



US 20250262326A1

(19) United States

(12) Patent Application Publication

Gersbach et al.

(10) Pub. No.: US 2025/0262326 A1

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

(54) EFFECTOR DOMAINS FOR CRISPR-CAS SYSTEMS

filed on Apr. 26, 2022, provisional application No. 63/330,691, filed on Apr. 13, 2022.

(71) Applicant: Duke University, Durham, NC (US)

Publication Classification

(72) Inventors: Charles A. Gersbach, Chapel Hill, NC (US); Gabriel Butterfield, Durham, NC (US); Dahlia Rohm, Durham, NC (US); Nahid Iglesias, Durham, NC (US)

(51) Int. Cl.

A61K 48/00 (2006.01)
C12N 9/22 (2006.01)
C12N 15/11 (2006.01)

(52) U.S. Cl.

CPC *A61K 48/0058* (2013.01); *C12N 9/226* (2025.05); *C12N 15/II* (2013.01); *C12N 2310/20* (2017.05); *C12N 2310/3513* (2013.01)

(21) Appl. No.: 18/856,563

(57)

ABSTRACT

(22) PCT Filed: Apr. 13, 2023

Disclosed herein are effector domains. The effector domains may be used with, for example, Cas proteins and CRISPR-Cas systems. The effectors may be used in combination with a Cas protein to form a fusion protein. The effectors may also be used in combination with an antibody that binds to a peptide epitope, wherein the peptide epitope is fused to a Cas protein. The compositions and methods comprising the effectors may be used to modulate gene expression.

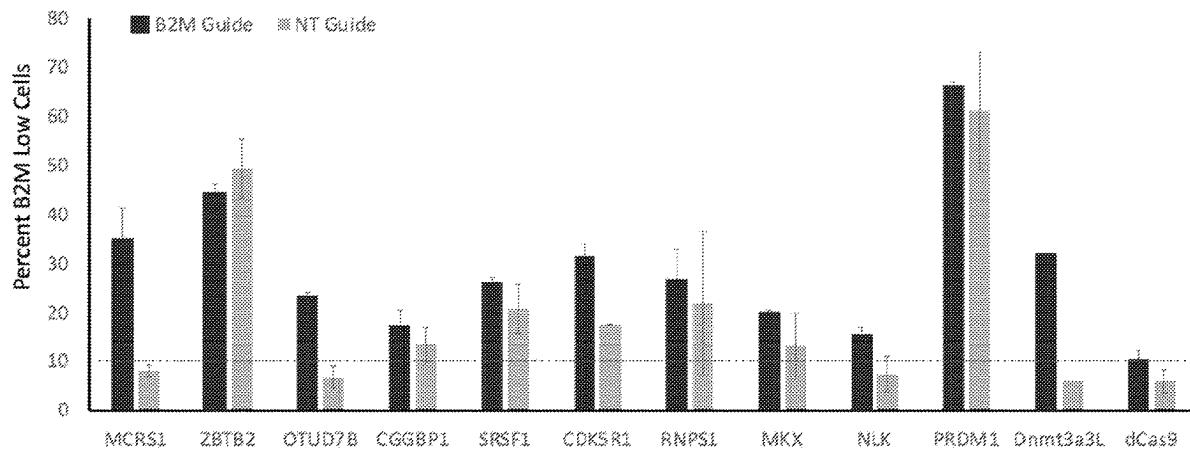
(86) PCT No.: PCT/US2023/018559

§ 371 (c)(1),
(2) Date: Oct. 11, 2024

Specification includes a Sequence Listing.

Related U.S. Application Data

(60) Provisional application No. 63/342,027, filed on May 13, 2022, provisional application No. 63/335,122,



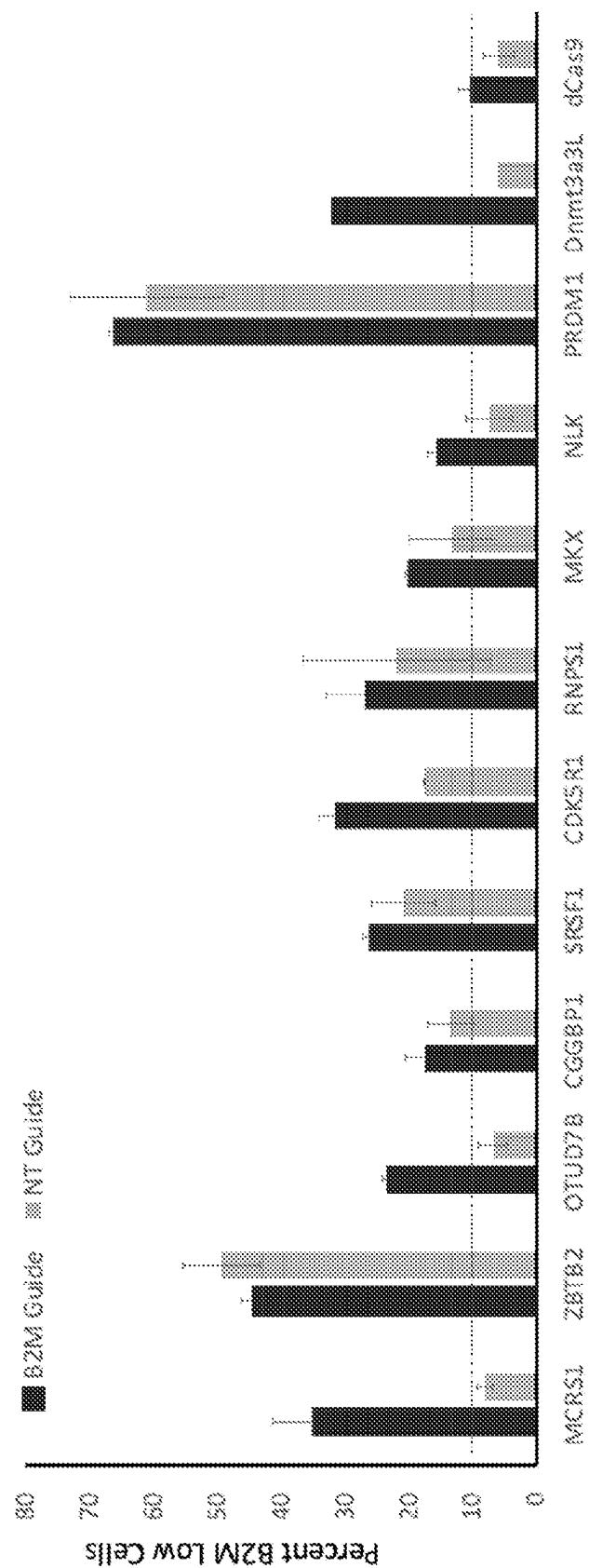


FIG. 1

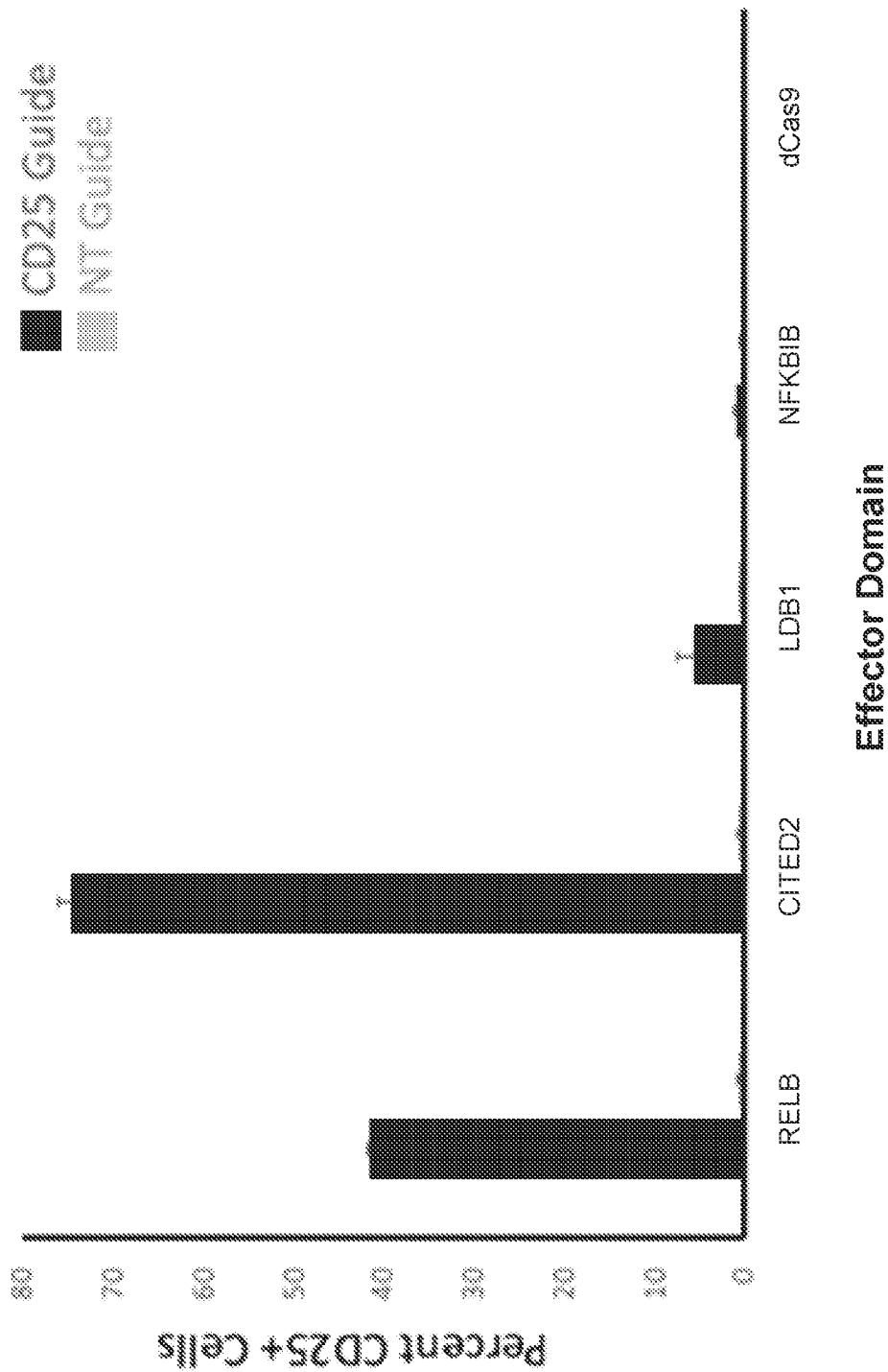


FIG. 2A

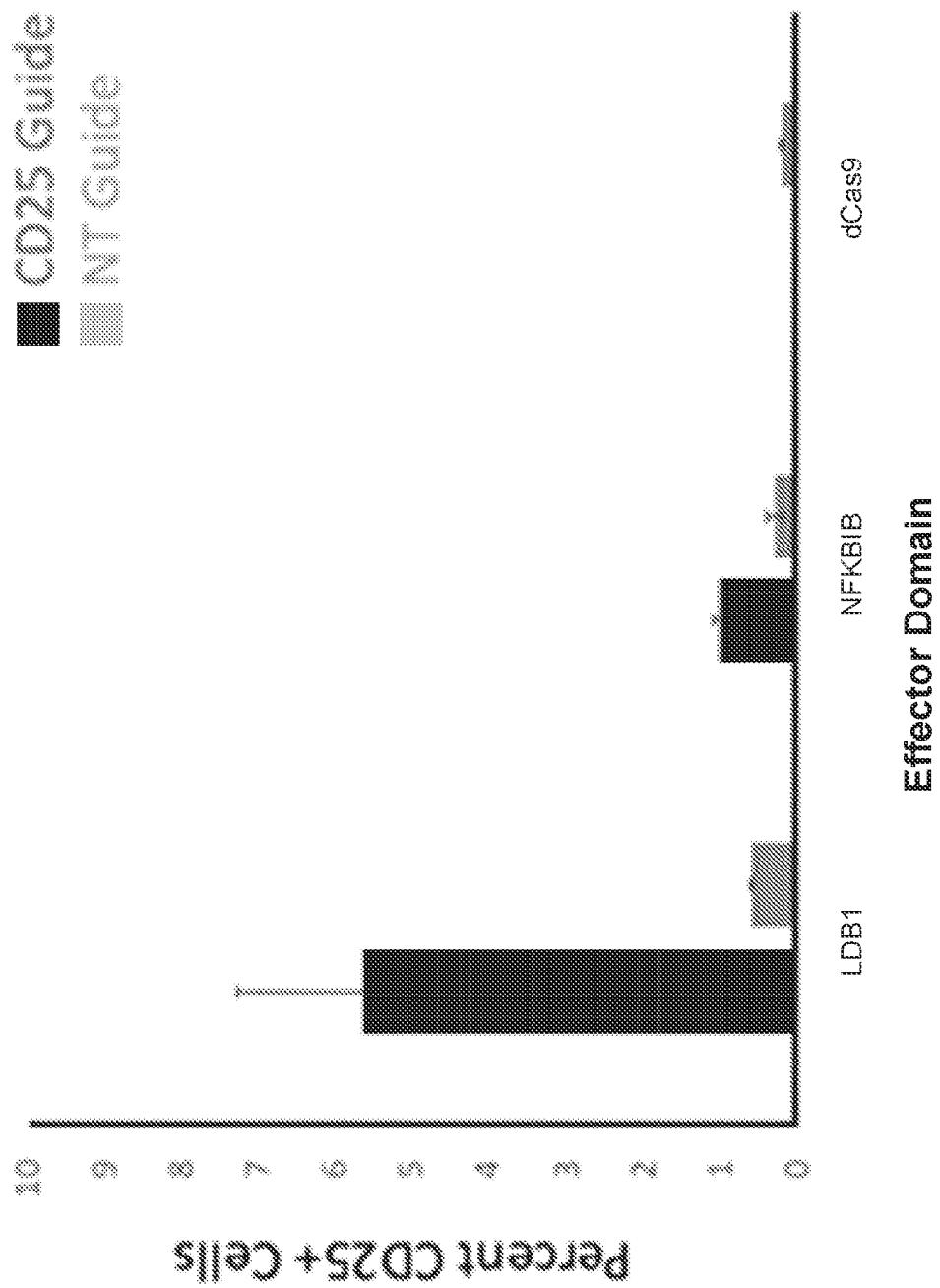


FIG. 2B

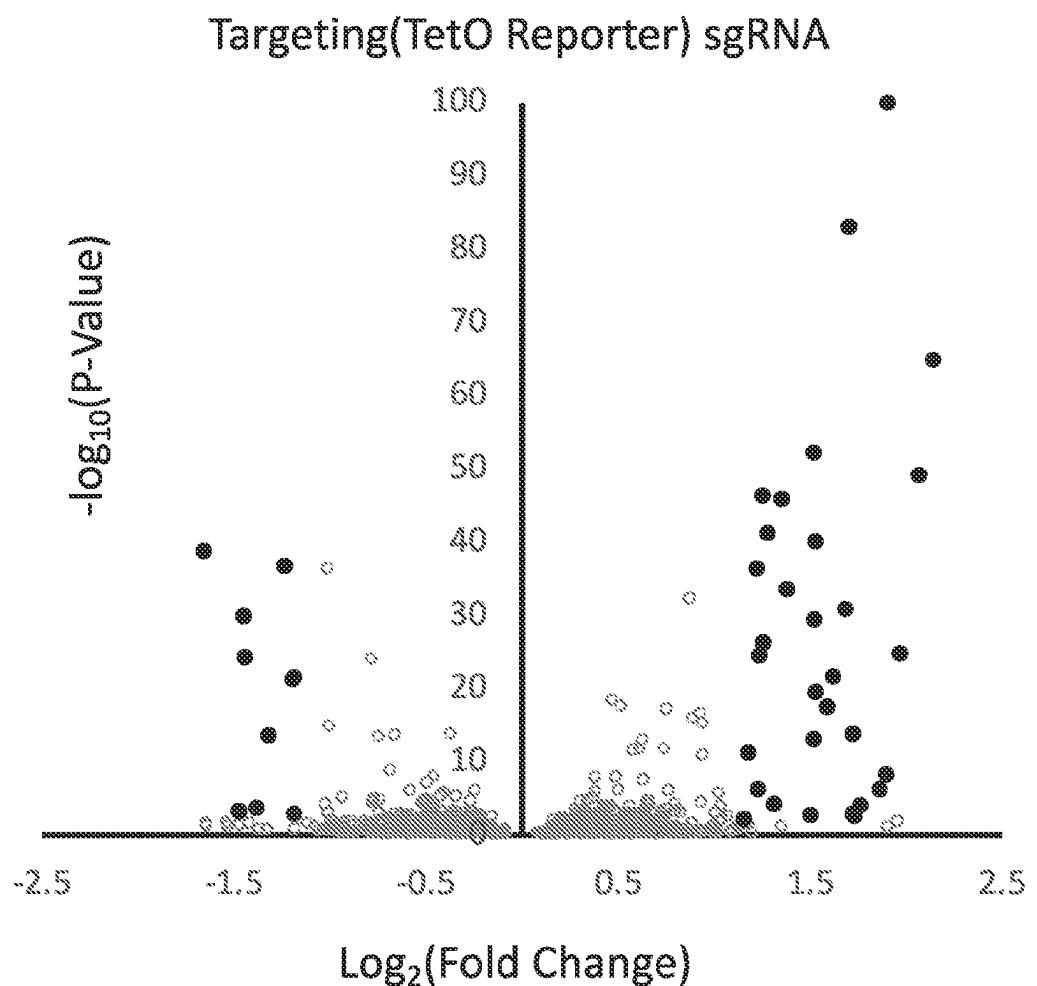


FIG. 3A

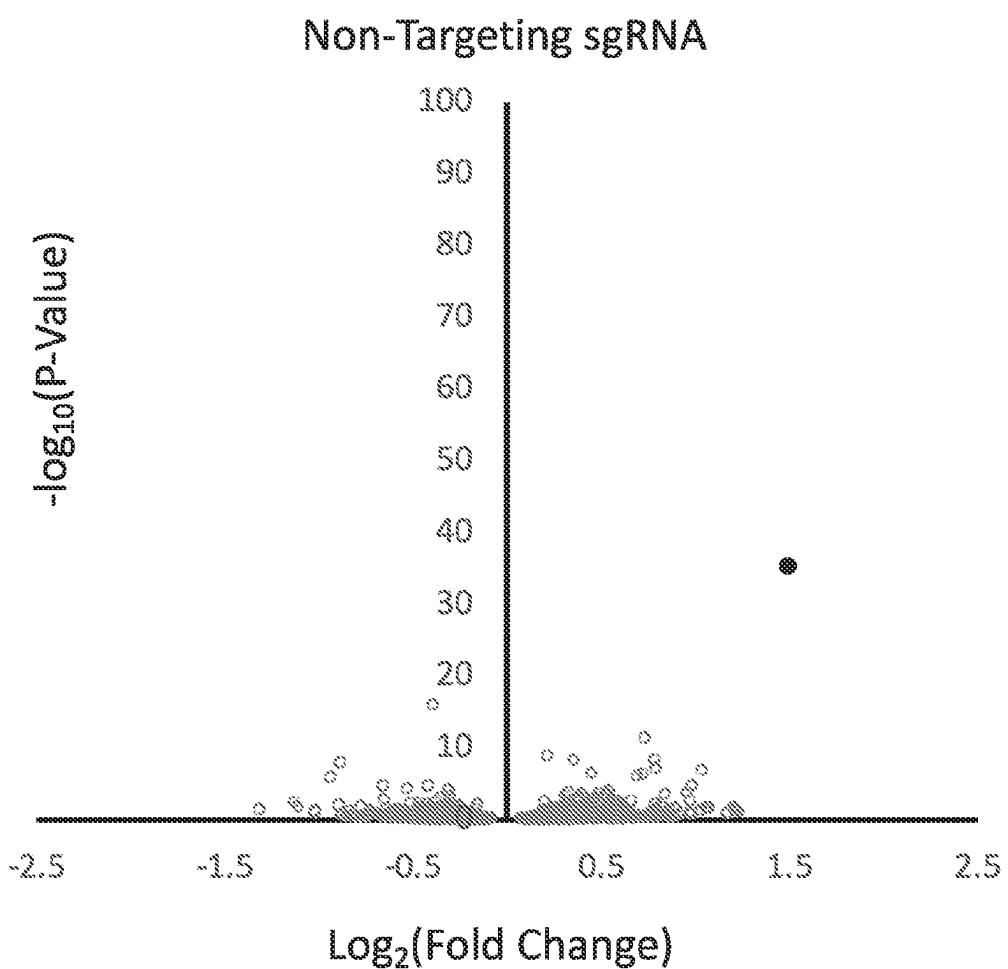


FIG. 3B

Activation of GFP Reporter Expression by Novel Effectors

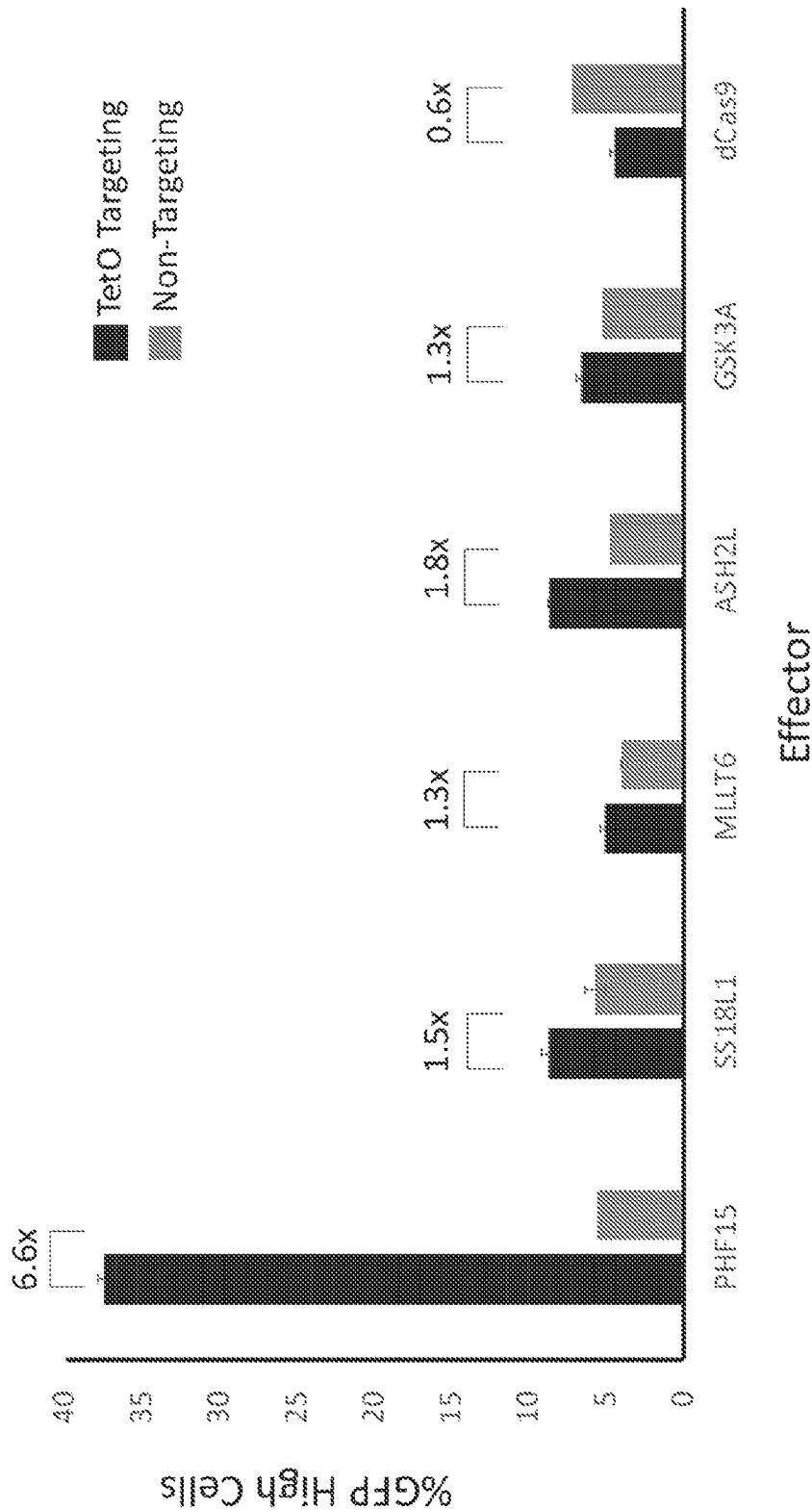


FIG. 4

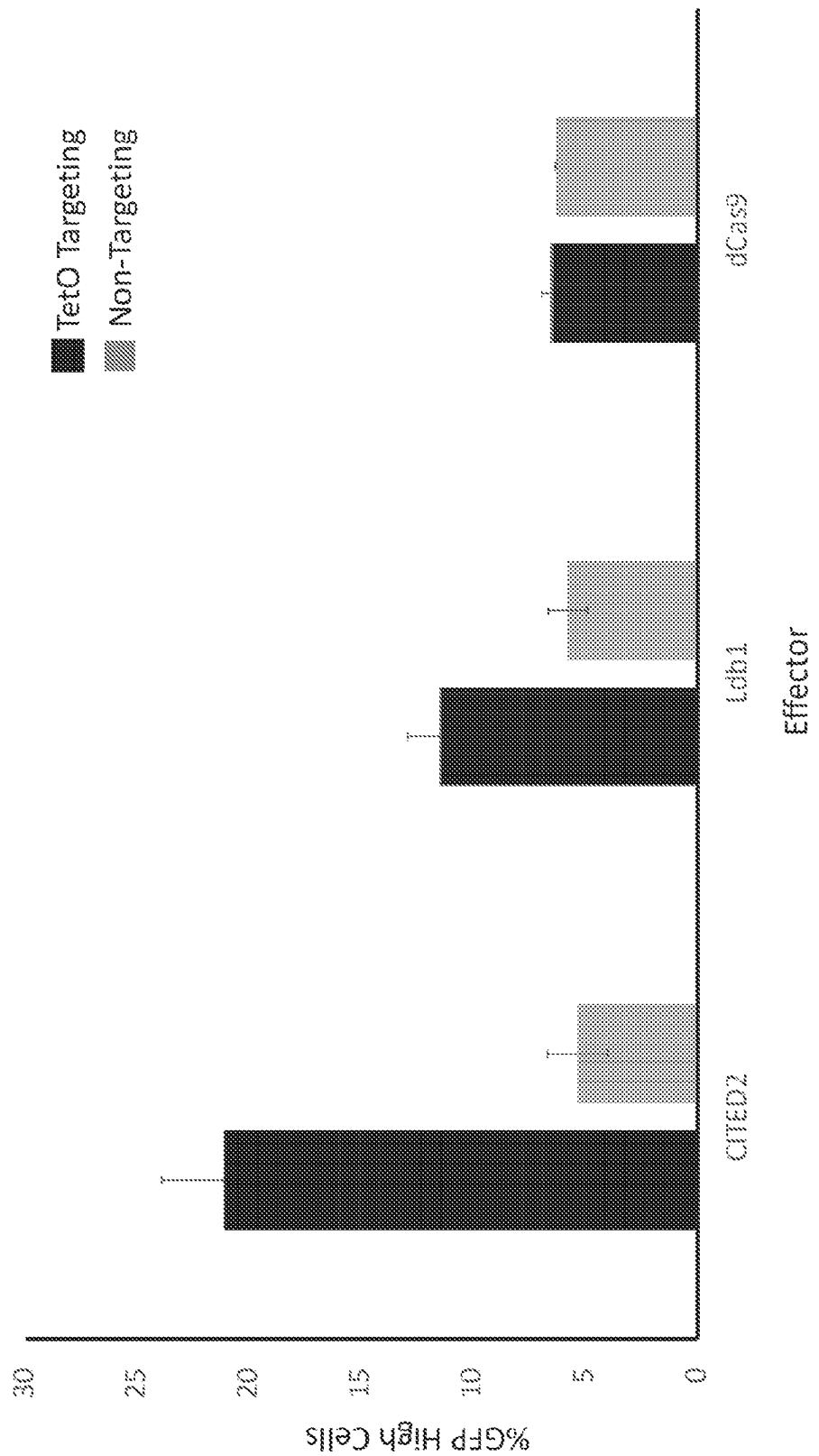


FIG. 5

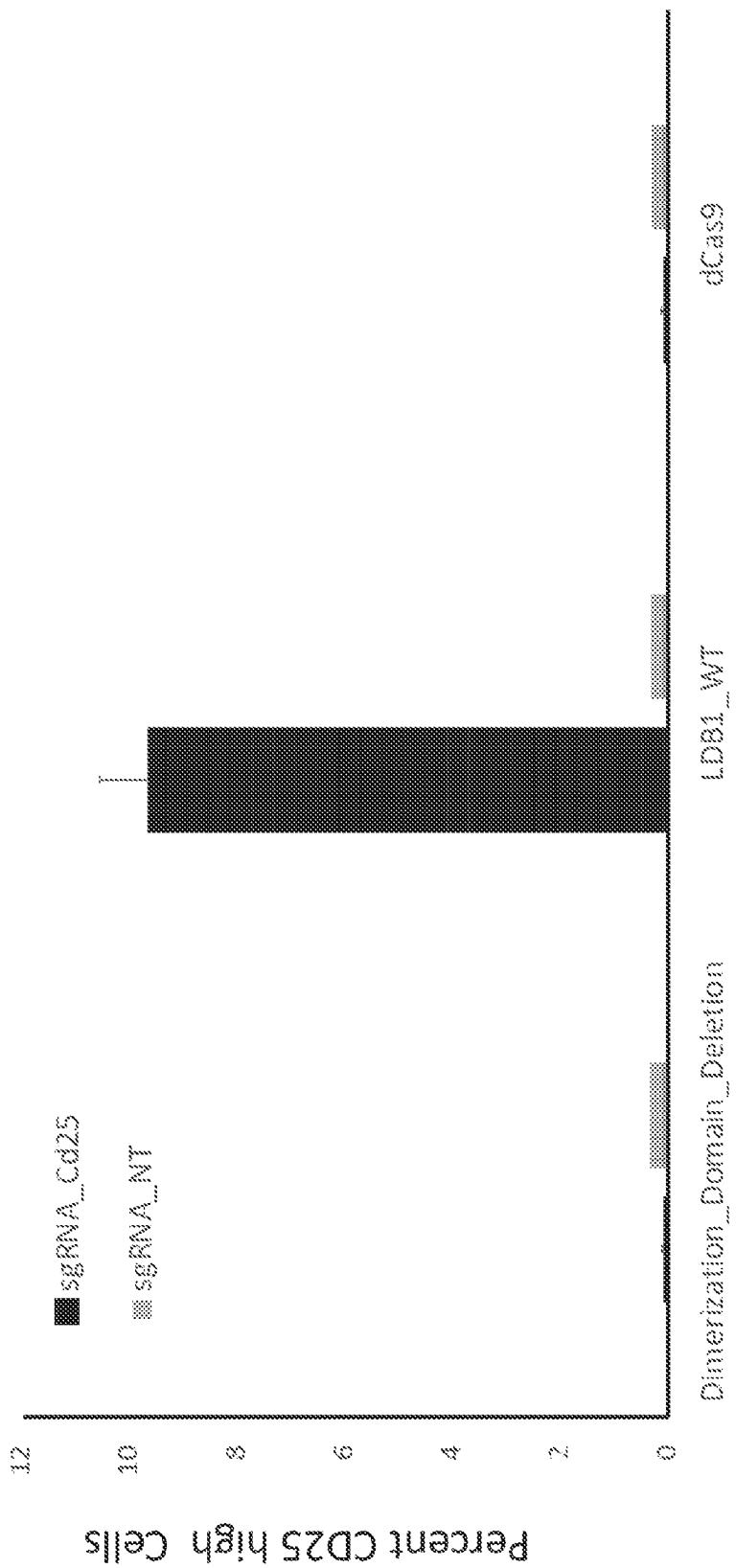


FIG. 6

EFFECTOR DOMAINS FOR CRISPR-CAS SYSTEMS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Patent Application No. 63/330,691 filed Apr. 13, 2022, U.S. Provisional Patent Application No. 63/335,122 filed Apr. 26, 2022, and U.S. Provisional Patent Application No. 63/342,027 filed May 13, 2022, the entire contents of each of which are hereby incorporated by reference.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under grant 5U01AI146356 awarded by the National Institutes of Health. The government has certain rights in the invention.

FIELD

[0003] This disclosure relates to compositions and methods including CRISPR-Cas systems with effector domains. The effector domains, which may be used, for example, in combination with a Cas protein, may be used to modulate gene expression.

INTRODUCTION

[0004] Synthetic transcription factors have been engineered to control gene expression for many different medical and scientific applications in mammalian systems, including stimulating tissue regeneration, drug screening, compensating for genetic defects, activating silenced tumor suppressors, controlling stem cell differentiation, performing genetic screens, and creating synthetic gene circuits. These transcription factors can target promoters or enhancers of endogenous genes or be purposefully designed to recognize sequences orthogonal to mammalian genomes for transgene regulation.

[0005] Further, these synthetic transcription factors rely on naturally occurring or designed effector protein domains which modulate gene expression. However, the full spectrum of regulatory mechanisms employed in mammalian cells cannot be programmed with currently described effector domains. Broadening the set of available effectors will enable both more potent and more specific gene activation and repression.

SUMMARY

[0006] In an aspect, the disclosure relates to a Cas effector. The Cas effector may include a first polypeptide comprising a Cas protein and at least one peptide epitope; and a second polypeptide comprising an effector selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, VPS72, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof, and an antibody to the peptide epitope. In some embodiments, the effector is selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB,

CITED2, PHF15, SS18L1, MLLT6, ASH2L, and GSK3A, or a combination thereof. In some embodiments, the effector is capable of increasing or decreasing expression of a gene. In some embodiments, the effector reduces expression of a target gene and is selected from MCRS1, OTUD7B, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof. In some embodiments, the effector increases expression of a target gene and is selected from RelB, LDB1, NFKBIB, CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, and VPS72, or a combination thereof. In some embodiments, the first polypeptide comprises about 2 to about 50 peptide epitopes. In some embodiments, the first polypeptide comprises more than one copy of the peptide epitope and further comprises at least one linker in between adjacent copies of the peptide epitope. In some embodiments, the peptide epitope is GCN4 and comprises the amino acid sequence of SEQ ID NO: 85. In some embodiments, the first polypeptide comprises at least one peptide epitope at the N-terminus and/or at the C-terminus of the Cas protein. In some embodiments, the first polypeptide comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 87 or 89, or any fragment thereof, or the first polypeptide comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 87 or 89, or any fragment thereof, or the first polypeptide comprises the amino acid sequence of SEQ ID NO: 87 or 89. In some embodiments, the antibody comprises the amino acid sequence of SEQ ID NO: 81. In some embodiments, the second polypeptide comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to a sequence selected from SEQ ID NOs: 69, 71, 73, 75, 77, and 79, or any fragment thereof, or the second polypeptide comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to a sequence selected from SEQ ID NOs: 69, 71, 73, 75, 77, and 79, or any fragment thereof, or the second polypeptide comprises an amino acid sequence selected from SEQ ID NOs: 69, 71, 73, 75, 77, and 79.

[0007] In a further aspect, the disclosure relates to a Cas fusion protein. The Cas fusion protein may include two heterologous polypeptide domains, wherein the first polypeptide domain comprises a Cas protein, and wherein the second polypeptide domain comprises an effector selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, and CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, VPS72, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof. In some embodiments, the effector is selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, CITED2, PHF15, SS18L1, MLLT6, ASH2L, and GSK3A, or a combination thereof. In some embodiments, the effector is capable of increasing or decreasing expression of a gene. In some embodiments, the effector reduces expression of a target gene and is selected

from MCRS1, OTUD7B, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof. In some embodiments, the effector increases expression of a target gene and is selected from RelB, LDB1, NFKBIB, CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, and VPS72, or a combination thereof. In some embodiments, the second polypeptide domain has transcription repression activity, transcription activation activity, de-ubiquitinase activity, p300 recruitment activity, enhancer looping mediation activity, or a combination thereof.

[0008] In some embodiments, the MCRS1 comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 57 or any fragment thereof, and/or the MCRS1 comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 57, or any fragment thereof, and/or the MCRS1 comprises the amino acid sequence of SEQ ID NO: 57, and/or the MCRS1 is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 58, or any fragment thereof, and/or the MCRS1 is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 58, or any fragment thereof, and/or the MCRS1 is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 58. In some embodiments, the OTUD7B comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to a sequence selected from SEQ ID NO: 59, amino acids 167-440 of SEQ ID NO: 59, or amino acids 792-831 of SEQ ID NO: 59, or any fragment thereof, and/or the OTUD7B comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to a sequence selected from SEQ ID NO: 59, amino acids 167-440 of SEQ ID NO: 59, or amino acids 792-831 of SEQ ID NO: 59, or any fragment thereof, and/or the OTUD7B comprises the amino acid sequence selected from SEQ ID NO: 59, amino acids 167-440 of SEQ ID NO: 59, or amino acids 792-831 of SEQ ID NO: 59, and/or the OTUD7B is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 60, or any fragment thereof, and/or the OTUD7B is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 60, or any fragment thereof, and/or the OTUD7B is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 60. In some embodiments, the RelB comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 65, or any fragment thereof, and/or the RelB comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 65, or any fragment thereof, and/or the RelB comprises the amino acid sequence of SEQ ID NO: 65, and/or the RelB is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%,

95%, or 98% or greater identity to SEQ ID NO: 66 or any fragment thereof, and/or the RelB is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 66, or any fragment thereof, and/or the RelB is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 66. In some embodiments, the LDB1 comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 61, or any fragment thereof, and/or the LDB1 comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 61, or any fragment thereof, and/or the LDB1 comprises the amino acid sequence of SEQ ID NO: 61, and/or the LDB1 is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 62, or any fragment thereof, and/or the LDB1 is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 62, or any fragment thereof, and/or the LDB1 is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 62. In some embodiments, the NFKBIB comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 63, or any fragment thereof, and/or the NFKBIB comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 63, or any fragment thereof, and/or the NFKBIB comprises the amino acid sequence of SEQ ID NO: 63, and/or the NFKBIB is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 64, or any fragment thereof, and/or the NFKBIB is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 64, or any fragment thereof, and/or the NFKBIB is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 64. In some embodiments, the CITED2 comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 67, or any fragment thereof, and/or the CITED2 comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 67, or any fragment thereof, and/or the CITED2 comprises the amino acid sequence of SEQ ID NO: 67, and/or the CITED2 is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 68, or any fragment thereof, and/or the CITED2 is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 68, or any fragment thereof, and/or the CITED2 is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 68. In some embodiments, the PHF15 comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 133, or any fragment thereof, and/or the PHF15 comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, inser-

tions, or deletions, relative to SEQ ID NO: 133, or any fragment thereof, and/or the PHF15 comprises the amino acid sequence of SEQ ID NO: 133, and/or the PHF15 is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 134, or any fragment thereof, and/or the PHF15 is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 134, or any fragment thereof, and/or the PHF15 is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 134. In some embodiments, the SS18L1 comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 149, or any fragment thereof, and/or the SS18L1 comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 149, or any fragment thereof, and/or the SS18L1 comprises the amino acid sequence of SEQ ID NO: 149, and/or the SS18L1 is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 150, or any fragment thereof, and/or the SS18L1 is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 150, or any fragment thereof, and/or the SS18L1 is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 150. In some embodiments, the MLLT6 comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 127, or any fragment thereof, and/or the MLLT6 comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 127, or any fragment thereof, and/or the MLLT6 comprises the amino acid sequence of SEQ ID NO: 127, and/or the MLLT6 is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 128, or any fragment thereof, and/or the MLLT6 is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 128, or any fragment thereof, and/or the MLLT6 is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 128. In some embodiments, the ASH2L comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 103, or any fragment thereof, and/or the ASH2L comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 103, or any fragment thereof, and/or the ASH2L comprises the amino acid sequence of SEQ ID NO: 103, and/or the ASH2L is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 104, or any fragment thereof, and/or the ASH2L is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 104, or any fragment thereof, and/or the ASH2L is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 104. In some embodiments, the GSK3A

comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 117, or any fragment thereof, and/or the GSK3A comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 117, or any fragment thereof, and/or the GSK3A comprises the amino acid sequence of SEQ ID NO: 117, and/or the GSK3A is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 118, or any fragment thereof, and/or the GSK3A is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 118, or any fragment thereof, and/or the GSK3A is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 118. In some embodiments, the effector is selected from BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, JAZF1, KAT7, KEAP1, MEAF6, MORF4L2, NFYC, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, TADA3, TAF6, TBPL1, VPS72, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, and wherein the effector comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to a sequence selected from SEQ ID NOS: 105, 107, 109, 111, 113, 115, 119, 121, 123, 125, 129, 131, 135, 137, 139, 141, 143, 145, 147, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, or 175, or any fragment thereof, and/or wherein the effector comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to a sequence selected from SEQ ID NOS: 105, 107, 109, 111, 113, 115, 119, 121, 123, 125, 129, 131, 135, 137, 139, 141, 143, 145, 147, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, or 175, or any fragment thereof, and/or wherein the effector comprises an amino acid sequence selected from SEQ ID NOS: 105, 107, 109, 111, 113, 115, 119, 121, 123, 125, 129, 131, 135, 137, 139, 141, 143, 145, 147, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, or 175, and/or wherein the effector is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to a sequence selected from SEQ ID NOS: 106, 108, 110, 112, 114, 116, 120, 122, 124, 126, 130, 132, 136, 138, 140, 142, 144, 146, 148, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, or 176, or any fragment thereof, and/or wherein the effector is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to a sequence selected from SEQ ID NOS: 106, 108, 110, 112, 114, 116, 120, 122, 124, 126, 130, 132, 136, 138, 140, 142, 144, 146, 148, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, or 176, or any fragment thereof, and/or wherein the effector is encoded by a polynucleotide comprising a sequence selected from SEQ ID NOS: 106, 108, 110, 112, 114, 116, 120, 122, 124, 126, 130, 132, 136, 138, 140, 142, 144, 146, 148, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, or 176. In some embodiments, the Cas protein comprises at least one amino acid mutation that knocks out nuclease activity of the Cas protein. In some embodiments, the at least one amino acid mutation is at least one of D10A and H840A. In some embodiments, the Cas protein comprises an amino acid

sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to one of SEQ ID NOS: 26-29, or any fragment thereof, or the Cas protein comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to one of SEQ ID NOS: 26-29, or any fragment thereof, or the Cas protein comprises the amino acid sequence of one of SEQ ID NOS: 26-29. In some embodiments, the Cas protein is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to one of SEQ ID NOS: 30-31, or any fragment thereof, or the Cas protein is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to one of SEQ ID NOS: 30-31, or any fragment thereof, or the Cas protein is encoded by a polynucleotide comprising the sequence of one of SEQ ID NOS: 30-31.

[0009] Another aspect of the disclosure provides a DNA targeting composition. The DNA targeting composition may include a Cas effector as detailed herein or a Cas fusion protein as detailed herein; and at least one guide RNA (gRNA) that targets the Cas protein to a target region of a target gene. In some embodiments, the gRNA targets the Cas protein to target region selected from a non-open chromatin region, an open chromatin region, a transcribed region of the target gene, a region upstream of a transcription start site of the target gene, a regulatory element of the target gene, an intron of the target gene, or an exon of the target gene. In some embodiments, the gRNA targets the Cas protein to a promoter of the target gene. In some embodiments, the target region is located between about 1 to about 1000 base pairs upstream of a transcription start site of the target gene. In some embodiments, the at least one gRNA comprises a sequence selected from SEQ ID NOS: 96-98 and 101-102, or the at least one gRNA is encoded by a polynucleotide comprising a sequence selected from SEQ ID NOS: 93-95 and 99-100, or the at least one gRNA targets and binds a polynucleotide comprising a sequence selected from SEQ ID NOS: 93-95 and 99-100 or a complement thereof, or a combination thereof. In some embodiments, the DNA targeting composition comprises two or more gRNAs, each gRNA binding to a different target region.

[0010] Another aspect of the disclosure provides an isolated polynucleotide sequence encoding a Cas effector as detailed herein or a Cas fusion protein as detailed herein, or a DNA targeting composition as detailed herein.

[0011] Another aspect of the disclosure provides a vector comprising an isolated polynucleotide sequence as detailed herein. In some embodiments, the vector is an adeno-associated virus (AAV) vector.

[0012] Another aspect of the disclosure provides a cell comprising a Cas effector as detailed herein or a Cas fusion protein as detailed herein, or a DNA targeting composition as detailed herein, or an isolated polynucleotide sequence as detailed herein, or a vector as detailed herein, or a combination thereof.

[0013] Another aspect of the disclosure provides a pharmaceutical composition. The pharmaceutical composition may include a Cas effector as detailed herein or a Cas fusion protein as detailed herein, or a DNA targeting composition as detailed herein, or an isolated polynucleotide sequence as detailed herein, or a vector as detailed herein, or a combination thereof.

[0014] Another aspect of the disclosure provides a method of modulating expression of a gene in a cell or in a subject. The method may include administering to the cell or the subject a DNA targeting composition as detailed herein, or an isolated polynucleotide sequence as detailed herein, or a vector as detailed herein, or a pharmaceutical composition as detailed herein, or a combination thereof. The method may include administering to the cell or the subject an effector selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, VPS72, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof, or a polynucleotide encoding the effector. In some embodiments, the effector is targeted to the gene. In some embodiments, the effector is selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, CITED2, PHF15, SS18L1, MLLT6, ASH2L, and GSK3A, or a combination thereof. In some embodiments, the effector is capable of increasing or decreasing expression of the gene. In some embodiments, the effector reduces expression of the gene and is selected from MCRS1, OTUD7B, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof. In some embodiments, the effector increases expression of the gene and is selected from RelB, LDB1, NFKBIB, CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, and VPS72, or a combination thereof. In some embodiments, the expression of the gene is increased relative to a control. In some embodiments, the expression of the gene is decreased relative to a control. In some embodiments, the gene comprises the dystrophin gene, the CD25 gene, the B2M gene, or the TRAC gene. In some embodiments, the cell is a muscle cell or a T cell.

[0015] Another aspect of the disclosure provides a method of treating a disease in a subject. The method may include administering to the subject a DNA targeting composition as detailed herein, or an isolated polynucleotide sequence as detailed herein, or a vector as detailed herein, or a cell as detailed herein, or a pharmaceutical composition as detailed herein, or a combination thereof. The method may include administering to the subject an effector selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, VPS72, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof, or a polynucleotide encoding the effector. In some embodiments, the effector is targeted to a gene. In some embodiments, the method treats a disease selected from Duchenne muscular dystrophy (DMD), Becker muscular dystrophy (BMD), and cancer.

[0016] The disclosure provides for other aspects and embodiments that will be apparent in light of the following detailed description and accompanying figures.

BRIEF DESCRIPTION OF THE DRAWINGS

[0017] FIG. 1 is a graph showing the results from the individual testing of top repressor effectors from B2M screen. The graph displays the percent of cells in the low B2M bin, with higher numbers suggesting more potent repression. A non-targeting guide was also included as a control for non-specific repression. MCRS1 and OTUD7B both showed repression that was both greater than the steric effects of dCas9 alone and largely dependent on dCas9 targeting, rather than a non-specific effect.

[0018] Shown in FIG. 2A is the level of CD25 activation after delivery of each effector domain recruited by dCas9 in Jurkat cells. A non-targeting guide (gray bars) showed no effect on CD25, suggesting that each effector was specifically activating CD25 upon recruitment by dCas9. Shown in FIG. 2B is a zoomed-in view of data in FIG. 2A to show the specific activation by LDB1 and NFKBIB.

[0019] FIGS. 3A-3B are graphs showing the results for each effector in a screen for the ability to modulate expression of TetO with a GFP reporter. Log 2(fold change) and Log 10(Adjusted P Value) for each effector in the screen are plotted. Shown in FIG. 3A are results with a gRNA targeting TetO, and shown in FIG. 3B are results with a non-targeting gRNA. Effectors with Log 2(fold change)>1.1 and Adjusted P Value <0.01 were considered to be hits and are shown in filled black circles, while non-hits are shown in open gray circles. This threshold gave 41 hits in the targeting condition and only 1 hit in the non-targeting condition.

[0020] FIG. 4 shows GFP reporter expression in the TetO-GFP reporter screen in 293T cells for a subset of effectors, including PHF15, SS18L1, MLLT6, ASH2L, and GSK3A. 293T cells containing a GFP reporter were transduced with Lentivirus encoding a subset of effectors found to be hits in the high-throughput screen along with a targeting (black) or non-targeting (gray) sgRNA. The fold activation of GFP (shown above each pair of bars) was found to be greater than 1 for all effectors tested, while the dCas9 alone control showed the opposite trend, supporting the idea that even the small effects seen for some effectors are likely meaningful. All hit effectors tested did modulate GFP to some degree.

[0021] FIG. 5 is a graph showing activation of TetO with a GFP reporter in 293T cells by CITED2 and LDB1. 293T cells previously transduced with a TetO-GFP reporter were transfected with the indicated effector. Both LDB1 and CITED2 robustly activated GFP expression, demonstrating that activation by these effectors is not limited to CD25.

[0022] FIG. 6 is a graph showing activation of CD25 expression with either wild-type LDB1 or LDB1 with a deletion in its dimerization domain. Jurkat cells expressing dCas9-GCN4 and a CD25-targeting gRNA or non-targeting gRNA were transduced with the indicated effector-scFv fusion, and CD25 expression was analyzed by flow cytometry 10 days later. Only the intact LDB1 effector was able to activate CD25 expression.

DETAILED DESCRIPTION

[0023] Disclosed herein is a set of novel effectors that may activate or repress gene expression when recruited to the gene, for example, via a Cas protein such as dCas9. As detailed herein, the human genome was screened for potential proteins that impact gene expression. The proteins may be referred to as effectors or effector domains. Several novel effectors were discovered, including MCRS1, OTUD7B,

RelB, LDB1, NFKBIB, CITED2, ASH2L, GSK3A, MLLT6, PHF15, SS18L1, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, JAZF1, KAT7, KEAP1, MEAF6, MORF4L2, NFYC, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, TADA3, TAF6, TBPL1, VPS72, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81. These effectors may be used in combination with a Cas protein, for example, to target a region of a gene or other DNA sequence. The effector and a Cas protein may form a fusion protein. In other embodiments, the effector is used in combination with an antibody, a peptide epitope is fused to a Cas protein, and binding of the antibody to the peptide epitope brings the effector proximal to the Cas protein. The effector and Cas protein may be used to modulate expression of a gene. The effector and Cas protein may also be used to treat various diseases.

1. DEFINITIONS

[0024] 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. In case of conflict, the present document, including definitions, will control. Preferred methods and materials are described below, although methods and materials similar or equivalent to those described herein can be used in practice or testing of the present invention. All publications, patent applications, patents and other references mentioned herein are incorporated by reference in their entirety. The materials, methods, and examples disclosed herein are illustrative only and not intended to be limiting.

[0025] The terms "comprise(s)," "include(s)," "having," "has," "can," "contain(s)," and variants thereof, as used herein, are intended to be open-ended transitional phrases, terms, or words that do not preclude the possibility of additional acts or structures. The singular forms "a," "and," and "the" include plural references unless the context clearly dictates otherwise. The present disclosure also contemplates other embodiments "comprising," "consisting of," and "consisting essentially of," the embodiments or elements presented herein, whether explicitly set forth or not.

[0026] For the recitation of numeric ranges herein, each intervening number there between with the same degree of precision is explicitly contemplated. For example, for the range of 6-9, the numbers 7 and 8 are contemplated in addition to 6 and 9, and for the range 6.0-7.0, the number 6.0, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, and 7.0 are explicitly contemplated.

[0027] The term "about" or "approximately" as used herein as applied to one or more values of interest, refers to a value that is similar to a stated reference value, or within an acceptable error range for the particular value as determined by one of ordinary skill in the art, which will depend in part on how the value is measured or determined, such as the limitations of the measurement system. In certain aspects, the term "about" refers to a range of values that fall within 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, or less in either direction (greater than or less than) of the stated reference value unless otherwise stated or otherwise evident from the context (except where such number would exceed 100% of a possible value). Alternatively, "about" can mean within 3 or more than 3 standard deviations, per the practice in the art. Alternatively, such as with respect to biological

systems or processes, the term "about" can mean within an order of magnitude, preferably within 5-fold, and more preferably within 2-fold, of a value.

[0028] "Adeno-associated virus" or "AAV" as used interchangeably herein refers to a small virus belonging to the genus Dependovirus of the Parvoviridae family that infects humans and some other primate species. AAV is not currently known to cause disease and consequently the virus causes a very mild immune response.

[0029] "Allogeneic" refers to any material derived from another subject of the same species. Allogeneic cells are genetically distinct and immunologically incompatible yet belong to the same species. Typically, "allogeneic" is used to define cells, such as stem cells, that are transplanted from a donor to a recipient of the same species.

[0030] "Amino acid" as used herein refers to naturally occurring and non-natural synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code. Amino acids can be referred to herein by either their commonly known three-letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Amino acids include the side chain and polypeptide backbone portions.

[0031] "Autologous" refers to any material derived from a subject and re-introduced to the same subject.

[0032] "Binding region" as used herein refers to the region within a target region that is recognized and bound by the CRISPR/Cas-based gene editing system.

[0033] The terms "cancer", "cancer cell", "tumor", and "tumor cell" are used interchangeably herein and refer generally to a group of diseases characterized by uncontrolled, abnormal growth of cells (e.g., a neoplasia). In some forms of cancer, the cancer cells can spread locally or through the bloodstream and lymphatic system to other parts of the body ("metastatic cancer"). "Cancer" refers to all types of cancer or neoplasm or malignant tumors found in animals, including carcinoma, adenoma, melanoma, sarcoma, lymphoma, leukemia, blastoma, glioma, astrocytoma, mesothelioma, or a germ cell tumor. Cancer may include cancer of, for example, the colon, rectum, stomach, bladder, cervix, uterus, skin, epithelium, muscle, kidney, liver, lymph, bone, blood, ovary, prostate, lung, brain, head and neck, and/or breast. Cancer may include medullablastoma, non-small cell lung cancer, and/or mesothelioma. In embodiments detailed herein, the cancer includes leukemia. The term "leukemia" refers to broadly progressive, malignant diseases of the hematopoietic organs/systems and is generally characterized by a distorted proliferation and development of leukocytes and their precursors in the blood and bone marrow. Leukemia diseases include, for example, acute nonlymphocytic leukemia, chronic lymphocytic leukemia, acute granulocytic leukemia, chronic granulocytic leukemia, acute promyelocytic leukemia, adult T-cell leukemia, aleukemic leukemia, a leukocytemic leukemia, basophilic leukemia, blast cell leukemia, bovine leukemia, chronic myelocytic leukemia, leukemia cutis, embryonal leukemia, eosinophilic leukemia, Gross' leukemia, Rieder cell leukemia, Schilling's leukemia, stem cell leukemia, subleukemic leukemia, undifferentiated cell leukemia, hairy-cell leukemia, hemoblastic leukemia, hemocytoblastic leukemia, histiocytic leukemia, stem cell leukemia, acute monocytic leukemia, leukopenic leukemia, lymphatic leu-

kemia, lymphoblastic leukemia, lymphocytic leukemia, lymphogenous leukemia, lymphoid leukemia, lymphosarcoma cell leukemia, mast cell leukemia, megakaryocytic leukemia, micromyeloblastic leukemia, monocytic leukemia, myeloblastic leukemia, myelocytic leukemia, myeloid leukemia, myeloid granulocytic leukemia, myelomonocytic leukemia, Naegeli leukemia, plasma cell leukemia, plasma-cytic leukemia, and promyelocytic leukemia. In some embodiments, the leukemia is chronic myeloid leukemia (CML). In some embodiments, the leukemia is acute myeloid leukemia (AML).

[0034] "Clustered Regularly Interspaced Short Palindromic Repeats" and "CRISPRs", as used interchangeably herein, refers to loci containing multiple short direct repeats that are found in the genomes of approximately 40% of sequenced bacteria and 90% of sequenced archaea.

[0035] "Coding sequence" or "encoding nucleic acid" as used herein means the nucleic acids (RNA or DNA molecule) that comprise a nucleotide sequence which encodes a protein. The coding sequence can further include initiation and termination signals operably linked to regulatory elements including a promoter and polyadenylation signal capable of directing expression in the cells of an individual or mammal to which the nucleic acid is administered. The regulatory elements may include, for example, a promoter, an enhancer, an initiation codon, a stop codon, or a polyadenylation signal. The coding sequence may be codon optimized.

[0036] "Complement" or "complementary" as used herein means a nucleic acid can mean Watson-Crick (e.g., A-T/U and C-G) or Hoogsteen base pairing between nucleotides or nucleotide analogs of nucleic acid molecules. "Complementarity" refers to a property shared between two nucleic acid sequences, such that when they are aligned antiparallel to each other, the nucleotide bases at each position will be complementary.

[0037] The terms "control," "reference level," and "reference" are used herein interchangeably. The reference level may be a predetermined value or range, which is employed as a benchmark against which to assess the measured result. "Control group" as used herein refers to a group of control subjects. The predetermined level may be a cutoff value from a control group. The predetermined level may be an average from a control group. Cutoff values (or predetermined cutoff values) may be determined by Adaptive Index Model (AIM) methodology. Cutoff values (or predetermined cutoff values) may be determined by a receiver operating curve (ROC) analysis from biological samples of the patient group. ROC analysis, as generally known in the biological arts, is a determination of the ability of a test to discriminate one condition from another, e.g., to determine the performance of each marker in identifying a patient having CRC. A description of ROC analysis is provided in P. J. Heagerty et al. (*Biometrics* 2000, 56, 337-44), the disclosure of which is hereby incorporated by reference in its entirety. Alternatively, cutoff values may be determined by a quartile analysis of biological samples of a patient group. For example, a cutoff value may be determined by selecting a value that corresponds to any value in the 25th-75th percentile range, preferably a value that corresponds to the 25th percentile, the 50th percentile or the 75th percentile, and more preferably the 75th percentile. Such statistical analyses may be performed using any method known in the art and can be implemented through any number of commercially available

software packages (e.g., from Analyse-it Software Ltd., Leeds, UK; StataCorp LP, College Station, TX; SAS Institute Inc., Cary, NC.). The healthy or normal levels or ranges for a target or for a protein activity may be defined in accordance with standard practice. A control may be a subject or cell without a composition as detailed herein. A control may be a subject, or a sample therefrom, whose disease state is known. The subject, or sample therefrom, may be healthy, diseased, diseased prior to treatment, diseased during treatment, or diseased after treatment, or a combination thereof.

[0038] “Correcting”, “gene editing,” and “restoring” as used herein refers to changing a mutant gene that encodes a dysfunctional protein or truncated protein or no protein at all, such that a full-length functional or partially full-length functional protein expression is obtained. Correcting or restoring a mutant gene may include replacing the region of the gene that has the mutation or replacing the entire mutant gene with a copy of the gene that does not have the mutation with a repair mechanism such as homology-directed repair (HDR). Correcting or restoring a mutant gene may also include repairing a frameshift mutation that causes a premature stop codon, an aberrant splice acceptor site or an aberrant splice donor site, by generating a double stranded break in the gene that is then repaired using non-homologous end joining (NHEJ). NHEJ may add or delete at least one base pair during repair which may restore the proper reading frame and eliminate the premature stop codon. Correcting or restoring a mutant gene may also include disrupting an aberrant splice acceptor site or splice donor sequence. Correcting or restoring a mutant gene may also include deleting a non-essential gene segment by the simultaneous action of two nucleases on the same DNA strand in order to restore the proper reading frame by removing the DNA between the two nuclease target sites and repairing the DNA break by NHEJ.

[0039] “Donor DNA”, “donor template,” and “repair template” as used interchangeably herein refers to a double-stranded DNA fragment or molecule that includes at least a portion of the gene of interest. The donor DNA may encode a full-functional protein or a partially functional protein.

[0040] “Duchenne Muscular Dystrophy” or “DMD” as used interchangeably herein refers to a recessive, fatal, X-linked disorder that results in muscle degeneration and eventual death. DMD is a common hereditary monogenic disease and occurs in 1 in 3500 males. DMD is the result of inherited or spontaneous mutations that cause nonsense or frame shift mutations in the dystrophin gene. The majority of dystrophin mutations that cause DMD are deletions of exons that disrupt the reading frame and cause premature translation termination in the dystrophin gene. DMD patients typically lose the ability to physically support themselves during childhood, become progressively weaker during the teenage years, and die in their twenties.

[0041] “Dystrophin” as used herein refers to a rod-shaped cytoplasmic protein which is a part of a protein complex that connects the cytoskeleton of a muscle fiber to the surrounding extracellular matrix through the cell membrane. Dystrophin provides structural stability to the dystroglycan complex of the cell membrane that is responsible for regulating muscle cell integrity and function. The dystrophin gene or “DMD gene” as used interchangeably herein is 2.2 megabases at locus Xp21. The primary transcription mea-

sures about 2,400 kb with the mature mRNA being about 14 kb. 79 exons code for the protein which is over 3500 amino acids.

[0042] “Enhancer” as used herein refers to non-coding DNA sequences containing multiple activator and repressor binding sites. Enhancers range from 200 bp to 1 kb in length and may be either proximal, 5' upstream to the promoter or within the first intron of the regulated gene, or distal, in introns of neighboring genes or intergenic regions far away from the locus. Through DNA looping, active enhancers contact the promoter independently of the core DNA binding motif promoter specificity. 4 to 5 enhancers may interact with a promoter. Similarly, enhancers may regulate more than one gene without linkage restriction and may “skip” neighboring genes to regulate more distant ones. Transcriptional regulation may involve elements located in a chromosome different to one where the promoter resides. Proximal enhancers or promoters of neighboring genes may serve as platforms to recruit more distal elements.

[0043] “Frameshift” or “frameshift mutation” as used interchangeably herein refers to a type of gene mutation wherein the addition or deletion of one or more nucleotides causes a shift in the reading frame of the codons in the mRNA. The shift in reading frame may lead to the alteration in the amino acid sequence at protein translation, such as a missense mutation or a premature stop codon.

[0044] “Functional” and “full-functional” as used herein describes protein that has biological activity. A “functional gene” refers to a gene transcribed to mRNA, which is translated to a functional protein.

[0045] “Fusion protein” as used herein refers to a chimeric protein created through the joining of two or more genes that originally coded for separate proteins. The translation of the fusion gene results in a single polypeptide with functional properties derived from each of the original proteins.

[0046] “Genetic construct” as used herein refers to the DNA or RNA molecules that comprise a polynucleotide that encodes a protein. The coding sequence includes initiation and termination signals operably linked to regulatory elements including a promoter and polyadenylation signal capable of directing expression in the cells of the individual to whom the nucleic acid molecule is administered. As used herein, the term “expressible form” refers to gene constructs that contain the necessary regulatory elements operable linked to a coding sequence that encodes a protein such that when present in the cell of the individual, the coding sequence will be expressed. The regulatory elements may include, for example, a promoter, an enhancer, an initiation codon, a stop codon, or a polyadenylation signal.

[0047] “Genome editing” or “gene editing” as used herein refers to changing the DNA sequence of a gene. Genome editing may include correcting or restoring a mutant gene or adding additional mutations. Genome editing may include knocking out a gene, such as a mutant gene or a normal gene. Genome editing may be used to treat disease or, for example, enhance muscle repair, by changing the gene of interest. In some embodiments, the compositions and methods detailed herein are for use in somatic cells and not germ line cells.

[0048] The term “heterologous” as used herein refers to nucleic acid comprising two or more subsequences that are not found in the same relationship to each other in nature. For instance, a nucleic acid that is recombinantly produced typically has two or more sequences from unrelated genes synthetically arranged to make a new functional nucleic

acid, for example, a promoter from one source and a coding region from another source. The two nucleic acids are thus heterologous to each other in this context. When added to a cell, the recombinant nucleic acids would also be heterologous to the endogenous genes of the cell. Thus, in a chromosome, a heterologous nucleic acid would include a non-native (non-naturally occurring) nucleic acid that has integrated into the chromosome, or a non-native (non-naturally occurring) extrachromosomal nucleic acid. Similarly, a heterologous protein indicates that the protein comprises two or more subsequences that are not found in the same relationship to each other in nature (for example, a "fusion protein," where the two subsequences are encoded by a single nucleic acid sequence).

[0049] "Homology-directed repair" or "HDR" as used interchangeably herein refers to a mechanism in cells to repair double strand DNA lesions when a homologous piece of DNA is present in the nucleus, mostly in G2 and S phase of the cell cycle. HDR uses a donor DNA template to guide repair and may be used to create specific sequence changes to the genome, including the targeted addition of whole genes. If a donor template is provided along with the CRISPR/Cas9-based gene editing system, then the cellular machinery will repair the break by homologous recombination, which is enhanced several orders of magnitude in the presence of DNA cleavage. When the homologous DNA piece is absent, non-homologous end joining may take place instead.

[0050] "Identical" or "identity" as a percentage as used herein in the context of two or more polynucleotide or polypeptide sequences means that the sequences have a specified percentage of residues that are the same over a specified region. The percentage may be calculated by optimally aligning the two sequences, comparing the two sequences over the specified region, determining the number of positions at which the identical residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the specified region, and multiplying the result by 100 to yield the percentage of sequence identity. In cases where the two sequences are of different lengths or the alignment produces one or more staggered ends and the specified region of comparison includes only a single sequence, the residues of single sequence are included in the denominator but not the numerator of the calculation. When comparing DNA and RNA, thymine (T) and uracil (U) may be considered equivalent. Identity may be performed manually or by using a computer sequence algorithm such as BLAST or BLAST 2.0.

[0051] "Mutant gene" or "mutated gene" as used interchangeably herein refers to a gene that has undergone a detectable mutation. A mutant gene has undergone a change, such as the loss, gain, or exchange of genetic material, which affects the normal transmission and expression of the gene. A "disrupted gene" as used herein refers to a mutant gene that has a mutation that causes a premature stop codon. The disrupted gene product is truncated relative to a full-length undisrupted gene product.

[0052] "Non-homologous end joining (NHEJ) pathway" as used herein refers to a pathway that repairs double-strand breaks in DNA by directly ligating the break ends without the need for a homologous template. The template-independent re-ligation of DNA ends by NHEJ is a stochastic, error-prone repair process that introduces random micro-

insertions and micro-deletions (indels) at the DNA break-point. This method may be used to intentionally disrupt, delete, or alter the reading frame of targeted gene sequences. NHEJ typically uses short homologous DNA sequences called microhomologies to guide repair. These microhomologies are often present in single-stranded overhangs on the end of double-strand breaks. When the overhangs are perfectly compatible, NHEJ usually repairs the break accurately, yet imprecise repair leading to loss of nucleotides may also occur, but is much more common when the overhangs are not compatible. "Nuclease mediated NHEJ" as used herein refers to NHEJ that is initiated after a nuclease cuts double stranded DNA.

[0053] "Normal gene" as used herein refers to a gene that has not undergone a change, such as a loss, gain, or exchange of genetic material. The normal gene undergoes normal gene transmission and gene expression. For example, a normal gene may be a wild-type gene.

[0054] "Nucleic acid" or "oligonucleotide" or "polynucleotide" as used herein means at least two nucleotides covalently linked together. The depiction of a single strand also defines the sequence of the complementary strand. Thus, a polynucleotide also encompasses the complementary strand of a depicted single strand. Many variants of a polynucleotide may be used for the same purpose as a given polynucleotide. Thus, a polynucleotide also encompasses substantially identical polynucleotides and complements thereof. A single strand provides a probe that may hybridize to a target sequence under stringent hybridization conditions. Thus, a polynucleotide also encompasses a probe that hybridizes under stringent hybridization conditions. Polynucleotides may be single stranded or double stranded or may contain portions of both double stranded and single stranded sequence. The polynucleotide can be nucleic acid, natural or synthetic, DNA, genomic DNA, cDNA, RNA, mRNA, or a hybrid, where the polynucleotide can contain combinations of deoxyribo- and ribo-nucleotides, and combinations of bases including, for example, uracil, adenine, thymine, cytosine, guanine, inosine, xanthine hypoxanthine, isocytosine, and isoguanine. Polynucleotides can be obtained by chemical synthesis methods or by recombinant methods.

[0055] "Open reading frame" refers to a stretch of codons that begins with a start codon and ends at a stop codon. In eukaryotic genes with multiple exons, introns are removed, and exons are then joined together after transcription to yield the final mRNA for protein translation. An open reading frame may be a continuous stretch of codons. In some embodiments, the open reading frame only applies to spliced mRNAs, not genomic DNA, for expression of a protein.

[0056] "Operably linked" as used herein means that expression of a gene is under the control of a promoter with which it is spatially connected. A promoter may be positioned 5' (upstream) or 3' (downstream) of a gene under its control. The distance between the promoter and a gene may be approximately the same as the distance between that promoter and the gene it controls in the gene from which the promoter is derived. As is known in the art, variation in this distance may be accommodated without loss of promoter function. Nucleic acid or amino acid sequences are "operably linked" (or "operatively linked") when placed into a functional relationship with one another. For instance, a promoter or enhancer is operably linked to a coding sequence if it regulates, or contributes to the modulation of,

the transcription of the coding sequence. Operably linked DNA sequences are typically contiguous, and operably linked amino acid sequences are typically contiguous and in the same reading frame. However, since enhancers generally function when separated from the promoter by up to several kilobases or more and intronic sequences may be of variable lengths, some polynucleotide elements may be operably linked but not contiguous. Similarly, certain amino acid sequences that are non-contiguous in a primary polypeptide sequence may nonetheless be operably linked due to, for example folding of a polypeptide chain. With respect to fusion polypeptides, the terms "operatively linked" and "operably linked" can refer to the fact that each of the components performs the same function in linkage to the other component as it would if it were not so linked.

[0057] "Partially-functional" as used herein describes a protein that is encoded by a mutant gene and has less biological activity than a functional protein but more than a non-functional protein.

[0058] A "peptide" or "polypeptide" is a linked sequence of two or more amino acids linked by peptide bonds. The polypeptide can be natural, synthetic, or a modification or combination of natural and synthetic. Peptides and polypeptides include proteins such as binding proteins, receptors, and antibodies. The terms "polypeptide", "protein," and "peptide" are used interchangeably herein. "Primary structure" refers to the amino acid sequence of a particular peptide. "Secondary structure" refers to locally ordered, three dimensional structures within a polypeptide. These structures are commonly known as domains, for example, enzymatic domains, extracellular domains, transmembrane domains, pore domains, and cytoplasmic tail domains. "Domains" are portions of a polypeptide that form a compact unit of the polypeptide and are typically 15 to 350 amino acids long. Exemplary domains include domains with enzymatic activity or ligand binding activity. Typical domains are made up of sections of lesser organization such as stretches of beta-sheet and alpha-helices. "Tertiary structure" refers to the complete three-dimensional structure of a polypeptide monomer. "Quaternary structure" refers to the three-dimensional structure formed by the noncovalent association of independent tertiary units. A "motif" is a portion of a polypeptide sequence and includes at least two amino acids. A motif may be 2 to 20, 2 to 15, or 2 to 10 amino acids in length. In some embodiments, a motif includes 3, 4, 5, 6, or 7 sequential amino acids. A domain may be comprised of a series of the same type of motif.

[0059] "Premature stop codon" or "out-of-frame stop codon" as used interchangeably herein refers to nonsense mutation in a sequence of DNA, which results in a stop codon at location not normally found in the wild-type gene. A premature stop codon may cause a protein to be truncated or shorter compared to the full-length version of the protein.

[0060] "Promoter" as used herein means a synthetic or naturally derived molecule which is capable of conferring, activating or enhancing expression of a nucleic acid in a cell. A promoter may comprise one or more specific transcriptional regulatory sequences to further enhance expression and/or to alter the spatial expression and/or temporal expression of same. A promoter may also comprise distal enhancer or repressor elements, which may be located as much as several thousand base pairs from the start site of transcription. A promoter may be derived from sources including viral, bacterial, fungal, plants, insects, and animals. A pro-

moter may regulate the expression of a gene component constitutively, or differentially with respect to cell, the tissue or organ in which expression occurs or, with respect to the developmental stage at which expression occurs, or in response to external stimuli such as physiological stresses, pathogens, metal ions, or inducing agents. Representative examples of promoters include the bacteriophage T7 promoter, bacteriophage T3 promoter, SP6 promoter, lac operator-promoter, tac promoter, SV40 late promoter, SV40 early promoter, RSV-LTR promoter, CMV IE promoter, SV40 early promoter or SV40 late promoter, human U6 (hU6) promoter, and CMV IE promoter. Promoters that target muscle-specific stem cells may include the CK8 promoter, the Spc5-12 promoter, and the MHCK7 promoter.

[0061] The term "recombinant" when used with reference to, for example, a cell, nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein, or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (naturally occurring) form of the cell or express a second copy of a native gene that is otherwise normally or abnormally expressed, under expressed, or not expressed at all.

[0062] "Sample" or "test sample" as used herein can mean any sample in which the presence and/or level of a target is to be detected or determined or any sample comprising a DNA targeting or gene editing system or component thereof as detailed herein. Samples may include liquids, solutions, emulsions, or suspensions. Samples may include a medical sample. Samples may include any biological fluid or tissue, such as blood, whole blood, fractions of blood such as plasma and serum, muscle, interstitial fluid, sweat, saliva, urine, tears, synovial fluid, bone marrow, cerebrospinal fluid, nasal secretions, sputum, amniotic fluid, bronchoalveolar lavage fluid, gastric lavage, emesis, fecal matter, lung tissue, peripheral blood mononuclear cells, total white blood cells, lymph node cells, spleen cells, tonsil cells, cancer cells, tumor cells, bile, digestive fluid, skin, or combinations thereof. In some embodiments, the sample comprises an aliquot. In other embodiments, the sample comprises a biological fluid. Samples can be obtained by any means known in the art. The sample can be used directly as obtained from a patient or can be pre-treated, such as by filtration, distillation, extraction, concentration, centrifugation, inactivation of interfering components, addition of reagents, and the like, to modify the character of the sample in some manner as discussed herein or otherwise as is known in the art.

[0063] "Subject" and "patient" as used herein interchangeably refers to any vertebrate, including, but not limited to, a mammal that wants or is in need of the herein described compositions or methods. The subject may be a human or a non-human. The subject may be a vertebrate. The subject may be a mammal. The mammal may be a primate or a non-primate. The mammal can be a non-primate such as, for example, cow, pig, camel, llama, hedgehog, anteater, platypus, elephant, alpaca, horse, goat, rabbit, sheep, hamster, guinea pig, cat, dog, rat, and mouse. The mammal can be a primate such as a human. The mammal can be a non-human primate such as, for example, monkey, cynomolgous monkey, rhesus monkey, chimpanzee, gorilla, orangutan, and gibbon. The subject may be of any age or stage of devel-

opment, such as, for example, an adult, an adolescent, a child, such as age 0-2, 2-4, 2-6, or 6-12 years, or an infant, such as age 0-1 years. The subject may be male. The subject may be female. In some embodiments, the subject has a specific genetic marker. The subject may be undergoing other forms of treatment.

[0064] “Substantially identical” can mean that a first and second amino acid or polynucleotide sequence are at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical over a region of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100 amino acids or nucleotides, respectively.

[0065] “Target gene” as used herein refers to any nucleotide sequence encoding a known or putative gene product. The target gene may be a mutated gene involved in a genetic disease. The target gene may encode a known or putative gene product that is intended to be corrected or for which its expression is intended to be modulated.

[0066] “Target region” as used herein refers to the region of the target gene to which the CRISPR/Cas9-based gene editing or targeting system is designed to bind.

[0067] “Transgene” as used herein refers to a gene or genetic material containing a gene sequence that has been isolated from one organism and is introduced into a different organism. This non-native segment of DNA may retain the ability to produce RNA or protein in the transgenic organism, or it may alter the normal function of the transgenic organism’s genetic code. The introduction of a transgene has the potential to change the phenotype of an organism.

[0068] “Transcriptional regulatory elements” or “regulatory elements” refers to a genetic element which can control the expression of nucleic acid sequences, such as activate, enhancer, or decrease expression, or alter the spatial and/or temporal expression of a nucleic acid sequence. Examples of regulatory elements include, for example, promoters, enhancers, splicing signals, polyadenylation signals, and termination signals. A regulatory element can be “endogenous,” “exogenous,” or “heterologous” with respect to the gene to which it is operably linked. An “endogenous” regulatory element is one which is naturally linked with a given gene in the genome. An “exogenous” or “heterologous” regulatory element is one which is not normally linked with a given gene but is placed in operable linkage with a gene by genetic manipulation.

[0069] “Treatment” or “treating” or “therapy” when referring to protection of a subject from a disease, means suppressing, repressing, reversing, alleviating, ameliorating, or inhibiting the progress of disease, or completely eliminating a disease. A treatment may be either performed in an acute or chronic way. The term also refers to reducing the severity of a disease or symptoms associated with such disease prior to affliction with the disease. Treatment may result in a reduction in the incidence, frequency, severity, and/or duration of symptoms of the disease. Preventing the disease involves administering a composition of the present invention to a subject prior to onset of the disease. Suppressing the disease involves administering a composition of the present invention to a subject after induction of the disease but before its clinical appearance. Repressing or ameliorating the disease involves administering a composition of the present invention to a subject after clinical appearance of the disease.

[0070] As used herein, the term “gene therapy” refers to a method of treating a patient wherein polypeptides or nucleic acid sequences are transferred into cells of a patient such that activity and/or the expression of a particular gene is modulated. In certain embodiments, the expression of the gene is suppressed. In certain embodiments, the expression of the gene is enhanced. In certain embodiments, the temporal or spatial pattern of the expression of the gene is modulated.

[0071] “Variant” used herein with respect to a polynucleotide means (i) a portion or fragment of a referenced nucleotide sequence; (ii) the complement of a referenced nucleotide sequence or portion thereof; (iii) a nucleic acid that is substantially identical to a referenced nucleic acid or the complement thereof; or (iv) a nucleic acid that hybridizes under stringent conditions to the referenced nucleic acid, complement thereof, or a sequence substantially identical thereto. A variant can be a polynucleotide sequence that is substantially identical over the full length of the full polynucleotide sequence or a fragment thereof. The polynucleotide sequence can be 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%, or less than 100% identical over the full length of the polynucleotide sequence or a fragment thereof.

[0072] “Variant” with respect to a peptide or polypeptide that differs in amino acid sequence by the insertion, deletion, or conservative substitution of amino acids, but retain at least one biological activity. Variant may also mean a protein with an amino acid sequence that is substantially identical to a referenced protein with an amino acid sequence that retains at least one biological activity. Representative examples of “biological activity” include the ability to be bound by a specific antibody or polypeptide or to promote an immune response. Variant can mean a functional fragment thereof. Variant can also mean multiple copies of a polypeptide. The multiple copies can be in tandem or separated by a linker. A conservative substitution of an amino acid, for example, replacing an amino acid with a different amino acid of similar properties (for example, hydrophilicity, degree and distribution of charged regions) is recognized in the art as typically involving a minor change. These minor changes may be identified, in part, by considering the hydrophilicity index of amino acids, as understood in the art (Kyte et al., *J. Mol. Biol.* 1982, 157, 105-132). The hydrophilicity index of an amino acid is based on a consideration of its hydrophobicity and charge. It is known in the art that amino acids of similar hydrophilicity indexes may be substituted and still retain protein function. In one aspect, amino acids having hydrophilicity indexes of ± 2 are substituted. The hydrophilicity of amino acids may also be used to reveal substitutions that would result in proteins retaining biological function. A consideration of the hydrophilicity of amino acids in the context of a peptide permits calculation of the greatest local average hydrophilicity of that peptide. Substitutions may be performed with amino acids having hydrophilicity values within ± 2 of each other. Both the hydrophobicity index and the hydrophilicity value of amino acids are influenced by the particular side chain of that amino acid. Consistent with that observation, amino acid substitutions that are compatible with biological function are understood to depend on the relative similarity of the amino acids, and particularly the side chains of those amino acids, as revealed by the hydrophobicity, hydrophilicity, charge, size, and other properties. A variant can be an amino acid sequence that is

substantially identical over the full length of the amino acid sequence or fragment thereof. The amino acid sequence can be 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%, or less than 100% identical over the full length of the amino acid sequence or a fragment thereof.

[0073] “Vector” as used herein means a nucleic acid sequence containing an origin of replication. A vector may be capable of directing the delivery or transfer of a polynucleotide sequence to target cells, where it can be replicated or expressed. A vector may contain an origin of replication, one or more regulatory elements, and/or one or more coding sequences. A vector may be a viral vector, bacteriophage, bacterial artificial chromosome, plasmid, cosmid, or yeast artificial chromosome. A vector may be a DNA or RNA vector. A vector may be a self-replicating extrachromosomal vector. Viral vectors include, but are not limited to, adenovirus vector, adeno-associated virus (AAV) vector, retrovirus vector, or lentivirus vector. A vector may be an adeno-associated virus (AAV) vector. The vector may encode, for example, a Cas9 protein or fusion protein and at least one gRNA molecule.

[0074] Unless otherwise defined herein, scientific and technical terms used in connection with the present disclosure shall have the meanings that are commonly understood by those of ordinary skill in the art. For example, any nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics, and protein and nucleic acid chemistry and hybridization described herein are those that are well known and commonly used in the art. The meaning and scope of the terms should be clear; in the event however of any latent ambiguity, definitions provided herein take precedent over any dictionary or extrinsic definition. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular.

2. DNA TARGETING SYSTEMS

[0075] Provided herein are DNA Targeting Systems that may be used, for example, to modulate gene expression. A “DNA Targeting System” as used herein is a system capable of specifically targeting a particular region of DNA and modulating gene expression by binding to that region. Non-limiting examples of these systems are CRISPR-Cas-based systems, zinc finger (ZF)-based systems, and/or transcription activator-like effector (TALE)-based systems. The DNA Targeting System may be a nuclease system that acts through mutating or editing the target region (such as by insertion, deletion or substitution) or it may be a system that delivers a functional second polypeptide domain, such as an activator or repressor, to the target region.

[0076] Each of these systems comprises a DNA-binding portion or domain, such as a guide RNA, a ZF, or a TALE, that specifically recognizes and binds to a particular target region of a target DNA. The DNA-binding portion (for example, Cas protein, ZF, or TALE) can be linked to a second protein domain, such as a polypeptide with transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, nuclease activity, nucleic acid association activity, methylase activity, demethylase activity, acetylation activity, or deacetylation activity, to form a fusion protein. In other embodiments, the DNA-binding portion is linked with a

second protein domain using an antibody and peptide epitope, such as the Suntag recruitment system (Tanenbaum et al., *Cell* 2014, 159, 635-646, incorporated herein by reference in its entirety). Exemplary second polypeptide domains are detailed further below (see “Cas Fusion Protein”). For example, the DNA-binding portion can be linked to an activator and thus guide the activator to a specific target region of the target DNA. Similarly, the DNA-binding portion can be linked to a repressor and thus guide the repressor to a specific target region of the target DNA.

[0077] In some embodiments, the DNA-binding portion comprises a Cas protein, such as a Cas9 protein. Some CRISPR-Cas-based systems can operate to activate or repress expression using the Cas protein alone, not linked to an activator or repressor. For example, a nuclease-null Cas9 can act as a repressor on its own, or a nuclease-active Cas9 can act as an activator when paired with an inactive (dead) guide RNA. In addition, RNA or DNA that hybridizes to a particular target region of the target DNA can be directly linked (covalently or non-covalently) to an activator or repressor. Some CRISPR-Cas-based systems can operate to activate or repress expression using the Cas protein linked to a second protein domain, such as, for example, an activator or repressor. Some embodiments include a Cas protein linked to a second polypeptide domain such as an effector (see “Cas Fusion Protein”).

[0078] In other embodiments, a first polypeptide comprising a DNA-binding portion further comprises at least one peptide epitope, and a second polypeptide comprises an activator or repressor and an antibody to the peptide epitope. For example, some embodiments include a first polypeptide comprising a Cas protein and at least one peptide epitope, and a second polypeptide comprising the effector domain and an antibody to the peptide epitope (see “Cas Effector”).

3. CRISPR/CAS-BASED GENE EDITING SYSTEM

[0079] Provided herein are CRISPR/Cas9-based gene editing systems. The CRISPR/Cas-based gene editing system may be used to modulate expression of a gene and/or treat a disease. The CRISPR/Cas-based gene editing system may include a Cas protein or a fusion protein, and at least one gRNA, and may also be referred to as a “CRISPR-Cas system.” Other embodiments include a first polypeptide comprising a Cas protein and at least one peptide epitope, at least one gRNA, and a second polypeptide comprising the effector domain and an antibody to the peptide epitope.

[0080] “Clustered Regularly Interspaced Short Palindromic Repeats” and “CRISPRs”, as used interchangeably herein, refers to loci containing multiple short direct repeats that are found in the genomes of approximately 40% of sequenced bacteria and 90% of sequenced archaea. The CRISPR system is a microbial nuclease system involved in defense against invading phages and plasmids that provides a form of acquired immunity. The CRISPR loci in microbial hosts contain a combination of CRISPR-associated (Cas) genes as well as non-coding RNA elements capable of programming the specificity of the CRISPR-mediated nucleic acid cleavage. Short segments of foreign DNA, called spacers, are incorporated into the genome between CRISPR repeats, and serve as a “memory” of past exposures. Cas proteins include, for example, Cas12a, Cas9, and Cascade proteins. Cas12a may also be referred to as “Cpf1.” Cas12a causes a staggered cut in double stranded DNA,

while Cas9 produces a blunt cut. In some embodiments, the Cas protein comprises Cas12a. In some embodiments, the Cas protein comprises Cas9. Cas9 forms a complex with the 3' end of the sgRNA (which may be referred interchangeably herein as "gRNA"), and the protein-RNA pair recognizes its genomic target by complementary base pairing between the 5' end of the gRNA sequence and a predefined 20 bp DNA sequence, known as the protospacer. This complex is directed to homologous loci of pathogen DNA via regions encoded within the crRNA, i.e., the protospacers, and protospacer-adjacent motifs (PAMs) within the pathogen genome. The non-coding CRISPR array is transcribed and cleaved within direct repeats into short crRNAs containing individual spacer sequences, which direct Cas nucleases to the target site (protospacer). By simply exchanging the 20 bp recognition sequence of the expressed gRNA, the Cas9 nuclease can be directed to new genomic targets. CRISPR spacers are used to recognize and silence exogenous genetic elements in a manner analogous to RNAi in eukaryotic organisms.

[0081] Three classes of CRISPR systems (Types I, II, and III effector systems) are known. The Type II effector system carries out targeted DNA double-strand break in four sequential steps, using a single effector enzyme, Cas9, to cleave dsDNA. Compared to the Type I and Type III effector systems, which require multiple distinct effectors acting as a complex, the Type II effector system may function in alternative contexts such as eukaryotic cells. The Type II effector system consists of a long pre-crRNA, which is transcribed from the spacer-containing CRISPR locus, the Cas9 protein, and a tracrRNA, which is involved in pre-crRNA processing. The tracrRNAs hybridize to the repeat regions separating the spacers of the pre-crRNA, thus initiating dsRNA cleavage by endogenous RNase III. This cleavage is followed by a second cleavage event within each spacer by Cas9, producing mature crRNAs that remain associated with the tracrRNA and Cas9, forming a Cas9:crRNA-tracrRNA complex. Cas12a systems include crRNA for successful targeting, whereas Cas9 systems include both crRNA and tracrRNA.

[0082] The Cas9:crRNA-tracrRNA complex unwinds the DNA duplex and searches for sequences matching the crRNA to cleave. Target recognition occurs upon detection of complementarity between a "protospacer" sequence in the target DNA and the remaining spacer sequence in the crRNA. Cas9 mediates cleavage of target DNA if a correct protospacer-adjacent motif (PAM) is also present at the 3' end of the protospacer. For protospacer targeting, the sequence must be immediately followed by the protospacer-adjacent motif (PAM), a short sequence recognized by the Cas9 nuclease that is required for DNA cleavage. Different Cas and Cas Type II systems have differing PAM requirements. For example, Cas12a may function with PAM sequences rich in thymine "T."

[0083] An engineered form of the Type II effector system of *S. pyogenes* was shown to function in human cells for genome engineering. In this system, the Cas9 protein was directed to genomic target sites by a synthetically reconstituted "guide RNA" ("gRNA", also used interchangeably herein as a chimeric single guide RNA ("sgRNA")), which is a crRNA-tracrRNA fusion that obviates the need for RNase III and crRNA processing in general. Provided herein are CRISPR/Cas9-based engineered systems for use in gene editing and treating genetic diseases. The CRISPR/Cas9-

based engineered systems can be designed to target any gene, including genes involved in, for example, a genetic disease, aging, tissue regeneration, or wound healing. The CRISPR/Cas9-based gene editing system can include a Cas9 protein or a Cas9 fusion protein.

[0084] In some embodiments, the Cas protein and/or the Cas fusion protein and/or Cas effector and/or gRNAs and/or Effector domains detailed herein may be used in compositions and methods for modulating expression of a gene. The Cas protein and/or the Cas fusion protein and/or Cas effector and/or Effector domains detailed herein may be targeted to the gene. The Cas protein and/or the Cas fusion protein and/or Cas effector and/or Effector domains detailed herein may be targeted to a regulatory element of the gene. Modulating may include, for example, increasing or enhancing expression of the gene, or reducing or inhibiting expression of the gene. The expression of the gene may be modulated by at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 1.5-fold, 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, or 10-fold, relative to a control. The expression of the gene may be modulated by less than about 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 1.5-fold, 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, or 10-fold, relative to a control. The expression of the gene may be modulated by about 5-95%, 10-90%, 15-85%, 20-80%, or 1.5-fold to 10-fold, relative to a control. The expression of the gene may be reduced by at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 1.5-fold, 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, or 10-fold, relative to a control. The expression of the gene may be reduced by less than about 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 1.5-fold, 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, or 10-fold, relative to a control. The expression of the gene may be reduced by about 5-95%, 10-90%, 15-85%, 20-80%, or 1.5-fold to 10-fold, relative to a control. The expression of the gene may be increased by at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 1.5-fold, 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, or 10-fold, relative to a control. The expression of the gene may be increased by less than about 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 1.5-fold, 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, or 10-fold, relative to a control. The expression of the gene may be increased by about 5-95%, 10-90%, 15-85%, 20-80%, or 1.5-fold to 10-fold, relative to a control.

a. Cas9 Protein

[0085] Cas9 protein is an endonuclease that cleaves nucleic acid and is encoded by the CRISPR loci and is involved in the Type II CRISPR system. The Cas9 protein can be from any bacterial or archaea species, including, but not limited to, *Streptococcus pyogenes*, *Staphylococcus aureus* (*S. aureus*), *Acidovorax avenae*, *Actinobacillus pleuropneumoniae*, *Actinobacillus succinogenes*, *Actinobacillus suis*, *Actinomyces* sp., *Cycliphilus denitrificans*, *Aminomonas paucivorans*, *Bacillus cereus*, *Bacillus smithii*, *Bacillus thuringiensis*, *Bacteroides* sp., *Blastopirellula marina*, *Bradyrhizobium* sp., *Brevibacillus laterosporus*, *Campylobacter coli*, *Campylobacter jejuni*, *Campylobacter lari*, *Candidatus*

puniceispirillum, *Clostridium cellulolyticum*, *Clostridium perfringens*, *Corynebacterium accolens*, *Corynebacterium diphtheriae*, *Corynebacterium matruchotii*, *Dinoroseobacter shibae*, *Eubacterium dolichum*, gamma proteobacterium, *Gluconacetobacter diazotrophicus*, *Haemophilus parainfluenzae*, *Haemophilus sputorum*, *Helicobacter canadensis*, *Helicobacter cinaedi*, *Helicobacter mustelae*, *Ilyobacter polytropis*, *Kingella kingae*, *Lactobacillus crispatus*, *Listeria ivanovii*, *Listeria monocytogenes*, *Listeriaceae bacterium*, *Methylocystis* sp., *Methylosinus trichosporium*, *Mobiluncus mulieris*, *Neisseria bacilliformis*, *Neisseria cinerea*, *Neisseria flavescens*, *Neisseria lactamica*, *Neisseria* sp., *Neisseria wadsworthii*, *Nitrosomonas* sp., *Parvibaculum lavamentivorans*, *Pasteurella multocida*, *Phascolarctobacterium succinatutens*, *Ralstonia syzygii*, *Rhodopseudomonas palustris*, *Rhodovulum* sp., *Simonsiella muelleri*, *Sphingomonas* sp., *Sporolactobacillus vineae*, *Staphylococcus lugdunensis*, *Streptococcus* sp., *Subdoligranulum* sp., *Tistrella mobilis*, *Treponema* sp., or *Verminephrobacter eiseniae*. In certain embodiments, the Cas9 molecule is a *Streptococcus pyogenes* Cas9 molecule (also referred herein as “SpCas9”). SpCas9 may comprise an amino acid sequence of SEQ ID NO: 26. In certain embodiments, the Cas9 molecule is a *Staphylococcus aureus* Cas9 molecule (also referred herein as “SaCas9”). SaCas9 may comprise an amino acid sequence of SEQ ID NO: 27.

[0086] A Cas9 molecule or a Cas9 fusion protein can interact with one or more gRNA molecule(s) and, in concert with the gRNA molecule(s), can localize to a site which comprises a target domain, and in certain embodiments, a PAM sequence. The Cas9 protein forms a complex with the 3' end of a gRNA. The ability of a Cas9 molecule or a Cas9 fusion protein to recognize a PAM sequence can be determined, for example, by using a transformation assay as known in the art.

[0087] The specificity of the CRISPR-based system may depend on two factors: the target sequence and the protospacer-adjacent motif (PAM). The target sequence is located on the 5' end of the gRNA and is designed to bond with base pairs on the host DNA at the correct DNA sequence known as the protospacer. By simply exchanging the recognition sequence of the gRNA, the Cas9 protein can be directed to new genomic targets. The PAM sequence is located on the DNA to be altered and is recognized by a Cas9 protein. PAM recognition sequences of the Cas9 protein can be species specific.

[0088] In certain embodiments, the ability of a Cas9 molecule or a Cas9 fusion protein to interact with and cleave a target nucleic acid is PAM sequence dependent. A PAM sequence is a sequence in the target nucleic acid. In certain embodiments, cleavage of the target nucleic acid occurs upstream from the PAM sequence. Cas9 molecules from different bacterial species can recognize different sequence motifs (for example, PAM sequences). A Cas9 molecule of *S. pyogenes* may recognize the PAM sequence of NRG (5'-NRG-3', where R is any nucleotide residue, and in some embodiments, R is either A or G, SEQ ID NO: 1). In certain embodiments, a Cas9 molecule of *S. pyogenes* may naturally prefer and recognize the sequence motif NGG (SEQ ID NO: 2) and directs cleavage of a target nucleic acid sequence 1 to 10, for example, 3 to 5, bp upstream from that sequence. In some embodiments, a Cas9 molecule of *S. pyogenes* accepts other PAM sequences, such as NAG (SEQ ID NO: 3) in engineered systems (Hsu et al., *Nature Biotechnology*

2013 doi:10.1038/nbt.2647). In certain embodiments, a Cas9 molecule of *S. thermophilus* recognizes the sequence motif NGNG (SEQ ID NO: 4) and/or NNAGAAW (W=A or T) (SEQ ID NO: 5) and directs cleavage of a target nucleic acid sequence 1 to 10, for example, 3 to 5, bp upstream from these sequences. In certain embodiments, a Cas9 molecule of *S. mutans* recognizes the sequence motif NGG (SEQ ID NO: 2) and/or NAAR (R=A or G) (SEQ ID NO: 6) and directs cleavage of a target nucleic acid sequence 1 to 10, for example, 3 to 5 bp, upstream from this sequence. In certain embodiments, a Cas9 molecule of *S. aureus* recognizes the sequence motif NNGRR (R=A or G) (SEQ ID NO: 7) and directs cleavage of a target nucleic acid sequence 1 to 10, for example, 3 to 5, bp upstream from that sequence. In certain embodiments, a Cas9 molecule of *S. aureus* recognizes the sequence motif NNGRRN (R=A or G) (SEQ ID NO: 8) and directs cleavage of a target nucleic acid sequence 1 to 10, for example, 3 to 5, bp upstream from that sequence. In certain embodiments, a Cas9 molecule of *S. aureus* recognizes the sequence motif NNGRRRN (R=A or G) (SEQ ID NO: 9) and directs cleavage of a target nucleic acid sequence 1 to 10, for example, 3 to 5, bp upstream from that sequence. In certain embodiments, a Cas9 molecule of *S. aureus* recognizes the sequence motif NNGRRRT (R=A or G) (SEQ ID NO: 10) and directs cleavage of a target nucleic acid sequence 1 to 10, for example, 3 to 5, bp upstream from that sequence. A Cas9 molecule derived from *Neisseria meningitidis* (NmCas9) normally has a native PAM of NNNNGATT (SEQ ID NO: 11), but may have activity across a variety of PAMs, including a highly degenerate NNNNGNNN PAM (SEQ ID NO: 12) (Esveld et al. *Nature Methods* 2013 doi:10.1038/nmeth.2681). In the aforementioned embodiments, N can be any nucleotide residue, for example, any of A, G, C, or T. Cas9 molecules can be engineered to alter the PAM specificity of the Cas9 molecule.

[0089] In some embodiments, the Cas9 protein recognizes a PAM sequence NGG (SEQ ID NO: 2) or NGA (SEQ ID NO: 13) or NNNRRT (R=A or G) (SEQ ID NO: 14) or ATTCCT (SEQ ID NO: 15) or NGAN (SEQ ID NO: 16) or NGNG (SEQ ID NO: 17). In some embodiments, the Cas9 protein is a Cas9 protein of *S. aureus* and recognizes the sequence motif NNGRR (R=A or G) (SEQ ID NO: 7), NNGRRN (R=A or G) (SEQ ID NO: 8), NNGRRRT (R=A or G) (SEQ ID NO: 9), or NNGRRV (R=A or G; V=A or C or G) (SEQ ID NO: 10). In the aforementioned embodiments, N can be any nucleotide residue, for example, any of A, G, C, or T.

[0090] Additionally or alternatively, a nucleic acid encoding a Cas9 molecule or Cas9 polypeptide may comprise a nuclear localization sequence (NLS). Nuclear localization sequences are known in the art, for example, SV40 NLS (Pro-Lys-Lys-Lys-Arg-Lys-Val; SEQ ID NO: 20).

[0091] In some embodiments, the at least one Cas9 molecule is a mutant Cas9 molecule. The Cas9 protein can be mutated so that the nuclease activity is inactivated. An inactivated Cas9 protein (“iCas9”, also referred to as “dCas9”) with no endonuclease activity has been targeted to genes in bacteria, yeast, and human cells by gRNAs to silence gene expression through steric hindrance. Exemplary mutations with reference to the *S. pyogenes* Cas9 sequence to inactivate the nuclease activity include: D10A, E762A, H840A, N854A, N863A and/or D986A. A *S. pyogenes* Cas9 protein with the D10A mutation may comprise an amino

acid sequence of SEQ ID NO: 28. A *S. pyogenes* Cas9 protein with D10A and H840A mutations may comprise an amino acid sequence of SEQ ID NO: 29. Exemplary mutations with reference to the *S. aureus* Cas9 sequence to inactivate the nuclease activity include D10A and N580A. In certain embodiments, the mutant *S. aureus* Cas9 molecule comprises a D10A mutation. The nucleotide sequence encoding this mutant *S. aureus* Cas9 is set forth in SEQ ID NO: 30. In certain embodiments, the mutant *S. aureus* Cas9 molecule comprises a N580A mutation. The nucleotide sequence encoding this mutant *S. aureus* Cas9 molecule is set forth in SEQ ID NO: 31.

[0092] In some embodiments, the Cas9 protein is a VQR variant. The VQR variant of Cas9 is a mutant with a different PAM recognition, as detailed in Kleinstiver, et al. (*Nature* 2015, 523, 481-485, incorporated herein by reference).

[0093] A polynucleotide encoding a Cas9 molecule can be a synthetic polynucleotide. For example, the synthetic polynucleotide can be chemically modified. The synthetic polynucleotide can be codon optimized, for example, at least one non-common codon or less-common codon has been replaced by a common codon. For example, the synthetic polynucleotide can direct the synthesis of an optimized messenger mRNA, for example, optimized for expression in a mammalian expression system, as described herein. An exemplary codon optimized nucleic acid sequence encoding a Cas9 molecule of *S. pyogenes* is set forth in SEQ ID NO: 32. Exemplary codon optimized nucleic acid sequences encoding a Cas9 molecule of *S. aureus*, and optionally containing nuclear localization sequences (NLSSs), are set forth in SEQ ID NOs: 33-39. Another exemplary codon optimized nucleic acid sequence encoding a Cas9 molecule of *S. aureus* comprises the nucleotides 1293-4451 of SEQ ID NO: 40.

b. Cas Fusion Protein

[0094] Alternatively or additionally, the CRISPR/Cas-based gene editing system can include a fusion protein. The fusion protein can comprise two heterologous polypeptide domains. The first polypeptide domain comprises a Cas protein or a mutated Cas protein. The first polypeptide domain is fused to at least one second polypeptide domain. The second polypeptide domain may comprise or also be referred to as an effector, or effector domain. The second polypeptide domain has a different activity than what is endogenous to Cas protein. For example, the second polypeptide domain may have an activity such as transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, nuclease activity, nucleic acid association activity, histone methylase activity, DNA methylase activity, histone demethylase activity, DNA demethylase activity, acetylation activity, and/or deacetylation activity. The activity of the second polypeptide domain may be direct or indirect. The second polypeptide domain may have this activity itself (direct), or it may recruit and/or interact with a polypeptide domain that has this activity (indirect). In some embodiments, the second polypeptide domain has transcription activation activity. In some embodiments, the second polypeptide domain has transcription repression activity. In some embodiments, the second polypeptide domain comprises a synthetic transcription factor. The second polypeptide domain may be at the C-terminal end of the first polypeptide domain, or at the N-terminal end of the first polypeptide domain, or a combination thereof. The fusion protein may

include one second polypeptide domain. In some embodiments, the fusion protein comprises more than one second polypeptide domain. The fusion protein may include two of the second polypeptide domains. For example, the fusion protein may include a second polypeptide domain at the N-terminal end of the first polypeptide domain as well as a second polypeptide domain at the C-terminal end of the first polypeptide domain. In other embodiments, the fusion protein may include a single first polypeptide domain and more than one (for example, two or three) second polypeptide domains in tandem.

[0095] The linkage from the first polypeptide domain to the second polypeptide domain can be through reversible or irreversible covalent linkage or through a non-covalent linkage, as long as the linker does not interfere with the function of the second polypeptide domain. For example, a Cas polypeptide can be linked to a second polypeptide domain as part of a fusion protein. As another example, they can be linked through reversible non-covalent interactions such as avidin (or streptavidin)-biotin interaction, histidine-divalent metal ion interaction (such as, Ni, Co, Cu, Fe), interactions between multimerization (such as, dimerization) domains, or glutathione S-transferase (GST)-glutathione interaction. As yet another example, they can be linked covalently but reversibly with linkers such as dibromomaleimide (DBM) or amino-thiol conjugation.

[0096] In some embodiments, the fusion protein includes at least one linker. A linker may be included anywhere in the polypeptide sequence of the fusion protein, for example, between the first and second polypeptide domains. A linker may be of any length and design to promote or restrict the mobility of components in the fusion protein. A linker may comprise any amino acid sequence of about 2 to about 100, about 5 to about 80, about 10 to about 60, or about 20 to about 50 amino acids. A linker may comprise an amino acid sequence of at least about 2, 3, 4, 5, 10, 15, 20, 25, or 30 amino acids. A linker may comprise an amino acid sequence of less than about 100, 90, 80, 70, 60, 50, or 40 amino acids. A linker may include sequential or tandem repeats of an amino acid sequence that is 2 to 20 amino acids in length. Linkers may include, for example, a GS linker (Gly-Gly-Gly-Gly-Ser)n, wherein n is an integer between 0 and 10 (SEQ ID NO: 21). In a GS linker, n can be adjusted to optimize the linker length and achieve appropriate separation of the functional domains. Other examples of linkers may include, for example, Gly-Gly-Gly-Gly-Gly (SEQ ID NO: 22), Gly-Gly-Ala-Gly-Gly (SEQ ID NO: 23), Gly/Ser rich linkers such as Gly-Gly-Gly-Gly-Ser-Ser-Ser (SEQ ID NO: 24) or GSGSG (SEQ ID NO: 91) or GSGGGSGSGGGSGSGGGSG (SEQ ID NO: 92), or Gly/Ala rich linkers such as Gly-Gly-Gly-Ala-Ala-Ala (SEQ ID NO: 25).

c. Cas Effector

[0097] Alternatively or additionally, the CRISPR/Cas-based gene editing system can include a Cas effector. The Cas effector can include a first polypeptide comprising a Cas protein and at least one peptide epitope, and a second polypeptide comprising an effector and an antibody to the peptide epitope. Such systems are described in, for example, in Tanenbaum et al. (*Cell* 2014, 159, 635-646, incorporated herein by reference in its entirety) with reference to, for example, the Suntag recruitment system. For the Cas effector, the first polypeptide and the second polypeptide may be two separate polypeptides or chains.

[0098] The first polypeptide of the Cas effector may comprise about 2 to about 50 peptide epitopes, about 2 to about 40 peptide epitopes, about 2 to about 30 peptide epitopes, or about 3 to about 25 peptide epitopes. The first polypeptide may comprise about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 peptide epitopes. In some embodiments, the first polypeptide comprises about 5 peptide epitopes. In some embodiments, the first polypeptide comprises about 24 peptide epitopes. The first polypeptide may comprise at least one peptide epitope at the N-terminus and/or at the C-terminus of the Cas protein.

[0099] The peptide epitope may comprise any amino acid sequence that the antibody binds. The antibody may bind specifically to the peptide epitope. The peptide epitope may comprise an amino acid sequence that is not found in humans. In some embodiments, the peptide epitope comprises GCN4. GCN4 may comprise a peptide having an amino acid sequence of SEQ ID NO: 85 and may be encoded by a polynucleotide comprising SEQ ID NO: 86. The first polypeptide may comprise at least one linker N-terminal or C-terminal to the peptide epitope. The first polypeptide may comprise more than one copy of the peptide epitope and at least one linker in between adjacent copies of the peptide epitope. The linker may be, for example, selected from SEQ ID NOs: 21-24 and 91-92, as detailed above.

[0100] In some embodiments, the first polypeptide comprises dCas9-5X-GCN4 (SEQ ID NO: 87). dCas9-5X-GCN4 may be encoded by a polynucleotide comprising SEQ ID NO: 88. In some embodiments, the first polypeptide comprises dCas9-24X-GCN4 (SEQ ID NO: 89). dCas9-24X-GCN4 may be encoded by a polynucleotide comprising SEQ ID NO: 90 or a variant thereof. The first polypeptide may comprise an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 87 or 89, or any fragment thereof. The first polypeptide may comprise an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 87 or 89, or any fragment thereof. The first polypeptide may comprise the amino acid sequence of SEQ ID NO: 87 or 89.

[0101] The second polypeptide of the Cas effector may comprise an effector (also referred to as an “effector domain”) and an antibody to the peptide epitope. The antibody may be any antibody that binds the peptide epitope. The antibody may specifically bind the peptide epitope. In some embodiments, the antibody comprises ScFv. ScFv may comprise the amino acid sequence of SEQ ID NO: 81 and may be encoded by a polynucleotide comprising SEQ ID NO: 82. The second polypeptide of the Cas effector may further comprise a reporter protein such as sfBFP. The sfBFP may comprise the amino acid sequence of SEQ ID NO: 83 and may be encoded by a polynucleotide comprising SEQ ID NO: 84. The reporter protein may be at the N-terminus and/or at the C-terminus of the effector. The reporter protein may be at the N-terminus and/or at the C-terminus of the antibody. The reporter protein may be in between the effector and the antibody in the polypeptide chain.

[0102] The effector has a different activity that what is endogenous to Cas protein. For example, the effector may have an activity such as transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, nuclease activity,

nucleic acid association activity, histone methylase activity, DNA methylase activity, histone demethylase activity, DNA demethylase activity, acetylation activity, and/or deacetylation activity. The activity of the effector may be direct or indirect. The effector may have this activity itself (direct), or it may recruit and/or interact with a polypeptide domain that has this activity (indirect). The effector may be at the C-terminal end of the antibody, or at the N-terminal end of the antibody, or a combination thereof. The second polypeptide of the Cas effector may include one or more than one effector. For example, the second polypeptide of the Cas effector may include an effector at the N-terminal end of the antibody as well as an effector at the C-terminal end of the antibody. In other embodiments, the second polypeptide of the Cas effector may include a single antibody and more than one (for example, two or three) effectors in tandem.

[0103] The linkage from the effector to the antibody, or from the Cas protein to the peptide epitope, can be through reversible or irreversible covalent linkage or through a non-covalent linkage, as long as the linker does not interfere with the function of the effector or antibody. For example, an antibody can be linked to an effector as part of a fusion protein. As another example, they can be linked through reversible non-covalent interactions such as avidin (or streptavidin)-biotin interaction, histidine-divalent metal ion interaction (such as, Ni, Co, Cu, Fe), interactions between multimerization (such as, dimerization) domains, or glutathione S-transferase (GST)-glutathione interaction. As yet another example, they can be linked covalently but reversibly with linkers such as dibromomaleimide (DBM) or amino-thiol conjugation.

[0104] In some embodiments, the second polypeptide of the Cas effector includes at least one linker. A linker may be included anywhere in the polypeptide sequence, for example, between the antibody and the effector. In some embodiments, the first polypeptide of the Cas effector includes at least one linker. A linker may be included anywhere in the polypeptide sequence, for example, between the Cas protein and the peptide epitope. A linker may be of any length and design to promote or restrict the mobility of components in the protein. A linker may comprise any amino acid sequence of about 2 to about 100, about 5 to about 80, about 10 to about 60, or about 20 to about 50 amino acids. A linker may comprise an amino acid sequence of at least about 2, 3, 4, 5, 10, 15, 20, 25, or 30 amino acids. A linker may comprise an amino acid sequence of less than about 100, 90, 80, 70, 60, 50, or 40 amino acids. A linker may include sequential or tandem repeats of an amino acid sequence that is 2 to 20 amino acids in length. Linkers may comprise a sequence, for example, selected from SEQ ID NOs: 21-24 and 91-92, as detailed above.

[0105] In some embodiments, the second polypeptide comprises ScFv-sfBFP-MCRS1 (amino acid sequence comprising SEQ ID NO: 69, polynucleotide sequence comprising SEQ ID NO: 70), or ScFv-sfBFP-OTUD7B (amino acid sequence comprising SEQ ID NO: 71, polynucleotide sequence comprising SEQ ID NO: 72), or ScFv-sfBFP-LDB1 (amino acid sequence comprising SEQ ID NO: 73, polynucleotide sequence comprising SEQ ID NO: 74), or ScFv-sfBFP-NFKBIB (amino acid sequence comprising SEQ ID NO: 75, polynucleotide sequence comprising SEQ ID NO: 76), or ScFv-sfBFP-RelB (amino acid sequence comprising SEQ ID NO: 77, polynucleotide sequence comprising SEQ ID NO: 78), or ScFv-sfBFP-CITED2 (amino

acid sequence comprising SEQ ID NO: 79, polynucleotide sequence comprising SEQ ID NO: 80). The first polypeptide may comprise an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to a sequence selected from SEQ ID NOs: 69, 71, 73, 75, 77, and 79, or any fragment thereof. The first polypeptide may comprise an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to a sequence selected from SEQ ID NOs: 69, 71, 73, 75, 77, and 79, or any fragment thereof. The first polypeptide may comprise an amino acid sequence selected from SEQ ID NOs: 69, 71, 73, 75, 77, and 79.

d. Effector Domains

[0106] Further provided herein are novel effector domains. An effector (or “effector domain”) may modulate expression of gene it is targeted to. An effector may increase, enhance, decrease, or reduce the expression of a gene. The expression of the gene may be modulated by at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 1.5-fold, 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, or 10-fold, relative to a control. The expression of the gene may be modulated by less than about 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 1.5-fold to 10-fold, relative to a control. The expression of the gene may be reduced by at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 1.5-fold, 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, or 10-fold, relative to a control. The expression of the gene may be reduced by about 5-95%, 10-90%, 15-85%, 20-80%, or 1.5-fold to 10-fold, relative to a control. The expression of the gene may be increased by at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 1.5-fold, 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, or 10-fold, relative to a control. The expression of the gene may be increased by less than about 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 1.5-fold to 10-fold, relative to a control. The expression of the gene may be increased by about 5-95%, 10-90%, 15-85%, 20-80%, or 1.5-fold to 10-fold, relative to a control.

[0107] As detailed above, a Cas fusion protein may comprise at least one effector as the second polypeptide. The second polypeptide of the Cas effector may comprise at least one effector. As also detailed above, at least one effector may be fused to at least one antibody for use in a Suntag recruitment system or a variation thereof. Effectors may include, for example, MCRS1, OTUD7B, RelB, LDB1, NFKBIB, CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, and VPS72. Effectors that increase or enhance expression of a gene may be referred to as activators.

ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, or ZNF81, or a combination thereof. In some embodiments, the effector is selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, CITED2, PHF15, SS18L1, MLLT6, ASH2L, and GSK3A. In some embodiments, the effector is selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, and CITED2. In some embodiments, the second polypeptide domain or the effector has transcription repression activity, transcription activation activity, de-ubiquitinase activity, p300 recruitment activity, enhancer looping mediation activity, methylation activity, demethylation activity, acetylation activity, deacetylation activity, histone modification activity, histone acetylase activity, histone deacetylase activity, chromatin remodeling activity, chromatin looping modification activity, or a combination thereof.

[0108] In some embodiments, the effector reduces expression of a gene. Effectors that reduce expression of a gene may include MCRS1, OTUD7B, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81. Effectors that reduce expression of a gene may be referred to as repressors.

[0109] In some embodiments, the effector increases or enhances expression of a gene. Effectors that increase or enhance expression of a gene may include RelB, LDB1, NFKBIB, CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, and VPS72. Effectors that increase or enhance expression of a gene may be referred to as activators.

[0110] MCRS1 may comprise the amino acid sequence of SEQ ID NO: 57, encoded by a polynucleotide comprising the sequence of SEQ ID NO: 58. In some embodiments, the MCRS1 may comprise an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 57, or any fragment thereof. In some embodiments, the MCRS1 comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 57, or any fragment thereof. In some embodiments, the MCRS1 is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 58, or any fragment thereof. In some embodiments, the MCRS1 is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 58, or any fragment thereof.

[0111] OTUD7B may comprise the amino acid sequence of SEQ ID NO: 59, encoded by a polynucleotide comprising the sequence of SEQ ID NO: 60. In some embodiments, the OTUD7B comprises all of SEQ ID NO: 60 (“full OTUD7B”). OTUD7B may also comprise a fragment of SEQ ID NO: 60, such as a fragment comprising amino acids 167-440 or SEQ ID NP: 60, or a fragment comprising amino acids 792-831 of SEQ ID NO: 59. In some embodiments, the OTUD7B may comprise an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 59, or any fragment thereof. In some embodiments, the OTUD7B comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, rela-

[0147] ZNF254 may comprise the amino acid sequence of SEQ ID NO: 165, encoded by a polynucleotide comprising the sequence of SEQ ID NO: 166. In some embodiments, the ZNF254 may comprise an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 165, or any fragment thereof. In some embodiments, the ZNF254 comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 165, or any fragment thereof. In some embodiments, the ZNF254 is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 166, or any fragment thereof. In some embodiments, the ZNF254 is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 166, or any fragment thereof.

[0148] ZNF566 may comprise the amino acid sequence of SEQ ID NO: 167, encoded by a polynucleotide comprising the sequence of SEQ ID NO: 168. In some embodiments, the ZNF56 may comprise an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 167 or any fragment thereof. In some embodiments, the ZNF56 comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 167, or any fragment thereof. In some embodiments, the ZNF56X is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 168, or any fragment thereof. In some embodiments, the ZNF56 is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 168, or any fragment thereof.

[0149] ZNF585A may comprise the amino acid sequence of SEQ ID NO: 169, encoded by a polynucleotide comprising the sequence of SEQ ID NO: 170. In some embodiments, the ZNF585A may comprise an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 169, or any fragment thereof. In some embodiments, the ZNF585A comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 169, or any fragment thereof. In some embodiments, the ZNF585A is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 170, or any fragment thereof. In some embodiments, the ZNF585A is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 170, or any fragment thereof.

[0150] ZNF689 may comprise the amino acid sequence of SEQ ID NO: 171, encoded by a polynucleotide comprising the sequence of SEQ ID NO: 172. In some embodiments, the ZNF689 may comprise an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 171, or any fragment thereof. In some embodiments, the ZNF689 comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 171, or any fragment thereof. In some

embodiments, the ZNF689 is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 172, or any fragment thereof. In some embodiments, the ZNF689 is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 172, or any fragment thereof.

[0151] ZNF765 may comprise the amino acid sequence of SEQ ID NO: 173, encoded by a polynucleotide comprising the sequence of SEQ ID NO: 174. In some embodiments, the ZNF765 may comprise an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 173, or any fragment thereof. In some embodiments, the ZNF765 comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 173, or any fragment thereof. In some embodiments, the ZNF765 is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 174, or any fragment thereof. In some embodiments, the ZNF765 is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 174, or any fragment thereof.

[0152] ZNF81 may comprise the amino acid sequence of SEQ ID NO: 175, encoded by a polynucleotide comprising the sequence of SEQ ID NO: 176. In some embodiments, the ZNF81 may comprise an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 175, or any fragment thereof. In some embodiments, the ZNF81 comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 175, or any fragment thereof. In some embodiments, the ZNF81 is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 176, or any fragment thereof. In some embodiments, the ZNF81 is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 176, or any fragment thereof.

[0153] Other examples of effectors, or second polypeptide domains of the Cas fusion protein, are detailed below.

i) Transcription Activation Activity

[0154] The second polypeptide domain, or the effector, can have transcription activation activity, for example, a transactivation domain. For example, gene expression of endogenous mammalian genes, such as human genes, can be achieved by targeting a fusion protein of a first polypeptide domain, such as dCas9, and a transactivation domain to mammalian promoters via combinations of gRNAs. The transactivation domain can include a VP16 protein, multiple VP16 proteins, such as a VP48 domain or VP64 domain, p65 domain of NF kappa B transcription activator activity, TET1, VPR, VPH, Rta, and/or p300. For example, the fusion protein may comprise dCas9-p300. In some embodiments, p300 comprises a polypeptide having the amino acid sequence of SEQ ID NO: 41 or SEQ ID NO: 42. In other embodiments, the fusion protein comprises dCas9-VP64. In other embodiments, the fusion protein comprises VP64-

dCas9-VP64. VP64-dCas9-VP64 may comprise a polypeptide having the amino acid sequence of SEQ ID NO: 43, encoded by the polynucleotide of SEQ ID NO: 44. VPH may comprise a polypeptide having the amino acid sequence of SEQ ID NO: 53, encoded by a polynucleotide comprising the sequence of SEQ ID NO: 54. VPR may comprise a polypeptide having the amino acid sequence of SEQ ID NO: 55, encoded by a polynucleotide comprising the sequence of SEQ ID NO: 56.

ii) Transcription Repression Activity

[0155] The second polypeptide domain, or the effector, can have transcription repression activity. Non-limiting examples of repressors include Kruppel associated box activity such as a KRAB domain or KRAB, MECP2, EED, ERF repressor domain (ERD), Mad mSIN3 interaction domain (SID) or Mad-SID repressor domain, SID4X repressor domain, Mxi1 repressor domain, SUV39H1, SUV39H2, G9A, ESET/SETBD1, Cir4, Su(var)3-9, Pr-SET7/8, SUV4-20H1, PR-set7, Suv4-20, Set9, EZH2, RIZ1, JMJD2A/JHDM3A, JMJD2B, JMJ2D2C/GASC1, JMJD2D, Rph1, JARID1A/RBP2, JARID1B/PLU-1, JARID1C/SMCX, JARID1D/SMCY, Lid, Jhn2, Jmj2, HDAC1, HDAC2, HDAC3, HDAC8, Rpd3, Hos1, Cir6, HDAC4, HDAC5, HDAC7, HDAC9, Hda1, Cir3, SIRT1, SIRT2, Sir2, Hst1, Hst2, Hst3, Hst4, HDAC11, DNMT1, DNMT3a/3b, DNMT3A-3L, MET1, DRM3, ZMET2, CMT1, CMT2, Laminin A, Laminin B, CTCF, and/or a domain having TATA box binding protein activity, or a combination thereof. In some embodiments, the second polypeptide domain, or the effector, has a KRAB domain activity, ERF repressor domain activity, Mxi1 repressor domain activity, SID4X repressor domain activity, Mad-SID repressor domain activity, DNMT3A or DNMT3L or fusion thereof activity, LSD1 histone demethylase activity, or TATA box binding protein activity. In some embodiments, the second polypeptide domain or the effector comprises KRAB. KRAB may comprise a polypeptide having the amino acid sequence of SEQ ID NO: 45, encoded by polynucleotide comprising the sequence of SEQ ID NO: 46. For example, the fusion protein may be *S. pyogenes* dCas9-KRAB (protein sequence comprising SEQ ID NO: 47; polynucleotide sequence comprising SEQ ID NO: 48). The fusion protein may be *S. aureus* dCas9-KRAB (protein sequence comprising SEQ ID NO: 49; polynucleotide sequence comprising SEQ ID NO: 50).

iii) Transcription Release Factor Activity

[0156] The second polypeptide domain, or the effector, can have transcription release factor activity. The second polypeptide domain, or the effector, can have eukaryotic release factor 1 (ERF1) activity or eukaryotic release factor 3 (ERF3) activity.

iv) Histone Modification Activity

[0157] The second polypeptide domain, or the effector, can have histone modification activity. The second polypeptide domain, or the effector, can have histone deacetylase, histone acetyltransferase, histone demethylase, or histone methyltransferase activity. The histone acetyltransferase may be p300 or CREB-binding protein (CBP) protein, or fragments thereof. For example, the fusion protein may be dCas9-p300. In some embodiments, p300 comprises a polypeptide of SEQ ID NO: 41 or SEQ ID NO: 42.

v) Nuclease Activity

[0158] The second polypeptide domain, or the effector, can have nuclease activity that is different from the nuclease activity of the Cas9 protein. A nuclease, or a protein having nuclease activity, is an enzyme capable of cleaving the phosphodiester bonds between the nucleotide subunits of nucleic acids. Nucleases are usually further divided into endonucleases and exonucleases, although some of the enzymes may fall in both categories. Well known nucleases include deoxyribonuclease and ribonuclease.

vi) Nucleic Acid Association Activity

[0159] The second polypeptide domain, or the effector, can have nucleic acid association activity or nucleic acid binding protein-DNA-binding domain (DBD). A DBD is an independently folded protein domain that contains at least one motif that recognizes double- or single-stranded DNA. A DBD can recognize a specific DNA sequence (a recognition sequence) or have a general affinity to DNA. A nucleic acid association region may be selected from helix-turn-helix region, leucine zipper region, winged helix region, winged helix-turn-helix region, helix-loop-helix region, immunoglobulin fold, B3 domain, Zinc finger, HMG-box, Wor3 domain, and TAL effector DNA-binding domain.

vii) Methylase Activity

[0160] The second polypeptide domain, or the effector, can have methylase activity, which involves transferring a methyl group to DNA, RNA, protein, small molecule, cytosine, or adenine. In some embodiments, the second polypeptide domain or the effector includes a DNA methyltransferase.

viii) Demethylase Activity

[0161] The second polypeptide domain, or the effector, can have demethylase activity. The second polypeptide domain or the effector can include an enzyme that removes methyl (CH₃-) groups from nucleic acids, proteins (in particular histones), and other molecules. Alternatively, the second polypeptide or the effector can convert the methyl group to hydroxymethylcytosine in a mechanism for demethylating DNA. The second polypeptide or the effector can catalyze this reaction. For example, a second polypeptide that catalyzes this reaction can be Tet1, also known as Tet1CD (Ten-eleven translocation methylcytosine dioxygenase 1; amino acid sequence comprising SEQ ID NO: 51; polynucleotide sequence comprising SEQ ID NO: 52). In some embodiments, the second polypeptide domain or the effector has histone demethylase activity. In some embodiments, the second polypeptide domain or the effector has DNA demethylase activity.

e. Guide RNA (gRNA)

[0162] The CRISPR/Cas-based gene editing system may include at least one gRNA molecule. For example, the CRISPR/Cas-based gene editing system may include two gRNA molecules. The at least one gRNA molecule can bind and recognize a target region. The gRNA is the part of the CRISPR-Cas system that provides DNA targeting specificity to the CRISPR/Cas-based gene editing system. The gRNA is a fusion of two noncoding RNAs: a crRNA and a tracrRNA. gRNA mimics the naturally occurring crRNA:tracrRNA duplex involved in the Type II Effector system. This duplex, which may include, for example, a 42-nucleotide crRNA and a 75-nucleotide tracrRNA, acts as a guide for the Cas9 to bind, and in some cases, cleave the target nucleic acid. The

gRNA may target any desired DNA sequence by exchanging the sequence encoding a 20 bp protospacer which confers targeting specificity through complementary base pairing with the desired DNA target. The “target region” or “target sequence” or “protospacer” refers to the region of the target gene to which the CRISPR/Cas9-based gene editing system targets and binds. The portion of the gRNA that targets the target sequence in the genome may be referred to as the “targeting sequence” or “targeting portion” or “targeting domain.” “Protospacer” or “gRNA spacer” may refer to the region of the target gene to which the CRISPR/Cas9-based gene editing system targets and binds; “protospacer” or “gRNA spacer” may also refer to the portion of the gRNA that is complementary to the targeted sequence in the genome. The gRNA may include a gRNA scaffold. A gRNA scaffold facilitates Cas9 binding to the gRNA and may facilitate endonuclease activity. The gRNA scaffold is a polynucleotide sequence that follows the portion of the gRNA corresponding to sequence that the gRNA targets. Together, the gRNA targeting portion and gRNA scaffold form one polynucleotide. The constant region of the gRNA may include the sequence of SEQ ID NO: 19 (RNA), which is encoded by a sequence comprising SEQ ID NO: 18 (DNA). The CRISPR/Cas9-based gene editing system may include at least one gRNA, wherein the gRNAs target different DNA sequences. The target DNA sequences may be overlapping. The gRNA may comprise at its 5' end the targeting domain that is sufficiently complementary to the target region to be able to hybridize to, for example, about 10 to about 20 nucleotides of the target region of the target gene, when it is followed by an appropriate Protospacer Adjacent Motif (PAM). The target region or protospacer is followed by a PAM sequence at the 3' end of the protospacer in the genome. Different Type II systems have differing PAM requirements, as detailed above.

[0163] The targeting domain of the gRNA does not need to be perfectly complementary to the target region of the target DNA. In some embodiments, the targeting domain of the gRNA is at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or at least 99% complementary to (or has 1, 2 or 3 mismatches compared to) the target region over a length of, such as, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides. For example, the DNA-targeting domain of the gRNA may be at least 80% complementary over at least 18 nucleotides of the target region. The target region may be on either strand of the target DNA.

[0164] The gRNA may target the Cas9 protein or fusion protein to a gene or a regulatory element thereof. The gRNA may target the Cas protein or fusion protein to a non-open chromatin region, an open chromatin region, a transcribed region of the target gene, a region upstream of a transcription start site of the target gene, a regulatory element of the target gene, an intron of the target gene, or an exon of the target gene, or a combination thereof. In some embodiments, the gRNA targets the Cas9 protein or fusion protein to a promoter of a gene. In some embodiments, the target region is located between about 1 to about 1000 base pairs upstream of a transcription start site of a target gene. In some embodiments, the DNA targeting composition comprises two or more gRNAs, each gRNA binding to a different target region.

[0165] The gRNA may target a region within or near a gene of interest. For example, the gRNA may target B2M or CD25 or TetO (see TABLE 3 and TABLE 4). The gRNA

may target or bind to a regulatory region of a gene of interest. The gRNA may comprise a polynucleotide sequence comprising at least one of SEQ ID NOS: 96-98 and 101-102, or a complement thereof, or a variant thereof, or a truncation thereof. The gRNA may be encoded by a polynucleotide sequence comprising at least one of SEQ ID NOS: 93-95 and 99-100, or a complement thereof, or a variant thereof, or a truncation thereof. The gRNA may bind and target a polynucleotide sequence comprising at least one of SEQ ID NOS: 93-95 and 99-100, or a complement thereof, or a variant thereof, or a truncation thereof. A truncation may be 1, 2, 3, 4, 5, 6, 7, 8, or 9 nucleotides shorter than the sequence of any one of SEQ ID NOS: 93-102. In some embodiments, the gRNA targets or binds to a gene or regulatory element thereof that is related to a disease, such as, for example, Duchenne muscular dystrophy (DMD), Becker muscular dystrophy (BMD), and/or cancer.

[0166] As described above, the gRNA molecule comprises a targeting domain (also referred to as targeted or targeting sequence), which is a polynucleotide sequence complementary to the target DNA sequence. The gRNA may comprise a “G” at the 5' end of the targeting domain or complementary polynucleotide sequence. The CRISPR/Cas9-based gene editing system may use gRNAs of varying sequences and lengths. The targeting domain of a gRNA molecule may comprise at least a 10 base pair, at least a 11 base pair, at least a 12 base pair, at least a 13 base pair, at least a 14 base pair, at least a 15 base pair, at least a 16 base pair, at least a 17 base pair, at least a 18 base pair, at least a 19 base pair, at least a 20 base pair, at least a 21 base pair, at least a 22 base pair, at least a 23 base pair, at least a 24 base pair, at least a 25 base pair, at least a 30 base pair, or at least a 35 base pair complementary polynucleotide sequence of the target DNA sequence followed by a PAM sequence. In certain embodiments, the targeting domain of a gRNA molecule has 19-25 nucleotides in length. In certain embodiments, the targeting domain of a gRNA molecule is 20 nucleotides in length. In certain embodiments, the targeting domain of a gRNA molecule is 21 nucleotides in length. In certain embodiments, the targeting domain of a gRNA molecule is 22 nucleotides in length. In certain embodiments, the targeting domain of a gRNA molecule is 23 nucleotides in length.

[0167] The number of gRNA molecules that may be included in the CRISPR/Cas9-based gene editing system can be at least 1 gRNA, at least 2 different gRNAs, at least 3 different gRNAs, at least 4 different gRNAs, at least 5 different gRNAs, at least 6 different gRNAs, at least 7 different gRNAs, at least 8 different gRNAs, at least 9 different gRNAs, at least 10 different gRNAs, at least 11 different gRNAs, at least 12 different gRNAs, at least 13 different gRNAs, at least 14 different gRNAs, at least 15 different gRNAs, at least 16 different gRNAs, at least 17 different gRNAs, at least 18 different gRNAs, at least 18 different gRNAs, at least 20 different gRNAs, at least 25 different gRNAs, at least 30 different gRNAs, at least 35 different gRNAs, at least 40 different gRNAs, at least 45 different gRNAs, or at least 50 different gRNAs. The number of gRNA molecules that may be included in the CRISPR/Cas9-based gene editing system can be less than 50 different gRNAs, less than 45 different gRNAs, less than 40 different gRNAs, less than 35 different gRNAs, less than 30 different gRNAs, less than 25 different gRNAs, less than 20 different gRNAs, less than 19 different gRNAs, less than 18

different gRNAs, less than 17 different gRNAs, less than 16 different gRNAs, less than 15 different gRNAs, less than 14 different gRNAs, less than 13 different gRNAs, less than 12 different gRNAs, less than 11 different gRNAs, less than 10 different gRNAs, less than 9 different gRNAs, less than 8 different gRNAs, less than 7 different gRNAs, less than 6 different gRNAs, less than 5 different gRNAs, less than 4 different gRNAs, less than 3 different gRNAs, or less than 2 different gRNAs. The number of gRNAs that may be included in the CRISPR/Cas9-based gene editing system can be between at least 1 gRNA to at least 50 different gRNAs, at least 1 gRNA to at least 45 different gRNAs, at least 1 gRNA to at least 40 different gRNAs, at least 1 gRNA to at least 35 different gRNAs, at least 1 gRNA to at least 30 different gRNAs, at least 1 gRNA to at least 25 different gRNAs, at least 1 gRNA to at least 20 different gRNAs, at least 1 gRNA to at least 16 different gRNAs, at least 1 gRNA to at least 12 different gRNAs, at least 1 gRNA to at least 8 different gRNAs, at least 8 different gRNAs to at least 50 different gRNAs, at least 8 different gRNAs to at least 45 different gRNAs, at least 8 different gRNAs to at least 40 different gRNAs, at least 8 different gRNAs to at least 35 different gRNAs, 8 different gRNAs to at least 30 different gRNAs, at least 8 different gRNAs to at least 25 different gRNAs, 8 different gRNAs to at least 20 different gRNAs, at least 8 different gRNAs to at least 16 different gRNAs, or 8 different gRNAs to at least 12 different gRNAs.

f. Repair Pathways

[0168] The CRISPR/Cas9-based gene editing system may be used to introduce site-specific double strand breaks at targeted genomic loci. Site-specific double-strand breaks are created when the CRISPR/Cas9-based gene editing system binds to a target DNA sequences, thereby permitting cleavage of the target DNA. This DNA cleavage may stimulate the natural DNA-repair machinery, leading to one of two possible repair pathways: homology-directed repair (HDR) or the non-homologous end joining (NHEJ) pathway.

i) Homology-Directed Repair (HDR)

[0169] Restoration of protein expression from a gene may involve homology-directed repair (HDR). A donor template may be administered to a cell. A donor sequence comprises a polynucleotide sequence to be inserted into a genome. The donor template may include a nucleotide sequence encoding a full-functional protein or a partially functional protein. In such embodiments, the donor template may include fully functional gene construct for restoring a mutant gene, or a fragment of the gene that after homology-directed repair, leads to restoration of the mutant gene. In other embodiments, the donor template may include a nucleotide sequence encoding a mutated version of an inhibitory regulatory element of a gene. Mutations may include, for example, nucleotide substitutions, insertions, deletions, or a

combination thereof. In such embodiments, introduced mutation(s) into the inhibitory regulatory element of the gene may reduce the transcription of or binding to the inhibitory regulatory element.

ii) Non-Homologous End Joining (NHEJ)

[0170] Restoration of protein expression from gene may be through template-free NHEJ-mediated DNA repair. In certain embodiments, NHEJ is a nuclease mediated NHEJ, which in certain embodiments, refers to NHEJ that is initiated a Cas9 molecule that cuts double stranded DNA. The method comprises administering a presently disclosed CRISPR/Cas9-based gene editing system or a composition comprising thereof to a subject for gene editing.

[0171] Nuclease mediated NHEJ may correct a mutated target gene and offer several potential advantages over the HDR pathway. For example, NHEJ does not require a donor template, which may cause nonspecific insertional mutagenesis. In contrast to HDR, NHEJ operates efficiently in all stages of the cell cycle and therefore may be effectively exploited in both cycling and post-mitotic cells, such as muscle fibers. This provides a robust, permanent gene restoration alternative to oligonucleotide-based exon skipping or pharmacologic forced read-through of stop codons and could theoretically require as few as one drug treatment.

4. REPORTER PROTEIN

[0172] In some embodiments, the DNA targeting compositions or CRISPR/Cas9 systems include at least one reporter protein. For example, and as detailed above, the second polypeptide of the Cas effector may comprise a reporter protein such as sfBFP. A polynucleotide sequence encoding the reporter protein may be operably linked to the polynucleotide sequence encoding the Cas9 protein and/or Cas9 fusion protein and/or antibody and/or effector. The reporter protein may include any protein or peptide that is suitably detectable, such as, by fluorescence, chemiluminescence, enzyme activity such as beta galactosidase or alkaline phosphatase, and/or antibody binding detection. The reporter protein may comprise a fluorescent protein. The reporter protein may comprise a protein or peptide detectable with an antibody. For example, the reporter protein may comprise sfBFP, GFP, YFP, RFP, CFP, DsRed, luciferase, and/or Thy1.

5. GENETIC CONSTRUCTS

[0173] The CRISPR/Cas9-based gene editing system or any component thereof may be encoded by or comprised within one or more genetic constructs. The CRISPR/Cas9-based gene editing system may comprise one or more genetic constructs. The genetic construct, such as a plasmid or expression vector, may comprise a nucleic acid that encodes the CRISPR/Cas9-based gene editing system and/or at least one component thereof such as at lease one gRNA. In some embodiments, a genetic construct encodes at least one effector domain. In certain embodiments, a genetic construct encodes one gRNA molecule, i.e., a first gRNA molecule, and optionally a Cas9 molecule or fusion protein. In some embodiments, a genetic construct encodes two gRNA molecules, i.e., a first gRNA molecule and a second gRNA molecule, and optionally a Cas9 molecule or fusion protein. In some embodiments, a first genetic construct encodes one gRNA molecule, i.e., a first gRNA molecule,

and optionally a Cas9 molecule or fusion protein, and a second genetic construct encodes one gRNA molecule, i.e., a second gRNA molecule, and optionally a Cas9 molecule or fusion protein. In some embodiments, a first genetic construct encodes one gRNA molecule and one donor sequence, and a second genetic construct encodes a Cas9 molecule or fusion protein. In some embodiments, a first genetic construct encodes one gRNA molecule and a Cas9 molecule or fusion protein, and a second genetic construct encodes one donor sequence. In some embodiments, a single genetic construct encodes at least one effector domain, at least one antibody, a Cas9 molecule or fusion protein, and at least one peptide epitope. In some embodiments, a first genetic construct encodes at least one effector domain and at least one antibody, and a second genetic construct encodes a Cas9 molecule or fusion protein and at least one peptide epitope.

[0174] Genetic constructs may include polynucleotides such as vectors and plasmids. The genetic construct may be a linear minichromosome including centromere, telomeres, or plasmids or cosmids. The vector may be an expression vectors or system to produce protein by routine techniques and readily available starting materials including Sambrook et al., Molecular Cloning and Laboratory Manual, Second Ed., Cold Spring Harbor (1989), which is incorporated fully by reference. The construct may be recombinant. The genetic construct may be part of a genome of a recombinant viral vector, including recombinant lentivirus, recombinant adenovirus, and recombinant adenovirus associated virus. The genetic construct may comprise regulatory elements for gene expression of the coding sequences of the nucleic acid. The regulatory elements may be a promoter, an enhancer, an initiation codon, a stop codon, or a polyadenylation signal.

[0175] The genetic construct may comprise heterologous nucleic acid encoding the CRISPR/Cas-based gene editing system and may further comprise an initiation codon, which may be upstream of the CRISPR/Cas-based gene editing system coding sequence, and a stop codon, which may be downstream of the CRISPR/Cas-based gene editing system coding sequence. The genetic construct may include more than one stop codon, which may be downstream of the CRISPR/Cas-based gene editing system coding sequence. In some embodiments, the genetic construct includes 1, 2, 3, 4, or 5 stop codons. In some embodiments, the genetic construct includes 1, 2, 3, 4, or 5 stop codons downstream of the sequence encoding the donor sequence. A stop codon may be in-frame with a coding sequence in the CRISPR/Cas-based gene editing system. For example, one or more stop codons may be in-frame with the donor sequence. The genetic construct may include one or more stop codons that are out of frame of a coding sequence in the CRISPR/Cas-based gene editing system. For example, one stop codon may be in-frame with the donor sequence, and two other stop codons may be included that are in the other two possible reading frames. A genetic construct may include a stop codon for all three potential reading frames. The initiation and termination codon may be in frame with the CRISPR/Cas-based gene editing system coding sequence.

[0176] The vector may also comprise a promoter that is operably linked to the CRISPR/Cas-based gene editing system coding sequence. The promoter may be a constitutive promoter, an inducible promoter, a repressible promoter, or a regulatable promoter. The promoter may be a ubiquitous promoter. The promoter may be a tissue-specific promoter. The tissue specific promoter may be a muscle specific

promoter. The tissue specific promoter may be a skin specific promoter. The CRISPR/Cas-based gene editing system may be under the light-inducible or chemically inducible control to enable the dynamic control of gene/genome editing in space and time. The promoter operably linked to the CRISPR/Cas-based gene editing system coding sequence may be a promoter from simian virus 40 (SV40), a mouse mammary tumor virus (MMTV) promoter, a human immunodeficiency virus (HIV) promoter such as the bovine immunodeficiency virus (BIV) long terminal repeat (LTR) promoter, a Moloney virus promoter, an avian leukosis virus (ALV) promoter, a cytomegalovirus (CMV) promoter such as the CMV immediate early promoter, Epstein Barr virus (EBV) promoter, or a Rous sarcoma virus (RSV) promoter. The promoter may also be a promoter from a human gene such as human ubiquitin C (hUbC), human actin, human myosin, human hemoglobin, human muscle creatine, or human metallothionein. Examples of a tissue specific promoter, such as a muscle or skin specific promoter, natural or synthetic, are described in U.S. Patent Application Publication No. US20040175727, the contents of which are incorporated herein in its entirety. The promoter may be a CK8 promoter, a Spc512 promoter, a MHCK7 promoter, for example.

[0177] The genetic construct may also comprise a polyadenylation signal, which may be downstream of the CRISPR/Cas-based gene editing system. The polyadenylation signal may be a SV40 polyadenylation signal, LTR polyadenylation signal, bovine growth hormone (bGH) polyadenylation signal, human growth hormone (hGH) polyadenylation signal, or human β -globin polyadenylation signal. The SV40 polyadenylation signal may be a polyadenylation signal from a pCEP4 vector (Invitrogen, San Diego, CA).

[0178] Coding sequences in the genetic construct may be optimized for stability and high levels of expression. In some instances, codons are selected to reduce secondary structure formation of the RNA such as that formed due to intramolecular bonding.

[0179] The genetic construct may also comprise an enhancer upstream of the CRISPR/Cas-based gene editing system or gRNAs. The enhancer may be necessary for DNA expression. The enhancer may be human actin, human myosin, human hemoglobin, human muscle creatine or a viral enhancer such as one from CMV, HA, RSV, or EBV. Polynucleotide function enhancers are described in U.S. Pat. Nos. 5,593,972, 5,962,428, and WO94/016737, the contents of each are fully incorporated by reference. The genetic construct may also comprise a mammalian origin of replication in order to maintain the vector extrachromosomally and produce multiple copies of the vector in a cell. The genetic construct may also comprise a regulatory sequence, which may be well suited for gene expression in a mammalian or human cell into which the vector is administered. The genetic construct may also comprise a reporter gene, such as green fluorescent protein ("GFP") and/or a selectable marker, such as hygromycin ("Hygro").

[0180] The genetic construct may be useful for transfecting cells with nucleic acid encoding the CRISPR/Cas-based gene editing system, which the transformed host cell is cultured and maintained under conditions wherein expression of the CRISPR/Cas-based gene editing system takes place. The genetic construct may be transformed or transduced into a cell. The genetic construct may be formulated

into any suitable type of delivery vehicle including, for example, a viral vector, lentiviral expression, mRNA electroporation, and lipid-mediated transfection for delivery into a cell. The genetic construct may be part of the genetic material in attenuated live microorganisms or recombinant microbial vectors which live in cells. The genetic construct may be present in the cell as a functioning extrachromosomal molecule.

[0181] Further provided herein is a cell transformed or transduced with a system or component thereof as detailed herein. Suitable cell types are detailed herein. In some embodiments, the cell is a stem cell. The stem cell may be a human stem cell. In some embodiments, the cell is an embryonic stem cell. The stem cell may be a human pluripotent stem cell (iPSCs). Further provided are stem cell-derived neurons, such as neurons derived from iPSCs transformed or transduced with a DNA targeting system or component thereof as detailed herein.

a. Viral Vectors

[0182] A genetic construct may be a viral vector. Further provided herein is a viral delivery system. Viral delivery systems may include, for example, lentivirus, retrovirus, adenovirus, mRNA electroporation, or nanoparticles. In some embodiments, the vector is a modified lentiviral vector. In some embodiments, the viral vector is an adeno-associated virus (AAV) vector. The AAV vector is a small virus belonging to the genus Dependovirus of the Parvoviridae family that infects humans and some other primate species.

[0183] AAV vectors may be used to deliver CRISPR/Cas9-based gene editing systems using various construct configurations. For example, AAV vectors may deliver Cas9 or fusion protein and gRNA expression cassettes on separate vectors or on the same vector. Alternatively, if the small Cas9 proteins or fusion proteins, derived from species such as *Staphylococcus aureus* or *Neisseria meningitidis*, are used then both the Cas9 and up to two gRNA expression cassettes may be combined in a single AAV vector. In some embodiments, the AAV vector has a 4.7 kb packaging limit.

[0184] In some embodiments, the AAV vector is a modified AAV vector. The modified AAV vector may have enhanced cardiac and/or skeletal muscle tissue tropism. The modified AAV vector may be capable of delivering and expressing the CRISPR/Cas9-based gene editing system in the cell of a mammal. For example, the modified AAV vector may be an AAV-SASTG vector (Piacentino et al. *Human Gene Therapy* 2012, 23, 635-646). The modified AAV vector may be based on one or more of several capsid types, including AAV1, AAV2, AAV5, AAV6, AAV8, and AAV9. The modified AAV vector may be based on AAV2 pseudotype with alternative muscle-tropic AAV capsids, such as AAV2/1, AAV2/6, AAV2/7, AAV2/8, AAV2/9, AAV2.5, and AAV/SASTG vectors that efficiently transduce skeletal muscle or cardiac muscle by systemic and local delivery (Seto et al. *Current Gene Therapy* 2012, 12, 139-151). The modified AAV vector may be AAV2i8G9 (Shen et al. *J. Biol. Chem.* 2013, 288, 28814-28823).

6. PHARMACEUTICAL COMPOSITIONS

[0185] Further provided herein are pharmaceutical compositions comprising the above-described genetic constructs or gene editing systems. In some embodiments, the pharmaceutical composition may comprise about 1 ng to about 10 mg of DNA encoding the CRISPR/Cas-based gene

editing system or at least one component thereof. The systems or genetic constructs as detailed herein, or at least one component thereof, may be formulated into pharmaceutical compositions in accordance with standard techniques well known to those skilled in the pharmaceutical art. The pharmaceutical compositions can be formulated according to the mode of administration to be used. In cases where pharmaceutical compositions are injectable pharmaceutical compositions, they are sterile, pyrogen free, and particulate free. An isotonic formulation is preferably used. Generally, additives for isotonicity may include sodium chloride, dextrose, mannitol, sorbitol and lactose. In some cases, isotonic solutions such as phosphate buffered saline are preferred. Stabilizers include gelatin and albumin. In some embodiments, a vasoconstriction agent is added to the formulation.

[0186] The composition may further comprise a pharmaceutically acceptable excipient. The pharmaceutically acceptable excipient may be functional molecules as vehicles, adjuvants, carriers, or diluents. The term "pharmaceutically acceptable carrier," may be a non-toxic, inert solid, semi-solid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. Pharmaceutically acceptable carriers include, for example, diluents, lubricants, binders, disintegrants, colorants, flavors, sweeteners, antioxidants, preservatives, glidants, solvents, suspending agents, wetting agents, surfactants, emollients, propellants, humectants, powders, pH adjusting agents, and combinations thereof. The pharmaceutically acceptable excipient may be a transfection facilitating agent, which may include surface active agents, such as immune-stimulating complexes (ISCOMS), Freunds incomplete adjuvant, LPS analog including monophosphoryl lipid A, muramyl peptides, quinone analogs, vesicles such as squalene and squalene, hyaluronic acid, lipids, liposomes, calcium ions, viral proteins, polyanions, polycations, or nanoparticles, or other known transfection facilitating agents. The transfection facilitating agent may be a polyanion, polycation, including poly-L-glutamate (LGS), or lipid. The transfection facilitating agent may be poly-L-glutamate, and more preferably, the poly-L-glutamate may be present in the composition for gene editing in skeletal muscle or cardiac muscle at a concentration less than 6 mg/mL.

7. ADMINISTRATION

[0187] The systems or genetic constructs as detailed herein, or at least one component thereof, may be administered or delivered to a cell. Methods of introducing a nucleic acid into a host cell are known in the art, and any known method can be used to introduce a nucleic acid (e.g., an expression construct) into a cell. Suitable methods include, for example, viral or bacteriophage infection, transfection, conjugation, protoplast fusion, polycation or lipid:nucleic acid conjugates, lipofection, electroporation, nucleofection, immunoliposomes, calcium phosphate precipitation, polyethyleneimine (PEI)-mediated transfection, DEAE-dextran mediated transfection, liposome-mediated transfection, particle gun technology, calcium phosphate precipitation, direct micro injection, nanoparticle-mediated nucleic acid delivery, and the like. In some embodiments, the composition may be delivered by mRNA delivery and ribonucleoprotein (RNP) complex delivery. The system, genetic construct, or composition comprising the same, may be electroporated using BioRad Gene Pulser Xcell or Amaxa Nucleofector IIb devices or other electroporation device. Several different

buffers may be used, including BioRad electroporation solution, Sigma phosphate-buffered saline product #D8537 (PBS), Invitrogen OptiMEM I (OM), or Amaxa Nucleofector solution V (N.V.). Transfections may include a transfection reagent, such as Lipofectamine 2000.

[0188] The systems or genetic constructs as detailed herein, or at least one component thereof, or the pharmaceutical compositions comprising the same, may be administered to a subject. Such compositions can be administered in dosages and by techniques well known to those skilled in the medical arts taking into consideration such factors as the age, sex, weight, and condition of the particular subject, and the route of administration. The presently disclosed systems, or at least one component thereof, genetic constructs, or compositions comprising the same, may be administered to a subject by different routes including orally, parenterally, sublingually, transdermally, rectally, transmucosally, topically, intranasal, intravaginal, via inhalation, via buccal administration, intrapleurally, intravenous, intraarterial, intraperitoneal, subcutaneous, intradermally, epidermally, intramuscular, intranasal, intrathecal, intracranial, and intraarticular or combinations thereof. In certain embodiments, the system, genetic construct, or composition comprising the same, is administered to a subject intramuscularly, intravenously, or a combination thereof. The systems, genetic constructs, or compositions comprising the same may be delivered to a subject by several technologies including DNA injection (also referred to as DNA vaccination) with and without in vivo electroporation, liposome mediated, nanoparticle facilitated, recombinant vectors such as recombinant lentivirus, recombinant adenovirus, and recombinant adenovirus associated virus. The composition may be injected into the brain or other component of the central nervous system. The composition may be injected into the skeletal muscle or cardiac muscle. For example, the composition may be injected into the tibialis anterior muscle or tail. For veterinary use, the systems, genetic constructs, or compositions comprising the same may be administered as a suitably acceptable formulation in accordance with normal veterinary practice. The veterinarian may readily determine the dosing regimen and route of administration that is most appropriate for a particular animal. The systems, genetic constructs, or compositions comprising the same may be administered by traditional syringes, needleless injection devices, "microprojectile bombardment gun guns," or other physical methods such as electroporation ("EP"), "hydrodynamic method", or ultrasound. Alternatively, transient in vivo delivery of CRISPR/Cas-based systems by non-viral or non-integrating viral gene transfer, or by direct delivery of purified proteins and gRNAs containing cell-penetrating motifs may enable highly specific correction and/or restoration in situ with minimal or no risk of exogenous DNA integration.

[0189] Upon delivery of the presently disclosed systems or genetic constructs as detailed herein, or at least one component thereof, or the pharmaceutical compositions comprising the same, and thereupon the vector into the cells of the subject, the transfected cells may express the gRNA molecule(s) and/or the Cas9 molecule or fusion protein and/or Cas effector and/or effector domain.

a. Cell Types

[0190] Any of the delivery methods and/or routes of administration detailed herein can be utilized with a myriad of cell types. Further provided herein is a cell transformed

or transduced with a system or component thereof as detailed herein. For example, provided herein is a cell comprising an isolated polynucleotide encoding a CRISPR/Cas9 system as detailed herein. Suitable cell types are detailed herein. In some embodiments, the cell is an immune cell. Immune cells may include, for example, lymphocytes such as T cells and B cells and natural killer (NK) cells. In some embodiments, the cell is a T cell. T cells may be divided into cytotoxic T cells and helper T cells, which are in turn categorized as TH1 or TH2 helper T cells. Immune cells may further include innate immune cells, adaptive immune cells, tumor-primed T cells, NKT cells, IFN- γ producing killer dendritic cells (IKDC), memory T cells (TCMs), and effector T cells (Tes). The cell may be a stem cell such as a human stem cell. In some embodiments, the cell is an embryonic stem cell or a hematopoietic stem cell. The stem cell may be a human induced pluripotent stem cell (iPSCs). Further provided are stem cell-derived neurons, such as neurons derived from iPSCs transformed or transduced with a DNA targeting system or component thereof as detailed herein. The cell may be a muscle cell. Cells may further include, but are not limited to, immortalized myoblast cells, dermal fibroblasts, bone marrow-derived progenitors, skeletal muscle progenitors, human skeletal myoblasts, CD 133+ cells, mesoangioblasts, cardiomyocytes, hepatocytes, chondrocytes, mesenchymal progenitor cells, hematopoietic stem cells, smooth muscle cells, and MyoD- or Pax7-transduced cells, or other myogenic progenitor cells.

8. KITS

[0191] Provided herein is a kit, which may be used to modulate gene expression. The kit comprises genetic constructs or a composition comprising the same, for modulating gene expression, as described above, and instructions for using said composition. In some embodiments, the kit includes at least one effector as detailed herein, or a polynucleotide encoding the at least one effector. The effector may be selected from, for example, MCRS1, OTUD7B, RelB, LDB1, NFKBIB, CITED2, ASH2L, GSK3A, MLLT6, PHF15, SS18L1, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, JAZF1, KAT7, KEAP1, MEAF6, MORF4L2, NFYC, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, TADA3, TAF6, TBPL1, VPS72, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81. In some embodiments, the kit comprises at least one gRNA as detailed herein. The kit may further include instructions for using the CRISPR/Cas-based gene editing system.

[0192] Instructions included in kits may be affixed to packaging material or may be included as a package insert. While the instructions are typically written on printed materials they are not limited to such. Any medium capable of storing such instructions and communicating them to an end user is contemplated by this disclosure. Such media include, but are not limited to, electronic storage media (e.g., magnetic discs, tapes, cartridges, chips), optical media (e.g., CD ROM), and the like. As used herein, the term "instructions" may include the address of an internet site that provides the instructions.

[0193] The genetic constructs or a composition comprising thereof for modulating gene expression may include a modified AAV vector that includes a gRNA molecule(s) and a Cas9 protein or fusion protein or Cas effector, as described

above. The CRISPR/Cas-based gene editing system, as described above, may be included in the kit to specifically bind and target a particular region in a gene.

9. METHODS

a. Methods of Modulating Expression of a Gene

[0194] Provided herein are methods of modulating expression of a gene in a cell or in a subject. The methods may include administering to the cell or the subject a DNA targeting composition as detailed herein or at least one component thereof, or an isolated polynucleotide sequence as detailed herein, or a vector as detailed herein, or a pharmaceutical composition as detailed herein, or a combination thereof. In some embodiments, the method includes administering to a cell or subject an effector selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, VPS72, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof, or a polynucleotide encoding the effector. In some embodiments, the effector is targeted to a gene or a regulatory element thereof.

[0195] In some embodiments, the expression of the gene is increased relative to a control. In some embodiments, wherein the expression of the gene is decreased relative to a control. In some embodiments, the gene comprises the dystrophin gene, or the CD25 gene, or the B2M gene, or the TRAC gene. In some embodiments, the cell is a muscle cell or a T cell.

[0196] In some embodiments, the gene is the dystrophin gene. Dystrophin is a rod-shaped cytoplasmic protein which is a part of a protein complex that connects the cytoskeleton of a muscle fiber to the surrounding extracellular matrix through the cell membrane. Dystrophin provides structural stability to the dystroglycan complex of the cell membrane. The dystrophin gene is 2.2 megabases at locus Xp21. The primary transcription measures about 2,400 kb with the mature mRNA being about 14 kb. 79 exons include approximately 2.2 million nucleotides and code for the protein which is over 3500 amino acids. Normal skeleton muscle tissue contains only small amounts of dystrophin, but its absence of abnormal expression leads to the development of severe and incurable symptoms. Some mutations in the dystrophin gene lead to the production of defective dystrophin and severe dystrophic phenotype in affected patients. Some mutations in the dystrophin gene lead to partially-functional dystrophin protein and a much milder dystrophic phenotype in affected patients.

[0197] Duchenne muscular dystrophy (DMD) is the result of inherited or X-linked recessive spontaneous mutation(s) that cause nonsense or frame shift mutations in the dystrophin gene. DMD is a severe, highly debilitating and incurable muscle disease and is the most prevalent lethal heritable childhood disease and affects approximately one in 5,000 newborn males. DMD is characterized by muscle deterioration, progressive muscle weakness, often leading to mortality in subjects at age mid-twenties and premature death, due to the lack of a functional dystrophin gene. Most mutations are deletions in the dystrophin gene that disrupt the reading frame. Naturally occurring mutations and their consequences are relatively well understood for DMD. In-

frame deletions that occur in the exon 45-55 regions contained within the rod domain can produce highly functional dystrophin proteins, and many carriers are asymptomatic or display mild symptoms. Exons 45-55 of dystrophin are a mutational hotspot. More than 60% of patients may be treated by targeting exons in this region of the dystrophin gene. Efforts have been made to restore the disrupted dystrophin reading frame in DMD patients by skipping non-essential exon(s) (e.g., exon 45 skipping) during mRNA splicing to produce internally deleted but functional dystrophin proteins. The deletion of internal dystrophin exon(s) (for example, deletion of exon 45) may retain the proper reading frame and can generate an internally truncated but partially functional dystrophin protein. Deletions between exons 45-55 of dystrophin can result in a phenotype that is much milder compared to DMD.

[0198] A dystrophin gene may be a mutant dystrophin gene. A dystrophin gene may be a wild-type dystrophin gene. A dystrophin gene may have a sequence that is functionally identical to a wild-type dystrophin gene, for example, the sequence may be codon-optimized but still encode for the same protein as the wild-type dystrophin. A mutant dystrophin gene may include one or more mutations relative to the wild-type dystrophin gene. Mutations may include, for example, nucleotide deletions, substitutions, additions, transversions, or combinations thereof. A mutation in the dystrophin gene may be a functional deletion of the dystrophin gene. In some embodiments, the mutation in the dystrophin gene comprises an insertion or deletion in the dystrophin gene that prevents protein expression from the dystrophin gene. Mutations may be in one or more exons and/or introns. Mutations may include deletions of all or parts of at least one intron and/or exon. An exon of a mutant dystrophin gene may be mutated or at least partially deleted from the dystrophin gene. An exon of a mutant dystrophin gene may be fully deleted. A mutant dystrophin gene may have a portion or fragment thereof that corresponds to the corresponding sequence in the wild-type dystrophin gene. In some embodiments, a disrupted dystrophin gene caused by a deleted or mutated exon can be restored in DMD patients by adding back the corresponding wild-type exon. In some embodiments, disrupted dystrophin caused by a deleted or mutated exon 52 can be restored in DMD patients by adding back in wild-type exon 52. In certain embodiments, addition of exon 52 to restore reading frame ameliorates the phenotype in DMD subjects, including DMD subjects with deletion mutations. In certain embodiments, one or more exons may be added and inserted into the disrupted dystrophin gene. The one or more exons may be added and inserted so as to restore the corresponding mutated or deleted exon(s) in dystrophin. The one or more exons may be added and inserted into the disrupted dystrophin gene in addition to adding back and inserting the exon 52. In certain embodiments, exon 52 of a dystrophin gene refers to the 52nd exon of the dystrophin gene. Exon 52 is frequently adjacent to frame-disrupting deletions in DMD patients.

b. Methods of Treating a Disease

[0199] Provided herein are methods of treating a disease in a subject. The methods may include administering to the cell or the subject a DNA targeting composition as detailed herein or at least one component thereof, or an isolated polynucleotide sequence as detailed herein, or a vector as detailed herein, or a pharmaceutical composition as detailed herein, or a combination thereof. In some embodiments, the

method includes administering to the subject an effector selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, VPS72, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof, or a polynucleotide encoding the effector. In some embodiments, the effector is targeted to a gene or a regulatory element thereof.

were then transduced at a minimum of 200-fold coverage (600,000 cells for 3000 effectors). Cells were cultured for 10 days after transduction with the library. Cells were then subjected to fluorescence-activated cell sorting, and the top and bottom 10% by antibody staining for the target protein (B2M) were collected. Genomic DNA was purified, and the barcode cassette was amplified and sequenced on an Illumina MiSeq (San Diego, CA) to generate Log 2Fold Change and P-values. Calculations were performed by first summing all mapped barcodes for each effector in each condition. The gRNAs used are shown in TABLE 3.

TABLE 3

gRNA sequences		
Target	DNA	RNA
B2M	GGGCCAGTCTGCAAAGCGAG (SEQ ID NO: 93)	GGGCCAGUCUGCAAAGCGAG (SEQ ID NO: 96)
CD25	TTATGGGCGTAGCTGAAGAA (SEQ ID NO: 94)	UUUAUGGGCGUAGCUGAAGAA (SEQ ID NO: 97)
Non-targeting	GTATGGAGGGCTGGATCTGC (SEQ ID NO: 95)	GUAUGGAGGGCUGGAUCUGC (SEQ ID NO: 98)

[0200] In some embodiments, the disease is selected from Duchenne muscular dystrophy (DMD), Becker muscular dystrophy (BMD), and cancer.

10. EXAMPLES

[0201] The foregoing may be better understood by reference to the following examples, which are presented for purposes of illustration and are not intended to limit the scope of the invention. The present disclosure has multiple aspects and embodiments, illustrated by the appended non-limiting examples.

Example 1

Effector Screen 1: Suntag System and B2M Expression

[0202] A library was generated including 3015 effector domains derived from a commercial ORFeome library. A version of the suntag system compatible with LR cloning to insert effectors was generated, and random barcodes were appended at high coverage. Effectors were then cloned in by LR cloning, intentionally bottlenecked at 100 k colonies to maintain a manageable number of barcodes. Barcodes were then mapped to effectors using nanopore sequencing.

[0203] The effect of each effector on gene expression was measured in pooled screens. Each effector from the library was recruited to dCas9 using a slightly modified version of the Suntag recruitment system (Tanenbaum et al., *Cell* 2014, 159, 635-646, incorporated herein by reference in its entirety). The modified version included a Cas9 protein fused to repeats of a GCN4 peptide epitope, a gRNA targeting the Cas9 to the target gene, and an antibody to the epitope fused to one effector from the library with the setup ScFV-sfBFP-[EFFECTOR]. For this experiment, the target gene was B2M. Lentivirus encoding the library was produced in 293T cells and titered based on sfBFP fluorescence of a dilution series in the cell type used in the screen. Cells

[0204] Novel effectors were discovered that activate or repress gene expression when recruited via dCas9 to a gene of interest. Hits shown in FIG. 1 were cloned individually, and lentivirus was produced. 293T cells encoding dCas9 and either a B2M-targeting gRNA or non-targeting gRNA were each transduced in duplicate. Cells were cultured for 10 days after transduction with the library. Cells were then stained for B2M and analyzed by flow cytometry.

[0205] The effectors resulting in significant increased or decreased expression of B2M with the targeting gRNA but not with the non-targeting gRNA included MCRS1, OTUD7B, RelB, LDB1, NFKBIB, and CITED2. Two novel hits were discovered in the first screen, MCRS1 and OTUD7B. Both appeared to repress gene expression when recruited to dCas9 at a target gene promoter, and both have not previously been used as dCas9 fusions. OTUD7B (also known as Cezanne) is a de-ubiquitinase which has previously been shown to be involved in DNA repair but not to repress gene expression (Mevissen et al, *Nature* 2016, 538, 402-405, incorporated herein by reference in its entirety). MCRS1 has been shown to bind the DAXX repressor, which may explain its repressive effect (Lin, D. Y. et al. *J. Biol. Chem.* 2002, 277, 25446-25456, incorporated herein by reference in its entirety).

[0206] FIG. 1 shows the percent of cells in the low B2M bin, with higher numbers suggesting more potent repression. The results shown in FIG. 1 were based on fold changes and p-values for all tested effectors targeted to B2M in 293T cells (TABLE 1). Cells were screened by B2M staining in flow cytometry, and fold changes were calculated between barcode counts recovered from cells collected in the top or bottom 10% B2M expression. A non-targeting guide was also included as a control for non-specific repression. MCRS1 and OTUD7B both showed repression that is both greater than the steric effects of dCas9 alone and largely dependent on dCas9 targeting, rather than a non-specific

effect. Although other effectors also repressed B2M as predicted from the screen, these effects appeared to be non-specific.

Example 2

Effector Screen 2: SunTag System and CD25 Expression

[0207] A second screening experiment as detailed in Example 1 was completed, except examining CD25 expression instead of B2M, and these further experiments were completed to determine the fold changes and p-values for all tested effectors targeted to CD25 in Jurkat cells (TABLE 2). Cells were screened by CD25 staining in flow cytometry, and fold changes were calculated between barcode counts recovered from cells collected in the top or bottom 10%. Jurkat cell lines were generated by first transducing with lentiviral vectors encoding an sgRNA and dCas9 fused to a gcn4 peptide array that recruits the effector. A cell line with a CD25 targeting guide or a non-targeting guide was generated. These cell lines were then transduced with the indicated effectors fused to an scFv for recruitment to dCas9 (Tanenbaum et al., *Cell* 2014, 159, 635-646, incorporated herein by reference in its entirety).

[0208] Cells were cultured for 7 days after transduction with effector virus and stained for CD25 expression using a

with the highest fold induction was chosen for the screen. This cell line was then transduced with lentivirus encoding both the TetO targeting and non-targeting (negative control) sgRNA along with iRFP. These transduced cells were then sorted for iRFP expression to generate pure populations expressing each sgRNA.

[0211] The TetO-GFP reporter cell lines (with either TetO targeting or non-targeting gRNA), were transduced at an MOI of 0.2 with lentivirus encoding the effector library. A total of 3.75 million cells were transduced with virus, giving 300-fold coverage (750,000 transductants) of the approximately 2500 effectors in the library. Cells were then cultured for three days, subjected to puromycin selection (0.5 µg/mL) for 3 days, and then allowed to expand for an additional 4 days before sorting the top 10% of GFP expressing cells. Genomic DNA was purified from the collected cells, the DNA encoding the effector barcodes was amplified by PCR, and the resulting amplicons were sequenced on an Illumina MiSeq (San Diego, CA). The barcode frequency in each sample was determined using custom python scripts, and the resulting barcode abundances were analyzed in the DESeq2 R package to calculate fold changes and p values between the input cells and the top 10% GFP expressing cells. This was performed for both the TetO-targeting gRNA and the non-targeting gRNA. The gRNAs used are shown in TABLE 4.

TABLE 4

gRNA sequences		
Target	DNA	RNA
TetO	TACGTTCTCTATCACTGATA (SEQ ID NO: 99)	UACGUUCUCUAUCACUGUA (SEQ ID NO: 101)
Non-targeting	TATGGAGGGCTGGATCTGCG (SEQ ID NO: 100)	UAUAGGAGGGCUGGAUCUGCG (SEQ ID NO: 102)

CD25 Monoclonal Antibody (BC96, PE-Cyanine7, eBioscience™, San Diego, CA). Only cells positive for the BFP fluorophore associated with the effector virus were included in the analysis of positive cells.

[0209] Shown in FIG. 2A is the level of CD25 activation after delivery of each effector domain recruited by dCas9 in Jurkat cells. A non-targeting guide (gray bars) showed no effect on CD25, suggesting that each effector is specifically activating CD25 upon recruitment by dCas9. Shown in FIG. 2B is a zoomed-in view of data in FIG. 2A to show the specific activation by LDB1 and NFKBIB.

Example 3

Effector Screen 3: High-Throughput TetO-GFP Screen

[0210] A cell line was constructed for use in a TetO-GFP reporter screen. 293T cells were first transduced with dCas9-GCN4, which recruited the ScFv fused to an effector, and subjected to blast selection (5 µg/mL). These cells were then transduced with lentivirus encoding a minimal CMV promoter driving GFP expression and flanked by 7 repeats of the Tet operator. Clonal cell lines were generated by plating of a limiting dilution in a 96-well plate. Twelve clonal cell lines were then tested for robust GFP induction upon delivery of ScFv-VPR (a known positive control), and the clone

[0212] Shown in FIGS. 3A-3B are plots showing results for each effector in a screen for the ability to modulate GFP reporter expression. Log 2 (fold change) and Log 10 (Adjusted P Value) for each effector in the screen are plotted. Effectors with Log 2(fold change)>1.1 and Adjusted P Value <0.01 were considered to be hits and are shown in filled black circles, while non-hits are shown in open gray circles. This threshold gave 41 hits in the targeting condition and only 1 hit in the non-targeting condition, suggesting that it accurately filtered for legitimate hits. The 40 effector hits in the targeting condition that are not hits in the non-targeting (NT) condition included ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, VPS72, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, which are disclosed herein as SEQ ID NOS: 103-176. These effectors showed significant increased or decreased expression of GFP with the targeting gRNA but not with the non-targeting gRNA.

Example 4

Effector Screen 4: Examining Subset of Effectors with TetO-GFP Reporter

[0213] A subset of the effectors discovered as described in Example 3 was further examined using the same TetO-GFP

reporter. As shown in FIG. 4, 293T cells containing a GFP reporter were transduced with Lentivirus encoding a subset of effectors (PHF15, SS18L1, MLLT6, ASH2L, and GSK3A) found to be hits in the high-throughput screen along with a targeting or non-targeting gRNA. The fold activation of GFP (shown above each pair of bars) was found to be greater than 1 for all effectors tested, while the dCas9 alone control showed the opposite trend, supporting the idea that even the small effects seen for some effectors are likely meaningful. All hit effectors tested did modulate GFP to some degree, suggesting that all effectors found to be hits in the high-throughput screen of Example 3 are likely to be modulators of gene expression.

[0214] CITED2 and LDB1 were also examined for activation of GFP expression in 293T cells, with results shown in FIG. 5. 293T cells previously transduced with a TetO-GFP reporter were transfected with the indicated effector. Both LDB1 and CITED2 were able to robustly activate GFP expression, demonstrating that activation by these effectors was not limited to CD25, as shown in Example 2.

Example 5

Effect of LDB1 Dimerization Domain on Activation of Gene Expression

[0215] The LDB1 effector was examined using the CD25 expression system detailed in Example 2. Wild-type LDB1, as well as a mutant LDB1 with the dimerization domain deleted, were tested. Jurkat cells expressing dCas9-GCN4 and a CD25-targeting or non-targeting gRNA were transduced with the indicated effector-scFv fusion, and CD25 expression was analyzed by flow cytometry 10 days later. Results are shown in FIG. 6. Only the intact LDB1 effector was able to activate CD25. Activation of CD25 by LDB1 was dependent on the LDB1 dimerization domain. The dimerization domain deletion was a small deletion in the dimerization domain that was shown to be necessary for chromatin looping (Ivan Krivega, et al. *Genes Dev.* 2014, 28, 1278-90, incorporated herein by reference in its entirety), which suggested that LDB1 activated CD25 expression via a mechanism involving chromatin looping.

[0216] The foregoing description of the specific aspects will so fully reveal the general nature of the invention that others can, by applying knowledge within the skill of the art, readily modify and/or adapt for various applications such specific aspects, without undue experimentation, without departing from the general concept of the present disclosure. Therefore, such adaptations and modifications are intended to be within the meaning and range of equivalents of the disclosed aspects, based on the teaching and guidance presented herein. It is to be understood that the phraseology or terminology herein is for the purpose of description and not of limitation, such that the terminology or phraseology of the present specification is to be interpreted by the skilled artisan in light of the teachings and guidance.

[0217] The breadth and scope of the present disclosure should not be limited by any of the above-described exemplary aspects, but should be defined only in accordance with the following claims and their equivalents.

[0218] All publications, patents, patent applications, and/or other documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication, patent, patent appli-

cation, and/or other document were individually indicated to be incorporated by reference for all purposes.

[0219] For reasons of completeness, various aspects of the invention are set out in the following numbered clauses:

[0220] Clause 1. A Cas effector comprising: a first polypeptide comprising a Cas protein and at least one peptide epitope; and a second polypeptide comprising an effector selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, VPS72, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof, and an antibody to the peptide epitope.

[0221] Clause 2. The Cas effector of clause 1, wherein the effector is selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, CITED2, PHF15, SS18L1, MLLT6, ASH2L, and GSK3A, or a combination thereof.

[0222] Clause 3. The Cas effector of clause 1 or 2, wherein the effector is capable of increasing or decreasing expression of a gene.

[0223] Clause 4. The Cas effector of clause 3, wherein the effector reduces expression of a target gene and is selected from MCRS1, OTUD7B, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof.

[0224] Clause 5. The Cas effector of clause 3, wherein the effector increases expression of a target gene and is selected from RelB, LDB1, NFKBIB, CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, and VPS72, or a combination thereof.

[0225] Clause 6. The Cas effector of any one of clauses 1-5, wherein the first polypeptide comprises about 2 to about 50 peptide epitopes.

[0226] Clause 7. The Cas effector of any one of clauses 1-6, wherein the first polypeptide comprises more than one copy of the peptide epitope and further comprises at least one linker in between adjacent copies of the peptide epitope.

[0227] Clause 8. The Cas effector of any one of clauses 1-7, wherein the peptide epitope is GCN4 and comprises the amino acid sequence of SEQ ID NO: 85.

[0228] Clause 9. The Cas effector of any one of clauses 1-8, wherein the first polypeptide comprises at least one peptide epitope at the N-terminus and/or at the C-terminus of the Cas protein.

[0229] Clause 10. The Cas effector of any one of clauses 1-9, wherein the first polypeptide comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 87 or 89, or any fragment thereof, or wherein the first polypeptide comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 87 or 89, or any fragment thereof, or wherein the first polypeptide comprises the amino acid sequence of SEQ ID NO: 87 or 89.

[0230] Clause 11. The Cas effector of any one of clauses 1-10, wherein the antibody comprises the amino acid sequence of SEQ ID NO: 81.

[0231] Clause 12. The Cas effector of any one of clauses 1-11, wherein the second polypeptide comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to a sequence selected from SEQ ID NOs: 69, 71, 73, 75, 77, and 79, or any fragment thereof, or wherein the second polypeptide comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to a sequence selected from SEQ ID NOs: 69, 71, 73, 75, 77, and 79, or any fragment thereof, or wherein the second polypeptide comprises an amino acid sequence selected from SEQ ID NOs: 69, 71, 73, 75, 77, and 79.

[0232] Clause 13. A Cas fusion protein comprising two heterologous polypeptide domains, wherein the first polypeptide domain comprises a Cas protein, and wherein the second polypeptide domain comprises an effector selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, and CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, VPS72, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof.

[0233] Clause 14. The Cas fusion protein of clause 13, wherein the effector is selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, CITED2, PHF15, SS18L1, MLLT6, ASH2L, and GSK3A, or a combination thereof.

[0234] Clause 15. The Cas fusion protein of clause 13 or 14, wherein the effector is capable of increasing or decreasing expression of a gene.

[0235] Clause 16. The Cas fusion protein of clause 15, wherein the effector reduces expression of a target gene and is selected from MCRS1, OTUD7B, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof.

[0236] Clause 17. The Cas fusion protein of clause 15, wherein the effector increases expression of a target gene and is selected from RelB, LDB1, NFKBIB, CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, and VPS72, or a combination thereof.

[0237] Clause 18. The Cas fusion protein of any one of clauses 13-17, wherein the second polypeptide domain has transcription repression activity, transcription activation activity, de-ubiquitinase activity, p300 recruitment activity, enhancer looping mediation activity, or a combination thereof.

[0238] Clause 19. The Cas effector of any one of clauses 1-12 or the Cas fusion protein of any one of clauses 13-18, wherein the MCRS1 comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 57 or any fragment thereof, and/or wherein the MCRS1 comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 57, or any fragment thereof, and/or wherein the MCRS1 comprises the amino acid sequence of SEQ ID NO: 57, and/or wherein the MCRS1 is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO:

58, or any fragment thereof, and/or wherein the MCRS1 is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 58, or any fragment thereof, and/or wherein the MCRS1 is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 58.

[0239] Clause 20. The Cas effector of any one of clauses 1-12 or the Cas fusion protein of any one of clauses 13-18, wherein the OTUD7B comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to a sequence selected from SEQ ID NO: 59, amino acids 167-440 of SEQ ID NO: 59, or amino acids 792-831 of SEQ ID NO: 59, or any fragment thereof, and/or wherein the OTUD7B comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to a sequence selected from SEQ ID NO: 59, amino acids 167-440 of SEQ ID NO: 59, or amino acids 792-831 of SEQ ID NO: 59, or any fragment thereof, and/or wherein the OTUD7B comprises the amino acid sequence selected from SEQ ID NO: 59, amino acids 167-440 of SEQ ID NO: 59, or amino acids 792-831 of SEQ ID NO: 59, and/or wherein the OTUD7B is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 60, or any fragment thereof, and/or wherein the OTUD7B is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 60, or any fragment thereof, and/or wherein the OTUD7B is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 60.

[0240] Clause 21. The Cas effector of any one of clauses 1-12 or the Cas fusion protein of any one of clauses 13-18, wherein the RelB comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 65, or any fragment thereof, and/or wherein the RelB comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 65, or any fragment thereof, and/or wherein the RelB comprises the amino acid sequence of SEQ ID NO: 65, and/or wherein the RelB is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 66 or any fragment thereof, and/or wherein the RelB is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 66, or any fragment thereof, and/or wherein the RelB is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 66.

[0241] Clause 22. The Cas effector of any one of clauses 1-12 or the Cas fusion protein of any one of clauses 13-18, wherein the LDB1 comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 61, or any fragment thereof, and/or wherein the LDB1 comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 61, or any fragment thereof, and/or wherein the LDB1 comprises the amino acid sequence of SEQ ID NO: 61, and/or wherein the LDB1 is encoded by a polynucleotide

the ASH2L is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 104.

[0248] Clause 29. The Cas effector of any one of clauses 1-12 or the Cas fusion protein of any one of clauses 13-18, wherein the GSK3A comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 117, or any fragment thereof, and/or wherein the GSK3A comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 117, or any fragment thereof, and/or wherein the GSK3A comprises the amino acid sequence of SEQ ID NO: 117, and/or wherein the GSK3A is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 118, or any fragment thereof, and/or wherein the GSK3A is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 118, or any fragment thereof, and/or wherein the GSK3A is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 118.

[0249] Clause 30. The Cas effector of any one of clauses 1-12 or the Cas fusion protein of any one of clauses 13-18, wherein the effector is selected from BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, JAZF1, KAT7, KEAP1, MEAF6, MORF4L2, NFYC, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, TADA3, TAF6, TBPL1, VPS72, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, and wherein the effector comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to a sequence selected from SEQ ID NOS: 105, 107, 109, 111, 113, 115, 119, 121, 123, 125, 129, 131, 135, 137, 139, 141, 143, 145, 147, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, or 175, or any fragment thereof, and/or wherein the effector comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to a sequence selected from SEQ ID NOS: 105, 107, 109, 111, 113, 115, 119, 121, 123, 125, 129, 131, 135, 137, 139, 141, 143, 145, 147, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, or 175, or any fragment thereof, and/or wherein the effector comprises an amino acid sequence selected from SEQ ID NOS: 105, 107, 109, 111, 113, 115, 119, 121, 123, 125, 129, 131, 135, 137, 139, 141, 143, 145, 147, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, or 175, and/or wherein the effector is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to a sequence selected from SEQ ID NOS: 106, 108, 110, 112, 114, 116, 120, 122, 124, 126, 130, 132, 136, 138, 140, 142, 144, 146, 148, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, or 176, or any fragment thereof, and/or wherein the effector is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to a sequence selected from SEQ ID NOS: 106, 108, 110, 112, 114, 116, 120, 122, 124, 126, 130, 132, 136, 138, 140, 142, 144, 146, 148, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, or 176, or any fragment thereof, and/or wherein the effector is encoded by a polynucleotide comprising a sequence selected from SEQ ID NOS: 106, 108, 110, 112, 114, 116, 120, 122, 124, 126,

130, 132, 136, 138, 140, 142, 144, 146, 148, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, or 176.

[0250] Clause 31. The Cas effector of any one of clauses 1-12 and 19-31 or the Cas fusion protein of clause any one of clauses 13-31, wherein the Cas protein comprises at least one amino acid mutation that knocks out nuclease activity of the Cas protein.

[0251] Clause 32. The Cas effector or the Cas fusion protein of clause 31, wherein the at least one amino acid mutation is at least one of D10A and H840A.

[0252] Clause 33. The Cas effector of any one of clauses 1-12 and 19-32 or the Cas fusion protein of any one of clauses 13-32, wherein the Cas protein comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to one of SEQ ID NOS: 26-29, or any fragment thereof, or wherein the Cas protein comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to one of SEQ ID NOS: 26-29, or any fragment thereof, or wherein the Cas protein comprises the amino acid sequence of one of SEQ ID NOS: 26-29.

[0253] Clause 34. The Cas effector of any one of clauses 1-12 and 19-33 or the Cas fusion protein of any one of clauses 13-33, wherein the Cas protein is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to one of SEQ ID NOS: 30-31, or any fragment thereof, or wherein the Cas protein is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to one of SEQ ID NOS: 30-31, or any fragment thereof, or wherein the Cas protein is encoded by a polynucleotide comprising the sequence of one of SEQ ID NOS: 30-31.

[0254] Clause 35. A DNA targeting composition comprising: the Cas effector of any one of clauses 1-12 and 19-34 or the Cas fusion protein of any one of clauses 13-34; and at least one guide RNA (gRNA) that targets the Cas protein to a target region of a target gene.

[0255] Clause 36. The DNA targeting composition of clause 35, wherein the gRNA targets the Cas protein to target region selected from a non-open chromatin region, an open chromatin region, a transcribed region of the target gene, a region upstream of a transcription start site of the target gene, a regulatory element of the target gene, an intron of the target gene, or an exon of the target gene.

[0256] Clause 37. The DNA targeting composition of clause 35 or 36, wherein the gRNA targets the Cas protein to a promoter of the target gene.

[0257] Clause 38. The DNA targeting composition of any one of clauses 35-37, wherein the target region is located between about 1 to about 1000 base pairs upstream of a transcription start site of the target gene.

[0258] Clause 39. The DNA targeting composition of any one of clauses 35-38, wherein the at least one gRNA comprises a sequence selected from SEQ ID NOS: 96-98 and 101-102, or wherein the at least one gRNA is encoded by a polynucleotide comprising a sequence selected from SEQ ID NOS: 93-95 and 99-100, or wherein the at least one gRNA targets and binds a polynucleotide comprising a sequence selected from SEQ ID NOS: 93-95 and 99-100 or a complement thereof, or a combination thereof.

[0259] Clause 40. The DNA targeting composition of any one of clauses 35-39, wherein the DNA targeting composition comprises two or more gRNAs, each gRNA binding to a different target region.

[0260] Clause 41. An isolated polynucleotide sequence encoding the Cas effector of any one of clauses 1-12 and 19-34 or the Cas fusion protein of any one of clauses 13-34, or the DNA targeting composition of any one of clauses 35-40.

[0261] Clause 42. A vector comprising: the isolated polynucleotide sequence of clause 41.

[0262] Clause 43. The vector of clause 42, wherein the vector is an adeno-associated virus (AAV) vector.

[0263] Clause 44. A cell comprising: the Cas effector of any one of clauses 1-12 and 19-34 or the Cas fusion protein of any one of clauses 13-34, or the DNA targeting composition of any one of clauses 35-40, or the isolated polynucleotide sequence of clause 41, or the vector of clause 42 or 43, or a combination thereof.

[0264] Clause 45. A pharmaceutical composition comprising: the Cas effector of any one of clauses 1-12 and 19-34 or the Cas fusion protein of any one of clauses 13-34, or the DNA targeting composition of any one of clauses 35-40, or the isolated polynucleotide sequence of clause 41, or the vector of clause 42 or 43, or a combination thereof.

[0265] Clause 46. A method of modulating expression of a gene in a cell or in a subject, the method comprising administering to the cell or the subject the DNA targeting composition of any one of clauses 35-40, or the isolated polynucleotide sequence of clause 41, or the vector of clause 42 or 43, or the pharmaceutical composition of clause 45, or a combination thereof.

[0266] Clause 47. A method of modulating expression of a gene in a cell or in a subject, the method comprising administering to the cell or the subject an effector selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, VPS72, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof, or a polynucleotide encoding the effector.

[0267] Clause 48. The method of clause 47, wherein the effector is targeted to the gene.

[0268] Clause 49. The method of clause 47 or 48, wherein the effector is selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, CITED2, PHF15, SS18L1, MLLT6, ASH2L, and GSK3A, or a combination thereof.

[0269] Clause 50. The method of any one of clauses 47-49, wherein the effector is capable of increasing or decreasing expression of the gene.

[0270] Clause 51. The method of clause 50, wherein the effector reduces expression of the gene and is selected from MCRS1, OTUD7B, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof.

[0271] Clause 52. The method of clause 50, wherein the effector increases expression of the gene and is selected from RelB, LDB1, NFKBIB, CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, and VPS72, or a combination thereof.

[0272] Clause 53. The method of any one of clauses 46-50 and 52, wherein the expression of the gene is increased relative to a control.

[0273] Clause 54. The method of any one of clauses 46-51, wherein the expression of the gene is decreased relative to a control.

[0274] Clause 55. The method of any one of clauses 46-54, wherein the gene comprises the dystrophin gene, the CD25 gene, the B2M gene, or the TRAC gene.

[0275] Clause 56. The method of any one of clauses 46-55, wherein the cell is a muscle cell or a T cell.

[0276] Clause 57. A method of treating a disease in a subject, the method comprising administering to the subject the DNA targeting composition of any one of clauses 35-40, or the isolated polynucleotide sequence of clause 41, or the vector of clause 42 or 43, or the cell of clause 44, or the pharmaceutical composition of clause 45, or a combination thereof.

[0277] Clause 58. A method of treating a disease in a subject, the method comprising administering to the subject an effector selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, VPS72, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof, or a polynucleotide encoding the effector.

[0278] Clause 59. The method of clause 58, wherein the effector is targeted to a gene.

[0279] Clause 60. The method of any one of clauses 46-59, wherein the method treats a disease selected from Duchenne muscular dystrophy (DMD), Becker muscular dystrophy (BMD), and cancer.

SEQUENCES

SEQ ID NO: 1
NRG (R = A or G; N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 2
NGG (N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 3
NAG (N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 4
NGGNG (N can be any nucleotide residue, e.g., any of A, G, C, or T)

-continued

SEQUENCES

SEQ ID NO: 5

NNAGAAW (W = A or T; N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 6

NAAR (R = A or G; N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 7

NNGRR (R = A or G; N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 8

NNGRRN (R = A or G; N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 9

NNGRRT (R = A or G; N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 10

NNNGRRV (R = A or G; N can be any nucleotide residue, e.g., any of A, G, C, or T; V = A or C or G)

SEQ ID NO: 11

NNNNNGATT (N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 12

NNNNNGNNN (N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 13

NGA (N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 14

NNNRRT (R = A or G; N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 15

ATTCCT

SEQ ID NO: 16

NGAN (N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 17

NGNG (N can be any nucleotide residue, e.g., any of A, G, C, or T)

SEQ ID NO: 18

DNA sequence of the gRNA constant region

gtttaagagctatgctggaaacagcatagcaagttaaataaggctagtcggtatcaacttgaaaaaa
gtggcaccggactcggtgc

SEQ ID NO: 19

RNA sequence of the gRNA constant region

guuuuaagagcuaugcuggaaacagcauagcaaguuaauaaggcuaguuccguuaucacuugaaaaaa
guggcacccgagucggugc

SEQ ID NO: 20

SV40 NLS (Pro-Lys-Lys-Ser-Arg-Lys-Val)

SEQ ID NO: 21

GS linker (Gly-Gly-Gly-Gly-Ser)_n, wherein n is an integer between 0 and 10

SEQ ID NO: 22

Gly-Gly-Gly-Gly-Gly

SEQ ID NO: 23

Gly-Gly-Ala-Gly-Gly

SEQ ID NO: 24

Gly-Gly-Gly-Gly-Ser-Ser-Ser

SEQ ID NO: 25

Gly-Gly-Gly-Gly-Ala-Ala-Ala

SEQ ID NO: 26

Streptococcus pyogenes Cas9

MDKYSIGLDISTNSVGAWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAETRLKRTA
RRRYTRRKNRICYLQEIFSNEAKVDDSFHRLLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY
HLRKKLVDSTDKAIDLRLIYLALAHMIKFRGHFLIEGDLNPNDSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSGLGLTPNFKSNEDEAKLQLSKDTYD

- continued

SEQUENCES

```
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFPDQSNGYAGYIDGGASQEEFYKFKIIPILEKMDGTEELLVVLNRREDLLRKORTEDNG
SIPHQIHLGELHAILRQRQEDFYPFLKDNRREKI EKILTFRIPIYYVGPLARGNSRFAMTRKSEETITPW
NFEVVVDKGASAQSFIGERMTNDKNLPEVKLPKHSLLYEFVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKKIKDKDFLDNEEN
EFILEDVLTLTLEFREMIEERLKTYAHLFDKVMQKLKRRTYTGWGRLSRKLINGIRDQKSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAKVSGQGDSLHEHIANLAGSPAIIKGILQTVVVDELV
KVMGRHKPENIVIEMARENTTQKGQKNSREMKRIEEGT KELGSQILKEHPVENTQLNEKLYYLY
QNGRDMYVQDQELEDINRLSDYDVDHIVPQSFLKDDSDINDKVLTRSDKNRGKSDNVPSSEEVVKMKNYWR
QLLNNAKLITORKFDNLTKAERGGLSELDKAGFIKROLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSCLVSDFRKDQFYKVREINNYHHADAYLNAVVTALIKKPKLESEFVYGDYKVDVRK
MIAKSEQIGKATAKYFYSNMNFFKTEITLANGEIRKRPLIETGETGEIIVWDKGDRPATVRKVL
MPQVNIVKKTETVQTGGFSKESIPLPKRNSDKLIAKCKDWPKYYGGFDSPVAYSVLVVAKEKGSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKMLASAGELOKGN
ELALPSKYVNFLYASHYEKLKGSPEDNEQKOLFVEQHKHYLDEIEQISEFSKRVILADANLDKVL
AYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLHQISITGLYETRI
DLSQLGGD
```

SEQ ID NO: 27

Staphylococcus aureus Cas9

```
MKRNYIILGLDIGITSVGYGIIDYETDRVIDAVGVLKFKEANVENNEGRRSKRGARRLKRRRRHRIQRVK
KLLFDYNNLLTDHSELSGINPYEARVKGLSQKLSEEFSAAALLHLAKRRGVHNVNEVEEDTGNEELSTKE
QISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSVDYKEAKQQLLKVKAYHOLDQSFIDTYIDL
LETTRTYEGPEGSPGFWKDIKEWYEMLMGHTYPEELRSVKYAYNADLYNALNDLNNLVITRDEN
EKLEYYBKFQJIEVNPQKQKKPTLKQIAKEILVNEEDIKGYRVTGKPEFTNLKVYHDIKDITARKE
IIENAELLDQIAKLTIQSSEDIQEELTNLNSELTQEEETQIISNLKGYGTTHNLSSLKAUNLILDELW
HTNDNQIAIFNRNLKLPVKVQDLSQQKEIPTTLVDDPILSPVVVKRSFIQSIVKINAIKKYGLPNID
ELAREKNSKDAQKMINEMQKRNQRTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLE
DILNNPPNVEVDHIIPRSVSFDSPMNKVLVQEEENSKKGNRTPQYLSSSDSKISYETFKKHLNLA
KGKGRISKTKKEYLLERDINFPSVQKDFINRNLVDTRYATRGLMNLRSYFRVNLDVKVKSINGGF
TSFLRRWKWFKKERNKGYKHHADALIIANADIFKEWKLDKAKKVMENQMFEEKQAESMPETEQ
EYKEIFITPHQIKHIDFKDYKSHRVDDKPNRELINDTLYSTRKDKGNLTIVNNLNGLYDKDNDKL
KLINKNSPEKRNKVKLSPYRFDVYLDNGVYKFTVKNLDVVIKENYYEVNSKCYEAKK
NKLNAHLDITDDYPSNSRNKVKLSPYRFDVYLDNGVYKFTVKNLDVVIKENYYEVNSKCYEAKK
LKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLNR1EVNMIDI TYREYLEMMNDKRPPRIKTI
ASKTQSICKYSTDILGNLYEVSKKKHPOIIKKG
```

SEQ ID NO: 28

Streptococcus pyogenes Cas9 (with D10A)

```
MDKYSIGLAIGTNSVGWAVITDEYKVPSSKKFVLGNTDRHSIKKNLIGALLFDGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEAKVDDSSFFHRLLEESLVEEDKKHERHPIFGNIVDEVAYHEKYTIY
HLRKKLVDSTDKDADRLLIYLAQKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAIISARLSKSRRLLENLIAQLGKNGLFGNLLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFPDQSNGYAGYIDGGASQEEFYKFKIIPILEKMDGTEELLVVLNRREDLLRKORTEDNG
SIPHQIHLGELHAILRQRQEDFYPFLKDNRREKI EKILTFRIPIYYVGPLARGNSRFAMTRKSEETITPW
NFEVVVDKGASAQSFIGERMTNDKNLPEVKLPKHSLLYEFVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKKIKDKDFLDNEEN
EFILEDVLTLTLEFREMIEERLKTYAHLFDKVMQKLKRRTYTGWGRLSRKLINGIRDQKSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAKVSGQGDSLHEHIANLAGSPAIIKGILQTVVVDELV
KVMGRHKPENIVIEMARENTTQKGQKNSREMKRIEEGT KELGSQILKEHPVENTQLNEKLYYLY
QNGRDMYVQDQELEDINRLSDYDVDHIVPQSFLKDDSDINDKVLTRSDKNRGKSDNVPSSEEVVKMKNYWR
QLLNNAKLITORKFDNLTKAERGGLSELDKAGFIKROLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSCLVSDFRKDQFYKVREINNYHHADAYLNAVVTALIKKPKLESEFVYGDYKVDVRK
MIAKSEQIGKATAKYFYSNMNFFKTEITLANGEIRKRPLIETGETGEIIVWDKGDRPATVRKVL
MPQVNIVKKTETVQTGGFSKESIPLPKRNSDKLIAKCKDWPKYYGGFDSPVAYSVLVVAKEKGSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKMLASAGELOKGN
ELALPSKYVNFLYASHYEKLKGSPEDNEQKOLFVEQHKHYLDEIEQISEFSKRVILADANLDKVL
AYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLHQISITGLYETRI
DLSQLGGD
```

SEQ ID NO: 29

Streptococcus pyogenes Cas9 (with D10A, H840A)

```
MDKYSIGLAIGTNSVGWAVITDEYKVPSSKKFVLGNTDRHSIKKNLIGALLFDGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEAKVDDSSFFHRLLEESLVEEDKKHERHPIFGNIVDEVAYHEKYTIY
HLRKKLVDSTDKDADRLLIYLAQKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAIISARLSKSRRLLENLIAQLGKNGLFGNLLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFPDQSNGYAGYIDGGASQEEFYKFKIIPILEKMDGTEELLVVLNRREDLLRKORTEDNG
SIPHQIHLGELHAILRQRQEDFYPFLKDNRREKI EKILTFRIPIYYVGPLARGNSRFAMTRKSEETITPW
NFEVVVDKGASAQSFIGERMTNDKNLPEVKLPKHSLLYEFVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKKIKDKDFLDNEEN
EFILEDVLTLTLEFREMIEERLKTYAHLFDKVMQKLKRRTYTGWGRLSRKLINGIRDQKSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAKVSGQGDSLHEHIANLAGSPAIIKGILQTVVVDELV
KVMGRHKPENIVIEMARENTTQKGQKNSREMKRIEEGT KELGSQILKEHPVENTQLNEKLYYLY
QNGRDMYVQDQELEDINRLSDYDVDHIVPQSFLKDDSDINDKVLTRSDKNRGKSDNVPSSEEVVKMKNYWR
QLLNNAKLITORKFDNLTKAERGGLSELDKAGFIKROLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSCLVSDFRKDQFYKVREINNYHHADAYLNAVVTALIKKPKLESEFVYGDYKVDVRK
MIAKSEQIGKATAKYFYSNMNFFKTEITLANGEIRKRPLIETGETGEIIVWDKGDRPATVRKVL
MPQVNIVKKTETVQTGGFSKESIPLPKRNSDKLIAKCKDWPKYYGGFDSPVAYSVLVVAKEKGSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKMLASAGELOKGN
ELALPSKYVNFLYASHYEKLKGSPEDNEQKOLFVEQHKHYLDEIEQISEFSKRVILADANLDKVL
AYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLHQISITGLYETRI
DLSQLGGD
```

-continued

SEQUENCES

KVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQLKEHPVENTQLQNEKLYLYYL
 QNGRDMYVDQEELINRLSDYDVAIVPQSFLKDDSDNKLITRSKDRNGKSNDVPSSEEVVKMKNYWR
 QLLNAKLTQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDLSRMNTKYDENDKLIRE
 VKVITLKSCLVSDFRKDQFYKREINNNYHHADAYLNAUVGTALIKKYPKLESFVYGDYKVYDVRK
 MIAKSEQEIGKATAKYFFYSNIMNNFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
 MPQVNIVVKKTEVOTGGFSKESILPKRNSDKLIARKKDWDPKYGGFDSPVAYSVLVVAKVEKGSKK
 LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIILPKYSLFELENGRKMLASAGELQKGN
 ELALPSKVNVFLASHYEKLGSPEDNEQKOLFVBQHKhYLDEIIEQISEFSKRVILADANLDKVL
 AYNKHRDKPIREAENI1HLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAHLIQSITGLEYTRI
 DLSQLGGD

SEQ ID NO: 30

Polynucleotide	sequence of D10A mutant of <i>S. aureus</i> Cas9				
atgaaaaaggaa	actacattctt	ggggctggcc	atcggttata	caagcgtggd	gtatgggatt
attgactatgt	aaacaaggga	cgtgatcgac	gcaggcgctca	gactgttcaa	ggaggccaac
gtggaaaaca	atgaggggacg	gagaagcaag	agggggagcc	ggccctgtaa	acgacggaga
aggcacagaa	tccagagggt	gaaaaactg	ctgttcgatt	acaacctgtct	gaccgaccat
tctgagctga	gtggaaattaa	tccttatgaa	gccagggtga	aaggccttag	tcagaagctg
tcagaggaaag	agttttccgc	agctctgtcg	cacctggcta	agcgcggagg	agtgcataac
gtcaatgagg	tggaaaggaa	cacccggcaac	gagctgtcta	caaaggaaaca	gatctcaacgc
aatagcaaaag	ctctggaaaga	gaagtatgtc	gcagacgtgc	agctggaaacg	gctgaagaaa
gtatggcagg	tgagagggtc	aattaatagg	ttcaagacaa	gcgactacgt	caaagaagcc
aagcagctgc	tggaaatgtca	gaaggcttac	caccagctgg	atcagagctt	catcgatact
tatatcgacc	tgctggagac	tccggagaacc	tactatgagg	gaccaggaga	aggagacccc
ttcggatggaa	aagacatcaa	ggaaatgtgtac	gagatgtcgta	tgggacattt	cacctatttt
ccagaagagc	tgagaaggcgt	caagtacgtc	tataacgcag	atctgtacaa	cgcctgaaat
gacactgaaaca	acctgttcat	cacccaggat	gaaaacggaga	aactgtgaaata	ctatgagaag
ttccagatca	tgcggaaacgt	gtttaagcag	aagaaaacgc	ctacactgaa	acagattgtct
aaggagatcc	tggtaacgca	agaggacatc	aagggttacc	gggtgacaag	cactggaaaa
ccagaggctca	ccaaatgtaa	agtgtatcac	gatattaagg	acatcacagc	acggaaagaa
atcattgaga	acgcccgaact	gtctggatcag	attgtcaaga	tcctgtactat	ctaccagagc
tccgaggaca	tccaggaaga	gtgtactaac	ctgaacacgc	agctgaccca	ggaagagatc
gaacagatcca	gtaaatctgaa	gggggtacacc	ggaacacaca	acctgtccct	gaaagctatc
aatctgatcc	tggatgtact	gtggcataca	aacgacatc	agattgtcaat	ctttaaccgg
ctgaagctgg	tccccaaaaa	gggtggacctg	agtcaacgca	aagagatccc	aaccacactg
gtggacgatt	tcatttctgtc	accctgggtc	aagcggagct	tcatccagag	catcaaagtg
atcaacgcoca	tcatacaagaa	gtacggctcg	cccaatgata	tcattatcgta	gctggcttag
gagaagaaca	gcaaggacgc	acagaagatg	atcaatgaga	tgcagaaacg	aaaccggcag
accaatqaac	gcatttgaaga	gattatccga	actaccggga	aagagaacgc	aaagtacctg
attgaaaaaa	tcaagctgca	cgatatgcag	gagggaaagt	gtctgttattc	tctggaggcc
atccccctgg	aggacgtgtc	gacacaatcca	ttcaactacg	aggtcgatca	tattatcccc
agaagcgtgt	ctttcgcacaa	ttcttttaac	aacaagggtc	tgtcaagca	ggaagagaac
tctaaaaaagg	gcaataggac	tctttttccag	tacctgtcta	gttcagattc	caagatctt
tacgaaacct	ttaaaaagca	cattctgaat	ctggccaaag	gaaagggccg	catcagcaag
acccaaaaaa	agtacccgtc	ggaaagacggg	gacatcaaca	gatttccgt	ccagaaggat
tttattaaacc	ggaatcttgtt	ggacaaagaga	tacgtactc	ggggctgtat	gaatctgtcg
cgtatccatt	tccgggtgaa	caatctggat	gtgaaatgtca	agtccatcaa	cgggggttc
acatctttc	tgaggcgccaa	atggaagttt	aaaaaggagc	gcaacaaagg	gtacaagcac
catggcaag	atgctctgtat	tatcgcaat	gccgacttca	tcttaagga	gtggaaaaaa
ctggacaaaag	ccaagaaagt	gatggagaac	cagatgtcg	aagagaagca	ggccgaatct
atgcccggaa	tccgagacaga	acaggatgtc	aaggagattt	tcatcactcc	tcaccagatc
aagcatatca	aggatttcaa	ggactacaag	tactctcacc	gggtggataa	aaagccaaac
agagagctgt	tcaatgacac	cctgtatagt	acaagaaaag	acgataagg	gaataccctg
atttgtaaaca	atctgaacgg	actgtatcgc	aaagataatg	acaagctgaa	aaagctgatc
aacaaaatgc	ccggagaact	gtctgtatgc	caccatgtc	ctcagacata	tcagaaactg
aagctgatta	tggagcgtta	cgccgacgag	aagaaccac	tgataaagta	ctatgaagag
actggaaact	acctgtacca	gtatagcaaa	aaggataatg	gccccgtat	caagaagatc
aagtactatg	ggaacaagct	gaatggccat	ctggacatca	cagacgatta	ccctaacagt
cgcacaacaagg	tggtcaagct	gtcactgtcg	ccatcatgat	tctggacaaac	
ggcgtgtata	aattttgtac	tgtcaagaaat	ctggatgtca	tcaaaaagga	gaactactat
gaagtgaata	gcaagtgtca	cgaagaggct	aaaaagctga	aaaagattag	caaccaggca
gagttcatcg	cctccattta	caacaacgc	ctgatataa	tcaatggcga	actgtatagg
gtcatcgcccc	tgaacaatga	tctgtgcac	cgcatgttgc	tgaatatgtat	tgacatcact
tacccgagatc	atctggaaaa	catgtatgt	aagcgcccc	ctcgaaattat	caaaaactt
gcctctaaga	ctcagagttat	caaaaagttac	tcaacogaca	ttctggaaa	cctgtatgag
gtgaagagca	aaaagccccc	tcagattatc	aaaaaggcc		

SEQ ID NO: 31

Polynucleotide	sequence of N580A mutant of <i>S. aureus</i> Cas9				
atgaaaaaggaa	actacattctt	ggggctggac	atcggttata	caagcgtggd	gtatgggatt
attgactatgt	aaacaaggga	cgtgatcgac	gcaggcgctca	gactgttcaa	ggaggccaac
gtggaaaaca	atgaggggacg	gagaagcaag	agggggagcc	ggccctgtaa	acgacggaga
aggcacagaa	tccagagggt	gaaaaactg	ctgttcgatt	acaacctgtct	gaccgaccat
tctgagctga	gtggaaattaa	tccttatgaa	gccagggtga	aaggccttag	tcagaagctg

-continued

SEQUENCES					
tca	gaggaa	agtttccgc	agctctgt	cac	tcggagg
gt	caatgagg	tggaaagga	caccggcaac	gag	ctgtcta
aat	agcaaa	ctctggaaa	gaagtatgtc	ca	aaaggaca
gat	ggcgagg	tgagagggtc	aattaatagg	g	actggaaac
aag	caggtgc	tgaaagtgtca	gaagggtac	cg	actgtacgt
tat	atcgacc	tgctggagac	tcggagaacc	act	caagactt
ttc	ggatgga	aagacatcaa	ggaatgtac	gag	atcgatcc
cc	agaagacg	tgagaacgt	caagtacgt	ata	actgtacaa
gac	ctgaaaca	acctgtcat	caccaggat	ac	atctgtacaa
ttc	caggatca	tcgaaaacgt	gtttaacgag	aa	actgtacaa
aagg	agatcc	tggtaacgaa	agaggacatc	aaa	ctacactgaa
cc	agatcca	ccaatctgaa	agtgtatcac	aaa	acatcacagc
atc	attgaga	acggcaact	gttggatcg	at	tctgtactat
tcc	gaggaca	tccaggaa	gtctactac	ctg	ctaccagagc
ga	acagatta	gtaatctgaa	gggttacacc	aa	ggttaccc
aat	ctgtatc	tggatgact	gttgcataca	act	gaaagctatc
ct	taagctgg	tccaaaaaaa	gggttgcctg	gtt	ctttaacccg
gt	tgacgtt	tcattctgtc	accctgtgtc	tc	aaccacactg
atc	aacaccca	tcatcaagaa	gtacggcctg	at	catcaaagtg
gaga	aaaca	gcaaggacgc	acagaagatg	ca	gctggctagg
acc	atgaaac	gcatttgaaga	gattatccga	at	gaaagctatc
at	tttttttttt	tcaagctgtca	cgatgtcg	ca	aaacccggag
at	ccccttgg	aggacctgt	gacaatcca	gg	aaagtacctg
aga	aaacgtgt	ccttcgacaa	ttcccttaac	gg	tctggaggcc
tct	aaaaagg	gcaatagac	tccttcccg	gt	ggaagaggcc
tac	ggaaac	ttaaaaagca	cattctgaat	tc	caagatctt
ac	aaaaagg	agtacctgt	ggaaagagcgg	gg	tttttttttt
tt	tatttaacc	ggaatctgtt	ggacacaaga	gg	ccagaaggat
cg	atcttatt	tccgggtgaa	caatctgtt	gg	gaatctgt
ac	atctttt	tgaggcgcac	atggaaagt	gg	cggegggtt
cat	gcccgaag	atgctctgt	tatcgcaat	gg	gtacaagcac
ct	ggacaaag	ccaagaaagt	gatggagaac	gg	gtggaaaaag
at	ggccgaa	ttcggatgtt	cgatgttcg	gg	ggccgaatct
a	agcatatca	aggatcttac	aaggagattt	gg	tcaccagatc
ag	agagactg	tcaatgtac	cctgtatagt	gg	aaagcccaac
at	tgtgaaca	atctgaacgg	actgtacgc	gg	gaataccctg
aa	aaaaagt	ccggagaact	gtgtatgtac	gg	aaagctgatc
aa	gcttggata	tggagacta	cggcgacg	gg	tcagaaactg
act	ggggact	acctgtacca	gtatagcaaa	gg	ctatgaagag
aa	gtactatg	ggaacaacgt	gaatgtccat	gg	caagaagatc
cg	caacaa	tggtaacgt	gtcaactgt	gg	ccctaacagt
gg	ctgttata	aatttgcac	tgtcaagaat	gg	tctggacaac
ga	aatgtata	gcaagtgtca	cgaaagggtc	gg	gaactactat
ga	gttcatcg	cctcctttta	caacaacgc	gg	caaccaggca
gt	atcgatggg	tgaacaatga	tctgtgtac	gg	actgtatagg
tac	ccgagat	atctggaaaa	catgtatgt	gg	tgacatact
gc	ctctaaaga	ctcagatgt	caaaaatgtac	gg	caaaaacatt
gt	gaagagca	aaaagcaccc	tcagattatc	gg	cctgtatgag
SEQ ID NO: 32					
codon optimized polynucleotide encoding <i>S. pyogenes Cas9</i>					
atggataaaa	agtacagcat	cgggctggac	atcggtacaa	actcagtgg	gtggccgtg
attacggacg	agtacaagg	accctccaaa	aaattnaaag	tgcgtggtaa	cacggacaga
cactctataa	agaaaaatct	tattggagcc	ttgcgtgtcg	actcaggcga	gacagccgaa
gccacaagg	tgaagcgac	cggccaggagg	cggtatacca	ggagaaagaaa	ccgcatatgc
tacctgtcaag	aaatcttcag	taacgagatg	gcaaaagggt	acgatagctt	tttccatcg
ctggaagaat	ccttttgtt	tgaggaaagac	aagaaggacg	aacggcacc	cattttggc
aatatttgc	acgaagtggc	atatcagcaaa	aagtaccgca	ctatctacca	cctcaggaaag
aa	gctgtgtgg	acttcacca	taaggcgac	tttatttggc	actcgcccc
atgattaaat	ttagaggaca	tttcttgatc	gagggcgacc	tgaacccgga	caacagtgt
gtcgataagc	tgttcatca	acttgtcgac	acctaatac	aactgttca	agaaaaaccc
ataatgttt	caggatgtca	cgctaaagca	atccgttccg	cgcgcctctc	aaaatctaga
agacttgaga	atctgtatgc	tcagttgtcc	ggggaaaaga	aaaatggatt	gtttggcaac
ctgatcgccc	tcagtcctgg	actgacccca	aatttcaaaa	gtaaatttcg	cctggccgaa
gacgctaagc	tccagtcgtc	caaggacaca	tacgtatgc	actcgcacaa	tctgtgtcc
cagattgggg	atcagatgtc	cgatctttt	ttggcagcaa	agaacctgtc	cgacgcac
ctgttgcgc	atatcttgc	agtgaacacc	gaaattacta	aagcaccct	tagcgcatt
atgatcaagc	ggtacgcacg	gatcatcg	gatctgaccc	tgcgtggac	tcttggagg
caacagtc	ccgaaaaata	caaggaaatc	ttctttgacc	agagaaaaaa	cggtacgt
ggctatata	atggtggggc	cagtcaggag	gaattctata	aatttcatca	gcccatttctc
gagaaaatgg	acggcacaga	ggagttgtc	gtcaaactt	acaggagga	cctgtgtccg
aagcagcgaa	cctttgacaa	cgggtctatc	ccccaccaga	ttcatctgg	cgaactgcac
gcaatcctga	ggaggcgac	ggattttat	ccttttctt	aagataaccg	cgagaaaata
aaaaagat	ttacattcg	gatcccgatc	tacgtggac	ctctcgcccc	ggcaattca

-continued

SEQUENCES					
cggtttgcct	ggatgacaag	gaagtcaag	gagactatta	caccttggaa	cttcgaagaa
gttgtggaca	agggtgcattc	tgcgcgcgtt	ttcatcgagc	ggatgacaaa	ttttgacaag
aacctcccta	atgagaagg	gctgccc	cattctctgc	tctacgagta	ctttaccgtc
tacaatgaaac	tgactaaatg	caagtacgtc	accgaggaa	tgaggaaagcc	ggcattcctt
agtggagaac	agaagaaggc	gattgttagac	ctgttgtca	agaccaacag	gaaggtgact
gtgaagcaac	ttaaagaaga	ctactttaag	aagatcgaa	gttttgacag	tgtgaaatt
tcaggggttg	aagaccgctt	caatgcgtca	ttggggactt	accatgatct	tctcaagatc
ataaaggaca	aagacttctt	ggacaacgaa	gaaaatgggg	atattctcg	agacatcg
ctcacctctg	ccctgttgc	agacaggaa	atgatagaag	agcgttgc	aacctatgoc
caccccttcg	acgataaaatg	tatgaagcag	ctgaagcgc	ggagatacac	aggatgggaa
agattgtcaa	ggaagctgtat	caatggaaatt	agggataaac	agagtggcaa	gaccataactg
gatttcttca	aatctgtatgg	cttcgc	aggacttca	tgcactgtat	tcacgtgac
tctcttacct	tcaaggaggaa	cattcaaaag	gctcagggtga	gcccgcagg	agactccctt
catgaacaca	tcgcqattt	ggcagggttcc	cccgtattt	aaaaggcat	ccttcaact
gtcaagggtgg	tggatgaaatt	ggtcaaggta	atggggagac	ataagccaga	aaatattgtg
atcgagatgg	cccgcgaaaa	ccagaccaca	cagaaggcc	aaaaaaatag	tagagagccg
atgaagagga	tcgaggagg	catcaaaag	ctgggatctc	agattctcaa	agaacacccc
gtgaaaaca	cacagctgc	gaacggaaaa	ttgtacttgc	actatctgc	gaacggcaga
gacatgtacg	tcgaccaaga	acttgcattt	aatagactgt	ccgactatga	cgtagaccat
atcggtcccc	agtcccttc	gaaggacgc	tccattgata	acaaagtctt	gacaagaagc
gacaagaaca	ggggtaaaatg	tgataatgt	cctagcagg	aggtgggtaa	aaaaaaatgag
aactactggc	gacagctgt	taatgc	ctcattacac	aacggaaat	cgataatctg
acgaaagcag	agagagggtg	cttgc	ttggacaagg	cagggtttat	taagccgcag
ctgggtggaaa	ctaggcagat	cacaaggac	gtggcgc	tttggacag	ccggatgaac
acaaaatacg	acgaaaatg	taaactgata	cgagggcgtca	aagtatc	gctgaaaagc
aagctgtgt	ccgattttcg	gaaagacttc	cagtcc	aagtgcgc	gattaataac
taccatcatg	ctcacatgc	gtac	gctgttgc	ggaccgc	gataaagaag
tacccaagc	tggatccg	gttcgtatac	ggggat	aagtgtac	tgtgaggaaa
atgatgacoca	agtccgc	ggagatgg	aaggccacag	ctaagtactt	cttttatct
aacatcatg	attttttt	gacggaaattt	accctggc	acggagagat	cagaagccg
cccttatag	agacaatgg	tggaaacagg	gaaatcgct	gggataagg	cagggat
gctactgtga	ggaaggtgc	gagtatgcca	caggtaata	tcgtgaaaa	aaccgaagta
cajacccgg	gattttccaa	ggggaaat	ttgcctaaa	gaaactcaga	caagtcatac
gccccgcaaga	aagatggg	ccctaa	tacggggat	ttgactc	caccgtagcc
tattctgtgc	tgggttgc	taagttgg	aaaggaaat	ctaqaagct	gaatccgt
aaggaactct	tgggaatc	tatcatgg	agatcatct	ttgaaaagaa	ccctatcgat
ttctctgggg	ctaagggtt	caaggag	aagaaagacc	tcat	actgcacaaa
tactctct	tcgagctg	aaatgg	aagagaatgt	tgccgc	cggagactg
caaaaggaa	acgacttgc	tctgc	aaatgtt	atttctct	tctcgcttcc
cactatgaa	agctgaaagg	gtctccg	gataacg	agaagcag	gttgcgaa
cajcaca	actatcg	tggaaata	gaacaaata	gcgajttc	caaaagggtt
atctctgggg	atgctattt	ggacaaatg	ctgtctgc	ataacaagca	ccggataaag
cctttaggg	aacaagcc	gaatataatt	cacctt	cactcac	tctcgagcc
ccgcgcgcct	tcaaatactt	tgat	atcgac	aacggtatac	cagtaccaaa
gaggctctcg	atgccac	catccac	tcaattact	gcgtgtac	aacacggatc
gaccccttc	aactggcg	cgact			

SEQ ID NO: 33

codon	optimized	nucleic acid	sequences	encoding	S. aureus	Cas9
atgaaaagg	actacattt	ggggctggac	atcggttata	caagcg	gtatgggatt	
attgactatg	aaacaagg	cgtgatgc	gcaggcg	tcgttca	ggaggccaa	
gtggaaaaca	atgaggggac	gagaagcaag	aggggagcc	ggcgcctg	acgacggaga	
aggcagacaa	tccaggggt	gaagaaatgc	ctgttgcatt	acaacctgt	gaccgacat	
tctgagatgt	gtggat	tcttgcata	gcccagggt	aaggcctg	tcaagactg	
tcagggagg	agtttcc	agctctgt	cacctgc	acgcggc	agtgcataac	
gtcaatgagg	tggagagg	cacccg	gagctgt	caaaggaa	gatctcac	
aatagcaaa	ctctggaa	gaatgt	gcagactgc	agctg	gctgaaaagaa	
gatggcgagg	tgaggggtc	aattaaatgg	ttcaagacaa	ggactacgt	caaagaagcc	
aacgcgtc	tggaaatgc	gaaggcttac	caccagctt	atcagat	catcgata	
tatatcgacc	tgctgg	tccgaaacc	tactatg	gaccagg	agggagcccc	
ttcggatgg	aagacat	ggaatgg	gagatgt	tgggacatt	cacctat	
ccagaagac	tgagaacg	caagtc	tataacgc	atctgt	cgcctgaat	
gaccgtaca	acctgg	caccagg	gaaaacg	aactgg	ctatgaaag	
ttccatgtca	tcgaaaatgc	gttta	aaagaaatgc	ctacactg	acagattgt	
aaggagatcc	tggtca	agggacatc	aagggttacc	gggtg	cactggaaa	
ccagagg	ccaaatgc	agtgtatc	gatatta	acatcac	acggaaagaa	
atcattgaga	acgcgc	gtgttgc	attgt	tcctgtact	ctaccagac	
tccgaggaca	tccagg	gtgtacta	ctgaac	agctg	ggaagagatc	
gaacagat	gtatgc	gggttac	ggaacaca	actgt	gaaagctatc	
aatctgtatc	tggatgt	gtggcata	aacgacaatc	agattg	ctttaaccgg	
ctgaagctgg	tcccaaaaa	ggtggac	agtca	aagagat	aaccacactg	
gtggacgatt	tcattctgt	accgcgt	aaggcag	tcat	catcaaagt	
atcaacgc	tcatca	gtacgc	cccaat	tcat	gctggctagg	
gagaagaaca	gcaagg	acagaagat	atcaat	tgcagaa	aaacccggc	

-continued

SEQUENCES					
accaatgaac	gcattgaaga	gattatccga	actaccggga	aagagaacgc	aaagtacctg
attgaaaaaa	tcaagctgca	cгататгаг	gagggaагт	gtctgttattc	tctggaggccc
tccccctgg	aggacctgct	gaacaatcca	ttcaactacg	aggtcgatca	tattatcccc
агаагсгтгт	ccttcgacaa	ttcccttaac	aacaаггтгс	tggtcaагca	ggaагагааac
tctaaaaagg	gcaataggac	tccttccag	tacctgtcta	gttcагаттc	caагатctct
tacgaaaact	ttaaaaагca	cattctgaat	ctggccaaаг	gaaагggccg	catcагааг
accaaaaagg	agtacctgt	ggaагагсgg	gacatcaaca	gatttccgt	ccагаагагат
tttattaaac	ggaatctgt	ggacacaагa	tacgtactc	gcccctgtat	гаатctgtg
cgatcttac	tccgggtgaa	caatctggat	gtgaaагtca	агtccatcaa	cgggggttcc
acatctttc	tgaggcgeaa	atggaaагtt	aaaагggагc	gcaacaагgg	gtacaагаac
catgcccag	atgctctgt	tatcgaaaat	gccgacttca	tcttaагга	gtggaaaагa
ctggacaaag	ccaагаагt	gatggagaac	cагatgtcg	aагагаагca	ggccгаатt
atgcccggaa	tсgagacaga	acaggagatc	aaggagattt	tcatcactcc	tcaccагатc
aacgatatac	aggatttca	ggaactacaag	tacttccacc	gggtgatataa	aaагcccac
agagagctga	tcaatgacac	cctgtatagt	acaагааааг	acgataaggg	gaataccctg
attgtgaaca	atctgacccg	actgtacgac	aaагатаатg	acaагctgaa	aaаготgатc
aacaaaагt	ccggagaагt	gотgatgtac	caccatgtac	ctcагаacata	tcагаaaactg
aaactgtatta	tggagacaga	cggcggacgg	aагаacccac	tgatataагt	ctatgагаг
actggggact	acctgaccaa	gtatagcaaa	aaggataatg	gcccctgtat	caагаагатc
aагtactatg	ggaacaагct	gaatgcccac	ctggacatca	cагacgatтa	ccctaaacагt
cгоaacaагgg	tggtcaагct	gtcactgag	ccatcагat	tctgtcttca	tctggacaac
ggcgtgtata	aatttggtac	tgtcaагаat	ctggatgtca	tcaaaaагга	gaactactat
gaагtgaата	gcaагtгct	cgaaагggct	aaaагctga	aaaагattag	caaccагgca
gagttcatcg	cctccittt	caacaacgc	ctgattaагa	tcaatggcga	actgtataаг
gtcatcgggg	tgaacaatga	tctgtgaaac	cggcattgag	tgaatatgtat	tgacatcaat
taccggagat	atctggaaaa	catgaatgtat	aагcggcccc	ctcgaatttat	caaаacaatt
gcccctaaaga	ctcagagat	caaааагtac	tcaacggaca	ttctgggaaa	cctgtatgag
gtgaagagca	aaaагcacc	tcagattatc	aaaагggc		
SEQ ID NO: 34					
codon optimized nucleic acid	sequences encoding <i>S. aureus</i> Cas9				
atgaagcgga	actacatct	gggcctggac	atcggcatca	ccagcgtggg	ctacggcatc
atcgactacg	agacacggga	cgtgtatcgat	gcccggctg	ggctgttcaa	agaggccaaac
gtggaaaaca	acggggcag	gcccggacag	agaggccca	gaaggctgaa	gcccggggagg
cggcatagaa	tccagagat	gaaагаагct	ctgttccact	acaacctgt	gaccgaccac
агcгагctga	gcccgtatcaa	cccctacgag	gccagagtga	aggcctgtag	ccагаагctg
агcгагggaa	агttctctgc	cccccctgt	cacctggcca	агаагаагgg	cgtgcaacaaac
gtgaacgagg	tggaaагга	caccggcaac	gagctgttca	ccaaагага	gatcагccgg
aacagcaagg	ccctggaaаг	gaaatacgt	gccgaacttgc	agctggaaacg	gctgaaагaaa
gacggcgaag	tgcggggcag	catcaacaga	ttcaagacca	gcgactacgt	gaaагаагcc
aaacagctgc	tgaagggtgca	gaaggcttac	caccacgtt	accaагtct	catcагacacc
tacatcgacc	tgctggaaac	ccggcggacc	tactatggg	gaccctggca	ggggагcccc
ttcggctgga	aggacatcaa	agaatggtac	gagatgtct	tggggactt	caccatactt
cccgaggaaac	tgcggagct	gaагtacg	tacaacccg	actgtacaa	cgccctgaaac
gactctgaaca	atctgtat	caccggac	gagaacggaa	агtggaaata	ttacgагаag
ttccagatca	tгgagaагct	gttcaагаг	aагаагааг	ccacccctgaa	gсагатcгcc
aaagaaatcc	tсgtgaacga	agaggatatt	aагggotaca	gагtгaccag	cacggcгаag
cccgagttca	ccaacctgaa	ggtgtaccac	gacatcaagg	acattaccgc	ccggaaагаг
attattgaga	acgcccagct	gtctgtat	atttccaaаг	tctgtaccat	ctaccагаг
agcggaggaca	tccaggaaаг	actggaccaat	ctgaaactcc	агtggaccat	ggaагагатc
gaggatct	ctaatctgaa	gggtatacc	ggcaccacc	actgtgacc	gaaагccatc
aacctgtatcc	tggacggact	gttggacacc	aacgacaacc	агtgcgtat	cttcaaccgg
ctgaagctgg	tggccaaагa	ggtggacc	tcccacg	aагагатccc	caccacccctg
gtggacgact	tcatctctg	cccccctgt	aагагаагct	tcatccagag	catcaaагt
atcaacgcoca	tcatcaagaa	gtacggccct	cccaaccc	tcatatccg	gctggcccg
gagaагаact	ccaaggac	ccaaaагat	atcaacgaga	tгcagaагcg	gaaaggccag
accaacgagc	ggatcggaga	aatcatccgg	accacccg	аагагаacgc	caагtactg
atcgagaaga	tcaagctgca	cgacatcg	gaaggcагt	gctgtacag	cctggaaагcc
atccctctgg	aaatctgtat	gaaacacc	ttcaactat	агtggatgg	catcatcccc
агaагсгtгt	ccttcgacaa	cагttcaac	aacaагgt	tгtgtaaагca	ggaагaaaac
агcагаагgg	gcaacccgg	cccattcc	tacctgtac	gcaaggcаг	caагatcаг
tacgaaaact	tcaagaагca	catctgtat	ctggccaaаг	gcaaggccag	aatcагааг
accaагааг	агtactctgt	ggaагаacgg	gacatcaaca	ggttctccgt	gсагааагаг
ttcatcaacc	ggaaaccc	ggtatacc	tacccccc	gaggcctgtat	gaaacctgt
cgaggactact	tсagagtgaa	caacctggac	gtgaaагtga	агtccatcaa	tggggctt
accagcttcc	tgcggggaa	gttggaaагtt	aагaaагагc	ggaacaагgg	gtacaагac
cacggcggagg	acggccctgat	cattcggcaac	gcccatttca	tотtcaaагa	gtggaaагaaa
ctggacaagg	ccaaaaагt	gatggaaaac	cагatgtcg	агаагаагca	ggccggagag
atgcccggaa	tggaaaccc	gcaaggatc	aaagagatct	tcatcacc	ccaccagatc
aacgacatt	aggacttca	ggactacaag	tacagocacc	gggtggacaa	gaaагctt
agagagctga	ttaacgacac	cctgtactcc	accggaaagg	acgacaагgg	caacaccctg
atcgtgaaca	atctgaaacgg	octgtacgac	aaggacaat	acaагctgaa	aaагctgatc
aacaагагcc	ccgaaaагt	gctgtatgtac	caccacgacc	ccacgac	ccагаaaactg
aагctgatta	tggaaacatg	cgggcggac	aагаatcccc	tgtacaагt	ctacgaggaa

-continued

SEQUENCES					
accgggaaact	acctgaccaa	gtactccaaa	aaggacaacg	gccccgttat	caagaagatt
aagtattacg	gcaacaact	gaacgcccatt	ctggacatca	ccgacgacta	ccccaacacgc
agaaacaagg	tcgtgaact	gtccctgaag	ccctacagat	tcgacgttta	cctggacaat
ggcgttaca	agtgcgtac	cgtgaagaat	ctggatgttga	tcaaaaaaaga	aaactactac
gaagtgaata	gcaagtgtca	tgaggaagct	aagaagctga	agaagatcag	caaccagggcc
gagtttatcg	cctccttcta	caacaacgtat	ctgatcaaga	tcaacggcga	gctgtataga
gtgtatccgg	tgaacaacga	cctgctgaac	cggatcgaag	tgaacatgtat	cgacatcacc
tacccggact	acctggaaaa	catgaacgac	aagaggcccc	ccaggatcat	taagacaatc
gcttcaaga	cccagagcat	taagaagtac	agcacagaca	ttctggcata	cctgtatgaa
gtgaaatcta	agaagcaccc	tcagatcatc	aaaaaggcc		
SEQ ID NO: 35					
codon optimized nucleic acid sequence encoding <i>S. aureus</i> Cas9					
atgaaggcga	actacatct	cggacttggac	atccggattat	cctccgttgg	atacggcattc
atcgattacg	aaacttaggg	tgtgtatgcac	gctggatgtca	ggctgttcaat	agaggcgaac
gtggagaaaca	acggggggcg	ggcgttcaaaag	aggggggccc	gcggcgtgaa	gcgcgcgcgc
agacatagaa	tccagcgtcg	gaagaagctg	ctgttgcact	acaacattct	gaccgacac
tccgaactt	ccggcatcaat	cccatatgat	gcttagatgt	agggattgtc	cctaaagctg
tccgaggaaag	agttctccgc	cgcgttgc	cacctcgcca	agcgcagggg	agtgcacaaat
gtgaacgaag	tggaaqaaga	tacccggaaac	gagctgttca	ccaaaggagca	gatcagccgg
aactccaaagg	cccttggaaaga	gaaatacgtt	ggcgaacttgc	aactggagcg	gctgaagaaa
gacgggaaag	tgcgcgttgc	gtatcaacgc	ttcaagacct	cgactactgt	gaaggaggcc
aagcagctcc	tgaaagtgc	aaaggtctat	caccaacttg	accagtctt	tatcgatacc
tacatcgatc	tgctcgagac	tccgcgtact	tactacgagg	gtccagggga	gggtccccca
tttgggttga	aggatataat	ggagtgttgc	gaaatgttgc	tgggacactg	cacatactt
cctggaggagc	tgcggaggcg	gaaatacgc	tacaacgc	actctgtacaa	cgcgttgcac
gacacttgcata	atctcgat	cacccggac	gagaacggaa	actctcgat	ttacgaaaag
ttccagatata	ttgagaacgt	gttcaaaacag	aagaagaagc	cgacactgtt	gcagatttgc
aaggaaatcc	tcgtgaacga	agaggacatc	aagggttac	gatgttgcct	aacgggaaag
ccggaggatca	ccaatctgaa	ggtcttaccac	gacatcaaa	acattaccgc	ccggaaaggag
atcattgaga	acgcggact	gttggaccag	attggaaaga	tttctgaccat	cttacaaatcc
tccgaggata	ttcaggaaga	actcaccac	ctcaacacgc	aactgaccat	ggaggagata
gaccaatct	ccaaaccttgc	gggttacacc	ggaaactata	actctgagct	gaaggccatc
aacttgcattc	tggacgtact	gtggcacacc	aacgataacc	agatcgctat	tttcaatcg
ctgaagctgg	tcccaaaatcc	agtggacactc	tcacaaatcaa	aggagatccc	tactaccctt
gtggacgtt	tcattctgtc	ccccgttgc	aagagaagct	tcatacgtt	aatcaaagt
atcaatgcata	ttatcaagaa	atacgttgc	cccaacgc	ttatcatgt	gctgcgcgc
gagaagaact	cgaaggacgc	ccagaaggat	attaacgaaa	tgcagaagag	gaaccgacag
actaacgac	ggatcgaaga	aatcatccgg	accacccgg	aggaaaacgc	gaqtactcg
atcgaaaaga	tcaagcttca	tgatcgac	gaaggaaatg	gtctgtact	gctggaggcc
atcccgctgg	aggacttgc	gacaacccct	ttaactacg	aatggatca	tatcattccg
aggagcgtgt	cattcgacaa	ttcccttcaac	aacaagggtt	tctgttgc	ggaggaaaac
tcaagaagg	gaaaccgcac	ggcgttccag	tacctgttgc	gcacggactc	caagattcc
tacgaaacat	tcaagaagca	catccttcaac	ctggcaaaag	ggaagggtcg	catcttcaag
accaaaagg	aatatctgt	ggaaaggaaaga	gacatcaaa	gatttccgt	gcaaaaggac
ttcatcaacc	gcaacccctgt	ggatactaga	tacgttact	ggggtctgt	gaacctctg
agaagctact	tttagtgc	caatctggac	gttgcgttca	atgcgtttaa	cggaggtttc
accttcttcc	tgcggcgc	gttgcgttca	aagaaggaaac	ggaacaagggg	ctacaagcac
cacggcgttgc	acggccctgt	cattgttcaac	gccgttctca	totttcaaa	atggaaagaaa
cttgcataagg	ctaagaagg	catggaaaac	cagatgttgc	aagaaaagca	ggccgttgc
atgccttgc	tcgagacttgc	acaggatgt	aaggaaatct	tttattacgc	acacccatc
aaacacatca	aggatttca	ggatttaca	tacttacatc	gctgttgc	aaagccgaaac
agggaacttgc	tcaacacac	cccttacttcc	acccggaaagg	atgacaaatgg	gaataccctt
atcgatcaaca	accttacatgg	ccctgttgcac	aaggacatc	ataagcttgc	gaagcttatt
aacaagtgc	ccggaaaatgt	gttgcgttgc	caccacgtt	cttgcactt	ccagaagctc
aagctgtatc	tggagacttgc	tggggacgg	aaaaaccgt	tgttacaatgt	ctacgaaagaa
actggaaatt	atctgttgc	gtactccaa	aaagataacg	gccccgttat	taagaagatt
aagacttgc	gcaacaact	gaacgcccatt	ctggatcatc	ccgtatgtt	ccctaattcc
cgcacaacagg	tcgttgcact	gaggcttca	cccttacccgt	tttgcgttgc	ccttgcataat
ggagtgttaca	agtgcgttgc	tgttgcgttgc	cttgcgttgc	tcaagaaggg	gaactactac
gaagtcaact	ccaaagtgttca	cgaggaaatc	aagaagtttgc	agaagatctc	gaaccaggcc
gagtttcatgt	cctcccttcta	taacaacgc	ctgttacaat	tcaacggcga	actgttaccc
gttcatttgc	tgaacaacgc	tcttgcgttgc	cgcatgttgc	tgaatgttgc	cgacatcatc
tacccggat	acctgggaaat	tatgttgcac	aaggccccc	cccgatcat	taagactatc
gccttcaaga	cccgatgttgc	caagaatgtt	agcacogaca	ttctggcata	cctgttacgg
gttcaatgc	agaagcaccc	ccagatcatc	aaaaaggcc		

SEQ ID NO: 36

codon optimized nucleic acid sequence encoding *S. aureus* Cas9

atggcccaaaaggaaagaaggcgaaagggttgcgttccacggatccggatccatgttgc

ggggcttgcgttccatgttgcgttccatgttgcgttccatgttgcgttccatgttgc

atgcggccgttgcgttccatgttgcgttccatgttgcgttccatgttgcgttccatgttgc

ggcagaagggttgcgttccatgttgcgttccatgttgcgttccatgttgcgttccatgttgc

cctgttgcgttccatgttgcgttccatgttgcgttccatgttgcgttccatgttgcgttccatgttgc

-continued

SEQUENCES

SEQ ID NO: 37

codon	optimized nucleic acid sequence	encoding S. aureus Cas9
accggtgcca	ccatgttaccc	atacgatgtt
aaggctgcaag	cgtccatgaa	aaggaaactac
tgggggtatg	ggattatgtta	ctatcaaaca
ttcaaggagg	ccaacgtgga	aaacaatgag
ctgaaacgcac	ggagaaggca	cagaatccag
ctgtcgaccg	accattctga	gctgagtgg
ctgactcaga	agctgtcaga	ggaagatgtt
cgaggagtgc	ataacgtcaa	tgagggtggaa
gaacagatct	caccaatacg	caaagctctg
gaacggctga	agaagatgg	cgagggtgaga
tagtc当地	aagccaaagca	gctgtgtaaa
agtttcatcg	atacttatat	cgaccctgtg
ggagaaggga	gcccttcgg	atggaaagac
cattgcactt	atttccaga	agagctgaga
tacaacgccc	tgaatgtacct	gaacaacctg
gaatactatg	agaagtccca	gatcatcgaa
ctggaaacaga	ttgctaaagga	gatcttggtc
acaaggactg	gaaaaccaga	gttcaccaa
acagcacgga	aagaaatcat	tgagaaacgc
actatctacc	agagtcgcga	ggacatccag
accggaggaa	agatcgaaaca	gatttagaat
tccctgtaaag	ctatcaatct	gattttgtgt
gcaatcttta	accggctgaa	gctggtccca
atccccaaacca	cactggtgg	cgatttcatt
cagagcatca	aagtgttcaa	cgccatcatc
atcgagctgg	ctaggggaaa	gaacgcaag
aaacgc当地	ggcagaccaa	tgaacgcatt
aaacgc当地	acctgttga	aaaaatcaag

-continued

SEQUENCES					
tattctctgg	aggccatccc	cctggaggac	ctgctgaaca	atccattcaa	ctacagggtc
gatcatatta	tccccagaag	cgtgtcttc	gacatttct	ttaacaacaa	ggtgtctggtc
aaggcaggaa	agaactctaa	aaaggggaaat	aggactcctt	tccagtacct	gtctagttca
gattccaaga	totcttacga	aacctttaaa	aagcacattc	tgaatctggc	caaaggaaag
ggccgcatca	gcaagaccaa	aaaggagttac	ctgctggaaag	agcgggacat	caacagattc
ctgttcacaga	aggattttat	taacccggaaat	ctggggaca	caagatacgc	tactccggc
ctgatgaatc	tgctcgatc	ctatccctgg	gtgaacatc	tggatgtgaa	agtcaagtcc
atacaacggcg	ggttcacatc	ttttctggagg	cgcaaatggaa	agttttaaaaaa	ggagcccaac
aaagggtaca	agcaccatgc	cgaagatgtc	ctgattatcg	caaatgcccga	cttcatcttt
aaggagttgga	aaaagctggaa	caaaggccaaag	aaatgtatgg	agaacccatgt	gttccgaagag
aaaggcccg	aatctatgcc	cgaatatcgag	acagaacagg	agtacaagga	gattttcatc
actccctacc	agatcaacgca	tatcaaggat	ttcaaggact	acaagtactc	tcaccgggtg
gataaaaaagc	ccaacagaga	gctgatcat	gacaccctgt	atagtacaag	aaaagacgt
aaggggaaata	ccctgtatgt	gaacaaatctg	aacggactgt	acgcaaaaga	taatccaaag
ctgaaaaaagc	tgatcaacaa	aagtcccgag	aagctgtgt	tgtaccacca	tgatcttcag
acatcatcaga	aactgtactgt	gattatggag	cagtagccgc	ccgagaaaga	cccactgtat
aagtactatg	aagagactgg	gaacttacctg	accaagtata	gcaaaaaggaa	taatggcccc
tggtatcaaga	agatcaagta	ctatgggaaac	aagctgtatg	cccatcttgg	catcacagac
gattacccta	acatgtccaa	caagggtgtc	aagctgttcac	tgaaggccata	cagattcgat
tgctatctgg	acaacccgcgt	gtatataattt	tgactgtca	aaatctggaa	tgtcatcaaa
aaggagaact	actatgtaaat	gaatagcaag	tgctacaaag	aggctaaaaaa	gctgaaaaag
attagcaacc	aggccaggtt	catcgccctc	tttttacaaaca	acgcggatgt	taagatcaat
ggcgaactgt	atagggtcat	cgggggtgaaac	aatgtatctgc	tgaaccggcat	tgaagtgaat
atgttgcaca	tcacttaccg	agatgtatctg	gaaaacatga	atgataaagcg	cccccttcga
attatcaaaa	caattgcctc	taagacttcg	agtatcaaaa	agtactcaac	cgacattctg
ggaaacccgtt	atgaggtgaa	gagcaaaaag	caccctcaga	ttatcaaaaa	gggctaaagaa

SEQ ID NO: 38

-continued

SEQUENCES

tccggcgtgaacaacgcacccgtctgaaccggatcgaaatgatcgacatcacctaccggcagtagcatcgaaaaatcatcgacacaaaggccccccaggatcataagaacatcgccctcaaaaggccacagacatcataagaagtacagcacacgacatcttgcggcaaccatgtatgaatgtaaagactatcataagaagccaccctcagatcatcaaaaaaggccaaaaggccggccggccacgaaaaaaggccggcccaaggccaaaaaaaagg

SEQ ID NO: 39

SEQ ID NO: 40

Vector (pD0242) encoding codon optimized nucleic acid sequence encoding *S. aureus* Cas9
ctaaatgttaacgcgttaatattttgttaaaattcgcgtaatttttgttaaatcagtcattttta
accataaggccgaatccggaaaatcccttataaatcaaaaagaatagaccgagatagggttgagtgtt
gttccaggatgttggacaagactccactattaaagaacgtggactccaaacgtcaaggccgaaaaacgtt
ctatcaggccatggccactacgtgaaccatcacccataatcaagttttgggtcgagggtgcgtta
aagcactaaatcggaaaccctaaggggagccccgatttagacttgcacggggaaagccggcaacgtg
ggcggaaaaggaaagggaagaaaggcggaaaggagggccgttagggcgctggcaagttagcggtcacgct
ggcgtcaaccacccacccggccgttaatggccgtacaggggcgccatccatggccatccatggc
tgcgcacttgtggaaaggcgatcggtggccctttcgcttacatggccgtacggggcgccatccatggc
tgtgtcgcaaggcgatataagtgtggtaacccggggctttccactacgtacgcgttgtaaaaacgcggc
cagttagcgccgtatatacgactactataggcgaaattgggtacCtttaattctgactatgcATg
cggtacattgattttagactatgttataatagaatcaattacggggtcatatgttcatggccata
tatggagtccgcgttatcaacttacgttaatggccggctggtgaccggccaaacgcacccccc
cattgcgtcaataatgcgtatgttccctttagtaacccggatggggactttccactgtcaatgg
gtgggatatttacgtttaacttgccttacttggcgactatcaagtgtatcatatgcggcc
tatttgcgtcaatgcgttaatggccggctggcattatggccactatgcggatcatgcac
ctacttggcgactatctacgtttagtgcattaccatggtgatgggttttggcgtatcacatc

-continued

SEQUENCES

-continued

SEQUENCES

gaggggcttaccatctggccccagtgtgcataatgatacccgagaccccacgtcaccggcttcaggatttt
atcagcaataaaaccggccaggccggaaaggccgagccgagaatgtgttcgtcaactttatccgcctccaa
tccagttctattaatgttgccgggaaactgttagatgttgcggcgttagtaatagtgttgcggcaacgtt
gttgcgttatgttgcacggcgttgtgtgtcagctgtcggttgtgttgcgttgcgttgcgttgcgttgcgtt
ccaaacgtcaaggcgagttacatgtccccatgtgtgtcaaaaaaaaggcggttagtgccttcgtgtcc
cgatcggtgtcagaagtaagtggccgcagtgttatcactcatgttgcgcactgtcataattctt
cttagtgcgtccatccgtcgttagtgcgttttgtgtactgtgttagtactcaaccaactgttgcgttgcgtt
atagtgtatggccgcaccggatgtgtgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
gaactttaaaagggtcgtatgtggaaaacgttttgcggggaaaactcttcgttgcgttgcgttgcgtt
ttgagatccgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
cggttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
gttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
ggatcatatgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
ggccac

SEQ ID NO: 41

Human p300 (with L553M mutation) protein
MAENVVPGPPSAKRKPLSSPALSASASDGTDFGSLELDHELDLPELINSTELGLINGGDINQLQTSL
GMVQDAASKHQKLSSELRSSGSPNLMVGPGPQVMASQAQQSSPGLGLINSMVKSPTQAGLTSPNM
GMGTSPGNQPGTQSTGCMNNSVPNQPMAMGNTMHNAGMNGPMLAAGNGQCIMPQMNVMSIGAQRGRNQ
MQYPNPGMGSAQNLLTEPLQQGSQMGQTGLRPQPKLMGMNNPNPYGSPYTNQNPQGQIAGASGLGL
QIQTKTVLVSNLSPFAMDKAQVPGGMPNMQQPAPQVQQPGLVTPVAQGMGSQAHADPEKRKLIIQQ
QLVLLNLHAAHKCORREQANEVRQCNLPHRCRTMKVNLNHMHQCQSKSVAHCAASSRQIIHSWKNCNTR
HDCPVCLPLNKAGDRNQPIITLTGAVPGLCNSLSSVGQGSQAPSNLTSVQDPPSIIERAYAALGLPYQ
VNQMPMTQPVQAKNQNQOPQGSQGMRPMNSMSAPMGVNGVGVTQPSLSDSMHLHSAINSNQNPQMS
SENASVPSMGPMTAAQPSTTGIRKQWHEIDTQDLRNHLVHLVQAIFFPTPDPAALKDRRMENLVAYA
RKVEGMDYESANRRAEYYHLLAEKTYKIQKELEKRRRLQKQMLLPNAAGAMVPSVMNPQGPMQGPQD
GMTSNGPLPDPSMIRGVSPNQMMPRITPQSGLNQFGQMSAQAPIVVRPQTPLQHNGQLAQGPALNQG
MGYGPQRMOPQSNGQFLPQTQPSQGMNVNTIPLAPSSQAOPIVQSPNQMSSSCPVNSIPMPPGSQGS
IHCPQLPQPALHQNSPSPVPSRTPTPHTPSIGAQQPPATTI PAPVPTPPAMPPGPQSQALHPPRQ
TPTPPTITQLPQVQPSLQPSAPSAQDQPOQFRQSSQTAASVPTPTAPLLPPQPATPLSQAPQ
NPPTSTSSTEVSQIAEQKPSQEVKMEAKWEDVQPEADPTQDIEISVKEVDCKMESTETEERSTELK
TEIKEEDEDQPSATQSSPAGQSKKKIIFPEELRQLAMPTEALYLRQDPESLFPFQVDPDQLLGI
YFDIVVKSPMDLTIKRKLDTGQYQEWPQYVDDIWLMMENNAWLYNRKTTSRVYKCSKLSSEVFQEIDPV
MQSLGYCCGRKLEFSPTQLCCTGGYQKLCITPFRDATTYSYSONRYHCECFKNEI QGESVSGLGDDPSQPT
TINKEQFSKRNKDTDLPELVEFECTECGRMKHQICVLHHEIWPGFVCDGLKKSARTKENFKSAK
LPSTRGLTFLENVRNDFLRQNHPESGEVTVRVVHASDKTVEVKPGMKARFWDGSEMAESFVYRTKAL
FAFEEIDGVLDCCFFGMHVQEYGSDCPPNPQRRVYISYLDSVHFRPKCLRATAVHEYIILIGLEYVKKL
GTTGTHIACWPSCEDDYIIFCHCPBDPKIKPCKRQLEWYKQMLDKAVSERI VHDYKDIKFQKATEDRLT
SAKELPYFEQGDFWPVNLEESI KELEQEEEERKTSNSETS DTVTKGDSKNAKKNNKKT SKNKSLL
RGNNKKPGMPNVSNLDSQKLYIATMEKHKEVFFVIRLIAGPAANSLPPVDPDPLIPCLCDMGRDAFLT
LARDKHLFESSLRRAQSTMCLVELHTQSQDRFVYTCNECKHHVETRWHCTCVDYLCTICYNTKN
HDHKMEKLGGLDDSESNQAAAATQSPGDSRRLS ITCRQ103LHVHACQCRNANC SLPSCQKMRVWVQH
KGCKRKRTNGCPICKQLIALCQYHAKHCQENKCPVFCLN1IKQKLRQQQHQLRQZQAMLRNRRMSAQ
RTGVVGQCGQLLPGSPPTQPTPTQPTQPTQPSQPOPTPPNSPMSPYPLRPTQAOAGPVSQGKAAGC
VTPPTPPQTAQPLPGLGPPPAVEMAMQI QRAAETQRQMAHVGQI FQRPQIHQMPMPTPMAPGMNPPPM
TRGPSSGLEPQMGPMQGPMQOPWPSQGGLPQOQQLQSGMPRPMAMSVQAHQGLNMAPQPGLQCVGQI SP
LKPGBTWSQQLNQNLRTLSPSSPLQQQQVLSIHLANPQLLAIFI KQRAAKYANSNQPI PGQPGM
QGPGLQPPTPMGPQGQHVHSNPAQMNQNMPMQAGVQORAGLPLQCPQOQOLQPPMGGMSQPAQMNHMHNTP
SQFRDI LRRQOMMQQQQQQGAGPGI GPGMANHNQFQOPQPGVGYPPQQQQRMQHHMQQMQQGNMKGQI GQ
LPQALGAEAGASLQAYQQLRQQQMGSPVOPNPMSPQQHMLPNQAOQS PHLQGQOIPNLSNQVRSPQ
VPSPRPQSPQSPHSSPSRPMQSPQSPHVSPTSPQSPHPGLVAAQANPMEQGHFASPQDNMSLSQLASNP
GMANLHGASATDGLSLTDNSLDLNSLSQLSTDIH

SEQ ID NO: 42

Human p300 Core Effector protein (aa 1048-1664 of SEQ ID NO: 41)
1FKPEELRQALMPTLEALYRQDPESLPFQRQPVDPQLLGIPDYFDIVKSPMDLSTIKRKLDTGQYQEWP
QYVDDIWLWMFNNAWLYNKRKTSRVYKCYCSLSEVFQEQTDPVMQMSLGYCCGRKLEFSPOTLCCYGKQLC
TIPRDRDATTYSYSQNRYKCFNEIQQGESVSLGDDPSQPQTINKEQFSRKNDTLDPELFVCECTKQH
RKHMHQCVLHHEIIWPAGFVCDGLKKSARTRKENFKSAKRLPSTRLFTGLENRVNDPLRQNHPESG
EVTVRVVHASDVTVEVKPGMKARFVDSGMAESFPYRTKALFAFEEDGVDLCCFGMHVQEYGSDCPP
PNQRRRVISYLDHSVHFRPKCLRTAVYHEILIGYLEVKKLGTTGHIWACPSEGDDYIFHCHPPDQ
K1PKPKRLQEWYKKMLDKAVSERIVHDYKD1FKQATEDRLTSAKELPYFEGDPWPVNVLSEEISTKELEQE
EEERKREENTSNESTDVTGDSKNAKKKNNKKTSKNKSSLSSRGNNKKKGMPNVNSNDLSQKLYATMEKH
KEVFFFVIRLJAGPAANSLPPIVDPDPLIPCDLMDGRDAFLTLARDKHLFEFSSLRRAQWSTMCMVLVELH
TOSOD

SEQ ID NO: 43

SEQ ID NO: 43
VP64-dCas9-VP64 protein

VP64-*Qcass*-VP64 protein
RADALDFDLDLMLGSDALDDFDLMLGSDALDDFDLMLGSDALDDDELDMVNPKKRKVGRGMDDKKY
SIGLAIGTNSGVWAVITDEYKVPSSKKFKVLGNTDRHSIKKNNLIGALLFDSGETAETRLLKRTARRRYTT
RRKNNIICYLQBIIFSNEMAKVDSFFHRLIEESFLVEEDKKHERRHPIFIQFGNIVDEVAHEYKPTTYHLRKKK
LVDSTDKDARLRLYIHALAHMKIFKRHHFLLIEEGDLNPDNSDVDKLFVQLVQTYNQLEFEENPINASGVDAK
A1SALSRSKSRLENLIAOLPGEKKNGLFGNLIALSLGLTPFNKSNDLEADAKL05KDTDYDLDLNN
A1SALSRSKSRLENLIAOLPGEKKNGLFGNLIALSLGLTPFNKSNDLEADAKL05KDTDYDLDLNN

-continued

SEQUENCES

L LAQIGDQYADLFLAAKNLSDAII LLS DILVRNTEITKAPLSASMI KRYDEHHQDLTLLKALV RQQLPE
KYKEIFFDQSXNGYAGYIDGGSQEFYFKIPKILEKMDGTEELVVKLNREDDLRKQRTFDNGSIPHQ
IHLGELHAIIRRQEDFYPFLKDNRKEIKLFTPIRYYVGPLARGNSRFWAMTRKSEETITPWNFEEV
VDKGASAQSFIERMTFDNKL PNEKVKPHSLLSEYYFTVNEYTKVKTGEMRKPFLSGEQKKAI
DLLFKTRNKT VTVKQLKEDYFPKKIECFDSVEISGVEDRPNASLGTYHDLLKI1KDKE LDNEEDEDILE
DIVLTTLTFEDREMIEERLKTYAHLFDDKVMQKLKRRRTGWRGLSRKLINGIRDQSGKTI1DFLKS
DGFANRNFMQLIHDDSLTFKEDIQKAQVSQGGSLDHEIANLAGSPAIKGILQTCKVVDLKVVMGR
HKPENIVIEMARENTQQKGQKNSRERMKLEIEGIGKELSGQSLKEHPVENTQLOQNEKLYLYLQNQRD
MVQDVLBDINRLSDYDVDA1VPQSFLKDDSDNKVLTRSDKRNQGKSDNPSEVVVKMKNYWRQQLNA
KLITQRKFNDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVIT
LKS KLVSDFRKD FQFYKVREIINYYHAHDAYLNAVVTGALI1KKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDGKRDFTVKRVLSPMQVN
IVKTEVQTGGFSKESLIPRNSKDLIARKHDWPKYGGFDSPTVASVVLVAKVEGKSKNLKSV
ELLGITIMERSSFEKNPIDPLEAKGYKEVKKD LI1KLPKYSFLERGKRM LASAGELQKGKNNELALP
SKYVNFLYLA SHYEKLKGSPEDNEQKQLFVEQHKHYLDEI1IEQISEFSKRVILADANLDKVL SAYNKH
RDKP1REQAENI IHLFTLMLGAPAAFKYEDTTIDRKY TSTKEVLDATLHQ SITGLYETRIDLSQL
GGDSRADPKK KKRKVASRADALDDFDLDMGLGSDALDDFDLDMGLGSDALDDFDLDMGLGSDALDDDELDML
I

-continued

SEQUENCES

SEQ ID NO: 45
Polypeptide sequence of KRAB protein
RTLVTFKDVFVDFTREEWKLLDTAQQILYRNVMLENYKNLVSLGYQLTKPDVILRLEKGEPP
WLV

SEQ ID NO: 46
Polynucleotide sequence for KRAB
cggacactggtgaccctcagaggatgtttgtggacttaccaggggaggagtggaaagctgtctggacactgtcgcagatctgtacagaaatgtgtcgatgtggagaactataagaacctgttgccttgggttatcagcttactaagccagatgtgatccctccgggttggagaaggggagaagcccctgtgtgt

SEQ ID NO: 47
Polypeptide sequence of *Streptococcus pyogenes* dCas9-KRAB protein
MDYKDHGDKYDHDIDYKDDDKMAPPKKRKVGRGMDKKSIGLAIGTNSVGWAVITDEYKVPSKKFK
VLGNTDRHSIJKKNLIGALLDSEGETAEATRLKRTARRRYTRRKNRICYLQEIEFSNEAKVDDFSHRL
EESFLVVEEDKKHERHPFGNIDEVAYHEKYPTIYHLRKKLVDSTDADLRLIYLALAHMIKERHEL
IEGDLNPNDNSDVKLFIQLQVQTYNQLFEENPINASGVDAKAISLARSLSKSRRLNLIQALGEKKNGL
FGNLLNLSSGLGLTPFNFSKNFDLAEDAKDQLKSKDYYDDDDNLQAQDQYADLFLAAKNLSDAILLSD
LRVNTEITKAPLASMISIYKDEHHQDLDLTLKVALQRQNLPEKYKEIFFPDQSQNKGYAGYIDGGSQEEFY
KFPIKPILEMDGTEELLVKNLNREDDLLRKQRTFDNGSIPHQIHLGELHAIIRRQEDFYPFLKDNRKIE
KILTFRIPYVYGVPLARGNSRFAMTRKSEETITPWNFEEVUDKGASQSFIERMTNFDKNLPLNEKVLP
KHSLLSEYYFTVNEYLTWKVYVTEGMRKPAFLSGEQOKKAIVDFFLTKRNKTVKQLKEDYFHKFIECFDS
WEISVGDEFDRNASSLGTYHDLKLICKDKDFLNEEDEDIILEDIVLTLLTFEDREMIEERLKTYAHLFDD
KVMKQLKRRRTGWRGLSRKLLINGIRDQSGKTIIDELKSDGFANRNFMQLIHDDSLTFKEDIQKAQV
SQGDSLHETIANLAGSPAIKKGLQJLQTVKVVDELVKVGMGRHKPENIVIEMARENQTTQKGQKNSRMR
KRIEIGEKELGSQILKEHPPVENTQLQNEKLYLYLQNQGRMVDQELDINRLSDYDVAUVPQSLFKD
DSIDNKVLTSDKRNQKCSNDNPSEEVVKMKMNYWRQQLNALKITQRKPDNLTKAERGGLSLEDKAGF
KRQLVERTRQITKHVAQILDSDRMNTYKEDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNNYHHAH
DAYLNAVVTGALIKKYPKLESEFVYGDYKVYDVVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLA
NGEIRKRPLIETNGTGEIWVDKRGRDFATVRKVLSPMQVINVKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYYGGFDSPTVAYSVLVVAKEVKGSKLSKLKSVEKLGLITIMERSSPKFNPIDFLEAKGYKE
VKKDLI1IKLPKYSLEFLENGKRMLASAGELKQGNELALPSKVNLYFLASHYEKLKGSPDENQKOL
FVEQKHHLDEI1IEQISEFSKRVILADANLDKVL SAYNKHRDKPKIREQAEINIIHLFTLNLGAPAAFK
YFDFT1DRKRTSTKEVLDATLHQISITGLYETRIDLSQGGDSRADPKKKRKVSDAKSLTAWSRSL
VTFKDVFVDFDTREEWKLLDAAQOQIYRNVMLENYKNLVSLGYQLTKPDVILRLKEGEPWLVEREIHQ
ETHPDSETAEIKSSVPKKRKV

SEQ ID NO: 48
 Polynucleotide sequence encoding *Streptococcus pyogenes* dCas9-KRAB
 atggactacaagacccatgcgggtattataaaagatcatgacatcgattacaaggatgcacatgcacaa
 gatggcccccaagaagaagaggaagggtggccgcggatggacaagaacttccatgggtccgc
 tcggcacaacacgcgtcggtggccgcgtattacggacggatcaagggtggcggagaaaaatcaa
 gttctggcaataccgcgtccacagcataaaaagaagaacccatggccctctgttcactccgg
 gggaaaccgcgaaggccacgcgcgtccaaaagaaacagcacggcgcagataccggcagaagaatcgga
 tctgtcacctgcggagatcttagatagatggatggcataagggtggatactttttccataggctgc
 gagggatctttttggggaggataaaaggacgcgcggccacccatcttggcaataatcggtgg
 cgagggtggcgttccatggatggggatggatggatggatggatggatggatggatggatggatgg
 ataaggctgacttgcgggttatctcgctggcgtatgcataatcttggggacacttcc
 atcgaggggggacccatggccacccaggacaaacacgcgtgcgacaaaactcttattccaaacttgtt
 cagaacttgcgttccatggccacccaggacaaacacgcgtgcgacaaaactcttggggacacttcc
 caatcagcttgcggatggccacccaggacaaacacgcgtgcgacaaaactcttggggacacttcc
 ggctgtccatggccacccaggacaaacacgcgtgcgacaaaactcttggggacacttcc
 ttggatatcttgcggcttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
 agatgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
 ggcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
 ctggccatggccatggccatggccatggccatggccatggccatggccatggccatggccatggccat
 ccaccaacggatggccatggccatggccatggccatggccatggccatggccatggccatggccat
 ttttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
 aaatatttaagccatcttggaaaaatggacggcaccggaggagctgtctgtttaaagcttaaacagaga

-continued

SEQUENCES

SEQ ID NO: 49

Polypeptide sequence of *Staphylococcus aureus* dCas9-KRAB protein
MAPKKKKRKGVIHGVPAAKRNLYILGLAIGTTSVGVGIDYETRVIDAGVRLFKEANVENNEGRRSKRG
ARRLKKRKKRHRQVKKLFDPEYDNLTDHSELSGGINPYEARVKGSLQKLSSEEEFAASLLHLAKRRGVHN
VNEVEEDTGNLSTKEQISRNSSKALEEKVVAEQLERLKKDGEVRGSINRFKTSVDYVKEAKQKLKVQK
AYHQLDQSFIGTIDYLLETRRTTYEGPGEGSPFGWWDIKEWYEMLHGCHTYPEELRSVKYAYNADLY
NALNDLNNLVITRDENEKLEYYEKFQIIEIVFKQKKPKTLQIAKEILVNNEIDIKGYRVSTGKPEFT
NLKVYHDIDITARKEIIENAEQLDQIAKILTITYQSSEDIQEELTNLNSLELTQEETBQISNLKGYSIT
HNLSLKLAINILDELDWHTNDNQIAIDNLRKLVPKVKDLBSLQSQKEIPTTLVDDFILSPVVKRSLFIQSIVK
INAIQYKGLPNLIDIELAREKNSKDAQKMINEQMOKRNRQNTNREBIEIIRTTGKENAKYLIEKIKLHD
MQEGKCLYSLEAIPLEDLNNPFPNEYEVDHIPRSVSFDNSFNNKVLVKQEEASKGNRTPFQYLSSSD
SKISYETFKKHILNLNAKGKRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMLNLLRSYF
RVNNNDLVKVKINGSQGGTSFLRRKFKKKFERNKGKYKHHADALI1ANADFIFKEWKLLDKAKKVMEQNM
FEEKQABESMPBIETEQYKEBIITPHQKIH1KDFDPTQYKLSKLIMEQYQGEDEKPNLKYLYKVEETGNYLTKY
IVNNLNGLYDNDKNDLKLKKLINKSPKELLMYHDQTYQKLSKLIMEQYQGEDEKPNLKYLYKVEETGNYLTKY
SKKDNGPVVIKKIYKGKLNMAHLDITDDDPNSRNKVKVLKSLPKYRFDVYLDNGVYKFVTVKVLNDV1KK
ENYYEVNSKCYEEAKLKKISIWNQAFIASFYNNLDL1KINGELYRV1GVMNDL1LNRIEVNMID1ITYREY
LENMNDKRPRI1KTIASKTQS1KYYKSTD1LGNYEVKS1KHPQ1IKKGKRPAATKKGQAKKKKGSD
AKSLTAWSRTLTVFKDVFVDTKLLD1TAQQLY1RNVMLENK1NLVSLGYQLTKPDV1LREKGE
EPWLVEREIHOPDSETAEI1KSSVPKKRKV

SEQ ID NO: 50

SEQ ID NO: 30
Polynucleotide sequence of *Staphylococcus aureus* dCas9-KRAB protein
atggcccccagaagaaggcggaaggtcggtatccacggagtcccgacgcggaaactacatcct

- continued

SEQUENCES

ggccctggccatcgcatcaccagctggctacggcatcatcgactacgagacacgggacgtgtcg
 atccggcgtggctgttcaaaggccaaactggaaaacaacgaggcaggcggagcaagagaggc
 gccagaaggctgaagcgccggaggccatagaatccagagactgagaagactgtgttcactacaa
 ctgtgcaccgaccacagcactgagcggcatcaaccctacgaggccagactgaagggccgtgaccc
 agaagctgagcggaggaagagtctcgccgeccctgcacctggcaagagaagaggcgtgcacaac
 gtgaacgagggtgaagaggacccggaaacgagactgtccacccaaagagacatcaccggaaacgaaa
 ggcctggaagagaatacgtggccgaactgcagctggaaacggctgaagaaagacggcgaagtgcggg
 gcacatcaaccatcggccggactgtaaagaacggcaaacacgtgtgaaggtgcagaag
 gcttaccacaggctggccggaggactgtttatcgacactacatcgactgtggaaacccggggactta
 ctatggggacttggcgaggcggcccttcggctgaaaggacatcaaagaatgttacgagatgtga
 tggggacttgcactacttccccggaggactgcggagcgtgaagtacgcctacaacgcggactgtac
 aacgcctgtgaacgacactgtaaatctgttatcaccaggggacggaaacgaaagactgttgaaatattacga
 gaagggttccatcgagaacgttcaagcagaagaagacggccaccctgtggacatgcggaaag
 aaatctctgtgaacgaaaggatattaaaggcttacagactgttacccggcgaacccggacttacc
 aacctgtaaagggttaccacgacatcaaggacattaccggccggaaagagattatttggaaacgcgcgact
 gctggatcagattggccaaatgttccgttacccggacttaccacgacgcggaggactccaggaaactgtacca
 atctgaatctccgagactgttccggaggactgttccgttacccggacttaccacgacgcggatc
 cacaaccttgacccgttacccggccatcaacactgttccgttacccggactgttccgttacccggacttacc
 cgacttccatcgaccgttacccggactgttccgttacccggacttaccacgacgcggatc
 ccaccctgttacccggacttaccctgttacccggacttaccacgacgcggatc
 atcaacccggccatcaacactgttccgttacccggacttaccacgacgcggatc
 ctccaaaggacccggccatcaacactgttccgttacccggacttaccacgacgcggatc
 agggaaatcatccggaccaccggccatcaacactgttccgttacccggacttaccacgacgcggatc
 atgcaggaaaggccatgttccgttacccggacttaccacgacgcggatc
 caactatcgagggttacccggacttaccacgacgcggatc
 tcgttacccggacttaccacgacgcggatc
 atccggccatcaacactgttccgttacccggacttaccacgacgcggatc
 agaagatcggacttaccacgacgcggatc
 caagaccaagaaaggatattctgttccgttacccggacttaccacgacgcggatc
 tcaacccggccatcaacactgttccgttacccggacttaccacgacgcggatc
 agaggtaacaccgttacccggacttaccacgacgcggatc
 gtgttacccggacttaccacgacgcggatc
 acgcccgttacccttccatcaacactgttccgttacccggacttaccacgacgcggatc
 ttcggaggaaaaggccatccggccatcaacactgttccgttacccggacttaccacgacgcggatc
 caccggccatcaacactgttccgttacccggacttaccacgacgcggatc
 agcctaataagaggacttaccacgacgcggatc
 atcgttacccggacttaccacgacgcggatc
 ccccgaaaaggccatcaacactgttccgttacccggacttaccacgacgcggatc
 agtacccggacttaccacgacgcggatc
 tccaaaaaggccatcaacactgttccgttacccggacttaccacgacgcggatc
 ggacatcaccggacttaccacgacgcggatc
 tccacgttacccttccatcaacactgttccgttacccggacttaccacgacgcggatc
 gaaaactactaaggacttaccacgacgcggatc
 gcccggatccatcaacactgttccgttacccggacttaccacgacgcggatc
 tccggcgttacccttccatcaacactgttccgttacccggacttaccacgacgcggatc
 ctggggatccatcaacactgttccgttacccggacttaccacgacgcggatc
 taaggacttaccacgacgcggatccatcaacactgttccgttacccggacttaccacgacgcggatc
 tccggggatccatcaacactgttccgttacccggacttaccacgacgcggatc
 gtaaggacttaccacgacgcggatccatcaacactgttccgttacccggacttaccacgacgcggatc
 ggaggacttaccacgacgcggatccatcaacactgttccgttacccggacttaccacgacgcggatc
 agaaccctgttacccttccgttacccggacttaccacgacgcggatccatcaacactgttccgttacccggacttaccacgacgcggatc
 nealnqipshkaltlthdnvvttvspyalthvgpynhwv
 caaatcatcagtccggaaaaggaaacgcggaaat

SEQ ID NO: 51

Polypeptide sequence of Tet1CD

LPTCSCLDRVIQKDKGPYYTHLGAGPSVAAREIMENRYQKGNAIRIEIVVYTKEGKSSHGCPIAK
 WVLRRSSDEEKVLCLVRORTGHICPTAVMVVLIMWDGIPLPMDRLYTELLENLKSYNHPTDRRCT
 LNEENRTFCQGIDPETCGASFSFGCSWMSYFNGCKFGRSPRFRIDPSSPLHEKNLEDNLQLSLATR
 LAPIYKQYAPVAYQNVEYENARECRLGSKEGRPFSGVTACLDFAHPRHDIHNMNNGSTVVCILTR
 EDNRSLGVIPQDEQLHLVPLPLYKLSLTDPEFGSKEGMEAKIKSGAIEVLAPRRKRTCTQPVPRSGKKR
 AAMMTEVLAHKIRAVEKKPPIPKRKNNSTTTNSKPSSLTLGSNTETVQPEVKSETEPHFILKSD
 NTKYLSLMPSPAHVPEASPGFWSPKTASATPAPLKNDATASCGFSERSSTPHCTMPSPGRLSGANAA
 AADNBADEPSPSDEPLSDDPLSPAEEKLPHIDEWSDEHIFLDANIGVVAIAPAHGSVLIECARRELHAT
 TPVEHPNRNHPTRLSLVFYQHKNLNKPQHGFELNKIKFEAKEAKNKMKAQEQDQAANEQPEQSSEV
 NELNQIPSHKALTlthdnvvttvspyalthvgpynhwv

SEQ ID NO: 52

Polynucleotide sequence of Tet1CD

CTGCCAACCTGCAGCTGTCTTGATCGAGTTATACAAAAGACAAGGCCATATTATAACACCTTGG
 GGCAGGACCAACTGTTGCTGTCAGGGAAATCATGGAGAATAGGTATGGTCAAAGGAAACGAA
 TAAGGATAGAAATAGTAGTGTACACCGGTAAGGAAGGGAAAAGCTCTCATGGGTGTCCAATTGCTAAG
 TGGGTTTAAGAAGAACGAGTGTAGAAGAAAAAGTTCTTGTGTTGGCCGGCAGCTACAGGCCACCA
 CTGTCACACTGCTGTGATGGTGTGCTCATCATGGTGTGGGATGGCATCCCTCTTCCAATGGCCGACC

- continued

SEQUENCES

GGCTATAACAGAGCTCACAGAGAATCTAAAGTCATACAATGGGCACCCCTACCGACAGAAGATGCA
 CTCATGAAATCGTACCTGTACATGTCAGGAATTGATCCAGAGACTTGTGGACTTCAATTCTCTTT
 TGGCTGTTCATGGAGTATGTACTTTAATGGCTGTAAGTTGGTAGAAGCCCCAGAAGATTAA
 GAATTGATCCAAGGCTCTCCCTACATGAAAAAAAACCTTGAAGATAACTTACAGACTTGGCTACACGA
 TTAGCTCCAATTATAAGCAGTGTCTCAGTAGCTTACCAAATCAGGTGGAAATATGAAAATGTTGC
 CCGAGAATGTCGGCTTGGCAGCAAGGAAGGTGACCCCTCTGGGGTCACTGCTTGCTGGACTTCT
 GTGCTCATCCCCACAGGGACATTACAACATGAATAATGGAAGCACTGTGGTTGACCTTAACCGA
 GAAGAGAACCGCTCTGGGTATTCCTCAAGAGTGGAGCACTGTGCTACCTTATAAGCT
 TTCAGACACAGATGAGTTGGCTTCAAGGAAGGAATGGAAGCCAGATCAAATCTGGGGCATCGAGG
 TCCCTGGCACCCGCCGAAAAAAAGAACGTTCACTCAGCTGTTCCCGTTCTGGAAAGAAGAGG
 GCTGCGATGATGACAGAGGTTCTGCACATAAGATAAGGGCAGTGGAAAAGAAACCTTCCCGAAT
 CAAGCAGAAGAAATACTCAACAAACAAACAGTAAGCCTCTGCACTGCCAACCTTAGGGAGTA
 ACACGTGAGACCGTGCACCTGAAAGTAAAGGTAAGGCAACCCCATTATCTAAAAGTTCAGAC
 AACACTAAAACCTATTGCTGATGCCATCCCTCCACCCAGTAAAGAGGCATCTCAGGGTCTC
 CTGGTCCCCGAAGACTGCTTCAGCCACACCAGCTTCACTGAAGAATGACGCAACGCCTATGCGGT
 TTTCAGAAAGAACGCACTTCCCTGTCAGATGCTTCGGGAAGACTCAGTGGTGCAATGCTGCA
 GCTGCTGATGGCCCTGGCATTCAGCTGGCGAAGTGCGCTCTCCCAACCCCTGTCCTGT
 GATGGAGCCCCCTATTAAATTCTGAGCTTCACTGGTGACTGACCGCTAACGGCTCATCAGCAA
 ACCACCAAGCCCTCTTCCACCTCTCCTCAAGACCTTGCCCTTCTCCAATGGAAGAAGATGAGCAG
 CATTCTGAAGCAGATGAGCCTCATCAGACGAAACCCCTATCTGATGACCCCTGTCACCTGCTGAGGA
 GAAATTGGCCCATCCACCTGTCACGGCTCGGTTTGATGGTGTGCCCCGGAGAGCTGACGCTACC
 ACTCTGTTGAGCACCCCAACCGTAATCATCAAACCGCCTCCCTGTCTTACAGCACAAAAAA
 CCTAAATAAGGCCAACATGGTTGAACTAAACAGATTAAGTTGAGGCTAAAGAGCTAAGAATA
 AGAAAATGAAGGCTCAGAGCAAAAGCAGCTAATGAAGGTCCAGAACAGTCTGAGTA
 ATGATTTGAACCAAATTCTCTCATAAAGCATTAACCCATGACAAATGTTGTCAACCGTGC
 CCCTTATGCTCTCACACAGTGTGCGGGCCCTATAACCATTGGTC

SEQ ID NO: 53

Protein sequence for VPH

DALDDFDLDMGLSDALDDFDLDMGLSDALDDFDLDMGLSPSASVEFEGSGGPG
 QISNQALALAPSSAPVLAQTMVSSAMVPLAQPPAPAPVLTGPQPSLSAPVPKSTQAGEGTLSALL
 HILQFDADEDLGLGNSTDPGVFTDLASVDNSEFQQLLNQGVSMSHSTAEPMLMEYPEAITLEVTSQ
 RPPDPAPTPGTGTLPNGLSGDEDFFSIADMDFSALLSQISSSGQGGGSGFSVDTSALEDFLSPSVT
 VPDMSLPDLDSLASIQELLSQEPPPRPEAENSPDSDGKQLVHYTAQPLFLDPGSVDTGSNDLPVL
 FELGEWSYFSEGDFGAEDPTISLLTGSEPPKAKDPTVS

SEQ ID NO: 54

DNA sequence for VPH

Gatgcttagacgatttgacttagatatgcttggtcagacgcgttagacgacttcgacccat
 gtttaggtcgatgcattggacactcgatttagatatgttggctccgtgccttagatgacttt
 atctagatatgtctagggtcaactccacccgcgcgcgtcgactgtcgacggcggggccctcagg
 cagatcgcacccaggccctgtgtcgcccttagtccgcgtccaggactgtggccaggactatgg
 ctctagtgtatgtgtgtctcgcccaacctgtcccgcccccgtgtgtgaccccaggaccccc
 agtgcactgacgcgcggccgttgcacacaggccggcgaggactctgtgtgtgcgttgc
 caccgtgcgtgcgtgtgcgtggactgtgtgtggaaacgcgcgcgttgcgttgcgttgc
 cacaatgtggccctcggtggaaactctgtgtttcgcgtgcgtgtgtgcgttgcgttgc
 atagtagcgcgaaccaatgtgtatgggtgtacccgcgttgcgtgtgtgcgttgcgttgc
 cggcccccgcgtccactccctggaaaccgcgcgtgtgtgtgtgtgtgtgtgtgtgt
 agacttctcaagcatcgctgtatggacttgtgtgtgtgtgtgtgtgtgtgtgtgt
 gaggagggtggaaacgcgttgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt
 gt
 ggaggcccccgcgtcccggtggcggcggcggcggcggcggcggcggcggcggcgg
 cggcgcgcgt
 tttagtgtggagagggtcttactttccgtggggacggcgttgcgtgtgtgtgtgt
 gtcgtacagggtcggtggacccacttcccaaggccaaggacccactgtgttgc

SEQ ID NO: 55

Protein sequence for VPR

DALDDFDLDMGLSDALDDFDLDMGLSDALDDFDLDMGLSPKKRKVGSQYLPDTD
 DRHRRIEKKRKYTYETFKSIMKSPFSGPTDPRPPPRIAVPSRSSASVPKPAAPQPYPTSSLSTINYD
 EFPFTMVFPSGQISQASALAPAPQVLPQAPAPAPAMVSLAQAQPAPVVLAPGPPQAVAPPKPT
 QAGEGTLSALLQFQDEDLGLGNSTDPGVFTDLASVDNSEFQQLLNQGIPVAPHTPEPMLMEYP
 EAITLEVTSQARPPDAPAPLPGAPGLPNGLLSGDEDFFSIADMDFSALLSQISSGGSGRSRREGMF
 LPKPEAGSAISDVFEGRREVQCQPKRIRPFHPPGSPWANRPLPASLAPTPGTGVHEPVGSLTAPAPVQPL
 DPAPAVTPEASHLLEDPDEETSQAVKALREMADTVIPQKEEAIICQMDLSHPPRGHLDELTTLES
 MTEDLNLDSPLTPELNEILDFTLNDECLLHAMHISTGLSIFDTSLF

SEQ ID NO: 56

DNA sequence for VPR

gtgcttagacgatttgacttagatatgcttggtcagacgcgttagacgacttcgacccat
 gtttaggtcgatgcattggacactcgatttagatatgttggctccgtgccttagatgacttt
 atctagatatgtctagggtcaactccacccgcgcgcgtcgactgtcgacggcggggccctcagg
 gatagacccaggccctgtgtcgcccttagtccgcgtccaggactgtggccaggacacagat

-continued

SEQUENCES

gcccccttcgggtccgaccgatcccaggccccccacccgagaaggattgcggcccgatccgtcccgatccgtcgccgg
ccagcgtgcggcaagccgtgcggccgcagccccatcccccgttcacgcgtccgcgtccgcgtccgcgtccgcgtccgcgt
gagttcccgacgatgggttccccctcgggacaataatctcacaaaggccgtccgcgtccgcgtccgcgtccgcgtccgcgt
ccaaatgtcccttccggcaaggccgtcccccggccaggccgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgt
ctggccgtcccgatccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgt
caggccggagagggaaacacttcggcaagcaacttccagttcaactccagttgtatgcgaggatcttggag
actcttggaaaactccgcacgacccctcgccgtttaaccgacttcgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgt
agcagcttttgaaaccagggtatcccggtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgt
gaagacatcaccgagacttgcggccaggccgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgt
acctggccctccataccgacttttgagccgcgcagggatccctccatccgcgtccgcgtccgcgtccgcgtccgcgtccgcgt
cagcccttgcgtcacaaggattccagccggctctggcagccggcagccgggatccaggaaaggatgttt
ttggccgaaggccgtggccggccgtccgcgtatccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgt
acgaatccggccatccatccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgt
caacaccaaccggccgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgt
gatccagccggccgcagtactcccgaggccagtcacccgttgcggaggatccgcgtccgcgtccgcgtccgcgtccgcgt
ggctgtcaaaggccctcgggagatggccgtactgtgatccccagaaggaaaggccgtccgcgtccgcgtccgcgt
gccaatggccatccatccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgt
atgcggccggatctgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgt
gaacgcgactgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgtccgcgt
tt

SEQ ID NO: 57

Protein sequence for MCRS1

YKKAGSTMDKSQGLDSSLMASGTTASRSEDEESLAGQKRASSQALGTIPFKRRSSSRFIKRK
KFDDDELVESSLAKSSTRAKGASGVEPGRCSGSEPSSEKKVSKAPSTPVPPSPAPAPGLTK
RVVKSQSPLQVTKDGLRWKPADDLILINAVLQTNDLTSVHLGVKESCRFTLREQRWYALL
YDPVVISKLACQAMROLHPEIAAIQSKALFSKAEBEQLLSKVGSTSQPTLETFQDLLHRHPDA
FYLARTAKALQAHWQLMKQYLYLLEDQTVPQLPKGDQVLNFSDADELDISSLKLDKMRDVELEH
ELMVADRRQKREIRQLEQELHKGWQVLVDSITGMSSPFDNQTLAVLRGRMVRYLMSREITL
GRATKDQNQIDVDSLSELEGPAWKISRKQGVILKLNNSGDFFIANEGRPRIVIDGRPVLCGSKWRL
SNNSVVEIASLRFVFLINQDLIALIRAEAKITPQQLDPAFL

SEQ ID NO: 58

DNA sequence for MCRS1

gtacaaaaaaaaaggcgtccacatggacaaagattctcgaaaaacttcgagggctgttagattcatccctga
tggcatcaggcactgcacccgctcagaggatgaggagtacttgcagggcagaagcgcagcc
tcccccggccatccataaaccggaaagctccatccaggatcatcaagaggaa
gaaggctcgatgtatcgctggggagggcagcgtccggaaaatcttccatccccggcaagggg
ccatgtgggtggaaacccaggcgctgttcggggaggtaaacccctcccaagtgtgaaagaagaa
gtatccaaaggccccagcacttcgtgcacccagccagccccagccctggactcacc
gcgtgtgaaagaagataaacaggccatctcagggtgaccaaggatctggccctggaaagectg
cagatgcaccttcgtcataatcgatgttgcacccaggccatggccatccatcttgc
ggcgtgaaattcagtcggccatcccttcgggggttcaggagcgttgtaacccctgt
ctacgatccatccaaatgtggccctgcaggcatacgaggcagatgtccacccagg
ttgcagccatccagacaaggccctgtttagcaaggctgaggagcagctgtgagcaagt
ggatcgaccaggccacccatctggagacccctccaggacactgtgcacagacacc
cttctacccggccatccggcgtaccggcgaaggccctcgaggccatggcactgt
acccatgtggaggaccagacagtgcagccgtgcacccaaaggggacaatgtgt
gatgcagaggactgtatgcacatgtcaaggatgcagatgtggactgttgc
tgagctgtatgttgcacccggccagaagcggagatccggcagctggaaacagg
ataagtggcagggtgtatggacacatcacaaggcatgtatgttgc
acacttgcgcgtgtccggggccgcgtatggcgggtacttgtgcgtcg
ggcgcaggacaaaccaggataaccaggatgtatgtggacctgt
agataccggaaacaagggttcatcaaggtaagaacaaagg
gagggtgcacggccatctacatcgatggacggccgggtgtctgt
cagcaacaacttgcgtggagatgcggcagccgtgc
tcattgcctcatcagggtgaggctgccaagatcac
tac

SEQ ID NO: 59

SEQ ID NO: 39

Protein sequence for CT07/B
MTLDMDAVLSDFVRSSTGAEPGLARDLLLEGKWNWDVNAALSDFQLRQVHAGNLPPSESEGSSG
SRTPEKGFSDRPTEPRPPRLILQRQDDIVQEKEKRLSRGISHASSIVSLARSHVSNNGGGGSN
EHPLEMPICAFCAQPLPDLTIVNEYEDERSFIERDLEI EQSMLVALEQAGRLNWVWSVDPTSQRLLPL
ATTGDNCLLHAASLGWGFHDRLMLRKALYALMEKGVEKEALKRRWRWQQTQQNKEGSLV
YTEDEWQKEWNELIKLASSEPRMHLIGINGANCGGVESSEEPVYESLEEFHVFLAHVLRRPI
VVVADTMLRDGSGEFAPIPFGGIYLPLPEVPASQCHRSPLVAYDQAHFSALSVMEQKENTK
EQAVIPLTDSEYKLLPLHFADVGPKGWEGWKDDDSNDVRLASVLISLEVKLHLLHSYMNVKWI
PLSSDQAQAPLAQPEPSATASAGDEPRSTPESGDSKDSECVSGSSSTSNGGRRKDRDKEK
KRADSVANKLGSFGKTLGSKLKKNNMGGLMHSKGSKGPGVGTGLGGSSGTETLEKKKKNSLKS

-continued

SEQUENCES

WKGGKEEAAGDPVSEKPPAESVGNGSKYSQEVMQSLSLIRLTAMQGEKFIFIVGTLKMGRHREHQYQEEMIQRYLSDAERFLAEQKQKEAERKLMNGGIGGGPAAKKPEPDAREEOPGTGPAAE~~SRAMAFTSTGYPGDFTIPRSPGGGVHQCPEPVRQLLAGGPCVGLPPTATFPRQCPCGPYPIHODNIPSLEPGSHSKDGLHGRALLPPYYRVADSYNSNGYREPPEDGWAGGLRLGPPTQTCKCQPNCSFYGFETNNFCSCCYREELRRREREPELDGELVHRFLDPAFLY~~

SEQ ID NO: 60

DNA sequence for OTUD7B

SEQ ID NO: 61

Protein sequence for LDB1
MLDRDVGTPMPYPPITLEPGIGRHTPYGNQTDYRIFELNKLQNWTTECDNLWWDAFTTEFF
EDDLAMTIEFICLDEGDPKRYTIGTRLIPRYFRSIFEGGATELYVVLKHPKEAFHSNSVSLDD
QGSVMTQHGPMPFTCQVEGRHLYLEFMFDMMRKTWHFSIRQHRELPIRSILAMHAQDQM
LDQLSKNITRCGLSNSTLNYLRLCVILEPMPQELMSRHKTYSLSPRDCLKTCFLPKWQRMVAP
PAEPTRRQQPSKRKRKRMSSGGNTNSNSKKSPASTFALSSQVPDVVMVGEPLTM
GGEFCDDEDELTIRLRENTQFDAANGIDDEDSENNSPALGANPWNNSKPPSQESKSENPTSQ
ASOLDAPFLY

SEQ ID NO: 62

DNA sequence for LDB1

DNA sequence for L31

```
atgtcgataggatggatgtggggccaactccatgtatccgcctacataccttgagccaggat  
tgggaggcacacaccatatggcaaccaaactgactacagaatatttgagcttaacaaacggc  
ttcagaacttggacagaggagtgtgacaatcttcgtggatgcattcacgactgagtttctt  
gaggatgtatggccatcaactttgcctggaggatggaccaaaagatataccat  
tggccggacccatgtccccacttccggacatctttgggggtgtacggagctgt  
actatgttctaaggcccccaaggaggcatccacagcaacttggtcctcgacttgtac  
cagggcagcatggtggaccacatggcaagccatgttccaccagggtgtgtggaggccg  
gttgtacccggatgttgcacatgtatggggataaaagacgtggacttcagcatcc  
ggcagaccggaggtatccccccggacatcttgcctgcacatgtatggccaaaccccgat  
ttggatcagcttccaaaacatctcggttggggctgtccaaatcttcaacttcagat  
ccgactctgtgtatctcqaaqcccatqaaqactctatqtcacqccacaqactacacc
```

- continued

SEQUENCES

tcagccccccgcgactgcctcaagacgccttccagaagtggcagcgcatggtagcaccc
cctgcggagccacacgtcagcagccagaaacgcggaaacggaaatgtcaggggcag
caccatagactctgggtggcaacaccaacaacagaacagaagaagagccactgata
gcacccctgcgcctctccagccaggtaactgtatgtatgtggggggagccaccctgtat
ggcggggagttgggggacgaggacgaggatcatcaccggctggagaacaccaggatgtga
cgccggccaaacggcattgacgacgaggacagcttaacaactccctgcactggcccaaca
gcccctggaaacagcaagccctccgtccagccaagaagcaaatcgagaacccacgtcacag
gcctcccaagttggacccagttcttgat

SEQ ID NO: 63

Protein sequence for NFKB1B
MAVGACLGKAADADEWDSGLGSLGPDAAAPGGPGLGAELGPGLSWAPLVFGYVTEDGDTAL
HLAVIHQHEPFLDFLLGFSAGTEYMDLQNDLQTAHLAAILGETSTVEKLYAAGAGLCVAE
RRGHTALHLHACRVRGAHACARALLQPRPRPRPREAPDTYLAQGPDRTPDTNHTPVALYPDSDE
KEEEESEEDWKLQLEAENYEGETHPLHVAVIHKDVMEMVRLRDAAGADLDKPEPTCRSPHLA
VEAQAAVDVLELLLRAGANPAARMYGGRTPLGSAMLRPNPILARLLRAHGAPEPEGEDEKSGP
CSSSSSDSGDEGDEYDDIVVHSRSQTRLPPTPASKPLPDPDPRV

SEQ ID NO: 64

DNA sequence for NFKB1B
ATGGCTGGGGTCCGTGCTTGGAAAAGCTGCCGACGCAGATGAATGGTGCGACAGCGGCT
GGGCTCCCTGGGTCGGACGCAGCGGCCCCGGAGGACCTGGTTGGCGGGAGTTGGCC
CGGGGCTGTCGGGGCTCCCTCGTCTCGGTACGTACTGAGGATGGGACACGGCACTG
CACTTGACTGTGATTCATCGATGAACCCCTCTCTGGATTCTCTTAGGCTTCTCGGCC
CACTGAGTACATGGACCTCGAGAATGACCTAGGCCAGACAGCCCCTGCACCTGGCAGGCCATCC
TGGGGGAGACATCACCGTGGAGAAGCTGTACGGCAGCAGGGCCGGCTGTGTGCGGAG
CGTAGGGGCCACACGGCGCTGCACCTGGCTGCCGTGGGGCACACGCCCTGTGCCGTGC
CCTGCTTCAGCCCCGG
CTGACCGTACTGGCCGACACCAACCCATACCCCTGTGACCCGATTCCGACTTGGAG
AAGGAAGAGAGGAGTGGAGGAGCTGGAAAGCTGAGCTGGAGGCTGA AAAATACAGAGGG
CCACACCCCACTCCACGTGGCGTTATCCACAAAGATGTGGAGATGGTCCGGCTGCTCGAG
ATGCTGGAGCTACCTGACAACCGGAGCCACGTGCGGGGGAGCCCCCTTATTGGCA
GTGGAGGGCCAGGGAGCGATGTGCTGGAGCTTCTGTGAGGGCAGGGCGAACCTGTG
CCGCGATGTACGTGGCCGACCCACTCGGAGTGCCTGCTCCGGCCAACCCATCTCG
CCGGCCTCTCGTGCACACGGAGCCCTGAGGCCGAGGGCGAGGACGAGAAATCGGCC
TGCAGCAGCAGTAGCGACAGCGACAGGGAGACGAGGGCGATGAATACGACGACATTGTGGT
TCACAGCAGCCGAGCAAACCCGGCTGCCCTCCACCCAGCCTAAACCTTCTCTGACG
ACCCCGCCCCGTGTGA

SEQ ID NO: 65

Protein sequence for RelB
MLRSGPASGPSPVTGRAMPSRVARPPAELPALGSPDLSSLALSRSTDLEI IDEYIK
ENGFLDGGQPSPGEGPLPRLVSRGAASLSTVTLGPVAPPATPPPWCPLGRLVSPAPGPQ
PHLVITEQPKQRGMFRYECERGSAGSILGESSTEAKSTLPAIELRDCGGLREVEVTACLVW
KDWPFRVHPHSLVGDKCTDGICRVRRLRPHVSPRHSENNLGIQCVRKKEIEAAIERKIQLGID
PYNAGSLKNHQEVDMNVVRICPQASYRDQQGMRRMDPVLSPEVYDKSTNTSELRICRINK
ESGPCTCGEEELYLLCDKVQKEDISVVFSSRASWEGRADESTQADVHRQIAIVEKTPPYEDLEIV
EPVTVNVLQRLTDGVCSEPLFTYLPRDHSDSYGVDDKKRKGMPDVGLELNSDPHGESKR
RKKKPAIDLHFLPNHGSGPFLPPSALLPDPDFSGTVSLPGLEPPPGPDLDDGFAYDPTAP
TLFTMLDLPAPPHASAVVCSGGAGAVVGETPGPEPLTLDSYQAPGPGDGGTASLVGSNMF
PNHYREAAGGGLLSPGPEAT

SEQ ID NO: 66

DNA sequence for RelB
ATGCTTCGGCTGGGCCAGCCTCTGGCCGTCCGCTCCCCTGGCCGGCATGCCGAGTCG
CCGCCTGCCAGACCGCCGGCTGCCGGAGCTGGGGCTTAAGGTCCCCGACCTCTCCT
CACTCTGGCTCCGCTTCCAGGAGCACAGATGAATTGGAGATCATCGACAGGTACATCAAG
GAGAACGGCTTGGGCTGGGGGGAGCAGCTGGGGGGGGGGGGGGGGGGGGGGGGGGGG
GTCTCGGGGGTGGCTCCCTGAGCACGGTACCGCTGGGGCTGTGGCGCCCCGGCCACGC
CGCCGCTTGGGCTGCCCTGGGGGACTAGTGTCCCCAGGCCGGGGGGGGGGGGGG
CCGCACCTGGTACACGGAGCAGGCCAACAGCAGCGCCATGCCCTCCGCTACGAGTGC
GGGGCCCTGG
CCATCGAGCTCCGGGGATGTGGGGAGGGCTGGGGGGAGGTGACTGCCCTGGTGTGG
AAGGACTGGCTCACCGAGTCACCCCTGAGCTGGGGAAAGACTGCACCGACGGCAT
CTGCAGGGTGGGGCTCCGGCTCACGTACGCCCTGGGGGACAGTTAAACACTGGGAC
AGTGTGTGAGGAAGGAGATTGAGGCTGCCATTGAGCGGAAGATTCAACTGGGATTGAC
CCCTACACGCTGGGCTCTGAGAACCATCAGGAAGTAGACATGAATGTGGTAGGGATCTG
CTTCCAGGGCTCATATCGGGAGCAGCGGGAGATGCGCCGGATGGATCTTGTGCTTCCG
AGCCCGTCTATGACAAGAACATCAGGCTGGGGATTTGGGGAAATTAACAAG
GAAAGCGGGCGGTGCACGGGGGGGGAGGAGCTACTTGCTCTGGCAGAACGGTGCAGAAGA
GACATACAGTGTGTTCAGCAGGGCTCTGGGGAGGTGGGGCTGACTTCTCCAGGGCG
ACGTGACCGCAGATTGCCATTGTGTTCAAGACGCCCTACGAGGACCTGGAGATTGTC
GAGCCCGTGAAGTCACACGTCTCTGAGCGGGCTACCGATGGGGCTGAGCGAGCATT
GCCTTCACTGACGCTCGCAGGACCATGACAGCTACGGCGTGGACAAGAGCGGAAACCGG

- continued

SEQUENCES

GGATGCCGACCTCCTGGGGAGCTGAACAGCTCTGACCCCCATGGCATCGAGAGCAAACGG
CGGAAGAAAAAGCCGGCATCTGGACACTTCCCTGCCAACCCACCGCTAGGCCGGTTCT
CCCGCGTCAGCCCTGCTGCCAGACCCCTGACTCTCTCTGGCACCGTGCCCTGCCCGGCC
TGGAGGCCCTGGCGGGCCCTGACCTCTGGACGATGGCTTGCCTACGCCCTACGGCCCC
ACACTCTTCAACCATGCTGGACCTGCTGCCCGGCCACCGCACACGCTAGCGCTGTTGTG
CAGCGGAGGTGCCGGGGCGTGGTGGGAGACCCCCGGCCCTGACCACTGACACTGGACT
CGTACCAAGGCCCGGGGGGGGGGGATGGAGGACCAGGCCAGCCTTGCGCAGAACATGTTC
CCAATATTACCGCGAGGCGGCCCTTGGGGCGGCCCTCTATCCCGGGGCTGAAGCCAC
GTAG

SEQ ID NO: 67

Protein sequence for CITED2
MADHMMAMNHGRFPDGTLHHHPAHRMGMGQFPSPHHQQQPQHAFNALMGEHIHYGAGN
MNATSGIRHAMPGPTVNGHPPSALAPAARENNSQFMGPPVASQGSSLPAQMQLKLNNOYF
NHHYPYPHNHYMPDLHPAAGHQMNNTQHERDCNPKHSGGSSTPGGSGGSSTPGGSSSGGG
AGSSNSGGSGSGSNMPASVAHPAAMLPPNVIDTFIDEELVMSLVIEMGLDRIKELPELWL
GQNEFDFTDFVCKQQPSRVSCLDPAFLY

SEQ ID NO: 68

DNA sequence for CITED2
atggcagacccatatgatggcaatgaaccacggcgcttcccgacggccaaatgggctgca
ccatcaccctgccacccgcatggcatgggcaggatcccgagccccatcacccaccagcagc
agcagccccagcacgccttaacgccttaatgggcagacatcacactacggcgccggcaac
atgaatgcacagcagccatgcgcgtggggccggggactgtgaacggaggccaccc
cccgagccgcgtggggccggccagggttaacaactccaggatgggtcccccgggtgg
ccagccaggagggtccctgcggccaggatgcagtcagaagctcaacaaccaggatattc
aaccatcacccttaccccaacaccatcacatgcggatgtgcacccctgtgcaggccacca
atgtggacgggacaaccaggacttcggagatggcaaccccaagcacagccggccaggca
ccccggccggccggccggccggccggccggccggccggccggccggccggccggcc
ggccggccggccggccggccggccggccggccggccggccggccggccggccggcc
cgccggccggccggccggccggccggccggccggccggccggccggccggccggcc
ttatgtcttggtatgggttggaccgcataaggagctggccgaactctgg
ggccaaaacgagttgtatggacttcgtgtcaacacgcggccaggccaggatgt
ctgtttggaccacccagcttcttgat

SEQ ID NO: 69

Protein sequence for ScFv-sfBFP-MCRS1

NLS-**ScFv**-linker-**sfBFP**-linker-**MCRS1**

MGPKKKRKVGGMGPDIVMTOSFSSLASAVGDRVTITCRSSTGAVTTSNYASW
QKPGKLFKGLVGGTNNRAPGVPSRFSGSLIGDKATLTISLQPEDFATYFC
LVVSNHWVFGQGTTKVELKRGGGGGGGGGGGGGGGSEVKLLESGGGLVQP
GGSLKLSCAVSGFSLTDYGVNWVRQAPGRGLEWIGVIWGDGITDYNALKRE
IISKDNKGNTVYLQMSKVRSDDTALYYCVTGLEDYWGQGTLLTVSSYPYDVPD
YAGGG
NGHKFSVRGEGEGLDATNGKLTLKFICTTGKLPVPWPTLVTLTHGVQCFSRYP
DHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKG
DFKEDGNILGHKLEYNPNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADH
YQONTPIGDGPVILLPDNHYSTQSVLSKDPNEKRDHMVILLEFVTAAGITHGMD
ELYKGGGRTGG
MDKDSQGLLDSSLMASGTAARSDEEESLAGQKRASSQALGTTPKRRSSSRFIK
RKFFDDDELVESSLAKSSTRAKGASVGPGRCSGSESSSEKKVSKAPSTVP
PSPPAPGLTKRVKKSQPLQVTKDLGRWKPADDLLINAVLQTNLTSVHLG
VKFSCRTFLREVQERWTALLYDPVISKLACQAMRQLHPEATAIAQSKALFSKA
EEQLLSKVGSTSQPTLETFQDLLHRHPDAFYLARTAKALQAHWQLMKQYYLLE
DQTVQPLPKGDQVLNFSDAEDLIDDSKLKDMDRDEVLEHELMVADRQKREIRQ
LEQELHKWQVVLDSITGMSSPDEDNQTLAVLRLGRMVRYLMRSREITLGRATKD
NQIDVDSLLEGPAWKISRKGVIKLKNNNGDFIANEGRRTYIDGRPVLCGSK
WRLSNNSNVVEIASLRFVFLINQDLIALTRAEEAKITPQLDPAFL

SEQ ID NO: 70

DNA sequence for ScFv-sfBFP-MCRS1

atgggtCCAAAGAAAAAGAGAAAGGTGcggtggcatggggcccgacatcggtatgaccagg
cccgaggccgtggccaggccgtggccggccggccgtggccaggatcacctggccggccagg
ggccgtggccaggccgtggccggccggccgtggccaggatcacctggccggccagg
ggccgtggccaggccgtggccggccggccgtggccaggatcacctggccggccagg
gtatggccggccaggccgtggccaggatcacctggccggccggccgtggccagg
tctggccgtgtggatcacggccaggccgtggccggccaggatcacctggccggccagg
cgccggccggccgtggccaggatcacggccaggccgtggccggccaggatcacctggcc
cgaggatcacggccaggccgtggccggccaggatcacggccaggccgtggccagg

- continued

SEQUENCES

gctgcgcggcgtgagccgttcagccgtaccgcgtactacggcgtaactgggtgcggccaggcccc
 ggccgcggccctggagtgatcgccgtatctggggcgacggcatcaccgactacaacagcgc
 cctgaaggaccgcttcatcatcagcaaggacaacggcaagaacaccgtgtacctgcagatga
 gcaagggtgcgcacgcacccggccctgtactactgcgtgaccggccctgttcgactactgg
 ggccaggccacccctggtagccgtgacgcgtaccatcacatcgatgttccagattacgctgg
 aggccggagggtctgggggaggaggtagtggcggtgtttcaggagggcgccggaaagctgg
 atccaggtggaggtgaaagcgtagcaaaggagaactttcactggagtgtcccatt
 ctgttgaattataggtgtatgttaatggcacaatatttctgtccgtggaggggtgaaagg
 ttagtgcataaaccatgttgcacttgcacccatgttcaatgtttccgttatccggat
 ccatgaaacccgcgtactttcaagagtgcacatggccaaagggttatgtacaggaaacgcac
 tatatcttcaaaatgtacgggacccatacgccgtgtcaagtttgcgttgcgttgcgttgc
 ccctgttgcattatgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc
 cacaactcgactacaactttaactcacaaatgtatacatcacggcaaaaaaaa
 tggaatcaaaggtaactcaaaattcccaacacgttgcgttgcgttgcgttgcgttgc
 accattatcaacaaaatactccaaatttgcgttgcgttgcgttgcgttgcgttgc
 ctgtcgacacaatctgttgcgttgcgttgcgttgcgttgcgttgcgttgc
 ttagttgtactctgtccgttgcgttgcgttgcgttgcgttgcgttgcgttgc
 ggaccgggtggcggtggcgggtggaggcggtGCTGACCCCAAGAAGAAGAGGTGGCT
 AGGATCACAAAGTTgtacaaaaggcaggctcaccgtacaaaaaaaggcaggctccaccatgg
 acaaaggatctcagggtgtctgttgcgttgcgttgcgttgcgttgcgttgc
 gaggataggacttcactggcggcagaaggcgactcccccaggcttggcaccatccc
 taacccggaaatcttccagggttgcgttgcgttgcgttgcgttgcgttgc
 gacccgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc
 tcggggagtgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc
 gccacccaggccccccggccctggactcccaaggcgtgttgcgttgcgttgc
 ttccagggttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc
 gtgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc
 ccttccgggggttccagggttgcgttgcgttgcgttgcgttgcgttgc
 ctgtcaggccatgcgttgcgttgcgttgcgttgcgttgcgttgc
 tttagcaaggcttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc
 gaccccccaggactgttgcgttgcgttgcgttgcgttgcgttgcgttgc
 ccttccaggactgttgcgttgcgttgcgttgcgttgcgttgcgttgc
 cctgtccggccaaatccctggcgttgcgttgcgttgcgttgcgttgc
 taaggcttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc
 agaaggcggaggatccggcgttgcgttgcgttgcgttgcgttgc
 atccacggccatgttgcgttgcgttgcgttgcgttgcgttgc
 ggttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc
 ttgtatgttgcgttgcgttgcgttgcgttgcgttgcgttgc
 aagcttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc
 tggacccggccggccgttgcgttgcgttgcgttgcgttgc
 ccaggccgttgcgttgcgttgcgttgcgttgcgttgcgttgc
 gccaaggatcacccatgttgcgttgcgttgcgttgcgttgc
 SEQ ID NO: 71
 Protein sequence for ScFv-sfBFP-OTUD7B
 NLS-**ScFV**-linker-**sfBFP**-linker-**OTUD7b**

MGPKKKRKVGGMGPDIVMTQSFSLSASVGDRVTICRSSTGAVTTSNYASWV
QEKPGKLFKGLITGGTNRNPAGVPSRFSGSILGDKATLTISSLQPEDFATYFC
LWYSNHWVFGQGTKVELKRGGGGSGGGGGGGGGGGSEVKLLESGGGLVQP
GSSLKLSCAVSGFSLTDYGVNWVRQAPGRGLEWIYIWGDGITDYNALKDRF
IISKDNGKNTVYLQMSKVRSDDTALYYCVTGLEDYWGQGLTVTVSSYPYDVPD
YAGGGGGSGGGGGSGGGGGGGGGSLDPGGGGSGSKGEELFTGVPLLVELDGV
NGHKFSVRGELEGDATNGKLTGKFICTTGKLPWPWTLVTLTHGVOCFSRYP
DHMKRHDFFKSAMPEGYVQERTISFKDDGYKTRAEVKFEGDTLVNRIELKG
DFKEDGNILGHKLEYNFSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADH
YQONTPIGDGPVLLPDNHYLSTQSVLSKDPNBKRDHMVLLFVTAAGITHGMD
EYKGGGRTGGGGGGGADPKKRKRVARITSLYKKAGSTM
TLDMDAVLSDFV
RSTGAEPGLARDLLEGKNWDVNAALSDFEQLRQVHAGNLPPSFSEGSGGSRTP
EKFGSDREPTTRPPRPILQRQDDIVQEKRRLSRGISHASSSIVSLARSHVSSNGG
GGGSNEHPLEMPICAFLPDLTVYNEDERSFIERDLIEQSMIVALEQAGRLNW
WVSVDPTSQRLLPLATTGDGNCLLHAASLGMWFHDRVLLMKRALYALMEKV
EKEALKRRWRWQOQTQQNKESEGLVYTEDEWQKEWNELIKLASSEPRMHLGTMGA
NCGGVESSEEPVYESLEEFHVFVLAHVLRPPIVVVADTMLRDSGGEEAFAPIPF
GGIYLPLEPVASOCHRSPVLAYDQAHFSALSMEQKENTKEQAVIPLTDSEY
KLLPLHFAVDPGKGWEWGKDDSDNVRLASVILSLEVKLHLLHSYMNWKIPLS
SDAQAPLAQPEPTASAGDEPRSTPESGDSDKESVGSSSTSNEGGRKEKSRR
DREKDKKRADSVANKLGSFGKTLGSKLKKNMGGLMHSKGSKPGGVGTGLGGSS
GTETLEKKKKNSLKSWKGGKEEAAGDGPVSEKPPAESVGNGGSKYSQEVMQSL

-continued

SEQUENCES

cagcaaacggcggaaacggaaagatgtcagggggcagcacatgagctctgggtggcaaca
ccaaacaaacggaaacggaaagaagggccacgtacggaccctgcgccttccaggccgac
cctgtatgtatggttggggggacccatgtatggggggagttcgggggacggacgaa
gaggctcatccccggctggagaacacccaggttcgacggccaaacggcattgacgagg
acagcttaacaactccccgtcactggggccaaacagccccctggaaacagcaagcccttc
agccaaagaacaaatcgggaaaccccacgtcacggcccccagtggtggaccaggatttt
gtac

SEQ ID NO: 75

Protein sequence for ScFv-sfBFP-NFKBIB

NLS-ScFV-linker-sfBFP-linker-*NFKBIB*

MGPKKKRKVCGMGPDIVMTQSPSSLASAVGDRVITCRSSGTAVTTSNAYASWV
QEKPGLKFLIGGJINNRAGPVSRFGSLIGDKATLTISSLOPEDFATYFC
LYWSNHWFVGQGTKEVLKGRRGGGSGGGGGGGGGGGGGGGGGGGGG
GGSLKLSCAVSGSLTDYGVNNWRQAPGRGLEWIIVIWGDGITDYNALSKDRF
IISKDNGKNVTYLQMSKVRSDDTALYYCTGLEDYWGQGTLVTVSSYPDV
YAGGG
NGHKFPSVRGEGEDATNGKLTLKFICTTGKLPVPWPWTLVTTLTHGVOCFSRY
DHMKRHDFFKSAMPEGVYQERTISFKDDGTYKTRAEVKFEGLDTLVNRIELKG
DFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADE
YQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLFVTAAGITHGMD
ELYKGGRRTGGGGSGGGGADPKKKRKRVARITSLYKKAGSTMAGVACLGKAAD
DEWCDSGLGSLGPDAAPGGPGLGAELGPGLSWAPLVFGYVTEDGDTAHLAV
IHQHEPFLDFLLESAGTEYMDLQNDLGQTALHAAILGETSTVEKLYAACAG
LCVAERRGHTALHACRGVAHACARALLQPRPRRPREAPDPTYLAQGPDRTPDT
NHTPVALYPSDLEKEEEESEEDWKLQLEAENYEGETPLHVAVIHKDVEVMVR
LRDAGADLKDPEFTCGRSPFLHAEQAQADVLLELLRAGANPAARMYGGRTPL
GSAMLRPNPILARLRAHGAPEPEGEDEKSGPCSSSSDSGSGDEGDEYDDIVV
HSSRSOTRDPPTPASKPLDPDRPV

SEQ ID NO: 76

DNA sequence for ScFv-sfBFP-NFKB1B

atgggtCCCAAGAAAAAGAGAAAGGTCGgtggcatgggccccgacatcggtatgaccaggccc
cccagcagccctgagcgccagcgtggcgaccgcgtaccatcacccgcgcagcaccgc
ggccgtgaccaccaactcgccagactgggtcgaggaaaggccggcgaactgttcaagg
ggcctgatcggccgccaacaaccgcgcggcgctggcccaaggcttcaggccgcgg
gatccgcgacaaggccaccctgaccatcagcagcgtcgagcccgaggacttcggccacact
tcgtcgccctgtgttacagcaaccactgggttccggccaggccacaagggtggactgttaagg
cgccggccggTggAaggcgAgggcgTggGTCTggTggAggcggcgaactCTTggcgAgggagg
cgagggtgaagctgtgtggagggcgccggcgctggcgtcagccggccggcgaactgttcaagg
gtgtccgggtgtcggccgttcaagctcagccactacggcgttaactgttggccggccccc
ggccgcggctggaggtggatcggcgtatctggggcgacggcatcaccgactacaaggcg
cttgcggaggccgttcatcatcagcaaggacaacggcagaagaacaccgtgtactcgatgaa
gcaagggtgcggcagcggcagacccggccctgtactactcggtgaccggcgttgcact
ccgcaggggccaccttgcgtaccggcgttgcgtactaccatcatacgatgttcaggatcgttgc
aggcggagggtttctggggggagggttgtggcggttgtttcaggaggccggaaactgtt
attcagggtggagggtggaaaggcgttgcacaaggagaacaatcttcactgggttgttccaa
cttggtaatagatgtgttgcataatggcacaatttttgcgttgcggagggttgaaagg
tgtatgttgcacaaggaaaactcacccttaatttttgcactactcggtttcccgat
cgtggccacacttgtactactctgaccatgggttcaatgtttcccgatcccgat
cacatgaaacccgtatgtttcaaggtgtccatggccgaagggttgtatcaggaaacgc
tatatatcttcaagatgacggccatcacaaggcgtgttgcgttgcgttgc
ccctgttgcataatcgtatcgatgttgcggatcacatcgatggatgagcttgc
cacaactcgatcacaactttaactcacacaatgtatacatcaccgcagacaaacaaaaggaa
tggatcaaggtcaacttccaaaattccggccacaacgttgcgttgcgttgc
accattatcaacaaaatcccaatttgcgttgcgttgcgttgcgttgcgttgc
ctgtccgacacaatctgttgcgttgcgttgcgttgcgttgcgttgcgttgc
tgagtgttgcatactgtgttgcgttgcgttgcgttgcgttgcgttgcgttgc
ggacccgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc
AGGATCACAAGTGTtgcataaaaaaggcaggccatccactGGCTGGGTGCGTCTGGAAA
AGCTGGCCACGGCAGATGAATGTGCGCACAGGCCCTGGCTCTGGTGGCCGACAGGGG
CCCCGGAGGACCTGGTTGGGCGCGAGGTGGGCCGGGCTGTGTTGGCTCCCTCGTC
TTGGGCTACTGCTACTGAGGATGGGACACCGCCTACTGCACTTGGCTGTGATT
ACCATCAGCATGAG
ACCCCTCTGGATTTTCTTAGGCTTCTCGGCGGCCACTGAGTACATGGACCTCGAGA
ACCTAGGCCAGCAGCCCTGACCGTCTGGCAGGCATCTGGGGAGACATCAGCGTGGAGAG
CTGTACGCCAGCAGGCCGGCTGTGTTGGGGAGCGTGAAGGGGCCACACGGCGTGCACCT
GGCCTGCCGTGTGGGGCACACGCCCTGTGCGCGTGCCTCAGCCCCGCCCCGGCGCC
CCAGGGAAGGCCCGACACACTACCTCGCTCAGGGCCCTGACCGTACTCCGACACAA
ACCCCTCTGGCTTGTACCCGGATTCTGGCAGAAGGAAGGAGGAGTGGAGGAG
ACCCCTCTGGCTTGTACCCGGATTCTGGCAGAAGGAAGGAGGAGTGGAGGAG

-continued

SEQUENCES

CTGGAAGCTGCAGCTGGAGGCTGAAAACACTAGAGGGCACACCCCCACTCCACGTGGCGTTA
TCCCAAAGATGTCGGAGATGTCGCCGCTGTCAGATGTTGGAGCTGACATTGCAAAACCG
GAGCCCCACGTGCGCCGGAGCCCCCTCATTTGGCAGTGAGGCCAGGAGCGATGTGCG
GGAGCTCTCTGAGGGCAGGGCGAACCTCTGCTCCGCAGTGAAGGTGGCCGACCCCCAC
TCGGCAGTGCATGCTCGGCCAACCCCATCTCTGCCCGCTCTCCGTGACACGGAGCC
CTCTGAGCCCCAGGGCGAGGAGCAGGAATCTGGGGCTGCAAGCAGTAGGCAGCGACAG
GGAGAGCAGGGGGCATGAATAGCAGCACATTGTGGTTCAAGCAGCGCAGCCAAACCCCC
TGCTCTCCACCCCCAGCCTAAACCCCTCTTCTGACGACCCCCGGCCGTGTA

SEQ ID NO: 77

Protein sequence for ScFv-sfBFP-RelB

NLS-ScFV-linker-**SfBFP**-linker-*RelB*

MGPKKRKGVGGMGPDIVMTQSPSSLSASVGDRVTITCRSSTGAUTTSNSYASW
QEKPGLKFLGLIIGGTNNRAPGVPGRFSGSLIGDKATLTISLQPEDFATYFC
LWYSNHNWFQGQTKVELRKRRGGGGGGGGGGGGGGGGGGGGGGGGGG
GGSLLKLSACVGFSLTDYGVNWRQAPGRGLEWIGVIWGDGTIDYNSALKDRF
IISKDNGKNTVYLQMSKVRSSDTALYYCTVGETDYGWQGTLTVTSSYPDV
YAGGG
NGHKFESVRGEGEGEDATAKGTLKFTTGKLPVFWPTLVTLTHGVQCFSRYP
DHMKRHDFFKSAMPEGYVQERTISFKDDCTYKTRAEVKFEQDPLVNRIELKG
DFKECDGNILGHKLEYNFNHNVYITADKQKNGIKANFKIRHNVEDGSVQLADH
YQONTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLFVTAAGITHGMD
ELKYKGGRRTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
TGRAMPSRRVARPPAAPELGALGSPDLSSLSSLAWSRSTDELEIIDEYIKENG
GLDGGOPGPGEGLPLRLVSRGAASLSTVTLGPVAPPATPPWCPLGRLVSPAP
GPGPQPHLVITEQPKQRGMFRYCEGRSAGSILGESSTEASKTLPAILRDC
GGLREVEVTA CLVWKDWDPRHVRPHSHLSVKGCDTGDICRVLRPHVSPRHSEN
GIQCVRKKEEAAIERKIQGLDIPYNAQSLKNHQEFYDMVNVRICFQASYRDQO
GQMRRMDPVLSEPVDPKKSNTSELRICRINKESGPCTGEEBLYLLDCDKVQE
DISVVFRSASWEGRADFSQADVHRQIAIVFKTPPYDELIEVPEVTVNMFLQR
TDGVCSEPLPFTYLRDHDSYGVDKKRKGMPDVGLELNSSDPHGIESKRRKK
KPAIDLWTBPHNGSGFPFLPSALLPDPDFTSGTWSLPCLEPPGPDLDDGFA
YDPTAPTLFTMLDLPPAPHSASA VCGGAGAVVGTGPPEPLTLDSYQAGC
PGDGDTATLFTMLDLPPAPHSASA VCGGAGAVVGTGPPEPLTLDSYQAGC

SEQ ID NO: 78

DNA sequence for ScFv-sfBFP-RelB

-continued

SEQUENCES

CGGTACCCCTGGGCCCCCTGGGCAGCCCCCAGCCACGCCGCGCCCTGGGGCTGCCCTGGGC
CGACTAGTGTCCCCAACGGCCGGGCCCCGGCCAGCCGACCTGGTCATCACGGAGCACCC
CAAGCAGCGGGCATGGCTTCCGCTACGAGTGCAGGGGGCTCGGGCCGACATCCTTG
GGAGAGCAGCACCGGGCAGCAAGCCTGGCCCGCATGGAGCTGGGGATTGGAGGG
CTGGGGAGGGTGGAGGTGACTGCCCTGGCTGGAGACTGGCCTACCGAGTCCACCC
CCACAGCCTCTGGGAAAGACTGCACCGAGCCATCTGCAGGGTGCCTGGGCTCAGG
TCAGCCCCCGCAGCTTAAACAACCTGGGCATCAGCTGGTGTGAGGAAGAGGATTTGAG
GTCGCTTGGAGCAGGGAAAGATTCAACTGGGCTATTGGCCCTAACACCTGGGCTCTGGAAAG
CCATCAGGAAGTAGACATGAATGTTGGTGGAGATCTGCTCCAGGGCTCATATGGGACAGC
AGGGAGACATGGCCGGGATGATCTGGTGGCTTCCAGGGCGCTATGAGAAGAAATCCAC
AACACATCAGACTGGGATTTGGCAATTAAAGGGAAAGGGGGGGCTGACCCGGTGGCA
GGAGCTACTCTGCTCGGCAAGAGCTGCAAGAAAGGACATATCAGTGGTTCTAGCAGGG
CCTCCTGGAAAGGTGGGCTGACTTCTCCAGGCCAGCTGCACCGCAGATTGCCATTG
TTCAAGACGGCCCTACAGGGACCTGGAGATTGTGCGAGGGCTGACAGTCACCTCTGG
GCAGCGGCTACCGGGTGTGCGAGGCCATTGCTTCAGCTACCTGGCTCGGCC
ATGACAGCTACCGCGTGGCAAGAAGCGGGAAAGGGGGATGGCCAGCTTCTGGGAGCTG
AACAGCTGACCCCCATGGCATCGAGAGAACCGGGAGAAAAAGCCGGCATCTGGG
CCACTTCTGGCCAACCCGGCTCACGGGGCTTCTCCGGCGCTAGCCCTGGCGACCC
CTGACTTCTCTGGCACCGTGTCTGGCCGGCTGGAGCCCCCTGGGGGCTGACCT
CTGGGAGCTGGCTTCTGGCTACGACCTACGGGCCCCCACACTCTCACCATGCTGGACCT
GCCCGGCAACGCCACCGCTAGCGCTGGTGTGCAAGGGAGGTGGGGGGCGTGGT
GGGAGACCCGGGGCTGAGAACCTGACACTGGACTCGTACAGGGGGGGGGGGGG
GGAGGACACCGGGCAGGGCTTGTGGCGACCAACATGTTCCCCAATCATACCGGAGGGGG
TGGGGGGGGCTCTATCCCCGGGGCTGAGAACCGTAG

SEQ ID NO: 79

Protein sequence for ScFv-sfBFP-CITED2

NLS-ScFV-linker-***sfBFP***-linker-CITED2

MGPKKRKGPGDIVMTQSPSSLASVGRVTITCRSSTGAVTTSNYASWV
QEKGPKLKFGLIGGINNRAFGVPVSRSFGSLIGDKATLTISLOPEDFATYFCA
LWYSNHWFVGOGTKVELKRGGGGSGGGSSGGGSEVLLESGGGLVQPP
GGSLKLSCASVGSFLSTDYGVNWVRQAPGRGLEWIYGWIWGDDITDYNALKRE
IISKDKNGNTVYLQMSVKRSDDTALYCTVGLEDWYQGGTTLTVSSYFYDWP
YAGGGCCSGGGGSGGGGSGGGS1DPGGGGSGSKGEELFTGUVPLIYELDGDW
YAGGGCCSGGGGSGGGGSGGGS1DPGGGGSGSKGEELFTGUVPLIYELDGDW

NGHKFSVRGEGEGDAATNGKLTALKFICTTGKLPPWPVTLVTTLTHGVOCFSRYP

DHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKG

DFKEDGNILGHKLEYNFNSHNVYITADKOKNGIKANFKIRHNVEDGSVQLADH

YQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMD

ELYGGGRTGGGSGGGADPKKRKVARITSLYKKAGSTMADHMMAMNHGRE

PDGTGNGLHHHPAHRGMGMGQPPSPHHHQQQQPQHAFNLMGEHIIHYGAGNMAT
SGIRHAMPGPTVNGHHPPSALAPAFARFNNSQFMGPVVASQGSSPLASMQLKL
NNQYHNHHNPYWHNPYHMDDPLPAAQCHGMNTNQHHERDCNDPKHSGGSSTPGGSD
SSTPGGSGSSGGGAGSSNSGGGSGSGNMPASVAHVPAAMLPPNVIDTFDFIGE
EVMLSLSVIEMLDRIKELPLEWLQGNEEDFMTDFVCKQQPSRSVSCLDPAFLY

SEQ ID NO: 80

DNA sequence for ScFv-sfBFP-CITED2

-continued

SEQUENCES

tggaaatcaaagctaacttcaaattccacacaacgttgaagatggttccgttcaactagcag
 accattatcaacaaaatactccaattggcgatggccctgtcctttaccagacaaccattac
 ctgtcgacacaatctgccttcgaaagatccaaacgaaaacgtgaccacatggtccttct
 ttagttttaactgtctgggattacacatggcatggatgagcttacaaagggtggaggctc
 ggaccgggtggcggtggcagcgggtggaggcggtGCTGACCCCAAGAAGAAGGGTGGCT
 AGGATCACAAGTTgtacaaaaaaagcaggctcaccatggcagaccatgtggcaatgaa
 ccacggggcgttcccgacggcaccatgggtcaccatccccccatccccccatgggca
 tggggcagtcccgagccccatcaccaccacgacggcagcagccccagcacgccttcaacggcc
 ctaatgggcgagcacatacactacggcgccggcaacatgaatgccacgagcggcatcaggca
 tgcgatgggcggggactgtgaaacggaggacccccccgagcgcgcgtggccccccggcca
 ggttaacaactcccaacttgcgggtggggactgtggggactgtggggactgtgggg
 agcatgcagctcagaagctcaacaaccaggatattcaaccatccccctaccccaaccaacca
 ctacatgcggatttgcaccttgcaggcaccaggatgacgggacaaacccagacttcc
 gagattgcaaccccaagcacagcggcgccagcagcaccccccggcgctcgccggcagcage
 acccccccgggtctggcagcagctggggcgccggcgccggcagcagcaacagcggcgccgg
 cagcggcagcggcaacatggccctcgtggcccccacgtgtcccgctgcacatgtgcggccca
 atgtcatagacatgatgttcatgcacggaggatgttctatgtccttggatagaatgggt
 ttggaccgcataaggagctggcaacttgcgtggggcaaaacggatgttttatttgcac
 ggacttcgtgtcggaaacagcagcccgacggagtgactgtttggaccaggatgttgcac

SEQ ID NO: 81

Protein sequence for scFv

MGPDIIVMTQSPSSLASVGDRVITCRSSTGAVENTSNYASWVQEKPGLPKLIGGTNNRAPH
 GVPSRFSGSLIGDKATLTISLQPEDFATYFCALWYSNHWFQGQTKVELKRGGGGGGGGG
 GGGGSSGGGSEVKLLESGGGLVQPGGLKLSCAVSGFSLTDYGVNWVRQAPGRGLEWIGVIW
 GDGITDYNALKRDIISKDNKGNTVYLQMSKVRSDDTALYYCVTGLFDYWQGTLVTVSS

SEQ ID NO: 82

DNA sequence for scFv

atggggcccccgcacatctgtatgacccagagccccagcggccctgagcgcggcggcggcgg
 cgttgccatcatctggcgccggcggcggccggccggccggccggccggccggccggccgg
 tgcaggagaacccggcggcggccggccggccggccggccggccggccggccggccggccgg
 ggcgtgcccggccggccggccggccggccggccggccggccggccggccggccggccgg
 cctgcagcccgaggacttcgcacacttctgcgcctgtgttgcacgcggccggccggccgg
 tcggccaggggcacaagggtggactgtgaaagcggccggccggccggccggccggccgg
 ggTggAgcggcggcggcggccggccggccggccggccggccggccggccggccggccgg
 ggtgcagcccgccggccggccggccggccggccggccggccggccggccggccggccgg
 acggcgtgaaacttgcgtggccggccggccggccggccggccggccggccggccggccgg
 ggccggccggccggccggccggccggccggccggccggccggccggccggccggccggccgg
 acggcgtgaaacttgcgtggccggccggccggccggccggccggccggccggccggccgg
 acttcgcgtggccggccggccggccggccggccggccggccggccggccggccggccgg
 ccatacgtttccagattacgcgtggccggccggccggccggccggccggccggccgg
 ttgtggttcaggaggccggccggccggccggccggccggccggccggccggccggccgg

SEQ ID NO: 83

Protein sequence for sfBFP

SKGEELFTGVVPILVELDGVNGHKFSVRGEGEGLATNGKLTLKFICTTGKLPWPWTLVTT
 LTHGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDCTYKTRAEVKFEGDTLVNRIEL
 KGIDFKEDGNILGHKLEYNFNSHNVYITADQKQNGIKANFKIRHNVEDGSVQLADHYQQNTP
 IGDGPVLLPDNHYLSTQSVLSKDPNEKRDMVLLFVTAAGITHGMDELYK

SEQ ID NO: 84

DNA sequence for sfBFP

agcaaaggagaacttttactggagttgtcccaattttgttgcatttttttttttttttttt
 taatgggcacaaaattttctgtccgtggaggggtgaagggtatgttgcacaaacggaaaactca
 cccttaaatatttgcactactggaaaactacacttgcgtggccaaacttgcgtactact
 ctgacccatgggttcaatgtttccgttatccggatcatcataacggcatgtacttttt
 caagagtgcgtccggccaaagggtatgtacaggaaacgcactatatttttttttttttt
 ctac
 aagggtatgttt
 ctcacacacatgtatatacatcaggccagacaaaacaaaggatggaatcaacttcaaaa
 ttggccacacacgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgt
 attggcgatggccctgtcccttaccagacaaccattacctgtcgacacaacttgcgtt
 gaaagatcccaacgaaaacgcgtgaccacatggcccttgcgttgcgttgcgttgcgt
 ttacacatggcatggatgagcttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt

SEQ ID NO: 85

Protein sequence for GCN4 Peptide (which is bound by ScFv)

EELLSKNYHLENEVARLKK

-continued

SEQUENCES

SEQ ID NO: 86
DNA sequence for GCN4 (one example of a sequence for GCN4)
gaggagcttctgagcaaaaactatcacctcgaaaaacgggttgcgcactgaagaaa

SEQ ID NO: 88
 DNA sequence for dCas9-5X-GCN4
 ATGGACAAGAAGTACTCCATTGGGCTGCCATCGCACAAACAGCGTCGCTGGGCCGTAT
 TAGGCCAGGAGTACAAGGTGCCAGAAAAAAATTCAAAAGTTCTGGGCAATACCGATGCCAAC
 GCATAAAAGAAGAACCTCATGGGCCCTCTGTTGCACCTGGGGAAACCGCGAAGGCCACG
 CGGCTAAAGAACACGCCACGGCGAGATATAACCGCAGAAAAGAATCGGATCTGCTACCTcgca
 GGAGATCTTTAGTAGATGGCTAAAGGGTGAECTCTTCCTCATAGGCTGGAGGAGT
 CTTTTGGTGGAGGAGATAAAAAGCACCGAGCGGCCAACATTCTGGCAATATCGTGAC
 GAGGTGGGTGACATGAAAAGTACCCAACCATATATCATCTGAGGAAGAGCTTGTAGACAG
 TACTGTAAAGGCTGACTTGGGTTGATCTATCTCGGCTGGCATATGATCAAATTTCGGG
 GACATCTCCATCGAGGGGACCTGAACCCAGAACACCGATCTGCAACAACTTTATC
 CAACHTGGTTCAGACTTACAATCAGGTTTCAAGAGAACCCGATCACGCGATCCGGAGTTG
 CGCCAAAGCAATTCTGAGCGTAGGCTGCAAAATCCGGGGCTGCAAAACCTATCGCAC
 AGCTCCGTGGGAGAAGAAGAACCGGCTGTTGGTAATTCTATCGCCCTGACTGGGTG
 ACCCCAACTTTAAACTCAACTCGACCTGGCAAGATGCCAACGTTCACTGACGAAAGA
 CACCTCAGATGTAGTCTGCAACATCTGCTGGCCAGATCGGCGAACAGTACGCGAACCTT
 TTTGGGCAAGAACCTGTCAAGCGCCATTCTGCTGAGTGTATTTCTGGAGTGAACACG
 GAGATCACCAGGCTCCGCTAGCGCTAGTATGATCAAGCGTATGATGAGCACCAACGAG
 CTGAGTCTTGTGAAGGGCTTGTAGAGACCAACTGCCCTGAGAAGTACAAGGAAATTCT
 TCGATCAGTCTAAAGGGTACCGGGATACTATGACGGGGAGAACGGAGGAAATT
 TACAAAATTATTAAGGCCATCTGGGAAAAAAATGGAGCGGCAACGGAGGAGCTCTGGTAAAGCT
 TAACAGAGAAGACTGTGCGAACAGCGCATTGCAATGGAGCATCCCCAACAGAG
 TTCACTTGGGCAACTGCACCTTATCCTCAAGCGGCAAGGAGGTTCTACCCCTTTTGGAAA
 GATAACAGGGAAAAGATTGGAGAAAATCTCACATTCCGATACCCCTACTATGTTAGGCCCCCT
 CGCCGGGGAAATTCCAGATTCTGGCTGGATACCTGCGAACATCGAGAGGACCATCACTCC
 GGAACATTGAGGAGTCGGATAAGGGGCTCTGCCAGTCCTCATGAAAGGATGACT
 AACTTGTATAAAACTGCTAACGAAAAGGTCTCTCTAAACACTCTCTGTCAGACTA
 CTTCACAGTTATAACGAGTCACCAAGGTCAAATACGTCACAGAACGGGATGAGAACAG
 CATTCTGTCTGGAGGAGCAGAAGAGGATCTGTCAGGCTTCTCTTCAAGACGAAAGGGAAA
 GTTACCGTGAACAGCTCAAAGAGACTTCTAAAGGATTGTAGATGTTGACTCTGTTGA
 AATCAGCGGAGTGGAGGATCGCTAACGATCCCTGGAAACGTATCACGATCTCTGAAA
 TCATTAAGACAAGGACTTCTGGACAATGAGGAGAACGAGGACATTCTGAGGACATTGTC
 CTCCACCTTACGGTTGGAGATAGGGAGATGTTGAAGAACGGCTTGAACAAACTACGCTCA
 TCTCTGTCTGGAGGAGCAGAAGGTCAAGACGTCAGAGGGCGCAGTATACAGGATGGGGGGC
 TGTCAGGAAACTGTCATGGatccGAGAACAGCAGAGTGGAAAGACATCTGGATT
 CTTAAGTCCGATGGATTGCCAACCGGAACCTCATGCACTGATGTCATGACTCTC
 CTTTAAGGAGGACATCCAGAAAGCACAAGTTCTGGCAGGGGACAGTCTTCACGAGCACA
 TCGCTAATCTTCAGGTGGCCAGTATCCTAAAGGAAACTCGCAGACCTTAAAGGTCTG
 GATGAACTCTGGTCAAGAATGGGAAGGCATAAGCCGAGAATATCGTATCGAGATGGCCC
 AGAGAACCAAAACTTACCCAGAAGGGCAGAACAGAACACTAGGGAAAGGATGAGGAGGATGAG

- continued

SEQUENCES

AGGGTATAAAAGAACTGGGTCCAAATCCTTAAGGAACACCCAGTTGAAAACACCCAGCTT
 CAGAATGAGAACGCTCACCTGTACTACCTGCAGAACGGCAGGGACATGTACGTGGATCAGGA
 ACTGGACATCAATCGGCTCTCGACTACGACGTGGATGCCATCGTCCCCAGTCTTCTCA
 AAAGATGATTCTATTGATAATAAAAGTGTGACAAGATCGATAAAAATAGAGGAAGAGTGAT
 AACGTCCCTCAGAAGAAGTTGCAAGAAAATGAAAATTATGGCGGCAGCTGCTGAACGC
 CAAACTGATCACACAAACGGAAGTTCGATAATCTGACTAAGGCTGAACGAGGTTGGCTGTCTG
 AGTTGGATAAACGCCGCTTCATCAAAGGCACTTGTGAGACACGCCAGATCACCAAgcac
 GTGGCCAAATCTCGATTCAAGCATGAACACCAAGTACGATGAAAATGACAACACTGATTG
 AGAGGTGAAAAGTTTACTCTGAAGTCTAGGTGGCTCAGATTTGAGAAAGGACTTTAGT
 TTATAAGGTGAGAGATCACAAATTACCAACATGGCATGATGCTTACCTGAAATGCACTG
 GTAGGCAGTCACTTATCAAATAATCCAAAGCTGAACTGAAATTGTTACGGAGACTA
 TAAAGTGTACGATGTTAGGAAATGATCGAAAGTCTGAGCAGGAAATAGGCAAGGCCACCG
 CTAAGTACTTCTTACAGCAATTATGAAATTTCAGACCGAGATTACACTGGCCAAT
 GGAGAGATTGCGAACCGGACCATATCGAAACAAACCGGAGAACAGGAGAAATCGTGTGGGA
 CAAGGGTAGGGATTCTCGCAGACGTCCGAAGGTCTGTCCATGCCAGGTGAACATCGTTA
 AAAAGACCGAAGTACAGACGGAGGCTCTCCAAGGAAAGTATCCTCCCAGAACAGC
 GACAAGCTGATGCCACCCAAAAAGATTGGACCCCAAGAAATACGGGATTGATTCTCC
 TACAGTCGCTTACAGTGTACTGGTTGTGGCAAAAGTGGAGAAAGGAGCTAAAAAAACTCA
 AAAGCGTCAAGGAACGTGGGCATCACATCGTGGCAGTCAGCTTCGAAAAAAACCC
 ATCGACTTCTCTGAGGCTGAAACGAGGATATAAGAGGTCAAAAAGACCTCATATTAGCTCC
 CAAGTACTCTCTTGTGAGCTGAAACGCGGGAAACGAATGCTCGTAGTGCAGGAGC
 TCCAGAAAGGTAACGAGCTGGCAGTCCCGTAAATACGTTAATTCTGTATCTGCCAGC
 CACTATGAAAGGCTAAAGGGTCTCCCGAAGATAATGAGCAGAAGCAGCTGTCGAAACA
 ACACAAACACTTGTGAGATCATCGAGCAAATAAGCGAATTCTCAAAGAGTGTATCC
 TCGCCCGAGCTAACCTGATAAGGTCTTCTGCTTACAATAAGCACAGGATAAGCCCATC
 AGGGAGCAGGAGAAAACATTATCCACTTGTGTTACTCTGACCAACTGGCGCGCTGCAGC
 CTTCAAGTACTCGACACCACCATAGACAGAAAGCGGTACACCTCTACAAAGGAGGTCTGG
 ACGCCACACTGATTGATCATCAGTCATTACGGGGCTATGAAACAAGAATCGACCTCTCAG
 CTCTGGTAGGAGACAGCAGGGCTGACCCCAAGAAGAAGAGGAAGGTGGtagCgacggcattgg
 tagtggagacaaggcagcaggatccaacggtccgactgacgcgcggaaagaggagcttc
 tgagcaaaaactatcacctcgaaaaacgagggttgcgcgactgagaagaaaggaaaggcgggtccgg
 ggaagtggctccggatctggagggttctggcagcggaggatggcggcgttgcggaaagagtgct
 tagtaagaactatcatcgaaaaataggatggcgcgcgttaaaagaaagggtcgggaagtggc
 gcagcggaaagttggagggtggaggagcgggttctggcggatccggcgttgcggcgttgc
 tctaagaactaccacttagaaaaacgaaatcgcaacgcgttttttttttttttttttt
 ctccggatctggagaagcgggggtcgggatcaggtggatctggatcaggagaggatgg
 caaaaaaaaactaccaccccttggagaatgggtggccaggtaaaagaaggggagcggctcgggg
 agtggatcgggtcggcgggtcaggaagcgggttgcggatctggggaggagctgtctc
 gaagaattaccatctggagaacgaatggcggactaaagaag

SEQ ID NO: 89

Protein sequence for dCas9-24X-GCN4 (GCN4 is underlined)
 MDKKYSIGLAIGTNCSVWAVITDEYKVPSSKKFKVLGNTRHSIKKNLIGALLFDGETAEAT
 RLKRTRARRYTRRKNRICYLQEIEFSNEAKVDDSFPHRLEESFLVEEDKKHERHPIFGNIVD
 EVAYHEKYPTIYHLRKLVDTDKADLRLIYLALAHMICKERGHFLIEGDLNPNDSDVDKLF
 QLVQTYNQLFEENPINASGVDAKAILSLRSRRLLENLIAQLPGEKKNGLFGNLIALSGL
 TPNFKSNFDLADAKLQLSKDTYDDDNLLAQIQDQYADLFLAANLSDAIIILSDILRVNT
 EITKAPLASMSI KRYDEHHQDLTLLKALVRQQLPKEKYKEIFFDQSNGYAGYIDGGASQEEF
 YKPIKPILEKMDGTEELLVKLNREDLLRKQRTEDNGSPHQSNGYAGYIDGGASQEEF
 DNREKIKIILTIPYVGPLARGNSRFAMWTRKSEETITPWNFEEVVDKGASAQSFIERTMT
 NFDKNIPLNEKVKLPKHSLLYEVFTVYELTKVYVTEGMRKPAFLSGEQQKAIIVDLLFKTNRK
 VTVQKLKEDYFKKI ECFDSVEISGVEDRENASLGYHDLLKIIKDKDLENEENEDILEDIV
 LTLTFEDREMIERLKTYAHLFDDKVMQQLKRRRYTGWGRSLRKLINGIRDQSGKTILDF
 LKSDGFANRNFMQLIHDDSILPKEDIQKAQVSQGDSSLHEHIANLAGSPAIIKGILQTVKV
 DELVKVMGRHKPEENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
 QNEKLYLYLQNGRDMYVQDQELDINRLSDYDVDAIPVQSFLLKDDSIDNKVLTSDKNRGKSD
 NVPSEEVVKKMKNYWRQLLNAKLI TQRKEDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
 VAQILD SRMNTKYDENDKLIREVKVITLKS KLVSDERKDFQFYKVR EINNNYHHAHDAYLNAV
 VGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQBIGKATAKYFFYSNIMNNFFTEITLAN
 GEIRKRPLIETGETGEIVWDKGRDFATVVKVL SMPQVNIVKKTEVQTGGFSKESI LPKRNS
 DKLIARKKDWDPKKYGGFDSPVAYS LVAVKVEKGSKKLKSVKELLGITIMERSSFEKNP
 IDFLEAKGYKEVVKD LIKLPKYSLPELENGRKRM LASAGELQKGNE LAPS KVNFLYLAS
 HYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRV ILADANLDKVLSAYNKH RDKPI
 REQAENIIHLFTLNLGAPAAFKYEDTTIDRKRYTSTKEVLDATLHQ SITGLYETRIDLSQ

-continued

SEQUENCES

SEQ ID NO: 90

DNA sequence for dCas9-2A-X-GCN4

-continued

SEQUENCES

SEQ ID NO: 91
Linker, peptide
GSGSG

SEQ ID NO: 92
Linker, peptide
GSGSGGSGSGSGGGSGSGGGSGSG

SEQ ID NO: 103
 ASH1L protein sequence
 MDTQAGSVDEENGRQLGEVELCQGICLKWTFTADTFGIDTSSCLPFMTNSYFHCNVCHSG
 NYTFLRKQANLKEKEMCLSALANLTWQSRTQDEHPKTMFSKDID1IPF1DKYWECKMTRQRP
 GKMTWPNNIVKTMSSKERDVFVKEHDPGSKDPEEYPKGLLDQDLSNIGPAYDNQKQS
 SAVSTSGNLNGIAAGSSKGGRGAKRQKQDGTTGTTKKARSDPLFLSAQRLPPHGYPLEH
 PPNKDKGYRLIALEPDPHAPDPEKLELDCAWGAKP1PGDLYRACTYLERVLLAHDRAPOLKI
 SDDRRLTVVGEKGYSMVRASHGVRGAWYFEITVDEMPPDTAARLGWSQPLGNLQAPLGYD
 KFSYSWRSKKGTKFHQS1GHYSSGGQGQDVLFVGFIINLPDTETAKSLPDTYDKDALIKE
 KSYLYFEEKDFVDKAESKLQKTPHSE1IFYKNGVNQGVAYKD1FEGVYFPASIYKSCTV
 SINFGCPFKYVKPDLTYPRMSDMGWGAVVHETTLADLVYHVEETEDGRSPWPWEP

SEQ ID NO: 104
ASH2L DNA sequence
ATGGTAACTCAGGGGGTCGGTGGATGAAGAGAATGGCCGACAGTTGGGTGAGGTAGAGCT
GCAATGGGATTGTCACAAAATGGTCAGGCTGACACATTGGCATAGATAACCTCATCT
GTCTACCTTCTATGCAACAACTACAGTTTCAATTGAACTGCTGCCATCACAGTGGAAATTAC
TATTTCTCCGAAGCAAGCAAACCTGAAGGAAATGTGCCTTAGTGCCTTGCCAAACCTGAC
ATGGCAGTCCGAACACAGGATGAACATCGAAGAACATGTTCTCAAAGATAAGGATATT
TACCACTTATTGATAAAATCTGGGACTGTCATGACAAACCGACAGGACACTGGAAATGACT
TGGCCAATTAACATTGTTAAACAACTAGTAAAGAAGAGATGTATTCTGGTAAGGAAAC
CCCAGATCAGGAGTAAAGATCCAGAAGAAGATTACCCAAATTGGCATTTGGATCAGG
ACCTTAGTAACATTGGCTCTGTTATGACAAACCAAAACAGAGCAGTGCCTGTCTACTAGT
GGGAATTAAATGGGGAAATGCAAGCAGGAAGCAGCGGAAAGGACGAGGAGCCAAGCGCAA
ACAGCAGGATGGAGGACCAACAGGACACCAGAAGAAGGCCGGAGTGACCCCTTGTCTTG
CTCGCGCTTCCCTCATGGTCACCTGGAAACCCGTTAACAAAGATGGCTATCTGG
TATATTCTAGGTAGGGCTGATCCGCAACCCCCCTGAGCCCGAACAGTCGAAGATTGACTGCG
GGCAGGAAACCTATCTGGAGACCTCAAGAGCCTGTTGATGAAAGGGTTTGTAGG

- continued

SEQUENCES

CCCTACATGATGAGCTCCCCAGTTAAAGATCTCAGATGACGGGCTGACTGTGGTGGAGAG
AAGGGCTACTCTATGGTGAGGGCCTCTCATGGAGTACGGAAGGGTGCCTGTATTGAAAT
CACTGTGGATGAGATGCCACCAAGATAACCGCTGCCAGACTGGGTGGTCCCAGCCCCTAGGAA
ACCTTCAGGCTCTTAAAGGTTATGATAAAATTAGCTATTCTGGCGAGCAAAGGGAAACC
AAGTTCCACCAAGCTCATGGCAAAACACTACTCTCTGGCTATGGACAGGGAGACGTCCCTGGG
ATTTTATATTAACTCTCTGAAGACACAGAGACAGCCAGACTCATGGCAGACACATACAAG
ATAAGGCTTGATAAAATTCAAGAGTTATTGTATTGAGGAAAAGACTTTGTGGATAAAAG
GCAGAGAAAGGCTGAAGCAGACTCCCATAGTGAGATAATATTGTATAAAATGGTGTCAA
TCAAGGTGTCCTACAAAGATAATTGTAGGGGTTTACTTCCAGCCATCTCACTGTACA
AGAGCTCACGGTTCCATTAACTTGGACCATGCTTCAGTATCTCCAGGGATCTCACT
TACCGCCCTATGAGTGACATGGGCTGGGGCGCGTGGTAGAGCACACCCCTGGCTGACGTCTT
GTATCACGTGGAGACAGAAGTGGATGGGAGGGCGAGTCCCCATGGAACCCCTGA

SEQ ID NO: 105

BCL7B protein sequence
MSGRSVAEAEERAKDDIKKVMMAIEKVRKWEKKWVTVGDTSRIFKWPVTDSEKEKS
KSNSSAAREPNGPSDASANSLLEFQDENSNQSSVSDVYQLKVDSSTNSSPQSES
LSPAHTEDERTDSDQPPTLGQHILEEPSLPSSEVADEPPTLTKEEPVPLETQVVEEEEDS
GAPPLKRFCVDQPTVPQTASESL

SEQ ID NO: 106

BCL7B DNA sequence
atgtcgggccggctcggtccggggcgagacccgcagccgggccaaggacgacatcaagaagg
atggcggcccatcgagaaagtgcggaaatgggagaagaagtgggtgactgtgggtgacacgt
ccctgaggatataatgggttcctgtgacagacagcaaggaaaaggtaaaaaatcg
aacagtccagcggccgagaacatggcttctgtgcgttcattctct
cctcttgaattccaggacgaaacacgcaaccagatgtccgtgtgcgtctatcagctt
aggtggacagcaccaactcaagccccagccccagcagactgtgagtccctgagccccagca
cacacccctcgactccgcacccgatgtcccgccccaaacgcgtggccaggagatccctgga
ggagccctccctccctcgaaatgtgcgtgatgaaaccttctaccctaccaaggaaagaa
cgttccacttagagacacaggctgttgaggaaagagactcagttgcggccctgtgaag
cgcttctgtgtggaccaacccacagtgcgcgcacggcgtcagaaagcttgc

SEQ ID NO: 107

C20orf20 protein sequence
MGBAEVGGGAAGDKGPGEAATSPAETVVWSPEVEVCLFHAMLGHKPVGVNRFHMIC
RDKFSQNIGRQPSKVIWDHLSTMQALHESEILPEPNPERNEVLPEEEIIQEVREGKV
MIEEMKEEMKEDVDPHNGADDVESSSGSGLKASEKSSKDKEKNSSDLCKEGADKRKRS
RVTDKVLTANSNPSSPAKRRRT

SEQ ID NO: 108

C20orf20 DNA sequence
ATGGGAGAGGCCAGGTGGGGCGGGGGCGCGCAGGCACAAGGGCCGGGGAGGCC
CACCAAGGCCGGAGGAGACAGTGGTGTGGAGCCCCGAGGTGGAGGTGTGCCTCTTCCACG
CCATGCTGGGACAACAGCCCTCGGTGAAACCGACACTCCACATGATTGTATTGGGAC
AAAGTCAGCCAGAACATCGGGCCAGGTCCCATCAAGGTCACTGGGACCATCTGAGCAC
CATGTACGACATGCAGCGCTGCATGAGTCAGATCTTCCATTCCGAATTCAGAGAGAGA
ACTTCGTCTCCAGAGAGACATTCAGGAGGTCCGAGAAGGAAAGTGTGATAGAAGAG
GAGATGAAAGAGGAGATGAAGGAGACGTGGACCCCCAACATGGGCTGACGATTTTTC
ATCTCAGGGAGTTGGGAAAGCATCAGAAAATCAGCAAAGACAAAGAGAAAGAAACTCCT
CAGACTTGGGTGCAAAGAAGGCCAGACAAGCGGAAGCGCAGCCGGTACCCGACAAAGTC
CTGACCGCAAACAGCAACCCTCAGTCCAGTGCCTGCAAGCGCGCGACGTAG

SEQ ID NO: 109

DMAP1 protein sequence
MATGADVRDILELGGPEGDAASGTISKDIINPDKKSKSSETLTFKRPEGMHREVYAL
LYSDKKDAPPPLPSDTQGYRTVKAKLGSKKVRPWKWMPTNPARKDGMFHWRRRAEE
GKDYPFARFNKTVQPVYSEQEYQLYLHDDAWTKAETDHLEDSLRSREDLRFVVIHDYDH
QQFKKRSEDLKERYHICAKLANVRAPVPGTLKIPVFDAGHERRKKEQLERLYNRTPEQ
VAEEYLLQELRKIEARKKEREKRSQDLQKLITAAADTTAEQRRTERKAPKKLPQKKEAE
KPAVPETAGIKEPDFKSAGVTLRSQRMKLPSSVGQKKIALEQMLLELGVELSPPTTEEL
VHMFNELRSDLVLLYELKQACANCEYELQMLRHRHEALARAGVLGGPATPASGPGPASAE
PAVTEPGLGPDPKDTIIDVVGAPLTPNSRKRESASSSSSVKAKKPL

SEQ ID NO: 110

DMAP1 DNA sequence
atggctacggggcgccatgtacgggacatttcaagaactcgggggtccagaaaggggatgc
ctctgggaccatcagcaagaaggacatttatcaacccggacaagaaaaatccaagaaggatc
ctggagacactgttcaagaggcccggggatgcacccggaaatgtctatgcctgtctac
tctgacaagaaggatgcacccactgtcatacccgatgcacactggccaggatccgtac
gaaggcacaaggatgggtccaaaggatggcggcccttggaaatgtggatgccattccaacccgg
cccgcaaggacggagcaatgttccactggcgacgtgcacggaggaggcaaggactac
cccttgcacggatcaataaaggactgtgcagggtgcctgtgtactcggagcaggatccgt
ttatctccacatgtgtggactaaggcacaaggactgaccaccccttgcacccgc

- continued

SEQUENCES

gcttgacctgcgtttgttattccatgaccggatgaccaccaggcagttcaagaaggcg
 tctgtggaaagactgaaggacggactaccacatctgtctaagctgccaacgtgcgggc
 tggccaggcacagacctaagataccagtatttgcgtggcacaacgcacggcgaaagg
 aacagcttggactcttacaacccggaccccggacggcggaggatctgtca
 caggagctgcgaaggatggaggccggaaagaaggacggggaaaacgcacccggactgc
 gaagctgatcacacggcagacaccactgcagacgcggccgcacggaaacggcaaggccccca
 aaaagaagtcaccccaaaaaaggaggctgagaagcggctttctgagactgcaggcatc
 aagttccatgcacacttcaggctgcaggactgcgtggcggaaacggatgtctggag
 ctctgtggacacaatggggacatggggacatgttgcataatgtcagatgtgggg
 tgagccgcacactacggggactgttgcacatgtcagatgtcgaacactgttgc
 ctgctctacgagactaaggcggcctgttgcacactgcggatgttgcacatgtcggca
 ccgtcatgaggcactggccgggctgttgcataggggccctgcacaccagcatcaggcc
 caggccggccctgttgcaggccggactgtactgaaccggacttgcaccccaaggac
 acatcatgtatgtggggccacccctcaggccaaatcgagaaaggcggggacttgc
 ctccagtcatccgtgaagaaaggccaaacggcg

SEQ ID NO: 111
 DYRK1B protein sequence
 MAVPPGHPFSGFPQEHITQVLPDVRLPPLAFRDATSLPLRKLSDVLIKYKHIN
 EVYYAKKKRAQQAPPDSSNKKEKVLNHYDDDNHDYIVRSGERWLEYEIDSILKG
 SFQGVVKAYDHQTLVVAIKI1KNNKAFLNQAQIELRLLELMNQHDEMKYYVHLKRHF
 MFRNHLCLVFELLSYNYLDLLRNTHFRGVSLNLTRKLAQQLCALLFLATPELSIIHCDL
 KPEENILLCNPKRSAIK1VDFGSSCQLGQRIYQYI0SRFYRSPVLLGTPYDIAIMWSLG
 CILVEMHTGEPLESGSNEVDQMNRIVEVLGIPPAAMLQAPKARYFERLPGGGWTLRRT
 KELRKDYQGPGTTRLQEVLGVTGGPGGRAGEPGHSADYLRFQDLVLRMLEYEPARI
 SPLGALQHGFRRTADEATNTGPAGSSASTSPAPLDTCPSSSTASSISSSGGSSSSDN
 RTYRSNRYCGGPPI TDCEMNSPQVPPSQPLRPWAGGDVPHKTQAPASASSLPGTGA
 QLPQPQYRLGRPSPTS PPPPELMDVSLVGGPADCSPPHPAPAPQHPAASALRTRMTGGR
 PLLPPPDDPATLGPLGLRGVPQSTAASSL

SEQ ID NO: 112
 DYRK1B DNA sequence
 atggccgtccaccggccatggcccttctggcttccaggggcccaggaggcacacgc
 gtatgcgtatgtggcgtactgcgtggaggctgcctggccctggccctggatgcacact
 cagcccccgtcgtcaagactctgtggacccatcaagacatcacaatgaggta
 tactatgcgaagaagaagcggccggccagcaggccgcaccccgaggatcggcaacaagaa
 ggagaagaaggctgttgcaccatggatgtatgcacgcacatgcactatcgtgcgcagtg
 gcggcgtgtggggaggcgtacggaaatgtactgcgtatggcaaggctcttggccag
 gtggtaaaggcatgtatcatcagacccaggatgtggcatcaagatcatcaagaacaa
 aaaggcttccttgcaccatggcccatgtggcatgtggcatgtggactcatcaaggatc
 acacggagatgaagtaactatatactacatgcacgcacttgcggccaccatccc
 tgctgttatggactgtgttgcataacatgcacgcacttgcggccacaccacttgc
 cggcgtctcgctgaccatggccggactgtggccacgcacgcacttgcacggactgc
 tggccgcgttgcacttgcaccatgcacgcacttgcggccatgcacgcacttgc
 aaccccaaggcgcgcacatcaaggatgtggccatggccatgcacgcacttgc
 acgacactggccatgcacatgcacatgcacgcacttgcggccatgcacgcacttgc
 cccctcttcagtggctcaatggatgcacgcacatgcacgcacttgcggggatgc
 cccaccggccgcacatgtggccaggcggcccaaggatgcacgcacttgc
 ggggtggctggccatcagacggcggacttgcacgcacttgcggccatgcacgcacttgc
 cggccgttgcggggatgcacgcacttgcggccatgcacgcacttgc
 gccggccacagcccccgcactacccgcgttccaggacgcacttgc
 atgagccgcgcgcgcacatgcacgcacttgcggccatgcacgcacttgc
 gccgcacgcacatgcacgcacttgcggccatgcacgcacttgc
 cgacacacttgcacgcacttgcacgcacttgcacgcacttgc
 ctcgcacttgcacgcacttgcacgcacttgcacgcacttgc
 atcacagactgttgcacgcacttgcacgcacttgc
 agggggatgtgtggccatcaaggacatcaaggccgcacatgc
 cccggccgcacatgcacgcacttgc
 ccaccccccggatgcacgcacttgc
 cccaccccccggatgcacgcacttgc
 cccaccccccggatgcacgcacttgc
 gtccaccccccgccttgcacgcacttgc
 gtaccccccggatgcacgcacttgc

SEQ ID NO: 113
 EAF1 protein sequence
 MNGTANPLLDREEHCLRLGESFEKRPRASFHTIRYDFKPASIDTSCEGELQVGKGDEVTI
 TLPHIPGSTPMITFKGNKRPYQKDCVLIINHDTGEFVLEKLSSSIQVKKTRAEGSSKIQ
 ARMEQQPTRLPPQTSQPPPPPPMPFRAPTKEPVGPKTSPLKDNPSEPEPQLDDIKRELRAE
 VDIEQMSSSSGSSSSDSESSSGSDDSSSSGEDNGPASPPQPSHQQPYNNSRPAVANGT
 SRPQGSNQLMNALRNDLQLESSESQSDSDD

- continued

SEQUENCES

SEQ ID NO: 114

EAF1 DNA sequence

```
ATGAATGGACCGCAAACCCGCTGCTGGACCCGAGGAACATTGCTGAGGCTCGGGGAGAG
CTTCGAGAACGGCCCGCGGGCCTCCACACTTCTGTTGATTTAAACCAGCATCTA
TAGACACTTCCGTGAAGGAGACTCAAGTGGCAAAGGAGATGAAGTCACAATTACACTT
CCACATATCCCTGGATCACACCACCATGACTGTGTTCAAGGGAAACAACGGCCTTACCA
GAAAGACTGTGCTTATTATTAATCATGACACTGGTGAATTGTGCTGGAAAAACTCAGTA
GCAGCATTCTCAGGTAAGAAAAGAGCTGAGGCCAGCAGTAAAATCCAGGCCAATGGAA
CAGCAGCCCACACTGCTCCACAGACCTCACGCCACCCACCTCACACCTATGCCATT
CAGAGCTCAACGAAGCCTCCAGTTGGACCCAAAATCTCCCTTGAAAGATAACCCCTCAC
CTGAACCTCAGTGGATGACATCAAAAGAGAGCTGAGGGCTGAAGTTGACATTATGAACAA
ATGAGCAGCAGCTGGAGCAGCTTCAGACTCTGAGAGCTTGGGAAGTGTGACAGA
TAGCTCAGCAGTGGAGGGAGGACATGCCAGCCTCTCCCGCAGCTTCACACAGC
AGCCCTAACACAGTAGGCCTGCCGTTGCAATGAAACCAAGCCGCCACAAGGAAGCAACCAG
CTCATGAACGCCCTCAGAAATGACTTGCAGTTGAGTGTGAGTCAGTGTGACTA
G
```

SEQ ID NO: 115

FOXR2 protein sequence

```
MDLKLKDCFWYSLHGQVPGLLDWDRNELFLPCTTDQCSLAEQILAKYRVGVMKPPEMP
QRKRPSPDGDGPCEPNLWNVDPNLLCPPLGSQEAPKPSKEDLTNIISPFPQQPKDEGS
NCSEDKVVESELPSSEQSPLQKQGIHSPSDFELTEEEAEEPDDNSLQSPEMKCYQSQKL
WQINNQEKSWQRPLNCSHLALALRNPNPHCGLSVQEINYNTRQHPFFWTAPDGWKSTI
HYNLCFLDSFEKVPDSLKDEDNARPRSLWKLTKEGHRRFWEETRVLAFQQRERIQECMS
QPELLTSLFDL
```

SEQ ID NO: 116

FOXR2 DNA sequence

```
ATGGACTAAACATAAGACTGTGAATTGGTATAGTCTCATGCCAGGTCCCAGGGCT
GCTGGACTGGGACATGAGGAATGAGTTATTCGCTTGTACACAGACAGTGCTCTTAG
CTGAGCAATCTTGCACATACAGACTCGGAGTAATGAAGCCCCAGAAATGCCCTCAGAAG
AGGAGACCGCAGTGGAGATGGCTCTCCCTGTGAACCCAATCTGTGGATGGGGTGBA
CCCCAATATCTGTGCCCTTGGCAGCCAGGCCAAAGCCAGTGAAAAGAGGATC
TGACAAACATTCTCTTCCCTCAGCCCCACAAAAAGCAGAAGGGTCTAAGTCTCAGAG
GACAAAGTGGTAGAGTCCTGCCATCTCCCTCAGTGAGCAGTCTCTTACAGAAGCAGGG
TATCCCATCCCTGGACTTGTGAGCTCACAGAAGAGGAGGCTGAGGAACCAAGCACA
CCTCCAGTCCCCGTAATGAAATGTTACCGAGGCCAGAAACTATGCCAAATCAACACAA
GAGAAGTCTGCCAAAGCCCCCTCTCAATTGTAGCCACCTTATTGCCCTAGCATTAAGAAA
CAACCCCCACTGTGGCTCAGTGTGAGGAGATCTACAAATTCAACCCGACAGCAATTCCCT
TTTCTGGACACTGGATGCTGAAAGACCAATTCTACACACCTCTGCTTCTGGAC
AGCTTGAGAAGGGCAGACGCCCTTAAGGATGAAGATAATGCAAGACCTCGCTTTGCT
TTGGAAGCTCACTAAGGAGGGCACCCGCTTGTGGAGGGACTCGTGTCTAGCCTTG
CTCAAAGGGAGAGAATCCAAGAGTCAGTGTGACCTCAGGAGTTGACCTCTCTTTGAT
CTTGAA
```

SEQ ID NO: 117

GSK3A protein sequence

```
MSGGGPSSGGPGGSGRARTSFAEPGGGGGGGGGGPGGSASGPGGTGGGKASVGAMGGV
GASSSSGGPGGGGGGGGGPGAGTSFPPGVKLGRDSKGKVTVVATLGQGPERSQEVAYT
DIKIVIGNGSFVVYQARLAETRELVAIKVLQDKRFKNRELQIMRKLDHCNIVRLRYFFY
SSGEKKDELYLNLVLEYVPETVYRVARHTKAALKTIPILYVKVYMQLFRSLAYIHSQGV
CHRDIKQNLVLPDPAVLKCDFGSAKQLVRGEPMVSYICSRYYRAPELIFGATDYTSS
IDVWSAGCVLAELLQGPIFPGDGVQDVLVEI1KVLGTPTRQIREMNPNTEFKFQPIK
AHFWTKVFKSRTPEIA1CSSLLEYTPSSRLSPLEACAHASFDELRCGLTQLPNNRPLP
PLENFSAGELSIQPSLNAILIPHLRSPAGTTLTPTSSQALETPTSSDWQSTDATPTLT
NSSL
```

SEQ ID NO: 118

GSK3A DNA sequence

```
atgagcggccggcccttcgggaggcgccctggggctcgcccagggcgccgactagctc
gttcgcggaggccggccggccggaggccggaggccggccggccggccatcgccctccg
ggccaggccggccggccggccatctgtcccccatgggtggggccgtcgccggcc
tcagactccgggggtggaccggccgcagccggccggaggccagccggccggccagg
cactagttcccgccgcggccgggtgaagactggccgtgacagccggaaagggtgaccacagt
tagccactctaggccaaggcccaaggccatggcttacacggacatcaaagt
atggcaatggctatgggtctgttacccggccggccgtgcggccaggactagt
cgccatcaagaagggttctccaggacaaggatctggccatcgccatcgccgt
agctggaccactcaatattgtggctgagatacttttactccagtgccgagaagaaa
gacgagcttaatctgggtggatatgtggccggccggccgtgatccgggtggccgg
ccacttcaccaaggccaaaggccatggccatccctatccctatgtcaagggtgatcatgtaccagc
tctccgcagctggccatccactccaggccggccgtgtcaccggccatcaaggccccag
aacctgtggccggccatctggccatccactgtgtccatcaagctgcaatggccggccggccat
gttgcgttccatctggccatccactccaggccggccgtgtcaccggccatcaaggccccag
```

- continued

SEQUENCES

```
tcatctttggagccactgattacacctatccatcgatgtttggcagctggctgtactg
gcagagctccttggccagccccatccctcctggggacagtgggtgaccagctgggaga
gatcatcaagggtctggaaacccaacccggaaacaatccgagagatgaaccccaactaca
cgaggtcaagtccctcagatcaaagctcacccctggacaagggttcaaatctcgaa
ccggccagaggccatcgctgtctgcgcacacttgcgatgtctggggatcacccatctcaagggtctc
cccaactagggctgtgcgcacacttgcgatgtctggggatcacccatctcaagggtctc
ctaacaaccgcacacttcccctcttcaacttcagtgctgtgtaacttcacatccaacc
tctctcaacgcacatctcacttgcgatgtggccggactaccacccctcac
cccgccatcacaagcttactgagactccgaccagctcagactggcagtcgaccatgc
cacttaccctcactaactcccttgc
```

SEQ ID NO: 119
JAZF1 protein sequence
MTGIAAAASFFSNTCRFGCGLHFPTLADLIEHIEDNHIDTPRVLEKQELQQPTYVALSY
INRFMTDAARREQESLKIKIQPKLSLTSSSVSRGNVSTPPRHSSGSLTPPVTPPSS
SFRSSTPTGSEYEEEVDYEEESDSDESWTTESAISSEAILSSMCNGEEKPFACPVPGC
KKRYKNVNGIKYHAKNGLHRTQIRVRPKRCRGKSYKTAQGLRHHTINFHPPVSAEIIRK
MQQL

SEQ ID NO: 120
JAZF1 DNA sequence
Atgacaggcatgccggccgccttcttccataacctgcgcattggggctgcggact
ccactccccccatcgccgacactatcgagcacatcgaggacaaccatcgatacagatc
caccgggttttagaaaaacaagaattacagcaggccaaacctatgttgccctgagttacataat
agattcatgcgatgtgcggcccgagagcaggacttcaaaaagaaagattcagccgaa
gtctcgctgcgtccagctcagtgctcgaggaaatgtgtccactccccacgcacca
gcagtggaaagcttactccccccgtgaccccaacccatccccctcttcatccgc
agcaactccgcacaggcagcgcgtatggcggaggagggtggactatggaggatcggacagcga
tgagtccgtggccacacagatgcgcacatccgcgatccgcgtccatgtgc
atggaggggaaagacttgcgtccgcgttctggatgtaaaaagagatacaagaat
gtgaatggcataaagtatcagctaaatggcacaagacacagattcgttccgcaaa
attcaagtgtcgtgtggaaagatgttacaagacagactcaggccctgcggcaccacacaatca
atttccatccccgggtcgccgtgagattacaggaaagatgcagcaattt

SEQ ID NO: 121
KAT7 protein sequence
MPRRKRNAGSSSDGTEDSDESTDEHTDSSESDEGTSSRSARVTRSSARLSQSQQDSSPVR
NLQSFGEVTEEPAYSTRRVRSPQQPTPVTPKYPLRQTRSSGSETEQVVDDESRETKNAD
HDESPPTPTGNAPSSESDIDISSPNVSHDESIAKDMSLKDSGDSLHRPKRRRFHESYN
FNMKCPPTGCNSLGHLTGKHERHESISGCPLYHNLSADECKVRAQS RDQI EERMLSHRQ
DDINNRHATRHQAPTEROLRYEKVAELRKRNGLSKEQKEKYMHEHQTYGNTREPLLEN
LTSEYDLDLERRAQARASEDLEKLRLQGQITEGSNMIKTIAFGRYELDTWHPYPEEYA
RLGRILYMEFCLKYMKSQTILRRHMAKCVWKPPGDEIYRKGSISVFEVGKKNKIYQCQ
LCLLAKLFLDHKTLYDVEPFLFYVMTEADNTGCHLIGYFSKEKNSELNYNVSCILTMPQ
YMRQGYGKMLIDFSYLLSKVEKVGSPERPLSDLGLISYRSWKEVLLRYLHNFGQKEIS
IKEISQETAVNPVDIVSTLQALQMLKYWKGHVLVRQDLIDEWIAKEAKRSNSNKTMDP
SCLKWTPPKGT

SEQ ID NO: 122
KAT7 DNA sequence
ATGCCGCGAAGGAAGAGGAATGCAGGCAGTAGTTCAAGATGGAACCGAAGATCCGATTTTC
TACAGATCTCGAGCACAGACAGTCAGAAAGTGTGGCACATCCGACGATCTGCTCGAG
TCACCCCTCCTCAGCGAGACTAGCCAGAGTCTCAAGATTCAGCTCTTGAAATCTG
CAGTCTTTGGCAGACTGGAGAGCTGCTTACTTACCCAGAGGTGACCCGTAGTCAGCAGCA
GCCCTACCCCAGTGCACACGGAAAAAAATACCCCTTCAGCAGACTCGTCATCGTTCAAGAAA
CTGAGCAAGTGGTTGATTTTCAGATAGAGAAAATACAGTCATGATGAGTC
CCGCCTCGAACCTCAACTGGAAATGCCCTCTCTGAGTCAGCATAGACATCTCCAGGCC
CAATGTATCTCGAGATGAGAGCATGGCAAGGACATGTCCTGAGGACTCAGGGCAGIGATC
TCTCTCATCGCCCCAACGGCGCTCGCTTCCATGAAAGCTACAATCTCAATATGAAGTGTCT
ACACCAGGCTGTAACCTCTAGGACACCTTACAGGAAAATGAGAGACATTCTCATCTC
AGGATGCCACTGTATCATAACCTCTCAGTCAGCAATGCAAGGTGAGAGCACAGAGCCGG
ATAAGCAGATAGAAGAAAGAGTGTCTCACAGGAAGATGACAACAAAGGCATGCAACC
AGGCACCAAGGCCAACACAGGAGAGCAGCTCGATATAAGGAAAAGTGGCTGAACTCAGGAA
AAAAAGAAATTCTGGACTGAGCAAAGAAACAGAAAGAGAAAATATATGAAACACAGACACT
ATGGGAACACACGGGAACCTCTTCTAGAAAACCTGACAAGCGAGTATGACTTGGATCTTTC
CGAAGAGCACAAGCCGGCTTCAGAGGATTTGGAGAAGTTAGGCTGCAAGGCCAAATCAC
AGAGGGAAAGCACTGATTAACAAACATTGCTTTGGCGCTATGAGCTTGTATCTGGTAC
ATTCTCCATATCTGAAGAATATGCAAGGCTGGAGCTCTCATATATGTGTGAATTCTGTTA
AAATATATGAAGAGCAAACGATACTCCGCCGGCACATGGCCAAATGTGTGTTGAAGTGGATG
ACCTGGTGTAGAGATATATCGCAAAGGTTCAATCTCTGTTGAAGTGGATGCAAGGAAA
ACAAAGATCTACTGCCAAACCTGTGCTGTGGCAAACCTTTCTGGACACAGACACTGGCTGTCA
TATTATGATGTGGAGGCCCTCTGTCTATGTATGACAGAGGGCGGACAACACTGGCTGTCA
CCTGATTGGATATTTCTAAGGAAAAGAATTCTTCAACTACAAACGTCCTGTATCC
TTACTATGCCCTCAGTACATGAGACAGGGCTATGGCAAGATGCTTATTGCTAGTT

- continued

SEQUENCES

CTTCCAAAGTCGAAGAAAAAGTTGGCTCCCCAGAACGTCCACTCTCAGATCTGGGCTTAT
AAGCTATCGCAGTTACTGGAAAGAAGTACTCTCCGCTACCTGCATAATTCAAGGCAAG
AGATTCTATCAAAGAAATCAGTCAGGAGACGGCTGTGAATCTGTGGACATGTGAGCACT
CTGCAAGCCCTCAGATGCTCAAATACTGGAAAGGGAAAACACTAGTTAAAGAGACAGGA
CTCTGATTGATGAGTGGATAGCCAAGAGGCCAAAGGTCCAACCTCCAATAAACCATGGATC
CCAGCTCTTAAATGGACCCCTCCAAGGGCACTTAA

SEQ ID NO: 123
KEAP1 protein sequence
MQDPDRPSGAGACCRELPLQSQCPPEGAGDAVMYASTECKAEVTPSQHGNRTFSYTLLEDHT
KQAFGIMNELRLSQQQLCDVTLQVKYQDAPAAQFMAMHKVVLASSSPVFKAMFTNGLREQGM
EVVSIEJHVKMVERLEFAYTASI SMGEKCVLHVVMNGAVMVQIDSVVRCASDELVQQLD
PSNAIGIANFAEIQGVCEHLQRAREYIYMHFGEVAKQEEFFNLSHCQLVTLISRDDLNV
CESEVFHACINWVKYDCEQRFFYVQALLRAVRCHSLLTPNFLQMLQKCEILQSDSRCKDY
LVKIFEELTLHKPTQVMPCRAPKVGRLIYTAGGYFRQSLSYLEAYNPDSGTVLRLADLQV
PRSGLAGCVVGLLIVAVGGRNNSPDGNTDSSALDCYNPMTNQWSPCAPMSVPRNRIGVGV
IDGHIAVGGSHGCIHNSVERYEPEERDEWLHVAPMLTRRIGVGVAVLNRLLYAVGGFDG
TNRLLNSAECYYPERNEWRMITAMNTIRSGAVCVLHNCIYAGGGYDQDQLNSVERYDVE
TETWTTFVAPMKHRRSALGITVHQRIYVLGGYDHTFLDSVECYDPDTDTWSEVTRMTSG
RSVGVGAVTMEPCRKQIDQQNCTCL

SEQ ID NO: 124
KEAP1 DNA sequence
atgcaggccatcccaggcctagccccggctggggctgtgcgcattctgcctgcggcgtcgc
acaggcgcctggggggcagggggacgggggtgtgtacgcctcactgagtgcgcaggcgagg
tgacgcctcccatgcgcacccatgcgcctcactgatcacccctggaggatcataccaggc
gcctttggccatgcgcacccatgcgcacccatgcgcctcactgatcacccctggaggat
caagtaccaggatgcaccggcccccaggatgcgcctccatggccacaagggtgtgtgc
gcgcctgtcttgcgcacccatgcgcacccatgcgcctcactgatcacccctggaggat
attgggggtatccaccccaaggatgcgcacccatgcgcacccatgcgcctcactgatcac
cttcatggcgagaaggatgtgcctccacgtcatgcgcacccatgcgcacccatgcgc
gcgttgtccgtgcgtgcgtgcgtgcgtgcgtgcgcacccatgcgcacccatgcgc
atcgccaaactcgctgcgcacccatgcgcacccatgcgcacccatgcgcacccatgcgc
atccatgcattttggggagggtggcccaaggatgcgcacccatgcgcacccatgcgc
tggtgaccctcatgcgcacccatgcgcacccatgcgcacccatgcgcacccatgcgc
tgcacatgcactgggtcaagtgcgcacccatgcgcacccatgcgcacccatgcgc
ggccgcgtgcgcgcacccatgcgcacccatgcgcacccatgcgcacccatgcgc
tccatgcgcacccatgcgcacccatgcgcacccatgcgcacccatgcgcacccatgc
cacaaggccacgcaggatgcgcacccatgcgcacccatgcgcacccatgcgc
ggggcggtacttcgcacgcacccatgcgcacccatgcgcacccatgcgcacccatgc
ggcgtccgtgcgcacccatgcgcacccatgcgcacccatgcgcacccatgcgc
ctgttgtacgcgcacccatgcgcacccatgcgcacccatgcgcacccatgcgc
ggactgttacacccatgcgcacccatgcgcacccatgcgcacccatgcgc
accgcacccatgcgcacccatgcgcacccatgcgcacccatgcgcacccatgc
atccacacacaaactcgatgcgcacccatgcgcacccatgcgcacccatgcgc
gggttacacccatgcgcacccatgcgcacccatgcgcacccatgcgcacccatgc
atccatgcacacaaactcgatgcgcacccatgcgcacccatgcgcacccatgc
aatgtgcacacaaactcgatgcgcacccatgcgcacccatgcgcacccatgc
ggggcttgcgcacccatgcgcacccatgcgcacccatgcgcacccatgcgc
tggcgaatgtatgcgcacccatgcgcacccatgcgcacccatgcgcacccatgc
ctgtatcgatgcgcacccatgcgcacccatgcgcacccatgcgcacccatgc
atcgatgcgcacccatgcgcacccatgcgcacccatgcgcacccatgcgc
atcaatgcgcacccatgcgcacccatgcgcacccatgcgcacccatgcgc
cagtgtggaggatgcgcacccatgcgcacccatgcgcacccatgcgcacccatgc
gc
aactgtacatgcgcacccatgcgcacccatgcgcacccatgcgcacccatgc
gc

SEQ ID NO: 125
MEAF6 protein sequence
MAMHNKAAPPQIPDTRRELALVRKRQKELAETLANLERQIYAFEGSYLEDTQMYGNIIRG
WDRLTNQKNSNSKNDRNRKPKEAERLFSKSVTSAAVSALAGVQDQLEKREPGSGT
ESDTSPDFHNQEENEPSQEDPEDLDGSVQGVKPQKAASSTSSGHHSSHKKRNKNRHRID
LKLNNKKPRADY

SEQ ID NO: 126
MEAF6 DNA sequence
ATGGCGATGCACAACAAGGGCGCGCCGCCAGATCCGGACACCCGGGGAGCTGGCGGA
GCTCGTAAGCGGAAGCAGGAGCTGGCGAACATGGCAATTGGAGCGACAGATCTATG
CTTTGAGGGAAAGCTACCTGGAGAGACACTCAGATGTATGGCAATTATTATCGTGGCTGGGAT
CGGTATCTGACCAACAAAAAAACTCCAAATAGCAAAATGATCGAAGGAACCGGAAGTTAA
GGAGAGCTGAGCGCCTTCAGTAATCTCGGTTACCTCAGCAGCTGCAGTAAGTCAGT
CAGGAGCTCAGGACCAGCTCATGGAAAGAGGGAGGCCAGGAAGTGGGACGGAAAGTGACACT
TCTCCAGACTTCCACATCAGGAAATGAGCCAGCCAGGAGGACCCCTGAGGATCTGGATGG
ATCTGTGCAAGGAGTGAAACCTCAGAAGGCTGCTTCTACTTCCTCAGGGAGTCACCA
GCAGCCATAAAAAGCGAAAGATAAAAACCGGCACAGGATTGATCTGAAGTTAACAAAAAA
CCACGAGCTGACTATTAG

- continued

SEQUENCES

SEQ ID NO: 127
 MLLT6 protein sequence
 MGAVNPILLSQAE~~S~~HTEPDLECSFRCRGTS~~P~~ESLSSMSP~~I~~SSLPALFDQTASAPCGGG
 QLDPAAPGTTNMEQ~~L~~LEKQGDGEAVNIVEMLKALHALQKENQRLQE~~Q~~ILSLTAKKERLQ
 ILNVQLSV~~P~~FP~~P~~ALPAALPAANGPVP~~G~~YGLPPQAGS~~S~~DSLTSKSPPGKSSLG~~D~~NLSLST
 SSEDPHSGCPSRSSSSLSFHSTPPPLLQQSPATLPLALPGAPAPLPPQPQNGLGRAPG
 AAGLGAMPMAEGLLGGLAGSGLPLNGLGG~~L~~NGAAAPNPASLSQAGGA~~T~~LQLPGCINS
 L~~T~~EQQRHL~~L~~QQQ~~Q~~QLQQLL~~A~~SPQLTPEHQTVVYQMIQOIQQKRELQRLQMAGGSQ~~L~~
 PMASLLAGS~~S~~TL~~L~~LSAGTPG~~G~~LPTASAPP~~L~~PAGALVAPS~~L~~GNNNTSLMAAAAAAAA~~V~~AAA
 GGPPVLTAQTNPFLSLSGAEGSGGP~~K~~GGTADKGASANQ~~E~~KG

SEQ ID NO: 128
 MLLT6 DNA sequence
 ATGGGTCCGTTAATCCCTCTCTCCAAGCTGAGAGCACGCCACACAGAGCCAGACCTGGA
 GGACTGCAGCTCCGGTGTGGGGGACCTCCCTCAGGAGAGTC~~T~~GTCTTCCATGTC~~CC~~CA
 TCAGCAGCCTCCCGC~~A~~CTCTTGACCAGACGCCCTCTGCACCC~~T~~GTGGGGCGGCCAGTTA
 GACCGCCGGGGCCCAGGACGACTAACATGGAGCAGCTCTGGAGAAGCAGGGCAGCGGG
 GGCGGGGGCAGCTACATGGAGAGTC~~T~~GTGAAGGCCCTG~~C~~ACGGCTGAGAAGGAGAAC~~C~~AGC
 GGC~~T~~GAAGAGCAGATCTGACGCCAAAAGGAGCCGCTGAGATTCTCAAC~~C~~GTG
 CAGCTCTCTG~~T~~GTGCCCTCCCTGCCCTG~~C~~CTG~~C~~CTGCC~~C~~CAACGCCCTG~~T~~CCC
 TGGGCC~~T~~ATGCCCTGCCCTCCCCAAGCCGGCAGCAGCAGACTCTTGAGCACCAGAACAGCC
 CTC~~C~~GGGAAAGAGCAGCCGCTGCCAGACATGGCTG~~T~~CCACTCTTGAGGACCCACAC
 TCAGGCTGCCCGAGCCGAGCAGCTCGTGTCTCTTCCACAGCACGCC~~C~~CCACGCC~~T~~GCC
 CCTCCCTCAGCAGGCC~~T~~GC~~C~~ACTCTGCCCTGGCCCTGCCCTGGGGCCCTGCC~~C~~ACTCC
 CGCC~~C~~CGAGCCGAGACGGGTTGGGCCGGCACCCGGGGCAGCGGGCTGGGGCATGCC
 ATGGCTGAGGGCTGTTGGGGGCTGGCAGGGCAGTGGGGCTG~~C~~CTGCC~~C~~CTCATGGCTCCT
 TGGGGCTTGTAGGGGCCAACCCGCCAGGCTGAGGCCAGGCTGGGGGGGCC
 CCACGCTGCCAGGCCAGGCTGCTCAACAGCCTACAGAGCAGAGAACATCTCTTCAG
 CAGCAAGAGCAGCAGCTCAGCAACTCCAGCAGCTCTGCCCTGCCAGCTGACCCCGGA
 ACACCAAGACTGTTGCTTACAGAGATGATCCAGCAGATCCAGCAGAAACGGGAGCTGAGCGCC
 TGCAGATGGCTGGGGCTCC~~C~~AGCTGCC~~C~~ATGCCAGCTG~~C~~CTGGCAGGAAGCTCCACCCCG
 CTGCTG~~T~~CTGCCGGTACCCCTGCCCTG~~C~~CTGCCACAGCGTCTGCTCCACCCCTG~~C~~CTGCC
 TGGAGGCTTAGGCTCTGCCAACACAAGCTCTCATGGCCAGCAGCTGCCAG
 CTG~~C~~AGCAGTAC~~C~~AGCAGCAGGCCAGCTCAGTCTCACTGCCAGACCAACCCCTTCTC
 AGCCTGTCGGGAGCAGAGGGCAGTGGCGGTGCC~~C~~AAAGGAGGGACC~~G~~TGACAAGGAGC
 CTAGCCAACCAGGAAAAGCTAA

SEQ ID NO: 129
 MORF4L2 protein sequence
 MSSRKQGSQPRGQ~~S~~AEENFKKPTRS~~N~~MORSKMRGASSGKKTAGPQQKNLEPALPGRWG
 GRSAENPPSGSVRKTRNKQKTPGNGDGGSTS~~E~~APQ~~P~~PRKKRARADPTVESEAFKNRME
 VIKV~~K~~IP~~E~~ELKPLV~~E~~D~~L~~V~~T~~RQ~~K~~Q~~L~~FOLPAK~~N~~DAI~~E~~YYANCKSQGNVDNKEYAVN
 EVVAGIKEYFNVMLGTQ~~L~~LYK~~F~~ER~~P~~Q~~Y~~AE~~I~~LLAHPDAPMSQVY~~G~~PH~~L~~RFVRIGAMLA
 YTPLDEKS~~L~~ALLG~~Y~~LHD~~F~~L~~K~~YLA~~K~~NSASLFTASDYKVASA~~E~~YHRKAL

SEQ ID NO: 130
 MORF4L2 DNA sequence
 ATGAGTTCCAGAAAGCAGGGTCTCAACCTCTG~~G~~ACAGCAATCTGAGAAAGAGAAC~~T~~T
 CAAA~~A~~AA~~C~~ACTAGAAGCAACATGCAGAGAA~~G~~TAA~~A~~ATGAGGGGCC~~T~~CTCAGGAAGA
 AGACAGCTGGT~~C~~ACAGCAGAAAATCTTGACCCAGCTCTCCAGGAAGATGGGTTGGT~~G~~CC
 TCTG~~C~~AGAGAACCC~~C~~CTTCAGGATCC~~G~~TGAGGAAGACAGAAAGAACAGCAGAACAGACTCC
 TGGAAACGGAGATGGT~~G~~CCAGTACCCAGCAGCAGC~~C~~CTCAGCC~~C~~CTCGGAAGAAAAGGGCC
 GGC~~C~~AGACCC~~C~~CTG~~T~~GAAGAAGT~~G~~AGGAGGGCTTAAGAATAGAATGGAGCTTAAAGTGAAG
 ATTCC~~T~~CTGAAGAATAAAACCATG~~T~~GGCTTGGAGGACTGGGACTTAGTTAC~~C~~AGG~~C~~AGAAGCA
 GCTGTTCAACTCC~~T~~GCCAACAAAATG~~T~~AGATG~~C~~AATTCTGGAGGAGTATG~~C~~AAATTG~~C~~A
 AGAAATCGAGGGAAATG~~T~~GTATAAAGGAATATGCC~~T~~TAATG~~A~~AGTTG~~T~~GGCAGGAATA
 AAAAGAATATT~~T~~CAATG~~T~~GTATG~~T~~GGC~~A~~CTCAGC~~G~~TCT~~C~~ACAAATTGAGAGGCC~~C~~AGTA
 TGCTGAATCTCTGCC~~T~~CAATG~~C~~TC~~A~~ATG~~T~~CCAGGTTATG~~G~~AGCACCACACC
 TACTGAGATTATTG~~T~~GAAGAATGGAGCAATG~~T~~GGC~~T~~CTATACGCC~~C~~CTTGATGAGAAAAGC
 CT~~T~~GCATTATTG~~T~~GGC~~T~~ATTG~~C~~ATG~~T~~TTCC~~A~~AAATATCTGCC~~C~~AAAGAATTCTGCATC
 TCTCTTACTGCCAGTGATTACAAAGTGGCTTCTG~~C~~TGAGTAC~~C~~CCGCCAACGCC~~T~~GTGA

SEQ ID NO: 131
 NFYC protein sequence
 M~~S~~TEGGFG~~G~~TSS~~D~~AQ~~S~~LQSFW~~P~~RV~~M~~EEIRNLTVK~~D~~ERVQ~~E~~PL~~P~~ARIKKIMKLDEDVKM
 ISAAEAPVLFAKAQ~~A~~Q~~I~~FE~~T~~EL~~T~~LR~~A~~W~~I~~HTEDNK~~R~~RTL~~Q~~RNDIA~~M~~AI~~T~~KFDQFD~~F~~L~~I~~DIV~~P~~R
 DELKPPKRQ~~Q~~EE~~V~~RS~~V~~TPA~~P~~V~~Q~~Y~~F~~TL~~A~~Q~~O~~PTAV~~V~~Q~~G~~Q~~Q~~Q~~G~~Q~~T~~TS~~S~~TT~~T~~IOPGQ~~I~~II
 A~~Q~~Q~~Q~~Q~~T~~TPV~~T~~M~~Q~~V~~G~~EG~~Q~~Q~~V~~Q~~I~~V~~Q~~A~~Q~~P~~Q~~Q~~Q~~Q~~A~~Q~~S~~GT~~G~~Q~~T~~M~~Q~~MQ~~Q~~I~~I~~TNTGE~~I~~QQ~~I~~P
 V~~Q~~LNAGQ~~Q~~QY~~I~~RLA~~Q~~P~~V~~SG~~T~~Q~~V~~V~~Q~~Q~~I~~Q~~T~~LA~~N~~QA~~Q~~Q~~I~~T~~Q~~TE~~V~~Q~~Q~~Q~~Q~~FSQ~~F~~D~~Q~~Q~~L~~Y~~Q~~
 I~~Q~~Q~~V~~TM~~P~~AGQ~~D~~LA~~Q~~P~~M~~FIQ~~S~~AN~~Q~~P~~S~~D~~Q~~Q~~A~~P~~Q~~V~~T~~GD

- continued

SEQUENCES

SEQ ID NO: 132

NFYC DNA sequence

ATGTCCACAGAAGGAGGATTTCGGTACTAGCAGCTAGTATGCCAGCAAAGCTTACAGTC
 GTTCTGGCTCGGTATGGAAGAAATCCGAATTAAACAGTGAAAGACTTCCGAGTCAGG
 AACTCCCACTGGCTCGTATTAAAGAAGTTATGAAACTGGATGAAGATGTAAAGATGATCAGT
 GCAGAAGGCCCTGACTCTTCCAAGGCAGCCAGATTTTATCACAGAGTTGACTCTTCG
 AGCCTGATTCAACAGAAGATAACAAGCAGGGACTCTACAGAGAAATGATATGCCATGG
 CAATTACAAAATTTGATCAGTTGATTTCTCATCGATATTGTTCCAAGAGATGAACTGAAA
 CCTCCAAAGCTCAGGAGGAGTGCAGCTGTAACCTCTGCCAGCAGCTCAGTACTA
 TTTCACGCTGGCTCAGAACCCACCCTGTCAGTCCAGGGCAGCAGCAAGGCCAGCAGA
 CCACCACTGCCAGCAGGACATCCAGCTGGCAGATCATCTGCACAGCCTCAGCAGGGC
 CAGACACACCTGTGACAATGAGGTGGAGAGGGTCAAGGGTGCAGCAGGTGCAAGTGTCCAGGCTCA
 GCCACAGGGTCAAGCCCACAGGCCAGAGTGGCACTGGCAGACCATGAGGTGATGCAGC
 AGATCATCACTAACACAGGAGAGATCCAGCAGATCCGGTGCAGCTGAATGCCGCCAGCTG
 CAGTATATCCGCTTAGCCAGCTGTCAAGTGGCAGGGACAGATCCAGAC
 ACTTGCACCAATGCTAACAGATTACACAGACAGGGTCCAGCAAGGGACAGCAGTTCA
 GCCAGTTCACAGATGGCACAGCAGCTTACCAAGATCCAGCAAGTCAACCATGCCGCCAG
 GACCTCGCCAGCCATGTTCAAGTCAGCCAACCAGCCCTCCAGGGCAGGCCAGGCCCCCA
 GGTGACCGGCAGTGA

SEQ ID NO: 133

PHF 15 protein sequence

MEEKRRKYSISSDNSTDHSATSTSASRCSKLPSSDKSGWPRQNEKKPSEVFRTDLITA
 MKIPDSYQLSPDYYILADPWRQEWEKGVQVPAGEAIPPEPVVRILPPLEGPPAQASPSS
 TMLGEGSQPDWPFGSRYLDLDEIADYLLELINSLEKEMERPELDLTLERLVELELTLCQ
 NMARAIETQEGLIEYEDEDVVCDVCRSPPEGEDGNEMVFCDKCNVCVHQACYGILKVPNTGS
 WLCRTCALGVQPKCLCPKRGALKPTRSGTKWVHSCALWIPEVSIGCFKMEPIITKIS
 HIPASRWALSCSLCKEKTGTCAIQCSMPSCVTAFHVTCAFHDGLEMRITILADNDEVFKFSF
 CQEBSHDGPPRNEPTEPTSEPSQAQEDLEKVTLRKQLQQLEDFYELVEPAEVABRLDIA
 EALVDFIYQYWKLKRKANANQFLTPKTDEVDNLAQEQDVLVYRRLKLFTHLRQDLERVR
 NLCYMVTRERTLKHAIKLQEIQFHLMQMLIEQDILCRAGLSTSFPIDGTFNNWSLAQSVQ
 IIAENNMASEWLNNNGHREDPAPGLLSEELLQDEETLLSFMRDPSLRPGDPARKGRTR
 LPAKKKPPPPPDPGPRSRTTPDAPKKTWTQDGSKGGGGPPTRKPPRTSSHLPSSP
 AAGDCPILATPESPPPLAPETPDEAASVAADSDVQVPGPAASPKPLGLRPPRESKVTRR
 LPGARPDAAGMGPSAVAERPVSLSHFDTETDGYFSDGEMSDDVEAEDGGVQRGPREAGA
 EEVVRMGVLAS

SEQ ID NO: 134

PHF 15 DNA sequence

ATGGAAGAGAACGGCGAAAATACTCCATCAGCAGTGAACACTCTGACACCCACTGACAGTCA
 TCGCACATCTACATCCGCATCAAGATGCTCCAAACTGCCAGCAGCACCAAGTCGGCTGG
 CCCGACAGAACAAAAGAGCCCTCCGAGGTTTCCGGACAGACTTGATCACAGCCATGAAG
 ATCCCGGACTCATACCAGCTCAGCCCGATGACTACTACATCTGCCAGACCCATGGCGACA
 GGAATGGGAGAACGGGTGCAAGGTGCTGCCAGGGCATCCAGAGGCCGGTGGTGA
 GGATCCTCCACCACTGGCAAGGCCCTGCCAGGCATCCCCAGCACACATGCTTGGT
 GAGGGCTCCACGCTGATTGGCAGGGGGCAGCCGTATGACTTGGACAGGATTGATGCTA
 CTGGCTGGAGCTCATCAACTCGGAGCTTAAGGAGATGGAGAGGCCGGAGCTGGACAGCTGA
 CATTAGAGCCTGGCTGGAGGAGCCTGTGCCACCCAGAATATGGCAGGGCCATT
 GAGACGCGAGGAGGGCTGGCATCGAGTACGAGCAGGAGATGGTCTCGCGACGTGTCGCTC
 TCCGTAGGGCGAGGATGGCAAGGAGATGGTCTCTGTGACAAGTCAACGTCTGTGTCATC
 AGGCATGCTACGGGATCCTCAAGGTGCCACGGGAGCTGGCTGTGCCGGAGCTGCCCTG
 GTGTCAGGCTTACAGTCAGCTGCTCTGCCAACGGGAGCTGGCTTGAAGCCACTAGAAG
 TGGGACCAAGTGGTGCATGTCAGCTGCTGCCATTGGATTCTGGAGGTCAAGCATCGGCTGCC
 CAGAGAAGATGGAGCCATCACCAAGATCTCCATATCCAGCCAGCGCTGGCTCTGTCC
 TGCAGCCTCTGCAAGGAATGCACAGGCACCTGCATCAGTGTCCATGCCCTCGTCAC
 AGGGTCCATGTACATGCGCTTGGACACGGCTGGAAATGCGACTATATTAGCAGACA
 ACAGATGGCTCAAGTTCAGTCTTCAGGAGCAGCAGTGCACGGGGGCCAGCTGAATGAG
 CCCACATCTGACGCCACGGACCCAGCAGCTGGCAGGGACCTGGAAAAGGTGACCGCTGCC
 CAAGCAGCGGCTGCACTGAGAGGAGCTTACAGAGCTGGTGGAGCGGGTGGAGGTGG
 CTGAGCGCTGGACCTGGCTGAGGACTGGTGCACCTTCATCACAGTACTGGAAAGCTGAAG
 AGGAAAGCCATGCCACCGCGCTGCTGACCCCCAACAGGAGCTGGACAACCTGGC
 CCAGCAGGAGGAGACGAGCTGGCTTACCGCCGCCGAGTGGACCTTCACCCATCTGCCAGGACC
 TAGAGAGGGTTAGAAATCTGCTCATGGTACAAGGCCAGAGAACGAAACAGCCATC
 TGCAAACCTCCAGGAGCAGATATTCCACCTGCAAGTGAACATTATTGAACAGGATCTGTGTC
 AGCAGGCTGTCCACCTCATTCCCCATCGATGGCACCTTCTCAACAGCTGCTGCCACAGT
 CGGTGAGATCACAGCAGAGAACATGGCATGAGCAGTGCCACTGAAACAATGGGCACCGC
 GAGGACCCCTGCTCCAGGGCTGCTGAGGAAACTGCTGAGGGACAGGAGACACTGCTCAG
 CTTCATCGGGACCCCTCGCTGCCACCTGGTGAACCTGCTAGGAAGGCCAGGGCCACCC
 GCCTGCTGCCAAGAAGAACACCACCAACCCAGGCCAGGAGCGGGCTGGTCAAGGACG
 ACTCCAGACAAAGCCCCAAGAGACCTGGGCCAGGATGAGCAGGAGTGGCAAGGGGGTCA
 AGGGCACCTACCGAGAACCTGGCACATCTCTCACTTGGCTCAGCCCTGCA
 CGGGGACTGCCCCATCAGGCCACCCCTGAAAGCCCCCGCACTGGCCCTGAGACCCCG
 GACGAGGAGCCTCAGTAGCTGACTCAGATGTCCAAGTGCCTGGCCCTGCAAGAGCCC

- continued

SEQUENCES

TAAGCCCTTGGGCCGGCTCCGCCACCCCGCAGAGACAAGGTAACCCGGAGATTGCCGGGTG
CCAGGCCTGATCTGGGATGGGACCACTTCAGCTGTGGCTGAGAGGCCAAGGTCAAGCTG
CATTTGACACTGAGACTGATGGCTACTTCTCTGATGGGAGATGAGCGACTCAGATGTAGA
GGCGAGGACGCGTGGGTGCAAGCGGGTCCCCGGAGGCAGGGCAGAGGAGGTGGTCCGA
TGGCGTACTGGCCTCTAA

SEQ ID NO: 135
PKIB protein sequence
MRTDSSKMTDVESGVANFASSARAGRNNALPDIQSSAATDGTSPLKLEALSVKEDAKE
KDEKTTQDQLEKPONEEKCPFLY

SEQ ID NO: 136
PKIB DNA sequence
Attaggacagattcataaaaatgactgacgtggagtctgggtcgccaatttgcatcttc
agcaaggcgcggccgcggaaatgccttaccagacatcccgagggtttcagtgccacacacggaa
cctcagatggccctcaaactggaggctctccgtgaaggaaatgcaaaagagaaatg
aaaaaaaaacaacacaagccaattggaaaagccctcaaatagaagaaaaatgcccacttctt
gtac

SEQ ID NO: 137
POLE4 protein sequence
MAAAAAGSGTPREEEVPAGEAAASQPAPTSVPGARLSRLPLARVKALVADPDVTLAG
QEAIFILARAELFVETIAKDAYCCAQQGKRKTLQRRLDNDIAEVDEFAFLGTLID

SEQ ID NO: 138
POLE4 DNA sequence
ATGGCGCGCGCGCGCGCGAGGAAGCGGGACGCCCGAGAGGGAGGGAGGTACCTGCTGGGGA
GGCAGCGCCCTCGCAGCCCCAACGGAGTGTGCTGGGCTCGTCTCTCGAGGGTTGC
CTCTGGCGAGTGAGGGCAGATCCCAGCTGACGCTAGGGACAGGAA
GCCATCTCATTCTGGCACGAGCCGGAACTGTTGTGGAGACATTGCAAAGATGCTA
CTGTTGCGCTCAGCAGGGAAAAGGAAAACCTTCAGAGGAGAGACTGGATAATGCAATAG
AAGCTGTGGATGAATTGCTTTCTGGAGGTACTTAGATTGA

SEQ ID NO: 139
PRKRIR protein sequence
MPNFCAAPNCTRKSTQSLSLAFFPRDPARCKWVNCRRADLEDKTPDQLNKHYRLCAK
HFETSMICRTSPYRTVLNLDNAIPTIFDLTSHLNPNHSRHRKRIKELSEDEIRTLKQKKID
ETSEQEQKHKETNTNSNAQNPSSEEEGEQDEDILPLTLEEKENKEYLKSLEFILILMGKQN
IPLDGHEADEIPEGLFTPDNFQALLECRINSSEEVLRKRFETTAVTNLFCSKTQQRQMLE
ICECSIREETLEVRDHSFFSII TDDVVDIAEHHLPVLVRFVDESHNLREFEFIGELPYE
ADEIILAVKFHTMITEKWLMEYCRQAYIVSSGESSKMKVVASRLLEKYPQAIYTLCS
SCALNMWLAKSPVMGVSVALTIEEVCSFFHRSQPLLLELDNVISVLFQNSKERGKELK
EICHSQLWTGRHDAFEILVELLQLALVLCLDGINSDTNIRWNNYIAGRAFVLCSAVSDFDFI
VTVVVLKVNLSPTRAFGNLQGQTSVDFFAAGSLTAVLHSLNEVMENIEVYHEFWFEEAT
NLATKLDIQMLPGKERRAHQNLNESQLTESYYKBTLSVPTVEHHI QELKDIFSEQHLK
ALKCLSLVPSVMDNLQKENTSEEHHADMYRSIDLNPDTLSAELHCHWRIKWKRKGDIELPS
TYEALHLPDIKFPNVYALLKVLCLLPVMKVENERYENGKRRLKAYLRNLTQRSSNL
ALLNINF DIKHLDLMDVDTYIKLYTSKSELPTDNSETVENT

SEQ ID NO: 140
PRKRIR DNA sequence
ATGCCGAACTTCTGCCTGCCCAACTGCACCGGAAGAGCACGCACTCGACTGGCCTT
CTTCAGGTTCCCGGGGGCTGCCAGATGCCAGAAGTGGGGAGAACTCTGAGGAGAGCAG
ACTTAGAAGATAAAACACTGTAGCTAGCTAAATAAACATTATCGATTATGTGCCAAACATT
GAGACCTCTATGATCTGTAGAACACTGCTCTATAGGACAGTTCTCGAGATAATGCAATACC
AACAAATTGATCTTACAGTCATTGACAACCCACATAGTAGACACAGAAAAGCAATAA
AAGAAGCTGAGTGAAGATGAAACTCAGGACACTGAAACAGAAAAAAATTGATGAAACTCTGAG
CAGGAACAAAAACATAAGAACAAATAGCAATGCTCAGAACCCAGCGGAAGAAGAGGAC
TGAAGGCGAAGATGAGGAGCATTTACCTCTAACCCCTGTAGAGGAAGGAAACAAAGAATACC
TAAAATCTCTATTGAAATCTTGATGGGAAAGCAAACATACCTCTGGATGGACAT
GAGGCTGATGAAATCCAGAAGGTCTTACCTCCAGATAACCTTCACTGCACTGCTGGAGCTG
TCGGAAATTCTGGTGAAGGAGTTCTGAGAAAGCGTTGAGACAACAGCACTGAAACAGCT
TGTGTTGTCAGGAAACACAGCGAGGAGATCTGAGAGCTGTGAGAGCTGTGAGAA
GAAACTCTCAGGGAAAGTGGAGACTCACACTCTTCTTCCATTACTGAGCATGCTAGTGGA
CATAGCAGGGGAAGAGCACCTACCTGTGTTGGTGGAGGTTGTGATGAATCTCATAACCTAA
GAGAGGAATTATAGGCTTCGCTTATGAAGCCGATGAGAAATTGGCTGTGAAATT
CACACTATGATAACTGAGAAGTGGGATTAATATGAGTATTGCTGTCAGGCTTACAT
TGTCTCTAGTGGATTCTTCTTCCAAAATGAAAGTTGTTGCTCTAGACTTTAGAGAAATATC
CCCAAGCTATCTACACACTCTGCTCTTCTGTGCTTAAATATGTTGCTGCAAAATCA
CCTGTTATGGGAGTATCTGTCAGTAGGAACAATTGAGGAAGTTGTTCTTTTCCATCG
ATCACCAACACTGCTTTAGAACATTGACAACGTTCTGTTCTTCACTGAGTGTGTT
AAAGGGTAAAGAAGTGAAGGAATCTGCCATCTCAGTGGACAGGCAGGCATGATGCTTT
AAATTTAGTGGAACTCTGCAAGCACTGTTATGTTAGTGGTATAAATAGTGACAC
AAATATTAGTGGAAATAACTATAGCTGGCGAGCATTGACTCTGCACTGAGTGTCA

-continued

SEQUENCES

ATTTGATTCTTGTACTATTGGTCTTAAAGTGCTATCTTACAAGAGCCTT
GGGAAAACCTCCAGGGGAAACCTCTGATGTCCTTGGGCCGGTAGCTTGAGT
ACTGCACTCACTAACGAAGTGATGGAAAATTAGTAAAGTTATCATGAAATTGGG
AGGCCAACAAATTGGCAACCAAACCTGTATACTTAAAGTAACCTCGGAAATTCCGCAGA
GCTCACCGGGTAACCTGGAAATCTAGCTAACCTCTGAGAGTTACTATAAAGAACCCCTAAG
TGCCCCAACAGTGGAGCACATTCTAGGAACTTAAAGATATATTCTCAGAACAGCACCTCA
AAAGCTTAAATGCTTATCTGGTACCCCTCAGTCAGGAGAACACTCAAATTCATAGTCG
GAGGAACACCATGCTGACATGTATAAGTGACTTACCCAATCTGACACGCTGTAGCTGA
GCTTCAATTGGTGGAGAACATGGAAACACAGGGGAAAGATATAGGCTTCCGCACCA
TCTATGAAGCCCTCACCTGCTGACATCAAGTTTTCTAATGTGTATGCTTGTGAAG
GTCTCTGTATTCCTCTGTGATGAAGGTTGAGAATGAGCGGTATGAAAATGGACGAAACGG
TCTTAAAGCATATTGGAGAACACTTGGACAGCACAAAGGTCAGTAACCTGGCTTGGCTTA
ACATAAATTGGTATAAAAACACGACTGGATTAATGGTGGACCATATATTAAACTCTAT
ACAAGTAAGTCAGAGCTTCCACAGATAATTCCGAACACTGGAAAATACCTAA

SEQ ID NO: 141
 PYGO2 protein sequence
 MAASAPPDPDKLEGGGGPAPPPAPGSTGRKQKGALQMKSPEKKRRKSNTQGPAYSHLTERFAPPPTPMVHDLVASNSPFEDDFGAPKVGVAAPPFLGSPVPFGGFRVQGGMAGQVPPGYSTGGGGGQPQLRRLQQPPPFPNPNGMGAFNMPQPGPYPPGNMNFPSQOPENQLGQNESPPSGQMMPGPVGGGPQPMISPTMGQOPRAELGPSPSLQRFAQPGAPFGGPSPLQRPGQGLPSLPNNTSPFPGPDPGPFPQPGGEDGKPKLNPPASTAFAQPGQEPHSGSAAAANGNQNSPPPNSSGRGGCTGPDTANDLAPPKGAGGGSGPQPPPLVYPCGACRSVEVNDDQDAILCEASCQKWFHRECTGMTESAYGLLTTEASAWWACDLCLKTKEIQSVIIREGGMQGLVAANDGL

SEQ ID NO: 143
RANBP1 protein sequence
MAAKADTTHEHDHTSTENTDESNDHPQFEPIVSLPQEIKTLEEDEEELFKMRALKERFAS
ENDLPEWKERGTGVDVLLKHKEKGAIIRLLMRRDKTLKICANHYITPMMEKLPNAGSDRAW
VWNTHADFADECPCPKPELLAIRFLNAENAQFKFTKPEECRKIEEREKKAGSGKNDHAEKV
AEKLEALSVKEEKTDAEEKOPTFLY

```

SEQ ID NO: 144
RANBP1 DNA sequence
atggccggccccaaggacactcatgaggaccatgatacttccactgagaatacagacgagtc
caacccatggacccctcaggatgttgccaaatagtttcttcttgcgaaagaattaaaacactgg
agaagatgtaaaaaggaaactttttaaatgcggccaaaactgttccgatttgcctcgagaa
gatctcccagaatggaaaggcgcggcactgtgcagctcaagctctgaagcacaaaggagaa
agggggccatccgcctctcatgcggaggacaagaccctgaagatctgtgcacaaccactaca
tcacgcgcatgtatggagctgaagcccaacgcgcggtagcgcaccgtgcctgggtctggaaacacc
cacgcgtacttcgcgcacgcgtgcggccaaagccgcggactgtgccttgcatttcgttgc
tgaaatgcacagaaaaatccaaaacaaaggatgttgcggaaaggatgcgaaaggagag
aaaaaaaggacggatcggccaaaatgtatgcgcggaaaaggatgttgcggaaaaggatgc
ctctcggtqaqqqaaqqqaccqaaqqqatctqadqqaqqcaaccaactttcttqta

```

SEQ ID NO: 145
 RPRD1B protein sequence
 MSSFSESALEKKLSELNSQHSVQTLSSLWLIIHHRKHAGPIVSVWHRELRAKSNRKLTFLYLANDVIQNSKRKGPFTRPESVLVDASFVAREADEGCKKPLERLLNWIQERSVYGGEFIQQQLKLSMEDSKSPPKATEKKSLKRTFQQTBEEDDDYPGSYSPQDSAGPLLTELIEKAAQDNAGCCTTNTWQKAGIPEQEVNLKIKITVQYAFSCKWYDQFQKQVLLKLV

- continued

SEQUENCES

NGRLAAELEDERRQLARMLVEYTONQKDVLSEKEKKLEEYKQKLARVTQVRKELKSHIQSL
 PDLSLLPNTGLAPLPSAGDLFSTD

SEQ ID NO: 146
 RPRD1B DNA sequence
 ATGCTCTCTTCTGAGTCGGCGCTGGAGAAGAAGCTCTGGAGCTGAGCAACTCTCAGCA
 CAGCGTGAGACCCCTGTCCCTTGGCTATCCACCACCGAAGCACGCCGGACCCATCGTCT
 CGCTGTGGCACCCGAGACCCAAATCAAATAGAAAGCTTACTTTCTGTATTAA
 GCGAATGATGTCATCAAACAGTAAAAGGAAAGGACCTGAATTCACTAGAGAATTGAATC
 TGTCTTGATGCTTTCTCATGTCAGAGAGGCAGATGAAGGCTGTAACACCTT
 TAGAAAGATTGCTGAACATCTGCAAGAACGAAGTGTGTATGGCGGAGGTTACAGCAG
 CTGAAGCTGTCTATGGAGGCTCAAGAGCCCCAACAGAACAGAAAGAGAAATC
 TCTGAAACGAACATTTCAGCAATTCTAGGAGGAGGATGACGACTACCCCTGGCAGCTACT
 CTCTCAGGATCCTCTGAGGACCCCTCTGAGGAGACTAACTCAAACCTTGAGGAT
 CTGAGAAATGCCGATCAGGGATGCTACTGTCGACAGAAAATTGCTTCTGCCAGGA
 ATGTCAGATGTTCTATTGAAAAAAACAGACAAAGGGCAGCTAACGTCCTTCAA
 AAACAGTAGATGAGACATGCTGTACTTAGCAGAATATAACGGGCGCTGGCAGCAGACTG
 GAGGACCTGCGCAGCTGGCTGGATGTTGGAGTATACCCAGAATCAGAAAGATGTTT
 GTCGGAGAAGGAGAAAAACTAGAGGAATAACACAGAAGCTGACGAGTAACCCAGGTCC
 GAAGGAACGAAATCCCATATCAGAGCTTGGCAGACCTCTACTGCTGCCAACGTACA
 GGGGCTTAGCCCCCTGCCCTGCTGGGACCTGTTTCAACTGACTAG

SEQ ID NO: 147
 SPIN1 protein sequence
 MTPFGKTPQGRSRADAGHAGVSANMMKKRTSHKKHRSVGPSKPVSQPRRNIVGCRIQH
 GWKEGNGPVTVQWKGTVLDQVPVNPSPLYLIKVDGFDCVYGLELNKDERVSALEVLPDRVAT
 SRISDAHLADTMIGKAVEHMFEDEDGSKDEWRGMVLARAPVMTWFYITYEKDPVLYMYQ
 LLDYKEDGLRIMPDSNDSPPAEREPGEVVDLSVGKQVEYAKEDGSKRTGMVIHQVEAKP
 SVYFIKFDDDFHIYVYDVLVKT

SEQ ID NO: 148
 SPIN1 DNA sequence
 ATGAAGACCCCATCGGAAAGACACCTGGCCAGCGGTCCAGAGCTGTCAGGCCATGCTGG
 AGTATCTGCCACATGATGAAAGAGGACATCCCACAAAAACATCGGAGCAGTGTGGCTC
 CGAGCAAACCTGTTCCAGCCCCGGGAACATCGTAGGCTGAGGATTCACTGGTAA
 AAAGAGGGAAATGCCCTGTGACCGTGGAAAGGAAACCGTCTGGACCAAGGTGCTGTAA
 TCCTCTTGTATCTTAAATACGATGGATTGTTACTGTGTTATGGACTAGAACTTAA
 AAAGATGAAAGACTTCTCGCTTGAAAGTCTCCCTGATAGGTTGCGACATCTGAATCAGC
 GATGCACTTGGCAGACACAAATGATGGCAAAGCAGTGGAACATATGTTGAGCAGAGGA
 TGTTCTAAAGATGAGTGGAGGGAAATGGCTTAGCACGTGACCTGTGATGAACACATGGT
 TTACATACCTGAGAACGACCTGCTCTGACATGACCAACTCTAGATGATTACAAA
 GAAGGCACCTTCGCATTATGCGTATTCCAATGATTCACTCCAGCAGAAAGGGAACCAGG
 AGAAAGTGTGGACAGCTGGTAGGCAACAAAGTGGAAATATGCCAAAGAAGATGGCTGAAA
 GGACTGGCATGGTCATTGATCAAGTAGAAGCCAAGCCCTCGCTATTTCATCAAGTTGAT
 GATGATTCCATATTATGTCAGATTGGTGAACACATCTTAG

SEQ ID NO: 149
 SS18L1 protein sequence
 MSVAFA SARPKGEVTTQQTQKMLDENHHLIQCILEYQSKGKTAECTQYQQILHRNLVY
 LATIADSQNMQSLPAPPQTQMNLLPGALTQSGSSQGLHSQGLSDAI STGLPPSSLQ
 GQIGNGPSHVSMQTAPNTLPTSMISGPGYSHAGPASQGVPMQQGTIGNYVSRTNIN
 MQSNPVSMIQQQAATSHYSSAQGGSQHYQGQSIAMMGGQGSQGSSMMQQRPMAPYRPSQQ
 GSSQQYLQGEEYGEQYSHSQAEPMGQOYYPDGHDYAYQQSSYTEQSYDRSFESTQ
 HYEGGNSQYSQQQQTYSQQQYPSQSQYPSQGQQGYGSAQGAPSQYPGYQQ
 GQQQYGSYRAPQTAPSQQQPYGYEQQQYQNL

SEQ ID NO: 150
 SS18L1 DNA sequence
 Atgtccgtggccctcgcgctgcccggccaagaggcaaaggggaggttacgcagcaaaccat
 ccagaagatgctggacgagaaccaccacactgatccagtgcacccctggagttacccagaccaagg
 gcaagacggccgagtgacgcacgcattaccaggacatctgcacccggacccctggatatacctggcc
 acgatcgcacgtcccaaccatgcacgtccctgtcttcgtccccccccccgcacccat
 gaaacctggccctggagccctgactcagacggccatccgcacccctcccttcgtccggc
 gcctgactgacgcacatcagcaccggccatccgcacccctcccttcgtccggc
 aacggccgcacgcacgtgtccatgcacgcacggccctaaacacgcgtcccoaccacccat
 gagcatctctggccggcgtacagccacgcgggacccgcctcgccggcgtcccatgcagg
 ggcacaggcaccatcgccactacgtgtctcgaccaactacaatgcacgtccaaaccacgtc
 tccatgatcagcagcaggccgcacgcacgtccgcacgcacgcacgcacgcac
 ctaccaggcccatcgccatcgccatgcacggccatccgcacggccatccgcac
 ggcacgcggccatggccctccgcacggccatccgcacggccatccgcac
 caggaggagtagtacatggcgacgcacgcacccgcggccggcggccatggcc
 ggcacgcacccgcacggccatggcgatccgcacgcacgcacgcacgcac
 gtcacgcacccgccttcgcaggatccgcacgcacgcacgcacgcac
 agccacgcacccgcgggttaccagcagggtgcgcgcacgcacgcacgcac

- continued

SEQUENCES

gtacccccagccagcagactaccccgccagcagcaggctacgggtctgccaggagccc
cgtcacagtaccggcttaccagaaggccaaaggccagcgtacggagactaccgagcc
cagacagcgcgtctgcccagcagcagcggccctacggctatgaacagggccagtaggaaa
ttaccagcagttg

SEQ ID NO: 151

TADA3 protein sequence
MSELKDCPLQFDKSVHLCPRYTAVLARSEDDGIGIEELDTLQLELETLLSSASRR
LRVLEAETQIILTWDQDKGDRLFLKLGRDHELGAPPKHGPKQKLEGKAGHGPGPGPR
PKSKNLQPKIQYEFTDIDPVPRIPKNDAPNRFWASVEPYCADITSEEVRTLEELLKPP
EDEAEHYKIPPLGKHYLSQRWAQEDLLEEQKDGAARAAVADKKGLMGPLTELDTKVDAL
LKKSEAQHEQPEPDGCPCFGALTQRLQALVEENIIISPMEDSPIDPMGKESGADGASTSPR
NQNKPFSVPHTKSKEELIAQGLLESEDRPAEDSEDEVLAELRKRAELKALSAHN
RTKHHDLRLAKEEVSRQELRQVRMADNEVMDAFRKIMAARQKKRTPTKKEKDQAWKTL
KERESILKLLDG

SEQ ID NO: 152

TADA3 DNA sequence
ATGAGTGAGTTGAAAGACTGCCCTTGCAGTTCCACGACTCAAGTCTGGATCACCTGAA
GGTCTGTCCCCCTACAGGGCAGTGGCACGCTTGAGGATGATGGCATCGGCATCGAGG
AGCTGGACACCTTCAGCTGGAGACCTGCTTCTGGCAGCAGGATAAGAAAAGTTGACAGACGATT
GTGCTTGAGGCCAAACCCAGATCCTCACCGACTGAGGATAAGAAAAGTTGACAGACGATT
CCTGAAGCTGGTCGAGACCATGAATTGGAGCTCCCCCAAACATGGGAAGGCCAAGAAC
AGAAAACCTGGAAGGAAAGGCAGACATGGCCGGCCCTGGCCAGGACGCCAAATCCAAA
AACCTTCAGCCCAGATCAGGAAATATGAAATTCACTGATGACCTTATCGACGTGCCACGGAT
CCCCAAAATGATGCCCTAACAGGTTCTGGGCTTCAGTGGAGCCCTACTGTGCTGACATCA
CCAGCGAGGAGGTCGACACTTGAGGAGTTACTGAAGCCCCCAGAAGATGAGGCTGAGCAT
TACAAGATCCACCCCTGGGAGACACTTCCAGGCTGGCCAGGAGGACCTGCTGGA
GGAGCAGAAGGATGGGGCCGGGAGCAGGAGCTGGCTGAGAAGAAGAAAGGCCCTATGGGC
CACTGACCGAACTTGGCACTAAAGATGTGGATGCCCTGCTGAAGAAGTCTGGCCAGCAT
GAACAGCGGAAGATGGATGCCCTTGGTGCCTGACGCAGCGCCTCTGCAGGCCCTGGT
GGAGGAATAATTATTTCCCTATGGAGATTCTCTTCTGACATGTCGGAAAGAAT
CAGGGGCTGACGGGCAAGCACCTCCCGCAATCAGAACAGGCCCTTCAGTGTGCCGCAT
ACTAAGTCCTGGAGAGCCGATCAAGGAGGCTAATTGCCAGGGCTTTGGAGTCGA
GGACCGCCCCGAGAGGACTCCGAGGATGAGGTCTTGTGCTGAGCTCGAACAGGCCAGGTG
AGTGAGGCACTTAGCACCACCGCACAGAACAGCACGACCTGCTGAGGCTGGCAAAG
GAGGAGGTGAGGGCAGGAGCTGAGGAGCTGGAGGCTGGCATGGCTGACAACAGGAGTCATGGA
CGCCTTCGCAAGATCATGGCTGCCGGCAGAAGAACGGACTCCACCAAGAAAGAAAAGG
ACCAAGGCTGGAGAGACTCTGAAGGAGGTGAGGAGCATCTGAAGCTGCTGGATGGTAG

SEQ ID NO: 153

TAF6 protein sequence
MAEEKKLKLNSNTVLPSESMKVVAESMGIAQIQEETCQLLTDEVSYRIKEIAQDALFKMHM
GKRQKLTTSDIDYALKLNVEPLYGFHAQEPIPFRFASGGRELYFYEEKEVDSLIDINT
PLPRVPLDVCLKAHWFIEGQPAIPENPPPAPKEQKAEATEPLKSAKPGQEEEDGPLKG
KGGGATTADGKGEKKAPLLEGAPLRLKPRS IHELSVEQOLYYKEITEACVGSCAKRA
EALQSTATDPGLYQMLPRESTFISEGVRVNVQNNLALLIYLMRMVKAALMDNPTLYLEKY
VHELIPAVMTCTIVSRLQLCLRPDVNDHWALRDFAAIRLVAQIICKHFSTTINNIQSRTKTFT
KSWVDEKTPWTRYGSIAGLAELGHDVIKTLIILPRLQQEGERIRSVDGPVLSNIDRIGA
DHVQSLLKKHCAPVLAKLRRPPDNQDAYRAEFGSLGPGPLLCSQVVKARAQAALQAOQVNRT
TLLITQPRPTLTLSQAPQPQGPRTPGPLLKVPGSIALPVQTLVSRARAAPPQSPPPPTKFIV
MSSSSSAPSTQQLSLSAPGSGTTTSPVTTVPSVQPIVKLVSTATTAPPSTAPSGP
GSVQKYIVVSLPPTGEGKGGBTSHPSPVPPASSPSPLSGSALCGGKQEADSPPPAPGT
PKANGSQPNNSGSPQAPL

SEQ ID NO: 154

TAF6 DNA sequence
atggctgaggagaagaagctgaagcttagacaactgtgtccctcgaggtccatgaaggat
ggggctgtatccatggccatccccccatccaggaggagactgtccatgtcaacggat
aggtcagctaccgcataaagagatcgcacaggatgccttgcgttcatgcacatggggaa
ccgcagaagctcaccaccaggatgcattactacgcgccttgcgttcatgcacatggggaa
ctatggctccacccggccaggacttcccttcggcttcgccttcgggtggggccgggg
tttacttctatggaggaggagggttgcgttcgcacatcatcaataccctctggccgg
gtggccctggcgcgttcgcctcaagcttcattggcgtgcgcacatcgaggctgcac
cccgagaaccggcccccagctccaaagagaacagaaggctgaaaggccacagaacc
agtcagccaaggccaggccaggagaagcggcccccgtgaaggccaaaggtaaggggcc
acagccgacggccaaaggaaagagaagaaggcccccctgtcgagggggcccccctgc
actgaagccccggagcatccacgcgcgttcgtggcgcacgcgtctactacaaggagatca
ccgaggccgtgcgtggcccttcgcgcaggccaaaggccggaaaggccctgcacatggccac
gaccctggactgtatcagatgtgcgcacgggttcgcgttcgcgttcgc
gaacgtgttcagaacaaccctggccctactcatctacccgtatgtgaaaggccgt
tggacaacccttcacgcgttatctagaaaaatactgtccatgcgttcgcgttatgt
tgcacatgtgcgcacaggatgtggcgcaccatgcgttcgcgttatgtgaaaggcc
cttgcgtcccgccctggcccgatctgcacgcattttaccaaccactaacaacatcc

-continued

SEQUENCES

agtccggat caccaggacccatccaaagacttccatgggatcgacgagaagacgcctggacgact
cgtagtgcgtccatcgccggatggcgtggatggacacgttatcaactgcgttt
gccccggctgcggcggaggggggggggatccggatcgccgtgtggacggccctgtgtggat
acatcgaccggatggcggacccatgtgcggacgacttcgtgtggaaactcgctgttgtt
ctggcaaagtgcgcacccgcctgacaatcaggacgcctatcgccggagaattcggttcc
tggggcccttcctgtcccaagggttcggggccagggtgtctgcgggttcgc
aggtecaacaggacccatcgaccatcgcggccggccggccacgtgcacccctcgccgg
ccacacgtccggccctcgccacccctgtggatgtggatccatcgacttcgtt
ccagacactgggtgtccgcacgcgggtggccaccacagcgttcccttcctccaaccac
ttatgttaatgtcatcgccatccaggccggccatccaccacgcgggttcgtccctcgacc
tcggggcccccgtcgatgtccaccacacttcggccgtcaccacccacgtccgggtga
gccccatcgccatcgatgtggatccaccacgcggccggccacggactgtcccttcgtt
ctggggatgtccaggatgtacatcggttctacttcccccacaggggaggggcaaggagg
ccccatcccttcatccatccatcgatgtcccccacactcgccggccatcgccgg
tgcctttgtggggggaaacggagggtggggacatcccccctccagctccaggagactccaa
aaggccatggatcccgcccaactccggatccctcgatcgatcgatcgatcgatcgat

SEQ ID NO: 155
TBPL1 protein sequence
MDADSDVALILITNVVCVERTRCHLNLRKIALEGANVIYKRDAKGVLMLKLRKPRTATI
WSSGKIICTGATSEEAKFGARRLARSLQKLGQVIFTDFKVNVNLAVCNPMPFEIRLPF
TKNNRPHASYEPELHPAVCYRIKSLRATLQIFSTGSITVTGPNVKAVATAVEQIYPFPVPE
SRKEILL

SEQ ID NO: 157
VPS72 protein sequence
MSLAGGRPKTAGNRLSGLLEAAEEDDEFYQTTYGGFTEESGDDEYQGQDSQDTEDEVSD
FDI D E G P S S D G E A E E P R K R V V T K A Y K E P L K S L R P K V N T P A G S S Q K A R E E K A L L P L
E L Q D D G S D S R K S M R Q S T A E H T R Q F L R V Q E R Q G Q S R R K G P H C E R P L T Q E E L L R E A K I T E
E L N L R S L E T Y E R L E A D K K Q V H K K R K C P G I I T Y H S V T V P L V G E P G P K E E N V D I E G L D P A
P S V A L S T P H A G T G P V N P A R C S R T F I T F S D D A T F E E W F P Q G R P P K V P R E V C P V T H R P A L
Y R D P V T D I P Y A T A R A F K I I R E A Y K K Y I T A H G L P P T A S A L G P G P P P E P L P G S G P R A L R Q K
I V I K L

SEQ ID NO: 159
ZNF133 protein sequence
MAFRDVAVDETQDEWRLLSPAQRTLYREVMLENYSNLVSLGISFSKPELITQLEQKGKETW
REKKCSPATCPDPPEPELYLDPFCPPGFSQSKFPQMQLVLCNHPPWIFTCLCAEGNIQPGD
PPGGDQEKKQQQAEGRPWSIDQAEGEPGEGAMPLGRTKRKTLLGAFSRPPQRQPVSSRNGL
PQHVEFASCPAATCTGMDTETTLLKPLRIVGCTTGTCAGCCLLGEGKMTNTLQHQDNUCCP

- continued

SEQUENCES

YVCVGCEKGFSLKSLRHKQKAHSGEKPIVCRECGRGENRKSTLIHERTHSGEKPYMC
ECGRGFSQKSNLIIHQRTHSGEKPYVCRECGRGFSQKSQSAVRHQRTHLEEKТИVCSDCGL
GFSDRSNLISHQRTHSGEKPYACKECGRCFRQRTTLVNHQRTHSKEKPYVCGVCGHSFSQ
NSTLISHRRTHTGEKPYVCGVCGRGFSLKSHLNHRHONIHSGEKPIVCKDCCRGFSSQSNL
IRHQRTHSGEKPMVCGEGRGFSQKSNLVAHQRTHSGERPYVCRECGRGFHQAGLIRHK
RKHSEKPYVCMRCQCLGFGNKSALITHKRAHSEEKPCVCRECGQFLQKSHLTLHQMTHT
GEKPYVCKTCGRGFSLKSHLSRHKTTSVHHLPVQPDPEPCAGQPSDSLYSYL

SEQ ID NO: 160

ZNF133 DNA sequence

ATGGCATTCAAGGGATGTGGCTGGATTTCACCCAGGATGAGTGGAGGCTGCTGAGCCCTGC
TCAAAGGACTCTGACAGAGGGTGTGAGAACTACAGCAACCTGGTCTACTGGGAA
TTTCATTCTAACCAACCATCACCAGCTGGAGCAAGGGAAAGACCTGGAGAGAG
GAAAAAAATGTTCACCGGAACCTGTCCAGATCCAGAGCAGAGCTACCTCGATCTTT
CTGCCCTCCGGTTCTCAGTAGAAATTCCCCATGCAGCATGTGCTGTGTAATCATCCCC
CTGGATCTTACATGCTTGCAAGGTTAACATCCAGCTGGGATCAGGGCCAGGG
GACCAGGAGAACGAGAACAAAGCCTCTGAGGGGAGACCTGGAGTGATCAAGCAGAACGGTCC
TGAGGGAGAGGTGCGATGCCATTGTTGGAGAACAAAAGGACTCTGGAGCGTTCT
CCAGGCCACCCCAGAGGCAGCCAGTCAGCTCTGGAACGGCTCAGAGGGGGGGAGTTAGAA
GCCAGGCCAGCTCAGAGGGAACCTGAGAACAGACAATTGTTGAAGAGGATAGAAGT
CTTAGGATTGGAAACAGTCACATGTTGGAGACTTGAGCTTCAGCAAGAGATGACAAC
TGCTCAGTCACCAAGCGGATACTCAGGGAGAACGCTTACGTGCTGGGTTATGTGAGAAG
GGCTCAGCCTAAAGAAGAGCCTGCCAGACACCAGAAGGCACACTCGGGGAGAACGAAAT
TGTGTCAGGGAGGTGAGCAGGGCTTAACCGGAAGTCAACGCTAATCATACAGAACGGA
CACACTCCGGTAGAGAACCTTACATGTCAGTGAGTTGGGGAGGGCTTACGCCAGAACGTA
AACCTCATACACAGAGGACACACTCAGGGAAAAGGCTTATGTGTCAGGGAAATGTGG
CAAAGGCTTCAGCCAGAAGTCAGCTGCTGTGAGACACAGAGGACACACTGGAGGAGAAGA
CCATGTGTCAGTGAACGACTGTGCGCTGGGCTTACGGCAGAGCTCAAACCTCATCTCCACCG
AGGACGCACTCTGGGGAGAACGGCTACAGCCTCAAGGGAGTGGGCGATGCTTCAGGCAGAG
GACCACCTTGTCAACACCAGAGGAACACTCAAAAGGAGAACGCTTATGTGTCGGGGGTG
GTGGGCACAGCTTCAGCCAGAAATTCAACCCATCTCACAGGGGGACACACACTGGGAG
AAGCGTAGTTGTGGGGTGTGGGGAGGGCTTATGTCTCAAGTCACACCTCAACAGACA
CCAGAACATACACTCAGGAGAACGGCTTACGGCAGAGCTGGGCGGGGCTTCAGCC
AGCAATCAACCTCATCACACAGAACGGCCACTCAGGGAGAACGCCATTGGTGTGG
GAGTGCAGGGAGGGCTTCAGCCAGAACCTTGTGACACACCAGAGGACGCACTCAGG
GGAGAGGGCGTAGTGTGCGAGAGTCGGGGAGGGCTTACGGCAGAGCTGGGCGGGTCTCATCA
GGCACAAGGGAGAACGACTCGAGGGAGAACGGCTTACATGTGCAAGGACTGTGGGACTGGGCTT
GGCAATAAGTCAGCTCTAATCACACACAGGGCTCACTCGGAAGAGAACGCTTGTGTGG
CAGAGAGTGTGCGCAAGGCTTCTCCAAAAGTCACACCTCACCTTACATCAAATGACACATA
CGGGGGAGAGGACATATGTGTCAGACGTTGGGGGGCTTCAGCTCAAGTCTCACCTC
ACGAGACACAGGAAGAACCCCTGTCACCCAGACTGCCAGTGCAGCCGACCTGAGCC
GTGTGCAAGGGCAACCTCGGATTCTTATACTCTCTGA

SEQ ID NO: 161

ZNF 140 protein sequence

MSQGSVTFRDVIAIDFSQEEWKLQPAQRDLYRCVMLENYGHVLVSLGLSISKPDVVSLLEQ
GKEPWLKREVKRDLFSVSESSGEKFDFSPKNVIYDSSQYLMERILSOPVYSSFKGG
WKCKDHTEMLQENQGCIKRKVTVSHQEALAHMNISTVERPYGCHECGKTFGRFRFLVLHQ
RTHTGEKPYECKAFSRSNLTRHQRIHIGKKQYICRKCGKAFSSGSELIRHQITHTGEKP
YECIECGKAFRESHLTRHQISHTTPYECNECRKAFRCHSFLLIKHQRTHAGEKLYECD
ECGKVFTWHASLIQHTKSHTGEKPYACAECDKAFAFSRSFLSLHQRTHTGEKPYVCKVCNK
SFSWSSNLAKHQRTHTLDNPYEEYENSFNYHSFLTEHQ

SEQ ID NO: 162

ZNF140 DNA sequence

ATGTCTCAGGGTCAGTCAGACATTAGAGATGTGGCCATAGACTTCTCCAGGGAGGTGGAA
ATGGCTTCAGCCTGCTCAAAGAGATTGTACAGATGTGTAATGTTGGAGAACATATGGCCATC
TGGTCTCACTGGGTCTTCTCAATTCTAACAGGAGATGTGTTCTTATTGGAGCAAGGGAAA
GAACCTGGCTGGGGAAAAGGGAGGTGAAAGAGACTGTGTTCTAGGGTCAAGTGG
TGAGATCAAAGACTTCTACCAAAATGTATTATGATGACTCATCCAGTATTGATCA
TGGAAAGAATTCTAAGTCAGGCCCTGTGTTAGCTGGGAGGCTGGAATGCAAG
GATCATACTGAGAGTCAGCAAGAAAATCAGGGATGTATTAGGAAAGTAACAGTCATCA
AGAAGCCCTGGCTCAACATATGAATATCAGTACTGTGGAGAGGCCATTGGATGCCATGAAT
GTGGAAAACCTTGGTCAGCCTTCCCTGGTTACACCCAGAGGACTCATACTGGAGAG
AAACCATATGCACTGTAAGGAATGTGGCAAAACCTTACGGAGATTCTCAAACCTTGTGAAACA
CCAATGATACTACTGAAAGAAACCCATGAGTCAGACTGTAAGGACTGTAATAAAACATTCACTT
ACCTTCATTCTTATTGAACACCCAGAGAACGCAACACTGGGGAGAACCTTATGAATGACT
GAGTGTGGAAAGGCCATTAGCGGTGCCAACCTCACTCGCACATCAAAGAATTCACATAGG
AAAGAAAACATATATGAGGAATGTGGTAAAGCATTAGCAGTGGCTCAGAACACTCATTC
GCCACAGATTACACACTGGAGAACCTTATGAATGCACTGTTACACCCAGTGTGAAAGCATT
CTGGCTTCTCACACCTACTCGACATCAGAGCATCCATAACACCAACCCCGTATGAAATG
TAATGAATGTAGGAAAGCTTCCGTTGCACTCATTCTTATAACATCAGAGAATTCTAG
CTGGAGAAAAGCTCATGATGAATGTGGTAAAGTTTCACTTGGCATGCATCCCTT

-continued

SEQUENCES

ATTCACACATAAGCAGAGTCACACTGGAGAGAAACCTATCGCTGTGCTGAATGTGATAAAGC
CTTCAGCGGAGCTTTCCCTCATTCACATCAGAGAACTCATACTGGAGAGAAACCTATG
TATGTAAGGTATGCACAAATCCTTCAGCTGGAGCTAACACCTGCTAACATCAGAGGACA
CACACTCTGAAACCCCATTGAAATATGAAAATTCAATTACCACTCATTCTTACTGAA
ACACCACTGAA

SEQ ID NO: 163
ZNF 169 protein sequence
MSPGLLTTRKEALMAFRDVAAVFTQEKWLSSAQRQLYREVMLENYSHLVSLGIAFSKP
KLIQELOEGDPPWREENEHILLDLCPGRRITRSGVRD

SEQ ID NO: 164
ZNF 169 DNA sequence
ATGTCACCAGGACTCTGACAAACCAGGAAGGGAGGCATTGATGGCCTTCCGGATGTGGCTGT
GGCCTTCACCCAGGAAGGGAGCTTGGAGTCTGAGTTCTGCTCAGAGGACCTGTACAGGGAGG
TGATGCTGGAGAACTACAGCCATCTGGTCTCCCTGGGAATTTGCAATTCAAACAAAAACTC
ATCGAACAGCTGGAGCAAGGGACGACAACTTGGAGAGGAGAGAACGAACATCTCTGGACCT
TTGTCAGGAGCAGCGATACAAGGTGGAGTTCTGAGACTAG

SEQ ID NO: 165
ZNF254 protein sequence
MPGPPRSLEMGLLTFRDVAIEFSLEEWQHLDIAQQNLYRNVMLENYRNLAFLGIAVSKPD
LITCLEGKPEWNMKRHEMDDEPPGLDESILL

SEQ ID NO: 166
ZNF254 DNA sequence
ATGCCAGGACCCCTAGAACGCTAGAAATGGACTGTTGACATTAGGGATGTGCCATAGA
ATTCTCTGGAGGACTGGCACACCTGGACATTGCACAGCAGAATTATATAGAACATGTGA
TGTAGAGACAATCAAGAAACCTGGCCTTCCTGGTATTGCTGTCTCTAAACCCAGACCTGATC
ACCTGTCTGGACAAAGGGAAGAGCGCTTGAAATATGAAGGCACATGAGATGGTGGATGAAAC
CCAGGAGTTGGATTTCATTAATCTGTGA

SEQ ID NO: 167
ZNF566 protein sequence
MAQESVMSFDSVSDFSQEEWCLNNDDQRDLYRDVMLENYSNLVSMGHSISKPNVISYLEQ
GKEPWLADETRGWPVLSRCETKKFLLKKEIYEIESTQWEIMEKLTRRDFQCSSER
DWECKNRQFKBELGSQGHFNQLVFTHEDLPTLSHHPSFTLQQIINSKKKFCAKEYRKTF
RHGSQFATHEIIHTIEKPYECKEGCKSFRHPSRLTHHQKIHTGKPKFECKEGCKTFIGS
DLTRHRIITGEKPYECKEGCKAFSGSNFTRHQRIITGEKPYECKEGCKAFSGSNFTQ
HQRIITGEKPYECKEGCKAFSGSNFTRHQRIITGEKPYECKEGCKAFSGSNFTQ
HTGEKPYECKEGCKAFSGSNFTRHQRIITGEKPYECKEGCKAFSGSNFTQ

SEQ ID NO: 169
ZNF585A protein sequence
MPANWTSQPQKSSALAPEDHGSSYEGWSVFRDVAIDFSREEWRLHDPSQRNLYRVDVMLEY
SHLJLSTGYOVRPAEVYMLFQCKEPWALQGFRPPROSCPAPCLVNNSHHLQESERG

- continued

SEQUENCES

SEQ ID NO: 170

ZNF585A DNA sequence

```
ATGCCAGCTTAATGGACCTAACCTCAGAAAATCTCAGCCCTGGCTCCAGAGGATCATGGCAG
CTCTATGAGGGATCAGTGTCTTCAGGGATGGCTATCGATTTCAGCAGAGAGGAATGGC
GGCACCTGGACCCCTCTCAGAGAAACCTGTACCGGGATGTGATGCTGGAGACTACAGCCAC
CTGCTCTCAATAGGATATCAAGTTCTGAAGCAGAGGTGGTATGTTGGAGCAAGGAAAGGA
ACCATGGGCACTGCAGGGTGAAGAGGCCACGTAGAGCTGCCAGCACGTCTTGTGAAC
CCCATCACCTCAAGAAAGCTTCGAGGGTGA
```

SEQ ID NO: 171

ZNP689 protein sequence

```
MAPPSPALPAQPGPKARPSKRGRPRPRAKFDVAVYFSPEEWGLRPAQRALYRDVMRE
TYGHLGALGCAFPKPALISWLERNTDDWEPAALDPQEYPRGLTVQRKSRTKRNKEKEVFP
PPKEAPRKKGKRRRPSKPRLIPRQTSGGPICPDCGCTFPDHQALESHKCAQNLKKPYCP
DCGRRFSPSYPLVSHRRRAHSGCPYVCDQCGKRFSSQRKNLSQLQHVINTGEKPYHCPDCGR
CFRRSRSLANHRTTHGEKPHQCPSCGRRFAYPSLLAIHQRTHTGEKPYTCLECNRRFRQ
RTALVIHQRIHTGEKPYPCPDCERFFSSSLRVSHRRVHSGERPYACEHCEARFSQRSTL
LQHQLLHTGEKPYPCPDCGRAFRRSGSLAIHRSTHTEEKHLACDDCGRRFAYPSLLASHR
RVHSGERPYACDLCSKRFAQWSHLAQHQLLHTGEKPFPCLECGRCSRQRWSLAHKCSPK
APNCSPRSAIGGSSQRGNAH
```

SEQ ID NO: 172

ZNP689 DNA sequence

```
ATGGCGCCACCTTCCGCTCCCCTGGCGAGGGACCAAGGAAGGCCAGACCCAGTCGGAA
AAGGGGCAGGAGGCCAGGGCTCTGAAGTTCTGGACGTGGCCGTGACTCTCCCGGAGG
AGTGGGGCTGCCCTGGGCCCGCGAGGGCCCTGTACCGGGACGTGATGCCGGAGACCTAC
GGTCACCTGGGCGCTGGTGCAGGGTCCAAACCCAGCCCTATCTCTGGGTGGAAAG
AAACACCGATGACTGGGAGGCTGTAGATCCGAGGAGTACCCGAGAGGGCTAACAG
TCCAGAGAAAAGCAGAACCAAAGAAGAATGGGAGAAGGAAGTATTCCGCTAACAG
GCACCCGAAAGGGGAGCGAGGCCAGCAAACCCCGACTGATTCTAGGCAGAC
GTCCGGGGGGCCCATCTGCCCTGACTCGGGCTGACTCTCCCTGATCATCAGGCCCTGGAGA
GCCACAAGTGGCCAGAACTCTAAAAGCCTAACCCCTGGCCAGACTGTGGGCGCCGCTT
TCCATCATCCCTGTTGGTCAAGTCACCGGGGGCACACTCCGGAGTGCCCCATGTTG
TGACCAAGTGTGGCAAACGTTCTCCAGCGCAAGAACCTCTCCAGCACCAAGTCATCCATA
CAGGGAGAAGCCCTACCTGCCCTGACTGTGGTCGCTGCTTCCGGAGGAGCCGGTCTTG
GCCAATCACCGGACCACACACACAGGTGAAAACCCACAGTGCCCTAGCTGTGGACGTCG
CTTCGCTCATCCCTCCCTGCTGAGCCATCCACCCAGCTACACACAGGGAGAGAACCCCTACA
CTTGCTCGAGTCAACCGCCCTCCGCCAGCGCACGGCCCTCGTACATCCACAGCGCATC
CACACGGCGAGAAGCCCTACCCGTGCCCGACTCGGAGCGCGCTTCTCCCTCTCTCG
CTCTGGTCACTGCCAGCGCTGACTGTGGGAGCGTCCCTATGCTGAGGACACTGTGAGG
CCCGCTTCTCCCGCAGCAGCAGCTGCTCCAGCAGCGCTTCTGACACCCGAGAGAACCC
TACCCCTGCCAGACTGTGGGCGCTTCCGGCGAGCGCTCCCTGGCCATCCATCGCAG
CACGCACACAGAGGAGAAGCTGCAACGCCCTGCGACGACTGTGGTCGCCCTTGCCTACCCCT
ACTGTGGCCAGCCACCGGGCGTCACTGGGGCGAGGCCCTATGCTGCGACCTTTC
TCCAAGGGTTTGCTCAAGTGGAGCCACTGGGCCAGCAGCACAGCTGCTGACACCGGGAGAA
GCCCTTCCCTGCTCGAGTGTGGCGTCTCCGCCAGAGGTGTTCTGGCTGTCCACA
AGTGTAGCCCCAAGGCCCAAAGTGTAGCCCTAGATCTGCTATCGGGGCTCAGTCAGAGG
GGCAACGCCCATAG
```

SEQ ID NO: 173

ZNF765 protein sequence

```
MALPQGLLTFRDVAIEFSQEEWKLDPAQRTLRYRDVMLENYRNLVSLELSGECPLAAPAS
LDPAFLC
```

SEQ ID NO: 174

ZNP765 DNA sequence

```
atggcttcctcgtcggttctattgcattcaggatgtggccatagaattctctcaggagga
gttgaaatgcctggaccctgtcgatggactctatacaggacgtatgtggaaattata
ggaacctgttccctggatgttgcaggatgtccattggcagcacctgccttggac
ccagcttcttgac
```

SEQ ID NO: 175

ZNP81 protein sequence

```
MPANEDAPQPGEHGSACEVSFSFEDVTDFSREEWQQLDSTQRRLYQDVMLENYSHLLSV
GFEVPKPEVIFKLEQGEGPWTLEGAEAPHQSCDGKFGIKPSQRRIISGKSTFHSEMEGDT
LCSGLM
```

SEQ ID NO: 176

ZNF81 DNA sequence

```
atgccagctaaccggaggacgtccccagccaggaaacatggcagtgcctgtgaggatcagt
gtcatttgaggatgtgactgtggacttcagtagagaggactgtggcagcaactggactctactc
aaagacgcctgtaccaggatgtatgtggagaactacagccacctgtctcagtgggttc
gaagttcttaaccaggatgtatctcaagtggagcaaggagaggggccatggacattggaa
```

-continued

SEQUENCES

aggggaagccccacatcagagctgttagatggaaatttggaaattaagecctcccagagga
gaatttctggaaatctacatttcatagtgaaatggagggtgaagacacactgtgttcaggc
ctcatggg

SEQUENCE LISTING

Sequence total quantity: 176
SEQ ID NO: 1 moltype = length =
SEQUENCE: 1
000

SEQ ID NO: 2 moltype = length =
SEQUENCE: 2
000

SEQ ID NO: 3 moltype = length =
SEQUENCE: 3
000

SEQ ID NO: 4 moltype = length =
SEQUENCE: 4
000

SEQ ID NO: 5 moltype = length =
SEQUENCE: 5
000

SEQ ID NO: 6 moltype = length =
SEQUENCE: 6
000

SEQ ID NO: 7 moltype = length =
SEQUENCE: 7
000

SEQ ID NO: 8 moltype = length =
SEQUENCE: 8
000

SEQ ID NO: 9 moltype = length =
SEQUENCE: 9
000

SEQ ID NO: 10 moltype = length =
SEQUENCE: 10
000

SEQ ID NO: 11 moltype = length =
SEQUENCE: 11
000

SEQ ID NO: 12 moltype = length =
SEQUENCE: 12
000

SEQ ID NO: 13 moltype = length =
SEQUENCE: 13
000

SEQ ID NO: 14 moltype = length =
SEQUENCE: 14
000

SEQ ID NO: 15 moltype = length =
SEQUENCE: 15
000

SEQ ID NO: 16 moltype = length =
SEQUENCE: 16
000

-continued

```

SEQ ID NO: 17      moltype = length =
SEQUENCE: 17
000

SEQ ID NO: 18      moltype = DNA length = 86
FEATURE
source
1..86
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 18
gtttaagagc tatgctggaa acagcatago aagttaaat aaggctagtc cgttatcaac 60
ttgaaaaagt ggcaccgagt cggtgc                         86

SEQ ID NO: 19      moltype = RNA length = 86
FEATURE
source
1..86
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 19
gtttaagagc tatgctggaa acagcatago aagttaaat aaggctagtc cgttatcaac 60
ttgaaaaagt ggcaccgagt cggtgc                         86

SEQ ID NO: 20      moltype = AA length = 7
FEATURE
source
1..7
mol_type = protein
organism = synthetic construct
SEQUENCE: 20
PKKKRKV                                         7

SEQ ID NO: 21      moltype = AA length = 5
FEATURE
source
1..5
mol_type = protein
organism = synthetic construct
REPEAT
1..5
note = repeat unit: repeats 0 to 10 times
SEQUENCE: 21
GGGGS                                         5

SEQ ID NO: 22      moltype = AA length = 5
FEATURE
source
1..5
mol_type = protein
organism = synthetic construct
SEQUENCE: 22
GGGGG                                         5

SEQ ID NO: 23      moltype = AA length = 5
FEATURE
source
1..5
mol_type = protein
organism = synthetic construct
SEQUENCE: 23
GGAGG                                         5

SEQ ID NO: 24      moltype = AA length = 7
FEATURE
source
1..7
mol_type = protein
organism = synthetic construct
SEQUENCE: 24
GGGGSSS                                         7

SEQ ID NO: 25      moltype = AA length = 7
FEATURE
source
1..7
mol_type = protein
organism = synthetic construct
SEQUENCE: 25
GGGGAAA                                         7

SEQ ID NO: 26      moltype = AA length = 1368
FEATURE
source
1..1368

```

-continued

```

mol_type = protein
organism = Streptococcus pyogenes
SEQUENCE: 26
MDKYSIGLD IGTNSVGWAV ITDEYKVPSK KFKVLGNTRD HSIKKNLIGA LLFDSGETA 60
ATRLKRTARR RYTRRKNRIC YLQEFSNEM AKVDDSFHRR LEESFLVEED KKHERHPIFG 120
NIVDEVAYHE KYPTIYHLRK KLVSTDKD LRILYLALAH MIKFGRHFLLI EGDLNPNDNSD 180
VDKLFIFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN 240
LIALSLGLTP NFKSNFDLAE DAKLQLSKDT YDDDLDNLLA QIGDQYADLF LAAKNLSDAI 300
LILSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLKALVR QQLPEKYKEI FFDQSKNGYA 360
GYIDGGASQE EFYKFIKIPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH 420
AIIIRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE 480
VVDKGASAQS FIERMTNFDK NLPNEKVLPK HSLLYEYFTV YNELTKVKYV TEGMRKP AFL 540
SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLKI 600
IKDKDFLDNE ENEDILEDIV LTTLTFEDRE MIEERLKTYA HLFDDKVMKQ LKRRRTGNG 660
RLSRKLINGI RDQSGKTTIL DFLKSDGFAN RNFMQLIHDD SLTFKEDIQK AQVSGQGDSDL 720
HEHIANLAGS PAIKKGILQI KVVDDELVKV MGRHKPENIV IEMARENQTT QKGOKNSRER 780
MKRIEEGIKE LGSQILKEHP VENTQLQNEK LYLYYLQNRR DMVYDQELDI NRSLSDYDVHD 840
IVPQSFLKDD SIDNKVLTRS DKNRGKSDNV PSEEVVKKMK NYWRQLLNAK LITQRKFNDNL 900
TKAERGGLSE LDKAGFIKRQ LVETRQITKH VAQILDSSRMN TKYDENDKLI REVKVITLKS 960
KLVSDFRKDF QFYKREINN YHHAHDAYLN AVVGTALIKK YPKLESEFVY GDYKVYDVRK 1020
MIAKSEQEIG KATAKYFFYS NIMNNFKTEI TLANGEIRRK PLIETNGETG EIVWDKGRDF 1080
ATVRKVLSMP QVNIVKKTVE QTGGFSKESI LPKRNRSKLLI ARKKDWDPKK YGGFDSPVVA 1140
YSVLVVAKVE KGKSKKLKS VELLKTIME RSSFEKNPID FLEAKGYKEV KKDLIILPK 1200
YSLFELENGR KMLASAGEL QKGNEALAPS KVYNFLYLAS HYEKLGSP E DNEQKQLFVE 1260
QHKHYLDEII EQISEFSKRV ILADANLDKV LSAYNKRDK PIREQAENII HLFTLTNLGA 1320
PAAFKYPDTT IDRKRYTSTK EVLDATLHQ SITGLYETRI DLSQLGDD 1368

```

```

SEQ ID NO: 27          moltype = AA length = 1053
FEATURE             Location/Qualifiers
source              1..1053
mol_type = protein
organism = Staphylococcus aureus
SEQUENCE: 27
MKRNYILGLD IGBTSGVYGI IDYETRVID AGVRLFKEAN VENNEGRRSK RGARRLKRR 60
RHRIQRVKKL LFIDYNNLTDH SELSGNPYE ARVKGLSQLK SEEFSAAALL HLAKRRGVHN 120
VNEVEEDTGN ELSTKEQISR NSKAEEKVY AELQLERLKK DGEVRGINSR FKTSVDYVKEA 180
KQLLKVKQKAY HQLDQSFIDT YIDLLETTRRT YYEGPGEGSP FGWKDIKEWY EMLMGHTYF 240
PEELRSVKYA YNADLYNALN DLNNLVITRD ENEKLEYYEK FQIIENVFQ KKKPTLKQIA 300
KEILVNEEDI LGYRVTSTGK PEFTNLKVKYH DIKIDTARKE IIENAELLDQ IAKILTIYQ 360
SEDIQEELTN LNSELTKQEEI EQISLNKGYT GTHNLSLSKAI NLILDELWHT NDNQIAIFNR 420
LKLVPKVKDL SQQKEIPPTL VDDFILSPVV KRFSIQSIKV INAIKKYGL PNDIIIELAR 480
EKNSKDAQKM INEMQKRNQ TNERIEEEIR TTGKENAKYL IEKIKLHDMQ EGKCLYSLEA 540
IPLEDLNNP FNYEVDDHIPI RSVSDFDNSPN NKVLVKQEEEN SKKGNRTPFQ YLSSSDSKIS 600
YETFKKHILN LAKGKGRISK TKKEYLLEER DINRFSVQKD FINRNLVDTR YATRGLMNLL 660
RSYFRVNLLD VKVKSINGGF TSFLRRKWFK KBERNKGYKH HAEDALIIAR ADFIFKEWKK 720
LDKAKKVMEN QMFEEKQAES MPEIETEQFY KEIFITPHQI KHIKDFKDYK YSHRVDKKP 780
RELINDTLYS TRKDDGNTL IVNNNLGLYD KDNDKLKLLI NKSPEKLLMT HHDPQTQYKQL 840
KLIMEQYDNE KNPLYQYEE TGNYLTQYKSK DNDGPVIKKY KYYGNKLNAH LDITDDYPPNS 900
RNKVKVLSLX PYRFDVYLDN GVYKFVTVKN LDVVKENNY EVNSKCYEEA KKLKKISNQA 960
EFIASFYNND LIKINGELYR VIGVNNDLLN RIEVNMDIT YREYLENMND KRPPRIIKTI 1020
ASKTQSICKY STIDLGNLYE VKSKKHPQII KKG 1053

```

```

SEQ ID NO: 28          moltype = AA length = 1368
FEATURE             Location/Qualifiers
source              1..1368
mol_type = protein
organism = synthetic construct
SEQUENCE: 28
MDKYSIGLD IGTNSVGWAV ITDEYKVPSK KFKVLGNTRD HSIKKNLIGA LLFDSGETA 60
ATRLKRTARR RYTRRKNRIC YLQEFSNEM AKVDDSFHRR LEESFLVEED KKHERHPIFG 120
NIVDEVAYHE KYPTIYHLRK KLVSTDKD LRILYLALAH MIKFGRHFLLI EGDLNPNDNSD 180
VDKLFIFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN 240
LIALSLGLTP NFKSNFDLAE DAKLQLSKDT YDDDLDNLLA QIGDQYADLF LAAKNLSDAI 300
LILSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLKALVR QQLPEKYKEI FFDQSKNGYA 360
GYIDGGASQE EFYKFIKIPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH 420
AIIIRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE 480
VVDKGASAQS FIERMTNFDK NLPNEKVLPK HSLLYEYFTV YNELTKVKYV TEGMRKP AFL 540
SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLKI 600
IKDKDFLDNE ENEDILEDIV LTTLTFEDRE MIEERLKTYA HLFDDKVMKQ LKRRRTGNG 660
RLSRKLINGI RDQSGKTTIL DFLKSDGFAN RNFMQLIHDD SLTFKEDIQK AQVSGQGDSDL 720
HEHIANLAGS PAIKKGILQI KVVDDELVKV MGRHKPENIV IEMARENQTT QKGOKNSRER 780
MKRIEEGIKE LGSQILKEHP VENTQLQNEK LYLYYLQNRR DMVYDQELDI NRSLSDYDVHD 840
IVPQSFLKDD SIDNKVLTRS DKNRGKSDNV PSEEVVKKMK NYWRQLLNAK LITQRKFNDNL 900
TKAERGGLSE LDKAGFIKRQ LVETRQITKH VAQILDSSRMN TKYDENDKLI REVKVITLKS 960
KLVSDFRKDF QFYKREINN YHHAHDAYLN AVVGTALIKK YPKLESEFVY GDYKVYDVRK 1020
MIAKSEQEIG KATAKYFFYS NIMNNFKTEI TLANGEIRRK PLIETNGETG EIVWDKGRDF 1080

```

-continued

ATVRKVLSMP	QVNIVKKTEV	QTGGFSKESI	LPKRNSDKLI	ARKKDWDPKK	YGGFDSPVVA	1140
YSVLVVAKVE	KGKSKKLKV	KELLGITIME	RSSFEKNPID	FLEAKGYKEV	KKDLI1KLPK	1200
YSLFELENGR	KRMLASAGEL	QKGNEALALPS	KVYNFLYLAS	HYEKLKGSPE	DNEQKQLFVE	1260
QHKHYLDEII	EQISEFSKRV	ILADANLDKV	LSAYNKHDRK	PIREQAENII	HLFTLTNLGA	1320
PAAFKYFDTT	IDRKRYTSTK	EVLDATLHQ	SITGLYETRI	DLSQLGGD		1368

SEQ ID NO: 29	moltype = AA	length = 1368				
FEATURE	Location/Qualifiers					
source	1..1368					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 29						
MDKKYSIGLA	IGTNSVGWAV	ITDEYKVPSK	KFKVLGNTDR	HSIKKNLIGA	LLFDGETAE	60
ATRLKRTARR	RYTRRKNRIC	YLQEIFSNEM	AKVDDDSFPHR	LEESFLVEED	KKHERHPIFG	120
NIVDEVAYHE	KYPTIYHLRK	KLVDSTDKA	LRLIYLALAH	MIKFRGHFL	EGDLNPDNSD	180
VDKLFIQLVQ	TYNQLFEENP	INASGVDAKA	IILSARLSKSR	RLENLIAQLP	GEKKNGLFGN	240
LIALSLGLTP	NFKSNFDLAE	DAKQLLSKDT	YDDLDNLLA	QIGDQYADLF	LAAKNLSDAI	300
LLSIDLRLVT	EITKAPLSAS	MIKRYDEHHQ	DLTLLKALVR	QOLPEKYKEI	FFDQSKNGYA	360
GYIDGGASQE	EFYKFKIPIL	EKMDGTTELL	KQRTFDNGSI	PHQIHLGELH	420	
AIRRRQEDFY	PFLKDNRKE	EKILTFRIPY	YVGPLARNS	RFAWMTRKSE	ETITPNFEE	480
VVDKGASAQS	PIERMNTFDK	NLPNEKVLPK	HSLLYEYFTV	YNELTKVKYV	TEGMRKP AFL	540
SGBQKKAIVD	LLFKTNRKV	VKQLKEDYFPK	KIECFDSVEI	SGVEDRPNAS	LGTYHDLIKI	600
IKDKDFLDNE	ENEDILEDIV	LTLLTKDRE	MIEERLKTYA	HLFDDKVMKQ	LKRRRTGWG	660
RLSRKLINGI	RDKQSGKTTIL	DFLKSDFGFAN	RNPMQLIHDD	SLTFKEDIQK	AQVSGGGDSL	720
HEHIANLAGS	PAIKKGILQT	VKVVDELVKV	MGRHKPENIV	IEMARENQTT	QKGQKNSRER	780
MKRIEEGIKE	LGSQILKEH	VENTOLQNEK	LYLYYLNQGR	DMYVDQELDI	NRLSDYDVDA	840
IVPQSFLKDD	SIDNKVLTRS	DKNRGKSDNV	PSEEVVKMKM	NYWRQLLNAK	LITQRKF DNL	900
TKAERGGLSE	LDDKAGPIKRQ	LVTERRQITKH	VAJILDLSRM	TKYDENDKLI	REVKVITLKS	960
KLVSDFRKDF	QFYKVREINN	YHHAHDAYLN	AVVGTALIKK	YPKLESEFVY	GDYKVDVRK	1020
MIAKSEQEIG	KATAKYFTEI	TLANGERIRKR	PLIETNGETG	EIVWDKGDRF	1080	
ATVRKVLSMP	QVNIVKKTEV	QTGGFSKESI	LPKRNSDKLI	ARKKDWDPKK	YGGFDSPVVA	1140
YSVLVVAKVE	KGKSKKLKV	KELLGITIME	RSSFEKNPID	FLEAKGYKEV	KKDLI1KLPK	1200
YSLFELENGR	KRMLASAGEL	QKGNEALALPS	KVYNFLYLAS	HYEKLKGSPE	DNEQKQLFVE	1260
QHKHYLDEII	EQISEFSKRV	ILADANLDKV	LSAYNKHDRK	PIREQAENII	HLFTLTNLGA	1320
PAAFKYFDTT	IDRKRYTSTK	EVLDATLHQ	SITGLYETRI	DLSQLGGD		1368

SEQ ID NO: 30	moltype = DNA	length = 3159				
FEATURE	Location/Qualifiers					
source	1..3159					
	mol_type = other DNA					
	organism = synthetic construct					
SEQUENCE: 30						
atgaaaagga	actacattct	ggggctggcc	atcgaggatta	caagcgtggg	gtatggatt	60
attgactatg	aaacaaggga	cgtgtatcgac	gcaggcgtca	gactgttcaa	ggaggccaa	120
gttgaaaaca	atgaggggacg	gagaagcaag	aggggagcca	ggcgccctgaa	acgcaggaga	180
aggcacagaa	tccagagggt	gaagaaactg	ctgttcgatt	acaacctgct	gaccgaccat	240
tctgagctga	gttggatata	tcctttatgg	gcccagggtt	aaggccctgg	tccaaactgt	300
tcagaggaga	agttttccgc	agctctgtct	cacttgcata	agcgcggagg	agtgcataa	360
gtcaatgagg	tggaaagagga	cacccggcaac	gagctgtct	caaaggaaaca	gatctcacgc	420
aatagcaaag	ctcttggaaaga	gaagtatgtc	gcagagctgc	agctggaaacg	gctgaagaaa	480
gtatggcagg	tgagagggtc	attaatatgg	ttaaagacaa	cgactactgt	caaagaagcc	540
aagcagctgc	tggaaatgtca	gaaggcttac	caccagctgg	atcagagctt	catcgataact	600
tatatcgacc	tgcttggagac	tcggagaaacc	tactatgagg	gaccaggaga	aggggacccc	660
ttcggatgga	aagacatcaa	ggaatggta	gagatgtca	tgggacatgt	cacctat	720
ccagaagagc	tgagaagacgt	caatgtacgt	tataacgcacg	atctgtacaa	cgccctgaaat	780
gacctgtaca	accttgttcat	caccggat	gaaaacgcga	aactggaaata	ctatggaaag	840
ttccatcatc	tccaaaaacgt	gtttaagcag	aaagaaaaacg	stacatgtaa	acagattgt	900
aaggagatcc	tggtaacacgt	agaggacatc	aagggttaccc	gggtgacaag	cactggaaaa	960
ccagagttca	ccaatgtcaa	agtgtatcc	gatattaaagg	acatcacagc	acggaaagaa	1020
atcatttggaa	acgcggcaact	gctggatcat	attgtactaa	tcctgtactat	ctaccagagc	1080
tccggggaca	tccaggaaaga	gtctgttcat	cttgcacggc	agctgttaccc	ggaagagatc	1140
gaacagatata	gtatctgtaa	gggggtacacc	ggaacacacaa	acctgtccct	gaaagctatc	1200
aatctgttcc	tggatgtatc	gttgcatacata	aacgacaatc	agattgtcaat	ctttaaccgg	1260
ctgttgcgtt	tcccaaaaagg	gggtggacccgt	agtgcggcaga	aaggatcc	aaccacactg	1320
gttggacgtt	tcattctgtc	acccgtggtc	aaggccggat	tcattccatg	catcaaaatg	1380
atcaacgcoca	tcatacaagaa	gtacggccgt	cccaatgtca	tcattatgtca	gttgcgttag	1440
gagaagaaca	gcaaggacgc	acagaagatg	atcaatgtca	tgcagaaacg	aaaccggcag	1500
accaatgaac	gattatccga	actaccggga	aaggagaacgc	aaagtactgt	1560	
attggaaaaaa	tcaagctgc	cgatgtacgt	gagggaaaatg	gtctgtatcc	tctggaggcc	1620
atccccctgg	aggacactgt	gaacaatccaa	ttcaactactgc	aggtcgatca	tattatcccc	1680
agaagcgtgt	ccttcgacaa	ttccctttaac	aacaagggtc	tgttgcacaa	ggaagagaaac	1740
tctaaaaagg	gcaataggac	tcctttccag	tacctgttca	gttcagatc	caagatctct	1800
tacgaaaacct	taaaaagca	cattctgtat	ctggccaaag	gaaaggcccg	catcgcaag	1860
accaaaaaagg	agtacccgt	ggaagagccg	gacatcaaca	gattctccgt	ccagaaggat	1920
tttattaacc	ggaatctgtt	ggacacaaga	tacgttactc	gcccgttat	gaatctgt	1980
cgatccattt	tccgggtgaa	caatctggat	gtgaaagtca	agttccatca	cgccgggttc	2040

-continued

acatcttttc	tgaggcgc当地	atggaagttt	aaaaaggagc	gcaacaagg	gtacaagcac	2100
catggc当地	atgc当地tgc当地	tatcgcaaaat	gccgacttca	tctttaagga	gtggaaaaaag	2160
ctggac当地	ccaagaaatg	gatggagaad	cagatgttc当地	aagagaagca	ggccgaaatct	2220
atgccc当地	tc当地gac当地	acaggagta	aaggagattt	tcatcactcc	tcaccagatc	2280
aagcatatca	aggatttca	ggactacaag	tactctcacc	gggtggataa	aaagcccaac	2340
agagagctga	tcaatgacac	cctgtatagt	acaagaaaatg	acgataaagg	gaataccctg	2400
attgtgaaca	atctgaacgg	actgtacgac	aaagataatg	aaagctgatc	2460	
aacaaaatgc	ccgagaaatgc	gctgtatgc当地	caccatgatc	ctcagacata	tcagaaactg	2520
aagctgatta	tggagcagta	cggc当地cag	aagaaccac	tgtataagta	ctatgaagag	2580
actgggact	actctgaccaa	gtatgcaaaat	aaggataatg	gccccgtgat	caagaagatc	2640
aagtaactatg	ggaacaatgc	gaatgccc当地	ccctaaacatg	cagacgatc	ccctaaacatg	2700
cgc当地aca	agggggggg	tggtaaagct	gtcactgaag	ccatacagat	tcgatgtcta	2760
ggcgtgtata	aattttgtgac	tgtcaagaat	ctggatgtca	tcaaaaagga	gaactactat	2820
gaagtgataa	gcaatgtcta	cgaagaggct	aaaagctgat	aaaagattag	caaccaggca	2880
gagttcatcg	cctcttttta	caacaacgc当地	ctgatataa	tcaatggc当地	actgtatagg	2940
gtcatcg	tgaacaatgc	tctgtc当地	cgccat	tgaaatgat	tgatcatc	3000
taccgagat	atctggaaaa	catgaatgat	aaggcccc	ctcgatattat	caaacaattt	3060
gcctctaaga	ctcagatgat	caaaaatgc	tcaaccgaca	ttctggaaa	cctgtatgag	3120
gtgaagagca	aaaagcacc	tcagattatc	aaaaggggc			3159

SEQ ID NO: 31	moltype = DNA	length = 3159					
FEATURE	Location/Qualifiers						
source	1..3159						
	mol_type = other DNA						
	organism = synthetic construct						
SEQUENCE: 31							
atgaaaagga	actacattct	ggggotggac	atcgggattta	caagcgtgg	gtatgggatt	60	
attgactatg	aaacaaggga	cgtgtatgc当地	gcaggcgtca	gactgttcaa	ggaggccaa	120	
gtggaaaaaca	atgagggacg	gagaagcaag	agggggagca	ggcgc当地gaa	acgacggaga	180	
aggc当地caaa	tccagagggt	gaagaaactg	ctgttgc当地	acaacctgt	gaccgaccat	240	
tctgagatg	ta	tcccttatgaa	gccc当地gtgaa	aaggcctgag	tc当地aaactg	300	
tc当地gaggaa	agttttccgc	agctgtctg	caccctggta	agcgc当地gg	agtgc当地aac	360	
gtcaatgagg	tggaqaaggaa	cacccggca	gagctgtcta	caaaggaaaca	gatctc当地cgc	420	
aatagcaaa	ctctggaa	gaagtatgtc	gcagagctg	actgtggaa	gctgaagaaa	480	
gatggc当地	tgaggggg	gat	aatatatagg	tcaagacaa	gcaactactg	caaqaagcc	540
aaggc当地tc	tgaaatgtca	gagggttac	caccatgtg	atcagatgtt	catcgatata	600	
tatatcgacc	tgctggagac	tc当地gagaacc	tactatgagg	gaccaggaga	aggggacccc	660	
ttcggatgga	aaagacatcaa	ggaatggta	gagatgtga	ttgggacat	gac	720	
ccagaagagc	tgagggcgt	caatgtacgt	tataacgc当地	atctgtacaa	cgccctgaa	780	
gaccgtgaa	acctgttca	caccatggat	gaaaacgaga	aactgttca	ctatgagaag	840	
ttccatgtca	tccggaaaatgt	gtttaagcag	aaagaaaatgc	ctacactgaa	acagatgtct	900	
aaggagatcc	ttgtcaacga	agaggacatc	aagggttacc	gggtgaca	aaactggaaaa	960	
ccagagttca	ccatctgaa	agtgtatc	gatattaa	ggatcaca	acggaaagaa	1020	
atcattgaga	acgc当地caact	gtcggatc	attgtca	tc当地tactat	acccatggac	1080	
ttccgaggaca	tccaggaa	gctactaa	ctgaa	acag	ccgatc	1140	
gaacagatta	gtatctgaa	ggggatcacc	ggaacaca	acctgttcc	gaaagctatc	1200	
aatctgat	ttgtatgtat	gttgc当地ata	aa	gacatc	atgttcaat	1260	
ctgaaatgtt	ttccaaaat	ggggatc	actgttca	aa	accacactg	1320	
gttgc当地	tc当地tctgtc	acccgttgc	aa	gccc当地gt	tc当地tccat	1380	
atcaacgc当地	tc当地caagaa	gtacggc当地	ccc当地atgata	tc当地tattat	gctggctt	1440	
gagaagaa	gcaaggacgc	acagaaatgt	atcaatgata	tgca	gagaaac	1500	
accaatgc当地	gat	tttgc当地ata	aa	accatgggg	aaatgtatc	1560	
attgaaaaaa	tcaatgtca	cgatgtc	gagg	ggaaaatgt	tctgtat	1620	
atccccctgg	aggac	ctgt	gat	ttcaactac	agg	1680	
agaagcg	gtt	ccat	ttcaactac	agg	tc当地tcatca	1740	
tctaaaatgg	gcaatgg	tc当地tcttca	ta	cttgc当地at	tttgc当地at	1800	
ta	ccat	tttgc当地at	ctggccaa	gaaagg	ggcc	1860	
accaaaaagg	agtatctgt	ggaagacgc当地	gacatcaaca	gat	tccctgt	1920	
tttattaa	cc	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	1980	
cgatcc	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	2040	
acatctttc	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	2100	
catggc当地	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	2160	
ctggac当地	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	2220	
atgccc当地	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	2280	
aagcatatca	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	2340	
agagagctga	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	2400	
attgtgaaca	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	2460	
aacaaaatgc	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	2520	
aaggc当地	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	2580	
actgggact	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	2640	
aagtactatg	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	2700	
cgcaacaagg	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	2760	
ggcgtgtata	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	2820	
gaagtgataa	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	2880	
gagttcatcg	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	2940	
gtcatcg	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	tttgc当地at	3000	

-continued

taccgagagt atctggaaaa catgaatgat aagcgccccc ctcgaattat caaaacaatt 3060
gcctctaaga ctcagatcat caaaaagtagc tcaaccgaca ttctggaaa cctgtatgag 3120
gtgaagagca aaaagcaccc tcagattatc aaaaaggcc 3159
SEQ ID NO: 32 moltype = DNA length = 4107
FEATURE Location/Qualifiers
source 1..4107
mol_type = other DNA
organism = synthetic construct
 SEQUENCE: 32
atggataaaa agtacagcat cgggtggac atcggtacaa actcagtgg gtggccgtg 60
attacggacg agtacaaggt accctccaaa aaatttaaag tgctggtaa cacggacaga 120
cactctataa agaaaaatct tattggagcc ttgctgttc actcaggcga gacagccaa 180
gccacaagggt tgaaggccgc cgccaggagg cggtataccca ggagaaaaaa ccgcataatgc 240
tacctgcaag aaatcttcag taacgagat gcaaagggtt acgtatgttt ttccatcgc 300
ctggaagaaat ctttttgttgc tgaggaaag aqaagcaccg aacggcaccc catcttggc 360
aatattgtcg acgaagtggc atatcacgaa aagtacccga ctatctacca cctcaggaaag 420
aagctgtggg actcttaccgc taaggccgc ctcagactta ttatgtgc actcgcccac 480
atgattaaat tttagggaca ttatgtgc gaggccgacc tgaaccggga caacagtgtac 540
gtcgataaagc tggtcatcca acttgcacg acctataac aactgttgcg agaaaaaccc 600
ataaaatgttt caggagtgcg cgctaaagca atcctgtccg cgccctctc aaaatctaga 660
agacttgaga atctgttgc tcagtgcgc gggaaaaaaga aaaaatggatt gtttggcaac 720
ctgatcgcccc tcagtcgtgg actgtgcgc aatttcggaa gtaacttcga ccttggccaa 780
gacgctaaaggc tcacatgttc caaggacaca tacgtacgc acctcgacaa tctgtggcc 840
cagattgggg atcgtacgc cgatctttt ttggcagecaa agaacctgtc cgacgcacate 900
ctgttgagcg atatcttgcg agtacacacca gaaattacta aagcaccctt tagccatct 960
atgatcaacgc ggtacgcacg gcatatcatc gatctgcaccc tgctgaaggc tcttgcagg 1020
caacagtcgc cggaaaaata caaggaaatc ttctttgcg agacaaaaaa cggctacgct 1080
ggctatatacg atgggggggc cagtcaggag gaattctata aattcatcaa gccattctc 1140
gagaaaaatgg acggcacaga ggagtgtctg gtcaaaacttta acagggagga cctgtgcgg 1200
aacgcacggg cttttgcacca cgggtctatc cccaccacca ttcatctgg cgaaactgcac 1260
gcaatcttgc ggaggccgg gatTTTCTTA ctttttttta aagataaccc cgagaaaaata 1320
gaaaagatttcc ttacatttcg gatcccgatc tacgtgggac ctctcgcccg gggcaattca 1380
cggtttgcct ggtatgacaag gaagtccagag gagactatta caccttgaa cttcgaagaa 1440
gttgcgtggaca agggtgcacat tggccatgtt ttcatcgacg ggtatgacaaa ttgtgacaaag 1500
aaccttcccta atgagaatgg gtcggccaaa cattctctgc tctacgatca ctttaccgtc 1560
tacaatgaac tgactaaatgc caagtacgtc accgggggaa tgaggaaaggc ggcattctt 1620
agtggagaac agaagaaggc gattgttagac ctgttgcgtca agaccaacag gaagggtact 1680
gtgaagacac ttaaaaagaatc ctactttaa gaaatcgat gtttgcacag ttttgcgg 1740
tcagggggttgc aagggtgcacat tggccatgtt ttcatcgacg ggtatgacaaa ttgtgacaaag 1800
ataaaggaca aagacttcct ggacaaacgaa gaaaatgggg atattctcgaa agacatcgac 1860
ctcacccttgc cctgttgcg agacaggggaa atgatagaag agcgttgcgaa aacctatgcc 1920
caaccttgc acgtatggatc tataaaggccatc ctgtggccgc ggagatcacac aggatgggg 1980
agattgtcaa gggatggatc caatggaaatc agggataaaac agatggccaa gaccatactg 2040
gatttccttca aatctgtatgg ctgcggccatc aggaacttcgatc tgcactgtatc tccatcgatc 2100
tctcttaccc tcaaggagga cattcaaaatgc gtcaggatgc gggggcagg agactccctt 2160
catgaacaca tcggcaatgtt ggcaggatcc cccgttataaaaaggccat ctttcaact 2220
gtcaagggtgg tggatggatc ggtcaaggatc atggggcagac ataaggccaga aaaaatgg 2280
atcgagatgg cccgcggaaaa ccagaccaca cagaaggccgc agaaaaatag tagagggccg 2340
atgaagagga tcgaggaggcatc catcaaaatgc ctggatctc agattctcaaa agaacacccc 2400
gtagaaaaacaa cacatcgatc gaacggaaaaaa ttgtactgtt actatctcgaa gaacggccaga 2460
gacatgtacg tcgaccaaga acttgcgtt aatagactgtt cgcactatgc cgttagacat 2520
atctgtccccc agtcttccttgc gaaaggccatc tccattgtata acaaaggctt gacaagaac 2580
gacaagaaca ggggtaaaatgc tgataatgtg ctcaggccggagg aggtgggtggaa aaaaatgg 2640
aactacttgcg gacatgtgtc taatgtccatc aacggaaatgc cgataatctgc 2700
acgaaaggccatc agagagggtgg ctgtgttgcgatc ttggacaagg cagggtttat taagccggcag 2760
ctgggtggaaa ctaggccatc cacaaggccatc gtggccggaa ttttggacag ccggatgaaac 2820
acaaaataatgc acgaaaatgc taaaactgtata cggaggccatc aagtttgcgtca gctggaaaac 2880
aagctgtgttgc cggatggatc gaaaggccatc ctttttttgcgatc aatgtccggatc gatataa 2940
taccatcatgc ctcacatgc gtcacatgc gtcgttgcgtcc ggaccgcctt gatataa 3000
tacccaaatgc tggatgttgc gtcgttgcgtcc ggaccgcctt gatataa 3060
atgatagccatc agtccgagca ggagatggaa aaggccacatc ctaatgttgcgtcc 3120
aacatcatgc attttttgcgatc gacggaaatgc accctggccatc acggggatcatc gacaaaggcc 3180
cccccattatgc agacaaaatgc tggatgttgc gaaatgttgcgtcc gggatgggg cggggatcc 3240
gctactgtgc ggaagggtgtc ggtatgtccatc ctttttttgcgatc aatgtccggatc gatataa 3300
caacccggatc gatTTTCTTA caaggccatc aagaaaggccatc tcatcattaa actggccaaa 3600
tactcttcctc tccatcgatc aatgtccggatc aagaaaggccatc tcatcattaa actggccaaa 3660
caaaaaggccatc acgacatgtgc tctggccatc aatgtccggatc aatgtccggatc tcatcattaa 3720
cactatcatgc aatgtccggatc gacggaaatgc gataacccatc acggggatcatc gacaaaggcc 3780
cacccatcgatc aatgtccggatc gacggaaatgc gataacccatc acggggatcatc gacaaaggcc 3840
atccttggccatc atgatgttgc ggtatgttgcgtcc gggatgggg cggggatcc 3900
cctttagggatc aacaaggccatc gatataa 3960

-continued

```

cccgccgcct tcaaatactt tgatcggact atcgaccggaa aacgggtatac cagttacccaa 4020
gagggttcctt atggccacctt atccacccag tcaatttactg gcctgtacga aacacggatc 4080
gaccccttcctt aactggggccg cgacttag 4107

```

```
SEQ ID NO: 33          moltype = DNA  length = 3158
FEATURE                Location/Qualifiers
source                 1..3158
                      mol_type = other DNA
                      organism = synthetic construct
```

SEQUENCE: 33
 atgaaaagga actacattt gggctggac atcgggatta caagcgtgg gtatggatt 60
 attgactat aaacaaggaa cgtatgcac gcaggcgtca gactgtcaa ggaggccaa 120
 gtggaaaaca atgaggggacg gagaacaaag aggggagcca ggcgcctgaa acgacggaga 180
 aggcacagaa tccaggggt gaagaaactg ctgttcatt acaactctgt gaccgaccat 240
 tctgactgta gtggaaataa tccttataa gccagggtga aaggcctgaa tcagaagctg 300
 tcagaggaaag agtttccgc agctctgtc cacctgtca agcgcggg agtgcataac 360
 tcatacgagg tggaaagaga cacggccaa gagetgtcta caaagaaca gatccacgc 420
 aatacgaaag ctcttggaaa gaagtatgtc gcagagctgc agcttggaaacg gcttggaaaa 480
 gatggcgagg tgaggggtc aattaaatgg ttcaagacaa gcgactactg caaagaagcc 540
 aaggcgtgc tggaaatggc gaaggcttc caccagggtt atcagaggctt catcgatact 600
 tatatacgacc tcgtggagac tcggaaatcc tactatggg gaccggaga aggaggcccc 660
 ttccggatggaa aagacatcaa ggaatggtag gagatgtca tgggacattt caccatttt 720
 ccagaagacg tgagaagcgt caagtacttataacgcg atctgttacaa cgcctgtaa 780
 gacctgtacca accttggtcat caccagggtt gaaaacgaga aacttggata ctatggaaag 840
 ttccagatca tcggaaaatgtttaaagc aaaaaaaaaaagc ctacatgtaa acatggatgt 900
 aaggagatcc tggtaacacg agaggacatc aagggttacc ggggtgacaag cactggaaa 960
 ccagagttca ccaatctgaa agtgttatcac gatatttaagg acatcacgc acggaaagaa 1020
 atatcttgaa acggcgaatc gctggatcatcattgttca gatctgtactt ctatccagacg 1080
 tccggaggaca tccggaaatggcacttgc aatcgacttgc aatcgacttgc ggaagatgc 1140
 gaacacattttaa gtaatctgaa ggggttacacc ggaacacaca acctgtccctt gaaatctt 1200
 aatctgttcc tggatggatgt gtggatcatca aacgacaaatc agatggcaatc cttaaacgg 1260
 ctgaaatctgg tcccaaaaaaa ggtggacactg agtcggcaga aagagatcc aaccacactg 1320
 gtggacatgttccatctgtc acccgtgttca aacggggact tcatccagag catcaatgg 1380
 atcaacccca tcatacgaaa gtacggctgttcccaatgtatcattatgtca gttggcttag 1440
 gagaagaaca gcaaggacgc acagaatgttcaatgtgatc tgcaagaaacg aaacccggc 1500
 accaaatgtca gcatgttcaaga gattatccga actatggggaa aagaaacgc aatgttccctg 1560
 atggaaaaaa tcaatctgca ctagatgtcag gaggggaaatgttccatctgttccatctgg 1620
 tcccccttggaa ggacctgttc aacaatccat tcaactatcg ggtcgatcat attatcccc 1680
 gaagcgtgtc ttccgtacaat tccctttaaca acaagggttgc ggtcaagcga gaaatggaa 1740
 cttaaaaaaggaa caatgttgcacttgcgttccatgttccatctgttccatctgttccatctgg 1800
 acggaaatctt taaaatggcacttctgttcaatgttccatctgttccatctgttccatctgg 1860
 ccaaaaaagga gtacccgttc gaaagacgggg acatcaacag atctccgtc cagaaggatt 1920
 ttatataaccg gaatctgggtt gacacaatgttccatctgttccatctgttccatctgg 1980
 gatccatctt cccgggttcaaaatggatgt tgaaatgttccatctgttccatctgttccatctgg 2040
 catcttcttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgg 2100
 atgcccggaaa tgctgttccatctgttccatctgttccatctgttccatctgttccatctgg 2160
 tggacaaatggc caagaaatgttccatctgttccatctgttccatctgttccatctgttccatctgg 2220
 tggccggaaatggc cgagacggaaa cagggttacca gggatgttccatctgttccatctgttccatctgg 2280
 agcatatccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgg 2340
 gggatgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgg 2400
 ttgttggaaatggc tctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgg 2460
 acaaaaaatgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgg 2520
 cgagaaatgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgg 2580
 agctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgg 2640
 tggggacttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgg 2700
 agtactatgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgg 2760
 gcaacaaatgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgg 2820
 ggggttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgg 2880
 aatgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgg 2940
 agttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgg 3000
 tcatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgg 3060
 accggatgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgg 3120
 ctcttcaatgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgg 3180
 tgaagagacaa aaagccacctt cagattatca aaaaggggc
 tcaatgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgttccatctgg 3158

SEQ ID NO: 34 moltype = DNA length = 3159
FEATURE Location/Qualifiers
source 1..3159
mol_type = other DNA
organism = synthetic construct

```

SEQUENCE: 34
          Organism = Synthetic construct

atgaaggcgga actacatcct gggcctggac atccggcatca ccagcgtggg ctacggcatc 60
atcgactacg agacacggga cgtgtatcgat gcccggctgc ggctgttcaa agaggccaa 120
gtggaaaaaa acggaggcgag gccggacaaag agaggccca gaaggctgaa gccgcggagg 180
ccggcatagaaa tccagagact gaagaaggct ctgttcgact aacaactgtg gaccgaccad 240
acgcagctga ggccgtatcaa cccctacagag gccagagtga agggcctgag ccagaagctg 300
acgcgaggaaag agttctctgc cgccccgtctg caccctggcca agagaagagg cgtgcacaac 360

```

-continued

gtgaacgagg	tggaagagga	cacggcaac	gagctgtcca	ccaaagagca	gatcagccgg	420			
aacagcaagg	ccctggaga	gaaatacg	tgccgaactgc	agctggaa	cg gctgaagaaa	480			
gacggcgaag	tgccccggc	catcaacaga	ttaaagacca	gcgactacgt	gaaagaagcc	540			
aaacagctgc	tgaagg	tgca	gaaggcctac	caccagctt	accagagctt	600			
tacatcgacc	tgctggaa	ac	ccggcggacc	tactatgagg	gac	ttcgccg	ggcagcccc	660	
ttcgccg	ttcg	gg	aggacatcaa	agaatggta	gagatgtca	tgccggccat	cacc	acttc	720
cccgaggaa	ac	tgccgg	ggcgt	gaagtac	gca	acccgc	ac	cttca	780
gac	cttca	ac	atctcg	tgat	ccac	ggaa	ac	ttacq	840
ttcc	ccat	ca	tc	taa	ccac	ggaa	ac	ttacq	900
aa	aa	aa	aa	aa	aa	aa	aa	aa	960
cccgagg	at	cc	cc	cc	cc	cc	cc	cc	1020
attattt	gaga	ac	cc	cc	cc	cc	cc	cc	1080
agc	gagg	ac	cc	cc	cc	cc	cc	cc	1140
gagg	ccat	cc	cc	cc	cc	cc	cc	cc	1200
aac	ctgt	at	cc	cc	cc	cc	cc	cc	1260
ct	taa	gg	gg	gg	gg	gg	gg	gg	1320
gtgg	gac	act	cc	cc	cc	cc	cc	cc	1380
atca	ac	cc	cc	cc	cc	cc	cc	cc	1440
gaga	aa	aa	aa	aa	aa	aa	aa	aa	1500
acca	aa	aa	aa	aa	aa	aa	aa	aa	1560
atc	gaga	aa	aa	aa	aa	aa	aa	aa	1620
atccc	ctt	gg	gg	gg	gg	gg	gg	gg	1680
aga	agg	ctt	cc	cc	cc	cc	cc	cc	1740
agca	aa	gg	cc	cc	cc	cc	cc	cc	1800
ta	cgaa	ac	cc	cc	cc	cc	cc	cc	1860
acca	aa	aa	aa	aa	aa	aa	aa	aa	1920
tt	catca	aa	cc	cc	cc	cc	cc	cc	1980
cg	gag	ct	cc	cc	cc	cc	cc	cc	2040
acc	ag	ttt	cc	cc	cc	cc	cc	cc	2100
cac	cc	cc	cc	cc	cc	cc	cc	cc	2160
ct	gg	ca	aa	aa	aa	aa	aa	aa	2220
at	ccc	cc	cc	cc	cc	cc	cc	cc	2280
a	ggc	at	cc	cc	cc	cc	cc	cc	2340
ag	ag	ct	cc	cc	cc	cc	cc	cc	2400
at	cgt	aa	cc	cc	cc	cc	cc	cc	2460
a	aa	aa	aa	aa	aa	aa	aa	aa	2520
aa	ctgtt	at	cc	cc	cc	cc	cc	cc	2580
acc	gg	aa	cc	cc	cc	cc	cc	cc	2640
a	gt	cc	cc	cc	cc	cc	cc	cc	2700
aga	aa	aa	aa	aa	aa	aa	aa	aa	2760
gg	cg	tg	cc	cc	cc	cc	cc	cc	2820
ga	gt	ta	cc	cc	cc	cc	cc	cc	2880
g	at	tt	cc	cc	cc	cc	cc	cc	2940
gt	tg	cc	cc	cc	cc	cc	cc	cc	3000
tac	cc	cc	cc	cc	cc	cc	cc	cc	3060
gc	ct	cc	cc	cc	cc	cc	cc	cc	3120
gt	aa	at	cc	cc	cc	cc	cc	cc	3159

```

SEQ ID NO: 35          moltype = DNA  length = 3159
FEATURE           Location/Qualifiers
source            1..3159
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 35
atgaaggcaca actacatcc cggactggac atcggcatta cctccgtgg atacggcatc 60
atcgattacg aaactaggga tttgtatcgac gctggagtca ggctgttcaa agaggcgaac 120
gtggagaaca acggggggcg ggcgtcaag agggggggcc gccggctgaa ggcgcgcgc 180
agacatagaa tccagcgctg gaagaagctg ctgttcgact acaaccttct gaccgaccac 240
tccgaacttt ccggcatcaa cccatatag gctagagtca agggattgtc cccaaagctg 300
tccgaggaaag atttccgcgc cgcgtgtcc cacctcgcaca agcgcagggg agtgcacaaat 360
gtgaacgaaat tccggaaaac gaggctgtca ccaaggagca gatcagccgg 420
aactccaagg ccctggaaaga gaaatacggt gccgaactgc aactggagcg gctgaagaaa 480
gacggagaag tgccggcgctc gatcaaccgc ttcaagacct cggaactacgt gaaggagggc 540
aaggcactcc tggaaatgtca aaaggccatca cccaaacttcc accagtccctt tatacgatcc 600
tacatcgatcc tgctcgagac tccggggact tactacgggg gtccaggggg gggctccca 660
tttgggttga aggatattaa ggaggctgtac gaaatgtca tggggacactg cacatcttc 720
cctgaggagc tgccggcgctg gaaatacgca tacaacgcgg acctgttacaa cgcgtgtac 780
gacctgaaca atctcgatcc caccggggac gagaacgaaa agctcgagta ttacgaaaag 840
ttccgatcc tggaaatgtca aaaggccatca cccaaacttcc accagtccctt tatacgatcc 900
aaggaaatcc tcgtgaacga aaggagacatc aagggtctatc gaggacatc aacggggaaag 960
ccggaggatca ccaatcttcc ggttccatcc gatcaacaaat cccatccgc cccggaaaggag 1020
atcatttggaa acgcggagct gttggaccatc atttgcgaaatc ttctgaccat cttacatcc 1080
tccgaggatcc tccggaaaga actcaccac ctcacacgcg aactggaccatc ggaggagata 1140
gacggaaatcc ccaacatgtca gggctacacc ggaactcatac acctgagctt gatggccatc 1200
aacttgcattcc tggacgatcc gttggaccatc aacgataacc agatcgatcc tttcaatccg 1260
ctgaagcttcc tcccaagaaat agtggaccatc tcacaacaaat aggatccct tactaccctt 1320

```

-continued

gtggacgatt	tcattctgtc	ccccgtggc	aagagaagct	tcatacagtc	aatcaaagtgc	1380
atcaatgoca	ttatcaagaa	atacggctcg	cccaacgcaca	ttatcatgta	gctcgcccg	1440
gagaagaact	cgaaggacgc	ccagaagatg	attaacgaaa	tgcagaagag	gaaccgcac	1500
actaacgcac	ggatcgaa	aatcatccgg	accaccggga	aggaaaaacgc	gaagtactcg	1560
atcgaaaaga	tcaagctcca	tgacatgcag	gaaggaaagt	gtctgtactc	gctggaggcc	1620
atcccgctgg	aggacttgc	gaacaacccc	ttaactacg	aagtggatca	tatcattccg	1680
aggagcgtgt	cattcgacaa	ttccctcaac	aacaagggtc	tgcgaa	ggggaaaac	1740
tgcagaaggg	gaaacccgac	gccgtccag	tacctgacca	gcagcactc	caagatttc	1800
tcgaaacct	tcaagaagca	catctcaac	ctggcaagg	ggaagggtcg	catctccaag	1860
accaagaagg	aatatctgtc	ggaagaaaqa	gacatcaaca	gattctcg	gcaaaaggac	1920
ttcatcaact	gcaacactgt	ggataacttgc	tagctactc	ggggtgtcg	gaacctctcg	1980
agaagctact	ttagagtga	caatctggac	gtgaaggc	agtgcattaa	cgagggttc	2040
acctccttcc	tgccgcgca	gttgaagttc	aagaaggac	ggaacaagg	ctacaagcac	2100
cacggccagg	acgcctctgt	cattggccaa	gccgacttc	tcttcaaga	atgaaagaaa	2160
cttgacaaagg	ctaagaagg	catggaaaac	catgttgc	aagaaaagca	ggccgagtc	2220
atcgcttgc	ttccatcaact	acaggagtc	aaggaaatct	tattatccgc	acacccatgc	2280
aaacacatca	aggatttca	ggattacaag	tactcacatc	gctgtggacaa	aaaggccgaa	2340
agggaaactcg	tcaacgcac	cctctactcg	acccggaa	atgacaagg	gaataccctc	2400
atcgtaaca	accttaacgg	cctgtaccc	aaggacaac	ataagctgaa	gaagctcatt	2460
aacaagtgc	ccggaaatgt	gtgtgtac	caccacgacc	ctcgactca	ccagaagctc	2520
aagctgtatca	tggagcgt	ttgggacgag	aaaacccgt	tgtacaagta	ctcagaagaa	2580
actggaaatt	atctgactaa	gtacttccaa	aaagataacg	gccccgtgt	taagaagatt	2640
aagtactacg	gcaacaagct	gaacgcctat	ctggacatca	cccgat	ccctaatcc	2700
cgcacaacaa	tcgtcaact	gagcttc	ccttaccgg	ttgtatgt	cttgcacaa	2760
ggagtgtaca	agtctgtac	tgtgaagaa	cttgacgt	tcaagaagga	gaactactac	2820
gaagtcaact	ccaaatgtca	cgaggaa	aagaatgtca	gaaatctc	gaaccaggcc	2880
gagttcattc	ctcccttcta	taaacaacg	ctgat	taaacaacg	actgtaccc	2940
gtcatttggc	tttgcacaa	tcctcttgc	cgcatcga	tgaatcat	cgacatcact	3000
tacccggaaat	acctggagaa	tatgaacac	aacgcggcc	ccggat	catat	3060
gcctcaaa	cccagtgc	caagaatgt	agcaccgaca	tcctggccaa	cctgtacgag	3120
gtcaaatcga	agaagcacc	ccagatcato	agaagg			3159

SEQ ID NO: 36	moltype = DNA	length = 3255				
FEATURE	Location/Qualifiers					
source	1..3255					
	mol_type = other DNA					
	organism = synthetic construct					
SEQUENCE: 36						
atggcccaaa	agaagaagcg	gaagggtcggt	atccacggag	tcccgacgc	caageggaa	60
tacatcctgg	gcctggacat	cgccatcacc	agcggtgg	acggcatcat	cgactacgag	120
acacgggac	tgatcgatgc	cggcggtcg	ctgttcaaa	aggccaa	ggaaaacaa	180
gagggcggc	ggagcaagag	aggcgccaga	aggctgaac	ggcgaggcg	gcata	240
caagagtg	agaagatgtc	gttcgactac	aacctgtcg	ccgaccacag	cgagctgag	300
ggcatcaacc	catacgaggc	cgagcttgc	ggcgttgc	agaagatgt	cgaggaagag	360
tttctgtcgc	coctgtgtca	cctggccaa	agaagaggcg	tgcacaacgt	gaacgagggt	420
gaagaggaca	ccggcaacga	gtgttccacc	agagagcaga	tcaagccgaa	cagcaaggcc	480
ctggaagaga	aataatgtcc	cgaatgtcg	ctggaacggc	tgaagaa	cgccgaaatgt	540
cggggcggc	caacatcgatt	caagaccatc	gactactgt	aaaagccaa	acagctgt	600
aagggtcaga	aggccatcca	ccatgttgc	cgatgttca	tcgacactca	catcgactc	660
ctggaaaccc	ggcggaccta	ctatgaggga	cctggcgagg	gcagcccc	cggtgttgc	720
gacatcaaa	aatgtgtacg	gatgtgtac	ggccacttc	cctacttccc	cgaggaaact	780
cggagctgc	agatcgctca	caacgcgc	ctgtacaac	ccttgcacaa	cttgcacaa	840
ctcgatgtca	ccaggggac	gaacggaa	ctgatatt	acgaaatgt	ccatgtatc	900
gagaacgtgt	tcaagcaga	gaagaagcc	accctgttgc	agatcgccaa	agaaaatctc	960
gtgaaacaa	aggatattaa	gggttacaca	gtgaccatc	ccggcaacggc	cgagttcacc	1020
aacctgttgc	tgttaccc	catcaaggac	attaccgc	ggaaagatgt	tatttgaaac	1080
cccgagctgc	tggatcagat	tgcaatgt	ctgaccatc	accagacgac	cgaggatc	1140
cagggaaac	tgaccaatct	gaactccgag	ctgacccagg	aagatgtca	gcagatctc	1200
aatctgttgc	gtataccgg	cacccacaa	ctgacccatc	aggccatcaa	cctgtatctc	1260
gacgagatgt	ggcacac	cgacaccc	atcgat	tatcgttatc	gtgatgtgt	1320
cccaaaatgg	tggatcgtc	ccacggaaa	gatgttccca	ccacccgttgc	ggacgacttc	1380
atccctgttgc	ccgttgtgt	gagaaggttc	atccatgtat	tcaaaatgt	caacgcctc	1440
atcaagaatgt	acggccgtcc	caacgcac	attatcgac	tggccgcga	gaagaactcc	1500
aaggacgc	aaaaatgtat	caacggat	caacggcg	accggcagac	caacgcgc	1560
atcgaggaaa	tcatccggac	cacccggaa	gaaacgcaca	agttatgt	cgagat	1620
aagctgtc	acatcgac	aggcaatgt	ctgtaccc	tggaccat	ccctcttgc	1680
gatctgtgt	acaacccctt	caactatgt	gtggaccaca	tcatccccc	aagctgttgc	1740
ttcgacaaaca	gttcaacaa	caagggtc	gtgatgt	aaagaaaac	caagaaggcc	1800
aacccggac	cattccat	cctgttgc	agatcagca	cgaaaccc	cgaaaccc	1860
aagaagcaca	tcctgtatc	ggccaa	aaaggcc	tcaagac	caagaagag	1920
tatctgttgc	aaaacgggca	catcaac	ttccctgtc	agaaatgt	catcaaccc	1980
acccgtgttgc	ataccagata	cgccacc	ggctgtat	acctgttgc	gagctacttc	2040
agagtgtac	acctggacgt	gaaatgt	tccatcaat	cgccgttcc	cagtttgc	2100
ccggcggaaat	ggaatgttac	gaaagagcgg	acaaggacca	cgccgaggac	2160	
ccctgtatc	ttgccaacgc	cgat	ttaaaatgt	ggaagaaact	ggacaaggcc	2220
aaaaaaatgt	tggaaaacca	gtatgttgc	gaaaggcagg	ccgagat	ccccggatc	2280

-continued

```

aaaaaccgagc aggagtacaa agagatcc tc atcacccccc accagatcaa gcacatta ag 2340
gacttcaagg actacaaga cagccacccg gtggacaaga agccttaatg agagctgatt 2400
aacgcacacc t gttactcc cccggaaggac gacaaggc acaccctgtat cgtgaaacata 2460
ctgaaggccg t gtagcaca ggacaaatgc aagctggaaa agtgcataa caagagcccc 2520
gaaaagtcg t gatgtacca ccacgacccc cagacatcc agaaactgaa gctgatatt 2580
gaacagtacg cgacgagaa gaatccctg tacaatgtact acggagaaac cgggaaactac 2640
ctgcaagaatg actccaaa ggacaaacggc cccgtatca aagaaggattaa gtatccggc 2700
aacaacta tgcggccatc gacatcacc gacgactacc ccaacagcag aaaacaggtc 2760
gtgaagctgt ccctgaaggc ctacatggc acgtgttacc tgacaaatgg cgtgtacaag 2820
ttcggtaccc tgaagaatct ggatgtatc aaaaaagaaa actactacga agtgaatagc 2880
aagtctatc aggaagctaa gagtcgttacg aagatcagca accaggccga gttagccatc 2940
tccttttata aacaacgtat gatcaatgc aacggcggc tgtagataggat gatccggctg 3000
aacaacgacc tgctgaaacctt gatcgaaatg aacatgtatcc acatcaccta cccggcgtac 3060
ctggaaaaca tgaacgacaa gaggcccccc agatcattt agacaatcgc ctccaaagacc 3120
cagagcattt agaagtatcag cacagacatt ctggccaaatc tgatgtaaatgaaatctaa 3180
aaggcccttc agatcatcaa aaaggccaaa aggccggcgg ccacggaaaaa ggccggccag 3240
qaaaaaaaaaaa qaaaq 3255

```

```
SEQ ID NO: 37          moltype = DNA    length = 3242
FEATURE              Location/Qualifiers
source               1..3242
                     mol_type = other DNA
                     organism = synthetic construct
```

-continued

SEQ ID NO: 38 moltype = DNA length = 3255
FEATURE Location/Qualifiers
source 1..3255
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 38

```
atggcccca agaagaagcg gaaggctcggt atccacggag tcccgccgcg caagcggAAC 60
tacatectgg gcctggacat cggcatcacc agcgtgggt acggcatcat cgactacgc 120
acacgggacg tgatcgatgc cggcgctcggt ctgttcaaaAG ggccaaacgt gaaaaaacac 180
gagggcggc ggagcaagag aggccocatc aggtcaagc ggcggaggcg gcataaaatc 240
cagagagtga agaagctgtt gttcgactac aacctgtctga ccgaccacag cgagctgago 300
ggcataccaccc cctacqgagc cagaqtgaag ggcctgaggc agaagctgtt cgaggaaag 360
ttctctggc ccctgtcga cctggccaaAG agaagggcg tcgcacaacgt gaacggatgt 420
aaagggacca cccgcaacca gctgtccacc aaagagccatc tcaggccggaa cggcaaggcc 480
ctggaaagaga aatacggtt cgaactgcag ctggaaeccc tgaagaaaga cggcgaagt 540
cggggcggc tcaacacattt caagacccg gactactgtt aagaaggccaa acatgtctg 600
aagggtcaga aggcttacca ccatgtcggtt cagatgttca tcgacactta catcgactt 660
ctggaaaccc gggccggactt ctatgaggga cctggccggg cgacggccctt cggctggaa 720
gacatcaaAG aatgggtacca gatgtctgtt ggccactgc cctacttccc cgaggaact 780
cggagcgttga agtacgccta caacccggccatc ctgttcaaccc ccctgttacatc cttgttacatc 840
ctcgatgttca ccaggggacca gaaacggatc ctggatattt acggaaatgtt ccagatcatc 900
gagaacgttga tcaacacattt caagacccg gactactgtt aagaaggccaa acatgtctt 960
gtgaacgttga aggatattttt gggctacaga gtgaccagca ccggcaagcc cgagtttcacc 1020
aacctgttaccc ttttggatgttca catcaaggac attaccggcc gggaaaatgtt tattttgttac 1080
ggccggatgttcc tggatgttca ctggatgttca ctggaccatctt accagggatc cgaggatcatc 1140
cggaggatgttcc tggatgttca gaaacggatc ctggatattt acggaaatgtt ccagatcatc 1200
gagaacgttga tcaacacattt caagacccg gactactgtt aagaaggccaa acatgtctt 1260
aatctgttaccc gtttggatgttca caccggacatc ctggactctgtt agggccatcaatc cttgtatctgtt 1320
gacgagctgtt gggcacccaa cggacacccatc atcgcttatct tcaacccgggtt gaagctgggt 1380
cccaaaacggatcc tggatgttcc cccggggatc gggatccccc ccacccctgtt ggacggactt 1440
atccctgttaccc cccgtgttca gaaacgttca ctggatattt acggaaatgtt ccacggccatc 1500
atcaacaaatgtt acggccgttcc caacggatcattt atccatgttca tggcccccgtt gaaqaacttcc 1560
aaggacgccc agaaaatgtt caacggatc gagaaggccg accggccatc caacggccatc 1620
atccggggatcc ttatccggatc caccggccaaAG gagaacggccatc agtactgtt cgagaagatc 1680
aagggtcacttcc acatgttaccc gggatgttca ctgttacatc tggaccatc ccctctggatc 1740
gtatgttca gaaacgttcc caatcatgttca gtttggatgttca tccatccatc acggctgttcc 1800
ttcgacaaatgtt gtttcaacccatc caaggtgttca ctggatgttca gtttggatgttca tccatccatc 1860
aaccggaccatc cattccggatc cctggatgttca gggatgttca acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 1920
aagaacgacatc ttctgttaccc tggccaaacggcc gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 1980
atctgttccatc aagaacggatcc catcaacccatc ttctccgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 2040
aacctgttccatc aatccatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 2100
agagtgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 2160
cggggatgttca tggatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 2220
ggccctgttca tggccaaacggcc gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 2280
aaaaaaatgtt gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 2340
aacctgttccatc aatccatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 2400
gacttcaacccatc actaaatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 2460
aacaacacccatc tggatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 2520
gtgaacgttca tggatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 2580
ggccctgttca tggatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 2640
gaaacggatcc agggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 2700
ctggatgttca tggatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 2760
aacaacacccatc tggatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 2820
gtgaacgttca tggatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 2880
ttctgttccatc aatccatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 2940
aagggttccatc aatccatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 3000
ttctgttccatc aatccatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 3060
aacaacacccatc tggatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 3120
ctggatgttca tggatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 3180
aagaacgttca tggatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 3240
gcaaaaaatgtt gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 3255
```

SEQ ID NO: 39 moltype = DNA length = 3156
FEATURE Location/Qualifiers
source 1..3156
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 39

```
aagcggaaactt acatctgttccatc cctggatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 60
gacttcaacccatc aatccatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 120
gaaaacgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 180
catagaatccatc aatccatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 240
aagggttccatc aatccatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 3000
ttctgttccatc aatccatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 3060
aacaacacccatc tggatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 3120
ctggatgttca tggatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 3180
aagaacgttca tggatgttca gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 3240
gcaaaaaatgtt gggatgttca tccatccatc acggccatcattt atccatgttca tggcccccgtt gaaqaacttcc 3255
```

-continued

SEQ ID NO: 40 moltype = DNA length = 7009
FEATURE Location/Qualifiers
source 1..7009
mol_type = other DNA
organism = synthetic construct

-continued

acaactccgc	cccattgacg	caaatggcg	gtagggcgtgt	acgggtggag	gtctataataa	1260
gcagagctct	ctggctaact	accggtgcca	ccatgaaaag	gaactacatt	ctggggctgg	1320
acatecggat	tacaaggcgt	gggtatggg	ttatgtacta	tgaaaacaagg	gacgtgatcg	1380
acgcaggcgt	cagactgttc	aaggaggca	acgtggaaa	caatgaggga	cggagaagca	1440
agaggggagc	caggcgctg	aaacgacgga	gaaggcacaag	aatccagagg	gtgaagaaac	1500
tgtgttgcg	ttacaacctg	ctgaccgaco	attctgagct	gagtggattt	aatctttag	1560
aagccagggt	gaaaggcctg	agtcaaga	tgtcagaggaa	agatgtttcc	gcagctctgc	1620
tgacacccgc	taagcgccga	ggagtgcata	acgtcaatga	gggtggaaag	gacaccggca	1680
acgagctgtc	tacaaaaggaa	cagatctac	gcaatagcaa	agctctggaa	gagaagtatg	1740
tcgcagacgt	gcagctggaa	cggctgaaqa	aaqatggcga	ggtgagagg	tcaattataa	1800
ggttcaagac	aagcgactac	gttcaaga	ccaaggcag	gctgaaaatg	cagaaggctt	1860
accaccacgt	ggatcagagc	ttcatcgata	cttataatcg	cctgtggag	actcggagaa	1920
cctactatga	gggaccagga	gaaggaggcc	ccttcggatg	gaaagacatc	aaggaatgg	1980
acgagatgtc	gttgggatc	tgcacccat	ttccagaaga	gctgagaagc	gtcaagtgac	2040
cttataacgc	agatctgtac	aacccctga	atgaccatgaa	caacctggc	atcaccagg	2100
atgaaaaacgc	gaaactggaa	tatctatg	agttccatgat	catcggaaac	qtgttaagc	2160
agaagaaaaa	gcctacactg	aaacagattt	ctaaggagat	cctggtaaac	gaagaggaca	2220
tcaagggcota	ccgggtgaca	agcactggaa	aaccaggat	ccccatctg	aaagtgtatc	2280
acgatattaa	ggacatcaca	gcacggaaag	aaatcattga	gaacccggaa	ctgtggatc	2340
agattgtca	gatctgtact	atcttccaga	gctccggag	catccaggaa	gagctgacta	2400
acctgaacag	cgagctgacc	caggaagaga	tcgaacagat	tagtatctg	aagggttaca	2460
ccggaacacca	caacctgtcc	ctgaaagta	tcaatctgt	tctggatgg	ctgtggata	2520
caaacgcacaa	tcagatttca	atctttaacc	ggctgaagct	gggtccaaa	aagggtggacc	2580
tgagtcaagc	gaaagagatc	ccaaacac	ttgtggacg	tttcattctg	tcacccgtgg	2640
tcaagcggag	cttcattccag	agcatcaaa	tgtatcaacgc	catcatcaag	aagtacggcc	2700
tgcccaatga	tatcttatt	gagctggta	gggagaagaa	cagcaaggac	gcacagaaga	2760
tgtatcaatga	gtatgttca	ggggatggc	aaqatggcga	agcggatagg	ttttttttcc	2820
gaactaccgg	gaaagagaa	cgaaagtacc	tgattgaaa	aatcaatgc	cacgatatac	2880
aggagggaaa	gttgtctgtat	tctctggagg	ccatccccct	ggaggacgt	ctgaacatac	2940
cattcaacta	cgaggtcgat	catatatacc	ccagaaggct	gtccctcgac	aattcttta	3000
acaacaaagg	gttggatccag	caggaagaga	actctaaaaa	gggcaatagg	actcttttcc	3060
agtacctgtc	tagttcagat	tccaaatgtt	cttacggaaac	ctttaaaag	cacattctga	3120
atctggccaa	aggaaaaggc	cgccatcaga	agacaaaaaa	ggagtaactg	ctggaaagac	3180
gggacatcaa	cgatcttcc	gtccagaagg	attttattaa	ccggatctg	gtggacacaa	3240
gatacgtcac	tcgcgggctg	atgatctgc	tgcgatctc	tttccgggt	aacaatctgg	3300
atgtgaaatg	caagtccatc	aacgggggtt	tccatcttt	tctgaggcgc	aatatggaa	3360
ttaaaaaagg	gogcaacaaa	gggttacaag	accatggca	agatgtctg	attatcgca	3420
atgccgactt	catcttaag	gagttggaaa	agctggacaa	agccaagaaa	gtgatggaga	3480
accagatgtt	cgaaagaga	caggccgat	ctatggccca	aatcgagaca	gaacaggagt	3540
acaaggagat	tttccatcact	cctccaggaa	tcaagcatat	caaggatttc	aaggactaca	3600
agtaactctca	ccgggtggat	aaaaacccca	acagagacgt	gatcaatgac	accctgtat	3660
gtacaagaaa	agacgataag	gggataatcc	tgattgtga	caatctgaac	ggactgtacg	3720
acaacaaat	tgacaaatgc	aaaaacgtca	tcaacaaaag	tcccgaaag	ctgtgtatgt	3780
accaccatgt	tccttcgaca	tatcagaaat	tgaatgttgc	tatggacg	tcacggcag	3840
agaagaaccc	actgtataag	tactatgtaa	agactggaa	ctacctgacc	aagtatagcc	3900
aaaaggatata	tggccccgt	atcaagaaga	tcaagtacta	tgggaacaag	ctgaatgccc	3960
atctggacat	cacagacat	tacccttaa	gtcgcaacaa	gggtggtaa	ctgtcactga	4020
agccatatac	atctgtatc	tatctggaca	acggccgtgt	taattttgc	actgtcaaga	4080
atctggatgt	cataaaaaa	gagaactact	atgaatgtaa	tagcaatgtc	tacgaagagg	4140
ctaaaaaaat	gaaaaaaat	agcaaccagg	cagatgtcat	cgcccttctt	tacaacaaac	4200
acctgtat	gatcaatggc	gaactgtata	gggtatctcg	gggtgaacaa	gatctgtctg	4260
accgcattaa	agtatgtat	attgacatca	cttaccggaa	gtatctggaa	aacatgtat	4320
ataaagcgccc	ccctcgaaat	atcaaaacaa	ttgccttata	gactcaatgt	atcaaaaat	4380
actcaaccga	cattctggaa	aacctgtatg	aggtgaagag	aaaaaagcac	ccttcagatta	4440
tcaaaaaagg	cagcgaggac	aagcgctcc	ctgctactaa	gaaagctgtt	caagctaa	4500
aaaagaaaagg	atcttaccca	taatgtatc	catggatgtt	catggatgtt	ctagatgtcg	4560
ctgatcgcc	tccactgtgc	tttctatgtt	ccagccatctt	gttggggcc	cctcccccgt	4620
gecccttgc	accctggaa	gtgcactcc	cactgtctt	ttctataataa	atgagaaat	4680
tgcatecgat	tgtgttgcg	gggttgcattc	tattctgggg	gggtgggggg	ggcaggacag	4740
caagggggag	gattggggag	agaatagcg	gcatgttgc	gaggtagccg	ccgccccgg	4800
tggagctcca	gtttttgttc	cctttagtga	gggttatttg	ccgcgttggc	gtatcatgg	4860
tcatagctgt	ttctctgtgt	aaattgttat	ccgctcaacaa	ttccacacaa	catacgagcc	4920
ggaagcataa	agtgtttaa	ctgggggtcc	taatgtgtga	gctaaatcac	attaatttgc	4980
tttgcgttac	tgcccgctt	ccatcgccg	aacctgtcg	gccatgtca	ttaatgtatc	5040
ggccaaacgc	cggggggggg	cggttttgcg	atggggctt	tttccgttctc	ctcgcttact	5100
gactcgctgc	gtcggtcg	tcggctcg	cgagccgtat	cagtcactc	aaaggccgt	5160
atacgggtat	ccacagaatc	agggataac	gcaggaaaga	acatgtgagc	aaaaggccag	5220
caaaaggcgc	ggaaccgtaa	aaaggccgc	ttgttgcgt	ttttccatag	gtctccccc	5280
cctgcacgc	atcacaaaaa	tcgacgtca	agtcagatgt	ggcgaaaccc	gacaggacta	5340
taaaagatacc	aggcggttcc	ccctggaa	tcctctgtgc	gtcttctgt	ttcgaccctg	5400
ccgccttacc	gatactgtc	ccgccttctc	ccttcggaa	gggtggccgt	ttctctatgc	5460
tcacgctgt	ggatctgtat	ttccgtgtat	gtcggttgcgt	ccaaatgtgg	ctgtgtgcac	5520
gaacccccc	ttcagccgca	ccgcgtcgcc	ttatccgtta	actatgtct	tgagtccaa	5580
ccggtaagac	acgacttatac	gccactggca	gcagccactg	gtaacaggat	tagcagagcg	5640
aggtatgtat	cgccgtgtatc	aggttcttgc	aaatgtgttgc	ctaactacgg	ctacactaga	5700
aggacagat	ttggatctgt	cgctctgtcg	aaggccgtta	ccttcggaaa	aaggttgggt	5760

-continued

actcttgcgat	ccggcaaaaca	aaccaccgct	ggtagcggtg	gttttttgt	ttgcaagcag	5820
cagattacgc	gcagaaaaaa	aggatctcaa	gaagatctt	tcatctttc	tacggggct	5880
gacgctagt	ggaacgaaaa	ctcacgtta	gggattttgg	tcatgagatt	atcaaaaagg	5940
atttcacct	agatctttt	aaattaaaaa	tgaatttta	aatcaatcta	aagtataat	6000
gagtaaaactt	ggtctgacag	ttaccatgc	ttaatcagtg	aggcacctat	ctcagcgatc	6060
tgcttatttc	gttcatccat	agtgcctga	ctccccgtc	tgttagataac	tacgatacgg	6120
gaggggcttac	catctggccc	cagtgtcga	atgataccgc	gagaccacag	ctcaccggct	6180
ccagatttat	cagcaataaa	ccagccagcc	ggaagggccg	agcgcagaag	tggctctgca	6240
actttatccg	cctccatcca	gtcttataat	tgttgcgggg	aagctagagt	aagtagttcg	6300
ccagtaataa	gttgcggca	cgtgttgc	attgtacag	gcatctgtgt	gtcacgcctg	6360
tcgtttggta	tggtttcatt	cagetccgt	tcccaacgat	caaggcgaat	tacatgatcc	6420
cccatgttgt	gcaaaaaaagc	ggtagtctc	ttccggcttc	cgatcgatgt	cagaagtaag	6480
ttggcccgag	tgtttatcact	catggttatg	gcagcaactgc	ataattctct	tactgtcatg	6540
ccatccgtaa	gatgttttgc	tgtgactgtt	gagtaactcaa	ccaagtcatt	ctgagaatag	6600
tgtatgcggc	gaccgagttg	ctcttccgg	gctgtcaatac	gggataatac	cgcgcacat	6660
agcagaactt	aaaaatgtct	catattgg	aaacgttctt	cggggcggaaa	actctcaagg	6720
atcttacgc	tgtttagatc	cagttcgat	taacccactc	gtgcacccaa	ctgatctca	6780
gcatctttt	cttcttgc	cgttctggg	tgagcaaaaa	caggaaggca	aatgcccga	6840
aaaaggggaa	taaggggcgac	acggaaatgt	tgaataactca	tactttctt	ttttcaat	6900
tatttgacg	tttatcaggg	ttatgtctc	atgagcggat	acatatttg	atgtatttag	6960
aaaaataaac	aaatagggggt	tccgogcaca	tttccccgaa	aagtgcac		7009

SEQ ID NO: 41 moltype = AA length = 2414
 FEATURE Location/Qualifiers
 source 1..2414
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 41

MAENVVVEPGP	PSAKRPKLSS	PALSASASDG	TDFGSLFDLE	HDLPDELINS	TELGLTNGGD	60
INQLQTSLGM	VODAASKHKQ	LSELLRSGSS	PNLNMGVGGP	GOVMASQAQQ	SSPGLGLINS	120
MVKSPMTQAG	LTSPNMGMT	SGPNQGPTQS	TGMMNSPVNQ	PAMGNTGMN	AGMNPGLMAA	180
GNGQGIMPNQ	VMNGSIGAGR	GRQNMQYPNP	GMGSAGNLIT	EPLQGSPQM	GGQTGLRGQP	240
PLKMGMMNNP	NPYGSPYTQN	PGQIGIGASGL	GLQIQTKTVEL	SNNLSPFAMD	KKAVPGGGMP	300
NMGQOPAQV	QPQGLVTPVA	QGMGSSGA	DPEKRKLIQQ	QLVLLLHAKK	CQRREQANGE	360
VRQCNLPHCR	TCATVNLNMHT	HQCSKGSCQV	AHCASSRQII	SHWKNCNTRHD	CPVCLPLKNA	420
GDKRNQNPIL	TGAPVGLGNP	SSLGVGQQSA	PNLSTVSQID	PSSEIRAYAA	LGLPYQVNQM	480
PTQPQVQAKN	QONQOPQGSP	QGMRPMNSMS	ASPMGVNGVG	GVQTPSLLSD	SMLHSAINSQ	540
NPMMSENASV	PSMGPMPAA	QPSTTGIRKQ	WHEDITQDLR	NHLVHKLVQA	IFPTPDPAAL	600
KDRRMENLVA	YARKVEGDMY	ESANNRRAEYY	HLLAEKIYKQ	QKELEEKRT	RLQKQNMLPN	660
AAGMPVPSMN	PGPNMGQPQP	GMTSNGPLPD	PSMIRGSPVN	QMMRPTQPS	GLNQFGQMSM	720
AQPPIVPRQT	PPLQIHQGOLA	QPGALNPPMG	YGRPMQOPSN	QGQFLPQTQF	PSQGMNVTNI	780
PLAPSSGQAP	VSQAQMSSSS	CPVNSPIMP	GSQGSHIHC	QLPQPALHQH	SPSPVPSRTP	840
TPHHTPPSIG	AQQPATTIP	APVPPTPAMP	PGPQSQLAHF	PPRQPTPTPT	TQLPQQVQPS	900
LPAAPSADQP	QQOPRQSQST	AASVPTPTAP	LLPPQPATPL	SQPAVSIEGQ	VSNPPSTSST	960
EVNSQATAEK	QPSQEVKMEA	KMEVDQPEPA	DTQPEDISES	KVEDCKMEST	ETEERSTELK	1020
TEIKEEEDQP	STSATQSSPA	PGQSKKKIFK	PEELRQALMP	TLEALYRQDP	ESLPFRQPVD	1080
PQLLGIIDFY	DIVKSPMDLS	TIKRKLDTGC	YQEPWQYVDD	IWL MFNNAWL	YNRKTSRVYK	1140
YCSKLSEVFE	QEIOPVQMSL	GYCCGRKLEP	SPOTLCCYKG	QLCTIPRDAT	YYSYQNRYHF	1200
CEKCFNEIQQ	ESVSLGDDPS	QPQTTINKEQ	FSRKNDTLD	PELFVCTEC	GRKMHQICV	1260
HHEIIWPAF	VCDGCLKSA	RTRKENKFS	KRLPSTRGLT	FLENRVNDFL	RRQNHPESGE	1320
TVTRVVFHASD	KTVEVKPGM	ARFVDSGEMA	ESFPYRTKAL	FAFEEIDGVD	LCFFGMHVQE	1380
YGSDCPPPNQ	RVRYISYLD	VHFFRCPCLR	TAVYHEILIG	YLEYVKLGY	TTGHIWACPP	1440
SEGDDYIFPHC	HPPDQKIPKP	KRLQEWYKKM	LDAKAVSERIV	HYDKDIFKQA	TEDRLTSAKE	1500
LPYFEGDFWP	NVLEESIKEL	EQEEEERKRE	ENTSNESTDV	TKGDSKNAKK	KNNKKTSKNK	1560
SSLSRGNKNN	PGMPNVSNDL	SQKLIAATMEK	HKEVFFFVIRL	IAGPAANSLP	PIVDDPDLIP	1620
CDLMDGRDAF	TLTARDKHL	FSSLNRAQWS	TMCMLVELTH	QSQDRFVYTC	NECKHHVETR	1680
WHTCVCEDYD	LCITCYNTK	HDHKMEKLG	GLLDESNNQQ	AAATQSPGD	RRLSIQRCIQ	1740
SLVHACQCRN	ANCSLPSCQK	MKRVVQHTKG	CKRKTNNGCP	ICKQLIALCC	YHAKHCQENK	1800
CPVPFCLNK	QKLRQOOLQH	RQOQAQMLRR	RMASMQRTGV	VGQQQGLPSP	TPATPTTPTG	1860
QQPTTPQTPQ	PTSQPQPTPP	NSMSPQPLRT	AAQGPVQSGK	AAGQVTPPTP	PQTAQPPPLPG	1920
PPPAAVEMAM	QIQRRAETQR	OMAHVQIFQR	P1QHQMPPTM	PMAPMGNNPP	PMTRGSPSGHL	1980
EPGMGPTGMQ	QQPPWSQGGL	PQPQQLQSGM	PRPAMMSVAQ	HGQPLNMAPQ	PGLGQVGISP	2040
LKPGTVSQA	LQNLLRTLRS	PSSPLQQQOV	LSTLHANPOL	LAIFIQRAA	KYANSNPQPI	2100
PGQPGMPQGQ	PGLQPPPTMPG	QGQVHSNPAM	QNMNPQMAGV	QRAGLPQQQP	QQQLQPPMGG	2160
MSPQAQQMNM	NHNTMPQSF	DILRRQQMMQ	QQQQQGAGPG	IGPGMANHNQ	FQQPQGVGYP	2220
PQQQQRMOHH	MQQMQQGNMG	QIGQLPQALG	AEAGASLQAY	QORLLQQQMG	SPVQPNPMSP	2280
QQHMLPNQAOQ	SPHLQGQQIP	NSLSNQVRSP	QPVPSPRPOS	QPPHSSPSPR	MQPQPSPHV	2340
SPQTSSPHPG	LVAAQANPME	QGHFASPDQN	SMLSQLASNP	GMANLHGASA	TDLGLSTDNS	2400
DLNSNLSQLST	LDIH					2414

SEQ ID NO: 42 moltype = AA length = 617
 FEATURE Location/Qualifiers
 source 1..617
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 42

-continued

IFKPEELRQA	LMPTLEALYR	QDPESLPPRQ	PVDPQLLGIP	DYFDIVKSPM	DLSTIKRKL	60
TGQYQEWPWQY	VDDIILMFNN	AWLYNRRKTSR	VYKYCSKLS	VFEQEIDPVM	QSLGYCCGRK	120
LEFSPTQLTCZ	YKGQLCTIPIR	DATYYSYQMR	YHFCCEKCFNE	IQGESVSLGD	DPSQPTTIN	180
KEQFSKRKN	TLDPELFVCE	TECGRKMHQI	CVLHHHEIIWP	AGFVCDGCLK	KSARTRKEN	240
FSAKRLPLSTR	LGTFLLENRVN	DFLERQNHPFE	SEGVTFVRVBS	ASDKTVEVKP	GMKFARVFDSG	300
EMAESFPYRT	KALFAFEEID	GVDLCCFFGMH	VQEYGSDCPP	PNQRVVYISY	LDSVHFFRPK	360
CLRATAVYHEI	LIGLYLEYVKK	LGYTGGHWA	CPSEGGDDYI	FHCHPDPQKI	PKPKRLQEWY	420
KKMLDKAVSE	RIVHDYKDIF	KQATEDRLTS	AKELPYFEGD	FWPNVLEESI	KELEQEEEER	480
KREENTSNES	TDVTKGDSKN	AKGNKNKTKS	KNKSSLRGN	KKPGPMNVS	NDLSQLKLY	540
MEHKHHVFFFY	IRLIAVGPAAN	SLPPIVDPDP	LIPCDLMGDR	DAFLTLARDK	HLEFSSLRRA	600
QWSTMCMVL	E LHTQSQD					617

```
SEQ ID NO: 43          moltype = AA    length = 1497
FEATURE                Location/Qualifiers
source                  1..1497
                        mol_type = protein
                        organism = synthetic construc
```

SEQUENCE : 43

SEQUENCE: 43	RADALDDFDL DMLGSDALDD FDLDLMLGSDA LDDFDLDMGL SDALDDDFDLD MVNPKKKRKV	60
GRGMDDKYSI	GLAIGTNSVG WAVITDEYKV PSKKFKVLGN TDRHSIKKNL IGALLFDSGE	120
TAAEATRLKR	ARRYYTRRKW RICLYQLEIFS NEMAKVUDSF FHRLEESFLV EEDDKHHERP	180
IFGNVINDVEA	YHEKYPTIYH RKLKLVSTD KadLRLIYL AHMFKRGH FLIEGDLNPD	240
NSDVLDLFIQ	LVQTVNTQLF EENPINSAGSD AILKSARLS KSRRLENLIA OLPGKEKGNL	300
FGNLIALSLG	LTPNFKSNFD LAEADAKLQLS KDTYDDDLDN LLAQIQDGQA DLFLAAKNLS	360
DAIILSDILR	VNTEITKAP SASKMIRKYE HQQDLTLKKA LVRQQOLPEKY KEIFFDQSOK	420
GYAGYIDGGA	SQEEFYKFIF PILEKMDGTE ELVVKLNRN DLLRKQRTFDN GSIPHQIHLG	480
ELHALIRRQE	DFYPFLKDNP EKLEKILTFR IPYVGVPLR GNSRFAMWTR KSEETITPW	540
FEEVVLDKGAS	AQSFIERMTN FDKNLPNEKV LPKHSLLYYE FTVYNELTKV KYVTEGMRKP	600
AFLSGBQKKA	IVDLFLKTNP KVTVQKLEKD YFKKIECPDS VEISVGVEDR NASLGTYHDL	660
LKIIKDKDFL	DNEENEDEDILE DIVLTTLFE DREMIEERLK TYAHLPPDKV MKQLKRRRTY	720
WGWRSLRKLII	NGIRDQKSGK TLDLFLSKD FANRNFMQLI HDDSLTFKED IQKAVQGSQG	780
DSLHEHIANL	AGSPAIIKKGI LQTVKVVDEL VKVMGRHKPE NIVIEMAREN QTTQKGQKNS	840
REMRMKRIEEG	IKELQGSQLI EHPVENTQLOQ NEKLYLYLQ NGDRMVTDQD LDINRRLSDY	900
VDAIVPQSF1	KDDSIDNKVL TRSDKNRKGK DNPVSEEVVK KMKNYWRQQL NAKLITQRKF	960
DNLTKAERGG	LSELDKAGFI KRLQVLTROI TKHVQAQLDS RMNTKYDEND KLIREVKVIT	1020
LKSKLVLSDFR	KDFQFQVKRE INNYHHADHA YLNNAVVTAL IKKYPKLESE FVYGDYKVYD	1080
VRKMIAKSEQ	EIGKATAKYF FYSNMNNFFK TEITLANGEI RKRPLIETNG ETGEIIVWDKG	1140
DRFATVRKVL	SMQPVNIVKK TEVQTGGFSK ESIPLKRNDS KLIARKKWDW PKKYGGFDSP	1200
TVAYSVLVVA	KVEKGKSKKL KSVKELLGIT IMERSSEFKN PIDFLEAKGY KEVKDLDLIIK	1260
LPKYSLFELE	NGRKMLASA GELQKGNEIA LPSKYVNFY LASHYEEKLKG SPEDNEQKQL	1320
FVEQHKHYL	EIIIQEISEFS KRVILADANL DKVLSAYNHK RDKPIRQEAE NIIHFLFTLN	1380
LGAPAAFKYF	DTTIDTRKRYT STKEVBLDATL IHQSITGLYE TRIDLSQLGG DSRADPKKKR	1440
KVASRADALD	DFDIDMLGSD ALDDDFDLDL MGSALDDDFDLD MGLGSDALDD FDLDMLI	1490

SEQ ID NO: 44 moltype = DNA length = 4491
FEATURE Location/Qualifiers
source 1..4491
mol_type = other DNA
organism = synthetic constru

-continued

ggaaattcca	gattcgcgtg	gatgactcg	aaatcagaag	agaccatcac	tccctggAAC	1620
ttcgaggaa	tcgttgataa	gggggcctt	gcccaagtcc	tcatgaaag	gatgactaac	1680
tttgataaaa	atctgcctaa	cgaaaagggt	cttccctaaac	actctctgt	gtacgagtt	1740
ttcacagttt	ataacgagct	caccaaggtc	aaatacgtca	cagaagggt	gagaaggcca	1800
gcattcctgt	ctggagagca	gaagaaagct	atcgtggacc	tccttcaa	gacgaaccgg	1860
aaagttaacc	tgaaaacagt	caaagaagac	tatttcaa	agattgaatg	tttcgactct	1920
gttgaatca	cgggagggt	ggatcgctt	aacgcatacc	tgggaaacgt	tcacgatct	1980
ctgaaaaatca	ttaaagacaa	ggacttcctg	gacaatgagg	agaacgagga	catttttag	2040
gacattgtcc	tcacccctac	gttggtaa	gatagggaga	tgattgaaga	acgcttggaa	2100
acttacgctc	atcttcgtc	cgacaaaggct	atgaaacagc	tcaagaggcg	ccgatataca	2160
ggatggggcc	ggctgtcaag	aaaactgatc	aatgggatc	gagacaaggc	gagtggaaag	2220
acaatcttgg	attttttaa	ttccgatgg	tttgcaccc	ggaacttcat	gcagtgtac	2280
catgtatgact	ctctcacc	taaggaggac	atccagaaag	cacaagttt	tggccagggg	2340
gacagtcttc	acggcacat	cgctaatctt	gcaggtagcc	cagctatcaa	aaagggaaata	2400
ctgcagaccg	ttaaggctgt	ggatgaaact	gtcaaaaggta	tgggaaggca	taagcccgg	2460
aatatcgta	ttcgatggc	ccqgagaa	caaaatcccc	agaaggdaca	gaagaaacagt	2520
agggaaagga	tgaagaggat	tgaagagggt	ataaaagaac	tggggtccca	aatcttaag	2580
gaacaccccg	ttgaaaacac	ccagctttag	aatgagaagc	tctacctgt	ctacctgcag	2640
aacggcaggg	acatgtacgt	ggatcaggaa	ctggacatca	atcggcttc	cgactacgac	2700
gttggatgcca	tcgtggccca	gtctttctc	aaagatgat	ttattgataa	taaagtgtt	2760
acaagatccg	ataaaaatag	agggaaaggt	gataacgtcc	cctcagaaga	agttgtcaag	2820
aaaatgtaaaa	attttggc	gcagctgtca	aacggccaaac	tgatcacaca	acggaaaggcc	2880
gataatctga	ctaaggctgt	acggaggctg	ctgtctgagt	ttggaaaggc	cggttctatc	2940
aaaaggcagg	ttgttgagac	acggccagatc	accaaggcag	tggcccaat	tctcgat	3000
cgcacatgaa	ccaagttacga	tggaaatgac	aaactgattt	gagagggtgaa	agtttattact	3060
ctgaagtctt	agctgttctc	agatttca	aaggacttcc	agttttataa	ggtgagagag	3120
atcaacaatt	accacatgc	gcatgtatc	tacatgtatc	cagtgttagg	cactgcactt	3180
atcaaaaaat	atcccaagat	tgaatctgg	tttggtttac	gagactataa	agttgtacat	3240
gttagggaaa	tgatcgcaaa	gtctgagcag	gaaataggca	aggccacccg	taagtactt	3300
ttttacagca	attatgtaa	ttttttcaag	accgagatca	cactggccaa	tggagagat	3360
cggaaaggcgc	cacttctgca	aacaaacgg	gaaacaggag	aatcgtgtt	ggacaagggt	3420
agggatttcg	cgacagtcgg	gaaaggctctg	tccatgcgc	agggtgaaat	cgttaaaaag	3480
accgaagttac	agacccggagg	cttcttca	gaaagatattc	tcccgaaaag	gaacagcgcac	3540
aagctgtatc	cacgcacaaa	agattgggac	cccaagaaat	acggcggatt	cgattcttct	3600
acagtgcgtt	acagtgtact	ggtgtggcc	aaatgtggaa	aagggaaggc	taaaaaactc	3660
aaaaggcgtca	aggaactgt	gggcacatca	atcatgggc	gatcaagctt	cggaaaaaac	3720
cccatcgact	ttctcgaggc	gaaaaggat	aaagaggatc	aaaagagact	catcattaag	3780
cttcccaagt	actctcttctt	tgagcttga	aacggccgg	aacgaatgt	cgcttagtgc	3840
ggcgagctgc	agaaaaggta	cgatgtggca	ctggccctct	aatactgtaa	tttcttgc	3900
ctggccaggcc	actatgtggaa	gctcaaaagg	tctccca	gatataggca	gaagcagct	3960
ttctgtggac	aacacaaaca	ctacccgtat	gagatcatcg	agcaaaataa	cgaattctcc	4020
aaaagagtga	tcctcgccga	cgctaacc	gataagggtc	tttctgttca	caataagcac	4080
agggataa	ccatcgagg	gcaaggcaggaa	aacattatcc	acttgttac	tctgaccaac	4140
ttggggcgcgc	ctgcacgctt	caagtactt	gacaccacca	tagacagaaa	cggttacacc	4200
tctacaaagg	aggtcttgc	cgccacact	attcatagt	caattacggg	gtctatgaa	4260
acaagaatcg	acctctctca	gctcggtgg	gacagcagg	ctgacccaa	gaagaagagg	4320
aagggtgttca	gocgcggccg	cgccgtgg	gatttgcgt	tgcacatgt	gggttctgt	4380
gcctcgatgt	actttgac	ggatgttt	ggaaggcagc	cattggatg	ctttgtatc	4440
gacatgtctgc	tctcgatgt	ttcgatctcg	atatgttaat	c		4491

SEQ ID NO: 45 moltype = AA length = 65
 FEATURE Location/Qualifiers
 source 1..65
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 45
 RTLVTFKDVF VDFTREEWKL LDTAQQILYR NVMLENYKNL VSLGYQLTKP DVILRLEKGE 60
 EPWLW 65

SEQ ID NO: 46 moltype = DNA length = 195
 FEATURE Location/Qualifiers
 source 1..195
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 46
 cggacactgg tgaccttcaa ggtatgttt gtggacttca ccagggagga gtggaaagt 60
 ctggacactg ctcagcagat cctgtacaga aatgtgtatgc tggagaacta taagaacct 120
 gtttccttgg gttatcagat tactaaggca gatgtgtatcc tccgggttga gaaggagaa 180
 gagccctggc tgggt 195

SEQ ID NO: 47 moltype = AA length = 1519
 FEATURE Location/Qualifiers
 source 1..1519
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 47

-continued

MDYKDHDGDY	KDHIDYKDD	DDKMAPKKR	KVGRGMDKY	SIGLAIGTNS	VGWAVITDEY	60
KVPSKKFVL	GNTDRHSIKI	NLIGALLFD	GETAEATRLK	RTARRRYTRR	KNRICYLQE	120
FSNEMAKVDD	SFFHRLEESF	LVEEDKKHER	HPIFGNVIDE	VAYHEKYPTI	YHLRKKLVDS	180
TDKADLRILY	LALAHMIKFR	GHFLIEGDLN	PDNSDVDFKL	IQLVQTYNQL	FEENPINASG	240
VDAKAILSAR	LSKSRRLENL	IAQLPGEKKN	GLFGNLIALS	LGLTPNFKSN	FDLAEDAKLQ	300
LSKDTYDDDL	DNLLAQIGDQ	YADLFLAAKN	LSDAIIILSDI	LRVNTEITKA	PLSASMICKY	360
DEHHQDLTLL	KALVRQQLPE	YKKEIFFDQS	KNGYAGYIDG	GASQEEFYKF	IKPILEKMDG	420
TEELLVKLNR	EDLLRKQRTF	DNGSIPHQIH	LGBELHAILRR	QEDFYFFLKD	NREKIEKILT	480
FRIPYVVGPL	ARGNSRFAWM	TRKSEETITP	WNFEEVVDKG	ASAQSFIERM	TNFDKNLPNE	540
KVLPKHSLLY	EYFTVYNELY	KVKVTEGMR	KPAFLSGEOK	KAIVDLLFKY	NRKVTVKQLK	600
EDYFKKIECF	DSVEISGVED	RFNASLGTYH	DLLKIIKDKD	FLDNEENEDI	LEDIVLITL	660
FEDREMIEER	LKTYAHLFDD	KVMQQLKRRR	YTGWGRLSRK	LINGIRDQKS	GKTILDFLKS	720
DGFANRNFMQ	LIHDDSLTFK	EDIQKAQVSG	QGDSLHHEHIA	NLAGSPAIKK	GILQTVKVVD	780
ELVKVGMGRHK	PENIVIEMAP	ENQTTQKGQK	NSRERMKRIE	EGIKELGQSQI	LKEHPVENTQ	840
LQNEKLYLY	LQNGRDMYVD	QELDINRQK	YDVDAIVPQS	FLKDDSIDNK	VLTRSDKNRG	900
KSDNPVPSEEV	VKMKMNYWRQ	LLNAKLITQR	KFDNLTKAER	GGLSELDKAG	FIKROLVETR	960
QITKHVAQIL	DSRMNTKYDE	NDKLIREVKV	ITLKSKLVS	FRKDFQFYKV	REINNNYHHAH	1020
DAYLNAVVG	ALIKKYPKLE	SEFVYGYDKV	YDVRKMITAK	EQEIGKATAK	YFFYSNIMNF	1080
FKTEITLANG	EIRKRPRIET	NGETGEW	KGRDFATVVK	VLSMPQVNIV	KKTEVQTCGGF	1140
SKESILPKRN	SDKLIAARKK	WDPKYGGFD	SPTVAYSVLV	VAKVEKGKSK	KLKSVELLG	1200
ITIMERSSFE	KNPIDFLEAK	GYKEVKKDLI	IKLPKYSLF	LENGRKMLA	SAGELQKGNE	1260
LALPSKVVF	LYLASHYEKL	KGSSEDNEQK	QLFVEQHKEY	LDEIIEQISE	FSKRVILADA	1320
NLDKVLSSAYN	KHRDKPIREQ	AENIILHFTL	TNLGAPAAFK	YFDTTIDRKR	YTSTKEVLDA	1380
TLIHQSITGL	YETRIDSQL	GGDSRADPKK	KRKVASDAKS	LTAWSRTLVT	FKDVFVDFTR	1440
EEWKLLDTAQ	QILYRNVMLE	NYKNLVSLGY	QLTKPDVILR	LEKGEWPWL	EREIHQETHP	1500
DSETAFEI	SVPKKKRKV					1519

SEQ ID NO: 48	moltype = DNA	length = 4560		
FEATURE	Location/Qualifiers			
source	1..4560			
	mol_type = other DNA			
	organism = synthetic construct			
SEQUENCE: 48				
atggactaca aagaccatga	cgggtattat	aaagatcatg acatcgatta	caaggatgac	60
gatggacaaga	tggccccc	gaagaaggg	aagggtggcgc	120
tccattggc	tcgcacatgg	caacaacagc	gtcggtctgg	180
aaagtgcgc	gaaaaaaaatt	caaagtctg	ggcaataacc	240
aacctcattg	gcgccttctt	gttcgactcc	ggggaaacc	300
agaaacagoc	ggcgacatgg	tacccggcga	aagaatcgg	360
tttagatag	atagttctaa	ggtggatgac	tcttcttcc	420
tttgtggagg	aggataaaaa	gcacggcg	cacccatct	480
gtggcgtacc	atgaaaagta	ccccaccata	tatcatctg	540
actgtaaa	ctgtactgcg	gttgatctat	ctcgcgctgg	600
ggacacttc	tcatcgagg	ggacacttgc	ccagacaaca	660
atccacttgc	ttcagactta	caatcagtt	ttcgaagaga	720
gttgacgcca	aagcaactct	gagcgctagg	ctgtccaaat	780
atcgacacgc	tccctgggaa	gaagaagaa	ggccgttgc	840
ctggggctg	cccccaactt	ttatcttac	ttcgacctgg	900
ctgagcaag	acacccatgc	tgatgtatc	gacaatctgc	960
ta	tttttttggc	ggcaagaac	ctgtcagac	1020
ctgcgactg	acacccatgc	caccaaaat	ccgctgagcg	1080
gatggagacc	accaagactt	gacttgcgt	ctagatgtat	1140
aatgtacaagg	aaattttctt	cgatcgtat	aaaatggct	1200
ggagcaagcc	aggagaaatt	tttacaaattt	attaagccca	1260
accgaggagc	tgctgtttaa	gcttacaaac	gaagatctgt	1320
gacaatggaa	tttttttttt	ccagatcc	tgccaaaca	1380
caagaggatt	ttaaaaaattt	tttggaaat	aacaggaaaa	1440
tttcggat	cctactatgt	aggccccctc	gccccgggaa	1500
actcgc	at	tggacacttc	tggacacttc	1560
gcctctgccc	atgttccat	cgaaaatct	actaacttgc	1620
aagggtgttc	ctaaacactc	tctgtgtad	gatgacttca	1680
aagggtcaat	acgtcacaga	agggtgaga	aagccagcat	1740
aaaggtatcg	tggacacttc	tttcaagac	aaccggaaa	1800
gaagactatt	tcaaaaat	tgtatgttc	gactctgtt	1860
cgettcaacg	catccctgg	aacgtatc	gaaatcgg	1920
ttcctggaca	atgaggagaa	cgaggacatt	tttgaggaca	1980
tttgaagata	gggagatgt	tgaagaac	tttgcactt	2040
aaagtcatga	aacacgtca	gaggcccg	tatacaggat	2100
ctgatcaatg	ggatccgg	caagcagat	ggaaagacaa	2160
gatggat	ccaacccgg	tttcatcgat	tttatccat	2220
gaggacatcc	agaaacgcac	atgttctgc	caggggaca	2280
aatcttgcac	gtagccgc	tatcaaaaag	ggaatactgc	2340
gaactcgtca	aagtaatggg	aaggcataag	cccggataa	2400
gagaacccaa	ctacccagaa	gggacagaag	aacagtaggg	2460
gagggtataa	aagaactggg	gtcccaatc	cttaaggaaac	2520
cttcagaatg	agaagctcta	cctgtactac	ctgcagaacg	2580

-continued

caggaactgg acatcaatcg	gctctccgac tacgacgtgg	atgcccattgt gccccagtt	2640
tttctcaag atgatcttat	tgataataaa gtgttgacaa	gatccgataa aaatagaggg	2700
aagggtgata acgtccccctc	agaagaagtgtcaagaaa	tgaaaaatatttggccag	2760
ctgctgaacgc ccaaactgtat	cacacaacgg aagttcgata	atctactaa ggctgaacga	2820
ggtggcctgt ctgagttgga	taaagccggttcatcaaa	ggcagettgt tgagacacgc	2880
cagatcacca agcacgtggc	ccaaattctc gattcacggc	tgaacaccaa gtacgtatggaa	2940
aatgacaaacatgttgcgaga	ggtgaaagtattactgt	atgtctaaatgttgcgat	3000
ttcagaaagg actttcagttt	ttataagggtt	agagagatca acaattacca	3060
gatgcctacc tgaatcgat	ggtaggcact gcacttata	aaaaatatcc caagttgaa	3120
tctgaattttt ttacggaga	ctataaaggat	tacgatgttggaaatgtat	3180
gaggcaggaa taggcaggc	caccgttaaacttgcgat	ttttttttt acagcaatattat	3240
ttaagaccg agattacact	ggccaatggaa	ttttttttt acagcaatattat	3300
aacggagaaa caggagaaat	cgtgtgggac	ttttttttt acagcaatattat	3360
gtctgttcca tgccgcaggat	gaacatcgat	ttttttttt acagcaatattat	3420
tccaaggaaa gtatccccc	aaaagggac	ttttttttt acagcaatattat	3480
tgggacccca agaaatacgg	ttttttttt acagcaatattat	ttttttttt acagcaatattat	3540
gtggccaaag tggagaaagg	ttttttttt acagcaatattat	ttttttttt acagcaatattat	3600
atcacaatcg tggacgcgat	ttttttttt acagcaatattat	ttttttttt acagcaatattat	3660
ggatcatcaaa aggtcaaaa	ttttttttt acagcaatattat	ttttttttt acagcaatattat	3720
cttggaaaacg gccggaaaacg	ttttttttt acagcaatattat	ttttttttt acagcaatattat	3780
ctggcactgc cttctaaata	ttttttttt acagcaatattat	ttttttttt acagcaatattat	3840
aaagggttcc cggaaagataa	ttttttttt acagcaatattat	ttttttttt acagcaatattat	3900
cttggatgaga tcatcgacaa	ttttttttt acagcaatattat	ttttttttt acagcaatattat	3960
aacctcgata aggtgtttc	ttttttttt acagcaatattat	ttttttttt acagcaatattat	4020
gcagaaaaca ttatccactt	ttttttttt acagcaatattat	ttttttttt acagcaatattat	4080
tacttcgaca ccaccataga	ttttttttt acagcaatattat	ttttttttt acagcaatattat	4140
acactgttacatcgttgcata	ttttttttt acagcaatattat	ttttttttt acagcaatattat	4200
gggtggagaca cgagggtgt	ttttttttt acagcaatattat	ttttttttt acagcaatattat	4260
ctgactgctt ggtccggac	ttttttttt acagcaatattat	ttttttttt acagcaatattat	4320
gaggaggttga agctgtgtt	ttttttttt acagcaatattat	ttttttttt acagcaatattat	4380
aactataaga actgtgttcc	ttttttttt acagcaatattat	ttttttttt acagcaatattat	4440
ttgggagaaagg gagaagggcc	ttttttttt acagcaatattat	ttttttttt acagcaatattat	4500
gattcagaga ctgcatttga	ttttttttt acagcaatattat	ttttttttt acagcaatattat	4560

SEQ ID NO: 49 moltype = AA length = 1190
 FEATURE Location/Qualifiers
 source 1..1190
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 49
 MAPKKKRKV YIGVPAAKRN YILGLAIGIT SVGYGIIDYE TRDVIDAGVR LFKEANVENN 60
 EGRRSKRGAR RLKRRRRHRI QRVKLLFDY NLLTDHSELS GINPYEARVK GLSQLSSEE 120
 FSAALLHLAK RRGVHNVNVE EEDTGNELST KEQISRNSKA LEEKYVAELQ LERLKKDGEV 180
 RGSINRKFQTS DYVKEAKQOLL VPKQKYHQLD QSFIDTYIDL LETRRTYYEG PGEGLSPFGWK 240
 DIKEWYEMLM GHCTYPEEL RSVKYAYNAD LYNAFLNDLN LVITRDENEK LEYYEKFQII 300
 ENVFKQKKP TLKQIAKEIL VNEEDIKGYR VTSTGKPEFT NLKVYHDIKD ITARKEIEN 360
 AEELLDQIAKI LTIYQSSEDI QEELTNLNE LTQEEIEQIS NLKGYTGTHN LSLKAINLIL 420
 DELWHTNDQ IAIIFNRLKLV PKVVDLSQOK EIPTTLVDFP ILSPVVKRSP IQSIKVINAI 480
 IKKYGLPNDI IIELAREKNS KDAQKMINEM QKRNQTMER IEEIIRTTGK ENAKYLIIEKI 540
 KLHDMQEKGK LYSLEAIPLE DLLNNPFNYE VDHIIIPRSVS FDNSFNNKVL VKQEEASKKG 600
 NRTPFQYLSS SDSKLYSIEFT KKHILNLAKG KGRISKTKKE YLLEERDINR FSVQKDFINR 660
 NLVLDTRYATR GLMNLLRSYF RVNNLNLVVKV SINGGFTSFL RRKWKFKKER NKGYKHHaed 720
 ALIJIANADFI PKEWKLLDKA KVMMENQMPE EKQAESMPEI ETEQEYKEIF ITPHOIKHIK 780
 DFKDQYSHR VDKKPNNRELI NDTLYSTRKD DKGNLTIVNN LNGLYDKDND KLKKLINKSP 840
 EKLLMYHHDQ QTYPQLKLIM EQYGDKEKNPL YKYYEETCNY LTKYSKKDNG PVIKKIKYYG 900
 NKLNAHLDIT DDYPNLSRNKV VKLSSLKPYRF DVYLDNGVYK FVTVKNLNDVI KKENYYEVNS 960
 KYCEEAKKL KISNQAEFIA SFYNNNDLKI NGELYRVIGV NNDLLNRREV NMIDITYREY 1020
 LENMNDKRPP RIIKTIASKT QSIKKYSTDI LGNLYEVKSK KHPQIICKGK RPAATKKAGQ 1080
 AKKKKGSDAK SLTAWSRTLW TFKDVFDFT REEWKLDDTA QQILYRNVML ENYKNLVSLG 1140
 YQLTKPDVIL RLEKGEWPWL VEREIHQETH PDSETAFIEIK SSVPKKKRKV 1190

SEQ ID NO: 50 moltype = DNA length = 3570
 FEATURE Location/Qualifiers
 source 1..3570
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 50
 atggcccaa agaagaagcg gaagggtcggt atccacggag tcccagcagc caagcggAAC 60
 tacatcttgcg gcttggccat cggcatcacc agcgtgggct acggcatcat cgactacag 120
 acacggacgc tgatcgatgc cggcggtcgcc ctgttcaaag aggccaaact ggaaaacaac 180
 gagggcaggc ggagcaagag aggcccaaga aggtgttgcggc ggcggaggcg gcatagaatc 240
 cagagagtga agaagctgttgcgttgcactac aacctgttgcggc accaccacag cgactgttgc 300
 ggcgttgcacc cttacggatc cttttttttt acagcaatattat 360
 ttctctgtccg cctgttgcga cttttttttt acagcaatattat 420
 gaagaggaca cggcaacga gttgtccacc aaagagcaga tcagccggaa cagcaaggcc 480
 ctggaaagaga aatactgttgcg cttttttttt acagcaatattat 540

-continued

ccccggcagca	tcaacagatt	caagaccago	gactacgtga	aagaagccaa	acagctgctg	600
aagggtcaga	aggccttacca	ccatcgac	cagagcttca	tcgacaccta	catcgactg	660
ctggaaacc	ggccggaccta	ctatgggg	cctggcgagg	gcagccccct	cggtcgaaag	720
gacatcaa	aatggtaacga	gatgtatg	ggccactgca	cctactccc	cgaggaaactg	780
cgaggcgtg	agtacgcctt	caacgcgcac	ctgtacaacg	ccctgaacga	cctgaacaat	840
ctcgtatca	ccagggacga	gaacgagaag	ctgaaatatt	acgagaagg	ccagatcate	900
gagaacgtgt	tcaagcagaa	gaagaagcc	accctgaa	agatcgccaa	agaaatctc	960
gtgaacgaa	aggatattaa	gggttacaga	gtgaccacga	ccggcaaggc	cgaggtaacc	1020
aacctgaagg	tgtaccacga	catcaaggac	attaccgccc	ggaaagagat	tattgagaac	1080
gcccggatgc	tggatcagat	tgcacatc	accagacgac	cgaggacat	1140	
caggaagaac	tgaccaatct	gaactcgag	ctgacccagg	aagagatcga	gcagatctt	1200
aatctgaa	gtataccgg	cacccacaa	ctgagctga	aggccatcaa	cctgatctg	1260
gacgagctgt	ggcacaccaa	cgacaaccag	atcgctatct	tcaaccggct	gaagctggtg	1320
cccaagaagg	tggatctgtc	ccagcagaaa	gagatcccc	ccacccctgt	ggacgactc	1380
atccctgagc	ccgtcgat	gagaagctt	atccagacga	tcaacgtat	caacgcctc	1440
atcaagaagt	acggctgccc	caacqacatc	attatcgac	tggcccgca	gaagaactcc	1500
aaggacgccc	agaaaaatgat	caacgagat	cagaagcgga	accggcagac	caacgagcg	1560
atcgaggaaa	tcatccggac	cacccgcac	gagaacgcac	agtacctgt	cgagaagatc	1620
aagctgcac	acatgcac	aggcaggg	ctgtacagtg	tggaaagccat	ccctctggaa	1680
gatctgtg	tcaacccctt	caactatg	gtggacacc	tcatccccag	aagctgttcc	1740
ttcgacaaca	gcttcaacaa	caagggtct	gtgaagcagg	aagaagccag	caagaaggc	1800
aaccggacc	cattccgat	cctgagcago	acggacacga	agatcaacta	cgaaaccctt	1860
agaagcaca	tcttgaatct	ggccaaaggc	aaggccagaa	tcaagcagac	caagaagag	1920
tatctgtgg	aaaacgggaa	catcaacagg	tttccgtgc	agaaagactt	catcaacccg	1980
aacctgggt	ataccagata	cgccaccaga	ggectgtat	acctgtcg	gagctactt	2040
agagttaaca	acctggacgt	gaaagtgaag	tccatcaat	gcccgttcc	cagctttctg	2100
cgccggaa	ggaagttaa	gaaagacggc	aacaagggg	acaagcacca	cgccggaggac	2160
gcccgtatca	ttggcaacgc	cgatccatc	ttcaaaagat	ggaaaact	ggacaaggcc	2220
aaaaaaatgt	tggaaaacca	gatgttcag	gaaaaggcag	ccgagacat	gcccgagatc	2280
gaaaaccgac	aggagtacaa	agagatctt	atcccccc	accagatcaa	gcacattaag	2340
gacttcaat	actcaatgta	cagccacccgg	gtggacacca	agcataatag	agagctgatt	2400
aacgacaccc	tgtactccac	cggaaggac	gacaaggc	acacccat	cgtgaacat	2460
ctgaacggcc	tgtacgacaa	ggacatgac	aagctgaaa	agctgtat	caagacccc	2520
gaaaagctg	tgtatgtac	ccacgacccc	cagacctacc	agaaactgaa	gctgtattat	2580
gaacagtac	cgacgacgaa	gaatccccc	tacaatgt	acgaggaaac	cgggaaact	2640
ctgaccaat	actccaaaaa	ggacaacgc	ccctgtatc	agaagat	tattacggc	2700
aacaactg	acgccccat	ggacatcacc	gacgactacc	ccaacacg	aaacaagg	2760
gtgaagctgt	ccctgaagcc	ctacagat	gacgtgtacc	tggacaatgg	cgtgtacaag	2820
ttctgtgacc	tgaatgtat	ggatgtat	aaaaaaagaa	actactacg	agttaatagc	2880
aagtgtat	aggaagactaa	gaaatgt	agatcaga	accaggcga	gtttatcgcc	2940
tccttctaca	acaacgatc	gatcaagatc	aacggcgac	tgtatagat	gatcgctgt	3000
aacaacgacc	tgctgaacc	gatcgaatg	aatatgtatc	acatcacca	cccgacgtac	3060
cttgaaaaaca	tgaacgacaa	gaggcccc	aggatcatt	agacaatcgc	ctccaaagacc	3120
cagagcatt	agaagtacag	cacagacatt	ctggcacaac	tgtatgtat	gaaatctaag	3180
aagcaccc	agatcatca	aaaggccaa	aggccggcgg	ccacaaaa	ggccggccag	3240
gcaaaaaa	aaaaggatc	cgatgtca	tcaatgtact	cctggcccg	gacactgtgt	3300
acccatca	atgttgtt	ggatgtatc	aggggaggat	ggaactgt	ggacactgt	3360
cagcagatc	tgtacgaaa	tgtatgt	gagaactat	agaacctgt	ttccttggt	3420
tatcgttca	ctaaaggc	tgtatctc	cggttggaga	aggggaga	gccctggct	3480
gtggagag	aaattcacca	agagaccc	cctgattca	agactgcatt	tgaaatcaa	3540
tcatcgttca	cgaaaaagaa	acgcaaaat				3570

```

SEQ ID NO: 51      moltype = AA  length = 718
FEATURE          Location/Qualifiers
source           1..718
mol_type = protein
organism = synthetic construct

SEQUENCE: 51
LPTCSCLDRV IOKDKGPYYT HLGAGPSVAA VREIMENRYG QKGNNAIRIEI VVYTKEGKS 60
SHGCPIKAWV LRRSSDEEVK LCLVRQRTGH HCPTAVMVLL IMVWDGIPPL MADRLYTEL 120
ENLKSYNGHP TDRRCNLNE RTCTCQGIDP ETCGASFVFG CSWSMYFNNG KFGRSPSPR 180
FRIDPSSPLH EKNLEDNLQS LATRLAPIYK QYAPVAYQMQ VEYENVAREC RLGSKEGRPF 240
SGVTACLDFC AHPHRDIHN MNGSTVVCTL TREDNRSLGV IPQDEQLHVL PLYKLSDTDE 300
FGSKEGMEAK IKSGAIEVLA PRRKRTCTQ PVPRSGKRR AAMMTEVLAH KIRAVEKPKI 360
PRIKRKNNST TTNNSKPSSL PTLGSNTETV HCFILKSSDNT KTYSLMPSAP 420
HPVKEASPGF SWSPKTASAT PAPLKNDATA SCGFERSST PHCTMPSGRL SGANAAAADG 480
PGISQLGEVA PLPTLSAPVM EPLINSEPST GVTPLTPHQ PNHQPSFLTS PQDLASSPM 540
EDBQHSEADE PPSDEPLSDD PLSPAEEKLH HIDYEWSDSE HIFLDANIIG VAIAPAHSV 600
LIECARRELH ATTVEHPNR NHPTRLSLVF YQHKNLNKPQ HGFEELNKIKF EAKEAKNKKM 660
KASEQKDQAA NEGPEQSSEV NELNQIPSHK ALTLTHDNVV TVSPYALTHV AGPYNHHW 718

```

```

SEQ ID NO: 52      moltype = DNA  length = 2154
FEATURE          Location/Qualifiers
source           1..2154
mol_type = other DNA
organism = synthetic construct

```

-continued

SEQUENCE: 52

```

ctgcccacact gcagctgtct ttagatcgattt atacaaaaaag acaaaggccc atattataca 60
cacccgggg caggacaaag ttgttcgtct gtcaggaaa tcatggagaa taggtatgg 120
caaaaaaggaa acgcaataag gatagaataa gtatgtaca ccggtaaaaa agggaaaagc 180
tctcatgggt gtccaaattgc taagtgggtt ttaagaagaaa gcagtatgtga agaaaaagtt 240
cttgcgttgg tccggcagcg tacaggccac cactgtccaa ctgtgtgtat ggtgggtgtc 300
atcatgggtt gggatggcat cccttccaa atggccgacc ggctatacac agagctcaca 360
gagaatctaa aatgtatcacaa tgggcacccct accgacaaatgatgcaccc caatgaaaaat 420
cgtagctgtaa catgtcaagg aattgtatcca gagacttggt gagetttcatt ctctttggc 480
tgttcatgttggatgtactt taatggctgt aqgtttggta gaagccccaa ccccaaaaaaaga 540
tttagaaatggatgtacttcaatccaaatccatcataatgaaatggatgaaatccatcataatg 600
ttggctcacac gattagctcc aattttaatgaaatggatgtacttcaatccatcataatgaaatggatg 660
gtggaaatatg aaaaatgttgc ccgagaatgt cggtttggca gcaaggaaagg tcgacccttc 720
tctgggttca ctgtgtgttgc ggatgttgcgt gtcatecccc acaggacat tcacaacatg 780
aataatgttca gactgtgttgc ttgtatccaa actcgagaag ataaccgttc tttgggtgtt 840
attcctcaag atgacgacgtt ccgtatgtca ctctttataa atgcttcaga cacatgtatg 900
tttggcttca aggaaggaat ggaagccaaatgatccaaatgttgc gggccatcga ggtcttgca 960
ccccccggca aaaaaaaagc gtgtttcaatc cagcgttgcgttcccccggccatcga ggtcttgca 1020
gtgtcgatgtatgttgc ttcacatgttgc ttcacatgttgc ttcacatgttgc 1080
ccccggatca aacggaaatggatgtatgttgc ttcacatgttgc ttcacatgttgc 1140
ccaacccatggatgtatgttgc ttcacatgttgc ttcacatgttgc ttcacatgttgc 1200
cattttatct taaaatgttgc agacaaatgttgc ttcacatgttgc ttcacatgttgc 1260
caccggatgtatgttgc ttcacatgttgc ttcacatgttgc ttcacatgttgc 1320
ccagcttccatgtatgttgc ttcacatgttgc ttcacatgttgc ttcacatgttgc 1380
ccccactgtatgttgc ttcacatgttgc ttcacatgttgc ttcacatgttgc 1440
cttggcattt ctttttttttgc ttcacatgttgc ttcacatgttgc ttcacatgttgc 1500
gagcccccattt ttttttttttgc ttcacatgttgc ttcacatgttgc ttcacatgttgc 1560
ccaaaccatccatgtatgttgc ttcacatgttgc ttcacatgttgc ttcacatgttgc 1620
gaagatgatgtatgttgc ttcacatgttgc ttcacatgttgc ttcacatgttgc 1680
ccccctgtatgttgc ttcacatgttgc ttcacatgttgc ttcacatgttgc 1740
cacatgtatgttgc ttcacatgttgc ttcacatgttgc ttcacatgttgc 1800
ttgttgcgtatgttgc ttcacatgttgc ttcacatgttgc ttcacatgttgc 1860
aatcatccatgtatgttgc ttcacatgttgc ttcacatgttgc ttcacatgttgc 1920
catgggttttgc ttcacatgttgc ttcacatgttgc ttcacatgttgc 1980
aaggccatccatgtatgttgc ttcacatgttgc ttcacatgttgc ttcacatgttgc 2040
aatgtatgtatgttgc ttcacatgttgc ttcacatgttgc ttcacatgttgc 2100
accgtgtccatgtatgttgc ttcacatgttgc ttcacatgttgc ttcacatgttgc 2154

```

SEQ ID NO: 53 moltype = AA length = 378
FEATURE Location/Qualifiers
source 1..378
mol_type = protein
organism = synthetic construct

SEQUENCE: 53

```

DALDDDFDLDLM LGSDALDDFD LDMLGSDALD DFDFDMLGLSD ALDDDFDLDLM GSLPSASVEF 60
EGSGGGPSPQI SNQALALAPS SAPVLAQTMV PSSAMVPLAQ PPAPAPVLTP GPPQSLSAPV 120
PKSTQAGEGT LSEALLHLQF DADEDLGALL GNSTDPGVFT DLASVDNSEF QQLLNQGVSM 180
SHSTAEPMLM EYPEAITLEV TGSQRPDPDA PTPLGTSGLP NGLSGDEDFS SIADMDSAL 240
LSQISSSSQG GGGSGFSVDT SALLDLFSPS VTVPDMSLPD LDSSLASIQE LLSPPEPPRP 300
PEAENSSPDS GKQLVHYTAQ PLFLDDPGSV DTGSNDLPLV FELGEGRSYFS EGDGFAEDPT 360
ISLLTGSEPP KAKDPTVS 378

```

SEQ ID NO: 54 moltype = DNA length = 1134
FEATURE Location/Qualifiers
source 1..1134
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 54

```

gtatgccttag acgtatgttgc ttagatgtatgttgc ttcgttgcgttgc acgtatgttgc ttcgttgcgttgc 60
cttagatgttgc ttcgttgcgttgc acgtatgttgc ttcgttgcgttgc ttcgttgcgttgc 120
ggcccttagatgttgc ttcgttgcgttgc acgtatgttgc ttcgttgcgttgc ttcgttgcgttgc 180
gaaggccatccatgtatgttgc ttcgttgcgttgc acgtatgttgc ttcgttgcgttgc ttcgttgcgttgc 240
tccgtcttgcgttgc ttcgttgcgttgc acgtatgttgc ttcgttgcgttgc ttcgttgcgttgc 300
ccacccatgtatgttgc ttcgttgcgttgc acgtatgttgc ttcgttgcgttgc ttcgttgcgttgc 360
cccaatgtatgttgc ttcgttgcgttgc acgtatgttgc ttcgttgcgttgc ttcgttgcgttgc 420
gacgtgtatgtatgttgc ttcgttgcgttgc acgtatgttgc ttcgttgcgttgc ttcgttgcgttgc 480
gatctggatgtatgttgc ttcgttgcgttgc acgtatgttgc ttcgttgcgttgc ttcgttgcgttgc 540
tctcatatgtatgttgc ttcgttgcgttgc acgtatgttgc ttcgttgcgttgc ttcgttgcgttgc 600
accggccatccatgtatgttgc ttcgttgcgttgc acgtatgttgc ttcgttgcgttgc ttcgttgcgttgc 660
aatgggtgtatgtatgttgc ttcgttgcgttgc acgtatgttgc ttcgttgcgttgc ttcgttgcgttgc 720
ctgttcacatgtatgttgc ttcgttgcgttgc acgtatgttgc ttcgttgcgttgc ttcgttgcgttgc 780
agtgccatgtatgttgc ttcgttgcgttgc acgtatgttgc ttcgttgcgttgc ttcgttgcgttgc 840
cttgacatgtatgttgc ttcgttgcgttgc acgtatgttgc ttcgttgcgttgc ttcgttgcgttgc 900
cccgaggccatccatgtatgttgc ttcgttgcgttgc acgtatgttgc ttcgttgcgttgc ttcgttgcgttgc 960
ccgctgttccatgtatgttgc ttcgttgcgttgc acgtatgttgc ttcgttgcgttgc ttcgttgcgttgc 1020
tttggatgtatgttgc ttcgttgcgttgc acgtatgttgc ttcgttgcgttgc ttcgttgcgttgc 1080

```

-continued

-continued

gaggaagaag	tgcgtatgt	agctgggtgga	gagcagcctg	gaaaaatctt	ctaccgggc	240
aaaggggccc	agtgggggtgg	aaccaggcg	ctgttcgggg	agtgaacct	cctccagtga	300
gaagaagaag	gtatccaaag	ccccagcad	tctctgtccca	cccagccag	ccccagcccc	360
tggactcacc	aacgtgtga	agaagagtaa	acagccactt	caggtgacca	aggatctggg	420
ccgcgttggaa	cctcgatgt	acccctgtct	cataaatgtt	gtgttgacaa	ccaacgaccc	480
gacccctggc	cacccctggcg	tgaaaattcag	ctggccgttt	acccttcggg	aggccaggaa	540
gctgttgtac	gccccgtct	acgatccctt	catctccaaag	ttggccctgtc	aggccatgg	600
gcagctgcac	ccagaggcta	ttgcgcattt	ccagagaacg	gccctgttta	gcaaggctga	660
ggagcagctg	ctgagccaaag	tgggatcgac	cagccagccc	accttggaga	ccttccagga	720
cctgtctgcac	agacacccctg	atgccttca	cctggccctgt	accgcgaagg	cccttgcaggc	780
caactggcag	ctcatgtaa	agttatccat	gttggaggaa	cagacatgtc	agccgtgtcc	840
caaaaggggaa	caatgtctga	actttctgt	tgcagaggac	ctgattgtat	acagtaagct	900
caaggacatg	cgagatgagg	tccttggaaaca	tgagctgtat	gtggctgacc	ggccggccagaa	960
gcgagagatt	ccgcagctgg	aacaggaaat	gcataatgtt	cagggtgtat	ttggacagcat	1020
cacaggatgt	agcttccgg	atcttcgttt	ccagacactt	gcagtgatgc	ggggccgcat	1080
ggttgcggta	ctgtatgtat	cgccgtgtat	caccctggcc	agagcaacca	aggataacca	1140
gattgtatgt	gacccgttctc	tgagggttcc	ggccctggaa	atatcccgaa	aacaagggtgt	1200
catcaagctg	aagaacaacg	gtgttttctt	catttgcata	gagggtgtac	ggcccatctt	1260
catcgatgg	ccgcgggtgc	tctgtgttcc	caaattggcg	ctcagcaaca	actctgttgg	1320
ggagatcgcc	accctcgat	tcgttttctt	tatccaaatgg	gaccatccat	cccttcatcg	1380
ggctgaggct	gcctaaatgt	caccacatgtt	ggaccatgtt	ttcttgcata		1429

SEQ ID NO: 59 moltype = AA length = 850
 FEATURE Location/Qualifiers
 source 1..850
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 59

MTLDMDAVL	DFVRSTGAEP	GLARDLLEGK	NWDVNAALSD	FEQLRQVHAG	NLPPSFSEGS	60
GGSRTPEKGF	SDREPTRPPR	PILQRQDDIV	QEKRRLSRGIS	HASSSIVSLA	RSHVSSNNGG	120
GGSNEHPLEM	PICAFQLPDL	TVYNEDFRSF	IERDLIEQSM	LVALEQAGR	NWWVSVDPTS	180
QRLPLPLATT	DGNCLLHAAS	LGMWGFHDRD	LMLRKALYAL	MEKGVEKEAL	KRRWRWQQTQ	240
QNKESEGLVYT	EDEWQKEWNE	LIKLIASSEPR	MHLGTTNGANC	GGVESSEEPV	YESLEEFHVF	300
VLAHVLRPPI	VVVADTMRL	SGGEFAPAPI	FGGIYLPLEV	PASQCHRSP	VLAYDQAHFS	360
ALVSMEQKEN	TKEQAVIPLT	DSEYKLLPLH	FAVDPGKGEW	WGKDDDSDNVR	LASVILSLEV	420
KLHLHLHSYMN	VKWIPLSSDA	QAPLAQPEP	TASAGDEPRS	TPESGDSDKE	SVGSSSTSNE	480
GGRRKEKSKR	DREKDKKRAD	SVANKLGSGF	KTLSKSLKKN	MGGLMHSKGS	KPGVGVTGLG	540
GSSGTETLEK	KKKNSLKSWK	GGKEEAAGDG	PVSEKPPAES	VGNNGSKYSQ	EVMQSLSIIR	600
TAMQGEKFI	FVGTTLKMGHR	HQYSEEMIQR	YLSDAEERFL	AEQKQKEAER	KIMNGGIGGG	660
PPPAKKPEPD	AREEQPTGP	AESRAMAFST	GYPQDFTIPIR	PSGGGVHQCQE	PRRQLLAGGPC	720
VGGLPPYATF	PRQCPGGRPY	PHQDSIPSLE	PGSHSKDGLH	RGALLPPPYR	VADSYNSNGYR	780
EPPEPDGWAG	GLRGLPPTQT	KCKQPNCSFY	GHPETNNFC	CCYREELRRR	EREPEPDGELLV	840
HRFLDPAFLY						850

SEQ ID NO: 60 moltype = DNA length = 2572
 FEATURE Location/Qualifiers
 source 1..2572
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 60

gtacaaaaaa	gcaggatcca	ccatgaccct	ggacatggat	gttgcgttgc	cagattttgt	60
ccgttccaca	ggaggcagac	cagggttagc	gcgcagatctc	ctagaaggaa	agaattggaa	120
tgtgtatgcc	gccctcaatgt	attttgcata	gttacgttca	gtccatgtct	gaaacctacc	180
cccatccctt	agtggggaa	gttggggctc	caggaccctt	aaaaaagggt	tttctgtacag	240
agagcctact	ccccccctccc	gaccatctt	ccagcggctt	gatgcacatcg	ttcaagaaaa	300
acgcctgtt	aggggcattct	ccccacccat	ctccagccat	gtttccctgg	cccggttcca	360
tgtctccccc	aatggtgggg	gtggggggag	caatggacac	ccccctggaaa	tgcccatctg	420
tgccttccat	ttccagatc	tcactgtata	caatgaagac	ttccgcagat	tcatagagag	480
agacccattt	gaggcgttca	tgctgggttc	cttggaaacag	gcagggccgtt	tgaactgtgt	540
ggtgagttgt	gttccatccat	cttcagggat	gtttcccttt	gcaactactg	gagatgggaa	600
ctgccttccat	catgcgttca	cccttggat	gtggggtttc	catgttggg	acttgcgtt	660
gcccggaaatgt	ttgtatgcac	tgatgggaa	ggggatgttgc	aaaggatgttgc	tgaaaaggcg	720
ctggaggatgg	caggcgttcc	agcgttccat	ttccatgttca	ttccatgttca	cagaagatgt	780
atggcgttcc	gaggatgttgc	ttccatgttca	ttccatgttca	ttccatgttca	ttccatgttca	840
aggttccat	ggggatgttgc	ttccatgttca	ttccatgttca	ttccatgttca	ttccatgttca	900
ccttgcgttcc	tttcacgtt	tttgcgttca	tttgcgttca	tttgcgttca	tttgcgttca	960
ggcgttccat	tttcacgtt	tttgcgttca	tttgcgttca	tttgcgttca	tttgcgttca	1020
aatcttatgt	cctttggagg	ttccatgttca	ttccatgttca	ttccatgttca	ttccatgttca	1080
ctatgtatcg	gccccctttt	ttccatgttca	ttccatgttca	ttccatgttca	ttccatgttca	1140
acaagctgtt	atccccat	ttccatgttca	ttccatgttca	ttccatgttca	ttccatgttca	1200
ggacccttgg	aaaggcttgg	ttccatgttca	ttccatgttca	ttccatgttca	ttccatgttca	1260
tgtatattgt	tcccttagagg	ttccatgttca	ttccatgttca	ttccatgttca	ttccatgttca	1320
gatccccactt	ttccatgttca	ttccatgttca	ttccatgttca	ttccatgttca	ttccatgttca	1380
agctggatgt	gagccccctt	ttccatgttca	ttccatgttca	ttccatgttca	ttccatgttca	1440
cacgcgttcc	accaccaac	ttccatgttca	ttccatgttca	ttccatgttca	ttccatgttca	1500
gaaggacaac	aaggagac	ttccatgttca	ttccatgttca	ttccatgttca	ttccatgttca	1560

-continued

ggcagcaga	ctcaagaaga	acatgggggg	cctgatgcac	agcaagggtt	caaagctgg	1620
agggggtggg	acagggtgg	gagaaagcag	cggcaactgag	acactggaga	agaagaagaa	1680
aaactca	ctg aagatcg	gggggtggca	ggaggaggca	gctggggatg	ggccctgtgtc	1740
tgagaagccc	ccagctgagt	ctgtttgtaa	cggaggggagc	aagtatagcc	aggaggtgat	1800
gcaaagctg	agcattctga	ggactgccc	gcaaggggg	ggaaagtta	tttttgttgg	1860
aaccctgaag	atgggtcacc	gtcaccatg	tcaggaggaa	atgateccagc	gtctaccttc	1920
tgatgctgag	gagagattcc	tgccagaaca	gaagcagaag	gaggcagaga	ggaagatcat	1980
gaatggagga	atagggggtg	gccttcctcc	agccaaaaaa	ccagagccag	atgttaggg	2040
agagcagccg	accggcccc	cageagagtc	caggcaatg	gcattttcca	ctggctaccc	2100
tggggactt	atactccctc	ggccgtctgg	ggggcggagg	cactgcagg	aaccccgag	2160
gcagttggca	gggggttccat	gtgtgggggg	cattaccacca	tatgcacact	tcccccacaa	2220
gtgcccctct	ggggcaccc	accccccacca	ggacagcatc	ccttcttgg	agccaggcag	2280
ccactctaag	gatggacttc	acaggggtgc	cttggtaacc	cccccttacc	gagttggctga	2340
ttctatagc	aatggctaca	gagggccccc	tgagccagat	ggatggggctg	gagggtctccg	2400
gggccttccc	ccaactcaga	ccaaatgc	acaacccgaa	tgcagettct	atggacaccc	2460
ttagacaaac	aactctgtt	cctgtgtt	caggaaqaa	ctgaggagg	gggagcggga	2520
accggatggg	gagctctgg	tgcacagg	ttggacccca	gttttcttgt	ac	2572

SEQ ID NO: 61 moltype = AA length = 382

FEATURE Location/Qualifiers
source 1..382
mol_type = protein
organism = synthetic construct

SEQUENCE: 61

MLDRDVGPTP	MYPPTYLEPG	IGRHTPYGNQ	TDYRIFELNK	RLQNWTTECD	NLWWDAFTTE	60
FFEDDAMLT	TFCLLEDGPKR	YTIGRTLIPR	YFRSIFEGGA	TELYYYVLKHP	KEAFHSNFVS	120
LDCDQGSMVT	QHGKPMFTQV	CVEGRYLIEF	MFDDMMRIKT	WHFSIRQHRE	LIPRSILAMH	180
AQDPQMLDQL	SKNITRCGLS	NSTLNLYRLRC	VILEPMQELM	SRHKTYSLSP	RDCLKTCLFQ	240
KWQRMVAPP	EPTRQQPSKR	RKRKMSGST	MSSGGGNTNN	SNSKKSPAS	TFALSSQVPD	300
VMVGEPTLM	GGEFGDEDER	LITRLENTQF	DAANGIDDED	SPNNNSPALGA	NSPWNSKPPS	360
QESKSENPT	SQASQLDPAF	LY				382

SEQ ID NO: 62 moltype = DNA length = 1146

FEATURE Location/Qualifiers
source 1..1146
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 62

atgctggata	gggatgtggg	cccaactccc	atgtatccgc	ctacataact	ggagccagg	60
attggggaggc	acacacata	tggcaaccaa	actgactaca	aatatttga	gcttaacaaa	120
cggttcaga	actggacaga	ggagtgtac	aatctctgg	ggatgcatt	cacgactgag	180
ttctttgagg	atgatgccat	gttgaccatc	actttctgc	tggaggatgg	accaaagaga	240
tataccattt	gcgggaccc	gatccaccc	tacttccgc	gcatcttga	gggggggtgt	300
acggaggtgt	atatgttct	taagcacccc	aaggaggcat	ccacagccaa	ctttgtgtcc	360
ctcgactgtg	accaggcag	catgttacc	cacgatggca	agcccatgtt	cacccagggt	420
tgtgtggagg	gcccgtgtt	cctggagtt	atgtttgacg	acatgtatgc	gataaagacg	480
tggcacttca	gatccggcgc	gcacccggag	ctcatcccc	gcagcatct	tgccatgtat	540
gccccaa	cccaatgtt	ggatacc	tccaaaaaaa	tcactcggt	tgggtgttcc	600
aattccact	tcaactaccc	ccgactctgt	gtgtatctcg	agcccatgca	agagctcatg	660
tcacggccaca	agacctacag	cctcagcccc	cgcgactgc	tcaagacctg	ccttttccag	720
aagtggcagc	gatgttgc	acccctcg	gagcccccac	gtcagcagcc	cagcaaacgg	780
cgaaaaacgg	agatgttgc	gggcaccc	atgactctg	gtgggtggca	caccaaaac	840
agcaacacg	agaagaagag	cccaatgt	acccatcgcc	tcccaacgg	ggtagctgt	900
gtgtatgtgg	tgggggagcc	caccctgtat	ggggggagg	tccggggacga	ggacgagagg	960
ctcatcaccc	ggctggagaa	cacccatgtt	gacgcagcca	acggcatgtt	cgacgaggac	1020
agctttaaca	actccctgc	actggggcc	aacagccccc	gaaacagcaa	gcctccgtcc	1080
agccaaagaaa	gcaaatcgga	gaacccac	tcacagg	cccaatgttgc	cccagtttcc	1140
ttgtac						1146

SEQ ID NO: 63 moltype = AA length = 356

FEATURE Location/Qualifiers
source 1..356
mol_type = protein
organism = synthetic construct

SEQUENCE: 63

MGAVACLGKA	ADADEWCDSG	LGSGLPDAAA	PGGPGGLGAEL	GPGLSWAPLV	FGYVTEDGDT	60
ALHLAVIHQH	EPFLDFLLGF	SAGTEYMDLQ	NDLGQTALHL	AAILGETSTV	EKLYAAGAGL	120
CVAAERRGHTA	LHLACRVGAH	ACARALLQPR	PRRPREADT	YLAQGPDRTP	DTNHTPVALY	180
PDSDLEKEEE	ESEEDWKLQL	EAENYEGHTP	LHVAVIHKDV	EMVRILLRDAG	ADLDKPEPTC	240
GRSPPLHLLAVE	AQAADVLELL	LRAGANPAAR	MYGGRTPLGS	AMLRPNPILA	LLLRAHGAPE	300
PEGEDEKSGP	CSSSSDSDSG	DEGDEYDDIV	VHSSRSQTRL	PPTPASKPLP	DDPRPV	356

SEQ ID NO: 64 moltype = DNA length = 1071

FEATURE Location/Qualifiers
source 1..1071
mol_type = other DNA

-continued

```

SEQUENCE: 64          organism = synthetic construct
atggctgggg tcgcgtgctt gggaaaagct gccgacgcag atgaatggtg cgacagccgc 60
ctgggctccc tgggttcgga cgcgcggcc cccggaggac ctgggttggg cgccgaggatc 120
ggcccggggc tgcgtgggc tccccctgctc ttccggctacg tcactgagga tggggacacg 180
gcactgcaact tggctgtatc tcatacgat gaacccttc tggatttttc tcttaggttc 240
tcggccggca ctgagatcat gcgcacgtcg aatgacactg gccaacgacg cctgcacctg 300
gcagccatcc tgggggagac atccacggtg gagaagctgt acgcacgagg cgccgggtcg 360
tgtgtgggg agcgtgggg ccacacggcg ctgcacctgg cttccgtgt gggggcacac 420
gcctgtgccg gtgcctgtct tcacccggcc cccggggcgcc ccaggaaagc cccggacacc 480
tacccctgctc agggccgtcg cctgcacccgacaccaacc atacccctgt cgccctgtac 540
cccgattccg acttggagaa ggaagaagag gagagtgggg aggactggaa gctgcacgtg 600
gaggctgaaa actacgggg ccacacccca ctccacgtgg ccgttatcca caaagatgtg 660
gagatgttcg ccgtctccg agatgtctgg gctgacccatc acaaaccggg gcccacgtgc 720
ggccggagcc cccttcattt ggccgtgggg gcccaggcgac cctgcacgtgt ggagtttctc 780
cttggggccg ccgcgtcgcc tgcgtgggg atgtacccgtt gccgcaccc acctggcgat 840
gcccattgtcc gggccaaacc catccctggcc cgcctctcc gtcacacgg agccctgtgg 900
cccgaggggc aggacggagaa atccggcccc tgcacggcgac gtagegacacg cgacagccg 960
gacgaggggc atgaatacgac cgacattgtg gttcacaacg gccgcacca aaccgggtcg 1020
cctcccaacc cagccctaaaa accttcttc gacgacccccc gccccgtgtg a 1071

SEQ ID NO: 65          moltype = AA length = 579
FEATURE           Location/Qualifiers
source            1..579
mol_type = protein
organism = synthetic construct

SEQUENCE: 65
MLRSGPASGP SVPTGRAMPS RRVARPPAAP ELGALGSPDL SSLSLAVSRS TDELEIIDEY 60
IKENGPGFLDG GQPGPGEGLP RLVSRGAASL STVTLGPVPAT PPPWGCP LGRLVSPAPG 120
PGPQPHLVIT EQPKQRGMRF RYECEGRSAG SILGESSTEE SKTLPAIELR DCGGLREVEV 180
TACLVWKDWPLVHRVPHSLVGC KDCTDGICRVL RLRPHVSPRH SFNNLGIQCV RKKEIEAAIE 240
RKIQLGIDPY NAGSLKNHQE VDMNVVRICF QASYRDQQQ MRRMDPVLSPE PVYDKSTNT 300
SELRICRINK ESGPCTGGEE LYLLCDKVQK EDISVVFSRA SWEGRADFSQ ADVHROIAIV 360
FKTPPYEDLE IVEPVTVNLF LQLRTDGVCSE EPLPFTYLPDR DHDSYGVDDKK RKRGMPDVLG 420
ELNSSDPHGI ESKRRKKKPAA ILDHFLPNHG SGPFLLPPSAL LPDPDFFSGT VSLPGLLEPPG 480
GPDLDDDGFA YDPTAPLTFT MLDDLPPAPP HASAVVCSGG AGAVVGETPG PEPLTLDSDYQ 540
APGPGDGGTA SLVGSNMFPN HYREAAGGGG LLSPGPEAT 579

SEQ ID NO: 66          moltype = DNA length = 1740
FEATURE           Location/Qualifiers
source            1..1740
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 66
atgcttcgggt ctggggcage ctctggccg tccgtccccca ctggccgggc catgcggagt 60
cgccgcgtcg ccgacccgcg ggctgcgcgg gagctggggg ctttaggttc cccgcaccc 120
tccctactcg ccgtcgccgt ttccaggagc acagatgtatc tggagatcat cgacggatc 180
atcaaggaga acgggttcgg cctggacgggg ggacagccgg gcccggcga ggggtgtcca 240
cgctctgggtt ctgcgggggc tgctgtccctg agcacgggtca ccctggggcc tttggggccc 300
ccagccacgc cgccgcctt ggggtgtcccc ctggggccac tagtgtcccc agcgcggggc 360
ccggggccgc acggccaccc ggtcatccg gggcggccggc aacggccggg catgcgttcc 420
cgctacggat ggcggggccg ctggggccggc agcatccctt gggagggccgac caccggggc 480
agcaagacgc tgccggccat ccgcgtccgg gattgtgggg ggtctggggg ggtgggggtg 540
actgcctgcc tgggtgtggaa ggactggctt caccggatcc accccacacg cctctgtggg 600
aaagactcga ccgcggcat ctgcgggtt cggctccggc ctcacgttcg ccccccggcac 660
agtttaaca accttggggcat ccagtgtgtt aggaagaagg agattgggg tgccatttgag 720
cgaaagatcc aacttggggcat tgacccttac aacgcgtggg ccctgtggggaa ccatcaggaa 780
gtagacatga atgtgttgat gatctgttc caggcctcat atcgggacca gcaggacac 840
atgcgcggga tggatctgtt gcttcgtatcg acaagaatc cacaacacaa 900
tcagagtcgc ggtattggcg aattaacaag gaaaggccggg cgtgcacccgg tggcgaggag 960
ctctacttgc tctgtcgacaa ggttcgtggaa gggacatcat cagttgttgc cagcggggcc 1020
tcttggggaaat gtcgggttgc ctgttcccg gccgcgtgc accggccatg tgccatttg 1080
ttcaagacgc cgccctacga ggacccgtggg attgtcgacgc cctgtacatc aaacgtttc 1140
ctgcgcggcc tcacccatgg ggtctggacg gggccatcgat ctttcacgtt cctgcctccgc 1200
gaccatgaca gtcacccgtt ggacaaaggcg cggaaacccggg ggtatccggg cgtcccttggg 1260
gagctgtacca gtcctgcaccc ccatcgatcc gggacaaacg gggccggggaa aaaggccggcc 1320
atccctggacc acttcttgcg caaccacggc tcaggccgtt tccctccggc gtcagccctg 1380
ctggccagacc ctgacttctt ctctggcacc gttgtccgtc ccggccgtgg gcccccttggc 1440
ggccctgacc tcctggacg tggctttggc tacgaccccta cggcccccac actcttcacc 1500
atgtctggacc tgcgtggccccc ggcacccgcca cacgtcgatcg ctgttgcgttgc cagcggggat 1560
ggccggggccg tgggtggggaa gaccccccgc cctgaacccatc tgacactggat ctcgttccac 1620
ggccggggccg cggggatgg aggcacccgc agcctgtgg gcaacatgttcccaat 1680
cattaccgcg aggccggccctt tggggccggc ctccatccc cggggccgtga agccacgttag 1740

SEQ ID NO: 67          moltype = AA length = 277
FEATURE           Location/Qualifiers

```

-continued

```

source          1..277
               mol_type = protein
               organism = synthetic construct

SEQUENCE: 67
MADHMMMNH GRFPDGTNGL HHHPAHRMGM GQFPSPHHQ QQQPQHAFNA LMGEHIHYGA 60
GNNNATSGIR HAMGPGTVNG GHPPSALAPA ARPNNSQPMG PPVASQGSSL PASMQLQKLN 120
NQYFNHHPYP HNHYMPDLHP AAGHMNGTN QHFRDCNPKH SGSSTPGGS GGSSTPGGSG 180
SSGGGGAGSS NSGGGGSGGN MPASVAHVP AMLPPNVIDT DFIDEELVMS LVIEMGLDRI 240
KELPELWLQG NEFDFTDFV CKQQPSRVSC LDPAFLY 277

SEQ ID NO: 68      moltype = DNA length = 831
FEATURE
source          1..831
               mol_type = other DNA
               organism = synthetic construct

SEQUENCE: 68
atggcagacc atatgatggc aatgaaccac gggcgttcc cgcacggcac caatggctg 60
caccatcacc ctgcccaccc catggcatg gggcagtcc cgagccccca tcaccacag 120
cagcagcage cccagcacgc cttcaacgc ctaatggcgc agcacataca ctacggcg 180
ggcaacatga atgccccacg cggcatcagg catcgatgg ggcggggac tggtaacgg 240
ggccacccccc cgagcgcgcg ggccccccgg gccagggtta acaactcca gttcatgg 300
ccccccgtgg ccagccaggc aggtcccttg cccggccagc tgcaagtcga gaagtcac 360
aaccatgtt tcaaccatca ccccttaccc cacaaccat acatgcggc ttgcaccc 420
gtgtcaggccc accagatgaa cggccaaac cggcacttc gagattgcgg ccccaagcac 480
agggggggca gcagcacccc cggggggctcg ggcggccagc gcaccccccgg cggctctgg 540
agcagctgg gggccggcgc gggcagcgc aacagcggcgc gggcagcgg cagcggcaac 600
atgccccgctt cgttgtggca cgtccctgtc gcaatgtcg cggcccaatgt catagacact 660
gatttcatcg cccatggggaaatgggtt ggaccgcata 720
aaggagctgc cccaaactctg gctggggcaa aacggatgg attttatgac ggactctgt 780
tgcaaacagc agcccccaggc agtgagctgt ttggaccacg ctttcttgc 831

SEQ ID NO: 69      moltype = AA length = 1066
FEATURE
source          1..1066
               mol_type = protein
               organism = synthetic construct

SEQUENCE: 69
NLSSCFVLLIN KERSFEPPLI NKERMRCSMG PKKKRKVGGM GDPIVMTQSP SSLSASVGDR 60
VTITCRSTS VATTNSNYASW VQECPGKLFK GLIGGTTNRA PGVPSRFSGS LIGDKATLTI 120
SSLQPEDFAT YFCALWSNHN WFGQGKTVE LKRGGGGGSGG GGSEVKLLES 180
GGGLVQPGGS LKLSCAVSGF SLTDYGVNWV RQAPGRGLEW IGVIWGDGIT DYNALKDRF 240
IISKDNGKNT VYLQMSKVRS DDTALYYCVT GLEPDYWGQGT LTVVSSYPD VPDYAGGGGG 300
SGGGGGGGGG SGGGGSLDPG GGGSGSKGEE LFTGVVPILV ELDGDVNNGHK FSVRGELEGD 360
ATNGKLTLKF ICTTGKLPV WPTLVTLLTH GVQCFSRYPD HMKRHDFFKS AMPEGYVQER 420
TISFKDDGTY KTRAEVKFEV DTLVNRIELK GIDFKEDCNI LGHKLEYNFN SHNVYITADK 480
QKNGIKANFK IRHNVEDGSV QLADHYQONT PIGDGVPVLLP DHNHYLTQSV LSKDPNEKRD 540
HMLVLEFVTA AGITHGMDEL YKGGGRTGGG GSGGGGGADPK KKRKVARIITS LYKKAGSTM 600
KDSQGLLDSS LMASGTASRS EDEESLAGOK RASSQALGTI PKRRSSSRFI KRKKPDDDELV 660
ESSLAKSSTR AKGASGVEPG RCSGSEPPSS EKKKVKAPS TPVPPSPAPA PGLTKRKKVS 720
KQPLQVTKDL GRWPKPADDLL LINAVLQTNL LTSVHLGVKF SCRFTRLETFQ ERWYALLYDP 780
VISKLACQAM RQLHPEAIA1 IQSKALFSKA EQLLSSVKGS TSQPTLETTFQ DLLHRHPDAF 840
YLARTAKALO AHWQLMKQYY LLEDQTVQPL PKGDQVLNFS DAEDLIDDSK LKDMRDEVLE 900
HELMVADRRQ KREIROLEQK LHKWQVLVDS ITGMSSPFD NQTLAVLRGR MVRYLMRSRE 960
ITLGGRATKDN QIDVDSLLEG PAWKISRKQG VIKLKNNGDF FIANEGRPRI YIDGRPVLCG 1020
SKWRLSNNSV VEIASLRFVF LINQDLIALI RAEAAKITPQ LDPAFL 1066

SEQ ID NO: 70      moltype = DNA length = 3139
FEATURE
source          1..3139
               mol_type = other DNA
               organism = synthetic construct

SEQUENCE: 70
atgggtccca agaaaaagag aaaggatcggtt ggcacatggcc ccgacatcgat gatgaccac 60
agccccacgc gctgtggcgc cggcggtggc gacccatcgat ccgtacccgtt ccgcacgc 120
acccggcgcg tgaccacccg caactacgcg agctgggtgc aggagaagcc cggcaagctg 180
ttcaaggccc ttatcgccgg cacaacaaac cggcccccgg cggcttccggc 240
ggcagcgtga tcggcgacaa ggccacccctg accatcgacg ggcgtccggc cggggacttc 300
gccacccatct tctcgccctt gtggatcgac aaccactggg tggttccggca gggccaccaag 360
gtggatcgatg acgcggcgcc cgggttggaa gggaggccgtt ggttctgggg aggccggc 420
tctggccggc gcaagcgatg gaagctgtcg gggaggccgtt gggccctgtt gcagccggc 480
ggcagcgtga agctgtcgat cggccgttgc ggttccggc tgaccgtacta cggccgttgc 540
tgggttgcggc aggccccccgg cggccggccgtt gatggatcg gggccgttgc gggccacccg 600
atcaccgact acaacacgcgc cctgttccatca tcagcaagga caacggcaag 660
aacaccgtgtt acctgtcgat gggccatggc cggccgttgc acaccgttgc gtttactgtc 720
gtgaccggcc tggatcgacta ctggggccggc gggccatggc tgaccgttgc cggccgttgc 780
tacgtatgttc cggatcgatcgc tggatcgacta gggccatggc tgaccgttgc gggccgttgc 840

```

-continued

ggtggttcag	gaggcggcgg	aagcttggat	ccaggtggag	gtggaaaggcg	tagcaaagga	900
gaagaacctt	tcactggagt	tgtcccaatt	cttggtaaat	tagatggta	tgttaatggg	960
cacaattttt	ctgtccgtgg	agaggggtgaa	ggtgatgcta	caaacggaaa	actcacccctt	1020
aaattttttt	gcaactactgg	aaaactacct	gttccgtggc	caacacttgt	cactactctg	1080
accatgggt	ttcaatgttt	ttcccgttat	ccggatcaca	tgaaacggca	tgacttttc	1140
aagagtgc	tgcccgaaagg	ttatgtacag	gaacgcacta	tatcttcaa	agatgacggg	1200
acatcacaaga	cgctgtgtga	agtcaagttt	gaaggtttaa	tcgtatcgag	1260	
ttaaagggtt	ttgatTTAA	agaagatgaa	aacattcttg	gacacaact	cgagtacaa	1320
ttaactcac	acaatgtata	catcacggca	gacaaacaaa	agaatggaa	caaagcta	1380
tccaaatatt	gcacaaacgtt	tgaagatgtt	tccgttcaac	tagcagacca	ttatcaacaa	1440
aataactcaa	ttggcgatgg	ccctgttcc	ttaccagaca	accattac	gtcgacacaa	1500
tctgtccccc	cgaaagatcc	caacggaaag	cgtgaccaca	tggtccttct	tgagtggta	1560
actgtgtctg	ggattacaca	tggeatggat	gagctctaca	aaagggtggg	tcggaccgg	1620
ggcggtggca	gggggtgggg	cgggtctgtac	cccaagaaga	agaggaaagg	ggcttaggatc	1680
acaaggtttgt	aaaaaaagc	aggetccacc	gtacaaaaaa	gcgatgtcca	ccatggacaa	1740
agattctcg	gggtgtcgat	attatcccc	gtatggatca	ggcactgtca	gcccgtcaga	1800
ggatgaggag	tcactggcag	ggcagaageg	agectctcc	caggccttgg	gcaccatccc	1860
taaacggaga	actcttccca	gggtcatca	gagggaaag	ttcgtatgt	agctgggtga	1920
gagcagcctg	gcaaaaattt	ctaccgggg	aaaggggggc	agttgggtgg	aaccaggggcg	1980
ctgttccggg	agtgaatccct	cctccagtga	gagaagaaag	gtatccaag	ccccccagcac	2040
tcctgtgtca	cccagcccc	ccccagcccc	ttgactcacc	aagcgtgtga	agaagagata	2100
acagccactt	cagggtgacca	aggatctggg	ccgctggaa	cctgcagatg	acctccctgt	2160
cataaaatgt	tggttgacga	ccaaacgcac	gacccctcg	cacccctggcg	tgaatttcag	2220
ctgcccgttc	acccttccggg	aggtcaggaa	ggcttggat	gcccgtct	acgatctgt	2280
catctccaa	ttggccgtgtc	aggccatgag	gcagctgcac	ccagaggcta	ttgcagccat	2340
ccagagcaag	gccctgttta	gcaaggctga	ggagcagctg	ctgagcaaaag	ttggatcgac	2400
cagccagcccc	accttggaga	ccttccaggaa	cctgtgcac	agacacccgt	atgccttcta	2460
cctggcccg	accgcgaagg	ccctgcagg	ccactggcg	ctcatgaagc	agtattacct	2520
gtcgaggag	cagacagtgc	agccgctgc	caaagggggc	caagtgtca	acttctctga	2580
tgcagaggac	ctgattgtat	acagtaagct	caaggacat	cgagatgagg	tccttggaaaca	2640
tgagctgtat	tggtgtgaccc	ggcgcacagaa	gctgagatgt	cgccgcgtgg	aacaggaaact	2700
gcataatgtt	cagggtgtat	tggacagat	caacaggat	actctccgg	acttcgacaa	2760
ccagacactg	gcagtgctgc	ggggcgcac	gggtgcgtac	ctgatgcgt	cgctgtgagat	2820
caccctggc	agagcaacca	aggataacca	gattgtatgt	gaccctgtctc	tggagggtcc	2880
ggccttggaa	atatcccgaa	aacaaggctt	catcaagctt	aagaacaacg	gtgatTTCTT	2940
cattggcaat	gagggtcgac	ggcccatctt	catcgatgga	cgcccggtgc	tctgtggctc	3000
caaatggcgc	cttcagcaaca	actctgtgtt	ggagatgcgc	agcctgcgtat	tcgttccct	3060
tatcaaccag	gaccttattt	ccctcatcg	ggctgaggct	gccaagatca	caccacagtt	3120
ggacccagct	ttctttgtac					3139

SEQ ID NO: 71	moltype = AA	length = 1423			
FEATURE	Location/Qualifiers				
source	1..1423				
	mol_type = protein				
	organism = synthetic construct				
SEQUENCE: 71					
MGPKKKRKVG MGPDPDIVMTQ	SPSSLSASVG	DRVTITCRSS	TGAVTTSNYA	SWVQEKPGL	60
FKGLIGGTNN RAPGVPSRFS	GSLIGDKATL	TISSLQPEDF	ATYFCALWYS	NHWVPGQGTK	120
VELKRGGGGS CGGGSGGGGS	SGGGSEVKLL	ESGGGLVQPG	GSLKLSCAVS	GFSLTIDYGVN	180
WVRQAPGRGL EWIGVIWGDG	ITDYNALSALKD	RFIISKDNGK	NTVYLQMSKV	RSDDTALYYC	240
VTGLFDYWQG CTLVTVSSYP	YDVPDYAGGG	GGSGGGGGGG	GGSGGGGGSLD	PGGGGGSGSKG	300
EELFTGVVPI LVELDGDVNG	KHFSVRLGE	EDATNGKLTL	KFICTTGKLP	VPWPTLVTL	360
THGVQCFRSY PDHMKRHDF	KSAMPEGVQ	ERTISFKDDG	TYKTRAEVVK	EGDTLVNR	420
LKGIDFKEDG NILGHKLEYN	FNSHNVYITA	DKQKNGIKAN	FKIRHNVEDG	SQVLADHYQQ	480
NTPIGDPVLP LPDNHYLSTQ	SVLSKDPNEK	RDHMVLLLEFV	TAAGITHGMD	ELYKGGRRTG	540
GGGSGGGGAD PKKKRKVARI	TSLYKAGEST	MTLDMDAVLS	DFVRSTGAEP	GLARDLLEKG	600
NWDVNALALSD FEQLRQVHAG	NLPPSFSEGS	GGSRTPKEGF	SDREPRTRPPR	PILQRQDDIV	660
QEKRLSRGIS HASSSIVSLA	RSHVSSNGGG	GGSNEHPLEM	PICAFQLPDL	TVYNEDFRSF	720
IERDPLQSM LVALEQAGRL	NWWVSDPNTS	QRLLPLATTG	DGNCLLHAAS	LGMNGFHDRD	780
LMLRKALYAL MEKGVEKEAL	KRRWRWQQTQ	QNKESGGLVYT	EDEWQKEWNE	LIKLASSEPR	840
MHLGTTNGANC GGVESSEEPV	YESLEEFHFV	VLAHVLRPPI	VVADTMRLD	SGGEAFAPIP	900
FGGIYLPLEV PASQCHRSP	VLAYDQAHFS	ALVSMEQKEN	TKEQAVIPLT	DSEYKLLPLH	960
FAVDPGKWE WGKDDSDNVR	LASVILSLEV	KLHLLHSYMN	VKWIPLSSDA	QAPLAQPESP	1020
TASAGDEPRS TPESGDSKTE	SVGSSSTSNE	GGGRKEKSRA	DREKDKKRAD	SVANKLGSGF	1080
KTLGSKLKKN MGGLMHSKGS	KPGGVGTGLG	GSSGTETLEK	KKKNLSKSWK	GGKEEAAGDG	1140
PVSEKPPAES VNGGSKYSQ	EVMQSLSILR	TAMQGEKGFI	FVGTLKMGRH	HQYQEMIQR	1200
YLSDAEERFL AEQKQKEAER	KIMNGGIGGG	PPPAKKPEPD	AREEQPTGPP	AESRAMAFST	1260
GYPGDFITPR PSGGGVHCQE	PRRQLAGGP	VGGLPPYATF	PRQCPPGRPY	PHQDSIPSLE	1320
PGSHSKDGLH RGALLPPPYR	VADSYNSNGYR	EPPEPDGWAG	GLRGLPPTQT	KCKQPNCFSY	1380
GHPETNNFCS CCYREELRRR	EREPEPDGELLV	HRFLDPAFLY	KVV		1423

SEQ ID NO: 72	moltype = DNA	length = 4282
FEATURE	Location/Qualifiers	
source	1..4282	
	mol_type = other DNA	
	organism = synthetic construct	

-continued

SEQUENCE: 72

```

atgggtccca agaaaaagag aaaggtcggtt ggcattggcc cgcacatcgatgaccagg 60
agccccagca gcctgagcgc cagctgggc gaccgcgtga ccataccctgcgcgcgc 120
acccggcgcgc tgaccaccag caactacgc agctgggtc aggagaaggcc cgcaagctg 180
ttcaaggccc tgatcgccgg cacaacaac cgcgcggcccg cgctgccccag ccgcttcagc 240
ggcagcgtga tccggaccaa ggccaccctgt accatcagca gcctgcggcc cgaggactt 300
gccaccaact tctggccct gtggtacago aaccactgggg tggctggcca gggcacaagg 360
gttgagctga agcggcgccg cgggtgaagc ggaggcggtt gggtctgggg aggccggcgc 420
tctggccgg gcaagcgaggta gaagctgtcg gagagcgccg gggccctgtt gcagccggc 480
ggcagcgtga acgtcgactg cgcgcgtggc ggcgtcaaggc tgaccgacta ccgcgtgaac 540
tgggtgcggc aggacccggc cggcggtt ggtggatcg ggcgtatctt gggcgacggc 600
atcaccgact acaacacgcgc cctaaaggac cgttcatca tcagcaaggaa caacggcaag 660
aacaccgtgt acctgcagat gagaagggtt cgcagcgacg acaccgcct gtactactgc 720
gtgaccggcc tggggacta ctggggccca ggccaccctgt tgaccgttgc cagctaccca 780
tacgatgttc cagattacgc tgggtggggc ggagggttgc ttggggaggagg tagtggccgt 840
ggtgtgttag gaggcgccgg aagttggat ccagggtggg ggtggaaaggc tagcaaaggaa 900
gaagaacttt tcaactggagt tggcccaattt cttgttgaat tagatggta tggtaatggg 960
cacaatattt ctgtccgtgg agagggttgc ggtgtatcta aaaaaggaaa actcaccctt 1020
aaattttttt gcaactactgg aaaactaccc tttccgtggc caacacttgc cactactctg 1080
acccatgttgc ttcaatgttcc cccgtatca tcgaaacggca tgactttttc 1140
aagagtgcga tgcccgaaagg ttatgtacag gaaacgcacta tatcttccaa agatgacggg 1200
acotacaaaga cgcgtgtga agtcaagggtt gaaagggttgc cccttgcattt tcgtatcgag 1260
ttaaagggtt tttttttttt aagaagatggaa aacatttttgc gacacaacttgc gggatcaac 1320
ttaacttcac acaatgttgc catacggca gacaaacaaa aagaatggaaat caaaggctaac 1380
ttcaaaattt gccacaacgt tggatgtgtt tccgttcaac tagcagacca ttatcaacaa 1440
aataactccaa ttggcgatgg ccctgttccattt ttaccagaca accattactt gtcgacacaa 1500
tctgttccattt cggaaaggatcc caacggaaaagg cgtgaccacca tgggttctt tggatgttgc 1560
actgctgttgc tgattacaca tggatggat gggatctaca aagggtgggg tcggaccgg 1620
ggcggtggca ggggtgggg cgggtgtac cccaaaggaaa agaggaaagggtt ggctggatgc 1680
acaagtttgtt acaaaaaaaaggc aggttccaccat ttttttttttgc gtcggccca ccatgaccct 1740
ggacatggat gctgttgc ttttttttttgc cggatggatccca gggacggacgc cagggttgc 1800
ggcggatgttcc ctggaaaggaa aaaaagggttgc tggatgttgc cccttgcattt attttgc 1860
gtcacgttca gtcctatgttcc gaaaccttccccc cccatccccc cccatccccc 1920
caggaccctt gaaaaagggtt tttctgttgc agagccatctt cgcgcctccccc gaccctatcc 1980
ccacggccatgttccatgttcc gttcaaggaaa acggccctgttgc agggccatctt cccacggccatgttcc 2040
ctccacggatcc ttttttttttgc tggatgttgc ttttttttttgc aatgggggggg 2100
caatgagoc cccctggaaa tggccatcttgc tggatgttgc cttccatgttccatgttca 2160
caatgaagac ttccgcagttt tccatgttgc agacccatgttcc tggatgttgc tggatgttgc 2220
cttggaaacag gcaaggccgtt tggatgttgc tggatgttgc ttttttttttgc aatgggggggg 2280
gtttcccttgc gcaactactgttcc gggatgttgc tggatgttgc ttttttttttgc aatgggggggg 2340
gttgggttttccatgttgc ttttttttttgc tggatgttgc ttttttttttgc aatgggggggg 2400
ggggatgttgc aaggaaaggcgtt tggatgttgc tggatgttgc ttttttttttgc aatgggggggg 2460
agagtcaaggatcc tggatgttgc tggatgttgc tggatgttgc ttttttttttgc aatgggggggg 2520
gtttccatgttca agtggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 2580
ggaggttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 2640
tcatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 2700
ggaaatgttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 2760
ceagtgatccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 2820
gttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 2880
gtataagcttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 2940
agatgtatgttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 3000
tctgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 3060
tctggccatgttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 3120
gttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 3180
ggcggtggatgttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 3240
taacaaactgttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 3300
cctgtatgttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 3360
cggtggatgttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 3420
ggggatgttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 3480
cggtggatgttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 3540
gcaaggccatgttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 3600
tcaggatgttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 3660
gaagcagaagatccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 3720
agccaaatgttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 3780
caggccatgttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 3840
ggggatgttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 3900
cctaccatgttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 3960
ggacacgtatccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 4020
cttggatgttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 4080
tgagccatgttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 4140
acaaacccatgttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 4200
caggatgttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 4260
cttggatgttccatgttca tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc 4282

```

-continued

-continued

SEQ ID NO: 75 moltype = AA length = 926
 FEATURE Location/Qualifiers
 source 1..926
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 75
 MGPCKKRRVGMGPDIVMTQ SPSSLSASVG DRVTITCRSS TGAVTTSNYA SWVQEKGKGL 60
 FKGLIGGTNN RAPVPSRSF GSLIGDKATL TISSLQPEDF ATYFCALWYS NHWFVFGQGTK 120
 VELKRGGGGS GGGGGGGGGG SGGGSEVKLL ESGGGLVQPG GSLSKLCASVS GFSLTDYGVN 180
 WVRQAPGRGL EWIGVIWGDD ITDYNALSKD RFIIKSNDNGK NTVYLQMSKV RSDDTALYYC 240
 VTGLFDYWGYQ GTLTVVSSYP YDVPDYAGGG GGSGGGGGGG GGSGGGGSLD PGGGGGSGKG 300
 EELFTGVVPI LVELDGDVNG HKFSVREGE GE DATNGKLTL FKIICLTGKLP VPWPTLVTTL 360
 THGVQFCFSR VPDHMKRHDF KSAMPEGVYQ ERTISFKDDG TYKTRAEVFK EGDTILVNRIE 420
 LKGIDPKEDG NILGHKLEYN FNSHNVYITA DKQKNGIKAN FKIRHNVEDG SVQLADHYQQ 480
 NTPIDGDPVL LPDNHYLSTQ SVLSKDPNEK RDHMVMLLEFV TAAGITHMD ELYKGGRGT 540
 GGSGGGGGAD PKKRKVARI TSLYKKGAGT MAGVACLGKA ADADEWDSG LSGLGFDAAA 600
 PGPGPLGAEL GPGLSWALPV FGVTEDGDL AHLAVIHOH EPFLDFLFLGF SAGTEYMDLQ 660
 NDLGQTALHI AAILGETSTV EKLYAAAGAGL CVAERRGHTA LHLACRVGAH ACARALLQPR 720
 PPRPRAEPDT YLAQGPDRTP DTNHTPVALY PDSDELEKEEE ESEEDDWKQL EAENYEGHTP 780
 LHVAVIHKHD EMVRLLRDAG ADLDDKPEPTC GRSPLHLAVE AQAAADVLELL LRAGANPAAR 840
 MYGGRTPLGS AMLRPNPILA RLLRAHGAPE PEGEDEKSGP CSSSSSDSDSG DEGDEYDDIV 900
 VHSSRSQTRL PPTPASKPLP DDPPRVP 926

SEQ ID NO: 76 moltype = DNA length = 2781
 FEATURE Location/Qualifiers
 source 1..2781
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 76
 atgggtccca agaaaaagag aaaggctcggt ggcatggcc cccgacatcg gatgaccagg 60
 agccccagca gcctgagcgc cagcgtggc gaccgcgtga ccatcacctg ccgcacgcgc 120
 accggcgcccg tgaccacccg caactacggc agctgggtgc aggagaacgc cggcaactg 180
 ttcaaggccg tgatcgccgg caccacaac cgcgcggcccg gctgcggccag ccgccttcagc 240
 ggcagcctga tcggcggacaa ggcacccctg accatcacgca gctgcgcggc cgaggactc 300
 gocaccaact tctgcgcctt gtggtagacg aaccactggg tggcggccca gggcacaag 360
 gtggagctga agcgcggcg gggtggaaac ggaggcggtg ggtctgggtt aggcggcagc 420
 tctggcggtg gcaaggcgatg gaaggctcg gagaacggcg ggcgccttgtt gcaaggccgc 480
 ggcagcctga agctcgatgt cgcgcgtgac ggcttcagcc tgaccgacta cggcgtgaaac 540
 tgggtcgccg aggcccccgg cgcgcgttgc gatgtatgc gctgtatctg gggcgcacggc 600
 atcaccgact acaaagacgc cttgaaggac cgcttcatca tcagcaaggaa acacggcaag 660
 aacacccgtt acctcgatgt gacaaggatc cgcacgcacg acacccgcctt gtactatgc 720
 gtgcacggcc tggccacta ctggggccac ggcacccctt tgaccgttgcg cagtcatacca 780
 tacgtatgtc cagattacgc tggtggggccac ggagggtttccggggaggtag tgtaggggt 840
 ggtggttcag gaggcgccgg aagcttgat ccagggttgc gttggaaacggg tagcaaggaa 900
 gaagaacttt tcactggagt tgcccaattt ctgttgcataat tagatgttgc tgtaatggg 960
 cacaatattt ctgcctgggtt aggggttgc ggtgtatgc aaaaacggaaa actcaccctt 1020
 aatatttttgcactatgggaaatactactt cggcttggc caacacttgcgtactacttgc 1080
 accccatggt ttcaatgttttccatggatcaca tggatggccca tgactttttc 1140
 aagagtgcaca tgcccgaaagg ttatgtatcg gaacgcacta tatctttcaaa agatgcacgg 1200
 acctaacaac cgcgtgtctt agtcaatgtt ggatgtata cccttgcataatcgtatccg 1260
 ttaaagggtt ttatgttttcaaa agaaggatgtt aacattcttgc gacacaaacttgcgtatcc 1320
 ttaactacaaatccatcatacatcggca gacaacaaaagaatggaaatcaacttgcgtatcc 1380
 ttcaaaaatccatcatacatcggca gacaacaaaagaatggaaatcaacttgcgtatcc 1440
 aatactccaaatccatcatacatcggca gacaacaaaagaatggaaatcaacttgcgtatcc 1500
 ttggcgatgtt ccctgtctt tttacccaca accattacttgcgtatcc 1560
 tctgtatccatcatacatcggca gcaaggatgtt ccgttgcataatcgtatcc 1620
 atctgtgtctt gggatccatcatacatcggca gacgttgcataatcgtatcc 1680
 ggcgggtggca gcggtggagg cgggtgtcgc cccaaagaaagagggatgc 1740
 acaagggttttgcactatgggaaatactactt cggcttggc caacacttgcgtactacttgc 1800
 ggcgcacggatcgtatccatcatacatcggca gcaaggatgtt ccgttgcataatcgtatcc 1860
 atgaatggatgtt ccgttgcataatcgtatccatcatacatcggca gcaaggatgtt ccgttgcataatcgtatcc 1920
 aatggacttagt gccagacacgc cttgcacccgttgc gacatggacatccatcatacatcggca 1980
 gagatgttgcgtatccatcatacatcggca gcaaggatgtt ccgttgcataatcgtatcc 2040
 atcgacccgttgc gacatggacatccatcatacatcggca gcaaggatgtt ccgttgcataatcgtatcc 2100
 ctgcacccgttgc gacatggacatccatcatacatcggca gcaaggatgtt ccgttgcataatcgtatcc 2160
 ccccgccggccatcatacatcggca gcaaggatgtt ccgttgcataatcgtatcc 2220
 gacaccaaccatcatacatcggca gcaaggatgtt ccgttgcataatcgtatcc 2280
 atacccttgcgtatccatcatacatcggca gcaaggatgtt ccgttgcataatcgtatcc 2340
 gagatgttgcgtatccatcatacatcggca gcaaggatgtt ccgttgcataatcgtatcc 2400
 ctccacccgttgc gacatggacatccatcatacatcggca gcaaggatgtt ccgttgcataatcgtatcc 2460
 tgcacccgttgc gacatggacatccatcatacatcggca gcaaggatgtt ccgttgcataatcgtatcc 2520
 atgtacccgttgc gacatggacatccatcatacatcggca gcaaggatgtt ccgttgcataatcgtatcc 2580
 gccgcacccgttgc gacatggacatccatcatacatcggca gcaaggatgtt ccgttgcataatcgtatcc 2640
 gtcacccgttgc gacatggacatccatcatacatcggca gcaaggatgtt ccgttgcataatcgtatcc 2700
 gtcacccgttgc gacatggacatccatcatacatcggca gcaaggatgtt ccgttgcataatcgtatcc 2760
 gacgacccgttgc gacatggacatccatcatacatcggca gcaaggatgtt ccgttgcataatcgtatcc 2820

-continued

```

SEQ ID NO: 77      moltype = AA length = 1149
FEATURE          Location/Qualifiers
source           1..1149
                 mol_type = protein
                 organism = synthetic construct
SEQUENCE: 77
MGPKKKRKVG GMGPDIVMTQ SPSSLSASVG DRVTITCRSS TGAVTTSNYA SWVQEKPGL 60
FKGLIGGTNN RAPGVPSRFS GSLIGDKATL TISSLQPEDF ATYFCALWYS NHWFVFGQGTK 120
VELKRGGGGS GGGGGGGGS SGGGEVKKL ESGGGLVQPG GSLKLSCAVS GFSLTDYGVN 180
WVRQAPGRGL EWIGVIWGDG ITDYNMSALK RFIISKDNKG NTVYLQMSKV RSDDTALYYC 240
VTGLFDYWGQ CTLVTVSSYP YDVPDYAGGG GGSGGGGGSGG GGSGGGGSLD PGGGGGSGKG 300
EELFTGVVPI LVELDGDVNG HKFSVSRGEGE GDATNGKLTL KFICTTGKLP VPWPTLVTTL 360
THGVQCFRSY PDHMKRHDFE KSAMPEVQVQ ERTISFKDDG TYKTRAEVKF EGDTLVRNRIE 420
LKGIIDFKEDG NILGHKLEYN FNSHNVIYITA DKQKNGIKAN FKIRHNVEDG SVQLADHYQQ 480
NTPIGDGPVL LPDNHNLSTQ SVLSKDPNEK RDHMVLLFV TAAIGITHGMD ELYKGGGRTG 540
GGGSGGGGAD PKKKRKVARI TSLYKKAGST MLRSGPASPQ SVPTGRAMPS RRVARPPAAP 600
ELGALGSPDL SSLSLAWSRS TDELEIIDEY IKENGFGLDG QOPGPGEGLP RLVSRAASL 660
STVTLGPVPAT PPPPWGCP LGRLVSPAPG PGPQPHLVIT EQPKQRGMRF RYECEGRSAG 720
SILGESSTEA SKTLPAIELR DCGGLREVEV TACLVWKDWP HRVPHPSLVG KDCTDGICRV 780
RLRPHVSPRH SFNNNLGIQCV RKKEIEAAAIE RKIQLGIDPY NAGSLKNHQE VDMNVVRICF 840
QASYRDOQGQ MRRMPDVLSR PVDYKKSTNT SELRICRINK ESGPCTGSEE LYLLCDKVQK 900
EDISVVFSRA SWEGRADFSQ ADVRHQIAIV FKTPPYEDLIE IVEPVTVNF LQRLTDGVCS 960
EPLPFTYLPR DHDSYGVDDK RRKRGMPDVLG ELNNSDPHGI ESKRRKKKPA ILDHFLPNHG 1020
SGPFLPPPSA LPDPDFSGT VSLPGLEPPG GPDLDDGFA YDPTAPTLFT MLDLLPPAPP 1080
HASAVVCSGG AGAVVGETPG PEPLTLDSYQ APGPGDGGTA SLVGSNMFPN HYREAAFGGG 1140
LLSPGPPEAT                                         1149

SEQ ID NO: 78      moltype = DNA length = 3450
FEATURE          Location/Qualifiers
source           1..3450
                 mol_type = other DNA
                 organism = synthetic construct
SEQUENCE: 78
atgggtccaa aaaaaaaaaaaaaagag aaagggtcggtt ggcatgggcc ccgacatcggt gatgaccagg 60
agcccccaaga gcctgagcgc cagcggtggc gaccgcgtga ccatcacctg ccgcgcgcgc 120
acccggcgcgc tgaccaccag caactacgcg agctgggtgc aggagaaggcc cggcaactgc 180
ttcaaggccc tgatcgccgg cacaacaac cgccgcggcccg gcgtgcggccag ccgttcagc 240
ggcagctcta tcggcgacaa ggcaccttcg accatcaagca gcctgcggcc cgaggacttc 300
gccacctaact tctgcggccct gtggtagcago aaccactggg ttgtcgccca gggcaccaaag 360
gttgagactga acgcgcgggg cgggtgaagc ggagcggtgg ggtctgtgtgg aggcgccgc 420
tctggcggag gcaagcgaggt gaagctgtgc gagagcgccg cggcccttgt gcagccccgc 480
ggcagctgaa agctgactgc cgccgtgagc gggttcagcc tgaccgacta cggcgtaaac 540
tgggtgcggc aggccccccgg ccggggcttg gagtggatcg cgtgtatctg gggcgacggc 600
atcaccgact acaacagcgc cttcaaggac cgcttcataca tcagcaaggaa caacggcaag 660
aacaccgtgt acctgcagat gagaagggtg cgcagcgacg acaccgcct gtactactgc 720
gtgaccggcc ttgttcgacta ctggggccca ggccacccctgg tgacccgtgg cagctaccca 780
tagatgttc cagattacgc ttggggggccgg ggggggttctgg ggggggggg tagtgccgg 840
gggtggttcg gagggggccgg aagtttggat ccagggtggag gtggaaaggcc tagcaaaagg 900
gaagaacttt tcaactggagt tgcattttttt cttgttgaat tagatggta tggtaatggg 960
cacaatattt ctgtccgtgg agagggtaac ggtgtatgata caaacggaaa actcaccctt 1020
aaatattttt gcaactactgg aaaactacatc gttccgtggc caaacttgt cactactgt 1080
acccatgtt gttcaatgtt tccccgttat ccggatcaca tggaaacggca tgacttttc 1140
aagagtgcga tggccgaagg ttatgtacag gaacgcacta tatcttcaa agatgacggg 1200
acctacaaaga cgcgtgtctga agtcacgtt gaaggtata cccttggtaa tcgtatcgag 1260
ttaaagggtt ttgattttaa agaagatggaa aacattcttgc gacacaacta cggatcaac 1320
ttaacttcac acaatgtata catcaggca gacaaacaaa agaatggaaat caaagctaacc 1380
ttcaaaaattt gccacaacgt tgaagatgtt tccgttcaac tagcagacca ttatcaacaa 1440
aataactccaa ttggcgatgg ccctgtccctt ttaccagaca accattactt gtgcacacaa 1500
tctgtccctt cggaaatgtcc caacaaaggaaat cgtgaccatc tggctcttgc tggtttgt 1560
actgctgtgtt ggattacaca tggatggat gacgttcataa aagggtgggg tcggaccgg 1620
ggcggtggca gcggtggagg cgggtgtgac cccaaagaaga agagggaaatgg ggtctggatc 1680
acaagtttgtt acaaaaaaaggc aggttccacc atgcttcgtt ctggggccgc ctctggggcc 1740
tccgtccca ctggcgccggc catgcgcgtt ccggccgcgc ggctggccgg 1800
gagctggggg ctttagggc ccccgaccc tccctactt cgtctgcggcgt ttccaggagc 1860
acagatggat ttggatgtatc cgcacgttcc atcaaggaga acggcttcgg cctggacccgg 1920
ggacagccgg gccccggcga ggggtgttgc cgcgtgggtt ctggcgccgg tggctccctg 1980
agcacgggtca ccctggggcc tttggcgccc ccagccacgc ccgcgcctt gggctgcggcc 2040
ctggggccac tagtgcggcc ccggccggcc acggccacccgtt ggtcatcaac 2100
gagcagccca agcagcgccg catgcgttcc cgttactgtt gggggccgc ctggccggc 2160
agcatccctt gggagagcag caccggggcc agcaagacgc tcggcccat cggatccgg 2220
gattgtggag ggctgggggaa ggtggagggtt actgcctgc tgggtggaa ggactggcc 2280
caccgagttcc accccccacag cctctgtgggg aaagactgca ccgacggcat ctgcgggtt 2340
cggtctccggc ctcacgttcc ccccccggcactt agttttaaaca acctggccat ccagtgtgt 2400
aggaagaagg agattggggc tgccatttag cggaaagattt aactggccat tgacccttac 2460
aacgctgggtt ccctgaagaa ccatcaggaa gttagacatga atgtgggtt gatctgtttc 2520

```

-continued

caggcctcat	atcgggacca	gcagggacag	atgcggccga	tggatctgt	gctttccgag	2580
cccgcttatc	acaagaatac	cacaacacaa	tcagagctgc	ggatttgcgg	aattaacaag	2640
gaaaggcgggc	cgtgcacccgg	tggcggaggag	ctctacttgc	tctgcacaa	ggtgtcggaaa	2700
gaggacatat	cagtgggttt	cagcaggccc	tccctggaaag	gtcgccgtga	cttctccca	2760
gccgacgtgc	accggccagat	tgccattgtg	ttcaagacgc	cgccctacga	ggacctggag	2820
attgtcggc	ccgtgacagt	caacgtctc	ctgcagccgc	tcacccatgg	ggttctgcgc	2880
gagccatgc	ctttcacgt	cetgcctcgc	gaccatgaca	getacggcgt	ggacaagaag	2940
cggaaacgggc	ggatgcccga	cgtcttgggg	gagctgacaa	gtctgaccc	ccatggcgc	3000
gagagcaaac	ggcggaaagaa	aaagccggcc	atctggacc	acttcttgc	caaccacggc	3060
tcagggccgt	tcctcccccc	gtcggccctt	ctgcccagacc	ctgacttctt	ctctggcacc	3120
gtgtccctgc	ccggcctgg	gcccccttgc	tcctggaccc	tcctggacca	tggttttgc	3180
tacgacccta	ccggccccccac	acttcttacc	atgtggacc	tgctgcccc	ggcaccggca	3240
cacgttagcg	ctgttgtgt	cageggaggt	gcccggggcc	tggttgggg	gaccggccgc	3300
cctgaacacc	tgacactgg	ctctgtacc	gccccggggc	ccggggatgg	aggcaccggc	3360
agecctgtgg	ycgcaacat	gttccccat	cattaccgc	aggcggccct	tggggccgc	3420
ctcttataccc	cggggctg	agccacgtag				3450

SEQ ID NO: 79 moltype = AA length = 847

FEATURE Location/Qualifiers
source 1..847
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 79

MGPKKKRKVG	GMGPDIVMTQ	SPSSLSASVG	DRVTITCRSS	TGAUTTSNYA	SWVQEKPGL	60
FKGLIGTNN	RAPGVPSRFS	GSLIGDKATL	TISSLQPEDF	ATYFCALWYS	NHWVFGQGTK	120
VELKRGGGGS	GGGGSGGGGS	SGGGSEVKLL	ESGGGLVPG	GSLKLSCAVS	GFSLTDYGVN	180
WVRQAPGRGL	EWIGVIWGDG	ITDYNALSKD	RFIISKDNGK	NTVYLQMSKV	RSDDTALYYC	240
VTGLFLDXWQQ	GTLLTVSSYP	YDVDPYAGGG	GGSGGGGGSGG	GGSGGGGSLD	PGGGGGSGSKG	300
EELFTGVVPI	LVELDGDVNG	HFKFSVRGEGE	GDATNGKLTL	KFICTTGKLP	VPWPVLTVTTL	360
THGVQCFSRY	PDHMKRHDF	KSAMPEGYVQ	ERTISFKDDG	TYKTRAEVKF	EGDTLVNRIE	420
LKGIDDFKEDG	NILGHKLEYN	FNSHNVYITA	DKQKNGIKAN	FKIRHNVEDG	SVQLADHYQQ	480
NTPIGDGPVL	LPDNHNLSTQ	SVLSKDPNEK	RDHMVLLFEV	TAAGITHGMD	ELYKGGRRTG	540
GGGGGGGGAD	PKKKRKVARI	TSLYKKAGST	MADHMMAMNH	GRFPDGTNGL	HHHPAHRMGM	600
GQFPSPHHHQ	QQQPOHAFNA	LMGEHIHYGA	GNNNATSGIR	HAMGPGTVNG	GHPPSALAPA	660
ARFNNSQFMG	PPVASQGSSL	PASMQLQLN	NOYFNHHPPY	HNHYMPDLHP	AAGHQMNQTN	720
QHFRDCPNKH	SGGSSTPGGS	GGGSTPGGS	SSSGGGGAGSS	NSGGGSGSGN	MPASVAHVPA	780
AMLPPNVIDT	DFIDEEVLMS	LVIEMGLDRI	KELPELWLQQ	NEFDFTDFV	CKQOPSRVSC	840
LDPAFLY						847

SEQ ID NO: 80 moltype = DNA length = 2541

FEATURE Location/Qualifiers
source 1..2541
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 80

atgggtccca	agaaaaaagag	aaaggctcggt	ggcatggccc	ccgacatctgt	gtatgaccag	60
agccccagca	gcgtgagcgc	cagcgctggc	gaccgcgtgc	ccatcacctg	ccgcacgc	120
accggcgcgc	tgaccaccc	caactacgc	agttgggtgc	aggaaagcc	ccgcaagctg	180
ttaaaggccc	tgatcggccg	cacaacaaac	cgccgccttc	cgctgcccc	ccgcttcage	240
ggcagcctga	tcggcgcacaa	ggccaccctg	accatcaga	gcctgcagcc	cgaggactc	300
gcacacact	tctggccctt	gttgcacgc	aaccactgg	tgttgcggca	gggcacccaag	360
gtggagctga	acgcggcggc	cgttgcacgc	ggggcgggttgc	gttgcgggttgc	aggcggcagc	420
tctggccgg	gcaacggagg	gaactgtctg	gagacggccgc	cgccgcctgt	gcagccccgc	480
ggcagcctga	agctgagctg	cgcgtgago	ggcttcagcc	tgaccgacta	cgccgtgaac	540
tgggtgcggc	aggccccccg	cgcgcgcctg	gagtggatcg	cgctgtatctg	gggcgcacggc	600
atcaccgact	acaacacgac	cctgtatcata	tcagcaagga	caacggcaag	660	
aacaccgtgt	acactcagat	gagaacggtg	cgcacgcacg	acaccgcct	gtactactc	720
gtgacccggc	tgttcacta	ctggggccag	ggcaccctgg	tgaccgtgag	cagctacca	780
tacgatgtt	cagatatacg	tggtgaggc	ggagggttctg	gggggaggagg	tagtggccgt	840
ggtggttcag	gaggcggcgg	aaagttgtat	ccaggtggag	gttggagccg	tagcaaagga	900
gaagaacttt	tcactggat	tgtcccaatt	cttgcgttata	tagatgttgc	tgttaatggg	960
cacaaatttt	ctgtccgtgg	agaggggtaa	ggtgcgtacta	caaacggaaa	actcaccctt	1020
aaatttttt	gtactactgg	aaaactacct	gttccgtggc	caacacttgt	cactactctg	1080
accctatgtt	ttaaatgttt	ccggatcaca	tgaacacggca	tgacttttt	1140	
aagagtgc	tggccgaagg	ttatgtacag	gaacgcacta	tatcttcaa	atagtgacgg	1200
acctacaaga	cgcgtgctga	agtcacgtt	gaaggtgata	ccctgtttaa	tcgtatcgag	1260
ttaaagggtt	ttgatattaa	agaagatgg	aacattcttgc	gacacaaact	cgagtacac	1320
ttaaactc	acaatgtata	catcacggca	gacaaacaaa	agaatggaaat	caaagctaac	1380
tccaaaattt	gcacacaacgt	tgaatgtgt	tccgttcaac	tagcagacca	ttatcaacaa	1440
aatacttca	ttggcgatgg	ccctgtctt	ttaccagaca	accattacct	gtcgacacaa	1500
tctgtccctt	cgaaaagatcc	caacggaaa	cgtgaccaca	tgttccctt	ttagtttgc	1560
actgtgtct	gatttacaca	tggtatggat	gagctctaca	aaagggtgggg	tcggaccgg	1620
ggcggtggca	gcgggtggagg	cgtgtctgac	ccccagaaga	agaggaaggt	ggcttaggatc	1680
acaagtttgt	acaaaaaaagc	aggctccacc	atggcagac	atatgtggc	aatgaaccac	1740
gggcgttcc	ccgacggcac	caatgggtc	caccaatc	ctgcccaccc	catggccat	1800
gggcagttcc	cgagccccca	tcaccaccc	cagcagcagc	ccccacacgc	tttcaacgccc	1860

-continued

ctaatggcg	agcacataca	ctacggcg	ggcaacatga	atgcccacgag	cggccatcagg	1920
catgcgttgc	ggcccgggac	tgtgaacgg	gggcacccccc	cgagccgcgt	ggcccccccg	1980
gccagggtta	acaactccca	tttcatgggt	ccccccgggt	ccagccagg	aggccctcg	2040
ccggccacca	tgcagctca	gaagtcac	aaccagtatt	tcaaccatca	cccttacccc	2100
cacaaccact	acatggcg	tttgcaccc	gctgcaggcc	accagatgaa	cgggacaaac	2160
cagcacttcc	gagatggcaa	ccccaaagcac	ageggcgccca	gcagcacccc	cgccggctcg	2220
ggccggcagca	gcaccccccgg	cggtctgg	agcagctcg	ggggccggc	gggcagecagc	2280
aacagccgc	ggggcagcgg	cageggcaac	atgcccgcct	ccgtggccca	cgtccccgc	2340
gcaatgtgc	cgcccaatgt	catagacact	gatttcatcg	acgaggaaat	tcttatgtcc	2400
tttgtatag	aatatggttt	ggacggcata	aaggagctgc	ccgaactctg	gctggggcaa	2460
aacgagttt	atttatgac	ggactctgt	tgcaaacagg	agcccaacagg	agttagctgt	2520
ttggaccagg	ctttttgtca	c				2541

SEQ ID NO: 81 moltype = AA length = 247

FEATURE Location/Qualifiers
source 1..247
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 81

MGPDIVMTQS	PSSLSASVGD	RVTITCRSST	GAVTTSNYAS	WVQEKPGLF	KGLIGGTNNR	60
APGVPSRFSG	SLIGDKATLT	ISSLQPEDFA	TYFCALWYSN	HWFQGQTKV	ELKRGGGSG	120
GGGGGGGGSS	GGGSEVKLLE	SGGLLVQPGG	SLKLSCAVSG	FSLTDYGVNW	VRQAPGRGLE	180
WIGVIWGDGI	TDYNSALKDR	FIISKDNGKN	TVYLQMSKVR	SDDTALYYCV	TGLFDYWGQG	240
TLTVSS						247

SEQ ID NO: 82 moltype = DNA length = 858

FEATURE Location/Qualifiers
source 1..858
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 82

atggggcccg	acatcggtat	gaccaggagc	cccaggcagg	tgagcgccag	cgtggggcag	60
cgcgtgacca	tcacccgtcg	cagcagcacc	ggggccgtga	ccaccaccaa	ctacggccacg	120
tgggtgcagg	agaagcccg	caagctgttc	aaggccgtga	tcggccgac	caacaaccgc	180
gccccccggc	tgcccgccgc	cttccggcgg	agccctgtatcg	gcgacaaggc	caccctgacc	240
atcagcggcc	tgcagccgg	ggactctcg	acactacttc	gcccctgtg	gtacagcaac	300
cactgggtgt	tccggccagg	cacaagggt	gagctgaacg	gcggcgccgg	tggaaaggaa	360
ggcggtgggt	ctgggtgg	cgccgactct	ggcgaggagc	gcgagggtaa	gctgtggag	420
agccggccgc	gctctgtca	gccccggcgg	agccctgaacg	tgagctcg	cgtgacggc	480
ttcagccgt	ccgactacgg	cgtgaactcg	gtgcggccgg	ccccggccg	cgccctggag	540
tggatcgccg	tgatctgggg	cgacggcata	accgactaca	acagccct	gaaggaccc	600
ttcatacatca	gcaaggacaa	cggaagaac	accgtgtacc	tgcagatgag	caagggtcgc	660
agccgacgaca	ccggccgtat	ctactcgctg	accggctctg	tgcactactg	ggccaggccc	720
accctgggtg	ccgtgagcag	ctaccatata	gatgttccag	attacgttgg	tggaggccgg	780
ggttctgggg	gaggaggtag	tggcggtgt	ggttcaggag	gcggcgaaag	cttggatcca	840
ggtggaggt	gaaaggcg					858

SEQ ID NO: 83 moltype = AA length = 237

FEATURE Location/Qualifiers
source 1..237
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 83

SKGEELFTGV	VPILVLDGD	VNGHKFSVRG	ECEGDATAKNG	LTLKFICTTG	KLPVPWPTLV	60
TTLTHGVQCF	SRYPDHMKRH	DFFKSAMPEG	YVQERTISFK	DDGTYKTRAE	VKFEGDTLVN	120
RIELKGIDFK	EDGNILGHKL	EYNFNNSHNV	ITADKQKNGI	KANFKIRHNV	EDGSVQLADH	180
YQQNTPIGDG	PVLLPDNHYL	STQSVLSKDP	NEKRDHMVLL	EFVTAAGITH	GMDELYK	237

SEQ ID NO: 84 moltype = DNA length = 711

FEATURE Location/Qualifiers
source 1..711
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 84

agcaaaaggag	agaactttt	cactggat	gtcccaattc	ttgttgaatt	agatggtgat	60
gttaatgggc	acaaaatttc	tgtccgtga	gagggtaa	gtgtatctac	aaacggaaaa	120
ctoaccctta	aatttttt	cactactga	aaactaccc	ttccgtggcc	aacacttgc	180
actactctga	cccatgggt	tcaatgttt	tcccgttatac	cggatcacat	gaaacggcat	240
gacttttca	agagtggcat	gccccaaagg	tatgtacagg	aacgcactat	atctttcaaa	300
gatgacggga	cctacaagac	cgctgtcg	gtcaagttt	aaaggatatac	ccttggtaat	360
cgtatcgagt	taaagggtat	tgatttaaa	gaagatgaa	acattctgg	acacaaactc	420
gagtacaact	ttaactcaca	caatgtatac	atcacggcag	acaaacaaa	gaatggaaatc	480
aaagctaact	tcaaaaattcg	ccacaacgtt	gaagatgtt	ccgttcaact	agcagaccat	540
tatcaacaaa	atactccat	tggcgatgg	cctgtccctt	taccagacaa	ccattacctg	600
tcgacacaat	ctgtccttc	gaaagatccc	aacgaaaagg	gtgaccacat	ggtccttctt	660
gagtttggtaa	ctgtcgatgg	gattacacat	ggcatggat	agcttacaca	a	711

-continued

```

SEQ ID NO: 85      moltype = AA length = 19
FEATURE
source          Location/Qualifiers
1..19           mol_type = protein
                organism = synthetic construct
SEQUENCE: 85
EEELSKNYHL ENEVARLKK                                         19

SEQ ID NO: 86      moltype = DNA length = 57
FEATURE
source          Location/Qualifiers
1..57           mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 86
gaggagcttc tgagcaaaaa ctatcacctc gaaaacgagg ttgcgcgact gaagaaa      57

SEQ ID NO: 87      moltype = AA length = 1607
FEATURE
source          Location/Qualifiers
1..1607          mol_type = protein
                organism = synthetic construct
SEQUENCE: 87
MDKYSIGLA IGTNSVGWAV ITDEYKVPSK KFKVLGNNTDR HSIKKNLIGA LLFDSGETAЕ 60
ATRLKRTARR RYTRRKNRIC YLQEIFSNEМ AKVDDSSFFH LEESFLVEED KKHERHPIFG 120
NIVDEVAYHE KYPTIYHLRK KLVSTDKAD LRILYLAHAL MIKFRGHFLI EGDLNPNSD 180
VDKLFIQQLVQ TYNQLFEENP INASGKDAK ILSARLSKSR RLENLIAQLP GEKKNGLFGN 240
LIALSLGLTP NFKSNPDLAЕ DAKLQLSKDT YDDDLDNLLA QIGDQYADLF LAAKNLSDAI 300
LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLKLALVR QQLPEKYKEI FFDQSKNGYA 360
GYIDGGASQE EFYKFIKPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH 420
AILRRQQEDFY PFLKDNRREKI EKILTFRIPP YVGPLARGNS RFAWMTRKSE ETITPWNFEE 480
VVDKGASAQS FIERMTNFDK NLPNEKVLPK HSLLYYEFTV YNELTKVYV TEGMRKP AFL 540
SGEQKKAAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLLKI 600
IKDKDFLDNE ENEDILEDIV LTTLTFLFEDR MIEERLKTYA HLFDDKVMKQ LKRRRTGNG 660
RLSRKLINGI RDKQSGKTIL DFLKSDGFPAN RNFMQLIHDD SLTFKEDIQK AQVSQGQDSL 720
HEHIANLAGS PAIKGILQT KVVDDELVKV MGRHKPENIV IEMARENQTT QKGQKNSRER 780
MKRIEEGIKE LGSQILKEHP VENTQLQNEK LYLYYLQNGR DMYVQELDI NRSLSDYDVDA 840
IVPQSFLLKDD SIDNKVLTRS DKNRGKSDNV PSEEVVKKM NYWRQLLNK LITQRKFNDL 900
TKAERGGLSE LDKAGFIKRDF LVETRQITKH VAQILDSRMN TKYDNLKLI REVKVITLKS 960
KLVSDFRKDF QFYKVKREINN YHHAHKTVLN AVVGTALIKK YPKLESEFVY GDYKVDVRK 1020
MIAKSEQQIG KATAKYFFYS NIMMFKTEI TLANGEIRRK PLIETNGETG EIVWDKGRDF 1080
ATVRKVLSMP QVNIVKKTEV QTGGFSKESI LPKRNNSDKLI ARKKDWDPKK YGGFDSPVVA 1140
YSLVLLVAKVE KGKSQKLKSV KELLGILTIME RSSFEKPNIP FLEAKGYKEV KKDLIIKLPK 1200
YSLFELENGR KMLASAGEL QKGNEALALPS KYVNFLYLAS HYEKLKGSPE DNEQKQLFVE 1260
QHKHYLDEII EQISEFSKRV ILADANLDKV LSAYNKHRDK PIREQAENII HLFTLTNLGA 1320
PAAFKYFDTT IDRKRYTSTK EVLDATLHQ SITGLYETRI DLSQLGGDSR ADPKKKRKVA 1380
SDIGIGGSNG SSSGNSNGPTDA AEEELLSKRN HLENEVARLKGSGSGSGS GSGGSQGGGS 1440
GSGGEELLSKRN HLENEVARLKGSGSGSGS GSGSGEELLSK NYHLENEVARLKGSGSGSGS 1500
LKKKGSGSGS GSGSGSGSGS GSGSGEELLSK NYHLENEVA RLKKKGSGSGG SGSGSGSGS 1560
GGSGSGEELL SKNYHLENEV ARLKKKGSGSG GSGSGSGGG SGSGSGSGS 1607

SEQ ID NO: 88      moltype = DNA length = 4755
FEATURE
source          Location/Qualifiers
1..4755          mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 88
atttacggacaaga agtactccat tgggtcgcc atcggcacaa acagcgtcgg ctggggccgtc 60
attacggacg agtacaaggt gcccggcaaa aaattcaaaat ttctggccaa taccgtcg 120
cacagcataaa agaaaaccc tattttcgcc cttctgttcg actccgggaa aaccggccaa 180
ggccacggccg tcaaaaacac agcacggccg agatatacccg cagaaaaaaa tcggatctcg 240
tacctgcagg agatcttag taatgagatg gctaagggtgg atgactctt cttccatagg 300
ctggaggagt ctttttttgtt ggaggaggat aaaaacgcg agcgccaccc aatcttggc 360
aatatctgg acgggtggc gtacccatgaa aagtacccaa ccataatatac tctgaggaaag 420
aagcttggtag acatgtactga taaggctgac ttgcgggtta tctatctcg gctggcgcatt 480
atgtaccaat ttccggggaca cttcttcattc gagggggacc tgaacccaga caacacgcatt 540
gtcgacaaac tctttatcca actgggttcg acttacaatc agcttttcga agagaacccg 600
atcaacgcatt ccggagggtga cgccaaagca atccctgacg ctaggctgtc caaatccgg 660
cggtcgaaa acctcatcgc acagctccctt ggggagaaga agaacggcctt gtttgtaat 720
cttatacgccc tgcactcggtt gctgaccccc aactttaaat ctaacttcga cctggccgaa 780
gatgccaacg ttcaacttgat ccaagacacc tacggatgtt atctcgacaa tctgtggcc 840
cagatcggcg accagttacgc agacctttt ttggcggcaaa agaacctgtc agacgcatt 900
ctgctgatgtt atattctgcg agtgaacacg gagatcacca aagctccgtt gaggcgctgt 960
atgatcaacg gctatgtga gcaccacaa gacttgactt tgctgaaggc ctttgcaga 1020
cagcaactgc ctgagaagta caaggaaatt ttcttcgatc agtctaaaaa tggctacgcc 1080
ggatacatttgc acggcggcggc aagccaggag gaattttaca aattttatcaa gccccatctt 1140

```

-continued

aaaaaaatgg	acggcaccga	ggagctgctg	gtaaagctta	acagagaaga	tctgtgcgc	1200
aaacagcgc	cttgcacaa	tggacatc	ccccaccaga	ttcacctgg	cgaactgcac	1260
gctatccca	ggcgccaaaga	ggatttc	cccttttg	aagataacag	ggaaaagatt	1320
gagaaaaatcc	tcacatttc	gataccctac	tatgttaggc	ccctcgcccg	gggaattcc	1380
agattcgcgt	ggatgactcg	caaatacgaa	gagaccatca	ctccctggaa	cttcgaggaa	1440
gtctgtggata	agggggcctc	tgcccagtcc	ttcatcgaaa	ggatgactaa	ctttgataaa	1500
aatctgccta	acgaaaaggt	gttctctaa	cactctgc	tgtacgaga	cttcacagtt	1560
tataacgacg	tcaccaaggt	caaatacgtc	acagaaggga	tgagaaaagcc	agcattctg	1620
tctggagagc	agaagaagc	tategtggac	ctctcttca	agacgaacg	gaaagtacc	1680
gtgaaacacg	tcaaqaaga	ctatccaa	aaqattgaa	gttgcactc	tgttqaatc	1740
acgggatgttgg	aggatcgtt	caacgcattc	ctggaaacgt	atcacgatc	cctgaaaatc	1800
attaaagaca	aggactctct	ggacaatgag	gagaacggg	acattcttg	ggacattgtc	1860
ctcaccctta	ctgtgttga	agatagggag	atgattgaag	aacgcttga	aacttacgt	1920
catctcttcg	acgacaaaggt	catgaaacac	ctcaagggc	gccgatatac	aggatgggg	1980
cggctgttca	gaaaactgt	caatggatc	cgagacaa	agatggaaa	gacaatctg	2040
gattttctaa	agtccgatgg	atttccaa	cgaaactca	tgcagtgt	ccatgtgac	2100
tctctcacct	ttaaggagga	catccagaa	gcacaagtt	ctggccagg	ggacagtctt	2160
cacgagcaca	tcgctaatct	tgcaaggtag	ccagctatca	aaaaggaaat	actgcagacc	2220
gttaaggctcg	ttgtatcg	cgtcaagta	atggaaaggc	ataaggccga	gaatatcg	2280
atcgagatgg	cccgagagaa	ccaaatcacc	cagaagggac	agaagaaacag	tagggaaagg	2340
atgaagagga	ttgaagaggg	tataaaagaa	ctggggtccc	aaatcttaa	ggaacaccca	2400
gttggaaaaca	cccgacttc	gaatgaga	cttacccgt	actacccgtc	gaacggcagg	2460
gacatgtacg	ttgtatcgagg	actggacatc	aatcggtct	cgcactacga	cgtggatg	2520
atcggtcccc	agtctttct	caaaatgt	tcttattgt	ataaaatgtt	gacaatgtc	2580
gataaaaata	gagggaaagag	tgataacgtc	ccctcagaag	aagtgtcaa	gaaaatgaaa	2640
aattatttgc	ggcactgtgt	gaacggccaa	ctgatcacac	aacggaaatt	cgataatctg	2700
actaaggctg	aacgaggatgg	cctgtctgt	tttgataa	cgggtttcat	aaaaaggcag	2760
cttgggtgaga	cacggccat	caccaagac	gtggcccaa	tttcgcattc	acgcgtgaa	2820
accaagtagc	atgaaaatga	caaactgatt	cgagaggtag	aagtattac	tctgaatgt	2880
aagctgttct	catgttc	aaaggactt	cattttat	aggtgagaga	gatcaacaat	2940
taccacatg	ccgcattc	ctaccatgt	gcaatgtt	gcaatgtt	gactgtcaat	3000
tatccccaa	ttgaatctg	atttgc	ggagactata	aagtgtacg	tgttggaaa	3060
atgatcgca	agtctgagca	ggaaaatggc	aaggccacc	ctaagtactt	cttttacacg	3120
aatattatg	atttttca	gaccgagatt	acactggcc	atggagagat	tcggaagcga	3180
ccacttatcg	aaacaaacacg	agaaaacagg	gaardatgt	gggacaagg	tagggattt	3240
gcgacagatcc	gaaaggctt	gtccatggc	caggtagaa	tcgtt	aaaaaa	3300
caagccgg	ggttccaa	ggaaatgtatc	ctcccgaaa	ggaacacgca	caagctgatc	3360
gcacgc	aaaaaa	aaagattgg	ccccaaagaa	tacggccgt	tcgattctc	3420
taacatgtac	ttgttgttgc	caaaatgt	aaaggaaatg	ctaaaaact	aaaaacgc	3480
aaggaaatgc	ttggccatc	aatcatgg	cgtcaaggt	tcgaaaaaa	ccccatcgac	3540
tttctcgagg	cgaaaagata	taaaaggatc	aaaaaaagacc	tcatactt	aaatccaa	3600
tactctct	ttgagcttgc	aaacggccgg	aaacgatgc	tcgtatgt	ggggcgact	3660
caaaaaatgt	acgactgttgc	actgcctt	aaatacg	tatttctgt	tctggccac	3720
cactatgt	aaatcgaa	gtctccgg	gataatgag	agaacgcgt	tttcgtgg	3780
caacacaaac	actacccgt	tgatgtatc	gaccaatata	gcaatttctc	aaaaagatg	3840
atccctcgcc	acgactaacct	cgataagg	tttctgtt	acaataagca	caggataag	3900
cccatcgagg	acgcggcaga	aaacattatc	cacttgc	ttctgcac	tttgcac	3960
cctgcacgc	tcactgtt	cgacaccac	atagacagaa	agcggat	tttgcac	4020
gaggctctgg	acggccacact	gatccat	tcaattacgg	ggctctatg	aacaagatc	4080
gacctcttc	acgtcggtt	agacacg	gctgacccca	agaagaagag	gaaggatg	4140
agcgacgc	ttgtgtatgg	gaccaacggc	agcagcggat	ccaaacgg	gactgcac	4200
gcggaaagagg	acgttctgt	caaaaactat	cacccgt	acggatgt	gcgact	4260
aaaggaaatgc	gttccggatgg	aaatgttgc	tttgcgtt	tttgcgtt	tttgcgtt	4320
ggcactgttgc	aaagacttcc	tagtaagaa	tatcatgtt	aaaatgagg	agcgcgtt	4380
aaagaaatgg	cgggaaatgg	cgccacgg	agtggttgc	gagggttgc	tttgcgtt	4440
tccggcgttgc	gaggatgttgc	gtgttctat	aacttacat	tagaaaaac	agtcgcac	4500
ctaaaaaaatgt	gttccggatgg	cgccggatcc	ggtttgcgtt	ggggatgttgc	gggttgcgtt	4560
ggatctgtat	caggagatgg	atttgc	aaaaactacc	accttgc	tttgcgtt	4620
agtttaaaatgt	aggggagccg	ctcggttgc	agtggttgc	ggttgcgtt	tttgcgtt	4680
gttccggatgg	gtatctgggg	ggatctgtat	tttgcgtt	tttgcgtt	tttgcgtt	4740
gcgagactaa	agaag					4755

SEQ ID NO: 89 moltype = AA length = 1997
 FEATURE Location/Qualifiers
 source 1..1997

mol_type = protein
 organism = synthetic construct

SEQUENCE: 89

MDKKYSIGLA	IGTNSVGWAV	ITDEYKVPSK	KFKVLGNTDR	HSIKKNLIGA	LLFDSETAE	60
ATRLKRTARR	RYTRRKNRIC	YLQEIFSNE	AKVDDSFHFR	LEESFLVEED	KKHERHPIFG	120
NIVDEVAYHE	KYPTIYHLRK	KLVSTDKDAD	LRLIYLALAH	MIKFRGHFLI	EGDLNPDNSD	180
VDKLFQILVQ	TYNQLFEENP	INASGVDAKA	ILSARLSKS	RLENLIAQLP	GEKKNGLFGN	240
LIALSLGLTP	NFKSNFDLAE	DAKLQLSKDT	YDDLDNLILA	QIGDQYADLF	LAACKNLSDAI	300
LISDILRVNT	EITKAPLSAS	MIKRYDEHHQ	DLTLLKALV	QOLPEKYKEI	FFDQSKNGYA	360
GYIDGGASQE	EFYKFIKPIL	EKMDGTEELL	VKLNREDLLR	KQRTFDNGSI	PHQIHLGELH	420
AILRRQEDFY	PFLKDNRREKI	EKILTFRIPY	YVGPLARGNS	RFAWMTRKSE	ETITPWNFEE	480

-continued

VVDKGASAQS	FIERMTNFDK	NLPNEKVLPK	HSLLYEYFTV	YNELTKVKVY	TEGMRKPAFL	540
SGBQKKAIVD	LLFKTMRKVT	VKQLKEDYFK	KIECFDSVEI	SGVEDRPNAS	LGTYHDLIKI	600
IKDKDFLDNE	ENEDILEDIV	LTLTLFEDRE	MIEERLKYA	HLFDDKVMKQ	LKRRRTGWG	660
RLSRKLINGI	RDKQSGKTAIL	DFLKSDGFAN	RNFMQLIHDD	SLTFKEDIQK	AQVSGQGDSDL	720
HEHIANLAGS	PAIKKGILQT	VKVVDELVKV	MGRHKPENIV	IEMARENQTT	QKGQKNSRER	780
MKRIEEGIKE	LGSQILKEHP	VENTQLQNKE	LYLYYLQNGR	DMYVDQELDI	NRLSDYDVDA	840
IVPQSFLKDD	SIDNKVLTRS	DKNRKSDNV	PSEEVVKKMS	NWWRQLLNAK	LITQRKFDSL	900
TKAERGLLSE	LDKAGFIKRQ	LVTROQITKH	VQJILDSRMN	TKYDENDKLI	REVKVTILKS	960
KLVSDFRKDF	QFYKVREINN	YHHAHDAYLN	AVVGTALIKK	YPKLESEFVY	GDYKVYDVRK	1020
MIAKSEQEIQIG	KATAKFRK	NIMMAFFKTEI	TLANGETRKR	PLIETNGETG	EIVWDKGDRF	1080
ATVRKVLSMP	QVNIVVKTEV	QTGGFSKESI	LPKRNSDKLI	ARKKDWDPKK	YGGFDSPVVA	1140
YSLVLVAKVE	KGKSKKLKV	KELLGITMCE	RSSFEKPNID	FLEAKGYKEV	KKDLI1KLPK	1200
YSLFELENGR	KRMLASAGEL	QKGNELALPS	KYVNFLYLAS	HYEKLKGSP	DNEQKQLFVE	1260
QHKHYLDEII	EQISEPKH	ILADANLDKV	LSAYNKHDRK	PIREQAENII	HLFTLTNLGA	1320
PAAFKYMFDTT	IDRKRYTSTK	EVLDATLQH	SITGLYETRI	DLSQLGGDAY	PYDVPDYASL	1380
GSSGPKKKRK	VEDPKKKRKV	DGIGGSNSG	SGSNGPGGSG	GGGSGGEELL	SKNYHLENEV	1440
ARLKKGSGSG	EELLSKNYHL	ENEVARLKKG	SGSGEELLSSK	NYHLENEVAR	LKKGSGSGEE	1500
LLSKNYHLEN	EVARLKKGSG	SGEELLSKNY	HLENEVARLKK	KGSGSGEELL	SKNYHLENEV	1560
ARLKKGSGSG	EELLSKNYHL	ENEVARLKKG	SGSGEELLSSK	NYHLENEVAR	LKKGSGSGEE	1620
LLSKNYHLEN	EVARLKKGSG	SGEELLSKNY	HLENEVARLKK	KGSGSGEELL	SKNYHLENEV	1680
ARLKKGSGSG	EELLSKNYHL	ENEVARLKKG	SGSGEELLSSK	NYHLENEVAR	LKKGSGSGEE	1740
LLSKNYHLEN	EVARLKKGSG	SGEELLSKNY	HLENEVARLKK	KGSGSGEELL	SKNYHLENEV	1800
ARLKKGSGSG	EELLSKNYHL	ENEVARLKKG	SGSGEELLSSK	NYHLENEVAR	LKKGSGSGEE	1860
LLSKNYHLEN	EVARLKKGSG	SGEELLSKNY	HLENEVARLKK	KGSGSGEELL	SKNYHLENEV	1920
ARLKKGSGSG	EELLSKNYHL	ENEVARLKKG	SGSGEELLSSK	DYHLENEVAR	LKKGSGSGEE	1980
LLSKNYHLEN	EVARLKK					1997

SEQ ID NO: 90	moltype = DNA	length = 5991	
FEATURE	Location/Qualifiers		
source	1..5991		
	mol_type = other DNA		
	organism = synthetic construct		
SEQUENCE: 90			
atggacaaga agtacagcat	cggccgtggcc	atcggcacca actctgtggg ctggggcg	60
atcacccgacg agtacaagg	gcggcaagca	aaattcaagg tgctggccaa caccgacccg	120
cacagcatca aagaagaacct	gatggccgc	ctgctgttc acagccggaaa aacagccgg	180
gccacccggc tgaagagaac	cgccagaaga	agatacacca gacggaaagaa ccggatctc	240
tatctgcaga agatcttcag	caacgagatg	gccaagggtgg acgacagctt cttccacaga	300
ctggaagagt ccttcgttg	ggaaggagat	aqaaggacag acggggcaccc catcttcggc	360
aacatcgtgg acgagggtgc	ctaccacg	aatgtacccca ccatctacca ccttgagaaag	420
aaactgtgtt acagcaccga	caaggccac	ctgccccgtga tctatctggc cctggccac	480
atgatcaagt tccggggcca	cttctgtatc	gaggggcacc tgaacccca caacagcgac	540
gtggacaaggc tgttcatcca	gctgtgtc	acccatcaacc agctgttca ggaaaaacccc	600
atcaacgcga	cgccgtgg	cgccaaaggcc atcctgtctg ccagactgag caagacgaga	660
cggtctggaa atctgtatgc	ccatgtcc	ggcgagaaga agaatggctt gttcggcaac	720
ctgattggcc	tgagcgtgg	cctgacccca aacttcaaga gcaacttca cctggccag	780
gatgcaaac tacgtgtat	caaggacacc	tacgacgacg acctggacaa cctgtgtggc	840
catatcgccg acaggatcgc	cgacccgt	ctggccgccta agaaacctgtc cgacgcctc	900
ctgtgtatcc acatcttgc	agtgttttt	gatgttccca agaaccttgc gggccctct	960
atgatcaaga gatacgtat	gcaccacc	gacctgttcc ttcttcgacc tgctgttca gtcgtgcgg	1020
caicgtatgc ctgtatgtt	caaagatgtt	tttttcgacc agagcaagaa cgggtatcgcc	1080
ggctatcatcg atggccggac	cggccggac	gatgttccatca agtttcatcaa gcccattctg	1140
aaaaatgttggcc accggcaccga	ggaaatgttgc	gtgttccatca agacagaggaa cttgtgtccg	1200
aaggcagegga ctttcgacaa	cgccatgtatc	ccccaccaga tccacccgg agagctgcac	1260
gcattatgtc	ggccggcggac	atgtttttccatca aggacaaaccg ggaaaatgtc	1320
gagaatgttgc tgacccatcg	cattttatcg	catctgtggcc ctgtgttccatca agggaaacacgc	1380
agatgtccgtt ggtatgttgc	ggatgttgc	gaaatgttgc gaaatccatca cccctgttca cttcgaggaa	1440
gtgttccatcg aagggttgc	cgccatgtatc	tttccatcgatc ggtatgttgc cttcgataag	1500
aaacctgttgc accggatccaa	gtgttccatcg	tttccatcgatc ggtatgttgc cttcgataag	1560
tacaacgtatc	tttccatcg	tttccatcgatc ggtatgttgc cttcgataag	1620
agccggcggac	tttccatcg	tttccatcgatc ggtatgttgc cttcgataag	1680
gtgttccatcg taaaatgttgc	tttccatcgatc	tttccatcgatc ggtatgttgc cttcgataag	1740
tccggcgtgg aagatgttgc	tttccatcgatc	tttccatcgatc ggtatgttgc cttcgataag	1800
atcaaggatca	tttccatcgatc	tttccatcgatc ggtatgttgc cttcgataag	1860
ctgaccatcg	tttccatcgatc	tttccatcgatc ggtatgttgc cttcgataag	1920
caccctgttc	tttccatcgatc	tttccatcgatc ggtatgttgc cttcgataag	1980
aggctgtatc	tttccatcgatc	tttccatcgatc ggtatgttgc cttcgataag	2040
gatttccatcg	tttccatcgatc	tttccatcgatc ggtatgttgc cttcgataag	2100
agcctgtatc	tttccatcgatc	tttccatcgatc ggtatgttgc cttcgataag	2160
cacggatcaca	tttccatcgatc	tttccatcgatc ggtatgttgc cttcgataag	2220
gtgttccatcg tggatgttgc	tttccatcgatc	tttccatcgatc ggtatgttgc cttcgataag	2280
atcgatgtatc	tttccatcgatc	tttccatcgatc ggtatgttgc cttcgataag	2340
atcgatgtatc	tttccatcgatc	tttccatcgatc ggtatgttgc cttcgataag	2400
gtggaaaaca cccatgtatc	tttccatcgatc	tttccatcgatc ggtatgttgc cttcgataag	2460
gatatgtatc	tttccatcgatc	tttccatcgatc ggtatgttgc cttcgataag	2520
atcgatgtatc	tttccatcgatc	tttccatcgatc ggtatgttgc cttcgataag	2580

-continued

gacaagaacc	ggggcaagag	cgacaacgtg	ccctccgaag	aggctcgaa	gaagatgaag	2640
aactactggc	gcacatcgct	gaatgccaag	ctgattacc	agaggaaagg	cgacaatctg	2700
accaaggccg	agagaggccg	cctgacgcga	ctggataagg	ccggcttcat	caagagacag	2760
cttgtggaaa	ccggcagat	cacaaggac	gtggcacaga	tcctggactc	ccggatgaa	2820
actaagtacg	acgagaacga	caaactgatc	cgggaaagtga	aagtgtatcac	cctgaagtcc	2880
aagctgtgt	ccgatttccg	gaaggattt	cagttttaca	aagtgcgcga	gatcaacaac	2940
taccaccacg	cccacgacgc	ctacccgtac	gcccgtcg	gaaggccct	gatcaaaaaag	3000
taccctaagc	tggaaagcga	gttgcgtac	ggcactaca	agggtacga	cgtgcggaa	3060
atgatcgcca	agagcgagca	ggaaatcg	aaggctacc	ccaagtactt	cttctacagc	3120
aacatcatg	actttttca	gaccggat	accctggcca	acggcgagat	ccggaaacgg	3180
cctctgtacg	acaaacaaacgg	cgaaacaggc	gagatcg	gggataaagg	ccgggactt	3240
gccaccgtgc	ggaaaagtgc	gtctatccc	caagtata	tcgtaaaaaa	gaccgagggt	3300
cagacaggcg	gtttcagca	agagtctatc	ctgccccaga	ggaacacgca	caagctgatc	3360
gccccaaaaa	aggactgggc	ccctaaaga	ta	ccgcacgccc	caccgtggcc	3420
tattctgtc	tggtgtggc	caaagtgg	aagggc	ccaagaaact	gaagagtgt	3480
aaagagctg	tgggatcac	catatggaa	aaagacq	tcgagaa	ttccatcgac	3540
tttctggaa	ccaaggc	caaagaatg	aaaaggacc	tgatcatca	gctgcctaa	3600
tactccctgt	tcgagctgg	aaacggccgg	aagagaatgc	tggccttc	cggegaact	3660
cagaaggccg	acgactggc	cctggccct	aaatatgt	acttctgt	cctggccagc	3720
cactatggaa	agctgaagg	ctcccccg	gataatgac	agaaacagct	gttggatgg	3780
cagcacaac	actactgg	cgagatcatc	gagcagatc	cgagatc	cgagatc	3840
atcctggccg	acgctaat	ggacaagg	ctgagcgc	acaacaa	cagagacaag	3900
cctatcagag	agcaggccg	gaatcatc	caccc	ccctgtt	tctggagcc	3960
cctggccgc	tcaatgtactt	tgaccc	atcgacc	aggat	caccc	4020
gagggtgt	acgccc	gatccacc	agcatc	ccctgt	gacacggatc	4080
gacctgttc	acttgggagg	cgacgc	ccctatgac	tgcc	ccgt	4140
ggcaggcgc	ccccaaag	aaaaggc	gttgg	ctaa	ggggaaatg	4200
gacggcatt	gttaggg	caacgg	ca	acggat	ccgg	4260
gggggggtt	ctgggg	agaat	ttt	atcat	ttg	4320
gctcgctt	agaaagg	tggc	gat	ggact	tttca	4380
gaaaatgg	tagtact	gaaaagg	ggcgg	gggagg	gttgg	4440
aattatcat	ttgg	gac	atg	ggatc	ttttca	4500
ttactctcg	aaaattatc	tctcg	aa	gaa	ttttcc	4560
tctggggaa	agcttat	taaaact	cac	ctt	ttt	4620
aaggaaatg	tgatgt	tttt	at	ggat	ttt	4680
gcccgtt	agaagg	ttt	cc	ggat	ttt	4740
gaaaatgg	ttcg	tttt	tt	ggat	ttt	4800
aattaccacc	tcgaa	ttaa	at	ggat	ttt	4860
ttactgt	aaaattat	tctcg	aa	gat	ttt	4920
tctgggg	agttgt	aaaa	at	ttt	ttt	4980
aaggcct	gtc	tttt	tt	ttt	ttt	5040
gcccgg	aaaagg	tttt	tt	ttt	ttt	5100
gaaaatgg	ttcg	tttt	tt	ttt	ttt	5160
aactaccacc	tcg	tttt	tt	ttt	ttt	5220
cttttgac	agaact	tttt	tt	ttt	ttt	5280
agcggcgg	attgtc	tttt	tt	ttt	ttt	5340
aaaggcttgc	gttgc	tttt	tt	ttt	ttt	5400
gcccgg	aaaagg	tttt	tt	ttt	ttt	5460
gaaaatgg	ttgg	tttt	tt	ttt	ttt	5520
aactaccact	ttgg	tttt	tt	ttt	ttt	5580
cttttgac	agaact	tttt	tt	ttt	ttt	5640
agcggcgg	attgttac	tttt	tt	ttt	ttt	5700
aaaggcttgc	gttgc	tttt	tt	ttt	ttt	5760
gcccgg	aaaagg	tttt	tt	ttt	ttt	5820
gaaaatgg	ttgc	tttt	tt	ttt	ttt	5880
gattatcat	ttgg	tttt	tt	ttt	ttt	5940
ttctgtcg	aaaattatc	tctcg	aa	gaa	tttt	5991

SEQ ID NO: 91 moltype = AA length = 5
 FEATURE Location/Qualifiers
 source 1..5
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 91 GSGSG
 5

SEQ ID NO: 92 moltype = AA length = 22
 FEATURE Location/Qualifiers
 source 1..22
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 92 GSGSGGGSGSG SG
 22

SEQ ID NO: 93 moltype = DNA length = 20
 FEATURE Location/Qualifiers

-continued

```

source          1..20
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 93
ggcccagtct gcaaagcgag                                20

SEQ ID NO: 94      moltype = DNA  length = 20
FEATURE
source          1..20
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 94
ttatgggcgt agctgaagaa                                20

SEQ ID NO: 95      moltype = DNA  length = 20
FEATURE
source          1..20
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 95
gtatggaggg ctggatctgc                                20

SEQ ID NO: 96      moltype = RNA   length = 20
FEATURE
source          1..20
               mol_type = other RNA
               organism = synthetic construct
SEQUENCE: 96
ggcccagtct gcaaagcgag                                20

SEQ ID NO: 97      moltype = RNA   length = 20
FEATURE
source          1..20
               mol_type = other RNA
               organism = synthetic construct
SEQUENCE: 97
ttatgggcgt agctgaagaa                                20

SEQ ID NO: 98      moltype = RNA   length = 20
FEATURE
source          1..20
               mol_type = other RNA
               organism = synthetic construct
SEQUENCE: 98
gtatggaggg ctggatctgc                                20

SEQ ID NO: 99      moltype = DNA   length = 20
FEATURE
source          1..20
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 99
tacgttctct atcaactgata                                20

SEQ ID NO: 100     moltype = DNA   length = 20
FEATURE
source          1..20
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 100
tatggagggg tggatctgcg                                20

SEQ ID NO: 101     moltype = RNA   length = 20
FEATURE
source          1..20
               mol_type = other RNA
               organism = synthetic construct
SEQUENCE: 101
tacgttctct atcaactgata                                20

SEQ ID NO: 102     moltype = RNA   length = 20
FEATURE
source          1..20
               mol_type = other RNA
               organism = synthetic construct
SEQUENCE: 102

```

-continued

-continued

caggagatcc tggaggagcc ctcccgtccc tcctcgaaat ttgctatgtt accttcattacc	480
ctcaccaaagg aagaaccagt tccacttagat acacaggctcg ttgaggaaat ggaagactca	540
ggtgccccgc ccctgtttt gaccaacccaa cagtccgcac gacggcgatca	600
aaaatgtttt	609
SEQ ID NO: 107 moltype = AA length = 204	
FEATURE Location/Qualifiers	
source 1..204	
mol_type = protein	
organism = synthetic construct	
 SEQUENCE: 107	
MGEAEVGGGG AAGDKGPGEA ATSPAETVV WSPEVEVCLF HAMLGHKPVG VNRHFMICL	60
RDKFSQNIGR QVPSKVIWDH LSTMYDMQAL HESEILPFVN PERNFVLPEE IIQEVRREGKV	120
MIEEEMKEEM KEDVDPHNGA DDFVSSSGSL GKASEKSSKD KEKNSSDLGC KEGADKRKRS	180
RTVDKVLTAN SNPSSPSAAK RRRT	204
SEQ ID NO: 108 moltype = DNA length = 615	
FEATURE Location/Qualifiers	
source 1..615	
mol_type = other DNA	
organism = synthetic construct	
 SEQUENCE: 108	
atggggagagg ccgaggtggg cgccgggggc gcccggggcg acaaggccc gggggaggcg	60
gccaccaccc cggccggagg gacagtgggtt tggagccccc aggtggaggt gtgcctttc	120
cacggcatgc tggggccacaa gcccgctggt gtgaaccgc acttccacat gatttttatt	180
cgggacaaat tcaagccagaa categggcg caggtcccat ccaaggatcat ctggggacat	240
ctggagcacca tgtaacatcat gcaggcgctg catggatgtt agatcttttcc attcccgaaat	300
ccagagggagg acttcgttcc tcccaaggatc atcattcagg aggtccggaa aggaaaaatg	360
atgatagaag aggatgttggaa agggggatgg aagggaaatgg tggaccccca caatggggct	420
gacgatgtttt ttcatcttc agggatgtttt gggaaaggatc cagaaaaatc cagaaagac	480
aaagagaaatc acttccttgc ctgggggtgc aaagaaaggcg cagacaaggcg gaagcgac	540
cgggttcaccc acaaaggctt gaccgaaac agcaaccctt ccagttccatc tgctgcac	600
cgccgcgcgc cgttag	615
SEQ ID NO: 109 moltype = AA length = 468	
FEATURE Location/Qualifiers	
source 1..468	
mol_type = protein	
organism = synthetic construct	
 SEQUENCE: 109	
MATGADVRDI LELGGPEGDA ASGTISKDI INPDKKSKK SSETLTFRP EGMHREVYAL	60
LYSDKKDAPP LLPSPDTGQGY RTVKAKLGSK KVRPWKWMFP TNPARKDGM FFHWRRRAEE	120
GKDYPFARFN KTVQVPVYSE QEQYQLYHDD ANTKAETDHL FDLSRRFDL RFFVHIDRYDH	180
QQFKKRSVED LKERYYYHICA KLANQVRAPVG TDLKIPVFDL GHERRKEQL ERLYNRTPEQ	240
VAEEEYLLQE LRKIEARKKE REKRSQDLQK LITAADTTAE QRTERKAPK KKLPOKKEAE	300
KPAVPETAGI KFPDFKSAGV TLRSQRMKLP SSVGQKKIKA LEQMLLELGV ELSPTPTEL	360
VHMFNELRSD LVLLYBLKQA CANCEYELQM LRHRHEALAR AGVLGGPATP ASGPGPASAE	420
PAVTEPGLGP DPKDTIIDVV GAPLTPNSRK RRESASSSSS VKKAKKPL	468
SEQ ID NO: 110 moltype = DNA length = 1404	
FEATURE Location/Qualifiers	
source 1..1404	
mol_type = other DNA	
organism = synthetic construct	
 SEQUENCE: 110	
atggctacgg gcggcgatgt acgggacatt ctggaaatcg ggggtccaga agggatgc	60
gcctctggaa ccatacgaa gaaggacatt atcaaccccg acaaaaaaa atcaagaag	120
tcctctgaga cactgacttt caagggccc gagggcatgc accggaaatg ctatccctt	180
ctctactctg acaagaaggat tgccatccca ctgttacccca gtgacactgg ccaggatcc	240
cgtacatgtt ggggtccaaat gggtggggcc cttggaaatgtt gatggccatcc	300
accaacccgg ccccaaggat cggggcaatg ttcttccactt ggcggatgtgc aegggaggag	360
ggcaaggact acccccttgc caggttcaat aagactgtgc aggtgectgt gtactcgag	420
caggaggatcc agctttatct ccacatgtat gcttggacta aggcagaaac tgaccac	480
tttggacttca cggccgcgtt tgacccgtt tttgttgtt tccatgtacccg gtatgtaccc	540
caggatgttca agaaggcttc tggggaaatc ctggggggc ggtactacca catctgtgt	600
aaccttgccca acgtgcgggc tggggccatcc acagaccatca agatccatgtt atttgtatgt	660
gggcacgaac gacggggaa ggaacagctt gagctgtctt acaaccggac cccagagcag	720
gtggcagagg aggatgttccctt gctacaggat ctggcgaaga ttgaggcccg gaagaaggag	780
cgggagaaac gcaggccaggat cctggccatcc ctgttccatcc cggccggatcc cactgcaggat	840
caggccggca cggaaacgcggaa ggcccccaaa aagaaggatccccc aacccggatcc ggaggctgat	900
aaggccggat ttccttgcac tggggccatcc acgttccatcc acttcaatgc tggggatgtt	960
acgtgcggat gccaacggat gaatgttccatcc agtctgttgg gacagaagaa gatcaaggcc	1020
ctggaaacaga tgctgttccatcc gcttggatgtt gatgttccatcc cggccggatcc ggaggatgtt	1080
gtggcactgt tcaatgttccatcc gcttggatgtt gatgttccatcc caagcaggcc	1140
tgtggccactt gcgatgttccatcc gcttggatgtt gatgttccatcc gtcatgttccatcc actggccgg	1200
gtggatgttccatcc tggggcccgcc tggccacacca gcatcaggcc caggccggcc ctctgtgtt	1260

-continued

ccggcagtga	ctgaaccgg	acttggctt	gaccccaagg	acaccatcat	tgatgtgg	1320
ggcgcccc	tcacgccc	ttcagaaag	cgcgggagt	cgccctccag	ctcatctcc	1380
gtgaagaaag	ccaagaagcc	gtt				1404

SEQ ID NO: 111	moltype = AA	length = 630
FEATURE	Location/Qualifiers	
source	1..630	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 111		
MAVPPGHGPF SGPGPQEHT QVLPDVRLLP RRLPLAFRDA TSAPLRKLSV DLIKTYKHIN	60	
EVYYAKKKRR QQAPPQDSS NKKKEKVNLH GYDDDNHDYI VRSGERWLER YEIDSILGKG	120	
SFGQVVKAYD HQTQELVAIK IIKNKKAFLN QAQIELRLLE LMNQHDTMK YYIVHLKRHF	180	
MFRNHLCLVF ELLSYNLYL LRNTRFLGRVS LNLTTRKLAAQ LCTALLFLAT PELSIIHCDL	240	
KPENILLRNKP KRSAKIVDF GSSCQLGQRI YQYIQSRFYR SPEVLLGTPY DLAIDMWSLG	300	
CILVEMHTGE PLFSGSNEVD OMNRIVEVLG IPPAAMLDQA PKARKYFERL PGGGWTLLRT	360	
KELRKDYQGP GTRRLQEVLG VQTGGPGGR AGEPEHSPAD YLRFQDLVLR MLEYEPAARI	420	
SPLGALQHGF FRRTADEATN TGPAQSSA STPAPLDTCPS SSTASSISS GGSSGSSSDN	480	
RTYRYSNRYC GGPGPPIITDC EMNSPQVAPP QPLRPWAGGD VPHKTHQAPA SASSLPGTGA	540	
QLPPQPRYLG RPPSPTSPPP PELMDVSLVG GPADCSPPHP APAPQHPPAS ALRTRMTGGR	600	
PPLPPPDDPA TLGPHGLRG VPQSTAASSL	630	

SEQ ID NO: 112	moltype = DNA	length = 1890
FEATURE	Location/Qualifiers	
source	1..1890	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 112		
atggccgtcc caccgggcca tggcccttc tctggcttc cagggccca ggagcacacg	60	
caggtagattgc ctgatgtgcg gctaactgc cgaggctgc ccctggccct ccggatgc	120	
acctcagccc cgctgcgtaa gctctgtgc gacctcatca agacataaa gcacatcaa	180	
gaggtagata atgcaagaa gaagcggcgg gcccagcgg cgccacccca ggattcagc	240	
aacaagaagg aagaaggat cctgaaccat gtttatgtt acgacaaacca tgactacatc	300	
gtgcgcagtgc gcgagcgcgtc gctggagcgc tacgaaattt actcgctcat tggcaaggc	360	
tcccttgcgc aggtggatgaa agcctatcatc gagacaggcc aggagttgt ggccatcaag	420	
atcatcaaga acaaaaaggc ttcttcgaac caggcccaaa ttgagctgcg gctgtggag	480	
ctgtatgaacc agcatgacac ggagatgaa tactatatac tacacatgaa gggcac	540	
atgttccgga accacatgtc cctgttattt gagctgtgtt cctacaactt gtacgactc	600	
ctgcgcacca cccactccg cggcgctcgc ctgaacctcg cccgaaactt ggcgcagc	660	
ctctgcaccc cactgtctt tctggccacg cctgagctca gcatcattca ctgcgactc	720	
aaggccgaaa acatctgtt gtcaacccca aagcgcacgg ccataaagat tggacttc	780	
ggcagctctt gccagcttgc ccagaggatc tccagatatac tccagatgcg ttcttaccgc	840	
tcacccgtgg tgctcttgcg cacccatca gacccatgtt ttgacatgtg gtccctggc	900	
tgcatcttgc tggagatgcg caccggagag ccctcttgcg ttggtccaa tgaggctgac	960	
caagatgacc gcaatgttgc ggtgttggg atccccccgg cccgatgtt ggacccaggc	1020	
cccaaggctc gcaagttactt tgaacggctg cttgggggtt gttggactt acgaaaggac	1080	
aaagaacttca ggaaggatca ccaggcccccc gggacacggc ggctgcaggaa ggtgttggc	1140	
gtcagacaggc ggg	1200	
tacccctggc tccaggaccc ggtgtgcgc atgttggatg atgagccgc cggccgc	1260	
agccccctgg gggctctgca gcacgggttc ttccggcgcg cggccgcacg ggcaccaac	1320	
acggggcccc caggcagcagc tgccctccac ccctcgacac ctgcccctt	1380	
tccagcaccg ccagctccat ctccatgtt ggaggcttca gtggcttc cagtgcacac	1440	
ccggacttacc gtcacacaa ccgtatattgt qggggcccttgg ggccttccatc cacaqactgt	1500	
gagatgaaca gcccccaagg cccacccctcc cagccgcgtc ggccctggc aggggggtgat	1560	
gtccccccaca agacacatca accccctgc tctgcctctg cactgcctgg gacccggggcc	1620	
cagttaaaaa ccaggcccccc atacccctggt cgtccccccat caccacaccc accaccaccc	1680	
cccgagatgc tggatgtgag cctgtgggg ggcctgtgtc actgcctccc acctccacca	1740	
gcgcctggcc cccagcaccc ggctgcctca gcccctccggaa ctccggatgc tggaggctgt	1800	
ccacccctcc cgcctcttgc tgaccctgc actctggggc ctcacctggg cctccgttgt	1860	
gtaccccaaga gcaacacggc cagtcgttgt	1890	

SEQ ID NO: 113	moltype = AA	length = 268
FEATURE	Location/Qualifiers	
source	1..268	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 113		

MNGTANPLLD REEHCLRLGE SFEKRPRASF HTIRYDFKPA SIDTSCEGEL QVGKGDEVTI	60	
TLPHIPGSTP PMTVFKGNKR PYQKDCVLII NHDTGEFVLE KLSSSIQVKK TRAEQSSKIQ	120	
ARMEQQPTRP PQTSQPPP PPPMPFRAPTK PPVGPKTSPN KDNPSPEPQL DDIKRELRAE	180	
VIIEQMSSES SGSSSSDSES SSGSDDDSSS SGGEDNGPAS PPQPSHQQQY NSRPAVANGT	240	
SRPQGSNQLM NALRNLDQLS ESGSDSDD	268	

SEQ ID NO: 114	moltype = DNA	length = 807
FEATURE	Location/Qualifiers	
source	1..807	

-continued

```

mol_type = other DNA
organism = synthetic construct

SEQUENCE: 114
atgaatggaa cgc当地accc gctgtggac cgcgaggaaattgcctgag gctcggggag 60
agcttcgaga agc当地cgcg ggc当地cttc cacacttattttatgatt taaaccagca 120
tctatagaca cttc当地gtga aggagagtt caagttggca aaggagatga agtcacaatt 180
acacttccac atatccctgg atccacacca ccatgactgtgttcaaggaa gaacaaacgg 240
ccttaccaga aagactgtgt gcttattttt aatcatgaca ctggtaatt tttgtggaa 300
aaactcagta gc当地cattca ggtgaagaaa acaagagctt agggcagcag taaaatccag 360
gccc当地atgg aacagcagcc cactc当地tctt ccacagactt cacagccacc accacccca 420
ccacccatgg cttc当地atgg cttccatgg gacccaaac ttctccctt 480
aaagataaacc cttc当地ctgaa acctcagttt gatgacatca aagagagctt gagggctgaa 540
gttgacatta ttgaacaaat gaggc当地gagc agtggggagca gcttcttca gcttggagac 600
tcttc当地ggaa gtgtatggc tagtccatgg agtggaggccagg aacgaaatgg cccagccctt 660
cctccgagc cttc当地acca gca当地cttcc aacagttaggc ctggc当地tgc caatggaaacc 720
agccccccac aaggaaagca cc当地ctatgg aaaaatgactt ggaggatgt 780
gagtctggca gtgacatgtga tgactag 807

SEQ ID NO: 115      moltype = AA length = 311
FEATURE
source          1..311
mol_type = protein
organism = synthetic construct

SEQUENCE: 115
MDLKLKDCEF WYSLHGQVPG LLDWDMRNEL FLPCTTDQCS LAEQILAKYR VGVMKPPEMP 60
QKRRPSPDGD GPCPEPNLWM WVDPNILCPL GSQEAPKPSG KEDLTNISPF PQPPQKDEGS 120
NCSEDKVVES LPSSSEQSP LQKQGIHSPS DFEELTEEEAE EPDDNSLQSP EMKCYQSQKL 180
WQINNNQEKSW QRPPLNCSHL IALALRNNPH CGLSVQEIYN FTRQHFFFW TAPDGWKSTI 240
HYNLCFLDSF EKVPDSLKDE DNARPRSLCW KLTKEGHRRF WEETRVLAF AQRERIQECMS 300
QPELLTSLFD L 311

SEQ ID NO: 116      moltype = DNA length = 936
FEATURE
source          1..936
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 116
atggacttaa aactaaaaga ctgtgaattt tggatagtc tccatggca ggtcccaggg 60
ctgctggact gggactatgg gaatggatgttttgcctt gtaccacaga cc当地gtct 120
ttagctggc aatcccttc当地caaaaatcaga gtc当地ggatgt tgaagcccccc agaaaatgcct 180
caaaagggaa gacc当地gtcc tgatggatgtt ggtcccttctt gtgaacccaa tctgtggatg 240
tgggtggacc ccaatatactt gtggccctt ggc当地ggcagg aggccccaaa gccc当地gtgga 300
aaagggatc tgacaaaatctt ctccatggcc caaaaaaaaa cgaagggtct 360
aactgcttag aggacaaaatgtt ggtatgtctt ctccatgtt gcaatgttctt 420
ttacagaaacg aggatgttca ttccatggatgtt gacttggc当地tca gagaaggctgag 480
gaaccaggacg acaactccctt cc当地atggatgtt gaaatgaaat gttaccaggag cc当地aaacta 540
tggcaatatac acaaccaaga gaaatggctgg caaaaggccctt ctctcaatgt tagccacctt 600
attggccctatg cattaaagaa caaccccaatgttggctca gttgtggatgtt gatctacaat 660
ttcacccggc agcattttccctt ctccatggcc acatggccgg atggctggaa gagcaccatt 720
cattacaacc tctgttcttgg gggatgtt gagaaggctg cagacagctt taaggatgaa 780
gataatgcaa gacccatggc ttgc当地ttgg aagctctacta aggaggggca cc当地ccctt 840
tgggaggaga ctgtgttctt gacccatgtt gaaaggggaa gaatccaaga gttgtatgtt 900
cagccatgtt tctttttttt gtttgc当地tctt 936

SEQ ID NO: 117      moltype = AA length = 484
FEATURE
source          1..484
mol_type = protein
organism = synthetic construct

SEQUENCE: 117
MSGGGPSGGG PGGSGRARTS SFAEPGGGGGG GGGGGPGGSSA SGPGGTGGGK ASVGAMGGV 60
GASSSSGGPG GSGGGGSGGP GAGTSFPPPG VKLGRDGSVK TTIVVATLGQG PERSQEVAYT 120
DIKVIGNGSF GVYYQARLAE TRELVAIKKV LDQKRFKNRE LQIMRKLDHC NIVRLRYFFY 180
SSGEKKDELY LNLVLEYVPV TVYVVARHPT KAKLTIPILY VKVYMYQLFR SLAYIHSQGV 240
CHRDIKPKQNL LVDPDTAVLK LCDFGSAKQL VRGEPNVSYI CSRYYRAPEL IFGATDYTSS 300
IDVWSAGCWL AELLLGQPIF PGDSGVDQLV EIIKVLGTPT REQIREMNPY YTEFKFPQIK 360
AHPWTKVFKS RTPPEAIALC SSLLEYPSS RLSPLEACAH SFFDELRCLG TQLPNNRPLP 420
PLPNFSAGEL SIQPSLNLAIL IPPHLRSPAG TTTLTPSSQA LTETPTSSDW QSTDATPTLT 480
NSSL 484

SEQ ID NO: 118      moltype = DNA length = 1452
FEATURE
source          1..1452
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 118

```

-continued

atgagcggcg	cgggcccttc	gggaggcggc	cctgggggt	cgggcaggc	gcggactagc	60
tgcgttcggg	agcccccggg	cggaggcggg	ggaggccccgg	ggggccccgg	aggctcgccc	120
tccggcccaag	cgccacccgg	cgccggaaag	gcacatctgcg	ggggcatggg	tggggggcgtc	180
ggggccctcg	gtccgggggg	tggaccggc	ggcagcggcg	gaggaggcag	cggaggcccc	240
ggcgcaggca	ctagtttccc	gcccgggggg	gtgaagctgg	gcgtgacag	cggaaaggtg	300
accacagtgc	tagccactct	agggcaaggc	ccagagcgct	ccaaagaagt	ggcttacacg	360
gacatcaaag	tgattggcaa	tggttcattt	ggggtcgtgt	accaggcaacg	gttgcagag	420
accaggaaac	tagtcggccat	caaaagggtt	cttcaggaca	agagggttcaa	gaaccggagag	480
ctgcagatca	tgcgtaaagct	ggaccactgc	aatattgtga	ggctgagata	cttttctac	540
tccagtggcg	agaagaaaga	cgggtttac	ctaaatctgg	tgctggaaa	tgtgcccgg	600
acagtgtacc	gggtggcccg	ccacttcac	aaggccaaatg	tgaccatccc	tatctcttat	660
gtcaagggtt	acatgtacca	gctttcccg	agttggccct	acatccactc	ccaggcgctg	720
tgtcaccgcg	acatcaagcc	ccagaacctg	ctgggtggacc	ctgacactgc	tgtctcaag	780
ctctgcgatt	ttggcagtgc	aaagcagttt	gtccgggggg	agcccaatgt	ctccatcac	840
tgttctcgct	actaccggcc	cccagagctc	atctttggag	ccactgatata	cacccatcc	900
atcgatgtt	ggtcagctgg	ctgtgtactg	gcagagctcc	tcttggccca	gccccatctt	960
cctggggaca	gtgggggttga	ccagctgggt	gagatcatca	aggtgtggg	aacaccaacc	1020
cgggaaacaaa	tccgagatg	gaaccccaatc	tacacggagt	tcaagttccc	tcagattaaa	1080
getcaccctt	ggacaaagggt	gttcaaatct	cgaaacgcggc	caggggccat	cgcgctctgc	1140
ttagcctgc	tgggatcac	ccatccctca	aggctctccc	cactagaggc	ctgtgcgcac	1200
agtttcttg	atgaactgcg	atgttggga	accagctgc	ctaacaaccc	cccacttccc	1260
cctctcttca	atccatgtgc	tgggtgaactc	tccatccaa	cgttctctaa	cgccatttctc	1320
atccctctt	acttgagggtc	cccagcgggc	actaccaccc	tcacccctgc	ctcacaagct	1380
ttaactgaga	ctccgaccag	ctcagactgg	cagtcgaccg	atgcccacacc	taccctca	1440
aactccctct	tg					1452

SEQ ID NO: 119	moltype = AA length = 244	
FEATURE	Location/Qualifiers	
source	1..244	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 119		
MTGIAASFF SNTCRFGGCG	LHFPTLADLI EHIEDNHIDT DPRVLEKQEL QQPTYVALSY	60
INRFMTDAAR REQESLKKI	QPQLSLTLSS SVSRGNVSTP PRHSSGSLTP PVTPIPSS	120
SFRSSTPTGS EYGEEEVDYE	ESDSDESWTT ESAISSEAIL SSNCMNGGEE KPFACPVPGC	180
KRYKNVNGI KYHAKNGHRT	QIRVRKPKFC RCGKSYKTAQ GLRHHTINFH PPVSAEIRK	240
MQQL		244

SEQ ID NO: 120	moltype = DNA length = 732					
FEATURE	Location/Qualifiers					
source	1..732					
	mol_type = other DNA					
	organism = synthetic construct					
SEQUENCE: 120						
atgacaggca	tcggccggc	ctccttcttc	tccaataacct	gcccatttcg	gggctcgccg	60
ctccacttcc	ccaccctggc	cgacccatc	gagcacatcg	aggacaacca	catcgataca	120
gatccacggg	ttttagaaaa	acaacaaat	cacggccggc	cctatgttgc	cctgagttac	180
ataaaatagat	tcatgacaga	tgctgccccg	cgagagcggg	agttccctaa	gaagaagatt	240
cacccggaa	tctcgctgac	tctgtccago	tcagtttctc	gagggaaatgt	gttccactcc	300
ccacggccaca	gcagtgaaag	ccttactccc	cccggtgacc	cacccatcac	ccccctctct	360
tcatcgccca	gcagcactcc	gacggccggc	gagttatggcc	aggaggaggt	ggactatagag	420
gagtcggaca	gcgatgatgc	ctggaccaca	gagatgtggca	tcagtcggc	agccatctc	480
agttccatgt	gcataatgg	aggggaaatgg	aaggcttttgc	cctgccccatgt	tcctgtatgt	540
aaaaagagat	acaagaatgt	gaatggcata	aagtatcacg	ctaagaatgg	tcacagaaca	600
caagattcg	tccgcaaaacc	atccaatgt	cgttggggaa	agagttacaa	gacagctcag	660
ggcctgeggc	accacacaat	caatccat	ccccctgt	cggctgagat	tatcggaaag	720
atgcagaaat	tg					732

SEQ ID NO: 121	moltype = AA length = 611	
FEATURE	Location/Qualifiers	
source	1..611	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 121		
MPRRKRNRAGS SSDGTEDSDF	STDLEHTDSS ESDGTSRRSA RVTRSSARLS QSSQDSPV	60
NLQSFGTEEP AYSTRRVTRS	QQQPTPVTPK KYPLRQTRSS GSETEQVDF SDRETQNTAD	120
HDESPPRPT GNAPSSESIDI	DISSPNVSHD ESIAKDMMSLK DSGSDLSHRP KRRRFHESYN	180
FNMKCPTPGC NSLGHLTGKH	ERHFSISGCP LYHNLSADEC KVRAQSRDKQ IEERMLSHRQ	240
DDNNRHATRH QAPTERQLRY	KEKVAELRK RNSGLSKEQK EKYMHEHQTY GNTREPLLEN	300
LTSEYDLDLF RRAQARASED	LEKLRLQGQI TEGSNMIKTI AFGRYELDTW YHSPYPPEYA	360
RLGRLYMCEF CLKYMKSQTT	LRRHMAKCVW KHPGDEIYR KGSISVFED GKKNKIYCQN	420
LCLLAFLKFLD HKTLYYDVEP	FLFYVMTAED NTGCHLIGYF SKEKNSFLNY NVSCILTMPQ	480
YMRQGYGKML IDFSYLLSKV	EKVGSPERP LSDLGLISYR SYWKEVLLRY LHNFGQKEIS	540
IKEISQETAV NPVDIVSTLQ	ALQMLKYWKG KHLVLKRQDL IDEWIAKEAK RSNSNKTMDP	600
SCLKWTPPKG T		611

-continued

-continued

```

ctgcagatgc agctgcagaa gtgcgagata ctgcagtcgg actcccgctg caaggactac 900
cttgtcaaga ttctcgagga gtcacccctg cacaaggccca cgcagggtat gcccgcgg 960
gcgcggcaagg tggggcgcct gatctacacc gcggggcggt acttccgaca gtcgccteagc 1020
taacctggagg cttaaacccc cagtgcaccc acctggctcc gggtggcgga cttgcagggtg 1080
ccgcggagcg gcctggccgg ctgcgtggtg ggccggctgt tgtacgcgt gggccgcagg 1140
aacaactccgc cccgacggcaa caccgactcc acggccctgg actgttacaa ccccatgacc 1200
aatcagtgggt cgcctcgccccc atccatgago gtgcggccgtt accgcattgg ggtgggggtc 1260
atcgatggcc acatctatgc cgtcgccgg tccccacggct gatccacca caacagtgt 1320
gagaggtatg agccagagcg ggatgagtgg cacttggtg ccacaatgtt gacacgaagg 1380
atcggtgggg gctgtggctgt cctcaatgtt ctctttatgtt cgctgggggg ctttgacggg 1440
acaacccggcc ttatttcgc tgatgttttacccagaga ggaacgagtg gctgttgc 1500
acaccaatgaa acaccatccg aacggggcca ggctgtcgcc ttctgcacaa ctgtatctat 1560
gtctgtgggg gctatgtgg tcaggaccag ctgaacagcg tggagcgcta cgatgtggaa 1620
acagagacgt ggacttctgtt acgcggccatgg aagcaccggc gaagtgcctt gggatcaact 1680
gtccaccagg ggagaatcta cgttggatgtt ggttatgtat gtcacacgtt cttggacagt 1740
gttggatgtt acgaccaggaa tacaacacc tgaggacggg tgacccggat gacatgggc 1800
cgaggatgggg tggggcggtt gttcaccatg gagccctgcc ggaaggat gtagccgg 1860
aactgttaccc tttttgc 1875

```

SEQ ID NO: 125 moltype = AA length = 191

FEATURE Location/Qualifiers
source 1..191
mol_type = protein
organism = synthetic construct

SEQUENCE: 125

```

MAMHNKAAPP QIPDTRRELA ELVKRKQELA ETLANLEROI YAFEGSYLED TQMYGNIIIRG 60
WDRYLTNQKN SNSKNDRRNR KFKEAERLFS KSSVTSAAV SALAGVQDQL IEKREPSSGT 120
ESDTSPDFHN QNEPQSQEDP EDLDGSVQGV KPQKAASSTS SGSHHSSHKK RKNKNRHRID 180
LKLNKKPRAD Y 191

```

SEQ ID NO: 126 moltype = DNA length = 576

FEATURE Location/Qualifiers
source 1..576
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 126

```

atgcgtatgc acaacaaggc ggccgcggccg cagatccggg acacccggcg ggagctggcg 60
gagctcgta acggaaagca ggagctggcg gaaacattgg caaatttgg ggcacagatc 120
tatgccttgg aggaaagcta cctggaaagacttccatgtt atggcaatatttgcgtggc 180
tggatcggtt atctggccaa cccaaaaaaaatccaaatagca aaaaatgtcg aaggacccgg 240
aaatggaaagg aagctgagcg gctttcaatggccatgtt aaatcctcggtt acgtgcagtg 300
agtgcattgg caggaggatca ggaccagatc attggaaaga gggagccagg aagtggacg 360
gaaatggaca ctttcggccatgtt cttccacaat caggaaaaatggccatgtt ggaggaccc 420
gaggatctgg atggatctgtt gcaaggatgtt aaacctcgatgtt ggaggaccc 480
tcaggagtc accacagcgccatggaaatggccatgtt aaaaaggatc aaaaaggatc 540
ctgaagttaa acaaaaaacc acgagctgatgtt 576

```

SEQ ID NO: 127 moltype = AA length = 462

FEATURE Location/Qualifiers
source 1..462
mol_type = protein
organism = synthetic construct

SEQUENCE: 127

```

MGAVNPPLSQ AEESHTEPDL EDCSFRCRGT SPQESLSSMS PISSLPALFD QTASAPCGGG 60
QLDPAAPGTT NMEQLLEKQK DGEAGVNIVE MLKALHALQK ENQRLOEQIL SLTAKKERLQ 120
ILNVQLSVPF PALPAALPAA NGPVPGPYGL PPQAGSSDSL STSKSPPGKS SLGLDNLSLT 180
SSEDDPHSGCP SRSSSSLSFH STPPPLPLLQ QSPATLPLAL PGAPAPLPPQ PQNGLGRAPG 240
AAGLGAMPMA EGLGGLAGS GGLPLNLGLG GLNGAAAPNP ASLSQAGGAP TLQLPGCLNS 300
LTBQQRHLQ QOEQQQLQQLQ QLLASPQLTP EHQTVVVYMI QQIQQQRELQ RLQMAGGSQL 360
PMASLLAGSS TPLLSAGTPG LLPTASAPPL LPAGALVAPS LGNNNTSLMAA AAAAAAVAAA 420
GGPPVLTAAQT NPFLSLSGAE GSGGGPKGGT ADKGASANQE KG 462

```

SEQ ID NO: 128 moltype = DNA length = 1389

FEATURE Location/Qualifiers
source 1..1389
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 128

```

atgggtgcgc ttaatccccctt cctctcccaa gctgagacca gccacacaga gccagacactg 60
gaggactgca gttccgggtg tggggggacc tccctcagg aggtctgtt ttccatgtcc 120
cccatcgac gctccccccgc acttttcgac cagacacggctt ctgcacccctg tggggggccg 180
cagtttagacc cggcgcccccc agggacgact aacatggaccc agttctggtaa gaagcaggcc 240
gacggggagg cccggcgatca catgtggatgtt gatgttgcggcc gctgcacccg 300
gagaaccaggc ggctgcaga gcaatccctg agccctgcggcc cccaaaaaggatc ggggtgcgg 360
attctcaaccc tgcagatcttc tttcccttc cttccatgtcc ctgcacccctt gcttgcggcc 420
aacggccctggcc tccctggccctt cttccatgtcc cccaaaaaggatc ggggtgcgg 480

```

-continued

agcaccaggca	agagccctcc	gggaaagago	agcctcgccc	tggacaactc	gtgttccact	540
tcttctgagg	acccacactc	aggctgccc	agccgcacgc	gctcgtcgct	gtccttcac	600
agcacgcccc	caccgctgcc	cctctccag	cagagccctg	ccactctgcc	cctggccctg	660
cctggggcc	ctgccccact	cccgcggcc	ccgcagaacg	ggttggcccg	ggcaccggg	720
gcagcgggcc	tggggggccat	gccccatggct	gaggggctgt	tggggggct	ggcaggcagt	780
ggggggctgc	ccctcaatgg	gtctcttggg	gggttgaatg	ggggccgtgc	ccccaacccc	840
gcaagctgtga	gcaggctgg	cggggccccc	acgtgcage	tgcaggctgt	tctcaacagc	900
cttacagagc	agcagagaca	tctcttcag	cagcaagagc	agcagctcca	gcaactccag	960
cagctctgg	ccctcccgca	gtgtaccccg	gaacaccaga	ctgttgtcta	ccagatgatc	1020
cagcagatcc	agcagaaaacg	ggagctgcag	cgccgtcaga	tggatgggg	ctcccaactg	1080
cccatggcca	gcctgtggc	aggaatggc	accccgctgc	tgtctgggg	tacccctggc	1140
ctgctgcccc	cagegtctgc	tccacccctg	ctgccccctg	gagccctagt	ggctccctcg	1200
cttggcaaca	acacaagtct	catggccgca	gcagctgcag	ctgcagcagt	agcagcagca	1260
ggcggaccc	cagctctcac	tgcctccagcc	aaccccttcc	tcagctgtc	gggaggcagag	1320
gcaactggcg	gtggcccaa	aggagggacc	gtgtacccaa	gagcctcagc	caaccaggaa	1380
aaaggctta						1389

SEQ ID NO: 129 moltype = AA length = 288
 FEATURE Location/Qualifiers
 source 1..288
 mol_type = protein
 organism = synthetic construct
 SEQUENCE: 129
 MSSRKQGSQP RGQQSAAEEN FKKPRTSNMQ RSKMRGASSK KKTAGPQQKN LEPALPGRWG 60
 GRSNAEPPSG SVRKTRKNKQ KTPGNGDGGS TSEAPQPPLR KRARADPTVE SEEAFKNRME 120
 VVKVKEPEELK PWLVEDWDLV TRQKOLFQLP AKKNVDAILE EYANCKKSQG NVDNKEYAVN 180
 EVVAGIKEYF NVMLGTQLLY KFERPQYABE LLAHPDAPMS QVYGAPHLLR LFVRIGAMLA 240
 YTPLDEKSLA LLGYLHDFL KYLAKNSASL FTASDYKVAS AEYHRKAL 288

SEQ ID NO: 130 moltype = DNA length = 867
 FEATURE Location/Qualifiers
 source 1..867
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 130
 atgagttcca gaaagcaggg ttctcaacct cgtggacagc aatctgcaga agaagagaac 60
 ttcaaaaaac caactagaag caacatgcag agaaataaa tgagagggc ctcctcagga 120
 aagaagacag ctggtcaca gcagaaaaat cttgaaccag ctctccagg aagatgggg 180
 ggtcgctctg cagagaaccc cccttcagg tccgtgagga agaccaaaa gaacaagcag 240
 aagactcttg gaaacggaga tgggtggcgt accagcgaag cacctcaggcc ccctcgaaag 300
 aaaaggggcc gggcajaccc cactgttga agtggaggagg cgtttaagaa tagaatggag 360
 gttaaagtga agattcctga agaattaaaa ccatggcttg ttgaggactg ggacttagtt 420
 accaggcaga acgactgtt tcaactccct gccaagaaaa atgtatgtc aatctggag 480
 gagtatgcaaa atgcaagaaa atcgcaggga aatgttgata ataaggaaata tgccgttaat 540
 gaagtttggc caggaataaa aagatatttcc aatgtgtatg tgggcactca gtcgtctac 600
 aaatttgaga ggcggccatgt tgctgaaatc ctcttggctc accctgtatgc tccaaatgtcc 660
 caagttttagt ggcggccatgt tgctgaaatc ttatgttga aatgtggagc aatgttggcc 720
 tatacgcggc ttgtatgaa aagcttgcgat tttatgttgg gctatgttca tgatttctca 780
 aaatatctgg caaagaaatc tgcatctctc tttactgcca gtgattacaa agtggcttct 840
 gctgagttacc accgcaaaatc cctgtga 867

SEQ ID NO: 131 moltype = AA length = 335
 FEATURE Location/Qualifiers
 source 1..335
 mol_type = protein
 organism = synthetic construct
 SEQUENCE: 131
 MSTE GGFFGGT SSSDAQQSLQ SFWPRVMEEI RNLTVKDFRV QELPLARIKK IMKLDEDVKM 60
 ISAEAPVLFKA KAAQIFITEL TLRAWIHTED NKRRTLQRND IAMAITKFQD FDFLIDIVPR 120
 DELKPKPQRE EVRQSVTPAE PVQYYFTLAQ QPTAVQVQGQ QQQQQTTST TTIQPGQII 180
 AQPQGQGTTP VTMQVGEQQQ VQIVQAQPQG QAQQAQSGTG QTMQVMQII TNTGBIQQIP 240
 VQLNAGQLQY IRLAQPVSGT QVVGQIQT ATNAQQITQT EVQOQQQFS QFTDGQQLYQ 300
 IQQVTMPAGQ DLAQPMFIQS ANQPSDGQAP QVTGD 335

SEQ ID NO: 132 moltype = DNA length = 1008
 FEATURE Location/Qualifiers
 source 1..1008
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 132
 atgtccacag aaggaggatt tgggttact agcagcagtg atgcccagca aagcttacag 60
 tctttctggc ctgggtcat ggaagaaatc cggatattaa cagtggaaat cttccggatgt 120
 cagggaaactcc cactggctcg tattaagaag attatgaaac tggatggaaat tggatggatgt 180
 atcagtgcag aagcgctgt actctttcc aaggcagccc agatgtttat cacagatgtt 240
 atcttcgag cctggattca cacagaatg aacaagcgcg ggactctaca gagaatgtat 300
 atcgccatgg caattacaaa atttgtatc tttactgatgt tttatgttcc 360

-continued

gatgaactga	aacctccaaa	gctgtcaggag	gagggtgcgc	agtctgtaac	tcctgtccggag	420
cacgtccagt	actatttcac	gctgtcgtcag	caacccaccc	ctgtccaagt	ccagggcccg	480
caccaaggcc	agaagaccac	cagtcacac	accaccatcc	agcctgggca	gatcatcata	540
gcacagcc	acgaggccca	gaccacac	gtgacaatgc	aggttgaaaa	aggccagcg	600
gtgcagattg	tccaggctca	gcccacagggt	caagccaaac	aggcccagag	tggcactgg	660
cacaccatgc	aggtgtatgc	gcaatcata	actaacacag	gagagatcca	gcagatcccg	720
gtgcagctga	atgcggccca	gtgcgtat	atccgcgtta	ccccactgt	atcaggcact	780
caagttgtgc	agggacagat	ccacacactt	gccaccaatg	ctcaacagat	tacacagaca	840
gagggtccagc	aaggacagca	gcagtccago	cagttcaca	atggacagca	gctctaccag	900
atccagcaag	tcacatgcc	tgccggccag	gacctcgccc	agcccatgtt	catccagtc	960
gccaaccaggc	cctccgacgg	gcaggcccc	caggtgaccc	gcaactgt		1008

SEQ ID NO: 133	moltype = AA	length = 791				
FEATURE	Location/Qualifiers					
source	1..791					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 133						
MEEKRRKYSI	SSDNSDTDS	HATSTSASRC	SKLPSSTKSG	WPRQNEKKPS	EVFRDLITA	60
MKIPDSYQLS	PDDYYILADP	WRQEWKEGVQ	VPAAGAEAIPE	PVVRILPPL	GPPAQASPSS	120
TMLGEGSQPD	WPGBSRYLDL	EIDAYWLELI	NSELKEMERP	ELDELTLE	LEELETLCQ	180
NMARAIETQE	GLGIEYDEDV	VCDVCRSP	EDGNEVMFCD	KCNVCVHQAC	YGILKVP	240
WLCRTCALGV	QPKCLLCPKR	GGALKPTCRSG	TKWVHVSCAL	WIPEVSIGCP	EKMEPITKIS	300
HIPASRWALS	CSLCKEKTGT	CIQCSMPSCV	TAPHVTCAFD	HGLEMRILA	DNDEVKFKSF	360
CQEHSDDGPR	NEPTSEPTEP	SQAGEDLEKV	TLRKQLRLQQL	EEDFYELVEP	AEVAERLDIA	420
EALVDFIYQY	WKLKRKANAN	QPLLTPKTDE	VDNLAQQBQD	VLYRRLKLFT	HLRQDLERVR	480
NLCYVMVTRRE	RTKHAICKLQ	EQIFHQLQML	IEQDLCRAGL	STSFPIDGTF	FNSWLAQSVQ	540
ITAENNMAMSE	WPLNNNGHRED	PAPGLLSEEL	LQDEETLLSF	MRDPSSLRPGD	PARKARTR	600
LPAKKKKPPP	PPQDGPGSRT	TPDKAPKKTW	GQDAGSGKGK	QGPPTRKPR	RTSSHLPSSP	660
AAGDCPILAT	PESPPPLAPE	TPDEAASVAA	DSDVQVPGPA	ASPKPLGR	PPRESKVTR	720
LPGARPDAGM	GPPSSAVAERP	KVSLHFDTET	DGYFSDGEMS	DSDVVAEDGG	VQRGPREAGA	780
EEVVRMVGVLAS						791

SEQ ID NO: 134	moltype = DNA	length = 2376				
FEATURE	Location/Qualifiers					
source	1..2376					
	mol_type = other DNA					
	organism = synthetic construct					
SEQUENCE: 134						
atgaaagaga	agaggcgaaa	atactccatc	agcagtgaca	actctgacac	cactgacagt	60
catgcacat	ctacatccgc	atcaagatgc	tccaaactgc	ccagcacac	caagtcggc	120
tggccccgac	agaacgaaaa	gaagccctcc	gagggtttcc	ggacagac	tttgcacagcc	180
atgaagatcc	ccggacttata	ccagctcagc	ccggatgtact	actacatct	ggcagaccca	240
tggcgacagg	aatggggaaa	aggtgtcg	gtgcgtccg	gggcagaggc	catccagag	300
cccggtgtga	ggatctccc	accactggaa	ggcccccttc	cccaggatc	cccggacgc	360
accatgttg	gtgagggtc	ccagctgtat	tggccagggg	gcagccgcta	tgacttggac	420
gagattgtat	cctactggct	ggagctatc	aactcgagg	ttaaggat	ggagaggccg	480
gagctggacg	agctgacatt	agagctgtgt	ctggaggaggc	ttggagaccc	gtgcacac	540
aatatggcoca	ggggccattgt	gacgacagg	gggttggca	tcgagatcga	cgaggatgtt	600
gtctgcgacg	tgtgtcgctc	tcctgagggc	gaggatggc	acgagatgtt	cttctgtgac	660
aagtgcacac	tctgtgtgc	tcagggatgc	tcagggatcc	tcaagggtgc	cacggccac	720
tggctgtgc	ggacgtgtgc	cctgggtgtc	cagccaaatg	gcctgtctg	ccccaaagcg	780
ggaggagcgt	tgaagccac	tagaagtgg	accaaatggg	tgcatgtcag	ctgtgcctta	840
tggatttcgt	aggtcagcat	cggtgtccca	gagaagatgg	agcccatcac	caagatctcg	900
catatatcccg	ccagccgtgt	ggctatgtcc	tgcaagcttc	gcaaggaaatg	cacaggcacc	960
tgcatccatgt	tttccatgtc	acagcggttc	atgtcacatg	cgcttttgac	1020	
cacggcttgg	aaatgcggac	tatattagca	gacaacatgt	aggtaatgtt	caagtcattc	1080
tgccaggagc	acagtgtacgg	ggggccacgt	aatgagccca	catctgagcc	cacggaaaccc	1140
agccaggcgt	ggggggacgt	ggaaaagggt	accctggccca	agcagccgt	gcagcagcta	1200
gaggaggact	tctacatgt	gggtggagcc	gctgggggtt	ctgggggggt	ggacctggct	1260
gaggcactgg	tccacttcat	tcacccatgt	tggaaatgt	agagggaaatg	caatgcacac	1320
cacggcgtgc	tgaccccaa	gaccgacgg	gtggacaacc	tggccagcga	ggaggcaggac	1380
gtctcttacc	ggccgtctgaa	gcttttacc	catctgccc	aggacataga	gagggttaga	1440
aatctgtgt	acatgtgtac	aaggccgtac	agaaacgaaatc	acgcacatctg	caaacttcac	1500
gaggcataat	tccacatgtca	gatgaaatct	attgaaacagg	atctgtgtcg	agcaggctcg	1560
tccacccat	tccccatgtca	tggcaccctc	tcaacacgt	ggctggcaca	gtcggtgc	1620
atcacacatgt	agaacatggc	catgagcgg	tggccactga	acaatgggca	ccggcaggac	1680
cctgcttcac	ggctgtgtgc	agaggaaatgt	ctgcaggac	aggagacact	gctcgttc	1740
atgcgggacc	cctcgctgc	acctgtgtac	ctgtctggaa	aggccccagg	ccgcacccgc	1800
ctgcctgtca	agaagaaacc	accaccatca	ccacccgcagg	acggggctgg	ttcacggac	1860
actccacatgt	aaggccccaa	gaagacgtgg	ggccaggatg	caggcgtgg	caaggggggt	1920
caagggccac	ctaccaggaa	gccaccacgt	cggtacatctt	ctcaactgtcc	gtccagccct	1980
gcagccgggg	actgtccat	cctagccacc	cctgaaagcc	ccccggccat	ggccccctgag	2040
accccgaggc	aggcagccctc	agtagctgt	gactcagatg	tccaaatgtcc	tggccctgca	2100
gcaagcccta	agcccttggg	ccggctccgg	ccaccccgcc	agagcaaggt	aaccggaga	2160
ttgcgggggt	ccaggcctgt	tgctgggtat	ggaccacatc	cagctgtggc	tgagaggccc	2220

-continued

```

aaggctcagcc tgcatttga cactgagact gatggctact tctctgtatgg ggagatgagc 2280
gactcagatg tagaggccga ggacgggtggg gtgcagccgg gtcggccggaa ggcaggggca 2340
gaggaggtgg tccgcatggg cgtactggcc tcctaa 2376

SEQ ID NO: 135      moltype = AA length = 84
FEATURE           Location/Qualifiers
source            1..84
mol_type = protein
organism = synthetic construct

SEQUENCE: 135
MRTDSSKMTD VESGVANFAS SARAGRRNAL PDIQSSAATD GTSDLPLKLE ALSVKEDAKE 60
KDEKTTQDQL EKPQNNEEKCP TFLY 84

SEQ ID NO: 136      moltype = DNA length = 252
FEATURE           Location/Qualifiers
source            1..252
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 136
atgaggacag attcatcaa aatgactgac gtggagtctg gggcgccaa ttttgcatt 60
tcagcaaggc caggccggc gaatgcctta ccagacatcc agagttcagc tgccacagac 120
ggaacacctg atttgcctt caaactggag gctctctccg tgaaggaa tgcaaaaagag 180
aaagatgaaa aaacaacaca agaccaattt gaaaagcctc aaaatgaga aaaaatgcca 240
actttcttgtt ac 252

SEQ ID NO: 137      moltype = AA length = 117
FEATURE           Location/Qualifiers
source            1..117
mol_type = protein
organism = synthetic construct

SEQUENCE: 137
MAAAAAGSG TPREEEVPG EAAASQPQAP TSVPGARLRS LPLARVKALV KADPDVTLAG 60
QEAIIFILARA AELFVETIAK DAYCCAQQGK RKTLCRRDLD NAIEAVDEFA FLEGTL 117

SEQ ID NO: 138      moltype = DNA length = 354
FEATURE           Location/Qualifiers
source            1..354
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 138
atgcggccgg cggccggccgc aggaagccggg acgccccggg aggaggaggt acctgctggg 60
gaggccagccg cctcgccagcc ccaggccccca acgagttgc ctggggctcg tctctcgagg 120
ttgcctctgg cgcgactgaa ggcgttggtg aaggccatgc ccgacgtgac gctacgggaa 180
caggaagccca ttccatttc ggcacgagcc gccggactgtt ttgtggagac cattgcaaaa 240
gtatgcctact gttgcgctca gcaggggaaaa agggaaaacccttcagaggag agacttggat 300
aatgcataat aagctgtgga tgaatttgc tttctggaa gtactttttaga ttga 354

SEQ ID NO: 139      moltype = AA length = 761
FEATURE           Location/Qualifiers
source            1..761
mol_type = protein
organism = synthetic construct

SEQUENCE: 139
MPNFCAAPNC TRKSTQSDLA FFRFPDRPAR CQKWWVNCRR ADLEDKTPDQ LNKHYRLCAK 60
HFTETSMICRT SPYRTVLRDNA APIFTFDLTS HLNNPHSRHR KRIKELSEDE IRTLQKKID 120
ETSEQEQKHK ETNNNSNAQNP SEEEGEGQDE DILPLTLEEK ENKEYLKSLF EILILMGKQN 180
IPLDGHEADE IPEGLFTPDM FQALLECRIN SGEEVLRKRF ETTAVNTLFC SKTQRQMLE 240
ICESCIREET LREVRDSSHFF SIITDDVVDI AGEELPVLV RFVDESHNLR EEFIGFLPYE 300
ADEAELAVFKH HTMITEKWKH NMEEYCRGQAY IVSSGFSSKM KVVASRLIEK YPQAIYTLC 360
SCALNMWLAK SPTVMGVSVL LGTIEEVCSF FHRSPQLLLE LDNVISVLFQ NSKERGKELK 420
EICHQSOWTGR HDAFEILVLAQ LQALVLCLDG INSDTNIRWN NYIAGRAFVL CSAVSDFDI 480
VTIVVLKNVL SFTRAFGKNL QGQTSVDFFA AGSLTAVLHS LNEVMENIEV YHEFWFEEAT 540
NLATKLIDIQM KLPGKEPRAH QGNLSQLTS ESYKETLSV PTVEHHIQEL KDIFSEQHLK 600
ALKCLSLVPS VMGQLKFNTS EEEHHADMYRS DLNPNDTLSA ELHCWRIKWK HRGKDIELPS 660
TIYEALHLPD IKFFPNVYAL LKVLCLIPVM KVENERYENG RKRKAYLRLN TLTDQRSSNL 720
ALLNINFEDIK HDLDMVDTY IKLYTSKSEL PTDNSETVEN T 761

SEQ ID NO: 140      moltype = DNA length = 2286
FEATURE           Location/Qualifiers
source            1..2286
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 140
atgccgaact tctgcgctgc ccccaactgc acggggaa gacgcgcgtc cgacttggcc 60
ttcttcaggat tcccgccggc ccctgcccaga tgccagaatgggtggagaa ctgtaggaga 120
gcagacttag aagataaaac acctgtatcg ctaaataaaatattatcgatt atgtgc 180

```

-continued

cattttgaga	cctcttatgtat	ctgtagaact	agtccctata	ggacagttct	tcgagataat	240
gcataatccaa	caatatttgc	tcttaccatgt	catttgaaca	accacatag	tagacacaga	300
aaacgaataaa	aagaacttag	tgaagatgaa	atcaggacac	tgaacagaa	aaaaatttgat	360
gaaaacttctg	agcaggaaaca	aaaacataaa	gaaaccaaca	atagcaatgc	tcagaacccc	420
agcgaagaag	agggtgaaagg	gcaagatgag	gacattttac	ctctaaccct	tgaagagaag	480
gaaaacaaag	aatacctaaat	atctcttattt	gaaatcttgc	ttctgtatgg	aaagcaaaac	540
atacatctgg	atggacatgt	ggctgtatgg	atccccagaag	gtctctttac	tccagataac	600
tttcaggcac	tgctggatgt	tcggataaat	tctggtaaag	agggttctgg	aaagcggtt	660
gagacaacag	cagttaacac	gttggggatgt	tcaaaaacac	agcagaggca	gatgttagag	720
atctgtgaga	gtgttatgt	agaagaaact	ctcaggaaag	tgagagactc	acacttctt	780
tcattatca	ctgacatgt	agtggacata	gcaggggaaag	agcacaatcc	tgtgttgg	840
agggttgg	atgaatctca	taacccaaga	gaggattta	taggttctt	gccttatgaa	900
gccgatcg	aaatttggc	tgtgaaattt	cacactatga	taactgagaa	gtggggatta	960
aatatggagt	atgtcggt	ccaggcttac	atgtctca	gtggattttc	ttccaaatg	1020
aaagttgtt	cttcatgtact	tttagagaaa	tatccccaaag	ctatctacac	actctgtct	1080
tcctgtgc	ttatgtgt	tttagatgtt	ataaataatgt	acacaaatat	tagatggaaat	1140
tttaggaaca	ttgagaaatgt	ttgttctttt	ttccatcgat	caccacaact	gtcttttagaa	1200
cttgacaa	caatttctgt	tccttttcgc	aacagtaaag	aaaggggtaa	agaactgaag	1260
gaaatctgc	attctcgatgt	gacaggcagg	catgtgtttt	ttgaaattttt	agtggaaactc	1320
ctgcaagcac	ttgtttatgt	tttagatgtt	ataaataatgt	acacaaatat	tagatggaaat	1380
aactatata	ctggccgagc	atttgtactc	tgcaatgtcg	tgtcagat	tgtatttgcatt	1440
gttactattt	ttgttctttaa	aaatgtctca	tcttttacaa	gagcctttgg	gaaaaacctc	1500
caggggcaaa	cctctgtatgt	tttcttttgcg	gcccgtatgt	tgactgtcg	actgtcatca	1560
ctcaacaa	tgatggaaaa	tatttggatgt	tatcatgaat	tttgggttgg	ggaagccacaa	1620
aatttggca	ccaaacttgc	tattooaaatgt	aaactccctg	gaaaaattccg	cagagctcac	1680
cagggttaact	ttgaaatctca	gctaaccctt	gagagttact	ataaagaaac	cctaagtgtc	1740
ccaaacatgg	agcacaatcc	tcaggaaactt	aaagatataat	tctcagaaca	gcacccatcaa	1800
gtctttaat	gttattctct	ggtagccatca	aactcaattt	caataacgtcg	1860	
gaggaacacc	atgctgacat	gtatagaatgt	gacttacca	atccgtacac	gtctgtcgat	1920
gagcttattt	gttggagaat	caaattggaa	cacaggggga	aaagatataaa	gtctccgtcc	1980
accatctat	tttgccttcc	cctgtctgtac	atcaatgtttt	tttgcataatgt	gtatgcattt	2040
ctgaagggtt	tgtgttattt	ttctgtatgt	aaagggttgg	atgagcggtt	tgaaaatgg	2100
cgaaggcgtt	ttaaaggcata	tttggggaaat	actttgacac	acccaaaggc	aagtaacttg	2160
gtcttgc	atacaaat	ttgtataaaa	cacgacatgt	atttatgtt	ggacacatata	2220
attaaactct	atacaatgtt	gtcagatgtt	cctacagata	attccgaaac	tgtggaaaat	2280
acctaa						2286

SEQ ID NO: 141	moltype = AA	length = 407
FEATURE	Location/Qualifiers	
source	1..407	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 141		
MAASAPPPD KLEGGGGPAP PPAPPSTGRK QKGAGLQMKNS PEKKRRKSNT QGPAYSHLTE	60	
FAPPPTPMVD HLVASNPFD DFGAPKVGA APPFLGSPVP FGGFRVQGGM AGQVPGYIST	120	
GGGGGPQPLR RQPPPFPNPN MGPAFNMPQQ GPGYPPPGNM NFPSQPFNQP LGQNFSPPSG	180	
QMMGPVQGF GPMISPTMGG PPRAEGLPPS LSQRFAQPGA PFGPSPLQRP GQGLPSSLPPN	240	
TSPPPGPDPG FPGPGEEDGG KPLNPPASTA FPQEPHSQSGP AAAVNGNQPS FPPNSSGRGG	300	
GTPDANSLAP PGKAGGGSGP QPPPGLVYPG GACRSEVNDD QDAILCEASC QKWFHRECTG	360	
MTESAYGLLT TEASAVWACD LCLKTKEIQS VYIREGMQGL VAANDGL	407	

SEQ ID NO: 142	moltype = DNA	length = 1221				
FEATURE	Location/Qualifiers					
source	1..1221					
	mol_type = other DNA					
	organism = synthetic construct					
SEQUENCE: 142						
atggccgcctt cggcgccgc	cccacccggac	aagctggagg	gagggtggcg	ccccgcaccc	60	
ccccctgcgc	cgcccgacac	cgggggaggaa	cggggcaagg	ccggctcgca	aatgaagat	120
ccagaaaaaa	agcgaaggaa	gtcaaaatact	caggcccttg	catactcaca	tctgtacggag	180
tttgcaccc	cccccaatcc	catgttgcatt	cacctgggtt	catccaaatcc	ttttgtggat	240
gacttcggag	cccccaatgt	gggggttgc	ccccctccat	tccttggcag	tcctgtggcc	300
ttcggaggct	tcctgtgtca	ggggggcatgt	gcggggccagg	taccccccagg	ctacagcact	360
ggagggtgg	ggggggcccca	gcacatccct	cgacaggccac	cccccttccc	tcccaatcc	420
atggggccct	tttcaat	gccccccatgt	ggtccctggct	acccacccccc	aggcaacat	480
aactttccca	gcaccaatcc	caacccatgt	ctgggtcaaa	actttatgtt	tcccaatgtgg	540
cagatgtgc	cgggcccaatgt	ggggggatgtt	ggtcccatgt	tctcacccac	catgggacac	600
cctcccccag	caagatgtgg	ccccatcttct	ctgtcccaac	gatttgtca	gccaggggct	660
ccttttggcc	tttccatcttct	ccagatgttgc	ggtcaggggc	tcccaatcc	ggccacttac	720
acaaggccat	tttctgtgtcc	ggacccatgttgc	tttctgtgtcc	ctgggtgtgt	ggatgggggg	780
aaggccatgtt	atccacatgt	tttccatgttgc	ggatggggatgtt	gggggggggg	gggggggggg	840
gctgtgtgtt	ttatatggaa	ccagccatgttgc	tttccatgttgc	acagcgttgc	gggggggggg	900
ggcaatccat	atgccaacatgt	tttccatgttgc	ggatggggatgtt	tttccatgttgc	tttccatgttgc	960
cagccatccc	cagggttgggt	gttccatgttgc	ggatggggatgtt	ggatggggatgtt	ggatggggatgtt	1020
caggatgtca	tttctgtgtca	ggatggggatgtt	ggatggggatgtt	ggatggggatgtt	ggatggggatgtt	1080
atgactgttgc	ggccatgttgc	actgttgc	tttccatgttgc	tttccatgttgc	tttccatgttgc	1140

-continued

```

ctctgcctca agaccaagga gatccagtct gtctacatcc gtgaggcat gggcagctg 1200
gtggctcta acgatgggtt g g 1221

SEQ ID NO: 143 moltype = AA length = 206
FEATURE Location/Qualifiers
source 1..206
mol_type = protein
organism = synthetic construct

SEQUENCE: 143
MAAAKDTTHED HDTSTENTDE SNHDPQFPEI VSLPEQEIKT LEEDEEELFK MRAKLFRFAS 60
ENDLPWKER GTGDKVLLKH KEKGAIILM RRDKTLKICA NYIITPMMEK KPNAGSDRAW 120
VNNTTHADFA ECPKPPELLAI RFLNAENAOK FTKFEECRK EIEEREKKAG SGKNDHAEKV 180
AEKLEALSVK EETKEDAEEK QPTFLY 206

SEQ ID NO: 144 moltype = DNA length = 618
FEATURE Location/Qualifiers
source 1..618
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 144
atggcgcccg ccaaggacac tcatgaggac catgatactt ccactgagaa tacagacgag 60
tccaaccatg accctcaggta tgagccaata gtttcttc ctgagaaga aattaaaaca 120
cttggaaagag acatggggaa atggggggaa aactgttccg atttgcctct 180
gagaacgatc tcccaaaatgaaaggagca gggactgttg acgtcaactt cctgaaagcac 240
aaggagaaag gggccatccg cctcttcatg cggaggggaca agaccctgaa gatctgtgcc 300
aaccactaca tcacggccatg gatggagctg aagcccaacg caggtagcga ccgtgcctgg 360
gtctggaaaca cccacgctga ctccggcgat gatggccca agccagact gctggccatc 420
cgcttcctgaa atgctgagaa tgccacggaa ttccaaacaaa agtttgaaga atgcaggaaa 480
gagatcgaag agagagaaaa gaaacggaga tcaggcaaaa atgatcatgc cgaaaaagt 540
gcggaaaaacg tagaagctct ctcgggtaaag gaggagaccatc aggaggatgc tgaggagaag 600
caaccaacctt tcttgtac 618

SEQ ID NO: 145 moltype = AA length = 326
FEATURE Location/Qualifiers
source 1..326
mol_type = protein
organism = synthetic construct

SEQUENCE: 145
MSSFSESALE KKLSELSNSQ HSVQTLSLWL IHHRKHAGPI VSVWHRELK AKSNRKLTF 60
YLANDVIQNS RKKGPEFTRE FESVLVDAFS HVAREADEGC KKPLERLLNI WQERSVYGG 120
FIQLQKLSME DSKSPPPQKAT EEEKSLKRTF QQIQEEEDD YPGSYSPQDP SAGPLLTTEEL 180
IKALQDLENA ASGDATVRQK IASLPQEVQD VSLLKEITDK EAAERLSKTV DEACLLAAY 240
NGRLAABLED RRQLARMLVE YTQNQKDVL S EKEKKLEEEYK QKLARVTQVR KELKSHIQSL 300
PDLSSLNPNTT GGLAPLPSAG DLFSTD 326

SEQ ID NO: 146 moltype = DNA length = 981
FEATURE Location/Qualifiers
source 1..981
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 146
atgttcctct tctctgagtc ggcgtggag aagaagctct cggagctgag caactcttag 60
cacagcgtgc agaccctgtc cctttggctc atccaccacc gcaaggacgc gggaccatc 120
gtctccgtgt ggcaccgcga gctccgcata gccaaatcaa atagaaagct tactttctg 180
tattttagcga atgatgtcat cccaaacagt aaaaggaaag gacctgaatt cactagagaa 240
tttgaatctg tccttgcata gctttttctc catgttgcata gagggcaga tgaaggctgt 300
aaaaaacctt tagaaagatt gctaaacatc tggcaaaac gaaatgttta tggcgccgag 360
ttcatacagc agctgaagct gtctatggag gactccaaga gccctcccc caaagcaaca 420
gaagagaaga aatctctgaa acgaaatctt cagcaattt aggaggaggaa ggatgacgac 480
taccctggca gctacttcac tcaggatct tctgcaggac ccctttgac tgaggaaacta 540
atcaaaggatt tgccaggatct gggaaatgcg gcatcagggg atgctactgt ccgacagaaaa 600
attgcttc tgcggccatg agtcaagat gtttctctat tggaaaaaat aacagacaaa 660
gaggcagctg aacgtcttc aaaaacagta gatgaagcat gtctgttaact agcagaatat 720
aacggggccg tggcggcaga actggaggac cgtcgccagc tggctggat gttggtgag 780
tataccccaga atcagaaga tggtttgtcg gagaaggaga aaaaactaga ggaataaaaa 840
cagaagcttgc cagcgtaaac ccaggccgcg aaggaactga aatccatata tcagagctt 900
ccagaccctt cactgtgcc caacgtcaca gggggcttag ccccccgtcc ctctgtggg 960
gacctgtttt caactgacta g 981

SEQ ID NO: 147 moltype = AA length = 262
FEATURE Location/Qualifiers
source 1..262
mol_type = protein
organism = synthetic construct

SEQUENCE: 147
MKTPFGKTPG QRSRADAGHA GVSANMMKKR TSHKKHRSSV GPSKPVSQPR RNIVGCRQH 60

```

-continued

GWKEGNGPVT	QWKGTVDQV	PVNPSLYLIK	YDGFDIVYGL	ELNKDERVSA	LEVLDRVAT	120
SRIISDAHLAD	TMIGKAVEHM	FETEDGSKDE	WRGMVLARAP	VMNTWFIITY	EKDPVLYMQ	180
LLDDYKEGDL	RIMPDSNDSP	PAEREPGEVV	DSLVGKQVEY	AKEDGSKRTG	MVIHQVEAKP	240
SVYFIKFDDD	PHIYVYDLVK	TS				262
 SEQ ID NO: 148		moltype = DNA length = 789				
FEATURE		Location/Qualifiers				
source		1..789				
		mol_type = other DNA				
		organism = synthetic construct				
SEQUENCE: 148						
atgaagaccc	cattcggaaa	gacacctggc	cagcggtcca	gagctgatgc	aggccatgct	60
ggagtatctg	ccaacatgtat	gaagaagagg	acatcccaca	aaaaacatcg	gaggcgtgtg	120
ggtccggacca	aacctgtttc	ccaggccccgg	cggacatcg	taggctgcag	gattcagcat	180
gggtggaaacg	aggggatgg	ccctgttacc	cagtggaaag	gaacatgttc	ggaccagggt	240
cctgttaatc	tttcttgcata	tcttataaaa	tacgatggat	ttgactgtgt	ttatqacta	300
gaacttaata	aagatgaaag	agtttctgcg	cttgaagtcc	tcctgtatag	agttgcgaca	360
tccgaatca	cgatgcaca	cttgcgcagac	acaatgttag	gcaaaaggagt	ggaacatatg	420
tttgagacag	aggatgttgc	taaagatgag	tggaggggaa	tggctttagc	acgtgcac	480
gtcatgaaca	catggttta	cattacctat	gagaaagacc	ctgtttgtat	catgtaccaa	540
ctcttagatg	attacaaaga	aggcgacatt	cgcattatgc	ctgattccaa	tgattcacct	600
ccacgacaaa	ggggacccagg	agaagttgtg	gacagcctgg	taggcaaaaca	agtggatat	660
gccaaagaag	atggctcgaa	aaggactggc	atggtcattc	atcaagttaga	agccaaggccc	720
tccgtctatt	tcatcaagt	tgatgtat	ttccatatatt	atgtctacga	tttggtgaaa	780
acatccat						789
 SEQ ID NO: 149		moltype = AA length = 397				
FEATURE		Location/Qualifiers				
source		1..397				
		mol_type = protein				
		organism = synthetic construct				
SEQUENCE: 149						
MSVAFASARP	RGKGEVTQQT	IQKMLDENHH	LIQCILEYQS	KGKTAECTQY	QQILHRLNLY	60
LATIADSQNQ	MOSLLPAPPT	QNMLNLGPGL	TQSGSSQGLH	SQGSLSDAIS	TGLPPSSLQ	120
GQIGNGPSHV	SMQQTAPNTL	PTTSMMSISGP	GYSHAGPASQ	GVPMQGQGTI	GNYVSRTNIN	180
MQSNPVSIMQ	QQAATSHYSS	AQGGSQHYQQ	QSSIAMMGQG	SQGSSMMGQR	PMAPYRPSQQ	240
GSSQQYLGQE	EYYGEQYSHS	QGAAEPMGQQ	YYPDGHGDYA	YQQSSYTEQS	YDRSFEESTQ	300
HYYEGGNNSQY	SQQQAGYQQG	AAQQQTYSQQ	QYPSQQSYPG	QQQGYGSAQG	APSQYPGYQQ	360
GQQQQYGSYR	APQTAPSQQ	QRPYGYEQQQ	YGNYQQL			397
 SEQ ID NO: 150		moltype = DNA length = 1191				
FEATURE		Location/Qualifiers				
source		1..1191				
		mol_type = other DNA				
		organism = synthetic construct				
SEQUENCE: 150						
atgtccgtgg	ctttcgcgtc	tgccggcga	agaggcaaaag	gggagggttac	gcagcaacc	60
atccagaaga	tgctggacga	gaaccaccac	ctgatccatg	gtatcttgcg	gtaccagac	120
aaggcgaaga	cgcccgagtg	cacgcgtac	cagcagatcc	tgcacccggaa	cctggatatac	180
ctggccacga	tcgcagactc	caaccagaac	atgcagttcc	tgttcttgc	cccggccacg	240
cagaacatga	acttggggcc	tggagccctg	actcagagcg	gttccacca	gggcctgcac	300
tctcaggggc	gcctgtat	cgccatcago	acggggcttg	cacccttc	cctctgtcag	360
ggccagattg	gcaacggggc	gagccacgtg	tccatgcacg	agacggccgc	taacacgtg	420
ccaccacact	ccatgagcat	ctctggggcc	ggctacagcc	acgcgggacc	cgcctcgac	480
ggcgtccccca	tgcggggcc	aggcaccatc	ggcaactacg	tgtctcgac	caacatcaac	540
atgcagttcc	acccagtttc	catgatcac	cagcaggccg	ccacgtcgca	ctacagtcg	600
ggcgaggccg	gcaaggccgca	tcaccaggcc	caatcgatca	tgcctatgt	ggggcaggcc	660
agccaggggc	gcagcatgt	ggggcagccg	ccatggccg	cctaccggcc	ctcccaagca	720
ggctcttccc	aggactatcc	ggggcaggag	gagtaactatg	ggcggcagata	cagccacac	780
caggcgcgcg	cgaggccat	ggggcaggcg	tactaccccg	acggccatgg	cgattacgac	840
taccagcagt	catcttacac	ggacagacg	tacggccgtt	ccttcggata	gtccacgcag	900
cactactatg	agggggggaa	cttccagttac	agccagccgg	aggccgggt	ccagcagggt	960
gccgcgcgcg	gcacagacgt	cttccagcc	cagtacccca	gccagcagag	ctacccgggg	1020
cacgcaggccg	gtacgggtc	tgcccaggga	gccccgtcac	agtaccccg	ctaccagca	1080
ggccaaaggcc	agcgtacccg	gacatccga	gcacccgcga	cagcgcgc	tgcccaggcc	1140
cgccggccat	acggctatga	acagggccat	tatggaaat	accaggcgtt	g	1191
 SEQ ID NO: 151		moltype = AA length = 432				
FEATURE		Location/Qualifiers				
source		1..432				
		mol_type = protein				
		organism = synthetic construct				
SEQUENCE: 151						
MSELKDPLQ	FHDFKSVDHL	KVCPRYTAVL	ARSEDDGIGI	EELDTLQLEL	ETLLSSASRR	60
LRVLEAFTQI	LTDWQDKKGD	RRFLKLGRDH	ELGAPPKHGK	PKKQKLEGKA	GHGPGPGPGR	120
PKSKNLQPKI	QEYEFTDDPI	DVPRIPKNDA	PNRFWASVEP	YCADITSEEV	RTLEELLKPP	180

-continued

EDEAEHYH	KIP	PLGKHYSQRW	AQEDLLEEQK	DGARAAAVAD	KKKGLMGPLT	ELDTKDVAL	240
LKKSEAQH	EQ	PEDGCPFGAL	TQRLLQALVE	ENIISPMEDS	PIPDMMSGKES	GADGASTSPR	300
NQMKPFSV	PK	TKSLESRIKE	ELIAQGLLES	EDRPAEDED	EVLAELRKQ	AELKALSAHN	360
RTKHKDLLR	L	AKEEVSRQEL	RQRVRMADNE	VMDAFRKIMA	ARQKKRTPTK	KEKDQAWKTL	420
KERESILKLK	L	DG					432

```

SEQ ID NO: 152 moltype = DNA length = 1299
FEATURE Location/Qualifiers
source 1..1299
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 152
atgagtgtat tggaaagactg ccccttgcgt ttcccaagact tcaagtctgt ggatcaccgt 60
aaggctgttc cccgcacac ggcagtgtcg gcacgtctg aggatgtatgg cattcgccatc 120
gaggagctgg acaccctgtca gctggagctg gagaccctgc tgtttctgc cagccggcgc 180
ctgcgtgtgc ttgaggccga aaccggatc ctcacccgact ggcaggatata gaaagggtac 240
agacgatttc tgaagctggg ttagagacat gaacttggag ctccccccaa acatggaaag 300
cccaaaagac agaaactgtga aaggaaaggca ggacatggc cggggccctgg cccaggacgg 360
cccaaaatcca aaaacccatca gcccaagatc caggaatatg aatttactgt tgaccctatc 420
gacgtgcccac ggatccccaa aaatgtatcc cccaaacaggat tctgggttctc aagtggagcc 480
tactgtgttc acatcacccg cggaggagtc cgcacacttg aggatgtact gaagccccca 540
qaagatgagg ctgagcatca caagatccca cccctgggaa acgcactactc ccacgcgtgg 600
gcccaggagg acctgtgtta ggagcagaag gatggggccc gggcagcggc tgggtgtac 660
aagaagaaagg gcttcatgtt gccactgtacc gaacttgtaca ctaaaatgtt ggtatccctg 720
ctgaagaaatg ctggggccca gcatgaacag cccggaaatgt gatggccctt tggtggccct 780
acgcacggcc tcctggccggc cttgtggggaa gaaaatatta ttcccttat ggaggattt 840
cttattctgtt acatgtgtttt gaaagaatca ggggtgtacg gggcaagcac ctcccttcgc 900
aatcagaaca agcccttcgt tggccgcatt actaagtccc tggagagccg cataaaggag 960
gagcttaatgg cccaggggct tttgggtct gaggacggcc cccggagggaa ctccggggat 1020
gggttcttgc ctggatgttc ccaaaacggcc gctgactgtc aggacttag tgccccacaac 1080
cgccaccaaga agcacgtact gctggggctg gcaaaaggagg aggtgagccg gcaggagctg 1140
aggccacggcc tgccgtatgc tgacaaacgg gtcatggacg cttttcccaa gatcatgtt 1200
gccccggcaga agaaacggac tcccaacaaag aaagaaaaagg accaggccctg gaagactctg 1260
aaqaaqccctq aqaaqccctt qaaqgctctq qatqqqtaq 1299

```

```

SEQ ID NO: 153          moltype = AA length = 678
FEATURE                Location/Qualifiers
source                 1..678
                      mol_type = protein
                      organism = synthetic construct
SEQUENCE: 153
MAEKKKLKLS NTVLPSEMSK VVAESMGIAQ IQEETCQLLT DEVSYRIKEI AQDALKFMMH 60
GKRQKLLTSD IDYALKLKVN EPLYGFHAQE FIPFRFASGG GRELYFYEEK EVDSLSDIINT 120
PLPVPVLDC LKHAHWLSIEG CQPAPENPP PAPKEQKAE ATEPLKSAKP GQEEDGPLKG 180
KGQGATTADG KGKEKKAPPL LEGAPLRLKP RSIHELSVEQ QLYYYKEITEA CVGSCEAKRA 240
EALQSIATDP GLYQMLPRFS TFISEGVVRN VVQNLLALLI YLMRMVKMLDNPTITLEKY 300
VHELIPAVMT CIVSRQLCLR PDVDNHWALR DFAARLVAQI CKHFSTTTNN IQSRTKTF 360
KSWVDEKTPW TTRYGSIAGL AELGHVDVIKT LILPRLQQEG ERIRSLVLDGP VLSNIDRIGA 420
DHQVSLLLKH CAPVLAKLRR PPDNDQDAYRA EFGSLGPLLC SQVVKARAQA ALAQAOVNRT 480
TLTITQPRPT LTLSQAPQPG PRTPGLLKVP GSIALPVQTL VSARAAAAPPQ PSPPPTKFIV 540
MSSSSSSAPST QQLVLSLSTA PGSGSTTTSV VTTTVPBVQD IVKLVSTAT APPSTAPSGP 600
GSVQKYIVVLS LPPTGEKGKG PTSHPSVPP PASSPSPLSG SALCGGKQEA GDSPPPPAGT 660
PKANGSOOPNS GSPQOAPL 678

```

```

SEQ ID NO: 154 moltype = DNA length = 2034
FEATURE Location/Qualifiers
source 1..2034
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 154
atggctgagg agaagaagct gaagcttagc aacactgtgc tgccctcgga gtccatgaag 60
gttgtgtggc aatccatggg catgcggccagg attcaggagg agacctgcga gctgtcaacg 120
gtatgggtca gctaccgcatt ccaagagtc gcacaggatg ctttgaaggct catgcacatc 180
ggaaaggcgcg agaagtcac caccatgtac atttgactactc ctttgaaggct aaagaatgtc 240
gagccactct atggcttcca cgccccaggag ttcatccctt tccgcttcgc ctctgggtggg 300
ggccggggaggc tttaacttcta tgaggagaag gaggttgcatt tgagcgacat catcaatacc 360
ctctctggccc gggtggccctt ggacgtctgc ctcaaaatgc atttgggttag catggaggcc 420
tgcacccaggcc statccccca gaaccgcggc ccacatccca aagagcaaca gaaggctgca 480
gcacacagaa cccttggatc agccaaaggcc gggccaggagg aagacggacc ccttggggc 540
aaaaggctcaag gggccaccac agcccgaggc aaaggaaagg agaagaaggc ggcggcttgc 600
ctggaggggggcccccttgcg actgaaggccc cggagcatcc acggatgttc tttggggcggc 660
cagctctact acaaggagat caccggggcc tgcgtgggtt ccttgcggggc caagggggc 720
gaaggccctgc aaggatgtc cacggggccctt ggactgtatc agatgtggcc acgggttgc 780
accccttatcatc cggggggggcccttgcg actgttgcggatc gtttttttttttccatccatc 840
tacctgtatcgttgcgttgcgaa acggccgtatc gtttttttttttccatccatc 900
gtccatggcgttgcgttgcgacggccgtatc gtttttttttttccatccatc 960

```

-continued

ccagatgtgg	acaatcactg	ggcactccga	gactttctgt	ccgccttgtt	ggcccagatc	1020
tgcgaaggatt	tttagcacaac	cactaacaad	atccagtcc	ggatcaccaa	gaccattcacc	1080
aagagctggg	ttggacgagaa	gacgcctgg	acgactcggt	atggctccat	cgcaggett	1140
gtctgagctgg	gacacgatgt	tataaagact	ctgattctgc	cccggtctca	gcaggaaagg	1200
gagcggatcc	gcagtggtct	ggacggccct	gtgttgagca	acattgaccc	gattggagca	1260
gaccatgtgc	agagcctctt	gctgaaacac	tgtgtctctt	ttctggcaaa	gtgtgcggca	1320
ccgcctgaca	atcaggacgc	ctatcgggca	gaattcggtt	cccttggggc	ccttcctctgc	1380
tcccaggatgg	tcaaggctcg	gccccaggct	gtctgtcagg	ctcagcagggt	caacaggacc	1440
actctgtacca	tcacgcagcc	ccggcccaac	ctgaccctct	cgccaggcccc	acagctggc	1500
cctcgcaccc	ctgggttctgt	gaaggttctc	gggtccatct	cacttcttgt	ccagacactg	1560
gtgtctgtgc	gagcggctgc	cccaccaac	cattttccctt	cttccaacca	gtttattgt	1620
atgtcatctgt	cctccagcgc	ccccatccac	cacgggttcc	tgtcccttca	cacccggcc	1680
cccggtctcg	gttccaccac	cacttcgccc	gtcaccacca	ccgtccccag	cgtgcagccc	1740
atcgtaaagt	ttgttccatcc	cgccaccacc	gcacccccc	gcaactgtcc	ctctgtgtct	1800
gggagtgtcc	agaagttacat	cgtgttctca	cttccccca	caggggagg	caaaggaggc	1860
cccacccatcc	atcccttctcc	agtttctccc	ccggcatct	ccccgttccc	actcagccgc	1920
agtgcctttt	gtggggggaa	gcaggaggct	ggggacatc	cccttccagc	tccagggtact	1980
ccaaaagcca	atggctccca	gccccactcc	ggccacttcc	agccgtctcc	gttg	2034

SEQ ID NO: 155 moltype = AA length = 187

FEATURE Location/Qualifiers
source 1..187
mol_type = protein
organism = synthetic construct

SEQUENCE: 155

MDADSDVALD	IILITNVVCVF	RTRCHLNLRK	IALEGANVIY	KRDAGKVLMK	LKPRITATI	60
WSSGKIICTG	ATSEEAKFG	ARRLARSLQK	LGFQVIFTDF	KVNVNLAVCN	MPFEIRLPEF	120
TKNNRPHASY	EPELHPAVCY	RIKSLRATLQ	IFSTGSITVT	GPNVKAVATA	VEQIYPFVFE	180
SRKEILL						187

SEQ ID NO: 156 moltype = DNA length = 561

FEATURE Location/Qualifiers
source 1..561
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 156

atggatgcag	acagtgtatgt	tgcattggac	attctaatta	caaatgttagt	ctgtgttttt	60
agaacaagat	gtcatttaaa	cttaaggaag	attgttcttg	aaggagcaaa	tgttaatttt	120
aaacacgtatg	ctggaaaaatg	attaaatgtt	cttagaaaaac	ctagaattac	agctacaatt	180
tggtcctctcg	gaaaaaaat	ttgcacttgg	gcaacaatgg	aagaagaagc	taaattttgg	240
gccagacgcct	tagccctgt	tctcagaaaa	ctaggttttc	aggttaattt	tacatgttt	300
aagggttta	acgttctggc	agtgtgttac	atgccatttg	aatccgtt	gccagaattc	360
acaaagaaca	atagaccta	tgccaggta	gaaacctgtt	ttccatcttc	tgtgtgttat	420
cggatataat	ctctaagac	tacattacag	atttttca	caggaaatgt	cacagtaaca	480
ggggccaaatg	taaaggctgt	tgtactgt	gttggaaacaga	tttaccatt	tgtgtttgaa	540
agcaggaaag	aaattttattt	g				561

SEQ ID NO: 157 moltype = AA length = 365

FEATURE Location/Qualifiers
source 1..365
mol_type = protein
organism = synthetic construct

SEQUENCE: 157

MSLAGGRAPR	KTAGNRLSGL	LEAAEEDFY	QTTYGGFTEE	SGDDEYQGDQ	SDTEDEVDS	60
FIDIDEDEPS	SDGEABEPRK	KRRVVTKAYK	EPLKSLRPRK	VNTPAGSSQK	AREEKALLPL	120
ELQDDGDSDR	KSMRQSTAEH	TRQTFLRVQE	RQGQSRRRKQ	PHCERPLTQE	ELLREAKITE	180
ELNLRSLETY	ERLEADKKKQ	VHKRKCPGP	IITYHSVTPV	LGVGPGKKEE	NVDIBGLDPA	240
PSVSALTPHA	GTGPVNPPAR	CSRFTIFSD	DATFEFWFPQ	GRPPKPVRE	VCPVTHRPAL	300
YRDPVTDIPY	ATARAFKIIR	EAYKKYITAH	GLPPTASALG	PGPPPPEPLP	GSGPRALRQK	360
IVIKL						365

SEQ ID NO: 158 moltype = DNA length = 1095

FEATURE Location/Qualifiers
source 1..1095
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 158

atgagtttgg	ctggggggccg	ggcaccccg	aagaccgctg	ggaaccggct	ttctgggtt	60
ttggaggccg	aggaggaaga	ttagttctac	cajacgactt	atgggggtt	cacagaggaa	120
tccggagatg	atgatgtatca	aggggaccag	tcajacacag	aggacgaatg	ggactctgac	180
tttgacattt	atgaagggg	tgaaccatcc	agtgtatgg	aagcagaaga	gccaagaagg	240
aagcgcgg	tagtccacaa	ggctataag	gaacctctca	agagcttaag	gcctcgaaag	300
gtcaacaccc	ccggctgttag	ctctcagaag	gcccggagaag	agaaggact	actgcattt	360
gaactacaag	atgacggctc	ttagtctatgc	gtcacttac	agctgagcat	420	
acacgacaaa	cgttccatcg	ggtacaggag	aggcaggcc	agtcacaaacg	gccaaggagg	480
ccccactgt	agcggccact	aaccaggag	gaactgtcc	ggggggccaa	gatcacagaa	540

-continued

```
gagcttaatt tacggtaact ggagacatat gagcggctcg aggctgataa aaagaagcag 600
gttcataaga agcggaaatg ccccgccccca ataatcacct atcattcaat gacagtggca 660
cttgggggg agecaggccc caaggaagaa aacgttgaca tagaaggact tgatctgtct 720
ccctcggtgt ctgcattgac tcctcatgt gggactggac ccgtcaaccc ccctgctcgc 780
tgctcacgtc cttcatcac tttagtgc gatgcaactt tcgaggaaatg gttcccccaa 840
ggccggcccc caaaatccc tgttctgtgag gtctgtcccg tgaccatcg tccagcccta 900
taccgggacc ctgttacaga cataccctat gccaactgtc gagecttcaa gatcattcgt 960
gaggcttaca agaagttacat taatgtccat ggactgccc gcaactgtc acgcttggc 1020
cccgccccgc caccccttga gcccctccct ggctctggc cccgagcctt ggcggaaaa 1080
attgtcatta aattt 1095
```

```
SEQ ID NO: 159 moltype = AA length = 653
FEATURE Location/Qualifiers
source 1..653
mol_type = protein
organism = synthetic construct
SEQUENCE: 159
MAFRDVAVDF TQDEWRLLSP AQRTLYREVM LENYSNLVSL GISFSKPELI TQLEOGKETW 60
REEKKCSPAT CPDPPEPELYL DPFCPPGFSS QKFPMPQHVLC NHPPWIFTCI CAEGNIQPGD 120
PGPGDQEKKQQ QASEGRPWSD QAEQEGEEGA MPLFGRTKRR TLGAFSRPPQ RQPVSSRNL 180
RGVELEASPA QTGNPEETDK LLKRIEVLGF GTVNCGECL SFSKMTNLIS HQRIHSGEKP 240
YVCVGVCCKGF SLKKSALARHQ KAHSGEKPIV CRECGRGEFNR KSTLIIHERT HSGEKPYMCS 300
ECGRGFSQKS NLIHQRTQHS GEKPYVCREC GKGFQSQKSAV VRHQRTHLEE KTIVCSDCGL 360
GFSDRSNLIS HQRTHSGEKP YACKECGRCP RQRTTLVNHQ RTHSKEKPYV CGVCGHSFSQ 420
NSTLISHRRT HTGEKPYVCG VCGRGFSLKS HLNRHQNIHS GEKPIVCKDC GRGFSQSQNL 480
IRHQRTHSGE KPMVCGECGB GFSQKSNLVA HQRTHSGERP YVCRECGRGF SHQAGLIRHK 540
RKHSREKPYM CRQGGLGFGN KSALITHKRA HSSEKPCVCR ECGQGFLQKS HLTLMQHT 600
GEKPYVCKTC GRGFSLKSHL SRHRKTTSVH HRLPVQPDPE PCAGQPSDSL YSL 653
```

```
SEQ ID NO: 160 moltype = DNA length = 1962
FEATURE Location/Qualifiers
source 1..1962
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 160
atggcattca gggatgtggc tggatgttcc acccaggatg agtggaggct gctggccct 60
gctcaaaaggc ctctgtacag agagggtatg ctggagaact acagcaacctt ggtctcaatg 120
ggaaatttcat ttcttaacc accaactcata acccagtg agcaaggaa agagacctgg 180
agagaggaaa aaaaatgttcc accggcaatc tgccatcagat cagagccaga gcttctaccc 240
gatcccttcc gccctccggg ttcttccttgc cagaaatcc ccatgcgcga tggctgtgt 300
aatcatcccc cttgtatctt ccatatgtt gttgcagaaat gtaatccca gcctggggat 360
ccaggccccag gggaccagga gaagcagcaa caagcctctg agggggagacc ctggatgtat 420
caagcagaag gtccttgcggg agaaatgtcc atgcctttgt ttggaaagaa caagaaagg 480
actcttgggg cggttccatc ggcacccatc aggcagccatc tcaatctcc gaacggcctc 540
agagggttgg agtttgcgc cagccatgc cagacaggaa acccttggaa aacagacaaa 600
tttgttgcaga ggtatgttgc tttaggtttt ggaacatgtca actgtggaaat gtgtggactg 660
agtttgcacca agatgtacaaa cttgtatgttgc caccatgcgaa tacactcagg ggagaaggccc 720
taatgttgcgtt gggatgttgc agcgttgcgtt agcgttgcgtt agagcctcgc cagacaccc 780
aaggcacactt cggggggagaa gccaatttgc tgcaggatgtt gtggacgggg cttaacccg 840
aagtcaacgc taatcatata cgaacggaca cactccgtt agaaacactt catgtgcagt 900
gagttgtggc gaggatgttgc ccagaatgtca aacccatgtca tacaccaggag gacacactc 960
ggggaaaaggc ttatgttgc cggggatgttgc ggcggaaatgttgc cttcggccatc 1020
gttgagacacc agaggacacaat cttggggatgttgc agaaggatgttgc ttttgcgtt 1080
ggcttcagcg acaggatcaaa cttatctcc caccatgttgc cgcactctgg ggagaaggccc 1140
taatgttgcgtt gggatgttgc tggatgttgc agggatgttgc ccacccatgtt caaccaccc 1200
aggacacactt caaaaggatgttgc gcccattgttgc tggggatgttgc gtggacccatc 1260
aattcaaccc tcatatctca cggggatgttgc caccatgttgc agaaggatgttgc ttttgcgtt 1320
gttgatgttgc gaggatgttgc ttttgcgttgc caccatgttgc agaaggatgttgc ttttgcgtt 1380
ggagagaaatc ccattgttgc caaggatgttgc ggcggggatgttgc ttttgcgttgc 1440
atcaggatgttgc aaggatgttgc cttggatgttgc aaggatgttgc ttttgcgttgc 1500
ggcttcagcc aaggatgttgc ctttgcgttgc caccatgttgc cgcactctgg ggagaaggccc 1560
tatgttgcgtt gaggatgttgc gggggatgttgc ttttgcgttgc caccatgttgc ttttgcgtt 1620
cggaaggactt cgaggatgttgc gggggatgttgc ttttgcgttgc caccatgttgc ttttgcgtt 1680
aaggatgttgc taatcatata cggggatgttgc ttttgcgttgc caccatgttgc ttttgcgtt 1740
gaggatgttgc aaggatgttgc cttggatgttgc ttttgcgttgc caccatgttgc ttttgcgtt 1800
ggggggatgttgc cttggatgttgc ttttgcgttgc caccatgttgc ttttgcgttgc 1860
agcagacacaat ggaaggaccat gtttgcgttgc ttttgcgttgc caccatgttgc ttttgcgtt 1920
ccgtgttgcgtt gggatgttgc ttttgcgttgc caccatgttgc ttttgcgttgc 1962
```

```
SEQ ID NO: 161 moltype = AA length = 457
FEATURE Location/Qualifiers
source 1..457
mol_type = protein
organism = synthetic construct
SEQUENCE: 161
MSQGSVTFRD VAIDFSQEEW KWLQPAQRDL YRCVMLENYG HLVSLGLSIS KPDVVSLLEQ 60
```

-continued

GKEPWLKGRE	VKRLDFSVSE	SSGEIKDFSP	KNVIYDDSSQ	YLIMERILSQ	GPVYSSFKGG	120
WKCKDHTMLEM	QENQCGIRKV	TIVSHQEALAQ	HNMISTVERP	YGCHECGKTF	GRRFLSVLHQ	180
RTHTGEKPYA	CKECGKTFQS	I3NLVKHQMI	HTGKKPKHEC	DCNKTFPSYL	FLEIHQRHTT	240
GEKPYECTEC	GKAFAFRASNL	TRHQRIHIGK	KQY1ICRKCGK	AFFSSGSELIR	HQITHTEGEK	300
YECIGECKAF	RRFSLHTRHO	SIHTTKTPYE	CNECRKAFCR	HSFLIHKHORI	HAGEKLYECD	360
ECGIVFWTHW	SLIGHTKSNT	GEKPYACAE	DCAFSAFRSFSL	ILHQRHTHG	KPYVCKVCNK	420
SPFWNSNLAK	HORTHTLDNQ	EATSENFSNHY	SFLTEHO			457

SEQ ID NO: 163 moltype = AA length = 96
FEATURE Location/Qualifiers
source 1..96
mol_type = protein
organism = synthetic construct
SEQUENCE: 163 MSPGLLITTRK EALMAFRDVA VAFTQKEWL LSSAQRTLYR EVMLENYSHL VSLGIAFSKP 60
KLIQOLFGCD EDWEEERNEUHL LBLGRGRDTT RSCVYRD 96

```

SEQ ID NO: 164      moltype = DNA    length = 291
FEATURE          Location/Qualifiers
source           1..291
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 164
atgtcaagg gactcctgac aaccggaaag gaggcattga tggcctccg ggatggct 60
gtggccttc cccagaagga gtggaaagcta tttagttctg ctcatggac cctgtacagg 120
gaggtgtatgc tggagaacta cagccatctg gtctccctgg gaattgcatt ttccaaacca 180
aaatcatcg aacagtgtgg acaaggcaca gaaaccttgg gagatggagaa cgaacatctt 240
ttggatccatc ttggatccatc acgtatcata acgttccatc ttccatcata c 291

```

SEQ ID NO: 165 moltype = AA length = 91
FEATURE Location/Qualifiers
source 1..91
mol_type = protein

SEQUENCE: 165 organism = synthetic construct
MPGPPRSLEMLGLTFRDVAI EFSLEEQWHL DIAQONLYRN VMLENYRNLA FLGIAVSKPD 60

```
SEQ ID NO: 166          moltype = DNA  length = 276
FEATURE                 Location/Qualifiers
source                  1..276
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 166
atggcaggac cccctagaag cctagaaatg ggactgttga catttagggg tggggccata 60
aaatcttc tggggaaatg qcaaacacctg qacatttgcac accadaattt atatagaaat 120
```

-continued

gtgatgttag agaactacag aaacctggcc ttcttggta ttgtgtctc taagccagac	180
ctgatcacct gtctgaaaca aggaaagag ccctggaata tgaagcgaca tgagatgtg	240
gatgaacccc caggattgga ttttcatta ctgtga	276
SEQ ID NO: 167	moltype = AA length = 419
FEATURE	Location/Qualifiers
source	1..419
	mol_type = protein
	organism = synthetic construct
 SEQUENCE: 167	
MAQESVMPSD VSVDFPSQEEW ECLNDDQRDL YRDVMLENYS NLVSMGHSIS KPNVISYLEQ	60
GKEPWLADRE LTRGQWPVLE SRCETKKLFL KKEIYEIEST QWEIMEKLTR RDFQCSSFRD	120
DWECNRQFKK ELGSQGGHFN QLVFTHELDLP TLSHHPSTL QQIINSKKF CASKEYRKT	180
RHGSQFATHE IHTIEKPYCE CKECGKSFRH PSRLTHHQKI HTGKKPFEC ECGKTFICGS	240
DLTRHHRIHT GEKPYECKEC KAFPSGSNF TRHQRIRHTGE KYECKECKG AFSSGSNFTQ	300
HORIHTGEKP YECKECGNAF SQSSOLIKHQ RIHTGEKPYCE CKECEKAFRS GSDLTRHQRI	360
HTGEKPYECK ICGKAYSQSS QLISHHRIHT SEKPYEYREC GKNFNYDPQL IQHQNLWYL	419
 SEQ ID NO: 168	moltype = DNA length = 1257
FEATURE	Location/Qualifiers
source	1..1257
	mol_type = other DNA
	organism = synthetic construct
 SEQUENCE: 168	
atggctcagg agtcagtgtat gttcgtgtat gtgtccgtat acttctctca ggaggaggatgg	60
gaatgcgtga atgatgtatc gagatgttata tacagatgtat tgatgttga gaattacagc	120
aacctgttta caatggggca ttcttatttttca aaaaaaatatc tgatctctca ttggagcaa	180
gggaaaggagc cctgggttgc tgacagatgtat aaaaaagaaa ttatgttata agaatcaacc	240
tcaagatgttgc agaccaagaa atttttgc aaaaaagaaa ttatgttata agaatcaacc	300
cagtggaaa taatggaaa acttcacaaga cgttattttc achtgtccag ttccatgtat	360
gattggatgtat gtaatccggc gtttataaaaaa gaaatccggc ttccatggggg acatttcaat	420
caatgttat tcaactgtatc agatgtccccccatccatccatc ttccatgtatc	480
cacaaatca ttaacgtaa aaaaaatccatc ttccatgtatc aaaaatataatc gaaaaaccc	540
agacatggct cacatgttgc tacatcatgtatc aaaaaatccatc ccatttgaaa gccttatgaa	600
tgttataatc gttggaaatgc ttccatgtatc ccctcaaaatc ttccatgtatc tcagaaaaat	660
catactggca agaaaaccctt tcaatgttgc gatgttgc gaaatccatc ttccatgtatc	720
gacccatctc gacatcacatc aatttgcactt ggttgc gatgttgc gaaatccatc ttccatgtatc	780
ggggaaagcc tttagtagtgg ttccaaacttc actcgacatc agagaattca cacaggtag	840
aacccatgtatc aatgttgc gttggatgttgc gttggatgttgc gttggatgttgc gttggatgttgc	900
catcagatgtatc ttcataactgg gggaaaaccctt tcaatgttgc gatgttgc gttggatgttgc	960
agtcagact cacaacttat taaatccatc aaaaatccatc ccatttgaaa acatccatc	1020
tgttataatc gttggaaatgc ttccatgtatc ggctcagacc ttccatgtatc tcagatgtatc	1080
catactgggg agaaaaccctt tcaatgttgc gttggatgttgc gttggatgttgc gttggatgttgc	1140
cagcttataatc gtcatcatgtatc aatttgcactt ggttgc gatgttgc gttggatgttgc	1200
ggggaaagcc tttagtagtgg ttccaaacttc actcgacatc agagaattca cacaggtag	1257
 SEQ ID NO: 169	moltype = AA length = 113
FEATURE	Location/Qualifiers
source	1..113
	mol_type = protein
	organism = synthetic construct
 SEQUENCE: 169	
MPANWTSPQK SSALAPEDHG SSYEGSVSFR DVAIDFSREE WRHLDPSQRN LYRDVMLETY	60
SHLLSIGYQV PEAEVVMLEQ GKEPWLQGE RPRQSCPAPC LVNSHHLQES FRG	113
 SEQ ID NO: 170	moltype = DNA length = 342
FEATURE	Location/Qualifiers
source	1..342
	mol_type = other DNA
	organism = synthetic construct
 SEQUENCE: 170	
atgccagcta attggacctc acctcagaaa tcctcagccc tggctccaga ggtatcatggc	60
agctcctatg agggatcagt gtccttcagg gatgtggctt tcgatttcag cagagaggaa	120
tggccggacc tggacccctc tcgatggaaa ctgttacccggg atgtgtatgttgc ggagacccat	180
agccacacttc ttcataatgg atataatgg ctcgttacccggg atgtgtatgttgc ggagacccat	240
ggaaaggaaac catggggact tcgatggaaa ctgttacccggg atgtgtatgttgc ggagacccat	300
cttggatgtatc tcaatgttgc gttggatgttgc gttggatgttgc gttggatgttgc	342
 SEQ ID NO: 171	moltype = AA length = 500
FEATURE	Location/Qualifiers
source	1..500
	mol_type = protein
	organism = synthetic construct
 SEQUENCE: 171	
MAPPSPALPA QGPGKARPSR KRGRPRALK FVDVAVYFSP EEWGCLRPAQ RALYRDVMRE	60
TYGHGLGALGC AGPKPALISW LERNTDDWEP AALDPQEYPR GLTVQRKSRT RKKNGEKEVF	120

-continued

PPKAEPRKKGK	RGRPSPKPL	I P R Q T S G G P I	C P D C G C T F P D	H Q A L E S H K C A	Q N L K K P Y P C P	180
DGCRRFSYPS	LLVSHRRRAHS	G E C P V C D Q C	G K R F S Q R K N L	S Q H Q V I H T G E	K P Y H C P D C G R	240
CFRRSRSLAN	H R T T H T G E K P	H Q C P S C G R R F	A Y P S L L A I H Q	R T H T G E K P Y T	C L E C N R R F R Q	300
R T A L V I H Q R I	H T G E K P Y P C P	D C E R F R F S S S	R L V S H R R V H S	G E R P Y A C E H	E A R F S Q R S T L	360
L Q H Q L L H T G E	K P Y P C P D C G	A F R R S G S L A I	H R S T H T E K L	H A C D D C G R R F	A Y P S L L A S H R	420
R V H S G E R P Y A	C D L C S K R F Q A Q	W S H L A Q H Q L L	H T G E K P F P C L	E C G R C F R Q R W	S L A V H K C S P K	480
A P N C S P R S A I	G G S S O R Q N H A H					500

```

SEQ ID NO: 173      moltype = AA length = 67
FEATURE          Location/Qualifiers
source           1..67
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 173
MALPQGLLTF RDVAIEFSQE EWKCLDPAQR TLYRVDVMLEN YRNLVSLELS GCPPLAAPAS 60
LDPAFLC          67

```

```

SEQ ID NO: 174      moltype = DNA    length = 201
FEATURE          Location/Qualifiers
source           1..201
                 mol_type = other DNA
                 organism = synthetic construct
SEQUENCE: 174
atggcttcc ctcagggtct attgacatc agggatgtgg ccatagaatt ctctcaggag 60
gagtggaaat gcctggaccc tgctcagagg actctataca gggacgtgtat gctggagaat 120
tatagaacc tggttccctt ggatgttca ggggagtgtc cattggcagc acctgcctcc 180
tttqccaqq ctttttttq c 201

```

```

SEQUENCE: 175
MPANEDAPQP GEHGSACEVS VSFEDTVDF SREWQQLDS TQRRLYQDVM LENYSHLLSV 60
MPANEDAPQP GEHGSACEVS VSFEDTVDF SREWQQLDS TQRRLYQDVM LENYSHLLSV 60
FKLQPKPEVI FKLEQGEGPW TLEGEAPHQS CSDGKFGIKP SQRRISGKST FHSEMEGEDT 120
LCSGLM 126

```

-continued

SEQUENCE: 176

atgcccagcta	acgaggacgc	tccccagcca	ggggaaacatg	gcagtgcctg	tgaggttatca	60
gtgtcatttg	aggatgtac	tgtggactt	agttagagagg	agtggcagca	actggactct	120
actctaaagac	gctgttacca	ggatgttaatg	ttggagaact	acagccacct	gctcttcagt	180
gggttcaag	ttccctaaacc	agaggatcatc	ttcaagttgg	agcaaggaga	ggggccatgg	240
acatttggaa	gggaagcccc	acatcagago	tgttcagatg	ggaaatttgg	aattaaggct	300
tcccagagga	gaatttctgg	gaaatctaca	tttcatatgt	aatggagggg	tgaagacaca	360
ctgtgttcag	gctcatggg					380

1. A Cas effector comprising:
 - a first polypeptide comprising a Cas protein and at least one peptide epitope; and
 - a second polypeptide comprising an effector selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, VPS72, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof, and an antibody to the peptide epitope.
2. The Cas effector of claim 1, wherein the effector is selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, CITED2, PHF15, SS18L1, MLLT6, ASH2L, and GSK3A, or a combination thereof.
3. The Cas effector of claim 1 or 2, wherein the effector is capable of increasing or decreasing expression of a gene.
4. The Cas effector of claim 3, wherein the effector reduces expression of a target gene and is selected from MCRS1, OTUD7B, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof.
5. The Cas effector of claim 3, wherein the effector increases expression of a target gene and is selected from RelB, LDB1, NFKBIB, CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, and VPS72, or a combination thereof.
6. The Cas effector of any one of claims 1-5, wherein the first polypeptide comprises about 2 to about 50 peptide epitopes.
7. The Cas effector of any one of claims 1-6, wherein the first polypeptide comprises more than one copy of the peptide epitope and further comprises at least one linker in between adjacent copies of the peptide epitope.
8. The Cas effector of any one of claims 1-7, wherein the peptide epitope is GCN4 and comprises the amino acid sequence of SEQ ID NO: 85.
9. The Cas effector of any one of claims 1-8, wherein the first polypeptide comprises at least one peptide epitope at the N-terminus and/or at the C-terminus of the Cas protein.
10. The Cas effector of any one of claims 1-9, wherein the first polypeptide comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 87 or 89, or any fragment thereof,
 - or wherein the first polypeptide comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 87 or 89, or any fragment thereof,
11. The Cas effector of any one of claims 1-10, wherein the antibody comprises the amino acid sequence of SEQ ID NO: 81.
12. The Cas effector of any one of claims 1-11, wherein the second polypeptide comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to a sequence selected from SEQ ID NOs: 69, 71, 73, 75, 77, and 79, or any fragment thereof,
 - or wherein the second polypeptide comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to a sequence selected from SEQ ID NOs: 69, 71, 73, 75, 77, and 79, or any fragment thereof,
13. A Cas fusion protein comprising two heterologous polypeptide domains, wherein the first polypeptide domain comprises a Cas protein, and wherein the second polypeptide domain comprises an effector selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, and CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, VPS72, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof.
14. The Cas fusion protein of claim 13, wherein the effector is selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, CITED2, PHF15, SS18L1, MLLT6, ASH2L, and GSK3A, or a combination thereof.
15. The Cas fusion protein of claim 13 or 14, wherein the effector is capable of increasing or decreasing expression of a gene.
16. The Cas fusion protein of claim 15, wherein the effector reduces expression of a target gene and is selected from MCRS1, OTUD7B, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof.
17. The Cas fusion protein of claim 15, wherein the effector increases expression of a target gene and is selected from RelB, LDB1, NFKBIB, CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2,

RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, and VPS72, or a combination thereof.

18. The Cas fusion protein of any one of claims **13-17**, wherein the second polypeptide domain has transcription repression activity, transcription activation activity, de-ubiquitinase activity, p300 recruitment activity, enhancer looping mediation activity, or a combination thereof.

19. The Cas effector of any one of claims **1-12** or the Cas fusion protein of any one of claims **13-18**, wherein the MCRS1 comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 57 or any fragment thereof,

and/or wherein the MCRS1 comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 57, or any fragment thereof,

and/or wherein the MCRS1 comprises the amino acid sequence of SEQ ID NO: 57,

and/or wherein the MCRS1 is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 58, or any fragment thereof,

and/or wherein the MCRS1 is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 58, or any fragment thereof,

and/or wherein the MCRS1 is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 58.

20. The Cas effector of any one of claims **1-12** or the Cas fusion protein of any one of claims **13-18**, wherein the OTUD7B comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to a sequence selected from SEQ ID NO: 59, amino acids 167-440 of SEQ ID NO: 59, or amino acids 792-831 of SEQ ID NO: 59, or any fragment thereof,

and/or wherein the OTUD7B comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to a sequence selected from SEQ ID NO: 59, amino acids 167-440 of SEQ ID NO: 59, or amino acids 792-831 of SEQ ID NO: 59, or any fragment thereof,

and/or wherein the OTUD7B comprises the amino acid sequence selected from SEQ ID NO: 59, amino acids 167-440 of SEQ ID NO: 59, or amino acids 792-831 of SEQ ID NO: 59,

and/or wherein the OTUD7B is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 60, or any fragment thereof,

and/or wherein the OTUD7B is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 60, or any fragment thereof,

and/or wherein the OTUD7B is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 60.

21. The Cas effector of any one of claims **1-12** or the Cas fusion protein of any one of claims **13-18**, wherein the RelB comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 65, or any fragment thereof,

and/or wherein the RelB comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 65, or any fragment thereof,

and/or wherein the RelB comprises the amino acid sequence of SEQ ID NO: 65,

and/or wherein the RelB is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 66 or any fragment thereof,

and/or wherein the RelB is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 66, or any fragment thereof,

and/or wherein the RelB is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 66.

22. The Cas effector of any one of claims **1-12** or the Cas fusion protein of any one of claims **13-18**, wherein the LDB1 comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 61, or any fragment thereof,

and/or wherein the LDB1 comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 61, or any fragment thereof,

and/or wherein the LDB1 comprises the amino acid sequence of SEQ ID NO: 61,

and/or wherein the LDB1 is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 62, or any fragment thereof,

and/or wherein the LDB1 is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 62, or any fragment thereof,

and/or wherein the LDB1 is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 62.

23. The Cas effector of any one of claims **1-12** or the Cas fusion protein of any one of claims **13-18**, wherein the NFKBIB comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 63, or any fragment thereof,

and/or wherein the NFKBIB comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 63, or any fragment thereof,

and/or wherein the NFKBIB comprises the amino acid sequence of SEQ ID NO: 63,

and/or wherein the NFKBIB is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 64, or any fragment thereof,

and/or wherein the NFKBIB is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 64, or any fragment thereof,

and/or wherein the NFKBIB is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 64.

24. The Cas effector of any one of claims **1-12** or the Cas fusion protein of any one of claims **13-18**, wherein the CITED2 comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 67, or any fragment thereof,
and/or wherein the CITED2 comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 67, or any fragment thereof,
and/or wherein the CITED2 comprises the amino acid sequence of SEQ ID NO: 67,
and/or wherein the CITED2 is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 68, or any fragment thereof,
and/or wherein the CITED2 is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 68, or any fragment thereof,
and/or wherein the CITED2 is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 68.

25. The Cas effector of any one of claims **1-12** or the Cas fusion protein of any one of claims **13-18**, wherein the PHF15 comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 133, or any fragment thereof,
and/or wherein the PHF15 comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 133, or any fragment thereof,
and/or wherein the PHF15 comprises the amino acid sequence of SEQ ID NO: 133,
and/or wherein the PHF15 is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 134, or any fragment thereof,
and/or wherein the PHF15 is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 134, or any fragment thereof,
and/or wherein the PHF15 is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 134.

26. The Cas effector of any one of claims **1-12** or the Cas fusion protein of any one of claims **13-18**, wherein the SS18L1 comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 149, or any fragment thereof,
and/or wherein the SS18L1 comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 149, or any fragment thereof,
and/or wherein the SS18L1 comprises the amino acid sequence of SEQ ID NO: 149,
and/or wherein the SS18L1 is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 150, or any fragment thereof,
and/or wherein the SS18L1 is encoded by a polynucleotide comprising a sequence having one, two, three,

four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 150, or any fragment thereof,
and/or wherein the SS18L1 is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 150.

27. The Cas effector of any one of claims **1-12** or the Cas fusion protein of any one of claims **13-18**, wherein the MLLT6 comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 127, or any fragment thereof,
and/or wherein the MLLT6 comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 127, or any fragment thereof,
and/or wherein the MLLT6 comprises the amino acid sequence of SEQ ID NO: 127,
and/or wherein the MLLT6 is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 128, or any fragment thereof,
and/or wherein the MLLT6 is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 128, or any fragment thereof,
and/or wherein the MLLT6 is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 128.

28. The Cas effector of any one of claims **1-12** or the Cas fusion protein of any one of claims **13-18**, wherein the ASH2L comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 103, or any fragment thereof,
and/or wherein the ASH2L comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 103, or any fragment thereof,
and/or wherein the ASH2L comprises the amino acid sequence of SEQ ID NO: 103,
and/or wherein the ASH2L is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 104, or any fragment thereof,
and/or wherein the ASH2L is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 104, or any fragment thereof,
and/or wherein the ASH2L is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 104.

29. The Cas effector of any one of claims **1-12** or the Cas fusion protein of any one of claims **13-18**, wherein the GSK3A comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 117, or any fragment thereof,
and/or wherein the GSK3A comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to SEQ ID NO: 117, or any fragment thereof,
and/or wherein the GSK3A comprises the amino acid sequence of SEQ ID NO: 117,

and/or wherein the GSK3A is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to SEQ ID NO: 118, or any fragment thereof,

and/or wherein the GSK3A is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to SEQ ID NO: 118, or any fragment thereof,

and/or wherein the GSK3A is encoded by a polynucleotide comprising the sequence of SEQ ID NO: 118.

30. The Cas effector of any one of claims **1-12** or the Cas fusion protein of any one of claims **13-18**, wherein the effector is selected from BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, JAZF1, KAT7, KEAP1, MEAF6, MORF4L2, NFYC, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, TADA3, TAF6, TBPL1, VPS72, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, and wherein the effector comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to a sequence selected from SEQ ID NOs: 105, 107, 109, 111, 113, 115, 119, 121, 123, 125, 129, 131, 135, 137, 139, 141, 143, 145, 147, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, or 175, or any fragment thereof,

and/or wherein the effector comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to a sequence selected from SEQ ID NOs: 105, 107, 109, 111, 113, 115, 119, 121, 123, 125, 129, 131, 135, 137, 139, 141, 143, 145, 147, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, or 175, or any fragment thereof,

and/or wherein the effector comprises an amino acid sequence selected from SEQ ID NOs: 105, 107, 109, 111, 113, 115, 119, 121, 123, 125, 129, 131, 135, 137, 139, 141, 143, 145, 147, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, or 175,

and/or wherein the effector is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to a sequence selected from SEQ ID NOs: 106, 108, 110, 112, 114, 116, 120, 122, 124, 126, 130, 132, 136, 138, 140, 142, 144, 146, 148, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, or 176, or any fragment thereof,

and/or wherein the effector is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to a sequence selected from SEQ ID NOs: 106, 108, 110, 112, 114, 116, 120, 122, 124, 126, 130, 132, 136, 138, 140, 142, 144, 146, 148, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, or 176, or any fragment thereof,

and/or wherein the effector is encoded by a polynucleotide comprising a sequence selected from SEQ ID NOs: 106, 108, 110, 112, 114, 116, 120, 122, 124, 126, 130, 132, 136, 138, 140, 142, 144, 146, 148, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, or 176.

31. The Cas effector of any one of claims **1-12** and **19-31** or the Cas fusion protein of claim any one of claims **13-31**,

wherein the Cas protein comprises at least one amino acid mutation that knocks out nuclease activity of the Cas protein.

32. The Cas effector or the Cas fusion protein of claim **31**, wherein the at least one amino acid mutation is at least one of D10A and H840A.

33. The Cas effector of any one of claims **1-12** and **19-32** or the Cas fusion protein of any one of claims **13-32**, wherein the Cas protein comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to one of SEQ ID NOs: 26-29, or any fragment thereof,

or wherein the Cas protein comprises an amino acid sequence having one, two, three, four, five or more changes selected from amino acid substitutions, insertions, or deletions, relative to one of SEQ ID NOs: 26-29, or any fragment thereof,

or wherein the Cas protein comprises the amino acid sequence of one of SEQ ID NOs: 26-29.

34. The Cas effector of any one of claims **1-12** and **19-33** or the Cas fusion protein of any one of claims **13-33**, wherein the Cas protein is encoded by a polynucleotide comprising a sequence having at least 80%, 85%, 90%, 95%, or 98% or greater identity to one of SEQ ID NOs: 30-31, or any fragment thereof,

or wherein the Cas protein is encoded by a polynucleotide comprising a sequence having one, two, three, four, five or more changes selected from nucleotide substitutions, insertions, or deletions, relative to one of SEQ ID NOs: 30-31, or any fragment thereof,

or wherein the Cas protein is encoded by a polynucleotide comprising the sequence of one of SEQ ID NOs: 30-31.

35. A DNA targeting composition comprising:
the Cas effector of any one of claims **1-12** and **19-34** or
the Cas fusion protein of any one of claims **13-34**; and
at least one guide RNA (gRNA) that targets the Cas protein to a target region of a target gene.

36. The DNA targeting composition of claim **35**, wherein the gRNA targets the Cas protein to target region selected from a non-open chromatin region, an open chromatin region, a transcribed region of the target gene, a region upstream of a transcription start site of the target gene, a regulatory element of the target gene, an intron of the target gene, or an exon of the target gene.

37. The DNA targeting composition of claim **35** or **36**, wherein the gRNA targets the Cas protein to a promoter of the target gene.

38. The DNA targeting composition of any one of claims **35-37**, wherein the target region is located between about 1 to about 1000 base pairs upstream of a transcription start site of the target gene.

39. The DNA targeting composition of any one of claims **35-38**, wherein the at least one gRNA comprises a sequence selected from SEQ ID NOs: 96-98 and 101-102, or wherein the at least one gRNA is encoded by a polynucleotide comprising a sequence selected from SEQ ID NOs: 93-95 and 99-100, or wherein the at least one gRNA targets and binds a polynucleotide comprising a sequence selected from SEQ ID NOs: 93-95 and 99-100 or a complement thereof, or a combination thereof.

40. The DNA targeting composition of any one of claims **35-39**, wherein the DNA targeting composition comprises two or more gRNAs, each gRNA binding to a different target region.

41. An isolated polynucleotide sequence encoding the Cas effector of any one of claims **1-12** and **19-34** or the Cas fusion protein of any one of claims **13-34**, or the DNA targeting composition of any one of claims **35-40**.

42. A vector comprising: the isolated polynucleotide sequence of claim **41**.

43. The vector of claim **42**, wherein the vector is an adeno-associated virus (AAV) vector.

44. A cell comprising: the Cas effector of any one of claims **1-12** and **19-34** or the Cas fusion protein of any one of claims **13-34**, or the DNA targeting composition of any one of claims **35-40**, or the isolated polynucleotide sequence of claim **41**, or the vector of claim **42** or **43**, or a combination thereof.

45. A pharmaceutical composition comprising: the Cas effector of any one of claims **1-12** and **19-34** or the Cas fusion protein of any one of claims **13-34**, or the DNA targeting composition of any one of claims **35-40**, or the isolated polynucleotide sequence of claim **41**, or the vector of claim **42** or **43**, or a combination thereof.

46. A method of modulating expression of a gene in a cell or in a subject, the method comprising administering to the cell or the subject the DNA targeting composition of any one of claims **35-40**, or the isolated polynucleotide sequence of claim **41**, or the vector of claim **42** or **43**, or the pharmaceutical composition of claim **45**, or a combination thereof.

47. A method of modulating expression of a gene in a cell or in a subject, the method comprising administering to the cell or the subject an effector selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, VPS72, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof, or a polynucleotide encoding the effector.

48. The method of claim **47**, wherein the effector is targeted to the gene.

49. The method of claim **47** or **48**, wherein the effector is selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, CITED2, PHF15, SS18L1, MLLT6, ASH2L, and GSK3A, or a combination thereof.

50. The method of any one of claims **47-49**, wherein the effector is capable of increasing or decreasing expression of the gene.

51. The method of claim **50**, wherein the effector reduces expression of the gene and is selected from MCRS1, OTUD7B, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof.

52. The method of claim **50**, wherein the effector increases expression of the gene and is selected from RelB, LDB1, NFKBIB, CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, and VPS72, or a combination thereof.

53. The method of any one of claims **46-50** and **52**, wherein the expression of the gene is increased relative to a control.

54. The method of any one of claims **46-51**, wherein the expression of the gene is decreased relative to a control.

55. The method of any one of claims **46-54**, wherein the gene comprises the dystrophin gene, the CD25 gene, the B2M gene, or the TRAC gene.

56. The method of any one of claims **46-55**, wherein the cell is a muscle cell or a T cell.

57. A method of treating a disease in a subject, the method comprising administering to the subject the DNA targeting composition of any one of claims **35-40**, or the isolated polynucleotide sequence of claim **41**, or the vector of claim **42** or **43**, or the cell of claim **44**, or the pharmaceutical composition of claim **45**, or a combination thereof.

58. A method of treating a disease in a subject, the method comprising administering to the subject an effector selected from MCRS1, OTUD7B, RelB, LDB1, NFKBIB, CITED2, ASH2L, BCL7B, C20orf20, DMAP1, DYRK1B, EAF1, FOXR2, GSK3A, JAZF1, KAT7, KEAP1, MEAF6, MLLT6, MORF4L2, NFYC, PHF15, PKIB, POLE4, PRKRIR, PYGO2, RANBP1, RPRD1B, SPIN1, SS18L1, TADA3, TAF6, TBPL1, VPS72, ZNF133, ZNF140, ZNF169, ZNF254, ZNF566, ZNF585A, ZNF689, ZNF765, and ZNF81, or a combination thereof, or a polynucleotide encoding the effector.

59. The method of claim **58**, wherein the effector is targeted to a gene.

60. The method of any one of claims **46-59**, wherein the method treats a disease selected from Duchenne muscular dystrophy (DMD), Becker muscular dystrophy (BMD), and cancer.

* * * * *