatccggaca actcggatgt cgacaaactg 960

ttcatccagt tagtacaaac ctataatcag ttgtttgaag agaaccctat aaatgcaagt 1020

ggcgtggatg cgaaggctat tcttagcgcc cgcctctcta aatcccgacg gctagaaaac 1080

ctgatcgcac aattacccgg agagaagaaa aatgggttgt tcggtaacct tatagcgctc 1140

tcactaggcc tgacaccaaa ttttaagtcg aacttcgact tagctgaaga tgccaaattg 1200

cagcttagta aggacacgta cgatgacgat ctcgacaatc tactggcaca aattggagat 1260

cagtatgcgg acttattttt ggctgccaaa aaccttagcg atgcaatcct cctatctgac 1320

atactgagag ttaatactga gattaccaag gcgccgttat ccgcttcaat gatcaaaagg 1380

tacgatgaac atcaccaaga cttgacactt ctcaaggccc tagtccgtca gcaactgcct 1440

gagaaatata aggaaatatt ctttgatcag tcgaaaaacg ggtacgcagg ttatattgac 1500

ggcggagcga gtcaagagga attctacaag tttatcaaac ccatattaga gaagatggat 1560

gggacggaag agttgcttgt aaaactcaat cgcgaagatc tactgcgaaa gcagcggact 1620

ttcgacaacg gtagcattcc acatcaaatc cacttaggcg aattgcatgc tatacttaga 1680

aggcaggagg atttttatcc gttcctcaaa gacaatcgtg aaaagattga gaaaatccta 1740

acctttcgca taccttacta tgtgggaccc ctggcccgag ggaactctcg gttcgcatgg 1800

atgacaagaa agtccgaaga aacgattact ccatggaatt ttgaggaagt tgtcgataaa 1860

ggtgcgtcag ctcaatcgtt catcgagagg atgaccaact ttgacaagaa tttaccgaac 1920

gaaaaagtat tgcctaagca cagtttactt tacgagtatt tcacagtgta caatgaactc 1980

acgaaagtta agtatgtcac tgagggcatg cgtaaacccg cctttctaag cggagaacag 2040

aagaaagcaa tagtagatct gttattcaag accaaccgca aagtgacagt taagcaattg 2100

aaagaggact actttaagaa aattgaatgc ttcgattctg tcgagatctc cggggtagaa 2160

gatcgattta atgcgtcact tggtacgtat catgacctcc taaagataat taaagataag 2220

gacttcctgg ataacgaaga gaatgaagat atcttagaag atatagtgtt gactcttacc 2280

ctctttgaag atcgggaaat gattgaggaa agactaaaaa catacgctca cctgttcgac 2340

gataaggtta tgaaacagtt aaagaggcgt cgctatacgg gctggggacg attgtcgcgg 2400

aaacttatca acgggataag agacaagcaa agtggtaaaa ctattctcga ttttctaaag 2460

agcgacggct tcgccaatag gaactttatg cagctgatcc atgatgactc tttaaccttc 2520

aaagaggata tacaaaaggc acaggtttcc ggacaagggg actcattgca cgaacatatt 2580

gcgaatcttg ctggttcgcc agccatcaaa aagggcatac tccagacagt caaagtagtg 2640

gatgagctag ttaaggtcat gggacgtcac aaaccggaaa acattgtaat cgagatggca 2700

cgcgaaaatc aaacgactca gaaggggcaa aaaaacagtc gagagcggat gaagagaata 2760

gaagagggta ttaaagaact gggcagccag atcttaaagg agcatcctgt ggaaaatacc 2820

caattgcaga acgagaaact ttacctctat tacctacaaa atggaaggga catgtatgtt 2880

gatcaggaac tggacataaa ccgtttatct gattacgacg tcgatgccat tgtaccccaa 2940

tccttttga aggacgattc aatcgacaat aaagtgctta cacgctcgga taagaaccga 3000

gggaaaagtg acaatgttcc aagcgaggaa gtcgtaaaga aaatgaagaa ctattggcgg 3060

cagctcctaa atgcgaaact gataacgcaa agaaagttcg ataacttaac taaagctgag 3120

aggggtggct tgtctgaact tgacaaggcc ggatttatta aacgtcagct cgtggaaacc 3180

cgccaaatca caaagcatgt tgcacagata ctagattccc gaatgaatac gaaatacgac 3240

gagaacgata agctgattcg ggaagtcaaa gtaatcactt taaagtcaaa attggtgtcg 3300

gacttcagaa aggattttca attctataaa gttagggaga taaataacta ccaccatgcg 3360

cacgacgctt atcttaatgc cgtcgtaggg accgcactca ttaagaaata cccgaagcta 3420

gaaagtgagt ttgtgtatgg tgattacaaa gtttatgacg tccgtaagat gatcgcgaaa 3480

agcgaacagg agataggcaa ggctacagcc aaatacttct tttattctaa cattatgaat 3540

ttctttaaga cggaaatcac tctggcaaac ggagagatac gcaaacgacc tttaattgaa 3600

accaatgggg agacaggtga aatcgtatgg gataagggcc gggacttcgc gacggtgaga 3660

aaagttttgt ccatgcccca agtcaacata gtaaagaaaa ctgaggtgca gaccggaggg 3720

ttttcaaagg aatcgattct tccaaaaagg aatagtgata agctcatcgc tcgtaaaaag 3780

gactgggacc cgaaaaagta cggtggcttc gatagcccta cagttgccta ttctgtccta 3840

gtagtggcaa aagttgagaa gggaaaatcc aagaaactga agtcagtcaa agaattattg 3900

gggataacga ttatggagcg ctcgtctttt gaaaagaacc ccatcgactt ccttgaggcg 3960

aaaggttaca aggaagtaaa aaaggatctc ataattaaac taccaaagta tagtctgttt 4020

gagttagaaa atggccgaaa acggatgttg gctagcgccg gagagcttca aaaggggaac 4080

gaactcgcac taccgtctaa atacgtgaat ttcctgtatt tagcgtccca ttacgagaag 4140

ttgaaaggtt cacctgaaga taacgaacag aagcaacttt ttgttgagca gcacaaacat 4200

tatctcgacg aaatcataga gcaaatttcg gaattcagta agagagtcat cctagctgat 4260

gccaatctgg acaaagtatt aagcgcatac aacaagcaca gggataaacc catacgtgag 4320

caggcggaaa atattatcca tttgtttact cttaccaacc tcggcgctcc agccgcattc 4380

aagtattttg acacaacgat agatcgcaaa cgatacactt ctaccaagga ggtgctagac 4440

gcgacactga ttcaccaatc catcacggga ttatatgaaa ctcggataga tttgtcacag 4500

cttgggggtg acggatcccc caagaagaag aggaaagtct cgagcgacta caaagaccat 4560

gacggtgatt ataaagatca tgacatcgat tacaaggatg acgatgacaa ggctgcagga 4620

ggcggtggaa gcgggcgcgc cgacgcgctg gacgatttcg atctcgacat gctgggttct 4680

gatgccctcg atgactttga cctggatatg ttgggaagcg acgcattgga tgactttgat 4740

ctggacatgc tcggctccga tgctctggac gatttcgatc tcgatatgtt ataaccggtc 4800

atcatcacca tcaccattga gtttaaaccc gctgatcagc ctcgactgtg ccttctagtt 4860

gccagccatc tgttgtttgc ccctcccccg tgccttcctt gaccctggaa ggtgccactc 4920

ccactgtcct ttcctaataa aatgaggaaa ttgcatcgca ttgtctgagt aggtgtcatt 4980

ctattctggg gggtggggtg gggcaggaca gcaaggggga ggattgggaa gacaatagca 5040

ggcatgctgg ggatgcggtg ggctctatgg cttctgaggc ggaaagaacc agctggggct 5100

cgataccgtc gacctctagc tagagcttgg cgtaatcatg gtcatagctg tttcctgtgt 5160

gaaattgtta tccgctcaca attccacaca acatacgagc cggaagcata aagtgtaaag 5220

cctagggtgc ctaatgagtg agctaactca cattaattgc gttgcgctca ctgcccgctt 5280

tccagtcggg aaacctgtcg tgccagctgc attaatgaat cggccaacgc gcggggagag 5340

gcggtttgcg tattgggcgc tcttccgctt cctcgctcac tgactcgctg cgctcggtcg 5400

ttcggctgcg gcgagcggta tcagctcact caaaggcggt aatacggtta tccacagaat 5460

caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta 5520

aaaaggccgc gttgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa 5580

atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc 5640

ccctggaag ctccctcgtg cgctctcctg ttccgaccct gccgcttacc ggatacctgt 5700

ccgcctttct cccttcggga agcgtggcgc tttctcaatg ctcacgctgt aggtatctca 5760

gttcggtgta ggtcgttcgc tccaagctgg gctgtgtgca cgaacccccc gttcagcccg 5820

accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggtaaga cacgacttat 5880

cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta 5940

cagagttctt gaagtggtgg cctaactacg gctacactag aaggacagta tttggtatct 6000

gcgctctgct gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac 6060

aaaccaccgc tggtagcggt ggtttttttg tttgcaagca gcagattacg cgcagaaaaa 6120

aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaacgaaa 6180

actcacgtta agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt 6240

taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact tggtctgaca 6300

gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcatcca 6360

tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggctta ccatctggcc 6420

ccagtgctgc aatgataccg cgagacccac gctcaccggc tccagattta tcagcaataa 6480

accagccagc cggaagggcc gagcgcagaa gtggtcctgc aactttatcc gcctccatcc 6540

agtctattaa ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca 6600

acgttgttgc cattgctaca ggcatcgtgg tgtcacgctc gtcgtttggt atggcttcat 6660

tcagctccgg ttcccaacga tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag 6720

cggttagctc cttcggtcct ccgatcgttg tcagaagtaa gttggccgca gtgttatcac 6780

tcatggttat ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt 6840

ctgtgactgg tgagtactca accaagtcat tctgagaata gtgtatgcgg cgaccgagtt 6900

gctcttgccc ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc 6960

tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat 7020

ccagttcgat gtaacccact cgtgcaccca actgatcttc agcatctttt actttcacca 7080

gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataagggcga 7140

cacggaaatg ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg 7200

gttattgtct catgagcgga tacatatttg aatgtattta gaaaaataaa caaatagggg 7260

ttccgcgcac atttccccga aaagtgccac ctgacgtcga cggatcggga gatcgatctc 7320

ccgatcccct agggtcgact ctcagtacaa tctgctctga tgccgcatag ttaagccagt 7380

atctgctccc tgcttgtgtg ttggaggtcg ctgagtagtg cgcgagcaaa atttaagcta 7440

caacaaggca aggcttgacc gacaattgca tgaagaatct gcttagggtt aggcgttttg 7500

cgctgcttcg cgatgtacgg gccagatata cgcgttgaca ttgattattg actagttatt 7560

aatagtaatc aattacgggg tcattagttc atagcccata tatggagttc cgcgttacat 7620

aacttacggt aaatggcccg cctggctgac cgcccaacga cccccgccca ttgacgtcaa 7680

taatgacgta tgttcccata gtaacgccaa tagggacttt ccattgacgt caatgggtgg 7740

actatttacg gtaaactgcc cacttggcag tacatcaagt gtatc 7785

<210> 110
<211> 1461
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
Cas9-activator polypeptide

<400> 110
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

```
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285
```

```
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495
```

```
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700
```

```
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910
```

```
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065

Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110
```

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

Gly Ser Pro Lys Lys Lys Arg Lys Val Ser Ser Asp Tyr Lys Asp
1370 1375 1380

His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp Tyr Lys Asp Asp
1385 1390 1395

Asp Asp Lys Ala Ala Gly Gly Gly Gly Ser Gly Arg Ala Asp Ala
1400 1405 1410

Leu Asp Asp Phe Asp Leu Asp Met Leu Gly Ser Asp Ala Leu Asp
1415 1420 1425

Asp Phe Asp Leu Asp Met Leu Gly Ser Asp Ala Leu Asp Asp Phe
1430 1435 1440

Asp Leu Asp Met Leu Gly Ser Asp Ala Leu Asp Asp Phe Asp Leu
1445 1450 1455

Asp Met Leu
1460

<210> 111
<211> 1527
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
dCas9-NLS-3xFLAG-HP1alpha polypeptide

<400> 111

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                  10                  15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

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

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205
```

```
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415
```

```
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620
```

```
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830
```

```
Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035
```

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230

```
Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                 1240                 1245

Pro Glu  Asp Asn Glu Gln Lys  Gln Leu Phe Val Glu  Gln His Lys
    1250                 1255                 1260

His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile Ser Glu  Phe Ser Lys
    1265                 1270                 1275

Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp Lys Val  Leu Ser Ala
    1280                 1285                 1290

Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg Glu Gln  Ala Glu Asn
    1295                 1300                 1305

Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu Gly Ala  Pro Ala Ala
    1310                 1315                 1320

Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg Lys Arg  Tyr Thr Ser
    1325                 1330                 1335

Thr Lys  Glu Val Leu Asp Ala  Thr Leu Ile His Gln  Ser Ile Thr
    1340                 1345                 1350

Gly Leu  Tyr Glu Thr Arg Ile  Asp Leu Ser Gln Leu  Gly Gly Asp
    1355                 1360                 1365

Gly Ser  Pro Lys Lys Lys Arg  Lys Val Ser Ser Asp  Tyr Lys Asp
    1370                 1375                 1380

His Asp  Gly Asp Tyr Lys Asp  His Asp Ile Asp Tyr  Lys Asp Asp
    1385                 1390                 1395

Asp Asp  Lys Ala Ala Gly Gly  Gly Gly Ser Met Lys  Glu Gly Glu
    1400                 1405                 1410

Asn Asn  Lys Pro Arg Glu Lys  Ser Glu Ser Asn Lys  Arg Lys Ser
    1415                 1420                 1425
```

Asn Phe Ser Asn Ser Ala Asp Asp Ile Lys Ser Lys Lys Lys Arg
1430 1435 1440

Glu Gln Ser Asn Asp Ile Ala Arg Gly Phe Glu Arg Gly Leu Glu
1445 1450 1455

Pro Glu Lys Ile Ile Gly Ala Thr Asp Ser Cys Gly Asp Leu Met
1460 1465 1470

Phe Leu Met Lys Trp Lys Asp Thr Asp Glu Ala Asp Leu Val Leu
1475 1480 1485

Ala Lys Glu Ala Asn Val Lys Cys Pro Gln Ile Val Ile Ala Phe
1490 1495 1500

Tyr Glu Glu Arg Leu Thr Trp His Ala Tyr Pro Glu Asp Ala Glu
1505 1510 1515

Asn Lys Glu Lys Glu Thr Ala Lys Ser
1520 1525

<210> 112
<211> 1521
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
dCas9-NLS-3xFLAG-HP1beta polypeptide

<400> 112
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

```
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

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

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255
```

```
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460
```

```
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670
```

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

```
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065

Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080
```

```
Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125

Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140

Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170

Phe Glu  Lys Asn Pro Ile Asp  Phe Leu Glu Ala Lys  Gly Tyr Lys
    1175                1180                1185

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                1240                1245

Pro Glu  Asp Asn Glu Gln Lys  Gln Leu Phe Val Glu  Gln His Lys
    1250                1255                1260

His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile Ser Glu  Phe Ser Lys
    1265                1270                1275
```

```
Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp Lys Val  Leu Ser Ala
    1280                1285                1290

Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg Glu Gln  Ala Glu Asn
    1295                1300                1305

Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu Gly Ala  Pro Ala Ala
    1310                1315                1320

Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg Lys Arg  Tyr Thr Ser
    1325                1330                1335

Thr Lys  Glu Val Leu Asp Ala  Thr Leu Ile His Gln  Ser Ile Thr
    1340                1345                1350

Gly Leu  Tyr Glu Thr Arg Ile  Asp Leu Ser Gln Leu  Gly Gly Asp
    1355                1360                1365

Gly Ser  Pro Lys Lys Lys Arg  Lys Val Ser Ser Asp  Tyr Lys Asp
    1370                1375                1380

His Asp  Gly Asp Tyr Lys Asp  His Asp Ile Asp Tyr  Lys Asp Asp
    1385                1390                1395

Asp Asp  Lys Ala Ala Gly Gly  Gly Gly Ser Thr Ala  His Glu Thr
    1400                1405                1410

Asp Lys  Ser Glu Gly Gly Lys  Arg Lys Ala Asp Ser  Asp Ser Glu
    1415                1420                1425

Asp Lys  Gly Glu Glu Ser Lys  Pro Lys Lys Lys Lys  Glu Glu Ser
    1430                1435                1440

Glu Lys  Pro Arg Gly Phe Ala  Arg Gly Leu Glu Pro  Glu Arg Ile
    1445                1450                1455

Ile Gly  Ala Thr Asp Ser Ser  Gly Glu Leu Met Phe  Leu Met Lys
    1460                1465                1470
```

Trp Lys Asn Ser Asp Glu Ala Asp Leu Val Pro Ala Lys Glu Ala
1475 1480 1485

Asn Val Lys Cys Pro Gln Val Val Ile Ser Phe Tyr Glu Glu Arg
1490 1495 1500

Leu Thr Trp His Ser Tyr Pro Ser Glu Asp Asp Asp Lys Lys Asp
1505 1510 1515

Asp Lys Asn
1520

<210> 113
<211> 2126
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic dCas9-3xFLAG-TET1CD polypeptide

<400> 113
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

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

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

```
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320
```

```
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525
```

```
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735
```

```
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940
```

```
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065

Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125

Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140
```

```
Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170

Phe Glu  Lys Asn Pro Ile Asp  Phe Leu Glu Ala Lys  Gly Tyr Lys
    1175                1180                1185

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                1240                1245

Pro Glu  Asp Asn Glu Gln Lys  Gln Leu Phe Val Glu  Gln His Lys
    1250                1255                1260

His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile Ser Glu  Phe Ser Lys
    1265                1270                1275

Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp Lys Val  Leu Ser Ala
    1280                1285                1290

Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg Glu Gln  Ala Glu Asn
    1295                1300                1305

Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu Gly Ala  Pro Ala Ala
    1310                1315                1320

Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg Lys Arg  Tyr Thr Ser
    1325                1330                1335
```

```
Thr Lys  Glu Val Leu Asp Ala  Thr Leu Ile His Gln  Ser Ile Thr
    1340                1345                1350

Gly Leu  Tyr Glu Thr Arg Ile  Asp Leu Ser Gln Leu  Gly Gly Asp
    1355                1360                1365

Gly Ser  Pro Lys Lys Lys Arg  Lys Val Ser Ser Asp  Tyr Lys Asp
    1370                1375                1380

His Asp  Gly Asp Tyr Lys Asp  His Asp Ile Asp Tyr  Lys Asp Asp
    1385                1390                1395

Asp Asp  Lys Ala Ala Gly Gly  Gly Gly Ser Leu Pro  Thr Cys Ser
    1400                1405                1410

Cys Leu  Asp Arg Val Ile Gln  Lys Asp Lys Gly Pro  Tyr Tyr Thr
    1415                1420                1425

His Leu  Gly Ala Gly Pro Ser  Val Ala Ala Val Arg  Glu Ile Met
    1430                1435                1440

Glu Asn  Arg Tyr Gly Gln Lys  Gly Asn Ala Ile Arg  Ile Glu Ile
    1445                1450                1455

Val Val  Tyr Thr Gly Lys Glu  Gly Lys Ser Ser His  Gly Cys Pro
    1460                1465                1470

Ile Ala  Lys Trp Val Leu Arg  Arg Ser Ser Asp Glu  Glu Lys Val
    1475                1480                1485

Leu Cys  Leu Val Arg Gln Arg  Thr Gly His His Cys  Pro Thr Ala
    1490                1495                1500

Val Met  Val Val Leu Ile Met  Val Trp Asp Gly Ile  Pro Leu Pro
    1505                1510                1515

Met Ala  Asp Arg Leu Tyr Thr  Glu Leu Thr Glu Asn  Leu Lys Ser
    1520                1525                1530
```

```
Tyr Asn  Gly His Pro Thr Asp  Arg Arg Cys Thr Leu  Asn Glu Asn
    1535                1540                1545

Arg Thr  Cys Thr Cys Gln Gly  Ile Asp Pro Glu Thr  Cys Gly Ala
    1550                1555                1560

Ser Phe  Ser Phe Gly Cys Ser  Trp Ser Met Tyr Phe  Asn Gly Cys
    1565                1570                1575

Lys Phe  Gly Arg Ser Pro Ser  Pro Arg Arg Phe Arg  Ile Asp Pro
    1580                1585                1590

Ser Ser  Pro Leu His Glu Lys  Asn Leu Glu Asp Asn  Leu Gln Ser
    1595                1600                1605

Leu Ala  Thr Arg Leu Ala Pro  Ile Tyr Lys Gln Tyr  Ala Pro Val
    1610                1615                1620

Ala Tyr  Gln Asn Gln Val Glu  Tyr Glu Asn Val Ala  Arg Glu Cys
    1625                1630                1635

Arg Leu  Gly Ser Lys Glu Gly  Arg Pro Phe Ser Gly  Val Thr Ala
    1640                1645                1650

Cys Leu  Asp Phe Cys Ala His  Pro His Arg Asp Ile  His Asn Met
    1655                1660                1665

Asn Asn  Gly Ser Thr Val Val  Cys Thr Leu Thr Arg  Glu Asp Asn
    1670                1675                1680

Arg Ser  Leu Gly Val Ile Pro  Gln Asp Glu Gln Leu  His Val Leu
    1685                1690                1695

Pro Leu  Tyr Lys Leu Ser Asp  Thr Asp Glu Phe Gly  Ser Lys Glu
    1700                1705                1710

Gly Met  Glu Ala Lys Ile Lys  Ser Gly Ala Ile Glu  Val Leu Ala
    1715                1720                1725
```

```
Pro Arg  Arg Lys Lys Arg Thr  Cys Phe Thr Gln Pro  Val Pro Arg
    1730                1735                1740

Ser Gly  Lys Lys Arg Ala Ala  Met Met Thr Glu Val  Leu Ala His
    1745                1750                1755

Lys Ile  Arg Ala Val Glu Lys  Lys Pro Ile Pro Arg  Ile Lys Arg
    1760                1765                1770

Lys Asn  Asn Ser Thr Thr Thr  Asn Asn Ser Lys Pro  Ser Ser Leu
    1775                1780                1785

Pro Thr  Leu Gly Ser Asn Thr  Glu Thr Val Gln Pro  Glu Val Lys
    1790                1795                1800

Ser Glu  Thr Glu Pro His Phe  Ile Leu Lys Ser Ser  Asp Asn Thr
    1805                1810                1815

Lys Thr  Tyr Ser Leu Met Pro  Ser Ala Pro His Pro  Val Lys Glu
    1820                1825                1830

Ala Ser  Pro Gly Phe Ser Trp  Ser Pro Lys Thr Ala  Ser Ala Thr
    1835                1840                1845

Pro Ala  Pro Leu Lys Asn Asp  Ala Thr Ala Ser Cys  Gly Phe Ser
    1850                1855                1860

Glu Arg  Ser Ser Thr Pro His  Cys Thr Met Pro Ser  Gly Arg Leu
    1865                1870                1875

Ser Gly  Ala Asn Ala Ala Ala  Ala Asp Gly Pro Gly  Ile Ser Gln
    1880                1885                1890

Leu Gly  Glu Val Ala Pro Leu  Pro Thr Leu Ser Ala  Pro Val Met
    1895                1900                1905

Glu Pro  Leu Ile Asn Ser Glu  Pro Ser Thr Gly Val  Thr Glu Pro
    1910                1915                1920
```

```
Leu Thr  Pro His Gln Pro Asn  His Gln Pro Ser Phe  Leu Thr Ser
    1925                1930                1935

Pro Gln  Asp Leu Ala Ser Ser  Pro Met Glu Glu Asp  Glu Gln His
    1940                1945                1950

Ser Glu  Ala Asp Glu Pro Pro  Ser Asp Glu Pro Leu  Ser Asp Asp
    1955                1960                1965

Pro Leu  Ser Pro Ala Glu Glu  Lys Leu Pro His Ile  Asp Glu Tyr
    1970                1975                1980

Trp Ser  Asp Ser Glu His Ile  Phe Leu Asp Ala Asn  Ile Gly Gly
    1985                1990                1995

Val Ala  Ile Ala Pro Ala His  Gly Ser Val Leu Ile  Glu Cys Ala
    2000                2005                2010

Arg Arg  Glu Leu His Ala Thr  Thr Pro Val Glu His  Pro Asn Arg
    2015                2020                2025

Asn His  Pro Thr Arg Leu Ser  Leu Val Phe Tyr Gln  His Lys Asn
    2030                2035                2040

Leu Asn  Lys Pro Gln His Gly  Phe Glu Leu Asn Lys  Ile Lys Phe
    2045                2050                2055

Glu Ala  Lys Glu Ala Lys Asn  Lys Lys Met Lys Ala  Ser Glu Gln
    2060                2065                2070

Lys Asp  Gln Ala Ala Asn Glu  Gly Pro Glu Gln Ser  Ser Glu Val
    2075                2080                2085

Asn Glu  Leu Asn Gln Ile Pro  Ser His Lys Ala Leu  Thr Leu Thr
    2090                2095                2100

His Asp  Asn Val Val Thr Val  Ser Pro Tyr Ala Leu  Thr His Val
    2105                2110                2115
```

Ala Gly Pro Tyr Asn His Trp Val
2120 2125