(12) STANDARD PATENT (11) Application No. **AU 2018290843 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
CRISPR/Cas-adenine deaminase based compositions, systems, and methods for targeted nucleic acid editing

(51) International Patent Classification(s)
A61K 48/00 (2006.01) ***C12N 15/00*** (2006.01)
C07H 21/04 (2006.01) ***C12N 15/82*** (2006.01)

(21) Application No: **2018290843** (22) Date of Filing: **2018.06.26**

(87) WIPO No: **WO19/005884**

(30) Priority Data

(31) Number	(32) Date	(33) Country
62/528,391	**2017.07.03**	**US**
62/525,181	**2017.06.26**	**US**
62/561,638	**2017.09.21**	**US**
62/534,016	**2017.07.18**	**US**
62/568,304	**2017.10.04**	**US**
62/574,158	**2017.10.18**	**US**
62/591,187	**2017.11.27**	**US**
62/610,105	**2017.12.22**	**US**

(43) Publication Date: **2019.01.03**
(44) Accepted Journal Date: **2025.04.24**

(71) Applicant(s)
The Broad Institute, Inc.;Massachusetts Institute of Technology;President and Fellows of Harvard College

(72) Inventor(s)
ZHANG, Feng;GOOTENBERG, Jonathan;COX, David Benjamin Turitz;ABUDAYYEH, Omar;KANNAN, Soumya

(74) Agent / Attorney
Phillips Ormonde Fitzpatrick, PO Box 323, COLLINS STREET WEST, VIC, 8007, AU

(56) Related Art
US 2016/0304846 A1
MATTHEWS ET AL.: "Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity", NAT STRUCT MOL BIOL., vol. 23, no. 5, May 2016 (2016-05-01), pages 426 - 33, XP055428412
MARIA FERNANDA MONTIEL-GONZ?LEZ ET AL: "An efficient system for selectively altering genetic information within mRNAs", NUCLEIC ACIDS RESEARCH, 23 August 2016 pages gkw738, ISSN: 0305-1048, DOI: 10.1093/nar/gkw738
SHMAKOV, S., ET AL.: ""Diversity and evolution of class 2 CRISPR-Cas systems."", NATURE REVIEWS MICROBIOLOGY, vol. 15, no. 3, JPN7022003260, March 2017 (2017-03-01), pages 169 - 182, DOI: 10.1038/nrmicro.2016.184

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
03 January 2019 (03.01.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/005884 A1

(51) International Patent Classification:
A61K 48/00 (2006.01) *C12N 15/82* (2006.01)
C12N 15/00 (2006.01) *C07H 21/04* (2006.01)

(21) International Application Number:
PCT/US2018/039616

(22) International Filing Date:
26 June 2018 (26.06.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

62/525,181	26 June 2017 (26.06.2017)	US
62/528,391	03 July 2017 (03.07.2017)	US
62/534,016	18 July 2017 (18.07.2017)	US
62/561,638	21 September 2017 (21.09.2017)	US
62/568,304	04 October 2017 (04.10.2017)	US
62/574,158	18 October 2017 (18.10.2017)	US
62/591,187	27 November 2017 (27.11.2017)	US
62/610,105	22 December 2017 (22.12.2017)	US

(71) Applicants: **THE BROAD INSTITUTE, INC.** [US/US]; 415 Main Street, Cambridge, Massachusetts 02142 (US). **MASSACHUSETTS INSTITUTE OF TECHNOLOGY** [US/US]; 77 Massachusetts Avenue, Cambridge, Massachusetts 02139 (US). **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, Massachusetts 02138 (US).

(72) Inventors: **ZHANG, Feng**; c/o 415 Main Street, Cambridge, Massachusetts 02142 (US). **GOOTENBERG, Jonathan**; c/o 17 Quincy Street, Cambridge, Massachusetts 02138 (US). **COX, David Benjamin Turitz**; c/o 77 Massachusetts Ave., Cambridge, Massachusetts 02139 (US). **ABUDAYYEH, Omar**; c/o 77 Massachusetts Ave, Cambridge, Massachusetts 02139 (US). **KANNAN, Soumya**; c/o 415 Main Street, Cambridge, Massachusetts 02142 (US).

(74) Agent: **NIX, F. Brent** et al.; Johnson, Marcou & Isaacs, LLC, P.O. Box 691, Hoschton, Georgia 30548 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH,

(54) Title: CRISPR/CAS-ADENINE DEAMINASE BASED COMPOSITIONS, SYSTEMS, AND METHODS FOR TARGETED NUCLEIC ACID EDITING

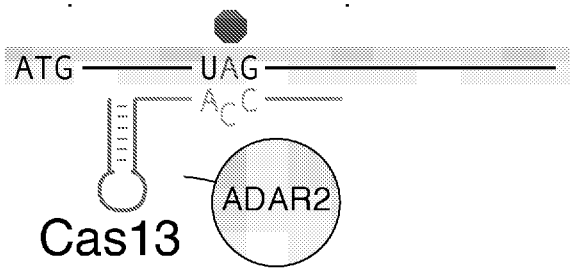


FIG. 1

(57) Abstract: The invention provides for systems, methods, and compositions for targeting and editing nucleic acids. In particular, the invention provides non-naturally occurring or engineered RNA-targeting systems comprising a RNA-targeting Cas13 protein, at least one guide molecule, and at least one adenosine deaminase protein or catalytic domain thereof.

WO 2019/005884 A1

[Continued on next page]

GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

CRISPR/CAS-ADENINE DEAMINASE BASED COMPOSITIONS, SYSTEMS, AND METHODS FOR TARGETED NUCLEIC ACID EDITING

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 62/525,181, filed June 26, 2017, U.S. Provisional Application No. 62/528,391, filed July 3, 2017, U.S. Provisional Application No. 62/534,016, filed July 18, 2017, U.S. Provisional Application No. 62/561,638, filed September 21, 2017, U.S. Provisional Application No. 62/568,304, filed October 4, 2017, U.S. Provisional Application No. 62/574,158, filed October 18, 2017, U.S. Provisional Application No. 62/591,187, filed November 27, 2017, and U.S. Provisional Application No. 62/610,105, filed December 22, 2017. The entire contents of the above-identified applications are hereby fully incorporated herein by reference.

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under grant numbers MH100706, MH110049, and HL141201 awarded by the National Institutes of Health. The government has certain rights in the invention

REFERENCE TO DOCUMENTS CO-FILED IN COMPUTER READABLE FORMAT

[0003] An ASCII compliant text file entitled "Clin_var_pathogenic_SNPS_TC.txt" created on July 3, 2017 and 891,043 bytes in size is filed herewith via EFS-WEB, the contents of which are hereby incorporated herein by reference.

FIELD OF THE INVENTION

[0004] The present invention generally relates to systems, methods, and compositions for targeting and editing nucleic acids, in particular for programmable deamination of adenine at a target locus of interest.

BACKGROUND

[0005] Recent advances in genome sequencing techniques and analysis methods have significantly accelerated the ability to catalog and map genetic factors associated with a diverse range of biological functions and diseases. Precise genome targeting technologies are needed

to enable systematic reverse engineering of causal genetic variations by allowing selective perturbation of individual genetic elements, as well as to advance synthetic biology, biotechnological, and medical applications. Although genome-editing techniques such as designer zinc fingers, transcription activator-like effectors (TALEs), or homing meganucleases are available for producing targeted genome perturbations, there remains a need for new genome engineering technologies that employ novel strategies and molecular mechanisms and are affordable, easy to set up, scalable, and amenable to targeting multiple positions within the eukaryotic genome. This would provide a major resource for new applications in genome engineering and biotechnology.

[0006] Programmable deamination of cytosine has been reported and may be used for correction of A→G and T→C point mutations. For example, Komor et al., Nature (2016) 533:420–424 reports targeted deamination of cytosine by APOBEC1 cytidine deaminase in a non-targeted DNA stranded displaced by the binding of a Cas9-guide RNA complex to a targeted DNA strand, which results in conversion of cytosine to uracil. See also Kim et al., Nature Biotechnology (2017) 35:371-376; Shimatani et al., Nature Biotechnology (2017) doi:10.1038/nbt.3833; Zong et al., Nature Biotechnology (2017) doi:10.1038/nbt.3811; Yang Nature Communication (2016) doi:10.1038/ncomms13330.

SUMMARY OF THE INVENTION

[0007] The present application relates to modifying a target RNA sequence of interest. Using RNA-targeting rather than DNA targeting offers several advantages relevant for therapeutic development. First, there are substantial safety benefits to targeting RNA: there will be fewer off-target events because the available sequence space in the transcriptome is significantly smaller than the genome, and if an off-target event does occur, it will be transient and less likely to induce negative side effects. Second, RNA-targeting therapeutics will be more efficient because they are cell-type independent and not have to enter the nucleus, making them easier to deliver.

[0008] At least a first aspect of the invention relates to a method of modifying an Adenine in a target RNA sequence of interest. In particular embodiments, the method comprises delivering to said target RNA: (a) a catalytically inactive (dead) Cas13 protein; (b) a guide molecule which comprises a guide sequence linked to a direct repeat sequence; and (c) an adenosine deaminase protein or catalytic domain thereof; wherein said adenosine deaminase protein or catalytic domain thereof is covalently or non-covalently linked to said dead Cas13

protein or said guide molecule or is adapted to link thereto after delivery; wherein guide molecule forms a complex with said dead Cas13 protein and directs said complex to bind said target RNA sequence of interest, wherein said guide sequence is capable of hybridizing with a target sequence comprising said Adenine to form an RNA duplex, wherein said guide sequence comprises a non-pairing Cytosine at a position corresponding to said Adenine resulting in an A-C mismatch in the RNA duplex formed; wherein said adenosine deaminase protein or catalytic domain thereof deaminates said Adenine in said RNA duplex.

[0009] In certain example embodiment the Cas13 protein is Cas13a, Cas13b or Cas 13c.

[0010] The adenosine deaminase protein or catalytic domain thereof is fused to N- or C-terminus of said dead Cas13 protein. In certain example embodiments, the adenosine deaminase protein or catalytic domain thereof is fused to said dead Cas13 protein by a linker. The linker may be $(GGGGS)_{3-11}$ (SEQ ID Nos. 1-9) GSG_5 (SEQ ID No. 10) or LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR (SEQ ID No. 11).

[0011] In certain example embodiments, the adenosine deaminase protein or catalytic domain thereof is linked to an adaptor protein and said guide molecule or said dead Cas13 protein comprises an aptamer sequence capable of binding to said adaptor protein. The adaptor sequence may be selected from MS2, PP7, Qβ, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕCb5, ϕCb8r, ϕCb12r, ϕCb23r, 7s and PRR1.

[0012] In certain example embodiments, the adenosine deaminase protein or catalytic domain thereof is inserted into an internal loop of said dead Cas13 protein. In certain example embodiments, the Cas13a protein comprises one or more mutations in the two HEPN domains, particularly at postion R474 and R1046 of Cas 13a protein originating from *Leptotrichia wadei* or amino acid positions corresponding thereto of a Cas13a ortholog.

[0013] In certain example embodiments, the Cas 13 protein is a Cas13b proteins, and the Cas13b comprises a mutation in one or more of positions R116, H121, R1177, H1182 of Cas13b protein originating from *Bergeyella zoohelcum* ATCC 43767 or amino acid positions corresponding thereto of a Cas13b ortholog. In certain other example embodiments, the mutation is one or more of R116A, H121A, R1177A, H1182A of Cas13b protein originating from *Bergeyella zoohelcum* ATCC 43767 or amino acid positions corresponding thereto of a Cas13b ortholog.

[0014] In certain example embodiments, the guide sequence has a length of about 29-53 nt capable of forming said RNA duplex with said target sequence. In certain other example embodiments, the guide sequence has a length of about 40-50 nt capable of forming said RNA

duplex with said target sequence. In certain example embodiments, the distance between said non-pairing C and the 5' end of said guide sequence is 20-30 nucleotides.

[0015] In certain example embodiments, the adenosine deaminase protein or catalytic domain thereof is a human, cephalopod, or Drosophila adenosine deaminase protein or catalytic domain thereof. In certain example embodiments, the adenosine deaminase protein or catalytic domain thereof has been modified to comprise a mutation at glutamic acid^{488} of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In certain example embodiments, the glutamic acid residue may be at position 488 or a corresponding position in a homologous ADAR protein is replaced by a glutamine residue (E488Q).

[0016] In certain other example embodiments, the adenosine deaminase protein or catalytic domain thereof is a mutated hADAR2d comprising mutation E488Q or a mutated hADAR1d comprising mutation E1008Q.

[0017] In certain example embodiments, the guide sequence comprises more than one mismatch corresponding to different adenosine sites in the target RNA sequence or wherein two guide molecules are used, each comprising a mismatch corresponding to a different adenosine sites in the target RNA sequence.

[0018] In certain example embodiments, the Cas13 protein and optionally said adenosine deaminase protein or catalytic domain thereof comprise one or more heterologous nuclear localization signal(s) (NLS(s)).

[0019] In certain example embodiments, the method further comprises, determining the target sequence of interest and selecting an adenosine deaminase protein or catalytic domain thereof which most efficiently deaminates said Adenine present in then target sequence.

[0020] The target RNA sequence of interest may be within a cell. The cell may be a eukaryotic cell, a non-human animal cell, a human cell, a plant cell. The target locus of interest may be within an animal or plant.

[0021] The target RNA sequence of interest may comprise in an RNA polynucleotide *in vitro*.

[0022] The components of the systems described herein may be delivered to said cell as a ribonucleoprotein complex or as one or more polynucleotide molecules. The one or more polynucleotide molecules may comprise one or more mRNA molecules encoding the components. The one or more polynucleotide molecules may be comprised within one or more vectors. The one or more polynucleotide molecules may further comprise one or more regulatory elements operably configured to express said Cas13 protein, said guide molecule, and said adenosine deaminase protein or catalytic domain thereof, optionally wherein said one

or more regulatory elements comprise inducible promoters. The one or more polynucleotide molecules or said ribonucleoprotein complex may be delivered via particles, vesicles, or one or more viral vectors. The particles may comprise a lipid, a sugar, a metal or a protein. The particles may comprise lipid nanoparticles. The vesicles may comprise exosomes or liposomes. The one or more viral vectors may comprise one or more of adenovirus, one or more lentivirus or one or more adeno-associated virus.

[0023] The methods disclosed herein may be used to modify a cell, a cell line or an organism by manipulation of one or more target RNA sequences.

[0024] In certain example embodiments, the deamination of said Adenine in said target RNA of interest remedies a disease caused by transcripts containing a pathogenic G→A or C→T point mutation.

[0025] The methods maybe be used to treat a disase. In certain example embodiments, the disease is selected from Meier-Gorlin syndrome, Seckel syndrome 4, Joubert syndrome 5, Leber congenital amaurosis 10; Charcot-Marie-Tooth disease, type 2; Charcot-Marie-Tooth disease, type 2; Usher syndrome, type 2C; Spinocerebellar ataxia 28; Spinocerebellar ataxia 28; Spinocerebellar ataxia 28; Long QT syndrome 2; Sjögren-Larsson syndrome; Hereditary fructosuria; Hereditary fructosuria; Neuroblastoma; Neuroblastoma; Kallmann syndrome 1; Kallmann syndrome 1; Kallmann syndrome 1; Metachromatic leukodystrophy, Rett syndrome, Amyotrophic lateral sclerosis type 10, Li-Fraumeni syndrome, or a disease listed in Table 5. The disease may be a premature termination disease.

[0026] The methods disclosed herein, may be used to make a modification that affects the fertility of an organism. The modification may affects splicing of said target RNA sequence. The modification mayintroduces a mutation in a transcript introducing an amino acid change and causing expression of a new antigen in a cancer cell.

[0027] In certain example embodiments, the target RNA may be a microRNA or comprised within a microRNA. In certain example embodiments, the deamination of said Adenine in said target RNA of interest causes a gain of function or a loss of function of a gene.In certain example embodiments, the gene is a gene expressed by a cancer cell.

[0028] In another aspect, the invention comprises a modified cell or progeny thereof that is obtained using the methods disclosed herein, wherein said cell comprises a hypoxanthine or a guanine in replace of said Adenine in said target RNA of interest compared to a corresponding cell not subjected to the method. The modified cell or progeny thereof may be a eukaryotic cell an animal cell, a human cell, a therapeutic T cell, an antibody-producing B cell, a plant cell.

[0029] In another aspect, the invention comprises a non-human animal comprising said modified cell or progeny therof. The modified may be a plant cell.

[0030] In another aspect, the invention comprises a method for cell therapy, comprising administering to a patient in need thereof the modified cells disclosed herein, wherein the presence of said modified cell remedies a disease in the patient.

[0031] In another aspect, the invention is directed to an engineered, non-naturally occurring system suitable for modifying an Adenine in a target locus of interest, comprising A) a guide molecule which comprises a guide sequence linked to a direct repeat sequence, or a nucleotide sequence encoding said guide molecule; B) a catalytically inactive Cas13 protein, or a nucleotide sequence encoding said catalytically inactive Cas13 protein; C) an adenosine deaminase protein or catalytic domain thereof, or a nucleotide sequence encoding said adenosine deaminase protein or catalytic domain thereof; wherein said adenosine deaminase protein or catalytic domain thereof is covalently or non-covalently linked to said Cas13 protein or said guide molecule or is adapted to link thereto after delivery; wherein said guide sequence is capable of hybridizing with a target RNA sequence comprising an Adenine to form an RNA duplex, wherein said guide sequence comprises a non-pairing Cytosine at a position corresponding to said Adenine resulting in an A-C mismatch in the RNA duplex formed.

[0032] In another aspect, the invention is directed to an engineered, non-naturally occurring vector system suitable for modifying an Adenine in a target locus of interest, comprising the nucleotide sequences of a), b) and c)

[0033] In another aspect, the invention is directed to an engineered, non-naturally occurring vector system, comprising one or more vectors comprising: a first regulatory element operably linked to a nucleotide sequence encoding said guide molecule which comprises said guide sequence, a second regulatory element operably linked to a nucleotide sequence encoding said catalytically inactive Cas13 protein; and a nucleotide sequence encoding an adenosine deaminase protein or catalytic domain thereof which is under control of said first or second regulatory element or operably linked to a third regulatory element; wherein, if said nucleotide sequence encoding an adenosine deaminase protein or catalytic domain thereof is operably linked to a third regulatory element, said adenosine deaminase protein or catalytic domain thereof is adapted to link to said guide molecule or said Cas13 protein after expression; wherein components A), B) and C) are located on the same or different vectors of the system.

[0034] As the methods disclosed herein demonstate the ability of Cas13 proteins to function in mammalian cells for binding and specificity of cleaving RNA, additional extended

2018290843 07 Apr 2025

applications include editing splice variants, and measuring how RNA-binding proteins interact with RNA.

[0035] In another aspect, the invention is directed to *in vitro* or *ex vivo* host cell or progeny thereof or cell line or progeny thereof comprising the systems disclosed herein. The host cell or progeny thereof may be a a eukaryotice cell, an animal cell, a human cell, or a plant cell.

[0036] In another aspect, the invention relates to an adenosine deaminase protein or catalytic domain thereof and comprising one or more mutations as described herein elsewhere.

[0037] In certain embodiments, such adenosine deaminase protein or catalytic domain thereof is covalently or non-covalently linked to a nucleic acid binding molecule or targeting domain as described herein elsewhere. Accordingly, the invention further relates to compositions comprising said adenosine deaminase protein or catalytic domain and a nucleic acid binding molecule and to fusion proteins of said adenosine deaminase protein or catalytic domain and said nucleic acid binding molecule.

[0038] In another aspect the invention relates to an engineered composition for site directed base editing comprising a targeting domain and an adenosine deaminase, or catalytic domain thereof. In particular embodiments, the targeting domain is an oligonucleotide targeting domain. In particular embodiments, the adenosine deaminase, or catalytic domain thereof, comprises one or more mutations that increase activity or specificity of the adenosine deaminase relative to wild type. In particular embodiments, the adenosine deaminase comprises one or more mutations that changes the functionality of the adenosine deaminase relative to wild type, preferably an ability of the adenosine deaminase to deaminate cytosine as described elsewhere herein. In particular embodiments, the targeting domain is a CRISPR system comprising a CRISPR effector protein, or functional domain thereof, and a guide molecule, more particularly the CRISPR system is catalytically inactive. In particular embodiments, the CRISPR system comprises an RNA-binding protein, preferably Cas13, preferably the Cas13 protein is Cas13a, Cas13b or Cas13c, preferably wherein said Cas13 a Cas13 listed in any of Tables 1, 2, 3, 4, or 6 or is from a bacterial species listed in any of Tables 1, 2, 3, 4, or 6, preferably wherein said Cas13 protein is Prevotella sp.P5-125 Cas13b, Porphyromas gulae Cas13b, or Riemerella anatipestifer Cas13b; preferably Prevotella sp.P5-125 Cas13b. Accordingly, embodiments of the invention include an engineered composition for site directed base editing comprising a Cas13 and an adenosine deaminase, or catalytic domain thereof. In a preferred embodiment,

2018290843 07 Apr 2025

the invention provides an engineered composition for site directed base editing comprising a Cas13 and an adenosine deaminase, or catalytic domain thereof, wherein the adenosine deaminase comprises one or more mutations that changes the functionality of the adenosine deaminase relative to wild type, wherein the one or more mutations comprise: (i) one or more mutations at one or more amino acid residues selected from T375, a S486, and a V351 in a human adenosine deaminase acting on RNA 2 (ADAR2) or deaminase domain thereof comprising an amino acid sequence SEQ ID NO: 650 or in a corresponding position thereto in a structural or functional homolog or ortholog ADAR or adenosine deaminase domain thereof, or (ii) an E448Q mutation and one or more mutations at one or more amino acid residues selected from T375, a S486, and a V351 in a human ADAR2 or deaminase domain thereof comprising an amino acid sequence SEQ ID NO: 650 or in a corresponding position thereto in a structural or functional homolog or ortholog ADAR or adenosine deaminase domain thereof, and the mutations in (i) or (ii) changes the ability of the adenosine deaminase to deaminate cytosine. In particular embodiments, the Cas13 protein is a Cas13a protein and said Cas13a comprises one or more mutations the two HEPN domains, particularly at position R474 and R1046 of Cas13a protein originating from *Leptotrichia wadei* or amino acid positions corresponding thereto of a Cas13a ortholog, or wherein said Cas13 protein is a Cas13b protein and said Cas13b comprises a mutation in one or more of positions R116, H121, R1177, H1182, preferably

R116A, H121A, R1177A, H1182A of Cas13b protein originating from *Bergeyella zoohelcum* ATCC 43767 or amino acid positions corresponding thereto of a Cas13b ortholog, or wherein said Cas13 protein is a Cas13b protein and said Cas13b comprises a mutation in one or more of positions R128, H133, R1053, H1058, preferably H133 and H1058, preferably H133A and H1058A, of a Cas13b protein originating from Prevotella sp. P5-125 or amino acid positions corresponding thereto of a Cas13b ortholog as described elsewhere herein or the Cas 13 is truncated, preferably C-terminally truncated, preferably wherein said Cas13 is a truncated functional variant of the corresponding wild type Cas13, optionally wherein said truncated Cas13b is encoded by nt 1-984 of Prevotella sp.P5-125 Cas13b or the corresponding nt of a Cas13b orthologue or homologue.

[0039] In particular embodiments, the guide molecule of the targeting domain comprises a guide sequence is capable of hybridizing with a target RNA sequence comprising an Adenine to form an RNA duplex, wherein said guide sequence comprises a non-pairing Cytosine at a position corresponding to said Adenine resulting in an A-C mismatch in the RNA duplex formed. In particular embodiments, the guide sequence has a length of about 20-53 nt, preferably 25-53 nt, more preferably 29-53 nt or 40-50 nt capable of forming said RNA duplex with said target sequence, and/or wherein the distance between said non-pairing C and the 5' end of said guide sequence is 20-30 nucleotides. In particular embodiments, the guide sequence comprises more than one mismatch corresponding to different adenosine sites in the target RNA sequence or wherein two guide molecules are used, each comprising a mismatch corresponding to a different adenosine sites in the target RNA sequence.

[0040] In particular embodiments, of the composition the adenosine deaminase protein or catalytic domain thereof is fused to a N- or C-terminus of said oligonucleotide targeting protein, optionally by a linker as described elsewhere herein. Alternatively, said adenosine deaminase protein or catalytic domain thereof is inserted into an internal loop of said dead Cas13 protein. In a further alternative embodiment, the adenosine deaminase protein or catalytic domain thereof is linked to an adaptor protein and said guide molecule or said dead Cas13 protein comprises an aptamer sequence capable of binding to said adaptor protein as described elsewhere herein.

[0041] In particular embodiments of the composition the adenosine deaminase protein or catalytic domain thereof capable of deaminating adenosine or cytosine in RNA or is an RNA specific adenosine deaminase and/or is a bacterial, human, cephalopod, or Drosophila adenosine deaminase protein or catalytic domain thereof, preferably TadA, more preferably

ADAR, optionally huADAR, optionally (hu)ADAR1 or (hu)ADAR2, preferably huADAR2 or catalytic domain thereof.

[0042] In particular embodiments of the composition, the targeting domain and optionally the adenosine protein or catalytic domain thereof comprise one or more heterologous nuclear export signal(s) (NES(s)) or nuclear localization signal(s) (NLS(s)), preferably an HIV Rev NES or MAPK NES, preferably C-terminal.

[0043] A further aspect of the invention relates to the composition as envisaged herein for use in prophylactic or therapeutic treatment, preferably wherein said target locus of interest is within a human or animal and to methods of modifying an Adenine or Cytidine in a target RNA sequence of interest, comprising delivering to said target RNA, the composition as described hereinabove. In particular embodiments, the CRISPR system and the adenonsine deaminase, or catalytic domain thereof, are delivered as one or more polynucleotide molecules, as a ribonucleoprotein complex, optionally via particles, vesicles, or one or more viral vectors. In particular embodiments, the composition is for use in the treatment or prevention of a disease caused by transcripts containing a pathogenic G→A or C→T point mutation. In particular embodiments, the invention thus comprises compositions for use in therapy. This implies that the methods can be performed in vivo, ex vivo or in vitro. In particular embodiments, the methods are not methods of treatment of the animal or human body or a method for modifying the germ line genetic identity of a human cell. In particular embodiments; when carrying out the method, the target RNa is not comprised within a human or animal cell. In particular embodiments, when the target is a human or animal target, the method is carried out ex vivo or in vitro

[0044] A further aspects relates to an isolated cell obtained or obtainable from the methods described above and/or comprising the composition described above or progeny of said modified cell, preferably wherein said cell comprises a hypoxanthine or a guanine in replace of said Adenine in said target RNA of interest compared to a corresponding cell not subjected to the method. In particular embodiments, the cell is a eukaryotic cell, preferably a human or non-human animal cell, optionally a therapeutic T cell or an antibody-producing B-cell or wherein said cell is a plant cell. A further aspect provides a non-human animal or a plant comprising said modified cell or progeny thereof . Yet a further aspect provides the modified cell as described hereinabove for use in therapy, preferably cell therapy.

BRIEF DESCRIPTION OF THE DRAWINGS

[0045] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0046] **FIG. 1** illustrates an example embodiment of the invention for targeted deamination of adenine at a target RNA sequence of interest, exemplified herein with a Cas13b protein.

[0047] **FIG. 2** illustrates the Development of RNA editing as a therapeutic strategy to treat human disease at the transcript level such as when using Cas13b. Schematic of RNA base editing by Cas13-ADAR2 fusion targeting an engineered pre-termination stop codon in the luciferase transcript.

[0048] **FIG. 3** Guide position and length optimization to restore luciferase expression.

[0049] **FIG. 4** Exemplary sequences of adenine deaminase proteins. (SEQ ID Nos. 650 - 656)

[0050] **FIG. 5** Guides used in an exemplary emodiment (SEQ ID Nos. 657 – 660 and 703)

[0051] **FIG. 6** : Editing efficiency correlates to edited base being further away from the DR and having a long RNA duplex, which is accomplished by extending the guide length

[0052] **FIG. 7** Greater editing efficiency the further the editing site is away from the DR/protein binding area.

[0053] **FIG. 8** Distance of edited site from DR

[0054] **FIG. 9A and B**: Fused ADAR1 or ADAR2 to Cas13b12 (double R HEPN mutant) on the N or C-terminus. Guides are perfect matches to the stop codon in luciferase. Signal appears correlated with distance between edited base and 5' end of the guide, with shorter distances providing better editing.

[0055] **FIG. 10**: Cluc/Gluc tiling for Cas13a/Cas13b interference

[0056] **FIG. 11**: ADAR editing quantification by NGS (luciferase reporter).

[0057] **FIG. 12**: ADAR editing quantification by NGS (KRAS and PPIB).

[0058] **FIG. 13**: Cas13a/b + shRNA specificity from RNA Seq

[0059] **FIG. 14**: Mismatch specificity to reduce off targets (A:A or A:G) (SEQ ID Nos. 661 - 668)

[0060] **FIG. 15**: Mismatch for on-target activity

[0061] **FIG. 16**: ADAR Motif preference

[0062] **FIG. 17**: Larger bubbles to enhance RNA editing efficiency

[0063] **FIG. 18**: Editing of multiple A's in a transcript (SEQ ID Nos. 669-672)

[0064] **FIG. 19**: Guide length titration for RNA editing

[0065] **FIG. 20**: Mammalian codon-optimized Cas13b orthologs mediate highly efficient RNA knockdown. (A) Schematic of representative Cas13a, Cas13b, and Cas13c loci and associated crRNAs. (B) Schematic of luciferase assay to measure Cas13a cleavage activity in HEK293FT cells. (C) RNA knockdown efficiency using two different guides targeting Cluc with 19 Cas13a, 15 Cas13b, and 5 Cas13c orthologs. Luciferase expression is normalized to the expression in non-targeting guide control conditions. (D) The top 7 orthologs performing in part C are assayed for activity with three different NLS and NES tags with two different guide RNAs targeting Cluc. (E) Cas13b12 and Cas13a2 (LwCas13a) are compared for knockdown activity against Gluc and Cluc. Guides are tiled along the transcripts and guides between Cas13b12 and Cas13a2 are position matched. (F) Guide knockdown for Cas13a2, Cas13b6, Cas13b11, and Cas13b12 against the endogenous KRAS transcript and are compared against corresponding shRNAs.

[0066] **FIG. 21**: Cas13 enzymes mediate specific RNA knockdown in mammalian cells. (A) Schematic of semi-degenerate target sequences for Cas13a/b mismatch specificity testing. (SEQ ID Nos. 673-694) (B) Heatmap of single mismatch knockdown data for Cas13 a/b. Knockdown is normalized to non-targeting (NT) guides for each enzyme. (C) Double mismatch knockdown data for Cas13a. The position of each mismatch is indicated on the X and Y axes. Knockdown data is the sum of all double mismatches for a given set of positions. Data is normalized to NT guides for each enzyme. (D) Double mismatch knockdown data for Cas13b. See C for description. (E) RNA-seq data comparing transcriptome-wide specificity for Cas13 a/b and shRNA for position-matched guides. The Y axis represents read counts for the targeting condition and the X axis represents counts for the non-targeting condition. (F) RNA expression as calculated from RNA-seq data for Cas13 a/b and shRNA. (G) Significant off-targets for Cas13 a/b and shRNA from RNA-seq data. Significant off-targets were calculated using FDR <0.05.

[0067] **FIG. 22**: Catalytically inactive Cas13b-ADAR fusions enable targeted RNA editing in mammalian cells. (A) Schematic of RNA editing with Cas13b-ADAR fusion proteins to remove stop codons on the Cypridina luciferase transcript. (B) RNA editing comparison between Cas13b fused with wild-type ADAR2 and Cas13b fused with the hyperactive ADAR2 E488Q mutant for multiple guide positions. Luciferase expression is normalized to Gaussia luciferase control values. (C) RNA editing comparisons between 30, 50, 70, and 84 nt guides designed to target various positions surrounding the editing site. (D) Effect of surrounding

motif sequence on ADAR editing efficiency on the Cypridina luciferase transcript. (SEQ ID No. 695) (E) Schematic showing the position and length of guides used for sequencing quantification relative to the stop codon on the Cypridina luciferase transcript. (F) On- and off-target editing efficiencies for each guide design at the corresponding adenine bases on the Cypridina luciferase transcript as quantified by sequencing. (G) Luciferase readout of guides with varied bases opposite to the targeted adenine.

[0068] **FIG. 23**: Endogenous RNA editing with Cas13b-ADAR fusions. (A) Next generation sequencing of endogenous Cas13b12-ADAR editing of endogenous KRAS and PPIB loci. Two different regions per transcript were targeted and A->G editing was quantified at all adenines in the vicinity of the targeted adenine.

[0069] **FIG. 24**: Strategy for determining optimal guide position.

[0070] **FIG. 25**: (A) Cas13b-huADAR2 promotes repair of mutated luciferase transcripts. (B) Cas13b-huADAR1 promotes repair of mutated luciferase transcripts. (C) Comparison of human ADAR1 and human ADAR2.

[0071] **FIG. 26**: Comparison of E488Q vs. wt dADAR2 editing. E488Q is a hyperactive mutant of dADAR2.

[0072] **FIG. 27**: Transcripts targeted by Cas13b-huADAR2-E488Q contain the expected A-G edit. (A) heatmap. (B) Positions in template. Only A sites are shown with the editing rate to G as in heatmap.

[0073] **FIG. 28**: Endogenous tiling of guides. (A) KRAS: heatmap. Only A sites are shown with the editing rate to G as in heatmap. (B) Positions in template (bottom). (C) PPIB: heatmap. Only A sites are shown with the editing rate to G as in heatmap. Positions in template (D).

[0074] **FIG. 29**: Non-targeting editing.

[0075] **FIG. 30**: Linker optimization.

[0076] **FIG. 31**: Cas13b ADAR can be used to correct pathogenic A>G mutations from patients in expressed cDNAs.

[0077] **FIG. 32**: Cas13b-ADAR has a slight restriction on 5' G motifs.

[0078] **FIG. 33**: Screening degenerate PFS locations for effect on editing efficiency. All PFS (4-N) identities have higher editing than non-targeting. Fig A. (SEQ ID Nos. 696 - 699)

[0079] **FIG. 34**: Reducing off-target editing in the target transcript.

[0080] **FIG. 35**: Reducing off-target editing in the target transcript.

[0081] **FIG. 36**: Cas13b-ADAR transcriptome specificity. On-target editing is 71%. (A) targeting guide; 482 significant sites. (B) non-targeting guide; 949 significant sites. Note that

chromosome 0 is Gluc and chromosome 1 is Cluc; human chromosomes are then in order after that.

[0082] **FIG. 37**: Cas13b-ADAR transcriptome specificity. (A) targeting guide. (B) non-targeting guide.

[0083] **FIG. 38**: Cas13b has the highest efficiency compared to competing ADAR editing strategies.

[0084] **FIG. 39**: Competing RNA editing systems. (A-B) BoxB; on-target editing is 63%; (A) targeting guide – 2020 significant sites; (B) non-targeting guide – 1805 significant sites. (C-D) Stafforst; on-target editing is 36%; (C) targeting guide – 176 significant sites; (D) non-targeting guide – 186 significant sites.

[0085] **FIG. 40**: Dose titration of ADAR. crRNA amount is constant.

[0086] **FIG. 41**: Dose response effect on specificity. (A-B) 150 ng Cas13-ADAR; on-target editing is 83%; (A) targeting guide – 1231 significant sites; (B) non-targeting guide – 520 significant sites. (C-D) 10 ng Cas13-ADAR; on-target editing is 80%; (C) targeting guide – 347 significant sites; (D) non-targeting guide – 223 significant sites.

[0087] **FIG. 42**: ADAR1 seems more specific than ADAR2. On-target editing is 29%. (A) targeting guide; 11 significant sites. (B) non-targeting guide; 6 significant sites. Note that chromosome 0 is Gluc and chromosome 1 is Cluc; human chromosomes are then in order after that.

[0088] **FIG. 43**: ADAR specificity mutants have enhanced specificity. (A) Targeting guide. (B) Non-targeting guide. (C) Targeting to non-targeting ratio. (D) Targeting and non-targeting guide.

[0089] **FIG. 44**: ADAR mutant luciferase results plotted along the contact points of each residue with the RNA target.

[0090] **FIG. 45**: ADAR specificity mutants have enhanced specificity. Purple points are mutants selected for whole transcriptome off-target NGS analysis. Red point is the starting point (i.e. E488Q mutant). Note that all additional mutants also have the E488Q mutation.

[0091] **FIG. 46**: ADAR mutants are more specific according to NGS. (A) on target. (B) Off-target.

[0092] **FIG. 47**: Luciferase data on ADAR specificity mutants matches the NGS. (A) Targeting guide selected for NGS. (B) Non-targeting guide selected for NGS. Luciferase data matches the NGS data in FIG.46. The orthologs that have fewer activity with non-targeting guide have fewer off-targets across the transcriptome and their on-target editing efficiency can be predicted by the targeting guide luciferase condition.

[0093] **FIG. 48**: C-terminal truncations of Cas13b 12 are still highly active in ADAR editing.

[0094] **FIG. 49**: Characterization of a highly active Cas13b ortholog for RNA knockdown A) Schematic of stereotypical Cas13 loci and corresponding crRNA structure. B) Evaluation of 19 Cas13a, 15 Cas13b, and 7 Cas13c orthologs for luciferase knockdown using two different guides. Orthologs with efficient knockdown using both guides are labeled with their host organism name. Values are normalized to a non-targeting guide with designed against the *E. coli LacZ* transcript, with no homology to the human transcriptome.C) PspCas13b and LwaCas13a knockdown activity are compared by tiling guides against Gluc and measuring luciferase expression. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig. 49B. D) PspCas13b and LwaCas13a knockdown activity are compared by tiling guides against Cluc and measuring luciferase expression. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig. 49B. E) Expression levels in log2(transcripts per million (TPM)) values of all genes detected in RNA-seq libraries of non-targeting control (x-axis) compared to Gluc-targeting condition (y-axis) for LwaCas13a (red) and shRNA (black). Shown is the mean of three biological replicates. The Gluc transcript data point is labeled. Non-targeting guide is the same as in Fig. 49B. F) Expression levels in log2(transcripts per million (TPM)) values of all genes detected in RNA-seq libraries of non-targeting control (x-axis) compared to Gluc-targeting condition (y-axis) for PspCas13b (blue) and shRNA (black). Shown is the mean of three biological replicates. The Gluc transcript data point is labeled. Non-targeting guide is the same as in Fig. 49B. G) Number of significant off-targets from Gluc knockdown for LwaCas13a, PspCas13b, and shRNA from the transcriptome wide analysis in E and F.

[0095] **FIG. 50**: Engineering dCas13b-ADAR fusions for RNA editing A) Schematic of RNA editing by dCas13b-ADAR fusion proteins. Catalytically dead Cas13b (dCas13b) is fused to the deaminase domain of human ADAR ($ADAR_{DD}$), which naturally deaminates adenosines to insosines in dsRNA. The crRNA specifies the target site by hybridizing to the bases surrounding the target adenosine, creating a dsRNA structure for editing, and recruiting the dCas13b-$ADAR_{DD}$ fusion. A mismatched cytidine in the crRNA opposite the target adenosine enhances the editing reaction, promoting target adenosine deamination to inosine, a base that functionally mimics guanosine in many cellular reactions. B) Schematic of *Cypridina* luciferase W85X target and targeting guide design. (SEQ ID Nos. 700 and 701) Deamination of the target adenosine restores the stop codon to the wildtype tryptophan. Spacer length is the region of the guide that contains homology to the target sequence. Mismatch distance is the

number of bases between the 3' end of the spacer and the mismatched cytidine. The cytidine mismatched base is included as part of the mismatch distance calculation. C) Quantification of luciferase activity restoration for Cas13b-dADAR1 (left) and Cas13b-ADAR2-cd (right) with tiling guides of length 30, 50, 70, or 84 nt. All guides with even mismatch distances are tested for each guide length. Values are background subtracted relative to a 30nt non-targeting guide that is randomized with no sequence homology to the human transcriptome. D) Schematic of target site for targeting *Cypridinia* luciferase W85X. (SEQ ID No. 702) E) Sequencing quantification of A->I editing for 50 nt guides targeting *Cypridinia* luciferase W85X. Blue triangle indicates the targeted adenosine. For each guide, the region of duplex RNA is outlined in red. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig. 50C.

[0096] **FIG. 51**: Measuring sequence flexibility for RNA editing by REPAIRv1 Schematic of screen for determining Protospacer Flanking Site (PFS) preferences of RNA editing by REPAIRv1. A randomized PFS sequence is cloned 5' to a target site for REPAIR editing. Following exposure to REPAIR, deep sequencing of reverse transcribed RNA from the target site and PFS is used to associate edited reads with PFS sequences. B) Distributions of RNA editing efficiencies for all 4-N PFS combinations at two different editing sites. C) Quantification of the percent editing of REPAIRv1 at Cluc W85 across all possible 3 base motifs. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig. 50C. D) Heatmap of 5' and 3' base preferences of RNA editing at Cluc W85 for all possible 3 base motifs

[0097] **FIG. 52**: Correction of disease-relevant mutations with REPAIRv1 A) Schematic of target and guide design for targeting *AVPR2* 878G>A. (SEQ ID Nos. 705-708) B) The 878G>A mutation in *AVPR2* is corrected to varying percentages using REPAIRv1 with three different guide designs. For each guide, the region of duplex RNA is outlined in red. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig. 50C. C) Schematic of target and guide design for targeting *FANCC* 1517G>A. (SEQ ID Nos. 709-712) D) The 1517G>A mutation in *FANCC* is corrected to varying percentages using REPAIRv1 with three different guide designs. For each guide, the region of duplex RNA is outlined in red. The heatmap scale bar is the same as in panel B. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig. 50C. E) Quantification of the percent editing of 34 different disease-relevant G>A mutations using REPAIRv1. Non-targeting guide is the same as in Fig. 50C. F) Analysis of all the possible G>A mutations that could be corrected as annotated by the ClinVar database. The distribution of editing motifs for all G>A mutations in ClinVar is shown versus the editing efficiency by REPAIRv1 per motif as quantified on the Gluc transcript. G) The

distribution of editing motifs for all G>A mutations in ClinVar is shown versus the editing efficiency by REPAIRv1 per motif as quantified on the *Gluc* transcript. Values represent mean +/– S.E.M.

[0098] **FIG. 53**: Characterizing specificity of REPAIRv1 A) Schematic of *KRAS* target site and guide design. (SEQ ID Nos. 713-720) B) Quantification of percent editing for tiled *KRAS*-targeting guides. Editing percentages are shown at the on-target and neighboring adenosine sites. For each guide, the region of duplex RNA is indicated by a red rectangle. Values represent mean +/– S.E.M. C) Transcriptome-wide sites of significant RNA editing by REPAIRv1 with Cluc targeting guide. The on-target site Cluc site (254 A>G) is highlighted in orange. D) Transcriptome-wide sites of significant RNA editing by REPAIRv1 (150ng REPAIR vector transfected) with non-targeting guide. Non-targeting guide is the same as in Fig. 50C.

[0099] **FIG. 54**: Rational mutagenesis of ADAR2 to improve the specificity of REPAIRv1 A) Quantification of luciferase signal restoration by various dCas13-ADAR2 mutants as well as their specificity score plotted along a schematic for the contacts between key ADAR2 deaminase residues and the dsRNA target. All deaminase mutations were made on the dCas13-$ADAR2_{DD}$(E488Q) background. The specificity score is defined as the ratio of the luciferase signal between targeting guide and non-targeting guide conditions. Schematic of ADAR2 deaminase domain contacts with dsRNA is adapted from ref (*20*) B) Quantification of luciferase signal restoration by various dCas13-ADAR2 mutants versus their specificity score. Non-targeting guide is the same as in Fig. 50C. C) Measurement of the on-target editing fraction as well as the number of significant off-targets for each dCas13-ADAR2 mutant by transcriptome wide sequencing of mRNAs. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig. 50C. D) Transcriptome-wide sites of significant RNA editing by REPAIRv1 and REPAIRv2 with a guide targeting a pretermination site in Cluc. The on-target Cluc site (254 A>G) is highlighted in orange. 10 ng of REPAIR vector was transfected for each condition. E) RNA sequencing reads surrounding the on-target Cluc editing site (SEQ ID No. 721) (254 A>G) highlighting the differences in off-target editing between REPAIRv1 and REPAIRv2. All A>G edits are highlighted in red while sequencing errors are highlighted in blue. Gaps reflect spaces between aligned reads. Non-targeting guide is the same as in Fig. 50C. F) RNA editing by REPAIRv1 and REPAIRv2 with guides targeting an out-of-frame UAG site in the endogenous *KRAS* and *PPIB* transcripts. The on-target editing fraction is shown as a sideways bar chart on the right for each condition row. The duplex region formed by the guide RNA is shown by a red outline box. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig. 50C.

[00100] **FIG. 55**: Bacterial screening of Cas13b orthologs for *in vivo* efficiency and PFS determination. A) Schematic of bacterial assay for determining the PFS of Cas13b orthologs. Cas13b orthologs with beta-lactamase targeting spacers (SEQ ID No. 722) are co-transformed with beta-lactamase expression plasmids containing randomized PFS sequences and subjected to double selection. PFS sequences that are depleted during co-transformation with Cas13b suggest targeting activity and are used to infer PFS preferences. B) Quantitation of interference activity of Cas13b orthologs targeting beta-lactamase as measured by colony forming units (cfu). Values represent mean +/– S.D. C) PFS logos for Cas13b orthologs as determined by depleted sequences from the bacterial assay. PFS preferences are derived from sequences depleted in the Cas13b condition relative to empty vector controls. Depletion values used to calculate PFS weblogos are listed in table 7.

[0100] **FIG. 56**: Optimization of Cas13b knockdown and further characterization of mismatch specificity. A) Gluc knockdown with two different guides is measured using the top 2 Cas13a and top 4 Cas13b orthologs fused to a variety of nuclear localization and nuclear export tags. B) Knockdown of *KRAS* is measured for LwaCas13a, RanCas13b, PguCas13b, and PspCas13b with four different guides and compared to four position-matched shRNA controls. Non-targeting guide is the same as in Figure 49B. shRNA non-targeting guide sequence is listed in table 11. C) Schematic of the single and double mismatch plasmid libraries used for evaluating the specificity of LwaCas13a and PspCas13b knockdown. Every possible single and double mismatch is present in the target sequence as well as in 3 positions directly flanking the 5' and 3' ends of the target site. (SEQ ID Nos. 723-734) D) The depletion level of transcripts with the indicated single mismatches are plotted as a heatmap for both the LwaCas13a and PspCas13b conditions. (SEQ ID Nos. 723 and 736) The wildtype base is outlined by a green box. E) The depletion level of transcripts with the indicated double mismatches are plotted as a heatmap for both the LwaCas13a and PspCas13b conditions (SEQ ID Nos. 723 and 736). Each box represents the average of all possible double mismatches for the indicated position.

[0101] **FIG. 57**: Characterization of design parameters for dCas13-ADAR2 RNA editing A) Knockdown efficiency of Gluc targeting for wildtype Cas13b and catalytically inactive H133A/H1058A Cas13b (dCas13b). B) Quantification of luciferase activity restoration by dCas13b fused to either the wildtype ADAR2 catalytic domain or the hyperactive E488Q mutant ADAR2 catalytic catalytic domain, tested with tiling Cluc targeting guides. C) Guide design and sequencing quantification of A->I editing for 30 nt guides targeting *Cypridinia* luciferase W85X D) Guide design and sequencing quantification of A->I editing for 50 nt

guides targeting *PPIB*. E) Influence of linker choice on luciferase activity restoration by REPAIRv1. F) Influence of base identify opposite the targeted adenosine on luciferase activity restoration by REPAIRv1 (SEQ ID Nos. 754 and 755). Values represent mean +/– S.E.M.

[0102] **FIG. 58**: ClinVar motif distribution for G>A mutations. The number of each possible triplet motif observed in the ClinVar database for all G>A mutations.

[0103] **FIG. 59**: Truncations of dCas13b still have functional RNA editing. Various N-terminal and C-terminal truncations of dCas13b allow for RNA editing as measured by restoration of luciferase signal for the *Cluc* W85X reporter. Values represent mean +/– S.E.M. The construct length refers to the coding sequence of the REPAIR constructs..

[0104] **FIG. 60**: Comparison of other programmable ADAR systems with the dCas13-ADAR2 editor. A) Schematic of two programmable ADAR schemes: BoxB-based targeting and full length ADAR2 targeting. In the BoxB scheme (top), the ADAR2 deaminase domain ($ADAR2_{DD}$(E488Q)) is fused to a small bacterial virus protein called lambda N (ƛN), which binds specifically a small RNA sequence called BoxB-ƛ, and the fusion protein is recruited to target adenosines by a guide RNA containing homology to the target site and hairpins that BoxB-ƛ binds to. Full length ADAR2 targeting utilizes a guide RNA with homology to the target site and a motif recognized by the double strand RNA binding domains of ADAR2.. A guide RNA containing two BoxB-ƛ hairpins can then guide the $ADAR2_{DD}$(E488Q), -ƛN for site specific editing. In the full length ADAR2 scheme (bottom), the dsRNA binding domains of ADAR2 bind a hairpin in the guide RNA, allowing for programmable ADAR2 editing (SEQ ID Nos. 756-760). B) Transcriptome-wide sites of significant RNA editing by BoxB-$ADAR2_{DD}$(E488Q) with a guide targeting Cluc and a non-targeting guide. The on-target Cluc site (254 A>G) is highlighted in orange. C) Transcriptome-wide sites of significant RNA editing by ADAR2 with a guide targeting Cluc and a non-targeting guide. The on-target Cluc site (254 A>G) is highlighted in orange. D) Transcriptome-wide sites of significant RNA editing by REPAIRv1 with a guide targeting Cluc and a non-targeting guide. The on-target Cluc site (254 A>G) is highlighted in orange. The non-targeting guide is the same as in Fig50C. E) Quantitation of on-target editing rate percentage for BoxB-$ADAR2_{DD}$(E488Q), ADAR2, and REPAIRv1 for targeting guides against Cluc. F) Overlap of off-target sites between different targeting and non-targeting conditions for programmable ADAR systems. The values plotted are the percent of the maximum possible intersection of the two off-target data sets.

[0105] **FIG. 61**: Efficiency and specificity of dCas13b-ADAR2 mutants A) Quantitation of luciferase activity restoration by dCas13b-$ADAR2_{DD}$(E488Q) mutants for Cluc-targeting and non-targeting guides. Non-targeting guide is the same as in Fig50C. B) Relationship

between the ratio of targeting and non-targeting guides and the number of RNA-editing off-targets as quantified by transcriptome-wide sequencing C) Quantification of number of transcriptome-wide off-target RNA editing sites versus on-target Cluc editing efficiency for dCas13b-ADAR2 $_{DD}$(E488Q) mutants.

[0106] **FIG. 62**: Transcriptome-wide specificity of RNA editing by dCas13b-ADAR2 $_{DD}$(E488Q) mutants A) Transcriptome-wide sites of significant RNA editing by dCas13b-ADAR2 $_{DD}$(E488Q) mutants with a guide targeting Cluc. The on-target Cluc site (254 A>G) is highlighted in orange. B) Transcriptome-wide sites of significant RNA editing by dCas13b-ADAR2 $_{DD}$(E488Q) mutants with a non-targeting guide.

[0107] **FIG. 63**: Characterization of motif biases in the off-targets of dCas13b-ADAR2 $_{DD}$(E488Q) editing. A) For each dCas13b-ADAR2 $_{DD}$(E488Q) mutant, the motif present across all A>G off-target edits in the transcriptome is shown. B) The distribution of off-target A>I edits per motif identity is shown for REPAIRv1 with targeting and non-targeting guide. C) The distribution of off-target A>I edits per motif identity is shown for REPAIRv2 with targeting and non-targeting guide.

[0108] **FIG. 64**: Further characterization of REPAIRv1 and REPAIRv2 off-targets. A) Histogram of the number of off-targets per transcript for REPAIRv1. B) Histogram of the number of off-targets per transcript for REPAIRv2.C) Variant effect prediction of REPAIRv1 off targets. D) Distribution of REPAIRv1 off targets in cancer-related genes. TSG, tumor suppressor gene.. E) Variant effect prediction of REPAIRv2 off targets. F) Distribution of REPAIRv2 off targets in cancer-related genes.

[0109] **FIG. 65**: RNA editing efficiency and specificity of REPAIRv1 and REPAIRv2. A) Quantification of percent editing of *KRAS* with *KRAS*-targeting guide 1 at the targeted adenosine and neighboring sites for REPAIRv1 and REPAIRv2. For each guide, the region of duplex RNA is outlined in red. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig. 50C. B) Quantification of percent editing of *KRAS* with *KRAS*-targeting guide 3 at the targeted adenosine and neighboring sites for REPAIRv1 and REPAIRv2. Non-targeting guide is the same as in Fig. 50C. C) Quantification of percent editing of *PPIB* with *PPIB*-targeting guide 2 at the targeted adenosine and neighboring sites for REPAIRv1 and REPAIRv2. Non-targeting guide is the same as in Fig. 50C.

[0110] **FIG. 66:** Demonstration of all potential codon changes with a A>I RNA editor. A) Table of all potential codon transitions enabled by A>I editing. B) A codon table demonstrating all the potential codon transitions enabled by A>I editing. Adapted and modified based on J. D. Watson, Molecular biology of the gene. (Pearson, Boston, ed. Seventh edition,

2014), pp. xxxiv, 872 pages.(38). C) Model of REPAIR A to I editing of a precisely encoded nucleotide via a mismatch in the guide sequence. The A to I transition is mediated by the catalytic activity of the ADAR2 deaminase domain and will be read as a guanosine by translational machinery. The base change does not rely on endogenous repair machinery and is permanent for as long as the RNA molecule exists in the cell. D) REPAIR can be used for correction of Mendelian disease mutations. E) REPAIR can be used for multiplexed A to I editing of multiple variants for engineering pathways or modifying disease. Multiplexed guide delivery can be achieved by delivering a single CRISPR array expression cassette since the Cas13b enzyme processes its own array. F) REPAIR can be used for modifying protein function through amino acid changes that affect enzyme domains, such as kinases. G) REPAIR can modulate splicing of transcripts by modifying the splice acceptor site.

[0111] **FIG. 67**: Additional truncations of Psp dCas13b.

[0112] **FIG. 68**: Potential effect of dosage on off target activity.

[0113] **FIG. 69**: Relative expression of Cas13 orthologs in mammalian cells and correlation of expression with interference activity. A) Expression of Cas13 orthologs as measured by msfGFP fluoresence. Cas13 orthologs C-terminally tagged with msfGFP were transfected into HEK293FT cells and their fluorescence measured 48 hours post transfection. B) Correlation of Cas13 expression to interference activity. The average RLU of two Gluc targeting guides for Cas13 orthologs, separated by subfamily, is plotted versus expression as determined by msfGFP fluoresence. The RLU for targeting guides are normalized to RLU for a non-targeting guide, whose value is set to 1. The non-targeting guide is the same as in Figure 49B for Cas13b.

[0114] **FIG. 70**: Comparison of RNA editing activity of dCas13b and REPAIRv1. A) Schematic of guides used to target the W85X mutation in the Cluc reporter (SEQ ID Nos. 911-917) B) Sequencing quantification of A to I editing for indicated guides transfected with dCas13b. For each guide, the region of duplex RNA is outlined in red. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig50C. C) Sequencing quantification of A to I editing for indicated guides transfected with REPAIRv1. For each guide, the region of duplex RNA is outlined in red. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig50C. D) Comparison of on-target A to I editing rates for dCas13b and dCas13b-ADAR2DD(E488Q) for guides tested in panel B and C. E) Influence of base identify opposite the targeted adenosine on luciferase activity restoration by REPAIRv1. Values represent mean +/– S.E.M. (SEQ ID Nos. 754 and 755)

[0115] **FIG. 71**: REPAIRv1 editing activity evaluated without a guide and in comparison to ADAR2 deaminase domain alone. A) Quantification of A to I editing of the Cluc W85X mutation by REPAIRv1 with and without guide as well as the ADAR2 deaminase domain only without guide. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig50C. B) Number of differentially expressed genes in the REPAIRv1 and ADAR2DD conditions from panel A. C) The number of significant off-targets from the REPAIRv1 and ADAR2DD conditions from panel A. D) Overlap of off-target A to I editing events between the REPAIRv1 and ADAR2DD conditions from panel A. The values plotted are the percent of the maximum possible intersection of the two off-target data sets.

[0116] **FIG. 72**: Evaluation of off-target sequence similarity to the guide sequence. A) Distribution of the number of mismatches (hamming distance) between the targeting guide sequence and the off-target editing sites for REPAIRv1 with a Cluc targeting guide. B) Distribution of the number of mismatches (hamming distance) between the targeting guide sequence and the off-target editing sites for REPAIRv2 with a Cluc targeting guide.

[0117] **FIG. 73**: Comparison of REPAIRv1, REPAIRv2, ADAR2 RNA targeting, and BoxB RNA targeting at two different doses of vector (150ng and 10ng effector). A) Quantification of RNA editing activity at the Cluc W85X (254 A>I) on-target editing site by REPAIRv1, REPAIRv2, ADAR2 RNA targeting, and BoxB RNA targeting approaches. Each of the four methods were tested with a targeting or non-targeting guide. Values shown are the mean of the three replicates. B) Quantification of RNA editing off-targets by REPAIRv1, REPAIRv2, ADAR2 RNA targeting, and BoxB RNA targeting approaches. Each of the four methods were tested with a targeting guide for the Cluc W85X (254 A>I) site or non-targeting guide. For REPAIR constructs, non-targeting guide is the same as in Fig. 50C.

[0118] **FIG. 74**: RNA editing efficiency and genome-wide specificity of REPAIRv1 and REPAIRv2. A) Quantification of RNA editing activity at the PPIB guide 1 on-target editing site by REPAIRv1, REPAIRv2 with targeting and non-targeting guides. Values represent mean +/– S.E.M. B) Quantification of RNA editing activity at the PPIB guide 2 on-target editing site by REPAIRv1, REPAIRv2 with targeting and non-targeting guides. Values represent mean +/– S.E.M. C) Quantification of RNA editing off-targets by REPAIRv1 or REPAIRv2 with PPIB guide 1, PPIB guide 2, or non-targeting guide. D) Overlap of off-targets between REPAIRv1 for PPIB targeting, Cluc targeting, and non-targeting guides. The values plotted are the percent of the maximum possible intersection of the two off-target data sets.

[0119] **FIG. 75**: High coverage sequencing of REPAIRv1 and REPAIRv2 off-targets. A) Quantitation of off-target edits for REPAIRv1 and REPAIRv2 as a function of read depth with

a total of 5 million reads (12.5x coverage), 15 million reads (37.5x coverage) and 50 million reads (125x coverage) per condition. B) Overlap of off-target sites at different read depths of the following conditions: REPAIRv1 versus REPAIRv1 (left), REPAIRv2 versus REPAIRv2 (middle), and REPAIRv1 versus REPAIRv2 (right). The values plotted are the percent of the maximum possible intersection of the two off-target data sets. C) Editing rate of off-target sites compared to the coverage (log2(number of reads)) of the off-target for REPAIRv1 and REPAIRv2 targeting conditions at different read depths. D) Editing rate of off-target sites compared to the log2(TPM+1) of the off-target gene expression for REPAIRv1 and REPAIRv2 targeting conditions at different read depths.

[0120] **FIG. 76**: Quantification of REPAIRv2 activity and off-targets in the U2OS cell line. A) Transcriptome-wide sites of significant RNA editing by REPAIRv2 with a guide targeting Cluc in the U2OS cell line. The on-target Cluc site (254 A>I) is highlighted in orange. B) Transcriptome-wide sites of significant RNA editing by REPAIRv2 with a non-targeting guide in the U2OS cell line. C) The on-target editing rate at the Cluc W85X (254 A>I) by REPAIRv2 with a targeting guide or non-targeting guide in the U2OS cell line. D) Quantification of off-targets by REPAIRv2 with a guide targeting Cluc or non-targeting guide in the U2OS cell line.

[0121] **FIG. 77:** Identifying additional ADAR mutants with increased efficiency and specificity. Cas13b-ADAR fusions with mutations in the ADAR deaminase domain, assayed on the luciferase target. Lower non-targeting RLU is indicative of more specificity.

[0122] **FIG. 78:** Identifying additional ADAR mutants with increased efficiency and specificity. Mutants were chosen from flow cytometry data for low, medium, and high-disrupting mutantions.

[0123] **FIG. 79:** Identifying additional ADAR mutants with increased efficiency and specificity.

[0124] **FIG. 80:** Identifying additional ADAR mutants with increased efficiency and specificity.

[0125] **FIG. 81:** Identifying additional ADAR mutants with increased efficiency and specificity through saturating mutagenesis on V351.

[0126] **FIG. 82:** Identifying additional ADAR mutants with increased efficiency and specificity through saturating mutagenesis on T375.

[0127] **FIG. 83:** Identifying additional ADAR mutants with increased efficiency and specificity through saturating mutagenesis on R455.

[0128] **FIG. 84:** Identifying additional ADAR mutants with increased efficiency and specificity through saturating mutagenesis.

[0129] **FIG. 85:** 3' binding loop residue saturation mutagenesis.

[0130] **FIG. 86:** Select ADAR mutants with increased efficiency and specificity. Screening has identified multiple mutants with increased specificity compared to REPAIRv1 and increased activity compared to REPAIRv1 and REPAIRv2.

[0131] **FIG. 87:** Second round saturating mutagenesis performed on promising residues with additional E488 mutations.

[0132] **FIG. 88:** Second round saturating mutagenesis performed on promising residues with additional E488 mutations.

[0133] **FIG. 89:** Combinations of ADAR mutants identified through screening.

[0134] **FIG. 90:** Combinations of ADAR mutants identified through screening.

[0135] **FIG. 91:** Testing most promising mutants by NGS.

[0136] **FIG. 92:** Testing most promising mutants by NGS.

[0137] **FIG. 93:** Testing most promising mutants by NGS.

[0138] **FIG. 94:** Testing most promising mutants by NGS.

[0139] **FIG. 95:** Finding most promising base flip for C-U activity on existing constructs.

[0140] **FIG. 96:** Testing ADAR mutants with best guide for C->U activity.

[0141] **FIG. 97:** Validation of V351 mutants for C>U activity.

[0142] **FIG. 98:** Testing Cas13b-cytidine deaminase fusions with testing panning guides across construct:

[0143] **FIG. 99:** Testing Cas13b-cytidine deaminase fusions with testing panning guides across construct.

[0144] **FIG. 100** is a graph depicting that Cas13b orthologs fused to ADAR exhibit variable protein recovery and off-target effects. 15 dCas13b orthologs were fused to ADAR and targeted to edit a Cypridina luciferase reporter with an introduced pretermination site that, when corrected, restores luciferase function. A nontargeting guide was additionally used to evaluate off target effects. REPAIRv1 and REPAIRv2 are as published in Cox et al. (2017). Different orthologs fused to ADAR exhibit different ability to recover functional luciferase, as well as different off-target effects. In particular, Cas12b6 (Riemerella anatipestifer (RanCas13b)) appears to have a better ability to recover functional luciferase as well as fewer off-target events than REPAIRv1. Points marked in red were selected for further engineering and analysis as these were the two orthologs that exhibited the highest functional protein recovery other than Cas13b12 (REPAIRv1).

[0145] **FIG. 101** is a graph showing targeted sequencing of editing locus for all orthologs. Targeted next generation sequencing of the editing locus shows that most Cas13b orthologs fused to ADAR mediate bona fide editing events at the target adenosine. Orthologs are ordered from lowest to highest editing percentage from top to bottom. In particular, although Cas13b6 is observed to exhibit higher functional luciferase recovery (FIG. 100), REPAIRv1 still shows a higher percentage of editing events at the target adenosine. Additionally, different orthologs show different percentages of off target edits at other adenosines within the sequencing window, and, in particular, Cas13b6 shows much lower editing at A33 both in the targeting and non-targeting condition than REPAIRv1, which is consistent with the lower off-target signal observed in the luciferase assay (FIG. 100). The ratio between on target and off-target editing is not consistent between orthologs, and in particular, Cas13b6 seems to maximize the amount of on-target edits per off-target edit.

[0146] **FIG. 102** is a schematic illustrating design constraints for delivery with Adeno-associated virus (AAV). AAV, a clinically relevant viral delivery vector, has a packaging limit of about 4.7 kilobases for efficient packaging and titering of the virus. However, REPAIR is much larger than this when the promoter is included. Additionally, it would be ideal to deliver the entire system (REPAIR fusion protein + guide RNA) in a single vector for ease of production and delivery. Therefore Cas13b orthologs are chosen to be truncated down

[0147] **FIG. 103A** is a graph showing results of truncating N-terminus of Cas13b6. Each ortholog was truncated down in 20 amino acid (60 base pair) intervals up to 300 amino acids (900 base pairs) from each of the N and C termini of the protein. RNA editing activity was then evaluated via the luciferase correction assay previously described. Luciferase recovery in the targeting guideRNA condition is shown on the y-axis, versus the size in amino acids of the truncated Cas13b ortholog on the x-axis. Truncating at different points changes the ability of the REPAIR fusion to recover luciferase function - some are better and some are worse than the full length Cas13b protein, and different patterns are observed with different orthologs. **FIG. 103B** is a graph showing results of truncating C-terminus of Cas13b6. For Cas13b6, the CΔ300 truncation was chosen as having the best activity with a sufficiently small size.

[0148] **FIG. 104A** is a graph showing results of truncating N-terminus of Cas13b11. **FIG. 104B** is a graph showing results of truncating C-terminus of Cas13b11. For Cas13b11, the NΔ280 truncation was chosen as having the best activity with a sufficiently small size.

[0149] **FIG. 105A** is a graph showing results of truncating N-terminus of Cas13b12. **FIG. 104B** is a graph showing results of truncating C-terminus of Cas13b12. For Cas13b12, the CΔ300 truncation was chosen as having the best activity with a sufficiently small size.

[0150] **FIG. 106** is a graph showing tiling guide RNAs across a single editing site. Editing is targeted to an adenosine in an introduced premature stop codon in a luciferase reporter, which, if corrected, will restore the amino acid at this position to a tryptophan and thus restore function of the luciferase. Guide RNAs with both 50 and 30 nucleotide spacers are tiled across this editing site such that the target adenosine is at a different position within the guide RNA. Each of these guides were evaluated with both the full length and best truncations previously noted on the preceding three slides. (SEQ ID Nos. 700 and 701)

[0151] **FIG. 107** is a graph showing Cas13b6 results with different guide RNAs. The results show that target adenosine position within the spacer sequence does have an effect on editing. Interestingly, both the full length and truncated Cas13b exhibit very similar patterns of which position within the guide is optimal, but different orthologs exhibit slightly different patterns, though still relatively similar (FIGs. 108 and 109). In general, 50 bp guides seem to be slightly better for A to I editing. shown here, B11 and B12 (REPAIRv1) on the following two slides.

[0152] **FIG. 108** is a graph showing Cas13b11 results with different guide RNAs.

[0153] **FIG. 109** is a graph showing Cas13b12 (REPAIRv1) with different guide RNAs.

[0154] **FIG. 110** is a graph showing results of Cas13b6-REPAIR targeting KRAS. In this figure, instead of moving the guide across a single editing site, the sequence of the guide is fixed and each guide RNA targets a different adenosine within the fixed sequence. Two sites were evaluated for both Cas13b6 and the Cas13b6CΔ300 truncation, with both 30 and 50 nucleotide guides as indicated in the schematic at the top (SEQ ID No. 918). Editing is evaluated by targeted next generation sequencing across the editing loci. Again, different target positions within the guide show different editing rates and patterns for both the full length and truncated Cas13b6s.

[0155] **FIG. 111** is a graph depicting that localization tags may affect on-target editing. Different localization tags (both nuclear localization and nuclear export tags) with Cas13b6 seem to affect the ability of Cas13b6-REPAIR to recover luciferase activity, but does not appear to affect off-target activity appreciably. Red points are REPAIRv1 and REPAIRv2, which are with the Cas13b12 ortholog and using the HIV NES, blue points with Cas13b6 ortholog.

[0156] **FIG. 112** is a graph showing results of RfxCas13d. Cas13d is a recently discovered class of Cas13 proteins that are on average smaller than Cas13b proteins. A characterized Cas13d ortholog known as RfxCas13d is tested in this figure for REPAIR activity using the same tiling guide scheme shown in Fig. 106. crRNA refers to mature CRISPR RNA and pre-

crRNA refers to unprocessed version. Although most guide RNAs with RfxCas13d-REPAIR show no RNA editing activity, there are a few that seem to mediate relatively good editing when compared to existing systems shown in black.

[0157] **FIG. 113** is a graph showing results of guide RNA-mediated editing with RfxCas13d. The data show that even without the RfxCas13d-REPAIR or even ADAR, the guide RNA (mismatch position 33) by itself is somehow able to mediate editing events (left-most condition), which is not the case with a Cas13b12 guide. Furthermore, it appears that the introduction of ADAR or RfxCas13d-REPAIR does not seem to have much effect on the editing mediated by this guide RNA.

[0158] **FIG. 114** is a schematic illustrating the dual vector system design for evaluating RNA editing in cultures of primary rat cortical neurons.

[0159] **FIG. 115** is a graph showing that up to 35% editing is achieved in neurons with dual vector system. Using two guides as indicated in the schematic at the top (SEQ ID No. 761, guide 1 has one base flip/targeted adenosine at the indicated position, while guide 2 has two targeted adenosine), REPAIR with B6/B11/B12 was packaged into AAV using the dual vector system in FIG. 114. Guide 2 was found to mediate up to 35% editing at A57 with B6-REPAIR (~30% for B11-REPAIR) with targeted next generation sequencing 14 days after transduction with AAV, showing that AAV-delivered REPAIR can mediate RNA base editing in post-mitotic cell types.

[0160] **FIG. 116** is a graph depicting that single vector AAV B6-REPAIR system is able to edit RNA in neuron cultures. Using the single vector system in FIG. 102 with the Cas13b6CΔ300 truncation, the guide that has two target adenosines in FIG. 115 was used, as well as a guide across the same sequence but only targeting A48 as indicated. 5 days after transduction with AAV, targeted next-generation sequencing shows approximately 6% editing with guide 2 at A24 (Same as A57 in FIG. 115), demonstrating the viability of the single vector approach.

[0161] **FIG. 117** is a graph is a graph depicting that different Cas13b orthologs fused to ADAR.

[0162] **FIG. 118** is a graph showing that V351G editing greatly increases REPAIR editing. The V351G mutation (pAB316) was introduced into the E488Q PspCas13b (Cas13b12) REPAIR construct (REPAIR v1, pAB0048) and tested for C-U activity on a gauss luciferase construct with a TCG motif (TCG). Editing was read out by next generation sequencing, revealing increased C-U activity.

[0163] **FIG. 119** is a graph showing endogenous KRAS and PPIB targeting. The V351G mutation (pAB316) was introduced into the E488Q PspCas13b REPAIR construct (REPAIR v1, pAB0048) and tested for C-U activity on a gauss four sites, two in each gene, with different motifs. Editing was read out by next generation sequencing, revealing increased C-U activity.

[0164] **FIG. 120** is a graph showing optimal V351G combination mutants. Selected sites (S486, G489) were mutagenized to all 20 possible residues and tested on a background of REPAIR[E488Q, V351G]. Constructs were tested on two luciferase motifs, TCG and GCG, and selected on the basis of luciferase activity.

[0165] **FIG. 121** is a graph showing S486A and V351G combination C-to-U activity. S486A was tested against the [V351G, E488Q] background and the E488Q background on all four motifs, with luciferase activity as a readout. S486A performs better on all motifs, especially ACG and TCG.

[0166] **FIG. 122** is a graph showing that S486A improves C-to-U editing across all motifs. S486A improves targeting over the [V351G, E488Q] background on all motifs, when measured by luciferase activity.

[0167] **FIG. 123A** is a graph showing S486 mutants C-to-U activity with both TCG and CCG targeting. **FIG. 123B** is a graph showing S486 mutants C-to-U activity with CCG targeting only. S486A was tested against the [V351G, E488Q] background and the E488Q background on all four motifs, with NGS as a readout. S486A performs better on all motifs, especially ACG and TCG.

[0168] **FIG. 124** is a graph showing S486A A-to-I activity. The data shows that S486A mutations maintain A-to-I activity of the previous constructs when measured on a luciferase reporter.

[0169] **FIG. 125** is a graph showing S486A A-to-I off-target activity. The data shows that S486A has comparable A-to-I off-target activity when measured on a luciferase reporter.

[0170] **FIG. 126A** is a graph showing that targeting by S486A/V351G/E488Q (pAB493), V351G/E488Q (pAB316), and E488Q (REPAIRv1) is comparable when read out by luciferase activity (Gluc/Cluc RLU). **FIG. 126B** is a graph showing that targeting by S486A/V351G/E488Q (pAB493), V351G/E488Q (pAB316), and E488Q (REPAIRv1) is comparable when assayed by NGS (fraction editing).

[0171] **FIG. 127A** is a graph showing S486A C-to-U activity by NGS on Cluc reporter constructs. **FIG. 127B** is a graph showing S486A C-to-U activity by NGS on endogenous gene PPIB.

[0172] **FIG. 128** is a graph depicting identification of new T375 and K376 mutants. Selected sites (T375, K376) were mutagenized to all 20 possible residues and tested on a background of REPAIR[E488Q, V351G]. Constructs were tested on the TCG luciferase motif and selected on the basis of luciferase activity.

[0173] **FIG. 129** is a graph showing that T375S has relaxed motif. T375S was tested against the [S486A,V351G, E488Q] background (pAB493), [V351G, E488Q] background (pAB316), and the E488Q background (pAB48) on all TCG and GCG motifs, with luciferase activity as a readout. T375S improves GCG motif.

[0174] **FIG. 130** is a graph showing that T375S has relaxed motif. T375S was tested against the [S486A,V351G, E488Q] background (pAB493), [V351G, E488Q] background (pAB316), and the E488Q background (pAB48) on GCG motifs, with luciferase activity as a readout. T375S improves GCG motif.

[0175] **FIG. 131** is a graph depicting that B6 and B11 orthologs show improved RESCUE activity. Cas13b orthologs Cas13b6 (RanCas13b) and Cas13b11 (PguCas13b) were tested with T375S mutation, and show improved activity as measured by luciferase assay. Mutations shows are on corresponding backgrounds (T375S = T375S/S486A/V351G/E448Q).

[0176] **FIG. 132** is a graph showing that DNA2.0 vectors has comparable luciferase to transient transfection vectors. RESCUE vectors based off of either DNA2.0 (now Atum) constructs compared to a non-lenti vector, with Cas13b11 (PguCas13b) show improved luciferase activity. The Atum vector map (https://benchling.com/s/seq-DENgx9izDhsRTFFgy71K) has additional EES elements for expression. Mutations shows are on corresponding backgrounds (V351G = V351G/E448Q, S486A = S486A/V351G/E448Q).

[0177] **FIG. 133A** is a graph showing luciferase results of testing truncations validated by REPAIR (B6 Cdelta300) with RESCUE using 30bp guides. **FIG. 133B** is a graph showing luciferase results of testing truncations validated by REPAIR (B6 Cdelta300) with RESCUE using 50bp guides. The 26 mismatch distance (as measured by the 5' end) shows the optimal activity with both full length and truncated versions).

[0178] **FIG. 134A** is a graph showing luciferase results of testing truncations validated by REPAIR (B11 Ndelta280) with RESCUE using 30bp guides. **FIG. 134B** is a graph showing luciferase results of testing truncations validated by REPAIR (B11 Ndelta280) with RESCUE using 50bp guides. The 26 mismatch distance (as measured by the 5' end) shows the optimal activity with both full length and truncated versions).

[0179] **FIG. 135** is a graph showing results of testing all B6 truncations. Iterative truncations were generated from the N and C termini on RanCas13b (B6), with the

T375S/S486A/V351G/E448Q mutation, with optimal activity up to C-delta 200, and activity at C-delta 320. Truncations are tested on luciferase, and editing is read out as luciferase activity. Missing bars indicate no data. The pAB0642 is an untruncated N-term control, T375S/S486A/V351G/E448Q. The pAB0440 is an untruncated C-term control, E448Q. All N-term constructs, and pAB0642, have an mark NES linker. All C-term constrcuts, and pAB0440, have a HIV-NES linker.

[0180] **FIG. 136** is a graph showing results of testing all B11 truncations. Iterative truncations were generated from the N and C termini on PguCas13b (B11), with the T375S/S486A/V351G/E448Q mutation. Truncations are tested on luciferase, and editing is read out as luciferase activity.

[0181] **FIG. 137A** is a graph showing Beta catenin modulation with REPAIR/RESCUE as measured by Beta-catenin activity via the TCF-LEF RE Wnt pathway reporter (Promega). **FIG. 137B** is a graph showing Beta catenin modulation with REPAIR/RESCUE as measured by the M50 Super 8x TOPFlash reporter (Addgene). Beta-catenin/Wnt pathway induction is tested by using RNA editing to remove phosphorylation sites on Beta catenin. Guides targeting beta-catenin for either REPAIR (RanCas13b ortholog, E488Q mutation) or RESCUE (RanCas13b ortholog, T375S/S486A/V351G/E448Q mutation) were tested for phenotypic activity. The T41A guide shows activity on both reporters.

[0182] **FIG. 138** is a graph showing NGS results of Beta catenin modulation. NGS readouts of either A-I (A) or C-U (C) activity at targeted sites by either REPAIR (RanCas13b ortholog, E488Q mutation) or RESCUE (RanCas13b ortholog, T375S/S486A/V351G/E448Q mutation. REPAIR was used on A targets, and RESCUE was used on C targets.

[0183] **FIG. 139** is a graph depicting that tiling different guides shows improved motif activity at the 30_5 mutation (mismatch is 26 nt away from the 5' of the guide). All four motifs were tested with various tiling guides for luciferase activity. Nomenclature corresponds to distance from the 3' end of the spacer (i.e., 26 nt mismatch is 30_5). The 26 mismatch distance (as measured by the 5' end) shows the optimal activity with most motifs. Guides were tested with RESCUE (RanCas13b ortholog, T375S/S486A/V351G/E448Q mutation.

[0184] **FIG. 140A** is a graph showing that REPAIR allows for editing residues associated with PTMs. **FIG. 140B** is a graph showing that RESCUE allows for editing residues associated with PTMs.

[0185] The appended claims are herein explicitly incorporated by reference.

[0186] The figures herein are for illustrative purposes only and are not necessarily drawn to scale.

DETAILED DESCRIPTION

General Definitions

[0187] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure pertains. Definitions of common terms and techniques in molecular biology may be found in Molecular Cloning: A Laboratory Manual, 2nd edition (1989) (Sambrook, Fritsch, and Maniatis); Molecular Cloning: A Laboratory Manual, 4th edition (2012) (Green and Sambrook); Current Protocols in Molecular Biology (1987) (F.M. Ausubel et al. eds.); the series Methods in Enzymology (Academic Press, Inc.): PCR 2: A Practical Approach (1995) (M.J. MacPherson, B.D. Hames, and G.R. Taylor eds.): Antibodies, A Laboraotry Manual (1988) (Harlow and Lane, eds.): Antibodies A Laboraotry Manual, 2nd edition 2013 (E.A. Greenfield ed.); Animal Cell Culture (1987) (R.I. Freshney, ed.); Benjamin Lewin, Genes IX, published by Jones and Bartlet, 2008 (ISBN 0763752223); Kendrew *et al.* (eds.), The Encyclopedia of Molecular Biology, published by Blackwell Science Ltd., 1994 (ISBN 0632021829); Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, published by VCH Publishers, Inc., 1995 (ISBN 9780471185710); Singleton *et al.*, Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley & Sons (New York, N.Y. 1994), March, Advanced Organic Chemistry Reactions, Mechanisms and Structure 4th ed., John Wiley & Sons (New York, N.Y. 1992); and Marten H. Hofker and Jan van Deursen, Transgenic Mouse Methods and Protocols, 2nd edition (2011).

[0188] Reference is made to US Provisional 62/351,662 and 62/351,803, filed on June 17, 2016, US Provisional 62/376,377, filed on August 17, 2016, US Provisional 62/410,366, filed October 19, 2016, US Provisional 62/432,240, filed December 9, 2016, US provisional 62/471,792 filed March 15, 2017, and US Provisional 62/484,786 filed April 12, 2017. Reference is made to International PCT application PCT/US2017/038154, filed June 19, 2017. Reference is made to US Provisional 62/471,710, filed March 15, 2017 (entitled, "Novel Cas13B Orthologues CRISPR Enzymes and Systems," Attorney Ref: BI-10157 VP 47627.04.2149). Reference is further made to US Provisional 62/432,553, filed December 9, 2016, US Provisional 62/456,645, filed February 8, 2017, and US Provisional 62/471,930, filed March 15, 2017 (entitled "CRISPR Effector System Based Diagnostics," Attorney Ref. BI-10121 BROD 0842P) and US Provisional To Be Assigned, filed April 12, 2017 (entitled "CRISPR Effector System Based Diagnostics," Attorney Ref. BI-10121 BROD 0843P)

[0189] As used herein, the singular forms "a", "an", and "the" include both singular and plural referents unless the context clearly dictates otherwise.

[0190] The term "optional" or "optionally" means that the subsequent described event, circumstance or substituent may or may not occur, and that the description includes instances where the event or circumstance occurs and instances where it does not.

[0191] The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

[0192] The terms "about" or "approximately" as used herein when referring to a measurable value such as a parameter, an amount, a temporal duration, and the like, are meant to encompass variations of and from the specified value, such as variations of +/-10% or less, +/-5% or less, +/-1% or less, and +/-0.1% or less of and from the specified value, insofar such variations are appropriate to perform in the disclosed invention. It is to be understood that the value to which the modifier "about" or "approximately" refers is itself also specifically, and preferably, disclosed.

[0193] Reference throughout this specification to "one embodiment", "an embodiment," "an example embodiment," means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, appearances of the phrases "in one embodiment," "in an embodiment," or "an example embodiment" in various places throughout this specification are not necessarily all referring to the same embodiment, but may. Furthermore, the particular features, structures or characteristics may be combined in any suitable manner, as would be apparent to a person skilled in the art from this disclosure, in one or more embodiments. Furthermore, while some embodiments described herein include some but not other features included in other embodiments, combinations of features of different embodiments are meant to be within the scope of the invention. For example, in the appended claims, any of the claimed embodiments can be used in any combination.

[0194] C2c2 is now known as Cas13a. It will be understood that the term "C2c2" herein is used interchangeably with "Cas13a".

[0195] All publications, published patent documents, and patent applications cited herein are hereby incorporated by reference to the same extent as though each individual publication, published patent document, or patent application was specifically and individually indicated as being incorporated by reference.

[0196] Various embodiments are described hereinafter. It should be noted that the specific embodiments are not intended as an exhaustive description or as a limitation to the broader

2018290843 07 Apr 2025

aspects discussed herein. One aspect described in conjunction with a particular embodiment is not necessarily limited to that embodiment and can be practiced with any other embodiment(s).

OVERVIEW

[0197] The embodiments disclosed herein provide systems, constructs, and methods for targeted base editing. In general the systems disclosed herein comprise a targeting component and a base editing component. The targeting component functions to specifically target the base editing component to a target nucleotide sequence in which one or more nucleotides are to be edited. The base editing component may then catalyze a chemical reaction to convert a first nucleotide in the target sequence to a second nucleotide. For example, the base editor may catalyze conversion of an adenine such that it is read as guanine by a cell's transcription or translation macchinery, or vice versa. Likewise, the base editing component may catalyze conversion of cytidine to a uracil, or vice versa. In certain example embodiments, the base editor may be derived by starting with a known base editor, such as an adenine deaminase or cytodine deaminase, and modified using methods such as directed evolution to derive new functionalities. Directed evolution techniques are known in the art and may include those described in WO 2015/184016 "High-Throughput Assembly of Genetic Permuatations." In will be understood that the present invention in certain aspects equally relates to deaminases per se as described herein and having undergone directed evolution, such as the mutated deaminases described herein elsewhere, as well as polynucleotides encoding such deaminases (including vectors and expression and/or delivery systems), as well as fusions between such mutated deaminases and targeting component, such as polynucleotide binding molecules or systems, as described herein elsewhere.

[0198] In one aspect the present invention provides methods for targeted deamination of adenine or cytosine in RNA or DNA by an adenosine deaminase or modified variant thereof. According to the methods of the invention, the adenosine deaminase (AD) protein is recruited specifically to the nucleic acid to be modified. The term "AD functionalized compositions" refers to the engineered compositions for site directed base editing disclosed herein, comprising a targeting domain complexed to an adenosine deaminase, or catalytic domain thereof.

[0199] In particular embodiments of the methods of the present invention, recruitment of the adenosine deaminase to the target locus is ensured by fusing the adenosine deaminase or catalytic domain thereof to the targeting domain. Methods of generating a fusion protein from two separate proteins are known in the art and typically involve the use of spacers or linkers. The target domain can be fused to the adenosine deaminase protein or catalytic domain thereof on either the N- or C-terminal end thereof.

[0200] The term "linker" as used in reference to a fusion protein refers to a molecule which joins the proteins to form a fusion protein. Generally, such molecules have no specific biological activity other than to join or to preserve some minimum distance or other spatial relationship between the proteins. However, in certain embodiments, the linker may be selected to influence some property of the linker and/or the fusion protein such as the folding, net charge, or hydrophobicity of the linker.

[0201] Suitable linkers for use in the methods of the present invention are well known to those of skill in the art and include, but are not limited to, straight or branched-chain carbon linkers, heterocyclic carbon linkers, or peptide linkers. However, as used herein the linker may also be a covalent bond (carbon-carbon bond or carbon-heteroatom bond). In particular embodiments, the linker is used to separate the targeting domain and the adenosine deaminase by a distance sufficient to ensure that each protein retains its required functional property. Preferred peptide linker sequences adopt a flexible extended conformation and do not exhibit a propensity for developing an ordered secondary structure. In certain embodiments, the linker can be a chemical moiety which can be monomeric, dimeric, multimeric or polymeric. Preferably, the linker comprises amino acids. Typical amino acids in flexible linkers include Gly, Asn and Ser. Accordingly, in particular embodiments, the linker comprises a combination of one or more of Gly, Asn and Ser amino acids. Other near neutral amino acids, such as Thr and Ala, also may be used in the linker sequence. Exemplary linkers are disclosed in Maratea et al. (1985), Gene 40: 39-46; Murphy et al. (1986) Proc. Nat'l. Acad. Sci. USA 83: 8258-62; U.S. Pat. No. 4,935,233; and U.S. Pat. No. 4,751,180. For example, GlySer linkers GGS, GGGS or GSG can be used. GGS, GSG, GGGS or GGGGS linkers can be used in repeats of 3 (such as (GGS)3 (SEQ ID No. 12), (GGGGS)3) or 5, 6, 7, 9 or even 12 (SEQ ID No. 13) or more, to provide suitable lengths. In particular embodiments, linkers such as (GGGGS)3 are preferably used herein. (GGGGS)6 (GGGGS)9 or (GGGGS)12 may preferably be used as alternatives. Other preferred alternatives are (GGGGS)1 (SEQ ID No 14), (GGGGS)2 (SEQ ID No. 15), (GGGGS)4, (GGGGS)5, (GGGGS)7, (GGGGS)8, (GGGGS)10, or (GGGGS)11. In yet a further embodiment, LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR (SEQ ID No:11) is used as a linker. In yet an additional embodiment, the linker is XTEN linker (SEQ ID No. 761). The invention also relates to a method for treating or preventing a disease by the targeted deamination or a disease causing variant using the AD-functionalized compositions. For example, the deamination of an A, may remedy a disease caused by transcripts containing a pathogenic G→A or C→T point mutation. Examples of disease that can be treated or prevented with the present invention include cancer, Meier-Gorlin syndrome, Seckel syndrome

4, Joubert syndrome 5, Leber congenital amaurosis 10; Charcot-Marie-Tooth disease, type 2; Charcot-Marie-Tooth disease, type 2; Usher syndrome, type 2C; Spinocerebellar ataxia 28; Spinocerebellar ataxia 28; Spinocerebellar ataxia 28; Long QT syndrome 2; Sjögren-Larsson syndrome; Hereditary fructosuria; Hereditary fructosuria; Neuroblastoma; Neuroblastoma; Kallmann syndrome 1; Kallmann syndrome 1; Kallmann syndrome 1; Metachromatic leukodystrophy.

[0202] In particular embodiments, the invention thus comprises compositions for use in therapy. This implies that the methods can be performed *in vivo, ex vivo* or *in vitro*. In particular embodiments, the methods are not methods of treatment of the animal or human body or a method for modifying the germ line genetic identity of a human cell. In particular embodiments; when carrying out the method, the target RNA is not comprised within a human or animal cell. In particular embodiments, when the target is a human or animal target, the method is carried out *ex vivo* or *in vitro*.

[0203] The invention also relates to a method for knocking-out or knocking-down an undesirable activity of a gene, wherein the deamination of an A or C at the transcript of the gene results in a loss of function. For example, in one embodiment, the targeted deamination by the AD-functionalized CRISPR system can cause a nonsense mutation resulting in a premature stop codon in an endogenous gene. This may alter the expression of the endogenous gene and can lead to a desirable trait in the edited cell. In another embodiment, the targeted deamination by the AD-functionalized compositions can cause a nonconservative missense mutation resulting in a code for a different amino acid residue in an endogenous gene. This may alter the function of the endogenous gene expressed and can also lead to a desirable trait in the edited cell.

[0204] The invention also relates to a modified cell obtained by the targeted deamination using the AD-functionalized composition, or progeny thereof, wherein the modified cell comprises an I or G in replace of the A, or a T in replace of the C in the target RNA sequence of interest compared to a corresponding cell before the targeted deamination. The modified cell can be a eukaryotic cell, such as an animal cell, a plant cell, an mammalian cell, or a human cell.

[0205] In some embodiments, the modified cell is a therapeutic T cell, such as a T cell sutiable for CAR-T therapies. The modification may result in one or more desirable traits in the therapeutic T cell, including but not limited to, reduced expression of an immune checkpoint receptor (e.g., PDA, CTLA4), reduced expression of HLA proteins (e.g., B2M, HLA-A), and reduced expression of an endogenous TCR.

2018290843 07 Apr 2025

[0206] In some embodiments, the modified cell is an antibody-producing B cell. The modification may results in one or more desirable traits in the B cell, including but not limited to, enhanced antibody production.

[0207] The invention also relates to a modified non-human animal or a modified plant. The modified non-human animal can be a farm animal. The modified plant can be an agricultural crop.

[0208] The invention further relates to a method for cell therapy, comprising administering to a patient in need thereof the modified cell described herein, wherein the presence of the modified cell remedies a disease in the patient. In one embodiment, the modified cell for cell therapy is a CAR-T cell capable of recognizing and/or attacking a tumor cell. In another embodiment, the modified cell for cell therapy is a stem cell, such as a neural stem cell, a mesenchymal stem cell, a hematopoietic stem cell, or an iPSC cell.

[0209] The invention additionally relates to an engineered, non-naturally occurring system suitable for modifying an Adenine or Cytosine in a target locus of interest, comprising: a targeteting domain; an adenosine deaminase protein or catalytic domain thereof, or one or more nucleotide sequences encoding; wherein the adenosine deaminase protein or catalytic domain thereof is covalently or non-covalently linked to the targeting domain or is adapted to link thereto after delivery; wherein the targeting domain is capable of hybridizing with a target sequence comprising an Adenine or Cytidine within an RNA or DNA polynucleotide of interest.

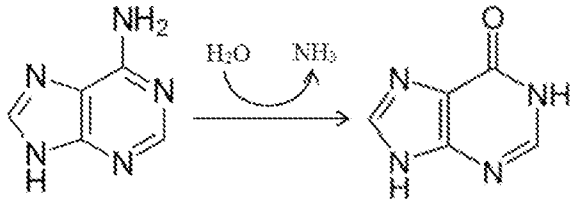
[0210] The invention additionally relates to an engineered, non-naturally occurring vector system suitable for modifying an Adenine or Cytosine in a target locus of interest, comprising one or more vectors comprising: (a) a first regulatory element operably linked to one or more nucleotide sequences encoding encoding a targeting domain; and (b) optionally a nucleotide sequence encoding an adenosine deaminase protein or catalytic domain thereof which is under control of the first or operably linked to a second regulatory element; wherein, if the nucleotide sequence encoding an adenosine deaminase protein or catalytic domain thereof is operably linked to a second regulatory element, the adenosine deaminase protein or catalytic domain thereof is adapted to link to the targeting domain after expression; wherein the targeting domain is capable of hybridizing with a target sequence comprising an Adenine or Cytosine within the target locus; wherein components (a) and (b) are located on the same or different vectors of the system.

[0211] The invention additionally relates to *in vitro*, *ex vivo* or *in vivo* host cell or cell line or progeny thereof comprising the engineered, non-naturally occurring system or vector system

described herein. The host cell can be a eukaryotic cell, such as an animal cell, a plant cell, an mammalian cell, or a human cell.

Adenosine Deaminase

[0212] The term "adenosine deaminase" or "adenosine deaminase protein" as used herein refers to a protein, a polypeptide, or one or more functional domain(s) of a protein or a polypeptide that is capable of catalyzing a hydrolytic deamination reaction that converts an adenine (or an adenine moiety of a molecule) to a hypoxanthine (or a hypoxanthine moiety of a molecule), as shown below. In some embodiments, the adenine-containing molecule is an adenosine (A), and the hypoxanthine-containing molecule is an inosine (I). The adenine-containing molecule can be deoxyribonucleic acid (DNA) or ribonucleic acid (RNA).



Adenine Hypoxanthine

[0213] According to the present disclosure, adenosine deaminases that can be used in connection with the present disclosure include, but are not limited to, members of the enzyme family known as adenosine deaminases that act on RNA (ADARs), members of the enzyme family known as adenosine deaminases that act on tRNA (ADATs), and other adenosine deaminase domain-containing (ADAD) family members. According to the present disclosure, the adenosine deaminase is capable of targeting adenine in a RNA/DNA and RNA duplexes. Indeed, Zheng et al. (Nucleic Acids Res. 2017, 45(6): 3369-3377) demonstrate that ADARs can cary out adenosine to inosine editing reactions on RNA/DNA and RNA/RNA duplexes. In particular embodiments, the adenosine deaminase has been modified to increase its ability to edit DNA in a RNA/DNAn RNA duplex as detailed herein below.

[0001] In some embodiments, the adenosine deaminase is derived from one or more metazoa species, including but not limited to, mammals, birds, frogs, squids, fish, flies and worms. In some embodiments, the adenosine deaminase is a human, squid or Drosophila adenosine deaminase.

[0214] In some embodiments, the adenosine deaminase is a human ADAR, including hADAR1, hADAR2, hADAR3. In some embodiments, the adenosine deaminase is a Caenorhabditis elegans ADAR protein, including ADR-1 and ADR-2. In some embodiments, the adenosine deaminase is a Drosophila ADAR protein, including dAdar. In some

embodiments, the adenosine deaminase is a squid *Loligo pealeii* ADAR protein, including sqADAR2a and sqADAR2b. In some embodiments, the adenosine deaminase is a human ADAT protein. In some embodiments, the adenosine deaminase is a Drosophila ADAT protein. In some embodiments, the adenosine deaminase is a human ADAD protein, including TENR (hADAD1) and TENRL (hADAD2).

[0215] In some embodiments, the adenosine deaminase is a TadA protein such as E. coli TadA. See Kim et al., Biochemistry 45:6407-6416 (2006); Wolf et al., EMBO J. 21:3841-3851 (2002). In some embodiments, the adenosine deaminase is mouse ADA. See Grunebaum et al., Curr. Opin. Allergy Clin. Immunol. 13:630-638 (2013). In some embodiments, the adenosine deaminase is human ADAT2. See Fukui et al., J. Nucleic Acids 2010:260512 (2010).

[0216] In some embodiments, the adenosine deaminase protein recognizes and converts one or more target adenosine residue(s) in a double-stranded nucleic acid substrate into inosine residues (s). In some embodiments, the double-stranded nucleic acid substrate is a RNA-DNA hybrid duplex. In some embodiments, the adenosine deaminase protein recognizes a binding window on the double-stranded substrate. In some embodiments, the binding window contains at least one target adenosine residue(s). In some embodiments, the binding window is in the range of about 3 bp to about 100 bp. In some embodiments, the binding window is in the range of about 5 bp to about 50 bp. In some embodiments, the binding window is in the range of about 10 bp to about 30 bp. In some embodiments, the binding window is about 1 bp, 2 bp, 3 bp, 5 bp, 7 bp, 10 bp, 15 bp, 20 bp, 25 bp, 30 bp, 40 bp, 45 bp, 50 bp, 55 bp, 60 bp, 65 bp, 70 bp, 75 bp, 80 bp, 85 bp, 90 bp, 95 bp, or 100 bp.

[0217] In some embodiments, the adenosine deaminase protein comprises one or more deaminase domains. Not intended to be bound by a particular theory, it is contemplated that the deaminase domain functions to recognize and convert one or more target adenosine (A) residue(s) contained in a double-stranded nucleic acid substrate into inosine (I) residue(s). In some embodiments, the deaminase domain comprises an active center. In some embodiments, the active center comprises a zinc ion. In some embodiments, during the A-to-I editing process, base pairing at the target adenosine residue is disrupted, and the target adenosine residue is "flipped" out of the double helix to become accessible by the adenosine deaminase. In some embodiments, amino acid residues in or near the active center interact with one or more nucleotide(s) 5' to a target adenosine residue. In some embodiments, amino acid residues in or near the active center interact with one or more nucleotide(s) 3' to a target adenosine residue. In some embodiments, amino acid residues in or near the active center further interact with the nucleotide complementary to the target adenosine residue on the opposite strand. In some

embodiments, the amino acid residues form hydrogen bonds with the 2' hydroxyl group of the nucleotides.

[0218] In some embodiments, the adenosine deaminase comprises human ADAR2 full protein (hADAR2) or the deaminase domain thereof (hADAR2-D). In some embodiments, the adenosine deaminase is an ADAR family member that is homologous to hADAR2 or hADAR2-D.

[0219] Particularly, in some embodiments, the homologous ADAR protein is human ADAR1 (hADAR1) or the deaminase domain thereof (hADAR1-D). In some embodiments, glycine 1007 of hADAR1-D corresponds to glycine 487 hADAR2-D, and glutamic Acid 1008 of hADAR1-D corresponds to glutamic acid 488 of hADAR2-D.

[0220] In some embodiments, the adenosine deaminase comprises the wild-type amino acid sequence of hADAR2-D. In some embodiments, the adenosine deaminase comprises one or more mutations in the hADAR2-D sequence, such that the editing efficiency, and/or substrate editing preference of hADAR2-D is changed according to specific needs.

[0221] Certain mutations of hADAR1 and hADAR2 proteins have been described in Kuttan et al., Proc Natl Acad Sci U S A. (2012) 109(48):E3295-304; Want et al. ACS Chem Biol. (2015) 10(11):2512-9; and Zheng et al. Nucleic Acids Res. (2017) 45(6):3369-337, each of which is incorporated herein by reference in its entirety.

[0222] In some embodiments, the adenosine deaminase comprises a mutation at glycine336 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glycine residue at position 336 is replaced by an aspartic acid residue (G336D).

[0223] In some embodiments, the adenosine deaminase comprises a mutation at Glycine487 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glycine residue at position 487 is replaced by a non-polar amino acid residue with relatively small side chains. For example, in some embodiments, the glycine residue at position 487 is replaced by an alanine residue (G487A). In some embodiments, the glycine residue at position 487 is replaced by a valine residue (G487V). In some embodiments, the glycine residue at position 487 is replaced by an amino acid residue with relatively large side chains. In some embodiments, the glycine residue at position 487 is replaced by a arginine residue (G487R). In some embodiments, the glycine residue at position 487 is replaced by a lysine residue (G487K). In some embodiments, the glycine residue at position 487 is replaced by a tryptophan residue (G487W). In some embodiments, the glycine residue at position 487 is replaced by a tyrosine residue (G487Y).

[0224] In some embodiments, the adenosine deaminase comprises a mutation at glutamic acid488 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glutamic acid residue at position 488 is replaced by a glutamine residue (E488Q). In some embodiments, the glutamic acid residue at position 488 is replaced by a histidine residue (E488H). In some embodiments, the glutamic acid residue at position 488 is replace by an arginine residue (E488R). In some embodiments, the glutamic acid residue at position 488 is replace by a lysine residue (E488K). In some embodiments, the glutamic acid residue at position 488 is replace by an asparagine residue (E488N). In some embodiments, the glutamic acid residue at position 488 is replace by an alanine residue (E488A). In some embodiments, the glutamic acid residue at position 488 is replace by a Methionine residue (E488M). In some embodiments, the glutamic acid residue at position 488 is replace by a serine residue (E488S). In some embodiments, the glutamic acid residue at position 488 is replace by a phenylalanine residue (E488F). In some embodiments, the glutamic acid residue at position 488 is replace by a lysine residue (E488L). In some embodiments, the glutamic acid residue at position 488 is replace by a tryptophan residue (E488W).

[0225] In some embodiments, the adenosine deaminase comprises a mutation at threonine490 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the threonine residue at position 490 is replaced by a cysteine residue (T490C). In some embodiments, the threonine residue at position 490 is replaced by a serine residue (T490S). In some embodiments, the threonine residue at position 490 is replaced by an alanine residue (T490A). In some embodiments, the threonine residue at position 490 is replaced by a phenylalanine residue (T490F). In some embodiments, the threonine residue at position 490 is replaced by a tyrosine residue (T490Y). In some embodiments, the threonine residue at position 490 is replaced by a serine residue (T490R). In some embodiments, the threonine residue at position 490 is replaced by an alanine residue (T490K). In some embodiments, the threonine residue at position 490 is replaced by a phenylalanine residue (T490P). In some embodiments, the threonine residue at position 490 is replaced by a tyrosine residue (T490E).

[0226] In some embodiments, the adenosine deaminase comprises a mutation at valine493 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the valine residue at position 493 is replaced by an alanine residue (V493A). In some embodiments, the valine residue at position 493 is replaced by a serine residue (V493S). In some embodiments, the valine residue at position 493 is replaced by a threonine residue (V493T). In some embodiments, the valine residue at position 493 is

replaced by an arginine residue (V493R). In some embodiments, the valine residue at position 493 is replaced by an aspartic acid residue (V493D). In some embodiments, the valine residue at position 493 is replaced by a proline residue (V493P). In some embodiments, the valine residue at position 493 is replaced by a glycine residue (V493G).

[0227] In some embodiments, the adenosine deaminase comprises a mutation at alanine589 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the alanine residue at position 589 is replaced by a valine residue (A589V).

[0228] In some embodiments, the adenosine deaminase comprises a mutation at asparagine597 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the asparagine residue at position 597 is replaced by a lysine residue (N597K). In some embodiments, the adenosine deaminase comprises a mutation at position 597 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 597 is replaced by an arginine residue (N597R). In some embodiments, the adenosine deaminase comprises a mutation at position 597 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 597 is replaced by an alanine residue (N597A). In some embodiments, the adenosine deaminase comprises a mutation at position 597 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 597 is replaced by a glutamic acid residue (N597E). In some embodiments, the adenosine deaminase comprises a mutation at position 597 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 597 is replaced by a histidine residue (N597H). In some embodiments, the adenosine deaminase comprises a mutation at position 597 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 597 is replaced by a glycine residue (N597G). In some embodiments, the adenosine deaminase comprises a mutation at position 597 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 597 is replaced by a tyrosine residue (N597Y). In some embodiments, the asparagine residue at position 597 is replaced by a phenylalanine residue (N597F). In some embodiments, the adenosine deaminase comprises mutation N597I. In some embodiments, the adenosine deaminase comprises mutation N597L. In some embodiments, the adenosine deaminase comprises mutation N597V. In some embodiments, the adenosine deaminase comprises

mutation N597M. In some embodiments, the adenosine deaminase comprises mutation N597C. In some embodiments, the adenosine deaminase comprises mutation N597P. In some embodiments, the adenosine deaminase comprises mutation N597T. In some embodiments, the adenosine deaminase comprises mutation N597S. In some embodiments, the adenosine deaminase comprises mutation N597W. In some embodiments, the adenosine deaminase comprises mutation N597Q. In some embodiments, the adenosine deaminase comprises mutation N597D. In certain example embodiments, the mutations at N597 described above are further made in the context of an E488Q background

[0229] In some embodiments, the adenosine deaminase comprises a mutation at serine599 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the serine residue at position 599 is replaced by a threonine residue (S599T).

[0230] In some embodiments, the adenosine deaminase comprises a mutation at asparagine613 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the asparagine residue at position 613 is replaced by a lysine residue (N613K). In some embodiments, the adenosine deaminase comprises a mutation at position 613 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 613 is replaced by an arginine residue (N613R). In some embodiments, the adenosine deaminase comprises a mutation at position 613 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 613 is replaced by an alanine residue (N613A) In some embodiments, the adenosine deaminase comprises a mutation at position 613 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 613 is replaced by a glutamic acid residue (N613E). In some embodiments, the adenosine deaminase comprises mutation N613I. In some embodiments, the adenosine deaminase comprises mutation N613L. In some embodiments, the adenosine deaminase comprises mutation N613V. In some embodiments, the adenosine deaminase comprises mutation N613F. In some embodiments, the adenosine deaminase comprises mutation N613M. In some embodiments, the adenosine deaminase comprises mutation N613C. In some embodiments, the adenosine deaminase comprises mutation N613G. In some embodiments, the adenosine deaminase comprises mutation N613P. In some embodiments, the adenosine deaminase comprises mutation N613T. In some embodiments, the adenosine deaminase comprises mutation N613S. In some embodiments, the adenosine deaminase comprises mutation N613Y.

In some embodiments, the adenosine deaminase comprises mutation N613W. In some embodiments, the adenosine deaminase comprises mutation N613Q. In some embodiments, the adenosine deaminase comprises mutation N613H. In some embodiments, the adenosine deaminase comprises mutation N613D. In some embodiments, the mutations at N613 described above are further made in combination with a E488Q mutation.

[0231] In some embodiments, to improve editing efficiency, the adenosine deaminase may comprise one or more of the mutations: G336D, G487A, G487V, E488Q, E488H, E488R, E488N, E488A, E488S, E488M, T490C, T490S, V493T, V493S, V493A, V493R, V493D, V493P, V493G, N597K, N597R, N597A, N597E, N597H, N597G, N597Y, A589V, S599T, N613K, N613R, N613A, N613E, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above.

[0232] In some embodiments, to reduce editing efficiency, the adenosine deaminase may comprise one or more of the mutations: E488F, E488L, E488W, T490A, T490F, T490Y, T490R, T490K, T490P, T490E, N597F, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In particular embodiments, it can be of interest to use an adenosine deaminase enzyme with reduced efficicay to reduce off-target effects.

[0233] In some embodiments, to reduce off-target effects, the adenosine deaminase comprises one or more of mutations at R348, V351, T375, K376, E396, C451, R455, N473, R474, K475, R477, R481, S486, E488, T490, S495, R510, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In some embodiments, the adenosine deaminase comprises mutation at E488 and one or more additional positions selected from R348, V351, T375, K376, E396, C451, R455, N473, R474, K475, R477, R481, S486, T490, S495, R510. In some embodiments, the adenosine deaminase comprises mutation at T375, and optionally at one or more additional positions. In some embodiments, the adenosine deaminase comprises mutation at N473, and optionally at one or more additional positions. In some embodiments, the adenosine deaminase comprises mutation at V351, and optionally at one or more additional positions. In some embodiments, the adenosine deaminase comprises mutation at E488 and T375, and optionally at one or more additional positions. In some embodiments, the adenosine deaminase comprises mutation at E488 and N473, and optionally at one or more additional positions. In some embodiments, the adenosine deaminase comprises mutation E488 and V351, and optionally at one or more additional positions. In some embodiments, the adenosine deaminase comprises mutation at E488 and one or more of T375, N473, and V351.

[0234] In some embodiments, to reduce off-target effects, the adenosine deaminase comprises one or more of mutations selected from R348E, V351L, T375G, T375S, R455G, R455S, R455E, N473D, R474E, K475Q, R477E, R481E, S486T, E488Q, T490A, T490S, S495T, and R510E, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In some embodiments, the adenosine deaminase comprises mutation E488Q and one or more additional mutations selected from R348E, V351L, T375G, T375S, R455G, R455S, R455E, N473D, R474E, K475Q, R477E, R481E, S486T, T490A, T490S, S495T, and R510E. In some embodiments, the adenosine deaminase comprises mutation T375G or T375S, and optionally one or more additional mutations. In some embodiments, the adenosine deaminase comprises mutation N473D, and optionally one or more additional mutations. In some embodiments, the adenosine deaminase comprises mutation V351L, and optionally one or more additional mutations. In some embodiments, the adenosine deaminase comprises mutation E488Q, and T375G or T375G, and optionally one or more additional mutations. In some embodiments, the adenosine deaminase comprises mutation E488Q and N473D, and optionally one or more additional mutations. In some embodiments, the adenosine deaminase comprises mutation E488Q and V351L, and optionally one or more additional mutations. In some embodiments, the adenosine deaminase comprises mutation E488Q and one or more of T375G/S, N473D and V351L.

[0235] Crystal structures of the human ADAR2 deaminase domain bound to duplex RNA reveal a protein loop that binds the RNA on the 5' side of the modification site. This 5' binding loop is one contributor to substrate specificity differences between ADAR family members. See Wang et al., Nucleic Acids Res., 44(20):9872-9880 (2016), the content of which is incorporated herein by reference in its entirety. In addition, an ADAR2-specific RNA-binding loop was identified near the enzyme active site. See Mathews et al., Nat. Struct. Mol. Biol., 23(5):426-33 (2016), the content of which is incorporated herein by reference in its entirety. In some embodiments, the adenosine deaminase comprises one or more mutations in the RNA binding loop to improve editing specificity and/or efficiency.

[0236] In some embodiments, the adenosine deaminase comprises a mutation at alanine454 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the alanine residue at position 454 is replaced by a serine residue (A454S). In some embodiments, the alanine residue at position 454 is replaced by a cysteine residue (A454C). In some embodiments, the alanine residue at position 454 is replaced by an aspartic acid residue (A454D).

[0237] In some embodiments, the adenosine deaminase comprises a mutation at arginine455 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the arginine residue at position 455 is replaced by an alanine residue (R455A). In some embodiments, the arginine residue at position 455 is replaced by a valine residue (R455V). In some embodiments, the arginine residue at position 455 is replaced by a histidine residue (R455H). In some embodiments, the arginine residue at position 455 is replaced by a glycine residue (R455G). In some embodiments, the arginine residue at position 455 is replaced by a serine residue (R455S). In some embodiments, the arginine residue at position 455 is replaced by a glutamic acid residue (R455E). In some embodiments, the adenosine deaminase comprises mutation R455C. In some embodiments, the adenosine deaminase comprises mutation R455I. In some embodiments, the adenosine deaminase comprises mutation R455K. In some embodiments, the adenosine deaminase comprises mutation R455L. In some embodiments, the adenosine deaminase comprises mutation R455M. In some embodiments, the adenosine deaminase comprises mutation R455N. In some embodiments, the adenosine deaminase comprises mutation R455Q. In some embodiments, the adenosine deaminase comprises mutation R455F. In some embodiments, the adenosine deaminase comprises mutation R455W. In some embodiments, the adenosine deaminase comprises mutation R455P. In some embodiments, the adenosine deaminase comprises mutation R455Y. In some embodiments, the adenosine deaminase comprises mutation R455E. In some embodiments, the adenosine deaminase comprises mutation R455D. In some embodiments, the mutations at at R455 described above are further made in combination with a E488Q mutation.

[0238] In some embodiments, the adenosine deaminase comprises a mutation at isoleucine456 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the isoleucine residue at position 456 is replaced by a valine residue (I456V). In some embodiments, the isoleucine residue at position 456 is replaced by a leucine residue (I456L). In some embodiments, the isoleucine residue at position 456 is replaced by an aspartic acid residue (I456D).

[0239] In some embodiments, the adenosine deaminase comprises a mutation at phenylalanine457 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the phenylalanine residue at position 457 is replaced by a tyrosine residue (F457Y). In some embodiments, the phenylalanine residue at position 457 is replaced by an arginine residue (F457R). In some embodiments, the phenylalanine residue at position 457 is replaced by a glutamic acid residue (F457E).

[0240] In some embodiments, the adenosine deaminase comprises a mutation at serine458 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the serine residue at position 458 is replaced by a valine residue (S458V). In some embodiments, the serine residue at position 458 is replaced by a phenylalanine residue (S458F). In some embodiments, the serine residue at position 458 is replaced by a proline residue (S458P). In some embodiments, the adenosine deaminase comprises mutation S458I. In some embodiments, the adenosine deaminase comprises mutation S458L. In some embodiments, the adenosine deaminase comprises mutation S458M. In some embodiments, the adenosine deaminase comprises mutation S458C. In some embodiments, the adenosine deaminase comprises mutation S458A. In some embodiments, the adenosine deaminase comprises mutation S458G. In some embodiments, the adenosine deaminase comprises mutation S458T. In some embodiments, the adenosine deaminase comprises mutation S458Y. In some embodiments, the adenosine deaminase comprises mutation S458W. In some embodiments, the adenosine deaminase comprises mutation S458Q. In some embodiments, the adenosine deaminase comprises mutation S458N. In some embodiments, the adenosine deaminase comprises mutation S458H. In some embodiments, the adenosine deaminase comprises mutation S458E. In some embodiments, the adenosine deaminase comprises mutation S458D. In some embodiments, the adenosine deaminase comprises mutation S458K. In some embodiments, the adenosine deaminase comprises mutation S458R. In some embodiments, the mutations at S458 described above are further made in combination with a E488Q mutation.

[0241] In some embodiments, the adenosine deaminase comprises a mutation at proline459 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the proline residue at position 459 is replaced by a cysteine residue (P459C). In some embodiments, the proline residue at position 459 is replaced by a histidine residue (P459H). In some embodiments, the proline residue at position 459 is replaced by a tryptophan residue (P459W).

[0242] In some embodiments, the adenosine deaminase comprises a mutation at histidine460 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the histidine residue at position 460 is replaced by an arginine residue (H460R). In some embodiments, the histidine residue at position 460 is replaced by an isoleucine residue (H460I). In some embodiments, the histidine residue at position 460 is replaced by a proline residue (H460P). In some embodiments, the adenosine deaminase comprises mutation H460L. In some embodiments, the adenosine

deaminase comprises mutation H460V. In some embodiments, the adenosine deaminase comprises mutation H460F. In some embodiments, the adenosine deaminase comprises mutation H460M. In some embodiments, the adenosine deaminase comprises mutation H460C. In some embodiments, the adenosine deaminase comprises mutation H460A. In some embodiments, the adenosine deaminase comprises mutation H460G. In some embodiments, the adenosine deaminase comprises mutation H460T. In some embodiments, the adenosine deaminase comprises mutation H460S. In some embodiments, the adenosine deaminase comprises mutation H460Y. In some embodiments, the adenosine deaminase comprises mutation H460W. In some embodiments, the adenosine deaminase comprises mutation H460Q. In some embodiments, the adenosine deaminase comprises mutation H460N. In some embodiments, the adenosine deaminase comprises mutation H460E. In some embodiments, the adenosine deaminase comprises mutation H460D. In some embodiments, the adenosine deaminase comprises mutation H460K. In some embodiments, the mutations at H460 described above are further made in combination with a E488Q mutation.

[0243] In some embodiments, the adenosine deaminase comprises a mutation at proline462 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the proline residue at position 462 is replaced by a serine residue (P462S). In some embodiments, the proline residue at position 462 is replaced by a tryptophan residue (P462W). In some embodiments, the proline residue at position 462 is replaced by a glutamic acid residue (P462E).

[0244] In some embodiments, the adenosine deaminase comprises a mutation at aspartic acid469 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the aspartic acid residue at position 469 is replaced by a glutamine residue (D469Q). In some embodiments, the aspartic acid residue at position 469 is replaced by a serine residue (D469S). In some embodiments, the aspartic acid residue at position 469 is replaced by a tyrosine residue (D469Y).

[0245] In some embodiments, the adenosine deaminase comprises a mutation at arginine470 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the arginine residue at position 470 is replaced by an alanine residue (R470A). In some embodiments, the arginine residue at position 470 is replaced by an isoleucine residue (R470I). In some embodiments, the arginine residue at position 470 is replaced by an aspartic acid residue (R470D).

[0246] In some embodiments, the adenosine deaminase comprises a mutation at histidine471 of the hADAR2-D amino acid sequence, or a corresponding position in a

homologous ADAR protein. In some embodiments, the histidine residue at position 471 is replaced by a lysine residue (H471K). In some embodiments, the histidine residue at position 471 is replaced by a threonine residue (H471T). In some embodiments, the histidine residue at position 471 is replaced by a valine residue (H471V).

[0247] In some embodiments, the adenosine deaminase comprises a mutation at proline472 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the proline residue at position 472 is replaced by a lysine residue (P472K). In some embodiments, the proline residue at position 472 is replaced by a threonine residue (P472T). In some embodiments, the proline residue at position 472 is replaced by an aspartic acid residue (P472D).

[0248] In some embodiments, the adenosine deaminase comprises a mutation at asparagine473 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the asparagine residue at position 473 is replaced by an arginine residue (N473R). In some embodiments, the asparagine residue at position 473 is replaced by a tryptophan residue (N473W). In some embodiments, the asparagine residue at position 473 is replaced by a proline residue (N473P). In some embodiments, the asparagine residue at position 473 is replaced by an aspartic acid residue (N473D).

[0249] In some embodiments, the adenosine deaminase comprises a mutation at arginine474 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the arginine residue at position 474 is replaced by a lysine residue (R474K). In some embodiments, the arginine residue at position 474 is replaced by a glycine residue (R474G). In some embodiments, the arginine residue at position 474 is replaced by an aspartic acid residue (R474D). In some embodiments, the arginine residue at position 474 is replaced by a glutamic acid residue (R474E).

[0250] In some embodiments, the adenosine deaminase comprises a mutation at lysine475 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the lysine residue at position 475 is replaced by a glutamine residue (K475Q). In some embodiments, the lysine residue at position 475 is replaced by an asparagine residue (K475N). In some embodiments, the lysine residue at position 475 is replaced by an aspartic acid residue (K475D).

[0251] In some embodiments, the adenosine deaminase comprises a mutation at alanine476 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the alanine residue at position 476 is

replaced by a serine residue (A476S). In some embodiments, the alanine residue at position 476 is replaced by an arginine residue (A476R). In some embodiments, the alanine residue at position 476 is replaced by a glutamic acid residue (A476E).

[0252] In some embodiments, the adenosine deaminase comprises a mutation at arginine477 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the arginine residue at position 477 is replaced by a lysine residue (R477K). In some embodiments, the arginine residue at position 477 is replaced by a threonine residue (R477T). In some embodiments, the arginine residue at position 477 is replaced by a phenylalanine residue (R477F). In some embodiments, the arginine residue at position 474 is replaced by a glutamic acid residue (R477E).

[0253] In some embodiments, the adenosine deaminase comprises a mutation at glycine478 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glycine residue at position 478 is replaced by an alanine residue (G478A). In some embodiments, the glycine residue at position 478 is replaced by an arginine residue (G478R). In some embodiments, the glycine residue at position 478 is replaced by a tyrosine residue (G478Y). In some embodiments, the adenosine deaminase comprises mutation G478I. In some embodiments, the adenosine deaminase comprises mutation G478L. In some embodiments, the adenosine deaminase comprises mutation G478V. In some embodiments, the adenosine deaminase comprises mutation G478F. In some embodiments, the adenosine deaminase comprises mutation G478M. In some embodiments, the adenosine deaminase comprises mutation G478C. In some embodiments, the adenosine deaminase comprises mutation G478P. In some embodiments, the adenosine deaminase comprises mutation G478T. In some embodiments, the adenosine deaminase comprises mutation G478S. In some embodiments, the adenosine deaminase comprises mutation G478W. In some embodiments, the adenosine deaminase comprises mutation G478Q. In some embodiments, the adenosine deaminase comprises mutation G478N. In some embodiments, the adenosine deaminase comprises mutation G478H. In some embodiments, the adenosine deaminase comprises mutation G478E. In some embodiments, the adenosine deaminase comprises mutation G478D. In some embodiments, the adenosine deaminase comprises mutation G478K. In some embodiments, the mutations at G478 described above are further made in combination with a E488Q mutation.

[0254] In some embodiments, the adenosine deaminase comprises a mutation at glutamine479 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glutamine residue at position 479 is

replaced by an asparagine residue (Q479N). In some embodiments, the glutamine residue at position 479 is replaced by a serine residue (Q479S). In some embodiments, the glutamine residue at position 479 is replaced by a proline residue (Q479P).

[0255] In some embodiments, the adenosine deaminase comprises a mutation at arginine348 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the arginine residue at position 348 is replaced by an alanine residue (R348A). In some embodiments, the arginine residue at position 348 is replaced by a glutamic acid residue (R348E).

[0256] In some embodiments, the adenosine deaminase comprises a mutation at valine351 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the valine residue at position 351 is replaced by a leucine residue (V351L). In some embodiments, the adenosine deaminase comprises mutation V351Y. In some embodiments, the adenosine deaminase comprises mutation V351M. In some embodiments, the adenosine deaminase comprises mutation V351T. In some embodiments, the adenosine deaminase comprises mutation V351G. In some embodiments, the adenosine deaminase comprises mutation V351A. In some embodiments, the adenosine deaminase comprises mutation V351F. In some embodiments, the adenosine deaminase comprises mutation V351E. In some embodiments, the adenosine deaminase comprises mutation V351I. In some embodiments, the adenosine deaminase comprises mutation V351C. In some embodiments, the adenosine deaminase comprises mutation V351H. In some embodiments, the adenosine deaminase comprises mutation V351P. In some embodiments, the adenosine deaminase comprises mutation V351S. In some embodiments, the adenosine deaminase comprises mutation V351K. In some embodiments, the adenosine deaminase comprises mutation V351N. In some embodiments, the adenosine deaminase comprises mutation V351W. In some embodiments, the adenosine deaminase comprises mutation V351Q. In some embodiments, the adenosine deaminase comprises mutation V351D. In some embodiments, the adenosine deaminase comprises mutation V351R. In some embodiments, the mutations at V351 described above are further made in combination with a E488Q mutation.

[0257] In some embodiments, the adenosine deaminase comprises a mutation at threonine375 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the threonine residue at position 375 is replaced by a glycine residue (T375G). In some embodiments, the threonine residue at position 375 is replaced by a serine residue (T375S). In some embodiments, the adenosine deaminase

comprises mutation T375H. In some embodiments, the adenosine deaminase comprises mutation T375Q. In some embodiments, the adenosine deaminase comprises mutation T375C. In some embodiments, the adenosine deaminase comprises mutation T375N. In some embodiments, the adenosine deaminase comprises mutation T375M. In some embodiments, the adenosine deaminase comprises mutation T375A. In some embodiments, the adenosine deaminase comprises mutation T375W. In some embodiments, the adenosine deaminase comprises mutation T375V. In some embodiments, the adenosine deaminase comprises mutation T375R. In some embodiments, the adenosine deaminase comprises mutation T375E. In some embodiments, the adenosine deaminase comprises mutation T375K. In some embodiments, the adenosine deaminase comprises mutation T375F. In some embodiments, the adenosine deaminase comprises mutation T375I. In some embodiments, the adenosine deaminase comprises mutation T375D. In some embodiments, the adenosine deaminase comprises mutation T375P. In some embodiments, the adenosine deaminase comprises mutation T375L. In some embodiments, the adenosine deaminase comprises mutation T375Y. In some embodiments, the mutations at T375Y described above are further made in combination with an E488Q mutation.

[0258] In some embodiments, the adenosine deaminase comprises a mutation at arginine481 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the arginine residue at position 481 is replaced by a glutamic acid residue (R481E).

[0259] In some embodiments, the adenosine deaminase comprises a mutation at serine486 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the serine residue at position 486 is replaced by a threonine residue (S486T).

[0260] In some embodiments, the adenosine deaminase comprises a mutation at threonine490 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the threonine residue at position 490 is replaced by an alanine residue (T490A). In some embodiments, the threonine residue at position 490 is replaced by a serine residue (T490S).

[0261] In some embodiments, the adenosine deaminase comprises a mutation at serine495 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the serine residue at position 495 is replaced by a threonine residue (S495T).

[0262] In some embodiments, the adenosine deaminase comprises a mutation at arginine510 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the arginine residue at position 510 is replaced by a glutamine residue (R510Q). In some embodiments, the arginine residue at position 510 is replaced by an alanine residue (R510A). In some embodiments, the arginine residue at position 510 is replaced by a glutamic acid residue (R510E).

[0263] In some embodiments, the adenosine deaminase comprises a mutation at glycine593 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glycine residue at position 593 is replaced by an alanine residue (G593A). In some embodiments, the glycine residue at position 593 is replaced by a glutamic acid residue (G593E).

[0264] In some embodiments, the adenosine deaminase comprises a mutation at lysine594 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the lysine residue at position 594 is replaced by an alanine residue (K594A).

[0265] In some embodiments, the adenosine deaminase comprises a mutation at any one or more of positions A454, R455, I456, F457, S458, P459, H460, P462, D469, R470, H471, P472, N473, R474, K475, A476, R477, G478, Q479, R348, R510, G593, K594 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein.

[0266] In some embodiments, the adenosine deaminase comprises any one or more of mutations A454S, A454C, A454D, R455A, R455V, R455H, I456V, I456L, I456D, F457Y, F457R, F457E, S458V, S458F, S458P, P459C, P459H, P459W, H460R, H460I, H460P, P462S, P462W, P462E, D469Q, D469S, D469Y, R470A, R470I, R470D, H471K, H471T, H471V, P472K, P472T, P472D, N473R, N473W, N473P, R474K, R474G, R474D, K475Q, K475N, K475D, A476S, A476R, A476E, R477K, R477T, R477F, G478A, G478R, G478Y, Q479N, Q479S, Q479P, R348A, R510Q, R510A, G593A, G593E, K594A of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein.

[0267] In some embodiments, the adenosine deaminase comprises a mutation at any one or more of positions T375, V351, G478, S458, H460 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein, optionally in combination a mutation at E488. In some embodiments, the adenosine deaminase comprises one or more of mutations selected from T375G, T375C, T375H, T375Q, V351M, V351T, V351Y, G478R, S458F, H460I, optionally in combination with E488Q.

[0268] In some embodiments, the adenosine deaminase comprises one or more of mutations selected from T375H, T375Q, V351M, V351Y, H460P, optionally in combination with E488Q.

[0269] In some embodiments, the adenosine deaminase comprises mutations T375S and S458F, optionally in combination with E488Q.

[0270] In some embodiments, the adenosine deaminase comprises a mutation at two or more of positions T375, N473, R474, G478, S458, P459, V351, R455, R455, T490, R348, Q479 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein, optionally in combination a mutation at E488. In some embodiments, the adenosine deaminase comprises two or more of mutations selected from T375G, T375S, N473D, R474E, G478R, S458F, P459W, V351L, R455G, R455S, T490A, R348E, Q479P, optionally in combination with E488Q.

[0271] In some embodiments, the adenosine deaminase comprises mutations T375G and V351L. In some embodiments, the adenosine deaminase comprises mutations T375G and R455G. In some embodiments, the adenosine deaminase comprises mutations T375G and R455S. In some embodiments, the adenosine deaminase comprises mutations T375G and T490A. In some embodiments, the adenosine deaminase comprises mutations T375G and R348E. In some embodiments, the adenosine deaminase comprises mutations T375S and V351L. In some embodiments, the adenosine deaminase comprises mutations T375S and R455G. In some embodiments, the adenosine deaminase comprises mutations T375S and R455S. In some embodiments, the adenosine deaminase comprises mutations T375S and T490A. In some embodiments, the adenosine deaminase comprises mutations T375S and R348E. In some embodiments, the adenosine deaminase comprises mutations N473D and V351L. In some embodiments, the adenosine deaminase comprises mutations N473D and R455G. In some embodiments, the adenosine deaminase comprises mutations N473D and R455S. In some embodiments, the adenosine deaminase comprises mutations N473D and T490A. In some embodiments, the adenosine deaminase comprises mutations N473D and R348E. In some embodiments, the adenosine deaminase comprises mutations R474E and V351L. In some embodiments, the adenosine deaminase comprises mutations R474E and R455G. In some embodiments, the adenosine deaminase comprises mutations R474E and R455S. In some embodiments, the adenosine deaminase comprises mutations R474E and T490A. In some embodiments, the adenosine deaminase comprises mutations R474E and R348E. In some embodiments, the adenosine deaminase comprises mutations S458F and T375G. In some embodiments, the adenosine deaminase comprises mutations S458F and

T375S. In some embodiments, the adenosine deaminase comprises mutations S458F and N473D. In some embodiments, the adenosine deaminase comprises mutations S458F and R474E. In some embodiments, the adenosine deaminase comprises mutations S458F and G478R. In some embodiments, the adenosine deaminase comprises mutations G478R and T375G. In some embodiments, the adenosine deaminase comprises mutations G478R and T375S. In some embodiments, the adenosine deaminase comprises mutations G478R and N473D. In some embodiments, the adenosine deaminase comprises mutations G478R and R474E. In some embodiments, the adenosine deaminase comprises mutations P459W and T375G. In some embodiments, the adenosine deaminase comprises mutations P459W and T375S. In some embodiments, the adenosine deaminase comprises mutations P459W and N473D. In some embodiments, the adenosine deaminase comprises mutations P459W and R474E. In some embodiments, the adenosine deaminase comprises mutations P459W and G478R. In some embodiments, the adenosine deaminase comprises mutations P459W and S458F. In some embodiments, the adenosine deaminase comprises mutations Q479P and T375G. In some embodiments, the adenosine deaminase comprises mutations Q479P and T375S. In some embodiments, the adenosine deaminase comprises mutations Q479P and N473D. In some embodiments, the adenosine deaminase comprises mutations Q479P and R474E. In some embodiments, the adenosine deaminase comprises mutations Q479P and G478R. In some embodiments, the adenosine deaminase comprises mutations Q479P and S458F. In some embodiments, the adenosine deaminase comprises mutations Q479P and P459W. All mutations described in this paragraph may also further be made in cominbation with a E488Q mutations.

[0272] In some embodiments, the adenosine deaminase comprises a mutation at any one or more of positions K475, Q479, P459, G478, S458of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein, optionally in combination a mutation at E488. In some embodiments, the adenosine deaminase comprises one or more of mutations selected from K475N, Q479N, P459W, G478R, S458P, S458F, optionally in combination with E488Q.

[0273] In some embodiments, the adenosine deaminase comprises a mutation at any one or more of positions T375, V351, R455, H460, A476 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein, optionally in combination a mutation at E488. In some embodiments, the adenosine deaminase comprises one or more of mutations selected from T375G, T375C, T375H, T375Q, V351M, V351T, V351Y, R455H, H460P, H460I, A476E, optionally in combination with E488Q.

[0274] In certain embodiments, improvement of editing and reduction of off-target modification is achieved by chemical modification of gRNAs. gRNAs which are chemically modified as exemplified in Vogel et al. (2014), Angew Chem Int Ed, 53:6267-6271, doi:10.1002/anie.201402634 (incorporated herein by reference in its entirety) reduce off-target activity and improve on-target efficiency. 2′-O-methyl and phosphothioate modified guide RNAs in general improve editing efficiency in cells.

[0275] ADAR has been known to demonstrate a preference for neighboring nucleotides on either side of the edited A (www.nature.com/nsmb/journal/v23/n5/full/nsmb.3203.html, Matthews et al. (2017), Nature Structural Mol Biol, 23(5): 426-433, incorporated herein by reference in its entirety). Accordingly, in certain embodiments, the gRNA, target, and/or ADAR is selected optimized for motif preference.

[0276] Intentional mismatches have been demonstrated in vitro to allow for editing of non-preferred motifs (https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gku272; Schneider et al (2014), Nucleic Acid Res, 42(10):e87); Fukuda et al. (2017), Scienticic Reports, 7, doi:10.1038/srep41478, incorporated herein by reference in its entirety). Accordingly, in certain embodiments, to enhance RNA editing efficiency on non-preferred 5’ or 3’ neighboring bases, intentional mismatches in neighboring bases are introduced.

[0277] Results suggest that A’s opposite C’s in the targeting window of the ADAR deaminase domain are preferentially edited over other bases. Additionally, A’s base-paired with U’s within a few bases of the targeted base show low levels of editing by Cas13b-ADAR fusions, suggesting that there is flexibility for the enzyme to edit multiple A’s. See e.g. FIG. 18. These two observations suggest that multiple A’s in the activity window of Cas13b-ADAR fusions could be specified for editing by mismatching all A’s to be edited with C’s. Accordingly, in certain embodiments, multiple A:C mismatches in the activity window are designed to create multiple A:I edits. In certain embodiments, to suppress potential off-target editing in the activity window, non-target A’s are paired with A’s or G’s.

[0278] The terms “editing specificity” and “editing preference” are used interchangeably herein to refer to the extent of A-to-I editing at a particular adenosine site in a double-stranded substrate. In some embodiment, the substrate editing preference is determined by the 5’ nearest neighbor and/or the 3’ nearest neighbor of the target adenosine residue. In some embodiments, the adenosine deaminase has preference for the 5’ nearest neighbor of the substrate ranked as U>A>C>G (“>” indicates greater preference). In some embodiments, the adenosine deaminase has preference for the 3’ nearest neighbor of the substrate ranked as G>C~A>U (“>” indicates greater preference; “~” indicates similar preference). In some embodiments, the adenosine

deaminase has preference for the 3' nearest neighbor of the substrate ranked as G>C>U~A (">" indicates greater preference; "~" indicates similar preference). In some embodiments, the adenosine deaminase has preference for the 3' nearest neighbor of the substrate ranked as G>C>A>U (">" indicates greater preference). In some embodiments, the adenosine deaminase has preference for the 3' nearest neighbor of the substrate ranked as C~G~A>U (">" indicates greater preference; "~" indicates similar preference). In some embodiments, the adenosine deaminase has preference for a triplet sequence containing the target adenosine residue ranked as TAG>AAG>CAC>AAT>GAA>GAC (">" indicates greater preference), the center A being the target adenosine residue.

[0279] In some embodiments, the substrate editing preference of an adenosine deaminase is affected by the presence or absence of a nucleic acid binding domain in the adenosine deaminase protein. In some embodiments, to modify substrate editing preference, the deaminase domain is connected with a double-strand RNA binding domain (dsRBD) or a double-strand RNA binding motif (dsRBM). In some embodiments, the dsRBD or dsRBM may be derived from an ADAR protein, such as hADAR1 or hADAR2. In some embodiments, a full length ADAR protein that comprises at least one dsRBD and a deaminase domain is used. In some embodiments, the one or more dsRBM or dsRBD is at the N-terminus of the deaminase domain. In other embodiments, the one or more dsRBM or dsRBD is at the C-terminus of the deaminase domain.

[0280] In some embodiments, the substrate editing preference of an adenosine deaminase is affected by amino acid residues near or in the active center of the enzyme. In some embodiments, to modify substrate editing preference, the adenosine deaminase may comprise one or more of the mutations: G336D, G487R, G487K, G487W, G487Y, E488Q, E488N, T490A, V493A, V493T, V493S, N597K, N597R, A589V, S599T, N613K, N613R, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above.

[0281] Particularly, in some embodiments, to reduce editing specificity, the adenosine deaminase can comprise one or more of mutations E488Q, V493A, N597K, N613K, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In some embodiments, to increase editing specificity, the adenosine deaminase can comprise mutation T490A.

[0282] In some embodiments, to increase editing preference for target adenosine (A) with an immediate 5' G, such as substrates comprising the triplet sequence GAC, the center A being the target adenosine residue, the adenosine deaminase can comprise one or more of mutations

G336D, E488Q, E488N, V493T, V493S, V493A, A589V, N597K, N597R, S599T, N613K, N613R, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above.

[0283] Particularly, in some embodiments, the adenosine deaminase comprises mutation E488Q or a corresponding mutation in a homologous ADAR protein for editing substrates comprising the following triplet sequences: GAC, GAA, GAU, GAG, CAU, AAU, UAC, the center A being the target adenosine residue.

[0284] In some embodiments, the adenosine deaminase comprises the wild-type amino acid sequence of hADAR1-D as defined in SEQ ID No. 761. In some embodiments, the adenosine deaminase comprises one or more mutations in the hADAR1-D sequence, such that the editing efficiency, and/or substrate editing preference of hADAR1-D is changed according to specific needs.

[0285] In some embodiments, the adenosine deaminase comprises a mutation at Glycine1007 of the hADAR1-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glycine residue at position 1007 is replaced by a non-polar amino acid residue with relatively small side chains. For example, in some embodiments, the glycine residue at position 1007 is replaced by an alanine residue (G1007A). In some embodiments, the glycine residue at position 1007 is replaced by a valine residue (G1007V). In some embodiments, the glycine residue at position 1007 is replaced by an amino acid residue with relatively large side chains. In some embodiments, the glycine residue at position 1007 is replaced by an arginine residue (G1007R). In some embodiments, the glycine residue at position 1007 is replaced by a lysine residue (G1007K). In some embodiments, the glycine residue at position 1007 is replaced by a tryptophan residue (G1007W). In some embodiments, the glycine residue at position 1007 is replaced by a tyrosine residue (G1007Y). Additionally, in other embodiments, the glycine residue at position 1007 is replaced by a leucine residue (G1007L). In other embodiments, the glycine residue at position 1007 is replaced by a threonine residue (G1007T). In other embodiments, the glycine residue at position 1007 is replaced by a serine residue (G1007S).

[0286] In some embodiments, the adenosine deaminase comprises a mutation at glutamic acid1008 of the hADAR1-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glutamic acid residue at position 1008 is replaced by a polar amino acid residue having a relatively large side chain. In some embodiments, the glutamic acid residue at position 1008 is replaced by a glutamine residue (E1008Q). In some embodiments, the glutamic acid residue at position 1008 is replaced by a histidine residue

(E1008H). In some embodiments, the glutamic acid residue at position 1008 is replaced by an arginine residue (E1008R). In some embodiments, the glutamic acid residue at position 1008 is replaced by a lysine residue (E1008K). In some embodiments, the glutamic acid residue at position 1008 is replaced by a nonpolar or small polar amino acid residue. In some embodiments, the glutamic acid residue at position 1008 is replaced by a phenylalanine residue (E1008F). In some embodiments, the glutamic acid residue at position 1008 is replaced by a tryptophan residue (E1008W). In some embodiments, the glutamic acid residue at position 1008 is replaced by a glycine residue (E1008G). In some embodiments, the glutamic acid residue at position 1008 is replaced by an isoleucine residue (E1008I). In some embodiments, the glutamic acid residue at position 1008 is replaced by a valine residue (E1008V). In some embodiments, the glutamic acid residue at position 1008 is replaced by a proline residue (E1008P). In some embodiments, the glutamic acid residue at position 1008 is replaced by a serine residue (E1008S). In other embodiments, the glutamic acid residue at position 1008 is replaced by an asparagine residue (E1008N). In other embodiments, the glutamic acid residue at position 1008 is replaced by an alanine residue (E1008A). In other embodiments, the glutamic acid residue at position 1008 is replaced by a Methionine residue (E1008M). In some embodiments, the glutamic acid residue at position 1008 is replaced by a leucine residue (E1008L).

[0287] In some embodiments, to improve editing efficiency, the adenosine deaminase may comprise one or more of the mutations: E1007S, E1007A, E1007V, E1008Q, E1008R, E1008H, E1008M, E1008N, E1008K, based on amino acid sequence positions of hADAR1-D, and mutations in a homologous ADAR protein corresponding to the above.

[0288] In some embodiments, to reduce editing efficiency, the adenosine deaminase may comprise one or more of the mutations: E1007R, E1007K, E1007Y, E1007L, E1007T, E1008G, E1008I, E1008P, E1008V, E1008F, E1008W, E1008S, E1008N, E1008K, based on amino acid sequence positions of hADAR1-D, and mutations in a homologous ADAR protein corresponding to the above.

[0289] In some embodiments, the substrate editing preference, efficiency and/or selectivity of an adenosine deaminase is affected by amino acid residues near or in the active center of the enzyme. In some embodiments, the adenosine deaminase comprises a mutation at the glutamic acid 1008 position in hADAR1-D sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the mutation is E1008R, or a corresponding mutation in a homologous ADAR protein. In some embodiments, the E1008R mutant has an increased

editing efficiency for target adenosine residue that has a mismatched G residue on the opposite strand.

[0290] In some embodiments, the adenosine deaminase protein further comprises or is connected to one or more double-stranded RNA (dsRNA) binding motifs (dsRBMs) or domains (dsRBDs) for recognizing and binding to double-stranded nucleic acid substrates. In some embodiments, the interaction between the adenosine deaminase and the double–stranded substrate is mediated by one or more additional protein factor(s), including a CRISPR/CAS protein factor. In some embodiments, the interaction between the adenosine deaminase and the double-stranded substrate is further mediated by one or more nucleic acid component(s), including a guide RNA.

Modified Adenosine Deaminase Having C-to U Deamination Activity

[0291] In certain example embodiments, directed evolution may be used to design modified ADAR proteins capable of catalyzing additional reactions besides deamination of an adenine to a hypoxanthine. For example, the modified ADAR protein may be capable of catalyzing deamination of a cytidine to a uracil. While not bound by a particular theory, mutations that improve C to U activity may alter the shape of the binding pocket to be more amenable to the smaller cytidine base.

[0292] In some embodiments, the modified adenosine deaminase having C-to-U deamination activity comprises a mutation at any one or more of positions V351, T375, R455, and E488 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the adenosine deaminase comprises mutation E488Q. In some embodiments, the adenosine deaminase comprises one or more of mutations selected from V351I, V351L, V351F, V351M, V351C, V351A, V351G, V351P, V351T, V351S, V351Y, V351W, V351Q, V351N, V351H, V351E, V351D, V351K, V351R, T375I, T375L, T375V, T375F, T375M, T375C, T375A, T375G, T375P, T375S, T375Y, T375W, T375Q, T375N, T375H, T375E, T375D, T375K, T375R, R455I, R455L, R455V, R455F, R455M, R455C, R455A, R455G, R455P, R455T, R455S, R455Y, R455W, R455Q, R455N, R455H, R455E, R455D, R455K. In some embodiments, the adenosine deaminase comprises mutation E488Q, and further comprises one or more of mutations selected from V351I, V351L, V351F, V351M, V351C, V351A, V351G, V351P, V351T, V351S, V351Y, V351W, V351Q, V351N, V351H, V351E, V351D, V351K, V351R, T375I, T375L, T375V, T375F, T375M, T375C, T375A, T375G, T375P, T375S, T375Y, T375W, T375Q, T375N, T375H, T375E, T375D, T375K, T375R, R455I, R455L, R455V, R455F, R455M, R455C, R455A, R455G, R455P, R455T, R455S, R455Y, R455W, R455Q, R455N, R455H, R455E, R455D, R455K.

[0293] In connection with the aforementioned modified ADAR protein having C-to-U deamination activity, the invention described herein also relates to a method for deaminating a C in a target RNA sequence of interest, comprising delivering to a target RNA or DNA an AD-functoinalized composition disclosed herein.

[0294] In certain example embodiments, the method for deaminating a C in a target RNA sequencecomprising delivering to said target RNA: (a) a catalytically inactive (dead) Cas; (b) a guide molecule which comprises a guide sequence linked to a direct repeat sequence; and (c) a modified ADAR protein having C-to-U deamination activity or catalytic domain thereof; wherein said modified ADAR protein or catalytic domain thereof is covalently or non-covalently linked to said dead Cas protein or said guide molecule or is adapted to link thereto after delivery; wherein guide molecule forms a complex with said dead Cas protein and directs said complex to bind said target RNA sequence of interest; wherein said guide sequence is capable of hybridizing with a target sequence comprising said C to form an RNA duplex; wherein, optionally, said guide sequence comprises a non-pairing A or U at a position corresponding to said C resulting in a mismatch in the RNA duplex formed; and wherein said modified ADAR protein or catalytic domain thereof deaminates said C in said RNA duplex.

[0295] In connection with the aforementioned modified ADAR protein having C-to-U deamination activity, the invention described herein further relates to an engineered, non-naturally occurring system suitable for deaminating a C in a target locus of interest, comprising: (a) a guide molecule which comprises a guide sequence linked to a direct repeat sequence, or a nucleotide sequence encoding said guide molecule; (b) a catalytically inactive Cas13 protein, or a nucleotide sequence encoding said catalytically inactive Cas13 protein; (c) a modified ADAR protein having C-to-U deamination activity or catalytic domain thereof, or a nucleotide sequence encoding said modified ADAR protein or catalytic domain thereof; wherein said modified ADAR protein or catalytic domain thereof is covalently or non-covalently linked to said Cas13 protein or said guide molecule or is adapted to link thereto after delivery; wherein said guide sequence is capable of hybridizing with a target RNA sequence comprising a C to form an RNA duplex; wherein, optionally, said guide sequence comprises a non-pairing A or U at a position corresponding to said C resulting in a mismatch in the RNA duplex formed; wherein, optionally, the system is a vector system comprising one or more vectors comprising: (a) a first regulatory element operably linked to a nucleotide sequence encoding said guide molecule which comprises said guide sequence, (b) a second regulatory element operably linked to a nucleotide sequence encoding said catalytically inactive Cas13 protein; and (c) a nucleotide sequence encoding a modified ADAR protein having C-to-U deamination activity

or catalytic domain thereof which is under control of said first or second regulatory element or operably linked to a third regulatory element; wherein, if said nucleotide sequence encoding a modified ADAR protein or catalytic domain thereof is operably linked to a third regulatory element, said modified ADAR protein or catalytic domain thereof is adapted to link to said guide molecule or said Cas13 protein after expression; wherein components (a), (b) and (c) are located on the same or different vectors of the system, optionally wherein said first, second, and/or third regulatory element is an inducible promoter.

[0296] According to the present invention, the substrate of the adenosine deaminase is an RNA/DNAn RNA duplex formed upon binding of the guide molecule to its DNA target which then forms the CRISPR-Cas complex with the CRISPR-Cas enzyme. The substrate of the adenosine deaminase can also be an RNA/RNA duplex formed upon binding of the guide molecule to its RNA target which then forms the CRISPR-Cas complex with the CRISPR-Cas enzyme. The RNA/DNA or DNA/RNAn RNA duplex is also referred to herein as the "RNA/DNA hybrid", "DNA/RNA hybrid" or "double-stranded substrate". The particular features of the guide molecule and CRISPR-Cas enzyme are detailed below.

[0297] The term "editing selectivity" as used herein refers to the fraction of all sites on a double-stranded substrate that is edited by an adenosine deaminase. Without being bound by theory, it is contemplated that editing selectivity of an adenosine deaminase is affected by the double-stranded substrate's length and secondary structures, such as the presence of mismatched bases, bulges and/or internal loops.

[0298] In some embodiments, when the substrate is a perfectly base-paired duplex longer than 50 bp, the adenosine deaminase may be able to deaminate multiple adenosine residues within the duplex (e.g., 50% of all adenosine residues). In some embodiments, when the substrate is shorter than 50 bp, the editing selectivity of an adenosine deaminase is affected by the presence of a mismatch at the target adenosine site. Particularly, in some embodiments, adenosine (A) residue having a mismatched cytidine (C) residue on the opposite strand is deaminated with high efficiency. In some embodiments, adenosine (A) residue having a mismatched guanosine (G) residue on the opposite strand is skipped without editing.

TARGETING DOMAIN

[0299] The methods, tools, and compositions of the invention comprise or make use of a targeting component which can be referred to as a targeting domain. The targeting domain is preferably a DNA or RNA targeting domain, more particularly an oligonucleotide targeting domain, or a variant or fragment theofe which retains DNA and/or RNA binding activity. The oligonucleotide targeting domain may bind a sequence, motif, or structural feature of the RNA

or DNA of interest at or adajacent to the target locus. A structural feature may include hairpins, tetraloops, or other secondary structural features of a nucleic acid. As used herein "adjacent" means within a distance and/or orientation of the target locus in which the adenosine deaminase can complete its base editing function. In certain example embodiments, the oligonucleotide binding protein may be a RNA-binding protein or functional domain thereof, or a DNA-binding protein or functional domain thereof.

[0300] In particular embodiments, the targeting domain further comprises a guide RNA (as will be detailed below). The nucleic acid binding protein can be an (endo)nuclease or any other (oligo)nucleotide binding protein. In particular embodiments, the nucleotide binding protein is modified to inactivate any other function not required for said DNA or RNA binding. In particular embodiments, where the nucleotide binding protein is an (endo)nuclease, preferably the (endo)nuclease has altered or modified activity (i.e. a modified nuclease, as described herein elsewhere) compared to the wildtype DNA or RNA binding protein. In certain embodiments, said nuclease is a targeted or site-specific or homing nuclease or a variant thereof having altered or modified activity. In certain embodiments, said (oligo)nucleotide binding protein is the (oligo)nucleotide binding domain of said (oligo)nucleotide binding protein and does not comprise one or more domains of said protein not required for DNA and/or RNA binding (more particular does not comprise one or more other functional domains).

RNA-binding proteins

[0301] In certain example embodiments, the oligonucleotide binding domain may comprise or consist of a RNA-binding protein, or functional domain thereof, that comprises a RNA recognition motif. Example RNA-binding proteins comprising a RNA recognition motif include, but are not limited to, A2BP1; ACF; BOLL; BRUNOL4; BRUNOL5; BRUNOL6; CCBL2; CGI96; CIRBP; CNOT 4; CPEB2; CPEB3; CPEB4; CPSF7; CSTF2; CSTF2T; CUGBP1; CUGBP2; D10S102; DAZ 1; DAZ2; DAZ3; DAZ4; DAZAP1; DAZL; DNAJC17; DND1; EIF3S4; EIF3S9; EIF4B; EI F4H; ELAVL1; ELAVL2; ELAVL3; ELAVL4; ENOX1; ENOX2; EWSR1; FUS; FUSIP1; G3BP; G3BP1; G3BP2; GRSF1; HNRNPL; HNRPA0; HNRPA1; HNRPA2B1; HNRPA3; H NRPAB; HNRPC; HNRPCL1; HNRPD; HNRPDL; HNRPF; HNRPH1; HNRPH2; HNRPH 3; HNRPL; HNRPLL; HNRPM; HNRPR; HRNBP1; HSU53209; HTATSF1; IGF2BP1; IGF 2BP2; IGF2BP3; LARP7; MKI67IP; MSI1; MSI2; MSSP2; MTHFSD; MYEF2; NCBP2; N CL; NOL8; NONO; P14; PABPC1; PABPC1L; PABPC3; PABPC4; PABPC5; PABPN1; PO LDIP3; PPARGC1; PPARGC1A; PPARGC1B; PPIE; PPIL4; PPRC1; PSPC1; PTBP1; PTB P2; PUF60; RALY; RALYL; RAVER1; RAVER2; RBM10; RBM11; RBM12; RBM12B; R

BM14; RBM15; RBM15B; RBM16; RBM17; RBM18; RBM19; RBM22; RBM23; RBM24; RBM25; RBM26; RBM27; RBM28; RBM3; RBM32B; RBM33; RBM34; RBM35A; RBM35B; RBM38; RBM39; RBM4; RBM41; RBM42; RBM44; RBM45; RBM46; RBM47; RBM4B; RBM5; RBM7; RBM8A; RBM9; RBMS1; RBMS2; RBMS3; RBMX; RBMX2; RBMXL2; RBMY1A1; RBMY1B; RBMY1E; RBMY1F; RBMY2FP; RBPMS; RBPMS2; RDBP; RNPC3; RNPC4; RNPS1; ROD1; SAFB; SAFB2; SART3; SETD1A; SF3B14; SF3B4; SFPQ; SFRS1; SFRS10; SFRS11; SFRS12; SFRS15; SFRS2; SFRS2B; SFRS3; SFRS4; SFRS5; SFRS6; SFRS7; SFRS9; SLIRP; SLTM; SNRP70; SNRPA; SNRPB2; SPEN; SR140; SRRP35; SSB; SYNCRIP; TAF15; TARDBP; THOC4; TIA1; TIAL1; TNRC4; TNRC6C; TRA2A; TRSPAP1; TUT1; U1SNRNPBP; U2AF1; U2AF2; UHMK1; ZCRB1; ZNF638; ZRSR1; and ZRSR2.

[0302] In certain example embodiments, the RNA-binding protein or function domain thereof may comprise a K homology domain. Example RNA-binding proteins comprising a K homology domain include, but are not limited to, AKAP1; ANKHD1; ANKRD17; ASCC1; BICC1; DDX43; DDX53; DPPA5; FMR1; FUBP1; FUBP3; FXR1; FXR2; GLD1; HDLBP; HNRPK; IGF2BP1; IGF2BP2; IGF2BP3; KHDRBS1; KHDRBS2; KHDRBS3; KHSRP; KRR1; MEX3A; MEX3B; MEX3C; MEX3D; NOVA1; NOVA2; PCBP1; PCBP2; PCBP3; PCBP4; PNO1; PNPT1; QKI; SF1; and TDRKH

[0303] In certain example embodiments, the RNA-binding protein comprises a zinc finger motif. RNA-binding proteins or functional domains thereof may comprise a Cys2-His2, Gag-knuckle, Treble-clet, Zinc ribbon, Zn2/Cys6 class motif.

[0304] In certain example embodiments, the RNA-binding protein may comprise a Pumilio homology domain.

TALENS

[0305] In certain embodiments, the nucleic acid binding protein is a (modified) transcription activator-like effector nuclease (TALEN) system. Transcription activator-like effectors (TALEs) can be engineered to bind practically any desired DNA sequence. Exemplary methods of genome editing using the TALEN system can be found for example in Cermak T. Doyle EL. Christian M. Wang L. Zhang Y. Schmidt C, et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. 2011;39:e82; Zhang F. Cong L. Lodato S. Kosuri S. Church GM. Arlotta P Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol. 2011;29:149–153 and US Patent Nos. 8,450,471, 8,440,431 and 8,440,432, all of which are specifically incorporated by reference. By means of further

guidance, and without limitation, naturally occurring TALEs or "wild type TALEs" are nucleic acid binding proteins secreted by numerous species of proteobacteria. TALE polypeptides contain a nucleic acid binding domain composed of tandem repeats of highly conserved monomer polypeptides that are predominantly 33, 34 or 35 amino acids in length and that differ from each other mainly in amino acid positions 12 and 13. In advantageous embodiments the nucleic acid is DNA. As used herein, the term "polypeptide monomers", or "TALE monomers" will be used to refer to the highly conserved repetitive polypeptide sequences within the TALE nucleic acid binding domain and the term "repeat variable di-residues" or "RVD" will be used to refer to the highly variable amino acids at positions 12 and 13 of the polypeptide monomers. As provided throughout the disclosure, the amino acid residues of the RVD are depicted using the IUPAC single letter code for amino acids. A general representation of a TALE monomer which is comprised within the DNA binding domain is X1-11-(X12X13)-X14-33 or 34 or 35, where the subscript indicates the amino acid position and X represents any amino acid. X12X13 indicate the RVDs. In some polypeptide monomers, the variable amino acid at position 13 is missing or absent and in such polypeptide monomers, the RVD consists of a single amino acid. In such cases the RVD may be alternatively represented as X*, where X represents X12 and (*) indicates that X13 is absent. The DNA binding domain comprises several repeats of TALE monomers and this may be represented as (X1-11-(X12X13)-X14-33 or 34 or 35)z, where in an advantageous embodiment, z is at least 5 to 40. In a further advantageous embodiment, z is at least 10 to 26. The TALE monomers have a nucleotide binding affinity that is determined by the identity of the amino acids in its RVD. For example, polypeptide monomers with an RVD of NI preferentially bind to adenine (A), polypeptide monomers with an RVD of NG preferentially bind to thymine (T), polypeptide monomers with an RVD of HD preferentially bind to cytosine (C) and polypeptide monomers with an RVD of NN preferentially bind to both adenine (A) and guanine (G). In yet another embodiment of the invention, polypeptide monomers with an RVD of IG preferentially bind to T. Thus, the number and order of the polypeptide monomer repeats in the nucleic acid binding domain of a TALE determines its nucleic acid target specificity. In still further embodiments of the invention, polypeptide monomers with an RVD of NS recognize all four base pairs and may bind to A, T, G or C. The structure and function of TALEs is further described in, for example, Moscou et al., Science 326:1501 (2009); Boch et al., Science 326:1509-1512 (2009); and Zhang et al., Nature Biotechnology 29:149-153 (2011), each of which is incorporated by reference in its entirety. In certain embodiments, targeting is effected by a polynucleic acid binding TALEN fragment.

In certain embodiments, the targeting domain comprises or consists of a catalytically inactive TALEN or nucleic acid binding fragment thereof.

Zn-Finger Nucleases

[0306] In certain embodiments, the targeting domain comprises or consists of a (modified) zinc-finger nuclease (ZFN) system. The ZFN system uses artificial restriction enzymes generated by fusing a zinc finger DNA-binding domain to a DNA-cleavage domain that can be engineered to target desired DNA sequences. Exemplary methods of genome editing using ZFNs can be found for example in U.S. Patent Nos. 6,534,261, 6,607,882, 6,746,838, 6,794,136, 6,824,978, 6,866,997, 6,933,113, 6,979,539, 7,013,219, 7,030,215, 7,220,719, 7,241,573, 7,241,574, 7,585,849, 7,595,376, 6,903,185, and 6,479,626, all of which are specifically incorporated by reference. By means of further guidance, and without limitation, artificial zinc-finger (ZF) technology involves arrays of ZF modules to target new DNA-binding sites in the genome. Each finger module in a ZF array targets three DNA bases. A customized array of individual zinc finger domains is assembled into a ZF protein (ZFP). ZFPs can comprise a functional domain. The first synthetic zinc finger nucleases (ZFNs) were developed by fusing a ZF protein to the catalytic domain of the Type IIS restriction enzyme FokI. (Kim, Y. G. et al., 1994, Chimeric restriction endonuclease, Proc. Natl. Acad. Sci. U.S.A. 91, 883−887; Kim, Y. G. et al., 1996, Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc. Natl. Acad. Sci. U.S.A. 93, 1156−1160). Increased cleavage specificity can be attained with decreased off target activity by use of paired ZFN heterodimers, each targeting different nucleotide sequences separated by a short spacer. (Doyon, Y. et al., 2011, Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. Nat. Methods 8, 74−79). ZFPs can also be designed as transcription activators and repressors and have been used to target many genes in a wide variety of organisms. In certain embodiments, the targeting domain comprises or consists of a nucleic acid binding zinc finger nuclease or a nucleic acid binding fragment thereof. In certain embodiments, the nucleic acid binding (fragment of) a zinc finger nuclease is catalytically inactive.

Meganuclease

[0307] In certain embodiments, the targeting domain comprises a (modified) meganuclease, which are endodeoxyribonucleases characterized by a large recognition site (double-stranded DNA sequences of 12 to 40 base pairs). Exemplary method for using meganucleases can be found in US Patent Nos: 8,163,514; 8,133,697; 8,021,867; 8,119,361; 8,119,381; 8,124,369; and 8,129,134, which are specifically incorporated by reference. In certain embodiments, targeting is effected by a polynucleic acid binding meganuclease

fragment. In certain embodiments, targeting is effected by a polynucleic acid binding catalytically inactive meganuclease (fragment). Accordingly in particular embodiments, the targeting domain comprises or consists of a nucleic acid binding meganuclease or a nucleic acid binding fragment thereof.

CRISPR-Cas Systems

[0308] In certain embodiments, the targeting domain comprises a (modified) CRISPR/Cas complex or system. General information on CRISPR/Cas Systems, components thereof, and delivery of such components, including methods, materials, delivery vehicles, vectors, particles, and making and using thereof, including as to amounts and formulations, as well as CRISPR/Cas-expressing eukaryotic cells, CRISPR/Cas expressing eukaryotes, such as a mouse, is described herein elsewhere. In certain embodiments, targeting is effected by an oligonucleic acid binding CRISPR protein fragment and/or a gRNA. In certain embodiments, targeting is effected by a nucleic acid binding catalytically inactive CRISPR protein (fragment). Accordingly in particular embodiments, the targeting domain comprises oligonucleic acid binding CRISPR protein or an oligonucleic acid binding fragment of a CRISPR protein and/or a gRNA.

[0309] As used herein, the term "Cas" generally refers to a (modified) effector protein of the CRISPR/Cas system or complex, and can be without limitation a (modified) Cas9, or other enzymes such as Cpf1, C2c1, C2c2, C2c3, group 29, or group 30 protein The term "Cas" may be used herein interchangeably with the terms "CRISPR" protein, "CRISPR/Cas protein", "CRISPR effector", "CRISPR/Cas effector", "CRISPR enzyme", "CRISPR/Cas enzyme" and the like, unless otherwise apparent, such as by specific and exclusive reference to Cas9. It is to be understood that the term "CRISPR protein" may be used interchangeably with "CRISPR enzyme", irrespective of whether the CRISPR protein has altered, such as increased or decreased (or no) enzymatic activity, compared to the wild type CRISPR protein. Likewise, as used herein, in certain embodiments, where appropriate and which will be apparent to the skilled person, the term "nuclease" may refer to a modified nuclease wherein catalytic activity has been altered, such as having increased or decreased nuclease activity, or no nuclease activity at all, as well as nickase activity, as well as otherwise modified nuclease as defined herein elsewhere, unless otherwise apparent, such as by specific and exclusive reference to unmodified nuclease.

[0310] In some embodiments, the CRISPR effector protein is Cas9, Cpf1, C2c1, C2c2, or Cas13a, Cas13b, Cas13c, or Cas13d. In some embodiments, the CRISPR effector protein is a DNA-targeting CRISPR effector protein. In some embodiments, the CRISPR effector protein

is a Type-II CRISPR effector protein such as Cas9. In some embodiments, the CRISPR effector protein is a Type-V CRISPR effector protein such as Cpf1 or C2c1. In some embodiments, the CRISPR effector protein is a RNA-targeting CRISPR effector protein. In some embodiments, the CRISPR effector protein is a Type-VI CRISPR effector protein such as Cas13a, Cas13b, Cas13c, or Cas13d.

[0311] In some embodiments, the CRISPR effector protein is a Cas9, for instance SaCas9, SpCas9, StCas9, CjCas9 and so forth – any ortholog is envisaged. In some embodiments, the CRISPR effector protein is a Cpf1, for instance AsCpf1, LbCpf1, FnCpf1 and so forth – any ortholog is envisaged.In certain embodiments, the targeting component as described herein according to the invention is a (endo)nuclease or a variant thereof having altered or modified activity (i.e. a modified nuclease, as described herein elsewhere). In certain embodiments, said nuclease is a targeted or site-specific or homing nuclease or a variant thereof having altered or modified activity. In certain embodiments, said nuclease or targeted/site-specific/homing nuclease is, comprises, consists essentially of, or consists of a (modified) CRISPR/Cas system or complex, a (modified) Cas protein, a (modified) zinc finger, a (modified) zinc finger nuclease (ZFN), a (modified) transcription factor-like effector (TALE), a (modified) transcription factor-like effector nuclease (TALEN), or a (modified) meganuclease. In certain embodiments, said (modified) nuclease or targeted/site-specific/homing nuclease is, comprises, consists essentially of, or consists of a (modified) RNA-guided nuclease.

[0312] In particular embodiments, more particularly where the nuclease is a CRISPR protein, the targeting domain further comprises a guide molecule which targets a selected nucleic acid. For instance, in the context of the CRISPR/Cas system, the guide RNA is capable of hybridizing with a selected nucleic acid sequence. As uses herein, "hybridization" or “hybridizing” refers to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding may occur by Watson Crick base pairing, Hoogstein binding, or in any other sequence specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi stranded complex, a single self-hybridizing strand, or any combination of these. A hybridization reaction may constitute a step in a more extensive process, such as the initiation of PGR, or the cleavage of a polynucleotide by an enzyme. A sequence capable of hybridizing with a given sequence is referred to as the "complement" of the given sequence

[0313] In the methods and systems of the present invention use is made of a CRISPR-Cas protein and corresponding guide molecule. More particularly, the CRISPR-Cas protein is a

class 2 CRISPR-Cas protein. In certain embodiments, said CRISPR-Cas protein is a Cas13. The CRISPR-Cas system does not require the generation of customized proteins to target specific sequences but rather a single Cas protein can be programmed by guide molecule to recognize a specific nucleic acid target, in other words the Cas enzyme protein can be recruited to a specific nucleic acid target locus of interest using said guide molecule.

[0314] The term "AD-functionalized CRISPR system" as used here refers to a nucleic acid targeting and editing system comprising (a) a CRISPR-Cas protein, more particularly a Cas13 protein which is catalytically inactive; (b) a guide molecule which comprises a guide sequence; and (c) an adenosine deaminase protein or catalytic domain thereof; wherein the adenosine deaminase protein or catalytic domain thereof is covalently or non-covalently linked to the CRISPR-Cas protein or the guide molecule or is adapted to link thereto after delivery; wherein the guide sequence is substantially complementary to the target sequence but comprises a non-pairing C corresponding to the A being targeted for deamination, resulting in an A-C mismatch in an RNA duplex formed by the guide sequence and the target sequence. For application in eukaryotic cells, the CRISPR-Cas protein and/or the adenosine deaminase are preferably NLS-tagged.

[0315] In particular embodiments, the targeting domain is a CRISPR-cas protein. In certain example embodiments, the CRISPR-cas protein is linked to the deaminase protein or its catalytic domain by means of an LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR (SEQ ID No. 11) linker. In further particular embodiments, the CRISPR-Cas protein is linked C-terminally to the N-terminus of a deaminase protein or its catalytic domain by means of an LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR (SEQ ID No. 11) linker. In addition, N- and C-terminal NLSs can also function as linker (e.g., PKKKRKVEASSPKKRKVEAS (SEQ ID No. 16)). In particular embodiments of the methods of the present invention, the adenosine deaminase protein or catalytic domain thereof is delivered to the cell or expressed within the cell as a separate protein, but is modified so as to be able to link to the targeting domain or the guide molecule. In those embodiments in which the targeting domain is a CRISPR-Cas system, the adenosine deaminase may link to either the Cas protein or the guide moledule. In particular embodiments, this is ensured by the use of orthogonal RNA-binding protein or adaptor protein / aptamer combinations that exist within the diversity of bacteriophage coat proteins. Examples of such coat proteins include but are not limited to: MS2, Qβ, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕCb5, ϕCb8r, ϕCb12r, ϕCb23r, 7s and PRR1. Aptamers can be naturally occurring or synthetic

oligonucleotides that have been engineered through repeated rounds of in vitro selection or SELEX (systematic evolution of ligands by exponential enrichment) to bind to a specific target.

[0316] In particular embodiments of the methods and systems of the present invention, the guide molecule is provided with one or more distinct RNA loop(s) or disctinct sequence(s) that can recruit an adaptor protein. For example, a guide molecule may be extended without colliding with the Cas protein by the insertion of distinct RNA loop(s) or distinct sequence(s) that may recruit adaptor proteins that can bind to the distinct RNA loop(s) or distinct sequence(s). Examples of modified guides and their use in recruiting effector domains to the CRISPR-Cas complex are provided in Konermann (Nature 2015, 517(7536): 583–588). In particular embodiments, the aptamer is a minimal hairpin aptamer which selectively binds dimerized MS2 bacteriophage coat proteins in mammalian cells and is introduced into the guide molecule, such as in the stemloop and/or in a tetraloop. In these embodiments, the adenosine deaminase protein is fused to MS2. The adenosine deaminase protein is then co-delivered together with the CRISPR-Cas protein and corresponding guide RNA.

[0317] In some embodiments, the components (a), (b) and (c) are delivered to the cell as a ribonucleoprotein complex. The ribonucleoprotein complex can be delivered via one or more lipid nanoparticles.

[0318] In some embodiments, the components (a), (b) and (c) are delivered to the cell as one or more RNA molecules, such as one or more guide RNAs and one or more mRNA molecules encoding the CRISPR-Cas protein, the adenosine deaminase protein, and optionally the adaptor protein. The RNA molecules can be delivered via one or more lipid nanoparticles.

[0319] In some embodiments, the components (a), (b) and (c) are delivered to the cell as one or more DNA molecules. In some embodiments, the one or more DNA molecules are comprised within one or more vectors such as viral vectors (e.g., AAV). In some embodiments, the one or more DNA molecules comprise one or more regulatory elements operably configured to express the CRISPR-Cas protein, the guide molecule, and the adenosine deaminase protein or catalytic domain thereof, optionally wherein the one or more regulatory elements comprise inducible promoters.

[0320] In some embodiments, the CRISPR-Cas protein is a dead Cas13. In some embodiments, the dead Cas13 is a dead Cas13a protein which comprises one or more mutations in the HEPN domain. In some embodiments, the dead Cas13a comprises a mutation corresponding to R474A and R1046A in Leptotrichia wadei (LwaCas13a). In some embodiments, the dead Cas13 is a dead Cas13b protein which comprises one or more of

R116A, H121A, R1177A, H1182A of a Cas13b protein originating from Bergeyella zoohelcum ATCC 43767 or amino acid positions corresponding thereto of a Cas13b ortholog.

[0321] In some embodiments of the guide molecule is capable of hybridizing with a target sequence comprising the Adenine to be deaminated within an RNA sequence to form an RNA duplex which comprises a non-pairing Cytosine opposite to said Adenine. Upon RNA duplex formation, the guide molecule forms a complex with the Cas13 protein and directs the complex to bind the RNA polynucleotide at the target RNA sequence of interest. Details on the aspect of the guide of the AD-functionalized CRISPR-Cas system are provided herein below.

[0322] In some embodiments, a Cas13 guide RNA having a canonical length of, e.g. LawCas13 is used to form an RNA duplex with the target DNA. In some embodiments, a Cas13 guide molecule longer than the canonical length for, e.g. LawCas13a is used to form an RNA duplex with the target DNA including outside of the Cas13-guide RNA-target DNA complex.

[0323] In at least a first design, the AD-functionalized CRISPR system comprises (a) an adenosine deaminase fused or linked to a CRISPR-Cas protein, wherein the CRISPR-Cas protein is catalytically inactive, and (b) a guide molecule comprising a guide sequence designed to introduce an A-C mismatch in an RNA duplex formed between the guide sequence and the target sequence. In some embodiments, the CRISPR-Cas protein and/or the adenosine deaminase are NLS-tagged, on either the N- or C-terminus or both.

[0324] In at least a second design, the AD-functionalized CRISPR system comprises (a) a CRISPR-Cas protein that is catalytically inactive, (b) a guide molecule comprising a guide sequence designed to introduce an A-C mismatch in an RNA duplex formed between the guide sequence and the target sequence, and an aptamer sequence (e.g., MS2 RNA motif or PP7 RNA motif) capable of binding to an adaptor protein (e.g., MS2 coating protein or PP7 coat protein), and (c) an adenosine deaminase fused or linked to an adaptor protein, wherein the binding of the aptamer and the adaptor protein recruits the adenosine deaminase to the RNA duplex formed between the guide sequence and the target sequence for targeted deamination at the A of the A-C mismatch. In some embodiments, the adaptor protein and/or the adenosine deaminase are NLS-tagged, on either the N- or C-terminus or both. The CRISPR-Cas protein can also be NLS-tagged.

[0325] The use of different aptamers and corresponding adaptor proteins also allows orthogonal gene editing to be implemented. In one example in which adenosine deaminase are used in combination with cytidine deaminase for orthogonal gene editing/deamination, sgRNA targeting different loci are modified with distinct RNA loops in order to recruit MS2-adenosine deaminase and PP7-cytidine deaminase (or PP7-adenosine deaminase and MS2-cytidine

deaminase), respectively, resulting in orthogonal deamination of A or C at the target loci of interested, respectively. PP7 is the RNA-binding coat protein of the bacteriophage Pseudomonas. Like MS2, it binds a specific RNA sequence and secondary structure. The PP7 RNA-recognition motif is distinct from that of MS2. Consequently, PP7 and MS2 can be multiplexed to mediate distinct effects at different genomic loci simultaneously. For example, an sgRNA targeting locus A can be modified with MS2 loops, recruiting MS2-adenosine deaminase, while another sgRNA targeting locus B can be modified with PP7 loops, recruiting PP7-cytidine deaminase. In the same cell, orthogonal, locus-specific modifications are thus realized. This principle can be extended to incorporate other orthogonal RNA-binding proteins.

[0326] In at least a third design, the AD-functionalized CRISPR system comprises (a) an adenosine deaminase inserted into an internal loop or unstructured region of a CRISPR-Cas protein, wherein the CRISPR-Cas protein is catalytically inactive or a nickase, and (b) a guide molecule comprising a guide sequence designed to introduce an A-C mismatch in an RNA duplex formed between the guide sequence and the target sequence.

[0327] CRISPR-Cas protein split sites that are suitable for inseration of adenosine deaminase can be identified with the help of a crystal structure. One can use the crystal structure of an ortholog if a relatively high degree of homology exists between the ortholog and the intended CRISPR-Cas protein.

[0328] The split position may be located within a region or loop. Preferably, the split position occurs where an interruption of the amino acid sequence does not result in the partial or full destruction of a structural feature (e.g. alpha-helixes or β-sheets). Unstructured regions (regions that did not show up in the crystal structure because these regions are not structured enough to be "frozen" in a crystal) are often preferred options. The positions within the unstructured regions or outside loops may not need to be exactly the numbers provided above, but may vary by, for example 1, 2, 3, 4, 5, 6, 7, 8, 9, or even 10 amino acids either side of the position given above, depending on the size of the loop, so long as the split position still falls within an unstructured region of outside loop.

[0329] The AD-functionalized CRISPR system described herein can be used to target a specific Adenine or Cytidine within an RNA polynucleotide sequence for deamination. For example, the guide molecule can form a complex with the CRISPR-Cas protein and directs the complex to bind a target RNA sequence in the RNA polynucleotide of interest. In certain example embodiments, because the guide sequence is designed to have a non-pairing C, the RNA duplex formed between the guide sequence and the target sequence comprises an A-C mismatch, which directs the adenosine deaminase to contact and deaminate the A opposite to

the non-pairing C, converting it to a Inosine (I). Since Inosine (I) base pairs with C and functions like G in cellular process, the targeted deamination of A described herein are useful for correction of undesirable G-A and C-T mutations, as well as for obtaining desirable A-G and T-C mutations.

[0330] In some embodiments, the AD-functionalized CRISPR system is used for targeted deamination in an RNA polynucleotide molecule in vitro. In some embodiments, the AD-functionalized CRISPR system is used for targeted deamination in a DNA molecule within a cell. The cell can be a eukaryotic cell, such as a animal cell, a mammalian cell, a human, or a plant cell.

Guide molecule

[0331] The guide molecule or guide RNA of a Class 2 type V CRISPR-Cas protein comprises a tracr-mate sequence (encompassing a "direct repeat" in the context of an endogenous CRISPR system) and a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system). Indeed, in contrast to the type II CRISPR-Cas proteins, the Cas13 protein does not rely on the presence of a tracr sequence. In some embodiments, the CRISPR-Cas system or complex as described herein does not comprise and/or does not rely on the presence of a tracr sequence (e.g. if the Cas protein is Cas13). In certain embodiments, the guide molecule may comprise, consist essentially of, or consist of a direct repeat sequence fused or linked to a guide sequence or spacer sequence.

[0332] In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence. In the context of formation of a CRISPR complex, "target sequence" refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target DNA sequence and a guide sequence promotes the formation of a CRISPR complex.

[0333] The terms "guide molecule" and "guide RNA" are used interchangeably herein to refer to RNA-based molecules that are capable of forming a complex with a CRISPR-Cas protein and comprises a guide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of the complex to the target nucleic acid sequence. The guide molecule or guide RNA specifically encompasses RNA-based molecules having one or more chemically modifications (e.g., by chemical linking two ribonucleotides or by replacement of one or more ribonucleotides with one or more deoxyribonucleotides), as described herein.

[0334] As used herein, the term "crRNA" or "guide RNA" or "single guide RNA" or "sgRNA" or "one or more nucleic acid components" of a Type V or Type VI CRISPR-Cas

locus effector protein comprises any polynucleotide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target nucleic acid sequence. In some embodiments, the degree of complementarity, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g., the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). The ability of a guide sequence (within a nucleic acid-targeting guide RNA) to direct sequence-specific binding of a nucleic acid-targeting complex to a target nucleic acid sequence may be assessed by any suitable assay. For example, the components of a nucleic acid-targeting CRISPR system sufficient to form a nucleic acid-targeting complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target nucleic acid sequence, such as by transfection with vectors encoding the components of the nucleic acid-targeting complex, followed by an assessment of preferential targeting (e.g., cleavage) within the target nucleic acid sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target nucleic acid sequence may be evaluated in a test tube by providing the target nucleic acid sequence, components of a nucleic acid-targeting complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art. A guide sequence, and hence a nucleic acid-targeting guide may be selected to target any target nucleic acid sequence. The target sequence may be DNA. The target sequence may be any RNA sequence. In some embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of messenger RNA (mRNA), pre-mRNA, ribosomal RNA (rRNA), transfer RNA (tRNA), micro-RNA (miRNA), small interfering RNA (siRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), double stranded RNA (dsRNA), non-coding RNA (ncRNA), long non-coding RNA (lncRNA), and small cytoplasmatic RNA (scRNA). In some preferred embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of mRNA, pre-mRNA, and rRNA. In some preferred embodiments, the target sequence may be a

sequence within a RNA molecule selected from the group consisting of ncRNA, and lncRNA. In some more preferred embodiments, the target sequence may be a sequence within an mRNA molecule or a pre-mRNA molecule.

[0335] In some embodiments, the guide molecule comprises a guide sequence that is designed to have at least one mismatch with the taret sequence, such that an RNA duplex formed between the guide sequence and the target sequence comprises a non-pairing C in the guide sequence opposite to the target A for deamination on the target sequence. In some embodiments, aside from this A-C mismatch, the degree of complementarity, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more.

[0336] As used herein, the term "crRNA" or "guide RNA" or "single guide RNA" or "sgRNA" or "one or more nucleic acid components" of a Type V or Type VI CRISPR-Cas locus effector protein comprises any polynucleotide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target nucleic acid sequence. In some embodiments, the degree of complementarity, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g., the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). The ability of a guide sequence (within a nucleic acid-targeting guide RNA) to direct sequence-specific binding of a nucleic acid-targeting complex to a target nucleic acid sequence may be assessed by any suitable assay. For example, the components of a nucleic acid-targeting CRISPR system sufficient to form a nucleic acid-targeting complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target nucleic acid sequence, such as by transfection with vectors encoding the components of the nucleic acid-targeting complex, followed by an assessment of preferential targeting (e.g., cleavage) within the target nucleic acid sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target nucleic acid sequence may be evaluated in a test tube by providing the target nucleic acid sequence, components of a nucleic acid-targeting complex, including the guide sequence to be tested and a control guide sequence different from

the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art. A guide sequence, and hence a nucleic acid-targeting guide may be selected to target any target nucleic acid sequence. The target sequence may be DNA. The target sequence may be any RNA sequence. In some embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of messenger RNA (mRNA), pre-mRNA, ribosomal RNA (rRNA), transfer RNA (tRNA), micro-RNA (miRNA), small interfering RNA (siRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), double stranded RNA (dsRNA), non-coding RNA (ncRNA), long non-coding RNA (lncRNA), and small cytoplasmatic RNA (scRNA). In some preferred embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of mRNA, pre-mRNA, and rRNA. In some preferred embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of ncRNA, and lncRNA. In some more preferred embodiments, the target sequence may be a sequence within an mRNA molecule or a pre-mRNA molecule.

[0337] In some embodiments, a nucleic acid-targeting guide is selected to reduce the degree secondary structure within the nucleic acid-targeting guide. In some embodiments, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of the nucleotides of the nucleic acid-targeting guide participate in self-complementary base pairing when optimally folded. Optimal folding may be determined by any suitable polynucleotide folding algorithm. Some programs are based on calculating the minimal Gibbs free energy. An example of one such algorithm is mFold, as described by Zuker and Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Another example folding algorithm is the online webserver RNAfold, developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (see e.g., A.R. Gruber et al., 2008, Cell 106(1): 23-24; and PA Carr and GM Church, 2009, Nature Biotechnology 27(12): 1151-62).

[0338] In certain embodiments, a guide RNA or crRNA may comprise, consist essentially of, or consist of a direct repeat (DR) sequence and a guide sequence or spacer sequence. In certain embodiments, the guide RNA or crRNA may comprise, consist essentially of, or consist of a direct repeat sequence fused or linked to a guide sequence or spacer sequence. In certain embodiments, the direct repeat sequence may be located upstream (i.e., 5') from the guide sequence or spacer sequence. In other embodiments, the direct repeat sequence may be located downstream (i.e., 3') from the guide sequence or spacer sequence.

[0339] In certain embodiments, the crRNA comprises a stem loop, preferably a single stem loop. In certain embodiments, the direct repeat sequence forms a stem loop, preferably a single stem loop.

[0340] In certain embodiments, the spacer length of the guide RNA is from 15 to 35 nt. In certain embodiments, the spacer length of the guide RNA is at least 15 nucleotides. In certain embodiments, the spacer length is from 15 to 17 nt, e.g., 15, 16, or 17 nt, from 17 to 20 nt, e.g., 17, 18, 19, or 20 nt, from 20 to 24 nt, e.g., 20, 21, 22, 23, or 24 nt, from 23 to 25 nt, e.g., 23, 24, or 25 nt, from 24 to 27 nt, e.g., 24, 25, 26, or 27 nt, from 27-30 nt, e.g., 27, 28, 29, or 30 nt, from 30-35 nt, e.g., 30, 31, 32, 33, 34, or 35 nt, or 35 nt or longer.

[0341] The "tracrRNA" sequence or analogous terms includes any polynucleotide sequence that has sufficient complementarity with a crRNA sequence to hybridize. In some embodiments, the degree of complementarity between the tracrRNA sequence and crRNA sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher. In some embodiments, the tracr sequence is about or more than about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or more nucleotides in length. In some embodiments, the tracr sequence and crRNA sequence are contained within a single transcript, such that hybridization between the two produces a transcript having a secondary structure, such as a hairpin. In an embodiment of the invention, the transcript or transcribed polynucleotide sequence has at least two or more hairpins. In preferred embodiments, the transcript has two, three, four or five hairpins. In a further embodiment of the invention, the transcript has at most five hairpins. In a hairpin structure the portion of the sequence 5' of the final "N" and upstream of the loop corresponds to the tracr mate sequence, and the portion of the sequence 3' of the loop corresponds to the tracr sequence.

[0342] In general, degree of complementarity is with reference to the optimal alignment of the sca sequence and tracr sequence, along the length of the shorter of the two sequences. Optimal alignment may be determined by any suitable alignment algorithm, and may further account for secondary structures, such as self-complementarity within either the sca sequence or tracr sequence. In some embodiments, the degree of complementarity between the tracr sequence and sca sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher.

[0343] In general, the CRISPR-Cas or CRISPR system may be as used in the foregoing documents, such as WO 2014/093622 (PCT/US2013/074667) and refers collectively to

transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, in particular a Cas13 gene in the case of CRISPR-Cas13, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence (encompassing a "direct repeat" and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system), or "RNA(s)" as that term is herein used (e.g., RNA(s) to guide Cas13, e.g. CRISPR RNA and transactivating (tracr) RNA or a single guide RNA (sgRNA) (chimeric RNA)) or other sequences and transcripts from a CRISPR locus. In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR system). In the context of formation of a CRISPR complex, "target sequence" refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. The section of the guide sequence through which complementarity to the target sequence is important for cleavage activity is referred to herein as the seed sequence. A target sequence may comprise any polynucleotide, such as DNA or RNA polynucleotides. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell, and may include nucleic acids in or from mitochondrial, organelles, vesicles, liposomes or particles present within the cell. In some embodiments, especially for non-nuclear uses, NLSs are not preferred. In some embodiments, a CRISPR system comprises one or more nuclear exports signals (NESs). In some embodiments, a CRISPR system comprises one or more NLSs and one or more NESs. In some embodiments, direct repeats may be identified in silico by searching for repetitive motifs that fulfill any or all of the following criteria: 1. found in a 2Kb window of genomic sequence flanking the type II CRISPR locus; 2. span from 20 to 50 bp; and 3. interspaced by 20 to 50 bp. In some embodiments, 2 of these criteria may be used, for instance 1 and 2, 2 and 3, or 1 and 3. In some embodiments, all 3 criteria may be used.

[0344] In embodiments of the invention the terms guide sequence and guide RNA, i.e. RNA capable of guiding Cas to a target genomic locus, are used interchangeably as in foregoing cited documents such as WO 2014/093622 (PCT/US2013/074667). In general, a guide sequence is any polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybridize with the target sequence and direct sequence-specific binding of a CRISPR complex to the target sequence. In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence, when

optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g. the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). In some embodiments, a guide sequence is about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length. In some embodiments, a guide sequence is less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length. Preferably the guide sequence is 10 30 nucleotides long. The ability of a guide sequence to direct sequence-specific binding of a CRISPR complex to a target sequence may be assessed by any suitable assay. For example, the components of a CRISPR system sufficient to form a CRISPR complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target sequence, such as by transfection with vectors encoding the components of the CRISPR sequence, followed by an assessment of preferential cleavage within the target sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target polynucleotide sequence may be evaluated in a test tube by providing the target sequence, components of a CRISPR complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art.

[0345] In some embodiments of CRISPR-Cas systems, the degree of complementarity between a guide sequence and its corresponding target sequence can be about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or 100%; a guide or RNA or sgRNA can be about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length; or guide or RNA or sgRNA can be less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length; and advantageously tracr RNA is 30 or 50 nucleotides in length. However, an aspect of the invention is to reduce off-target interactions, e.g., reduce the guide interacting with a target sequence having low complementarity. Indeed, in the examples, it is shown that the invention involves mutations that result in the CRISPR-Cas system being able to distinguish between target and off-target sequences that have greater than 80% to about 95%

complementarity, e.g., 83%-84% or 88-89% or 94-95% complementarity (for instance, distinguishing between a target having 18 nucleotides from an off-target of 18 nucleotides having 1, 2 or 3 mismatches). Accordingly, in the context of the present invention the degree of complementarity between a guide sequence and its corresponding target sequence is greater than 94.5% or 95% or 95.5% or 96% or 96.5% or 97% or 97.5% or 98% or 98.5% or 99% or 99.5% or 99.9%, or 100%. Off target is less than 100% or 99.9% or 99.5% or 99% or 99% or 98.5% or 98% or 97.5% or 97% or 96.5% or 96% or 95.5% or 95% or 94.5% or 94% or 93% or 92% or 91% or 90% or 89% or 88% or 87% or 86% or 85% or 84% or 83% or 82% or 81% or 80% complementarity between the sequence and the guide, with it advantageous that off target is 100% or 99.9% or 99.5% or 99% or 99% or 98.5% or 98% or 97.5% or 97% or 96.5% or 96% or 95.5% or 95% or 94.5% complementarity between the sequence and the guide.

[0346] In particularly preferred embodiments according to the invention, the guide RNA (capable of guiding Cas to a target locus) may comprise (1) a guide sequence capable of hybridizing to a genomic target locus in the eukaryotic cell; (2) a tracr sequence; and (3) a tracr mate sequence. All (1) to (3) may reside in a single RNA, i.e. an sgRNA (arranged in a 5' to 3' orientation), or the tracr RNA may be a different RNA than the RNA containing the guide and tracr sequence. The tracr hybridizes to the tracr mate sequence and directs the CRISPR/Cas complex to the target sequence. Where the tracr RNA is on a different RNA than the RNA containing the guide and tracr sequence, the length of each RNA may be optimized to be shortened from their respective native lengths, and each may be independently chemically modified to protect from degradation by cellular RNase or otherwise increase stability.

[0347] The methods according to the invention as described herein comprehend inducing one or more mutations in a eukaryotic cell (in vitro, i.e. in an isolated eukaryotic cell) as herein discussed comprising delivering to cell a vector as herein discussed. The mutation(s) can include the introduction, deletion, or substitution of one or more nucleotides at each target sequence of cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 1-75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 1, 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations include the introduction, deletion, or substitution of 10, 11, 12, 13,

14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 40, 45, 50, 75, 100, 200, 300, 400 or 500 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s).

[0348] For minimization of toxicity and off-target effect, it may be important to control the concentration of Cas mRNA and guide RNA delivered. Optimal concentrations of Cas mRNA and guide RNA can be determined by testing different concentrations in a cellular or non-human eukaryote animal model and using deep sequencing the analyze the extent of modification at potential off-target genomic loci. Alternatively, to minimize the level of toxicity and off-target effect, Cas nickase mRNA (for example S. pyogenes Cas9 with the D10A mutation) can be delivered with a pair of guide RNAs targeting a site of interest. Guide sequences and strategies to minimize toxicity and off-target effects can be as in WO 2014/093622 (PCT/US2013/074667); or, via mutation as herein.

[0349] Typically, in the context of an endogenous CRISPR system, formation of a CRISPR complex (comprising a guide sequence hybridized to a target sequence and complexed with one or more Cas proteins) results in cleavage of one or both strands in or near (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target sequence. Without wishing to be bound by theory, the tracr sequence, which may comprise or consist of all or a portion of a wild-type tracr sequence (e.g. about or more than about 20, 26, 32, 45, 48, 54, 63, 67, 85, or more nucleotides of a wild-type tracr sequence), may also form part of a CRISPR complex, such as by hybridization along at least a portion of the tracr sequence to all or a portion of a tracr mate sequence that is operably linked to the guide sequence.

Guide Modifications

[0350] In certain embodiments, guides of the invention comprise non-naturally occurring nucleic acids and/or non-naturally occurring nucleotides and/or nucleotide analogs, and/or chemically modifications. Non-naturally occurring nucleic acids can include, for example, mixtures of naturally and non-naturally occurring nucleotides. Non-naturally occurring nucleotides and/or nucleotide analogs may be modified at the ribose, phosphate, and/or base moiety. In an embodiment of the invention, a guide nucleic acid comprises ribonucleotides and non-ribonucleotides. In one such embodiment, a guide comprises one or more ribonucleotides and one or more deoxyribonucleotides. In an embodiment of the invention, the guide

comprises one or more non-naturally occurring nucleotide or nucleotide analog such as a nucleotide with phosphorothioate linkage, boranophosphate linkage, a locked nucleic acid (LNA) nucleotides comprising a methylene bridge between the 2,Ä≤ and 4,Ä≤ carbons of the ribose ring, peptide nucleic acids (PNA), or bridged nucleic acids (BNA). Other examples of modified nucleotides include 2'-O-methyl analogs, 2'-deoxy analogs, 2-thiouridine analogs, N6-methyladenosine analogs, or 2'-fluoro analogs. Further examples of modified nucleotides include linkage of chemical moieties at the 2' position, including but not limited to peptides, nuclear localization sequence (NLS), peptide nucleic acid (PNA), polyethylene glycol (PEG), triethylene glycol, or tetraethyleneglycol (TEG). Further examples of modified bases include, but are not limited to, 2-aminopurine, 5-bromo-uridine, pseudouridine (Œ®), N1-methylpseudouridine (me1Œ®), 5-methoxyuridine(5moU), inosine, 7-methylguanosine. Examples of guide RNA chemical modifications include, without limitation, incorporation of 2'-O-methyl (M), 2'-O-methyl-3'-phosphorothioate (MS), phosphorothioate (PS), S-constrained ethyl(cEt), 2'-O-methyl-3'-thioPACE (MSP), or 2'-O-methyl-3'-phosphonoacetate (MP) at one or more terminal nucleotides. Such chemically modified guides can comprise increased stability and increased activity as compared to unmodified guides, though on-target vs. off-target specificity is not predictable. (See, Hendel, 2015, Nat Biotechnol. 33(9):985-9, doi: 10.1038/nbt.3290, published online 29 June 2015; Ragdarm et al., 0215, PNAS, E7110-E7111; Allerson et al., J. Med. Chem. 2005, 48:901-904; Bramsen et al., Front. Genet., 2012, 3:154; Deng et al., PNAS, 2015, 112:11870-11875; Sharma et al., MedChemComm., 2014, 5:1454-1471; Hendel et al., Nat. Biotechnol. (2015) 33(9): 985-989; Li et al., Nature Biomedical Engineering, 2017, 1, 0066 DOI:10.1038/s41551-017-0066; Ryan et al., Nucleic Acids Res. (2018) 46(2): 792-803). In some embodiments, the 5' and/or 3' end of a guide RNA is modified by a variety of functional moieties including fluorescent dyes, polyethylene glycol, cholesterol, proteins, or detection tags. (See Kelly et al., 2016, J. Biotech. 233:74-83). In certain embodients, a guide comprises ribonucleotides in a region that binds to a target DNA and one or more deoxyribonucletides and/or nucleotide analogs in a region that binds to Cas9, Cpf1, C2c1, or Cas13. In an embodiment of the invention, deoxyribonucleotides and/or nucleotide analogs are incorporated in engineered guide structures, such as, without limitation, 5' and/or 3' end, stem-loop regions, and the seed region. In certain embodiments, the modification is not in the 5'-handle of the stem-loop regions. Chemical modification in the 5'-handle of the stem-loop region of a guide may abolish its function (see Li, et al., Nature Biomedical Engineering, 2017, 1:0066). In certain embodiments, at least 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, 35, 40, 45, 50, or 75 nucleotides of a

guide is chemically modified. In some embodiments, 3-5 nucleotides at either the 3' or the 5' end of a guide is chemically modified. In some embodiments, only minor modifications are introduced in the seed region, such as 2'-F modifications. In some embodiments, 2'-F modification is introduced at the 3' end of a guide. In certain embodiments, three to five nucleotides at the 5' and/or the 3' end of the guide are chemically modified with 2'-O-methyl (M), 2'-O-methyl-3'-phosphorothioate (MS), S-constrained ethyl(cEt), 2'-O-methyl-3'-thioPACE (MSP), or 2'-O-methyl-3'-phosphonoacetate (MP). Such modification can enhance genome editing efficiency (see Hendel et al., Nat. Biotechnol. (2015) 33(9): 985-989; Ryan et al., Nucleic Acids Res. (2018) 46(2): 792-803). In certain embodiments, all of the phosphodiester bonds of a guide are substituted with phosphorothioates (PS) for enhancing levels of gene disruption. In certain embodiments, more than five nucleotides at the 5' and/or the 3' end of the guide are chemically modified with 2'-O-Me, 2'-F or S-constrained ethyl(cEt). Such chemically modified guide can mediate enhanced levels of gene disruption (see Ragdarm et al., 0215, PNAS, E7110-E7111). In an embodiment of the invention, a guide is modified to comprise a chemical moiety at its 3' and/or 5' end. Such moieties include, but are not limited to amine, azide, alkyne, thio, dibenzocyclooctyne (DBCO), Rhodamine, peptides, nuclear localization sequence (NLS), peptide nucleic acid (PNA), polyethylene glycol (PEG), triethylene glycol, or tetraethyleneglycol (TEG). In certain embodiment, the chemical moiety is conjugated to the guide by a linker, such as an alkyl chain. In certain embodiments, the chemical moiety of the modified guide can be used to attach the guide to another molecule, such as DNA, RNA, protein, or nanoparticles. Such chemically modified guide can be used to identify or enrich cells generically edited by a CRISPR system (see Lee et al., eLife, 2017, 6:e25312, DOI:10.7554). In some embodiments, 3 nucleotides at each of the 3' and 5' ends are chemically modified. In a specific embodiment, the modifications comprise 2'-O-methyl or phosphorothioate analogs. In a specific embodiment, 12 nucleotides in the tetraloop and 16 nucleotides in the stem-loop region are replaced with 2'-O-methyl analogs. Such chemical modifications improve in vivo editing and stability (see Finn et al., Cell Reports (2018), 22: 2227-2235). In some embodiments, more than 60 or 70 nucleotides of the guide are chemically modified. In some embodiments, this modification comprises replacement of nucleotides with 2'-O-methyl or 2'-fluoro nucleotide analogs or phosphorothioate (PS) modification of phosphodiester bonds. In some embodiments, the chemical modification comprises 2'-O-methyl or 2'-fluoro modification of guide nucleotides extending outside of the nuclease protein when the CRISPR complex is formed or PS modification of 20 to 30 or more nucleotides of the 3'-terminus of the guide. In a particular embodiment, the chemical modification further

comprises 2'-O-methyl analogs at the 5' end of the guide or 2'-fluoro analogs in the seed and tail regions. Such chemical modifications improve stability to nuclease degradation and maintain or enhance genome-editing activity or efficiency, but modification of all nucleotides may abolish the function of the guide (see Yin et al., Nat. Biotech. (2018), 35(12): 1179-1187). Such chemical modifications may be guided by knowledge of the structure of the CRISPR complex, including knowledge of the limited number of nuclease and RNA 2'-OH interactions (see Yin et al., Nat. Biotech. (2018), 35(12): 1179-1187). In some embodiments, one or more guide RNA nucleotides may be replaced with DNA nucleotides. In some embodiments, up to 2, 4, 6, 8, 10, or 12 RNA nucleotides of the 5'-end tail/seed guide region are replaced with DNA nucleotides. In certain embodiments, the majority of guide RNA nucleotides at the 3' end are replaced with DNA nucleotides. In particular embodiments, 16 guide RNA nucleotides at the 3' end are replaced with DNA nucleotides. In particular embodiments, 8 guide RNA nucleotides of the 5'-end tail/seed region and 16 RNA nucleotides at the 3' end are replaced with DNA nucleotides. In particular embodiments, guide RNA nucleotides that extend outside of the nuclease protein when the CRISPR complex is formed are replaced with DNA nucleotides. Such replacement of multiple RNA nucleotides with DNA nucleotides leads to decreased off-target activity but similar on-target activity compared to an unmodified guide; however, replacement of all RNA nucleotides at the 3' end may abolish the function of the guide (see Yin et al., Nat. Chem. Biol. (2018) 14, 311-316). Such modifications may be guided by knowledge of the structure of the CRISPR complex, including knowledge of the limited number of nuclease and RNA 2'-OH interactions (see Yin et al., Nat. Chem. Biol. (2018) 14, 311-316).

[0351] In one aspect of the invention, the guide comprises a modified crRNA for Cpf1, having a 5'-handle and a guide segment further comprising a seed region and a 3'-terminus. In some embodiments, the modified guide can be used with a Cpf1 of any one of Acidaminococcus sp. BV3L6 Cpf1 (AsCpf1); Francisella tularensis subsp. Novicida U112 Cpf1 (FnCpf1); L. bacterium MC2017 Cpf1 (Lb3Cpf1); Butyrivibrio proteoclasticus Cpf1 (BpCpf1); Parcubacteria bacterium GWC2011_GWC2_44_17 Cpf1 (PbCpf1); Peregrinibacteria bacterium GW2011_GWA_33_10 Cpf1 (PeCpf1); Leptospira inadai Cpf1 (LiCpf1); Smithella sp. SC_K08D17 Cpf1 (SsCpf1); L. bacterium MA2020 Cpf1 (Lb2Cpf1); Porphyromonas crevioricanis Cpf1 (PcCpf1); Porphyromonas macacae Cpf1 (PmCpf1); Candidatus Methanoplasma termitum Cpf1 (CMtCpf1); Eubacterium eligens Cpf1 (EeCpf1); Moraxella bovoculi 237 Cpf1 (MbCpf1); Prevotella disiens Cpf1 (PdCpf1); or L. bacterium ND2006 Cpf1 (LbCpf1).

[0352] In some embodiments, the modification to the guide is a chemical modification, an insertion, a deletion or a split. In some embodiments, the chemical modification includes, but is not limited to, incorporation of 2'-O-methyl (M) analogs, 2'-deoxy analogs, 2-thiouridine analogs, N6-methyladenosine analogs, 2'-fluoro analogs, 2-aminopurine, 5-bromo-uridine, pseudouridine (Œ®), N1-methylpseudouridine (me1Œ®), 5-methoxyuridine(5moU), inosine, 7-methylguanosine, 2'-O-methyl-3'-phosphorothioate (MS), S-constrained ethyl(cEt), phosphorothioate (PS), 2'-O-methyl-3'-thioPACE (MSP), or 2'-O-methyl-3'-phosphonoacetate (MP). In some embodiments, the guide comprises one or more of phosphorothioate modifications. In certain embodiments, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 25 nucleotides of the guide are chemically modified. In some embodiments, all nucleotides are chemically modified. In certain embodiments, one or more nucleotides in the seed region are chemically modified. In certain embodiments, one or more nucleotides in the 3'-terminus are chemically modified. In certain embodiments, none of the nucleotides in the 5'-handle is chemically modified. In some embodiments, the chemical modification in the seed region is a minor modification, such as incorporation of a 2'-fluoro analog. In a specific embodiment, one nucleotide of the seed region is replaced with a 2'-fluoro analog. In some embodiments, 5 or 10 nucleotides in the 3'-terminus are chemically modified. Such chemical modifications at the 3'-terminus of the Cpf1 CrRNA improve gene cutting efficiency (see Li, et al., Nature Biomedical Engineering, 2017, 1:0066). In a specific embodiment, 5 nucleotides in the 3'-terminus are replaced with 2'-fluoro analogues. In a specific embodiment, 10 nucleotides in the 3'-terminus are replaced with 2'-fluoro analogues. In a specific embodiment, 5 nucleotides in the 3'-terminus are replaced with 2'- O-methyl (M) analogs. In some embodiments, 3 nucleotides at each of the 3' and 5' ends are chemically modified. In a specific embodiment, the modifications comprise 2'-O-methyl or phosphorothioate analogs. In a specific embodiment, 12 nucleotides in the tetraloop and 16 nucleotides in the stem-loop region are replaced with 2'-O-methyl analogs. Such chemical modifications improve in vivo editing and stability (see Finn et al., Cell Reports (2018), 22: 2227-2235).

[0353] In some embodiments, the loop of the 5'-handle of the guide is modified. In some embodiments, the loop of the 5'-handle of the guide is modified to have a deletion, an insertion, a split, or chemical modifications. In certain embodiments, the loop comprises 3, 4, or 5 nucleotides. In certain embodiments, the loop comprises the sequence of UCUU, UUUU, UAUU, or UGUU. In some embodiments, the guide molecule forms a stemloop with a separate non-covalently linked sequence, which can be DNA or RNA.

Synthetically linked guide

[0354] In one aspect, the guide comprises a tracr sequence and a tracr mate sequence that are chemically linked or conjugated via a non-phosphodiester bond. In one aspect, the guide comprises a tracr sequence and a tracr mate sequence that are chemically linked or conjugated via a non-nucleotide loop. In some embodiments, the tracr and tracr mate sequences are joined via a non-phosphodiester covalent linker. Examples of the covalent linker include but are not limited to a chemical moiety selected from the group consisting of carbamates, ethers, esters, amides, imines, amidines, aminotrizines, hydrozone, disulfides, thioethers, thioesters, phosphorothioates, phosphorodithioates, sulfonamides, sulfonates, fulfones, sulfoxides, ureas, thioureas, hydrazide, oxime, triazole, photolabile linkages, C-C bond forming groups such as Diels-Alder cyclo-addition pairs or ring-closing metathesis pairs, and Michael reaction pairs.

[0355] In some embodiments, the tracr and tracr mate sequences are first synthesized using the standard phosphoramidite synthetic protocol (Herdewijn, P., ed., Methods in Molecular Biology Col 288, Oligonucleotide Synthesis: Methods and Applications, Humana Press, New Jersey (2012)). In some embodiments, the tracr or tracr mate sequences can be functionalized to contain an appropriate functional group for ligation using the standard protocol known in the art (Hermanson, G. T., Bioconjugate Techniques, Academic Press (2013)). Examples of functional groups include, but are not limited to, hydroxyl, amine, carboxylic acid, carboxylic acid halide, carboxylic acid active ester, aldehyde, carbonyl, chlorocarbonyl, imidazolylcarbonyl, hydrozide, semicarbazide, thio semicarbazide, thiol, maleimide, haloalkyl, sufonyl, ally, propargyl, diene, alkyne, and azide. Once the tracr and the tracr mate sequences are functionalized, a covalent chemical bond or linkage can be formed between the two oligonucleotides. Examples of chemical bonds include, but are not limited to, those based on carbamates, ethers, esters, amides, imines, amidines, aminotrizines, hydrozone, disulfides, thioethers, thioesters, phosphorothioates, phosphorodithioates, sulfonamides, sulfonates, fulfones, sulfoxides, ureas, thioureas, hydrazide, oxime, triazole, photolabile linkages, C-C bond forming groups such as Diels-Alder cyclo-addition pairs or ring-closing metathesis pairs, and Michael reaction pairs.

[0356] In some embodiments, the tracr and tracr mate sequences can be chemically synthesized. In some embodiments, the chemical synthesis uses automated, solid-phase oligonucleotide synthesis machines with 2'-acetoxyethyl orthoester (2'-ACE) (Scaringe et al., J. Am. Chem. Soc. (1998) 120: 11820-11821; Scaringe, Methods Enzymol. (2000) 317: 3-18) or 2'-thionocarbamate (2'-TC) chemistry (Dellinger et al., J. Am. Chem. Soc. (2011) 133: 11540-11546; Hendel et al., Nat. Biotechnol. (2015) 33:985-989).

[0357] In some embodiments, the tracr and tracr mate sequences can be covalently linked using various bioconjugation reactions, loops, bridges, and non-nucleotide links via modifications of sugar, internucleotide phosphodiester bonds, purine and pyrimidine residues. Sletten et al., Angew. Chem. Int. Ed. (2009) 48:6974-6998; Manoharan, M. Curr. Opin. Chem. Biol. (2004) 8: 570-9; Behlke et al., Oligonucleotides (2008) 18: 305-19; Watts, et al., Drug. Discov. Today (2008) 13: 842-55; Shukla, et al., ChemMedChem (2010) 5: 328-49.

[0358] In some embodiments, the tracr and tracr mate sequences can be covalently linked using click chemistry. In some embodiments, the tracr and tracr mate sequences can be covalently linked using a triazole linker. In some embodiments, the tracr and tracr mate sequences can be covalently linked using Huisgen 1,3-dipolar cycloaddition reaction involving an alkyne and azide to yield a highly stable triazole linker (He et al., ChemBioChem (2015) 17: 1809-1812; WO 2016/186745). In some embodiments, the tracr and tracr mate sequences are covalently linked by ligating a 5'-hexyne tracrRNA and a 3'-azide crRNA. In some embodiments, either or both of the 5'-hexyne tracrRNA and a 3'-azide crRNA can be protected with 2'-acetoxyethl orthoester (2'-ACE) group, which can be subsequently removed using Dharmacon protocol (Scaringe et al., J. Am. Chem. Soc. (1998) 120: 11820-11821; Scaringe, Methods Enzymol. (2000) 317: 3-18).

[0359] In some embodiments, the tracr and tracr mate sequences can be covalently linked via a linker (e.g., a non-nucleotide loop) that comprises a moiety such as spacers, attachments, bioconjugates, chromophores, reporter groups, dye labeled RNAs, and non-naturally occurring nucleotide analogues. More specifically, suitable spacers for purposes of this invention include, but are not limited to, polyethers (e.g., polyethylene glycols, polyalcohols, polypropylene glycol or mixtures of efhylene and propylene glycols), polyamines group (e.g., spennine, spermidine and polymeric derivatives thereof), polyesters (e.g., poly(ethyl acrylate)), polyphosphodiesters, alkylenes, and combinations thereof. Suitable attachments include any moiety that can be added to the linker to add additional properties to the linker, such as but not limited to, fluorescent labels. Suitable bioconjugates include, but are not limited to, peptides, glycosides, lipids, cholesterol, phospholipids, diacyl glycerols and dialkyl glycerols, fatty acids, hydrocarbons, enzyme substrates, steroids, biotin, digoxigenin, carbohydrates, polysaccharides. Suitable chromophores, reporter groups, and dye-labeled RNAs include, but are not limited to, fluorescent dyes such as fluorescein and rhodamine, chemiluminescent, electrochemiluminescent, and bioluminescent marker compounds. The design of example linkers conjugating two RNA components are also described in WO 2004/015075.

[0360] The linker (e.g., a non-nucleotide loop) can be of any length. In some embodiments, the linker has a length equivalent to about 0-16 nucleotides. In some embodiments, the linker has a length equivalent to about 0-8 nucleotides. In some embodiments, the linker has a length equivalent to about 0-4 nucleotides. In some embodiments, the linker has a length equivalent to about 2 nucleotides. Example linker design is also described in WO2011/008730.

[0361] A typical Type II Cas sgRNA comprises (in 5' to 3' direction): a guide sequence, a poly U tract, a first complimentary stretch (the "repeat"), a loop (tetraloop), a second complimentary stretch (the "anti-repeat" being complimentary to the repeat), a stem, and further stem loops and stems and a poly A (often poly U in RNA) tail (terminator). In preferred embodiments, certain aspects of guide architecture are retained, certain aspect of guide architecture cam be modified, for example by addition, subtraction, or substitution of features, whereas certain other aspects of guide architecture are maintained. Preferred locations for engineered sgRNA modifications, including but not limited to insertions, deletions, and substitutions include guide termini and regions of the sgRNA that are exposed when complexed with CRISPR protein and/or target, for example the tetraloop and/or loop2.

[0362] In certain embodiments, guides of the invention comprise specific binding sites (e.g. aptamers) for adapter proteins, which may comprise one or more functional domains (e.g. via fusion protein). When such a guide forms a CRISPR complex (i.e. CRISPR enzyme binding to guide and target) the adapter proteins bind and, the functional domain associated with the adapter protein is positioned in a spatial orientation which is advantageous for the attributed function to be effective. For example, if the functional domain is a transcription activator (e.g. VP64 or p65), the transcription activator is placed in a spatial orientation which allows it to affect the transcription of the target. Likewise, a transcription repressor will be advantageously positioned to affect the transcription of the target and a nuclease (e.g. Fok1) will be advantageously positioned to cleave or partially cleave the target.

[0363] The skilled person will understand that modifications to the guide which allow for binding of the adapter + functional domain but not proper positioning of the adapter + functional domain (e.g. due to steric hindrance within the three dimensional structure of the CRISPR complex) are modifications which are not intended. The one or more modified guide may be modified at the tetra loop, the stem loop 1, stem loop 2, or stem loop 3, as described herein, preferably at either the tetra loop or stem loop 2, and most preferably at both the tetra loop and stem loop 2.

[0364] The repeat:anti repeat duplex will be apparent from the secondary structure of the sgRNA. It may be typically a first complimentary stretch after (in 5' to 3' direction) the poly U tract and before the tetraloop; and a second complimentary stretch after (in 5' to 3' direction) the tetraloop and before the poly A tract. The first complimentary stretch (the "repeat") is complimentary to the second complimentary stretch (the "anti-repeat"). As such, they Watson-Crick base pair to form a duplex of dsRNA when folded back on one another. As such, the anti-repeat sequence is the complimentary sequence of the repeat and in terms to A-U or C-G base pairing, but also in terms of the fact that the anti-repeat is in the reverse orientation due to the tetraloop.

[0365] In an embodiment of the invention, modification of guide architecture comprises replacing bases in stemloop 2. For example, in some embodiments, "actt" ("acuu" in RNA) and "aagt" ("aagu" in RNA) bases in stemloop2 are replaced with "cgcc" and "gcgg". In some embodiments, "actt" and "aagt" bases in stemloop2 are replaced with complimentary GC-rich regions of 4 nucleotides. In some embodiments, the complimentary GC-rich regions of 4 nucleotides are "cgcc" and "gcgg" (both in 5' to 3' direction). In some embodiments, the complimentary GC-rich regions of 4 nucleotides are "gcgg" and "cgcc" (both in 5' to 3' direction). Other combination of C and G in the complimentary GC-rich regions of 4 nucleotides will be apparent including CCCC and GGGG.

[0366] In one aspect, the stemloop 2, e.g., "ACTTgtttAAGT" can be replaced by any "XXXXgtttYYYY", e.g., where XXXX and YYYY represent any complementary sets of nucleotides that together will base pair to each other to create a stem.

[0367] In one aspect, the stem comprises at least about 4bp comprising complementary X and Y sequences, although stems of more, e.g., 5, 6, 7, 8, 9, 10, 11 or 12 or fewer, e.g., 3, 2, base pairs are also contemplated. Thus, for example X2-12 and Y2-12 (wherein X and Y represent any complementary set of nucleotides) may be contemplated. In one aspect, the stem made of the X and Y nucleotides, together with the "gttt," will form a complete hairpin in the overall secondary structure; and, this may be advantageous and the amount of base pairs can be any amount that forms a complete hairpin. In one aspect, any complementary X:Y basepairing sequence (e.g., as to length) is tolerated, so long as the secondary structure of the entire sgRNA is preserved. In one aspect, the stem can be a form of X:Y basepairing that does not disrupt the secondary structure of the whole sgRNA in that it has a DR:tracr duplex, and 3 stemloops. In one aspect, the "gttt" tetraloop that connects ACTT and AAGT (or any alternative stem made of X:Y basepairs) can be any sequence of the same length (e.g., 4 basepair) or longer that does not interrupt the overall secondary structure of the sgRNA. In one aspect, the

stemloop can be something that further lengthens stemloop2, e.g. can be MS2 aptamer. In one aspect, the stemloop3 "GGCACCGagtCGGTGC" can likewise take on a "XXXXXXXagtYYYYYYY" form, e.g., wherein X7 and Y7 represent any complementary sets of nucleotides that together will base pair to each other to create a stem. In one aspect, the stem comprises about 7bp comprising complementary X and Y sequences, although stems of more or fewer basepairs are also contemplated. In one aspect, the stem made of the X and Y nucleotides, together with the "agt", will form a complete hairpin in the overall secondary structure. In one aspect, any complementary X:Y basepairing sequence is tolerated, so long as the secondary structure of the entire sgRNA is preserved. In one aspect, the stem can be a form of X:Y basepairing that doesn't disrupt the secondary structure of the whole sgRNA in that it has a DR:tracr duplex, and 3 stemloops. In one aspect, the "agt" sequence of the stemloop 3 can be extended or be replaced by an aptamer, e.g., a MS2 aptamer or sequence that otherwise generally preserves the architecture of stemloop3. In one aspect for alternative Stemloops 2 and/or 3, each X and Y pair can refer to any basepair. In one aspect, non-Watson Crick basepairing is contemplated, where such pairing otherwise generally preserves the architecture of the stemloop at that position.

[0368] In one aspect, the DR:tracrRNA duplex can be replaced with the form: gYYYYag(N)NNNNxxxxNNNN(AAN)uuRRRRu (using standard IUPAC nomenclature for nucleotides), wherein (N) and (AAN) represent part of the bulge in the duplex, and "xxxx" represents a linker sequence. NNNN on the direct repeat can be anything so long as it basepairs with the corresponding NNNN portion of the tracrRNA. In one aspect, the DR:tracrRNA duplex can be connected by a linker of any length (xxxx...), any base composition, as long as it doesn't alter the overall structure.

[0369] In one aspect, the sgRNA structural requirement is to have a duplex and 3 stemloops. In most aspects, the actual sequence requirement for many of the particular base requirements are lax, in that the architecture of the DR:tracrRNA duplex should be preserved, but the sequence that creates the architecture, i.e., the stems, loops, bulges, etc., may be alterred.

Aptamers

[0370] One guide with a first aptamer/RNA-binding protein pair can be linked or fused to an activator, whilst a second guide with a second aptamer/RNA-binding protein pair can be linked or fused to a repressor. The guides are for different targets (loci), so this allows one gene to be activated and one repressed. For example, the following schematic shows such an approach:

[0371] Guide 1- MS2 aptamer-------MS2 RNA-binding protein-------VP64 activator; and

[0372] Guide 2 - PP7 aptamer-------PP7 RNA-binding protein-------SID4x repressor.

[0373] The present invention also relates to orthogonal PP7/MS2 gene targeting. In this example, sgRNA targeting different loci are modified with distinct RNA loops in order to recruit MS2-VP64 or PP7-SID4X, which activate and repress their target loci, respectively. PP7 is the RNA-binding coat protein of the bacteriophage Pseudomonas. Like MS2, it binds a specific RNA sequence and secondary structure. The PP7 RNA-recognition motif is distinct from that of MS2. Consequently, PP7 and MS2 can be multiplexed to mediate distinct effects at different genomic loci simultaneously. For example, an sgRNA targeting locus A can be modified with MS2 loops, recruiting MS2-VP64 activators, while another sgRNA targeting locus B can be modified with PP7 loops, recruiting PP7-SID4X repressor domains. In the same cell, dCas13 can thus mediate orthogonal, locus-specific modifications. This principle can be extended to incorporate other orthogonal RNA-binding proteins such as Q-beta.

[0374] An alternative option for orthogonal repression includes incorporating non-coding RNA loops with transactive repressive function into the guide (either at similar positions to the MS2/PP7 loops integrated into the guide or at the 3' terminus of the guide). For instance, guides were designed with non-coding (but known to be repressive) RNA loops (e.g. using the Alu repressor (in RNA) that interferes with RNA polymerase II in mammalian cells). The Alu RNA sequence was located: in place of the MS2 RNA sequences as used herein (e.g. at tetraloop and/or stem loop 2); and/or at 3' terminus of the guide. This gives possible combinations of MS2, PP7 or Alu at the tetraloop and/or stemloop 2 positions, as well as, optionally, addition of Alu at the 3' end of the guide (with or without a linker).

[0375] The use of two different aptamers (distinct RNA) allows an activator-adaptor protein fusion and a repressor-adaptor protein fusion to be used, with different guides, to activate expression of one gene, whilst repressing another. They, along with their different guides can be administered together, or substantially together, in a multiplexed approach. A large number of such modified guides can be used all at the same time, for example 10 or 20 or 30 and so forth, whilst only one (or at least a minimal number) of Cas13s to be delivered, as a comparatively small number of Cas13s can be used with a large number modified guides. The adaptor protein may be associated (preferably linked or fused to) one or more activators or one or more repressors. For example, the adaptor protein may be associated with a first activator and a second activator. The first and second activators may be the same, but they are preferably different activators. For example, one might be VP64, whilst the other might be p65, although these are just examples and other transcriptional activators are envisaged. Three or more or even four or more activators (or repressors) may be used, but package size may limit

the number being higher than 5 different functional domains. Linkers are preferably used, over a direct fusion to the adaptor protein, where two or more functional domains are associated with the adaptor protein. Suitable linkers might include the GlySer linker.

[0376] It is also envisaged that the enzyme-guide complex as a whole may be associated with two or more functional domains. For example, there may be two or more functional domains associated with the enzyme, or there may be two or more functional domains associated with the guide (via one or more adaptor proteins), or there may be one or more functional domains associated with the enzyme and one or more functional domains associated with the guide (via one or more adaptor proteins).

[0377] The fusion between the adaptor protein and the activator or repressor may include a linker. For example, GlySer linkers GGGS can be used. They can be used in repeats of 3 ((GGGGS)3) or 6, 9 or even 12 or more, to provide suitable lengths, as required. Linkers can be used between the RNA-binding protein and the functional domain (activator or repressor), or between the CRISPR Enzyme (Cas13) and the functional domain (activator or repressor). The linkers the user to engineer appropriate amounts of "mechanical flexibility".

[0378] *Dead guides: Guide RNAs comprising a dead guide sequence may be used in the present invention*

[0379] In one aspect, the invention provides guide sequences which are modified in a manner which allows for formation of the CRISPR complex and successful binding to the target, while at the same time, not allowing for successful nuclease activity (i.e. without nuclease activity / without indel activity). For matters of explanation such modified guide sequences are referred to as "dead guides" or "dead guide sequences". These dead guides or dead guide sequences can be thought of as catalytically inactive or conformationally inactive with regard to nuclease activity. Nuclease activity may be measured using surveyor analysis or deep sequencing as commonly used in the art, preferably surveyor analysis. Similarly, dead guide sequences may not sufficiently engage in productive base pairing with respect to the ability to promote catalytic activity or to distinguish on-target and off-target binding activity. Briefly, the surveyor assay involves purifying and amplifying a CRISPR target site for a gene and forming heteroduplexes with primers amplifying the CRISPR target site. After re-anneal, the products are treated with SURVEYOR nuclease and SURVEYOR enhancer S (Transgenomics) following the manufacturer's recommended protocols, analyzed on gels, and quantified based upon relative band intensities.

[0380] Hence, in a related aspect, the invention provides a non-naturally occurring or engineered composition Cas13 CRISPR-Cas system comprising a functional Cas13 as

described herein, and guide RNA (gRNA) wherein the gRNA comprises a dead guide sequence whereby the gRNA is capable of hybridizing to a target sequence such that the Cas13 CRISPR-Cas system is directed to a genomic locus of interest in a cell without detectable indel activity resultant from nuclease activity of a non-mutant Cas13 enzyme of the system as detected by a SURVEYOR assay. For shorthand purposes, a gRNA comprising a dead guide sequence whereby the gRNA is capable of hybridizing to a target sequence such that the Cas13 CRISPR-Cas system is directed to a genomic locus of interest in a cell without detectable indel activity resultant from nuclease activity of a non-mutant Cas13 enzyme of the system as detected by a SURVEYOR assay is herein termed a "dead gRNA". It is to be understood that any of the gRNAs according to the invention as described herein elsewhere may be used as dead gRNAs / gRNAs comprising a dead guide sequence as described herein below. Any of the methods, products, compositions and uses as described herein elsewhere is equally applicable with the dead gRNAs / gRNAs comprising a dead guide sequence as further detailed below. By means of further guidance, the following particular aspects and embodiments are provided.

[0381] The ability of a dead guide sequence to direct sequence-specific binding of a CRISPR complex to a target sequence may be assessed by any suitable assay. For example, the components of a CRISPR system sufficient to form a CRISPR complex, including the dead guide sequence to be tested, may be provided to a host cell having the corresponding target sequence, such as by transfection with vectors encoding the components of the CRISPR sequence, followed by an assessment of preferential cleavage within the target sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target polynucleotide sequence may be evaluated in a test tube by providing the target sequence, components of a CRISPR complex, including the dead guide sequence to be tested and a control guide sequence different from the test dead guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art. A dead guide sequence may be selected to target any target sequence. In some embodiments, the target sequence is a sequence within a genome of a cell.

[0382] As explained further herein, several structural parameters allow for a proper framework to arrive at such dead guides. Dead guide sequences are shorter than respective guide sequences which result in active Cas13-specific indel formation. Dead guides are 5%, 10%, 20%, 30%, 40%, 50%, shorter than respective guides directed to the same Cas13 leading to active Cas13-specific indel formation.

[0383] As explained below and known in the art, one aspect of gRNA - Cas specificity is the direct repeat sequence, which is to be appropriately linked to such guides. In particular, this implies that the direct repeat sequences are designed dependent on the origin of the Cas. Thus, structural data available for validated dead guide sequences may be used for designing Cas specific equivalents. Structural similarity between, e.g., the orthologous nuclease domains RuvC of two or more Cas effector proteins may be used to transfer design equivalent dead guides. Thus, the dead guide herein may be appropriately modified in length and sequence to reflect such Cas specific equivalents, allowing for formation of the CRISPR complex and successful binding to the target, while at the same time, not allowing for successful nuclease activity.

[0384] The use of dead guides in the context herein as well as the state of the art provides a surprising and unexpected platform for network biology and/or systems biology in both in vitro, ex vivo, and in vivo applications, allowing for multiplex gene targeting, and in particular bidirectional multiplex gene targeting. Prior to the use of dead guides, addressing multiple targets, for example for activation, repression and/or silencing of gene activity, has been challenging and in some cases not possible. With the use of dead guides, multiple targets, and thus multiple activities, may be addressed, for example, in the same cell, in the same animal, or in the same patient. Such multiplexing may occur at the same time or staggered for a desired timeframe.

[0385] For example, the dead guides now allow for the first time to use gRNA as a means for gene targeting, without the consequence of nuclease activity, while at the same time providing directed means for activation or repression. Guide RNA comprising a dead guide may be modified to further include elements in a manner which allow for activation or repression of gene activity, in particular protein adaptors (e.g. aptamers) as described herein elsewhere allowing for functional placement of gene effectors (e.g. activators or repressors of gene activity). One example is the incorporation of aptamers, as explained herein and in the state of the art. By engineering the gRNA comprising a dead guide to incorporate protein-interacting aptamers (Konermann et al., "Genome-scale transcription activation by an engineered CRISPR-Cas9 complex," doi:10.1038/nature14136, incorporated herein by reference), one may assemble a synthetic transcription activation complex consisting of multiple distinct effector domains. Such may be modeled after natural transcription activation processes. For example, an aptamer, which selectively binds an effector (e.g. an activator or repressor; dimerized MS2 bacteriophage coat proteins as fusion proteins with an activator or repressor), or a protein which itself binds an effector (e.g. activator or repressor) may be

appended to a dead gRNA tetraloop and/or a stem-loop 2. In the case of MS2, the fusion protein MS2-VP64 binds to the tetraloop and/or stem-loop 2 and in turn mediates transcriptional up-regulation, for example for Neurog2. Other transcriptional activators are, for example, VP64. P65, HSF1, and MyoD1. By mere example of this concept, replacement of the MS2 stem-loops with PP7-interacting stem-loops may be used to recruit repressive elements.

[0386] Thus, one aspect is a gRNA of the invention which comprises a dead guide, wherein the gRNA further comprises modifications which provide for gene activation or repression, as described herein. The dead gRNA may comprise one or more aptamers. The aptamers may be specific to gene effectors, gene activators or gene repressors. Alternatively, the aptamers may be specific to a protein which in turn is specific to and recruits / binds a specific gene effector, gene activator or gene repressor. If there are multiple sites for activator or repressor recruitment, it is preferred that the sites are specific to either activators or repressors. If there are multiple sites for activator or repressor binding, the sites may be specific to the same activators or same repressors. The sites may also be specific to different activators or different repressors. The gene effectors, gene activators, gene repressors may be present in the form of fusion proteins.

[0387] In an embodiment, the dead gRNA as described herein or the Cas13 CRISPR-Cas complex as described herein includes a non-naturally occurring or engineered composition comprising two or more adaptor proteins, wherein each protein is associated with one or more functional domains and wherein the adaptor protein binds to the distinct RNA sequence(s) inserted into the at least one loop of the dead gRNA.

[0388] Hence, an aspect provides a non-naturally occurring or engineered composition comprising a guide RNA (gRNA) comprising a dead guide sequence capable of hybridizing to a target sequence in a genomic locus of interest in a cell, wherein the dead guide sequence is as defined herein, a Cas13 comprising at least one or more nuclear localization sequences, wherein the Cas13 optionally comprises at least one mutation wherein at least one loop of the dead gRNA is modified by the insertion of distinct RNA sequence(s) that bind to one or more adaptor proteins, and wherein the adaptor protein is associated with one or more functional domains; or, wherein the dead gRNA is modified to have at least one non-coding functional loop, and wherein the composition comprises two or more adaptor proteins, wherein the each protein is associated with one or more functional domains.

[0389] In certain embodiments, the adaptor protein is a fusion protein comprising the functional domain, the fusion protein optionally comprising a linker between the adaptor protein and the functional domain, the linker optionally including a GlySer linker.

[0390] In certain embodiments, the at least one loop of the dead gRNA is not modified by the insertion of distinct RNA sequence(s) that bind to the two or more adaptor proteins.

[0391] In certain embodiments, the one or more functional domains associated with the adaptor protein is a transcriptional activation domain.

[0392] In certain embodiments, the one or more functional domains associated with the adaptor protein is a transcriptional activation domain comprising VP64, p65, MyoD1, HSF1, RTA or SET7/9.

[0393] In certain embodiments, the one or more functional domains associated with the adaptor protein is a transcriptional repressor domain.

[0394] In certain embodiments, the transcriptional repressor domain is a KRAB domain.

[0395] In certain embodiments, the transcriptional repressor domain is a NuE domain, NcoR domain, SID domain or a SID4X domain.

[0396] In certain embodiments, at least one of the one or more functional domains associated with the adaptor protein have one or more activities comprising methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, DNA integration activity RNA cleavage activity, DNA cleavage activity or nucleic acid binding activity.

[0397] In certain embodiments, the DNA cleavage activity is due to a Fok1 nuclease.

[0398] In certain embodiments, the dead gRNA is modified so that, after dead gRNA binds the adaptor protein and further binds to the Cas13 and target, the functional domain is in a spatial orientation allowing for the functional domain to function in its attributed function.

[0399] In certain embodiments, the at least one loop of the dead gRNA is tetra loop and/or loop2. In certain embodiments, the tetra loop and loop 2 of the dead gRNA are modified by the insertion of the distinct RNA sequence(s).

[0400] In certain embodiments, the insertion of distinct RNA sequence(s) that bind to one or more adaptor proteins is an aptamer sequence. In certain embodiments, the aptamer sequence is two or more aptamer sequences specific to the same adaptor protein. In certain embodiments, the aptamer sequence is two or more aptamer sequences specific to different adaptor protein.

[0001] In certain embodiments, the adaptor protein comprises MS2, PP7, Qβ, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕCb5, ϕCb8r, ϕCb12r, ϕCb23r, 7s, PRR1.

[0401] In certain embodiments, the cell is a eukaryotic cell. In certain embodiments, the eukaryotic cell is a mammalian cell, optionally a mouse cell. In certain embodiments, the mammalian cell is a human cell.

[0402] In certain embodiments, a first adaptor protein is associated with a p65 domain and a second adaptor protein is associated with a HSF1 domain.

[0403] In certain embodiments, the composition comprises a Cas13 CRISPR-Cas complex having at least three functional domains, at least one of which is associated with the Cas13 and at least two of which are associated with dead gRNA.

[0404] In certain embodiments, the composition further comprises a second gRNA, wherein the second gRNA is a live gRNA capable of hybridizing to a second target sequence such that a second Cas13 CRISPR-Cas system is directed to a second genomic locus of interest in a cell with detectable indel activity at the second genomic locus resultant from nuclease activity of the Cas13 enzyme of the system.

[0405] In certain embodiments, the composition further comprises a plurality of dead gRNAs and/or a plurality of live gRNAs.

[0406] One aspect of the invention is to take advantage of the modularity and customizability of the gRNA scaffold to establish a series of gRNA scaffolds with different binding sites (in particular aptamers) for recruiting distinct types of effectors in an orthogonal manner. Again, for matters of example and illustration of the broader concept, replacement of the MS2 stem-loops with PP7-interacting stem-loops may be used to bind / recruit repressive elements, enabling multiplexed bidirectional transcriptional control. Thus, in general, gRNA comprising a dead guide may be employed to provide for multiplex transcriptional control and preferred bidirectional transcriptional control. This transcriptional control is most preferred of genes. For example, one or more gRNA comprising dead guide(s) may be employed in targeting the activation of one or more target genes. At the same time, one or more gRNA comprising dead guide(s) may be employed in targeting the repression of one or more target genes. Such a sequence may be applied in a variety of different combinations, for example the target genes are first repressed and then at an appropriate period other targets are activated, or select genes are repressed at the same time as select genes are activated, followed by further activation and/or repression. As a result, multiple components of one or more biological systems may advantageously be addressed together.

[0407] In an aspect, the invention provides nucleic acid molecule(s) encoding dead gRNA or the Cas13 CRISPR-Cas complex or the composition as described herein.

[0408] In an aspect, the invention provides a vector system comprising: a nucleic acid molecule encoding dead guide RNA as defined herein. In certain embodiments, the vector system further comprises a nucleic acid molecule(s) encoding Cas13. In certain embodiments, the vector system further comprises a nucleic acid molecule(s) encoding (live) gRNA. In

certain embodiments, the nucleic acid molecule or the vector further comprises regulatory element(s) operable in a eukaryotic cell operably linked to the nucleic acid molecule encoding the guide sequence (gRNA) and/or the nucleic acid molecule encoding Cas13 and/or the optional nuclear localization sequence(s).

[0409] In another aspect, structural analysis may also be used to study interactions between the dead guide and the active Cas nuclease that enable DNA binding, but no DNA cutting. In this way amino acids important for nuclease activity of Cas are determined. Modification of such amino acids allows for improved Cas enzymes used for gene editing.

[0410] A further aspect is combining the use of dead guides as explained herein with other applications of CRISPR, as explained herein as well as known in the art. For example, gRNA comprising dead guide(s) for targeted multiplex gene activation or repression or targeted multiplex bidirectional gene activation / repression may be combined with gRNA comprising guides which maintain nuclease activity, as explained herein. Such gRNA comprising guides which maintain nuclease activity may or may not further include modifications which allow for repression of gene activity (e.g. aptamers). Such gRNA comprising guides which maintain nuclease activity may or may not further include modifications which allow for activation of gene activity (e.g. aptamers). In such a manner, a further means for multiplex gene control is introduced (e.g. multiplex gene targeted activation without nuclease activity / without indel activity may be provided at the same time or in combination with gene targeted repression with nuclease activity).

[0411] For example, 1) using one or more gRNA (e.g. 1-50, 1-40, 1-30, 1-20, preferably 1-10, more preferably 1-5) comprising dead guide(s) targeted to one or more genes and further modified with appropriate aptamers for the recruitment of gene activators; 2) may be combined with one or more gRNA (e.g. 1-50, 1-40, 1-30, 1-20, preferably 1-10, more preferably 1-5) comprising dead guide(s) targeted to one or more genes and further modified with appropriate aptamers for the recruitment of gene repressors. 1) and/or 2) may then be combined with 3) one or more gRNA (e.g. 1-50, 1-40, 1-30, 1-20, preferably 1-10, more preferably 1-5) targeted to one or more genes. This combination can then be carried out in turn with 1) + 2) + 3) with 4) one or more gRNA (e.g. 1-50, 1-40, 1-30, 1-20, preferably 1-10, more preferably 1-5) targeted to one or more genes and further modified with appropriate aptamers for the recruitment of gene activators. This combination can then be carried in turn with 1) + 2) + 3) + 4) with 5) one or more gRNA (e.g. 1-50, 1-40, 1-30, 1-20, preferably 1-10, more preferably 1-5) targeted to one or more genes and further modified with appropriate aptamers for the recruitment of gene repressors. As a result various uses and combinations are included in the

invention. For example, combination 1) + 2); combination 1) + 3); combination 2) + 3); combination 1) + 2) + 3); combination 1) + 2) +3) +4); combination 1) + 3) + 4); combination 2) + 3) +4); combination 1) + 2) + 4); combination 1) + 2) +3) +4) + 5); combination 1) + 3) + 4) +5); combination 2) + 3) +4) +5); combination 1) + 2) + 4) +5); combination 1) + 2) +3) + 5); combination 1) + 3) +5); combination 2) + 3) +5); combination 1) + 2) +5).

[0412] In an aspect, the invention provides an algorithm for designing, evaluating, or selecting a dead guide RNA targeting sequence (dead guide sequence) for guiding a Cas13 CRISPR-Cas system to a target gene locus. In particular, it has been determined that dead guide RNA specificity relates to and can be optimized by varying i) GC content and ii) targeting sequence length. In an aspect, the invention provides an algorithm for designing or evaluating a dead guide RNA targeting sequence that minimizes off-target binding or interaction of the dead guide RNA. In an embodiment of the invention, the algorithm for selecting a dead guide RNA targeting sequence for directing a CRISPR system to a gene locus in an organism comprises a) locating one or more CRISPR motifs in the gene locus, analyzing the 20 nt sequence downstream of each CRISPR motif by i) determining the GC content of the sequence; and ii) determining whether there are off-target matches of the 15 downstream nucleotides nearest to the CRISPR motif in the genome of the organism, and c) selecting the 15 nucleotide sequence for use in a dead guide RNA if the GC content of the sequence is 70% or less and no off-target matches are identified. In an embodiment, the sequence is selected for a targeting sequence if the GC content is 60% or less. In certain embodiments, the sequence is selected for a targeting sequence if the GC content is 55% or less, 50% or less, 45% or less, 40% or less, 35% or less or 30% or less. In an embodiment, two or more sequences of the gene locus are analyzed and the sequence having the lowest GC content, or the next lowest GC content, or the next lowest GC content is selected. In an embodiment, the sequence is selected for a targeting sequence if no off-target matches are identified in the genome of the organism. In an embodiment, the targeting sequence is selected if no off-target matches are identified in regulatory sequences of the genome.

[0413] In an aspect, the invention provides a method of selecting a dead guide RNA targeting sequence for directing a functionalized CRISPR system to a gene locus in an organism, which comprises: a) locating one or more CRISPR motifs in the gene locus; b) analyzing the 20 nt sequence downstream of each CRISPR motif by: i) determining the GC content of the sequence; and ii) determining whether there are off-target matches of the first 15 nt of the sequence in the genome of the organism; c) selecting the sequence for use in a guide RNA if the GC content of the sequence is 70% or less and no off-target matches are identified.

In an embodiment, the sequence is selected if the GC content is 50% or less. In an embodiment, the sequence is selected if the GC content is 40% or less. In an embodiment, the sequence is selected if the GC content is 30% or less. In an embodiment, two or more sequences are analyzed and the sequence having the lowest GC content is selected. In an embodiment, off-target matches are determined in regulatory sequences of the organism. In an embodiment, the gene locus is a regulatory region. An aspect provides a dead guide RNA comprising the targeting sequence selected according to the aforementioned methods.

[0414] In an aspect, the invention provides a dead guide RNA for targeting a functionalized CRISPR system to a gene locus in an organism. In an embodiment of the invention, the dead guide RNA comprises a targeting sequence wherein the CG content of the target sequence is 70% or less, and the first 15 nt of the targeting sequence does not match an off-target sequence downstream from a CRISPR motif in the regulatory sequence of another gene locus in the organism. In certain embodiments, the GC content of the targeting sequence 60% or less, 55% or less, 50% or less, 45% or less, 40% or less, 35% or less or 30% or less. In certain embodiments, the GC content of the targeting sequence is from 70% to 60% or from 60% to 50% or from 50% to 40% or from 40% to 30%. In an embodiment, the targeting sequence has the lowest CG content among potential targeting sequences of the locus.

[0415] In an embodiment of the invention, the first 15 nt of the dead guide match the target sequence. In another embodiment, first 14 nt of the dead guide match the target sequence. In another embodiment, the first 13 nt of the dead guide match the target sequence. In another embodiment first 12 nt of the dead guide match the target sequence. In another embodiment, first 11 nt of the dead guide match the target sequence. In another embodiment, the first 10 nt of the dead guide match the target sequence. In an embodiment of the invention the first 15 nt of the dead guide does not match an off-target sequence downstream from a CRISPR motif in the regulatory region of another gene locus. In other embodiments, the first 14 nt, or the first 13 nt of the dead guide, or the first 12 nt of the guide, or the first 11 nt of the dead guide, or the first 10 nt of the dead guide, does not match an off-target sequence downstream from a CRISPR motif in the regulatory region of another gene locus. In other embodiments, the first 15 nt, or 14 nt, or 13 nt, or 12 nt, or 11 nt of the dead guide do not match an off-target sequence downstream from a CRISPR motif in the genome.

[0416] In certain embodiments, the dead guide RNA includes additional nucleotides at the 3'-end that do not match the target sequence. Thus, a dead guide RNA that includes the first 15 nt, or 14 nt, or 13 nt, or 12 nt, or 11 nt downstream of a CRISPR motif can be extended in length at the 3' end to 12 nt, 13 nt, 14 nt, 15 nt, 16 nt, 17 nt, 18 nt, 19 nt, 20 nt, or longer.

[0417] The invention provides a method for directing a Cas13 CRISPR-Cas system, including but not limited to a dead Cas13 (dCas13) or functionalized Cas13 system (which may comprise a functionalized Cas13 or functionalized guide) to a gene locus. In an aspect, the invention provides a method for selecting a dead guide RNA targeting sequence and directing a functionalized CRISPR system to a gene locus in an organism. In an aspect, the invention provides a method for selecting a dead guide RNA targeting sequence and effecting gene regulation of a target gene locus by a functionalized Cas13 CRISPR-Cas system. In certain embodiments, the method is used to effect target gene regulation while minimizing off-target effects. In an aspect, the invention provides a method for selecting two or more dead guide RNA targeting sequences and effecting gene regulation of two or more target gene loci by a functionalized Cas13 CRISPR-Cas system. In certain embodiments, the method is used to effect regulation of two or more target gene loci while minimizing off-target effects.

[0418] In an aspect, the invention provides a method of selecting a dead guide RNA targeting sequence for directing a functionalized Cas13 to a gene locus in an organism, which comprises: a) locating one or more CRISPR motifs in the gene locus; b) analyzing the sequence downstream of each CRISPR motif by: i) selecting 10 to 15 nt adjacent to the CRISPR motif, ii) determining the GC content of the sequence; and c) selecting the 10 to 15 nt sequence as a targeting sequence for use in a guide RNA if the GC content of the sequence is 40% or more. In an embodiment, the sequence is selected if the GC content is 50% or more. In an embodiment, the sequence is selected if the GC content is 60% or more. In an embodiment, the sequence is selected if the GC content is 70% or more. In an embodiment, two or more sequences are analyzed and the sequence having the highest GC content is selected. In an embodiment, the method further comprises adding nucleotides to the 3' end of the selected sequence which do not match the sequence downstream of the CRISPR motif. An aspect provides a dead guide RNA comprising the targeting sequence selected according to the aforementioned methods.

[0419] In an aspect, the invention provides a dead guide RNA for directing a functionalized CRISPR system to a gene locus in an organism wherein the targeting sequence of the dead guide RNA consists of 10 to 15 nucleotides adjacent to the CRISPR motif of the gene locus, wherein the CG content of the target sequence is 50% or more. In certain embodiments, the dead guide RNA further comprises nucleotides added to the 3' end of the targeting sequence which do not match the sequence downstream of the CRISPR motif of the gene locus.

[0420] In an aspect, the invention provides for a single effector to be directed to one or more, or two or more gene loci. In certain embodiments, the effector is associated with a Cas13, and one or more, or two or more selected dead guide RNAs are used to direct the Cas13-associated effector to one or more, or two or more selected target gene loci. In certain embodiments, the effector is associated with one or more, or two or more selected dead guide RNAs, each selected dead guide RNA, when complexed with a Cas13 enzyme, causing its associated effector to localize to the dead guide RNA target. One non-limiting example of such CRISPR systems modulates activity of one or more, or two or more gene loci subject to regulation by the same transcription factor.

[0421] In an aspect, the invention provides for two or more effectors to be directed to one or more gene loci. In certain embodiments, two or more dead guide RNAs are employed, each of the two or more effectors being associated with a selected dead guide RNA, with each of the two or more effectors being localized to the selected target of its dead guide RNA. One non-limiting example of such CRISPR systems modulates activity of one or more, or two or more gene loci subject to regulation by different transcription factors. Thus, in one non-limiting embodiment, two or more transcription factors are localized to different regulatory sequences of a single gene. In another non-limiting embodiment, two or more transcription factors are localized to different regulatory sequences of different genes. In certain embodiments, one transcription factor is an activator. In certain embodiments, one transcription factor is an inhibitor. In certain embodiments, one transcription factor is an activator and another transcription factor is an inhibitor. In certain embodiments, gene loci expressing different components of the same regulatory pathway are regulated. In certain embodiments, gene loci expressing components of different regulatory pathways are regulated.

[0422] In an aspect, the invention also provides a method and algorithm for designing and selecting dead guide RNAs that are specific for target DNA cleavage or target binding and gene regulation mediated by an active Cas13 CRISPR-Cas system. In certain embodiments, the Cas13 CRISPR-Cas system provides orthogonal gene control using an active Cas13 which cleaves target DNA at one gene locus while at the same time binds to and promotes regulation of another gene locus.

[0423] In an aspect, the invention provides an method of selecting a dead guide RNA targeting sequence for directing a functionalized Cas13 to a gene locus in an organism, without cleavage, which comprises a) locating one or more CRISPR motifs in the gene locus; b) analyzing the sequence downstream of each CRISPR motif by i) selecting 10 to 15 nt adjacent to the CRISPR motif, ii) determining the GC content of the sequence, and c) selecting the 10

to 15 nt sequence as a targeting sequence for use in a dead guide RNA if the GC content of the sequence is 30% more, 40% or more. In certain embodiments, the GC content of the targeting sequence is 35% or more, 40% or more, 45% or more, 50% or more, 55% or more, 60% or more, 65% or more, or 70% or more. In certain embodiments, the GC content of the targeting sequence is from 30% to 40% or from 40% to 50% or from 50% to 60% or from 60% to 70%. In an embodiment of the invention, two or more sequences in a gene locus are analyzed and the sequence having the highest GC content is selected.

[0424] In an embodiment of the invention, the portion of the targeting sequence in which GC content is evaluated is 10 to 15 contiguous nucleotides of the 15 target nucleotides nearest to the PAM. In an embodiment of the invention, the portion of the guide in which GC content is considered is the 10 to 11 nucleotides or 11 to 12 nucleotides or 12 to 13 nucleotides or 13, or 14, or 15 contiguous nucleotides of the 15 nucleotides nearest to the PAM.

[0425] In an aspect, the invention further provides an algorithm for identifying dead guide RNAs which promote CRISPR system gene locus cleavage while avoiding functional activation or inhibition. It is observed that increased GC content in dead guide RNAs of 16 to 20 nucleotides coincides with increased DNA cleavage and reduced functional activation.

[0426] It is also demonstrated herein that efficiency of functionalized Cas13 can be increased by addition of nucleotides to the 3' end of a guide RNA which do not match a target sequence downstream of the CRISPR motif. For example, of dead guide RNA 11 to 15 nt in length, shorter guides may be less likely to promote target cleavage, but are also less efficient at promoting CRISPR system binding and functional control. In certain embodiments, addition of nucleotides that don't match the target sequence to the 3' end of the dead guide RNA increase activation efficiency while not increasing undesired target cleavage. In an aspect, the invention also provides a method and algorithm for identifying improved dead guide RNAs that effectively promote CRISPRP system function in DNA binding and gene regulation while not promoting DNA cleavage. Thus, in certain embodiments, the invention provides a dead guide RNA that includes the first 15 nt, or 14 nt, or 13 nt, or 12 nt, or 11 nt downstream of a CRISPR motif and is extended in length at the 3' end by nucleotides that mismatch the target to 12 nt, 13 nt, 14 nt, 15 nt, 16 nt, 17 nt, 18 nt, 19 nt, 20 nt, or longer.

[0427] In an aspect, the invention provides a method for effecting selective orthogonal gene control. As will be appreciated from the disclosure herein, dead guide selection according to the invention, taking into account guide length and GC content, provides effective and selective transcription control by a functional Cas13 CRISPR-Cas system, for example to regulate transcription of a gene locus by activation or inhibition and minimize off-target effects.

Accordingly, by providing effective regulation of individual target loci, the invention also provides effective orthogonal regulation of two or more target loci.

[0428] In certain embodiments, orthogonal gene control is by activation or inhibition of two or more target loci. In certain embodiments, orthogonal gene control is by activation or inhibition of one or more target locus and cleavage of one or more target locus.

[0429] In one aspect, the invention provides a cell comprising a non-naturally occurring Cas13 CRISPR-Cas system comprising one or more dead guide RNAs disclosed or made according to a method or algorithm described herein wherein the expression of one or more gene products has been altered. In an embodiment of the invention, the expression in the cell of two or more gene products has been altered. The invention also provides a cell line from such a cell.

[0430] In one aspect, the invention provides a multicellular organism comprising one or more cells comprising a non-naturally occurring Cas13 CRISPR-Cas system comprising one or more dead guide RNAs disclosed or made according to a method or algorithm described herein. In one aspect, the invention provides a product from a cell, cell line, or multicellular organism comprising a non-naturally occurring Cas13 CRISPR-Cas system comprising one or more dead guide RNAs disclosed or made according to a method or algorithm described herein.

[0431] A further aspect of this invention is the use of gRNA comprising dead guide(s) as described herein, optionally in combination with gRNA comprising guide(s) as described herein or in the state of the art, in combination with systems e.g. cells, transgenic animals, transgenic mice, inducible transgenic animals, inducible transgenic mice) which are engineered for either overexpression of Cas13 or preferably knock in Cas13. As a result a single system (e.g. transgenic animal, cell) can serve as a basis for multiplex gene modifications in systems / network biology. On account of the dead guides, this is now possible in both in vitro, ex vivo, and in vivo.

[0432] For example, once the Cas13 is provided for, one or more dead gRNAs may be provided to direct multiplex gene regulation, and preferably multiplex bidirectional gene regulation. The one or more dead gRNAs may be provided in a spatially and temporally appropriate manner if necessary or desired (for example tissue specific induction of Cas13 expression). On account that the transgenic / inducible Cas13 is provided for (e.g. expressed) in the cell, tissue, animal of interest, both gRNAs comprising dead guides or gRNAs comprising guides are equally effective. In the same manner, a further aspect of this invention is the use of gRNA comprising dead guide(s) as described herein, optionally in combination with gRNA comprising guide(s) as described herein or in the state of the art, in combination

with systems (e.g. cells, transgenic animals, transgenic mice, inducible transgenic animals, inducible transgenic mice) which are engineered for knockout Cas13 CRISPR-Cas.

[0433] As a result, the combination of dead guides as described herein with CRISPR applications described herein and CRISPR applications known in the art results in a highly efficient and accurate means for multiplex screening of systems (e.g. network biology). Such screening allows, for example, identification of specific combinations of gene activities for identifying genes responsible for diseases (e.g. on/off combinations), in particular gene related diseases. A preferred application of such screening is cancer. In the same manner, screening for treatment for such diseases is included in the invention. Cells or animals may be exposed to aberrant conditions resulting in disease or disease like effects. Candidate compositions may be provided and screened for an effect in the desired multiplex environment. For example a patient's cancer cells may be screened for which gene combinations will cause them to die, and then use this information to establish appropriate therapies.

[0434] In one aspect, the invention provides a kit comprising one or more of the components described herein. The kit may include dead guides as described herein with or without guides as described herein.

[0435] The structural information provided herein allows for interrogation of dead gRNA interaction with the target DNA and the Cas13 permitting engineering or alteration of dead gRNA structure to optimize functionality of the entire Cas13 CRISPR-Cas system. For example, loops of the dead gRNA may be extended, without colliding with the Cas13 protein by the insertion of adaptor proteins that can bind to RNA. These adaptor proteins can further recruit effector proteins or fusions which comprise one or more functional domains.

[0436] In some preferred embodiments, the functional domain is a transcriptional activation domain, preferably VP64. In some embodiments, the functional domain is a transcription repression domain, preferably KRAB. In some embodiments, the transcription repression domain is SID, or concatemers of SID (e.g. SID4X). In some embodiments, the functional domain is an epigenetic modifying domain, such that an epigenetic modifying enzyme is provided. In some embodiments, the functional domain is an activation domain, which may be the P65 activation domain.

[0437] An aspect of the invention is that the above elements are comprised in a single composition or comprised in individual compositions. These compositions may advantageously be applied to a host to elicit a functional effect on the genomic level.

[0438] In general, the dead gRNA is modified in a manner that provides specific binding sites (e.g. aptamers) for adapter proteins comprising one or more functional domains (e.g. via

fusion protein) to bind to. The modified dead gRNA is modified such that once the dead gRNA forms a CRISPR complex (i.e. Cas13 binding to dead gRNA and target) the adapter proteins bind and, the functional domain on the adapter protein is positioned in a spatial orientation which is advantageous for the attributed function to be effective. For example, if the functional domain is a transcription activator (e.g. VP64 or p65), the transcription activator is placed in a spatial orientation which allows it to affect the transcription of the target. Likewise, a transcription repressor will be advantageously positioned to affect the transcription of the target and a nuclease (e.g. Fok1) will be advantageously positioned to cleave or partially cleave the target.

[0439] The skilled person will understand that modifications to the dead gRNA which allow for binding of the adapter + functional domain but not proper positioning of the adapter + functional domain (e.g. due to steric hindrance within the three dimensional structure of the CRISPR complex) are modifications which are not intended. The one or more modified dead gRNA may be modified at the tetra loop, the stem loop 1, stem loop 2, or stem loop 3, as described herein, preferably at either the tetra loop or stem loop 2, and most preferably at both the tetra loop and stem loop 2.

[0440] As explained herein the functional domains may be, for example, one or more domains from the group consisting of methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, DNA cleavage activity, nucleic acid binding activity, and molecular switches (e.g. light inducible). In some cases it is advantageous that additionally at least one NLS is provided. In some instances, it is advantageous to position the NLS at the N terminus. When more than one functional domain is included, the functional domains may be the same or different.

[0441] The dead gRNA may be designed to include multiple binding recognition sites (e.g. aptamers) specific to the same or different adapter protein. The dead gRNA may be designed to bind to the promoter region -1000 - +1 nucleic acids upstream of the transcription start site (i.e. TSS), preferably -200 nucleic acids. This positioning improves functional domains which affect gene activation (e.g. transcription activators) or gene inhibition (e.g. transcription repressors). The modified dead gRNA may be one or more modified dead gRNAs targeted to one or more target loci (e.g. at least 1 gRNA, at least 2 gRNA, at least 5 gRNA, at least 10 gRNA, at least 20 gRNA, at least 30 gRNA, at least 50 gRNA) comprised in a composition.

[0442] The adaptor protein may be any number of proteins that binds to an aptamer or recognition site introduced into the modified dead gRNA and which allows proper positioning of one or more functional domains, once the dead gRNA has been incorporated into the CRISPR complex, to affect the target with the attributed function. As explained in detail in this application such may be coat proteins, preferably bacteriophage coat proteins. The functional domains associated with such adaptor proteins (e.g. in the form of fusion protein) may include, for example, one or more domains from the group consisting of methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, DNA cleavage activity, nucleic acid binding activity, and molecular switches (e.g. light inducible). Preferred domains are Fok1, VP64, P65, HSF1, MyoD1. In the event that the functional domain is a transcription activator or transcription repressor it is advantageous that additionally at least an NLS is provided and preferably at the N terminus. When more than one functional domain is included, the functional domains may be the same or different. The adaptor protein may utilize known linkers to attach such functional domains.

[0443] Thus, the modified dead gRNA, the (inactivated) Cas13 (with or without functional domains), and the binding protein with one or more functional domains, may each individually be comprised in a composition and administered to a host individually or collectively. Alternatively, these components may be provided in a single composition for administration to a host. Administration to a host may be performed via viral vectors known to the skilled person or described herein for delivery to a host (e.g. lentiviral vector, adenoviral vector, AAV vector). As explained herein, use of different selection markers (e.g. for lentiviral gRNA selection) and concentration of gRNA (e.g. dependent on whether multiple gRNAs are used) may be advantageous for eliciting an improved effect.

[0444] On the basis of this concept, several variations are appropriate to elicit a genomic locus event, including DNA cleavage, gene activation, or gene deactivation. Using the provided compositions, the person skilled in the art can advantageously and specifically target single or multiple loci with the same or different functional domains to elicit one or more genomic locus events. The compositions may be applied in a wide variety of methods for screening in libraries in cells and functional modeling in vivo (e.g. gene activation of lincRNA and identification of function; gain-of-function modeling; loss-of-function modeling; the use the compositions of the invention to establish cell lines and transgenic animals for optimization and screening purposes).

[0445] The current invention comprehends the use of the compositions of the current invention to establish and utilize conditional or inducible CRISPR transgenic cell /animals, which are not believed prior to the present invention or application. For example, the target cell comprises Cas13 conditionally or inducibly (e.g. in the form of Cre dependent constructs) and/or the adapter protein conditionally or inducibly and, on expression of a vector introduced into the target cell, the vector expresses that which induces or gives rise to the condition of Cas13 expression and/or adaptor expression in the target cell. By applying the teaching and compositions of the current invention with the known method of creating a CRISPR complex, inducible genomic events affected by functional domains are also an aspect of the current invention. One example of this is the creation of a CRISPR knock-in / conditional transgenic animal (e.g. mouse comprising e.g. a Lox-Stop-polyA-Lox(LSL) cassette) and subsequent delivery of one or more compositions providing one or more modified dead gRNA (e.g. -200 nucleotides to TSS of a target gene of interest for gene activation purposes) as described herein (e.g. modified dead gRNA with one or more aptamers recognized by coat proteins, e.g. MS2), one or more adapter proteins as described herein (MS2 binding protein linked to one or more VP64) and means for inducing the conditional animal (e.g. Cre recombinase for rendering Cas13 expression inducible). Alternatively, the adaptor protein may be provided as a conditional or inducible element with a conditional or inducible Cas13 to provide an effective model for screening purposes, which advantageously only requires minimal design and administration of specific dead gRNAs for a broad number of applications.

[0446] In another aspect the dead guides are further modified to improve specificity. Protected dead guides may be synthesized, whereby secondary structure is introduced into the 3' end of the dead guide to improve its specificity. A protected guide RNA (pgRNA) comprises a guide sequence capable of hybridizing to a target sequence in a genomic locus of interest in a cell and a protector strand, wherein the protector strand is optionally complementary to the guide sequence and wherein the guide sequence may in part be hybridizable to the protector strand. The pgRNA optionally includes an extension sequence. The thermodynamics of the pgRNA-target DNA hybridization is determined by the number of bases complementary between the guide RNA and target DNA. By employing 'thermodynamic protection', specificity of dead gRNA can be improved by adding a protector sequence. For example, one method adds a complementary protector strand of varying lengths to the 3' end of the guide sequence within the dead gRNA. As a result, the protector strand is bound to at least a portion of the dead gRNA and provides for a protected gRNA (pgRNA). In turn, the dead gRNA references herein may be easily protected using the described embodiments, resulting in

pgRNA. The protector strand can be either a separate RNA transcript or strand or a chimeric version joined to the 3' end of the dead gRNA guide sequence.

Tandem guides and uses in a multiplex (tandem) targeting approach

[0447] The inventors have shown that CRISPR enzymes as defined herein can employ more than one RNA guide without losing activity. This enables the use of the CRISPR enzymes, systems or complexes as defined herein for targeting multiple DNA targets, genes or gene loci, with a single enzyme, system or complex as defined herein. The guide RNAs may be tandemly arranged, optionally separated by a nucleotide sequence such as a direct repeat as defined herein. The position of the different guide RNAs is the tandem does not influence the activity. It is noted that the terms "CRISPR-Cas system", "CRISP-Cas complex" "CRISPR complex" and "CRISPR system" are used interchangeably. Also the terms "CRISPR enzyme", "Cas enzyme", or "CRISPR-Cas enzyme", can be used interchangeably. In preferred embodiments, said CRISPR enzyme, CRISP-Cas enzyme or Cas enzyme is Cas13, or any one of the modified or mutated variants thereof described herein elsewhere.

[0448] In one aspect, the invention provides a non-naturally occurring or engineered CRISPR enzyme, preferably a class 2 CRISPR enzyme, preferably a Type V or VI CRISPR enzyme as described herein, such as without limitation Cas13 as described herein elsewhere, used for tandem or multiplex targeting. It is to be understood that any of the CRISPR (or CRISPR-Cas or Cas) enzymes, complexes, or systems according to the invention as described herein elsewhere may be used in such an approach. Any of the methods, products, compositions and uses as described herein elsewhere are equally applicable with the multiplex or tandem targeting approach further detailed below. By means of further guidance, the following particular aspects and embodiments are provided.

[0449] In one aspect, the invention provides for the use of a Cas13 enzyme, complex or system as defined herein for targeting multiple gene loci. In one embodiment, this can be established by using multiple (tandem or multiplex) guide RNA (gRNA) sequences.

[0450] In one aspect, the invention provides methods for using one or more elements of a Cas13 enzyme, complex or system as defined herein for tandem or multiplex targeting, wherein said CRISP system comprises multiple guide RNA sequences. Preferably, said gRNA sequences are separated by a nucleotide sequence, such as a direct repeat as defined herein elsewhere.

[0451] The Cas13 enzyme, system or complex as defined herein provides an effective means for modifying multiple target polynucleotides. The Cas13 enzyme, system or complex as defined herein has a wide variety of utility including modifying (e.g., deleting, inserting,

translocating, inactivating, activating) one or more target polynucleotides in a multiplicity of cell types. As such the Cas13 enzyme, system or complex as defined herein of the invention has a broad spectrum of applications in, e.g., gene therapy, drug screening, disease diagnosis, and prognosis, including targeting multiple gene loci within a single CRISPR system.

[0452] In one aspect, the invention provides a Cas13 enzyme, system or complex as defined herein, i.e. a Cas13 CRISPR-Cas complex having a Cas13 protein having at least one destabilization domain associated therewith, and multiple guide RNAs that target multiple nucleic acid molecules such as DNA molecules, whereby each of said multiple guide RNAs specifically targets its corresponding nucleic acid molecule, e.g., DNA molecule. Each nucleic acid molecule target, e.g., DNA molecule can encode a gene product or encompass a gene locus. Using multiple guide RNAs hence enables the targeting of multiple gene loci or multiple genes. In some embodiments the Cas13 enzyme may cleave the RNA molecule encoding the gene product. In some embodiments expression of the gene product is altered. The Cas13 protein and the guide RNAs do not naturally occur together. The invention comprehends the guide RNAs comprising tandemly arranged guide sequences. The invention further comprehends coding sequences for the Cas13 protein being codon optimized for expression in a eukaryotic cell. In a preferred embodiment the eukaryotic cell is a mammalian cell, a plant cell or a yeast cell and in a more preferred embodiment the mammalian cell is a human cell. Expression of the gene product may be decreased. The Cas13 enzyme may form part of a CRISPR system or complex, which further comprises tandemly arranged guide RNAs (gRNAs) comprising a series of 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 25, 25, 30, or more than 30 guide sequences, each capable of specifically hybridizing to a target sequence in a genomic locus of interest in a cell. In some embodiments, the functional Cas13 CRISPR system or complex binds to the multiple target sequences. In some embodiments, the functional CRISPR system or complex may edit the multiple target sequences, e.g., the target sequences may comprise a genomic locus, and in some embodiments there may be an alteration of gene expression. In some embodiments, the functional CRISPR system or complex may comprise further functional domains. In some embodiments, the invention provides a method for altering or modifying expression of multiple gene products. The method may comprise introducing into a cell containing said target nucleic acids, e.g., DNA molecules, or containing and expressing target nucleic acid, e.g., DNA molecules; for instance, the target nucleic acids may encode gene products or provide for expression of gene products (e.g., regulatory sequences).

[0453] In preferred embodiments the CRISPR enzyme used for multiplex targeting is Cas13, or the CRISPR system or complex comprises Cas13. In some embodiments, the

CRISPR enzyme used for multiplex targeting is AsCas13, or the CRISPR system or complex used for multiplex targeting comprises an AsCas13. In some embodiments, the CRISPR enzyme is an LbCas13, or the CRISPR system or complex comprises LbCas13. In some embodiments, the Cas enzyme used for multiplex targeting cleaves both strands of DNA to produce a double strand break (DSB). In some embodiments, the CRISPR enzyme used for multiplex targeting is a nickase. In some embodiments, the Cas13 enzyme used for multiplex targeting is a dual nickase. In some embodiments, the Cas13 enzyme used for multiplex targeting is a Cas13 enzyme such as a DD Cas13 enzyme as defined herein elsewhere.

[0454] In some general embodiments, the Cas13 enzyme used for multiplex targeting is associated with one or more functional domains. In some more specific embodiments, the CRISPR enzyme used for multiplex targeting is a deadCas13 as defined herein elsewhere.

[0455] In an aspect, the present invention provides a means for delivering the Cas13 enzyme, system or complex for use in multiple targeting as defined herein or the polynucleotides defined herein. Non-limiting examples of such delivery means are e.g. particle(s) delivering component(s) of the complex, vector(s) comprising the polynucleotide(s) discussed herein (e.g., encoding the CRISPR enzyme, providing the nucleotides encoding the CRISPR complex). In some embodiments, the vector may be a plasmid or a viral vector such as AAV, or lentivirus. Transient transfection with plasmids, e.g., into HEK cells may be advantageous, especially given the size limitations of AAV and that while Cas13 fits into AAV, one may reach an upper limit with additional guide RNAs.

[0456] Also provided is a model that constitutively expresses the Cas13 enzyme, complex or system as used herein for use in multiplex targeting. The organism may be transgenic and may have been transfected with the present vectors or may be the offspring of an organism so transfected. In a further aspect, the present invention provides compositions comprising the CRISPR enzyme, system and complex as defined herein or the polynucleotides or vectors described herein. Also provides are Cas13 CRISPR systems or complexes comprising multiple guide RNAs, preferably in a tandemly arranged format. Said different guide RNAs may be separated by nucleotide sequences such as direct repeats.

[0457] Also provided is a method of treating a subject, e.g., a subject in need thereof, comprising inducing gene editing by transforming the subject with the polynucleotide encoding the Cas13 CRISPR system or complex or any of polynucleotides or vectors described herein and administering them to the subject. A suitable repair template may also be provided, for example delivered by a vector comprising said repair template. Also provided is a method of treating a subject, e.g., a subject in need thereof, comprising inducing transcriptional activation

or repression of multiple target gene loci by transforming the subject with the polynucleotides or vectors described herein, wherein said polynucleotide or vector encodes or comprises the Cas13 enzyme, complex or system comprising multiple guide RNAs, preferably tandemly arranged. Where any treatment is occurring ex vivo, for example in a cell culture, then it will be appreciated that the term 'subject' may be replaced by the phrase "cell or cell culture."

[0458] Compositions comprising Cas13 enzyme, complex or system comprising multiple guide RNAs, preferably tandemly arranged, or the polynucleotide or vector encoding or comprising said Cas13 enzyme, complex or system comprising multiple guide RNAs, preferably tandemly arranged, for use in the methods of treatment as defined herein elsewhere are also provided. A kit of parts may be provided including such compositions. Use of said composition in the manufacture of a medicament for such methods of treatment are also provided. Use of a Cas13 CRISPR system in screening is also provided by the present invention, e.g., gain of function screens. Cells which are artificially forced to overexpress a gene are be able to down regulate the gene over time (re-establishing equilibrium) e.g. by negative feedback loops. By the time the screen starts the unregulated gene might be reduced again. Using an inducible Cas13 activator allows one to induce transcription right before the screen and therefore minimizes the chance of false negative hits. Accordingly, by use of the instant invention in screening, e.g., gain of function screens, the chance of false negative results may be minimized.

[0459] In one aspect, the invention provides an engineered, non-naturally occurring CRISPR system comprising a Cas13 protein and multiple guide RNAs that each specifically target a DNA molecule encoding a gene product in a cell, whereby the multiple guide RNAs each target their specific DNA molecule encoding the gene product and the Cas13 protein cleaves the target DNA molecule encoding the gene product, whereby expression of the gene product is altered; and, wherein the CRISPR protein and the guide RNAs do not naturally occur together. The invention comprehends the multiple guide RNAs comprising multiple guide sequences, preferably separated by a nucleotide sequence such as a direct repeat and optionally fused to a tracr sequence. In an embodiment of the invention the CRISPR protein is a type V or VI CRISPR-Cas protein and in a more preferred embodiment the CRISPR protein is a Cas13 protein. The invention further comprehends a Cas13 protein being codon optimized for expression in a eukaryotic cell. In a preferred embodiment the eukaryotic cell is a mammalian cell and in a more preferred embodiment the mammalian cell is a human cell. In a further embodiment of the invention, the expression of the gene product is decreased.

[0460] In another aspect, the invention provides an engineered, non-naturally occurring vector system comprising one or more vectors comprising a first regulatory element operably linked to the multiple Cas13 CRISPR system guide RNAs that each specifically target a DNA molecule encoding a gene product and a second regulatory element operably linked coding for a CRISPR protein. Both regulatory elements may be located on the same vector or on different vectors of the system. The multiple guide RNAs target the multiple DNA molecules encoding the multiple gene products in a cell and the CRISPR protein may cleave the multiple DNA molecules encoding the gene products (it may cleave one or both strands or have substantially no nuclease activity), whereby expression of the multiple gene products is altered; and, wherein the CRISPR protein and the multiple guide RNAs do not naturally occur together. In a preferred embodiment the CRISPR protein is Cas13 protein, optionally codon optimized for expression in a eukaryotic cell. In a preferred embodiment the eukaryotic cell is a mammalian cell, a plant cell or a yeast cell and in a more preferred embodiment the mammalian cell is a human cell. In a further embodiment of the invention, the expression of each of the multiple gene products is altered, preferably decreased.

[0461] In one aspect, the invention provides a vector system comprising one or more vectors. In some embodiments, the system comprises: (a) a first regulatory element operably linked to a direct repeat sequence and one or more insertion sites for inserting one or more guide sequences up- or downstream (whichever applicable) of the direct repeat sequence, wherein when expressed, the one or more guide sequence(s) direct(s) sequence-specific binding of the CRISPR complex to the one or more target sequence(s) in a eukaryotic cell, wherein the CRISPR complex comprises a Cas13 enzyme complexed with the one or more guide sequence(s) that is hybridized to the one or more target sequence(s); and (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said Cas13 enzyme, preferably comprising at least one nuclear localization sequence and/or at least one NES; wherein components (a) and (b) are located on the same or different vectors of the system. Where applicable, a tracr sequence may also be provided. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a Cas13 CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the CRISPR complex comprises one or more nuclear localization sequences and/or one or more NES of sufficient strength to drive accumulation of said Cas13 CRISPR complex in a detectable amount in or out of the nucleus of a eukaryotic cell. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments,

the second regulatory element is a polymerase II promoter. In some embodiments, each of the guide sequences is at least 16, 17, 18, 19, 20, 25 nucleotides, or between 16-30, or between 16-25, or between 16-20 nucleotides in length.

[0462] Recombinant expression vectors can comprise the polynucleotides encoding the Cas13 enzyme, system or complex for use in multiple targeting as defined herein in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g., in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell).

[0463] In some embodiments, a host cell is transiently or non-transiently transfected with one or more vectors comprising the polynucleotides encoding the Cas13 enzyme, system or complex for use in multiple targeting as defined herein. In some embodiments, a cell is transfected as it naturally occurs in a subject. In some embodiments, a cell that is transfected is taken from a subject. In some embodiments, the cell is derived from cells taken from a subject, such as a cell line. A wide variety of cell lines for tissue culture are known in the art and exemplified herein elsewhere. Cell lines are available from a variety of sources known to those with skill in the art (see, e.g., the American Type Culture Collection (ATCC) (Manassus, Va.)). In some embodiments, a cell transfected with one or more vectors comprising the polynucleotides encoding the Cas13 enzyme, system or complex for use in multiple targeting as defined herein is used to establish a new cell line comprising one or more vector-derived sequences. In some embodiments, a cell transiently transfected with the components of a Cas13 CRISPR system or complex for use in multiple targeting as described herein (such as by transient transfection of one or more vectors, or transfection with RNA), and modified through the activity of a Cas13 CRISPR system or complex, is used to establish a new cell line comprising cells containing the modification but lacking any other exogenous sequence. In some embodiments, cells transiently or non-transiently transfected with one or more vectors comprising the polynucleotides encoding the Cas13 enzyme, system or complex for use in multiple targeting as defined herein, or cell lines derived from such cells are used in assessing one or more test compounds.

[0464] The term "regulatory element" is as defined herein elsewhere.

[0465] Advantageous vectors include lentiviruses and adeno-associated viruses, and types of such vectors can also be selected for targeting particular types of cells.

[0466] In one aspect, the invention provides a eukaryotic host cell comprising (a) a first regulatory element operably linked to a direct repeat sequence and one or more insertion sites for inserting one or more guide RNA sequences up- or downstream (whichever applicable) of the direct repeat sequence, wherein when expressed, the guide sequence(s) direct(s) sequence-specific binding of the Cas13 CRISPR complex to the respective target sequence(s) in a eukaryotic cell, wherein the Cas13 CRISPR complex comprises a Cas13 enzyme complexed with the one or more guide sequence(s) that is hybridized to the respective target sequence(s); and/or (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said Cas13 enzyme comprising preferably at least one nuclear localization sequence and/or NES. In some embodiments, the host cell comprises components (a) and (b). Where applicable, a tracr sequence may also be provided. In some embodiments, component (a), component (b), or components (a) and (b) are stably integrated into a genome of the host eukaryotic cell. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, and optionally separated by a direct repeat, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a Cas13 CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the Cas13 enzyme comprises one or more nuclear localization sequences and/or nuclear export sequences or NES of sufficient strength to drive accumulation of said CRISPR enzyme in a detectable amount in and/or out of the nucleus of a eukaryotic cell.

[0467] In some embodiments, the Cas13 enzyme is a type V or VI CRISPR system enzyme. In some embodiments, the Cas enzyme is a Cas13 enzyme. In some embodiments, the Cas13 enzyme is derived from Francisella tularensis 1, Francisella tularensis subsp. novicida, Prevotella albensis, Lachnospiraceae bacterium MC2017 1, Butyrivibrio proteoclasticus, Peregrinibacteria bacterium GW2011_GWA2_33_10, Parcubacteria bacterium GW2011_GWC2_44_17, Smithella sp. SCADC, Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020, Candidatus Methanoplasma termitum, Eubacterium eligens, Moraxella bovoculi 237, Leptospira inadai, Lachnospiraceae bacterium ND2006, Porphyromonas crevioricanis 3, Prevotella disiens, or Porphyromonas macacae Cas13, and may include further alterations or mutations of the Cas13 as defined herein elsewhere, and can be a chimeric Cas13. In some embodiments, the Cas13 enzyme is codon-optimized for expression in a eukaryotic cell. In some embodiments, the CRISPR enzyme directs cleavage

of one or two strands at the location of the target sequence. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In some embodiments, the one or more guide sequence(s) is (are each) at least 16, 17, 18, 19, 20, 25 nucleotides, or between 16-30, or between 16-25, or between 16-20 nucleotides in length. When multiple guide RNAs are used, they are preferably separated by a direct repeat sequence. In an aspect, the invention provides a non-human eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. In other aspects, the invention provides a eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. The organism in some embodiments of these aspects may be an animal; for example a mammal. Also, the organism may be an arthropod such as an insect. The organism also may be a plant. Further, the organism may be a fungus.

[0468] In one aspect, the invention provides a kit comprising one or more of the components described herein. In some embodiments, the kit comprises a vector system and instructions for using the kit. In some embodiments, the vector system comprises (a) a first regulatory element operably linked to a direct repeat sequence and one or more insertion sites for inserting one or more guide sequences up- or downstream (whichever applicable) of the direct repeat sequence, wherein when expressed, the guide sequence directs sequence-specific binding of a Cas13 CRISPR complex to a target sequence in a eukaryotic cell, wherein the Cas13 CRISPR complex comprises a Cas13 enzyme complexed with the guide sequence that is hybridized to the target sequence; and/or (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said Cas13 enzyme comprising a nuclear localization sequence. Where applicable, a tracr sequence may also be provided. In some embodiments, the kit comprises components (a) and (b) located on the same or different vectors of the system. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the Cas13 enzyme comprises one or more nuclear localization sequences of sufficient strength to drive accumulation of said CRISPR enzyme in a detectable amount in the nucleus of a eukaryotic cell. In some embodiments, the CRISPR enzyme is a type V or VI CRISPR system enzyme. In some embodiments, the CRISPR enzyme is a Cas13 enzyme. In some embodiments, the Cas13 enzyme is derived from Francisella tularensis 1, Francisella tularensis subsp. novicida, Prevotella albensis, Lachnospiraceae

bacterium MC2017 1, Butyrivibrio proteoclasticus, Peregrinibacteria bacterium GW2011_GWA2_33_10, Parcubacteria bacterium GW2011_GWC2_44_17, Smithella sp. SCADC, Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020, Candidatus Methanoplasma termitum, Eubacterium eligens, Moraxella bovoculi 237, Leptospira inadai, Lachnospiraceae bacterium ND2006, Porphyromonas crevioricanis 3, Prevotella disiens, or Porphyromonas macacae Cas13 (e.g., modified to have or be associated with at least one DD), and may include further alteration or mutation of the Cas13, and can be a chimeric Cas13. In some embodiments, the DD-CRISPR enzyme is codon-optimized for expression in a eukaryotic cell. In some embodiments, the DD-CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the DD-CRISPR enzyme lacks or substantially DNA strand cleavage activity (e.g., no more than 5% nuclease activity as compared with a wild type enzyme or enzyme not having the mutation or alteration that decreases nuclease activity). In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In some embodiments, the guide sequence is at least 16, 17, 18, 19, 20, 25 nucleotides, or between 16-30, or between 16-25, or between 16-20 nucleotides in length.

[0469] In one aspect, the invention provides a method of modifying multiple target polynucleotides in a host cell such as a eukaryotic cell. In some embodiments, the method comprises allowing a Cas13 CRISPR complex to bind to multiple target polynucleotides, e.g., to effect cleavage of said multiple target polynucleotides, thereby modifying multiple target polynucleotides, wherein the Cas13 CRISPR complex comprises a Cas13 enzyme complexed with multiple guide sequences each of the being hybridized to a specific target sequence within said target polynucleotide, wherein said multiple guide sequences are linked to a direct repeat sequence. Where applicable, a tracr sequence may also be provided (e.g. to provide a single guide RNA, sgRNA). In some embodiments, said cleavage comprises cleaving one or two strands at the location of each of the target sequence by said Cas13 enzyme. In some embodiments, said cleavage results in decreased transcription of the multiple target genes. In some embodiments, the method further comprises repairing one or more of said cleaved target polynucleotide by homologous recombination with an exogenous template polynucleotide, wherein said repair results in a mutation comprising an insertion, deletion, or substitution of one or more nucleotides of one or more of said target polynucleotides. In some embodiments, said mutation results in one or more amino acid changes in a protein expressed from a gene comprising one or more of the target sequence(s). In some embodiments, the method further comprises delivering one or more vectors to said eukaryotic cell, wherein the one or more

vectors drive expression of one or more of: the Cas13 enzyme and the multiple guide RNA sequence linked to a direct repeat sequence. Where applicable, a tracr sequence may also be provided. In some embodiments, said vectors are delivered to the eukaryotic cell in a subject. In some embodiments, said modifying takes place in said eukaryotic cell in a cell culture. In some embodiments, the method further comprises isolating said eukaryotic cell from a subject prior to said modifying. In some embodiments, the method further comprises returning said eukaryotic cell and/or cells derived therefrom to said subject.

[0470] In one aspect, the invention provides a method of modifying expression of multiple polynucleotides in a eukaryotic cell. In some embodiments, the method comprises allowing a Cas13 CRISPR complex to bind to multiple polynucleotides such that said binding results in increased or decreased expression of said polynucleotides; wherein the Cas13 CRISPR complex comprises a Cas13 enzyme complexed with multiple guide sequences each specifically hybridized to its own target sequence within said polynucleotide, wherein said guide sequences are linked to a direct repeat sequence. Where applicable, a tracr sequence may also be provided. In some embodiments, the method further comprises delivering one or more vectors to said eukaryotic cells, wherein the one or more vectors drive expression of one or more of: the Cas13 enzyme and the multiple guide sequences linked to the direct repeat sequences. Where applicable, a tracr sequence may also be provided.

[0471] In one aspect, the invention provides a recombinant polynucleotide comprising multiple guide RNA sequences up- or downstream (whichever applicable) of a direct repeat sequence, wherein each of the guide sequences when expressed directs sequence-specific binding of a Cas13 CRISPR complex to its corresponding target sequence present in a eukaryotic cell. In some embodiments, the target sequence is a viral sequence present in a eukaryotic cell. Where applicable, a tracr sequence may also be provided. In some embodiments, the target sequence is a proto-oncogene or an oncogene.

[0472] Aspects of the invention encompass a non-naturally occurring or engineered composition that may comprise a guide RNA (gRNA) comprising a guide sequence capable of hybridizing to a target sequence in a genomic locus of interest in a cell and a Cas13 enzyme as defined herein that may comprise at least one or more nuclear localization sequences.

[0473] An aspect of the invention encompasses methods of modifying a genomic locus of interest to change gene expression in a cell by introducing into the cell any of the compositions described herein.

[0474] An aspect of the invention is that the above elements are comprised in a single composition or comprised in individual compositions. These compositions may advantageously be applied to a host to elicit a functional effect on the genomic level.

[0475] As used herein, the term "guide RNA" or "gRNA" has the leaning as used herein elsewhere and comprises any polynucleotide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target nucleic acid sequence. Each gRNA may be designed to include multiple binding recognition sites (e.g., aptamers) specific to the same or different adapter protein. Each gRNA may be designed to bind to the promoter region -1000 - +1 nucleic acids upstream of the transcription start site (i.e. TSS), preferably -200 nucleic acids. This positioning improves functional domains which affect gene activation (e.g., transcription activators) or gene inhibition (e.g., transcription repressors). The modified gRNA may be one or more modified gRNAs targeted to one or more target loci (e.g., at least 1 gRNA, at least 2 gRNA, at least 5 gRNA, at least 10 gRNA, at least 20 gRNA, at least 30 g RNA, at least 50 gRNA) comprised in a composition. Said multiple gRNA sequences can be tandemly arranged and are preferably separated by a direct repeat.

[0476] Thus, gRNA, the CRISPR enzyme as defined herein may each individually be comprised in a composition and administered to a host individually or collectively. Alternatively, these components may be provided in a single composition for administration to a host. Administration to a host may be performed via viral vectors known to the skilled person or described herein for delivery to a host (e.g., lentiviral vector, adenoviral vector, AAV vector). As explained herein, use of different selection markers (e.g., for lentiviral sgRNA selection) and concentration of gRNA (e.g., dependent on whether multiple gRNAs are used) may be advantageous for eliciting an improved effect. On the basis of this concept, several variations are appropriate to elicit a genomic locus event, including DNA cleavage, gene activation, or gene deactivation. Using the provided compositions, the person skilled in the art can advantageously and specifically target single or multiple loci with the same or different functional domains to elicit one or more genomic locus events. The compositions may be applied in a wide variety of methods for screening in libraries in cells and functional modeling in vivo (e.g., gene activation of lincRNA and identification of function; gain-of-function modeling; loss-of-function modeling; the use the compositions of the invention to establish cell lines and transgenic animals for optimization and screening purposes).

[0477] The current invention comprehends the use of the compositions of the current invention to establish and utilize conditional or inducible CRISPR transgenic cell /animals;

see, e.g., Platt et al., Cell (2014), 159(2): 440-455, or PCT patent publications cited herein, such as WO 2014/093622 (PCT/US2013/074667). For example, cells or animals such as non-human animals, e.g., vertebrates or mammals, such as rodents, e.g., mice, rats, or other laboratory or field animals, e.g., cats, dogs, sheep, etc., may be 'knock-in' whereby the animal conditionally or inducibly expresses Cas13 akin to Platt et al. The target cell or animal thus comprises the CRISPR enzyme (e.g., Cas13) conditionally or inducibly (e.g., in the form of Cre dependent constructs), on expression of a vector introduced into the target cell, the vector expresses that which induces or gives rise to the condition of the CRISPR enzyme (e.g., Cas13) expression in the target cell. By applying the teaching and compositions as defined herein with the known method of creating a CRISPR complex, inducible genomic events are also an aspect of the current invention. Examples of such inducible events have been described herein elsewhere.

[0478] In some embodiments, phenotypic alteration is preferably the result of genome modification when a genetic disease is targeted, especially in methods of therapy and preferably where a repair template is provided to correct or alter the phenotype.

[0479] In some embodiments diseases that may be targeted include those concerned with disease-causing splice defects.

[0480] In some embodiments, cellular targets include Hemopoietic Stem/Progenitor Cells (CD34+); Human T cells; and Eye (retinal cells) - for example photoreceptor precursor cells.

[0481] In some embodiments Gene targets include: Human Beta Globin - HBB (for treating Sickle Cell Anemia, including by stimulating gene-conversion (using closely related HBD gene as an endogenous template)); CD3 (T-Cells); and CEP920 - retina (eye).

[0482] In some embodiments disease targets also include: cancer; Sickle Cell Anemia (based on a point mutation); HBV, HIV; Beta-Thalassemia; and ophthalmic or ocular disease - for example Leber Congenital Amaurosis (LCA)-causing Splice Defect.

[0483] In some embodiments delivery methods include: Cationic Lipid Mediated "direct" delivery of Enzyme-Guide complex (RiboNucleoProtein) and electroporation of plasmid DNA.

[0484] Methods, products and uses described herein may be used for non-therapeutic purposes. Furthermore, any of the methods described herein may be applied in vitro and ex vivo.

[0485] In an aspect, provided is a non-naturally occurring or engineered composition comprising:

I. two or more CRISPR-Cas system polynucleotide sequences comprising

(a) a first guide sequence capable of hybridizing to a first target sequence in a polynucleotide locus,

(b) a second guide sequence capable of hybridizing to a second target sequence in a polynucleotide locus,

(c) a direct repeat sequence,

and

II. a Cas13 enzyme or a second polynucleotide sequence encoding it,

wherein when transcribed, the first and the second guide sequences direct sequence-specific binding of a first and a second Cas13 CRISPR complex to the first and second target sequences respectively,

wherein the first CRISPR complex comprises the Cas13 enzyme complexed with the first guide sequence that is hybridizable to the first target sequence,

wherein the second CRISPR complex comprises the Cas13 enzyme complexed with the second guide sequence that is hybridizable to the second target sequence, and

wherein the first guide sequence directs cleavage of one strand of the DNA duplex near the first target sequence and the second guide sequence directs cleavage of the other strand near the second target sequence inducing a double strand break, thereby modifying the organism or the non-human or non-animal organism. Similarly, compositions comprising more than two guide RNAs can be envisaged e.g. each specific for one target, and arranged tandemly in the composition or CRISPR system or complex as described herein.

[0486] In another embodiment, the Cas13 is delivered into the cell as a protein. In another and particularly preferred embodiment, the Cas13 is delivered into the cell as a protein or as a nucleotide sequence encoding it. Delivery to the cell as a protein may include delivery of a Ribonucleoprotein (RNP) complex, where the protein is complexed with the multiple guides.

[0487] In an aspect, host cells and cell lines modified by or comprising the compositions, systems or modified enzymes of present invention are provided, including stem cells, and progeny thereof.

[0488] In an aspect, methods of cellular therapy are provided, where, for example, a single cell or a population of cells is sampled or cultured, wherein that cell or cells is or has been modified ex vivo as described herein, and is then re-introduced (sampled cells) or introduced (cultured cells) into the organism. Stem cells, whether embryonic or induce pluripotent or totipotent stem cells, are also particularly preferred in this regard. But, of course, in vivo embodiments are also envisaged.

[0489] Inventive methods can further comprise delivery of templates, such as repair templates, which may be dsODN or ssODN, see below. Delivery of templates may be via the cotemporaneous or separate from delivery of any or all the CRISPR enzyme or guide RNAs and via the same delivery mechanism or different. In some embodiments, it is preferred that the template is delivered together with the guide RNAs and, preferably, also the CRISPR enzyme. An example may be an AAV vector where the CRISPR enzyme is AsCas or LbCas.

[0490] Inventive methods can further comprise: (a) delivering to the cell a double-stranded oligodeoxynucleotide (dsODN) comprising overhangs complimentary to the overhangs created by said double strand break, wherein said dsODN is integrated into the locus of interest; or -(b) delivering to the cell a single-stranded oligodeoxynucleotide (ssODN), wherein said ssODN acts as a template for homology directed repair of said double strand break. Inventive methods can be for the prevention or treatment of disease in an individual, optionally wherein said disease is caused by a defect in said locus of interest. Inventive methods can be conducted in vivo in the individual or ex vivo on a cell taken from the individual, optionally wherein said cell is returned to the individual.

[0491] The invention also comprehends products obtained from using CRISPR enzyme or Cas enzyme or Cas13 enzyme or CRISPR-CRISPR enzyme or CRISPR-Cas system or CRISPR-Cas13 system for use in tandem or multiple targeting as defined herein.

Escorted guides for the Cas13 CRISPR-Cas system according to the invention

[0492] In one aspect the invention provides escorted Cas13 CRISPR-Cas systems or complexes, especially such a system involving an escorted Cas13 CRISPR-Cas system guide. By "escorted" is meant that the Cas13 CRISPR-Cas system or complex or guide is delivered to a selected time or place within a cell, so that activity of the Cas13 CRISPR-Cas system or complex or guide is spatially or temporally controlled. For example, the activity and destination of the Cas13 CRISPR-Cas system or complex or guide may be controlled by an escort RNA aptamer sequence that has binding affinity for an aptamer ligand, such as a cell surface protein or other localized cellular component. Alternatively, the escort aptamer may for example be responsive to an aptamer effector on or in the cell, such as a transient effector, such as an external energy source that is applied to the cell at a particular time.

[0493] The escorted Cas13 CRISPR-Cas systems or complexes have a gRNA with a functional structure designed to improve gRNA structure, architecture, stability, genetic expression, or any combination thereof. Such a structure can include an aptamer.

[0494] Aptamers are biomolecules that can be designed or selected to bind tightly to other ligands, for example using a technique called systematic evolution of ligands by exponential

enrichment (SELEX; Tuerk C, Gold L: "Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase." Science 1990, 249:505-510). Nucleic acid aptamers can for example be selected from pools of random-sequence oligonucleotides, with high binding affinities and specificities for a wide range of biomedically relevant targets, suggesting a wide range of therapeutic utilities for aptamers (Keefe, Anthony D., Supriya Pai, and Andrew Ellington. "Aptamers as therapeutics." Nature Reviews Drug Discovery 9.7 (2010): 537-550). These characteristics also suggest a wide range of uses for aptamers as drug delivery vehicles (Levy-Nissenbaum, Etgar, et al. "Nanotechnology and aptamers: applications in drug delivery." Trends in biotechnology 26.8 (2008): 442-449; and, Hicke BJ, Stephens AW. "Escort aptamers: a delivery service for diagnosis and therapy." J Clin Invest 2000, 106:923-928.). Aptamers may also be constructed that function as molecular switches, responding to a que by changing properties, such as RNA aptamers that bind fluorophores to mimic the activity of green fluorescent protein (Paige, Jeremy S., Karen Y. Wu, and Samie R. Jaffrey. "RNA mimics of green fluorescent protein." Science 333.6042 (2011): 642-646). It has also been suggested that aptamers may be used as components of targeted siRNA therapeutic delivery systems, for example targeting cell surface proteins (Zhou, Jiehua, and John J. Rossi. "Aptamer-targeted cell-specific RNA interference." Silence 1.1 (2010): 4).

[0495] Accordingly, provided herein is a gRNA modified, e.g., by one or more aptamer(s) designed to improve gRNA delivery, including delivery across the cellular membrane, to intracellular compartments, or into the nucleus. Such a structure can include, either in addition to the one or more aptamer(s) or without such one or more aptamer(s), moiety(ies) so as to render the guide deliverable, inducible or responsive to a selected effector. The invention accordingly comprehends an gRNA that responds to normal or pathological physiological conditions, including without limitation pH, hypoxia, O2 concentration, temperature, protein concentration, enzymatic concentration, lipid structure, light exposure, mechanical disruption (e.g. ultrasound waves), magnetic fields, electric fields, or electromagnetic radiation.

[0496] An aspect of the invention provides non-naturally occurring or engineered composition comprising an escorted guide RNA (egRNA) comprising:

an RNA guide sequence capable of hybridizing to a target sequence in a genomic locus of interest in a cell; and,

an escort RNA aptamer sequence, wherein the escort aptamer has binding affinity for an aptamer ligand on or in the cell, or the escort aptamer is responsive to a localized aptamer

effector on or in the cell, wherein the presence of the aptamer ligand or effector on or in the cell is spatially or temporally restricted.

[0497] The escort aptamer may for example change conformation in response to an interaction with the aptamer ligand or effector in the cell.

[0498] The escort aptamer may have specific binding affinity for the aptamer ligand.

[0499] The aptamer ligand may be localized in a location or compartment of the cell, for example on or in a membrane of the cell. Binding of the escort aptamer to the aptamer ligand may accordingly direct the egRNA to a location of interest in the cell, such as the interior of the cell by way of binding to an aptamer ligand that is a cell surface ligand. In this way, a variety of spatially restricted locations within the cell may be targeted, such as the cell nucleus or mitochondria.

[0500] Once intended alterations have been introduced, such as by editing intended copies of a gene in the genome of a cell, continued CRISPR/Cas13 expression in that cell is no longer necessary. Indeed, sustained expression would be undesirable in certain casein case of off-target effects at unintended genomic sites, etc. Thus time-limited expression would be useful. Inducible expression offers one approach, but in addition Applicants have engineered a Self-Inactivating Cas13 CRISPR-Cas system that relies on the use of a non-coding guide target sequence within the CRISPR vector itself. Thus, after expression begins, the CRISPR system will lead to its own destruction, but before destruction is complete it will have time to edit the genomic copies of the target gene (which, with a normal point mutation in a diploid cell, requires at most two edits). Simply, the self inactivating Cas13 CRISPR-Cas system includes additional RNA (i.e., guide RNA) that targets the coding sequence for the CRISPR enzyme itself or that targets one or more non-coding guide target sequences complementary to unique sequences present in one or more of the following: (a) within the promoter driving expression of the non-coding RNA elements, (b) within the promoter driving expression of the Cas13 gene, (c) within 100bp of the ATG translational start codon in the Cas13 coding sequence, (d) within the inverted terminal repeat (iTR) of a viral delivery vector, e.g., in an AAV genome.

[0501] The egRNA may include an RNA aptamer linking sequence, operably linking the escort RNA sequence to the RNA guide sequence.

[0502] In embodiments, the egRNA may include one or more photolabile bonds or non-naturally occurring residues.

[0503] In one aspect, the escort RNA aptamer sequence may be complementary to a target miRNA, which may or may not be present within a cell, so that only when the target miRNA is present is there binding of the escort RNA aptamer sequence to the target miRNA which

results in cleavage of the egRNA by an RNA-induced silencing complex (RISC) within the cell.

[0504] In embodiments, the escort RNA aptamer sequence may for example be from 10 to 200 nucleotides in length, and the egRNA may include more than one escort RNA aptamer sequence.

[0505] It is to be understood that any of the RNA guide sequences as described herein elsewhere can be used in the egRNA described herein. In certain embodiments of the invention, the guide RNA or mature crRNA comprises, consists essentially of, or consists of a direct repeat sequence and a guide sequence or spacer sequence. In certain embodiments, the guide RNA or mature crRNA comprises, consists essentially of, or consists of a direct repeat sequence linked to a guide sequence or spacer sequence. In certain embodiments the guide RNA or mature crRNA comprises 19 nts of partial direct repeat followed by 23-25 nt of guide sequence or spacer sequence. In certain embodiments, the effector protein is a FnCas13 effector protein and requires at least 16 nt of guide sequence to achieve detectable DNA cleavage and a minimum of 17 nt of guide sequence to achieve efficient DNA cleavage in vitro. In certain embodiments, the direct repeat sequence is located upstream (i.e., 5') from the guide sequence or spacer sequence. In a preferred embodiment the seed sequence (i.e. the sequence essential critical for recognition and/or hybridization to the sequence at the target locus) of the FnCas13 guide RNA is approximately within the first 5 nt on the 5' end of the guide sequence or spacer sequence.

[0506] The egRNA may be included in a non-naturally occurring or engineered Cas13 CRISPR-Cas complex composition, together with a Cas13 which may include at least one mutation, for example a mutation so that the Cas13 has no more than 5% of the nuclease activity of a Cas13 not having the at least one mutation, for example having a diminished nuclease activity of at least 97%, or 100% as compared with the Cas13 not having the at least one mutation. The Cas13 may also include one or more nuclear localization sequences. Mutated Cas13 enzymes having modulated activity such as diminished nuclease activity are described herein elsewhere.

[0507] The engineered Cas13 CRISPR-Cas composition may be provided in a cell, such as a eukaryotic cell, a mammalian cell, or a human cell.

[0508] In embodiments, the compositions described herein comprise a Cas13 CRISPR-Cas complex having at least three functional domains, at least one of which is associated with Cas13 and at least two of which are associated with egRNA.

[0509] The compositions described herein may be used to introduce a genomic locus event in a host cell, such as a eukaryotic cell, in particular a mammalian cell, or a non-human eukaryote, in particular a non-human mammal such as a mouse, in vivo. The genomic locus event may comprise affecting gene activation, gene inhibition, or cleavage in a locus. The compositions described herein may also be used to modify a genomic locus of interest to change gene expression in a cell. Methods of introducing a genomic locus event in a host cell using the Cas13 enzyme provided herein are described herein in detail elsewhere. Delivery of the composition may for example be by way of delivery of a nucleic acid molecule(s) coding for the composition, which nucleic acid molecule(s) is operatively linked to regulatory sequence(s), and expression of the nucleic acid molecule(s) in vivo, for example by way of a lentivirus, an adenovirus, or an AAV.

[0510] The present invention provides compositions and methods by which gRNA-mediated gene editing activity can be adapted. The invention provides gRNA secondary structures that improve cutting efficiency by increasing gRNA and/or increasing the amount of RNA delivered into the cell. The gRNA may include light labile or inducible nucleotides.

[0511] To increase the effectiveness of gRNA, for example gRNA delivered with viral or non-viral technologies, Applicants added secondary structures into the gRNA that enhance its stability and improve gene editing. Separately, to overcome the lack of effective delivery, Applicants modified gRNAs with cell penetrating RNA aptamers; the aptamers bind to cell surface receptors and promote the entry of gRNAs into cells. Notably, the cell-penetrating aptamers can be designed to target specific cell receptors, in order to mediate cell-specific delivery. Applicants also have created guides that are inducible.

[0512] Light responsiveness of an inducible system may be achieved via the activation and binding of cryptochrome-2 and CIB1. Blue light stimulation induces an activating conformational change in cryptochrome-2, resulting in recruitment of its binding partner CIB1. This binding is fast and reversible, achieving saturation in <15 sec following pulsed stimulation and returning to baseline <15 min after the end of stimulation. These rapid binding kinetics result in a system temporally bound only by the speed of transcription/translation and transcript/protein degradation, rather than uptake and clearance of inducing agents. Crytochrome-2 activation is also highly sensitive, allowing for the use of low light intensity stimulation and mitigating the risks of phototoxicity. Further, in a context such as the intact mammalian brain, variable light intensity may be used to control the size of a stimulated region, allowing for greater precision than vector delivery alone may offer.

[0513] The invention contemplates energy sources such as electromagnetic radiation, sound energy or thermal energy to induce the guide. Advantageously, the electromagnetic radiation is a component of visible light. In a preferred embodiment, the light is a blue light with a wavelength of about 450 to about 495 nm. In an especially preferred embodiment, the wavelength is about 488 nm. In another preferred embodiment, the light stimulation is via pulses. The light power may range from about 0-9 mW/cm2. In a preferred embodiment, a stimulation paradigm of as low as 0.25 sec every 15 sec should result in maximal activation.

[0514] Cells involved in the practice of the present invention may be a prokaryotic cell or a eukaryotic cell, advantageously an animal cell a plant cell or a yeast cell, more advantageously a mammalian cell.

[0515] The chemical or energy sensitive guide may undergo a conformational change upon induction by the binding of a chemical source or by the energy allowing it act as a guide and have the Cas13 CRISPR-Cas system or complex function. The invention can involve applying the chemical source or energy so as to have the guide function and the Cas13 CRISPR-Cas system or complex function; and optionally further determining that the expression of the genomic locus is altered.

[0516] There are several different designs of this chemical inducible system: 1. ABI-PYL based system inducible by Abscisic Acid (ABA) (see, e.g., http://stke.sciencemag.org/cgi/content/abstract/sigtrans;4/164/rs2), 2. FKBP-FRB based system inducible by rapamycin (or related chemicals based on rapamycin) (see, e.g., http://www.nature.com/nmeth/journal/v2/n6/full/nmeth763.html), 3. GID1-GAI based system inducible by Gibberellin (GA) (see, e.g., http://www.nature.com/nchembio/journal/v8/n5/full/nchembio.922.html).

[0517] Another system contemplated by the present invention is a chemical inducible system based on change in sub-cellular localization. Applicants also developed a system in which the polypeptide include a DNA binding domain comprising at least five or more Transcription activator-like effector (TALE) monomers and at least one or more half-monomers specifically ordered to target the genomic locus of interest linked to at least one or more effector domains are further linker to a chemical or energy sensitive protein. This protein will lead to a change in the sub-cellular localization of the entire polypeptide (i.e. transportation of the entire polypeptide from cytoplasm into the nucleus of the cells) upon the binding of a chemical or energy transfer to the chemical or energy sensitive protein. This transportation of the entire polypeptide from one sub-cellular compartments or organelles, in which its activity is sequestered due to lack of substrate for the effector domain, into another one in which the

substrate is present would allow the entire polypeptide to come in contact with its desired substrate (i.e. genomic DNA in the mammalian nucleus) and result in activation or repression of target gene expression.

[0518] This type of system could also be used to induce the cleavage of a genomic locus of interest in a cell when the effector domain is a nuclease.

[0519] A chemical inducible system can be an estrogen receptor (ER) based system inducible by 4-hydroxytamoxifen (4OHT) (see, e.g., http://www.pnas.org/content/104/3/1027.abstract). A mutated ligand-binding domain of the estrogen receptor called ERT2 translocates into the nucleus of cells upon binding of 4-hydroxytamoxifen. In further embodiments of the invention any naturally occurring or engineered derivative of any nuclear receptor, thyroid hormone receptor, retinoic acid receptor, estrogen receptor, estrogen-related receptor, glucocorticoid receptor, progesterone receptor, androgen receptor may be used in inducible systems analogous to the ER based inducible system.

[0520] Another inducible system is based on the design using Transient receptor potential (TRP) ion channel based system inducible by energy, heat or radio-wave (see, e.g., http://www.sciencemag.org/content/336/6081/604). These TRP family proteins respond to different stimuli, including light and heat. When this protein is activated by light or heat, the ion channel will open and allow the entering of ions such as calcium into the plasma membrane. This influx of ions will bind to intracellular ion interacting partners linked to a polypeptide including the guide and the other components of the Cas13 CRISPR-Cas complex or system, and the binding will induce the change of sub-cellular localization of the polypeptide, leading to the entire polypeptide entering the nucleus of cells. Once inside the nucleus, the guide protein and the other components of the Cas13 CRISPR-Cas complex will be active and modulating target gene expression in cells.

[0521] This type of system could also be used to induce the cleavage of a genomic locus of interest in a cell; and, in this regard, it is noted that the Cas13 enzyme is a nuclease. The light could be generated with a laser or other forms of energy sources. The heat could be generated by raise of temperature results from an energy source, or from nano-particles that release heat after absorbing energy from an energy source delivered in the form of radio-wave.

[0522] While light activation may be an advantageous embodiment, sometimes it may be disadvantageous especially for in vivo applications in which the light may not penetrate the skin or other organs. In this instance, other methods of energy activation are contemplated, in particular, electric field energy and/or ultrasound which have a similar effect.

[0523] Electric field energy is preferably administered substantially as described in the art, using one or more electric pulses of from about 1 Volt/cm to about 10 kVolts/cm under in vivo conditions. Instead of or in addition to the pulses, the electric field may be delivered in a continuous manner. The electric pulse may be applied for between 1 ¬µs and 500 milliseconds, preferably between 1 ¬µs and 100 milliseconds. The electric field may be applied continuously or in a pulsed manner for 5 about minutes.

[0524] As used herein, 'electric field energy' is the electrical energy to which a cell is exposed. Preferably the electric field has a strength of from about 1 Volt/cm to about 10 kVolts/cm or more under in vivo conditions (see WO97/49450).

[0525] As used herein, the term "electric field" includes one or more pulses at variable capacitance and voltage and including exponential and/or square wave and/or modulated wave and/or modulated square wave forms. References to electric fields and electricity should be taken to include reference the presence of an electric potential difference in the environment of a cell. Such an environment may be set up by way of static electricity, alternating current (AC), direct current (DC), etc, as known in the art. The electric field may be uniform, non-uniform or otherwise, and may vary in strength and/or direction in a time dependent manner.

[0526] Single or multiple applications of electric field, as well as single or multiple applications of ultrasound are also possible, in any order and in any combination. The ultrasound and/or the electric field may be delivered as single or multiple continuous applications, or as pulses (pulsatile delivery).

[0527] Electroporation has been used in both in vitro and in vivo procedures to introduce foreign material into living cells. With in vitro applications, a sample of live cells is first mixed with the agent of interest and placed between electrodes such as parallel plates. Then, the electrodes apply an electrical field to the cell/implant mixture. Examples of systems that perform in vitro electroporation include the Electro Cell Manipulator ECM600 product, and the Electro Square Porator T820, both made by the BTX Division of Genetronics, Inc (see U.S. Pat. No 5,869,326).

[0528] The known electroporation techniques (both in vitro and in vivo) function by applying a brief high voltage pulse to electrodes positioned around the treatment region. The electric field generated between the electrodes causes the cell membranes to temporarily become porous, whereupon molecules of the agent of interest enter the cells. In known electroporation applications, this electric field comprises a single square wave pulse on the order of 1000 V/cm, of about 100 .mu.s duration. Such a pulse may be generated, for example, in known applications of the Electro Square Porator T820.

[0529] Preferably, the electric field has a strength of from about 1 V/cm to about 10 kV/cm under in vitro conditions. Thus, the electric field may have a strength of 1 V/cm, 2 V/cm, 3 V/cm, 4 V/cm, 5 V/cm, 6 V/cm, 7 V/cm, 8 V/cm, 9 V/cm, 10 V/cm, 20 V/cm, 50 V/cm, 100 V/cm, 200 V/cm, 300 V/cm, 400 V/cm, 500 V/cm, 600 V/cm, 700 V/cm, 800 V/cm, 900 V/cm, 1 kV/cm, 2 kV/cm, 5 kV/cm, 10 kV/cm, 20 kV/cm, 50 kV/cm or more. More preferably from about 0.5 kV/cm to about 4.0 kV/cm under in vitro conditions. Preferably the electric field has a strength of from about 1 V/cm to about 10 kV/cm under in vivo conditions. However, the electric field strengths may be lowered where the number of pulses delivered to the target site are increased. Thus, pulsatile delivery of electric fields at lower field strengths is envisaged.

[0530] Preferably the application of the electric field is in the form of multiple pulses such as double pulses of the same strength and capacitance or sequential pulses of varying strength and/or capacitance. As used herein, the term "pulse" includes one or more electric pulses at variable capacitance and voltage and including exponential and/or square wave and/or modulated wave/square wave forms.

[0531] Preferably the electric pulse is delivered as a waveform selected from an exponential wave form, a square wave form, a modulated wave form and a modulated square wave form.

[0532] A preferred embodiment employs direct current at low voltage. Thus, Applicants disclose the use of an electric field which is applied to the cell, tissue or tissue mass at a field strength of between 1V/cm and 20V/cm, for a period of 100 milliseconds or more, preferably 15 minutes or more.

[0533] Ultrasound is advantageously administered at a power level of from about 0.05 W/cm2 to about 100 W/cm2. Diagnostic or therapeutic ultrasound may be used, or combinations thereof.

[0534] As used herein, the term "ultrasound" refers to a form of energy which consists of mechanical vibrations the frequencies of which are so high they are above the range of human hearing. Lower frequency limit of the ultrasonic spectrum may generally be taken as about 20 kHz. Most diagnostic applications of ultrasound employ frequencies in the range 1 and 15 MHz' (From Ultrasonics in Clinical Diagnosis, P. N. T. Wells, ed., 2nd. Edition, Publ. Churchill Livingstone [Edinburgh, London & NY, 1977]).

[0535] Ultrasound has been used in both diagnostic and therapeutic applications. When used as a diagnostic tool ("diagnostic ultrasound"), ultrasound is typically used in an energy density range of up to about 100 mW/cm2 (FDA recommendation), although energy densities of up to 750 mW/cm2 have been used. In physiotherapy, ultrasound is typically used as an

energy source in a range up to about 3 to 4 W/cm2 (WHO recommendation). In other therapeutic applications, higher intensities of ultrasound may be employed, for example, HIFU at 100 W/cm up to 1 kW/cm2 (or even higher) for short periods of time. The term "ultrasound" as used in this specification is intended to encompass diagnostic, therapeutic and focused ultrasound.

[0536] Focused ultrasound (FUS) allows thermal energy to be delivered without an invasive probe (see Morocz et al 1998 Journal of Magnetic Resonance Imaging Vol.8, No. 1, pp.136-142. Another form of focused ultrasound is high intensity focused ultrasound (HIFU) which is reviewed by Moussatov et al in Ultrasonics (1998) Vol.36, No.8, pp.893-900 and TranHuuHue et al in Acustica (1997) Vol.83, No.6, pp.1103-1106.

[0537] Preferably, a combination of diagnostic ultrasound and a therapeutic ultrasound is employed. This combination is not intended to be limiting, however, and the skilled reader will appreciate that any variety of combinations of ultrasound may be used. Additionally, the energy density, frequency of ultrasound, and period of exposure may be varied.

[0538] Preferably the exposure to an ultrasound energy source is at a power density of from about 0.05 to about 100 Wcm-2. Even more preferably, the exposure to an ultrasound energy source is at a power density of from about 1 to about 15 Wcm-2.

[0539] Preferably the exposure to an ultrasound energy source is at a frequency of from about 0.015 to about 10.0 MHz. More preferably the exposure to an ultrasound energy source is at a frequency of from about 0.02 to about 5.0 MHz or about 6.0 MHz. Most preferably, the ultrasound is applied at a frequency of 3 MHz.

[0540] Preferably the exposure is for periods of from about 10 milliseconds to about 60 minutes. Preferably the exposure is for periods of from about 1 second to about 5 minutes. More preferably, the ultrasound is applied for about 2 minutes. Depending on the particular target cell to be disrupted, however, the exposure may be for a longer duration, for example, for 15 minutes.

[0541] Advantageously, the target tissue is exposed to an ultrasound energy source at an acoustic power density of from about 0.05 Wcm-2 to about 10 Wcm-2 with a frequency ranging from about 0.015 to about 10 MHz (see WO 98/52609). However, alternatives are also possible, for example, exposure to an ultrasound energy source at an acoustic power density of above 100 Wcm-2, but for reduced periods of time, for example, 1000 Wcm-2 for periods in the millisecond range or less.

[0542] Preferably the application of the ultrasound is in the form of multiple pulses; thus, both continuous wave and pulsed wave (pulsatile delivery of ultrasound) may be employed in

any combination. For example, continuous wave ultrasound may be applied, followed by pulsed wave ultrasound, or vice versa. This may be repeated any number of times, in any order and combination. The pulsed wave ultrasound may be applied against a background of continuous wave ultrasound, and any number of pulses may be used in any number of groups.

[0543] Preferably, the ultrasound may comprise pulsed wave ultrasound. In a highly preferred embodiment, the ultrasound is applied at a power density of 0.7 Wcm-2 or 1.25 Wcm-2 as a continuous wave. Higher power densities may be employed if pulsed wave ultrasound is used.

[0544] Use of ultrasound is advantageous as, like light, it may be focused accurately on a target. Moreover, ultrasound is advantageous as it may be focused more deeply into tissues unlike light. It is therefore better suited to whole-tissue penetration (such as but not limited to a lobe of the liver) or whole organ (such as but not limited to the entire liver or an entire muscle, such as the heart) therapy. Another important advantage is that ultrasound is a non-invasive stimulus which is used in a wide variety of diagnostic and therapeutic applications. By way of example, ultrasound is well known in medical imaging techniques and, additionally, in orthopedic therapy. Furthermore, instruments suitable for the application of ultrasound to a subject vertebrate are widely available and their use is well known in the art.

[0545] The rapid transcriptional response and endogenous targeting of the instant invention make for an ideal system for the study of transcriptional dynamics. For example, the instant invention may be used to study the dynamics of variant production upon induced expression of a target gene. On the other end of the transcription cycle, mRNA degradation studies are often performed in response to a strong extracellular stimulus, causing expression level changes in a plethora of genes. The instant invention may be utilized to reversibly induce transcription of an endogenous target, after which point stimulation may be stopped and the degradation kinetics of the unique target may be tracked.

[0546] The temporal precision of the instant invention may provide the power to time genetic regulation in concert with experimental interventions. For example, targets with suspected involvement in long-term potentiation (LTP) may be modulated in organotypic or dissociated neuronal cultures, but only during stimulus to induce LTP, so as to avoid interfering with the normal development of the cells. Similarly, in cellular models exhibiting disease phenotypes, targets suspected to be involved in the effectiveness of a particular therapy may be modulated only during treatment. Conversely, genetic targets may be modulated only during a pathological stimulus. Any number of experiments in which timing of genetic cues to external

experimental stimuli is of relevance may potentially benefit from the utility of the instant invention.

[0547] The in vivo context offers equally rich opportunities for the instant invention to control gene expression. Photoinducibility provides the potential for spatial precision. Taking advantage of the development of optrode technology, a stimulating fiber optic lead may be placed in a precise brain region. Stimulation region size may then be tuned by light intensity. This may be done in conjunction with the delivery of the Cas13 CRISPR-Cas system or complex of the invention, or, in the case of transgenic Cas13 animals, guide RNA of the invention may be delivered and the optrode technology can allow for the modulation of gene expression in precise brain regions. A transparent Cas13 expressing organism, can have guide RNA of the invention administered to it and then there can be extremely precise laser induced local gene expression changes.

[0548] A culture medium for culturing host cells includes a medium commonly used for tissue culture, such as M199-earle base, Eagle MEM (E-MEM), Dulbecco MEM (DMEM), SC-UCM102, UP-SFM (GIBCO BRL), EX-CELL302 (Nichirei), EX-CELL293-S (Nichirei), TFBM-01 (Nichirei), ASF104, among others. Suitable culture media for specific cell types may be found at the American Type Culture Collection (ATCC) or the European Collection of Cell Cultures (ECACC). Culture media may be supplemented with amino acids such as L-glutamine, salts, anti-fungal or anti-bacterial agents such as Fungizone¬Æ, penicillin-streptomycin, animal serum, and the like. The cell culture medium may optionally be serum-free.

[0549] The invention may also offer valuable temporal precision in vivo. The invention may be used to alter gene expression during a particular stage of development. The invention may be used to time a genetic cue to a particular experimental window. For example, genes implicated in learning may be overexpressed or repressed only during the learning stimulus in a precise region of the intact rodent or primate brain. Further, the invention may be used to induce gene expression changes only during particular stages of disease development. For example, an oncogene may be overexpressed only once a tumor reaches a particular size or metastatic stage. Conversely, proteins suspected in the development of Alzheimer's may be knocked down only at defined time points in the animal's life and within a particular brain region. Although these examples do not exhaustively list the potential applications of the invention, they highlight some of the areas in which the invention may be a powerful technology.

Protected guides: Enzymes according to the invention can be used in combination with protected guide RNAs

[0550] In one aspect, an object of the current invention is to further enhance the specificity of Cas13 given individual guide RNAs through thermodynamic tuning of the binding specificity of the guide RNA to target DNA. This is a general approach of introducing mismatches, elongation or truncation of the guide sequence to increase / decrease the number of complimentary bases vs. mismatched bases shared between a genomic target and its potential off-target loci, in order to give thermodynamic advantage to targeted genomic loci over genomic off-targets.

[0551] In one aspect, the invention provides for the guide sequence being modified by secondary structure to increase the specificity of the Cas13 CRISPR-Cas system and whereby the secondary structure can protect against exonuclease activity and allow for 3' additions to the guide sequence.

[0552] In one aspect, the invention provides for hybridizing a "protector RNA" to a guide sequence, wherein the "protector RNA" is an RNA strand complementary to the 5' end of the guide RNA (gRNA), to thereby generate a partially double-stranded gRNA. In an embodiment of the invention, protecting the mismatched bases with a perfectly complementary protector sequence decreases the likelihood of target DNA binding to the mismatched base pairs at the 3' end. In embodiments of the invention, additional sequences comprising an extended length may also be present.

[0553] Guide RNA (gRNA) extensions matching the genomic target provide gRNA protection and enhance specificity. Extension of the gRNA with matching sequence distal to the end of the spacer seed for individual genomic targets is envisaged to provide enhanced specificity. Matching gRNA extensions that enhance specificity have been observed in cells without truncation. Prediction of gRNA structure accompanying these stable length extensions has shown that stable forms arise from protective states, where the extension forms a closed loop with the gRNA seed due to complimentary sequences in the spacer extension and the spacer seed. These results demonstrate that the protected guide concept also includes sequences matching the genomic target sequence distal of the 20mer spacer-binding region. Thermodynamic prediction can be used to predict completely matching or partially matching guide extensions that result in protected gRNA states. This extends the concept of protected gRNAs to interaction between X and Z, where X will generally be of length 17-20nt and Z is of length 1-30nt. Thermodynamic prediction can be used to determine the optimal extension state for Z, potentially introducing small numbers of mismatches in Z to promote the formation

of protected conformations between X and Z. Throughout the present application, the terms "X" and seed length (SL) are used interchangeably with the term exposed length (EpL) which denotes the number of nucleotides available for target DNA to bind; the terms "Y" and protector length (PL) are used interchangeably to represent the length of the protector; and the terms "Z", "E", "E'" and "EL" are used interchangeably to correspond to the term extended length (ExL) which represents the number of nucleotides by which the target sequence is extended.

[0554] An extension sequence which corresponds to the extended length (ExL) may optionally be attached directly to the guide sequence at the 3' end of the protected guide sequence. The extension sequence may be 2 to 12 nucleotides in length. Preferably ExL may be denoted as 0, 2, 4, 6, 8, 10 or 12 nucleotides in length.. In a preferred embodiment the ExL is denoted as 0 or 4 nucleotides in length. In a more preferred embodiment the ExL is 4 nucleotides in length. The extension sequence may or may not be complementary to the target sequence.

[0555] An extension sequence may further optionally be attached directly to the guide sequence at the 5' end of the protected guide sequence as well as to the 3' end of a protecting sequence. As a result, the extension sequence serves as a linking sequence between the protected sequence and the protecting sequence. Without wishing to be bound by theory, such a link may position the protecting sequence near the protected sequence for improved binding of the protecting sequence to the protected sequence. It will be understood that the above-described relationship of seed, protector, and extension applies where the distal end (i.e., the targeting end) of the guide is the 5' end, e.g. a guide that functions is a Cas13 system. In an embodiment wherein the distal end of the guide is the 3' end, the relationship will be the reverse. In such an embodiment, the invention provides for hybridizing a "protector RNA" to a guide sequence, wherein the "protector RNA" is an RNA strand complementary to the 3' end of the guide RNA (gRNA), to thereby generate a partially double-stranded gRNA.

[0556] Addition of gRNA mismatches to the distal end of the gRNA can demonstrate enhanced specificity. The introduction of unprotected distal mismatches in Y or extension of the gRNA with distal mismatches (Z) can demonstrate enhanced specificity. This concept as mentioned is tied to X, Y, and Z components used in protected gRNAs. The unprotected mismatch concept may be further generalized to the concepts of X, Y, and Z described for protected guide RNAs.

[0557] Cas13. In one aspect, the invention provides for enhanced Cas13 specificity wherein the double stranded 3' end of the protected guide RNA (pgRNA) allows for two

possible outcomes: (1) the guide RNA-protector RNA to guide RNA-target DNA strand exchange will occur and the guide will fully bind the target, or (2) the guide RNA will fail to fully bind the target and because Cas13 target cleavage is a multiple step kinetic reaction that requires guide RNA:target DNA binding to activate Cas13-catalyzed DSBs, wherein Cas13 cleavage does not occur if the guide RNA does not properly bind. According to particular embodiments, the protected guide RNA improves specificity of target binding as compared to a naturally occurring CRISPR-Cas system. According to particular embodiments the protected modified guide RNA improves stability as compared to a naturally occurring CRISPR-Cas. According to particular embodiments the protector sequence has a length between 3 and 120 nucleotides and comprises 3 or more contiguous nucleotides complementary to another sequence of guide or protector. According to particular embodiments, the protector sequence forms a hairpin. According to particular embodiments the guide RNA further comprises a protected sequence and an exposed sequence. According to particular embodiments the exposed sequence is 1 to 19 nucleotides. More particularly, the exposed sequence is at least 75%, at least 90% or about 100% complementary to the target sequence. According to particular embodiments the guide sequence is at least 90% or about 100% complementary to the protector strand. According to particular embodiments the guide sequence is at least 75%, at least 90% or about 100% complementary to the target sequence. According to particular embodiments, the guide RNA further comprises an extension sequence. More particularly, when the distal end of the guide is the 3' end, the extension sequence is operably linked to the 3' end of the protected guide sequence, and optionally directly linked to the 3' end of the protected guide sequence. According to particular embodiments the extension sequence is 1-12 nucleotides. According to particular embodiments the extension sequence is operably linked to the guide sequence at the 3' end of the protected guide sequence and the 5' end of the protector strand and optionally directly linked to the 3' end of the protected guide sequence and the 53' end of the protector strand, wherein the extension sequence is a linking sequence between the protected sequence and the protector strand. According to particular embodiments the extension sequence is 100% not complementary to the protector strand, optionally at least 95%, at least 90%, at least 80%, at least 70%, at least 60%, or at least 50% not complementary to the protector strand. According to particular embodiments the guide sequence further comprises mismatches appended to the end of the guide sequence, wherein the mismatches thermodynamically optimize specificity.

[0558] According to the invention, in certain embodiments, guide modifications that impede strand invasion will be desireable. For example, to minimize off-target actifity, in

certain embodiments, it will be desireable to design or modify a guide to impede strand invasiom at off-target sites. In certain such embodiments, it may be acceptable or useful to design or modify a guide at the expense of on-target binding efficiency. In certain embodiments, guide-target mismatches at the target site may be tolerated that substantially reduce off-target activity.

[0559] In certain embodiments of the invention, it is desirable to adjust the binding characteristics of the protected guide to minimize off-target CRISPR activity. Accordingly, thermodynamic prediction algoithms are used to predict strengths of binding on target and off target. Alternatively or in addition, selection methods are used to reduce or minimize off-target effects, by absolute measures or relative to on-target effects.

[0560] Design options include, without limitation, i) adjusting the length of protector strand that binds to the protected strand, ii) adjusting the length of the portion of the protected strand that is exposed, iii) extending the protected strand with a stem-loop located external (distal) to the protected strand (i.e. designed so that the stem loop is external to the protected strand at the distal end), iv) extending the protected strand by addition of a protector strand to form a stem-loop with all or part of the protected strand, v) adjusting binding of the protector strand to the protected strand by designing in one or more base mismatches and/or one or more non-canonical base pairings, vi) adjusting the location of the stem formed by hybridization of the protector strand to the protected strand, and vii) addition of a non-structured protector to the end of the protected strand.

[0561] In one aspect, the invention provides an engineered, non-naturally occurring CRISPR-Cas system comprising a Cas13 protein and a protected guide RNA that targets a DNA molecule encoding a gene product in a cell, whereby the protected guide RNA targets the DNA molecule encoding the gene product and the Cas13 protein cleaves the DNA molecule encoding the gene product, whereby expression of the gene product is altered; and, wherein the Cas13 protein and the protected guide RNA do not naturally occur together. The invention comprehends the protected guide RNA comprising a guide sequence fused 3' to a direct repeat sequence. The invention further comprehends the Cas13 CRISPR protein being codon optimized for expression in a eEukaryotic cell. In a preferred embodiment the eEukaryotic cell is a mammalian cell, a plant cell or a yeast cell and in a more preferred embodiment the mammalian cell is a human cell. In a further embodiment of the invention, the expression of the gene product is decreased. In some embodiments the CRISPR protein is Cas13. In some embodiments the CRISPR protein is Cas12a. In some embodiments,, the Cas13 or Cas12a enzyme protein is Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium or Francisella

Novicida Cas13 or Cas12a, and may include mutated Cas13 or Cas12a derived from these organisms. The enzyme protein may be a further Cas13 or Cas12a homolog or ortholog. In some embodiments, the nucleotide sequence encoding the Cfp1 Csa13 or Cas12a enzyme protein is codon-optimized for expression in a eukaryotic cell. In some embodiments, the Cas13 or Cas12a enzyme protein directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In general, and throughout this specification, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Vectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g., circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g., retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors." Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[0562] Recombinant expression vectors can comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g., in an in vitro

transcription/translation system or in a host cell when the vector is introduced into the host cell).

[0563] Advantageous vectors include lentiviruses and adeno-associated viruses, and types of such vectors can also be selected for targeting particular types of cells.

[0564] In one aspect, the invention provides a eukaryotic host cell comprising (a) a first regulatory element operably linked to a direct repeat sequence and one or more insertion sites for inserting one or more guide sequences downstream of the direct repeat sequence, wherein when expressed, the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a eukaryotic cell, wherein the CRISPR complex comprises a CRISPR enzyme complexed with the guide RNA comprising the guide sequence that is hybridized to the target sequence and/or (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said Cas13 enzyme comprising a nuclear localization sequence. In some embodiments, the host cell comprises components (a) and (b). In some embodiments, component (a), component (b), or components (a) and (b) are stably integrated into a genome of the host eukaryotic cell. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the Cas13 enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the Cas13 enzyme lacks RNA strand cleavage activity. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter.

[0565] In an aspect, the invention provides a non-human eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. In other aspects, the invention provides a eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. The organism in some embodiments of these aspects may be an animal; for example a mammal. Also, the organism may be an arthropod such as an insect. The organism also may be a plant or a yeast. Further, the organism may be a fungus.

[0566] In one aspect, the invention provides a kit comprising one or more of the components described herein above. In some embodiments, the kit comprises a vector system and instructions for using the kit. In some embodiments, the vector system comprises (a) a first regulatory element operably linked to a direct repeat sequence and one or more insertion sites for inserting one or more guide sequences downstream of the direct repeat sequence,

wherein when expressed, the guide sequence directs sequence-specific binding of a Cas13 CRISPR complex to a target sequence in a eukaryotic cell, wherein the CRISPR complex comprises a Cas13 enzyme complexed with the protected guide RNA comprising the guide sequence that is hybridized to the target sequence and/or (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said Cas13 enzyme comprising a nuclear localization sequence. In some embodiments, the kit comprises components (a) and (b) located on the same or different vectors of the system. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the Cas13 enzyme comprises one or more nuclear localization sequences of sufficient strength to drive accumulation of said Cas13 enzyme in a detectable amount in the nucleus of a eukaryotic cell. In some embodiments, the Cas13 enzyme is Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020 or Francisella tularensis 1 Novicida Cas13, and may include mutated Cas13 derived from these organisms. The enzyme may be a Cas13 homolog or ortholog. In some embodiments, the CRISPR enzyme is codon-optimized for expression in a eukaryotic cell. In some embodiments, the CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the CRISPR enzyme lacks DNA strand cleavage activity. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter.

[0567] In one aspect, the invention provides a method of modifying a target polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the target polynucleotide to effect cleavage of said target polynucleotide thereby modifying the target polynucleotide, wherein the CRISPR complex comprises a Cas13 enzyme complexed with protected guide RNA comprising a guide sequence hybridized to a target sequence within said target polynucleotide. In some embodiments, said cleavage comprises cleaving one or two strands at the location of the target sequence by said Cas13 enzyme. In some embodiments, said cleavage results in decreased transcription of a target gene. In some embodiments, the method further comprises repairing said cleaved target polynucleotide by non-homologous end joining (NHEJ)-based gene insertion mechanisms, more particularly with an exogenous template polynucleotide, wherein said repair results in a mutation comprising an insertion, deletion, or substitution of one or more nucleotides of said target polynucleotide. In some embodiments, said mutation results in one or more amino acid

changes in a protein expressed from a gene comprising the target sequence. In some embodiments, the method further comprises delivering one or more vectors to said eukaryotic cell, wherein the one or more vectors drive expression of one or more of: the Cas13 enzyme, the protected guide RNA comprising the guide sequence linked to direct repeat sequence. In some embodiments, said vectors are delivered to the eukaryotic cell in a subject. In some embodiments, said modifying takes place in said eukaryotic cell in a cell culture. In some embodiments, the method further comprises isolating said eukaryotic cell from a subject prior to said modifying. In some embodiments, the method further comprises returning said eukaryotic cell and/or cells derived therefrom to said subject.

[0568] In one aspect, the invention provides a method of modifying expression of a polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a Cas13 CRISPR complex to bind to the polynucleotide such that said binding results in increased or decreased expression of said polynucleotide; wherein the CRISPR complex comprises a Cas13 enzyme complexed with a protected guide RNA comprising a guide sequence hybridized to a target sequence within said polynucleotide. In some embodiments, the method further comprises delivering one or more vectors to said eukaryotic cells, wherein the one or more vectors drive expression of one or more of: the Cas13 enzyme and the protected guide RNA.

[0569] In one aspect, the invention provides a method of generating a model eukaryotic cell comprising a mutated disease gene. In some embodiments, a disease gene is any gene associated an increase in the risk of having or developing a disease. In some embodiments, the method comprises (a) introducing one or more vectors into a eukaryotic cell, wherein the one or more vectors drive expression of one or more of: a Cas13 enzyme and a protected guide RNA comprising a guide sequence linked to a direct repeat sequence; and (b) allowing a CRISPR complex to bind to a target polynucleotide to effect cleavage of the target polynucleotide within said disease gene, wherein the CRISPR complex comprises the Cas13 enzyme complexed with the guide RNA comprising the sequence that is hybridized to the target sequence within the target polynucleotide, thereby generating a model eukaryotic cell comprising a mutated disease gene. In some embodiments, said cleavage comprises cleaving one or two strands at the location of the target sequence by said Cas13 enzyme. In some embodiments, said cleavage results in decreased transcription of a target gene. In some embodiments, the method further comprises repairing said cleaved target polynucleotide by non-homologous end joining (NHEJ)-based gene insertion mechanisms with an exogenous template polynucleotide, wherein said repair results in a mutation comprising an insertion,

deletion, or substitution of one or more nucleotides of said target polynucleotide. In some embodiments, said mutation results in one or more amino acid changes in a protein expression from a gene comprising the target sequence.

[0570] In one aspect, the invention provides a method for developing a biologically active agent that modulates a cell signaling event associated with a disease gene. In some embodiments, a disease gene is any gene associated an increase in the risk of having or developing a disease. In some embodiments, the method comprises (a) contacting a test compound with a model cell of any one of the described embodiments; and (b) detecting a change in a readout that is indicative of a reduction or an augmentation of a cell signaling event associated with said mutation in said disease gene, thereby developing said biologically active agent that modulates said cell signaling event associated with said disease gene.

[0571] In one aspect, the invention provides a recombinant polynucleotide comprising a protected guide sequence downstream of a direct repeat sequence, wherein the protected guide sequence when expressed directs sequence-specific binding of a CRISPR complex to a corresponding target sequence present in a eukaryotic cell. In some embodiments, the target sequence is a viral sequence present in a eukaryotic cell. In some embodiments, the target sequence is a proto-oncogene or an oncogene.

[0572] In one aspect the invention provides for a method of selecting one or more cell(s) by introducing one or more mutations in a gene in the one or more cell (s), the method comprising: introducing one or more vectors into the cell (s), wherein the one or more vectors drive expression of one or more of: a Cas13 enzyme, a protected guide RNA comprising a guide sequence, and an editing template; wherein the editing template comprises the one or more mutations that abolish Cas13 enzyme cleavage; allowing non-homologous end joining (NHEJ)-based gene insertion mechanisms of the editing template with the target polynucleotide in the cell(s) to be selected; allowing a CRISPR complex to bind to a target polynucleotide to effect cleavage of the target polynucleotide within said gene, wherein the CRISPR complex comprises the Cas13 enzyme complexed with the protected guide RNA comprising a guide sequence that is hybridized to the target sequence within the target polynucleotide, wherein binding of the CRISPR complex to the target polynucleotide induces cell death, thereby allowing one or more cell(s) in which one or more mutations have been introduced to be selected. In a preferred embodiment of the invention the cell to be selected may be a eukaryotic cell. Aspects of the invention allow for selection of specific cells without requiring a selection marker or a two-step process that may include a counter-selection system.

[0573] With respect to mutations of the Cas13 enzyme, when the enzyme is not FnCas13, mutations may be as described herein elsewhere; conservative substitution for any of the replacement amino acids is also envisaged. In an aspect the invention provides as to any or each or all embodiments herein-discussed wherein the CRISPR enzyme comprises at least one or more, or at least two or more mutations, wherein the at least one or more mutation or the at least two or more mutations are selected from those described herein elsewhere.

[0574] In a further aspect, the invention involves a computer-assisted method for identifying or designing potential compounds to fit within or bind to CRISPR-Cas13 system or a functional portion thereof or vice versa (a computer-assisted method for identifying or designing potential CRISPR-Cas13 systems or a functional portion thereof for binding to desired compounds) or a computer-assisted method for identifying or designing potential CRISPR-Cas13 systems (e.g., with regard to predicting areas of the CRISPR-Cas13 system to be able to be manipulated-for instance, based on crystal structure data or based on data of Cas13 orthologs, or with respect to where a functional group such as an activator or repressor can be attached to the CRISPR-Cas13 system, or as to Cas13 truncations or as to designing nickases), said method comprising:

using a computer system, e.g., a programmed computer comprising a processor, a data storage system, an input device, and an output device, the steps of:

(a) inputting into the programmed computer through said input device data comprising the three-dimensional co-ordinates of a subset of the atoms from or pertaining to the CRISPR-Cas13 crystal structure, e.g., in the CRISPR-Cas13 system binding domain or alternatively or additionally in domains that vary based on variance among Cas13 orthologs or as to Cas13s or as to nickases or as to functional groups, optionally with structural information from CRISPR-Cas13 system complex(es), thereby generating a data set;

(b) comparing, using said processor, said data set to a computer database of structures stored in said computer data storage system, e.g., structures of compounds that bind or putatively bind or that are desired to bind to a CRISPR-Cas13 system or as to Cas13 orthologs (e.g., as Cas13s or as to domains or regions that vary amongst Cas13 orthologs) or as to the CRISPR-Cas13 crystal structure or as to nickases or as to functional groups;

(c) selecting from said database, using computer methods, structure(s)-e.g., CRISPR-Cas13 structures that may bind to desired structures, desired structures that may bind to certain CRISPR-Cas13 structures, portions of the CRISPR-Cas13 system that may be manipulated, e.g., based on data from other portions of the CRISPR-Cas13 crystal structure and/or from

Cas13 orthologs, truncated Cas13s, novel nickases or particular functional groups, or positions for attaching functional groups or functional-group-CRISPR-Cas13 systems;

(d) constructing, using computer methods, a model of the selected structure(s); and

(e) outputting to said output device the selected structure(s);

and optionally synthesizing one or more of the selected structure(s);

and further optionally testing said synthesized selected structure(s) as or in a CRISPR-Cas13 system;

or, said method comprising: providing the co-ordinates of at least two atoms of the CRISPR-Cas13 crystal structure, e.g., at least two atoms of the herein Crystal Structure Table of the CRISPR-Cas13 crystal structure or co-ordinates of at least a sub-domain of the CRISPR-Cas13 crystal structure ("selected co-ordinates"), providing the structure of a candidate comprising a binding molecule or of portions of the CRISPR-Cas13 system that may be manipulated, e.g., based on data from other portions of the CRISPR-Cas13 crystal structure and/or from Cas13 orthologs, or the structure of functional groups, and fitting the structure of the candidate to the selected co-ordinates, to thereby obtain product data comprising CRISPR-Cas13 structures that may bind to desired structures, desired structures that may bind to certain CRISPR-Cas13 structures, portions of the CRISPR-Cas13 system that may be manipulated, truncated Cas13s, novel nickases, or particular functional groups, or positions for attaching functional groups or functional-group-CRISPR-Cas13 systems, with output thereof; and optionally synthesizing compound(s) from said product data and further optionally comprising testing said synthesized compound(s) as or in a CRISPR-Cas13 system.

[0575] The testing can comprise analyzing the CRISPR-Cas13 system resulting from said synthesized selected structure(s), e.g., with respect to binding, or performing a desired function.

[0576] The output in the foregoing methods can comprise data transmission, e.g., transmission of information via telecommunication, telephone, video conference, mass communication, e.g., presentation such as a computer presentation (e.g. POWERPOINT), internet, email, documentary communication such as a computer program (e.g. WORD) document and the like. Accordingly, the invention also comprehends computer readable media containing: atomic co-ordinate data according to the herein-referenced Crystal Structure, said data defining the three dimensional structure of CRISPR-Cas13 or at least one sub-domain thereof, or structure factor data for CRISPR-Cas13, said structure factor data being derivable from the atomic co-ordinate data of herein-referenced Crystal Structure. The computer readable media can also contain any data of the foregoing methods. The invention further comprehends methods a computer system for generating or performing rational design as in

the foregoing methods containing either: atomic co-ordinate data according to herein-referenced Crystal Structure, said data defining the three dimensional structure of CRISPR-Cas13 or at least one sub-domain thereof, or structure factor data for CRISPR-Cas13, said structure factor data being derivable from the atomic co-ordinate data of herein-referenced Crystal Structure. The invention further comprehends a method of doing business comprising providing to a user the computer system or the media or the three dimensional structure of CRISPR-Cas13 or at least one sub-domain thereof, or structure factor data for CRISPR-Cas13, said structure set forth in and said structure factor data being derivable from the atomic co-ordinate data of herein-referenced Crystal Structure, or the herein computer media or a herein data transmission.

[0577] A "binding site" or an "active site" comprises or consists essentially of or consists of a site (such as an atom, a functional group of an amino acid residue or a plurality of such atoms and/or groups) in a binding cavity or region, which may bind to a compound such as a nucleic acid molecule, which is/are involved in binding.

[0578] By "fitting", is meant determining by automatic, or semi-automatic means, interactions between one or more atoms of a candidate molecule and at least one atom of a structure of the invention, and calculating the extent to which such interactions are stable. Interactions include attraction and repulsion, brought about by charge, steric considerations and the like. Various computer-based methods for fitting are described further

[0579] By "root mean square (or rms) deviation", we mean the square root of the arithmetic mean of the squares of the deviations from the mean.

[0580] By a "computer system", is meant the hardware means, software means and data storage means used to analyze atomic coordinate data. The minimum hardware means of the computer-based systems of the present invention typically comprises a central processing unit (CPU), input means, output means and data storage means. Desirably a display or monitor is provided to visualize structure data. The data storage means may be RAM or means for accessing computer readable media of the invention. Examples of such systems are computer and tablet devices running Unix, Windows or Apple operating systems.

[0581] By "computer readable media", is meant any medium or media, which can be read and accessed directly or indirectly by a computer e.g., so that the media is suitable for use in the above-mentioned computer system. Such media include, but are not limited to: magnetic storage media such as floppy discs, hard disc storage medium and magnetic tape; optical storage media such as optical discs or CD-ROM; electrical storage media such as RAM and

ROM; thumb drive devices; cloud storage devices and hybrids of these categories such as magnetic/optical storage media.

[0582] The invention comprehends the use of the protected guides described herein above in the optimized functional CRISPR-Cas enzyme systems described herein.

[0583] In some embodiments, the guide RNA is a toehold based guide RNA. The toehold based guide RNAs allows for guide RNAs only becoming activated based on the RNA levels of other transcripts in a cell. In certain embodiments, the guide RNA has an extension that includes a loop and a complementary sequence that fold over onto the guide and block the guide. The loop can be complementary to transcripts or miRNA in the cell and bind these transcripts if present. This will unfold the guide RNA allowing it to bind a Cas13 molecule. This bound complex can then knockdown transcripts or edit transcripts depending on the application.

CRISPR-Cas Enzyme

[0584] In its unmodified form, a CRISPR-Cas protein is a catalytically active protein. This implies that upon formation of a nucleic acid-targeting complex (comprising a guide RNA hybridized to a target sequence one or both DNA strands in or near (e.g., within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target sequence is modified (e.g. cleaved). As used herein the term "sequence(s) associated with a target locus of interest" refers to sequences near the vicinity of the target sequence (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from the target sequence, wherein the target sequence is comprised within a target locus of interest). The unmodified catalytically active Cas13 protein generates a staggered cut, whereby the cut sites are typically within the target sequence. More particularly, the staggered cut is typically 13–23 nucleotides distal to the PAM. In particular embodiments, the cut on the non-target strand is 17 nucleotides downstream of the PAM (i.e. between nucleotide 17 and 18 downstream of the PAM), while the cut on the target strand (i.e. strand hybridizing with the guide sequence) occurs a further 4 nucleotides further from the sequence complementary to the PAM (this is 21 nucleotides upstream of the complement of the PAM on the 3' strand or between nucleotide 21 and 22 upstream of the complement of the PAM).

[0585] In the methods according to the present invention, the CRISPR-Cas protein is preferably mutated with respect to a corresponding wild-type enzyme such that the mutated CRISPR-Cas protein lacks the ability to cleave one or both DNA strands of a target locus containing a target sequence. In particular embodiments, one or more catalytic domains of the Cas13 protein are mutated to produce a mutated Cas protein which cleaves only one DNA strand of a target sequence.

[0586] In particular embodiments, the CRISPR-Cas protein may be mutated with respect to a corresponding wild-type enzyme such that the mutated CRISPR-Cas protein lacks substantially all DNA cleavage activity. In some embodiments, a CRISPR-Cas protein may be considered to substantially lack all DNA and/or RNA cleavage activity when the cleavage activity of the mutated enzyme is about no more than 25%, 10%, 5%, 1%, 0.1%, 0.01%, or less of the nucleic acid cleavage activity of the non-mutated form of the enzyme; an example can be when the nucleic acid cleavage activity of the mutated form is nil or negligible as compared with the non-mutated form.

[0587] In certain embodiments of the methods provided herein the CRISPR-Cas protein is a mutated CRISPR-Cas protein which cleaves only one DNA strand, i.e. a nickase. More particularly, in the context of the present invention, the nickase ensures cleavage within the non-target sequence, i.e. the sequence which is on the opposite DNA strand of the target sequence and which is 3' of the PAM sequence. By means of further guidance, and without limitation, an arginine-to-alanine substitution (R1226A) in the Nuc domain of Cas13 from Acidaminococcus sp. converts Cas13 from a nuclease that cleaves both strands to a nickase (cleaves a single strand). It will be understood by the skilled person that where the enzyme is not AsCas13, a mutation may be made at a residue in a corresponding position. In particular embodiments, the Cas13 is FnCas13 and the mutation is at the arginine at position R1218. In particular embodiments, the Cas13 is LbCas13 and the mutation is at the arginine at position R1138. In particular embodiments, the Cas13 is MbCas13 and the mutation is at the arginine at position R1293.

[0588] In certain embodiments of the methods provided herein the CRISPR-Cas protein has reduced or no catalytic activity. Where the CRISPR-Cas protein is a Cas13 protein, the mutations may include but are not limited to one or more mutations in the catalytic RuvC-like domain, such as D908A or E993A with reference to the positions in AsCas13.

[0589] In some embodiments, a CRISPR-Cas protein is considered to substantially lack all DNA cleavage activity when the DNA cleavage activity of the mutated enzyme is about no more than 25%, 10%, 5%, 1%, 0.1%, 0.01%, or less of the DNA cleavage activity of the non-mutated form of the enzyme; an example can be when the DNA cleavage activity of the mutated form is nil or negligible as compared with the non-mutated form. In these embodiments, the CRISPR-Cas protein is used as a generic DNA binding protein. The mutations may be artificially introduced mutations or gain- or loss-of-function mutations.

[0590] In addition to the mutations described above, the CRISPR-Cas protein may be additionally modified. As used herein, the term "modified" with regard to a CRISPR-Cas

protein generally refers to a CRISPR-Cas protein having one or more modifications or mutations (including point mutations, truncations, insertions, deletions, chimeras, fusion proteins, etc.) compared to the wild type Cas protein from which it is derived. By derived is meant that the derived enzyme is largely based, in the sense of having a high degree of sequence homology with, a wildtype enzyme, but that it has been mutated (modified) in some way as known in the art or as described herein.

[0591] In some embodiments, to reduce the size of a fusion protein of the Cas13b effector and the one or more functional domains, the C-terminus of the Cas13b effector can be truncated while still maintaining its RNA binding function. For example, at least 20 amino acids, at least 50 amino acids, at least 80 amino acids, or at least 100 amino acids, or at least 150 amino acids, or at least 200 amino acids, or at least 250 amino acids, or at least 300 amino acids, or at least 350 amino acids, or up to 120 amino acids, or up to 140 amino acids, or up to 160 amino acids, or up to 180 amino acids, or up to 200 amino acids, or up to 250 amino acids, or up to 300 amino acids, or up to 350 amino acids, or up to 400 amino acids, may be truncated at the C-terminus of the Cas13b effector. Specific examples of Cas13b truncations include C-terminal Δ984-1090, C-terminal Δ1026-1090, and C-terminal Δ1053-1090, C-terminal Δ934-1090, C-terminal Δ884-1090, C-terminal Δ834-1090, C-terminal Δ784-1090, and C-terminal Δ734-1090, wherein amino acid positions correspond to amino acid positions of Prevotella sp. P5-125 Cas13b protein. See also FIG. 67.

[0592] The additional modifications of the CRISPR-Cas protein may or may not cause an altered functionality. By means of example, and in particular with reference to CRISPR-Cas protein, modifications which do not result in an altered functionality include for instance codon optimization for expression into a particular host, or providing the nuclease with a particular marker (e.g. for visualization). Modifications with may result in altered functionality may also include mutations, including point mutations, insertions, deletions, truncations (including split nucleases), etc.. Fusion proteins may without limitation include for instance fusions with heterologous domains or functional domains (e.g. localization signals, catalytic domains, etc.). In certain embodiments, various different modifications may be combined (e.g. a mutated nuclease which is catalytically inactive and which further is fused to a functional domain, such as for instance to induce DNA methylation or another nucleic acid modification, such as including without limitation a break (e.g. by a different nuclease (domain)), a mutation, a deletion, an insertion, a replacement, a ligation, a digestion, a break or a recombination). As used herein, "altered functionality" includes without limitation an altered specificity (e.g. altered target recognition, increased (e.g. "enhanced" Cas proteins) or decreased specificity, or

altered PAM recognition), altered activity (e.g. increased or decreased catalytic activity, including catalytically inactive nucleases or nickases), and/or altered stability (e.g. fusions with destalilization domains). Suitable heterologous domains include without limitation a nuclease, a ligase, a repair protein, a methyltransferase, (viral) integrase, a recombinase, a transposase, an argonaute, a cytidine deaminase, a retron, a group II intron, a phosphatase, a phosphorylase, a sulpfurylase, a kinase, a polymerase, an exonuclease, etc.. Examples of all these modifications are known in the art. It will be understood that a "modified" nuclease as referred to herein, and in particular a "modified" Cas or "modified" CRISPR-Cas system or complex preferably still has the capacity to interact with or bind to the polynucleic acid (e.g. in complex with theguide molecule). Such modified Cas protein can be combined with the deaminase protein or active domain thereof as described herein.

[0593] In certain embodiments, CRISPR-Cas protein may comprise one or more modifications resulting in enhanced activity and/or specificity, such as including mutating residues that stabilize the targeted or non-targeted strand (e.g. eCas9; "Rationally engineered Cas9 nucleases with improved specificity", Slaymaker et al. (2016), Science, 351(6268):84-88, incorporated herewith in its entirety by reference). In certain embodiments, the altered or modified activity of the engineered CRISPR protein comprises increased targeting efficiency or decreased off-target binding. In certain embodiments, the altered activity of the engineered CRISPR protein comprises modified cleavage activity. In certain embodiments, the altered activity comprises increased cleavage activity as to the target polynucleotide loci. In certain embodiments, the altered activity comprises decreased cleavage activity as to the target polynucleotide loci. In certain embodiments, the altered activity comprises decreased cleavage activity as to off-target polynucleotide loci. In certain embodiments, the altered or modified activity of the modified nuclease comprises altered helicase kinetics. In certain embodiments, the modified nuclease comprises a modification that alters association of the protein with the nucleic acid molecule comprising RNA (in the case of a Cas protein), or a strand of the target polynucleotide loci, or a strand of off-target polynucleotide loci. In an aspect of the invention, the engineered CRISPR protein comprises a modification that alters formation of the CRISPR complex. In certain embodiments, the altered activity comprises increased cleavage activity as to off-target polynucleotide loci. Accordingly, in certain embodiments, there is increased specificity for target polynucleotide loci as compared to off-target polynucleotide loci. In other embodiments, there is reduced specificity for target polynucleotide loci as compared to off-target polynucleotide loci. In certain embodiments, the mutations result in decreased off-target effects (e.g. cleavage or binding properties, activity, or kinetics), such as in case for Cas

proteins for instance resulting in a lower tolerance for mismatches between target and guide RNA. Other mutations may lead to increased off-target effects (e.g. cleavage or binding properties, activity, or kinetics). Other mutations may lead to increased or decreased on-target effects (e.g. cleavage or binding properties, activity, or kinetics). In certain embodiments, the mutations result in altered (e.g. increased or decreased) helicase activity, association or formation of the functional nuclease complex (e.g. CRISPR-Cas complex). In certain embodiments, as described above, the mutations result in an altered PAM recognition, i.e. a different PAM may be (in addition or in the alternative) be recognized, compared to the unmodified Cas protein. Particularly preferred mutations include positively charged residues and/or (evolutionary) conserved residues, such as conserved positively charged residues, in order to enhance specificity. In certain embodiments, such residues may be mutated to uncharged residues, such as alanine.

[0594] In certain embodiments, the methods, products, and uses as described herein can be expanded or adapted to implement any type of CRISPR effector.

[0595] In certain embodiments, the CRISPR effector is a class 2 CRISPR-Cas system effector. It is to be understood that the term "CRISPR effector" preferably refers to an RNA-guided endonuclease. The skilled person will understand that the CRISPR effector may be modified, as described herein elsewhere, and as known in the art. By means of example, and without limitation, CRISPR effector modifications include modifications affecting CRISPR effector functionality or nuclease activity (e.g. catalytically inactive variants (optionally fused or otherwise associated with heterologous functional domains), nickases, altered PAM specificity/recognition, split CRISPR effectors,...), specificity (e.g. enhanced specificity mutants), stability (e.g. destabilized variants), etc.

[0596] In certain embodiments, the CRISPR effector cleaves, binds to, or associates with RNA. In certain embodiments, the CRISPR effector cleaves, binds to, or associates with DNA. In certain embodiments, the CRISPR effector cleaves, binds to, or associates with single stranded RNA. In certain embodiments, the CRISPR effector cleaves, binds to, or associates with single stranded DNA. In certain embodiments, the CRISPR effector cleaves, binds to, or associates with double stranded RNA. In certain embodiments, the CRISPR effector cleaves, binds to, or associates with Double stranded DNA. In certain embodiments, the CRISPR effector cleaves, binds to, or associates with DNA/RNA hybrids.

[0597] In certain embodiments, the CRISPR effector is a class 2, type II CRISPR effector. In certain embodiments, the CRISPR effector is a class 2, type II-A CRISPR effector. In certain embodiments, the CRISPR effector is a class 2, type II-B CRISPR effector. In certain

embodiments, the CRISPR effector is a class 2, type II-C CRISPR effector. In certain embodiments, the CRISPR effector is Cas9.

[0598] In certain embodiments, the CRISPR effector is a class 2, type V CRISPR effector. In certain embodiments, the CRISPR effector is a class 2, type V-A CRISPR effector. In certain embodiments, the CRISPR effector is a class 2, type V-B CRISPR effector. In certain embodiments, the CRISPR effector is a class 2, type V-C CRISPR effector. In certain embodiments, the CRISPR effector is Cas12a (Cpf1). In certain embodiments, the CRISPR effector is Cas12b (C2c1). In certain embodiments, the CRISPR effector is Cas12c (C2c3). In certain embodiments, the CRISPR effector is a class 2, type V-U CRISPR effector. In certain embodiments, the CRISPR effector is a class 2, type V-U1 CRISPR effector (e.g. C2c4). In certain embodiments, the CRISPR effector is a class 2, type V-U2 CRISPR effector (e.g. C2c8). In certain embodiments, the CRISPR effector is a class 2, type V-U3 CRISPR effector (e.g. C2c10). In certain embodiments, the CRISPR effector is a class 2, type V-U4 CRISPR effector (e.g. C2c9). In certain embodiments, the CRISPR effector is a class 2, type V-U5 CRISPR effector (e.g. C2c5).

[0599] In certain embodiments, the CRISPR effector is a class 2, type VI CRISPR effector. In certain embodiments, the CRISPR effector is a class 2, type VI-A CRISPR effector. In certain embodiments, the CRISPR effector is a class 2, type VI-B CRISPR effector. In certain embodiments, the CRISPR effector is a class 2, type VI-B1 CRISPR effector. In certain embodiments, the CRISPR effector is a class 2, type VI-B2 CRISPR effector. In certain embodiments, the CRISPR effector is a class 2, type VI-C CRISPR effector. In certain embodiments, the CRISPR effector is Cas13a (C2c2). In certain embodiments, the CRISPR effector is Cas13b (C2c6). In certain embodiments, the CRISPR effector is Cas13c (C2c7).

[0600] In certain embodiments, the CRISPR effector comprises one or more RuvC domain. In certain embodiments, the CRISPR effector comprises a RuvC-I domain. In certain embodiments, the CRISPR effector comprises a RuvC-II domain. In certain embodiments, the CRISPR effector comprises a RuvC-III domain. In certain embodiments, the CRISPR effector comprises a RuvC-I, RuvC-II, and RuvC-III domain. In certain embodiments, one or more of RuvC-I, II, and/or III are contiguous motifs. In certain embodiments, one or more of RuvC-I, II, and/or III are non-contiguous or discrete motifs. In certain embodiments, the CRISPR effector comprises one or more HNH domain. In certain embodiments, the CRISPR effector comprises one or more RuvC domain and one or more HNH domain. In certain embodiments, the CRISPR effector comprises a RuvC-I domain and an HNH domain. In certain embodiments, the CRISPR effector comprises a RuvC-II domain and an HNH domain. In

certain embodiments, the CRISPR effector comprises a RuvC-III domain and an HNH domain. In certain embodiments, the CRISPR effector comprises a RuvC-I, RuvC-II, and RuvC-III domain and an HNH domain. In certain embodiments, the CRISPR effector comprises one or more Nuc (nuclease) domain. In certain embodiments, the CRISPR effector comprises one or more RuvC domain and one or more Nuc domain. In certain embodiments, the CRISPR effector comprises a RuvC-I domain and a Nuc domain. In certain embodiments, the CRISPR effector comprises a RuvC-II domain and a Nuc domain. In certain embodiments, the CRISPR effector comprises a RuvC-III domain and a Nuc domain.

[0601] In certain embodiments, the CRISPR effector comprises one or more HEPN domain. In certaim embodiments, the CRISPR effector comprises a HEPN I domain. In certain embodiments, the CRISPR effector comprises a HEPN II domain. In certain embodiments, the CRISPR effector comprises a HEPN I domain and a HEPN II domain. In certain embodiments, one or more of the HEPN domains are contiguous domains. In certain embodiments, one or more of the HEPN domains comprise non-contiguous or discrete motifs.

[0602] In certain embodiments, the CRISPR effector is a CRISPR effector as disclosed for instance in Shmakov et al. (2017), "Diversity and evolution of class 2 CRISPR-Cas systems", Nature Rev Microbiol, 15(3):169-182; Shmakov et al. (2015) "Discovery and functional characterization of diverse class 2 CRISPR-Cas systems", Mol Cell, 60(3):385-397; Makarova et al. (2015), "An updated evolutionary classification of CRISPR-Cas systems", Nat Rev Microbiol, 13(11):722-736; or Koonin et al. (2017), "Diversity, classification and evolution of CRISPR-Cas systems", Curr Opin Microbiol, 37:67-78. All are incorporated herein by reference in their entirety, as well as the references cited therein.

[0603] The skilled person will understand that the choice of CRISPR effector may depend on the application (e.g. knockout or suppression, activation,…) , as well as the target (e.g. RNA or DNA, single or double stranded, as well as target sequence, including associated PAM sequence and/or specificity,…). It will be understood, that the choice of CRISPR effector may determine the particulars of other CRISPR-Cas system components (e.g. spacer (or guide sequence) length, direct repeat (or tracr mate) sequence or length, the presence or absence of a tracr, as well as tracr sequence or length, etc.).

[0604] CRISPR-Cas systems have been identified in numerous archaeal and bacterial species. The skilled person will understand that CRISPR effector homologues or orthologues from any of the identified CRISPR-Cas systems may advantageously be used in certain embodiments. It will be understood that further homologues (e.g. additional class 2 types of CRISPR-Cas systems and CRISPR effectors) or orthologues (e.g. known or unknown

CRISPR-Cas systems or CRISPR effectors from additional archaeal or bacterial species) can be identified. Such may suitably be used in certain embodiments and aspects of the invention.

[0605] By means of example, CRISPR-Cas systems (and CRISPR effectors) may be identified for instance and without limitation as described in Shmakov et al. (2017), "Diversity and evolution of class 2 CRISPR-Cas systems", Nature Rev Microbiol, 15(3):169-182 or Shmakov et al. (2015) "Discovery and functional characterization of diverse class 2 CRISPR-Cas systems", Mol Cell, 60(3):385-397. The methodology for identifying CRISPR-Cas systems and effectors is explicitly incorporated herein by reference.

[0606] In certain embodiments, a method for the systematic detection of class 2 CRISPR-Cas systems may begin with the identification of a 'seed' that signifies the likely presence of a CRISPR-Cas locus in a given nucleotide sequence. For instance, Cas1 may be used as the seed, as it is the most common Cas protein in CRISPR-Cas systems and is most highly conserved at the sequence level. Sequence databases may be searched with this seed. To ensure the maximum sensitivity of detection, the search may be carried out by comparing a Cas1 sequence profile to translated genomic and metagenomic sequences. After the Cas1 genes are detected, their respective 'neighbourhoods' are examined for the presence of other Cas genes by searching with previously developed profiles for Cas proteins and applying the criteria for the classification of the CRISPR-Cas loci. In a complementary approach, to extend the search to non-autonomous CRISPR-Cas systems, the same procedure may be repeated using the CRISPR array as the seed. To ensure that the CRISPR array is detected at a high level of sensitivity, the predictions can be made for instance using the Piler-CR72 and CRISPRfinder methods, which predictions can be pooled and taken as the final CRISPR set. As illustrated in Shmakov et al. (2017), "Diversity and evolution of class 2 CRISPR-Cas systems", Nature Rev Microbiol, 15(3):169-182, this latter procedure (i.e. using the CRISPR array as seed) yielded 47,174 CRISPR arrays, which is more than twice the number of Cas1 genes that were detected, reflecting the fact that many CRISPR-Cas loci lack the adaptation module and that numerous 'orphan' arrays, some of which seem to be functional, also exist.

[0607] All loci can either subsequently be assigned to known CRISPR-Cas subtypes through the Cas protein profile search or alternatively can be assigned to new subtypes. In certain embodiments, among the Cas1 or CRISPR neighborhoods, those that encode large proteins (>500 amino acids) can be analyzed in detail, given that Cas9 and Cpf1 are large proteins (typically >1000 amino acids) and that their protein structures suggest that this large size is required to accommodate the CRISPR RNA (crRNA)-target DNA complex. The sequences of such large proteins can then be screened for known protein domains using

sensitive profile-based methods, such as HHpred, secondary structure prediction and manual examination of multiple alignments. Under the premise that class 2 effector proteins contain nuclease domains, even if they are distantly related or unrelated to known families of nucleases, the proteins that contain domains that are deemed irrelevant in the context of the CRISPR-Cas function (for example, membrane transporters or metabolic enzymes) can be discarded. The retained proteins either contain readily identifiable, or completely unknown, nuclease domains. The sequences of these proteins can then be analyzed using the most sensitive methods for domain detection, such as HHpred, with a curated multiple alignment of the respective protein sequences that can be used as the query. The use of sensitive methods is essential because proteins that are involved in antiviral defense, and the Cas proteins in particular, typically evolve extremely fast. The above procedure for the discovery of class 2 CRISPR-Cas systems, at least in principle, is expected to be exhaustive, because all loci that contain a gene that encodes a large protein (that is, a putative class 2 effector) in the vicinity of cas1 and/or CRISPR are analyzed in detail. The assumption of the structural requirements for a class 2 effector, which underlie the protein size cut-off that is used, and the precision of Cas1 and CRISPR detection, are the only limitations of this approach.

[0608] In certain embodiments, the CRISPR effector is a CRISPR effector as identified for instance according to the methodology presented above. It will be understood that functionality of the identified CRISPR effectors can be readily evaluated and validated by the skilled person.

Base Excision Repair Inhibitor

[0609] In some embodiments, the AD-functionalized CRISPR system further comprises a base excision repair (BER) inhibitor. Without wishing to be bound by any particular theory, cellular DNA-repair response to the presence of I:T pairing may be responsible for a decrease in nucleobase editing efficiency in cells. Alkyladenine DNA glycosylase (also known as DNA-3-methyladenine glycosylase, 3-alkyladenine DNA glycosylase, or N-methylpurine DNA glycosylase) catalyzes removal of hypoxanthine from DNA in cells, which may initiate base excision repair, with reversion of the I:T pair to a A:T pair as outcome.

[0610] In some embodiments, the BER inhibitor is an inhibitor of alkyladenine DNA glycosylase. In some embodiments, the BER inhibitor is an inhibitor of human alkyladenine DNA glycosylase. In some embodiments, the BER inhibitor is a polypeptide inhibitor. In some embodiments, the BER inhibitor is a protein that binds hypoxanthine. In some embodiments, the BER inhibitor is a protein that binds hypoxanthine in DNA. In some embodiments, the BER inhibitor is a catalytically inactive alkyladenine DNA glycosylase protein or binding domain

thereof. In some embodiments, the BER inhibitor is a catalytically inactive alkyladenine DNA glycosylase protein or binding domain thereof that does not excise hypoxanthine from the DNA. Other proteins that are capable of inhibiting (e.g., sterically blocking) an alkyladenine DNA glycosylase base-excision repair enzyme are within the scope of this disclosure. Additionally, any proteins that block or inhibit base-excision repair as also within the scope of this disclosure.

[0611] Without wishing to be bound by any particular theory, base excision repair may be inhibited by molecules that bind the edited strand, block the edited base, inhibit alkyladenine DNA glycosylase, inhibit base excision repair, protect the edited base, and/or promote fixing of the non-edited strand. It is believed that the use of the BER inhibitor described herein can increase the editing efficiency of an adenosine deaminase that is capable of catalyzing a A to I change.

[0612] Accordingly, in the first design of the AD-functionalized CRISPR system discussed above, the CRISPR-Cas protein or the adenosine deaminase can be fused to or linked to a BER inhibitor (e.g., an inhibitor of alkyladenine DNA glycosylase). In some embodiments, the BER inhibitor can be comprised in one of the following structures (nCas13=Cas13 nickase; dCas13=dead Cas13);
[AD]-[optional linker]-[nCas13/dCas13]-[optional linker]-[BER inhibitor];
[AD]-[optional linker]-[BER inhibitor]-[optional linker]-[nCas13/dCas13];
[BER inhibitor]-[optional linker]-[AD]-[optional linker]-[nCas13/dCas13];
[BER inhibitor]-[optional linker]-[nCas13/dCas13]-[optional linker]-[AD];
[nCas13/dCas13]-[optional linker]-[AD]-[optional linker]-[BER inhibitor];
[nCas13/dCas13]-[optional linker]-[BER inhibitor]-[optional linker]-[AD].

[0613] Similarly, in the second design of the AD-functionalized CRISPR system discussed above, the CRISPR-Cas protein, the adenosine deaminase, or the adaptor protein can be fused to or linked to a BER inhibitor (e.g., an inhibitor of alkyladenine DNA glycosylase). In some embodiments, the BER inhibitor can be comprised in one of the following structures (nCas13=Cas13 nickase; dCas13=dead Cas13):
[nCas13/dCas13]-[optional linker]-[BER inhibitor];
[BER inhibitor]-[optional linker]-[nCas13/dCas13];
[AD]-[optional linker]-[Adaptor]-[optional linker]-[BER inhibitor];
[AD]-[optional linker]-[BER inhibitor]-[optional linker]-[Adaptor];
[BER inhibitor]-[optional linker]-[AD]-[optional linker]-[Adaptor];
[BER inhibitor]-[optional linker]-[Adaptor]-[optional linker]-[AD];

[Adaptor]-[optional linker]-[AD]-[optional linker]-[BER inhibitor]; [Adaptor]-[optional linker]-[BER inhibitor]-[optional linker]-[AD].

[0614] In the third design of the AD-functionalized CRISPR system discussed above, the BER inhibitor can be inserted into an internal loop or unstructured region of a CRISPR-Cas protein.

Targeting to the Nucleus

[0615] In some embodiments, the methods of the present invention relate to modifying an Adenine in a target locus of interest, whereby the target locus is within a cell. In order to improve targeting of the CRISPR-Cas protein and/or the adenosine deaminase protein or catalytic domain thereof used in the methods of the present invention to the nucleus, it may be advantageous to provide one or both of these components with one or more nuclear localization sequences (NLSs).

[0616] In preferred embodiments, the NLSs used in the context of the present invention are heterologous to the proteins. Non-limiting examples of NLSs include an NLS sequence derived from: the NLS of the SV40 virus large T-antigen, having the amino acid sequence PKKKRKV (SEQ ID No. 17) or PKKKRKVEAS (SEQ ID No. 18); the NLS from nucleoplasmin (e.g., the nucleoplasmin bipartite NLS with the sequence KRPAATKKAGQAKKKK (SEQ ID No. 19)); the c-myc NLS having the amino acid sequence PAAKRVKLD (SEQ ID No. 20) or RQRRNELKRSP (SEQ ID No. 21); the hRNPA1 M9 NLS having the sequence NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID No. 22); the sequence RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID No. 23) of the IBB domain from importin-alpha; the sequences VSRKRPRP (SEQ ID No. 24) and PPKKARED (SEQ ID No. 25) of the myoma T protein; the sequence PQPKKKPL (SEQ ID No. 26) of human p53; the sequence SALIKKKKKMAP (SEQ ID No. 27) of mouse c-abl IV; the sequences DRLRR (SEQ ID No. 28) and PKQKKRK (SEQ ID No. 29) of the influenza virus NS1; the sequence RKLKKKIKKL (SEQ ID No. 30) of the Hepatitis virus delta antigen; the sequence REKKKFLKRR (SEQ ID No. 31) of the mouse Mx1 protein; the sequence KRKGDEVDGVDEVAKKKSKK (SEQ ID No. 32) of the human poly(ADP-ribose) polymerase; and the sequence RKCLQAGMNLEARKTKK (SEQ ID No. 33) of the steroid hormone receptors (human) glucocorticoid. In general, the one or more NLSs are of sufficient strength to drive accumulation of the DNA-targeting Cas protein in a detectable amount in the nucleus of a eukaryotic cell. In general, strength of nuclear localization activity may derive from the number of NLSs in the CRISPR-Cas protein, the particular NLS(s) used, or a combination of these factors. Detection of accumulation in the nucleus may be performed by

any suitable technique. For example, a detectable marker may be fused to the nucleic acid-targeting protein, such that location within a cell may be visualized, such as in combination with a means for detecting the location of the nucleus (e.g., a stain specific for the nucleus such as DAPI). Cell nuclei may also be isolated from cells, the contents of which may then be analyzed by any suitable process for detecting protein, such as immunohistochemistry, Western blot, or enzyme activity assay. Accumulation in the nucleus may also be determined indirectly, such as by an assay for the effect of nucleic acid-targeting complex formation (e.g., assay for deaminase activity) at the target sequence, or assay for altered gene expression activity affected by DNA-targeting complex formation and/or DNA-targeting), as compared to a control not exposed to the CRISPR-Cas protein and deaminase protein, or exposed to a CRISPR-Cas and/or deaminase protein lacking the one or more NLSs.

[0617] The CRISPR-Cas and/or adenosine deaminase proteins may be provided with 1 or more, such as with, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more heterologous NLSs. In some embodiments, the proteins comprises about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the amino-terminus, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the carboxy-terminus, or a combination of these (e.g., zero or at least one or more NLS at the amino-terminus and zero or at one or more NLS at the carboxy terminus). When more than one NLS is present, each may be selected independently of the others, such that a single NLS may be present in more than one copy and/or in combination with one or more other NLSs present in one or more copies. In some embodiments, an NLS is considered near the N- or C-terminus when the nearest amino acid of the NLS is within about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, or more amino acids along the polypeptide chain from the N- or C-terminus. In preferred embodiments of the CRISPR-Cas proteins, an NLS attached to the C-terminal of the protein.

[0618] In certain embodiments of the methods provided herein, the CRISPR-Cas protein and the deaminase protein are delivered to the cell or expressed within the cell as separate proteins. In these embodiments, each of the CRISPR-Cas and deaminase protein can be provided with one or more NLSs as described herein. In certain embodiments, the CRISPR-Cas and deaminase proteins are delivered to the cell or expressed with the cell as a fusion protein. In these embodiments one or both of the CRISPR-Cas and deaminase protein is provided with one or more NLSs. Where the adenosine deaminase is fused to an adaptor protein (such as MS2) as described above, the one or more NLS can be provided on the adaptor protein, provided that this does not interfere with aptamer binding. In particular embodiments, the one

or more NLS sequences may also function as linker sequences between the adenosine deaminase and the CRISPR-Cas protein.

[0619] In certain embodiments, guides of the invention comprise specific binding sites (e.g. aptamers) for adapter proteins, which may be linked to or fused to an adenosine deaminase or catalytic domain thereof. When such a guides forms a CRISPR complex (i.e. CRISPR-Cas protein binding to guide and target) the adapter proteins bind and, the adenosine deaminase or catalytic domain thereof associated with the adapter protein is positioned in a spatial orientation which is advantageous for the attributed function to be effective.

[0620] The skilled person will understand that modifications to the guide which allow for binding of the adapter + adenosine deaminase, but not proper positioning of the adapter + adenosine deaminase (e.g. due to steric hindrance within the three dimensional structure of the CRISPR complex) are modifications which are not intended. The one or more modified guide may be modified at the tetra loop, the stem loop 1, stem loop 2, or stem loop 3, as described herein, preferably at either the tetra loop or stem loop 2, and most preferably at both the tetra loop and stem loop 2.

Use of orthogonal catalytically inactive CRISPR-Cas proteins

[0621] In particular embodiments, the Cas13 nickase is used in combination with an orthogonal catalytically inactive CRISPR-Cas protein to increase efficiency of said Cas13 nickase (as described in Chen et al. 2017, Nature Communications 8:14958; doi:10.1038/ncomms14958). More particularly, the orthogonal catalytically inactive CRISPR-Cas protein is characterized by a different PAM recognition site than the Cas13 nickase used in the AD-functionalized CRISPR system and the corresponding guide sequence is selected to bind to a target sequence proximal to that of the Cas13 nickase of the AD-functionalized CRISPR system. The orthogonal catalytically inactive CRISPR-Cas protein as used in the context of the present invention does not form part of the AD-functionalized CRISPR system but merely functions to increase the efficiency of said Cas13 nickase and is used in combination with a standard guide molecule as described in the art for said CRISPR-Cas protein. In particular embodiments, said orthogonal catalytically inactive CRISPR-Cas protein is a dead CRISPR-Cas protein, i.e. comprising one or more mutations which abolishes the nuclease activity of said CRISPR-Cas protein. In particular embodiments, the catalytically inactive orthogonal CRISPR-Cas protein is provided with two or more guide molecules which are capable of hybridizing to target sequences which are proximal to the target sequence of the Cas13 nickase. In particular embodiments, at least two guide molecules are used to target said catalytically inactive CRISPR-Cas protein, of which at least one guide molecule is capable of

hybridizing to a target sequence 5" of the target sequence of the Cas13 nickase and at least one guide molecule is capable of hybridizing to a target sequence 3' of the target sequence of the Cas13 nickase of the AD-functionalized CRISPR system, whereby said one or more target sequences may be on the same or the opposite DNA strand as the target sequence of the Cas13 nickase. In particular embodiments, the guide sequences for the one or more guide molecules of the orthogonal catalytically inactive CRISPR-Cas protein are selected such that the target sequences are proximal to that of the guide molecule for the targeting of the AD-functionalized CRISPR, i.e. for the targeting of the Cas13 nickase. In particular embodiments, the one or more target sequences of the orthogonal catalytically inactive CRISPR-Cas enzyme are each separated from the target sequence of the Cas13 nickase by more than 5 but less than 450 basepairs. Optimal distances between the target sequences of the guides for use with the orthogonal catalytically inactive CRISPR-Cas protein and the target sequence of the AD-functionalized CRISPR system can be determined by the skilled person. In particular embodiments, the orthogonal CRISPR-Cas protein is a Class II, type II CRISPR protein. In particular embodiments, the orthogonal CRISPR-Cas protein is a Class II, type V CRISPR protein. In particular embodiments, the catalytically inactive orthogonal CRISPR-Cas protein In particular embodiments, the catalytically inactive orthogonal CRISPR-Cas protein has been modified to alter its PAM specificity as described elsewhere herein. In particular embodiments, the Cas13 protein nickase is a nickase which, by itself has limited activity in human cells, but which, in combination with an inactive orthogonal CRISPR-Cas protein and one or more corresponding proximal guides ensures the required nickase activity.

CRISPR Development and Use

[0622] The present invention may be further illustrated and extended based on aspects of CRISPR-Cas development and use as set forth in the following articles and particularly as relates to delivery of a CRISPR protein complex and uses of an RNA guided endonuclease in cells and organisms:

- Multiplex genome engineering using CRISPR-Cas systems. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., & Zhang, F. Science Feb 15;339(6121):819-23 (2013);
- RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Jiang W., Bikard D., Cox D., Zhang F, Marraffini LA. Nat Biotechnol Mar;31(3):233-9 (2013);
- One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR-Cas-Mediated Genome Engineering. Wang H., Yang H., Shivalila CS., Dawlaty MM., Cheng AW., Zhang F., Jaenisch R. Cell May 9;153(4):910-8 (2013);

- Optical control of mammalian endogenous transcription and epigenetic states. Konermann S, Brigham MD, Trevino AE, Hsu PD, Heidenreich M, Cong L, Platt RJ, Scott DA, Church GM, Zhang F. Nature. Aug 22;500(7463):472-6. doi: 10.1038/Nature12466. Epub 2013 Aug 23 (2013);
- Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Ran, FA., Hsu, PD., Lin, CY., Gootenberg, JS., Konermann, S., Trevino, AE., Scott, DA., Inoue, A., Matoba, S., Zhang, Y., & Zhang, F. Cell Aug 28. pii: S0092-8674(13)01015-5 (2013-A);
- DNA targeting specificity of RNA-guided Cas9 nucleases. Hsu, P., Scott, D., Weinstein, J., Ran, FA., Konermann, S., Agarwala, V., Li, Y., Fine, E., Wu, X., Shalem, O., Cradick, TJ., Marraffini, LA., Bao, G., & Zhang, F. Nat Biotechnol doi:10.1038/nbt.2647 (2013);
- Genome engineering using the CRISPR-Cas9 system. Ran, FA., Hsu, PD., Wright, J., Agarwala, V., Scott, DA., Zhang, F. Nature Protocols Nov;8(11):2281-308 (2013-B);
- Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Shalem, O., Sanjana, NE., Hartenian, E., Shi, X., Scott, DA., Mikkelson, T., Heckl, D., Ebert, BL., Root, DE., Doench, JG., Zhang, F. Science Dec 12. (2013);
- Crystal structure of cas9 in complex with guide RNA and target DNA. Nishimasu, H., Ran, FA., Hsu, PD., Konermann, S., Shehata, SI., Dohmae, N., Ishitani, R., Zhang, F., Nureki, O. Cell Feb 27, 156(5):935-49 (2014);
- Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Wu X., Scott DA., Kriz AJ., Chiu AC., Hsu PD., Dadon DB., Cheng AW., Trevino AE., Konermann S., Chen S., Jaenisch R., Zhang F., Sharp PA. Nat Biotechnol. Apr 20. doi: 10.1038/nbt.2889 (2014);
- CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. Platt RJ, Chen S, Zhou Y, Yim MJ, Swiech L, Kempton HR, Dahlman JE, Parnas O, Eisenhaure TM, Jovanovic M, Graham DB, Jhunjhunwala S, Heidenreich M, Xavier RJ, Langer R, Anderson DG, Hacohen N, Regev A, Feng G, Sharp PA, Zhang F. Cell 159(2): 440-455 DOI: 10.1016/j.cell.2014.09.014(2014);
- Development and Applications of CRISPR-Cas9 for Genome Engineering, Hsu PD, Lander ES, Zhang F., Cell. Jun 5;157(6):1262-78 (2014).
- Genetic screens in human cells using the CRISPR-Cas9 system, Wang T, Wei JJ, Sabatini DM, Lander ES., Science. January 3; 343(6166): 80–84. doi:10.1126/science.1246981 (2014);

- Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation, Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE., (published online 3 September 2014) Nat Biotechnol. Dec;32(12):1262-7 (2014);
- *In vivo* interrogation of gene function in the mammalian brain using CRISPR-Cas9, Swiech L, Heidenreich M, Banerjee A, Habib N, Li Y, Trombetta J, Sur M, Zhang F., (published online 19 October 2014) Nat Biotechnol. Jan;33(1):102-6 (2015);
- Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex, Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F., Nature. Jan 29;517(7536):583-8 (2015).
- A split-Cas9 architecture for inducible genome editing and transcription modulation, Zetsche B, Volz SE, Zhang F., (published online 02 February 2015) Nat Biotechnol. Feb;33(2):139-42 (2015);
- Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis, Chen S, Sanjana NE, Zheng K, Shalem O, Lee K, Shi X, Scott DA, Song J, Pan JQ, Weissleder R, Lee H, Zhang F, Sharp PA. Cell 160, 1246–1260, March 12, 2015 (multiplex screen in mouse), and
- In *vivo* genome editing using *Staphylococcus aureus* Cas9, Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, Koonin EV, Sharp PA, Zhang F., (published online 01 April 2015), Nature. Apr 9;520(7546):186-91 (2015).
- Shalem et al., "High-throughput functional genomics using CRISPR-Cas9," Nature Reviews Genetics 16, 299-311 (May 2015).
- Xu et al., "Sequence determinants of improved CRISPR sgRNA design," Genome Research 25, 1147-1157 (August 2015).
- Parnas et al., "A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks," Cell 162, 675-686 (July 30, 2015).
- Ramanan et al., CRISPR-Cas9 cleavage of viral DNA efficiently suppresses hepatitis B virus," Scientific Reports 5:10833. doi: 10.1038/srep10833 (June 2, 2015)
- Nishimasu et al., Crystal Structure of Staphylococcus aureus Cas9," Cell 162, 1113-1126 (Aug. 27, 2015)

- BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis, Canver et al., Nature 527(7577):192-7 (Nov. 12, 2015) doi: 10.1038/nature15521. Epub 2015 Sep 16.
- *Cas13 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System*, Zetsche et al., Cell 163, 759-71 (Sep 25, 2015).
- *Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems*, Shmakov et al., Molecular Cell, 60(3), 385–397 doi: 10.1016/j.molcel.2015.10.008 Epub October 22, 2015.
- *Rationally engineered Cas9 nucleases with improved specificity*, Slaymaker et al., Science 2016 Jan 1 351(6268): 84-88 doi: 10.1126/science.aad5227. Epub 2015 Dec 1.
- Gao *et al*, "Engineered Cas13 Enzymes with Altered PAM Specificities," bioRxiv 091611; doi: http://dx.doi.org/10.1101/091611 (Dec. 4, 2016).

each of which is incorporated herein by reference, may be considered in the practice of the instant invention, and discussed briefly below:

- Cong *et al.* engineered type II CRISPR-Cas systems for use in eukaryotic cells based on both *Streptococcus thermophilus* Cas9 and also *Streptococcus pyogenes* Cas9 and demonstrated that Cas9 nucleases can be directed by short RNAs to induce precise cleavage of DNA in human and mouse cells. Their study further showed that Cas9 as converted into a nicking enzyme can be used to facilitate homology-directed repair in eukaryotic cells with minimal mutagenic activity. Additionally, their study demonstrated that multiple guide sequences can be encoded into a single CRISPR array to enable simultaneous editing of several at endogenous genomic loci sites within the mammalian genome, demonstrating easy programmability and wide applicability of the RNA-guided nuclease technology. This ability to use RNA to program sequence specific DNA cleavage in cells defined a new class of genome engineering tools. These studies further showed that other CRISPR loci are likely to be transplantable into mammalian cells and can also mediate mammalian genome cleavage. Importantly, it can be envisaged that several aspects of the CRISPR-Cas system can be further improved to increase its efficiency and versatility.
- Jiang *et al.* used the clustered, regularly interspaced, short palindromic repeats (CRISPR)–associated Cas9 endonuclease complexed with dual-RNAs to introduce precise mutations in the genomes of *Streptococcus pneumoniae* and *Escherichia coli.* The approach relied on dual-RNA:Cas9-directed cleavage at the targeted genomic site to kill unmutated cells and circumvents the need for selectable markers or counter-

selection systems. The study reported reprogramming dual-RNA:Cas9 specificity by changing the sequence of short CRISPR RNA (crRNA) to make single- and multinucleotide changes carried on editing templates. The study showed that simultaneous use of two crRNAs enabled multiplex mutagenesis. Furthermore, when the approach was used in combination with recombineering, in *S. pneumoniae*, nearly 100% of cells that were recovered using the described approach contained the desired mutation, and in *E. coli,* 65% that were recovered contained the mutation.

- Wang *et al.* (2013) used the CRISPR-Cas system for the one-step generation of mice carrying mutations in multiple genes which were traditionally generated in multiple steps by sequential recombination in embryonic stem cells and/or time-consuming intercrossing of mice with a single mutation. The CRISPR-Cas system will greatly accelerate the *in vivo* study of functionally redundant genes and of epistatic gene interactions.
- Konermann *et al.* (2013) addressed the need in the art for versatile and robust technologies that enable optical and chemical modulation of DNA-binding domains based CRISPR Cas9 enzyme and also Transcriptional Activator Like Effectors
- Ran *et al.* (2013-A) described an approach that combined a Cas9 nickase mutant with paired guide RNAs to introduce targeted double-strand breaks. This addresses the issue of the Cas9 nuclease from the microbial CRISPR-Cas system being targeted to specific genomic loci by a guide sequence, which can tolerate certain mismatches to the DNA target and thereby promote undesired off-target mutagenesis. Because individual nicks in the genome are repaired with high fidelity, simultaneous nicking *via* appropriately offset guide RNAs is required for double-stranded breaks and extends the number of specifically recognized bases for target cleavage. The authors demonstrated that using paired nicking can reduce off-target activity by 50- to 1,500-fold in cell lines and to facilitate gene knockout in mouse zygotes without sacrificing on-target cleavage efficiency. This versatile strategy enables a wide variety of genome editing applications that require high specificity.
- Hsu *et al.* (2013) characterized SpCas9 targeting specificity in human cells to inform the selection of target sites and avoid off-target effects. The study evaluated >700 guide RNA variants and SpCas9-induced indel mutation levels at >100 predicted genomic off-target loci in 293T and 293FT cells. The authors that SpCas9 tolerates mismatches between guide RNA and target DNA at different positions in a sequence-dependent manner, sensitive to the number, position and distribution of mismatches. The authors

further showed that SpCas9-mediated cleavage is unaffected by DNA methylation and that the dosage of SpCas9 and guide RNA can be titrated to minimize off-target modification. Additionally, to facilitate mammalian genome engineering applications, the authors reported providing a web-based software tool to guide the selection and validation of target sequences as well as off-target analyses.

- Ran *et al.* (2013-B) described a set of tools for Cas9-mediated genome editing *via* non-homologous end joining (NHEJ) or homology-directed repair (HDR) in mammalian cells, as well as generation of modified cell lines for downstream functional studies. To minimize off-target cleavage, the authors further described a double-nicking strategy using the Cas9 nickase mutant with paired guide RNAs. The protocol provided by the authors experimentally derived guidelines for the selection of target sites, evaluation of cleavage efficiency and analysis of off-target activity. The studies showed that beginning with target design, gene modifications can be achieved within as little as 1–2 weeks, and modified clonal cell lines can be derived within 2–3 weeks.
- Shalem *et al.* described a new way to interrogate gene function on a genome-wide scale. Their studies showed that delivery of a genome-scale CRISPR-Cas9 knockout (GeCKO) library targeted 18,080 genes with 64,751 unique guide sequences enabled both negative and positive selection screening in human cells. First, the authors showed use of the GeCKO library to identify genes essential for cell viability in cancer and pluripotent stem cells. Next, in a melanoma model, the authors screened for genes whose loss is involved in resistance to vemurafenib, a therapeutic that inhibits mutant protein kinase BRAF. Their studies showed that the highest-ranking candidates included previously validated genes NF1 and MED12 as well as novel hits NF2, CUL3, TADA2B, and TADA1. The authors observed a high level of consistency between independent guide RNAs targeting the same gene and a high rate of hit confirmation, and thus demonstrated the promise of genome-scale screening with Cas9.
- Nishimasu *et al.* reported the crystal structure of *Streptococcus pyogenes* Cas9 in complex with sgRNA and its target DNA at 2.5 A° resolution. The structure revealed a bilobed architecture composed of target recognition and nuclease lobes, accommodating the sgRNA:DNAn RNA duplex in a positively charged groove at their interface. Whereas the recognition lobe is essential for binding sgRNA and DNA, the nuclease lobe contains the HNH and RuvC nuclease domains, which are properly positioned for cleavage of the complementary and non-complementary strands of the target DNA, respectively. The nuclease lobe also contains a carboxyl-terminal domain

responsible for the interaction with the protospacer adjacent motif (PAM). This high-resolution structure and accompanying functional analyses have revealed the molecular mechanism of RNA-guided DNA targeting by Cas9, thus paving the way for the rational design of new, versatile genome-editing technologies.

- Wu *et al.* mapped genome-wide binding sites of a catalytically inactive Cas9 (dCas9) from *Streptococcus pyogenes* loaded with single guide RNAs (sgRNAs) in mouse embryonic stem cells (mESCs). The authors showed that each of the four sgRNAs tested targets dCas9 to between tens and thousands of genomic sites, frequently characterized by a 5-nucleotide seed region in the sgRNA and an NGG protospacer adjacent motif (PAM). Chromatin inaccessibility decreases dCas9 binding to other sites with matching seed sequences; thus 70% of off-target sites are associated with genes. The authors showed that targeted sequencing of 295 dCas9 binding sites in mESCs transfected with catalytically active Cas9 identified only one site mutated above background levels. The authors proposed a two-state model for Cas9 binding and cleavage, in which a seed match triggers binding but extensive pairing with target DNA is required for cleavage.
- Platt *et al.* established a Cre-dependent Cas9 knockin mouse. The authors demonstrated *in vivo* as well as *ex vivo* genome editing using adeno-associated virus (AAV)-, lentivirus-, or particle-mediated delivery of guide RNA in neurons, immune cells, and endothelial cells.
- Hsu *et al.* (2014) is a review article that discusses generally CRISPR-Cas9 history from yogurt to genome editing, including genetic screening of cells.
- Wang *et al.* (2014) relates to a pooled, loss-of-function genetic screening approach suitable for both positive and negative selection that uses a genome-scale lentiviral single guide RNA (sgRNA) library.
- Doench *et al.* created a pool of sgRNAs, tiling across all possible target sites of a panel of six endogenous mouse and three endogenous human genes and quantitatively assessed their ability to produce null alleles of their target gene by antibody staining and flow cytometry. The authors showed that optimization of the PAM improved activity and also provided an on-line tool for designing sgRNAs.
- Swiech *et al.* demonstrate that AAV-mediated SpCas9 genome editing can enable reverse genetic studies of gene function in the brain.

- Konermann *et al.* (2015) discusses the ability to attach multiple effector domains, e.g., transcriptional activator, functional and epigenomic regulators at appropriate positions on the guide such as stem or tetraloop with and without linkers.
- Zetsche *et al.* demonstrates that the Cas9 enzyme can be split into two and hence the assembly of Cas9 for activation can be controlled.
- Chen *et al.* relates to multiplex screening by demonstrating that a genome-wide *in vivo* CRISPR-Cas9 screen in mice reveals genes regulating lung metastasis.
- Ran *et al.* (2015) relates to SaCas9 and its ability to edit genomes and demonstrates that one cannot extrapolate from biochemical assays.
- Shalem *et al.* (2015) described ways in which catalytically inactive Cas9 (dCas9) fusions are used to synthetically repress (CRISPRi) or activate (CRISPRa) expression, showing. advances using Cas9 for genome-scale screens, including arrayed and pooled screens, knockout approaches that inactivate genomic loci and strategies that modulate transcriptional activity.
- Xu *et al.* (2015) assessed the DNA sequence features that contribute to single guide RNA (sgRNA) efficiency in CRISPR-based screens. The authors explored efficiency of CRISPR-Cas9 knockout and nucleotide preference at the cleavage site. The authors also found that the sequence preference for CRISPRi/a is substantially different from that for CRISPR-Cas9 knockout.
- Parnas *et al.* (2015) introduced genome-wide pooled CRISPR-Cas9 libraries into dendritic cells (DCs) to identify genes that control the induction of tumor necrosis factor (Tnf) by bacterial lipopolysaccharide (LPS). Known regulators of Tlr4 signaling and previously unknown candidates were identified and classified into three functional modules with distinct effects on the canonical responses to LPS.
- Ramanan *et al* (2015) demonstrated cleavage of viral episomal DNA (cccDNA) in infected cells. The HBV genome exists in the nuclei of infected hepatocytes as a 3.2kb double-stranded episomal DNA species called covalently closed circular DNA (cccDNA), which is a key component in the HBV life cycle whose replication is not inhibited by current therapies. The authors showed that sgRNAs specifically targeting highly conserved regions of HBV robustly suppresses viral replication and depleted cccDNA.
- Nishimasu *et al.* (2015) reported the crystal structures of SaCas9 in complex with a single guide RNA (sgRNA) and its double-stranded DNA targets, containing the 5'-TTGAAT-3' PAM and the 5'-TTGGGT-3' PAM. A structural comparison of SaCas9

with SpCas9 highlighted both structural conservation and divergence, explaining their distinct PAM specificities and orthologous sgRNA recognition.

- Canver *et al.* (2015) demonstrated a CRISPR-Cas9-based functional investigation of non-coding genomic elements. The authors we developed pooled CRISPR-Cas9 guide RNA libraries to perform *in situ* saturating mutagenesis of the human and mouse BCL11A enhancers which revealed critical features of the enhancers.
- Zetsche et al. (2015) reported characterization of Cas13, a class 2 CRISPR nuclease from Francisella novicida U112 having features distinct from Cas9. Cas13 is a single RNA-guided endonuclease lacking tracrRNA, utilizes a T-rich protospacer-adjacent motif, and cleaves DNA via a staggered DNA double-stranded break.
- Shmakov et al. (2015) reported three distinct Class 2 CRISPR-Cas systems. Two system CRISPR enzymes (C2c1 and C2c3) contain RuvC-like endonuclease domains distantly related to Cas13. Unlike Cas13, C2c1 depends on both crRNA and tracrRNA for DNA cleavage. The third enzyme (C2c2) contains two predicted HEPN RNase domains and is tracrRNA independent.
- Slaymaker et al (2016) reported the use of structure-guided protein engineering to improve the specificity of Streptococcus pyogenes Cas9 (SpCas9). The authors developed "enhanced specificity" SpCas9 (eSpCas9) variants which maintained robust on-target cleavage with reduced off-target effects.

[0623] The methods and tools provided herein are exemplified for Cas13, a type II nuclease that does not make use of tracrRNA. Orthologs of Cas13 have been identified in different bacterial species as described herein. Further type II nucleases with similar properties can be identified using methods described in the art (Shmakov et al. 2015, 60:385–397; Abudayeh et al. 2016, Science, 5;353(6299)) . In particular embodiments, such methods for identifying novel CRISPR effector proteins may comprise the steps of selecting sequences from the database encoding a seed which identifies the presence of a CRISPR Cas locus, identifying loci located within 10 kb of the seed comprising Open Reading Frames (ORFs) in the selected sequences, selecting therefrom loci comprising ORFs of which only a single ORF encodes a novel CRISPR effector having greater than 700 amino acids and no more than 90% homology to a known CRISPR effector. In particular embodiments, the seed is a protein that is common to the CRISPR-Cas system, such as Cas1. In further embodiments, the CRISPR array is used as a seed to identify new effector proteins.

[0624] The effectiveness of the present invention has been demonstrated. Preassembled recombinant CRISPR-Cas13 complexes comprising Cas13 and crRNA may be transfected, for

example by electroporation, resulting in high mutation rates and absence of detectable off-target mutations. Hur, J.K. et al, Targeted mutagenesis in mice by electroporation of Cas13 ribonucleoproteins, Nat Biotechnol. 2016 Jun 6. doi: 10.1038/nbt.3596. Genome-wide analyses shows that Cas13 is highly specific. By one measure, in vitro cleavage sites determined for Cas13 in human HEK293T cells were significantly fewer that for SpCas9. Kim, D. et al., Genome-wide analysis reveals specificities of Cas13 endonucleases in human cells, Nat Biotechnol. 2016 Jun 6. doi: 10.1038/nbt.3609. An efficient multiplexed system employing Cas13 has been demonstrated in Drosophila employing gRNAs processed from an array containing inventing tRNAs. Port, F. et al, Expansion of the CRISPR toolbox in an animal with tRNA-flanked Cas9 and Cas13 gRNAs. doi: http://dx.doi.org/10.1101/046417.

[0625] Also, "Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing", Shengdar Q. Tsai, Nicolas Wyvekens, Cyd Khayter, Jennifer A. Foden, Vishal Thapar, Deepak Reyon, Mathew J. Goodwin, Martin J. Aryee, J. Keith Joung Nature Biotechnology 32(6): 569-77 (2014), relates to dimeric RNA-guided FokI Nucleases that recognize extended sequences and can edit endogenous genes with high efficiencies in human cells.

[0626] With respect to general information on CRISPR/Cas Systems, components thereof, and delivery of such components, including methods, materials, delivery vehicles, vectors, particles, and making and using thereof, including as to amounts and formulations, as well as CRISPR-Cas-expressing eukaryotic cells, CRISPR-Cas expressing eukaryotes, such as a mouse, reference is made to: US Patents Nos. 8,999,641, 8,993,233, 8,697,359, 8,771,945, 8,795,965, 8,865,406, 8,871,445, 8,889,356, 8,889,418, 8,895,308, 8,906,616, 8,932,814, and 8,945,839; US Patent Publications US 2014-0310830 (US App. Ser. No. 14/105,031), US 2014-0287938 A1 (U.S. App. Ser. No. 14/213,991), US 2014-0273234 A1 (U.S. App. Ser. No. 14/293,674), US2014-0273232 A1 (U.S. App. Ser. No. 14/290,575), US 2014-0273231 (U.S. App. Ser. No. 14/259,420), US 2014-0256046 A1 (U.S. App. Ser. No. 14/226,274), US 2014-0248702 A1 (U.S. App. Ser. No. 14/258,458), US 2014-0242700 A1 (U.S. App. Ser. No. 14/222,930), US 2014-0242699 A1 (U.S. App. Ser. No. 14/183,512), US 2014-0242664 A1 (U.S. App. Ser. No. 14/104,990), US 2014-0234972 A1 (U.S. App. Ser. No. 14/183,471), US 2014-0227787 A1 (U.S. App. Ser. No. 14/256,912), US 2014-0189896 A1 (U.S. App. Ser. No. 14/105,035), US 2014-0186958 (U.S. App. Ser. No. 14/105,017), US 2014-0186919 A1 (U.S. App. Ser. No. 14/104,977), US 2014-0186843 A1 (U.S. App. Ser. No. 14/104,900), US 2014-0179770 A1 (U.S. App. Ser. No. 14/104,837) and US 2014-0179006 A1 (U.S. App. Ser. No. 14/183,486), US 2014-0170753 (US App Ser No 14/183,429); US 2015-0184139 (U.S. App.

Ser. No. 14/324,960); 14/054,414 European Patent Applications EP 2 771 468 (EP13818570.7), EP 2 764 103 (EP13824232.6), and EP 2 784 162 (EP14170383.5); and PCT Patent Publications WO2014/093661 (PCT/US2013/074743), WO2014/093694 (PCT/US2013/074790), WO2014/093595 (PCT/US2013/074611), WO2014/093718 (PCT/US2013/074825), WO2014/093709 (PCT/US2013/074812), WO2014/093622 (PCT/US2013/074667), WO2014/093635 (PCT/US2013/074691), WO2014/093655 (PCT/US2013/074736), WO2014/093712 (PCT/US2013/074819), WO2014/093701 (PCT/US2013/074800), WO2014/018423 (PCT/US2013/051418), WO2014/204723 (PCT/US2014/041790), WO2014/204724 (PCT/US2014/041800), WO2014/204725 (PCT/US2014/041803), WO2014/204726 (PCT/US2014/041804), WO2014/204727 (PCT/US2014/041806), WO2014/204728 (PCT/US2014/041808), WO2014/204729 (PCT/US2014/041809), WO2015/089351 (PCT/US2014/069897), WO2015/089354 (PCT/US2014/069902), WO2015/089364 (PCT/US2014/069925), WO2015/089427 (PCT/US2014/070068), WO2015/089462 (PCT/US2014/070127), WO2015/089419 (PCT/US2014/070057), WO2015/089465 (PCT/US2014/070135), WO2015/089486 (PCT/US2014/070175), WO2015/058052 (PCT/US2014/061077), WO2015/070083 (PCT/US2014/064663), WO2015/089354 (PCT/US2014/069902), WO2015/089351 (PCT/US2014/069897), WO2015/089364 (PCT/US2014/069925), WO2015/089427 (PCT/US2014/070068), WO2015/089473 (PCT/US2014/070152), WO2015/089486 (PCT/US2014/070175), WO2016/049258 (PCT/US2015/051830), WO2016/094867 (PCT/US2015/065385), WO2016/094872 (PCT/US2015/065393), WO2016/094874 (PCT/US2015/065396), WO2016/106244 (PCT/US2015/067177).

[0627] Mention is also made of US application 62/180,709, 17-Jun-15, PROTECTED GUIDE RNAS (PGRNAS); US application 62/091,455, filed, 12-Dec-14, PROTECTED GUIDE RNAS (PGRNAS); US application 62/096,708, 24-Dec-14, PROTECTED GUIDE RNAS (PGRNAS); US applications 62/091,462, 12-Dec-14, 62/096,324, 23-Dec-14, 62/180,681, 17-Jun-2015, and 62/237,496, 5-Oct-2015, DEAD GUIDES FOR CRISPR TRANSCRIPTION FACTORS; US application 62/091,456, 12-Dec-14 and 62/180,692, 17-Jun-2015, ESCORTED AND FUNCTIONALIZED GUIDES FOR CRISPR-CAS SYSTEMS; US application 62/091,461, 12-Dec-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR GENOME EDITING AS TO HEMATOPOETIC STEM CELLS (HSCs); US application 62/094,903, 19-Dec-14, UNBIASED IDENTIFICATION OF DOUBLE-STRAND BREAKS AND GENOMIC REARRANGEMENT BY GENOME-WISE INSERT CAPTURE

SEQUENCING; US application 62/096,761, 24-Dec-14, ENGINEERING OF SYSTEMS, METHODS AND OPTIMIZED ENZYME AND GUIDE SCAFFOLDS FOR SEQUENCE MANIPULATION; US application 62/098,059, 30-Dec-14, 62/181,641, 18-Jun-2015, and 62/181,667, 18-Jun-2015, RNA-TARGETING SYSTEM; US application 62/096,656, 24-Dec-14 and 62/181,151, 17-Jun-2015, CRISPR HAVING OR ASSOCIATED WITH DESTABILIZATION DOMAINS; US application 62/096,697, 24-Dec-14, CRISPR HAVING OR ASSOCIATED WITH AAV; US application 62/098,158, 30-Dec-14, ENGINEERED CRISPR COMPLEX INSERTIONAL TARGETING SYSTEMS; US application 62/151,052, 22-Apr-15, CELLULAR TARGETING FOR EXTRACELLULAR EXOSOMAL REPORTING; US application 62/054,490, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING PARTICLE DELIVERY COMPONENTS; US application 61/939,154, 12-F EB-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/055,484, 25-Sep-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/087,537, 4-Dec-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/054,651, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO; US application 62/067,886, 23-Oct-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO; US applications 62/054,675, 24-Sep-14 and 62/181,002, 17-Jun-2015, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN NEURONAL CELLS/TISSUES; US application 62/054,528, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN IMMUNE DISEASES OR DISORDERS; US application 62/055,454, 25-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING CELL PENETRATION PEPTIDES (CPP); US application 62/055,460, 25-Sep-14, MULTIFUNCTIONAL-CRISPR

COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; US application 62/087,475, 4-Dec-14 and 62/181,690, 18-Jun-2015, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/055,487, 25-Sep-14, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/087,546, 4-Dec-14 and 62/181,687, 18-Jun-2015, MULTIFUNCTIONAL CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; and US application 62/098,285, 30-Dec-14, CRISPR MEDIATED IN VIVO MODELING AND GENETIC SCREENING OF TUMOR GROWTH AND METASTASIS.

[0628] Mention is made of US applications 62/181,659, 18-Jun-2015 and 62/207,318, 19-Aug-2015, ENGINEERING AND OPTIMIZATION OF SYSTEMS, METHODS, ENZYME AND GUIDE SCAFFOLDS OF CAS9 ORTHOLOGS AND VARIANTS FOR SEQUENCE MANIPULATION. Mention is made of US applications 62/181,663, 18-Jun-2015 and 62/245,264, 22-Oct-2015, NOVEL CRISPR ENZYMES AND SYSTEMS, US applications 62/181,675, 18-Jun-2015, 62/285,349, 22-Oct-2015, 62/296,522, 17-Feb-2016, and 62/320,231, 8-Apr-2016, NOVEL CRISPR ENZYMES AND SYSTEMS, US application 62/232,067, 24-Sep-2015, US Application 14/975,085, 18-Dec-2015, European application No. 16150428.7, US application 62/205,733, 16-Aug-2015, US application 62/201,542, 5-Aug-2015, US application 62/193,507, 16-Jul-2015, and US application 62/181,739, 18-Jun-2015, each entitled NOVEL CRISPR ENZYMES AND SYSTEMS and of US application 62/245,270, 22-Oct-2015, NOVEL CRISPR ENZYMES AND SYSTEMS. Mention is also made of US application 61/939,256, 12-Feb-2014, and WO 2015/089473 (PCT/US2014/070152), 12-Dec-2014, each entitled ENGINEERING OF SYSTEMS, METHODS AND OPTIMIZED GUIDE COMPOSITIONS WITH NEW ARCHITECTURES FOR SEQUENCE MANIPULATION. Mention is also made of PCT/US2015/045504, 15-Aug-2015, US application 62/180,699, 17-Jun-2015, and US application 62/038,358, 17-Aug-2014, each entitled GENOME EDITING USING CAS9 NICKASES.

[0629] Each of these patents, patent publications, and applications, and all documents cited therein or during their prosecution ("appln cited documents") and all documents cited or referenced in the appln cited documents, together with any instructions, descriptions, product specifications, and product sheets for any products mentioned therein or in any document therein and incorporated by reference herein, are hereby incorporated herein by reference, and may be employed in the practice of the invention. All documents (e.g., these patents, patent publications and applications and the appln cited documents) are incorporated herein by

reference to the same extent as if each individual document was specifically and individually indicated to be incorporated by reference.

Type-V CRISPR-Cas Protein

[0630] The application describes methods using Type-V CRISPR-Cas proteins. This is exemplified herein with Cas13, whereby a number of orthologs or homologs have been identified. It will be apparent to the skilled person that further orthologs or homologs can be identified and that any of the functionalities described herein may be engineered into other orthologs, incuding chimeric enzymes comprising fragments from multiple orthologs.

[0631] Computational methods of identifying novel CRISPR-Cas loci are described in EP3009511 or US2016208243 and may comprise the following steps: detecting all contigs encoding the Cas1 protein; identifying all predicted protein coding genes within 20kB of the cas1 gene; comparing the identified genes with Cas protein-specific profiles and predicting CRISPR arrays; selecting unclassified candidate CRISPR-Cas loci containing proteins larger than 500 amino acids (>500 aa); analyzing selected candidates using methods such as PSI-BLAST and HHPred to screen for known protein domains, thereby identifying novel Class 2 CRISPR-Cas loci (see also Schmakov et al. 2015, Mol Cell. 60(3):385-97). In addition to the above mentioned steps, additional analysis of the candidates may be conducted by searching metagenomics databases for additional homologs. Additionally or alternatively, to expand the search to non-autonomous CRISPR-Cas systems, the same procedure can be performed with the CRISPR array used as the seed.

[0632] In one aspect the detecting all contigs encoding the Cas1 protein is performed by GenemarkS which a gene prediction program as further described in "GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions." John Besemer, Alexandre Lomsadze and Mark Borodovsky, Nucleic Acids Research (2001) 29, pp 2607-2618, herein incorporated by reference.

[0633] In one aspect the identifying all predicted protein coding genes is carried out by comparing the identified genes with Cas protein-specific profiles and annotating them according to NCBI Conserved Domain Database (CDD) which is a protein annotation resource that consists of a collection of well-annotated multiple sequence alignment models for ancient domains and full-length proteins. These are available as position-specific score matrices (PSSMs) for fast identification of conserved domains in protein sequences via RPS-BLAST. CDD content includes NCBI-curated domains, which use 3D-structure information to explicitly define domain boundaries and provide insights into sequence/structure/function

relationships, as well as domain models imported from a number of external source databases (Pfam, SMART, COG, PRK, TIGRFAM). In a further aspect, CRISPR arrays were predicted using a PILER-CR program which is a public domain software for finding CRISPR repeats as described in "PILER-CR: fast and accurate identification of CRISPR repeats", Edgar, R.C., BMC Bioinformatics, Jan 20;8:18(2007), herein incorporated by reference.

[0634] In a further aspect, the case by case analysis is performed using PSI-BLAST (Position-Specific Iterative Basic Local Alignment Search Tool). PSI-BLAST derives a position-specific scoring matrix (PSSM) or profile from the multiple sequence alignment of sequences detected above a given score threshold using protein–protein BLAST. This PSSM is used to further search the database for new matches, and is updated for subsequent iterations with these newly detected sequences. Thus, PSI-BLAST provides a means of detecting distant relationships between proteins.

[0635] In another aspect, the case by case analysis is performed using HHpred, a method for sequence database searching and structure prediction that is as easy to use as BLAST or PSI-BLAST and that is at the same time much more sensitive in finding remote homologs. In fact, HHpred's sensitivity is competitive with the most powerful servers for structure prediction currently available. HHpred is the first server that is based on the pairwise comparison of profile hidden Markov models (HMMs). Whereas most conventional sequence search methods search sequence databases such as UniProt or the NR, HHpred searches alignment databases, like Pfam or SMART. This greatly simplifies the list of hits to a number of sequence families instead of a clutter of single sequences. All major publicly available profile and alignment databases are available through HHpred. HHpred accepts a single query sequence or a multiple alignment as input. Within only a few minutes it returns the search results in an easy-to-read format similar to that of PSI-BLAST. Search options include local or global alignment and scoring secondary structure similarity. HHpred can produce pairwise query-template sequence alignments, merged query-template multiple alignments (e.g. for transitive searches), as well as 3D structural models calculated by the MODELLER software from HHpred alignments.

Orthologs of Cas13

[0636] The terms "orthologue" (also referred to as "ortholog" herein) and "homologue" (also referred to as "homolog" herein) are well known in the art. By means of further guidance, a "homologue" of a protein as used herein is a protein of the same species which performs the same or a similar function as the protein it is a homologue of. Homologous proteins may but need not be structurally related, or are only partially structurally related. An "orthologue" of a protein as used herein is a protein of a different species which performs the same or a similar

function as the protein it is an orthologue of. Orthologous proteins may but need not be structurally related, or are only partially structurally related. Homologs and orthologs may be identified by homology modelling (see, e.g., Greer, Science vol. 228 (1985) 1055, and Blundell et al. Eur J Biochem vol 172 (1988), 513) or "structural BLAST" (Dey F, Cliff Zhang Q, Petrey D, Honig B. Toward a "structural BLAST": using structural relationships to infer function. Protein Sci. 2013 Apr;22(4):359-66. doi: 10.1002/pro.2225.). See also Shmakov et al. (2015) for application in the field of CRISPR-Cas loci. Homologous proteins may but need not be structurally related, or are only partially structurally related.

[0637] The Cas13 gene is found in several diverse bacterial genomes, typically in the same locus with cas1, cas2, and cas4 genes and a CRISPR cassette (for example, FNFX1_1431-FNFX1_1428 of Francisella cf. novicida Fx1). Thus, the layout of this putative novel CRISPR-Cas system appears to be similar to that of type II-B. Furthermore, similar to Cas9, the Cas13 protein contains a readily identifiable C-terminal region that is homologous to the transposon ORF-B and includes an active RuvC-like nuclease, an arginine-rich region, and a Zn finger (absent in Cas9). However, unlike Cas9, Cas13 is also present in several genomes without a CRISPR-Cas context and its relatively high similarity with ORF-B suggests that it might be a transposon component. It was suggested that if this was a genuine CRISPR-Cas system and Cas13 is a functional analog of Cas9 it would be a novel CRISPR-Cas type, namely type V (See Annotation and Classification of CRISPR-Cas Systems. Makarova KS, Koonin EV. Methods Mol Biol. 2015;1311:47-75). However, as described herein, Cas13 is denoted to be in subtype V-A to distinguish it from C2c1p which does not have an identical domain structure and is hence denoted to be in subtype V-B.

[0638] The present invention encompasses the use of a Cas13 effector protein, derived from a Cas13 locus denoted as subtype V-A. Herein such effector proteins are also referred to as "Cas13p", e.g., a Cas13 protein (and such effector protein or Cas13 protein or protein derived from a Cas13 locus is also called "CRISPR-Cas protein").

[0639] In particular embodiments, the effector protein is a Cas13 effector protein from an organism from a genus comprising Streptococcus, Campylobacter, Nitratifractor, Staphylococcus, Parvibaculum, Roseburia, Neisseria, Gluconacetobacter, Azospirillum, Sphaerochaeta, Lactobacillus, Eubacterium, Corynebacter, Carnobacterium, Rhodobacter, Listeria, Paludibacter, Clostridium, Lachnospiraceae, Clostridiaridium, Leptotrichia, Francisella, Legionella, Alicyclobacillus, Methanomethyophilus, Porphyromonas, Prevotella, Bacteroidetes, Helcococcus, Leptospira, Desulfovibrio, Desulfonatronum, Opitutaceae, Tuberibacillus, Bacillus, Brevibacilus, Methylobacterium, Butyvibrio, Perigrinibacterium,

Pareubacterium, Moraxella, Thiomicrospira or Acidaminococcus. In particular embodiments, the Cas13 effector protein is selected from an organism from a genus selected from Eubacterium, Lachnospiraceae, Leptotrichia, Francisella, Methanomethyophilus, Porphyromonas, Prevotella, Leptospira, Butyvibrio, Perigrinibacterium, Pareubacterium, Moraxella, Thiomicrospira or Acidaminococcus

[0640] In further particular embodiments, the Cas13 effector protein is from an organism selected from S. mutans, S. agalactiae, S. equisimilis, S. sanguinis, S. pneumonia; C. jejuni, C. coli; N. salsuginis, N. tergarcus; S. auricularis, S. carnosus; N. meningitides, N. gonorrhoeae; L. monocytogenes, L. ivanovii; C. botulinum, C. difficile, C. tetani, C. sordellii, L inadai, F. tularensis 1, P. albensis, L. bacterium, B. proteoclasticus, P. bacterium, P. crevioricanis, P. disiens and P. macacae .

[0641] The effector protein may comprise a chimeric effector protein comprising a first fragment from a first effector protein (e.g., a Cas13) ortholog and a second fragment from a second effector (e.g., a Cas13) protein ortholog, and wherein the first and second effector protein orthologs are different. At least one of the first and second effector protein (e.g., a Cas13) orthologs may comprise an effector protein (e.g., a Cas13) from an organism comprising Streptococcus, Campylobacter, Nitratifractor, Staphylococcus, Parvibaculum, Roseburia, Neisseria, Gluconacetobacter, Azospirillum, Sphaerochaeta, Lactobacillus, Eubacterium, Corynebacter, Carnobacterium, Rhodobacter, Listeria, Paludibacter, Clostridium, Lachnospiraceae, Clostridiaridium, Leptotrichia, Francisella, Legionella, Alicyclobacillus, Methanomethyophilus, Porphyromonas, Prevotella, Bacteroidetes, Helcococcus, Letospira, Desulfovibrio, Desulfonatronum, Opitutaceae, Tuberibacillus, Bacillus, Brevibacilus, Methylobacterium, Butyvibrio, Perigrinibacterium, Pareubacterium, Moraxella, Thiomicrospira or Acidaminococcus; e.g., a chimeric effector protein comprising a first fragment and a second fragment wherein each of the first and second fragments is selected from a Cas13 of an organism comprising Streptococcus, Campylobacter, Nitratifractor, Staphylococcus, Parvibaculum, Roseburia, Neisseria, Gluconacetobacter, Azospirillum, Sphaerochaeta, Lactobacillus, Eubacterium, Corynebacter, Carnobacterium, Rhodobacter, Listeria, Paludibacter, Clostridium, Lachnospiraceae, Clostridiaridium, Leptotrichia, Francisella, Legionella, Alicyclobacillus, Methanomethyophilus, Porphyromonas, Prevotella, Bacteroidetes, Helcococcus, Letospira, Desulfovibrio, Desulfonatronum, Opitutaceae, Tuberibacillus, Bacillus, Brevibacilus, Methylobacterium, Butyvibrio, Perigrinibacterium, Pareubacterium, Moraxella, Thiomicrospira or Acidaminococcus wherein the first and second fragments are not from the same bacteria; for instance a chimeric effector protein comprising

a first fragment and a second fragment wherein each of the first and second fragments is selected from a Cas13 of S. mutans, S. agalactiae, S. equisimilis, S. sanguinis, S. pneumonia; C. jejuni, C. coli; N. salsuginis, N. tergarcus; S. auricularis, S. carnosus; N. meningitides, N. gonorrhoeae; L. monocytogenes, L. ivanovii; C. botulinum, C. difficile, C. tetani, C. sordellii; Francisella tularensis 1, Prevotella albensis, Lachnospiraceae bacterium MC2017 1, Butyrivibrio proteoclasticus, Peregrinibacteria bacterium GW2011_GWA2_33_10, Parcubacteria bacterium GW2011_GWC2_44_17, Smithella sp. SCADC, Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020, Candidatus Methanoplasma termitum, Eubacterium eligens, Moraxella bovoculi 237, Leptospira inadai, Lachnospiraceae bacterium ND2006, Porphyromonas crevioricanis 3, Prevotella disiens and Porphyromonas macacae, wherein the first and second fragments are not from the same bacteria.

[0642] In a more preferred embodiment, the Cas13p is derived from a bacterial species selected from Francisella tularensis 1, Prevotella albensis, Lachnospiraceae bacterium MC2017 1, Butyrivibrio proteoclasticus, Peregrinibacteria bacterium GW2011_GWA2_33_10, Parcubacteria bacterium GW2011_GWC2_44_17, Smithella sp. SCADC, Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020, Candidatus Methanoplasma termitum, Eubacterium eligens, Moraxella bovoculi 237, Moraxella bovoculi AAX08_00205, Moraxella bovoculi AAX11_00205, Butyrivibrio sp. NC3005, Thiomicrospira sp. XS5, Leptospira inadai, Lachnospiraceae bacterium ND2006, Porphyromonas crevioricanis 3, Prevotella disiens and Porphyromonas macacae. In certain embodiments, the Cas13p is derived from a bacterial species selected from Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020.In certain embodiments, the effector protein is derived from a subspecies of Francisella tularensis 1, including but not limited to Francisella tularensis subsp. Novicida. In certain preferred embodiments, the Cas13p is derived from a bacterial species selected from Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium ND2006, Lachnospiraceae bacterium MA2020, Moraxella bovoculi AAX08_00205, Moraxella bovoculi AAX11_00205, Butyrivibrio sp. NC3005, or Thiomicrospira sp. XS5.

[0643] In particular embodiments, the homologue or orthologue of Cas13 as referred to herein has a sequence homology or identity of at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with the example Cas13 proteins disclosed herein. In further embodiments, the homologue or orthologue of Cas13 as referred to herein has a sequence identity of at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with the wild type Cas13. Where the Cas13 has one or more mutations (mutated), the homologue or orthologue of said Cas13 as

referred to herein has a sequence identity of at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with the mutated Cas13.

[0644] In an ambodiment, the Cas13 protein may be an ortholog of an organism of a genus which includes, but is not limited to Acidaminococcus sp, Lachnospiraceae bacterium or Moraxella bovoculi; in particular embodiments, the type V Cas protein may be an ortholog of an organism of a species which includes, but is not limited to Acidaminococcus sp. BV3L6; Lachnospiraceae bacterium ND2006 (LbCas13) or Moraxella bovoculi 237.In particular embodiments, the homologue or orthologue of Cas13 as referred to herein has a sequence homology or identity of at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with one or more of the Cas13 sequences disclosed herein. In further embodiments, the homologue or orthologue of Cas13 as referred to herein has a sequence identity of at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with the wild type FnCas13, AsCas13 or LbCas13.

[0645] In particular embodiments, the Cas13 protein of the invention has a sequence homology or identity of at least 60%, more particularly at least 70, such as at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with FnCas13, AsCas13 or LbCas13. In further embodiments, the Cas13 protein as referred to herein has a sequence identity of at least 60%, such as at least 70%, more particularly at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with the wild type AsCas13 or LbCas13. In particular embodiments, the Cas13 protein of the present invention has less than 60% sequence identity with FnCas13. The skilled person will understand that this includes truncated forms of the Cas13 protein whereby the sequence identity is determined over the length of the truncated form. In particular embodiments, the Cas13 enzyme is not FnCas13.

Modified Cas13 enzymes

[0646] In particular embodiments, it is of interest to make use of an engineered Cas13 protein as defined herein, such as Cas13, wherein the protein complexes with a nucleic acid molecule comprising RNA to form a CRISPR complex, wherein when in the CRISPR complex, the nucleic acid molecule targets one or more target polynucleotide loci, the protein comprises at least one modification compared to unmodified Cas13 protein, and wherein the CRISPR complex comprising the modified protein has altered activity as compared to the complex comprising the unmodified Cas13 protein. It is to be understood that when referring herein to CRISPR “protein”, the Cas13 protein preferably is a modified CRISPR-Cas protein (e.g. having increased or decreased (or no) enzymatic activity, such as without limitation including

Cas13. The term "CRISPR protein" may be used interchangeably with "CRISPR-Cas protein", irrespective of whether the CRISPR protein has altered, such as increased or decreased (or no) enzymatic activity, compared to the wild type CRISPR protein.

[0647] Computational analysis of the primary structure of Cas13 nucleases reveals three distinct regions. First a C-terminal RuvC like domain, which is the only functional characterized domain. Second a N-terminal alpha-helical region and thirst a mixed alpha and beta region, located between the RuvC like domain and the alpha-helical region.

[0648] Several small stretches of unstructured regions are predicted within the Cas13 primary structure. Unstructured regions, which are exposed to the solvent and not conserved within different Cas13 orthologs, are preferred sides for splits and insertions of small protein sequences . In addition, these sides can be used to generate chimeric proteins between Cas13 orthologs.

[0649] Based on the above information, mutants can be generated which lead to inactivation of the enzyme or which modify the double strand nuclease to nickase activity. In alternative embodiments, this information is used to develop enzymes with reduced off-target effects (described elsewhere herein)

[0650] In certain of the above-described Cas13 enzymes, the enzyme is modified by mutation of one or more residues (in the RuvC domain) including but not limited to positions R909, R912, R930, R947, K949, R951, R955, K965, K968, K1000, K1002, R1003, K1009, K1017, K1022, K1029, K1035, K1054, K1072, K1086, R1094, K1095, K1109, K1118, K1142, K1150, K1158, K1159, R1220, R1226, R1242, and/or R1252 with reference to amino acid position numbering of AsCas13 (Acidaminococcus sp. BV3L6). In certain embodiments, the Cas13 enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[0651] In certain of the above-described non-naturally-occurring CRISPR-Cas proteins, the enzyme is modified by mutation of one or more residues (in the RAD50) domain including but not limited positions K324, K335, K337, R331, K369, K370, R386, R392, R393, K400, K404, K406, K408, K414, K429, K436, K438, K459, K460, K464, R670, K675, R681, K686, K689, R699, K705, R725, K729, K739, K748, and/or K752 with reference to amino acid position numbering of AsCas13 (Acidaminococcus sp. BV3L6). In certain embodiments, the Cas13 enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[0652] In certain of the Cas13 enzymes, the enzyme is modified by mutation of one or more residues including but not limited positions R912, T923, R947, K949, R951, R955, K965,

K968, K1000, R1003, K1009, K1017, K1022, K1029, K1072, K1086, F1103, R1226, and/or R1252 with reference to amino acid position numbering of AsCas13 (Acidaminococcus sp. BV3L6). In certain embodiments, the Cas13 enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[0653] In certain embodiments, the Cas13 enzyme is modified by mutation of one or more residues including but not limited positions R833, R836, K847, K879, K881, R883, R887, K897, K900, K932, R935, K940, K948, K953, K960, K984, K1003, K1017, R1033, R1138, R1165, and/or R1252 with reference to amino acid position numbering of LbCas13 (Lachnospiraceae bacterium ND2006). In certain embodiments, the Cas13 enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[0654] In certain embodiments, the Cas13 enzyme is modified by mutation of one or more residues including but not limited positions K15, R18, K26, Q34, R43, K48, K51, R56, R84, K85, K87, N93, R103, N104, T118, K123, K134, R176, K177, R192, K200, K226, K273, K275, T291, R301, K307, K369, S404, V409, K414, K436, K438, K468, D482, K516, R518, K524, K530, K532, K548, K559, K570, R574, K592, D596, K603, K607, K613, C647, R681, K686, H720, K739, K748, K757, T766, K780, R790, P791, K796, K809, K815, T816, K860, R862, R863, K868, K897, R909, R912, T923, R947, K949, R951, R955, K965, K968, K1000, R1003, K1009, K1017, K1022, K1029, A1053, K1072, K1086, F1103, S1209, R1226, R1252, K1273, K1282, and/or K1288 with reference to amino acid position numbering of AsCas13 (Acidaminococcus sp. BV3L6). In certain embodiments, the Cas13 enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[0655] In certain embodiments, the enzyme is modified by mutation of one or more residues including but not limited positions K15, R18, K26, R34, R43, K48, K51, K56, K87, K88, D90, K96, K106, K107, K120, Q125, K143, R186, K187, R202, K210, K235, K296, K298, K314, K320, K326, K397, K444, K449, E454, A483, E491, K527, K541, K581, R583, K589, K595, K597, K613, K624, K635, K639, K656, K660, K667, K671, K677, K719, K725, K730, K763, K782, K791, R800, K809, K823, R833, K834, K839, K852, K858, K859, K869, K871, R872, K877, K905, R918, R921, K932, I960, K962, R964, R968, K978, K981, K1013, R1016, K1021, K1029, K1034, K1041, K1065, K1084, and/or K1098 with reference to amino acid position numbering of FnCas13 (Francisella novicida U112). In certain embodiments, the Cas13 enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[0656] In certain embodiments, the enzyme is modified by mutation of one or more residues including but not limited positions K15, R18, K26, K34, R43, K48, K51, R56, K83, K84, R86, K92, R102, K103, K116, K121, R158, E159, R174, R182, K206, K251, K253, K269, K271, K278, P342, K380, R385, K390, K415, K421, K457, K471, A506, R508, K514, K520, K522, K538, Y548, K560, K564, K580, K584, K591, K595, K601, K634, K640, R645, K679, K689, K707, T716, K725, R737, R747, R748, K753, K768, K774, K775, K785, K787, R788, Q793, K821, R833, R836, K847, K879, K881, R883, R887, K897, K900, K932, R935, K940, K948, K953, K960, K984, K1003, K1017, R1033, K1121, R1138, R1165, K1190, K1199, and/or K1208 with reference to amino acid position numbering of LbCas13 (Lachnospiraceae bacterium ND2006). In certain embodiments, the Cas13 enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[0657] In certain embodiments, the enzyme is modified by mutation of one or more residues including but not limited positions K14, R17, R25, K33, M42, Q47, K50, D55, K85, N86, K88, K94, R104, K105, K118, K123, K131, R174, K175, R190, R198, I221, K267, Q269, K285, K291, K297, K357, K403, K409, K414, K448, K460, K501, K515, K550, R552, K558, K564, K566, K582, K593, K604, K608, K623, K627, K633, K637, E643, K780, Y787, K792, K830, Q846, K858, K867, K876, K890, R900, K901, M906, K921, K927, K928, K937, K939, R940, K945, Q975, R987, R990, K1001, R1034, I1036, R1038, R1042, K1052, K1055, K1087, R1090, K1095, N1103, K1108, K1115, K1139, K1158, R1172, K1188, K1276, R1293, A1319, K1340, K1349, and/or K1356 with reference to amino acid position numbering of MbCas13 (Moraxella bovoculi 237). In certain embodiments, the Cas13 enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[0658] In one embodiment, the Cas13 protein is modified with a mutation at S1228 (e.g., S1228A) with reference to amino acid position numbering of AsCas13. See Yamano et al., Cell 165:949–962 (2016), which is incorporated herein by reference in its entirety.

[0659] In certain embodiments, the Cas13 protein has been modified to recognize a non-natural PAM, such as recognizing a PAM having a sequence or comprising a sequence YCN, YCV, AYV, TYV, RYN, RCN, TGYV, NTTN, TTN, TRTN, TYTV, TYCT, TYCN, TRTN, NTTN, TACT, TYCC, TRTC, TATV, NTTV, TTV, TSTG, TVTS, TYYS, TCYS, TBYS, TCYS, TNYS, TYYS, TNTN, TSTG, TTCC, TCCC, TATC, TGTG, TCTG, TYCV, or TCTC. In particular embodiments, said mutated Cas13 comprises one or more mutated amino acid residue at position 11, 12, 13, 14, 15, 16, 17, 34, 36, 39, 40, 43, 46, 47, 50, 54, 57, 58, 111, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 157, 158, 159, 160, 161, 162, 163, 164,

165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 642, 643, 644, 645, 646, 647, 648, 649, 651, 652, 653, 654, 655, 656, 676, 679, 680, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 707, 711, 714, 715, 716, 717, 718, 719, 720, 721, 722, 739, 765, 768, 769, 773, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, or 1048 of AsCas13 or a position corresponding thereto in a Cas13 ortholog; preferably, one or more mutated amino acid residue at position 130, 131, 132, 133, 134, 135, 136, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 570, 571, 572, 573, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 630, 631, 632, 646, 647, 648, 649, 650, 651, 652, 653, 683, 684, 685, 686, 687, 688, 689, or 690;

[0660] In certain embodiments, the Cas13 protein is modified to have increased activity, i.e. wider PAM specificity. In particular embodiments, the Cas13 protein is modified by mutation of one or more residues including but not limited positions 539, 542, 547, 548, 550, 551, 552, 167, 604, and/or 607 of AsCas13, or the corresponding position of an AsCas13 orthologue, homologue, or variant, preferably mutated amino acid residues at positions 542 or 542 and 607, wherein said mutations preferably are 542R and 607R, such as S542R and K607R; or preferably mutated amino acid residues at positions 542 and 548 (and optionally 552), wherein said mutations preferably are 542R and 548V (and optionally 552R), such as S542R and K548V (and optionally N552R); or at position 532, 538, 542, and/or 595 of LbCas13, or the corresponding position of an AsCas13 orthologue, homologue, or variant, preferably mutated amino acid residues at positions 532 or 532 and 595, wherein said mutations preferably are 532R and 595R, such as G532R and K595R; or preferably mutated amino acid residues at positions 532 and 538 (and optionally 542), wherein said mutations preferably are 532R and 538V (and optionally 542R), such as G532R and K538V (and optionally Y542R), most preferably wherein said mutations are S542R and K607R, S542R and K548V, or S542R, K548V and N552R of AsCas13.

Deactivated / inactivated Cas13 protein

[0661] Where the Cas13 protein has nuclease activity, the Cas13 protein may be modified to have diminished nuclease activity e.g., nuclease inactivation of at least 70%, at least 80%,

at least 90%, at least 95%, at least 97%, or 100% as compared with the wild type enzyme; or to put in another way, a Cas13 enzyme having advantageously about 0% of the nuclease activity of the non-mutated or wild type Cas13 enzyme or CRISPR-Cas protein, or no more than about 3% or about 5% or about 10% of the nuclease activity of the non-mutated or wild type Cas13 enzyme, e.g. of the non-mutated or wild type Francisella novicida U112 (FnCas13), Acidaminococcus sp. BV3L6 (AsCas13), Lachnospiraceae bacterium ND2006 (LbCas13) or Moraxella bovoculi 237 (MbCas13 Cas13 enzyme or CRISPR-Cas protein. This is possible by introducing mutations into the nuclease domains of the Cas13 and orthologs thereof.

[0662] In preferred embodiments of the present invention at least one Cas13 protein is used which is a Cas13 nickase. More particularly, a Cas13 nickase is used which does not cleave the target strand but is capable of cleaving only the strand which is complementary to the target strand, i.e. the non-target DNA strand also referred to herein as the strand which is not complementary to the guide sequence. More particularly the Cas13 nickase is a Cas13 protein which comprises a mutation in the arginine at position 1226A in the Nuc domain of Cas13 from Acidaminococcus sp., or a corresponding position in a Cas13 ortholog. In further particular embodiments, the enzyme comprises an arginine-to-alanine substitution or an R1226A mutation. It will be understood by the skilled person that where the enzyme is not AsCas13, a mutation may be made at a residue in a corresponding position. In particular embodiments, the Cas13 is FnCas13 and the mutation is at the arginine at position R1218. In particular embodiments, the Cas13 is LbCas13 and the mutation is at the arginine at position R1138. In particular embodiments, the Cas13 is MbCas13 and the mutation is at the arginine at position R1293.

[0663] In certain embodiments, use is made additionally or alternatively of a CRISPR-Cas protein which is is engineered and can comprise one or more mutations that reduce or eliminate a nuclease activity. The amino acid positions in the FnCas13p RuvC domain include but are not limited to D917A, E1006A, E1028A, D1227A, D1255A, N1257A, D917A, E1006A, E1028A, D1227A, D1255A and N1257A. Applicants have also identified a putative second nuclease domain which is most similar to PD-(D/E)XK nuclease superfamily and HincII endonuclease like. The point mutations to be generated in this putative nuclease domain to substantially reduce nuclease activity include but are not limited to N580A, N584A, T587A, W609A, D610A, K613A, E614A, D616A, K624A, D625A, K627A and Y629A. In a preferred embodiment, the mutation in the FnCas13p RuvC domain is D917A or E1006A, wherein the D917A or E1006A mutation completely inactivates the DNA cleavage activity of the FnCas13 effector protein. In another embodiment, the mutation in the FnCas13p RuvC domain is

D1255A, wherein the mutated FnCas13 effector protein has significantly reduced nucleolytic activity.

[0664] More particularly, the inactivated Cas13 enzymes include enzymes mutated in amino acid positions As908, As993, As1263 of AsCas13 or corresponding positions in Cas13 orthologs. Additionally, the inactivated Cas13 enzymes include enzymes mutated in amino acid position Lb832, 925, 947 or 1180 of LbCas13 or corresponding positions in Cas13 orthologs. More particularly, the inactivated Cas13 enzymes include enzymes comprising one or more of mutations AsD908A, AsE993A, AsD1263A of AsCas13 or corresponding mutations in Cas13 orthologs. Additionally, the inactivated Cas13 enzymes include enzymes comprising one or more of mutations LbD832A, E925A, D947A or D1180A of LbCas13 or corresponding mutations in Cas13 orthologs.

[0665] Mutations can also be made at neighboring residues, e.g., at amino acids near those indicated above that participate in the nuclease acrivity. In some embodiments, only the RuvC domain is inactivated, and in other embodiments, another putative nuclease domain is inactivated, wherein the effector protein complex functions as a nickase and cleaves only one DNA strand. In a preferred embodiment, the other putative nuclease domain is a HincII-like endonuclease domain.

[0666] The inactivated Cas13 or Cas13 nickase may have associated (e.g., via fusion protein) one or more functional domains, including for example, an adenosine deaminase or catalytic domain thereof. In some cases it is advantageous that additionally at least one heterologous NLS is provided. In some instances, it is advantageous to position the NLS at the N terminus. In general, the positioning of the one or more functional domain on the inactivated Cas13 or Cas13 nickase is one which allows for correct spatial orientation for the functional domain to affect the target with the attributed functional effect. For example, when the functional domain is an adenosine deaminase catalytic domain thereof, the adenosine deaminase catalytic domain is placed in a spatial orientation which allows it to contact and deaminate a target adenine. This may include positions other than the N- / C- terminus of Cas13. In some embodiments, the adenosine deaminase protein or catalytic domain thereof is inserted into an internal loop of Cas13.

Determination of PAM

[0667] Determination of PAM can be ensured as follows. This experiment closely parallels similar work in *E. coli* for the heterologous expression of StCas9 (Sapranauskas, R. et al. Nucleic Acids Res 39, 9275–9282 (2011)). Applicants introduce a plasmid containing both a

PAM and a resistance gene into the heterologous E. coli, and then plate on the corresponding antibiotic. If there is DNA cleavage of the plasmid, Applicants observe no viable colonies.

[0668] In further detail, the assay is as follows for a DNA target. Two E.coli strains are used in this assay. One carries a plasmid that encodes the endogenous effector protein locus from the bacterial strain. The other strain carries an empty plasmid (e.g.pACYC184, control strain). All possible 7 or 8 bp PAM sequences are presented on an antibiotic resistance plasmid (pUC19 with ampicillin resistance gene). The PAM is located next to the sequence of proto-spacer 1 (the DNA target to the first spacer in the endogenous effector protein locus). Two PAM libraries were cloned. One has a 8 random bp 5' of the proto-spacer (e.g. total of 65536 different PAM sequences = complexity). The other library has 7 random bp 3' of the proto-spacer (e.g. total complexity is 16384 different PAMs). Both libraries were cloned to have in average 500 plasmids per possible PAM. Test strain and control strain were transformed with 5'PAM and 3'PAM library in separate transformations and transformed cells were plated separately on ampicillin plates. Recognition and subsequent cutting/interference with the plasmid renders a cell vulnerable to ampicillin and prevents growth. Approximately 12h after transformation, all colonies formed by the test and control strains where harvested and plasmid DNA was isolated. Plasmid DNA was used as template for PCR amplification and subsequent deep sequencing. Representation of all PAMs in the untransfomed libraries showed the expected representation of PAMs in transformed cells. Representation of all PAMs found in control strains showed the actual representation. Representation of all PAMs in test strain showed which PAMs are not recognized by the enzyme and comparison to the control strain allows extracting the sequence of the depleted PAM.

[0669] The following PAMs have been identified for certain wild-type Cas13 orthologues: the Acidaminococcus sp. BV3L6 Cas13 (AsCas13), Lachnospiraceae bacterium ND2006 Cas13 (LbCas13) and Prevotella albensis (PaCas13) can cleave target sites preceded by a TTTV PAM, where V is A/C or G, FnCas13p, can cleave sites preceded by TTN, where N is A/C/G or T. The Moraxella bovoculi AAX08_00205, Moraxella bovoculi AAX11_00205, Butyrivibrio sp. NC3005, Thiomicrospira sp. XS5, or Lachnospiraceae bacterium MA2020 PAM is 5' TTN, where N is A/C/G or T. The natural PAM sequence is TTTV or BTTV, wherein B is T/C or G and V is A/C or G and the effector protein is Moraxella lacunata Cas13.

Codon optimized nucleic acid sequences

[0670] Where the effector protein is to be administered as a nucleic acid, the application envisages the use of codon-optimized CRISPR-Cas type V protein, and more particularly Cas13-encoding nucleic acid sequences (and optionally protein sequences). An example of a

codon optimized sequence, is in this instance a sequence optimized for expression in a eukaryote, e.g., humans (i.e. being optimized for expression in humans), or for another eukaryote, animal or mammal as herein discussed; see, e.g., SaCas9 human codon optimized sequence in WO 2014/093622 (PCT/US2013/074667) as an example of a codon optimized sequence (from knowledge in the art and this disclosure, codon optimizing coding nucleic acid molecule(s), especially as to effector protein (e.g., Cas13) is within the ambit of the skilled artisan). Whilst this is preferred, it will be appreciated that other examples are possible and codon optimization for a host species other than human, or for codon optimization for specific organs is known. In some embodiments, an enzyme coding sequence encoding a DNA/RNA-targeting Cas protein is codon optimized for expression in particular cells, such as eukaryotic cells. The eukaryotic cells may be those of or derived from a particular organism, such as a plant or a mammal, including but not limited to human, or non-human eukaryote or animal or mammal as herein discussed, e.g., mouse, rat, rabbit, dog, livestock, or non-human mammal or primate. In some embodiments, processes for modifying the germ line genetic identity of human beings and/or processes for modifying the genetic identity of animals which are likely to cause them suffering without any substantial medical benefit to man or animal, and also animals resulting from such processes, may be excluded. In general, codon optimization refers to a process of modifying a nucleic acid sequence for enhanced expression in the host cells of interest by replacing at least one codon (e.g., about or more than about 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more codons) of the native sequence with codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence. Various species exhibit particular bias for certain codons of a particular amino acid. Codon bias (differences in codon usage between organisms) often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, among other things, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization. Codon usage tables are readily available, for example, at the “Codon Usage Database” available at www.kazusa.orjp/codon/ and these tables can be adapted in a number of ways. See Nakamura, Y., et al. “Codon usage tabulated from the international DNA sequence databases: status for the year 2000” Nucl. Acids Res. 28:292 (2000). Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as Gene Forge (Aptagen; Jacobus, PA), are also available. In some embodiments, one or more codons

(e.g., 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more, or all codons) in a sequence encoding a DNA/RNA-targeting Cas protein corresponds to the most frequently used codon for a particular amino acid. As to codon usage in yeast, reference is made to the online Yeast Genome database available at http://www.yeastgenome.org/community/codon_usage.shtml, or Codon selection in yeast, Bennetzen and Hall, J Biol Chem. 1982 Mar 25;257(6):3026-31. As to codon usage in plants including algae, reference is made to Codon usage in higher plants, green algae, and cyanobacteria, Campbell and Gowri, Plant Physiol. 1990 Jan; 92(1): 1–11.; as well as Codon usage in plant genes, Murray et al, Nucleic Acids Res. 1989 Jan 25;17(2):477-98; or Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages, Morton BR, J Mol Evol. 1998 Apr;46(4):449-59.

[0671] In certain example embodiments, the CRISPR Cas protein is selected from Table 1.

Table 1

C2c2 orthologue	Code	Multi Letter
Leptotrichia shahii	C2-2	Lsh
L wadei F0279 (Lw2)	C2-3	Lw2
Listeria seeligeri	C2-4	Lse
Lachnospiraceae bacterium MA2020	C2-5	LbM
Lachnospiraceae bacterium NK4A179	C2-6	LbNK179
[Clostridium] aminophilum DSM 10710	C2-7	Ca
Carnobacterium gallinarum DSM 4847	C2-8	Cg
Carnobacterium gallinarum DSM 4847	C2-9	Cg2
Paludibacter propionicigenes WB4	C2-10	Pp
Listeria weihenstephanensis FSL R9-0317	C2-11	Lwei
Listeriaceae bacterium FSL M6-0635	C2-12	LbFSL
Leptotrichia wadei F0279	C2-13	Lw
Rhodobacter capsulatus SB 1003	C2-14	Rc
Rhodobacter capsulatus R121	C2-15	Rc
Rhodobacter capsulatus DE442	C2-16	Rc

[0672] In certain example embodiments, the CRISPR effector protein is a Cas13a protein selected from Table 2

Table 2

c2c2-5	1	Lachnospiraceae bacterium MA2020 (SEQ ID No. 34)	mqiskvnhkhvavgqkdreritgfiyndpvgdeksledvvakrandtkvlfnvfnt kdlydsqesdksekdkeiiskgakfvaksfnsaitilkkqnkiystltsqqvikelkdk fggariydddieealtetlkksfrkenvrnsikvlienaagirsslskdeeeliqeyfvk qlveeytktklqknvvksiknqnmviqpdsdsqvlslsesrrekqssavssdtlvnc kekdvlkafltdyavldedernsllwklrnlvnlyfygsesirdysytkeksvwkeh deqkanktlfideichitkigkngkeqkvldyeenrsrcrkqninyyrsalnyaknnt sgifenedsnhfwihlieneverlyngiengeefkfetgyisekvwkavinhlsikyi algkavynyamkelsspgdiepgkiddsyingitsfdyeiikaeeslqrdismnvvf atnylacatvdtdkdfllfskedirsctkkdgnlcknimqfwggystwknfceeylk ddkdalellyslksmlysmrnssfhfstenvdngswdteligklfeedcnraarieke kfynnnlhmfysssllekvlerlysshherasqvpsfnrvfvrknfpsslseqritpkft dskdeqiwqsavyylckeiyyndflqskeayklfregvknldkndinnqkaadsfk qavvyygkaignatlsqvcqaimteynrqnndglkkksayaekqnsnkykhyplf lkqvlqsafweyldenkeiygfisaqihksnveikaedfianyssqqykklvdkvk ktpelqkwytlgrlinprqanqflgsirnyvqfvkdiqrrakengnpirnyyevlesd siikilemctklngttsndihdyfrdedeyaeyisqfvnfgdvhsgaalnafcnsese gkkngiyydginpivnrnwvlcklygspdliskiisrvnenmihdfhkqedlireyq ikgicsnkkeqqdlrtfqvlknrvelrdiveyseiinelygqlikwcylrerdlmyfql gfhylclnnasskeadyikinvddrnisgailyqiaamyinglpvyykkddmyval ksgkkasdelnsneqtskkinyflkygnnilgdkkdqlylaglelfenvaeheniiifr neidhfhyfydrdrsmldlysevfdrfftydmklrknvvnmlynilldhnivssfvf etgekkvgrgdsevikpsakirlranngvssdvftykvgskdelkiatlpakneeflln varliyypdmeavsenmvregvvkveksndkkgkisrgsntrssnqskynnksk nrmnysmgsifekmdlkfd
c2c2-6	2	Lachnospiraceae bacterium NK4A179 (SEQ ID No. 35)	mkiskvreenrgakltvnaktavvsenrsqegilyndpsrygksrkndedrdryies rlkssgklyrifnedknkretdelqwflseivkkinrrnglvlsdmlsvddrafekafe kyaelsytnrrnkvsgspafetcgvdaataerlkgiisetnfinriknnidnkvsediid riiakylkkslcrervkrglkkllmnafdlpysdpdidvqrdfidyvledfyhvraks qvsrsiknmnmpvqpegdgkfaitvskggtesgnkrsaekeafkkflsdyaslder vrddmlrrmrrlvvlyfygsddsklsdvnekfdvwedhaarrvdnrefiklplenkl angktdkdaerirkntvkelyrnqnigcyrqavkaveednngryfddkmlnmffi

			hrieygvekiyanlkqvtefkartgylsekiwkdlinyisikyiamgkavynyamd elnasdkkeielgkiseeylsgissfdyelikaeemlqretavyvafaarhlssqtveld sensdflllkpkgtmdkndknklasnnilnflkdketlrdtilqyfgghslwtdfpfdk ylaggkddvdfltdlkdviysmrndsfhyatenhnngkwnkelisamfeheterm tvvmkdkfysnnlpmfyknddlkkllidlykdnverasqvpsfnkvfvrknfpalv rdkdnlgieldlkadadkgenelkfynalyymfkeiyynaflndknvrerfitkatkv adnydrnkernlkdriksagsdekkklreqlqnyiaendfgqriknivqvnpdytla qicqlimteynqqnngcmqkksaarkdinkdsyqhykmlllvnlrkaflefikeny afvlkpykhdlcdkadfvpdfakyvkpyaglisrvagsselqkwyivsrflspaqan hmlgflhsykqyvwdiyrrasetgteinhsiaedkiagvditdvdavidlsvklcgti sseisdyfkddevyaeyissyldfeydggnykdslnrfcnsdavndqkvalyydge hpklnrniilsklygerrflekitdrvsrsdiveyyklkketsqyqtkgifdsedeqkni kkfqemknivefrdlmdyseiadelqgqlinwiylrerdlmnfqlgyhyaclnnds nkqatyvtldyqgkknrkingailyqicamyinglplyyvdkdssewtvsdgkest gakigefyryaksfentsdcyasgleifenisehdnitelrnyiehfryyssfdrsflgiy sevfdrfftydlkyrknvptilynillqhfvnvrfefvsgkkmigidkkdrkiakekec aritireknggvyseqftyklkngtvyvdardkrylqsiirllfypekvnmdemievke kkkpsdnntgkgyskrdrqqdrkeydkykekkkkegnflsgmggninwdeina qlkn
c2c2-7	3	[Clostridium] aminophilum DSM 10710 SEQ ID No. 36)	mkfskvdhtrsavgiqkatdsvhgmlytdpkkqevndldkrfdqlnvkakrlynv fnqskaeedddekrfgkvvkklnrelkdllfhrevsrynsignakynyygiksnpee ivsnlgmveslkgerdpqkviskllyylrkglkpgtdglrmileascglrklsgdeke lkvflqtldedfekktfkknlirsienqnmavqpsnegdpiigitqgrfnsqkneeks aiermmsmyadlnedhredvlrklrrlnvlyfnvdtekteeptlpgevdtnpvfev whdhekgkendrqfatfakiltedretrkkeklavkealndlksairdhnimayrcsi kvteqdkdglffedqrinrfwihhiesaverilasinpeklyklrigylgekvwkdlln ylsikyiavgkavfhfamedlgktgqdielgklsnsvsggltsfdyeqiradetlqrql svevafaannlfravvgqtgkkieqskseeneedfllwkaekiaesikkegegntlks ilqffggasswdlnhfcaaygnessalgyetkfaddlrkaiyslrnetfhfttlnkgsfd wnakligdmfsheaatgiavertrfysnnlpmfyresdlkrimdhlyntyhprasqv psfnsvfvrknfrlflsntlntntsfdtevyqkwesgvyylfkeiyynsflpsgdahhlf feglrrirkeadnlpivgkeakkrnavqdfgrrcdelknlslsaicqmimteyneqnn

			gnrkvkstredkrkpdifqhykmlllrtlqeafaiyirreefkfifdlpktlyvmkpve eflpnwksgmfdslvervkqspdlqrwyvlckflngrllnqlsgvirsyiqfagdiqr rakanhnrlymdntqrveyysnvlevvdfcikgtsrfsnvfsdyfrdedayadyldn ylqfkdekiaevssfaalktfcneeevkagiymdgenpvmqrnivmaklfgpdev lknvvpkvtreeieeyyqlekqiapyrqngyckseedqkkllrfqriknrvefqtitef seiinellgqliswsflrerdllyfqlgfhylclhndtekpaeykeisredgtvirnailhq vaamyvgglpvytladkklaafekgeadcklsiskdtagagkkikdffryskyvlik drmltdqnqkytiylaglelfentdehdnitdvrkyvdhfkyyatsdenamsildlys eihdrfftydmkyqknvanmlenillrhfvlirpefftgskkvgegkkitckaraqiei aengmrsedftyklsdgkknistcmiaardqkylntvarllyypheakksivdtrek knnkktnrgdgtfnkqkgtarkekdngprefndtgfsntpfagfdpfrns
c2c2-8	5	Carnobacterium gallinarum DSM 4847 (SEQ ID No. 37)	mritkvkikldnklyqvtmqkeekygtlklneesrkstaeilrlkkasfnksfhsktin sqkenknatikkngdyisqifeklvgvdtnknirkpkmsltdlkdlpkkdlalfikrk fknddiveiknldlislfynalqkvpgehftdeswadfcqemmpyreyknkfierk iillansieqnkgfsinpetfskrkrvlhqwaievqergdfsildeklsklaeiynfkk mckrvqdelndleksmkkgknpekekeaykkqknfkiktiwkdypykthiglie kikeneelnqfnieigkyfehyfpikkerctedepyylnsetiattvnyqlknalisyl mqigkykqfglenqvldskklqeigiyegfqtkfmdacvfatsslkniiepmrsgdi lgkrefkeaiatssfvnyhhffpyfpfelkgmkdreselipfgeqteakqmqniwal rgsvqqirneifhsfdknqkfnlpqldksnfefdasenstgksqsyietdykflfeaek nqleqffierikssgaleyyplksleklfakkemkfslgsqvvafapsykklvkkghs yqtategtanylglsyynryelkeesfqaqyyllkliyqyvflpnfsqgnspafretvk ailrinkdearkkmkknkkflrkyafeqvremefketpdqymsylqsemreekvr kaekndkgfeknitmnfekllmqifvkgfdvflttfagkelllsseekviketeislsk kinerektlkasiqvehqlvatnsaisywlfcklldsrhlnelrnemikfkqsrikfnht qhaeliqnllpiveltilsndydekndsqnvdvsayfedkslyetapyvqtddrtrvsf rpilklekyhtkslieallkdnpqfrvaatdiqewmhkreeigelvekrknlhtewae gqqtlgaekreeyrdyckkidrfnwkankvtltylsqlhylitdllgrmvgfsalferd lvyfsrsfselggetyhisdyknlsgvlrlnaevkpikiknikvidneenpykgnepe vkpfldrlhaylenvigikavhgkirnqtahlsvlqlelsmiesmnnlrdlmaydrkl knavtksmikildkhgmilklkidenhknfeieslipkeiihlkdkaiktnqvseeyc qlvlallttnpgnqln

c2c2-9	6	Carnobacterium gallinarum DSM 4847 (SEQ ID No. 38)	mrmtkvkingspvsmnrsklnghlvwngttntvniltkkeqsfaasflnktlvkad qvkgykvlaenifiifeqleksnsekpsvylnnirrlkeaglkrffkskyheeikytse knqsvptklnliplffnavdriqedkfdeknwsyfckemspyldykksylnrkkeil ansiqqnrgfsmptaeepnllskrkqlfqqwamkfqespliqqnnfaveqfnkefa nkinelaavynvdelctaiteklmnfdkdksnktrnfeikklwkqhphnkdkalikl fnqegnealnqfnielgkyfehyfpktgkkesaesyylnpqtiiktvgyqlrnafvqy llqvgklhqynkgvldsqtlqeigmyegfqtkfmdacvfassslrniiqattnediltr ekfkkeleknvelkhdlffkteiveerdenpakkiamtpneldlwairgavqrvrnq ifhqqinkrhepnqlkvgsfengdlgnvsyqktiyqklfdaeikdieiyfaekikssg aleqysmkdleklfsnkeltlslggqvvafapsykklykqgyfyqnektieleqftdy dfsndvfkanyylikliyhyvflpqfsqannklfkdtvhyviqqnkelnttekdkkn nkkirkyafeqvklmknespekymqylqremqeertikeakktneekpnynfek lliqifikgfdtflrnfdlnlnpaeelvgtvkekaeglrkrkeriakilnvdeqiktgdeei afwifaklldarhlselrnemikfkqssvkkglikngdlieqmqpilelcilsndses mekesfdkievflekvelaknepymqedkltpvkfrfmkqlekyqtrnfienlvie npefkvsekivlnwheekekiadlvdkrtklheewaskareieeynekikknkskk ldkpaefakfaeykiiceaienfnrldhkvrltylknlhylmidlmgrmvgfsvlfer dfvymgrsysalkkqsiylndydtfanirdwevnenkhlfgtsssdltfqetaefknl kkpmenqlkallgvtnhsfeirnniahlhvlrndgkgegvsllscmndlrklmsydr klknavtkaiikildkhgmilkltnndhtkpfeieslkpkkiihleksnhsfpmdqvs qeycdlvkkmlvftn
c2c2-10	7	Paludibacter propionicigenes WB4 (SEQ ID No. 39)	mrvskvkvkdggkdkmvlvhrkttgaqlvysgqpvsnetsnilpekkrqsfdlstl nktiikfdtakkqklnvdqykivekifkypkqelpkqikaeeilpflnhkfqepvky wkngkeesfnltlliveavqaqdkrklqpyydwktwyiqtksdllkksiennridlte nlskrkkallaweteftasgsidlthyhkvymtdvlckmlqdvkpltddkgkintna yhrglkkalqnhqpaifgtrevpneanradnqlsiyhlevvkylehyfpiktskrrnt addiahylkaqtlkttiekqlvnairaniiqqgktnhhelkadttsndliriktneafvln ltgtcafaannirnmvdneqtndilgkgdfiksllkdntnsqlysfffgeglstnkaek etqlwgirgavqqirnnvnhykkdalktvfnisnfenptitdpkqqtnyadtiykarf inelekipeafaqqlktggavsyytienlkslltfqfslcrstipfapgfkkvfngginy qnakqdesfyelmleqylrkenfaeesynaryfmlkliynnlflpgfttdrkafadsv gfvqmqnkkqaekvnprkkeayafeavrpmtaadsiadymayvqselmqeqn

			kkeekvaeetrinfekfvlqvfikgfdsflrakefdfvqmpqpqltatasnqqkadkl nqleasitadckltpqyakaddathiafyvfcklldaahlsnlrnelikfresvnefkfh hlleiieiclsadvvptdyrdlysseadclarlrpfieqgaditnwsdlfvqsdkhspvi hanielsvkygttkllleqiinkdtqfktteanftawntaqksieqlikqredhheqwvk aknaddkekqerkreksnfaqkfiekhgddyldicdyintynwldnkmhfvhlnr lhgltiellgrmagfvalfdrdfqffdeqqiadefklhgfvnlhsidkklnevptkkike iydirnkiiqingnkinesvranliqfisskrnyynnaflhvsndeikekqmydirnh iahfnyltkdaadfslidlinelrellhydrklknavskafidlfdkhgmilklklnadh klkveslepkkiyhlgssakdkpeyqyctnqvmmaycnmcrsllemkk
c2c2-11	9	Listeria weihenstephanensis FSL R9-0317 (SEQ ID No. 40)	mlallhqevpsqklhnlkslntesltklfkpkfqnmisyppskgaehvqfcltdiavp airdldeikpdwgiffeklkpytdwaesyihykqttiqksieqnkiqspdsprklvlq kyvtaflngeplgldlvakkykladlaesfkvvdlnedksanykikaclqqhqrnild elkedpelnqygievkkyiqryfpikrapnrskharadflkkeliestveqqfknavy hyvleqgkmeayeltdpktkdlqdirsgeafsfkfinacafasnnlkmilnpecekd ilgkgdfkknlpnsttqsdvvkkmipffsdeiqnvnfdeaiwairgsiqqirnevyh ckkhswksilkikgfefepnnmkytdsdmqklmdkdiakipdfieeklkssgiirf yshdklqsiwemkqgfsllttnapfvpsfkrvyakghdyqtsknryydlglttfdile ygeedfraryfltklvyyqqfmpwftadnnafrdaanfvlrlnknrqqdakafinire veegemprdymgyvqgqiaihedstedtpnhfekfisqvfikgfdshmrsadlkfi knprnqgleqseieemsfdikvepsflknkddyiafwtfckmldarhlselrnemi kydghltgeqeiiglallgvdsrendwkqffssereyekimkgyvgeelyqrepyrq sdgktpilfrgveqarkygtetviqrlfdaspefkvskcnitewerqketieetierrkel hneweknpkkpqnnaffkeykeccdaidaynwhknkttlvyvnelhhllieilgry vgyvaiadrdfqcmanqyfkhsgiterveywgdnrlksikkldtflkkeglfvsekn arnhiahlnylslksectllylserlreifkydrklknavskslidildrhgmsvvfanlk enkhrlvikslepkklrhlgekkidngyietnqvseeycgivkrllei
c2c2-12	10	Listeriaceae bacterium FSL M6-0635 = Listeria newyorken	mkitkmrvdgrtivmertskegqlgyegidgnktteiifdkkkesfyksilnktvrkp dekeknrrkqainkainkeitelmlavlhqevpsqklhnlkslntesltklfkpkfqn misyppskgaehvqfcltdiavpairdldeikpdwgiffeklkpytdwaesyihyk qttiqksieqnkiqspdsprklvlqkyvtaflngeplgldlvakkykladlaesfklvdl nedksanykikaclqqhqrnildelkedpelnqygievkkyiqryfpikrapnrskh aradflkkeliestveqqfknavyhyvleqgkmeayeltdpktkdlqdirsgeafsfk

		sis FSL M6-0635 (SEQ ID No. 41)	finacafasnnlkmilnpecekdilgkgnfkknlpnsttrsdvvkkmipffsdelqn vnfdeaiwairgsiqqirnevyhckkhswksilkikgfefepnnmkyadsdmqkl mdkdiakipefieeklkssgvvrfyrhdelqsiwemkqgfsllttnapfvpsfkrvy akghdyqtsknryynldlttfdileygeedfraryfltklvyyqqfmpwftadnnafr daanfvlrlnknrqqdakafinireveegemprdymgyvqgqiaihedsiedtpnh fekfisqvfikgfdrhmrsanlkfiknprnqgleqseieemsfdikvepsflknkdd yiafwifckmldarhlselrnemikydghltgeqeiiglallgvdsrendwkqffsse reyekimkgyvveelyqrepyrqsdgktpilfrgveqarkygtetviqrlfdanpefk vskcnlaewerqketieetikrrkelhnewaknpkkpqnnaffkeykeccdaiday nwhknkttlayvnelhhllieilgryvgyvaiadrdfqcmanqyfkhsgiterveyw gdnrlksikkldtflkkeglfvseknarnhiahlnylslksectllylserlreifkydrkl knavskslidildrhgmsvvfanlkenkhrlvikslepkklrhlggkkidggyietnq vseeycgivkrllem
c2c2-13	12	Leptotrichia wadei F0279 (SEQ ID No. 42)	mkvtkvdgishkkyieegklvkstseenrtserlsellsirldiyiknpdnaseeenrir renlkkffsnkvlhlkdsvlylknrkeknavqdknyseediseydlknknsfsvlkk illnedvnseeleifrkdveaklnkinslkysfeenkanyqkinennvekvggkskr niiydyyresakrndyinnvqeafdklykkedieklfflienskkhekykireyyhki igrkndkenfakiiyeeiqnvnnikeliekipdmselkksqvfykyyldkeelndkn ikyafchfveiemsqllknyvykrlsnisndkikrifeyqnlkklienkllnkldtyvr ncgkynyylqvgeiatsdfiarnrqneaflrniigvssvayfslrniletenenditgrm rgktvknnkgeekyvsgevdkiynenkqnevkenlkmfysydfnmdnkneied ffanideaissirhgivhfnlelegkdifafkniapseiskkmfqneinekklklkifkq lnsanvfnyyekdviikylkntkfnfvnknipfvpsftklynkiedlrntlkffwsvp kdkeekdaqiyllkniyygeflnkfvknskvffkitnevikinkqrnqktghykyqk feniektvpveylaiiqsreminnqdkeekntyidfiqqiflkgfidylnknnlkyies nnnndnndifskikikkdnkekydkilknyekhnrnkeipheinefvreiklgkilk ytenlnmfylilkllnhkeltnlkgslekyqsankeetfsdelelinllnldnnrvtedfe leaneigkfldfnenkikdrkelkkfdtnkiyfdgeniikhrafynikkygmlnlleki adkakykislkelkeysnkkneieknytmqqnlhrkyarpkkdekfndedykeye kaigniqkythlknkvefnelnllqglllkilhrlvgytsiwerdlrfrlkgefpenhyie eifnfdnsknvkyksgqivekyinfykelykdnvekrsiysdkkvkklkqekkdly

			irnyiahfnyiphaeisllevlenlrkllsydrklknaimksivdilkeygfvatfkiga dkkieiqtlesekivhlknlkkkklmtdrnseelcelvkvmfeykale
c2c2-14	15	Rhodobacter capsulatus SB 1003 (SEQ ID No. 43)	mqigkvqgrtisefgdpagglkrkistdgknrkelpahlssdpkaligqwisgidkiy rkpdsrksdgkaihsptpskmqfdarddlgeafwklvseaglaqdsdydqfkrrlh pygdkfqpadsgaklkfeadppepqafhgrwygamskrgndakelaaalyehlh vdekridgqpkrnpktdkfapglvvaralgiessvlprgmarlarnwgeeeiqtyfv vdvaasvkevakaavsaaqafdpprqvsgrslspkvgfalaehlervtgskrcsfdp aagpsvlalhdevkktykrlcargknaarafpadktellalmrhthenrvrnqmvr mgrvseyrgqqagdlaqshywtsagqteikeseifvrlwvgafalagrsmkawid pmgkivnteknderdltaavnirqvisnkemvaeamarrgiyfgetpeldrlgaegn egfvfallrylrgcrnqtfhlgaragflkeirkelektrwgkakeaehvvltdktvaair aiidndakalgarlladlsgafvahyaskehfstlyseivkavkdapevssglprlklll kradgvrgyvhglrdtrkhafatklppppaprelddpatkaryiallrlydgpfrayas gitgtalagpaarakeaatalaqsvnvtkaysdvmegrtsrlrppndgetlreylsaltg etatefrvqigyesdsenarkqaefienyrrdmlafmfedyirakgfdwilkiepgat amtrapvlpepidtrgqyehwqaalylvmhfvpasdvsnllhqlrkwealqgkyel vqdgdatdqadarrealdlvkrfrdvlvlflktgearfegraapfdlkpfralfanpatf drlfmatpttarpaeddpegdgasepelrvartlrglrqiarynhmavlsdlfakhkvr deevarlaeiedetqeksqivaaqelrtdlhdkvmkchpktispeerqsyaaaiktie ehrflvgrvylgdhlrlhrlmmdvigrlidyagayerdtgtflinaskqlgagadwav tiagaantdartqtrkdlahfnvldradgtpdltalvnraremmaydrkrknavprsil dmlarlgltlkwqmkdhllqdatitqaaikhldkvrltvggpaavtearfsqdylqm vaavfngsvqnpkprrrddgdawhkppkpataqsqpdqkppnkapsagsrlppp qvgevyegvvvkvidtgslgflavegvagniglhisrlrriredaiivgrryrfrveiyv ppksntsklnaadlvrid
c2c2-15	16	Rhodobacter capsulatus R121 (SEQ ID No. 44)	mqigkvqgrtisefgdpagglkrkistdgknrkelpahlssdpkaligqwisgidkiy rkpdsrksdgkaihsptpskmqfdarddlgeafwklvseaglaqdsdydqfkrrlh pygdkfqpadsgaklkfeadppepqafhgrwygamskrgndakelaaalyehlh vdekridgqpkrnpktdkfapglvvaralgiessvlprgmarlarnwgeeeiqtyfv vdvaasvkevakaavsaaqafdpprqvsgrslspkvgfalaehlervtgskrcsfdp aagpsvlalhdevkktykrlcargknaarafpadktellalmrhthenrvrnqmvr mgrvseyrgqqagdlaqshywtsagqteikeseifvrlwvgafalagrsmkawid

			pmgkivntekndrdltaavnirqvisnkemvaeamarrgiyfgetpeldrlgaegn egfvfallrylrgcrnqtfhlgaragflkeirkelektrwgkakeaehvvltdktvaair aiidndakalgarlladlsgafvahyaskehfstlyseivkavkdapevssglprlklll kradgvrgyvhglrdtrkhafatklppppaprelddpatkaryiallrlydgpfrayas gitgtalagpaarakeaatalaqsvnvtkaysdvmegrssrlrppndgetlreylsalt getatefrvqigyesdsenarkqaefienyrrdmlafmfedyirakgfdwilkiepga tamtrapvlpepidtrgqyehwqaalylvmhfvpasdvsnllhqlrkwealqgkye lvqdgdatdqadarrealdlvkrfrdvlvlflktgearfegraapfdlkpfralfanpatf drlfmatpttarpaeddpegdgasepelrvartlrglrqiarynhmavlsdlfakhkvr deevarlaeiedetqeksqivaaqelrtdlhdkvmkchpktispeerqsyaaaiktie ehrflvgrvylgdhlrlhrlmmdvigrlidyagayerdtgtflinaskqlgagadwav tiagaantdartqtrkdlahfnvldradgtpdltalvnraremmaydrkrknavprsil dmlarlgltlkwqmkdhllqdatitqaaikhldkvrltvggpaavtearfsqdylqm vaavfngsvqnpkprrrddgdawhkppkpataqsqpdqkppnkapsagsrlppp qvgevyegvvvkvidtgslgflavegvagniglhisrlrriredaiivgrryrfrveiyv ppksntsklnaadlvrid
c2c2-16	17	Rhodobacter capsulatus DE442 (SEQ ID No. 45)	mqigkvqgrtisefgdpagglkrkistdgknrkelpahlssdpkaligqwisgidkiy rkpdsrksdgkaihsptpskmqfdarddlgeafwklvseaglaqdsdydqfkrrlh pygdkfqpadsgaklkfeadppepqafhgrwygamskrgndakelaaalyehlh vdekridgqpkrnpktdkfapglvvaralgiessvlprgmarlarnwgeeeiqtyfv vdvaasvkevakaavsaaqafdpprqvsgrslspkvgfalaehlervtgskrcsfdp aagpsvlalhdevkktykrlcargknaarafpadktellalmrhthenrvrnqmvr mgrvseyrgqqagdlaqshywtsagqteikeseifvrlwvgafalagrsmkawid pmgkivntekndrdltaavnirqvisnkemvaeamarrgiyfgetpeldrlgaegn egfvfallrylrgcrnqtfhlgaragflkeirkelektrwgkakeaehvvltdktvaair aiidndakalgarlladlsgafvahyaskehfstlyseivkavkdapevssglprlklll kradgvrgyvhglrdtrkhafatklppppaprelddpatkaryiallrlydgpfrayas gitgtalagpaarakeaatalaqsvnvtkaysdvmegrssrlrppndgetlreylsalt getatefrvqigyesdsenarkqaefienyrrdmlafmfedyirakgfdwilkiepga tamtrapvlpepidtrgqyehwqaalylvmhfvpasdvsnllhqlrkwealqgkye lvqdgdatdqadarrealdlvkrfrdvlvlflktgearfegraapfdlkpfralfanpatf drlfmatpttarpaeddpegdgasepelrvartlrglrqiarynhmavlsdlfakhkvr

			deevarlaeiedetqeksqivaaqelrtdlhdkvmkchpktispeerqsyaaaiktie ehrflvgrvylgdhlrlhrlmmdvigrlidyagayerdtgtflinaskqlgagadwav tiagaantdartqtrkdlahfnvldradgtpdltalvnraremmaydrkrknavprsil dmlarlgltlkwqmkdhllqdatitqaaikhldkvrltvggpaavtearfsqdylqm vaavfngsvqnpkprrrddgdawhkppkpataqsqpdqkppnkapsagsrlppp qvgevyegvvvkvidtgslgflavegvagniglhisrlrriredaiivgrryrfrveiyv ppksntsklnaadlvrid
c2c2-2		(SEQ ID No. 46)	mgnlfghkrwyevrdkkdfkikrkvkvkrnydgnkyilninennnkekidnnkfi rkyinykkndnilkeftrkfhagnilfklkgkegiirienndfleteevvlyieaygks eklkalgitkkkiideairqgitkddkkieikrqeneeeieidirdeytnktlndcsiilri iendeletkksiyeifkninmslykiiekiienetekvfenryyeehlrekllkddkid viltnfmeirekiksnleilgfvkfylnvggdkkksknkkmlvekilninvdltvedi adfvikelefwnitkriekvkkvnneflekrrnrtyiksyvlldkhekfkierenkkd kivkffveniknnsikekiekilaefkidelikklekelkkgncdteifgifkkhykvn fdskkfskksdeekelykiiyrylkgriekilvneqkvrlkkmekieiekilnesilse kilkrvkqytlehimylgklrhndidmttvntddfsrlhakeeldlelitffastnmeln kifsreninndenidffggdreknyvldkkilnskikiirdldfidnknnitnnfirkftk igtnernrilhaiskerdlqgtqddynkviniiqnlkisdeevskalnldvvfkdkknii tkindikiseennndikylpsfskvlpeilnlyrnnpknepfdtietekivlnaliyvnk elykklileddleeneskniflqelkktlgnideideniienyyknaqisaskgnnkai kkyqkkviecyigylrknyeelfdfsdfkmniqeikkqikdindnktyeritvktsd ktivinddfeyiisifallnsnavinkirnrffatsvwlntseyqniidildeimqlntlrn ecitenwnlnleefiqkmkeiekdfddfkiqtkkeifnnyyediknniltefkdding cdvlekklekivifddetkfeidkksnilqdeqrklsninkkdlkkkvdqyikdkdq eikskilcriifnsdflkkykkeidnliedmesenenkfqeiyypkerknelyiykkn lflnignpnfdkiyglisndikmadakflfnidgknirknkiseidailknlndklngy skeykekyikklkenddffakniqnknyksfekdynrvseykkirdlvefnylnki esylidinwklaiqmarferdmhyivnglrelgiiklsgyntgisraypkrngsdgfy tttayykffdeesykkfekicygfgidlsenseinkpenesirnyishfyivrnpfady siaeqidrvsnllsystrynnstyasvfevfkkdvnldydelkkkfklignndilerlm kpkkvsvlelesynsdyiknliielltkientndtl

c2c2-3		L wadei (Lw2) (SEQ ID No. 47)	mkvtkvdgishkkyieegklvkstseenrtserlsellsirldiyiknpdnaseeenrir renlkkffsnkvlhlkdsvlylknrkeknavqdknyseediseydlknknsfsvlkk illnedvnseeleifrkdveaklnkinslkysfeenkanyqkinennvekvggkskr niiydyyresakrndyinnvqeafdklykkedieklfflienskkhekykireyyhki igrkndkenfakiiyeeiqnvnnikeliekipdmselkksqvfykyyldkeelndkn ikyafchfveiemsqllknyvykrlsnisndkikrifeyqnlkklienkllnkldtyvr ncgkynyylqvgeiatsdfiarnrqneaflrniigvssvayfslrniletenenditgrm rgktvknnkgeekyvsgevdkiynenkqnevkenlkmfysydfnmdnkneied ffanideaissirhgivhfnlelegkdifafkniapseiskkmfqneinekklklkifkq lnsanvfnyyekdviikylkntkfnfvnknipfvpsftklynkiedlrntlkffwsvp kdkeekdaqiyllkniyygeflnkfvknskvffkitnevikinkqrnqktghykyqk feniektvpveylaiiqsreminnqdkeekntyidfiqqiflkgfidylnknnlkyies nnnndnndifskikikkdnkekydkilknyekhnrnkeipheinefvreiklgkilk ytenlnmfylilkllnhkeltnlkgslekyqsankeetfsdelelinllnldnnrvtedfe leaneigkfldfnenkikdrkelkkfdtnkiyfdgeniikhrafynikkygmlnlleki adkakykislkelkeysnkkneieknytmqqnlhrkyarpkkdekfndedykeye kaigniqkythlknkvefnelnllqglllkilhrlvgytsiwerdlrfrlkgefpenhyie eifnfdnsknvkyksgqivekyinfykelykdnvekrsiysdkkvkklkqekkdly irnyiahfnyiphaeisllevlenlrkllsydrklknaimksivdilkeygfvatfkiga dkkieiqtlesekivhlknlkkkklmtdrnseelcelvkvmfeykalekrpaatkka gqakkkkgsypydvpdyaypydvpdyaypydvpdya*
c2c2-4		Listeria seeligeri (SEQ ID No. 48)	mwisiktlihhlgvlffcdymynrrekkiievktmritkvevdrkkvlisrdknggkl vyenemqdnteqimhhkkssfyksvvnkticrpeqkqmkklvhgllqensqeki kvsdvtklnisnflnhrfkkslyyfpenspdkseeyrieinlsqlledslkkqqgtfic wesfskdmelyinwaenyissktklikksirnnriqstesrsgqlmdrymkdilnkn kpfdiqsvsekyqlekltsalkatfkeakkndkeinyklkstlqnherqiieelkense lnqfnieirkhletyfpikktnrkvgdirnleigeiqkivnhrlknkivqrilqegklasy eiestvnsnslqkikieeafalkfinaclfasnnlrnmvypvckkdilmigefknsfk eikhkkfirqwsqffsqeitvddielaswglrgaiapirneiihlkkhswkkffnnptf kvkkskiingktkdvtseflyketlfkdyfyseldsvpeliinkmesskildyyssdql nqvftipnfelslltsavpfapsfkrvylkgfdyqnqdeaqpdynlklniynekafns eafqaqyslfkmvyyqvflpqfttnndlfkssvdfiltlnkerkgyakafqdirkmn

			kdekpseymsyiqsqlmlyqkkqeekekinhfekfinqvfikgfnsfieknrltyic hptkntvpendnieipfhtdmddsniafwlmcklldakqlselrnemikfscslqst eeistftkareviglallngekgcndwkelfddkeawkknmslyvseellqslpytq edgqtpvinrsidlvkkygtetileklfsssddykvsakdiaklheydvtekiaqqesl hkqwiekpglardsawtkkyqnvindisnyqwaktkveltqvrhlhqltidllsrla gymsiadrdfqfssnyilerenseyrvtswillsenknknkyndyelynlknasikv sskndpqlkvdlkqlrltleylelfdnrlkekrnnishfnylngqlgnsilelfddardvl sydrklknavskslkeilsshgmevtfkplyqtnhhlkidklqpkkihhlgekstvss nqvsneycqlvrtlltmk

C2-17		Leptotrichia buccalis C-1013-b (SEQ ID No. 49)	mkvtkvggishkkytsegrlvkseseenrtderlsallnmrldmyiknpsstetken qkrigklkkffsnkmvylkdntlslkngkkenidreysetdilesdvrdkknfavlkk iylnenvnseelevfrndikkklnkinslkysfeknkanyqkinenniekvegkskr niiydyyresakrdayvsnvkeafdklykeediaklvleienltklekykirefyheii grkndkenfakiiyeeiqnvnnmkeliekvpdmselkksqvfykyyldkeelndk nikyafchfveiemsqllknyvykrlsnisndkikrifeyqnlkklienkllnkldtyv rncgkynyylqdgeiatsdfiarnrqneaflrniigvssvayfslrniletenenditgr mrgktvknnkgeekyvsgevdkiynenkknevkenlkmfysydfnmdnknei edffanideaissirhgivhfnlelegkdifafkniapseiskkmfqneinekklklkif rqlnsanvfrylekykilnylkrtrfefvnknipfvpsftklysriddlknslgiywktp ktnddnktkeiidaqiyllkniyygeflnyfmsnngnffeiskeiielnkndkrnlktg fyklqkfediqekipkeylaniqslyminagnqdeeekdtyidfiqkiflkgfmtyla nngrlsliyigsdeetntslaekkqefdkflkkyeqnnnikipyeineflreiklgnilk yterlnmfylilkllnhkeltnlkgslekyqsankeeafsdqlelinllnldnnrvtedfe leadeigkfldfngnkvkdnkelkkfdtnkiyfdgeniikhrafynikkygmlnlle kiadkagykisieelkkysnkkneieknhkmqenlhrkyarprkdekftdedyesy kqaienieeythlknkvefnelnllqglllrilhrlvgytsiwerdlrfrlkgefpenqyi eeifnfenkknvkykggqivekyikfykelhqndevkinkyssanikvlkqekkdl yirnyiahfnyiphaeisllevlenlrkllsydrklknavmksvvdilkeygfvatfki gadkkigiqtlesekivhlknlkkkklmtdrnseelcklvkimfeykmeekksen
C2-18		Herbinix hemicellulo	mkltrrrisgnsvdqkitaafyrdmsqgllyydsedndctdkviesmdferswrgril kngeddknpfymfvkglvgsndkivcepidvdsdpdnldilinknltgfgrnlkap

		silytica (SEQ ID No. 50)	dsndtlenlirkiqagipeeevlpelkkikemiqkdivnrkeqllksiknnripfslegs klvpstkkmkwlfklidvpnktfnekmlekyweiydydklkanitnrldktdkkar sisravseelreyhknlrtnynrfvsgdrpaagldnggsakynpdkeefllflkeveq yfkkyfpvkskhsnkskdkslvdkyknycsykvvkkevnrsiinqlvagliqqgkl lyyfyyndtwqedflnsyglsyiqveeafkksvmtslswginrltsffiddsntvkfd dittkkakeaiesnyfnklrtcsrmqdhfkeklaffypvyvkdkkdrpdddienlivl vknaiesvsylrnrtfhfkessllellkelddknsgqnkidysvaaefikrdienlydvf reqirslgiaeyykadmisdcfktcglefalyspknslmpafknvykrganlnkayir dkgpketgdqgqnsykaleeyreltwyievknndqsynayknllqliyyhaflpev renealitdfinrtkewnrketeerlntknnkkhknfdendditvntyryesipdyqg eslddylkvlqrkqmarakevnekeegnnnyiqfirdvvvwafgaylenklknyk nelqpplskeniglndtlkelfpeekvkspfnikcrfsistfidnkgkstdntsaeavkt dgkedekdkknikrkdllcfylflrlldeneicklqhqfikyrcslkerrfpgnrtklek etellaeleelmelvrftmpsipeisakaesgydtmikkyfkdfiekkvfknpktsnl yyhsdsktpvtrkymallmrsaplhlykdifkgyylitkkecleyiklsniikdyqns lnelheqleriklksekqngkdslyldkkdfykvkeyvenleqvarykhlqhkinfe slyrifrihvdiaarmvgytqdwerdmhflfkalvyngvleerrfeaifnnnddnnd grivkkiqnnlnnknrelvsmlcwnkklnknefgaiiwkrnpiahlnhftqteqns kssleslinslrillaydrkrqnavtktindlllndyhirikwegrvdegqiyfnikeked ienepiihlkhlhkkdcyiyknsymfdkqkewicngikeevydksilkcignlfkf dyedknkssanpkht
C2-19		[Eubacterium] rectale (SEQ ID No. 51)	mlrrdkevkklynvfnqiqvgtkpkkwnndeklspeenerraqqknikmknyk wreacskyvessqriindvifysyrkaknklrymrknedilkkmqeaeklskfsgg kledfvaytlrkslvvskydtqefdslaamvvflecigknnisdhereivckllelirkd fskldpnvkgsqganivrsvrnqnmivqpqgdrflfpqvyakenetvtnknveke glnefllnyanlddekraeslrklrrildvyfsapnhyekdmditlsdniekekfnvw ekhecgkketglfvdipdvlmeaeaenikldavvekrerkvlndrvrkqniicyryt ravvekynsneplffennainqywihhienaverilknckagklfklrkgylaekv wkdainlisikyialgkavynfalddiwkdkknkelgivderirngitsfdyemika henlqrelavdiafsvnnlaravcdmsnlgnkesdfllwkrndiadklknkddmas vsavlqffggksswdinifkdaykgkkkynyevrfiddlrkaiycarnenfhfktal vndekwntelfgkiferetefclnvekdrfysnnlymfyqvselrnmldhlysrsvsr

			aaqvpsynsvivrtafpeyitnvlgyqkpsydadtlgkwysacyyllkeiyynsflq sdralqlfeksvktlswddkkqqravdnfkdhfsdiksactslaqvcqiymteynqq nnqikkvrssndsifdqpvyqhykvllkkaianafadylknnkdlfgfigkpfkane ireidkeqflpdwtsrkyealcievsgsqelqkwyivgkflnarslnlmvgsmrsyi qyvtdikrraasignelhvsvhdvekvekwvqvievcsllasrtsnqfedyfndkdd yarylksyvdfsnvdmpseysalvdfsneeqsdlyvdpknpkvnrnivhsklfaa dhilrdivepvskdnieefysqkaeiayckikgkeitaeeqkavlkyqklknrvelrd iveygeiinellgqlinwsfmrerdllyfqlgfhydclrndskkpegyknikvdensi kdailyqiigmyvngvtvyapekdgdklkeqcvkggvgvkvsafhryskylglne ktlynagleifevvaehediinlrngidhfkyylgdyrsmlsiysevfdrfftydikyq knvlnllqnillrhnvivepilesgfktigeqtkpgaklsirsiksdtfqykvkggtlitd akderyletirkilyyaeneednlkksvvvtnadkyeknkesddqnkqkekknkd nkgkkneetksdaeknnnerlsynpfanlnfklsn
C2-20		Eubacteriaceae bacterium CHKCI004 (SEQ ID No. 52)	mkiskeshkrtavavmedrvggvvyvpggsgidlsnnlkkrsmdtkslynvfnqi qagtapseyewkdylseaenkkreaqkmiqkanyelrrecedyakkanlavsriif skkpkkifsdddiishmkkqrlskfkgrmedfvlialrkslvvstynqevfdsrkaat vflknigkknisadderqikqlmaliredydkwnpdkdssdkkessgtkvirsiehq nmviqpeknklslskisnvgkktktkqkekagldaflkeyaqidensrmeylkklrr lldtyfaapssyikgaavslpeninfsselnvwerheaakkvninfveipesllnaeq nnnkinkveqehsleqlrtdirrrnitcyhfanalaaderyhtlffenmamnqfwihh menaverilkkcnvgtlfklrigylsekvwkdmlnllsikyialgkavyhfalddiw kadiwkdasdknsgkindltlkgissfdyemvkaqedlqremavgvafstnnlarv tckmddlsdaesdfllwnkeairrhvkytekgeilsailqffggrslwdeslfekaysd snyelkflddlkraiyaarnetfhfktaaidggswntrlfgslfekeaglclnveknkfy snnlvlfykqedlrvfldklygkecsraaqipsyntilprksfsdfmkqllglkepvyg saildqwysacyylfkevyynlflqdssakalfekavkalkgadkkqekavesfrkr yweisknaslaeicqsyiteynqqnnkerkvrsandgmfnepiyqhykmllkeal kmafasyikndkelkfvykpteklfevsqdnflpnwnsekyntlisevknspdlqk wyivgkfmnarmlnlllgsmrsylqyvsdiqkraaglgenqlhlsaenvgqvkkw iqvlevcllsvrisdkftdyfkdeeeyasylkeyvdfedsampsdysallafsnegki dlyvdasnpkvnrniiqaklyapdmvlkkvvkkisqdeckefnekkeqimqfkn kgdevsweeqqkileyqklknrvelrdlseygelinellgqlinwsylrerdllyfqlg

			fhysclmneskkpdayktirrgtvsienavlyqiiamyingfpvyapekgelkpqc ktgsagqkirafcqwasmvekkkyelynaglelfevvkehdniidlrnkidhfkyy qgndsilalygeifdrfftydmkyrnnvlnhlqnillrhnviikpiiskdkkevgrgk mkdraaflleevssdrftykvkegerkidaknrlyletvrdilyfpnravndkgedvii cskkaqdlnekkadrdknhdkskdtnqkkegknqeeksenkepysdrmtwkpf agikle
C2-21		Blautia sp. Marseille-P2398 (SEQ ID No. 53)	mkiskvdhvksgidqklssqrgmlykqpqkkyegkqleehvrnlsrkakalyqvf pvsgnskmekelqiinsfiknillrldsgktseeivgyintysvasqisgdhiqelvdq hlkeslrkytcvgdkriyvpdiivallkskfnsetlqydnselkilidfiredylkekqik qivhsiennstplriaeingqkrlipanvdnpkksyifeflkeyaqsdpkgqesllqh mrylillylygpdkitddyceeieawnfgsivmdneqlfseeasmliqdriyvnqqi eegrqskdtakvkknkskyrmlgdkiehsinesvvkhyqeackaveekdipwik yisdhvmsvyssknrvdldklslpylakntwntwisfiamkyvdmgkgvyhfa msdvdkvgkqdnliigqidpkfsdgissfdyerikaeddlhrsmsgyiafavnnfar aicsdefrkknrkedvltvgldeiplydnvkrkllqyfggasnwddsiidiiddkdlv acikenlyvarnvnfhfagsekvqkkqddileeivrketrdigkhyrkvfysnnvav fycdediiklmnhlyqrekpyqaqipsynkvisktylpdlifmllkgknrtkisdpsi mnmfrgtfyfllkeiyyndflqasnlkemfceglknnvknkksekpyqnfmrrfe elenmgmdfgeicqqimtdyeqqnkqkkktatavmsekdkkirtldndtqkykh frtllyiglreafiiylkdeknkewyeflrepvkreqpeekefvnkwklnqysdcseli lkdslaaawyvvahfinqaqlnhligdiknyiqfisdidrrakstgnpvsesteiqier yrkilrvlefakffcgqitnvltdyyqdendfsthvghyvkfekknmepahalqafs nslyacgkekkkagfyydgmnpivnrnitlasmygnkkllenamnpvteqdirk yyslmaeldsvlkngavcksedeqknlrhfqnlknrielvdvltlselvndlvaqlig wvyirerdmmylqlglhyiklyftdsvaedsylrtldleegsiadgavlyqiaslysfn lpmyvkpnkssvyckkhvnsvatkfdifekeycngdetvienglrlfeninlhkdm vkfrdylahfkyfakldesilelyskaydfffsyniklkksvsyvltnvllsyfinaklsf stykssgnktvqhrttkisvvaqtdyftyklrsivknkngvesienddrrcevvniaar dkefvdevcnvinynsdk
C2-22		Leptotrichia sp. oral taxon 879	mgnlfghkrwyevrdkkdfkikrkvkvkrnydgnkyilninennnkekidnnkfi gefvnykknnnvlkefkrkfhagnilfklkgkeeiirienndfleteevvlyievyg kseklkaleitkkkiideairqgitkddkkieikrqeneeeieidirdeytnktlndcsiil

		str. F0557 (SEQ ID No. 54)	riiendeletkksiyeifkninmslykiiekiienetekvfenryyeehlrekllkdnki dviltnfmeirekiksnleimgfvkfylnvsgdkkksenkkmfvekilntnvdltve divdfivkelkfwnitkriekvkkfnneflenrrnrtyiksyvlldkhekfkierenkk dkivkffveniknnsikekiekilaefkinelikklekelkkgncdteifgifkkhykv nfdskkfsnksdeekelykiiyrylkgriekilvneqkvrlkkmekieiekilnesils ekilkrvkqytlehimylgklrhndivkmtvntddfsrlhakeeldlelitffastnme lnkifngkekvtdffgfnlngqkitlkekvpsfklnilkklnfinnennideklshfysf qkegyllrnkilhnsygniqetknlkgeyenveklikelkvsdeeiskslsldvifegk vdiinkinslkigeykdkkylpsfskivleitrkfreinkdklfdiesekiilnavkyvn kilyekitsneeneflktlpdklvkksnnkkenknllsieeyyknaqvssskgdkkai kkyqnkvtnayleylentfteiidfskfnlnydeiktkieerkdnkskiiidsistninit ndieyiisifallnsntyinkirnrffatsvwlekqngtkeydyeniisildevllinllre nnitdildlknaiidakivendetyiknyifesneeklkkrlfceelvdkedirkifede nfkfksfikkneignfkinfgilsnlecnseveakkiigknskklesfiqniideyksni rtlfsseflekykeeidnlvedtesenknkfekiyypkehknelyiykknlflnignpn fdkiygliskdiknvdtkilfdddikknkiseidailknlndklngysndykakyvnk lkenddffakniqnenyssfgefekdynkvseykkirdlvefnylnkiesylidinw klaiqmarferdmhyivnglrelgiiklsgyntgisraypkrngsdgfytttayykffd eesykkfekicygfgidlsenseinkpenesirnyishfyivrnpfadysiaeqidrvs nllsystrynnstyasvfevfkkdvnldydelkkkfrlignndilerlmkpkkvsvlel esynsdyiknliielltkientndtl
C2-23		Lachnospiraceae bacterium NK4A144 (SEQ ID No. 55)	mkiskvdhtrmavakgnqhrrdeisgilykdptktgsidfderfkklncsakilyhv fngiaegsnkyknivdkvnnnldrvlftgksydrksiididtvlrnvekinafdriste ereqiiddlleiqlrkglrkgkaglrevlligagvivrtdkkqeiadfleildedfnktnq akniklsienqglvvspvsrgeerifdvsgaqkgksskkaqekealsaflldyadldk nvrfeylrkirrlinlyfyvknddvmslteipaevnlekdfdiwrdheqrkeengdfv gcpdilladrdvkksnskqvkiaerqlresireknikryrfsiktiekddgtyffankqi svfwihrienaverilgsindkklyrlrlgylgekvwkdilnflsikyiavgkavfnfa mddlqekdrdiepgkisenavngltsfdyeqikademlqrevavnvafaannlarv tvdipqngekedillwnksdikkykknskkgilksilqffggastwnmkmfeiayh dqpgdyeenylydiiqiiyslrnksfhfktydhgdknwnreligkmiehdaervisv erekfhsnnlpmfykdadlkkildllysdyagrasqvpafntvlvrknfpeflrkdm

			gykvhfnnpevenqwhsavyylykeiyynlflrdkevknlfytslknirsevsdkk qklasddfasrceeiedrslpeicqiimteynaqnfgnrkvksqrvieknkdifrhyk mlliktlagafslylkqerfafigkatpipyettdvknflpewksgmyasfveeiknnl dlqewyivgrflngrmlnqlagslrsyiqyaedierraaenrnklfskpdekieackk avrvldlcikistrisaeftdyfdseddyadylekylkyqddaikelsgssyaaldhfc nkddlkfdiyvnagqkpilqrnivmaklfgpdnilsevmekvtesaireyydylkk vsgyrvrgkcstekeqedllkfqrlknavefrdvteyaevinellgqliswsylrerdll yfqlgfhymclknksfkpaeyvdirrnngtiihnailyqivsmyingldfyscdkeg ktlkpietgkgvgskigqfikysqylyndpsykleiynaglevfenidehdnitdlrk yvdhfkyyaygnkmslldlyseffdrfftydmkyqknvvnvlenillrhfvifypkf gsgkkdvgirdckkeraqieiseqsltsedfmfkldkageeakkfparderylqtia kllyypneiedmnrfmkkgetinkkvqfnrkkkitrkqknnssnevlsstmgylfk nikl
C2-24		Chloroflexus aggregans (SEQ ID No. 56)	mtdqvrreevaageladtplaaaqtpaadaavaatpapaeavaptpeqavdqpattg eseapvttaqaaaheaepaeatgasftpvseqqpqkprrlkdlqpgmelegkvtsial ygifvdvgvgrdglvhisemsdrridtpselvqigdtvkvwvksvdldarrisltml npsrgekprrsrqsqpaqpqprrqevdreklaslkvgeivegvitgfapfgafadigv gkdglihiselsegrvekpedavkvgeryqfkvleidgegtrislslrraqrtqrmqql epgqiiegtvsgiatfgafvdigvgrdglvhisalaphrvakvedvvkvgdkvkvk vlgvdpqskrisltmrleeeqpattagdeaaepaeevtptrrgnlerfaaaaqtarerse rgersergerrerrerrpaqsspdtyivgedddesfegnatiedlltkfggsssrrdrdrr rrhedddeemerpsnrrqreairrtlqqigyde
C2-25		Demequina aurantiaca (SEQ ID No. 57)	mdltwhallilfivallagfldtlaggggllvpallltgipplqalgtnklqssfgtgmat yqvirkkrvhwrdvrwpmvwaflgsaagavavqfidtdalliiipvvlalvaayflf vpkshlpppeprmsdpayeatlvpiigaydgafgpgtgslyalsgvalraktlvqsta iaktlnfatnfaallvfafaghmlwtvgavmiagqligayagshmlfrvnplvlrvli vvmslgmlirvlld
C2-26		Thalassospira sp. TSL5-1	mriikpygrshvegvatqeprrklrlnsspdisrdipgfaqshdaliiaqwisaidkiat kpkpdkkptqaqinlrttlgdaawqhvmaenllpaatdpaireklhliwqskiapw gtarpqaekdgkptpkggwyerfcgvlspeaitqnvarqiakdiydhlhvaakrkg repakqgessnkpgkfkpdrkrglieeraesiaknalrpgshapcpwgpddqatye

		(SEQ ID No. 58)	qagdvagqiyaaardcleekkrrsgnrntssvqylprdlaakilyaqygrvfgpdtti kaaldeqpslfalhkaikdcyhrlindarkrdilrilprnmaalfrlvraqydnrdinali rlgkvihyhaseqgksehhgirdywpsqqdiqnsrfwgsdgqadikrheafsriwr hiialasrtlhdwadphsqkfsgenddilllakdaieddvfkaghyerkcdvlfgaqa slfcgaedfekailkqaitgtgnlrnatfhfkgkvrfekelqeltkdvpvevqsaiaal wqkdaegrtrqiaetlqavlaghflteeqnrhifaaltaamaqpgdvplprlrrvlarh dsicqrgrilplspcpdrakleespaltcqytvlkmlydgpfrawlaqqnstilnhyid stiartdkaardmngrklaqaekdlitsraadlprlsvdekmgdflarltaatatemrv qrgyqsdgenaqkqaafigqfecdvigrafadflnqsgfdfvlklkadtpqpdaaqc dvtaliapddisvsppqawqqvlyfilhlvpvddashllhqirkwqvlegkekpaqi ahdvqsvlmlyldmhdakftggaalhgiekfaeffahaadfravfppqslqdqdrsi prrglreivrfghlpllqhmsgtvqithdnvvawqaartagatgmspiarrqkqreel halavertarfrnadlqnymhalvdvikhrqlsaqvtlsdqvrlhrlmmgvlgrlvd yaglwerdlyfvvlallyhhgatpddvfkgqgkknladgqvvaalkpknrkaaap vgvfddldhygiyqddrqsirnglshfnmlrggkapdlshwvnqtrslvahdrklk navaksviemlaregfdldwgiqtdrgqhilshgkirtrqaqhfqksrlhivkksakp dkndtvkirenlhgdamvervvqlfaaqvqkryditvekrldhlflkpqdqkgkng ihthngwsktekkrrpsrenrkgnhen
C2-27		SAMN04487830_13920 [Pseudobutyrivibrio sp. OR37] (SEQ ID No. 59)	mkfskeshrktavgvtesngiigllykdplnekekiedvvnqranstkrlfnlfgteat skdisraskdlakvvnkaignlkgnkkfnkkeqitkglntkiiveelknvlkdekkli vnkdiideacsrllktsfrtaktkqavkmiltavlientnlskedeafvheyfvkklvne ynktsvkkqipvalsnqnmviqpnsvngtleisetkksketkttekdafraflrdyatl denrrhkmrlclrnlvnlyfygetsvskddfdewrdhedkkqndelfvkkivsiktd rkgnvkevldvdatidairtnniacyrralayanenpdvffsdtmlnkfwihhvene veriyghinnntgdykyqlgylsekvwkgiinylsikyiaegkavynyamnalak dnnsnafgkldekfvngitsfeyerikaeetlqrecavniafaanhlanatvdlnekds dflllkhednkdtlgavarpnilrnilqffggksrwndfdfsgideiqlldlrkmiysl rnssfhfktenidndswntkligdmfaydfnmagnvqkdkmysnnvpmfysts diekmldrlyaevherasqvpsfnsvfvrknfpdylkndlkitsafgvddalkwqsa vyyvckeiyyndflqnpetftmlkdyvqclpididksmdqklksernahknfkeaf atyckecdslsaicqmimteynnqnkgnrkvisartkdgdkliykhykmilfealk nvftiyleknintygflkkpklinnvpaieeflpnyngrqyetlvnriteetelqkwyi

			vgrllnpkqvnqlignfrsyvqyvndvarrakqtgnnlsndniawdvkniiqifdvc tklngvtsniledyfddgddyarylknfvdytnknndhsatllgdfcakeidgikigi yhdgtnpivnrniiqcklygatgiisdltkdgsilsvdyeiikkymqmqkeikvyqq kgicktkeeqqnlkkyqelknivelrniidyseildelqgqlinwgylrerdlmyfql gfhylclhneskkpvgynnagdisgavlyqivamytnglslidangkskknakasa gakvgsfcsyskeirgvdkdtkedddpiylagvelfeninehqqcinlrnyiehfhy yakhdrsmldlysevfdrfftydmkytknvpnmmynillqhlvvpafefgssekrl ddndeqtkpramftlrekngllsseqftyrlgdgnstvklsargddylravasllyypdr apeglirdaeaedkfakinhsnpksdnrnnrgnfknpkvqwynnktkrk
C2-28		SAMN029 10398_000 08 [Butyrivibri o sp. YAB3001] (SEQ ID No. 60)	mkiskvdhrktavkitdnkgaegfiyqdptrdsstmeqiisnrarsskvlfnifgdtk kskdlnkytesliiyvnkaikslkgdkrnnkyeeiteslktervlnaliqagneftcsen niedalnkylkksfrvgntksalkkllmaaycgyklsieekeeiqnyfvdklvkeyn kdtvlkytakslkhqnmvvqpdtdnhvflpsriagatqnkmsekealteflkayavl deekrhnlriilrklvnlyfyespdfiypennewkehddrknktetfvspvkvneek ngktfvkidvpatkdlirlkniecyrrsvaetagnpityftdhniskfwihhienevek ifallksnwkdyqfsvgyisekvwkeiinylsikyiaigkavynyaledikkndgtl nfgvidpsfydginsfeyekikaeetfqrevavyvsfavnhlssatvklseaqsdmlv lnkndiekiaygntkrnilqffggqskwkefdfdryinpvnytdidflfdikkmvysl rnesfhftttdtesdwnknlisamfeyecrristvqknkffsnnlplfygenslervlhk lyddyvdrmsqvpsfgnvfvrkkfpdymkeigikhnlssednlklqgalyflykei yynafissekamkifvdlvnkldtnarddkgritheamahknfkdaishymthdcs ladicqkimteynqqntghrkkqttysseknpeifrhykmilfmllqkamteyisse eifdfimkpnspktdikeeeflpqykscaydnlikliadnvelqkwyitarllsprev nqligsfrsykqfvsdierraketnnslsksgmtvdvenitkvldlctklngrfsneltd yfdskddyavyvskfldfgfkidekfpaallgefcnkeengkkigiyhngtepilns niiksklygitdvvsravkpvseklireylqqevkikpylengvcknkeeqaalrky qelknriefrdiveyseiinelmgqlinfsylrerdlmyfqlgfhylclnnygakpegy ysivndkrtikgailyqivamytyglpiyhyvdgtisdrrknkktvldtlnssetvgak ikyfiyysdelfndslilynaglelfenineheniynlrkyidhfkyyvsqdrslldiys evfdryftydrkykknvmnlfsnimlkhfiitdfefstgektigekntakkecakvri krgglssdkftykfkdakpielsaknteftdgvarilyypenvvltdlvrnsevedekr

			iekydrnhnssptrkdktykqdvkknynkktskafdsskldtksvgnnlsdnpvlk qflseskkkr
C2-29		Blautia sp. Marseille-P2398 (SEQ ID No. 61)	mkiskvdhvksgidqklssqrgmlykqpqkkyegkqleehvrnlsrkakalyqvf pvsgnskmekelqiinsfiknillrldsgktseeivgyintysvasqisgdhiqelvdq hlkeslrkytcvgdkriyvpdiivallkskfnsetlqydnselkilidfiredylkekqik qivhsiennstplriaeingqkrlipanvdnpkksyifeflkeyaqsdpkgqesllqh mrylillylygpdkitddyceeieawnfgsivmdneqlfseeasmliqdriyvnqqi eegrqskdtakvkknkskyrmlgdkiehsinesvvkhyqeackaveekdipwik yisdhvmsvyssknrvdldklslpylakntwntwisfiamkyvdmgkgvyhfa msdvdkvgkqdnliigqidpkfsdgissfdyerikaeddlhrsmsgyiafavnnfar aicsdefrkknrkedvltvgldeiplydnvkrkllqyfggasnwddsiidiiddkdlv acikenlyvarnvnfhfagsekvqkkqddileeivrketrdigkhyrkvfysnnvav fycdediiklmnhlyqrekpyqaqipsynkvisktylpdlifmllkgknrtkisdpsi mnmfrgtfyfllkeiyyndflqasnlkemfceglknnvknkksekpyqnfmrrfe elenmgmdfgeicqqimtdyeqqnkqkkktatavmsekdkkirtldndtqkykh frtllyiglreafiiylkdeknkewyeflrepvkreqpeekefvnkwklnqysdcseli lkdslaaawyvvahfinqaqlnhligdiknyiqfisdidrrakstgnpvsesteiqier yrkilrvlefakffcgqitnvltdyyqdendfsthvghyvkfekknmepahalqafs nslyacgkekkkagfyydgmnpivnrnitlasmygnkkllenamnpvteqdirk yyslmaeldsvlkngavcksedeqknlrhfqnlknrielvdvltlselvndlvaqlig wvyirerdmmylqlglhyiklyftdsvaedsylrtldleegsiadgavlyqiaslysfn lpmyvkpnkssvyckkhvnsvatkfdifekeycngdetvienglrlfeninlhkdm vkfrdylahfkyfakldesilelyskaydfffsyniklkksvsyvltnvllsyfinaklsf stykssgnktvqhrttkisvvaqtdyftyklrsivknkngvesienddrrcevvniaar dkefvdevcnvinynsdk
C2-30		Leptotrichia sp. Marseille-P3007 (SEQ ID No. 62)	mkitkidgishkkyikegklvkstseenktderlselltirldtyiknpdnaseeenrirr enlkeffsnkvlylkdgilylkdrreknqlqnknyseediseydlknknnflvlkkill nedinseeleifrndfekkldkinslkysleenkanyqkinennikkvegkskrnify nyykdsakrndyinniqeafdklykkedienlfflienskkhekykirecyhkiigrk ndkenfatiiyeeiqnvnnmkeliekvpnvselkksqvfykyylnkeklndeniky vfchfveiemskllknyvykkpsnisndkvkrifeyqslkklienkllnkldtyvrnc gkysfylqdgeiatsdfivgnrqneaflrniigvsstayfslrniletenenditgrmrg

			ktvknnkgeekyisgeidklydnnkqnevkknlkmfysydfnmnskkeiedffs nideaissirhgivhfnlelegkdiftfknivpsqiskkmfhdeinekklklkifkqlns anvfrylekykilnylnrtrfefvnknipfvpsftklysriddlknslgiywktpktndd nktkeitdaqiyllkniyygeflnyfmsnngnffeitkeiielnkndkrnlktgfyklq kfenlqektpkeylaniqslyminagnqdeeekdtyidfiqkiflkgfmtylanngrl sliyigsdeetntslaekkqefdkflkkyeqnnnieipyeinefvreiklgkilkyterl nmfylilkllnhkeltnlkgslekyqsankeeafsdqlelinllnldnnrvtedfelead eigkfldfngnkvkdnkelkkfdtnkiyfdgeniikhrafynikkygmlnllekisd eakykisieelknyskkkneieenhttqenlhrkyarprkdekftdedykkyekair niqqythlknkvefnelnllqslllrilhrlvgytsiwerdlrfrlkgefpenqyieeifnf dnsknvkykngqivekyinfykelykddtekisiysdkkvkelkkekkdlyirnyi ahfnyipnaeisllemlenlrkllsydrklknaimksivdilkeygfvvtfkiekdkki rieslkseevvhlkklklkdndkkkepiktyrnskelcklvkvmfeykmkekksen
C2-31		Bacteroides ihuae (SEQ ID No. 63)	mritkvkvkessdqkdkmvlihrkvgegtlvldenladltapiidkykdksfelsllk qtlvsekemnipkcdkctakerclsckqrekrlkevrgaiektigaviagrdiiprlnif nedeicwlikpklrneftfkdvnkqvvklnlpkvlveyskkndptlflayqqwiaay lknkkghikksilnnrvvidysdesklskrkqalelwgeeyetnqrialesyhtsyni gelvtllpnpeeyvsdkgeirpafhyklknvlqmhqstvfgtneilcinpifnenrani qlsaynlevvkyfehyfpikkkkknlslnqaiyylkvetlkerlslqlenalrmnllqk gkikkhefdkntcsntlsqikrdeffvlnlvemcafaannirnivdkeqvneilskkd lcnslskntidkelctkfygadfsqipvaiwamrgsvqqirneivhykaeaidkifal ktfeyddmekdysdtpfkqylelsiekidsffieqlssndvlnyyctedvnkllnkck lslrrtsipfapgfktiyelgchlqdssntyrighylmliggrvanstvtkaskaypayrf mlkliynhlflnkfldnhnkrffmkavafvlkdnrenarnkfqyafkeirmmnnde siasymsyihslsvqeqekkgdkndkvryntekfiekvfvkgfddflswlgvefils pnqeerdktvtreeyenlmikdrvehsinsnqeshiafftfcklldanhlsdlrnewik frssgdkegfsynfaidiielclltvdrveqrrdgykeqtelkeylsffikgnesentvw kgfyfqqdnytpvlyspielirkygtlellkliivdedkitqgefeewqtlkkvvedkv trrnelhqewedmknkssfsqekcsiyqklcrdidrynwldnklhlvhlrklhnlvi qilsrmarfialwdrdfvlldasranddykllsffnfrdfinakktktddellaefgskie kknapfikaedvplmvecieakrsfyqkvffrnnlqvladrnfiahynyisktakcsl

			femiiklrtlmyydrklrnavvksianvfdqngmvlqlslddshelkvdkviskriv hlknnnimtdqvpeeyykicrrllemkk
C2-32		SAMN05216357_1045 [Porphyromonadaceae bacterium KH3CP3RA] (SEQ ID No. 64)	mefrdsifksllqkeiekaplcfaeklisggvfsyypserlkefvgnhpfslfrktmpf spgfkrvmksggnyqnanrdgrfydldigvylpkdgfgdeewnaryflmkliyn qlflpyfadaenhlfrecvdfvkrvnrdyncknnnseeqafidirsmredesiadyla fiqsniiieenkkketnkegqinfnkfllqvfvkgfdsflkdrtelnflqlpelqgdgtrg ddlesldklgavvavdlkldatgidadlnenisfytfcklldsnhlsrlrneiikyqsans dfshnedfdydriisiielcmlsadhvstndnesifpnndkdfsgirpylstdakvetf edlyvhsdaktpitnatmvlnwkygtdklferlmisdqdflvtekdyfvwkelkkd ieekiklreelhslwvntpkgkkgakkkngrettgefseenkkeylevcreidryvnl dnklhfvhlkrmhslliellgrfvgftylferdyqyyhleirsrrnkdagvvdkleynk ikdqnkydkddffactflyekankvrnfiahfnyltmwnspqeeehnsnlsgakns sgrqnlkcsltelinelrevmsydrklknavtkavidlfdkhgmvikfrivnnnnnd nknkhhlelddivpkkimhlrgiklkrqdgkpipiqtdsvdplycrmwkklldlkp tpf
C2-33		Listeria riparia (SEQ ID No. 65)	mhdawaenpkkpqsdaflkeykacceaidtynwhknkatlvyvnelhhllidilg rlvgyvaiadrdfqcmanqylkssghtervdswintirknrpdyiekldifmnkagl fvseknggrnyiahlnylspkhkysllylfeklremlkydrklknavtkslidlldkhg mcvvfanlknnkhrlviaslkpkkietfkwkkik
C2-34		Insolitispirillum peregrinum (SEQ ID No. 66)	mriirpygsstvaspspqdaqplrslqrqngtfdvaefsrrhpelvlaqwvamldkii rkpapgknstalprptaeqrrlrqqvgaalwaemqrhtpvppelkavwdskvhpy skdnapataktpshrgrwydrfgdpetsaatvaegvrrhlldsaqpfranggqpkgk gviehraltiqngtllhhhqsekagplpedwstyradelvstigkdarwikvaaslyq hygrifgpttpiseaqtrpefvlhtavkayyrrlfkerklpaerlerllprtgealrhavtv qhgnrsladavrigkilhygwlqngepdpwpddaalyssrywgsdgqtdikhsea vsrvwrraltaaqrtltswlypagtdagdilligqkpdsidrnrlpllygdstrhwtrsp gdvwlflkqtlenlrnssfhfktlsaftshldgtcesepaeqqaaqalwqddrqqdhq qvflslraldattylptgplhrivnavqstdatlplprfrrvvtraantrlkgfpvepvnrrt meddpllrcrygvlkllyergfrawletrpsiascldqslkrstkaaqtingknspqgv eilsratkllqaeggggghgihdlfdrlyaataremrvqvgyhhdaeaarqqaefiedl kcevvarafcaylktlgiqgdtfrrqpeplptwpdlpdlpsstigtaqaalysvlhlmp

			vedvgsllhqlrrwlvalqarggedgtaitatipllelylnrhdakfsgggagtglrwd dwqvffdcqatfdrvfppgpaldshrlplrglrevlrfgrvndlaaligqdkitaaevd rwhtaeqtiaaqqqrrealheqlsrkkgtdaevdeyralvtaiadhrhltahvtlsnvv rlhrlmttvlgrlvdygglwerdltfvtlyeahrlgglrnllsesrvnkfldgqtpaalsk knnaeengmiskvlgdkarrqirndfahfnmlqqgkktinltdeinnarklmahdr klknaitrsvttllqqdgldivwtmdashrltdakidsrnaihlhkthnranireplhgk sycrwvaalfgatstpsatkksdkir

[0673] In certain example embodiments, the CRISPR effector protein is a Cas13b protein selected from Table 3.

Table 3

Bergeyella zoohelcum (SEQ ID No. 67)	1	menktslgnniyynpfkpqdksyfagyfnaamentdsvfrelgkrlkgkeytsenf fdaifkenislveyeryvkllsdyfpmarlldkkevpikerkenfkknfkgiikavrd lrnfythkehgeveitdeifgvldemlkstvltvkkkkvktdktkeilkksiekqldil cqkkleylrdtarkieekrrnqrergekelvapfkysdkrddliaaiyndafdvyidk kkdslkesskakyntksdpqqeegdlkipiskngvvfllslfltkqeihafkskiagfk atvideatvseatvshgknsicfmatheifshlaykklkrkvrtaeinygeaenaeqls vyaketlmmqmldelskvpdvvyqnlsedvqktfiedwneylkenngdvgtme eeqvihpvirkryedkfnyfairfldefaqfptlrfqvhlgnylhdsrpkenlisdrrik ekitvfgrlselehkkalfikntetnedrehyweifpnpnydfpkenisvndkdfpia gsildrekqpvagkigikvkllnqqyvsevdkavkahqlkqrkaskpsiqniieeiv pinesnpkeaivfggqptaylsmndihsilyeffdkwekkkeklekkgekelrkei gkelekkivgkiqaqiqqiidkdtnakilkpyqdgnstaidkeklikdlkqeqnilqk lkdeqtvrekeyndfiayqdknreinkvrdrnhkqylkdnlkrkypeaparkevly yrekgkvavwlandikrfmptdfknewkgeqhsllqkslayyeqckeelknllpe kvfqhlpfklggyfqqkylyqfytcyldkrleyisglvqqaenfksenkvfkkvene cfkflkkqnythkeldarvqsilgypiflergfmdekptiikgktfkgnealfadwfr yykeyqnfqtfydtenyplvelekkqadrkrktkiyqqkkndvftllmakhifksvf kqdsidqfsledlyqsreerlgnqerarqtgerntnyiwnktvdlklcdgkitvenvkl knvgdfikyeydqrvqaflkyeeniewqaflikeskeeenypyvvereieqyekvr reellkevhlieeyilekvkdkeilkkgdnqnfkyyilngllkqlknedvesykvfnl ntepedvninqlkqeatdleqkafvltyirnkfahnqlpkkefwdycqekygkiek ektyaeyfaevfkkekealik

Prevotella intermedia (SEQ ID No. 68)	2	meddkkttdsiryelkdkhfwaaflnlarhnvyitvnhinkileegeinrdgyettlk ntwneikdinkkdrlskliikhfpfleaatyrlnptdttkqkeekqaeaqsleslrksff vfiyklrdlrnhyshykhskslerpkfeegllekmynifnasirlvkedyqynkdin pdedfkhldrteeefnyyftkdnegnitesgllffvslflekkdaiwmqqklrgfkdn renkkkmtnevfcrsrmllpklrlqstqtqdwilldmlnelircpkslyerlreedrek frvpieiadedydaeqepfkntlvrhqdrfpyfalryfdyneiftnlrfqidlgtyhfsi ykkqigdykeshhlthklygferiqeftkqnrpdewrkfvktfnsfetskepyipett phyhlenqkigirfrndndkiwpslktnsekneksky kldksfqaeaflsvhellpm mfyylllktentdndneietkkkenkndkqekhkieeiienkiteiyalydtfangei ksideleeyckgkdieighlpkqmiailkdehkvmateaerkqeemlvdvqksle sldnqineeienverknsslksgkiaswlvndmmrfqpvqkdnegkplnnskans teyqllqrtlaffgseherlapyfkqtkliessnphpflkdtewekcnnilsfyrsylea kknfleslkpedweknqyflklkepktkpktlvqgwkngfnlprgiftepirkwfm khrenitvaelkrvglvakviplffseeykdsvqpfynyhfnvgninkpdeknflnc eerrellrkkkdefkkmtdkekeenpsylefkswnkferelrlvrnqdivtwllcme lfnkkkikelnvekiylknintnttkkeknteekngeeknikeknnilnrimpmrlpi kvygrenfsknkkkkirrntfftvyieekgtkllkqgnfkalerdrrlgglfsfvktpsk aesksntisklrveyelgeyqkarieiikdmlalektlidkynsldtdnfnkmltdwle lkgepdkasfqndvdlliavrnafshnqypmrnriafaninpfslssantseekglgi anqlkdkthktiekiieiekpietke
Prevotella buccae (SEQ ID No. 69)	3	mqkqdklfvdrkknaifafpkyitimenkekpepiyyeltdkhfwaaflnlarhnv yttinhinrrleiaelkddgymmgikgswneqakkldkkvrlrdlimkhfpfleaaa yemtnskspnnkeqrekeqsealslnnlknvlfifleklqvlrnyyshykyseespk pifetsllknmykvfdanvrlvkrdymhhenidmqrdfthlnrkkqvgrtkniids pnfhyhfadkegnmtiagllffvslfldkkdaiwmqkklkgfkdgrnlreqmtnev fcrsrislpklklenvqtkdwmqldmlnelvrcpkslyerlrekdresfkvpfdifsd dynaeeepfkntlvrhqdrfpyfvlryfdlneifeqlrfqidlgtyhfsiynkrigdede vrhlthhlygfariqdfapqnqpeewrklvkdldhfetsqepyisktaphyhleneki gikfcsahnnlfpslqtdktcngrskfnlgtqftaeaflsvhellpmmfyyllltkdysr kesadkvegiirkeisniyaiydafanneinsiadltrrlqntnilqghlpkqmisilkg rqkdmgkeaerkigemiddtqrrldllckqtnqkirigkrnagllksgkiadwlvnd mmrfqpvqkdqnnipinnskansteyrmlqralalfgsenfrlkayfnqmnlvgn dnphpflaetqwehqtnilsfyrnylearkkylkglkpqnwkqyqhflilkvqktnr

		ntlvtgwknsfnlprgiftqpirewfekhnnskriydqilsfdrvgfvakaiplyfaee ykdnvqpfydypfnignrlkpkkrqfldkkervelwqknkelfknypsekkktdl ayldflswkkferelrliknqdivtwlmfkelfnmatveglkigeihlrdidtntanee snnilnrimpmklpvktyetdnkgnilkerplatfyieetetkvlkqgnfkalvkdrrl nglfsfaettdlnleehpisklsvdlelikyqttrisifemtlglekklidkystlptdsfrn mlerwlqckanrpelknyvnsliavrnafshnqypmydatlfaevkkftlfpsvdtk kielniapqlleivgkaikeieksenkn
Porphyromonas gingivalis (SEQ ID No. 70)	4	mntvpasenkgqsrtveddpqyfglylnlarenlieveshvrikfgkkklneeslkq sllcdhllsvdrwtkvyghsrrylpflhyfdpdsqiekdhdsktgvdpdsaqrlirely slldflrndfshnrldgttfehlevspdissfitgtyslacgraqsrfavffkpddfvlakn rkeqlisvadgkecltvsgfafficlfldreqasgmlsrirgfkrtdenwaravhetfcd lcirhphdrlessntkeallldmlnelnrcprilydmlpeeeraqflpaldensmnnls ensldeesrllwdgssdwaealtkrirhqdrfpylmlrfieemdllkgirfrvdlgeiel dsyskkvgrngeydrtitdhalafgklsdfqneeevsrmisgeasypvrfslfaprya iydnkigychtsdpvypksktgekralsnpqsmgfisvhdlrklllmellcegsfsr mqsdflrkanrildetaegklqfsalfpemrhrfippqnpkskdrrekaettlekykq eikgrkdklnsqllsafdmdqrqlpsrlldewmnirpashsvklrtyvkqlnedcrlr lrkfrkdgdgkaraiplvgematflsqdivrmiiseetkklitsayynemqrslaqya geenrrqfraivaelrlldpssghpflsatmetahrytegfykcylekkrewlakifyr peqdentkrrisvffvpdgearkllpllirrrmkeqndlqdwirnkqahpidlpshlf dskvmellkvkdgkkkwneafkdwwstkypdgmqpfyglrrelnihgksvsyi psdgkkfadcythlmektvrdkkrelrtagkpvppdlaadikrsfhravnerefmlr lvqeddrlmlmainkmmtdreedilpglknidsildeenqfslavhakvlekegeg gdnslslvpatieikskrkdwskyiryrydrrvpglmshfpehkatldevktllgeyd rcrikifdwafalegaimsdrdlkpylhesssregksgehstlvkmlvekkgcltpde sqylilirnkaahnqfpcaaempliyrdvsakvgsiegssakdlpegsslvdslwkk yemiirkilpildpenrffgkllnnmsqpindl
Bacteroides pyogenes (SEQ ID No. 71)	5	mesiknsqkstgktlqkdppyfglylnmallnvrkvenhirkwlgdvallpeksgf hsllttdnlssakwtrfyyksrkflpflemfdsdkksyenrretaecldtidrqkissllk evygklqdirnafshyhiddqsvkhtaliissemhrfienaysfalqktrarftgvfvet dflqaeekgdnkkffaiggnegiklkdnalifliclfldreeafkflsratgfkstkekgf lavretfcalccrqpherllsvnpreallmdmlnelnrcpdilfemldekdqksflpll geeeqahilenslndelceaiddpfemiaslskrvryknrfpylmlryieeknllpfir

		fridlgclelasypkkmgeennyersvtdhamafgrltdfhnedavlqqitkgitdev rfslyapryaiynnkigfvrtsgsdkisfptlkkkggeghcvaytlqntksfgfisiydl rkilllsfldkdkaknivsglleqcekhwkdlsenlfdairtelqkefpvplirytlprsk ggklvsskladkqekyeseferrkeklteilsekdfdlsqiprrmidewlnvlptsrek klkgyvetlkldcrerlrvfekrekgehplpprigematdlakdiirmvidqgvkqri tsayyseiqrclaqyagddnrrhldsiirelrlkdtknghpflgkvlrpglghteklyqr yfeekkewleatfypaaspkrvprfvnpptgkqkelpliirnlmkerpewrdwkqr knshpidlpsqlfeneicrllkdkigkepsgklkwnemfklywdkefpngmqrfy rckrrvevfdkvveyeyseeggnykkyyealidevvrqkissskeksklqvedltls vrrvfkrainekeyqlrllceddrllfmavrdlydwkeaqldldkidnmlgepvsvs qviqleggqpdavikaecklkdvsklmrycydgrvkglmpyfanheatqeqvem elrhyedhrrrvfnwvfaleksvlkneklrrfyeesqggcehrrcidalrkaslvseee yeflvhirnksahnqfpdleigklppnvtsgfceciwskykaiicriipfidperrffgk lleqk
Alistipes sp. ZOR0009 (SEQ ID No. 72)	6	msneigafrehqfayapgnekqeeatfatyfnlalsnvegmmfgevesnpdkiek sldtlppailrqiasfiwlskedhpdkaysteevkvivtdlvrrlcfyrnyfshcfyldtq yfysdelvdttaigeklpynfhhfitnrlfryslpeitlfrwnegerkyeilrdgliffcclf lkrgqaerflnelrffkrtdeegrikrtiftkyctreshkhigieeqdflifqdiigdlnrvp kvcdgvvdlskeneryiknretsnesdenkaryrllirekdkfpyylmryivdfgvl pcitfkqndystkegrgqfhyqdaavaqeercynfvvrngnvyysympqaqnvv riselqgtisveelrnmvyasingkdvnksveqylyhlhllyekiltisgqtikegrvd vedyrplldkllrpasngeelrrelrkllpkrvcdllsnrfdcsegvsavekrlkaillrh eqlllsqnpalhidkiksvidylylffsddekfrqqptekahrglkdeefqmyhylvg dydshplalwkeleasgrlkpemrkltsatslhglymlclkgtvewcrkqlmsigk gtakveaiadrvglklydklkeytpeqlerevklvvmhgyaaaatpkpkaqaaips kltelrfysflgkremsfaafirqdkkaqklwlrnfytveniktlqkrqaaadaackkl ynlvgevervhtndkvlvlvaqryrerllnvgskcavtldnperqqkladvyevqna wlsirfddldftlthvnlsnlrkaynliprkhilafkeyldnrvkqklceecrnvrrkedl ctccsprysnltswlkenhsessiereaatmmlldverkllsfllderrkaiieygkfip fsalvkecrladaglcgirndvlhdnvisyadaigklsayfpkeaseaveyirrtkevr eqrreelmanssq
Prevotella sp.	7a	mskeckkqrqekkrrlqkanfsisltgkhvfgayfnmartnfvktinyilpiagvrg nysenqinkmlhalfliqagrneeltteqkqwekklrlnpeqqtkfqkllfkhfpvlg

MA2016 (SEQ ID No. 73)		pmmadvadhkaylnkkkstvqtedetfamlkgvsladcldiiclmadtltecrnfy thkdpynkpsqladqylhqemiakkldkvvvasrrilkdreglsvnevefltgidhl hqevlkdefgnakvkdgkvmktfveyddfyfkisgkrlvngytvttkddkpvnvn tmlpalsdfgllyfcvlflskpyaklfidevrlfeyspfddkenmimsemlsiyrirtp rlhkidshdskatlamdifgelrrcpmelynlldknagqpffhdevkhpnshtpdvs krlryddrfptlalryidetelfkrirfqlqlgsfrykfydkencidgrvrvrriqkeingy grmqevadkrmdkwgdliqkreersvkleheelyinldqfledtadstpyvtdrrp aynihanriglywedsqnpkqykvfdengmyipelvvtedkkapikmpaprcal svydlpamlfyeylreqqdnefpsaeqviieyeddyrkffkavaegklkpfkrpkef rdflkkeypklrmadipkklqlflcshglcynnkpetvyerldrltlqhleerelhiqnr lehyqkdrdmignkdnqygkksfsdvrhgalarylaqsmmewqptklkdkekg hdkltglnynvltaylatyghpqvpeegftprtleqvlinahliggsnphpfinkvlal gnrnieelylhyleeelkhirsriqslssnpsdkalsalpfihhdrmryhertseemm alaaryttiqlpdglftpyileilqkhytensdlqnalsqdvpvklnptcnaaylitlfyq tvlkdnaqpfylsdktytrnkdgekaesfsfkrayelfsvlnnnkkdtfpfemiplflt sdeiqerlsaklldgdgnpvpevgekgkpatdsqgntiwkrriysevddyaekltdr dmkisfkgeweklprwkqdkiikrrdetrrqmrdellqrmpryirdikdnertlrry ktqdmvlfllaekmftniiseqssefnwkqmrlskvcneaflrqtltfrvpvtvgetti yveqenmslknygefyrfltddrlmsllnnivetlkpnengdlvirhtdlmselaay dqyrstifmliqsienliitnnavlddpdadgfwvredlpkrnnfasllelinqlnnvel tdderkllvairnafshnsynidfslikdvkhlpevakgilqhlqsmlgveitk
Prevotella sp. MA2016 (SEQ ID No. 74)	7b	mskeckkqrqekkrrlqkanfsisltgkhvfgayfnmartnfvktinyilpiagvrg nysenqinkmlhalfliqagrneeltteqkqwekklrlnpeqqtkfqkllfkhfpvlg pmmadvadhkaylnkkkstvqtedetfamlkgvsladcldiiclmadtltecrnfy thkdpynkpsqladqylhqemiakkldkvvvasrrilkdreglsvnevefltgidhl hqevlkdefgnakvkdgkvmktfveyddfyfkisgkrlvngytvttkddkpvnvn tmlpalsdfgllyfcvlflskpyaklfidevrlfeyspfddkenmimsemlsiyrirtp rlhkidshdskatlamdifgelrrcpmelynlldknagqpffhdevkhpnshtpdvs krlryddrfptlalryidetelfkrirfqlqlgsfrykfydkencidgrvrvrriqkeingy grmqevadkrmdkwgdliqkreersvkleheelyinldqfledtadstpyvtdrrp aynihanriglywedsqnpkqykvfdengmyipelvvtedkkapikmpaprcal svydlpamlfyeylreqqdnefpsaeqviieyeddyrkffkavaegklkpfkrpkef rdflkkeypklrmadipkklqlflcshglcynnkpetvyerldrltlqhleerelhiqnr

		lehyqkdrdmignkdnqygkksfsdvrhgalarylaqsmmewqptklkdkekg hdkltglnynvltaylatyghpqvpeegftprtleqvlinahliggsnphpfinkvlal gnrnieelylhyleeelkhirsriqslssnpsdkalsalpfihhdrmryhertseemm alaaryttiqlpdglftpyileilqkhytensdlqnalsqdvpvklnptcnaaylitlfyq tvlkdnaqpfylsdktytrnkdgekaesfsfkrayelfsvlnnnkkdtfpfemiplflt sdeiqerlsaklldgdgnpvpevgekgkpatdsqgntiwkrriysevddyaekltdr dmkisfkgeweklprwkqdkiikrrdetrrqmrdellqrmpryirdikdnertlrry ktqdmvlfllaekmftniiseqssefnwkqmrlskvcneaflrqtltfrvpvtvgetti yveqenmslknygefyrfltddrlmsllnnivetlkpnengdlvirhtdlmselaay dqyrstifmliqsienliitnnavlddpdadgfwvredlpkrnnfasllelinqlnnvel tdderkllvairnafshnsynidfslikdvkhlpevakgilqhlqsmlgveitk
Riemerella anatipestifer (SEQ ID No. 75)	8	mekpllpnvytlkhkffwgaflniarhnafitichineqlglktpsnddkivdvvcet wnnilnndhdllkksqltelilkhfpfltamcyhppkkegkkkghqkeqqkekese aqsqaealnpskliealeilvnqlhslrnyyshykhkkpdaekdifkhlykafdaslr mvkedykahftvnltrdfahlnrkgknkqdnpdfnryrfekdgfftesgllfftnlfld krdaywmlkkvsgfkashkqrekmttevfcrsrillpklrlesrydhnqmlldmls elsrcpkllyeklseenkkhfqveadgfldeieeeqnpfkdtlirhqdrfpyfalryldl nesfksirfqvdlgtyhyciydkkigdeqekrhltrtllsfgrlqdfteinrpqewkalt kdldyketsnqpfiskttphyhitdnkigfrlgtskelypsleikdganriakypynsg fvahafisvhellplmfyqhltgksedllketvrhiqriykdfeeerintiedlekanqg rlplgafpkqmlgllqnkqpdlsekakikiekliaetkllshrlntklksspklgkrrek liktgvladwlvkdfmrfqpvaydaqnqpiksskanstefwfirralalyggeknrl egyfkqtnligntnphpflnkfnwkacrnlvdfyqqyleqrekfleaiknqpwepy qycllllkipkenrknlvkgweqggislprglfteairetlsedlmlskpirkeikkhgr vgfisraitlyfkekyqdkhqsfynlsykleakapllkreehyeywqqnkpqsptes qrlelhtsdrwkdyllykrwqhlekklrlyrnqdvmlwlmtleltknhfkelnlnyh qlklenlavnvqeadaklnplnqtlpmvlpvkvypatafgevqyhktpirtvyiree htkalkmgnfkalvkdrrlnglfsfikeendtqkhpisqlrlrreleiyqslrvdafketl sleekllnkhtslsslenefralleewkkeyaassmvtdehiafiasvrnafchnqypf ykealhapiplftvaqptteekdglgiaeallkvlreyceivksqi
Prevotella aurantiaca	9	meddkkttgsisyelkdkhfwaaflnlarhnvyitinhinkllleireidndekvldikt lwqkgnkdlnqkarlrelmtkhfpfletaiytknkedkkevkqekqaeaqsleslkd clflfldklqearnyyshykysefskepefeegllekmynifgnniqlvindyqhnk

(SEQ ID No. 76)		dinpdedfkhldrkgqfkysfadnegnitesgllffvslflekkdaiwmqqklngfk dnlenkkkmthevfcrsrilmpklrlestqtqdwilldmlnelircpkslyerlqgdd rekfkvpfdpadedynaeqepfkntlirhqdrfpyfvlryfdyneifknlrfqidlgty hfsiykkliggqkedrhlthklygferiqefakqnrpdewkaivkdldtyetsnkryi settphyhlenqkigirfrngnkeiwpslktndennekskykldkqyqaeaflsvhe llpmmfyylllkkekpnndeinasivegfikreirnifklydafangeinniddleky cadkgipkrhlpkqmvailydehkdmvkeakrkqkemvkdtkkllatlekqtqk ekeddgrnvkllksgeiarwlvndmmrfqpvqkdnegkplnnskansteyqml qrslalynneekptryfrqvnliesnnphpflkwtkweecnniltfyysyltkkiefln klkpedwkknqyflklkepktnretlvqgwkngfnlprgiftepirewfkrhqnns keyekvealdrvglvtkviplffkeeyfkdkeenfkedtqkeindcvqpfynfpyn vgnihkpkekdflhreerielwdkkkdkfkgykekikskkltekdkeefrsylefqs wnkferelrlvrnqdivtwllckelidklkidelnieelkklrlnnidtdtakkeknnil nrvmpmelpvtvyeiddshkivkdkplhtiyikeaetkllkqgnfkalvkdrrlngl fsfvktnseaeskrnpisklrveyelgeyqearieiiqdmlaleeklinkykdlptnkf semlnswlegkdeadkarfqndvdfliavrnafshnqypmhnkiefanikpfslyt annseekglgianqlkdktkettdkikkiekpietke
Prevotella saccharolytica (SEQ ID No. 77)	10	medkpfwaaffnlarhnvyltvnhinklldleklydegkhkeiferedifnisddvm ndansngkkrkldikkiwddldtdltrkyqlrelilkhfpfiqpaiigaqtkerttidkd krststsndslkqtgegdindllslsnvksmffrllqileqlrnyyshvkhsksatmpn fdedllnwmryifidsvnkvkedyssnsvidpntsfshliykdeqgkikpcrypfts kdgsinafgllffvslflekqdsiwmqkkipgfkkasenymkmtnevfcrnhillp kirletvydkdwmlldmlnevvrcplslykrltpaaqnkfkvpekssdnanrqedd npfsrilvrhqnrfpyfvlrffdlnevfttlrfqinlgcyhfaickkqigdkkevhhlirtl ygfsrlqnftqntrpeewntlvkttepssgndgktvqgvplpyisytiphyqieneki gikifdgdtavdtdiwpsvstekqlnkpdkytltpgfkadvflsvhellpmmfyyql llcegmlktdagnavekvlidtrnaifnlydafvqekintitdlenylqdkpilighlpk qmidllkghqrdmlkaveqkkamlikdterrlklldkqlkqetdvaakntgtllkng qiadwlvndmmrfqpvkrdkegnpincskansteyqmlqrafafyatdscrlsryf tqlhlihsdnshlflsrfeydkqpnliafyaaylkakleflnelqpqnwasdnyflllra pkndrqklaegwkngfnlprglftekiktwfnehktivdisdcdifknrvgqvarlip vffdkkfkdhsqpfyrydfnvgnvskpteanylskgkreelfksyqnkfknnipae ktkeyreyknfslwkkferelrliknqdiliwlmcknlfdekikpkkdilepriavsyi

		kldslqtntstagslnalakvvpmtlaihidspkpkgkagnnekenkeftvyikeegt kllkwgnfktlladrrikglfsyiehddidlkqhpltkrrvdleldlyqtcridifqqtlgl eaqlldkysdlntdnfyqmligwrkkegiprnikedtdflkdvrnafshnqypdsk kiafrrirkfnpkelileeeeglgiatqmykevekvvnrikrielfd
HMPREF9712_03108 [Myroides odoratimimus CCUG 10230] (SEQ ID No. 78)	11	mkdilttdttekqnrfyshkiadkyffggyfnlasnniyevfeevnkrntfgklakrd ngnlknyiihvfkdelsisdfekrvaifasyfpiletvdkksikernrtidltlsqrirqfr emlislvtavdqlrnfythyhhsdivienkvldflnssfvstalhvkdkylktdktkefl ketiaaeldilieaykkkqiekkntrfkankredilnaiyneafwsfindkdkdkdke tvvakgadayfeknhhksndpdfalnisekgivyllsffltnkemdslkanltgfkg kvdresgnsikymatqriysfhtyrglkqkirtseegvketllmqmidelskvpnvv yqhlsttqqnsfiedwneyykdyeddvetddlsrvihpvirkryedrfnyfairflde ffdfptlrfqvhlgdyvhdrrtkqlgkvesdriikekvtvfarlkdinsakasyfhslee qdkeeldnkwtlfpnpsydfpkehtlqhqgeqknagkigiyvklrdtqykekaale earkslnpkersatkaskydiitqiieandnvksekplvftgqpiaylsmndihsmlf slltdnaelkktpeeveaklidqigkqineilskdtdtkilkkykdndlketdtdkitrdl ardkeeieklileqkqraddynytsstkfnidksrkrkhllfnaekgkigvwlandikr fmfkeskskwkgyqhtelqklfayfdtsksdlelilsnmvmvkdypielidlvkks rtlvdflnkylearleyienvitrvknsigtpqfktvrkecftflkksnytvvsldkqver ilsmplfiergfmddkptmlegksykqhkekfadwfvhykensnyqnfydtevy eittedkrekakvtkkikqqqkndvftlmmvnymleevlklssndrlslnelyqtke erivnkqvakdtqernknyiwnkvvdlqlcdglvhidnvklkdignfrkyendsrv kefltyqsdivwsaylsnevdsnklyvierqldnyesirskellkevqeiecsvynqv ankeslkqsgnenfkqyvlqgllpigmdvremlilstdvkfkkeeiiqlgqageveq dlysliyirnkfahnqlpikeffdfcennyrsisdneyyaeyymeifrsikekyan
Prevotella intermedia (SEQ ID No. 79)	12	meddkkttdsiryelkdkhfwaaflnlarhnvyitvnhinkileedeinrdgyentle nswneikdinkkdrlskliikhfpfleattyrqnptdttkqkeekqaeaqsleslkksff vfiyklrdlrnhyshykhskslerpkfeedlqnkmynifdvsiqfvkedykhntdin pkkdfkhldrkrkgkfhysfadnegnitesgllffvslflekkdaiwvqkklegfkcs nksyqkmtnevfcrsrmllpklrlestqtqdwilldmlnelircpkslyerlqgvnrk kfyvsfdpadedydaeqepfkntlvrhqdrfpyfalryfdynevfanlrfqidlgtyh fsiykkliggqkedrhlthklygferiqefdkqnrpdewkaivkdsdtfkkkeekee ekpyisettphyhlenkkigiafknhniwpstqteltnnkrkkynlgtsikaeaflsvh ellpmmfyylllktentkndnkvggkketkkqgkhkieaiieskikdiyalydafan

		geinsedelkeylkgkdikivhlpkqmiailknehkdmaekaeakqekmklaten rlktldkqlkgkiqngkrynsapksgeiaswlvndmmrfqpvqkdengeslnnsk ansteyqllqrtlaffgseherlapyfkqtkliessnphpflndtewekcsnilsfyrsyl karknfleslkpedweknqyflmlkepktnretlvqgwkngfnlprgfftepirkwf mehwksikvddlkrvglvakvtplffsekykdsvqpfynypfnvgdvnkpkeed flhreerielwdkkkdkfkgykakkkfkemtdkekeehrsylefqswnkferelrl vrnqdivtwllctelidklkidelnikelkklrlkdintdtakkeknnilnrvmpmelp vtvykvnkggyiiknkplhtiyikeaetkllkqgnfkalvkdrrlnglfsfvktpseae sesnpisklrveyelgkyqnarldiiedmlalekklidkynsldtdnfhnmltgwlel kgeakkarfqndvklltavrnafshnqypmydenlfgnierfslsssniieskgldia aklkeevskaakkiqneednkkeket
Capnocyto phaga canimorsus (SEQ ID No. 80)	13	mkniqrlgkgnefspfkkedkfyfggflnlannniedffkeiitrfgivitdenkkpk etfgekilneifkkdisivdyekwvnifadyfpftkylslyleemqfknrvicfrdvm kellktvealrnfythydhepikiedrvfyfldkvlldvsltvknkylktdktkeflnqh igeelkelckqrkdylvgkgkridkeseiingiynnafkdfickrekqddkenhnsv ekilcnkepqnkkqkssatvwelcskssskyteksfpnrendkhclevpisqkgivf llsfflnkgeiyaltsnikgfkakitkeepvtydknsirymathrmfsflaykglkrkir tseinynedgqasstyeketlmlqmldelnkvpdvvyqnlsedvqktfiedwneyl kenngdvgtmeeeqvihpvirkryedkfnyfairfldefaqfptlrfqvhlgnylcd krtkqicdttterevkkkitvfgrlselenkkaiflnereeikgwevfpnpsydfpken isvnykdfpivgsildrekqpvsnkigirvkiadelqreidkaikekklrnpknrkan qdekqkerlvneivstnsneqgepvvfigqptaylsmndihsvlyeflinkisgeale tkivekietqikqiigkdattkilkpytnansnsinrekllrdleqeqqilktlleeqqqre kdkkdkkskrkhelypsekgkvavwlandikrfmpkafkeqwrgyhhsllqkyl ayyeqskeelknllpkevfkhfpfklkgyfqqqylnqfytdylkrrlsyvnelllniq nfkndkdalkatekecfkffrkqnyiinpiniqiqsilvypiflkrgfldekptmidre kfkenkdteladwfmhyknykednyqkfyayplekveekekfkrnkqinkqkk ndvytlmmveyiiqkifgdkfveenplvlkgifqskaerqqnnthaattqernlngil nqpkdikiqgkitvkgvklkdignfrkyeidqrvntfldyeprkewmaylpndwk ekekqgqlppnnvidrqiskyetvrskillkdvqelekiisdeikeehrhdlkqgkyy nfkyyilngllrqlknenvenykvfklntnpekvnitqlkqeatdleqkafvltyirnk fahnqlpkkefwdycqekygkiekektyaeyfaevfkrekealik

Porphyromonas gulae (SEQ ID No. 81)	14	mteqserpyngtyytledkhfwaaflnlarhnayitlthidrqlayskaditndqdvls fkalwknfdndlerksrlrslilkhfsflegaaygkklfeskssgnkssknkeltkkek eelqanalsldnlksilfdflqklkdfrnyyshyrhsgsselplfdgnmlqrlynvfdv svqrvkidhehndevdphyhfnhlvrkgkkdryghndnpsfkhhfvdgegmvt eagllffvslflekrdaiwmqkkirgfkggtetyqqmtnevfcrsrislpklkleslrm ddwmlldmlnelvrcpkplydrlreddracfrvpvdilpdeddtdgggedpfkntl vrhqdrfpyfalryfdlkkvftslrfhidlgtyhfaiykkmigeqpedrhltrnlygfgr iqdfaeehrpeewkrlvrdldyfetgdkpyisqtsphyhiekgkiglrfmpegqhl wpspevgttrtgrskyaqdkrltaeaflsvhelmpmmfyyfllrekyseevsaervq grikrviedvyavydafardeintrdeldacladkgirrghlprqmiailsqehkdme ekirkklqemmadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrf qpvakdasgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphp flhetrweshtnilsfyrsylrarkaflerigrsdrvenrpflllkepktdrqtlvagwkg efhlprgifteavrdcliemghdevasykevgfmakavplyferacedrvqpfyds pfnvgnslkpkkgrflskeeraeewergkerfrdleawsysaarriedafagieyasp gnkkkieqllrdlslweafesklkvradrinlaklkkeileaqehpyhdfkswqkfer elrlvknqdiitwmmcrdlmeenkvegldtgtlylkdirpnvqeqgslnvlnrvkp mrlpvvvyradsrghvhkeeaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvd tgglameqypisklrveyelakyqtarvcvfeltlrleeslltryphlpdesfremles wsdpllakwpelhgkvrlliavrnafshnqypmydeavfssirkydpsspdaieer mglniahrlseevkqaketveriiqa
Prevotella sp. P5-125 (SEQ ID No. 82)	15	mnipalvenqkkyfgtysvmamlnaqtvldhiqkvadiegeqnennenlwfhp vmshlynakngydkqpektmfiierlqsyfpflkimaenqreysngkykqnrvev nsndifevlkrafgvlkmyrdltnhyktyeeklndgcefltsteqplsgminnyytva lrnmnerygyktedlafiqdkrfkfvkdaygkkksqvntgfflslqdyngdtqkklh lsgvgialliclfldkqyiniflsrlpifssynaqseerriiirsfginsiklpkdrihseksn ksvamdmlnevkrcpdelfttlsaekqsrfriisddhnevlmkrssdrfvplllqyid ygklfdhirfhvnmgklryllkadktcidgqtrvrvieqplngfgrleeaetmrkqen gtfgnsgirirdfenmkrddanpanypyivdtythyilennkvemfindkedsapll pvieddryvvktipscrmstleipamafhmflfgskkteklivdvhnrykrlfqam qkeevtaeniasfgiaesdlpqkildlisgnahgkdvdafirltvddmltdterrikrfk ddrksirsadnkmgkrgfkqistgkladflakdivlfqpsvndgenkitglnyrimq saiavydsgddyeakqqfklmfekarligkgttephpflykvfarsipanavefyer

		ylierkfyltglsneikkgnrvdvpfirrdqnkwktpamktlgriysedlpvelprqm fdneikshlkslpqmegidfnnanvtyliaeymkrvldddfqtfyqwnrnyrymd mlkgeydrkgslqhcftsveereglwkerasrteryrkqasnkirsnrqmrnassee ietildkrlsnsrneyqksekvirryrvqdallfllakktlteladfdgerfklkeimpda ekgilseimpmsftfekggkkytitsegmklknygdffvlasdkrignllelvgsdiv skedimeefnkydqcrpeissivfnlekwafdtypelsarvdreekvdfksilkilln nkninkeqsdilrkirnafdhnnypdkgvveikalpeiamsikkafgeyaimk
Flavobacterium branchiophilum (SEQ ID No. 83)	16	menlnkildkeneiciskifntkgiaapitekaldnikskqkndlnkearlhyfsighs fkqidtkkvfdyvlieelkdekplkfitlqkdfftkefsiklqklinsirninnhyvhnf ndinlnkidsnvfhflkesfelaiiekyykvnkkypldneivlflkelfikdentallny ftnlskdeaieyiltftitenkiwninnehnilniekgkyltfeamlflitiflykneanhl lpklydfknnkskqelftffskkftsqdidaeeghlikfrdmiqylnhyptawnndlk lesenknkimttklidsiiefelnsnypsfatdiqfkkeakaflfasnkkrnqtsfsnks yneeirhnphikqyrdeiasaltpisfnvkedkfkifvkkhvleeyfpnsigyekfle yndftekekedfglklysnpktnklieridnhklvkshgrnqdrfmdfsmrflaenn yfgkdaffkcykfydtqeqdeflqsnennddvkfhkgkvttyikyeehlknysyw dcpfveennsmsvkisigseekilkiqrnlmiyflenalynenvenqgyklvnnyy relkkdveesiasldliksnpdfkskykkilpkrllhnyapakqdkapenafetllkk adfreeqykkllkkaeheknkedfvkrnkgkqfklhfirkacqmmyfkekyntlk egnaafekkdpviekrknkehefghhknlnitreefndyckwmfafngndsykk ylrdlfsekhffdnqeyknlfessvnleafyaktkelfkkwietnkptnnenrytleny knlilqkqvfinvyhfskylidknllnsennviqykslenveylisdfyfqsklsidqy ktcgklfnklksnkledcllyeiaynyidkknvhkidiqkiltskiiltindantpykis vpfnklerytemiaiknqnnlkarflidlplylsknkikkgkdsagyeiiikndleied intinnkiindsvkftevlmelekyfilkdkcilsknyidnseipslkqfskvwikene neiinyrniachfhlplletfdnlllnveqkfikeelqnvstindlskpqeylillfikfkh nnfylnlfnknesktikndkevkknrvlqkfinqvilkkk
Myroides odoratimimus (SEQ ID No. 84)	17	mkdilttdttekqnrfyshkiadkyffggyfnlasnniyevfeevnkrntfgklakrd ngnlknyiihvfkdelsisdfekrvaifasyfpiletvdkksikernrtidltlsqrirqfr emlislvtavdqlrnfythyhhsdivienkvldflnssfvstalhvkdkylktdktkefl ketiaaeldilieaykkkqiekkntrfkankredilnaiyneafwsfindkdkdkdke tvvakgadayfeknhhksndpdfalnisekgivyllsffltnkemdslkanltgfkg kvdresgnsikymatqriysfhtyrglkqkirtseegvketllmqmidelskvpnvv

		yqhlsttqqnsfiedwneyykdyeddvetddlsrvthpvirkryedrfnyfairflde ffdfptlrfqvhlgdyvhdrrtkqlgkvesdriikekvtvfarlkdinsakasyfhslee qdkeeldnkwtlfpnpsydfpkehtlqhqgeqknagkigiyvklrdtqykekaale earkslnpkersatkaskydiitqiieandnvksekplvftgqpiaylsmndihsmlf slltdnaelkktpeeveaklidqigkqineilskdtdtkilkkykdndlketdtdkitrdl ardkeeieklileqkqraddynytsstkfnidksrkrkhllfnaekgkigvwlandikr fmfkeskskwkgyqhielqklfayfdtsksdlelilsnmvmvkdypielidlvkks rtlvdflnkylearleyienvitrvknsigtpqfktvrkecftflkksnytvvsldkqver ilsmplfiergfmddkptmlegksykqhkekfadwfvhykensnyqnfydtevy eittedkrekakvtkkikqqqkndvftlmmvnymleevlklssndrlslnelyqtke erivnkqvakdtqernknyiwnkvvdlqlcdglvhidnvklkdignfrkyendsrv kefltyqsdivwsaylsnevdsnklyvierqldnyesirskellkevqeiecsvynqv ankeslkqsgnenfkqyvlqgllpigmdvremlilstdvkfkkeeiiqlgqageveq dlysliyirnkfahnqlpikeffdfcennyrsisdneyyaeyymeifrsikekyan
Flavobacterium columnare (SEQ ID No. 85)	18	mssknesynkqktfnhykqedkyffggflnnaddnlrqvgkefktrinfnhnnnel asvfkdyfnkeksvakrehalnllsnyfpvleriqkhtnhnfeqtreifellldtikklrd yythhyhkpitinpkiydflddtlldvlitikkkkvkndtsrellkeklrpeltqlknqk reelikkgkklleenlenavfnhclipfleenktddkqnktvslrkyrkskpneetsitl tqsglvflmsfflhrkefqvftsglerfkakvntikeeeislnknnivymithwsysyy nfkglkhriktdqgvstleqnntthsltntntkealltqivdylskvpneiyetlsekqq kefeedineymrenpenedstfssivshkvirkryenkfnyfamrfldeyaelptlrf mvnfgdyikdrqkkilesiqfdseriikkeihlfeklslvteykknvylketsnidlsrf plfpnpsyvmannnipfyidsrsnnldeylnqkkkaqsqnkkrnltfekynkeqsk daiiamlqkeigvkdlqqrstigllscnelpsmlyevivkdikgaelenkiaqkireq yqsirdftldspqkdnipttliktintdssvtfenqpidiprlknalqkeltltqekllnvk eheievdnynrnkntykfknqpknkvddkklqrkyvfyrneirqeanwlasdlihf mknkslwkgymhnelqsflaffedkkndcialletvfnlkedciltkglknlflkhg nfidfykeylklkedflstestflengfiglppkilkkelskrlkyifivfqkrqfiikelee kknnlyadainlsrgifdekptmipfkkpnpdefaswfvasyqynnyqsfyeltpd iverdkkkkyknlrainkvkiqdyylklmvdtlyqdlfnqpldkslsdfyvskaere kikadakayqklndsslwnkvihlslqnnritanpklkdigkykralqdekiatllty dartwtyalqkpekenendykelhytalnmelqeyekvrskellkqvqelekkild kfydfsnnashpedleiedkkgkrhpnfklyitkallkneseiinlenidieillkyyd

		ynteelkekiknmdedekakiintkenynkitnvlikkalvliiirnkmahnqyppk fiydlanrfvpkkeeeyfatyfnrvfetitkelwenkekkdktqv
Porphyromonas gingivalis (SEQ ID No. 86)	19	mteqnekpyngtyytledkhfwaaflnlarhnayitlahidrqlayskaditndedil ffkgqwknldndlerkarlrslilkhfsflegaaygkklfesqssgnkssskkkelskke keelqanalsldnlksilfdflqklkdfrnyyshyrhpesselplfdgnmlqrlynvfd vsvqrvkrdhehndkvdphrhfnhlvrkgkkdkygnndnpffkhhfvdregtvte agllffvslflekrdaiwmqkkirgfkggteayqqmtnevfcrsrislpklkleslrtd dwmlldmlnelvrcpkslydrlreedrarfrvpvdilsdeddtdgteedpfkntlvrh qdrfpyfalryfdlkkvftslrfhidlgtyhfaiykknigeqpedrhltrnlygfgriqdf aeehrpeewkrlvrdldyfetgdkpyitqttphyhiekgkiglrfvpegqhlwpspe vgatrtgrskyaqdkrltaeaflsvhelmpmmfyyfllrekyseevsaekvqgrikr viedvyavydafardeintrdeldacladkgirrghlprqmiailsqehkdmeekvr kklqemiadtdhrldmldrqtdrkirigrknaglpksgvvadwlvrdmmrfqpva kdtsgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpflhetr weshtnilsfyrsylearkaflqsigrsdrvenhrflllkepktdrqtlvagwkgefhlp rgifteavrdcliemgydevgsykevgfmakavplyferaskdrvqpfydypfnv gnslkpkkgrflskekraeewesgkerfrlaklkkeileakehpyhdfkswqkfere lrlvknqdiitwmmcrdlmeenkvegldtgtlylkdirtdvqeqgslnvlnrvkpm rlpvvvyradsrghvhkeqaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvdtg alameqypisklrveyelakyqtarvcafeqtleleeslltryphlpdknfrkmlesw sdplldkwpdlhgnvrlliavrnafshnqypmydetlfssirkydpsspdaieermg lniahrlseevkqakemveriiqa
Porphyromonas sp. COT-052 OH4946 (SEQ ID No. 87)	20	mteqserpyngtyytledkhfwaaflnlarhnayitlthidrqlayskaditndqdvls fkalwknfdndlerksrlrslilkhfsflegaaygkklfeskssgnkssknkeltkkek eelqanalsldnlksilfdflqklkdfrnyyshyrhsesselplfdgnmlqrlynvfdv svqrvkrdhehndkvdphrhfnhlvrkgkkdryghndnpsfkhhfvdsegmvte agllffvslflekrdaiwmqkkirgfkggtetyqqmtnevfcrsrislpklkleslrtdd wmlldmlnelvrcpkplydrlreddracfrvpvdilpdeddtdgggedpfkntlvr hqdrfpyfalryfdlkkvftslrfhidlgtyhfaiykkmigeqpedrhltrnlygfgriq dfaeehrpeewkrlvrdldyfetgdkpyisqttphyhiekgkiglrfvpegqhlwps pevgttrtgrskyaqdkrltaeaflsvhelmpmmfyyfllrekyseevsaekvqgri krviedvyaiydafardeintlkeldacladkgirrghlpkqmigilsqerkdmeek vrkklqemiadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqpv

		akdtsgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpflhet rweshtnilsfyrsylrarkaflerigrsdrvencpflllkepktdrqtlvagwkgefhl prgifteavrdcliemgydevgsyrevgfmakavplyferacedrvqpfydspfnv gnslkpkkgrflskedraeewergkerfrdleawshsaarrikdafagieyaspgnk kkieqllrdlslweafesklkvradkinlaklkkeileaqehpyhdfkswqkferelrl vknqdiitwmmcrdlmeenkvegldtgtlylkdirpnvqeqgslnvlnrvkpmrl pvvvyradsrghvhkeeaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvdtggl ameqypisklrveyelakyqtarvcvfeltlrleesllsryphlpdesfremleswsdp llakwpelhgkvrlliavrnafshnqypmydeavfssirkydpsspdaieermglni ahrlseevkqaketveriiqa
Prevotella intermedia (SEQ ID No. 88)	21	meddkktkestnmldnkhfwaaflnlarhnvyitvnhinkvlelknkkdqdiiidn dqdilaikthwekvngdlnkterlrelmtkhfpfletaiytknkedkeevkqekqak aqsfdslkhclflfleklqearnyyshykysestkepmlekellkkmynifddniqlv ikdyqhnkdinpdedfkhldrteeefnyyfttnkkgnitasgllffvslflekkdaiw mqqklrgfkdnreskkkmthevfcrsrmllpklrlestqtqdwilldmlnelircpk slyerlqgeyrkkfnvpfdsadedydaeqepfkntlvrhqdrfpyfalryfdyneift nlrfqidlgtyhfsiykkliggqkedrhlthklygferiqefakqnrtdewkaivkdfd tyetseepyisetaphyhlenqkigirfrndndeiwpslktngennekrkykldkqy qaeaflsvhellpmmfyylllkkeepnndkknasivegfikreirdiyklydafang einniddlekycedkgipkrhlpkqmvailydehkdmaeeakrkqkemvkdtk kllatlekqtqgeiedggrnirllksgeiarwlvndmmrfqpvqkdnegnplnnsk ansteyqmlqrslalynkeekptryfrqvnlinssnphpflkwtkweecnnilsfyrs yltkkieflnklkpedweknqyflklkepktnretlvqgwkngfnlprgiftepirew fkrhqndseeyekvetldrvglvtkviplffkkedskdkeeylkkdaqkeinncvq pfygfpynvgnihkpdekdflpseerkklwgdkkykfkgykakvkskkltdkek eeyrsylefqswnkferelrlvrnqdivtwllctelidklkveglnveelkklrlkdidt dtakqeknnilnrvmpmqlpvtvyeiddshnivkdrplhtvyieetktkllkqgnfk alvkdrrlnglfsfvdtssetelksnpiskslveyelgeyqnarietikdmllleetliek yktlptdnfsdmlngwlegkdeadkarfqndvkllvavrnafshnqypmrnriafa ninpfslssadtseekkldianqlkdkthkiikriieiekpietke
PIN17_0200 [Prevotella	AFJ07523	mkmeddkktkestnmldnkhfwaaflnlarhnvyitvnhinkvlelknkkdqdii idndqdilaikthwekvngdlnkterlrelmtkhfpfletaiytknkedkeevkqek qakaqsfdslkhclflfleklqearnyyshykysestkepmlekellkkmynifddn

intermedia 17] (SEQ ID No. 89)		iqlvikdyqhnkdinpdedfkhldrteeefnyyfttnkkgnitasgllffvslflekkd aiwmqqklrgfkdnreskkkmthevfcrsrmllpklrlestqtqdwilldmlnelir cpkslyerlqgeyrkkfnvpfdsadedydaeqepfkntlvrhqdrfpyfalryfdyn eiftnlrfqidlgtyhfsiykkliggqkedrhlthklygferiqefakqnrtdewkaivk dfdtyetseepyisetaphyhlenqkigirfrndndeiwpslktngennekrkykld kqyqaeaflsvhellpmmfyylllkkeepnndkknasivegfikreirdiyklydaf angeinniddlekycedkgipkrhlpkqmvailydehkdmaeeakrkqkemvk dtkkllatlekqtqgeiedggrnirllksgeiarwlvndmmrfqpvqkdnegnplnn skansteyqmlqrslalynkeekptryfrqvnlinssnphpflkwtkweecnnilsf yrsyltkkieflnklkpedweknqyflklkepktnretlvqgwkngfnlprgiftepir ewfkrhqndseeyekvetldrvglvtkviplffkkedskdkeeylkkdaqkeinnc vqpfygfpynvgnihkpdekdflpseerkklwgdkkykfkgykakvkskkltdk ekeeyrsylefqswnkferelrlvrnqdivtwllctelidklkveglnveelkklrlkdi dtdtakqeknnilnrvmpmqlpvtvyeiddshnivkdrplhtvyieetktkllkqgn fkalvkdrrlnglfsfvdtssetelksnpiskslveyelgeyqnarietikdmllleetlie kyktlptdnfsdmlngwlegkdeadkarfqndvkllvavrnafshnqypmrnriaf aninpfslssadtseekkldianqlkdkthkiikriieiekpietke
Prevotella intermedia (SEQ ID No. 90)	BAU18623	meddkkttdsisyelkdkhfwaaflnlarhnvyitvnhinkvlelknkkdqdiiidn dqdilaikthwekvngdlnkterlrelmtkhfpfletaiysknkedkeevkqekqak aqsfdslkhclflfleklqetrnyyshykysestkepmlekellkkmynifddniqlv ikdyqhnkdinpdedfkhldrteedfnyyftrnkkgnitesgllffvslflekkdaiw mqqklrgfkdnreskkkmthevfcrsrmllpklrlestqtqdwilldmlnelircpk slyerlqgedrekfkvpfdpadedydaeqepfkntlvrhqdrfpyfalryfdyneift nlrfqidlgtfhfsiykkliggqkedrhlthklygferiqefakqnrpdewkaivkdld tyetsneryisettphyhlenqkigirfrndndeiwpslktngennekskykldkqyq aeaflsvhellpmmfyylllkkeepnndkknasivegfikreirdmyklydafang einniddlekycedkgipkrhlpkqmvailydehkdmvkeakrkqrkmvkdtek llaalekqtqektedggrnirllksgeiarwlvndmmrfqpvqkdnegnplnnska nsteyqmlqrslalynkeekptryfrqvnlinssnphpflkwtkweecnnilsfyrsy ltkkieflnklkpedweknqyflklkepktnretlvqgwkngfnlprgiftepirewf krhqndskeyekvealdrvglvtkviplffkkedskdkeedlkkdaqkeinncvqp fysfpynvgnihkpdekdflhreerielwdkkkdkfkgykakvkskkltdkekee yrsylefqswnkferelrlvrnqdivtwllctelidklkveglnveelkklrlkdidtdta

		kqeknnilnrvmpmqlpvtvyeiddshnivkdrplhtvyieetktkllkqgnfkalv kdrrlnglfsfvdtsseaelksnpiskslveyelgeyqnarietikdmllleetliekyk nlptdnfsdmlngwlegkdeadkarfqndvkllvavrnafshnqypmrnriafani npfslssadtseekkldianqlkdkthkiikriieiekpietke
HMPREF6 485_0083 [Prevotella buccae ATCC 33574] (SEQ ID No. 91)	EFU3198 1	mqkqdklfvdrkknaifafpkyitimenkekpepiyyeltdkhfwaaflnlarhnv yttinhinrrleiaelkddgymmgikgswneqakkldkkvrlrdlimkhfpfleaaa yemtnskspnnkeqrekeqsealslnnlknvlfifleklqvlrnyyshykyseespk pifetsllknmykvfdanvrlvkrdymhhenidmqrdfthlnrkkqvgrtkniids pnfhyhfadkegnmtiagllffvslfldkkdaiwmqkklkgfkdgrnlreqmtnev fcrsrislpklklenvqtkdwmqldmlnelvrcpkslyerlrekdresfkvpfdifsd dynaeeepfkntlvrhqdrfpyfvlryfdlneifeqlrfqidlgtyhfsiynkrigdede vrhlthhlygfariqdfapqnqpeewrklvkdldhfetsqepyisktaphyhleneki gikfcsahnnlfpslqtdktcngrskfnlgtqftaeaflsvhellpmmfyylllkdysr kesadkvegiirkeisniyaiydafanneinsiadltrrlqntnilqghlpkqmisilkg rqkdmgkeaerkigemiddtqrrldllckqtnqkirigkrnagllksgkiadwlvnd mmrfqpvqkdqnnipinnskansteyrmlqralalfgsenfrlkayfnqmnlvgn dnphpflaetqwehqtnilsfyrnylearkkylkglkpqnwkqyqhflilkvqktnr ntlvtgwknsfnlprgiftqpirewfekhnnskriydqilsfdrvgfvakaiplyfaee ykdnvqpfydypfnignrlkpkkrqfldkkervelwqknkelfknypsekkktdl ayldflswkkferelrliknqdivtwlmfkelfnmatveglkigeihlrdidtntanee snnilnrimpmklpvktyetdnkgnilkerplatfyieetetkvlkqgnfkalvkdrrl nglfsfaettdlnleehpisklsvdlelikyqttrisifemtlglekklidkystlptdsfrn mlerwlqckanrpelknyvnsliavrnafshnqypmydatlfaevkkftlfpsvdtk kielniapqlleivgkaikeieksenkn
HMPREF9 144_1146 [Prevotella pallens ATCC 700821] (SEQ ID No. 92)	EGQ1844 4	mkeeekgktpvvstynkddkhfwaaflnlarhnvyitvnhinkilgegeinrdgye ntlekswneikdinkkdrlskliikhfpflevttyqrnsadttkqkeekqaeaqslesl kksffvfiyklrdlrnhyshykhskslerpkfeedlqekmynifdasiqlvkedykh ntdikteedfkhldrkgqfkysfadnegnitesgllffvslflekkdaiwvqkklegfk csnesyqkmtnevfcrsrmllpklrlqstqtqdwilldmlnelircpkslyerlreedr kkfrvpieiadedydaeqepfknalvrhqdrfpyfalryfdyneiftnlrfqidlgtyh fsiykkqigdykeshhlthklygferiqeftkqnrpdewrkfvktfnsfetskepyip ettphyhlenqkigirfrndndkiwpslktnseknekskykldksfqaeaflsvhell pmmfyylllktentdndneietkkkenkndkqekhkieeiienkiteiyalydafan

		gkinsidkleeyckgkdieighlpkqmiailksehkdmateakrkqeemladvqk slesldnqineeienverknsslksgeiaswlvndmmrfqpvqkdnegnplnnsk ansteyqmlqrslalynkeekptryfrqvnliessnphpflnntewekcnnilsfyrs yleakknfleslkpedweknqyflmlkepktncetlvqgwkngfnlprgiftepirk wfmehrknitvaelkrvglvakviplffseeykdsvqpfynylfnvgninkpdekn flnceerrellrkkkdefkkmtdkekeenpsylefqswnkferelrlvrnqdivtwll cmelfnkkkikelnvekiylknintnttkkeknteekngeekiikeknnilnrimp mrlpikvygrenfsknkkkkirrntfftvyieekgtkllkqgnfkalerdrrlgglfsfv kthskaesksntisksrveyelgeyqkarieiikdmlaleetlidkynsldtdnfhnml tgwlklkdepdkasfqndvdlliavrnafshnqypmrnriafaninpfslssantsee kglgianqlkdkthktiekiieiekpietke
HMPREF9714_02132 [Myroides odoratimimus CCUG 12901] (SEQ ID No. 93)	EHO08761	mkdilttdttekqnrfyshkiadkyffggyfnlasnniyevfeevnkrntfgklakrd ngnlknyiihvfkdelsisdfekrvaifasyfpiletvdkksikernrtidltlsqrirqfr emlislvtavdqlrnfythyhhseivienkvldflnsslvstalhvkdkylktdktkefl ketiaaeldilieaykkkqiekkntrfkankredilnaiyneafwsfindkdkdketv vakgadayfeknhhksndpdfalnisekgivyllsffltnkemdslkanltgfkgkv dresgnsikymatqriysfhtyrglkqkirtseegvketllmqmidelskvpnvvyq hlsttqqnsfiedwneyykdyeddvetddlsrvihpvirkryedrfnyfairfldeffd fptlrfqvhlgdyvhdrrtkqlgkvesdriikekvtvfarlkdinsakanyfhsleeqd keeldnkwtlfpnpsydfpkehtlqhqgeqknagkigiyvklrdtqykekaaleea rkslnpkersatkaskydiitqiieandnvksekplvftgqpiaylsmndihsmlfsll tdnaelkktpeeveaklidqigkqineilskdtdtkilkkykdndlketdtdkitrdlar dkeeieklileqkqraddynytsstkfnidksrkrkhllfnaekgkigvwlandikrf mteefkskwkgyqhtelqklfayydtsksdldlilsdmvmvkdypielialvkksrt lvdflnkylearlgymenvitrvknsigtpqfktvrkecftflkksnytvvsldkqver ilsmplfiergfmddkptmlegksyqqhkekfadwfvhykensnyqnfydtevy eittedkrekakvtkkikqqqkndvftlmmvnymleevlklssndrlslnelyqtke erivnkqvakdtqernknyiwnkvvdlqlceglvridkvklkdignfrkyendsrv kefltyqsdivwsaylsnevdsnklyvierqldnyesirskellkevqeiecsvynqv ankeslkqsgnenfkqyvlqglvpigmdvremlilstdvkfikeeiiqlgqageveq dlysliyirnkfahnqlpikeffdfcennyrsisdneyyaeyymeifrsikekyts
HMPREF9711_00870	EKB06014	mkdilttdttekqnrfyshkiadkyffggyfnlasnniyevfeevnkrntfgklakrd ngnlknyiihvfkdelsisdfekrvaifasyfpiletvdkksikernrtidltlsqrirqfr

[Myroides odoratimimus CCUG 3837] (SEQ ID No. 94)		emlislvtavdqlrnfythyhhseivienkvldflnsslvstalhvkdkylktdktkefl ketiaaeldilieaykkkqiekkntrfkankredilnaiyneafwsfindkdkdketv vakgadayfeknhhksndpdfalnisekgivyllsffltnkemdslkanltgfkgkv dresgnsikymatqriysfhtyrglkqkirtseegvketllmqmidelskvpnvvyq hlsttqqnsfiedwneyykdyeddvetddlsrvihpvirkryedrfnyfairfldeffd fptlrfqvhlgdyvhdrrtkqlgkvesdriikekvtvfarlkdinsakasyfhsleeqd keeldnkwtlfpnpsydfpkehtlqhqgeqknagkigiyvklrdtqykekaaleea rkslnpkersatkaskydiitqiieandnvksekplvftgqpiaylsmndihsmlfsll tdnaelkktpeeveaklidqigkqineilskdtdtkilkkykdndlketdtdkitrdlar dkeeieklileqkqraddynytsstkfnidksrkrkhllfnaekgkigvwlandikrf mfkeskskwkgyqhtelqklfayfdtsksdlelilsdmvmvkdypielidlvrksrt lvdflnkylearlgyienvitrvknsigtpqfktvrkecfaflkesnytvasldkqieril smplfiergfmdskptmlegksyqqhkedfadwfvhykensnyqnfydtevyei itedkreqakvtkkikqqqkndvftlmmvnymleevlklpsndrlslnelyqtkee rivnkqvakdtqernknyiwnkvvdlqlceglvridkvklkdignfrkyendsrvk efltyqsdivwsgylsnevdsnklyvierqldnyesirskellkevqeiecivynqva nkeslkqsgnenfkqyvlqgllprgtdvremlilstdvkfkkeeimqlgqvreveqd lysliyirnkfahnqlpikeffdfcennyrpisdneyyaeyymeifrsikekyas
HMPREF9699_02005 [Bergeyella zoohelcum ATCC 43767] (SEQ ID No. 95)	EKB54193	menktslgnniyynpfkpqdksyfagyfnaamentdsvfrelgkrlkgkeytsenf fdaifkenislveyeryvkllsdyfpmarlldkkevpikerkenfkknfkgiikavrd lrnfythkehgeveitdeifgvldemlkstvltvkkkkvktdktkeilkksiekqldil cqkkleylrdtarkieekrrnqrergekelvapfkysdkrddliaaiyndafdvyidk kkdslkesskakyntksdpqqeegdlkipiskngvvfllslfltkqeihafkskiagfk atvideatvseatvshgknsicfmatheifshlaykklkrkvrtaeinygeaenaeqls vyaketlmmqmldelskvpdvvyqnlsedvqktfiedwneylkenngdvgtme eeqvihpvirkryedkfnyfairfldefaqfptlrfqvhlgnylhdsrpkenlisdrrik ekitvfgrlselehkkalfikntetnedrehyweifpnpnydfpkenisvndkdfpia gsildrekqpvagkigikvkllnqqyvsevdkavkahqlkqrkaskpsiqniieeiv pinesnpkeaivfggqptaylsmndihsilyeffdkwekkkeklekkgekelrkei gkelekkivgkiqaqiqqiidkdtnakilkpyqdgnstaidkeklikdlkqeqnilqk lkdeqtvrekeyndfiayqdknreinkvrdrnhkqylkdnlkrkypeaparkevly yrekgkvavwlandikrfmptdfknewkgeqhsllqkslayyeqckeelknllpe kvfqhlpfklggyfqqkylyqfytcyldkrleyisglvqqaenfksenkvfkkvene

		cfkflkkqnythkeldarvqsilgypiflergfmdekptiikgktfkgnealfadwfr yykeyqnfqtfydtenyplvelekkqadrkrktkiyqqkkndvftllmakhifksvf kqdsidqfsledlyqsreerlgnqerarqtgerntnyiwnktvdlklcdgkitvenvkl knvgdfikyeydqrvqaflkyeeniewqaflikeskeeenypyvvereieqyekvr reellkevhlieeyilekvkdkeilkkgdnqnfkyyilngllkqlknedvesykvfnl ntepedvninqlkqeatdleqkafvltyirnkfahnqlpkkefwdycqekygkiek ektyaeyfaevfkkekealik
HMPREF9151_01387 [Prevotella saccharolytica F0055] (SEQ ID No. 96)	EKY00089	mmekenvqgshiyyeptdkcfwaafynlarhnayltiahinsfvnskkginnddk vldiiddwskfdndllmgarlnklilkhfpflkaplyqlakrktrkqqgkeqqdyek kgdedpeviqeaianafkmanvrktlhaflkqledlrnhfshynynspakkmevk fddgfcnklyyvfdaalqmvkddnrmnpeinmqtdfehlvrlgrnrkipntfkyn ftnsdgtinnngllffvslflekrdaiwmqkkikgfkggtenymrmtnevfcrnrm vipklrletdydnhqlmfdmlnelvrcplslykrlkqedqdkfrvpiefldednead npyqenansdenpteetdplkntlvrhqhrfpyfvlryfdlnevfkqlrfqinlgcyh fsiydktigertekrhltrtlfgfdrlqnfsvklqpehwknmvkhldteessdkpylsd amphyqienekigihflktdtekketvwpsleveevssnrnkykseknltadaflst hellpmmfyyqllsseektraaagdkvqgvlqsyrkkifdiyddfangtinsmqkl derlakdnllrgnmpqqmlailehqepdmeqkakekldrlitetkkrigkledqfkq kvrigkrradlpkvgsiadwlvndmmrfqpakrnadntgvpdskansteyrllqea lafysaykdrlepyfrqvnliggtnphpflhrvdwkkcnhllsfyhdyleakeqyls hlspadwqkhqhfllkvrkdiqnekkdwkkslvagwkngfnlprglftesiktwf stdadkvqitdtklfenrvgliakliplyydkvyndkpqpfyqypfnindrykpedtr krftaassklwnekkmlyknaqpdssdkieypqyldflswkklerelrmlrnqdm mvwlmckdlfaqctvegvefadlklsqlevdvnvqdnlnvlnnvssmilplsvyp sdaqgnvlrnskplhtvyvqenntkllkqgnfksllkdrrlnglfsfiaaegedlqqhp ltknrleyelsiyqtmrisvfeqtlqlekailtrnktlcgnnfnnllnswsehrtdkktlq pdidfliavrnafshnqypmstntvmqgiekfniqtpkleekdglgiasqlakktkd aasrlqniinggtn
A343_1752 [Porphyromonas gingivalis	EOA10535	mteqnekpyngtyytledkhfwaaffnlarhnayitlthidrqlayskaditndedilf fkgqwknldndlerkarlrslilkhfsflegaaygkklfesqssgnksskkkeltkke keelqanalsldnlksilfdflqklkdfrnyyshyrhpesselplfdgnmlqrlynvfd vsvqrvkrdhehndkvdphrhfnhlvrkgkkdrcgnndnpffkhhfvdreekvte agllffvslflekrdaiwmqkkirgfkggtetyqqmtnevfcrsrislpklkleslrtdd

JCVI SC001] (SEQ ID No. 97)		wmlldmlnelvrcpkslydrlreedrarfrvpvdilsdeddtdgteedpfkntlvrhq drfpyfalryfdlkkvftslrfhidlgtyhfaiykknigeqpedrhltrnlygfgriqdfa eehrpeewkrlvrdldyfetgdkpyitqttphyhiekgkiglrfvpegqllwpspev gatrtgrskyaqdkrftaeaflsvhelmpmmfyyfllrekyseeasaervqgrikrvi edvyavydafargeidtldrldacladkgirrghlprqmiailsqehkdmeekvrkk lqemiadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqpvakdt sgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpflhetrwe shtnilsfyrsylkarkaflqsigrsdrvenhrflllkepktdrqtlvagwkgefhlprgi fteavrdcliemgldevgsykevgfmakavplyferackdrvqpfydypfnvgnsl kpkkgrflskekraeewesgkerfrdleawshsaarriedafagienasrenkkkie qllqdlslwetfesklkvkadkiniaklkkeileakehpyldfkswqkferelrlvkn qdiitwmmcrdlmeenkvegldtgtlylkdirtdvheqgslnvlnrvkpmrlpvv vyradsrghvhkeqaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvdtgalam eqypisklrveyelakyqtarvcafeqtleleeslltryphlpdknfrkmleswsdpll dkwpdlhgnvrlliavrnafshnqypmydetlfssirkydpsspdaieermglnia hrlseevkqakemveriiqa
HMPREF1 981_03090 [Bacteroid es pyogenes F0041] (SEQ ID No. 98)	ERI81700	mesiknsqkstgktlqkdppyfglylnmallnvrkvenhirkwlgdvallpeksgf hsllttdnlssakwtrfyyksrkflpflemfdsdkksyenrretteclдtidrqkissllk evygklqdirnafshyhiddqsvkhtaliissemhrfienaysfalqktrarftgvfvet dflqaeekgdnkkffaiggnegiklkdnalifliclfldreeafkflsratgfkstkekgf lavretfcalccrqpherllsvnpreallmdmlnelnrcpdilfemldekdqksflpll geeeqahilenslndelceaiddpfemiaslskrvryknrfpylmlryieeknllpfir fridlgclelasypkkmgeennyersvtdhamafgrltdfhnedavlqqitkgitdev rfslyapryaiynnkigfvrtggsdkisfptlkkkggeghcvaytlqntksfgfisiydl rkilllsfldkdkaknivsglleqcekhwkdlsenlfdairtelqkefpvplirytlprsk ggklvsskladkqekyeseferrkeklteilsekdfdlsqiprrmidewlnvlptsrek klkgyvetlkldcrerlrvfekrekgehpvpprigematdlakdiirmvidqgvkqri tsayyseiqrclaqyagddnrrhldsiirelrlkdtknghpflgkvlrpglghteklyqr yfeekkewleatfypaaspkrvprfvnpptgkqkelpliirnlmkerpewrdwkqr knshpidlpsqlfeneicrllkdkigkepsgklkwnemfklywdkefpngmqrfy rckrrvevfdkvveyeyseeggnykkyyealidevvrqkissskeksklqvedltls vrrvfkrainekeyqlrllceddrllfmavrdlydwkeaqldldkidnmlgepvsvs qviqleggqpdavikaecklkdvsklmrycydgrvkglmpyfanheatqeqvem

		elrhyedhrrrvfnwvfaleksvlkneklrrfyeesqggcehrrcidalrkaslvseee yeflvhirnksahnqfpdleigklppnvtsgfceciwskykaiicriipfidperrffgk lleqk
HMPREF1 553_02065 [Porphyro monas gingivalis F0568] (SEQ ID No. 99)	ERJ65637	mntvpasenkgqsrtveddpqyfglylnlarenlieveshvrikfgkkklneeslkq sllcdhllsvdrwtkvyghsrrylpflhyfdpdsqiekdhdsktgvdpdsaqrlirely slldflrndfshnrldgttfehlevspdissfitgtyslacgraqsrfadffkpddfvlakn rkeqlisvadgkecltvsglafficlfldreqasgmlsrirgfkrtdenwaravhetfcd lcirhphdrlessntkeallldmlnelnrcprilydmlpeeeraqflpaldensmnnls enslneesrllwdgssdwaealtkrirhqdrfpylmlrfieemdllkgirfrvdlgeiel dsyskkvgrngeydrtitdhalafgklsdfqneeevsrmisgeasypvrfslfaprya iydnkigychtsdpvypksktgekralsnprsmgfisvhdlrklllmellcegsfsrm qsdflrkanrildetaegklqfsalfpemrhrfippqnpkskdrrekaettlekykqei kgrkdklnsqllsafdmdqrqlpsrlldewmnirpashsvklrtyvkqlnedcrlrlq kfrkdgdgkaraiplvgematflsqdivrmiiseetkklitsayynemqrslaqyag eenrhqfraivaelrlldpssghpflsatmetahrytedfykcylekkrewlaktfyrp eqdentkrrisvffvpdgearkllpllirrrmkeqndlqdwirnkqahpidlpshlfds kimellkvkdgkkkwneafkdwwstkypdgmqpfyglrrelnihgksvsyips dgkkfadcythlmektvqdkkrelrtagkpvppdlaadikrsfhravnerefmlrlv qeddrlmlmainkmmtdreedilpglknidsildeenqfslavhakvlekegegg dnslslvpatieikskrkdwskyiryrydrrvpglmshfpehkatldevktllgeydr crikifdwafalegaimsdrdlkpylhesssregksgehstlvkmlvekkgcltpde sqylilirnkaahnqfpcaaempliyrdvsakvgsiegssakdlpegsslvdslwkk yemiirkilpildpenrffgkllnnmsqpindl
HMPREF1 988_01768 [Porphyro monas gingivalis F0185] (SEQ ID No. 100)	ERJ81987	mntvpasenkgqsrtveddpqyfglylnlarenlieveshvrikfgkkklneeslkq sllcdhllsvdrwtkvyghsrrylpflhyfdpdsqiekdhdsktgvdpdsaqrlirely slldflrndfshnrldgttfehlevspdissfitgtyslacgraqsrfadffkpddfvlakn rkeqlisvadgkecltvsglafficlfldreqasgmlsrirgfkrtdenwaravhetfcd lcirhphdrlessntkeallldmlnelnrcprilydmlpeeeraqflpaldensmnnls enslneesrllwdgssdwaealtkrirhqdrfpylmlrfieemdllkgirfrvdlgeiel dsyskkvgrngeydrtitdhalafgklsdfqneeevsrmisgeasypvrfslfaprya iydnkigychtsdpvypksktgekralsnpqsmgfisvhdlrklllmellcegsfsr mqsgflrkanrildetaegklqfsalfpemrhrfippqnpkskdrrekaettlekykq eikgrkdklnsqllsafdmnqrqlpsrlldewmnirpashsvklrtyvkqlnedcrlr

		lrkfrkdgdgkaraiplvgematflsqdivrmiiseetkklitsayynemqrslaqya geenrrqfraivaelhlldpssghpflsatmetahrytedfykcylekkrewlaktfyr peqdentkrrisvffvpdgearkllpllirrrmkeqndlqdwirnkqahpidlpshlf dskimellkvkdgkkkwneafkdwwstkypdgmqpfyglrrelnihgksvsyi psdgkkfadcythlmektvqdkkrelrtagkpvppdlaadikrsfhravnerefmlr lvqeddrlmlmainkmmtdreedilpglknidsildeenqfslavhakvlekegeg gdnslslvpatieikskrkdwskyiryrydrrvpglmshfpehkatldevktllgeyd rcrikifdwafalegaimsdrdlkpylhesssregksgehstlvkmlvekkgcltpde sqylilirnkaahnqfpcaaempliyrdvsakvgsiegssakdlpegsslvdslwkk yemiirkilpildhenrffgkllnnmsqpindl
HMPREF1 990_01800 [Porphyro monas gingivalis W4087] (SEQ ID No. 101)	ERJ87335	mntvpasenkgqsrtveddpqyfglylnlarenlieveshvrikfgkkklneeslkq sllcdhllsvdrwtkvyghsrrylpflhyfdpdsqiekdhdsktgvdpdsaqrlirely slldflrndfshnrldgttfehlevspdissfitgtyslacgraqsrfadffkpddfvlakn rkeqlisvadgkecltvsglafficlfldreqasgmlsrirgfkrtdenwaravhetfcd lcirhphdrlessntkeallldmlnelnrcprilydmlpeeeraqflpaldensmnnls enslneesrllwdgssdwaealtkrirhqdrfpylmlrfieemdllkgirfrvdlgeiel dsyskkvgrngeydrtitdhalafgklsdfqneeevsrmisgeasypvrfslfaprya iydnkigychtsdpvypksktgekralsnprsmgfisvhdlrklllmellcegsfsrm qsdflrkanrildetaegklqfsalfpemrhrfippqnpkskdrrekaettlekykqei kgrkdklnsqllsafdmdqrqlpsrlldewmnirpashsvklrtyvkqlnedcrlrlq kfrkdgdgkaraiplvgematflsqdivrmiiseetkklitsayynemqrslaqyag eenrhqfraivaelrlldpssghpflsatmetahrytedfykcylekkrewlaktfyrp eqdentkrrisvffvpdgearkllpllirrrmkeqndlqdwirnkqahpidlpshlfds kvmellkvkdgkkkwneafkdwwstkypdgmqpfyglrrelnihgksvsyips dgkkfadcythlmektvrdkkrelrtagkpvppdlaayikrsfhravnerefmlrlv qeddrlmlmainkimtdreedilpglknidsildkenqfslavhakvlekegeggd nslslvpatieikskrkdwskyiryrydrrvpglmshfpehkatldevktllgeydrcr ikifdwafalegaimsdrdlkpylhesssregksgehstlvkmlvekkgcltpdesq ylilirnkaahnqfpcaaeipliyrdvsakvgsiegssakdlpegsslvdslwkkye miirkilpildpenrffgkllnnmsqpindl
M573_117 042 [Prevotella	KJJ86756	mkmeddkkttestnmldnkhfwaaflnlarhnvyitvnhinkvlelknkkdqdiii dndqdilaikthwekvngdlnkterlrelmtkhfpfletaiytknkedkeevkqekq aeaqsleslkdclflfleklqearnyyshykysestkepmleegllekmynifddniq

intermediaZT] (SEQ ID No. 102)		lvikdyqhnkdinpdedfkhldrkgqfkysfadnegnitesgllffvslflekkdaiwmqqkltgfkdnreskkkmthevfcrrrmllpklrlestqtqdwilldmlnelircpkslyerlqgeyrkkfnvpfdsadedydaeqepfkntlvrhqdrfpyfalryfdyneiftnlrfqidlgtyhfsiykkliggqkedrhlthklygferiqefakqnrpdewkalvkdldtyetsneryisettphyhlenqkigirfrngnkeiwpslktngennekskykldkpyqaeaflsvhellpmmfyylllkkeepnndkknasivegfikreirdmyklydafangeinnigdlekycedkgipkrhlpkqmvailydepkdmvkeakrkqkemvkdtkkllatlekqtqeeiedggrnirllksgeiarwlvndmmrfqpvqkdnegnplnnskansteyqmlqrslalynkeekptryfrqvnlinssnphpflkwtkweecnnilsfyrnyltkkieflnklkpedweknqyflklkepktnretlvqgwkngfnlprgiftepirewfkrhqndskeyekvealkrvglvtkviplffkeeyfkedaqkeinncvqpfysfpynvgnihkpdekdflpseerkklwgdkkdkfkgykakvkskkltdkekeeyrsylefqswnkferelrlvrnqdivtwllctelidkmkveglnveelqklrlkdidtdtakqekennilnrimpmqlpvtvyeiddshnivkdrplhtvyieetktkllkqgnfkalvkdrrlnglfsfvdtsskaelkdkpisksvveyelgeyqnarietikdmlllektlikkyeklptdnfsdmlngwlegkdesdkarfqndvkllvavrnafshnqypmrnriafaninpfslssadiseekkldianqlkdkthkiikkiieiekpietke
A2033_10205 [Bacteroidetes bacterium GWA2_31_9] (SEQ ID No. 103)	OFX18020.1	menqtqkgkgiyyyytknedkhyfgsflnlannnieqiieefrirlslkdeknikeiinnyftdkksytdwerginilkeylpvidyldlaitdkefekidlkqketakrkyfrtnfsllidtiidlrnfythyfhkpisinpdvakfldknllnvcldikkqkmktdktkqalkdgldkelkklielkkaelkekkiktwnitenvegavyndafnhmvyknnagvtilkdyhksilpddkidselklnfsisglvfllsmflskkeieqfksnlegfkgkvigengeyeiskfnnslkymathwifsyltfkglkqrvkntfdketllmqmidelnkvphevyqtlskeqqnefledineyvqdneenkksmensivvhpvirkryddkfnyfairfldefanfptlkffvtagnfvhdkrekqiqgsmltsdrmikekinvfgklteiakyksdyfsnentletsewelfpnpsylliqnnipvhidlihnteeakqcqiaidrikcttnpakkrntrkskeeiikiiyqknknikygdptallssnelpaliyellvnkksgkeleniivekivnqyktiagfekgqnlsnslitkklkksepnedkinaekiilainreleitenklniiknnraeftrtgakrkhifyskelgqeatwiaydlkrfmpeasrkewkgfhhselqkflafydrnkndakallnmfwnfdndqlignd lnsafrefhfdkfyekylikrdeilegfksfisnfkdepkllkkgikdiyrvfdkryyiikstnaqkeqllskpiclprgifdnkptyiegvkvesnsalfadwyqytysdkhefqsfydmprdykeqfekfelnniksiqnkknlnksdkfiyfrykqdlkikqiksqdlfiklmvdelfnvvfknnielnlkklyqtsderfknqli

		advqknrekgdtsdnkmnenfiwnmtiplslcngqieepkvklkdigkfrkletdd kviqlleydkskvwkkleiedelenmpnsyerirrekllkgiqefehfllekekfdgi nhpkhfeqdlnpnfktyvingvlrknsklnyteidkllldlehisikdietsakeihlayf lihvrnkfghnqlpkleafelmkkyykknneetyaeyfhkvssqivnefknslekh s
SAMN054 21542_066 6 [Chryseoba cterium jejuense] (SEQ ID No. 104)	SDI27289 .1	mektqtglgiyydhtklqdkyffggffnlaqnnidnvikafiikffperkdkdiniaq fldicfkdndadsdfqkknkflrihfpvigfltsdndkagfkkkfalllktiselrnfyth yyhksiefpselfellddifvkttseikklkkkddktqqllnknlseeydiryqqqierl kelkaqgkrvsltdetairngvfnaafnhliyrdgenvkpsrlyqssysepdpaengi slsqnsilfllsmflerketedlksrvkgfkakiikqgeeqisglkfmathwvfsylcf kgikqklstefheetlliqiidelskvpdevysafdsktkekfledineymkegnadls ledskvihpvirkryenkfnyfairfldeylsstslkfqvhvgnyvhdrrvkhingtgf qterivkdrikvfgrlsnisnlkadyikeqlelpndsngweifpnpsyifidnnvpih vladeatkkgielfkdkrrkeqpeelqkrkgkiskynivsmiykeakgkdklridep lallslneipallyqilekgatpkdieliiknklterfekiknydpetpapasqiskrlrnn ttakgqealnaeklsllereientetklssieekrlkakkeqrrntpqrsifsnsdlgriaa wladdikrfmpaeqrknwkgyqhsqlqqslayfekrpqeafllllkegwdtsdgss ywnnwvmnsflennhfekfyknylmkrvkyfselagnikqhthntkflrkfikqq mpadlfpkrhyilkdleteknkvlskplvfsrglfdnnptfikgvkvtenpelfaewy sygyktehvfqhfygwerdynelldselqkgnsfaknsiyynresqldliklkqdlki kkikiqdlflkriaeklfenvfnypttlsldefyltqeeraekerialaqslreegdnspni ikddfiwsktiafrskqiyepaiklkdigkfnrfvlddeeskaskllsydknkiwnke qlerelsigensyevirreklfkeiqnlelqilsnwswdginhprefemedqkntrhp nfkmylvngilrkninlykededfwleslkendfktlpsevletksemvqllflvilir nqfahnqlpeiqfynfirknypeiqnntvaelylnliklavqklkdns
SAMN054 44360_113 66 [Chryseoba cterium carnipullor um] (SEQ	SHM5281 2.1	mntrvtgmgvsydhtkkedkhffggflnlaqdnitavikafcikfdknpmssvqfa escftdkdsdtdfqnkvryvrthlpvigylnyggdrntfrqklstllkavdslrnfythy yhsplalstelfelldtvfasvavevkqhkmkddktrqllskslaeeldirykqqlerlk elkeqgknidlrdeagirngvlnaafnhliykegeiakptlsyssfyygadsaengiti sqsgllfllsmflgkkeiedlksrirgfkakivrdgeenisglkfmathwifsylsfkg mkqrlstdfheetlliqiidelskvpdevyhdfdtatrekfvedineyiregnedfslg dstiihpvirkryenkfnyfavrfldefikfpslrfqvhlgnfvhdrrikdihgtgfqter vvkdrikvfgklseisslkteyiekeldldsdtgweifpnpsyvfidnnipiyistnktf

ID No. 105)		kngssefiklrrkekpeemkmrgedkkekrdiasmignagslnsktplamlslne mpallyeilvkkttpeeieliikekldshfeniknydpekplpasqiskrlrnnttdkg kkvinpeklihlinkeidateakfallaknrkelkekfrgkplrqtifsnmelgreatwl addikrfmpdilrknwkgyqhnqlqqslaffnsrpkeaftilqdgwdfadgssfwn gwiinsfvknrsfeyfyeayfegrkeyfsslaenikqhtsnhrnlrrfidqqmpkglf enrhyllenleteknkilskplvfprglfdtkptfikgikvdeqpelfaewyqygyste hvfqnfygwerdyndlleselekdndfsknsihysrtsqleliklkqdlkikkikiqdl flkliaghifenifkypasfsldelyltqeerlnkeqealiqsqrkegdhsdniikdnfig sktvtyeskqisepnvklkdigkfnrflldkvktllsynedkvwnkndldlelsige nsyevirreklfkkiqnfelqtltdwpwngtdhpeefgttdnkgvnhpnfkmyvv ngilrkhtdwfkegednwlenlnethfknlsfqeletksksiqtafliimirnqfahnq lpavqffefiqkkypeiqgsttselylnfinlavvellellek
SAMN05421786_1011119 [Chryseobacterium ureilyticum] (SEQ ID No. 106)	SIS70481.1	metqilgngisydhtktedkhffggflntaqnnidllikayiskfessprklnsvqfpd vcfkkndsdadfqhklqfirkhlpviqylkyggnrevlkekfrlllqavdslrnfythf yhkpiqlpnelltlldtifgeignevrqnkmkddktrhllkknlseeldfryqeqlerlr klksegkkvdlrdteairngvlnaafnhlifkdaedfkptvsyssyyydsdtaengisi sqsgllfllsmflgrremedlksrvrgfkariikheeqhvsglkfmathwvfsefcfk giktrlnadyheetlliqlidelskvpdelyrsfdvatrerfiedineyirdgkedkslies kivhpvirkryeskfnyfairfldefvnfptlrfqvhagnyvhdrriksiegtgfkterl vkdrikvfgklstisslkaeylakavnitddtgwellphpsyvfidnnipihltvdpsf kngvkeyqekrklqkpeemknrqggdkmhkpaisskigkskdinpespvalls mneipallyeilvkkaspeeveakirqkltavferirdydpkvplpasqvskrlrnnt dtlsynkeklvelankeveqterklalitknrrecrekvkgkfkrqkvfknaelgteat wlandikrfmpeeqkknwkgyqhsqlqqslaffesrpgearsllqagwdfsdgssf wngwvmnsfardntfdgfyesylngrmkyflrladniaqqsstnklisnfikqqm pkglfdrrlymledlateknkilskplifprgifddkptfkkgvqvseepeafadwys ygydvkhkfqefyawdrdyeellreelekdtaftknsihysresqiellakkqdlkvk kvriqdlylklmaeflfenvfghelalpldqfyltqeerlkqeqeaivqsqrpkgdds pnivkenfiwsktipfksgrvfepnvklkdigkfrnlltdekvdillsynnteigkqvi eneliigagsyefirreqlfkeiqqmkrlslrsvrgmgvpirlnlk
Prevotella buccae	WP_004343581	mqkqdklfvdrkknaifafpkyitimenqekpepiyyeltdkhfwaaflnlarhnv yttinhinrrleiaelkddgymmdikgswneqakkldkkvrlrdlimkhfpfleaaa yeitnskspnnkeqrekeqsealslnnlknvlfifleklqvlrnyyshykyseespkp

(SEQ ID No. 107)		ifetsllknmykvfdanvrlvkrdymhhenidmqrdfthlnrkkqvgrtkniidsp nfhyhfadkegnmtiagllffvslfldkkdaiwmqkklkgfkdgrnlreqmtnevf crsrislpklklenvqtkdwmqldmlnelvrcpkslyerlrekdresfkvpfdifsdd ydaeeepfkntlvrhqdrfpyfvlryfdlneifeqlrfqidlgtyhfsiynkrigdedev rhlthhlygfariqdfaqqnqpevwrklvkdldyfeasqepyipktaphyhleneki gikfcsthnnlfpslktektcngrskfnlgtqftaeaflsvhellpmmfyylllkdysr kesadkvegiirkeisniyaiydafangeinsiadltcrlqktnilqghlpkqmisileg rqkdmekeaerkigemiddtqrrldllckqtnqkirigkrnagllksgkiadwlvnd mmrfqpvqkdqnnipinnskansteyrmlqralalfgsenfrlkayfnqmnlvgn dnphpflaetqwehqtnilsfyrnylearkkylkglkpqnwkqyqhflilkvqktnr ntlvtgwknsfnlprgiftqpirewfekhnnskriydqilsfdrvgfvakaiplyfaee ykdnvqpfydypfnignklkpqkgqfldkkervelwqknkelfknypsekkktdl ayldflswkkferelrliknqdivtwlmfkelfnmatveglkigeihlrdidtntanee snnilnrimpmklpvktyetdnkgnilkerplatfyieetetkvlkqgnfkvlakdrrl ngllsfaettdidleknpitklsvdhelikyqttrisifemtlglekklinkyptlptdsfrn mlerwlqckanrpelknyvnsliavrnafshnqypmydatlfaevkkftlfpsvdtk kielniapqlleivgkaikeieksenkn
Porphyromonas gingivalis (SEQ ID No. 108)	WP_005873511	mntvpasenkgqsrtveddpqyfglylnlarenlieveshvrikfgkkklneeslkq sllcdhllsvdrwtkvyghsrrylpflhyfdpdsqiekdhdsktgvdpdsaqrlirely slldflrndfshnrldgttfehlevspdissfitgtyslacgraqsrfadffkpddfvlakn rkeqlisvadgkecltvsglafficlfldreqasgmlsrirgfkrtdenwaravhetfcd lcirhphdrlessntkeallldmlnelnrcprilydmlpeeeraqflpaldensmnnls enslneesrllwdgssdwaealtkrirhqdrfpylmlrfieemdllkgirfrvdlgeiel dsyskkvgrngeydrtitdhalafgklsdfqneeevsrmisgeasypvrfslfaprya iydnkigychtsdpvypksktgekralsnpqsmgfisvhnlrklllmellcegsfsr mqsdflrkanrildetaegklqfsalfpemrhrfippqnpkskdrrekaettlekykq eikgrkdklnsqllsafdmnqrqlpsrlldewmnirpashsvklrtyvkqlnedcrlr lrkfrkdgdgkaraiplvgematflsqdivrmiiseetkklitsayynemqrslaqya geenrrqfraivaelhlldpssghpflsatmetahrytedfykcylekkrewlaktfyr peqdentkrrisvffvpdgearkllpllirrrmkeqndlqdwirnkqahpidlpshlf dskimellkvkdgkkkwneafkdwwstkypdgmqpfyglrrelnihgksvsyi psdgkkfadcythlmektvqdkkrelrtagkpvppdlaadikrsfhravnerefmlr lvqeddrlmlmainkmmtdreedilpglknidsildeenqfslavhakvlekegeg

		gdnslslvpatieikskrkdwskyiryrydrrvpglmshfpehkatldevktllgeyd rcrikifdwafalegaimsdrdlkpylhesssregksgehstlvkmlvekkgcltpde sqylilirnkaahnqfpcaaempliyrdvsakvgsiegssakdlpegsslvdslwkk yemiirkilpildpenrffgkllnnmsqpindl
Porphyromonas gingivalis (SEQ ID No. 109)	WP_005874195	mteqnekpyngtyytledkhfwaaffnlarhnayitlahidrqlayskaditndedil ffkgqwknldndlerkarlrslilkhfsflegaaygkklfesqssgnkssskkkeltkke keelqanalsldnlksilfdflqklkdfrnyyshyrhpesselplfdgnmlqrlynvfd vsvqrvkrdhehndkvdphrhfnhlvrkgkkdkygnndnpffkhhfvdreekvt eagllffvslflekrdaiwmqkkirgfkggteayqqmtnevfcrsrislpklkleslrtd dwmlldmlnelvrcpkslydrlreedrarfrvpvdilsdeddtdgteedpfkntlvrh qdrfpyfalryfdlkkvftslrfhidlgtyhfaiykknigeqpedrhltrnlygfgriqdf aeehrpeewkrlvrdldyfetgdkpyitqttphyhiekgkiglrfvpegqllwpspe vgatrtgrskyaqdkrftaeaflsvhelmpmmfyyfllrekyseeasaekvqgrikr viedvyavydafardeintrdeldacladkgirrghlprqmiailsqehkdmeekvr kklqemiadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqpva kdtsgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpflhetr weshtnilsfyrsylkarkaflqsigrsdreenhrflllkepktdrqtlvagwksefhlp rgifteavrdcliemgydevgsykevgfmakavplyferackdrvqpfydypfnv gnslkpkkgrflskekraeewesgkerfrdleawshsaarriedafvgieyaswenk kkieqllqdlslwetfesklkvkadkiniaklkkeileakehpyhdfkswqkferelrl vknqdiitwmmcrdlmeenkvegldtgtlylkdirtdvqeqgslnvlnhvkpmrl pvvvyradsrghvhkeeaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvdtgal ameqypisklrveyelakyqtarvcafeqtleleeslltryphlpdesfremleswsd plldkwpdlqrevrlliavrnafshnqypmydetifssirkydpssldaieermglni ahrlseevklakemveriiqa
Prevotella pallens (SEQ ID No. 110)	WP_006044833	mkeeekgktpvvstynkddkhfwaaflnlarhnvyitvnhinkilgegeinrdgye ntlekswneikdinkkdrlskliikhfpflevttyqrnsadttkqkeekqaeaqslesl kksffvfiyklrdlrnhyshykhskslerpkfeedlqekmynifdasiqlvkedykh ntdikteedfkhldrkgqfkysfadnegnitesgllffvslflekkdaiwvqkklegfk csnesyqkmtnevfcrsrmllpklrlqstqtqdwilldmlnelircpkslyerlreedr kkfrvpieiadedydaeqepfknalvrhqdrfpyfalryfdyneiftnlrfqidlgtyh fsiykkqigdykeshhlthklygferiqeftkqnrpdewrkfvktfnsfetskepyip ettphyhlenqkigirfrndndkiwpslktnseknekskykldksfqaeaflsvhell

		pmmfyylllktentdndneietkkkenkndkqekhkieeiienkiteiyalydafan gkinsidkleeyckgkdieighlpkqmiailksehkdmateakrkqeemladvqk slesldnqineeienverknsslksgeiaswlvndmmrfqpvqkdnegnplnnsk ansteyqmlqrslalynkeekptryfrqvnliessnphpflnntewekcnnilsfyrs yleakknfleslkpedweknqyflmlkepktncetlvqgwkngfnlprgiftepirk wfmehrknitvaelkrvglvakviplffseeykdsvqpfynylfnvgninkpdekn flnceerrellrkkkdefkkmtdkekeenpsylefqswnkferelrlvrnqdivtwll cmelfnkkkikelnvekiylknintnttkkeknteekngeekiikeknnilnrimp mrlpikvygrenfsknkkkkirrntfftvyieekgtkllkqgnfkalerdrrlgglfsfv kthskaesksntisksrveyelgeyqkarieiikdmlaleetlidkynsldtdnfhnml tgwlklkdepdkasfqndvdlliavrnafshnqypmrnriafaninpfslssantsee kglgianqlkdkthktiekiieiekpietke
Myroides odoratimimus (SEQ ID No. 111)	WP_006261414	mkdilttdttekqnrfyshkiadkyffggyfnlasnniyevfeevnkrntfgklakrd ngnlknyiihvfkdelsisdfekrvaifasyfpiletvdkksikernrtidltlsqrirqfr emlislvtavdqlrnfythyhhseivienkvldflnsslvstalhvkdkylktdktkefl ketiaaeldilieaykkkqiekkntrfkankredilnaiyneafwsfindkdkdketv vakgadayfeknhhksndpdfalnisekgivyllsffltnkemdslkanltgfkgkv dresgnsikymatqriysfhtyrglkqkirtseegvketllmqmidelskvpnvvyq hlsttqqnsfiedwneyykdyeddvetddlsrvihpvirkryedrfnyfairfldeffd fptlrfqvhlgdyvhdrrtkqlgkvesdriikekvtvfarlkdinsakanyfhsleeqd keeldnkwtlfpnpsydfpkehtlqhqgeqknagkigiyvklrdtqykekaaleea rkslnpkersatkaskydiitqiieandnvksekplvftgqpiaylsmndihsmlfsll tdnaelkktpeeveaklidqigkqineilskdtdtkilkkykdndlketdtdkitrdlar dkeeieklileqkqraddynytsstkfnidksrkrkhllfnaekgkigvwlandikrf mteefkskwkgyqhtelqklfayydtsksdldlilsdmvmvkdypielialvkksrt lvdflnkylearlgymenvitrvknsigtpqfktvrkecftflkksnytvvsldkqver ilsmplfiergfmddkptmlegksyqqhkekfadwfvhykensnyqnfydtevy eittedkrekakvtkkikqqqkndvftlmmvnymleevlklssndrlslnelyqtke erivnkqvakdtqernknyiwnkvvdlqlceglvridkvklkdignfrkyendsrv kefltyqsdivwsaylsnevdsnklyvierqldnyesirskellkevqeiecsvynqv ankeslkqsgnenfkqyvlqglvpigmdvremlilstdvkfikeeiiqlgqageveq dlysliyirnkfahnqlpikeffdfcennyrsisdneyyaeyymeifrsikekyts

Myroides odoratimimus (SEQ ID No. 112)	WP_006265509	mkdilttdttekqnrfyshkiadkyffggyfnlasnniyevfeevnkrntfgklakrd ngnlknyiihvfkdelsisdfekrvaifasyfpiletvdkksikernrtidltlsqrirqfr emlislvtavdqlrnfythyhhseivienkvldflnsslvstalhvkdkylktdktkefl ketiaaeldilieaykkkqiekkntrfkankredilnaiyneafwsfindkdkdketv vakgadayfeknhhksndpdfalnisekgivyllsffltnkemdslkanltgfkgkv dresgnsikymatqriysfhtyrglkqkirtseegvketllmqmidelskvpnvvyq hlsttqqnsfiedwneyykdyeddvetddlsrvihpvirkryedrfnyfairfldeffd fptlrfqvhlgdyvhdrrtkqlgkvesdriikekvtvfarlkdinsakasyfhsleeqd keeldnkwtlfpnpsydfpkehtlqhqgeqknagkigiyvklrdtqykekaaleea rkslnpkersatkaskydiitqiieandnvksekplvftgqpiaylsmndihsmlfsll tdnaelkktpeeveaklidqigkqineilskdtdtkilkkykdndlketdtdkitrdlar dkeeieklileqkqraddynytsstkfnidksrkrkhllfnaekgkigvwlandikrf mfkeskskwkgyqhtelqklfayfdtsksdlelilsdmvmvkdypielidlvrksrt lvdflnkylearlgyienvitrvknsigtpqfktvrkecfaflkesnytvasldkqieril smplfiergfmdskptmlegksyqqhkedfadwfvhykensnyqnfydtevyei itedkreqakvtkkikqqqkndvftlmmvnymleevlklpsndrlslnelyqtkee rivnkqvakdtqernknyiwnkvvdlqlceglvridkvklkdignfrkyendsrvk efltyqsdivwsgylsnevdsnklyvierqldnyesirskellkevqeiecivynqva nkeslkqsgnenfkqyvlqgllprgtdvremlilstdvkfkkeeimqlgqvreveqd lysliyirnkfahnqlpikeffdfcennyrpisdneyyaeyymeifrsikekyas
Prevotella sp. MSX73 (SEQ ID No. 113)	WP_007412163	mqkqdklfvdrkknaifafpkyitimenqekpepiyyeltdkhfwaaflnlarhnv yttinhinrrleiaelkddgymmgikgswneqakkldkkvrlrdlimkhfpfleaaa yeitnskspnnkeqrekeqsealslnnlknvlfifleklqvlrnyyshykyseespkp ifetsllknmykvfdanvrlvkrdymhhenidmqrdfthlnrkkqvgrtkniidsp nfhyhfadkegnmtiagllffvslfldkkdaiwmqkklkgfkdgrnlreqmtnevf crsrislpklklenvqtkdwmqldmlnelvrcpkslyerlrekdresfkvpfdifsdd ydaeeepfkntlvrhqdrfpyfvlryfdlneifeqlrfqidlgtyhfsiynkrigdedev rhlthhlygfariqdfapqnqpeewrklvkdldhfetsqepyisktaphyhlenekig ikfcsthnnlfpslkrektcngrskfnlgtqftaeaflsvhellpmmfyyllltkdysrk esadkvegiirkeisniyaiydafanneinsiadltcrlqktnilqghlpkqmisilegr qkdmekeaerkigemiddtqrrldllckqtnqkirigkrnagllksgkiadwlvsd mmrfqpvqkdtnnapinnskansteyrmlqhalalfgsessrlkayfrqmnlvgn anphpflaetqwehqtnilsfyrnylearkkylkglkpqnwkqyqhflilkvqktnr

		ntlvtgwknsfnlprgiftqpirewfekhnnskriydqilsfdrvgfvakaiplyfaee ykdnvqpfydypfnignklkpqkgqfldkkervelwqknkelfknypseknktdl ayldflswkkferelrliknqdivtwlmfkelfktttveglkigeihlrdidtntanees nnilnrimpmklpvktyetdnkgnilkerplatfyieetetkvlkqgnfkvlakdrrl ngllsfaettdidleknpitklsvdyelikyqttrisifemtlglekklidkystlptdsfrn mlerwlqckanrpelknyvnsliavrnafshnqypmydatlfaevkkftlfpsvdtk kielniapqlleivgkaikeieksenkn
Porphyromonas gingivalis (SEQ ID No. 114)	WP_012458414	mteqnerpyngtyytledkhfwaaffnlarhnayitlahidrqlayskaditndedilf fkgqwknldndlerkarlrslilkhfsflegaaygkklfesqssgnksskkkeltkke keelqanalsldnlksilfdflqklkdfrnyyshyrhpesselplfdgnmlqrlynvfd vsvqrvkrdhehndkvdphrhfnhlvrkgkkdrygnndnpffkhhfvdreekvte agllffvslflekrdaiwmqkkirgfkggtetyqqmtnevfcrsrislpklkleslrtdd wmlldmlnelvrcpkslydrlreedrarfrvpvdilsdeddtdgteedpfkntlvrhq drfpyfalryfdlkkvftslrfhidlgtyhfaiykknigeqpedrhltrnlygfgriqdfa eehrpeewkrlvrdldyfetgdkpyitqttphyhiekgkiglrfvpegqhlwpspev gatrtgrskyaqdkrltaeaflsvhelmpmmfyyfllrekysdeasaervqgrikrvi edvyavydafargeintrdeldacladkgirrghlprqmigilsqehkdmeekvrk klqemivdtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqpvak dtsgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpflhetrw eshtnilsfyrsylkarkaflqsigrsdrvenhrflllkepktdrqtlvagwkgefhlprg ifteavrdcliemgldevgsykevgfmakavplyferackdrvqpfydypfnvgns lkpkkgrflskekraeewesgkerfrlaklkkeileakehpyldfkswqkferelrlv knqdiitwmicrdlmeenkvegldtgtlylkdirtdvqeqgnlnvlnrvkpmrlpv vvyradsrghvhkeqaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvdtgala meqypisklrveyelakyqtarvcafeqtleleeslltryphlpdknfrkmleswsdp lldkwpdlhgnvrlliavrnafshnqypmydeavfssirkydpsspdaieermgln iahrlseevkqakemaeriiqa
Paludibacter propionicigenes (SEQ ID No. 115)	WP_013446107	mktsanniyfnginsfkkifdskgaiapiaekscrnfdikaqndvnkeqrihyfavg htfkqldtenlfeyvldenlrakrptrfislqqfdkefienikrlisdirninshyihrfdpl kidavptniidflkesfelaviqiylkekginylqfsenphadqklvaflhdkflplde kktsmlqnetpqlkeykeyrkyfktlskqaaidqllfaeketdyiwnlfdshpvltisa gkylsfysclfllsmflykseanqliskikgfkkntteeekskreiftffskrfnsmdid seenqlvkfrdlilylnhypvawnkdleldssnpamtdklkskiieleinrsfplyeg

		nerfatfakyqiwgkkhlgksiekeyinasftdeeitaytyetdtcpelkdahkkladl kaakglfgkrkeknesdikktetsirelqhepnpikdkliqrieknlltvsygrnqdrf mdfsarflaeinyfgqdasfkmyhfyatdeqnselekyelpkdkkkydslkfhqg klvhfisykehlkryeswddafviennaiqlklsfdgventvtiqralliylledalrni qnntaenagkqllqeyyshnkadlsafkqiltqqdsiepqqktefkkllprrllnnysp ainhlqtphsslplilekallaekrycslvvkakaegnyddfikrnkgkqfklqfirka wnlmyfrnsylqnvqaaghhksfhierdefndfsrymfafeelsqykyylnemfe kkgffennefkilfqsgtslenlyektkqkfeiwlasntaktnkpdnyhlnnyeqqfs nqlffinlshfinylkstgklqtdangqiiyealnnvqylipeyyytdkpersesksgn klynklkatkledallyemamcylkadkqiadkakhpitklltsdvefnitnkegiql yhllvpfkkidafiglkmhkeqqdkkhptsflanivnylelvkndkdirktyeafstn pvkrtltyddlakidghlisksikftnvtleleryfifkeslivkkgnnidfkyikglrny ynnekkknegirnkafhfgipdsksydqlirdaevmfianevkpthatkytdlnkql htvcdklmetvhndyfskegdgkkkreaagqkyfeniisak
Porphyromonas gingivalis (SEQ ID No. 116)	WP_013816155	mteqnekpyngtyytledkhfwaaffnlarhnayitlahidrqlayskaditndedil ffkgqwknldndlerkarlrslilkhfsflegaaygkklfesqssgnkssknkeltkke keelqanalsldnlksilfdflqklkdfrnyyshyrhpesselplfdgnmlqrlynvfd vsvqrvkrdhehndkvdphrhfnhlvrkgkkdrygnndnpffkhhfvdregtvte agllffvslflekrdaiwmqkkirgfkggtetyqqmtnevfcrsrislpklkleslrtdd wmlldmlnelvrcpkslydrlreedrarfrvpvdilsdeedtdgaeedpfkntlvrhq drfpyfalryfdlkkvftslrfqidlgtyhfaiykknigeqpedrhltrnlygfgriqdfa eehrpeewkrlvrdldyfetgdkpyitqttphyhiekgkiglrfvpegqhlwpspev gatrtgrskyaqdkrftaeaflsahelmpmmfyyfllrekyseeasaervqgrikrvi edvyavydafardeintrdeldacladkgirrghlprqmigilsqehkdmeekirkk lqemmadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqpvak dtsgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpflhetrw eshtnilsfyrsylkarkaflqsigrsdrvenhrflllkepktdrqtlvagwkgefhlprg ifteavrdcliemgldevgsykevgfmakavplyferackdwvqpfynypfnvgn slkpkkgrflskekraeewesgkerfrlaklkkeileakehpyldfkswqkferelrl vknqdiitwmicgdlmeenkvegldtgtlylkdirtdvqeqgslnvlnrvkpmrlp vvvyradsrghvhkeqaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvdtgala meqypisklrveyelakyqtarvcafeqtleleeslltrcphlpdknfrkmleswsdp

		lldkwpdlhrkvrlliavrnafshnqypmydeavfssirkydpsfpdaieermglni ahrlseevkqaketveriiqa
Flavobacterium columnare (SEQ ID No. 117)	WP_014165541	mssknesynkqktfnhykqedkyffggflnnaddnlrqvgkefktrinfnhnnnel asvfkdyfnkeksvakrehalnllsnyfpvleriqkhtnhnfeqtreifellldtikklrd yythhyhkpitinpkiydfldtlldvlitikkkkvkndtsrellkeklrpeltqlknqk reelikkgkklleenlenavfnhclrpfleenktddkqnktvslrkyrkskpneetsitl tqsglvflmsfflhrkefqvftsglegfkakvntikeeeislnknnivymithwsysy ynfkglkhriktdqgvstleqnntthsltntntkealltqivdylskvpneiyetlsekq qkefeedineymrenpenedstfssivshkvirkryenkfnyfamrfldeyaelptlr fmvnfgdyikdrqkkilesiqfdseriikkeihlfeklslvteykknvylketsnidlsr fplfpnpsyvmannnipfyidsrsnnldeylnqkkkaqsqnkkrnltfekynkeqs kdaiiamlqkeigvkdlqqrstigllscnelpsmlyevivkdikgaelenkiaqkire qyqsirdftldspqkdnipttliktintdssvtfenqpidiprlknaiqkeltltqekllnv keheievdnynrnkntykfknqpknkvddkklqrkyvfyrneirqeanwlasdli hfmknkslwkgymhnelqsflaffedkkndcialletvfnlkedciltkglknlflkh gnfidfykeylklkedflntestflengliglppkilkkelskrfkyifivfqkrqfiikel eekknnlyadainlsrgifdekptmipfkkpnpdefaswfvasyqynnyqsfyelt pdiverdkkkkyknlrainkvkiqdyylklmvdtlyqdlfnqpldkslsdfyvska erekikadakayqkrndsslwnkvihlslqnnritanpklkdigkykralqdekiatl ltyddrtwtyalqkpekenendykelhytalnmelqeyekvrskellkqvqelekqi leeytdflstqihpadferegnpnfkkylahsileneddldklpekveamreldetitn piikkaivliiirnkmahnqyppkfiydlanrfvpkkeeeyfatyfnrvfetitkelwe nkekkdktqv
Psychroflexus torquis (SEQ ID No. 118)	WP_015024765	mesiiglglsfnpyktadkhyfgsflnlvennlnavfaefkerisykakdenissliek hfidnmsivdyekkisilngylpiidflddelennlntrvknfkknfiilaeaieklrdy ythfyhdpitfednkepllelldevllktildvkkkylktdktkeilkdslreemdllvir ktdelrekkktnpkiqhtdssqiknsifndafqgllyedkgnnkktqvshraktrlnp kdihkqeerdfeiplstsglvflmslflskkeiedfksnikgfkgkvvkdenhnslky mathrvysilafkglkyriktdtfsketlmmqmidelskvpdcvyqnlsetkqkdfi edwneyfkdneentenlensrvvhpvirkryedkfnyfairfldefanfktlkfqvf mgyyihdqrtktigttnittertvkekinvfgklskmdnlkkhffsqlsddentdwef fpnpsynfltqadnspannipiylelknqqiikekdaikaevnqtqnrnpnkpskrd llnkilktyedfhqgdptailslneipallhlflvkpnnktgqqieniirikiekqfkain

		hpsknnkgipkslfadtnvrvnaiklkkdleaeldmlnkkhiafkenqkassnydk llkehqftpknkrpelrkyvfyksekgeeatwlandikrfmpkdfktkwkgcqhse lqrklafydrhtkqdikellsgcefdhslldinayfqkdnfedffskylenrietlegvlk klhdfkneptplkgvfkncfkflkrqnyvtespeiikkrilakptflprgvfderptmk kgknplkdknefaewfveylenkdyqkfynaeeyrmrdadfkknavikkqklkd fytlqmvnyllkevfgkdemnlqlselfqtrqerlklqgiakkqmnketgdssentr nqtyiwnkdvpvsffngkvtidkvklknigkykryerdervktfigyevdekwm mylphnwkdrysvkpinvidlqiqeyeeirshellkeiqnleqyiydhttdknillqd gnpnfkmyvlnglligikqvnipdfivlkqntnfdkidftgiascselekktiiliairn kfahnqlpnkmiydlaneflkieknetyanyylkvlkkmisdla
Riemerella anatipestifer (SEQ ID No. 119)	WP_015345620	mffsfhnaqrvifkhlykafdaslrmvkedykahftvnltrdfahlnrkgknkqdn pdfnryrfekdgfftesgllfftnlfldkrdaywmlkkvsgfkashkqrekmttevfc rsrillpklrlesrydhnqmlldmlselsrcpkllyeklseenkkhfqveadgfldeie eeqnpfkdtlirhqdrfpyfalryldlnesfksirfqvdlgtyhyciydkkigdeqekr hltrtllsfgrlqdfteinrpqewkaltkdldyketsnqpfiskttphyhitdnkigfrlgt skelypsleikdganriakypynsgfvahafisvhellplmfyqhltgksedllketvr hiqriykdfeeerintiedlekanqgrlplgafpkqmlgllqnkqpdlsekakikiekl iaetkllshrlntklksspklgkrrekliktgvladwlvkdfmrfqpvaydaqnqpik sskanstefwfirralalyggeknrlegyfkqtnligntnphpflnkfnwkacrnlvdf yqqyleqrekfleaikhqpwepyqyclllkvpkenrknlvkgweqggislprglfte airetlskdltlskpirkeikkhgrvgfisraitlyfkekyqdkhqsfynlsykleakapl lkkeehyeywqqnkpqsptesqrlelhtsdrwkdyllykrwqhlekklrlyrnqdi mlwlmtleltknhfkelnlnyhqlklenlavnvqeadaklnplnqtlpmvlpvkvy pttafgevqyhetpirtvyireeqtkalkmgnfkalvkdrrlnglfsfikeendtqkhp isqlrlrreleiyqslrvdafketlsleekllnkhaslsslenefrtlleewkkkyaassm vtdkhiafiasvrnafchnqypfyketlhapillftvaqptteekdglgiaeallkvlre yceivksqi
Prevotella pleuritidis (SEQ ID No. 120)	WP_021584635	mendkrleesacytlndkhfwaaflnlarhnvyitvnhinktlelknkknqeiiidnd qdilaikthwakvngdlnktdrlrelmikhfpfleaaiysnnkedkeevkeekqaka qsfkslkdclflfleklqearnyyshykysesskepefeegllekmyntfdasirlvke dyqynkdidpekdfkhlerkedfnylftdkdnkgkitknglllffvslflekkdaiwm qqkfrgfkdnrgnkekmthevfcrsrmllpkirlestqtqdwilldmlnelircpksl yerlqgayrekfkvpfdsidedydaeqepfrntlvrhqdrfpyfalryfdyneifknlr

		fqidlgtyhfsiykkliggkkedrhlthklygferiqeftkqnrpdkwqaiikdldtye tsneryisettphyhlenqkigirfrndnndiwpslktngeknekskynldkpyqae aflsvhellpmmfyylllkmentdndkednevgtkkkgnknnkqekhkieeiien kikdiyalydaftngeinsidelaeqregkdieighlpkqlivilknkskdmaekanr kqkemikdtkkrlatldkqvkgeiedggrnirllksgeiarwlvndmmrfqpvqk dnegkplnnskansteyqmlqrslalynkeekptryfrqvnlikssnphpfledtkw eecynilsfyrnylkakikflnklkpedwkknqyflmlkepktnrktlvqgwkngf nlprgiftepikewfkrhqndseeykkvealdrvglvakviplffkeeyfkedaqke inncvqpfysfpynvgnihkpeeknflhceerrklwdkkkdkfkgykakekskk mtdkekeehrsylefqswnkferelrlvrnqdiltwllctklidklkidelnieelqklrl kdidtdtakkeknnilnrvmpmrlpvtvyeidksfnivkdkplhtvyieetgtkllk qgnfkalvkdrrlnglfsfvktsseaeskskpisklrveyelgayqkaridiikdmlal ektlidndenlptnkfsdmlkswlkgkgeankarlqndvgllvavrnafshnqyp mynsevfkgmkllslssdipekeglgiakqlkdkiketieriieiekeirn
Porphyromonas gingivalis (SEQ ID No. 121)	WP_021663197	mntvpasenkgqsrtveddpqyfglylnlarenlieveshvrikfgkkklneeslkq sllcdhllsvdrwtkvyghsrrylpflhyfdpdsqiekdhdsktgvdpdsaqrlirely slldflrndfshnrldgttfehlevspdissfitgtyslacgraqsrfadffkpddfvlakn rkeqlisvadgkecltvsglafficlfldreqasgmlsrirgfkrtdenwaravhetfcd lcirhphdrlessntkeallldmlnelnrcprilydmlpeeeraqflpaldensmnnls enslneesrllwdgssdwaealtkrirhqdrfpylmlrfieemdllkgirfrvdlgeiel dsyskkvgrngeydrtitdhalafgklsdfqneeevsrmisgeasypvrfslfaprya iydnkigychtsdpvypksktgekralsnprsmgfisvhdlrklllmellcegsfsrm qsdflrkanrildetaegklqfsalfpemrhrfippqnpkskdrrekaettlekykqei kgrkdklnsqllsafdmdqrqlpsrlldewmnirpashsvklrtyvkqlnedcrlrlq kfrkdgdgkaraiplvgematflsqdivrmiiseetkklitsayynemqrslaqyag eenrhqfraivaelrlldpssghpflsatmetahrytedfykcylekkrewlaktfyrp eqdentkrrisvffvpdgearkllpllirrrmkeqndlqdwirnkqahpidlpshlfds kimellkvkdgkkkwneafkdwwstkypdgmqpfyglrrelnihgksvsyips dgkkfadcythlmektvqdkkrelrtagkpvppdlaadikrsfhravnerefmlrlv qeddrlmlmainkmmtdreedilpglknidsildeenqfslavhakvlekegegg dnslslvpatieikskrkdwskyiryrydrrvpglmshfpehkatldevktllgeydr crikifdwafalegaimsdrdlkpylhesssregksgehstlvkmlvekkgcltpde

		sqylilirnkaahnqfpcaaempliyrdvsakvgsiegssakdlpegsslvdslwkk yemiirkilpildpenrffgkllnnmsqpindl
Porphyromonas gingivalis (SEQ ID No. 122)	WP_021665475	mntvpasenkgqsrtveddpqyfglylnlarenlieveshvrikfgkkklneeslkq sllcdhllsvdrwtkvyghsrrylpflhyfdpdsqiekdhdsktgvdpdsaqrlirely slldflrndfshnrldgttfehlevspdissfitgtyslacgraqsrfadffkpddfvlakn rkeqlisvadgkecltvsglafficlfldreqasgmlsrirgfkrtnenwaravhetfcd lcirhphdrlessntkeallldmlnelnrcprilydmlpeeeraqflpaldensmnnls enslneesrllwdgssdwaealtkrirhqdrfpylmlrfieemdllkgirfrvdlgeiel dsyskkvgrngeydrtitdhalafgklsdfqneeevsrmisgeasypvrfslfaprya iydnkigychtsdpvypksktgekralsnpqsmgfisvhdlrklllmellcegsfsr mqsgflrkanrildetaegklqfsalfpemrhrfippqnpkskdrrekaettlekykq eikgrkdklnsqllsafdmnqrqlpsrlldewmnirpashsvklrtyvkqlnedcrlr lrkfrkdgdgkaraiplvgematflsqdivrmiiseetkklitsayynemqrslaqya geenrrqfraivaelhlldpssghpflsatmetahrytedfykcylekkrewlaktfyr peqdentkrrisvffvpdgearkllpllirrrmkeqndlqdwirnkqahpidlpshlf dskimellkvkdgkkkwneafkdwwstkypdgmqpfyglrrelnihgksvsyi psdgkkfadcythlmektvqdkkrelrtagkpvppdlaadikrsfhravnerefmlr lvqeddrlmlmainkmmtdreedilpglknidsildkenqfslavhakvlekegeg gdnslslvpatieikskrkdwskyiryrydrrvpglmshfpehkatldevktllgeyd rcrikifdwafalegaimsdrdlkpylhesssregksgehstlvkmlvekkgcltpde sqylilirnkaahnqfpcaaempliyrdvsakvgsiegssakdlpegsslvdslwkk yemiirkilpildhenrffgkllnnmsqpindl
Porphyromonas gingivalis (SEQ ID No. 123)	WP_021677657	mntvpasenkgqsrtveddpqyfglylnlarenlieveshvrikfgkkklneeslkq sllcdhllsvdrwtkvyghsrrylpflhyfdpdsqiekdhdsktgvdpdsaqrlirely slldflrndfshnrldgttfehlevspdissfitgtyslacgraqsrfadffkpddfvlakn rkeqlisvadgkecltvsglafficlfldreqasgmlsrirgfkrtdenwaravhetfcd lcirhphdrlessntkeallldmlnelnrcprilydmlpeeeraqflpaldensmnnls enslneesrllwdgssdwaealtkrirhqdrfpylmlrfieemdllkgirfrvdlgeiel dsyskkvgrngeydrtitdhalafgklsdfqneeevsrmisgeasypvrfslfaprya iydnkigychtsdpvypksktgekralsnpqsmgfisvhdlrklllmellcegsfsr mqsgflrkanrildetaegklqfsalfpemrhrfippqnpkskdrrekaettlekykq eikgrkdklnsqllsafdmnqrqlpsrlldewmnirpashsvklrtyvkqlnedcrlr lrkfrkdgdgkaraiplvgematflsqdivrmiiseetkklitsayynemqrslaqya

		geenrrqfraivaelhlldpssghpflsatmetahrytedfykcylekkrewlaktfyr peqdentkrrisvffvpdgearkllpllirrrmkeqndlqdwirnkqahpidlpshlf dskimellkvkdgkkkwneafkdwwstkypdgmqpfyglrrelnihgksvsyi psdgkkfadcythlmektvqdkkrelrtagkpvppdlaadikrsfhravnerefmlr lvqeddrlmlmainkmmtdreedilpglknidsildeenqfslavhakvlekegeg gdnslslvpatieikskrkdwskyiryrydrrvpglmshfpehkatldevktllgeyd rcrikifdwafalegaimsdrdlkpylhesssregksgehstlvkmlvekkgcltpde sqylilirnkaahnqfpcaaempliyrdvsakvgsiegssakdlpegsslvdslwkk yemiirkilpildhenrffgkllnnmsqpindl
Porphyromonas gingivalis (SEQ ID No. 124)	WP_021680012	mntvpasenkgqsrtveddpqyfglylnlarenlieveshvrikfgkkklneeslkq sllcdhllsvdrwtkvyghsrrylpflhyfdpdsqiekdhdsktgvdpdsaqrlirely slldflrndfshnrldgttfehlevspdissfitgtyslacgraqsrfadffkpddfvlakn rkeqlisvadgkecltvsglafficlfldreqasgmlsrirgfkrtdenwaravhetfcd lcirhphdrlessntkeallldmlnelnrcprilydmlpeeeraqflpaldensmnnls enslneesrllwdgssdwaealtkrirhqdrfpylmlrfieemdllkgirfrvdlgeiel dsyskkvgrngeydrtitdhalafgklsdfqneeevsrmisgeasypvrfslfaprya iydnkigychtsdpvypksktgekralsnprsmgfisvhdlrklllmellcegsfsrm qsdflrkanrildetaegklqfsalfpemrhrfippqnpkskdrrekaettlekykqei kgrkdklnsqllsafdmdqrqlpsrlldewmnirpashsvklrtyvkqlnedcrlrlq kfrkdgdgkaraiplvgematflsqdivrmiiseetkklitsayynemqrslaqyag eenrhqfraivaelrlldpssghpflsatmetahrytedfykcylekkrewlaktfyrp eqdentkrrisvffvpdgearkllpllirrrmkeqndlqdwirnkqahpidlpshlfds kvmellkvkdgkkkwneafkdwwstkypdgmqpfyglrrelnihgksvsyips dgkkfadcythlmektvrdkkrelrtagkpvppdlaayikrsfhravnerefmlrlv qeddrlmlmainkimtdreedilpglknidsildkenqfslavhakvlekegeggd nslslvpatieikskrkdwskyiryrydrrvpglmshfpehkatldevktllgeydrcr ikifdwafalegaimsdrdlkpylhesssregksgehstlvkmlvekkgcltpdesq ylilirnkaahnqfpcaaeipliyrdvsakvgsiegssakdlpegsslvdslwkkye miirkilpildpenrffgkllnnmsqpindl
Porphyromonas gingivalis	WP_023846767	mntvpasenkgqsrtveddpqyfglylnlarenlieveshvrikfgkkklneeslkq sllcdhllsvdrwtkvyghsrrylpflhyfdpdsqiekdhdsktgvdpdsaqrlirely slldflrndfshnrldgttfehlevspdissfitgtyslacgraqsrfadffkpddfvlakn rkeqlisvadgkecltvsglafficlfldreqasgmlsrirgfkrtdenwaravhetfcd

(SEQ ID No. 125)		lcirhphdrlessntkeallldmlnelnrcprilydmlpeeeraqflpaldensmnnls enslneesrllwdgssdwaealtkrirhqdrfpylmlrfieemdllkgirfrvdlgeiel dsyskkvgrngeydrtitdhalafgklsdfqneeevsrmisgeasypvrfslfaprya iydnkigychtsdpvypksktgekralsnprsmgfisvhdlrkllmellcegsfsrm qsdflrkanrildetaegklqfsalfpemrhrfippqnpkskdrrekaettlekykqei kgrkdklnsqllsafdmnqrqlpsrlldewmnirpashsvklrtyvkqlnedcrlrlr kfrkdgdgkaraiplvgematflsqdivrmiiseetkklitsayynemqrslaqyag eenrrqfraivaelhlldpssghpflsatmetahrytedfykcylekkrewlaktfyrp eqdentkrrisvffvpdgearkllpllirrrmkeqndlqdwirnkqahpidlpshlfds kimellkvkdgkkkwneafkdwwstkypdgmqpfyglrrelnihgksvsyips dgkkfadcythlmektvqdkkrelrtagkpvppdlaadikrsfhravnerefmlrlv qeddrlmlmainkmmtdreedilpglknidsildeenqfslavhakvlekegegg dnslslvpatieikskrkdwskyiryrydrrvpglmshfpehkatldevktllgeydr crikifdwafalegaimsdrdlkpylhesssregksgehstlvkmlvekkgcltpde sqylilirnkaahnqfpcaaempliyrdvsakvgsiegssakdlpegsslvdslwkk yemiirkilpildpenrffgkllnnmsqpindl
Prevotella falsenii (SEQ ID No. 126)	WP_0368 84929	mkndnnstkstdytlgdkhfwaaflnlarhnvyitvnhinkvlelknkkdqeiiidn dqdilaiktlwgkvdtdinkkdrlrelimkhfpfleaatyqqsstnntkqkeeeqaka qsfeslkdclflfleklrearnyyshykhsksleepkleekllenmynifdtnvqlvik dyehnkdinpeedfkhlgraegefnyyftrnkkgnitesgllffvslflekkdaiwaq tkikgfkdnrenkqkmthevfcrsrmllpklrlestqtqdwilldmlnelircpksly krlqgekrekfrvpfdpadedydaeqepfkntlvrhqdrfpyfalryfdyneiftnlrf qidlgtyhfsiykkqigdkkedrhlthklygferiqefakenrpdewkalvkdldtfe esnepyisettphyhlenqkigirnknkkkkktiwpsletkttvnerskynlgksfka eaflsvhellpmmfyylllnkeepnngkinaskvegiiekkirdiyklygafaneei nneeelkeycegkdiairhlpkqmiailkneykdmakkaedkqkkmikdtkkrla aldkqvkgevedggrnikplksgriaswlvndmmrfqpvqrdrdgyplnnskan steyqllqrtlalfgsererlapyfrqmnligkdnphpflkdtkwkehnnilsfyrsyle akknflgslkpedwkknqyflklkepktnretlvqgwkngfnlprgiftepirewfir hqneseeykkvkdfdriglvakviplffkedyqkeiedyvqpfygypfnvgnihns qegtflnkkereelwkgnktkfkdyktkeknkektnkdkfkkktdeekeefrsyld fqswkkferelrlvrnqdivtwllcmelidklkidelnieelqklrlkdidtdtakkek nnilnrimpmelpvtvyetddsnniikdkplhtiyikeaetkllkqgnfkalvkdrrl

		nglfsfvetsseaelkskpiskslveyelgeyqrarveiikdmlrleetligndeklptn kfrqmldkwlehkketddtdlkndvkllevrnafshnqypmrdriafanikpfsls santsneeglgiakklkdktketidriieieeqtatkr
Prevotella pleuritidis (SEQ ID No. 127)	WP_0369 31485	mendkrleestcytlndkhfwaaflnlarhnvyitinhinkllleirqidndekvldika lwqkvdkdinqkarlrelmikhfpfleaaiysnnkedkeevkeekqakaqsfkslk dclflfleklqearnyyshykssesskepefeegllekmyntfgvsirlvkedyqynk didpekdfkhlerkedfnylftdkdnkgkitknglllffvslflekkdaiwmqqklrgf kdnrgnkekmthevfcrsrmllpkirlestqtqdwilldmlnelircpkslyerlqga yrekfkvpfdsidedydaeqepfrntlvrhqdrfpyfalryfdyneifknlrfqidlgt yhfsiykkligdnkedrhlthklygferiqefakqkrpnewqalvkdldiyetsneq yisettphyhlenqkigirfknkkdkiwpsletngkenekskynldksfqaeaflsih ellpmmfydlllkkeepnndeknasivegfikkeikrmyaiydafaneeinskegl eeycknkgfqerhlpkqmiailtnksknmaekakrkqkemikdtkkrlatldkqv kgeiedggrnirllksgeiarwlvndmmrfqsvqkdkegkplnnskansteyqml qrslalynkeqkptpyfiqvnlikssnphpfleetkweecnnilsfyrsyleakknfle slkpedwkknqyflmlkepktnrktlvqgwkngfnlprgiftepikewfkrhqnd seeykkvealdrvglvakviplffkeeyfkedaqkeinncvqpfysfpynvgnihk peeknflhceerrklwdkkkdkfkgykakekskkmtdkekeehrsylefqswnk ferelrlvrnqdivtwllctelidklkidelnieelqklrlkdidtdtakkeknnilnrim pmqlpvtvyeidksfnivkdkplhtiyieetgtkllkqgnfkalvkdrrlnglfsfvkt sseaeskskpisklrveyelgayqkaridiikdmlalektlidndenlptnkfsdmlk swlkgkgeankarlqndvdllvairnafshnqypmynsevfkgmkllslssdipek eglgiakqlkdkiketieriieiekeirn
[Porphyro monas gingivalis (SEQ ID No. 128)	WP_0394 17390	mteqnerpyngtyytledkhfwaaffnlarhnayitlahidrqlayskaditndedilf fkgqwknldndlerkarlrslilkhfsflegaaygkklfesqssgnksskkkeltkke keelqanalsldnlksilfdflqklkdfrnyyshyrhpesselplfdgnmlqrlynvfd vsvqrvkrdhehndkvdphrhfnhlvrkgkkdrygnndnpffkhhfvdregtvte agllffvslflekrdaiwmqkkirgfkggteayqqmtnevfcrsrislpklkleslrtd dwmlldmlnelvrcpkslydrlreedrarfrvpidilsdeddtdgteedpfkntlvrh qdrfpyfalryfdlkkvftslrfhidlgtyhfaiykknigeqpedrhltrnlygfgriqdf aeehrpeewkrlvrdldyfetgdkpyitqttphyhiekgkiglrfvpegqhlwpspe vgatrtgrskyaqdkrltaeaflsvhelmpmmfyyfllrekyseevsaekvqgrikr viedvyavydafargeidtldrldacladkgirrghlprqmiailsqehkdmeekvr

		kklqemiadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqpva kdtsgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpflhetr weshtnilsfyrsylkarkaflqsigrsdreenhrflllkepktdrqtlvagwksefhlp rgifteavrdcliemgydevgsykevgfmakavplyferackdrvqpfydypfnv gnslkpkkgrflskekraeewesgkerfrlaklkkeileakehpyldfkswqkferel rlvknqdiitwmmcrdlmeenkvegldtgtlylkdirtdvheqgslnvlnrvkpmr lpvvvyradsrghvhkeqaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvdtga lameqypisklrveyelakyqtarvcafeqtleleeslltryphlpdknfrkmlesws dplldkwpdlhrkvrlliavrnafshnqypmydeavfssirkydpsspdaieermg lniahrlseevkqakemaeriiqv
Porphyromonas gulae (SEQ ID No. 129)	WP_039418912	mteqserpyngtyytledkhfwaaflnlarhnayitlthidrqlayskaditndqdvls fkalwknldndlerksrlrslilkhfsflegaaygkklfeskssgnkssknkeltkkek eelqanalsldnlksilfdflqklkdfrnyyshyrhsgsselplfdgnmlqrlynvfdv svqrvkrdhehndkvdphrhfnhlvrkgkkdryghndnpsfkhhfvdsegmvte agllffvslflekrdaiwmqkkirgfkggtetyqqmtnevfcrsrislpklkleslrmd dwmlldmlnelvrcpkplydrlreddracfrvpvdilpdeddtdgggedpfkntlv rhqdrfpyfalryfdlkkvftslrfhidlgtyhfaiykkmigeqpedrhltrnlygfgri qdfaeehrpeewkrlvrdldyfetgdkpyisqtsphyhiekgkiglrfmpegqhlw pspevgttrtgrskyaqdkrltaeaflsvhelmpmmfyyfllrekyseevsaekvqg rikrviedvyaiydafardeintlkeldacladkgirrghlpkqmiailsqehknmee kvrkklqemiadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqp vakdasgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpflh dtrweshtnilsfyrsylrarkaflerigrsdrmenrpflllkepktdrqtlvagwksef hlprgifteavrdcliemgydevgsyrevgfmakavplyferacedrvqpfydspf nvgnslkpkkgrflskeeraeewergkerfrdleawshsaarriedafagieyaspg nkkkieqllrdlslweafesklkvradkinlaklkkeileaqehpyhdfkswqkfere lrlvknqdiitwmmcrdlmeenkvegldtgtlylkdirtnvqeqgslnvlnhvkp mrlpvvvyradsrghvhkeeaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvd tgglameqypisklrveyelakyqtarvcafeqtleleeslltryphlpdknfrkmles wsdpllakwpelhgkvrlliavrnafshnqypmydeavfssirkydpsspdaieer mglniahrlseevkqaketveriiqa
Porphyromonas gulae	WP_039419792	mteqserpyngtyytledkhfwaaflnlarhnayitlthidrqlayskaditndqdvls fkalwknldndlerksrlrslilkhfsflegaaygkklfeskssgnkssknkeltkkek

(SEQ ID No. 130)		eelqanalsldnlksilfdflqklkdfrnyyshyrhsgsselplfdgnmlqrlynvfdv svqrvkrdhehndkvdphrhfnhlvrkgkkdryghndnpsfkhhfvdgegmvt eagllffvslflekrdaiwmqkkirgfkggtetyqqmtnevfcrsrislpklkleslrtd dwmlldmlnelvrcpkplydrlrekdrarfrvpvdilpdeddtdgggedpfkntlvr hqdrfpyfalryfdlkkvftslrfhidlgtyhfaiykkvigeqpedrhltrnlygfgriq dfaeehrpeewkrlvrdldyfetgdkpyisqttphyhiekgkiglrfvpegqhlwps pevgttrtgrskyaqdkrltaeaflsvhelmpmmfyyfllrekyseevsaekvqgri krviedvyaiydafardeintrdeldacladkgirrghlpkqmigilsqehknmeek vrkklqemiadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqpv akdtsgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpfldet rweshtnilsfyrsylrarkaflerigrsdrvenrpfllllkepktdrqtlvagwksefhlp rgifteavrdcliemgydevgsykevgfmakavplyferackdrvqpfydspfnv gnslkpkkgrflskekraeewesgkerfrlaklkkeileaqehpyhdfkswqkfere lrlvknqdiitwmmcrdlmeenkvegldtgtlylkdirpnvqeqgslnvlnrvkp mrlpvvvyradsrghvhkeeaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvd tgglameqypisklrveyelakyqtarvcvfeltlrleesllsryphlpdesfremles wsdpllakwpelhgkvrlliavrnafshnqypmydeavfssirkydpsspdaieer mglniahrlseevkqaketveriiqa
Porphyromonas gulae (SEQ ID No. 131)	WP_039426176	mteqserpyngtyytledkhfwaaflnlarhnayitlthidrqlayskaditndqdvls fkalwknfdndlerksrlrslilkhfsflegaaygkklfeskssgnkssknkeltkkek eelqanalsldnlksilfdflqklkdfrnyyshyrhsgsselplfdgnmlqrlynvfdv svqrvkrdhehndkvdphyhfnhlvrkgkkdryghndnpsfkhhfvdsegmvt eagllffvslflekrdaiwmqkkirgfkggtgpyeqmtnevfcrsrislpklkleslrt ddwmlldmlnelvrcpkplydrlrekdracfrvpvdilpdeddtdgggedpfkntl vrhqdrfpyfalryfdlkkvftslrfhidlgtyhfaiykkmigeqpedrhltrnlygfgr iqdfaeehrpeewkrlvrdldyfetgdkpyisqttphyhiekgkiglrfmpegqhlw pspevgttrtgrskyaqdkrltaeaflsvhelmpmmfyyfllrekyseevsaekvqg rikrvikdvyaiydafardeintlkeldacsadkgirrghlpkqmigilsqehknmee kvrkklqemiadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqp vakdtsgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpfld etrweshtnilsfyrsylrarkaflerigrsdrvenrpfllllkepkndrqtlvagwksefh lprgifteavrdcliemgydevgsykevgfmakavplyferackdrvqpfydspfn vgnslkpkkgrflskekraeewesgkerfrlaklkkeileakehpyhdfkswqkfer

		elrlvknqdiitwmmcrdlmeenkvegldtgtlylkdirtdvheqgslnvlnrvkp mrlpvvvyradsrghvhkeqaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvd tgglameqypisklrveyelakyqtarvcafeqtleleeslltryphlpdenfremles wsdpllgkwpdlhgkvrlliavrnafshnqypmydeavfssirkydpsspdaieer mglniahrlseevkqaketveriiqa
Porphyromonas gulae (SEQ ID No. 132)	WP_039431778	mteqserpyngtyytledkhfwaaflnlarhnayitlthidrqlayskaditndqdvls fkalwknfdndlerksrlrslilkhfsflegaaygkklfeskssgnkssknkeltkkek eelqanalsldnlksilfdflqklkdfrnyyshyrhsesselplfdgnmlqrlynvfdv svqrvkrdhehndkvdphrhfnhlvrkgkkdryghndnpsfkhhfvdgegmvt eagllffvslflekrdaiwmqkkirgfkggtetyqqmtnevfcrsrislpklkleslrtd dwmlldmlnelvrcpkplydrlreddracfrvpvdilpdeddtdgggedpfkntlv rhqdrfpyfalryfdlkkvftslrfhidlgtyhfaiykkmigeqpedrhltrnlygfgri qdfaeehrpeewkrlvrdldyfetgdkpyisqtsphyhiekgkiglrfmpegqhlw pspevgttrtgrskyaqdkrltaeaflsvhelmpmmfyyfllrekyseevsaekvqg rikrviedvyaiydafardeintlkeldacladkgirrghlpkqmiailsqehkdmee kirkklqemiadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqp vakdtsgkplnnskansteyrmlqralalfggekkrltpyfrqmnltggnnphpflh etrweshtnilsfyrsylrarkaflerigrsdrmenrpflllkepktdrqtlvagwksef hlprgifteavrdcliemgydevgsyrevgfmakavplyferacedrvqpfydspf nvgnslkpkkgrflskeeraeewergkerfrdleawshsaarriedafagieyaspg nkkkieqllrdlslweafesklkvradkinlaklkkeileaqehpyhdfkswqkfere lrlvknqdiitwmmcrdlmeenkvegldtgtlylkdirpnvqeqgslnvlnrvkp mrlpvvvyradsrghvhkeeaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvd tgglameqypisklrveyelakyqtarvcvfeltlrleeslltryphlpdesfrkmles wsdpllakwpelhgkvrlliavrnafshnqypmydeavfssirkydpsspdaieer mglniahrlseevkqaketveriiqv
Porphyromonas gulae (SEQ ID No. 133)	WP_039437199	mteqserpyngtyytledkhfwaaflnlarhnayitlthidrqlayskaditndedilff kgqwknldndlerksrlrslilkhfsflegaaygkkffeskssgnkssknkeltkkek eelqanalsldnlksilfdflqklkdfrnyyshyrhsgsselplfdgnmlqrlynvfdv svqrvkrdhehndevdphyhfnhlvrkgkkdryghndnpsfkhhfvdgegmvt eagllffvslflekrdaiwmqkkirgfkggtepyeqmtnevfcrsrislpklkleslrtd dwmlldmlnelvrcpkplydrlrekdracfrvpvdilpdeddtdgggedpfkntlv rhqdrfpyfalryfdlkkvftslrfhidlgtyhfaiykkmigeqpedrhltrnlygfgri

		qdfaeehrpeewkrlvrdldyfetgdkpyisqttphyhiekgkiglrfvpegqhlwp spevgttrtgrskyaqdkrltaeaflsvhelmpmmfyyfllrekyseevsaekvqgr ikrviedvyaiydafardeintlkeldacladkgirrghlpkqmigilsqerkdmeek vrkklqemiadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqpv akdtsgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpflhet rweshtnilsfyrsylrarkaflerigrsdrvencpflllkepktdrqtlvagwkgefhl prgifteavrdcliemgydevgsyrevgfmakavplyferacedrvqpfydspfnv gnslkpkkgrflskekraeewesgkerfrlaklkkeileaqehpyhdfkswqkfere lrlvknqdiitwmmcrdlmeenkvegldtgtlylkdirpnvqeqgslnvlnrvkp mrlpvvvyradsrghvhkeeaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvd tgalameqypisklrveyelakyqtarvcafeqtleleeslltryphlpdesfremles wsdplltkwpelhgkvrlliavrnafshnqypmydeavfssiwkydpsspdaieer mglniahrlseevkqaketieriiqa
Porphyromonas gulae (SEQ ID No. 134)	WP_039442171	mteqserpyngtyytledkhfwaaflnlarhnayitlthidrqlayskaditndqdvls fkalwknldndlerksrlrslilkhfsflegaaygkklfeskssgnkssknkeltkkek eelqanalsldnlksilfdflqklkdfrnyyshyrhsgsselplfdgnmlqrlynvfdv svqrvkrdhehndkvdphyhfnhlvrkgkkdryghndnpsfkhhfvdsegmvt eagllffvslflekrdaiwmqkkirgfkggtgpyeqmtnevfcrsrislpklkleslrt ddwmlldmlnelvrcpkplydrlrekdracfrvpvdilpdeddtdgggedpfkntl vrhqdrfpyfalryfdlkkvftslrfhidlgtyhfaiykkmigeqpedrhltrnlygfgr iqdfaeehrpeewkrlvrdldyletgdkpyisqttphyhiekgkiglrfvpegqhlw pspevgttrtgrskcaqdkrltaeaflsvhelmpmmfyyfllrekyseevsaekvqg rikrviedvyaiydafardeintlkeldtcladkgirrghlpkqmitilsqerkdmkek irkklqemiadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqpv akdasgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpflhe trweshtnilsfyrsylrarkaflerigrsdrvencpflllkepktdrqtlvagwkdefhl prgifteavrdcliemgydevgsyrevgfmakavplyferacedrvqpfydspfnv gnslkpkkgrflskedraeewergmerfrdleawshsaarrikdafagieyaspgn kkkieqllrdlslweafesklkvradkinlaklkkeileaqehpyhdfkswqkferel rlvknqdiitwmmcrdlmeenkvegldtgtlylkdirpnvqeqgslnvlnrvkpm rlpvvvyradsrghvhkeaplatvyieerntkllkqgnfksfvkdrrlnglfsfvdtgg lameqypisklrveyelakyqtarvcvfeltlrleesllsryphlpdesfremleswsd

		pllakwpelhgkvrlliavrnafshnqypmydeavfssirkydpsspdaieermgl niahrlseevkqaketveriiqa
Porphyromonas gulae (SEQ ID No. 135)	WP_039445055	mntvpatenkgqsrtveddpqyfglylnlarenlieveshvrikfgkkklneeslkqs llcdhllsidrwtkvyghsrrylpflhcfdpdsgiekdhdsktgvdpdsaqrlirelysl ldflrndfshnrldgttfehlkvspdissfitgaytfaceraqsrfadffkpddfllaknrk eqlisvadgkecltvsgfafficlfldreqasgmlsrirgfkrtdenwaravhetfcdlci rhphdrlessntkeallldmlnelnrcprilydmlpeeeraqflpaldensmnnlsen slneesrllwdgssdwaealtkrirhqdrfpylmlrfieemdllkgirfrvdlgeields yskkvgrngeydrtitdhalafgklsdfqneeevsrmisgeasypvrfslfapryaiy dnkigychtsdpvypksktgekralsnpqsmgfisvhdlrklllmellcegsfsrmq sdflrkanrildetaegklqfsalfpemrhrfippqnpkskdrrekaettlekykqeik grkdklnsqllsafdmnqrqlpsrlldewmnirpashsvklrtyvkqlnedcrlrlrk frkdgdgkaraiplvgematflsqdivrmiiseetkklitsayynemqrslaqyagee nrrqfraivaelhlldpssghpflsatmetahrytedfykcylekkrewlaktfyrpeq dentkrrisvffvpdgearkllpllirrrmkeqndlqdwirnkqahpidlpshlfdski mellkvkdgkkkwneafkdwwstkypdgmqpfyglrrelnihgksvsyipsdg kkfadcythlmektvrdkkrelrtagkpvppdlaayikrsfhravnerefmlrlvqe ddrlmlmainkmmtdreedilpglknidsildeenqfslavhakvlekegeggdns lslvpatieikskrkdwskyiryrydrrvpglmshfpehkatldevktllgeydrcrik ifdwafalegaimsdrdlkpylhesssregksgehstlvkmlvekkgcltpdesqyl ilirnkaahnqfpcaaempliyrdvsakvgsiegssakdlpegsslvdslwkkyem iirkilpildhenrffgkllnnmsqpindl
Capnocytophaga cynodegmi (SEQ ID No. 136)	WP_041989581	menktslgnniyynpfkpqdksyfagylnaamenidsvfrelgkrlkgkeytsenf fdaifkenislveyeryvkllsdyfpmarlldkkevpikerkenfkknfrgiikavrdl rnfythkehgeveitdeifgvldemlkstvltvkkkkiktdktkeilkksiekqldilc qkkleylkdtarkieekrrnqrergekklvprfeysdrrddliaaiyndafdvyidkk kdslkessktkyntesypqqeegdlkipiskngvvfllslflskqevhafkskiagfka tvideatvshrknsicfmatheifshlaykklkrkvrtaeinyseaenaeqlsiyaketl mmqmldelskvpdvvyqnlsedvqktfiedwneylkenngdvgtmeeeqvih pvirkryedkfnyfairfldefaqfptlrfqvhlgnylhdsrpkehlisdrrikekitvfg rlselehkkalfikntetnedrkhywevfpnpnydfpkenisvndkdfpiagsildre kqptagkigikvnllnqkyisevdkavkahqlkqrnnkpsiqniieeivpingsnpk eiivfggqptaylsmndihsilyeffdkwekkkeklekkgekelrkeigkeleekiv

		gkiqtqiqqiidkdinakilkpyqdddstaidkeklikdlkqeqkilqklkneqtarek eyqeciayqeesrkikrsdksrqkylrnqlkrkypevptrkeilyyqekgkvavwla ndikrfmptdfknewkgeqhsllqkslayyeqckeelknllpqqkvfkhlpfelgg hfqqkylyqfytryldkrlehisglvqqaenfknenkvfkkvenecfkflkkqnyth kgldaqaqsvlgypiflergfmdekptiikgktfkgneslftdwfryykeyqnfqtfy dtenyplvelekkqadrkretkiyqqkkndvftllmakhifksvfkqdsidrfsledl yqsreerlenqekakqtgerntnyiwnktvdlnlcdgkvtvenvklknvgnfikye ydqrvqtflkyeenikwqaflikeskeeenypyivereieqyekvrreellkevhlie eyilekvkdkeilkkgdnqnfkyyilngllkqlknedvesykvfnlntkpedvninq lkqeatdleqkafvltyirnkfahnqlpkkefwdycqekygkiekektyaeyfaevf krekealmk
Prevotella sp. P5-119 (SEQ ID No. 137)	WP_0425 18169	mnipalvenqkkyfgtysvmamlnaqtvldhiqkvadiegeqnennenlwfhp vmshlynakngydkqpektmfiierlqsyfpflkimaenqreysngkykqnrvev nsndifevlkrafgvlkmyrdltnhyktyeeklidgcefltsteqplsgmiskyytval rntkerygyktedlafiqdnikkitkdaygkrksqvntgfflslqdyngdtqkklhlsg vgialliclfldkqyiniflsrlpifssynaqseerriiirsfginsiklpkdrihseksnks vamdmlnevkrcpdelfttlsaekqsrfriisddhnevlmkrstdrfvplllqyidyg klfdhirfhvnmgklryllkadktcidgqtrvrvieqplngfgrleeaetmrkqengtf gnsgirirdfenvkrddanpanypyivdtythyilennkvemfisdkgssapllplie ddryvvktipscrmstleipamafhmflfgskkteklivdvhnrykrlfqamqkee vtaeniasfgiaesdlpqkildlisgnahgkdvdafirltvddmltdterrikrfkddrk sirsadnkmgkrgfkqistgkladflakdivlfqpsvndgenkitglnyrimqsaiav ydsgddyeakqqfklmfekarligkgttephpflykvfarsipanavdfyerylierk fyltglcneikrgnrvdvpfirrdqnkwktpamktlgriysedlpvelprqmfdnei kshlkslpqmegidfnnanvtyliaeymkrvlnddfqtfyqwkrnyhymdmlkg eydrkgslqhcftsveereglwkerasrteryrklasnkirsnrqmrnasseeietild krlsncrneyqksekvirryrvqdallfllakktlteladfdgerfklkeimpdaekgil seimpmsftfekggkkytitsegmklknygdffvlasdkrignllelvgsdivsked imeefnkydqcrpeissivfnlekwafdtypelsarvdreekvdfksilkillnnkni nkeqsdilrkirnafdhnnypdkgiveikalpeiamsikkafgeyaimk
Prevotella sp. P4-76	WP_0440 72147	mnipalvenqkkyfgtysvmamlnaqtvldhiqkvadiegeqnennenlwfhp vmshlynakngydkqpektmfiierlqsyfpflkimaenqreysngkykqnrvev nsndifevlkrafgvlkmyrdqashyktydeklidgcefltsteqplsgminnyytv

(SEQ ID No. 138)		alrnmnerygyktedlafiqdkrfkfvkdaygkkksqvntgfflslqdyngdtqkkl hlsgvgialliclfldkqyiniflsrlpifssynaqseerriiirsfginsikqpkdrihsek snksvamdmlneikrcpnelfetlsaekqsrfriisndhnevlmkrssdrfvplllqy idygklfdhirfhvnmgklryllkadktcidgqtrvrvieqplngfgrleevetmrkq engtfgnsgirirdfenmkrddanpanypyivdtythyilennkvemfisdeetpa pllpvieddryvvktipscrmstleipamafhmflfgskkteklivdvhnrykrlfka mqkeevtaeniasfgiaesdlpqkiidlisgnahgkdvdafirltvddmladterrikr fkddrksirsadnkmgkrgfkqistgkladflakdivlfqpsvndgenkitglnyri mqsaiavynsgddyeakqqfklmfekarligkgttephpflykvfvrsipanavdf yerylierkfyliglsneikkgnrvdvpfirrdqnkwktpamktlgriydedlpvelp rqmfdneikshlkslpqmegidfnnanvtyliaeymkrvlnddfqtfyqwkrnyr ymdmlrgeydrkgslqscftsveereglwkerasrteryrklasnkirsnrqmrnas seeietildkrlsnsrneyqksekvirryrvqdallfllakktlteladfdgerfklkeim pdaekgilseimpmsftfekggkkytitsegmklknygdffvlasdkrignllelvg sdtvskedimeefkkydqcrpeissivfnlekwafdtypelsarvdreekvdfksilk illnnkninkeqsdilrkirnafdhnnypdkgvveiralpeiamsikkafgeyaimk
Prevotella sp. P5-60 (SEQ ID No. 139)	WP_0440 74780	mnipalvenqkkyfgtysvmamlnaqtvldhiqkvadiegeqnennenlwfhp vmshlynakngydkqpektmfiierlqsyfpflkimaenqreysngkykqnrvev nsndifevlkrafgvlkmyrdltnhyktyeeklidgcefltsteqpfsgmiskyytval rntkerygykaedlafiqdnrykftkdaygkrksqvntgsflslqdyngdttkklhls gvgialliclfldkqyinlflsrlpifssynaqseerriiirsfginsikqpkdrihseksnk svamdmlnevkrcpdelfttlsaekqsrfriisddhnevlmkrssdrfvplllqyidy gklfdhirfhvnmgklryllkadktcidgqtrvrvieqplngfgrleevetmrkqeng tfgnsgirirdfenmkrddanpanypyivetythyilennkvemfisdeenptpllp vieddryvvktipscrmstleipamafhmflfgsektekliidvhdrykrlfqamqk eevtaeniasfgiaesdlpqkimdlisgnahgkdvdafirltvddmltdterrikrfkd drksirsadnkmgkrgfkqistgkladflakdivlfqpsvndgenkitglnyrimqs aiavydsgddyeakqqfklmfekarligkgttephpflykvfvrsipanavdfyery lierkfyliglsneikkgnrvdvpfirrdqnkwktpamktlgriysedlpvelprqmf dneikshlkslpqmegidfnnanvtyliaeymkrvlnddfqtfyqwkrnyrymd mlrgeydrkgslqhcftsieereglwkerasrteryrklasnkirsnrqmrnasseeie tildkrlsncrneyqksekiirryrvqdallfllakktlteladfdgerfklkeimpdaek gilseimpmsftfekggkiytitsggmklknygdffvlasdkrignllelvgsntvsk

		edimeefkkydqcrpeissivfnlekwafdtypelparvdrkekvdfwsildvlsnn kdinneqsyilrkirnafdhnnypdkgiveikalpeiamsikkafgeyaimk
Phaeodactylibacter xiamenensis (SEQ ID No. 140)	WP_044218239	mtntpkrrtlhrhpsyfgaflniarhnafmimehlstkydmedkntldeaqlpnakl fgclkkrygkpdvtegvsrdlrryfpflnyplflhlekqqnaeqaatydinpedieftl kgffrllnqmrnnyshyisntdygkfdklpvqdiyeaaifrlldrgkhtkrfdvfesk htrhlesnnseyrprslanspdhentvafvtclflerkyafpflsrldcfrstndaaegd plirkashecytmfccrlpqpklessdilldmvnelgrcpsalynllseedqarfhikr eeitgfeedpdeeleqeivlkrhsdrfpyfalryfddteafqtlrfdvylgrwrtkpvyk kriygqerdrvltqsirtftrlsrllpiyenvkhdavrqneedgklvnpdvtsqfhksw iqiesddraflsdriehfsphynfgdqviglkfinpdryaaiqnvfpklpgeekkdkd aklvnetadaiistheirslflyhylskkpisagderrfiqvdtetfikqyidtiklffedik sgelqpiadppnyqkneplpyvrgdkektqeeraqyrerqkeikerrkelntllqnr yglsiqyipsrlreyllgykkvpyeklalqklraqrkevkkrikdiekmrtprvgeqat wlaedivfltppkmhtperkttkhpqklnndqfrimqsslayfsvnkkaikkffqke tgiglsnretshpflyridvgrcrgildfytgylkykmdwlddaikkvdnrkhgkke akkyekylpssiqhktpleldytrlpvylprglfkkaivkalaahadfqvepeednvi fcldqlldgdtqdfynwqryyrsalteketdnqlvlahpyaeqilgtiktlegkqknn klgnkakqkikdelidlkrakrrlldreqylravqaedralwlmiqerqkqkaeheei afdqldlknitkiltesidarlripdtkvditdklplrrygdlrrvakdrrlvnlasyyhva glseipydlvkkeleeydrrrvaffehvyqfekevydryaaelrnenpkgestyfsh weyvavavkhsadthfnelfkekvmqlrnkfhhnefpyfdwllpevekasaalya drvfdvaegyyqkmrklmrq
Flavobacterium sp. 316 (SEQ ID No. 141)	WP_045968377	mdnnitvektelglgitynhdkvedkhyfggffnlaqnnidlvaqefkkrlliqgkd sinifanyfsdqcsitnlergikilaeyfpvvsyidldeknksksirehlillletinnlrn yythyyhkkiiidgslfplldtillkvvleikkkklkedktkqllkkglekemtilfnlm kaeqkekkikgwnidenikgavlnrafshllyndelsdyrkskyntedetlkdtltes gilfllsfflnkkeqeqlkanikgykgkiasipdeeitlknnslrnmathwtyshltyk glkhriktdheketllvnmvdylskvpheiyqnlseqnkslfledineymrdneen hdsseasrvihpvirkryenkfayfairfldefaefptlrfmvnvgnyihdnrkkdig gtslitnrtikqqinvfgnlteihkkkndyfekeenkektlewelfpnpsyhfqkeni pifidleksketndlakeyakekkkifgssrkkqqntakknretiinlvfdkyktsdrk tvtfeqptallsfnelnsflyaflvenktgkelekiiiekianqyqilkncsstvdktndn ipksikkivntttdsfyfegkkidieklekditieiektnekletikeneesaqnykrner

		ntqkrklyrkyvfftneigieatwitndilrfldnkenwkgyqhselqkfisqydnyk kealgllesewnlesdaffgqnlkrmfqsnstfetfykkyldnrkntletylsaienlkt mtdvrpkvlkkkwtelfrffdkkiyllstietkinelitkpinlsrgifeekptfingknp nkennqhlfanwfiyakkqtilqdfynlpleqpkaitnlkkhkyklersinnlkiedi yikqmvdflyqklfeqsfigslqdlytskekreiekgkakneqtpdesfiwkkqvei nthngriiaktkikdigkfknlltdnkiahlisyddriwdfslnndgditkklysintele syetirrekllkqiqqfeqflleqeteysaerkhpekfekdcnpnfkkyiiegvlnkiip nheieeieilkskedvfkinfsdililnndnikkgyllimirnkfahnqlidknlfnfsl qlysknenenfseylnkvcqniiqefkeklk
Porphyromonas gulae (SEQ ID No. 142)	WP_046201018	mteqserpyngtyytledkhfwaaflnlarhnayitlthidrqlayskaditndqdvls fkalwknfdndlerksrlrslilkhfsflegaaygkklfeskssgnkssknkeltkkek eelqanalsldnlksilfdflqklkdfrnyyshyrhsesselplfdgnmlqrlynvfdv svqrvkrdhehndkvdphrhfnhlvrkgkkdryghndnpsfkhhfvdsegmvte agllffvslflekrdaiwmqkkirgfkggtetyqqmtnevfcrsrislpklkleslrtdd wmlldmlnelvrcpkplydrlrekdrarfrvpvdilpdeddtdgggedpfkntlvrh qdrfpyfalryfdlkkvftslrfhidlgtyhfaiykkmigeqpedrhltrnlygfgriqd faeehrpeewkrlvrdldyfetgdkpyisqttphyhiekgkiglrfmpegqhlwps pevgttrtgrskyaqdkrltaeaflsvhelmpmmfyyfllrekyseevsaekvqgri krviedvyaiydafardeintlkeldacladkgirrghlpkqmiailsqehkdmeeki rkklqemiadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqpva kdtsgkplnnskansteyrmlqralalfggekkrltpyfrqmnltggnnphpflhetr weshtnilsfyrsylrarkaflerigrsdrmenrpflllkepktdrqtlvagwksefhlp rgifteavrdcliemgydevgsyrevgfmakavplyferacedrvqpfydspfnvg nslkpkkgrflskeeraeewergkerfrdleawshsaarriedafagieyaspgnkk kieqllrdlslweafesklkvradkinlaklkkeileaqehpyhdfkswqkferelrlv knqdiitwmmcrdlmeenkvegldtgtlylkdirpnvqeqgslnvlnrvkpmrlp vvvyradsrghvhkeeaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvdtggla meqypisklrveyelakyqtarvcvfeltlrleeslltryphlpdesfrkmleswsdpl lakwpelhgkvrlliavrnafshnqypmydeavfssirkydpsspdaieermglni ahrlseevkqaketveriiqv
WP_047431796 (SEQ	Chryseobacterium	metqtighgiaydhskiqdkhffggflnlaennikavlkafsekfnvgnvdvkqfa dvslkdnlpdndfqkrvsflkmyfpvvdfinipnnrakfrsdlttlfksvdqlrnfyth yyhkpldfdaslfilddifartakevrdqkmkddktrqllskslseelqkgyelqlerl

ID No. 143)	sp. YR477	kelnrlgkkvnihdqlgikngvlnnafnhliykdgesfktkltyssaltsfesaengiei sqsgllfllsmflkrkeiedlknrnkgfkakvvidedgkvnglkfmathwvfsylcf kglksklstefheetlliqiidelskvpdelycafdketrdkfiedineyvkeghqdfsl edakvihpvirkryenkfnyfairfldefvkfpslrfqvhvgnyvhdrriknidgttfe tervvkdrikvfgrlseissykaqylssvsdkhdetgweifpnpsyvfinnnipihis vdtsfkkeiadfkklrraqvpdelkirgaekkrkfeitqmigsksvlnqeepiallsln eipallyeilingkepaeieriikdklnerqdviknynpenwlpasqisrrlrsnkgeri intdkllqlvtkellvteqklkiisdnrealkqkkegkyirkfiftnselgreaiwladdi krfmpadvrkewkgyqhsqlqqslafynsrpkealailesswnlkdekiiwnewil ksftqnkffdafyneylkgrkkyfaflsehivqytsnaknlqkfikqqmpkdlfekr hyiiedlqteknkilskpfifprgifdkkptfikgvkvedspesfanwyqygyqkdh qfqkfydwkrdysdvflehlgkpfinngdrrtlgmeelkeriiikqdlkikkikiqdlf lrliaenlfqkvfkysaklplsdfyltqeermekenmaalqnvreegdkspniikdnf iwskmipykkgqiienavklkdigklnvlslddkvqtllsyddakpwskialenefs igensyevirreklfkeiqqfeseilfrsgwdginhpaqlednrnpkfkmyivngilr ksaglysqgediwfeynadfnnldadvletkselvqlaflvtairnkfahnqlpakef yfyirakygfadepsvalvylnftkyainefkkvmi
Riemerella anatipestifer (SEQ ID No. 144)	WP_049354263	mffsfhnaqrvifkhlykafdaslrmvkedykahftvnltrdfahlnrkgknkqdn pdfnryrfekdgfftesgllfftnlfldkrdaywmlkkvsgfkashkqrekmttevfc rsrillpklrlesrydhnqmlldmlselsrcpkllyeklseenkkhfqveadgfldeie eeqnpfkdtlirhqdrfpyfalryldlnesfksirfqvdlgtyhyciydkkigdeqekr hltrtllsfgrlqdfteinrpqewkaltkdldyketsnqpfiskttphyhitdnkigfrlgt skelypsleikdganriakypynsgfvahafisvhellplmfyqhltgksedllketvr hiqriykdfeeerintiedlekanqgrlplgafpkqmlgllqnkqpdlsekakikiekl iaetkllshrlntklksspklgkrreklktgvladwlvkdfmrfqpvaydaqnqpik sskanstefwfirralalyggeknrlegyfkqtnligntnphpflnkfnwkacrnlvdf yqqyleqrekfleaiknqpwepyqycllllkipkenrknlvkgweqggislprglfte airetlsedlmlskpirkeikkhgrvgfisraitlyfkekyqdkhqsfynlsykleaka pllkreehyeywqqnkpqsptesqrlelhtsdrwkdyllykrwqhlekklrlyrnqd vmlwlmtleltknhfkelnlnyhqlklenlavnvqeadaklnplnqtlpmvlpvkv ypatafgevqyhktpirtvyireehtkalkmgnfkalvkdrrlnglfsfikeendtqk hpisqlrlrreleiyqslrvdafketlsleekllnkhtslsslenefralleewkkeyaass

		mvtdehiafiasvrnafchnqypfykealhapiplftvaqptteekdglgiaeallkvl reyceivksqi
Porphyromonas gingivalis (SEQ ID No. 145)	WP_052912312	mteqnekpyngtyytledkhfwaaffnlarhnayitlahidrqlayskaditndedil ffkgqwknldndlerkarlrslilkhfsflegaaygkklfesqssgnksskkkeltkke keelqanalsldnlksilfdflqklkdfrnyyshyrhpesselplfdgnmlqrlynvfd vsvqrvkrdhehndkvdphrhfnhlvrkgkkdkygnndnpffkhhfvdreekvt eagllffvslflekrdaiwmqkkirgfkggteayqqmtnevfcrsrislpklkleslrtd dwmlldmlnelvrcpkllydrlreedrarfrvpvdilsdeddtdgteedpfkntlvrh qdrfpyfalryfdlkkvftslrfhidlgtyhfaiykknigeqpedrhltrnlygfgriqdf aeehrpeewkrlvrdldyfetgdkpyitqttphyhiekgkiglrfvpegqllwpspe vgatrtgrskyaqdkrftaeaflsvhelmpmmfyyfllrekyseeasaekvqgrikr viedvyavydafardeintrdeldacladkgirrghlprqmiailsqehkdmeekvr kklqemiadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqpva kdtsgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpflhetr weshtnilsfyrsylkarkaflqsigrsdreenhrflllkepktdrqtlvagwksefhlp rgifteavrdcliemgydevgsykevgfmakavplyferackdrvqpfydypfnv gnslkpkkgrflskekraeewesgkerfrdleawshsaarriedafvgieyaswenk kkieqllqdlslwetfesklkvkadkiniaklkkeileakehpyhdfkswqkferelrl vknqdiitwmmcrdlmeenkvegldtgtlylkdirtdvqeqgslnvlnhvkpmrl pvvvyradsrghvhkeeaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvdtgal ameqypisklrveyelakyqtarvcafeqtleleeslltryphlpdesfremleswsd plldkwpdlqrevrlliavrnafshnqypmydetifssirkydpssldaieermglni ahrlseevklakemveriiqa
Porphyromonas gingivalis (SEQ ID No. 146)	WP_058019250	mteqnekpyngtyytlkdkhfwaaffnlarhnayitlthidrqlayskaditndedilf fkgqwknldndlerkarlrslilkhfsflegaaygkklfesqssgnksskkkeltkke keelqanalsldnlksilfdflqklkdfrnyyshyrhpesselpmfdgnmlqrlynvf dvsvqrvkrdhehndkvdphrhfnhlvrkgkkdrcgnndnpffkhhfvdregkv teagllffvslflekrdaiwmqkkirgfkggtetyqqmtnevfcrsrislpklkleslrt ddwmlldmlnelvrcpkslydrlreedracfrvpvdilsdeddtdgaeedpfkntlv rhqdrfpyfalryfdlkkvftslrfhidlgtyhfaiykknigeqpedrhltrnlygfgriq dfaeehrpeewkrlvrdldcfetgdkpyitqttphyhiekgkiglrfvpegqhlwps pevgatrtgrskyaqdkrftaeaflsvhelmpmmfyyfllrekyseevsaervqgri krviedvyavydafardeintrdeldacladkgirrghlprqmiailsqkhkdmeek

		vrkklqemiadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqpv akdtsgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpflhet rweshtnilsfyrsylkarkaflqsigrsdrvenhrflllkepktdrqtlvagwkgefhl prgifteavrdcliemgldevgsykevgfmakavplyferackdrvqpfydypfnv gnslkpkkgrflskekraeewesgkerfrdleawshsaarriedafagienasrenk kkieqllqdlslwetfesklkvkadkiniaklkkeileakehpyldfkswqkferelrl vknqdiitwmmcrdlmeenkvegldtgtlylkdirtdvqeqgslnvlnhvkpmrl pvvvyradsrghvhkeqaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvdtgal ameqypisklrveyelakyqtarvcafeqtleleeslltryphlpdenfrkmleswsd plldkwpdlhrkvrlliavrnafshnqypmydeavfssirkydpsspdaieermgl niahrlseevkqakemaeriiqa
Flavobacterium columnare (SEQ ID No. 147)	WP_060381855	mssknesynkqktfnhykqedkyffggflnnaddnlrqvgkefktrinfnhnnnel asvfkdyfnkeksvakrehalnllsnyfpvleriqkhtnhnfeqtreifellldtikklrd yythhyhkpitinpkvydflddtlldvlitikkkkvkndtsrellkekfrpeltqlknqk reelikkgkklleenlenavfnhclrpfleenktddkqnktvslrkyrkskpneetsitl tqsglvflisfflhrkefqvftsglegfkakvntikeeeislnknnivymithwsysyy nfkglkhriktdqgvstleqnntthsltntntkealltqivdylskvpneiyetlsekqq kefeedineymrenpenedstfssivshkvirkryenkfnyfamrfldeyaelptlrf mvnfgdyikdrqkkilesiqfdseriikkeihlfeklglvteykknvylketsnidlsrf plfpspsyvmannnipfyidsrsnnldeylnqkkkaqsqnrkrnltfekynkeqsk daiiamlqkeigvkdlqqrstigllscnelpsmlyevivkdikgaelenkiaqkireq yqsirdftldspqkdnipttltktistdtsvtfenqpidiprlknalqkeltltqekllnvkq heievdnynrnkntykfknqpkdkvddnklqrkyvfyrneigqeanwlasdlihf mknkslwkgymhnelqsflaffedkkndcialletvfnlkedciltkdlknlflkhg nfidfykeylklkedflntestflengfiglppkilkkelskrlnyifivfqkrqfiikele ekknnlyadainlsrgifdekptmipfkkpnpdefaswfvasyqynnyqsfyeltp dkiendkkkkyknlrainkvkiqdyylklmvdtlyqdlfnqpldkslsdfyvsktdr ekikadakayqkrndsflwnkvihlslqnnritanpklkdigkykralqdekiatllty ddrtwtyalqkpekenendykelhytalnmelqeyekvrskkllkqvqelekqild kfydfsnnathpedleiedkkgkrhpnfklyitkallkneseiinlenidieilikyydy nteklkekiknmdedekakivntkenynkitnvlikkalvliiirnkmahnqyppk fiydlatrfvpkkeeeyfacyfnrvfetittelwenkkkakeiv

Porphyromonas gingivalis (SEQ ID No. 148)	WP_061156470	mteqnerpyngtyytledkhfwaaffnlarhnayitlthidrqlayskaditndedilf fkgqwknldndlerkarlrslilkhfsflegaaygkklfenkssgnksskkkeltkke keelqanalsldnlksilfdflqklkdfrnyyshyrhpesselplfdgnmlqrlynvfd vsvqrvkrdhehndkvdphrhfnhlvrkgkkdrcgnndnpffkhhfvdregkvte agllffvslflekrdaiwmqkkirgfkggteayqqmtnevfcrsrislpklkleslrtd dwmlldmlnelvrcpkslydrlreedrarfrvpvdilsdeddtdgteedpfkntlvrh qdrfpyfalryfdlkkvftslrfhidlgtyhfaiykknigeqpedrhltrnlygfgriqdf aeehrpeewkrlvrdldyfetgdkpyitqttphyhiekgkiglrfvpegqhlwpspe vgatrtgrskyaqdkrltaeaflsvhelmpmmfyyfllrekyseevsaekvqgrikr viedvyavydafargeidtldrldacladkgirrghlprqmiailsqehkdmeekvr kklqemiadtdhrldmldrqtdrkirigrknaglpksgviadwlvrdmmrfqpva kdtsgkplnnskansteyrmlqralalfggekerltpyfrqmnltggnnphpflhetr weshtnilsfyrsylkarkaflqsigrsdreenhrflllkepktdrqtlvagwksefhlp rgifteavrdcliemgydevgsykevgfmakavplyferackdrvqpfydypfnv gnslkpkkgrflskekraeewesgkerfrlaklkkeileakehpyldfkswqkferel rlvknqdiitwmmcrdlmeenkvegldtgtlylkdirtevqeqgslnvlnrvkpmr lpvvvyradsrghvhkeqaplatvyieerdtkllkqgnfksfvkdrrlnglfsfvdtg glameqypisklrveyelakyqtarvcafeqtleleeslltrcphlpdknfrkmlesw sdplldkwpdlqrevwlliavrnafshnqypmydeavfssirkydpsspdaieer mglniahrlseevkqakemaeriiqa
Porphyromonas gingivalis (SEQ ID No. 149)	WP_061156637	mntvpasenkgqsrtveddpqyfglylnlarenlieveshvrikfgkkklneeslkq sllcdhllsvdrwtkvyghsrrylpflhyfdpdsqiekdhdsktgvdpdsaqrlirely slldflrndfshnrldgttfehlevspdissfitgtyslacgraqsrfadffkpddfvlakn rkeqlisvadgkecltvsglafficlfldreqasgmlsrirgfkrtdenwaravhetfcd lcirhphdrlessntkeallldmlnelnrcprilydmlpeeeraqflpaldensmnnls enslneesrllwdgssdwaealtkrirhqdrfpylmlrfieemdllkgirfrvdlgeiel dsyskkvgrngeydrtitdhalafgklsdfqneeevsrmisgeasypvrfslfaprya iydnkigychtsdpvypksktgekralsnpqsmgfisvhdlrklllmellcegsfsr mqsgflrkanrildetaegklqfsalfpemrhrfippqnpkskdrrekaettlekykq eikgrkdklnsqllsafdmnqrqlpsrlldewmnirpashsvklrtyvkqlnedcrlr lrkfrkdgdgkaraiplvgematflsqdivrmiiseetkklitsayynemqrslaqya geenrrqfraivaelhlldpssghpflsatmetahrytedfykcylekkrewlaktfyr peqdentkrrisvffvpdgearkllpllirrrmkeqndlqdwirnkqahpidlpshlf

		dskimellkvkdgkkkwneafkdwwstkypdgmqpfyglrrelnihgksvsyi psdgkkfadcythlmektvqdkkrelrtagkpvppdlaadikrsfhravnerefmlr lvqeddrlmlmainkmmtdreedilpglknidsildkenqfslavhakvlekegeg gdnslslvpatieikskrkdwskyiryrydrrvpglmshfpehkatldevktllgeyd rcrikifdwafalegaimsdrdlkpylhesssregksgehstlvkmlvekkgcltpde sqylilirnkaahnqfpcaaempliyrdvsakvgsiegssakdlpegsslvdslwkk yemiirkilpildpenrffgkllnnmsqpindl
Riemerella anatipestifer (SEQ ID No. 150)	WP_061710138	mffsfhnaqrvifkhlykafdaslrmvkedykahftvnltrdfahlnrkgknkqdn pdfnryrfekdgfftesgllfftnlfldkrdaywmlkkvsgfkashkqsekmttevfc rsrillpklrlesrydhnqmlldmlselsrcpkllyeklsekdkkcfqveadgfldeie eeqnpfkdtlirhqdrfpyfalryldlnesfksirfqvdlgtyhyciydkkigyeqekr hltrtllnfgrlqdfteinrpqewkaltkdldynetsnqpfiskttphyhitdnkigfrlrt skelypslevkdganriakypynsdfvahafisisvhellplmfyqhltgksedllket vrhiqriykdfeeerintiedlekanqgrlplgafpkqmlgllqnkqpdlsekakikie kliaetkllshrlntklksspklgkrrekliktgvladwlvkdfmrfqpvvydaqnqpi ksskanstesrlirralalyggeknrlegyfkqtnligntnphpflnkfnwkacrnlvd fyqqyleqrekfleaikhqpwepyqycllkvpkenrknlvkgweqggislprglft eairetlskdltlskpirkeikkhgrvgfisraitlyfkekyqdkhqsfynlsykleaka pllkkeehyeywqqnkpqsptesqrlelhtsdrwkdyllykrwqhlekklrlyrnq dimlwlmtleltknhfkelnlnyhqlklenlavnvqeadaklnplnqtlpmvlpvk vypttafgevqyhetpirtvyireeqtkalkmgnfkalvkdrhlnglfsfikeendtq khpisqlrlrreleiyqslrvdafketlsleekllnkhaslsslenefrtlleewkkkyaas smvtdkhiafiasvrnafchnqypfyketlhapillftvaqptteekdglgiaeallrvl reyceivksqi
Flavobacterium columnare (SEQ ID No. 151)	WP_063744070	mssknesynkqktfnhykqedkyffggflnnaddnlrqvgkefktrinfnhnnnel asvfkdyfnkeksvakrehalnllsnyfpvleriqkhtnhnfeqtreifellldtikklrd yythhyhkpitinpkiydfldddtlldvlitikkkkvkndtsrellkeklrpeltqlknqk reelikkgkkllleenlenavfnhclrpfleenktddkqnktvslrkyrkskpneetsitl tqsglvflmsfflhrkefqvftsglegfkakvntikeekislnknnivymithwsysy ynfkglkhriktdqgvstleqnntthsltntntkealltqivdylskvpneiyetlsekq qkefeedineymrenpenedstfssivshkvirkryenkfnyfamrfldeyaelptlr fmvnfgdyikdrqkkilesiqfdseriikkeihlfeklglvteykknvylketsnidlsr fplfpspsyvmannnipfyidsrsnnldeylnqkkkaqsqnrkrnltfekynkeqs

		kdaiiamlqkeigvkdlqqrstigllscnelpsmlyevivkdikgaelenkiaqkire qyqsirdftlnspqkdniptt liktistdtsvtfenqpidiprlknaiqkelaltqekllnv kqheievnnynrnkntykfknqpkdkvddnklqrkyvfyrneigqeanwlasdli hfmknkslwkgymhnelqsflaffedkkndcialletvfnlkedciltkdlknlflkh gnfidfykeylklkedflntestflengfiglppkilkkelskrlnyifivfqkrqfiikel eekknnlyadainlsrgifdekptmipfkkpnpdefaswfvasyqynnyqsfyelt pdkiendkkkkyknlrainkvkiqdyylklmvdtlyqdlfnqpldkslsdfyvskt drekikadakayqkrndsflwnkvihlslqnnritanpklkdigkykralqdekiatl ltyddrtwtyalqkpekenendykelhytalnmelqeyekvrskkllkqvqelekqi ldkfydfsnnathpedleiedkkgkrhpnfklyitkallkneseiinlenidieilikyy dynteklkekiknmdedekakivntkenynkitnvlikkalvliiirnkmahnqyp pkfiydlatrfvpkkeeeyfacyfnrvfetittelwenkkkakeiv
Riemerella anatipestifer (SEQ ID No. 152)	WP_064970887	mekplppnvytlkhkffwgaflniarhnafitichineqlglttppnddkiadvvcgt wnnilnndhdllkksqltelilkhfpflaamcyhppkkegkkkgsqkeqqkeken eaqsqaealnpselikvlktlvkqlrtlrnyyshhshkkpdaekdifkhlykafdaslr mvkedykahftvnltqdfahlnrkgknkqdnpdfdryrfekdgfftesgllfftnlfl dkrdaywmlkkvsgfkashkqsekmttevfcrsrillpklrlesrydhnqmlldml selsrypkllyeklseedkkrfqveadgfldeieeeqnpfkdtlirhqdrfpyfalryld lnesfksirfqvdlgtyhyciydkkigdeqekrhltrtllsfgrlqdfteinrpqewkalt kdldyketskqpfiskttphyhitdnkigfrlgtskelypslevkdganriaqypyns dfvahafisvhellplmfyqhltgksedllketvrhiqriykdfeeerintiedlekanq grlplgafpkqmlgllqnkqpdlsekakikiekliaetkllshrlntklksspklgkrre kliktgvladwlvkdfmrfqpvaydaqnqpiesskanstefqliqralalyggeknrl egyfkqtnligntnphpflnkfnwkacrnlvdfyqqyleqrekfleaiknqpwepy qycllllkipkenrknlvkgweqggislprglfteairetlskdltlskpirkeikkhgrv gfisraitlyfrekyqddhqsfydlpykleakasplpkkehyeywqqnkpqsptelq rlelhtsdrwkdyllykrwqhlekklrlyrnqdvmlwlmtleltknhfkelnlnyhq lklenlavnvqeadaklnplnqtlpmvlpvkvypatafgevqyqetpirtvyireeqt kalkmgnfkalvkdrrlnglfsfikeendtqkhpisqlrlrreleiyqslrvdafketlnl eekllkkhtslssvenkfrilleewkkeyaassmvtdehiafiasvrnafchnqypfy eealhapiplftvaqqtteekdglgiaeallrvlreyceivksqi
Sinomicrobium	WP_072319476.1	mestttlglhlkyqhdlfedkhyfgggvnlavqniesifqafaerygiqnplrkngvp ainnifhdnisisnykeylkflkqylpvvgfleksneinifefredfeilinaiyklrhfy

oceani (SEQ ID No. 153)		thyyhspikledrfytclnelfvavaiqvkkhkmksdktrqllnknlhqllqqlieqkr eklkdkkaegekvsldtksienavlndafvhlldkdenirlnyssrlsediitkngitlsi sgllfllslflqrkeaedlrsriegfkgkgnelrfmathwvfsylnvkrikhrlntdfqke tlliqiadelskvpdevyktldhenrskfledineyiregnedaslnestvvhgvirkry enkfhylvlryldefvdfpslrfqvhlgnyihdrrdkvidgtnfitnrvikepikvfgkl shvsklksdymeslsrehkngwdvfpnpsynfvghnipifinlrsasskgkelyrdl mkiksekkkksreegipmerrdgkptkieisnqidrnikdnnfkdiypgeplamls lnelpallfellrrpsitpqdiedrmveklyerfqiirdykpgdglstskiskklrkadns trldgkkllraiqtetrnareklhtleenkalqknrkrrtvyttreqgreaswlaqdlkrf mpiasrkewrgyhhsqlqqilafydqnpkqplelleqfwdlkedtyvwnswihks lsqhngfvpmyegylkgrlgyykklesdiigfleehkvlkryytqqhlnvifrerlyfi ktetkqklellarplvfprgifddkptfvqdkkvvdhpelfadwyvysykddhsfqe fyhykrdyneifetelswdidfkdnkrqlnpseqmdlfrmkwdlkikkikiqdiflk ivaediylkifghkiplslsdfyisrqerltldeqavaqsmrlpgdtsenqikesnlwqt tvpyekeqirepkiklkdigkfkyflqqqkvlnllkydpqhvwtkaeleeelyigkh syevvrremllqkchqlekhileqfrfdgsnhpreleqgnhpnfkmyivngiltkrg eleieaenwwlelgnsknsldkvevelltmktipeqkafllilirnkfahnqlpadny fhyasnlmnlkksdtyslfwftvadtivqefmsl
Reichenbachiella agariperforans (SEQ ID No. 154)	WP_073124441.1	mktnpliassgekpnykkfntesdksfkkifqnkgsiapiaekacknfeikskspvn rdgrlhyfsvghafknidsknvfryeldesqmdmkptqflalqkeffdfqgalngll khirnvnshyvhtfekleiqsinqklitflieafelavihsylneeelsyeaykddpqsg qklvqflcdkfypnkeheveerktilaknkrqalehllfievtsdidwklfekhkvfti sngkylsfhaclfllslflykseanqliskikgfkrnddnqyrskrqiftffskkftsqdv nseeqhlvkfrdviqylnhypsawnkhlelksgypqmtdklmryiveaeiyrsfp dqtdnhrfllfaireffgqscldtwtgntpinfsnqeqkgfsyeintsaeikdietklkal vlkgplnfkekkeqnrlekdlrrekkeqptnrvkeklltriqhnmlyvsygrnqdrf mdfaarflaetdyfgkdakfkmyqfytsdeqrdhlkeqkkelpkkefeklkyhqs klvdyftyaeqqarypdwdtpfvvennaiqikvtlfngakkivsvqrnlmlylleda lysekrenagkglisgyfvhhqkelkdqldileketeisreqkrefkkllpkrllhrysp aqindttewnpmevileeakaqeqryqlllekailhqteedflkrnkgkqfklrfvrk awhlmylkelymnkvaehghhksfhitkeefndfcrwmfafdevpkykeylcd yfsqkgffnnaefkdliesstslndlyektkqrfegwskdltkqsdenkyllanyesm lkddmlyvnishfisyleskgkinrnahghiaykalnnvphlieeyyykdrlapeey

		kshgklynklktvkledallyemamhylslepalvpkvktkvkdilssniafdikda aghhlyhllipfhkidsfvalinhqsqqekdpdktsflakiqpylekvknskdlkavy hyykdtphtlryedlnmihshivsqsvqftkvalkleeyfiakksitlqiarqisyseia dlsnyftdevrntafhfdvpetaysmilqgieseffldreikpqkpkslselstqqvsvc tafletlhnnlfdrkddkkerlskareryfeqin

[0674] In certain example embodiments, the RNA-targeting effector protein is a Cas13c effector protein as disclosed in PCT Application No. US18/39595 filed June 26, 2018 , and PCT Application No. US 2017/047193 filed August 16, 2017. Example wildtype orthologue sequences of Cas13c are provided in Table 4B below. In certain example embodiments, the CRISPR effector protein is a Cas13c protein from Table 4a or 4b.

Table 4a

Fusobacterium necrophorum subsp. funduliforme ATCC 51357 contig00003 (SEQ ID No. 155)		mekfrrqnrnsiikiiisnydtkgikelkvryrkqaqldtfiikteivnndifiksiiekareky rysflfdgeekyhfknkssveivkkdifsqtpdnmirnykitlkiseknprvveaeiedlm nstilkdgrrsarreksmterklieekvaknysllancpmeevdsikiykikrfltyrsnmll yfasinsflcegikgkdneteeiwhlkdndvrkekvrenfknkliqstenynsslknqiee kekllrkefkkgafyrtiikklqqerikelseksltedcekiiklysklrhslmhydyqyfenl fenkknddlmkdlnldlfkslplirkmklnnkvnyledgdtlfvlqktkkaktlyqiydal ceqkngfnkfindffvsdgeentvfkqiinekfqsemeflekrisesekkneklkkklds mkahfrninsedtkeayfwdihssrnyktkynerknlvneytellgsskekkllreeitkin rqllklkqemeeitkknslfrleykmkiafgflfcefdgniskfkdefdasnqekiiqyhkn gekyltsflkeeekekfnlekmqkiiqkteeedwllpetknnlfkfylltylllpyelkgdfl gfvkkhyydiknvdfidenqnniqvsqtvekqedyfyhkirlfekntkkyeivkysivp neklkqyfedlgidikyltveqksevseeknkkvslknngmfnktillfvfkyyqiafklf ndielyslfflreksgkpleifrkeleskmkdgylnfgqllyvvyevlvknkdldkilskki dyrkdksfspeiaylrnflshlnyskfldnfmkintnksdenkevlipsikiqkmiqfiekc nlqnqidfdfnfvndfymrkekmffiqlkqifpdinstekqkmnekeeilrnryhltdkk neqikdeheaqsqlyekilslqkiyssdknnfygrlkeekllflekqgkkklsmeeikdki agdisdllgilkkeitrdikdkltekfryceekllnlsfynhqdkkkeesirvflirdknsdnf kfesilddgsnkifiskngkeitiqccdkvletliiekntlkissngkiisliphysysidvky
Fusobacterium necrophoru		mekfrrqnrssiikiiisnydtkgikelkvryrkqaqldtfiikteivnndifiksiiekareky rysflfdgeekyhfknkssveivkkdifsqtpdnmirnykitlkiseknprvveaeiedlm nstilkdgrrsarreksmterklieekvaenysllancpmeevdsikiykikrfltyrsnmll

m DJ-2 contig0065, whole genome shotgun sequence (SEQ ID No. 156)		yfasinsflcegikgkdneteeiwhlkdndvrkekvkenfknkliqstenynsslknqiee kekllrkeskkgafyrtiikklqqerikelseksltedcekiiklyselrhplmhydyqyfen lfenkenseltknlnldifkslplvrkmklnnkvnyledndtlfvlqktkkaktlyqiydalc eqkngfnkfindffvsdgeentvfkqiinekfqseieflekrisesekkneklkkkldsmk ahfrninsedtkeayfwdihssrnyktkynerknlvneytellgsskekkllreeitkinrql lklkqemeeitkknslfrleykmkmafgflfcefdgnisrfkdefdasnqekiiqyhkng ekyltyflkeeekekfnlkklqetiqktgeenwllpqnknnlfkfylltylllpyelkgdflgf vkkhyydiknvdfmdenqsskiieskeddfyhkirlfekntkkyeivkysivpdkklkq yfkdlgidtkylildqksevsgeknkkvslknngmfnktillfvfkyyqiafklfndielysl fflreksgkpfevflkelkdkmigkqlnfgqllyvvyevlvknkdlseilseridyrkdmc fsaeiadlrnflshlnyskfldnfmkintnksdenkevlipsikiqkmikfieecnlqsqidf dfnfvndfymrkekmffiqlkqifpdinstekqkmnekeeilrnryhltdkkneqikde heaqsqlyekilslqkiyssdknnfygrlkeekllflekqekkklsmeeikdkiagdisdll gilkkeitrdikdkltekfryceekllnlsfynhqdkkkeesirvflirdknsdnfkfesildd gsnkifiskngkeitiqccdkvletliiekntlkissngkiisliphysysidvky
Fusobacterium necrophorum BFTR-1 contig0068 (SEQ ID No. 157)		mkvryrkqaqldtfiikteivnndifiksiiekarekyrysflfdgeekyhfknkssveivk ndifsqtpdnmirnykitlkiseknprvveaeiedlmnstilkdgrrsarreksmterkliee kvaenysllancpieevdsikiykikrfltyrsnmllyfasinsflcegikgkdneteeiwhl kdndvrkekvkenfknkliqstenynsslknqieekeklsskefkkgafyrtiikklqqeri kelseksltedcekiiklyselrhplmhydyqyfenlfenkenseltknlnldifkslplvrk mklnnkvnyledndtlfvlqktkkaktlyqiydalceqkngfnkfindffvsdgeentvf kqiinekfqsemeflekrisesekkneklkkkldsmkahfrninsedtkeayfwdihssr nyktkynerknlvneytkllgsskekkllreeitkinrqllklkqemeeitkknslfrleyk mkiafgflfcefdgniskfkdefdasnqekiiqyhkngekyltsflkeeekekfnlekmq kiiqkteeedwllpetknnlfkfylltylllpyelkgdflgfvkkhyydiknvdfmdenqn niqvsqtvekqedyfyhkirlfekntkkyeivkysivpneklkqyfedlgidikyltgsve sgekwlgenlgidikyltveqksevseeknkkvslknngmfnktillfvfkyyqiafklfn dielyslfflreksekpfevfleelkdkmigkqlnfgqllyvvyevlvknkdldkilskkid yrkdksfspeiaylrnflshlnyskfldnfmkintnksdenkevlipsikiqkmiqfiekcn lqnqidfdfnfvndfymrkekmffiqlkqifpdinstekqkksekeeilrkryhlinkkne qikdeheaqsqlyekilslqkifscdknnfyrrlkeekllflekqgkkkismkeikdkiasd isdllgilkkeitrdikdkltekfryceekllnisfynhqdkkkeegirvflirdknsdnfkfes ilddgsnkifiskngkeitiqccdkvletlmiekntlkissngkiisliphysysidvky

Fusobacterium necrophorum subsp. funduliforme 1_1_36S cont1.14 (SEQ ID No. 158)		mtekksiifknkssveivkkdifsqtpdnmirnykitlkiseknprvveaeiedlmnstilk dgrrsarreksmterklieekvaenysllancpmeevdsikiykikrfltyrsnmllyfasi nsflcegikgkdneteeiwhlkdndvrkekvkenfknkliqstenynsslknqieekekll rkeskkgafyrtiikklqqerikelseksltedcekiiklyselrhplmhydyqyfenlfenk enseltknlnldifkslplvrkmklnnkvnyledndtlfvlqktkkaktlyqiydalceqkn gfnkfindffvsdgeentvfkqiinekfqsemeflekrisesekkneklkkkfdsmkahf hninsedtkeayfwdihsssnyktkynerknlvneytellgsskekkllreeitqinrkllkl kqemeeitkknslfrleykmkiafgflfcefdgniskfkdefdasnqekiiqyhkngekyl tyflkeeekekfnlekmqkiiqkteeedwllpetknnlfkfylltylllpyelkgdflgfvkk hyydiknvdfmdenqnniqvsqtvekqedyfyhkirlfekntkkyeivkysivpneklk qyfedlgidikyltgsvesgekwlgenlgidikyltveqksevseekikkfl
Fusobacterium perfoetens ATCC 29250 T364DRAFT_scaffold00009.9_C (SEQ ID No. 159)		mgkpnrssiikiiisnydnkgikevkvrynkqaqldtflikselkdgkfilysivdkareky rysfeidktninkneiliikkdiysnkedkvirkyilsfevsekndrtivtkikdcletqkkek ferentrrliseterkllseetqktyskiaccspedidsvkiykikrylayrsnmllffslindif vkgvvkdngeevgeiwriidskeidekktydllvenfkkrmsqefinykqsienkieknt nkikeieqklkkekykkeinrlkkqlielnrendllekdkielsdeeirediekilkiysdlrh klmhynyqyfenlfenkkiskeknedvnltelldlnlfrylplvrqlklenktnylekedki tvlgvsdsaikyysyynflceqkngfnnfinsffsndgeenksfkekinlslekeieimek etnekikeinknelqlmkeqkelgtayvldihslndykishnernknvklqndimngnr dknaldkinkklvelkikmdkitkrnsilrlkyklqvaygflmeeykgnikkfkdefdisk ekiksykskgekylevksekkyitkilnsiedihnitwlknqeennlfkfyvltyillpfefr gdflgfvkkhyydiknvefldenndrltpeqlekmkndsffnkirlfeknskkydilkesi ltserigkyfsllntgakyfeyggeenrgifnkniiipifkyyqivlklyndvelamlltlses dekdinkikelvtlkekvspkkidyekkykfsvlldcfnriinlgkkdflaseevkevaktf tnlaylrnkichlnyskfiddlltidtnksttdsegkllindrirklikfirennqkmnisidyn yindyymkkekfifgqrkqaktiidsgkkankrnkaeellkmyrvkkeninliyelskkl neltkselflldkkllkdidftdvkiknksffelkndvkevanikqalqkhsseligiykkev imaikrsivskliydeekvlsiiiydktnkkyedfllleirrerdinkfqflidekkeklgyekii etkekkkvvvkiqnnselvsepriiknkdkkkaktpeeisklgildltnhycfnlkitl
Fusobacterium ulcerans ATCC 49185		menkgnnkkidfdenynilvaqikeyftkeienynnridniidkkellkysekkeesekn kkleelnklksqklkiltdeeikadvikiikifsdlrhslmhyeykyfenlfenkkneelael lnlnlfknltllrqmkienktnylegreefniigknikakevlghynllaeqkngfnnfinsf fvqdgtenlefkklidehfvnakkrlernikkskklekelekmeqhyqrlncayvwdiht

cont2.38 (SEQ ID No. 160)		sttykklynkrkslieeynkqineikdkevitainvellrikkemeeitksnslfrlkykmqi ayafleiefggniakfkdefdcskmeevqkylkkgvkylkyykdkeaqknyefpfeeif enkdthneewlentsennlfkfyiltylllpmefkgdflgvvkkhyydiknvdftdeseke lsqvqldkmigdsffhkirlfekntkryeiikysiltsdeikryfrlleldvpyfeyekgtdei gifnkniiltifkyyqiifrlyndleihglfnissdldkilrdlksygnkninfreflyvikqnn nssteeeyrkiwenleakylrlhlltpekeeiktktkeeleklneisnlrngichlnykeiieei lkteiseknkeatlnekirkvinfikeneldkvelgfnfindffmkkeqfmfgqikqvkeg nsdsittererkeknnkklketyelncdnlsefyetsnnlreransslledsaflkkiglykv knnkvnskvkdeekrienikrkllkdssdimgmykaevvkklkeklilifkhdeekriy vtvydtskavpeniskeilvkrnnskeeyffednnkkyvteyytleitetnelkvipakkle gkefkteknkenklmlnnhycfnvkiiy
Anaerosalibacter sp. ND1 genome assembly Anaerosalibacter massiliensis ND1 (SEQ ID No. 161)		mksgrrekaksnkssivrviisnfddkqvkeikvlytkqggidvikfkstekdekgrmkf nfdcaynrleeeefnsfggkgkqsffvttnedltelhvtkrhkttgeiikdytiqgkytpikq drtkvtvsitdnkdhfdsndlgdkirlsrsltqytnrilldadvmknyreivcsdsekvdeti nidsqeiykinrflsyrsnmiiyyqminnfllhydgeedkggndsinlineiwkyenkkn dekekiiersyksieksinqyilnhntevesgdkekkidiseerikedlkktfilfsrlrhym vhynykfyenlysgknfiiynkdksksrrfselldlnifkelskiklvknravsnyldkktti hvlnkninaiklldiyrdicetkngfnnfinnmmtisgeedkeykemvtkhfnenmnkl siylenfkkhsdfktnnkkketynllkqeldeqkklrlwfnapyvydihsskkykelyve rkkyvdihsklieaginndnkkklneinvklcelntemkemtklnskyrlqyklqlafgfi leefnldidkfvsafdkdnnltiskfmekretylsksldrrdnrfkklikdykfrdtedifcsd rennlvklyilmyillpveirgdflgfvkknyydlkhvdfidkrnndnkdtffhdlrlfekn vkrlevtsyslsdgflgkksrekfgkelekfiyknvsialptnidikefnkslvlpmmkny qiifkllndieisalfliakkegnegsitfkkvidkvrkedmngninfsqvmkmalnekv ncqirnsiahinmkqlyieplniyinnnqnkktiseqmeeiidicitkgltgkelnkniind yymkkeklvfnlklrkrnnlvsidaqqknmkeksilnkydlnykdenlnikeiilkvndl nnkqkllkettegesnyknalskdilllngiirkninfkikemilgiiqqneyryvniniydk irkedhnidlkinnkyieiscyenksnestderinfkikymdlkvknellvpscyediyik kkidleiryienckvvyidiyykkyninlefdgktlfvkfnkdvkknnqkvnlesnyiqni kfivs

Table 4B

Name	sequence
EHO19081 (SEQ ID No. 762)	mtekksiifknkssveivkkdifsqtpdnmirnykitlkiseknprvveaeiedlmnstilkdgrrsarreksmterklieekvaenysllancpmeevdsikiykikrfltyrsnmllyfasinsflcegikgkdneteeiwhlkdndvrkekvkenfknkliqstenynsslknqieekekllrkeskkgafyrtiikklqqerikelseksltedcekiiklyselrhplmhydyqyfenlfenkenseltknlnldifkslplvrkmklnnkvnyledndtlfvlqktkkaktlyqiydalceqkngfnkfindffvsdgeentvfkqiinekfqsemeflekrisesekkneklkkkfdsmkahfhninsedtkeayfwdihsssnyktkynerknlvneytellgsskekkllreeitqinrkllklkqemeeitkknslfrleykmkiafgflfcefdgniskfkdefdasnqekiiqyhkngekyltyflkeeekekfnlekmqkiiqkteeedwllpetknnlfkfyllttylllpyelkgdflgfvkkhyydiknvdfmdenqnniqvsqtvekqedyfyhkirlfekntkkyeivkysivpneklkqyfedlgidikyltgsvesgekwlgenlgidikyltveqksevseekikkfl
WP_094899336 (SEQ ID No. 763)	mekdkkgekidisqemieedlrkililfsrlrhsmvhydyefyqalysgkdfvisdknnlenrmisqlldlnifkelskvklikdkaisnyldknttihvlgqdikairlldiyrdicgskngfnkfintmitisgeedreykekviehfnkkmenlstyleklekqdnakrnnkrvynllkqklieqqklkewfggpyvydihsskrykelyierkklvdrhsklfeegldeknkkeltkindelsklnsemkemtklnskyrlqyklqlafgfileefdlnidtfinnfdkdkdliisnfmkkrdiylnrvldrgdnrlkniikeykfrdtedifcndrdnnlvklyilmyillpveirgdflgfvkknyydmkhvdfidkkdkedkdtffhdlrlfeknirkleitdyslssgflskehkvdiekkindfinrngamklpeditieefnkslilpimknyqinfkllndieisalfkiakdrsitfkqaideiknedikknskkndknnhkdkninftqlmkralhekipykagmyqirnnishidmeqlyidplnsymnsnknnitiseqiekiidvcvtggvtgkelnnniindyymkkeklvfnlklrkqndivsiesqeknkreefvfkkygldykdgeiniieviqkvnslqeelrniketskeklknketlfrdislingtirkninfkikemvldivrmdeirhinihiyykgenytrsniikfkyaidgenkkyylkqheindinlelkdkfvtlicnmdkhpnknkqtinlesnyiqnvkfiip
WP_040490876 (SEQ ID No. 764)	menkgnnkkidfdenynilvaqikeyftkeienynnridniidkkellkysekkeeseknkkleelnklksqklkiltdeeikadvikiikifsdlrhslmhyeykyfenlfenkkneelaellnlnlfknltllrqmkienktnylegreefnigknikakevlghynllaeqkngfnnfinsffvqdgtenlefkklidehfvnakkrlernikkskklekelekmeqhyqrlncayvwdihtsttykklynkrkslieeynkqineikdkevitainvellrikkemeeitksnslfrlkykmqiayafleiefggniakfkdefdcskmeevqkylkkgvkylkyykdkeaqknyefpfeeifenkdthneewlentsennlfkfyiltylllpmefkgdflgvvkkhyydiknvdftdesekelsqvqldkmigdsffhkirlfekntkryeiikysiltsdeikryfrlleldvpyfeyekgtdeigifnkniiltifkyyqiifrlyndleihglfnissdldkilrdlksygnkninfreflyvikqnnnssteeeyrkiwenleakylrlhlltpekeeiktktkeeleklneisnlrngichlnykeiieeilkteiseknkeatlnekirkvinfikeneldkvelgfnfindffmkkeqfmfgqikqvkegnsdsittererkeknnkklketyelncdnlsefyetsnnlreransslledsaflkkiglykvknnkvnskvkdeekrienikrkllkdssdimgmykaevvkklkeklilifkhdeekriyvtvydtskavpeniskeilvkrnnskeeyffednnkkyvteyytleitetnelkvipakklegkefkteknkenklmlnnhycfnvkiiy
WP_047396607 (SEQ ID No. 765)	meeikhkknkssiirvivsnydmtgikeikvlyqkqggvdtfnlktiinlesgnleiisckpkerekyryefnckteintisitkkdkvlkkeirkyslelyfknekkdtvvakvtdllkapdkiegernhlrklsssterkllsktlcknyseisktpieeidsikiykikrflnyrsnfliyfalindflcagvkeddinevwliqdkehtaflenriekitdyifdklskdienkknqfekrikkyktsleelktetleknktfyidsiktkitnlenkitelslynskeslkedlikiisiftnlrhslmhydyksfenlfenieneelknlldlnlfksirmsdefktknrtnyldgtesftivkkhqnlkklytyynnlcdkkngfntfinsffvtdgientdfknliilhfekemeeykksieyykikisneknkskkeklkekidllqselinmrehknllkqiyffdihnsikykelyserknlieqynlqingvkdvtainhintkllslknkmdkitkqnslyrlkyklkiaysflmiefdgdvskfknnfdptnlekrveyldkkeeylnytapknkfnfakleeelqkiqstsemgadylnvspennlfkfyiltyimlpvefkgdflgfvknhyyniknvdfmdeslldenevdsnklnekienlkdssffnkirlfeknikkyeivkysvstqenmkeyfkqlnldipyldykstdeigifnknmilpifkyyqnvfklcndieihallalankkqqnleyaiyccskknslnynellktfnrktyqnlsfirnkiahlnykelfsdlfnneldlntkvrcliefsqnnkfdqidlgmnfindyymkktrfifnqrrlrdlnvpskekiidgkrkqqndsnnellkkyglsrtnikdifnkawy

WP_035935671 (SEQ ID No. 766)	mkvryrkqaqldtfiikteivnndifiksiiekarekyrysflfdgeekyhfknkssveivkndifsqtpdnmirnykitlkiseknprvveaeiedlmnstilkdgrrsarreksmterklieekvaenysllancpieevdsikiykikrfltyrsnmllyfasinsflcegikgkdneteeiwhlkdndvrkekvkenfknkliqstenynsslknqieekeklsskefkkgafyrtiikklqqerikelseksltedcekiiklyselrhplmhydyqyfenlfenkenseltknlnldifkslplvrkmklnnkvnyledndtlfvlqktkkaktlyqiydalceqkngfnkfindffvsdgeentvfkqiinekfqsemeflekrisesekkneklkkkldsmkahfrninsedtkeayfwdihssrnyktkynerknlvneytkllgsskekkllreeitkinrqllklkqemeeitkknslfrleykmkiafgflfcefdgniskfkdefdasnqekiiqyhkngekyltsflkeeekekfnlekmqkiiqkteeedwllpetknnlfkfylltylllpyelkgdflgfvkkhyydiknvdfmdenqnniqvsqtvekqedyfyhkirlfekntkkyeivkysivpneklkqyfedlgidikyltgsvesgekwlgenlgidikyltveqksevseeknkkvslknngmfnktillfvfkyyqiafklfndielyslfflreksekpfevfleelkdkmigkqlnfgqllyvvyevlvknkdldkilskkidyrkdksfspeiaylrnflshlnyskfldnfmkintnksdenkevlipsikiqkmiqfiekcnlqnqidfdfnfvndfymrkekmffiqlkqifpdinstekqkksekeeilrkryhlinkkneqikdeheaqsqlyekilslqkifscdknnfyrrlkeekllflekqgkkkismkeikdkiasdisdllgilkkeitrdikdkltekfryceekllnisfynhqdkkkeegirvflirdknsdnfkfesilddgsnkifiskngkeitiqccdkvletlmiekntlkissngkiisliphysysidvky
WP_035906563 (SEQ ID No. 767)	mekfrrqnrssiikiiisnydtkgikelkvryrkqaqldtfiikteivnndifiksiiekarekyrysflfdgeekyhfknkssveivkkdifsqtpdnmirnykitlkiseknprvveaeiedlmnstilkdgrrsarreksmterklieekvaenysllancpmeevdsikiykikrfltyrsnmllyfasinsflcegikgkdneteeiwhlkdndvrkekvkenfknkliqstenynsslknqieekekllrkeskkgafyrtiikklqqerikelseksltedcekiiklyselrhplmhydyqyfenlfenkenseltknlnldifkslplvrkmklnnkvnyledndtlfvlqktkkaktlyqiydalceqkngfnkffindffvsdgeentvfkqiinekfqseieflekrisesekkneklkkkldsmkahfrninsedtkeayfwdihssrnyktkynerknlvneytellgsskekkllreeitkinrqllklkqemeeitkknslfrleykmkmafgflfcefdgnisrfkdefdasnqekiiqyhkngekyltyflkeeekekfnlkklqetiqktgeenwllpqnknnlfkfylltylllpyelkgdflgfvkkhyydiknvdfmdenqsskiieskeddfyhkirlfekntkkyeivkysivpdkklkqyfkdlgidtkylildqksevsgeknkkvslknngmfnktillfvfkyyqiafklfndielyslfflreksgkpfevflkelkdkmigkqlnfgqllyvvyevlvknkdlseilseridyrkdmcfsaeiadlrnflshnyskfldnfmkintnksdenkevlipsikiqkmikfieecnlqsqidfdfnfvndfymrkekmffiqlkqifpdinstekqkmnekeeilrnryhltdkkneqikdeheaqsqlyekilslqkiyssdknnfygrlkeekllflekqekkklsmeeikdkiagdisdllgilkkeitrdikdkltekfryceekllnlsfynhqdkkkeesirvflirdknsdnfkfesilddgsnkifiskngkeitiqccdkvletliiekntlkissngkiisliphysysidvky
WP_042678931 (SEQ ID No. 768)	mksgrrekaksnkssivrviisnfddkqvkeikvlytkqggidvikfkstekdekgrmkfnfdcaynrleeeefnsfggkgkqsffvttnedltelhvtkrhkttgeiikdytiqgkytpikqdrtkvtvsitdnkdhfdsndlgdkirlsrsltqytnrilldadvmknyreivcsdsekvdetinidsqeiykinrflsyrsnmiiyyqminnfllhydgeedkggndsinlineiwkyenkkndekekiiersyksieksinqyilnhntevesgdkekkidiseerikedlkktfilfsrlrhymvhynykfyenlysgknfiiynkdksksrrfselldlnifkelskiklvknravsnyldkkttihvlnkninaiklldiyrdicetkngfnnfinnmmtisgeedkeykemvtkhfnenmnklsiylenfkkhsdfktnnkkketynllkqeldeqkklrlwfnapyvydihsskkykelyverkkyvdihsklieaginndnkkklneinvklcelntemkemtklnskyrlqyklqlafgfileefnldidkfvsafdkdnnltiskfmekretylsksldrrdnrfkklikdykfrdtedifcsdrennlvklyilmyillpveirgdflgfvkknyydlkhvdfidkrnndnkdtffhdlrlfeknvkrlevtsyslsdgflgkksrekfgkelekfiyknvsialptnidikefnkslvlpmmknyqiifkllndieisalfliakkegnegsitfkkvidkvrkedmngninfsqvmkmalnekvncqirnsiahinmkqlyieplniyinnnqnkktiseqmeeiidicitkgltgkelnkniindyymkkeklvfnlklrkrnnlvsidaqqknmkeksilnkydlnykdenlnikeiilkvndlnnkqkllkettegesnyknalskdilllngiirkninfkikemilgiiqqneyryvniniydkirkedhnidlkinnkyieiscyenksnestderinfkikymdlkvknellvpscyediyikkkidleiryienckvvyidiyykkyninlefdgktlfvkfnkdvkknnqkvnlesnyiqnikfivs

WP_062627846 (SEQ ID No. 769	mekfrrqnrnsiikiiisnydtkgikelkvryrkqaqldtfiikteivnndifiksiiekarekyrysflfdgeekyhfk nkssveivkkdifsqtpdnmirnykitlkiseknprvveaeiedlmnstilkdgrrsarreksvterklieekvae nysllancpmeevdsikiykikrfltyrsnmllyfasinsflcegikgkeneteeiwhlkdndvrkekvkenfknk liqstenynsslknqieekekllrkeskkgafyrtiikklqqerikelseksltedcekiiklysklrhslmhydyqyf enlfenketpelkdkldlhlfkslplirkmklnnkvnyledgdtlfvlqktkkaktlyqiydalceqkngfnkfindf fvsdgeentvfkqiinekfqsemeflgkriseseeknpklkkkfdsmkahfhninsedtkeayfwdihsssny ktkynerknlvneytellgsskekkllreeitqinrkllklkqemeeitkknslfrleykmkmafgflfcefdgnisr fkdefdasnqekiiqyhkngekyltyflkeeekekfnlkklqetiqktgkenwllpqnknnlfkfylltylllpyelk gdflgfvkkhyydiknvdfmdenqsskiieskeddfyhkirlfekntkkyeivkysivpdeklkqyfkdlgidtky lileqksevsgeknkkvslknngmfnktillfvfkyyqiafklfndielyslfflreksgkpfevflkelkdkmigkql nfgqllyviyevlvknkdlseilseridyrkdmcfsaeiadlrnflshlnyskfldnfmkintnksdenkevlipsiki qkmikfieecnlqsqidfdfnfvndfymrkekmffiqlkqifpdinstekqkmnekeeilrnryhltdkkneqi kdeheaqsqlyekilslqkiyssdknnfygrlkeekllflgkqgkkklsmeeikdkiagdisdllgilkkeitrdikdk ltekfryceekllnlsfynhqdkkkeesirvflirdknsdnfkfesilddgsnkifiskngkeitiqccdkvletlmie kntlkissngkiislvphysysidvky
WP_005959231 (SEQ ID No. 770)	mekfrrqnrnsiikiiisnydtkgikelkvryrkqaqldtfiikteivnndifiksiiekarekyrysflfdgeekyhfk nkssveivkkdifsqtpdnmirnykitlkiseknprvveaeiedlmnstilkdgrrsarreksmterklieekvak nysllancpmeevdsikiykikrfltyrsnmllyfasinsflcegikgkdneteeiwhlkdndvrkekvrenfknk liqstenynsslknqieekekllrkefkkgafyrtiikklqqerikelseksltedcekiiklysklrhslmhydyqyf enlfenkknddlmkdlnldlfkslplirkmklnnkvnyledgdtlfvlqktkkaktlyqiydalceqkngfnkfin dffvsdgeentvfkqiinekfqsemeflekrisesekkneklkkkldsmkahfrninsedtkeayfwdihssrn yktkynerknlvneytellgsskekkllreeitkinrqllklkqemeeitkknslfrleykmkiafgflfcefdgnisk fkdefdasnqekiiqyhkngekyltsflkeeekekfnlekmqkiiqkteeedwllpetknnlfkfylltylllpyel kgdflgfvkkhyydiknvdfidenqnniqvsqtvekqedyfyhkirlfekntkkyeivkysivpneklkqyfedl gidikyltveqksevseeknkkvslknngmfnktillfvfkyyqiafklfndielyslfflreksgkpleifrkelesk mkdgylnfgqllyvvyevlvknkdldkilskkidyrkdksfspeiaylrnflshlnyskfldnfmkintnksdenk evlipsikiqkmiqfiekcnlqnqidfdfnfvndfymrkekmffiqlkqifpdinstekqkmnekeeilrnryhlt dkkneqikdeheaqsqlyekilslqkiyssdknnfygrlkeekllflekqgkkklsmeeikdkiagdisdllgilkk eitrdikdkltekfryceekllnlsfynhqdkkkeesirvflirdknsdnfkfesilddgsnkifiskngkeitiqccdk vletliiekntlkissngkiisliphysysidvky
WP_027128616 (SEQ ID No. 771)	mgkpnrssiikiiisnydnkgikevkvrynkqaqldtflikselkdgkfilysivdkarekyrysfeidktninkneil iikkdiysnkedkvirkyilsfevsekndrtivtkikdcletqkkekferentrrliseterkllseetqktyskiaccs pedidsvkiykikrylayrsnmllffslindifvkgvvkdngeevgeiwriidskeidekktydllvenfkkrmsqe finykqsienkiekntnkikeieqklkkekykkeinrlkkqlielnrendllekdkielsdeeirediekilkiysdlr hklmhynyqyfenlfenkkiskeknedvnltelldlnlfrylplvrqlklenktnylekedkitvlgvsdsaikyys yynflceqkngfnnfinsffsndgeenksfkekinlslekeieimeketnekikeinknelqlmkeqkelgtayv ldihslndykishnernknvklqndimngnrdknaldkinkklvelkikmdkitkrnsilrlkyklqvaygflme eykgnikkfkdefdiskekiksykskgekylevksekkyitkilnsiedihnitwlknqeennlfkfyvltyillpfef rgdflgfvkkhyydiknvefldenndrltpeqlekmkndsffnkirlfeknskkydilkesiltserigkyfsllntga kyfeyggeenrgifnkniiipifkyyqivlklyndvelamlltlsesdekdinkikelvtlkekvspkkidyekkykf svlldcfnriinlgkkdflaseevkevaktftnlaylrnkichlnyskfiddllticltnksttdsegkllindrirklikfire nnqkmnisidynyindyymkkekfifgqrkqaktiidsgkkankrnkaeellkmyrvkkeninliyelskklne ltkselflldkkllkdidftdvkiknksffelkndvkevanikqalqkhsseligiykkevimaikrsivskliydeekv lsiiiydktnkkyedflleirrerdinkfqflidekkeklgyekiietkekkkvvvkiqnnselvsepriiknkdkkka ktpeeisklgildltnhycfnlkitl

WP_062624740 (SEQ ID No. 772)	mekfrrqnrnsiikiiisnydtkgikelkvryrkqaqldtfiikteivnndifiksiiekarekyrysflfdgeekyhfk nkssveivkkdifsqtpdnmirnykitlkiseknprvveaeiedlmnstilkdgrrsarreksmterklieekvak nysllancpmeevdsikiykikrfltyrsnmllyfasinsflcegikgkdneteeiwhlkdndvrkekvrenfknk liqstenynsslknqieekekllrkefkkgafyrtiikklqqerikelseksltedcekiiklysklrhslmhydyqyf enlfenkknddlmkdlnldlfkslplirkmklnnkvnyledgdtlfvlqktkkaktlyqiydalceqkngfnkfin dffvsdgeentvfkqiinekfqsemeflekrisesekkneklkkkldsmkahfrninsedtkeayfwdihssrn yktkynerknlvneytellgsskekkllreeitkinrqllklkqemeeitkknslfrleykmkiafgflfcefdgnisk fkdefdasnqekiiqyhkngekyltsflkeeekekfnlekmqkiiqkteeedwllpetknnlfkfylltylllpyel kgdflgfvkkhyydiknvdfidenqnniqvsqtvekqedyfyhkirlfekntkkyeivkysivpneklkqyfedl gidikyltgsvesgekwlgenlgidikyltveqksevseeknkkvslknngmfnktillfvfkyyqiafklfndiely slfflreksgkpleifrkeleskmkdgylnfgqllyvvyevlvknkdldkilskkidyrkdksfspeiaylrnflshlny skfldnfmkintnksdenkevlipsikiqkmiqfiekcnlqnqidfdfnfvndfymrkekmffiqlkqifpdinst ekqkmnekeeilrnryhltdkkneqikdeheaqsqlyekilslqkiyssdknnfygrlkeekllflekqgkkkls meeikdkiagdisdllgilkkeitrdikdkltekfryceekllnlsfynhqdkkkeesirvflirdknsdnfkfesildd gsnkifiskngkeitiqccdkvletliiekntlkissngkiisliphysysidvky
WP_096402050 SEQ ID No. 773	menknkpnrgsivriiisnydmkgikelkvryrkqaqldtfilqttldksnnsilindfrvkarekyrysftydgke kfsvpsnsiivtkidnaapekskeirkykitlgidekcktgsmitaaiedlleddrvregirnprrkaskterklites ichnyaqitqcpveeidavkiykvkrflsyrsnmllffalindflcknlknekgekineiwemenkgnnkkidfd enynilvaqikeyftkeienynnridniidkkellkyseekeeseknkkleelnklesqklkiltdeeikadvikiiki fsdlrhslmhyeykyfenlfenkkneelaellnlnlfknltllrqmkienktnylegdekfnilgkdvraknalghy dllveqkngfnnfinsffvqdgtenlefkkfidenfikaqkeleediknckesvkklekklkenpkksedlekkle kkqkklkelkkelekmkqhykrlncayvwdihsstvykklynerknliekynkqlnglqdknaitginaqllrik kemeeitksnslfrlkykmqiayaflemeyegniakfknefdcsktekiqewlekseeylnycmekeedgkn ykfhfkeiseikdthneewlentsennlfkfyiltylllpmefkgdflgvvkkhyydiknvdftdesekelsqeqi dkmigdsffhkirlfekntkryeiikysiltsdeikkyfellelkvpyleykgideigifnkniilpifkyyqiifrlyndle ihglfnvsfdinkilsdlksygneninfreflyvikqnnnssteeeyqkiwekleskylkeplltpekkeinkktek elkkldgisflrnkishleyekiiegvlktavngenkktsetnadkvflnekikkiinfikeneldkielgfnfindff mkkeqfmfgqikqvkegnsdsitterkrkeennkrlkityglnynnlskiyefsntlreivnsplflkdstllkkvdl skvmlkekpicslqyenntkleddikrillkdssdimgiykaevvkklkeklvlifkydeekkiyvtvydtskavp eniskeilvkrnnskeeyffednkkkyttqyytleitkenelkvipakklegkefktekkeenklmlnnhycfnv kiiy

[0675] In some embodiments, the Cas13 protein is a Cas13d protein. Yan *et al.* Molecular Cell, 70, 327-339 (2018).

[0676] In some embodiments, the components of the AD-functionalized CRISPR-Cas system may be delivered in various form, such as combinations of DNA/RNA or RNA/RNA or protein RNA. For example, the Cas13 protein may be delivered as a DNA-coding polynucleotide or an RNA-coding polynucleotide or as a protein. The guide may be delivered may be delivered as a DNA-coding polynucleotide or an RNA. All possible combinations are envisioned, including mixed forms of delivery.

DELIVERY OF ENGINEERED COMPOSITIONS

[0677] In some aspects, the invention provides methods comprising delivering one or more polynucleotides, such as or one or more vectors as described herein, one or more transcripts thereof, and/or one or proteins transcribed therefrom, to a host cell.

Vectors

[0678] In general, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. It is a replicon, such as a plasmid, phage, or cosmid, into which another DNA segment may be inserted so as to bring about the replication of the inserted segment. Generally, a vector is capable of replication when associated with the proper control elementsVectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g., circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g., retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors." Vectors for and that result in expression in a eukaryotic cell can be referred to herein as "eukaryotic expression vectors." Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[0679] Recombinant expression vectors can comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g., in an in vitro

transcription/translation system or in a host cell when the vector is introduced into the host cell). Advantageous vectors include lentiviruses and adeno-associated viruses, and types of such vectors can also be selected for targeting particular types of cells.

[0680] With regards to recombination and cloning methods, mention is made of U.S. patent application 10/815,730, published September 2, 2004 as US 2004-0171156 A1, the contents of which are herein incorporated by reference in their entirety.

[0681] The term "regulatory element" is intended to include promoters, enhancers, internal ribosomal entry sites (IRES), and other expression control elements (e.g., transcription termination signals, such as polyadenylation signals and poly-U sequences). Such regulatory elements are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory elements include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). A tissue-specific promoter may direct expression primarily in a desired tissue of interest, such as muscle, neuron, bone, skin, blood, specific organs (e.g., liver, pancreas), or particular cell types (e.g., lymphocytes). Regulatory elements may also direct expression in a temporal-dependent manner, such as in a cell-cycle dependent or developmental stage-dependent manner, which may or may not also be tissue or cell-type specific. In some embodiments, a vector comprises one or more pol III promoter (e.g., 1, 2, 3, 4, 5, or more pol III promoters), one or more pol II promoters (e.g., 1, 2, 3, 4, 5, or more pol II promoters), one or more pol I promoters (e.g., 1, 2, 3, 4, 5, or more pol I promoters), or combinations thereof. Examples of pol III promoters include, but are not limited to, U6 and H1 promoters. Examples of pol II promoters include, but are not limited to, the retroviral Rous sarcoma virus (RSV) LTR promoter (optionally with the RSV enhancer), the cytomegalovirus (CMV) promoter (optionally with the CMV enhancer) [see, e.g., Boshart et al, Cell, 41:521-530 (1985)], the SV40 promoter, the dihydrofolate reductase promoter, the β-actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1α promoter. Also encompassed by the term "regulatory element" are enhancer elements, such as WPRE; CMV enhancers; the R-U5' segment in LTR of HTLV-I (Mol. Cell. Biol., Vol. 8(1), p. 466-472, 1988); SV40 enhancer; and the intron sequence between exons 2 and 3 of rabbit β-globin (Proc. Natl. Acad. Sci. USA., Vol. 78(3), p. 1527-31, 1981). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression desired, etc. A vector can be introduced into host cells to thereby produce transcripts, proteins, or peptides, including fusion proteins or peptides,

encoded by nucleic acids as described herein (e.g., clustered regularly interspersed short palindromic repeats (CRISPR) transcripts, proteins, enzymes, mutant forms thereof, fusion proteins thereof, etc.). With regards to regulatory sequences, mention is made of U.S. patent application 10/491,026, the contents of which are incorporated by reference herein in their entirety. With regards to promoters, mention is made of PCT publication WO 2011/028929 and U.S. application 12/511,940, the contents of which are incorporated by reference herein in their entirety.

[0682] Advantageous vectors include lentiviruses and adeno-associated viruses, and types of such vectors can also be selected for targeting particular types of cells.

[0683] In particular embodiments, use is made of bicistronic vectors for the guide RNA and (optionally modified or mutated) the CRISPR-Cas protein fused to adenosine deaminase. Bicistronic expression vectors for guide RNA and (optionally modified or mutated) CRISPR-Cas protein fused to adenosine deaminase are preferred. In general and particularly in this embodiment, (optionally modified or mutated) CRISPR-Cas protein fused to adenosine deaminase is preferably driven by the CBh promoter. The RNA may preferably be driven by a Pol III promoter, such as a U6 promoter. Ideally the two are combined.

[0684] Vectors can be designed for expression of CRISPR transcripts (e.g. nucleic acid transcripts, proteins, or enzymes) in prokaryotic or eukaryotic cells. For example, CRISPR transcripts can be expressed in bacterial cells such as Escherichia coli, insect cells (using baculovirus expression vectors), yeast cells, or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated in vitro, for example using T7 promoter regulatory sequences and T7 polymerase.

[0685] Vectors may be introduced and propagated in a prokaryote or prokaryotic cell. In some embodiments, a prokaryote is used to amplify copies of a vector to be introduced into a eukaryotic cell or as an intermediate vector in the production of a vector to be introduced into a eukaryotic cell (e.g. amplifying a plasmid as part of a viral vector packaging system). In some embodiments, a prokaryote is used to amplify copies of a vector and express one or more nucleic acids, such as to provide a source of one or more proteins for delivery to a host cell or host organism. Expression of proteins in prokaryotes is most often carried out in Escherichia coli with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, such as to the amino terminus of the recombinant protein. Such fusion vectors

may serve one or more purposes, such as: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Example fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. Gene 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion E. coli expression vectors include pTrc (Amrann et al., (1988) Gene 69:301-315) and pET 11d (Studier et al., GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89). In some embodiments, a vector is a yeast expression vector. Examples of vectors for expression in yeast Saccharomyces cerivisae include pYepSec1 (Baldari, et al., 1987. EMBO J. 6: 229-234), pMFa (Kuijan and Herskowitz, 1982. Cell 30: 933-943), pJRY88 (Schultz et al., 1987. Gene 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.). In some embodiments, a vector drives protein expression in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., SF9 cells) include the pAc series (Smith, et al., 1983. Mol. Cell. Biol. 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. Virology 170: 31-39).

[0686] In some embodiments, a vector is capable of driving expression of one or more sequences in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. Nature 329: 840) and pMT2PC (Kaufman, et al., 1987. EMBO J. 6: 187-195). When used in mammalian cells, the expression vector's control functions are typically provided by one or more regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, simian virus 40, and others disclosed herein and known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells see, e.g., Chapters 16 and 17 of Sambrook, et al., MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

[0687] In some embodiments, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-

specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, et al., 1987. Genes Dev. 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. Adv. Immunol. 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. EMBO J. 8: 729-733) and immunoglobulins (Baneiji, et al., 1983. Cell 33: 729-740; Queen and Baltimore, 1983. Cell 33: 741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle, 1989. Proc. Natl. Acad. Sci. USA 86: 5473-5477), pancreas-specific promoters (Edlund, et al., 1985. Science 230: 912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, e.g., the murine hox promoters (Kessel and Gruss, 1990. Science 249: 374-379) and the α-fetoprotein promoter (Campes and Tilghman, 1989. Genes Dev. 3: 537-546). With regards to these prokaryotic and eukaryotic vectors, mention is made of U.S. Patent 6,750,059, the contents of which are incorporated by reference herein in their entirety. Other embodiments of the invention may relate to the use of viral vectors, with regards to which mention is made of U.S. Patent application 13/092,085, the contents of which are incorporated by reference herein in their entirety. Tissue-specific regulatory elements are known in the art and in this regard, mention is made of U.S. Patent 7,776,321, the contents of which are incorporated by reference herein in their entirety. In some embodiments, a regulatory element is operably linked to one or more elements of a CRISPR system so as to drive expression of the one or more elements of the CRISPR system.

[0688] In some embodiments, one or more vectors driving expression of one or more elements of a nucleic acid-targeting system are introduced into a host cell such that expression of the elements of the nucleic acid-targeting system direct formation of a nucleic acid-targeting complex at one or more target sites. For example, a nucleic acid-targeting effector enzyme and a nucleic acid-targeting guide RNA could each be operably linked to separate regulatory elements on separate vectors. RNA(s) of the nucleic acid-targeting system can be delivered to a transgenic nucleic acid-targeting effector protein animal or mammal, e.g., an animal or mammal that constitutively or inducibly or conditionally expresses nucleic acid-targeting effector protein; or an animal or mammal that is otherwise expressing nucleic acid-targeting effector proteins or has cells containing nucleic acid-targeting effector proteins, such as by way of prior administration thereto of a vector or vectors that code for and express in vivo nucleic acid-targeting effector proteins. Alternatively, two or more of the elements expressed from the same or different regulatory elements, may be combined in a single vector, with one or more

additional vectors providing any components of the nucleic acid-targeting system not included in the first vector. nucleic acid-targeting system elements that are combined in a single vector may be arranged in any suitable orientation, such as one element located 5' with respect to ("upstream" of) or 3' with respect to ("downstream" of) a second element. The coding sequence of one element may be located on the same or opposite strand of the coding sequence of a second element, and oriented in the same or opposite direction. In some embodiments, a single promoter drives expression of a transcript encoding a nucleic acid-targeting effector protein and the nucleic acid-targeting guide RNA, embedded within one or more intron sequences (e.g., each in a different intron, two or more in at least one intron, or all in a single intron). In some embodiments, the nucleic acid-targeting effector protein and the nucleic acid-targeting guide RNA may be operably linked to and expressed from the same promoter. Delivery vehicles, vectors, particles, nanoparticles, formulations and components thereof for expression of one or more elements of a nucleic acid-targeting system are as used in the foregoing documents, such as WO 2014/093622 (PCT/US2013/074667). In some embodiments, a vector comprises one or more insertion sites, such as a restriction endonuclease recognition sequence (also referred to as a "cloning site"). In some embodiments, one or more insertion sites (e.g., about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more insertion sites) are located upstream and/or downstream of one or more sequence elements of one or more vectors. When multiple different guide sequences are used, a single expression construct may be used to target nucleic acid-targeting activity to multiple different, corresponding target sequences within a cell. For example, a single vector may comprise about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, or more guide sequences. In some embodiments, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more such guide-sequence-containing vectors may be provided, and optionally delivered to a cell. In some embodiments, a vector comprises a regulatory element operably linked to an enzyme-coding sequence encoding a a nucleic acid-targeting effector protein. Nucleic acid-targeting effector protein or nucleic acid-targeting guide RNA or RNA(s) can be delivered separately; and advantageously at least one of these is delivered via a particle complex. nucleic acid-targeting effector protein mRNA can be delivered prior to the nucleic acid-targeting guide RNA to give time for nucleic acid-targeting effector protein to be expressed. Nucleic acid-targeting effector protein mRNA might be administered 1-12 hours (preferably around 2-6 hours) prior to the administration of nucleic acid-targeting guide RNA. Alternatively, nucleic acid-targeting effector protein mRNA and nucleic acid-targeting guide RNA can be administered together. Advantageously, a second booster dose of guide RNA can be administered 1-12 hours (preferably around 2-6 hours) after the initial administration of nucleic

acid-targeting effector protein mRNA + guide RNA. Additional administrations of nucleic acid-targeting effector protein mRNA and/or guide RNA might be useful to achieve the most efficient levels of genome modification.

[0689] Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids in mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding components of a nucleic acid-targeting system to cells in culture, or in a host organism. Non-viral vector delivery systems include DNA plasmids, RNA (e.g. a transcript of a vector described herein), naked nucleic acid, and nucleic acid complexed with a delivery vehicle, such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see Anderson, Science 256:808-813 (1992); Nabel & Felgner, TIBTECH 11:211-217 (1993); Mitani & Caskey, TIBTECH 11:162-166 (1993); Dillon, TIBTECH 11:167-175 (1993); Miller, Nature 357:455-460 (1992); Van Brunt, Biotechnology 6(10):1149-1154 (1988); Vigne, Restorative Neurology and Neuroscience 8:35-36 (1995); Kremer & Perricaudet, British Medical Bulletin 51(1):31-44 (1995); Haddada et al., in Current Topics in Microbiology and Immunology, Doerfler and Böhm (eds) (1995); and Yu et al., Gene Therapy 1:13-26 (1994).

[0690] Methods of non-viral delivery of nucleic acids include lipofection, nucleofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in e.g., U.S. Pat. Nos. 5,049,386, 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424; WO 91/16024. Delivery can be to cells (e.g. in vitro or ex vivo administration) or target tissues (e.g. in vivo administration).

[0691] Plasmid delivery involves the cloning of a guide RNA into a CRISPR-Cas protein expressing plasmid and transfecting the DNA in cell culture. Plasmid backbones are available commercially and no specific equipment is required. They have the advantage of being modular, capable of carrying different sizes of CRISPR-Cas coding sequences (including those encoding larger sized proteins) as well as selection markers. Both an advantage of plasmids is that they can ensure transient, but sustained expression. However, delivery of plasmids is not straightforward such that in vivo efficiency is often low. The sustained expression can also be disadvantageous in that it can increase off-target editing. In addition excess build-up of the CRISPR-Cas protein can be toxic to the cells. Finally, plasmids always hold the risk of random

integration of the dsDNA in the host genome, more particularly in view of the double-stranded breaks being generated (on and off-target).

[0692] The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (see, e.g., Crystal, Science 270:404-410 (1995); Blaese et al., Cancer Gene Ther. 2:291-297 (1995); Behr et al., Bioconjugate Chem. 5:382-389 (1994); Remy et al., Bioconjugate Chem. 5:647-654 (1994); Gao et al., Gene Therapy 2:710-722 (1995); Ahmad et al., Cancer Res. 52:4817-4820 (1992); U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787). This is discussed more in detail below.

[0693] The use of RNA or DNA viral based systems for the delivery of nucleic acids takes advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (in vivo) or they can be used to treat cells in vitro, and the modified cells may optionally be administered to patients (ex vivo). Conventional viral based systems could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

[0694] The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of cis-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (see, e.g., Buchscher et al., J. Virol. 66:2731-2739 (1992); Johann et al., J. Virol. 66:1635-1640 (1992); Sommnerfelt et al., Virol. 176:58-59 (1990); Wilson et al., J. Virol. 63:2374-2378 (1989); Miller et al., J. Virol. 65:2220-2224 (1991); PCT/US94/05700).

[0695] In applications where transient expression is preferred, adenoviral based systems may be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors may also be used to transduce cells with target nucleic acids, e.g., in the in vitro production of nucleic acids and peptides, and for in vivo and ex vivo gene therapy procedures (see, e.g., West et al., Virology 160:38-47 (1987); U.S. Pat. No. 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994). Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin et al., Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, et al., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); and Samulski et al., J. Virol. 63:03822-3828 (1989).

[0696] The invention provides AAV that contains or consists essentially of an exogenous nucleic acid molecule encoding a CRISPR system, e.g., a plurality of cassettes comprising or consisting a first cassette comprising or consisting essentially of a promoter, a nucleic acid molecule encoding a CRISPR-associated (Cas) protein (putative nuclease or helicase proteins), e.g., Cas13 and a terminator, andone or more, advantageously up to the packaging size limit of the vector, e.g., in total (including the first cassette) five, cassettes comprising or consisting essentially of a promoter, nucleic acid molecule encoding guide RNA (gRNA) and a terminator (e.g., each cassette schematically represented as Promoter-gRNA1-terminator, Promoter-gRNA2-terminator ... Promoter-gRNA(N)-terminator, where N is a number that can be inserted that is at an upper limit of the packaging size limit of the vector), or two or more individual rAAVs, each containing one or more than one cassette of a CRISPR system, e.g., a first rAAV containing the first cassette comprising or consisting essentially of a promoter, a nucleic acid molecule encoding Cas, e.g., Cas (Cas13) and a terminator, and a second rAAV containing one or more cassettes each comprising or consisting essentially of a promoter, nucleic acid molecule encoding guide RNA (gRNA) and a terminator (e.g., each cassette schematically represented as Promoter-gRNA1-terminator, Promoter-gRNA2-terminator ... Promoter-gRNA(N)-terminator, where N is a number that can be inserted that is at an upper limit of the packaging size limit of the vector). Alternatively, because Cas13 can process its own crRNA/gRNA, a single crRNA/gRNA array can be used for multiplex gene editing. Hence, instead of including multiple cassettes to deliver the gRNAs, the rAAV may contain a single cassette comprising or consisting essentially of a promoter, a plurality of crRNA/gRNA, and a

terminator (e.g., schematically represented as Promoter-gRNA1-gRNA2 ...gRNA(N)-terminator, where N is a number that can be inserted that is at an upper limit of the packaging size limit of the vector). See Zetsche et al Nature Biotechnology 35, 31–34 (2017), which is incorporated herein by reference in its entirety. As rAAV is a DNA virus, the nucleic acid molecules in the herein discussion concerning AAV or rAAV are advantageously DNA. The promoter is in some embodiments advantageously human Synapsin I promoter (hSyn). Additional methods for the delivery of nucleic acids to cells are known to those skilled in the art. See, for example, US20030087817, incorporated herein by reference.

[0697] In another embodiment, Cocal vesiculovirus envelope pseudotyped retroviral vector particles are contemplated (see, e.g., US Patent Publication No. 20120164118 assigned to the Fred Hutchinson Cancer Research Center). Cocal virus is in the Vesiculovirus genus, and is a causative agent of vesicular stomatitis in mammals. Cocal virus was originally isolated from mites in Trinidad (Jonkers et al., Am. J. Vet. Res. 25:236-242 (1964)), and infections have been identified in Trinidad, Brazil, and Argentina from insects, cattle, and horses. Many of the vesiculoviruses that infect mammals have been isolated from naturally infected arthropods, suggesting that they are vector-borne. Antibodies to vesiculoviruses are common among people living in rural areas where the viruses are endemic and laboratory-acquired; infections in humans usually result in influenza-like symptoms. The Cocal virus envelope glycoprotein shares 71.5% identity at the amino acid level with VSV-G Indiana, and phylogenetic comparison of the envelope gene of vesiculoviruses shows that Cocal virus is serologically distinct from, but most closely related to, VSV-G Indiana strains among the vesiculoviruses. Jonkers et al., Am. J. Vet. Res. 25:236-242 (1964) and Travassos da Rosa et al., Am. J. Tropical Med. & Hygiene 33:999-1006 (1984). The Cocal vesiculovirus envelope pseudotyped retroviral vector particles may include for example, lentiviral, alpharetroviral, betaretroviral, gammaretroviral, deltaretroviral, and epsilonretroviral vector particles that may comprise retroviral Gag, Pol, and/or one or more accessory protein(s) and a Cocal vesiculovirus envelope protein. Within certain aspects of these embodiments, the Gag, Pol, and accessory proteins are lentiviral and/or gammaretroviral.

[0698] In some embodiments, a host cell is transiently or non-transiently transfected with one or more vectors described herein. In some embodiments, a cell is transfected as it naturally occurs in a subject optionally to be reintroduced therein. In some embodiments, a cell that is transfected is taken from a subject. In some embodiments, the cell is derived from cells taken from a subject, such as a cell line. A wide variety of cell lines for tissue culture are known in the art. Examples of cell lines include, but are not limited to, C8161, CCRF-CEM, MOLT,

mIMCD-3, NHDF, HeLa-S3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEKn, HEKa, MiaPaCell, Panc1, PC-3, TF1, CTLL-2, C1R, Rat6, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calu1, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J45.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, BS-C-1 monkey kidney epithelial, BALB/ 3T3 mouse embryo fibroblast, 3T3 Swiss, 3T3-L1, 132-d5 human fetal fibroblasts; 10.1 mouse fibroblasts, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, BCP-1 cells, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr -/-, COR-L23, COR-L23/CPR, COR-L23/5010, COR-L23/R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepa1c1c7, HL-60, HMEC, HT-29, Jurkat, JY cells, K562 cells, Ku812, KCL22, KG1, KYO1, LNCap, Ma-Mel 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK II, MOR/0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN / OPCT cell lines, Peer, PNT-1A / PNT 2, RenCa, RIN-5F, RMA/RMAS, Saos-2 cells, Sf-9, SkBr3, T2, T-47D, T84, THP1 cell line, U373, U87, U937, VCaP, Vero cells, WM39, WT-49, X63, YAC-1, YAR, and transgenic varieties thereof. Cell lines are available from a variety of sources known to those with skill in the art (see, e.g., the American Type Culture Collection (ATCC) (Manassus, Va.)).

[0699] In particular embodiments, transient expression and/or presence of one or more of the components of the AD-functionalized CRISPR system can be of interest, such as to reduce off-target effects. In some embodiments, a cell transfected with one or more vectors described herein is used to establish a new cell line comprising one or more vector-derived sequences. In some embodiments, a cell transiently transfected with the components of a AD-functionalized CRISPR system as described herein (such as by transient transfection of one or more vectors, or transfection with RNA), and modified through the activity of a CRISPR complex, is used to establish a new cell line comprising cells containing the modification but lacking any other exogenous sequence. In some embodiments, cells transiently or non-transiently transfected with one or more vectors described herein, or cell lines derived from such cells are used in assessing one or more test compounds.

[0700] In some embodiments it is envisaged to introduce the RNA and/or protein directly to the host cell. For instance, the CRISPR-Cas protein can be delivered as encoding mRNA

together with an in vitro transcribed guide RNA. Such methods can reduce the time to ensure effect of the CRISPR-Cas protein and further prevents long-term expression of the CRISPR system components.

[0701] In some embodiments the RNA molecules of the invention are delivered in liposome or lipofectin formulations and the like and can be prepared by methods well known to those skilled in the art. Such methods are described, for example, in U.S. Pat. Nos. 5,593,972, 5,589,466, and 5,580,859, which are herein incorporated by reference. Delivery systems aimed specifically at the enhanced and improved delivery of siRNA into mammalian cells have been developed, (see, for example, Shen et al FEBS Let. 2003, 539:111-114; Xia et al., Nat. Biotech. 2002, 20:1006-1010; Reich et al., Mol. Vision. 2003, 9: 210-216; Sorensen et al., J. Mol. Biol. 2003, 327: 761-766; Lewis et al., Nat. Gen. 2002, 32: 107-108 and Simeoni et al., NAR 2003, 31, 11: 2717-2724) and may be applied to the present invention. siRNA has recently been successfully used for inhibition of gene expression in primates (see for example. Tolentino et al., Retina 24(4):660 which may also be applied to the present invention.

[0702] Indeed, RNA delivery is a useful method of in vivo delivery. It is possible to deliver Cas13, adenosine deaminase, and guide RNA into cells using liposomes or nanoparticles. Thus delivery of the CRISPR-Cas protein, such as a Cas13, the delivery of the adenosine deaminase (which may be fused to the CRISPR-Cas protein or an adaptor protein), and/or delivery of the RNAs of the invention may be in RNA form and via microvesicles, liposomes or particle or particles. For example, Cas13 mRNA, adenosine deaminase mRNA, and guide RNA can be packaged into liposomal particles for delivery in vivo. Liposomal transfection reagents such as lipofectamine from Life Technologies and other reagents on the market can effectively deliver RNA molecules into the liver.

[0703] Means of delivery of RNA also preferred include delivery of RNA via particles (Cho, S., Goldberg, M., Son, S., Xu, Q., Yang, F., Mei, Y., Bogatyrev, S., Langer, R. and Anderson, D., Lipid-like nanoparticles for small interfering RNA delivery to endothelial cells, Advanced Functional Materials, 19: 3112-3118, 2010) or exosomes (Schroeder, A., Levins, C., Cortez, C., Langer, R., and Anderson, D., Lipid-based nanotherapeutics for siRNA delivery, Journal of Internal Medicine, 267: 9-21, 2010, PMID: 20059641). Indeed, exosomes have been shown to be particularly useful in delivery siRNA, a system with some parallels to the CRISPR system. For instance, El-Andaloussi S, et al. ("Exosome-mediated delivery of siRNA in vitro and in vivo." Nat Protoc. 2012 Dec;7(12):2112-26. doi: 10.1038/nprot.2012.131. Epub 2012 Nov 15.) describe how exosomes are promising tools for drug delivery across different biological barriers and can be harnessed for delivery of siRNA in vitro and in vivo. Their

approach is to generate targeted exosomes through transfection of an expression vector, comprising an exosomal protein fused with a peptide ligand. The exosomes are then purify and characterized from transfected cell supernatant, then RNA is loaded into the exosomes. Delivery or administration according to the invention can be performed with exosomes, in particular but not limited to the brain. Vitamin E (α-tocopherol) may be conjugated with CRISPR Cas and delivered to the brain along with high density lipoprotein (HDL), for example in a similar manner as was done by Uno et al. (HUMAN GENE THERAPY 22:711–719 (June 2011)) for delivering short-interfering RNA (siRNA) to the brain. Mice were infused via Osmotic minipumps (model 1007D; Alzet, Cupertino, CA) filled with phosphate-buffered saline (PBS) or free TocsiBACE or Toc-siBACE/HDL and connected with Brain Infusion Kit 3 (Alzet). A brain-infusion cannula was placed about 0.5mm posterior to the bregma at midline for infusion into the dorsal third ventricle. Uno et al. found that as little as 3 nmol of Toc-siRNA with HDL could induce a target reduction in comparable degree by the same ICV infusion method. A similar dosage of CRISPR Cas conjugated to α-tocopherol and co-administered with HDL targeted to the brain may be contemplated for humans in the present invention, for example, about 3 nmol to about 3 μmol of CRISPR Cas targeted to the brain may be contemplated. Zou et al. ((HUMAN GENE THERAPY 22:465-475 (April 2011)) describes a method of lentiviral-mediated delivery of short-hairpin RNAs targeting PKCγ for in vivo gene silencing in the spinal cord of rats. Zou et al. administered about 10 μl of a recombinant lentivirus having a titer of 1 x 109 transducing units (TU)/ml by an intrathecal catheter. A similar dosage of CRISPR Cas expressed in a lentiviral vector targeted to the brain may be contemplated for humans in the present invention, for example, about 10-50 ml of CRISPR Cas targeted to the brain in a lentivirus having a titer of 1 x 10^9 transducing units (TU)/ml may be contemplated.

Dosage of vectors

[0704] In some embodiments, the vector, e.g., plasmid or viral vector is delivered to the tissue of interest by, for example, an intramuscular injection, while other times the delivery is via intravenous, transdermal, intranasal, oral, mucosal, or other delivery methods. Such delivery may be either via a single dose, or multiple doses. One skilled in the art understands that the actual dosage to be delivered herein may vary greatly depending upon a variety of factors, such as the vector choice, the target cell, organism, or tissue, the general condition of the subject to be treated, the degree of transformation/modification sought, the administration route, the administration mode, the type of transformation/modification sought, etc.

[0705] Such a dosage may further contain, for example, a carrier (water, saline, ethanol, glycerol, lactose, sucrose, calcium phosphate, gelatin, dextran, agar, pectin, peanut oil, sesame oil, etc.), a diluent, a pharmaceutically-acceptable carrier (e.g., phosphate-buffered saline), a pharmaceutically-acceptable excipient, and/or other compounds known in the art. The dosage may further contain one or more pharmaceutically acceptable salts such as, for example, a mineral acid salt such as a hydrochloride, a hydrobromide, a phosphate, a sulfate, etc.; and the salts of organic acids such as acetates, propionates, malonates, benzoates, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, gels or gelling materials, flavorings, colorants, microspheres, polymers, suspension agents, etc. may also be present herein. In addition, one or more other conventional pharmaceutical ingredients, such as preservatives, humectants, suspending agents, surfactants, antioxidants, anticaking agents, fillers, chelating agents, coating agents, chemical stabilizers, etc. may also be present, especially if the dosage form is a reconstitutable form. Suitable exemplary ingredients include microcrystalline cellulose, carboxymethylcellulose sodium, polysorbate 80, phenylethyl alcohol, chlorobutanol, potassium sorbate, sorbic acid, sulfur dioxide, propyl gallate, the parabens, ethyl vanillin, glycerin, phenol, parachlorophenol, gelatin, albumin and a combination thereof. A thorough discussion of pharmaceutically acceptable excipients is available in REMINGTON'S PHARMACEUTICAL SCIENCES (Mack Pub. Co., N.J. 1991) which is incorporated by reference herein.

[0706] In an embodiment herein the delivery is via an adenovirus, which may be at a single booster dose containing at least 1 x 105 particles (also referred to as particle units, pu) of adenoviral vector. In an embodiment herein, the dose preferably is at least about 1 x 106 particles (for example, about 1 x 106-1 x 1012 particles), more preferably at least about 1 x 107 particles, more preferably at least about 1 x 108 particles (e.g., about 1 x 108-1 x 1011 particles or about 1 x 108-1 x 1012 particles), and most preferably at least about 1 x 100 particles (e.g., about 1 x 109-1 x 1010 particles or about 1 x 109-1 x 1012 particles), or even at least about 1 x 1010 particles (e.g., about 1 x 1010-1 x 1012 particles) of the adenoviral vector. Alternatively, the dose comprises no more than about 1 x 1014 particles, preferably no more than about 1 x 1013 particles, even more preferably no more than about 1 x 1012 particles, even more preferably no more than about 1 x 1011 particles, and most preferably no more than about 1 x 1010 particles (e.g., no more than about 1 x 109 articles). Thus, the dose may contain a single dose of adenoviral vector with, for example, about 1 x 106 particle units (pu), about 2 x 106 pu, about 4 x 106 pu, about 1 x 107 pu, about 2 x 107 pu, about 4 x 107 pu, about 1 x 108 pu, about 2 x 108 pu, about 4 x 108 pu, about 1 x 109 pu, about 2 x 109 pu, about 4 x 109

pu, about 1 x 1010 pu, about 2 x 1010 pu, about 4 x 1010 pu, about 1 x 1011 pu, about 2 x 1011 pu, about 4 x 1011 pu, about 1 x 1012 pu, about 2 x 1012 pu, or about 4 x 1012 pu of adenoviral vector. See, for example, the adenoviral vectors in U.S. Patent No. 8,454,972 B2 to Nabel, et. al., granted on June 4, 2013; incorporated by reference herein, and the dosages at col 29, lines 36-58 thereof. In an embodiment herein, the adenovirus is delivered via multiple doses.

[0707] In an embodiment herein, the delivery is via an AAV. A therapeutically effective dosage for in vivo delivery of the AAV to a human is believed to be in the range of from about 20 to about 50 ml of saline solution containing from about 1 x 1010 to about 1 x 1010 functional AAV/ml solution. The dosage may be adjusted to balance the therapeutic benefit against any side effects. In an embodiment herein, the AAV dose is generally in the range of concentrations of from about 1 x 105 to 1 x 1050 genomes AAV, from about 1 x 108 to 1 x 1020 genomes AAV, from about 1 x 1010 to about 1 x 1016 genomes, or about 1 x 1011 to about 1 x 1016 genomes AAV. A human dosage may be about 1 x 1013 genomes AAV. Such concentrations may be delivered in from about 0.001 ml to about 100 ml, about 0.05 to about 50 ml, or about 10 to about 25 ml of a carrier solution. Other effective dosages can be readily established by one of ordinary skill in the art through routine trials establishing dose response curves. See, for example, U.S. Patent No. 8,404,658 B2 to Hajjar, et al., granted on March 26, 2013, at col. 27, lines 45-60.

[0708] In an embodiment herein the delivery is via a plasmid. In such plasmid compositions, the dosage should be a sufficient amount of plasmid to elicit a response. For instance, suitable quantities of plasmid DNA in plasmid compositions can be from about 0.1 to about 2 mg, or from about 1 μg to about 10 μg per 70 kg individual. Plasmids of the invention will generally comprise (i) a promoter; (ii) a sequence encoding a CRISPR-Cas protein, operably linked to said promoter; (iii) a selectable marker; (iv) an origin of replication; and (v) a transcription terminator downstream of and operably linked to (ii). The plasmid can also encode the RNA components of a CRISPR complex, but one or more of these may instead be encoded on a different vector.

[0709] The doses herein are based on an average 70 kg individual. The frequency of administration is within the ambit of the medical or veterinary practitioner (e.g., physician, veterinarian), or scientist skilled in the art. It is also noted that mice used in experiments are typically about 20g and from mice experiments one can scale up to a 70 kg individual.

[0710] The dosage used for the compositions provided herein include dosages for repeated administration or repeat dosing. In particular embodiments, the administration is repeated within a period of several weeks, months, or years. Suitable assays can be performed to obtain

an optimal dosage regime. Repeated administration can allow the use of lower dosage, which can positively affect off-target modifications.

RNA delivery

[0711] In particular embodiments, RNA based delivery is used. In these embodiments, mRNA of the CRISPR-Cas protein, mRNA of the adenosine deaminase (which may be fused to a CRISPR-Cas protein or an adaptor), are delivered together with in vitro transcribed guide RNA. Liang et al. describes efficient genome editing using RNA based delivery (Protein Cell. 2015 May; 6(5): 363–372). In some embodiments, the mRNA(s) encoding Cas13 and/or adenosine deaminase can be chemically modified, which may lead to improved activity compared to plasmid-encoded Cas13 and/or adenosine deaminase. For example, uridines in the mRNA(s) can be partially or fully substituted with pseudouridine (Ψ), N1-methylpseudouridine (me1Ψ), 5-methoxyuridine(5moU). See Li et al., Nature Biomedical Engineering 1, 0066 DOI:10.1038/s41551-017-0066 (2017), which is incorporated herein by reference in its entirety.

RNP Delivery

[0712] In particular embodiments, pre-complexed guide RNA, CRISPR-Cas protein, and adenosine deaminase (which may be fused to a CRISPR-Cas protein or an adaptor) are delived as a ribonucleoprotein (RNP). RNPs have the advantage that they lead to rapid editing effects even more so than the RNA method because this process avoids the need for transcription. An important advantage is that both RNP delivery is transient, reducing off-target effects and toxicity issues. Efficient genome editing in different cell types has been observed by Kim et al. (2014, Genome Res. 24(6):1012-9), Paix et al. (2015, Genetics 204(1):47-54), Chu et al. (2016, BMC Biotechnol. 16:4), and Wang et al. (2013, Cell. 9;153(4):910-8).

[0713] In particular embodiments, the ribonucleoprotein is delivered by way of a polypeptide-based shuttle agent as described in WO2016161516. WO2016161516 describes efficient transduction of polypeptide cargos using synthetic peptides comprising an endosome leakage domain (ELD) operably linked to a cell penetrating domain (CPD), to a histidine-rich domain and a CPD. Similarly these polypeptides can be used for the delivery of CRISPR-effector based RNPs in eukaryotic cells.

Particles

[0714] In some aspects or embodiments, a composition comprising a delivery particle formulation may be used. In some aspects or embodiments, the formulation comprises a CRISPR complex, the complex comprising a CRISPR protein and a guide which directs sequence-specific binding of the CRISPR complex to a target sequence. In some embodiments, the delivery particle comprises a lipid-based particle, optionally a lipid nanoparticle, or cationic

lipid and optionally biodegradable polymer. In some embodiments, the cationic lipid comprises 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP). In some embodiments, the hydrophilic polymer comprises ethylene glycol or polyethylene glycol. In some embodiments, the delivery particle further comprises a lipoprotein, preferably cholesterol. In some embodiments, the delivery particles are less than 500 nm in diameter, optionally less than 250 nm in diameter, optionally less than 100 nm in diameter, optionally about 35 nm to about 60 nm in diameter.

[0715] Example particle delivery complexes are further disclosed in U.S. Provisional Application entitled "Nove Delivery of Large Payloads" filed 62/485,625 filed April 14, 2017.

[0716] Several types of particle delivery systems and/or formulations are known to be useful in a diverse spectrum of biomedical applications. In general, a particle is defined as a small object that behaves as a whole unit with respect to its transport and properties. Particles are further classified according to diameter. Coarse particles cover a range between 2,500 and 10,000 nanometers. Fine particles are sized between 100 and 2,500 nanometers. Ultrafine particles, or nanoparticles, are generally between 1 and 100 nanometers in size. The basis of the 100-nm limit is the fact that novel properties that differentiate particles from the bulk material typically develop at a critical length scale of under 100 nm.

[0717] As used herein, a particle delivery system/formulation is defined as any biological delivery system/formulation which includes a particle in accordance with the present invention. A particle in accordance with the present invention is any entity having a greatest dimension (e.g. diameter) of less than 100 microns (μm). In some embodiments, inventive particles have a greatest dimension of less than 10 μ m. In some embodiments, inventive particles have a greatest dimension of less than 2000 nanometers (nm). In some embodiments, inventive particles have a greatest dimension of less than 1000 nanometers (nm). In some embodiments, inventive particles have a greatest dimension of less than 900 nm, 800 nm, 700 nm, 600 nm, 500 nm, 400 nm, 300 nm, 200 nm, or 100 nm. Typically, inventive particles have a greatest dimension (e.g., diameter) of 500 nm or less. In some embodiments, inventive particles have a greatest dimension (e.g., diameter) of 250 nm or less. In some embodiments, inventive particles have a greatest dimension (e.g., diameter) of 200 nm or less. In some embodiments, inventive particles have a greatest dimension (e.g., diameter) of 150 nm or less. In some embodiments, inventive particles have a greatest dimension (e.g., diameter) of 100 nm or less. Smaller particles, e.g., having a greatest dimension of 50 nm or less are used in some embodiments of the invention. In some embodiments, inventive particles have a greatest dimension ranging between 25 nm and 200 nm.

[0718] In terms of this invention, it is preferred to have one or more components of CRISPR complex, e.g., CRISPR-Cas protein or mRNA, or adenosine deaminase (which may be fused to a CRISPR-Cas protein or an adaptor) or mRNA, or guide RNA delivered using nanoparticles or lipid envelopes. Other delivery systems or vectors are may be used in conjunction with the nanoparticle aspects of the invention.

[0719] In general, a "nanoparticle" refers to any particle having a diameter of less than 1000 nm. In certain preferred embodiments, nanoparticles of the invention have a greatest dimension (e.g., diameter) of 500 nm or less. In other preferred embodiments, nanoparticles of the invention have a greatest dimension ranging between 25 nm and 200 nm. In other preferred embodiments, nanoparticles of the invention have a greatest dimension of 100 nm or less. In other preferred embodiments, nanoparticles of the invention have a greatest dimension ranging between 35 nm and 60 nm. It will be appreciated that reference made herein to particles or nanoparticles can be interchangeable, where appropriate.

[0720] It will be understood that the size of the particle will differ depending as to whether it is measured before or after loading. Accordingly, in particular embodiments, the term “nanoparticles” may apply only to the particles pre loading.

[0721] Nanoparticles encompassed in the present invention may be provided in different forms, e.g., as solid nanoparticles (e.g., metal such as silver, gold, iron, titanium), non-metal, lipid-based solids, polymers), suspensions of nanoparticles, or combinations thereof. Metal, dielectric, and semiconductor nanoparticles may be prepared, as well as hybrid structures (e.g., core–shell nanoparticles). Nanoparticles made of semiconducting material may also be labeled quantum dots if they are small enough (typically sub 10 nm) that quantization of electronic energy levels occurs. Such nanoscale particles are used in biomedical applications as drug carriers or imaging agents and may be adapted for similar purposes in the present invention.

[0722] Semi-solid and soft nanoparticles have been manufactured, and are within the scope of the present invention. A prototype nanoparticle of semi-solid nature is the liposome. Various types of liposome nanoparticles are currently used clinically as delivery systems for anticancer drugs and vaccines. Nanoparticles with one half hydrophilic and the other half hydrophobic are termed Janus particles and are particularly effective for stabilizing emulsions. They can self-assemble at water/oil interfaces and act as solid surfactants.

[0723] Particle characterization (including e.g., characterizing morphology, dimension, etc.) is done using a variety of different techniques. Common techniques are electron microscopy (TEM, SEM), atomic force microscopy (AFM), dynamic light scattering (DLS), X-ray photoelectron spectroscopy (XPS), powder X-ray diffraction (XRD), Fourier transform

infrared spectroscopy (FTIR), matrix-assisted laser desorption/ionization time-of-flight mass spectrometry(MALDI-TOF), ultraviolet-visible spectroscopy, dual polarization interferometry and nuclear magnetic resonance (NMR). Characterization (dimension measurements) may be made as to native particles (i.e., preloading) or after loading of the cargo (herein cargo refers to e.g., one or more components of CRISPR-Cas system e.g., CRISPR-Cas protein or mRNA, adenosine deaminase (which may be fused to a CRISPR-Cas protein or an adaptor) or mRNA, or guide RNA, or any combination thereof, and may include additional carriers and/or excipients) to provide particles of an optimal size for delivery for any in vitro, ex vivo and/or in vivo application of the present invention. In certain preferred embodiments, particle dimension (e.g., diameter) characterization is based on measurements using dynamic laser scattering (DLS). Mention is made of US Patent No. 8,709,843; US Patent No. 6,007,845; US Patent No. 5,855,913; US Patent No. 5,985,309; US. Patent No. 5,543,158; and the publication by James E. Dahlman and Carmen Barnes et al. Nature Nanotechnology (2014) published online 11 May 2014, doi:10.1038/nnano.2014.84, concerning particles, methods of making and using them and measurements thereof.

[0724] Particles delivery systems within the scope of the present invention may be provided in any form, including but not limited to solid, semi-solid, emulsion, or colloidal particles. As such any of the delivery systems described herein, including but not limited to, e.g., lipid-based systems, liposomes, micelles, microvesicles, exosomes, or gene gun may be provided as particle delivery systems within the scope of the present invention.

[0725] CRISPR-Cas protein mRNA, adenosine deaminase (which may be fused to a CRISPR-Cas protein or an adaptor) or mRNA, and guide RNA may be delivered simultaneously using particles or lipid envelopes; for instance, CRISPR-Cas protein and RNA of the invention, e.g., as a complex, can be delivered via a particle as in Dahlman et al., WO2015089419 A2 and documents cited therein, such as 7C1 (see, e.g., James E. Dahlman and Carmen Barnes et al. Nature Nanotechnology (2014) published online 11 May 2014, doi:10.1038/nnano.2014.84), e.g., delivery particle comprising lipid or lipidoid and hydrophilic polymer, e.g., cationic lipid and hydrophilic polymer, for instance wherein the cationic lipid comprises 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP) or 1,2-ditetradecanoyl-sn-glycero-3-phosphocholine (DMPC) and/or wherein the hydrophilic polymer comprises ethylene glycol or polyethylene glycol (PEG); and/or wherein the particle further comprises cholesterol (e.g., particle from formulation 1 = DOTAP 100, DMPC 0, PEG 0, Cholesterol 0; formulation number 2 = DOTAP 90, DMPC 0, PEG 10, Cholesterol 0; formulation number 3 = DOTAP 90, DMPC 0, PEG 5, Cholesterol 5), wherein particles are formed using an efficient,

multistep process wherein first, effector protein and RNA are mixed together, e.g., at a 1:1 molar ratio, e.g., at room temperature, e.g., for 30 minutes, e.g., in sterile, nuclease free 1X PBS; and separately, DOTAP, DMPC, PEG, and cholesterol as applicable for the formulation are dissolved in alcohol, e.g., 100% ethanol; and, the two solutions are mixed together to form particles containing the complexes).

[0726] Nucleic acid-targeting effector proteins (e.g., a Type V protein such as Cas13) mRNA and guide RNA may be delivered simultaneously using particles or lipid envelopes. Examples of suitable particles include but are not limited to those described in US 9,301,923.

[0727] For example, Su X, Fricke J, Kavanagh DG, Irvine DJ ("In vitro and in vivo mRNA delivery using lipid-enveloped pH-responsive polymer nanoparticles" Mol Pharm. 2011 Jun 6;8(3):774-87. doi: 10.1021/mp100390w. Epub 2011 Apr 1) describes biodegradable core-shell structured nanoparticles with a poly(β-amino ester) (PBAE) core enveloped by a phospholipid bilayer shell. These were developed for in vivo mRNA delivery. The pH-responsive PBAE component was chosen to promote endosome disruption, while the lipid surface layer was selected to minimize toxicity of the polycation core. Such are, therefore, preferred for delivering RNA of the present invention.

[0728] In one embodiment, particles/nanoparticles based on self assembling bioadhesive polymers are contemplated, which may be applied to oral delivery of peptides, intravenous delivery of peptides and nasal delivery of peptides, all to the brain. Other embodiments, such as oral absorption and ocular delivery of hydrophobic drugs are also contemplated. The molecular envelope technology involves an engineered polymer envelope which is protected and delivered to the site of the disease (see, e.g., Mazza, M. et al. ACSNano, 2013. 7(2): 1016-1026; Siew, A., et al. Mol Pharm, 2012. 9(1):14-28; Lalatsa, A., et al. J Contr Rel, 2012. 161(2):523-36; Lalatsa, A., et al., Mol Pharm, 2012. 9(6):1665-80; Lalatsa, A., et al. Mol Pharm, 2012. 9(6):1764-74; Garrett, N.L., et al. J Biophotonics, 2012. 5(5-6):458-68; Garrett, N.L., et al. J Raman Spect, 2012. 43(5):681-688; Ahmad, S., et al. J Royal Soc Interface 2010. 7:S423-33; Uchegbu, I.F. Expert Opin Drug Deliv, 2006. 3(5):629-40; Qu, X.,et al. Biomacromolecules, 2006. 7(12):3452-9 and Uchegbu, I.F., et al. Int J Pharm, 2001. 224:185-199). Doses of about 5 mg/kg are contemplated, with single or multiple doses, depending on the target tissue.

[0729] In one embodiment, particles/nanoparticles that can deliver RNA to a cancer cell to stop tumor growth developed by Dan Anderson's lab at MIT may be used/and or adapted to the AD-functionalized CRISPR-Cas system of the present invention. In particular, the Anderson lab developed fully automated, combinatorial systems for the synthesis, purification,

characterization, and formulation of new biomaterials and nanoformulations. See, e.g., Alabi et al., Proc Natl Acad Sci U S A. 2013 Aug 6;110(32):12881-6; Zhang et al., Adv Mater. 2013 Sep 6;25(33):4641-5; Jiang et al., Nano Lett. 2013 Mar 13;13(3):1059-64; Karagiannis et al., ACS Nano. 2012 Oct 23;6(10):8484-7; Whitehead et al., ACS Nano. 2012 Aug 28;6(8):6922-9 and Lee et al., Nat Nanotechnol. 2012 Jun 3;7(6):389-93.

[0730] US patent application 20110293703 relates to lipidoid compounds are also particularly useful in the administration of polynucleotides, which may be applied to deliver the AD-functionalized CRISPR-Cas system of the present invention. In one aspect, the aminoalcohol lipidoid compounds are combined with an agent to be delivered to a cell or a subject to form microparticles, nanoparticles, liposomes, or micelles. The agent to be delivered by the particles, liposomes, or micelles may be in the form of a gas, liquid, or solid, and the agent may be a polynucleotide, protein, peptide, or small molecule. The aminoalcohol lipidoid compounds may be combined with other aminoalcohol lipidoid compounds, polymers (synthetic or natural), surfactants, cholesterol, carbohydrates, proteins, lipids, etc. to form the particles. These particles may then optionally be combined with a pharmaceutical excipient to form a pharmaceutical composition.

[0731] US Patent Publication No. 20110293703 also provides methods of preparing the aminoalcohol lipidoid compounds. One or more equivalents of an amine are allowed to react with one or more equivalents of an epoxide-terminated compound under suitable conditions to form an aminoalcohol lipidoid compound of the present invention. In certain embodiments, all the amino groups of the amine are fully reacted with the epoxide-terminated compound to form tertiary amines. In other embodiments, all the amino groups of the amine are not fully reacted with the epoxide-terminated compound to form tertiary amines thereby resulting in primary or secondary amines in the aminoalcohol lipidoid compound. These primary or secondary amines are left as is or may be reacted with another electrophile such as a different epoxide-terminated compound. As will be appreciated by one skilled in the art, reacting an amine with less than excess of epoxide-terminated compound will result in a plurality of different aminoalcohol lipidoid compounds with various numbers of tails. Certain amines may be fully functionalized with two epoxide-derived compound tails while other molecules will not be completely functionalized with epoxide-derived compound tails. For example, a diamine or polyamine may include one, two, three, or four epoxide-derived compound tails off the various amino moieties of the molecule resulting in primary, secondary, and tertiary amines. In certain embodiments, all the amino groups are not fully functionalized. In certain embodiments, two of the same types of epoxide-terminated compounds are used. In other embodiments, two or

more different epoxide-terminated compounds are used. The synthesis of the aminoalcohol lipidoid compounds is performed with or without solvent, and the synthesis may be performed at higher temperatures ranging from 30-100 °C., preferably at approximately 50-90 °C. The prepared aminoalcohol lipidoid compounds may be optionally purified. For example, the mixture of aminoalcohol lipidoid compounds may be purified to yield an aminoalcohol lipidoid compound with a particular number of epoxide-derived compound tails. Or the mixture may be purified to yield a particular stereo- or regioisomer. The aminoalcohol lipidoid compounds may also be alkylated using an alkyl halide (e.g., methyl iodide) or other alkylating agent, and/or they may be acylated.

[0732] US Patent Publication No. 20110293703 also provides libraries of aminoalcohol lipidoid compounds prepared by the inventive methods. These aminoalcohol lipidoid compounds may be prepared and/or screened using high-throughput techniques involving liquid handlers, robots, microtiter plates, computers, etc. In certain embodiments, the aminoalcohol lipidoid compounds are screened for their ability to transfect polynucleotides or other agents (e.g., proteins, peptides, small molecules) into the cell.

[0733] US Patent Publication No. 20130302401 relates to a class of poly(beta-amino alcohols) (PBAAs) has been prepared using combinatorial polymerization. The inventive PBAAs may be used in biotechnology and biomedical applications as coatings (such as coatings of films or multilayer films for medical devices or implants), additives, materials, excipients, non-biofouling agents, micropatterning agents, and cellular encapsulation agents. When used as surface coatings, these PBAAs elicited different levels of inflammation, both in vitro and in vivo, depending on their chemical structures. The large chemical diversity of this class of materials allowed us to identify polymer coatings that inhibit macrophage activation in vitro. Furthermore, these coatings reduce the recruitment of inflammatory cells, and reduce fibrosis, following the subcutaneous implantation of carboxylated polystyrene microparticles. These polymers may be used to form polyelectrolyte complex capsules for cell encapsulation. The invention may also have many other biological applications such as antimicrobial coatings, DNA or siRNA delivery, and stem cell tissue engineering. The teachings of US Patent Publication No. 20130302401 may be applied to the AD-functionalized CRISPR-Cas system of the present invention.

[0734] Preassembled recombinant CRISPR-Cas complexes comprising Cas13, adenosine deaminase (which may be fused to Cas13 or an adaptor protein), and guide RNA may be transfected, for example by electroporation, resulting in high mutation rates and absence of detectable off-target mutations. Hur, J.K. et al, Targeted mutagenesis in mice by

electroporation of Cas13 ribonucleoproteins, Nat Biotechnol. 2016 Jun 6. doi: 10.1038/nbt.3596.

[0735] In terms of local delivery to the brain, this can be achieved in various ways. For instance, material can be delivered intrastriatally e.g. by injection. Injection can be performed stereotactically via a craniotomy.

[0736] In some embodiments, sugar-based particles may be used, for example GalNAc, as described herein and with reference to WO2014118272 (incorporated herein by reference) and Nair, JK et al., 2014, Journal of the American Chemical Society 136 (49), 16958-16961) and the teaching herein, especially in respect of delivery applies to all particles unless otherwise apparent. This may be considered to be a sugar-based particle and further details on other particle delivery systems and/or formulations are provided herein. GalNAc can therefore be considered to be a particle in the sense of the other particles described herein, such that general uses and other considerations, for instance delivery of said particles, apply to GalNAc particles as well. A solution-phase conjugation strategy may for example be used to attach triantennary GalNAc clusters (mol. wt. ~2000) activated as PFP (pentafluorophenyl) esters onto 5′-hexylamino modified oligonucleotides (5′-HA ASOs, mol. wt. ~8000 Da; Østergaard et al., Bioconjugate Chem., 2015, 26 (8), pp 1451–1455). Similarly, poly(acrylate) polymers have been described for in vivo nucleic acid delivery (see WO2013158141 incorporated herein by reference). In further alternative embodiments, pre-mixing CRISPR nanoparticles (or protein complexes) with naturally occurring serum proteins may be used in order to improve delivery (Akinc A et al, 2010, Molecular Therapy vol. 18 no. 7, 1357–1364).

Nanoclews

[0737] Further, the AD-functionalized CRISPR system may be delivered using nanoclews, for example as described in Sun W et al, *Cocoon-like self-degradable DNA nanoclew for anticancer drug delivery.,* J Am Chem Soc. 2014 Oct 22;136(42):14722-5. doi: 10.1021/ja5088024. Epub 2014 Oct 13. ; or in Sun W et al, *Self-Assembled DNA Nanoclews for the Efficient Delivery of CRISPR-Cas9 for Genome Editing.,* Angew Chem Int Ed Engl. 2015 Oct 5;54(41):12029-33. doi: 10.1002/anie.201506030. Epub 2015 Aug 27.

LNP

[0738] In some embodiments, delivery is by encapsulation of the Cas13 protein or mRNA form in a lipid particle such as an LNP. In some embodiments, therefore, lipid nanoparticles (LNPs) are contemplated. An antitransthyretin small interfering RNA has been encapsulated in lipid nanoparticles and delivered to humans (see, e.g., Coelho et al., N Engl J Med 2013;369:819-29), and such a system may be adapted and applied to the CRISPR Cas system

of the present invention. Doses of about 0.01 to about 1 mg per kg of body weight administered intravenously are contemplated. Medications to reduce the risk of infusion-related reactions are contemplated, such as dexamethasone, acetampinophen, diphenhydramine or cetirizine, and ranitidine are contemplated. Multiple doses of about 0.3 mg per kilogram every 4 weeks for five doses are also contemplated.

[0739] LNPs have been shown to be highly effective in delivering siRNAs to the liver (see, e.g., Tabernero et al., Cancer Discovery, April 2013, Vol. 3, No. 4, pages 363-470) and are therefore contemplated for delivering RNA encoding CRISPR Cas to the liver. A dosage of about four doses of 6 mg/kg of the LNP every two weeks may be contemplated. Tabernero et al. demonstrated that tumor regression was observed after the first 2 cycles of LNPs dosed at 0.7 mg/kg, and by the end of 6 cycles the patient had achieved a partial response with complete regression of the lymph node metastasis and substantial shrinkage of the liver tumors. A complete response was obtained after 40 doses in this patient, who has remained in remission and completed treatment after receiving doses over 26 months. Two patients with RCC and extrahepatic sites of disease including kidney, lung, and lymph nodes that were progressing following prior therapy with VEGF pathway inhibitors had stable disease at all sites for approximately 8 to 12 months, and a patient with PNET and liver metastases continued on the extension study for 18 months (36 doses) with stable disease.

[0740] However, the charge of the LNP must be taken into consideration. As cationic lipids combined with negatively charged lipids to induce nonbilayer structures that facilitate intracellular delivery. Because charged LNPs are rapidly cleared from circulation following intravenous injection, ionizable cationic lipids with pKa values below 7 were developed (see, e.g., Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011). Negatively charged polymers such as RNA may be loaded into LNPs at low pH values (e.g., pH 4) where the ionizable lipids display a positive charge. However, at physiological pH values, the LNPs exhibit a low surface charge compatible with longer circulation times. Four species of ionizable cationic lipids have been focused upon, namely 1,2-dilineoyl-3-dimethylammonium-propane (DLinDAP), 1,2-dilinoleyloxy-3-N,N-dimethylaminopropane (DLinDMA), 1,2-dilinoleyloxy-keto-N,N-dimethyl-3-aminopropane (DLinKDMA), and 1,2-dilinoleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLinKC2-DMA). It has been shown that LNP siRNA systems containing these lipids exhibit remarkably different gene silencing properties in hepatocytes in vivo, with potencies varying according to the series DLinKC2-DMA>DLinKDMA>DLinDMA>>DLinDAP employing a Factor VII gene silencing model (see, e.g., Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011). A

dosage of 1 μg/ml of LNP or CRISPR-Cas RNA in or associated with the LNP may be contemplated, especially for a formulation containing DLinKC2-DMA.

[0741] Preparation of LNPs and CRISPR Cas encapsulation may be used/and or adapted from Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011). The cationic lipids 1,2-dilineoyl-3-dimethylammonium-propane (DLinDAP), 1,2-dilinoleyloxy-3-N,N-dimethylaminopropane (DLinDMA), 1,2-dilinoleyloxyketo-N,N-dimethyl-3-aminopropane (DLinK-DMA), 1,2-dilinoleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLinKC2-DMA), (3-o-[2″-(methoxypolyethyleneglycol 2000) succinoyl]-1,2-dimyristoyl-sn-glycol (PEG-S-DMG), and R-3-[(ω-methoxy-poly(ethylene glycol)2000) carbamoyl]-1,2-dimyristyloxlpropyl-3-amine (PEG-C-DOMG) may be provided by Tekmira Pharmaceuticals (Vancouver, Canada) or synthesized. Cholesterol may be purchased from Sigma (St Louis, MO). The specific CRISPR Cas RNA may be encapsulated in LNPs containing DLinDAP, DLinDMA, DLinK-DMA, and DLinKC2-DMA (cationic lipid:DSPC:CHOL: PEGS-DMG or PEG-C-DOMG at 40:10:40:10 molar ratios). When required, 0.2% SP-DiOC18 (Invitrogen, Burlington, Canada) may be incorporated to assess cellular uptake, intracellular delivery, and biodistribution. Encapsulation may be performed by dissolving lipid mixtures comprised of cationic lipid:DSPC:cholesterol:PEG-c-DOMG (40:10:40:10 molar ratio) in ethanol to a final lipid concentration of 10 mmol/l. This ethanol solution of lipid may be added drop-wise to 50 mmol/l citrate, pH 4.0 to form multilamellar vesicles to produce a final concentration of 30% ethanol vol/vol. Large unilamellar vesicles may be formed following extrusion of multilamellar vesicles through two stacked 80 nm Nuclepore polycarbonate filters using the Extruder (Northern Lipids, Vancouver, Canada). Encapsulation may be achieved by adding RNA dissolved at 2 mg/ml in 50 mmol/l citrate, pH 4.0 containing 30% ethanol vol/vol drop-wise to extruded preformed large unilamellar vesicles and incubation at 31 °C for 30 minutes with constant mixing to a final RNA/lipid weight ratio of 0.06/1 wt/wt. Removal of ethanol and neutralization of formulation buffer were performed by dialysis against phosphate-buffered saline (PBS), pH 7.4 for 16 hours using Spectra/Por 2 regenerated cellulose dialysis membranes. Nanoparticle size distribution may be determined by dynamic light scattering using a NICOMP 370 particle sizer, the vesicle/intensity modes, and Gaussian fitting (Nicomp Particle Sizing, Santa Barbara, CA). The particle size for all three LNP systems may be ~70 nm in diameter. RNA encapsulation efficiency may be determined by removal of free RNA using VivaPureD MiniH columns (Sartorius Stedim Biotech) from samples collected before and after dialysis. The encapsulated RNA may be extracted from the eluted nanoparticles and quantified at 260 nm. RNA to lipid ratio was determined by measurement of cholesterol content

in vesicles using the Cholesterol E enzymatic assay from Wako Chemicals USA (Richmond, VA). In conjunction with the herein discussion of LNPs and PEG lipids, PEGylated liposomes or LNPs are likewise suitable for delivery of a CRISPR-Cas system or components thereof.

[0742] A lipid premix solution (20.4 mg/ml total lipid concentration) may be prepared in ethanol containing DLinKC2-DMA, DSPC, and cholesterol at 50:10:38.5 molar ratios. Sodium acetate may be added to the lipid premix at a molar ratio of 0.75:1 (sodium acetate:DLinKC2-DMA). The lipids may be subsequently hydrated by combining the mixture with 1.85 volumes of citrate buffer (10 mmol/l, pH 3.0) with vigorous stirring, resulting in spontaneous liposome formation in aqueous buffer containing 35% ethanol. The liposome solution may be incubated at 37 °C to allow for time-dependent increase in particle size. Aliquots may be removed at various times during incubation to investigate changes in liposome size by dynamic light scattering (Zetasizer Nano ZS, Malvern Instruments, Worcestershire, UK). Once the desired particle size is achieved, an aqueous PEG lipid solution (stock = 10 mg/ml PEG-DMG in 35% (vol/vol) ethanol) may be added to the liposome mixture to yield a final PEG molar concentration of 3.5% of total lipid. Upon addition of PEG-lipids, the liposomes should their size, effectively quenching further growth. RNA may then be added to the empty liposomes at an RNA to total lipid ratio of approximately 1:10 (wt:wt), followed by incubation for 30 minutes at 37 °C to form loaded LNPs. The mixture may be subsequently dialyzed overnight in PBS and filtered with a 0.45-μm syringe filter.

[0743] Spherical Nucleic Acid (SNA™) constructs and other nanoparticles (particularly gold nanoparticles) are also contemplated as a means to delivery CRISPR-Cas system to intended targets. Significant data show that AuraSense Therapeutics' Spherical Nucleic Acid (SNA™) constructs, based upon nucleic acid-functionalized gold nanoparticles, are useful.

[0744] Literature that may be employed in conjunction with herein teachings include: Cutler et al., J. Am. Chem. Soc. 2011 133:9254-9257, Hao et al., Small. 2011 7:3158-3162, Zhang et al., ACS Nano. 2011 5:6962-6970, Cutler et al., J. Am. Chem. Soc. 2012 134:1376-1391, Young et al., Nano Lett. 2012 12:3867-71, Zheng et al., Proc. Natl. Acad. Sci. USA. 2012 109:11975-80, Mirkin, Nanomedicine 2012 7:635-638 Zhang et al., J. Am. Chem. Soc. 2012 134:16488-1691, Weintraub, Nature 2013 495:S14-S16, Choi et al., Proc. Natl. Acad. Sci. USA. 2013 110(19):7625-7630, Jensen et al., Sci. Transl. Med. 5, 209ra152 (2013) and Mirkin, et al., Small, 10:186-192.

[0745] Self-assembling nanoparticles with RNA may be constructed with polyethyleneimine (PEI) that is PEGylated with an Arg-Gly-Asp (RGD) peptide ligand attached at the distal end of the polyethylene glycol (PEG). This system has been used, for

example, as a means to target tumor neovasculature expressing integrins and deliver siRNA inhibiting vascular endothelial growth factor receptor-2 (VEGF R2) expression and thereby achieve tumor angiogenesis (see, e.g., Schiffelers et al., Nucleic Acids Research, 2004, Vol. 32, No. 19). Nanoplexes may be prepared by mixing equal volumes of aqueous solutions of cationic polymer and nucleic acid to give a net molar excess of ionizable nitrogen (polymer) to phosphate (nucleic acid) over the range of 2 to 6. The electrostatic interactions between cationic polymers and nucleic acid resulted in the formation of polyplexes with average particle size distribution of about 100 nm, hence referred to here as nanoplexes. A dosage of about 100 to 200 mg of CRISPR Cas is envisioned for delivery in the self-assembling nanoparticles of Schiffelers et al.

[0746] The nanoplexes of Bartlett et al. (PNAS, September 25, 2007,vol. 104, no. 39) may also be applied to the present invention. The nanoplexes of Bartlett et al. are prepared by mixing equal volumes of aqueous solutions of cationic polymer and nucleic acid to give a net molar excess of ionizable nitrogen (polymer) to phosphate (nucleic acid) over the range of 2 to 6. The electrostatic interactions between cationic polymers and nucleic acid resulted in the formation of polyplexes with average particle size distribution of about 100 nm, hence referred to here as nanoplexes. The DOTA-siRNA of Bartlett et al. was synthesized as follows: 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid mono(N-hydroxysuccinimide ester) (DOTA-NHSester) was ordered from Macrocyclics (Dallas, TX). The amine modified RNA sense strand with a 100-fold molar excess of DOTA-NHS-ester in carbonate buffer (pH 9) was added to a microcentrifuge tube. The contents were reacted by stirring for 4 h at room temperature. The DOTA-RNAsense conjugate was ethanol-precipitated, resuspended in water, and annealed to the unmodified antisense strand to yield DOTA-siRNA. All liquids were pretreated with Chelex-100 (Bio-Rad, Hercules, CA) to remove trace metal contaminants. Tf-targeted and nontargeted siRNA nanoparticles may be formed by using cyclodextrin-containing polycations. Typically, nanoparticles were formed in water at a charge ratio of 3 (+/-) and an siRNA concentration of 0.5 g/liter. One percent of the adamantane-PEG molecules on the surface of the targeted nanoparticles were modified with Tf (adamantane-PEG-Tf). The nanoparticles were suspended in a 5% (wt/vol) glucose carrier solution for injection.

[0747] Davis et al. (Nature, Vol 464, 15 April 2010) conducts a RNA clinical trial that uses a targeted nanoparticle-delivery system (clinical trial registration number NCT00689065). Patients with solid cancers refractory to standard-of-care therapies are administered doses of targeted nanoparticles on days 1, 3, 8 and 10 of a 21-day cycle by a 30-min intravenous infusion. The nanoparticles consist of a synthetic delivery system containing: (1) a linear,

cyclodextrin-based polymer (CDP), (2) a human transferrin protein (TF) targeting ligand displayed on the exterior of the nanoparticle to engage TF receptors (TFR) on the surface of the cancer cells, (3) a hydrophilic polymer (polyethylene glycol (PEG) used to promote nanoparticle stability in biological fluids), and (4) siRNA designed to reduce the expression of the RRM2 (sequence used in the clinic was previously denoted siR2B+5). The TFR has long been known to be upregulated in malignant cells, and RRM2 is an established anti-cancer target. These nanoparticles (clinical version denoted as CALAA-01) have been shown to be well tolerated in multi-dosing studies in non-human primates. Although a single patient with chronic myeloid leukaemia has been administered siRNA by liposomal delivery, Davis et al.'s clinical trial is the initial human trial to systemically deliver siRNA with a targeted delivery system and to treat patients with solid cancer. To ascertain whether the targeted delivery system can provide effective delivery of functional siRNA to human tumors, Davis et al. investigated biopsies from three patients from three different dosing cohorts; patients A, B and C, all of whom had metastatic melanoma and received CALAA-01 doses of 18, 24 and 30 mg m-2 siRNA, respectively. Similar doses may also be contemplated for the CRISPR Cas system of the present invention. The delivery of the invention may be achieved with nanoparticles containing a linear, cyclodextrin-based polymer (CDP), a human transferrin protein (TF) targeting ligand displayed on the exterior of the nanoparticle to engage TF receptors (TFR) on the surface of the cancer cells and/or a hydrophilic polymer (for example, polyethylene glycol (PEG) used to promote nanoparticle stability in biological fluids).

[0748] US Patent No. 8,709,843, incorporated herein by reference, provides a drug delivery system for targeted delivery of therapeutic agent-containing particles to tissues, cells, and intracellular compartments. The invention provides targeted particles comprising comprising polymer conjugated to a surfactant, hydrophilic polymer or lipid. US Patent No. 6,007,845, incorporated herein by reference, provides particles which have a core of a multiblock copolymer formed by covalently linking a multifunctional compound with one or more hydrophobic polymers and one or more hydrophilic polymers, and conatin a biologically active material. US Patent No. 5,855,913, incorporated herein by reference, provides a particulate composition having aerodynamically light particles having a tap density of less than 0.4 g/cm3 with a mean diameter of between 5 μm and 30 μ m, incorporating a surfactant on the surface thereof for drug delivery to the pulmonary system. US Patent No. 5,985,309, incorporated herein by reference, provides particles incorporating a surfactant and/or a hydrophilic or hydrophobic complex of a positively or negatively charged therapeutic or diagnostic agent and a charged molecule of opposite charge for delivery to the pulmonary system. US. Patent No.

5,543,158, incorporated herein by reference, provides biodegradable injectable particles having a biodegradable solid core containing a biologically active material and poly(alkylene glycol) moieties on the surface. WO2012135025 (also published as US20120251560), incorporated herein by reference, describes conjugated polyethyleneimine (PEI) polymers and conjugated aza-macrocycles (collectively referred to as "conjugated lipomer" or "lipomers"). In certain embodiments, it can envisioned that such conjugated lipomers can be used in the context of the CRISPR-Cas system to achieve in vitro, ex vivo and in vivo genomic perturbations to modify gene expression, including modulation of protein expression.

[0749] In one embodiment, the nanoparticle may be epoxide-modified lipid–polymer, advantageously 7C1 (see, e.g., James E. Dahlman and Carmen Barnes et al. Nature Nanotechnology (2014) published online 11 May 2014, doi:10.1038/nnano.2014.84). C71 was synthesized by reacting C15 epoxide-terminated lipids with PEI600 at a 14:1 molar ratio, and was formulated with C14PEG2000 to produce nanoparticles (diameter between 35 and 60 nm) that were stable in PBS solution for at least 40 days.

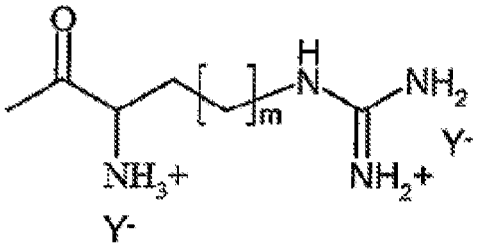
[0750] An epoxide-modified lipid-polymer may be utilized to deliver the CRISPR-Cas system of the present invention to pulmonary, cardiovascular or renal cells, however, one of skill in the art may adapt the system to deliver to other target organs. Dosage ranging from about 0.05 to about 0.6 mg/kg are envisioned. Dosages over several days or weeks are also envisioned, with a total dosage of about 2 mg/kg.

[0751] In some embodiments, the LNP for deliverting the RNA molecules is prepared by methods known in the art, such as those described in, for example, WO 2005/105152 (PCT/EP2005/004920), WO 2006/069782 (PCT/EP2005/014074), WO 2007/121947 (PCT/EP2007/003496), and WO 2015/082080 (PCT/EP2014/003274), which are herein incorporated by reference. LNPs aimed specifically at the enhanced and improved delivery of siRNA into mammalian cells are described in, for example, Aleku et al., Cancer Res., 68(23): 9788-98 (Dec. 1, 2008), Strumberg et al., Int. J. Clin. Pharmacol. Ther., 50(1): 76-8 (Jan. 2012), Schultheis et al., J. Clin. Oncol., 32(36): 4141-48 (Dec. 20, 2014), and Fehring et al., Mol. Ther., 22(4): 811-20 (Apr. 22, 2014), which are herein incorporated by reference and may be applied to the present technology.

[0752] In some embodiments, the LNP includes any LNP disclosed in WO 2005/105152 (PCT/EP2005/004920), WO 2006/069782 (PCT/EP2005/014074), WO 2007/121947 (PCT/EP2007/003496), and WO 2015/082080 (PCT/EP2014/003274).

[0753] In some embodiments, the LNP includes at least one lipid having Formula I:

(Formula I), wherein R1 and R2 are each and independently selected from the group comprising alkyl, n is any integer between 1 and 4, and R3 is an acyl selected from the group comprising lysyl, ornithyl, 2,4-diaminobutyryl, histidyl and an acyl moiety according to Formula II:



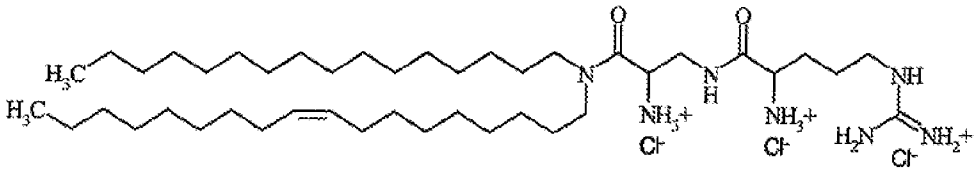
(Formula II),

wherein m is any integer from 1 to 3 and Y^- is a pharmaceutically acceptable anion. In some embodiments, a lipid according to Formula I includes at least two asymmetric C atoms. In some embodiments, enantiomers of Formula I include, but are not limited to, R-R; S-S; R-S and S-R enantiomer.

[0754] In some embodiments, R1 is lauryl and R2 is myristyl. In another embodiment, R1 is palmityl and R2 is oleyl. In some embodiments, m is 1 or 2. In some embodiments, Y- is selected from halogenids, acetate or trifluoroacetate.

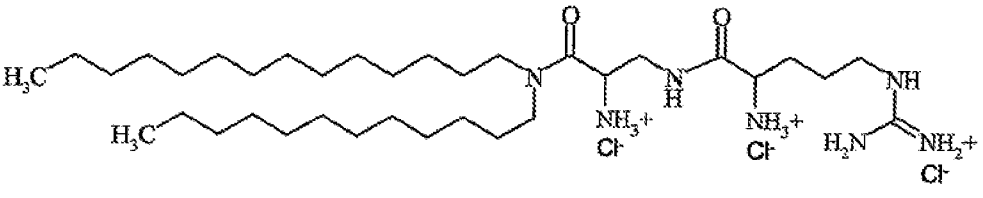
[0755] In some embodiments, the LNP comprises one or more lipids select from:

β-arginyl-2,3-diamino propionic acid-N-palmityl-N-oleyl-amide trihydrochloride (Formula III):



(Formula III);

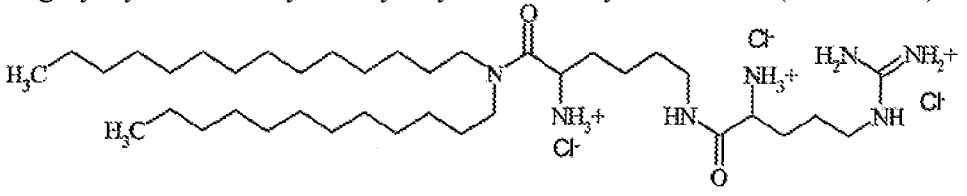
β-arginyl-2,3-diamino propionic acid-N-lauryl-N-myristyl-amide trihydrochloride (Formula IV):



(Formula IV);

and

ε-arginyl-lysine-N-lauryl-N-myristyl-amide trihydrochloride (Formula V):



(Formula V).

[0756] In some embodiments, the LNP also includes a constituent. By way of example, but not by way of limitation, in some embodiments, the constituent is selected from peptides, proteins, oligonucleotides, polynucleotides, nucleic acids, or a combination thereof. In some

embodiments, the constituent is an antibody, e.g., a monoclonal antibody. In some embodiments, the constituent is a nucleic acid selected from, e.g., ribozymes, aptamers, spiegelmers, DNA, RNA, PNA, LNA, or a combination thereof. In some embodiments, the nucleic acid is guide RNA and/or mRNA.

[0757] In some embodiments, the constituent of the LNP comprises an mRNA encoding a CRIPSR-Cas protein. In some embodiments, the constituent of the LNP comprises an mRNA encoding a Type-II or Type-V CRIPSR-Cas protein. In some embodiments, the constituent of the LNP comprises an mRNA encoding an adenosine deaminase (which may be fused to a CRISPR-Cas protein or an adaptor protein).

[0758] In some embodiments, the constituent of the LNP further comprises one or more guide RNA. In some embodiments, the LNP is configured to deliver the aforementioned mRNA and guide RNA to vascular endothelium. In some embodiments, the LNP is configured to deliver the aforementioned mRNA and guide RNA to pulmonary endothelium. In some embodiments, the LNP is configured to deliver the aforementioned mRNA and guide RNA to liver. In some embodiments, the LNP is configured to deliver the aforementioned mRNA and guide RNA to lung. In some embodiments, the LNP is configured to deliver the aforementioned mRNA and guide RNA to hearts. In some embodiments, the LNP is configured to deliver the aforementioned mRNA and guide RNA to spleen. In some embodiments, the LNP is configured to deliver the aforementioned mRNA and guide RNA to kidney. In some embodiments, the LNP is configured to deliver the aforementioned mRNA and guide RNA to pancrea. In some embodiments, the LNP is configured to deliver the aforementioned mRNA and guide RNA to brain. In some embodiments, the LNP is configured to deliver the aforementioned mRNA and guide RNA to macrophages.

[0759] In some embodiments, the LNP also includes at least one helper lipid. In some embodiments, the helper lipid is selected from phospholipids and steroids. In some embodiments, the phospholipids are di- and /or monoester of the phosphoric acid. In some embodiments, the phospholipids are phosphoglycerides and /or sphingolipids. In some embodiments, the steroids are naturally occurring and/or synthetic compounds based on the partially hydrogenated cyclopenta[a]phenanthrene. In some embodiments, the steroids contain 21 to 30 C atoms. In some embodiments, the steroid is cholesterol. In some embodiments, the helper lipid is selected from 1,2-diphytanoyl-sn-glycero-3-phosphoethanolamine (DPhyPE), ceramide, and 1,2-dioleylsn-glycero-3-phosphoethanolamine (DOPE).

[0760] In some embodiments, the at least one helper lipid comprises a moiety selected from the group comprising a PEG moiety, a HEG moiety, a polyhydroxyethyl starch (polyHES)

moiety and a polypropylene moiety. In some embodiments, the moiety has a molecule weight between about 500 to 10,000 Da or between about 2,000 to 5,000 Da. In some embodiments, the PEG moiety is selected from l,2-distearoyl-sn-glycero-3 phosphoethanolamine, l,2-dialkyl-sn-glycero-3-phosphoethanolamine, and Ceramide-PEG. In some embodiments, the PEG moiety has a molecular weight between about 500 to 10,000 Da or between about 2,000 to 5,000 Da. In some embodiments, the PEG moiety has a molecular weight of 2,000 Da.

[0761] In some embodiments, the helper lipid is between about 20 mol % to 80 mol % of the total lipid content of the composition. In some embodiments, the helper lipid component is between about 35 mol % to 65 mol % of the total lipid content of the LNP. In some embodiments, the LNP includes lipids at 50 mol% and the helper lipid at 50 mol% of the total lipid content of the LNP.

[0762] In some embodiments, the LNP includes any of β-3-arginyl-2,3-diaminopropionic acid-N-palmityl-N-oleyl-amide trihydrochloride, β-arginyl-2,3-diaminopropionic acid-N-lauryl-N-myristyl-amide trihydrochloride or arginyl-lysine-N-lauryl-N-myristyl-amide trihydrochloride in combination with DPhyPE, wherein the content of DPhyPE is about 80 mol %, 65 mol %, 50 mol % and 35 mol % of the overall lipid content of the LNP. In some embodiments, the LNP includes β-arginyl-2,3-diamino propionic acid-N-pahnityl-N-oleyl-amide trihydrochloride (lipid) and 1,2-diphytanoyl-sn-glycero-3-phosphoethanolamine (helper lipid). In some embodiments, the LNP includes β-arginyl-2,3-diamino propionic acid-N-palmityl-N-oleyl-amide trihydrochloride (lipid), l,2-diphytanoyl-sn-glycero-3-phosphoethanolamine (first helper lipid), and l,2-disteroyl-sn-glycero-3-phosphoethanolamine-PEG2000 (second helper lipid).

[0763] In some embodiments, the second helper lipid is between about 0.05 mol% to 4.9 mol% or between about 1 mol% to 3 mol% of the total lipid content. In some embodiments, the LNP includes lipids at between about 45 mol% to 50 mol% of the total lipid content, a first helper lipid between about 45 mol% to 50 mol% of the total lipid content, under the proviso that there is a PEGylated second helper lipid between about 0.1 mol% to 5 mol %, between about 1 mol% to 4 mol%, or at about 2 mol% of the total lipid content, wherein the sum of the content of the lipids, the first helper lipid, and of the second helper lipid is 100 mol% of the total lipid content and wherein the sum of the first helper lipid and the second helper lipid is 50 mol% of the total lipid content. In some embodiments, the LNP comprises: (a) 50 mol% of β-arginyl-2,3-diamino propionic acid-N-palmityl-N-oleyl-amide trihydrochloride, 48 mol% of l,2-diphytanoyl-sn-glycero-3-phosphoethanolamine; and 2 mol% l,2-distearoyl-sn-glycero-3-

phosphoethanolamine-PEG2000; or (b) 50 mol% of β-arginyl-2,3-diamino propionic acid-N-palmityl-N-oleyl-amide trihydrocloride, 49 mol% l,2-diphytanoyl-sn-glycero-3-phosphoethanolamine; and 1 mol% N(Carbonyl-methoxypolyethylenglycol-2000)-1,2-distearoyl-sn-glycero3-phosphoethanolamine, or a sodium salt thereof.

[0764] In some embodiments, the LNP contains a nucleic acid, wherein the charge ratio of nucleic acid backbone phosphates to cationic lipid nitrogen atoms is about 1: 1.5 – 7 or about 1:4.

[0765] In some embodiments, the LNP also includes a shielding compound, which is removable from the lipid composition under in vivo conditions. In some embodiments, the shielding compound is a biologically inert compound. In some embodiments, the shielding compound does not carry any charge on its surface or on the molecule as such. In some embodiments, the shielding compounds are polyethylenglycoles (PEGs), hydroxyethylglucose (HEG) based polymers, polyhydroxyethyl starch (polyHES) and polypropylene. In some embodiments, the PEG, HEG, polyHES, and a polypropylene weight between about 500 to 10,000 Da or between about 2000 to 5000 Da. In some embodiments, the shielding compound is PEG2000 or PEG5000.

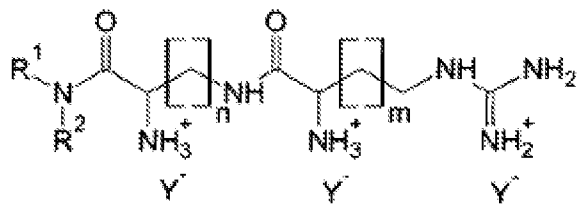
[0766] In some embodiments, the LNP includes at least one lipid, a first helper lipid, and a shielding compound that is removable from the lipid composition under in vivo conditions. In some embodiments, the LNP also includes a second helper lipid. In some embodiments, the first helper lipid is ceramide. In some embodiments, the second helper lipid is ceramide. In some embodiments, the ceramide comprises at least one short carbon chain substituent of from 6 to 10 carbon atoms. In some embodiments, the ceramide comprises 8 carbon atoms. In some embodiments, the shielding compound is attached to a ceramide. In some embodiments, the shielding compound is attached to a ceramide. In some embodiments, the shielding compound is covalently attached to the ceramide. In some embodiments, the shielding compound is attached to a nucleic acid in the LNP. In some embodiments, the shielding compound is covalently attached to the nucleic acid. In some embodiments, the shielding compound is attached to the nucleic acid by a linker. In some embodiments, the linker is cleaved under physiological conditions. In some embodiments, the linker is selected from ssRNA, ssDNA, dsRNA, dsDNA, peptide, S-S-linkers and pH sensitive linkers. In some embodiments, the linker moiety is attached to the 3' end of the sense strand of the nucleic acid. In some embodiments, the shielding compound comprises a pH-sensitive linker or a pH-sensitive moiety. In some embodiments, the pH-sensitive linker or pH-sensitive moiety is an anionic

linker or an anionic moiety. In some embodiments, the anionic linker or anionic moiety is less anionic or neutral in an acidic environment. In some embodiments, the pH-sensitive linker or the pH-sensitive moiety is selected from the oligo (glutamic acid), oligophenolate(s) and diethylene triamine penta acetic acid.

[0767] In any of the LNP embodiments in the previous paragraph, the LNP can have an osmolality between about 50 to 600 mosmole/kg, between about 250 to 350 mosmole/kg, or between about 280 to 320 mosmole/kg, and/or wherein the LNP formed by the lipid and/or one or two helper lipids and the shielding compound have a particle size between about 20 to 200 nm, between about 30 to 100 nm, or between about 40 to 80 nm.

[0768] In some embodiments, the shielding compound provides for a longer circulation time in vivo and allows for a better biodistribution of the nucleic acid containing LNP. In some embodiments, the shielding compound prevents immediate interaction of the LNP with serum compounds or compounds of other bodily fluids or cytoplasma membranes, e.g., cytoplasma membranes of the endothelial lining of the vasculature, into which the LNP is administered. Additionally or alternatively, in some embodiments, the shielding compounds also prevent elements of the immune system from immediately interacting with the LNP. Additionally or alternatively, in some embodiments, the shielding compound acts as an anti-opsonizing compound. Without wishing to be bound by any mechanism or theory, in some embodiments, the shielding compound forms a cover or coat that reduces the surface area of the LNP available for interaction with its environment. Additionally or alternatively, in some embodiments, the shielding compound shields the overall charge of the LNP.

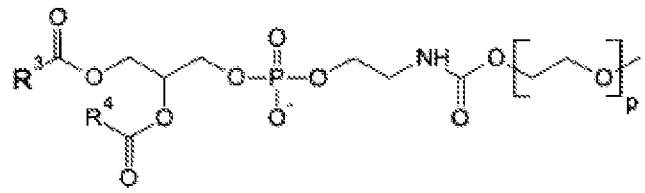
[0769] In another embodiment, the LNP includes at least one cationic lipid having Formula VI:



(Formula VI),

wherein n is 1, 2, 3, or 4, wherein m is 1, 2, or 3, wherein Y^- is anion, wherein each of R^1 and R^2 is individually and independently selected from the group consisting of linear C12-C18 alkyl and linear C12-C18 alkenyl, a sterol compound, wherein the sterol compound is selected from the group consisting of cholesterol and stigmasterol, and a PEGylated lipid, wherein the PEGylated lipid comprises a PEG moiety, wherein the PEGylated lipid is selected from the group consisting of:

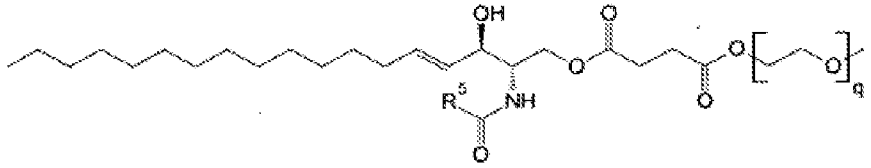
a PEGylated phosphoethanolamine of Formula VII:



(Formula VII),

wherein R^3 and R^4 are individually and independently linear C13-C17 alkyl, and p is any integer between 15 to 130;

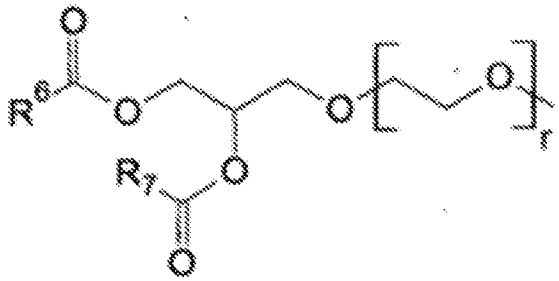
a PEGylated ceramide of Formula VIII:



(Formula VIII);

wherein R^5 is linear C7-C15 alkyl, and q is any number between 15 to 130; and

a PEGylated diacylglycerol of Formula IX:

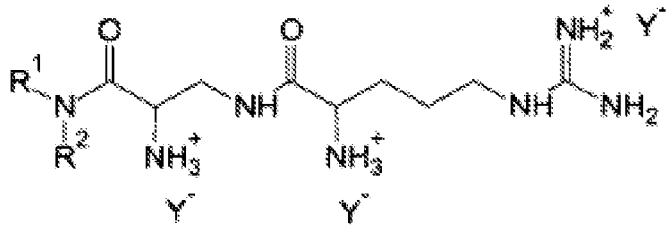


(Formula IX),

wherein each of R^6 and R^7 is individually and independently linear C11-C17 alkyl, and r is any integer from 15 to 130.

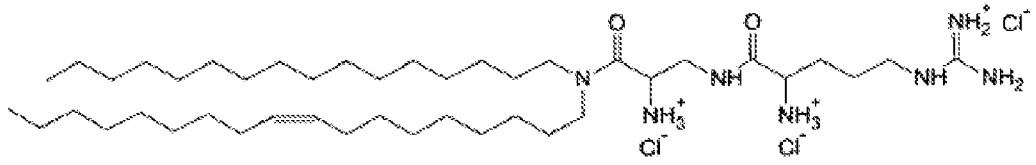
[0770] In some embodiments, R^1 and R^2 are different from each other. In some embodiments, R^1 is palmityl and R^2 is oleyl. In some embodiments, R^1 is lauryl and R^2 is myristyl. In some embodiments, R^1 and R^2 are the same. In some embodiments, each of R^1 and R^2 is individually and independently selected from the group consisting of C12 alkyl, C14 alkyl, Cl6 alkyl, Cl8 alkyl, Cl2 alkenyl, C14 alkenyl, Cl6 alkenyl and Cl8 alkenyl. In some embodiments, each of C12 alkenyl, C14 alkenyl, Cl6 alkenyl and Cl 8 alkenyl comprises one or two double bonds. In some embodiments, Cl8 alkenyl is Cl8 alkenyl with one double bond between C9 and C10. In some embodiments, C18 alkenyl is cis-9-octadecyl.

[0771] In some embodiments, the cationic lipid is a compound of Formula X:

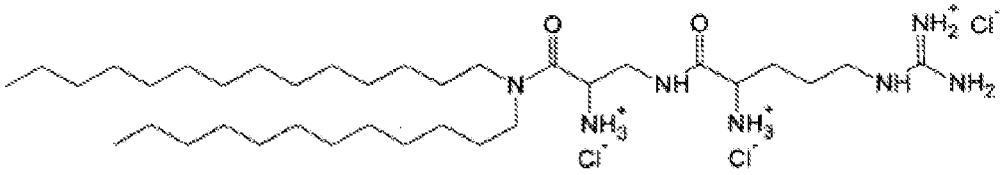


(Formula X). In some embodiments, Y^- is selected from halogenids, acetate and trifluoroacetate. In some

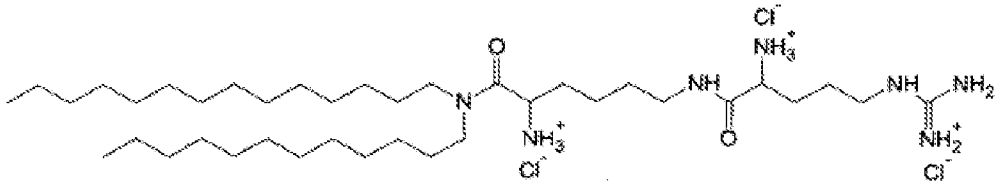
embodiments, the cationic lipid is β-arginyl-2,3-diamino propionic acid-N-palmityl-N-oleyl-amide trihydrochloride of Formula III:



(Formula III). In some embodiments, the cationic lipid is β-arginyl-2,3-diamino propionic acid-N-lauryl-N-myristyl-amide trihydrochloride of Formula IV:



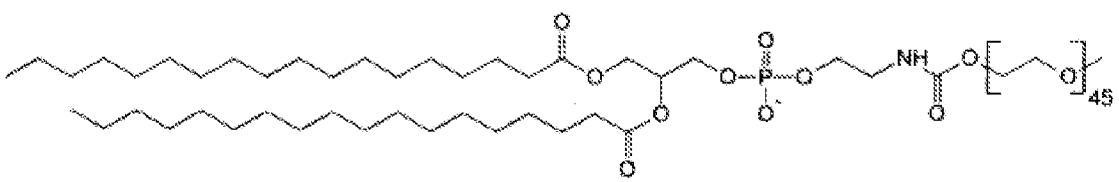
(Formula IV). In some embodiments, the cationic lipid is arginyl-lysine-N-lauryl-N-myristyl-amide trihydrochloride of Formula V:



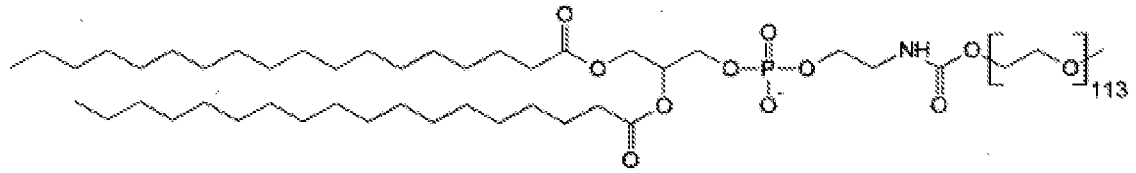
(Formula V).

[0772] In some embodiments, the sterol compound is cholesterol. In some embodiments, the sterol compound is stigmasterin.

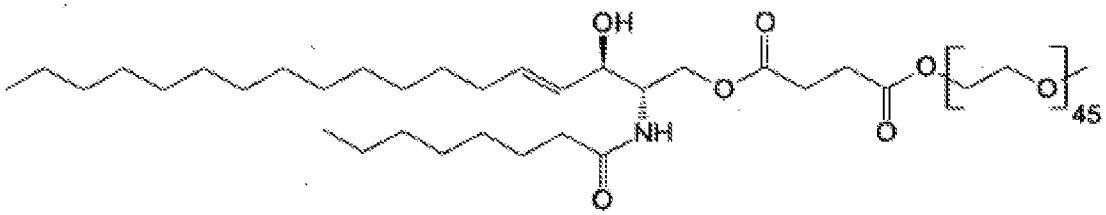
[0773] In some embodiments, the PEG moiety of the PEGylated lipid has a molecular weight from about 800 to 5,000 Da. In some embodiments, the molecular weight of the PEG moiety of the PEGylated lipid is about 800 Da. In some embodiments, the molecular weight of the PEG moiety of the PEGylated lipid is about 2,000 Da. In some embodiments, the molecular weight of the PEG moiety of the PEGylated lipid is about 5,000 Da.In some embodiments, the PEGylated lipid is a PEGylated phosphoethanolamine of Formula VII, wherein each of R^3 and R^4 is individually and independently linear C13-C17 alkyl, and p is any integer from 18, 19 or 20, or from 44, 45 or 46 or from 113, 114 or 115. In some embodiments, R^3 and R^4 are the same. In some embodiments, R^3 and R^4 are different. In some embodiments, each of R^3 and R^4 is individually and independently selected from the group consisting of C13 alkyl, Cl5 alkyl and Cl7 alkyl. In some embodiments, the PEGylated phosphoethanolamine of Formula VII is l,2-distearoyl-*sn*-glycero-3-phosphoethanolamine-N-[methoxy(polyethylene glycol)-2000] (ammonium salt):



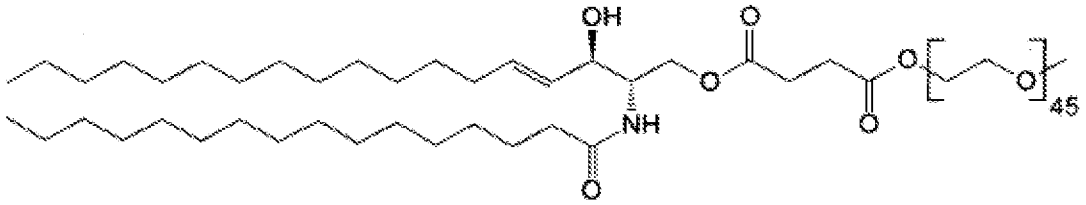
(Formula XI). In some embodiments, the PEGylated phosphoethanolamine of Formula VII is l,2-distearoyl-*sn*-glycero-3-phosphoethanolamine-N-[methoxy(polyethylene glycol)-5000] (ammonium salt):



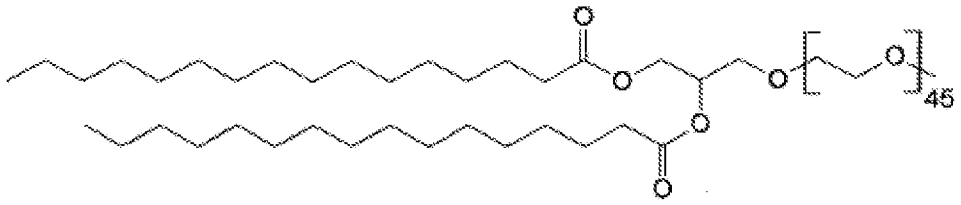
(Formula XII). In some embodiments, the PEGylated lipid is a PEGylated ceramide of Formula VIII, wherein R^5 is linear C7-C15 alkyl, and q is any integer from 18, 19 or 20, or from 44, 45 or 46 or from 113, 114 or 115. In some embodiments, R^5 is linear C7 alkyl. In some embodiments, R^5 is linear Cl5 alkyl.In some embodiments, the PEGylated ceramide of Formula VIII is N-octanoyl-sphingosine-1- {succinyl[methoxy(polyethylene glycol)2000]}:



(Formula XIII). In some embodiments, the PEGylated ceramide of Formula VIII is N-palmitoyl-sphingosine-1- {succinyl[methoxy(polyethylene glycol)2000]}

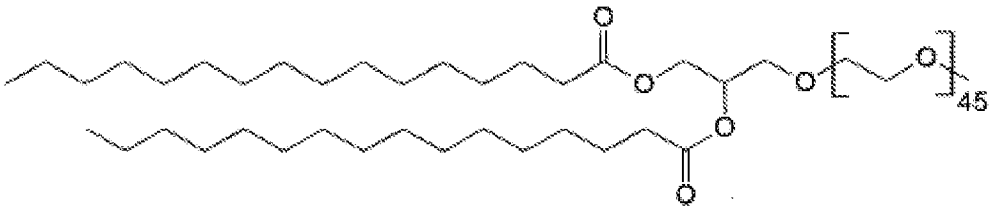


(Formula XIV). In some embodiments, the PEGylated lipid is a PEGylated diacylglycerol of Formula IX, wherein each of R^6 and R^7 is individually and independently linear Cl1-Cl7 alkyl, and r is any integer from 18, 19 or 20, or from 44, 45 or 46 or from 113, 114 or 115. In some embodiments, R^6 and R^7 are the same. In some embodiments, R^6 and R^7 are different. In some embodiments, each of R^6 and R^7 is individually and independently selected from the group consisting of linear Cl7 alkyl, linear C15 alkyl and linear Cl3 alkyl. In some embodiments, the PEGylated diacylglycerol of Formula IX 1,2-Distearoyl-sn-glycerol [methoxy(polyethylene glycol)2000]:

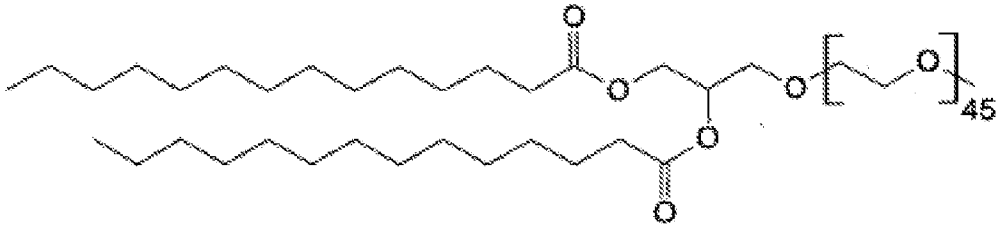


(Formula XV).

In some embodiments, the PEGylated diacylglycerol of Formula IX is 1,2-Dipalmitoyl-sn-glycerol [methoxy(polyethylene glycol)2000]:



(Formula XVI). In some embodiments, the PEGylated diacylglycerol of Formula IX is:



(Formula XVII). In some embodiments, the LNP includes at least one cationic lipid selected from of Formulas III, IV, and V, at least one sterol compound selected from a cholesterol and stigmasterin, and wherein the PEGylated lipid is at least one selected from Formulas XI and XII. In some embodiments, the LNP includes at least one cationic lipid selected from Formulas III, IV, and V, at least one sterol compound selected from a cholesterol and stigmasterin, and wherein the PEGylated lipid is at least one selected from Formulas XIII and XIV. In some embodiments, the LNP includes at least one cationic lipid selected from Formulas III, IV, and V, at least one sterol compound selected from a cholesterol and stigmasterin, and wherein the PEGylated lipid is at least one selected from Formulas XV and XVI. In some embodiments, the LNP includes a cationic lipid of Formula III, a cholesterol as the sterol compound, and wherein the PEGylated lipid is Formula XI.

[0774] In any of the LNP embodiments in the previous paragraph, wherein the content of the cationic lipid composition is between about 65 mole% to 75 mole%, the content of the sterol compound is between about 24 mole% to 34 mole% and the content of the PEGylated lipid is between about 0.5 mole% to 1.5 mole%, wherein the sum of the content of the cationic lipid, of the sterol compound and of the PEGylated lipid for the lipid composition is 100 mole%. In some embodiments, the cationic lipid is about 70 mole%, the content of the sterol compound is about 29 mole% and the content of the PEGylated lipid is about 1 mole%. In some embodiments, the LNP is 70 mole% of Formula III, 29 mole% of cholesterol, and 1 mole% of Formula XI.

Exosomes

[0775] Exosomes are endogenous nano-vesicles that transport RNAs and proteins, and which can deliver RNA to the brain and other target organs. To reduce immunogenicity, Alvarez-Erviti et al. (2011, Nat Biotechnol 29: 341) used self-derived dendritic cells for exosome production. Targeting to the brain was achieved by engineering the dendritic cells to express Lamp2b, an exosomal membrane protein, fused to the neuron-specific RVG peptide. Purified exosomes were loaded with exogenous RNA by electroporation. Intravenously injected RVG-targeted exosomes delivered GAPDH siRNA specifically to neurons, microglia,

oligodendrocytes in the brain, resulting in a specific gene knockdown. Pre-exposure to RVG exosomes did not attenuate knockdown, and non-specific uptake in other tissues was not observed. The therapeutic potential of exosome-mediated siRNA delivery was demonstrated by the strong mRNA (60%) and protein (62%) knockdown of BACE1, a therapeutic target in Alzheimer's disease.

[0776] To obtain a pool of immunologically inert exosomes, Alvarez-Erviti et al. harvested bone marrow from inbred C57BL/6 mice with a homogenous major histocompatibility complex (MHC) haplotype. As immature dendritic cells produce large quantities of exosomes devoid of T-cell activators such as MHC-II and CD86, Alvarez-Erviti et al. selected for dendritic cells with granulocyte/macrophage-colony stimulating factor (GM-CSF) for 7 d. Exosomes were purified from the culture supernatant the following day using well-established ultracentrifugation protocols. The exosomes produced were physically homogenous, with a size distribution peaking at 80 nm in diameter as determined by nanoparticle tracking analysis (NTA) and electron microscopy. Alvarez-Erviti et al. obtained 6-12 μg of exosomes (measured based on protein concentration) per 106 cells.

[0777] Next, Alvarez-Erviti et al. investigated the possibility of loading modified exosomes with exogenous cargoes using electroporation protocols adapted for nanoscale applications. As electroporation for membrane particles at the nanometer scale is not well-characterized, nonspecific Cy5-labeled RNA was used for the empirical optimization of the electroporation protocol. The amount of encapsulated RNA was assayed after ultracentrifugation and lysis of exosomes. Electroporation at 400 V and 125 μF resulted in the greatest retention of RNA and was used for all subsequent experiments.

[0778] Alvarez-Erviti et al. administered 150 μg of each BACE1 siRNA encapsulated in 150 μg of RVG exosomes to normal C57BL/6 mice and compared the knockdown efficiency to four controls: untreated mice, mice injected with RVG exosomes only, mice injected with BACE1 siRNA complexed to an in vivo cationic liposome reagent and mice injected with BACE1 siRNA complexed to RVG-9R, the RVG peptide conjugated to 9 D-arginines that electrostatically binds to the siRNA. Cortical tissue samples were analyzed 3 d after administration and a significant protein knockdown (45%, $P < 0.05$, versus 62%, $P < 0.01$) in both siRNA-RVG-9R-treated and siRNARVG exosome-treated mice was observed, resulting from a significant decrease in BACE1 mRNA levels (66% [+ or -] 15%, $P < 0.001$ and 61% [+ or -] 13% respectively, $P < 0.01$). Moreover, Applicants demonstrated a significant decrease (55%, $P < 0.05$) in the total [beta]-amyloid 1-42 levels, a main component of the amyloid plaques in Alzheimer's pathology, in the RVG-exosome-treated animals. The decrease

observed was greater than the β-amyloid 1-40 decrease demonstrated in normal mice after intraventricular injection of BACE1 inhibitors. Alvarez-Erviti et al. carried out 5'-rapid amplification of cDNA ends (RACE) on BACE1 cleavage product, which provided evidence of RNAi-mediated knockdown by the siRNA.

[0779] Finally, Alvarez-Erviti et al. investigated whether RNA-RVG exosomes induced immune responses in vivo by assessing IL-6, IP-10, TNFα and IFN-α serum concentrations. Following exosome treatment, nonsignificant changes in all cytokines were registered similar to siRNA-transfection reagent treatment in contrast to siRNA-RVG-9R, which potently stimulated IL-6 secretion, confirming the immunologically inert profile of the exosome treatment. Given that exosomes encapsulate only 20% of siRNA, delivery with RVG-exosome appears to be more efficient than RVG-9R delivery as comparable mRNA knockdown and greater protein knockdown was achieved with fivefold less siRNA without the corresponding level of immune stimulation. This experiment demonstrated the therapeutic potential of RVG-exosome technology, which is potentially suited for long-term silencing of genes related to neurodegenerative diseases. The exosome delivery system of Alvarez-Erviti et al. may be applied to deliver the AD-functionalized CRISPR-Cas system of the present invention to therapeutic targets, especially neurodegenerative diseases. A dosage of about 100 to 1000 mg of CRISPR Cas encapsulated in about 100 to 1000 mg of RVG exosomes may be contemplated for the present invention.

[0780] El-Andaloussi et al. (Nature Protocols 7,2112–2126(2012)) discloses how exosomes derived from cultured cells can be harnessed for delivery of RNA in vitro and in vivo. This protocol first describes the generation of targeted exosomes through transfection of an expression vector, comprising an exosomal protein fused with a peptide ligand. Next, El-Andaloussi et al. explain how to purify and characterize exosomes from transfected cell supernatant. Next, El-Andaloussi et al. detail crucial steps for loading RNA into exosomes. Finally, El-Andaloussi et al. outline how to use exosomes to efficiently deliver RNA in vitro and in vivo in mouse brain. Examples of anticipated results in which exosome-mediated RNA delivery is evaluated by functional assays and imaging are also provided. The entire protocol takes ~3 weeks. Delivery or administration according to the invention may be performed using exosomes produced from self-derived dendritic cells. From the herein teachings, this can be employed in the practice of the invention.

[0781] In another embodiment, the plasma exosomes of Wahlgren et al. (Nucleic Acids Research, 2012, Vol. 40, No. 17 e130) are contemplated. Exosomes are nano-sized vesicles

(30–90nm in size) produced by many cell types, including dendritic cells (DC), B cells, T cells, mast cells, epithelial cells and tumor cells. These vesicles are formed by inward budding of late endosomes and are then released to the extracellular environment upon fusion with the plasma membrane. Because exosomes naturally carry RNA between cells, this property may be useful in gene therapy, and from this disclosure can be employed in the practice of the instant invention.

[0782] Exosomes from plasma can be prepared by centrifugation of buffy coat at 900g for 20 min to isolate the plasma followed by harvesting cell supernatants, centrifuging at 300g for 10 min to eliminate cells and at 16 500g for 30 min followed by filtration through a 0.22 mm filter. Exosomes are pelleted by ultracentrifugation at 120 000g for70 min. Chemical transfection of siRNA into exosomes is carried out according to the manufacturer's instructions in RNAi Human/Mouse Starter Kit (Quiagen, Hilden, Germany). siRNA is added to 100 ml PBS at a final concentration of 2 mmol/ml. After adding HiPerFect transfection reagent, the mixture is incubated for 10 min at RT. In order to remove the excess of micelles, the exosomes are re-isolated using aldehyde/sulfate latex beads. The chemical transfection of CRISPR Cas into exosomes may be conducted similarly to siRNA. The exosomes may be co-cultured with monocytes and lymphocytes isolated from the peripheral blood of healthy donors. Therefore, it may be contemplated that exosomes containing CRISPR Cas may be introduced to monocytes and lymphocytes of and autologously reintroduced into a human. Accordingly, delivery or administration according to the invention may be performed using plasma exosomes.

Liposomes

[0783] Delivery or administration according to the invention can be performed with liposomes. Liposomes are spherical vesicle structures composed of a uni- or multilamellar lipid bilayer surrounding internal aqueous compartments and a relatively impermeable outer lipophilic phospholipid bilayer. Liposomes have gained considerable attention as drug delivery carriers because they are biocompatible, nontoxic, can deliver both hydrophilic and lipophilic drug molecules, protect their cargo from degradation by plasma enzymes, and transport their load across biological membranes and the blood brain barrier (BBB) (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review).

[0784] Liposomes can be made from several different types of lipids; however, phospholipids are most commonly used to generate liposomes as drug carriers. Although liposome formation is spontaneous when a lipid film is mixed with an aqueous solution, it can

also be expedited by applying force in the form of shaking by using a homogenizer, sonicator, or an extrusion apparatus (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review).

[0785] Several other additives may be added to liposomes in order to modify their structure and properties. For instance, either cholesterol or sphingomyelin may be added to the liposomal mixture in order to help stabilize the liposomal structure and to prevent the leakage of the liposomal inner cargo. Further, liposomes are prepared from hydrogenated egg phosphatidylcholine or egg phosphatidylcholine, cholesterol, and dicetyl phosphate, and their mean vesicle sizes were adjusted to about 50 and 100 nm. (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review).

[0786] A liposome formulation may be mainly comprised of natural phospholipids and lipids such as 1,2-distearoryl-sn-glycero-3-phosphatidyl choline (DSPC), sphingomyelin, egg phosphatidylcholines and monosialoganglioside. Since this formulation is made up of phospholipids only, liposomal formulations have encountered many challenges, one of the ones being the instability in plasma. Several attempts to overcome these challenges have been made, specifically in the manipulation of the lipid membrane. One of these attempts focused on the manipulation of cholesterol. Addition of cholesterol to conventional formulations reduces rapid release of the encapsulated bioactive compound into the plasma or 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) increases the stability (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review).

[0787] In a particularly advantageous embodiment, Trojan Horse liposomes (also known as Molecular Trojan Horses) are desirable and protocols may be found at http://cshprotocols.cshlp.org/content/2010/4/pdb.prot5407.long. These particles allow delivery of a transgene to the entire brain after an intravascular injection. Without being bound by limitation, it is believed that neutral lipid particles with specific antibodies conjugated to surface allow crossing of the blood brain barrier via endocytosis. Trojan Horse Liposomes may be used to deliver the CRISPR family of nucleases to the brain via an intravascular injection, which would allow whole brain transgenic animals without the need for embryonic manipulation. About 1-5 g of DNA or RNA may be contemplated for in vivo administration in liposomes.

[0788] In another embodiment, the AD-functionalized CRISPR Cas system or components thereof may be administered in liposomes, such as a stable nucleic-acid-lipid particle (SNALP)

(see, e.g., Morrissey et al., Nature Biotechnology, Vol. 23, No. 8, August 2005). Daily intravenous injections of about 1, 3 or 5 mg/kg/day of a specific CRISPR Cas targeted in a SNALP are contemplated. The daily treatment may be over about three days and then weekly for about five weeks. In another embodiment, a specific CRISPR Cas encapsulated SNALP) administered by intravenous injection to at doses of about 1 or 2.5 mg/kg are also contemplated (see, e.g., Zimmerman et al., Nature Letters, Vol. 441, 4 May 2006). The SNALP formulation may contain the lipids 3-N-[(wmethoxypoly(ethylene glycol) 2000) carbamoyl] -1,2-dimyristyloxy-propylamine (PEG-C-DMA), 1,2-dilinoleyloxy-N,N-dimethyl-3-aminopropane (DLinDMA), 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC) and cholesterol, in a 2:40:10:48 molar per cent ratio (see, e.g., Zimmerman et al., Nature Letters, Vol. 441, 4 May 2006).

[0789] In another embodiment, stable nucleic-acid-lipid particles (SNALPs) have proven to be effective delivery molecules to highly vascularized HepG2-derived liver tumors but not in poorly vascularized HCT-116 derived liver tumors (see, e.g., Li, Gene Therapy (2012) 19, 775–780). The SNALP liposomes may be prepared by formulating D-Lin-DMA and PEG-C-DMA with distearoylphosphatidylcholine (DSPC), Cholesterol and siRNA using a 25:1 lipid/siRNA ratio and a 48/40/10/2 molar ratio of Cholesterol/D-Lin-DMA/DSPC/PEG-C-DMA. The resulted SNALP liposomes are about 80–100 nm in size.

[0790] In yet another embodiment, a SNALP may comprise synthetic cholesterol (Sigma-Aldrich, St Louis, MO, USA), dipalmitoylphosphatidylcholine (Avanti Polar Lipids, Alabaster, AL, USA), 3-N-[(w-methoxy poly(ethylene glycol)2000)carbamoyl]-1,2-dimyrestyloxypropylamine, and cationic 1,2-dilinoleyloxy-3-N,Ndimethylaminopropane (see, e.g., Geisbert et al., Lancet 2010; 375: 1896-905). A dosage of about 2 mg/kg total CRISPR Cas per dose administered as, for example, a bolus intravenous infusion may be contemplated.

[0791] In yet another embodiment, a SNALP may comprise synthetic cholesterol (Sigma-Aldrich), 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC; Avanti Polar Lipids Inc.), PEG-cDMA, and 1,2-dilinoleyloxy-3-(N;N-dimethyl)aminopropane (DLinDMA) (see, e.g., Judge, J. Clin. Invest. 119:661-673 (2009)). Formulations used for in vivo studies may comprise a final lipid/RNA mass ratio of about 9:1.

[0792] The safety profile of RNAi nanomedicines has been reviewed by Barros and Gollob of Alnylam Pharmaceuticals (see, e.g., Advanced Drug Delivery Reviews 64 (2012) 1730–1737). The stable nucleic acid lipid particle (SNALP) is comprised of four different lipids — an ionizable lipid (DLinDMA) that is cationic at low pH, a neutral helper lipid, cholesterol, and a diffusible polyethylene glycol (PEG)-lipid. The particle is approximately 80 nm in

diameter and is charge-neutral at physiologic pH. During formulation, the ionizable lipid serves to condense lipid with the anionic RNA during particle formation. When positively charged under increasingly acidic endosomal conditions, the ionizable lipid also mediates the fusion of SNALP with the endosomal membrane enabling release of RNA into the cytoplasm. The PEG-lipid stabilizes the particle and reduces aggregation during formulation, and subsequently provides a neutral hydrophilic exterior that improves pharmacokinetic properties.

[0793] To date, two clinical programs have been initiated using SNALP formulations with RNA. Tekmira Pharmaceuticals recently completed a phase I single-dose study of SNALP-ApoB in adult volunteers with elevated LDL cholesterol. ApoB is predominantly expressed in the liver and jejunum and is essential for the assembly and secretion of VLDL and LDL. Seventeen subjects received a single dose of SNALP-ApoB (dose escalation across 7 dose levels). There was no evidence of liver toxicity (anticipated as the potential dose-limiting toxicity based on preclinical studies). One (of two) subjects at the highest dose experienced flu-like symptoms consistent with immune system stimulation, and the decision was made to conclude the trial.

[0794] Alnylam Pharmaceuticals has similarly advanced ALN-TTR01, which employs the SNALP technology described above and targets hepatocyte production of both mutant and wild-type TTR to treat TTR amyloidosis (ATTR). Three ATTR syndromes have been described: familial amyloidotic polyneuropathy (FAP) and familial amyloidotic cardiomyopathy (FAC) — both caused by autosomal dominant mutations in TTR; and senile systemic amyloidosis (SSA) cause by wildtype TTR. A placebo-controlled, single dose-escalation phase I trial of ALN-TTR01 was recently completed in patients with ATTR. ALN-TTR01 was administered as a 15-minute IV infusion to 31 patients (23 with study drug and 8 with placebo) within a dose range of 0.01 to 1.0 mg/kg (based on siRNA). Treatment was well tolerated with no significant increases in liver function tests. Infusion-related reactions were noted in 3 of 23 patients at≥0.4 mg/kg; all responded to slowing of the infusion rate and all continued on study. Minimal and transient elevations of serum cytokines IL-6, IP-10 and IL-1ra were noted in two patients at the highest dose of 1 mg/kg (as anticipated from preclinical and NHP studies). Lowering of serum TTR, the expected pharmacodynamics effect of ALN-TTR01, was observed at 1 mg/kg.

[0795] In yet another embodiment, a SNALP may be made by solubilizing a cationic lipid, DSPC, cholesterol and PEG-lipid e.g., in ethanol, e.g., at a molar ratio of 40:10:40:10, respectively (see, Semple et al., Nature Niotechnology, Volume 28 Number 2 February 2010, pp. 172-177). The lipid mixture was added to an aqueous buffer (50 mM citrate, pH 4) with

mixing to a final ethanol and lipid concentration of 30% (vol/vol) and 6.1 mg/ml, respectively, and allowed to equilibrate at 22 °C for 2 min before extrusion. The hydrated lipids were extruded through two stacked 80 nm pore-sized filters (Nuclepore) at 22 °C using a Lipex Extruder (Northern Lipids) until a vesicle diameter of 70–90 nm, as determined by dynamic light scattering analysis, was obtained. This generally required 1–3 passes. The siRNA (solubilized in a 50 mM citrate, pH 4 aqueous solution containing 30% ethanol) was added to the pre-equilibrated (35 °C) vesicles at a rate of ~5 ml/min with mixing. After a final target siRNA/lipid ratio of 0.06 (wt/wt) was reached, the mixture was incubated for a further 30 min at 35 °C to allow vesicle reorganization and encapsulation of the siRNA. The ethanol was then removed and the external buffer replaced with PBS (155 mM NaCl, 3 mM Na2HPO4, 1 mM KH2PO4, pH 7.5) by either dialysis or tangential flow diafiltration. siRNA were encapsulated in SNALP using a controlled step-wise dilution method process. The lipid constituents of KC2-SNALP were DLin-KC2-DMA (cationic lipid), dipalmitoylphosphatidylcholine (DPPC; Avanti Polar Lipids), synthetic cholesterol (Sigma) and PEG-C-DMA used at a molar ratio of 57.1:7.1:34.3:1.4. Upon formation of the loaded particles, SNALP were dialyzed against PBS and filter sterilized through a 0.2 μm filter before use. Mean particle sizes were 75–85 nm and 90–95% of the siRNA was encapsulated within the lipid particles. The final siRNA/lipid ratio in formulations used for in vivo testing was ~0.15 (wt/wt). LNP-siRNA systems containing Factor VII siRNA were diluted to the appropriate concentrations in sterile PBS immediately before use and the formulations were administered intravenously through the lateral tail vein in a total volume of 10 ml/kg. This method and these delivery systems may be extrapolated to the AD-functionalized CRISPR Cas system of the present invention.

Other Lipids

[0796] Other cationic lipids, such as amino lipid 2,2-dilinoleyl-4-dimethylaminoethyl-[1,3]-dioxolane (DLin-KC2-DMA) may be utilized to encapsulate CRISPR Cas or components thereof or nucleic acid molecule(s) coding therefor e.g., similar to SiRNA (see, e.g., Jayaraman, Angew. Chem. Int. Ed. 2012, 51, 8529 –8533), and hence may be employed in the practice of the invention. A preformed vesicle with the following lipid composition may be contemplated: amino lipid, distearoylphosphatidylcholine (DSPC), cholesterol and (R)-2,3-bis(octadecyloxy) propyl-1-(methoxy poly(ethylene glycol)2000)propylcarbamate (PEG-lipid) in the molar ratio 40/10/40/10, respectively, and a FVII siRNA/total lipid ratio of approximately 0.05 (w/w). To ensure a narrow particle size distribution in the range of 70–90 nm and a low polydispersity index of 0.11+0.04 (n=56), the particles may be extruded up to three times through 80 nm membranes prior to adding the guide RNA. Particles containing the highly potent amino lipid

16 may be used, in which the molar ratio of the four lipid components 16, DSPC, cholesterol and PEG-lipid (50/10/38.5/1.5) which may be further optimized to enhance in vivo activity.

[0797] Michael S D Kormann et al. ("Expression of therapeutic proteins after delivery of chemically modified mRNA in mice: Nature Biotechnology, Volume:29, Pages: 154–157 (2011)) describes the use of lipid envelopes to deliver RNA. Use of lipid envelopes is also preferred in the present invention.

[0798] In another embodiment, lipids may be formulated with the AD-functionalized CRISPR Cas system of the present invention or component(s) thereof or nucleic acid molecule(s) coding therefor to form lipid nanoparticles (LNPs). Lipids include, but are not limited to, DLin-KC2-DMA4, C12-200 and colipids disteroylphosphatidyl choline, cholesterol, and PEG-DMG may be formulated with CRISPR Cas instead of siRNA (see, e.g., Novobrantseva, Molecular Therapy–Nucleic Acids (2012) 1, e4; doi:10.1038/mtna.2011.3) using a spontaneous vesicle formation procedure. The component molar ratio may be about 50/10/38.5/1.5 (DLin-KC2-DMA or C12-200/disteroylphosphatidyl choline/cholesterol/PEG-DMG). The final lipid:siRNA weight ratio may be ~12:1 and 9:1 in the case of DLin-KC2-DMA and C12-200 lipid nanoparticles (LNPs), respectively. The formulations may have mean particle diameters of ~80 nm with >90% entrapment efficiency. A 3 mg/kg dose may be contemplated.

[0799] Tekmira has a portfolio of approximately 95 patent families, in the U.S. and abroad, that are directed to various aspects of LNPs and LNP formulations (see, e.g., U.S. Pat. Nos. 7,982,027; 7,799,565; 8,058,069; 8,283,333; 7,901,708; 7,745,651; 7,803,397; 8,101,741; 8,188,263; 7,915,399; 8,236,943 and 7,838,658 and European Pat. Nos 1766035; 1519714; 1781593 and 1664316), all of which may be used and/or adapted to the present invention.

[0800] The AD-functionalized CRISPR Cas system or components thereof or nucleic acid molecule(s) coding therefor may be delivered encapsulated in PLGA Microspheres such as that further described in US published applications 20130252281 and 20130245107 and 20130244279 (assigned to Moderna Therapeutics) which relate to aspects of formulation of compositions comprising modified nucleic acid molecules which may encode a protein, a protein precursor, or a partially or fully processed form of the protein or a protein precursor. The formulation may have a molar ratio 50:10:38.5:1.5-3.0 (cationic lipid:fusogenic lipid:cholesterol:PEG lipid). The PEG lipid may be selected from, but is not limited to PEG-c-DOMG, PEG-DMG. The fusogenic lipid may be DSPC. See also, Schrum et al., Delivery and Formulation of Engineered Nucleic Acids, US published application 20120251618.

[0801] Nanomerics' technology addresses bioavailability challenges for a broad range of therapeutics, including low molecular weight hydrophobic drugs, peptides, and nucleic acid based therapeutics (plasmid, siRNA, miRNA). Specific administration routes for which the technology has demonstrated clear advantages include the oral route, transport across the blood-brain-barrier, delivery to solid tumours, as well as to the eye. See, e.g., Mazza et al., 2013, ACS Nano. 2013 Feb 26;7(2):1016-26; Uchegbu and Siew, 2013, J Pharm Sci. 102(2):305-10 and Lalatsa et al., 2012, J Control Release. 2012 Jul 20; 161(2):523-36.

[0802] US Patent Publication No. 20050019923 describes cationic dendrimers for delivering bioactive molecules, such as polynucleotide molecules, peptides and polypeptides and/or pharmaceutical agents, to a mammalian body. The dendrimers are suitable for targeting the delivery of the bioactive molecules to, for example, the liver, spleen, lung, kidney or heart (or even the brain). Dendrimers are synthetic 3-dimensional macromolecules that are prepared in a step-wise fashion from simple branched monomer units, the nature and functionality of which can be easily controlled and varied. Dendrimers are synthesised from the repeated addition of building blocks to a multifunctional core (divergent approach to synthesis), or towards a multifunctional core (convergent approach to synthesis) and each addition of a 3-dimensional shell of building blocks leads to the formation of a higher generation of the dendrimers. Polypropylenimine dendrimers start from a diaminobutane core to which is added twice the number of amino groups by a double Michael addition of acrylonitrile to the primary amines followed by the hydrogenation of the nitriles. This results in a doubling of the amino groups. Polypropylenimine dendrimers contain 100% protonable nitrogens and up to 64 terminal amino groups (generation 5, DAB 64). Protonable groups are usually amine groups which are able to accept protons at neutral pH. The use of dendrimers as gene delivery agents has largely focused on the use of the polyamidoamine. and phosphorous containing compounds with a mixture of amine/amide or N--P(O2)S as the conjugating units respectively with no work being reported on the use of the lower generation polypropylenimine dendrimers for gene delivery. Polypropylenimine dendrimers have also been studied as pH sensitive controlled release systems for drug delivery and for their encapsulation of guest molecules when chemically modified by peripheral amino acid groups. The cytotoxicity and interaction of polypropylenimine dendrimers with DNA as well as the transfection efficacy of DAB 64 has also been studied.

[0803] US Patent Publication No. 20050019923 is based upon the observation that, contrary to earlier reports, cationic dendrimers, such as polypropylenimine dendrimers, display suitable properties, such as specific targeting and low toxicity, for use in the targeted delivery

of bioactive molecules, such as genetic material. In addition, derivatives of the cationic dendrimer also display suitable properties for the targeted delivery of bioactive molecules. See also, Bioactive Polymers, US published application 20080267903, which discloses "Various polymers, including cationic polyamine polymers and dendrimeric polymers, are shown to possess anti-proliferative activity, and may therefore be useful for treatment of disorders characterised by undesirable cellular proliferation such as neoplasms and tumours, inflammatory disorders (including autoimmune disorders), psoriasis and atherosclerosis. The polymers may be used alone as active agents, or as delivery vehicles for other therapeutic agents, such as drug molecules or nucleic acids for gene therapy. In such cases, the polymers' own intrinsic anti-tumour activity may complement the activity of the agent to be delivered." The disclosures of these patent publications may be employed in conjunction with herein teachings for delivery of AD-functionalized CRISPR Cas system(s) or component(s) thereof or nucleic acid molecule(s) coding therefor.

Supercharged proteins

[0804] Supercharged proteins are a class of engineered or naturally occurring proteins with unusually high positive or negative net theoretical charge and may be employed in delivery of AD-functionalized CRISPR Cas system(s) or component(s) thereof or nucleic acid molecule(s) coding therefor. Both supernegatively and superpositively charged proteins exhibit a remarkable ability to withstand thermally or chemically induced aggregation. Superpositively charged proteins are also able to penetrate mammalian cells. Associating cargo with these proteins, such as plasmid DNA, RNA, or other proteins, can enable the functional delivery of these macromolecules into mammalian cells both in vitro and in vivo. The creation and characterization of supercharged proteins has been reported in 2007 (Lawrence et al., 2007, Journal of the American Chemical Society 129, 10110–10112).

[0805] The nonviral delivery of RNA and plasmid DNA into mammalian cells are valuable both for research and therapeutic applications (Akinc et al., 2010, Nat. Biotech. 26, 561–569). Purified +36 GFP protein (or other superpositively charged protein) is mixed with RNAs in the appropriate serum-free media and allowed to complex prior addition to cells. Inclusion of serum at this stage inhibits formation of the supercharged protein-RNA complexes and reduces the effectiveness of the treatment. The following protocol has been found to be effective for a variety of cell lines (McNaughton et al., 2009, Proc. Natl. Acad. Sci. USA 106, 6111–6116) (However, pilot experiments varying the dose of protein and RNA should be performed to optimize the procedure for specific cell lines): (1) One day before treatment, plate 1 x 105 cells per well in a 48-well plate. (2) On the day of treatment, dilute purified +36 GFP protein in

serumfree media to a final concentration 200nM. Add RNA to a final concentration of 50nM. Vortex to mix and incubate at room temperature for 10min. (3) During incubation, aspirate media from cells and wash once with PBS. (4) Following incubation of +36 GFP and RNA, add the protein-RNA complexes to cells. (5) Incubate cells with complexes at 37 °C for 4h. (6) Following incubation, aspirate the media and wash three times with 20 U/mL heparin PBS. Incubate cells with serum-containing media for a further 48h or longer depending upon the assay for activity. (7) Analyze cells by immunoblot, qPCR, phenotypic assay, or other appropriate method.

[0806] It has been further found +36 GFP to be an effective plasmid delivery reagent in a range of cells. As plasmid DNA is a larger cargo than siRNA, proportionately more +36 GFP protein is required to effectively complex plasmids. For effective plasmid delivery Applicants have developed a variant of +36 GFP bearing a C-terminal HA2 peptide tag, a known endosome-disrupting peptide derived from the influenza virus hemagglutinin protein. The following protocol has been effective in a variety of cells, but as above it is advised that plasmid DNA and supercharged protein doses be optimized for specific cell lines and delivery applications: (1) One day before treatment, plate 1 x 105 per well in a 48-well plate. (2) On the day of treatment, dilute purified þ36 GFP protein in serumfree media to a final concentration 2 mM. Add 1mg of plasmid DNA. Vortex to mix and incubate at room temperature for 10min. (3) During incubation, aspirate media from cells and wash once with PBS. (4) Following incubation of þ36 GFP and plasmid DNA, gently add the protein-DNA complexes to cells. (5) Incubate cells with complexes at 37 C for 4h. (6) Following incubation, aspirate the media and wash with PBS. Incubate cells in serum-containing media and incubate for a further 24–48h. (7) Analyze plasmid delivery (e.g., by plasmid-driven gene expression) as appropriate.

[0807] See also, e.g., McNaughton et al., Proc. Natl. Acad. Sci. USA 106, 6111-6116 (2009); Cronican et al., ACS Chemical Biology 5, 747-752 (2010); Cronican et al., Chemistry & Biology 18, 833-838 (2011); Thompson et al., Methods in Enzymology 503, 293-319 (2012); Thompson, D.B., et al., Chemistry & Biology 19 (7), 831-843 (2012). The methods of the super charged proteins may be used and/or adapted for delivery of the AD-functionalized CRISPR Cas system of the present invention. These systems in conjunction with herein teaching can be employed in the delivery of AD-functionalized CRISPR Cas system(s) or component(s) thereof or nucleic acid molecule(s) coding therefor

Cell Penetrating Peptides (CPPs)

[0808] In yet another embodiment, cell penetrating peptides (CPPs) are contemplated for the delivery of the AD-functionalized CRISPR Cas system. CPPs are short peptides that

facilitate cellular uptake of various molecular cargo (from nanosize particles to small chemical molecules and large fragments of DNA). The term "cargo" as used herein includes but is not limited to the group consisting of therapeutic agents, diagnostic probes, peptides, nucleic acids, antisense oligonucleotides, plasmids, proteins, particles, including nanoparticles, liposomes, chromophores, small molecules and radioactive materials. In aspects of the invention, the cargo may also comprise any component of the AD-functionalized CRISPR Cas system or the entire AD-functionalized functional CRISPR Cas system. Aspects of the present invention further provide methods for delivering a desired cargo into a subject comprising: (a) preparing a complex comprising the cell penetrating peptide of the present invention and a desired cargo, and (b) orally, intraarticularly, intraperitoneally, intrathecally, intrarterially, intranasally, intraparenchymally, subcutaneously, intramuscularly, intravenously, dermally, intrarectally, or topically administering the complex to a subject. The cargo is associated with the peptides either through chemical linkage via covalent bonds or through non-covalent interactions.

[0809] The function of the CPPs are to deliver the cargo into cells, a process that commonly occurs through endocytosis with the cargo delivered to the endosomes of living mammalian cells. Cell-penetrating peptides are of different sizes, amino acid sequences, and charges but all CPPs have one distinct characteristic, which is the ability to translocate the plasma membrane and facilitate the delivery of various molecular cargoes to the cytoplasm or an organelle. CPP translocation may be classified into three main entry mechanisms: direct penetration in the membrane, endocytosis-mediated entry, and translocation through the formation of a transitory structure. CPPs have found numerous applications in medicine as drug delivery agents in the treatment of different diseases including cancer and virus inhibitors, as well as contrast agents for cell labeling. Examples of the latter include acting as a carrier for GFP, MRI contrast agents, or quantum dots. CPPs hold great potential as in vitro and in vivo delivery vectors for use in research and medicine. CPPs typically have an amino acid composition that either contains a high relative abundance of positively charged amino acids such as lysine or arginine or has sequences that contain an alternating pattern of polar/charged amino acids and non-polar, hydrophobic amino acids. These two types of structures are referred to as polycationic or amphipathic, respectively. A third class of CPPs are the hydrophobic peptides, containing only apolar residues, with low net charge or have hydrophobic amino acid groups that are crucial for cellular uptake. One of the initial CPPs discovered was the trans-activating transcriptional activator (Tat) from Human Immunodeficiency Virus 1 (HIV-1) which was found to be efficiently taken up from the surrounding media by numerous cell types in culture. Since then, the number of known CPPs has expanded considerably and small

molecule synthetic analogues with more effective protein transduction properties have been generated. CPPs include but are not limited to Penetratin, Tat (48-60), Transportan, and (R-AhX-R4) (Ahx=aminohexanoyl).

[0810] US Patent 8,372,951, provides a CPP derived from eosinophil cationic protein (ECP) which exhibits highly cell-penetrating efficiency and low toxicity. Aspects of delivering the CPP with its cargo into a vertebrate subject are also provided. Further aspects of CPPs and their delivery are described in U. S. patents 8,575,305; 8;614,194 and 8,044,019. CPPs can be used to deliver the AD-functionalized CRISPR-Cas system or components thereof. That CPPs can be employed to deliver the AD-functionalized CRISPR-Cas system or components thereof is also provided in the manuscript "Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA", by Suresh Ramakrishna, Abu-Bonsrah Kwaku Dad, Jagadish Beloor, et al. Genome Res. 2014 Apr 2, incorporated by reference in its entirety, wherein it is demonstrated that treatment with CPP-conjugated recombinant Cas9 protein and CPP-complexed guide RNAs lead to endogenous gene disruptions in human cell lines. In the paper the Cas9 protein was conjugated to CPP via a thioether bond, whereas the guide RNA was complexed with CPP, forming condensed, positively charged particles. It was shown that simultaneous and sequential treatment of human cells, including embryonic stem cells, dermal fibroblasts, HEK293T cells, HeLa cells, and embryonic carcinoma cells, with the modified Cas9 and guide RNA led to efficient gene disruptions with reduced off-target mutations relative to plasmid transfections.

Aerosol delivery

[0811] Subjects treated for a lung disease may for example receive pharmaceutically effective amount of aerosolized AAV vector system per lung endobronchially delivered while spontaneously breathing. As such, aerosolized delivery is preferred for AAV delivery in general. An adenovirus or an AAV particle may be used for delivery. Suitable gene constructs, each operably linked to one or more regulatory sequences, may be cloned into the delivery vector.

Packaging and Promoters

[0812] The promoter used to drive CRISPR-Cas protein and adenosine deaminase coding nucleic acid molecule expression can include AAV ITR, which can serve as a promoter. This is advantageous for eliminating the need for an additional promoter element (which can take up space in the vector). The additional space freed up can be used to drive the expression of additional elements (gRNA, etc.). Also, ITR activity is relatively weaker, so can be used to reduce potential toxicity due to over expression of Cas13.

[0813] For ubiquitous expression, promoters that can be used include: CMV, CAG, CBh, PGK, SV40, Ferritin heavy or light chains, etc. For brain or other CNS expression, SynapsinI can be used for all neurons, CaMKIIalpha can be used for excitatory neurons, GAD67 or GAD65 or VGAT can be used for GABAergic neurons. For liver expression, Albumin promoter can be used. For lung expression, SP-B can be used. For endothelial cells, ICAM can be used. For hematopoietic cells, IFNbeta or CD45 can be used. For Osteoblasts, the OG-2 can be used.

[0814] The promoter used to drive guide RNA can include Pol III promoters such as U6 or H1, as well as use of Pol II promoter and intronic cassettes to express guide RNA.

Adeno associated virus (AAV)

[0815] The targeting domain, adenosine deaminase, and one or more guide RNA can be delivered using adeno associated virus (AAV), lentivirus, adenovirus or other plasmid or viral vector types, in particular, using formulations and doses from, for example, US Patents Nos. 8,454,972 (formulations, doses for adenovirus), 8,404,658 (formulations, doses for AAV) and 5,846,946 (formulations, doses for DNA plasmids) and from clinical trials and publications regarding the clinical trials involving lentivirus, AAV and adenovirus. For examples, for AAV, the route of administration, formulation and dose can be as in US Patent No. 8,454,972 and as in clinical trials involving AAV. For Adenovirus, the route of administration, formulation and dose can be as in US Patent No. 8,404,658 and as in clinical trials involving adenovirus. For plasmid delivery, the route of administration, formulation and dose can be as in US Patent No 5,846,946 and as in clinical studies involving plasmids. Doses may be based on or extrapolated to an average 70 kg individual (e.g. a male adult human), and can be adjusted for patients, subjects, mammals of different weight and species. Frequency of administration is within the ambit of the medical or veterinary practitioner (e.g., physician, veterinarian), depending on usual factors including the age, sex, general health, other conditions of the patient or subject and the particular condition or symptoms being addressed. The viral vectors can be injected into the tissue of interest. For cell-type specific genome modification, the expression of Cas13 and adenosine deaminase can be driven by a cell-type specific promoter. For example, liver-specific expression might use the Albumin promoter and neuron-specific expression (e.g. for targeting CNS disorders) might use the Synapsin I promoter.

[0816] In terms of in vivo delivery, AAV is advantageous over other viral vectors for a couple of reasons: low toxicity (this may be due to the purification method not requiring ultra centrifugation of cell particles that can activate the immune response); and low probability of causing insertional mutagenesis because it doesn't integrate into the host genome.

[0817] AAV has a packaging limit of 4.5 or 4.75 Kb. This means that Cas13 as well as a promoter and transcription terminator have to be all fit into the same viral vector. Constructs larger than 4.5 or 4.75 Kb will lead to significantly reduced virus production. SpCas9 is quite large, the gene itself is over 4.1 Kb, which makes it difficult for packing into AAV. Therefore embodiments of the invention include utilizing homologs of Cas13 that are shorter.

[0818] As to AAV, the AAV can be AAV1, AAV2, AAV5 or any combination thereof. One can select the AAV of the AAV with regard to the cells to be targeted; e.g., one can select AAV serotypes 1, 2, 5 or a hybrid capsid AAV1, AAV2, AAV5 or any combination thereof for targeting brain or neuronal cells; and one can select AAV4 for targeting cardiac tissue. AAV8 is useful for delivery to the liver. The herein promoters and vectors are preferred individually. A tabulation of certain AAV serotypes as to these cells (see Grimm, D. et al, J. Virol. 82: 5887-5911 (2008)) is as follows:

Cell Line	AAV-1	AAV-2	AAV-3	AAV-4	AAV-5	AAV-6	AAV-8	AAV-9
Huh-7	13	100	2.5	0.0	0.1	10	0.7	0.0
HEK293	25	100	2.5	0.1	0.1	5	0.7	0.1
HeLa	3	100	2.0	0.1	6.7	1	0.2	0.1
HepG2	3	100	16.7	0.3	1.7	5	0.3	ND
Hep1A	20	100	0.2	1.0	0.1	1	0.2	0.0
911	17	100	11	0.2	0.1	17	0.1	ND
CHO	100	100	14	1.4	333	50	10	1.0
COS	33	100	33	3.3	5.0	14	2.0	0.5
MeWo	10	100	20	0.3	6.7	10	1.0	0.2
NIH3T3	10	100	2.9	2.9	0.3	10	0.3	ND
A549	14	100	20	ND	0.5	10	0.5	0.1
HT1180	20	100	10	0.1	0.3	33	0.5	0.1
Monocytes	1111	100	ND	ND	125	1429	ND	ND
Immature DC	2500	100	ND	ND	222	2857	ND	ND
Mature DC	2222	100	ND	ND	333	3333	ND	ND

Lentiviruses

[0819] Lentiviruses are complex retroviruses that have the ability to infect and express their genes in both mitotic and post-mitotic cells. The most commonly known lentivirus is the human immunodeficiency virus (HIV), which uses the envelope glycoproteins of other viruses to target a broad range of cell types.

[0820] Lentiviruses may be prepared as follows. After cloning pCasES10 (which contains a lentiviral transfer plasmid backbone), HEK293FT at low passage (p=5) were seeded in a T-75 flask to 50% confluence the day before transfection in DMEM with 10% fetal bovine serum

and without antibiotics. After 20 hours, media was changed to OptiMEM (serum-free) media and transfection was done 4 hours later. Cells were transfected with 10 μg of lentiviral transfer plasmid (pCasES10) and the following packaging plasmids: 5 μg of pMD2.G (VSV-g pseudotype), and 7.5ug of psPAX2 (gag/pol/rev/tat). Transfection was done in 4mL OptiMEM with a cationic lipid delivery agent (50uL Lipofectamine 2000 and 100ul Plus reagent). After 6 hours, the media was changed to antibiotic-free DMEM with 10% fetal bovine serum. These methods use serum during cell culture, but serum-free methods are preferred.

[0821] Lentivirus may be purified as follows. Viral supernatants were harvested after 48 hours. Supernatants were first cleared of debris and filtered through a 0.45um low protein binding (PVDF) filter. They were then spun in a ultracentrifuge for 2 hours at 24,000 rpm. Viral pellets were resuspended in 50ul of DMEM overnight at 4C. They were then aliquotted and immediately frozen at -80°C.

[0822] In another embodiment, minimal non-primate lentiviral vectors based on the equine infectious anemia virus (EIAV) are also contemplated, especially for ocular gene therapy (see, e.g., Balagaan, J Gene Med 2006; 8: 275 – 285). In another embodiment, RetinoStat®, an equine infectious anemia virus-based lentiviral gene therapy vector that expresses angiostatic proteins endostatin and angiostatin that is delivered via a subretinal injection for the treatment of the web form of age-related macular degeneration is also contemplated (see, e.g., Binley et al., HUMAN GENE THERAPY 23:980–991 (September 2012)) and this vector may be modified for the AD-functionalized CRISPR-Cas system of the present invention.

[0823] In another embodiment, self-inactivating lentiviral vectors with an siRNA targeting a common exon shared by HIV tat/rev, a nucleolar-localizing TAR decoy, and an anti–CCR5-specific hammerhead ribozyme (see, e.g., DiGiusto et al. (2010) Sci Transl Med 2:36ra43) may be used/and or adapted to the AD-functionalized CRISPR-Cas system of the present invention. A minimum of 2.5 × 106 CD34+ cells per kilogram patient weight may be collected and prestimulated for 16 to 20 hours in X-VIVO 15 medium (Lonza) containing 2 μmol/L-glutamine, stem cell factor (100 ng/ml), Flt-3 ligand (Flt-3L) (100 ng/ml), and thrombopoietin (10 ng/ml) (CellGenix) at a density of 2 × 106 cells/ml. Prestimulated cells may be transduced with lentiviral at a multiplicity of infection of 5 for 16 to 24 hours in 75-cm2 tissue culture flasks coated with fibronectin (25 mg/cm2) (RetroNectin,Takara Bio Inc.).

[0824] Lentiviral vectors have been disclosed as in the treatment for Parkinson's Disease, see, e.g., US Patent Publication No. 20120295960 and US Patent Nos. 7303910 and 7351585. Lentiviral vectors have also been disclosed for the treatment of ocular diseases, see e.g., US Patent Publication Nos. 20060281180, 20090007284, US20110117189; US20090017543;

US20070054961, US20100317109. Lentiviral vectors have also been disclosed for delivery to the brain, see, e.g., US Patent Publication Nos. US20110293571; US20110293571, US20040013648, US20070025970, US20090111106 and US Patent No. US7259015.

APPLICATION IN NON-ANIMAL ORGANISMS

[0825] The AD-functionalized CRISPR system(s) (e.g., single or multiplexed) can be used in conjunction with recent advances in crop genomics. The systems described herein can be used to perform efficient and cost effective plant gene or genome interrogation or editing or manipulation—for instance, for rapid investigation and/or selection and/or interrogations and/or comparison and/or manipulations and/or transformation of plant genes or genomes; e.g., to create, identify, develop, optimize, or confer trait(s) or characteristic(s) to plant(s) or to transform a plant genome. There can accordingly be improved production of plants, new plants with new combinations of traits or characteristics or new plants with enhanced traits. The AD-functionalized CRISPR system can be used with regard to plants in Site-Directed Integration (SDI) or Gene Editing (GE) or any Near Reverse Breeding (NRB) or Reverse Breeding (RB) techniques. Aspects of utilizing the herein described Cas13 effector protein systems may be analogous to the use of the CRISPR-Cas (e.g. CRISPR-Cas9) system in plants, and mention is made of the University of Arizona website "CRISPR-PLANT" (http://www.genome.arizona.edu/crispr/) (supported by Penn State and AGI). Emodiments of the invention can be used in genome editing in plants or where RNAi or similar genome editing techniques have been used previously; see, e.g., Nekrasov, "Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR-Cas system," Plant Methods 2013, 9:39 (doi:10.1186/1746-4811-9-39); Brooks, "Efficient gene editing in tomato in the first generation using the CRISPR-Cas9 system," Plant Physiology September 2014 pp 114.247577; Shan, "Targeted genome modification of crop plants using a CRISPR-Cas system," Nature Biotechnology 31, 686-688 (2013); Feng, "Efficient genome editing in plants using a CRISPR-Cas system," Cell Research (2013) 23:1229–1232. doi:10.1038/cr.2013.114; published online 20 August 2013; Xie, "RNA-guided genome editing in plants using a CRISPR-Cas system," Mol Plant. 2013 Nov;6(6):1975-83. doi: 10.1093/mp/sst119. Epub 2013 Aug 17; Xu, "Gene targeting using the Agrobacterium tumefaciens-mediated CRISPR-Cas system in rice," Rice 2014, 7:5 (2014), Zhou et al., "Exploiting SNPs for biallelic CRISPR mutations in the outcrossing woody perennial Populus reveals 4-coumarate: CoA ligase specificity and Redundancy," New Phytologist (2015) (Forum) 1-4 (available online only at www.newphytologist.com); Caliando et al, "Targeted DNA degradation using a CRISPR device stably carried in the host genome, NATURE COMMUNICATIONS 6:6989, DOI:

10.1038/ncomms7989, www.nature.com/naturecommunications DOI: 10.1038/ncomms7989; US Patent No. 6,603,061 - Agrobacterium-Mediated Plant Transformation Method; US Patent No. 7,868,149 - Plant Genome Sequences and Uses Thereof and US 2009/0100536 - Transgenic Plants with Enhanced Agronomic Traits, all the contents and disclosure of each of which are herein incorporated by reference in their entirety. In the practice of the invention, the contents and disclosure of Morrell et al "Crop genomics: advances and applications," Nat Rev Genet. 2011 Dec 29;13(2):85-96; each of which is incorporated by reference herein including as to how herein embodiments may be used as to plants. Accordingly, reference herein to animal cells may also apply, mutatis mutandis, to plant cells unless otherwise apparent; and, the enzymes herein having reduced off-target effects and systems employing such enzymes can be used in plant applciations, including those mentioned herein.

Application of Site Directed Base Editing to plants and yeast

[0826] In general, the term "plant" relates to any various photosynthetic, eukaryotic, unicellular or multicellular organism of the kingdom Plantae characteristically growing by cell division, containing chloroplasts, and having cell walls comprised of cellulose. The term plant encompasses monocotyledonous and dicotyledonous plants. Specifically, the plants are intended to comprise without limitation angiosperm and gymnosperm plants such as acacia, alfalfa, amaranth, apple, apricot, artichoke, ash tree, asparagus, avocado, banana, barley, beans, beet, birch, beech, blackberry, blueberry, broccoli, Brussel's sprouts, cabbage, canola, cantaloupe, carrot, cassava, cauliflower, cedar, a cereal, celery, chestnut, cherry, Chinese cabbage, citrus, clementine, clover, coffee, corn, cotton, cowpea, cucumber, cypress, eggplant, elm, endive, eucalyptus, fennel, figs, fir, geranium, grape, grapefruit, groundnuts, ground cherry, gum hemlock, hickory, kale, kiwifruit, kohlrabi, larch, lettuce, leek, lemon, lime, locust, pine, maidenhair, maize, mango, maple, melon, millet, mushroom, mustard, nuts, oak, oats, oil palm, okra, onion, orange, an ornamental plant or flower or tree, papaya, palm, parsley, parsnip, pea, peach, peanut, pear, peat, pepper, persimmon, pigeon pea, pine, pineapple, plantain, plum, pomegranate, potato, pumpkin, radicchio, radish, rapeseed, raspberry, rice, rye, sorghum, safflower, sallow, soybean, spinach, spruce, squash, strawberry, sugar beet, sugarcane, sunflower, sweet potato, sweet corn, tangerine, tea, tobacco, tomato, trees, triticale, turf grasses, turnips, vine, walnut, watercress, watermelon, wheat, yams, yew, and zucchini. The term plant also encompasses Algae, which are mainly photoautotrophs unified primarily by their lack of roots, leaves and other organs that characterize higher plants.

[0827] The methods for genome editing using the AD-functionalized CRISPR system as described herein can be used to confer desired traits on essentially any plant. A wide variety of

plants and plant cell systems may be engineered for the desired physiological and agronomic characteristics described herein using the nucleic acid constructs of the present disclosure and the various transformation methods mentioned above. In preferred embodiments, target plants and plant cells for engineering include, but are not limited to, those monocotyledonous and dicotyledonous plants, such as crops including grain crops (e.g., wheat, maize, rice, millet, barley), fruit crops (e.g., tomato, apple, pear, strawberry, orange), forage crops (e.g., alfalfa), root vegetable crops (e.g., carrot, potato, sugar beets, yam), leafy vegetable crops (e.g., lettuce, spinach); flowering plants (e.g., petunia, rose, chrysanthemum), conifers and pine trees (e.g., pine fir, spruce); plants used in phytoremediation (e.g., heavy metal accumulating plants); oil crops (e.g., sunflower, rape seed) and plants used for experimental purposes (e.g., Arabidopsis). Thus, the methods and systems can be used over a broad range of plants, such as for example with dicotyledonous plants belonging to the orders Magniolales, Illiciales, Laurales, Piperales, Aristochiales, Nymphaeales, Ranunculales, Papeverales, Sarraceniaceae, Trochodendrales, Hamamelidales, Eucomiales, Leitneriales, Myricales, Fagales, Casuarinales, Caryophyllales, Batales, Polygonales, Plumbaginales, Dilleniales, Theales, Malvales, Urticales, Lecythidales, Violales, Salicales, Capparales, Ericales, Diapensales, Ebenales, Primulales, Rosales, Fabales, Podostemales, Haloragales, Myrtales, Cornales, Proteales, San tales, Rafflesiales, Celastrales, Euphorbiales, Rhamnales, Sapindales, Juglandales, Geraniales, Polygalales, Umbellales, Gentianales, Polemoniales, Lamiales, Plantaginales, Scrophulariales, Campanulales, Rubiales, Dipsacales, and Asterales; the methods and CRISPR-Cas systems can be used with monocotyledonous plants such as those belonging to the orders Alismatales, Hydrocharitales, Najadales, Triuridales, Commelinales, Eriocaulales, Restionales, Poales, Juncales, Cyperales, Typhales, Bromeliales, Zingiberales, Arecales, Cyclanthales, Pandanales, Arales, Lilliales, and Orchid ales, or with plants belonging to Gymnospermae, e.g those belonging to the orders Pinales, Ginkgoales, Cycadales, Araucariales, Cupressales and Gnetales.

[0828] The AD-functionalized CRISPR systems and methods of use described herein can be used over a broad range of plant species, included in the non-limitative list of dicot, monocot or gymnosperm genera hereunder: Atropa, Alseodaphne, Anacardium, Arachis, Beilschmiedia, Brassica, Carthamus, Cocculus, Croton, Cucumis, Citrus, Citrullus, Capsicum, Catharanthus, Cocos, Coffea, Cucurbita, Daucus, Duguetia, Eschscholzia, Ficus, Fragaria, Glaucium, Glycine, Gossypium, Helianthus, Hevea, Hyoscyamus, Lactuca, Landolphia, Linum, Litsea, Lycopersicon, Lupinus, Manihot, Majorana, Malus, Medicago, Nicotiana, Olea, Parthenium, Papaver, Persea, Phaseolus, Pistacia, Pisum, Pyrus, Prunus, Raphanus, Ricinus, Senecio, Sinomenium, Stephania, Sinapis, Solanum, Theobroma, Trifolium, Trigonella, Vicia, Vinca,

Vilis, and Vigna; and the genera Allium, Andropogon, Aragrostis, Asparagus, Avena, Cynodon, Elaeis, Festuca, Festulolium, Heterocallis, Hordeum, Lemna, Lolium, Musa, Oryza, Panicum, Pannesetum, Phleum, Poa, Secale, Sorghum, Triticum, Zea, Abies, Cunninghamia, Ephedra, Picea, Pinus, and Pseudotsuga.

[0829] The AD-functionalized CRISPR systems and methods of use can also be used over a broad range of "algae" or "algae cells"; including for example algea selected from several eukaryotic phyla, including the Rhodophyta (red algae), Chlorophyta (green algae), Phaeophyta (brown algae), Bacillariophyta (diatoms), Eustigmatophyta and dinoflagellates as well as the prokaryotic phylum Cyanobacteria (blue-green algae). The term "algae" includes for example algae selected from : Amphora, Anabaena, Anikstrodesmis, Botryococcus, Chaetoceros, Chlamydomonas, Chlorella, Chlorococcum, Cyclotella, Cylindrotheca, Dunaliella, Emiliana, Euglena, Hematococcus, Isochrysis, Monochrysis, Monoraphidium, Nannochloris, Nannnochloropsis, Navicula, Nephrochloris, Nephroselmis, Nitzschia, Nodularia, Nostoc, Oochromonas, Oocystis, Oscillartoria, Pavlova, Phaeodactylum, Playtmonas, Pleurochrysis, Porhyra, Pseudoanabaena, Pyramimonas, Stichococcus, Synechococcus, Synechocystis, Tetraselmis, Thalassiosira, and Trichodesmium.

[0830] A part of a plant, i.e., a "plant tissue" may be treated according to the methods of the present invention to produce an improved plant. Plant tissue also encompasses plant cells.The term “plant cell” as used herein refers to individual units of a living plant, either in an intact whole plant or in an isolated form grown in in vitro tissue cultures, on media or agar, in suspension in a growth media or buffer or as a part of higher organized unites, such as, for example, plant tissue, a plant organ, or a whole plant.

[0831] A “protoplast” refers to a plant cell that has had its protective cell wall completely or partially removed using, for example, mechanical or enzymatic means resulting in an intact biochemical competent unit of living plant that can reform their cell wall, proliferate and regenerate grow into a whole plant under proper growing conditions.

[0832] The term "transformation" broadly refers to the process by which a plant host is genetically modified by the introduction of DNA by means of Agrobacteria or one of a variety of chemical or physical methods. As used herein, the term "plant host" refers to plants, including any cells, tissues, organs, or progeny of the plants. Many suitable plant tissues or plant cells can be transformed and include, but are not limited to, protoplasts, somatic embryos, pollen, leaves, seedlings, stems, calli, stolons, microtubers, and shoots. A plant tissue also refers to any clone of such a plant, seed, progeny, propagule whether generated sexually or asexually, and descendents of any of these, such as cuttings or seed.

[0833] The term "transformed" as used herein, refers to a cell, tissue, organ, or organism into which a foreign DNA molecule, such as a construct, has been introduced. The introduced DNA molecule may be integrated into the genomic DNA of the recipient cell, tissue, organ, or organism such that the introduced DNA molecule is transmitted to the subsequent progeny. In these embodiments, the "transformed" or "transgenic" cell or plant may also include progeny of the cell or plant and progeny produced from a breeding program employing such a transformed plant as a parent in a cross and exhibiting an altered phenotype resulting from the presence of the introduced DNA molecule. Preferably, the transgenic plant is fertile and capable of transmitting the introduced DNA to progeny through sexual reproduction.

[0834] The term "progeny", such as the progeny of a transgenic plant, is one that is born of, begotten by, or derived from a plant or the transgenic plant. The introduced DNA molecule may also be transiently introduced into the recipient cell such that the introduced DNA molecule is not inherited by subsequent progeny and thus not considered "transgenic". Accordingly, as used herein, a "non-transgenic" plant or plant cell is a plant which does not contain a foreign DNA stably integrated into its genome.

[0835] The term "plant promoter" as used herein is a promoter capable of initiating transcription in plant cells, whether or not its origin is a plant cell. Exemplary suitable plant promoters include, but are not limited to, those that are obtained from plants, plant viruses, and bacteria such as Agrobacterium or Rhizobium which comprise genes expressed in plant cells.

[0836] As used herein, a "fungal cell" refers to any type of eukaryotic cell within the kingdom of fungi. Phyla within the kingdom of fungi include Ascomycota, Basidiomycota, Blastocladiomycota, Chytridiomycota, Glomeromycota, Microsporidia, and Neocallimastigomycota. Fungal cells may include yeasts, molds, and filamentous fungi. In some embodiments, the fungal cell is a yeast cell.

[0837] As used herein, the term "yeast cell" refers to any fungal cell within the phyla Ascomycota and Basidiomycota. Yeast cells may include budding yeast cells, fission yeast cells, and mold cells. Without being limited to these organisms, many types of yeast used in laboratory and industrial settings are part of the phylum Ascomycota. In some embodiments, the yeast cell is an S. cerervisiae, Kluyveromyces marxianus, or Issatchenkia orientalis cell. Other yeast cells may include without limitation Candida spp. (e.g., Candida albicans), Yarrowia spp. (e.g., Yarrowia lipolytica), Pichia spp. (e.g., Pichia pastoris), Kluyveromyces spp. (e.g., Kluyveromyces lactis and Kluyveromyces marxianus), Neurospora spp. (e.g., Neurospora crassa), Fusarium spp. (e.g., Fusarium oxysporum), and Issatchenkia spp. (e.g., Issatchenkia orientalis, a.k.a. Pichia kudriavzevii and Candida acidothermophilum). In some

embodiments, the fungal cell is a filamentous fungal cell. As used herein, the term "filamentous fungal cell" refers to any type of fungal cell that grows in filaments, i.e., hyphae or mycelia. Examples of filamentous fungal cells may include without limitation Aspergillus spp. (e.g., Aspergillus niger), Trichoderma spp. (e.g., Trichoderma reesei), Rhizopus spp. (e.g., Rhizopus oryzae), and Mortierella spp. (e.g., Mortierella isabellina).

[0838] In some embodiments, the fungal cell is an industrial strain. As used herein, "industrial strain" refers to any strain of fungal cell used in or isolated from an industrial process, e.g., production of a product on a commercial or industrial scale. Industrial strain may refer to a fungal species that is typically used in an industrial process, or it may refer to an isolate of a fungal species that may be also used for non-industrial purposes (e.g., laboratory research). Examples of industrial processes may include fermentation (e.g., in production of food or beverage products), distillation, biofuel production, production of a compound, and production of a polypeptide. Examples of industrial strains may include, without limitation, JAY270 and ATCC4124.

[0839] In some embodiments, the fungal cell is a polyploid cell. As used herein, a "polyploid" cell may refer to any cell whose genome is present in more than one copy. A polyploid cell may refer to a type of cell that is naturally found in a polyploid state, or it may refer to a cell that has been induced to exist in a polyploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). A polyploid cell may refer to a cell whose entire genome is polyploid, or it may refer to a cell that is polyploid in a particular genomic locus of interest. Without wishing to be bound to theory, it is thought that the abundance of guideRNA may more often be a rate-limiting component in genome engineering of polyploid cells than in haploid cells, and thus the methods using the AD-functionalized CRISPR system described herein may take advantage of using a certain fungal cell type.

[0840] In some embodiments, the fungal cell is a diploid cell. As used herein, a "diploid" cell may refer to any cell whose genome is present in two copies. A diploid cell may refer to a type of cell that is naturally found in a diploid state, or it may refer to a cell that has been induced to exist in a diploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). For example, the S. cerevisiae strain S228C may be maintained in a haploid or diploid state. A diploid cell may refer to a cell whose entire genome is diploid, or it may refer to a cell that is diploid in a particular genomic locus of interest. In some embodiments, the fungal cell is a haploid cell. As used herein, a "haploid" cell may refer to any cell whose genome is present in one copy. A

haploid cell may refer to a type of cell that is naturally found in a haploid state, or it may refer to a cell that has been induced to exist in a haploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). For example, the S. cerevisiae strain S228C may be maintained in a haploid or diploid state. A haploid cell may refer to a cell whose entire genome is haploid, or it may refer to a cell that is haploid in a particular genomic locus of interest.

[0841] As used herein, a "yeast expression vector" refers to a nucleic acid that contains one or more sequences encoding an RNA and/or polypeptide and may further contain any desired elements that control the expression of the nucleic acid(s), as well as any elements that enable the replication and maintenance of the expression vector inside the yeast cell. Many suitable yeast expression vectors and features thereof are known in the art; for example, various vectors and techniques are illustrated in in Yeast Protocols, 2nd edition, Xiao, W., ed. (Humana Press, New York, 2007) and Buckholz, R.G. and Gleeson, M.A. (1991) Biotechnology (NY) 9(11): 1067-72. Yeast vectors may contain, without limitation, a centromeric (CEN) sequence, an autonomous replication sequence (ARS), a promoter, such as an RNA Polymerase III promoter, operably linked to a sequence or gene of interest, a terminator such as an RNA polymerase III terminator, an origin of replication, and a marker gene (e.g., auxotrophic, antibiotic, or other selectable markers). Examples of expression vectors for use in yeast may include plasmids, yeast artificial chromosomes, 2μ plasmids, yeast integrative plasmids, yeast replicative plasmids, shuttle vectors, and episomal plasmids.

Stable integration of AD-functionalized CRISPR system components in the genome of plants and plant cells

[0842] In particular embodiments, it is envisaged that the polynucleotides encoding the components of the AD-functionalized CRISPR system are introduced for stable integration into the genome of a plant cell. In these embodiments, the design of the transformation vector or the expression system can be adjusted depending on for when, where and under what conditions the guide RNA and/or fusion protein of adenosine deaminase and Cas13 are expressed.

[0843] In particular embodiments, it is envisaged to introduce the components of the AD-functionalized CRISPR system stably into the genomic DNA of a plant cell. Additionally or alternatively, it is envisaged to introduce the components of the AD-functionalized CRISPR system for stable integration into the DNA of a plant organelle such as, but not limited to a plastid, e mitochondrion or a chloroplast.

[0844] The expression system for stable integration into the genome of a plant cell may contain one or more of the following elements: a promoter element that can be used to express

the RNA and/or fusion protein of adenosine deaminase and Cas13 in a plant cell; a 5' untranslated region to enhance expression ; an intron element to further enhance expression in certain cells, such as monocot cells; a multiple-cloning site to provide convenient restriction sites for inserting the guide RNA and/or the fusion protein of adenosine deaminase and Cas13 encoding sequences and other desired elements; and a 3' untranslated region to provide for efficient termination of the expressed transcript.

[0845] The elements of the expression system may be on one or more expression constructs which are either circular such as a plasmid or transformation vector, or non-circular such as linear double stranded DNA.

[0846] In a particular embodiment, a AD-functionalized CRISPR expression system comprises at least: a nucleotide sequence encoding a guide RNA (gRNA) that hybridizes with a target sequence in a plant, and wherein the guide RNA comprises a guide sequence and a direct repeat sequence, and a nucleotide sequence encoding a fusion protein of adenosine deaminase and Cas13, wherein components (a) or (b) are located on the same or on different constructs, and whereby the different nucleotide sequences can be under control of the same or a different regulatory element operable in a plant cell.

[0847] DNA construct(s) containing the components of the AD-functionalized CRISPR system, and, where applicable, template sequence may be introduced into the genome of a plant, plant part, or plant cell by a variety of conventional techniques. The process generally comprises the steps of selecting a suitable host cell or host tissue, introducing the construct(s) into the host cell or host tissue, and regenerating plant cells or plants therefrom.

[0848] In particular embodiments, the DNA construct may be introduced into the plant cell using techniques such as but not limited to electroporation, microinjection, aerosol beam injection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using biolistic methods, such as DNA particle bombardment (see also Fu et al., Transgenic Res. 2000 Feb;9(1):11-9). The basis of particle bombardment is the acceleration of particles coated with gene/s of interest toward cells, resulting in the penetration of the protoplasm by the particles and typically stable integration into the genome. (see e.g. Klein et al, Nature (1987), Klein et ah, Bio/Technology (1992), Casas et ah, Proc. Natl. Acad. Sci. USA (1993).).

[0849] In particular embodiments, the DNA constructs containing components of the AD-functionalized CRISPR system may be introduced into the plant by Agrobacterium-mediated transformation. The DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional Agrobacterium tumefaciens host vector. The foreign DNA

can be incorporated into the genome of plants by infecting the plants or by incubating plant protoplasts with Agrobacterium bacteria, containing one or more Ti (tumor-inducing) plasmids. (see e.g. Fraley et al., (1985), Rogers et al., (1987) and U.S. Pat. No. 5,563,055).

Plant promoters

[0850] In order to ensure appropriate expression in a plant cell, the components of the AD-functionalized CRISPR system described herein are typically placed under control of a plant promoter, i.e. a promoter operable in plant cells. The use of different types of promoters is envisaged.

[0851] A constitutive plant promoter is a promoter that is able to express the open reading frame (ORF) that it controls in all or nearly all of the plant tissues during all or nearly all developmental stages of the plant (referred to as "constitutive expression"). One non-limiting example of a constitutive promoter is the cauliflower mosaic virus 35S promoter. "Regulated promoter" refers to promoters that direct gene expression not constitutively, but in a temporally- and/or spatially-regulated manner, and includes tissue-specific, tissue-preferred and inducible promoters. Different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. In particular embodiments, one or more of the AD-functionalized CRISPR components are expressed under the control of a constitutive promoter, such as the cauliflower mosaic virus 35S promoter issue-preferred promoters can be utilized to target enhanced expression in certain cell types within a particular plant tissue, for instance vascular cells in leaves or roots or in specific cells of the seed. Examples of particular promoters for use in the AD-functionalized CRISPR system are found in Kawamata et al., (1997) Plant Cell Physiol 38:792-803; Yamamoto et al., (1997) Plant J 12:255-65; Hire et al, (1992) Plant Mol Biol 20:207-18,Kuster et al, (1995) Plant Mol Biol 29:759-72, and Capana et al., (1994) Plant Mol Biol 25:681 -91.

[0852] Inducible promoters can be of interest to express one or more of the components of the AD-functionalized CRISPR system under limited circumstances to avoid non-specific activity of the deaminase. In particular embodiments, one or more elements of the AD-functionalized CRISPR system are expressed under control of an inducible promoter. Examples of promoters that are inducible and that allow for spatiotemporal control of gene editing or gene expression may use a form of energy. The form of energy may include but is not limited to sound energy, electromagnetic radiation, chemical energy and/or thermal energy. Examples of inducible systems include tetracycline inducible promoters (Tet-On or Tet-Off), small molecule two-hybrid transcription activations systems (FKBP, ABA, etc), or light

inducible systems (Phytochrome, LOV domains, or cryptochrome)., such as a Light Inducible Transcriptional Effector (LITE) that direct changes in transcriptional activity in a sequence-specific manner. The components of a light inducible system may include a fusion protein of adenosine deaminase and Cas13, a light-responsive cytochrome heterodimer (e.g. from Arabidopsis thaliana). Further examples of inducible DNA binding proteins and methods for their use are provided in US 61/736465 and US 61/721,283, which is hereby incorporated by reference in its entirety.

[0853] In particular embodiments, transient or inducible expression can be achieved by using, for example, chemical-regulated promotors, i.e. whereby the application of an exogenous chemical induces gene expression. Modulating of gene expression can also be obtained by a chemical-repressible promoter, where application of the chemical represses gene expression. Chemical-inducible promoters include, but are not limited to, the maize ln2-2 promoter, activated by benzene sulfonamide herbicide safeners (De Veylder et al., (1997) Plant Cell Physiol 38:568-77), the maize GST promoter (GST-ll-27, WO93/01294), activated by hydrophobic electrophilic compounds used as pre-emergent herbicides, and the tobacco PR-1 a promoter (Ono et al., (2004) Biosci Biotechnol Biochem 68:803-7) activated by salicylic acid. Promoters which are regulated by antibiotics, such as tetracycline-inducible and tetracycline-repressible promoters (Gatz et al., (1991) Mol Gen Genet 227:229-37; U.S. Patent Nos. 5,814,618 and 5,789,156) can also be used herein.

Translocation to and/or expression in specific plant organelles

[0854] The expression system may comprise elements for translocation to and/or expression in a specific plant organelle.

Chloroplast targeting

[0855] In particular embodiments, it is envisaged that the AD-functionalized CRISPR system is used to specifically modify chloroplast genes or to ensure expression in the chloroplast. For this purpose use is made of chloroplast transformation methods or compartimentalization of the AD-functionalized CRISPR components to the chloroplast. For instance, the introduction of genetic modifications in the plastid genome can reduce biosafety issues such as gene flow through pollen.

[0856] Methods of chloroplast transformation are known in the art and include Particle bombardment, PEG treatment, and microinjection. Additionally, methods involving the translocation of transformation cassettes from the nuclear genome to the pastid can be used as described in WO2010061186.

[0857] Alternatively, it is envisaged to target one or more of the AD-functionalized CRISPR components to the plant chloroplast. This is achieved by incorporating in the expression construct a sequence encoding a chloroplast transit peptide (CTP) or plastid transit peptide, operably linked to the 5' region of the sequence encoding the fusion protein of adenosine deaminase and Cas13. The CTP is removed in a processing step during translocation into the chloroplast. Chloroplast targeting of expressed proteins is well known to the skilled artisan (see for instance Protein Transport into Chloroplasts, 2010, Annual Review of Plant Biology,Vol. 61: 157-180) . In such embodiments it is also desired to target the guide RNA to the plant chloroplast. Methods and constructs which can be used for translocating guide RNA into the chloroplast by means of a chloroplast localization sequence are described, for instance, in US 20040142476, incorporated herein by reference. Such variations of constructs can be incorporated into the expression systems of the invention to efficiently translocate the AD-functionalized CRISPR system components.

Introduction of polynucleotides encoding the AD-functionalized CRISPR system in Algae cells.

[0858] Transgenic algae (or other plants such as rape) may be particularly useful in the production of vegetable oils or biofuels such as alcohols (especially methanol and ethanol) or other products. These may be engineered to express or overexpress high levels of oil or alcohols for use in the oil or biofuel industries.

[0859] US 8945839 describes a method for engineering Micro-Algae (Chlamydomonas reinhardtii cells) species) using Cas9. Using similar tools, the methods of the AD-functionalized CRISPR system described herein can be applied on Chlamydomonas species and other algae. In particular embodiments, a CRISPR-Cas protein (e.g., Cas13), adenosine deaminase (which may be fused to the CRISPR-Cas protein or an aptamer-binding adaptor protein), and guide RNA are introduced in algae expressed using a vector that expresses the fusion protein of adenosine deaminase and Cas13 under the control of a constitutive promoter such as Hsp70A-Rbc S2 or Beta2 -tubulin. Guide RNA is optionally delivered using a vector containing T7 promoter. Alternatively, Cas13 mRNA and in vitro transcribed guide RNA can be delivered to algal cells. Electroporation protocols are available to the skilled person such as the standard recommended protocol from the GeneArt Chlamydomonas Engineering kit.

Introduction of AD-functionalized Compositions in yeast cells

[0860] In particular embodiments, the invention relates to the use of the AD-functionalized CRISPR system for genome editing of yeast cells. Methods for transforming yeast cells which can be used to introduce polynucleotides encoding the AD-functionalized CRISPR system components are described in Kawai et al., 2010, Bioeng Bugs. 2010 Nov-Dec; 1(6): 395–403).

Non-limiting examples include transformation of yeast cells by lithium acetate treatment (which may further include carrier DNA and PEG treatment), bombardment or by electroporation.

Transient expression of AD-functionalized CRISPR system components in plants and plant cell

[0861] In particular embodiments, it is envisaged that the guide RNA and/or CRISPR-Cas gene are transiently expressed in the plant cell. In these embodiments, the AD-functionalized CRISPR system can ensure modification of a target gene only when both the guide RNA, the CRISPR-Cas protein (e.g., Cas13), and adenosine deaminase (which may be fused to the CRISPR-Cas protein or an aptamer-binding adaptor protein), are present in a cell, such that genomic modification can further be controlled. As the expression of the CRISPR-Cas protein is transient, plants regenerated from such plant cells typically contain no foreign DNA. In particular embodiments the CRISPR-Cas protein is stably expressed by the plant cell and the guide sequence is transiently expressed.

[0862] In particular embodiments, the AD-functionalized CRISPR system components can be introduced in the plant cells using a plant viral vector (Scholthof et al. 1996, Annu Rev Phytopathol. 1996;34:299-323). In further particular embodiments, said viral vector is a vector from a DNA virus. For example, geminivirus (e.g., cabbage leaf curl virus, bean yellow dwarf virus, wheat dwarf virus, tomato leaf curl virus, maize streak virus, tobacco leaf curl virus, or tomato golden mosaic virus) or nanovirus (e.g., Faba bean necrotic yellow virus). In other particular embodiments, said viral vector is a vector from an RNA virus. For example, tobravirus (e.g., tobacco rattle virus, tobacco mosaic virus), potexvirus (e.g., potato virus X), or hordeivirus (e.g., barley stripe mosaic virus). The replicating genomes of plant viruses are non-integrative vectors.

[0863] In particular embodiments, the vector used for transient expression of AD-functionalized CRISPR system is for instance a pEAQ vector, which is tailored for Agrobacterium-mediated transient expression (Sainsbury F. et al., Plant Biotechnol J. 2009 Sep;7(7):682-93) in the protoplast. Precise targeting of genomic locations was demonstrated using a modified Cabbage Leaf Curl virus (CaLCuV) vector to express guide RNAs in stable transgenic plants expressing a CRISPR enzyme (Scientific Reports 5, Article number: 14926 (2015), doi:10.1038/srep14926).

[0864] In particular embodiments, double-stranded DNA fragments encoding the guide RNA and/or the CRISPR-Cas gene can be transiently introduced into the plant cell. In such embodiments, the introduced double-stranded DNA fragments are provided in sufficient

quantity to modify the cell but do not persist after a contemplated period of time has passed or after one or more cell divisions. Methods for direct DNA transfer in plants are known by the skilled artisan (see for instance Davey et al. Plant Mol Biol. 1989 Sep;13(3):273-85.)

[0865] In other embodiments, an RNA polynucleotide encoding the CRISPR-Cas protein (e.g., Cas13) and/or adenosine deaminase (which may be fused to the CRISPR-Cas protein or an aptamer-binding adaptor protein) is introduced into the plant cell, which is then translated and processed by the host cell generating the protein in sufficient quantity to modify the cell (in the presence of at least one guide RNA) but which does not persist after a contemplated period of time has passed or after one or more cell divisions. Methods for introducing mRNA to plant protoplasts for transient expression are known by the skilled artisan (see for instance in Gallie, Plant Cell Reports (1993), 13;119-122).

[0866] Combinations of the different methods described above are also envisaged.

Delivery of AD-functionalized Compositions to the plant cell

[0867] In particular embodiments, it is of interest to deliver one or more components of the AD-functionalized CRISPR system directly to the plant cell. This is of interest, inter alia, for the generation of non-transgenic plants (see below). In particular embodiments, one or more of the AD-functionalized CRISPR system components is prepared outside the plant or plant cell and delivered to the cell. For instance in particular embodiments, the CRISPR-Cas protein is prepared in vitro prior to introduction to the plant cell. The CRISPR-Cas protein can be prepared by various methods known by one of skill in the art and include recombinant production. After expression, the CRISPR-Cas protein is isolated, refolded if needed, purified and optionally treated to remove any purification tags, such as a His-tag. Once crude, partially purified, or more completely purified CRISPR-Cas protein is obtained, the protein may be introduced to the plant cell.

[0868] In particular embodiments, the CRISPR-Cas protein is mixed with guide RNA targeting the gene of interest to form a pre-assembled ribonucleoprotein.

[0869] The individual components or pre-assembled ribonucleoprotein can be introduced into the plant cell via electroporation, by bombardment with CRISPR-Cas-associated gene product coated particles, by chemical transfection or by some other means of transport across a cell membrane. For instance, transfection of a plant protoplast with a pre-assembled CRISPR ribonucleoprotein has been demonstrated to ensure targeted modification of the plant genome (as described by Woo et al. Nature Biotechnology, 2015; DOI: 10.1038/nbt.3389).

[0870] In particular embodiments, the AD-functionalized CRISPR system components are introduced into the plant cells using nanoparticles. The components, either as protein or nucleic

acid or in a combination thereof, can be uploaded onto or packaged in nanoparticles and applied to the plants (such as for instance described in WO 2008042156 and US 20130185823). In particular, embodiments of the invention comprise nanoparticles uploaded with or packed with DNA molecule(s) encoding the CRISPR-Cas protein (e.g., Cas13), DNA molecule(s) encoding adenosine deaminase (which may be fused to the CRISPR-Cas protein or an aptamer-binding adaptor protein), and DNA molecules encoding the guide RNA and/or isolated guide RNA as described in WO2015089419.

[0871] Further means of introducing one or more components of the AD-functionalized CRISPR system to the plant cell is by using cell penetrating peptides (CPP). Accordingly, in particular, embodiments the invention comprises compositions comprising a cell penetrating peptide linked to the CRISPR-Cas protein. In particular embodiments of the present invention, the CRISPR-Cas protein and/or guide RNA is coupled to one or more CPPs to effectively transport them inside plant protoplasts. Ramakrishna (Genome Res. 2014 Jun;24(6):1020-7 for Cas9 in human cells). In other embodiments, the CRISPR-Cas gene and/or guide RNA are encoded by one or more circular or non-circular DNA molecule(s) which are coupled to one or more CPPs for plant protoplast delivery. The plant protoplasts are then regenerated to plant cells and further to plants. CPPs are generally described as short peptides of fewer than 35 amino acids either derived from proteins or from chimeric sequences which are capable of transporting biomolecules across cell membrane in a receptor independent manner. CPP can be cationic peptides, peptides having hydrophobic sequences, amphipatic peptides, peptides having proline-rich and anti-microbial sequence, and chimeric or bipartite peptides (Pooga and Langel 2005). CPPs are able to penetrate biological membranes and as such trigger the movement of various biomolecules across cell membranes into the cytoplasm and to improve their intracellular routing, and hence facilitate interaction of the biolomolecule with the target. Examples of CPP include amongst others: Tat, a nuclear transcriptional activator protein required for viral replication by HIV type1, penetratin, Kaposi fibroblast growth factor (FGF) signal peptide sequence, integrin β3 signal peptide sequence; polyarginine peptide Args sequence, Guanine rich-molecular transporters, sweet arrow peptide, etc.

Use of the AD-functionalized Compositions to make genetically modified non-transgenic plants

[0872] In particular embodiments, the methods described herein are used to modify endogenous genes or to modify their expression without the permanent introduction into the genome of the plant of any foreign gene, including those encoding CRISPR components, so as

to avoid the presence of foreign DNA in the genome of the plant. This can be of interest as the regulatory requirements for non-transgenic plants are less rigorous.

[0873] In particular embodiments, this is ensured by transient expression of the AD-functionalized CRISPR system components . In particular embodiments one or more of the components are expressed on one or more viral vectors which produce sufficient CRISPR-Cas protein, adenosine deaminase, and guide RNA to consistently steadily ensure modification of a gene of interest according to a method described herein.

[0874] In particular embodiments, transient expression of AD-functionalized CRISPR system constructs is ensured in plant protoplasts and thus not integrated into the genome. The limited window of expression can be sufficient to allow the AD-functionalized CRISPR system to ensure modification of a target gene as described herein.

[0875] In particular embodiments, the different components of the AD-functionalized CRISPR system are introduced in the plant cell, protoplast or plant tissue either separately or in mixture, with the aid of pariculate delivering molecules such as nanoparticles or CPP molecules as described herein above.

[0876] The expression of the AD-functionalized CRISPR system components can induce targeted modification of the genome, by deaminase activity of the adenosine deaminase. The different strategies described herein above allow CRISPR-mediated targeted genome editing without requiring the introduction of the AD-functionalized CRISPR systemt components into the plant genome. Components which are transiently introduced into the plant cell are typically removed upon crossing.

Plant cultures and regeneration

[0877] In particular embodiments, plant cells which have a modified genome and that are produced or obtained by any of the methods described herein, can be cultured to regenerate a whole plant which possesses the transformed or modified genotype and thus the desired phenotype. Conventional regeneration techniques are well known to those skilled in the art. Particular examples of such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, and typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. In further particular embodiments, plant regeneration is obtained from cultured protoplasts, plant callus, explants, organs, pollens, embryos or parts thereof (see e.g. Evans et al. (1983), Handbook of Plant Cell Culture, Klee et al (1987) Ann. Rev. of Plant Phys.).

[0878] In particular embodiments, transformed or improved plants as described herein can be self-pollinated to provide seed for homozygous improved plants of the invention

(homozygous for the DNA modification) or crossed with non-transgenic plants or different improved plants to provide seed for heterozygous plants. Where a recombinant DNA was introduced into the plant cell, the resulting plant of such a crossing is a plant which is heterozygous for the recombinant DNA molecule. Both such homozygous and heterozygous plants obtained by crossing from the improved plants and comprising the genetic modification (which can be a recombinant DNA) are referred to herein as "progeny". Progeny plants are plants descended from the original transgenic plant and containing the genome modification or recombinant DNA molecule introduced by the methods provided herein. Alternatively, genetically modified plants can be obtained by one of the methods described supra using the AD-functionalized CRISPR system whereby no foreign DNA is incorporated into the genome. Progeny of such plants, obtained by further breeding may also contain the genetic modification. Breedings are performed by any breeding methods that are commonly used for different crops (e.g., Allard, Principles of Plant Breeding, John Wiley & Sons, NY, U. of CA, Davis, CA, 50-98 (1960).

Generation of plants with enhanced agronomic traits

[0879] The AD-functionalized CRISPR systems provided herein can be used to introduce targeted A-G and T-C mutations. By co-expression of multiple targeting RNAs directed to achieve multiple modifications in a single cell, multiplexed genome modification can be ensured. This technology can be used to high-precision engineering of plants with improved characteristics, including enhanced nutritional quality, increased resistance to diseases and resistance to biotic and abiotic stress, and increased production of commercially valuable plant products or heterologous compounds.

[0880] In particular embodiments, the AD-functionalized CRISPR system as described herein is used to introduce targeted A-G and T-C mutations. Such mutation can be a nonsense mutation (e.g., premature stop codon) or a missense mutation (e.g., encoding different amino acid residue). This is of interest where the A-G and T-C mutations in certain endogenous genes can confer or contribute to a desired trait.

[0881] The methods described herein generally result in the generation of "improved plants" in that they have one or more desirable traits compared to the wildtype plant. In particular embodiments, the plants, plant cells or plant parts obtained are transgenic plants, comprising an exogenous DNA sequence incorporated into the genome of all or part of the cells of the plant. In particular embodiments, non-transgenic genetically modified plants, plant parts or cells are obtained, in that no exogenous DNA sequence is incorporated into the genome of any of the plant cells of the plant. In such embodiments, the improved plants are non-

transgenic. Where only the modification of an endogenous gene is ensured and no foreign genes are introduced or maintained in the plant genome, the resulting genetically modified crops contain no foreign genes and can thus basically be considered non-transgenic.

[0882] In particular embodiments, the polynucleotides are delivered into the cell by a DNA virus (e.g., a geminivirus) or an RNA virus (e.g., a tobravirus). In particular embodiments, the introducing steps include delivering to the plant cell a T-DNA containing one or more polynucleotide sequences encoding the CRISPR-Cas protein, the adenosine deaminase, and the guide RNA, where the delivering is via Agrobacterium. The polynucleotide sequence encoding the components of the AD-functionalized CRISPR system can be operably linked to a promoter, such as a constitutive promoter (e.g., a cauliflower mosaic virus 35S promoter), or a cell specific or inducible promoter. In particular embodiments, the polynucleotide is introduced by microprojectile bombardment. In particular embodiments, the method further includes screening the plant cell after the introducing steps to determine whether the expression of the gene of interest has been modified. In particular embodiments, the methods include the step of regenerating a plant from the plant cell. In further embodiments, the methods include cross breeding the plant to obtain a genetically desired plant lineage.

[0883] In particular embodiments of the methods described above, disease resistant crops are obtained by targeted mutation of disease susceptibility genes or genes encoding negative regulators (e.g. Mlo gene) of plant defense genes. In a particular embodiment, herbicide-tolerant crops are generated by targeted substitution of specific nucleotides in plant genes such as those encoding acetolactate synthase (ALS) and protoporphyrinogen oxidase (PPO). In particular embodiments drought and salt tolerant crops by targeted mutation of genes encoding negative regulators of abiotic stress tolerance, low amylose grains by targeted mutation of Waxy gene, rice or other grains with reduced rancidity by targeted mutation of major lipase genes in aleurone layer, etc. In particular embodiments. A more extensive list of endogenous genes encoding a traits of interest are listed below.

Use of AD-functionalized Compositions to modify polyploid plants

[0884] Many plants are polyploid, which means they carry duplicate copies of their genomes—sometimes as many as six, as in wheat. The methods according to the present invention, which make use of the AD-functionalized CRISPR system can be "multiplexed" to affect all copies of a gene, or to target dozens of genes at once. For instance, in particular embodiments, the methods of the present invention are used to simultaneously ensure a loss of function mutation in different genes responsible for suppressing defences against a disease. In particular embodiments, the methods of the present invention are used to simultaneously

suppress the expression of the TaMLO-Al, TaMLO-Bl and TaMLO-Dl nucleic acid sequence in a wheat plant cell and regenerating a wheat plant therefrom, in order to ensure that the wheat plant is resistant to powdery mildew (see also WO2015109752).

Examplary genes conferring agronomic traits

[0885] In particular embodiments, the invention encompasses methods which involve targeted A-G and T-C mutations in endogenous genes and their regulatory elements, such as listed below:

1. Genes that confer resistance to pests or diseases:

[0886] Plant disease resistance genes. A plant can be transformed with cloned resistance genes to engineer plants that are resistant to specific pathogen strains. See, e.g., Jones et al., Science 266:789 (1994) (cloning of the tomato Cf- 9 gene for resistance to Cladosporium fulvum); Martin et al., Science 262:1432 (1993) (tomato Pto gene for resistance to Pseudomonas syringae pv. tomato encodes a protein kinase); Mindrinos et al., Cell 78:1089 (1994) (Arabidopsmay be RSP2 gene for resistance to Pseudomonas syringae). A plant gene that is upregulated or down regulated during pathogen infection can be engineered for pathogen resistance. See, e.g., Thomazella et al., bioRxiv 064824; doi: https://doi.org/10.1101/064824 Epub. July 23, 2016 (tomato plants with deletions in the SlDMR6-1 which is normally upregulated during pathogen infection).

[0887] Genes conferring resistance to a pest, such as soybean cyst nematode. See e.g., PCT Application WO 96/30517; PCT Application WO 93/19181.

[0888] Bacillus thuringiensis proteins see, e.g., Geiser et al., Gene 48:109 (1986).

[0889] Lectins, see, for example, Van Damme et al., Plant Molec. Biol. 24:25 (1994.

[0890] Vitamin-binding protein, such as avidin, see PCT application US93/06487, teaching the use of avidin and avidin homologues as larvicides against insect pests.

[0891] Enzyme inhibitors such as protease or proteinase inhibitors or amylase inhibitors. See, e.g., Abe et al., J. Biol. Chem. 262:16793 (1987), Huub et al., Plant Molec. Biol. 21:985 (1993)), Sumitani et al., Biosci. Biotech. Biochem. 57:1243 (1993) and U.S. Pat. No. 5,494,813.

[0892] Insect-specific hormones or pheromones such as ecdysteroid or juvenile hormone, a variant thereof, a mimetic based thereon, or an antagonist or agonist thereof. See, for example Hammock et al., Nature 344:458 (1990).

[0893] Insect-specific peptides or neuropeptides which, upon expression, disrupts the physiology of the affected pest. For example Regan, J. Biol. Chem. 269:9 (1994) and Pratt et al., Biochem. Biophys. Res. Comm. 163:1243 (1989). See also U.S. Pat. No. 5,266,317.

[0894] Insect-specific venom produced in nature by a snake, a wasp, or any other organism. For example, see Pang et al., Gene 116: 165 (1992).

[0895] Enzymes responsible for a hyperaccumulation of a monoterpene, a sesquiterpene, a steroid, hydroxamic acid, a phenylpropanoid derivative or another nonprotein molecule with insecticidal activity.

[0896] Enzymes involved in the modification, including the post-translational modification, of a biologically active molecule; for example, a glycolytic enzyme, a proteolytic enzyme, a lipolytic enzyme, a nuclease, a cyclase, a transaminase, an esterase, a hydrolase, a phosphatase, a kinase, a phosphorylase, a polymerase, an elastase, a chitinase and a glucanase, whether natural or synthetic. See PCT application WO93/02197, Kramer et al., Insect Biochem. Molec. Biol. 23:691 (1993) and Kawalleck et al., Plant Molec. Biol. 21 :673 (1993).

[0897] Molecules that stimulates signal transduction. For example, see Botella et al., Plant Molec. Biol. 24:757 (1994), and Griess et al., Plant Physiol. 104:1467 (1994).

[0898] Viral-invasive proteins or a complex toxin derived therefrom.See Beachy et al., Ann. rev. Phytopathol. 28:451 (1990).

[0899] Developmental-arrestive proteins produced in nature by a pathogen or a parasite. See Lamb et al., Bio/Technology 10:1436 (1992) and Toubart et al., Plant J. 2:367 (1992).

[0900] A developmental-arrestive protein produced in nature by a plant. For example, Logemann et al., Bio/Technology 10:305 (1992).

[0901] In plants, pathogens are often host-specific. For example, some Fusarium species will causes tomato wilt but attacks only tomato, and other Fusarium species attack only wheat. Plants have existing and induced defenses to resist most pathogens. Mutations and recombination events across plant generations lead to genetic variability that gives rise to susceptibility, especially as pathogens reproduce with more frequency than plants. In plants there can be non-host resistance, e.g., the host and pathogen are incompatible or there can be partial resistance against all races of a pathogen, typically controlled by many genes and/or also complete resistance to some races of a pathogen but not to other races. Such resistance is typically controlled by a few genes. Using methods and components of the AD-functionalized CRISPR system, a new tool now exists to induce specific mutations in anticipation hereon. Accordingly, one can analyze the genome of sources of resistance genes, and in plants having desired characteristics or traits, use the method and components of the AD-functionalized CRISPR system to induce the rise of resistance genes. The present systems can do so with more precision than previous mutagenic agents and hence accelerate and improve plant breeding programs.

2. Genes involved in plant diseases, such as those listed in WO 2013046247

[0902] Rice diseases: Magnaporthe grisea, Cochliobolus miyabeanus, Rhizoctonia solani, Gibberella fujikuroi; Wheat diseases: Erysiphe graminis, Fusarium graminearum, F. avenaceum, F. culmorum, Microdochium nivale, Puccinia striiformis, P. graminis, P. recondita, Micronectriella nivale, Typhula sp., Ustilago tritici, Tilletia caries, Pseudocercosporella herpotrichoides, Mycosphaerella graminicola, Stagonospora nodorum, Pyrenophora tritici-repentis;Barley diseases: Erysiphe graminis, Fusarium graminearum, F. avenaceum, F. culmorum, Microdochium nivale, Puccinia striiformis, P. graminis, P. hordei, Ustilago nuda, Rhynchosporium secalis, Pyrenophora teres, Cochliobolus sativus, Pyrenophora graminea, Rhizoctonia solani;Maize diseases: Ustilago maydis, Cochliobolus heterostrophus, Gloeocercospora sorghi, Puccinia polysora, Cercospora zeae-maydis, Rhizoctonia solani;

[0903] Citrus diseases: Diaporthe citri, Elsinoe fawcetti, Penicillium digitatum, P. italicum, Phytophthora parasitica, Phytophthora citrophthora;Apple diseases: Monilinia mali, Valsa ceratosperma, Podosphaera leucotricha, Alternaria alternata apple pathotype, Venturia inaequalis, Colletotrichum acutatum, Phytophtora cactorum;

[0904] Pear diseases: Venturia nashicola, V. pirina, Alternaria alternata Japanese pear pathotype, Gymnosporangium haraeanum, Phytophtora cactorum;

[0905] Peach diseases: Monilinia fructicola, Cladosporium carpophilum, Phomopsis sp.;

[0906] Grape diseases: Elsinoe ampelina, Glomerella cingulata, Uninula necator, Phakopsora ampelopsidis, Guignardia bidwellii, Plasmopara viticola;

[0907] Persimmon diseases: Gloesporium kaki, Cercospora kaki, Mycosphaerela nawae;

[0908] Gourd diseases: Colletotrichum lagenarium, Sphaerotheca fuliginea, Mycosphaerella melonis, Fusarium oxysporum, Pseudoperonospora cubensis, Phytophthora sp., Pythium sp.;

[0909] Tomato diseases: Alternaria solani, Cladosporium fulvum, Phytophthora infestans; Pseudomonas syringae pv. Tomato; Phytophthora capsici; Xanthomonas

[0910] Eggplant diseases: Phomopsis vexans, Erysiphe cichoracearum; Brassicaceous vegetable diseases: Alternaria japonica, Cercosporella brassicae, Plasmodiophora brassicae, Peronospora parasitica;

[0911] Welsh onion diseases: Puccinia allii, Peronospora destructor;

[0912] Soybean diseases: Cercospora kikuchii, Elsinoe glycines, Diaporthe phaseolorum var. sojae, Septoria glycines, Cercospora sojina, Phakopsora pachyrhizi, Phytophthora sojae, Rhizoctonia solani, Corynespora casiicola, Sclerotinia sclerotiorum;

[0913] Kidney bean diseases: Colletrichum lindemthianum;

[0914] Peanut diseases: Cercospora personata, Cercospora arachidicola, Sclerotium rolfsii;

[0915] Pea diseases pea: Erysiphe pisi;

[0916] Potato diseases: Alternaria solani, Phytophthora infestans, Phytophthora erythroseptica, Spongospora subterranean, f. sp. Subterranean;

[0917] Strawberry diseases: Sphaerotheca humuli, Glomerella cingulata;

[0918] Tea diseases: Exobasidium reticulatum, Elsinoe leucospila, Pestalotiopsis sp., Colletotrichum theae-sinensis;

[0919] Tobacco diseases: Alternaria longipes, Erysiphe cichoracearum, Colletotrichum tabacum, Peronospora tabacina, Phytophthora nicotianae;

[0920] Rapeseed diseases: Sclerotinia sclerotiorum, Rhizoctonia solani;

[0921] Cotton diseases: Rhizoctonia solani;

[0922] Beet diseases: Cercospora beticola, Thanatephorus cucumeris, Thanatephorus cucumeris, Aphanomyces cochlioides;

[0923] Rose diseases: Diplocarpon rosae, Sphaerotheca pannosa, Peronospora sparsa;

[0924] Diseases of chrysanthemum and asteraceae: Bremia lactuca, Septoria chrysanthemi-indici, Puccinia horiana;

[0925] Diseases of various plants: Pythium aphanidermatum, Pythium debarianum, Pythium graminicola, Pythium irregulare, Pythium ultimum, Botrytis cinerea, Sclerotinia sclerotiorum;

[0926] Radish diseases: Alternaria brassicicola;

[0927] Zoysia diseases: Sclerotinia homeocarpa, Rhizoctonia solani;

[0928] Banana diseases: Mycosphaerella fijiensis, Mycosphaerella musicola;

[0929] Sunflower diseases: Plasmopara halstedii;

[0930] Seed diseases or diseases in the initial stage of growth of various plants caused by Aspergillus spp., Penicillium spp., Fusarium spp., Gibberella spp., Tricoderma spp., Thielaviopsis spp., Rhizopus spp., Mucor spp., Corticium spp., Rhoma spp., Rhizoctonia spp., Diplodia spp., or the like;

[0931] Virus diseases of various plants mediated by Polymixa spp., Olpidium spp., or the like.

3. Examples of genes that confer resistance to herbicides:

[0932] Resistance to herbicides that inhibit the growing point or meristem, such as an imidazolinone or a sulfonylurea, for example, by Lee et al., EMBO J. 7:1241 (1988), and Miki et al., Theor. Appl. Genet. 80:449 (1990), respectively.

[0933] Glyphosate tolerance (resistance conferred by, e.g., mutant 5-enolpyruvylshikimate-3- phosphate synthase (EPSPs) genes, aroA genes and glyphosate acetyl transferase (GAT) genes, respectively), or resistance to other phosphono compounds such as by glufosinate (phosphinothricin acetyl transferase (PAT) genes from Streptomyces species, including Streptomyces hygroscopicus and Streptomyces viridichromogenes), and to pyridinoxy or phenoxy proprionic acids and cyclohexones by ACCase inhibitor-encoding genes. See, for example, U.S. Pat. No. 4,940,835 and U.S. Pat. 6,248,876 , U.S. Pat. No. 4,769,061 , EP No. 0 333 033 and U.S. Pat No. 4,975,374. See also EP No. 0242246, DeGreef et al., Bio/Technology 7:61 (1989), Marshall et al., Theor. Appl. Genet. 83:435 (1992), WO 2005012515 to Castle et. al. and WO 2005107437.

[0934] Resistance to herbicides that inhibit photosynthesis, such as a triazine (psbA and gs+ genes) or a benzonitrile (nitrilase gene), and glutathione S-transferase in Przibila et al., Plant Cell 3:169 (1991), U.S. Pat. No. 4,810,648, and Hayes et al., Biochem. J. 285: 173 (1992).

[0935] Genes encoding Enzymes detoxifying the herbicide or a mutant glutamine synthase enzyme that is resistant to inhibition, e.g. n U.S. patent application Ser. No. 11/760,602. Or a detoxifying enzyme is an enzyme encoding a phosphinothricin acetyltransferase (such as the bar or pat protein from Streptomyces species). Phosphinothricin acetyltransferases are for example described in U.S. Pat. Nos. 5,561,236; 5,648,477; 5,646,024; 5,273,894; 5,637,489; 5,276,268; 5,739,082; 5,908,810 and 7,112,665.

[0936] Hydroxyphenylpyruvatedioxygenases (HPPD) inhibitors, naturally occuring HPPD resistant enzymes, or genes encoding a mutated or chimeric HPPD enzyme as described in WO 96/38567, WO 99/24585, and WO 99/24586, WO 2009/144079, WO 2002/046387, or U.S. Pat. No. 6,768,044.

4. Examples of genes involved in Abiotic stress tolerance:

[0937] Transgene capable of reducing the expression and/or the activity of poly(ADP-ribose) polymerase (PARP) gene in the plant cells or plants as described in WO 00/04173 or, WO/2006/045633.

[0938] Transgenes capable of reducing the expression and/or the activity of the PARG encoding genes of the plants or plants cells, as described e.g. in WO 2004/090140.

[0939] Transgenes coding for a plant-functional enzyme of the nicotineamide adenine dinucleotide salvage synthesis pathway including nicotinamidase, nicotinate phosphoribosyltransferase, nicotinic acid mononucleotide adenyl transferase, nicotinamide adenine dinucleotide synthetase or nicotine amide phosphorybosyltransferase as described e.g. in EP 04077624.7, WO 2006/133827, PCT/EP07/002,433, EP 1999263, or WO 2007/107326.

[0940] Enzymes involved in carbohydrate biosynthesis include those described in e.g. EP 0571427, WO 95/04826, EP 0719338, WO 96/15248, WO 96/19581, WO 96/27674, WO 97/11188, WO 97/26362, WO 97/32985, WO 97/42328, WO 97/44472, WO 97/45545, WO 98/27212, WO 98/40503, WO99/58688, WO 99/58690, WO 99/58654, WO 00/08184, WO 00/08185, WO 00/08175, WO 00/28052, WO 00/77229, WO 01/12782, WO 01/12826, WO 02/101059, WO 03/071860, WO 2004/056999, WO 2005/030942, WO 2005/030941, WO 2005/095632, WO 2005/095617, WO 2005/095619, WO 2005/095618, WO 2005/123927, WO 2006/018319, WO 2006/103107, WO 2006/108702, WO 2007/009823, WO 00/22140, WO 2006/063862, WO 2006/072603, WO 02/034923, EP 06090134.5, EP 06090228.5, EP 06090227.7, EP 07090007.1, EP 07090009.7, WO 01/14569, WO 02/79410, WO 03/33540, WO 2004/078983, WO 01/19975, WO 95/26407, WO 96/34968, WO 98/20145, WO 99/12950, WO 99/66050, WO 99/53072, U.S. Pat. No. 6,734,341, WO 00/11192, WO 98/22604, WO 98/32326, WO 01/98509, WO 01/98509, WO 2005/002359, U.S. Pat. No. 5,824,790, U.S. Pat. No. 6,013,861, WO 94/04693, WO 94/09144, WO 94/11520, WO 95/35026 or WO 97/20936 or enzymes involved in the production of polyfructose, especially of the inulin and levan-type, as disclosed in EP 0663956, WO 96/01904, WO 96/21023, WO 98/39460, and WO 99/24593, the production of alpha-1,4-glucans as disclosed in WO 95/31553, US 2002031826, U.S. Pat. No. 6,284,479, U.S. Pat. No. 5,712,107, WO 97/47806, WO 97/47807, WO 97/47808 and WO 00/14249, the production of alpha-1,6 branched alpha-1,4-glucans, as disclosed in WO 00/73422, the production of alternan, as disclosed in e.g. WO 00/47727, WO 00/73422, EP 06077301.7, U.S. Pat. No. 5,908,975 and EP 0728213, the production of hyaluronan, as for example disclosed in WO 2006/032538, WO 2007/039314, WO 2007/039315, WO 2007/039316, JP 2006304779, and WO 2005/012529.

[0941] Genes that improve drought resistance. For example, WO 2013122472 discloses that the absence or reduced level of functional Ubiquitin Protein Ligase protein (UPL) protein, more specifically, UPL3, leads to a decreased need for water or improved resistance to drought of said plant. Other examples of transgenic plants with increased drought tolerance are disclosed in, for example, US 2009/0144850, US 2007/0266453, and WO 2002/083911. US2009/0144850 describes a plant displaying a drought tolerance phenotype due to altered expression of a DR02 nucleic acid. US 2007/0266453 describes a plant displaying a drought tolerance phenotype due to altered expression of a DR03 nucleic acid and WO 2002/08391 1 describes a plant having an increased tolerance to drought stress due to a reduced activity of an ABC transporter which is expressed in guard cells. Another example is the work by Kasuga and co-authors (1999), who describe that overexpression of cDNA encoding DREB1 A in

transgenic plants activated the expression of many stress tolerance genes under normal growing conditions and resulted in improved tolerance to drought, salt loading, and freezing. However, the expression of DREB1A also resulted in severe growth retardation under normal growing conditions (Kasuga (1999) Nat Biotechnol 17(3) 287-291).

[0942] In further particular embodiments, crop plants can be improved by influencing specific plant traits. For example, by developing pesticide-resistant plants, improving disease resistance in plants, improving plant insect and nematode resistance, improving plant resistance against parasitic weeds, improving plant drought tolerance, improving plant nutritional value, improving plant stress tolerance, avoiding self-pollination, plant forage digestibility biomass, grain yield etc. A few specific non-limiting examples are provided hereinbelow.

[0943] In addition to targeted mutation of single genes, AD-functionalized CRISPR system can be designed to allow targeted mutation of multiple genes, deletion of chromosomal fragment, site-specific integration of transgene, site-directed mutagenesis in vivo, and precise gene replacement or allele swapping in plants. Therefore, the methods described herein have broad applications in gene discovery and validation, mutational and cisgenic breeding, and hybrid breeding. These applications facilitate the production of a new generation of genetically modified crops with various improved agronomic traits such as herbicide resistance, disease resistance, abiotic stress tolerance, high yield, and superior quality.

Use of AD-functionalized Compositions to create male sterile plants

[0944] Hybrid plants typically have advantageous agronomic traits compared to inbred plants. However, for self-pollinating plants, the generation of hybrids can be challenging. In different plant types, genes have been identified which are important for plant fertility, more particularly male fertility. For instance, in maize, at least two genes have been identified which are important in fertility (Amitabh Mohanty International Conference on New Plant Breeding Molecular Technologies Technology Development And Regulation, Oct 9-10, 2014, Jaipur, India; Svitashev et al. Plant Physiol. 2015 Oct;169(2):931-45; Djukanovic et al. Plant J. 2013 Dec;76(5):888-99). The methods and systems provided herein can be used to target genes required for male fertility so as to generate male sterile plants which can easily be crossed to generate hybrids. In particular embodiments, the AD-functionalized CRISPR system provided herein is used for targeted mutagenesis of the cytochrome P450-like gene (MS26) or the meganuclease gene (MS45) thereby conferring male sterility to the maize plant. Maize plants which are as such genetically altered can be used in hybrid breeding programs.

Increasing the fertility stage in plants

[0945] In particular embodiments, the methods and systems provided herein are used to prolong the fertility stage of a plant such as of a rice plant. For instance, a rice fertility stage gene such as Ehd3 can be targeted in order to generate a mutation in the gene and plantlets can be selected for a prolonged regeneration plant fertility stage (as described in CN 104004782)

Use of AD-functionalized Compositions to generate genetic variation in a crop of interest

[0946] The availability of wild germplasm and genetic variations in crop plants is the key to crop improvement programs, but the available diversity in germplasms from crop plants is limited. The present invention envisages methods for generating a diversity of genetic variations in a germplasm of interest. In this application of the AD-functionalized CRISPR system a library of guide RNAs targeting different locations in the plant genome is provided and is introduced into plant cells together with the CRISPR-Cas protein and adenosine deaminase. In this way a collection of genome-scale point mutations and gene knock-outs can be generated. In particular embodiments, the methods comprise generating a plant part or plant from the cells so obtained and screening the cells for a trait of interest. The target genes can include both coding and non-coding regions. In particular embodiments, the trait is stress tolerance and the method is a method for the generation of stress-tolerant crop varieties

Use of AD-functionalized Compositions to affect fruit-ripening

[0947] Ripening is a normal phase in the maturation process of fruits and vegetables. Only a few days after it starts it renders a fruit or vegetable inedible. This process brings significant losses to both farmers and consumers. In particular embodiments, the methods of the present invention are used to reduce ethylene production. This is ensured by ensuring one or more of the following: a. Suppression of ACC synthase gene expression. ACC (1-aminocyclopropane-1-carboxylic acid) synthase is the enzyme responsible for the conversion of S-adenosylmethionine (SAM) to ACC; the second to the last step in ethylene biosynthesis. Enzyme expression is hindered when an antisense ("mirror-image") or truncated copy of the synthase gene is inserted into the plant's genome; b. Insertion of the ACC deaminase gene. The gene coding for the enzyme is obtained from Pseudomonas chlororaphis, a common nonpathogenic soil bacterium. It converts ACC to a different compound thereby reducing the amount of ACC available for ethylene production; c. Insertion of the SAM hydrolase gene. This approach is similar to ACC deaminase wherein ethylene production is hindered when the amount of its precursor metabolite is reduced; in this case SAM is converted to homoserine. The gene coding for the enzyme is obtained from E. coli T3 bacteriophage and d. Suppression of ACC oxidase gene expression. ACC oxidase is the enzyme which catalyzes the oxidation of ACC to ethylene, the last step in the ethylene biosynthetic pathway. Using the methods

described herein, down regulation of the ACC oxidase gene results in the suppression of ethylene production, thereby delaying fruit ripening. In particular embodiments, additionally or alternatively to the modifications described above, the methods described herein are used to modify ethylene receptors, so as to interfere with ethylene signals obtained by the fruit. In particular embodiments, expression of the ETR1 gene, encoding an ethylene binding protein is modified, more particularly suppressed. In particular embodiments, additionally or alternatively to the modifications described above, the methods described herein are used to modify expression of the gene encoding Polygalacturonase (PG), which is the enzyme responsible for the breakdown of pectin, the substance that maintains the integrity of plant cell walls. Pectin breakdown occurs at the start of the ripening process resulting in the softening of the fruit. Accordingly, in particular embodiments, the methods described herein are used to introduce a mutation in the PG gene or to suppress activation of the PG gene in order to reduce the amount of PG enzyme produced thereby delaying pectin degradation.

[0948] Thus in particular embodiments, the methods comprise the use of the AD-functionalized CRISPR system to ensure one or more modifications of the genome of a plant cell such as described above, and regenerating a plant therefrom. In particular embodiments, the plant is a tomato plant.

Increasing storage life of plants

[0949] In particular embodiments, the methods of the present invention are used to modify genes involved in the production of compounds which affect storage life of the plant or plant part. More particularly, the modification is in a gene that prevents the accumulation of reducing sugars in potato tubers. Upon high-temperature processing, these reducing sugars react with free amino acids, resulting in brown, bitter-tasting products and elevated levels of acrylamide, which is a potential carcinogen. In particular embodiments, the methods provided herein are used to reduce or inhibit expression of the vacuolar invertase gene (VInv), which encodes a protein that breaks down sucrose to glucose and fructose (Clasen et al. DOI: 10.1111/pbi.12370).

The use of the AD-functionalized Compositions to ensure a value added trait

[0950] In particular embodiments the AD-functionalized CRISPR system is used to produce nutritionally improved agricultural crops. In particular embodiments, the methods provided herein are adapted to generate “functional foods”, i.e. a modified food or food ingredient that may provide a health benefit beyond the traditional nutrients it contains and or “nutraceutical”, i.e. substances that may be considered a food or part of a food and provides health benefits, including the prevention and treatment of disease. In particular embodiments,

the nutraceutical is useful in the prevention and/or treatment of one or more of cancer, diabetes, cardiovascular disease, and hypertension.

[0951] Examples of nutritionally improved crops include (Newell-McGloughlin, Plant Physiology, July 2008, Vol. 147, pp. 939–953):

[0952] Modified protein quality, content and/or amino acid composition, such as have been described for Bahiagrass (Luciani et al. 2005, Florida Genetics Conference Poster), Canola (Roesler et al., 1997, Plant Physiol 113 75–81), Maize (Cromwell et al, 1967, 1969 J Anim Sci 26 1325–1331, O'Quin et al. 2000 J Anim Sci 78 2144–2149, Yang et al. 2002, Transgenic Res 11 11–20, Young et al. 2004, Plant J 38 910–922), Potato (Yu J and Ao, 1997 Acta Bot Sin 39 329–334; Chakraborty et al. 2000, Proc Natl Acad Sci USA 97 3724–3729; Li et al. 2001) Chin Sci Bull 46 482–484, Rice (Katsube et al. 1999, Plant Physiol 120 1063–1074), Soybean (Dinkins et al. 2001, Rapp 2002, In Vitro Cell Dev Biol Plant 37 742–747), Sweet Potato (Egnin and Prakash 1997, In Vitro Cell Dev Biol 33 52A).

[0953] Essential amino acid content, such as has been described for Canola (Falco et al. 1995, Bio/Technology 13 577–582), Lupin (White et al. 2001, J Sci Food Agric 81 147–154), Maize (Lai and Messing, 2002, Agbios 2008 GM crop database (March 11, 2008)), Potato (Zeh et al. 2001, Plant Physiol 127 792–802), Sorghum (Zhao et al. 2003, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 413–416), Soybean (Falco et al. 1995 Bio/Technology 13 577–582; Galili et al. 2002 Crit Rev Plant Sci 21 167–204).

[0954] Oils and Fatty acids such as for Canola (Dehesh et al. (1996) Plant J 9 167–172 [PubMed] ; Del Vecchio (1996) INFORM International News on Fats, Oils and Related Materials 7 230–243; Roesler et al. (1997) Plant Physiol 113 75–81 [PMC free article] [PubMed]; Froman and Ursin (2002, 2003) Abstracts of Papers of the American Chemical Society 223 U35; James et al. (2003) Am J Clin Nutr 77 1140–1145 [PubMed]; Agbios (2008, above); coton (Chapman et al. (2001) . J Am Oil Chem Soc 78 941–947; Liu et al. (2002) J Am Coll Nutr 21 205S–211S [PubMed]; O'Neill (2007) Australian Life Scientist. http://www.biotechnews.com.au/index.php/id;866694817;fp;4;fpid;2 (June 17, 2008), Linseed (Abbadi et al., 2004, Plant Cell 16: 2734–2748), Maize (Young et al., 2004, Plant J 38 910–922), oil palm (Jalani et al. 1997, J Am Oil Chem Soc 74 1451–1455; Parveez, 2003, AgBiotechNet 113 1–8), Rice (Anai et al., 2003, Plant Cell Rep 21 988–992), Soybean (Reddy and Thomas, 1996, Nat Biotechnol 14 639–642; Kinney and Kwolton, 1998, Blackie Academic and Professional, London, pp 193–213), Sunflower (Arcadia, Biosciences 2008)

[0955] Carbohydrates, such as Fructans described for Chicory (Smeekens (1997) Trends Plant Sci 2 286–287, Sprenger et al. (1997) FEBS Lett 400 355–358, Sévenier et al. (1998) Nat

Biotechnol 16 843–846), Maize (Caimi et al. (1996) Plant Physiol 110 355–363), Potato (Hellwege et al. ,1997 Plant J 12 1057–1065), Sugar Beet (Smeekens et al. 1997, above), Inulin, such as described for Potato (Hellewege et al. 2000, Proc Natl Acad Sci USA 97 8699–8704), Starch, such as described for Rice (Schwall et al. (2000) Nat Biotechnol 18 551–554, Chiang et al. (2005) Mol Breed 15 125–143),

[0956] Vitamins and carotenoids, such as described for Canola (Shintani and DellaPenna (1998) Science 282 2098–2100), Maize (Rocheford et al. (2002) . J Am Coll Nutr 21 191S–198S, Cahoon et al. (2003) Nat Biotechnol 21 1082–1087, Chen et al. (2003) Proc Natl Acad Sci USA 100 3525–3530), Mustardseed (Shewmaker et al. (1999) Plant J 20 401–412, Potato (Ducreux et al., 2005, J Exp Bot 56 81–89), Rice (Ye et al. (2000) Science 287 303–305, Strawberry (Agius et al. (2003), Nat Biotechnol 21 177–181), Tomato (Rosati et al. (2000) Plant J 24 413–419, Fraser et al. (2001) J Sci Food Agric 81 822–827, Mehta et al. (2002) Nat Biotechnol 20 613–618, Díaz de la Garza et al. (2004) Proc Natl Acad Sci USA 101 13720–13725, Enfissi et al. (2005) Plant Biotechnol J 3 17–27, DellaPenna (2007) Proc Natl Acad Sci USA 104 3675–3676.

[0957] Functional secondary metabolites, such as described for Apple (stilbenes, Szankowski et al. (2003) Plant Cell Rep 22: 141–149), Alfalfa (resveratrol, Hipskind and Paiva (2000) Mol Plant Microbe Interact 13 551–562), Kiwi (resveratrol, Kobayashi et al. (2000) Plant Cell Rep 19 904–910), Maize and Soybean (flavonoids, Yu et al. (2000) Plant Physiol 124 781–794), Potato (anthocyanin and alkaloid glycoside, Lukaszewicz et al. (2004) J Agric Food Chem 52 1526–1533), Rice (flavonoids & resveratrol, Stark-Lorenzen et al. (1997) Plant Cell Rep 16 668–673, Shin et al. (2006) Plant Biotechnol J 4 303–315), Tomato (+resveratrol, chlorogenic acid, flavonoids, stilbene; Rosati et al. (2000) above, Muir et al. (2001) Nature 19 470–474, Niggeweg et al. (2004) Nat Biotechnol 22 746–754, Giovinazzo et al. (2005) Plant Biotechnol J 3 57–69), wheat (caffeic and ferulic acids, resveratrol; United Press International (2002)); and

[0958] Mineral availabilities such as described for Alfalfa (phytase, Austin-Phillips et al. (1999) http://www.molecularfarming.com/nonmedical.html), Lettuse (iron, Goto et al. (2000) Theor Appl Genet 100 658–664), Rice (iron, Lucca et al. (2002) J Am Coll Nutr 21 184S–190S), Maize, Soybean and wheate (phytase, Drakakaki et al. (2005) Plant Mol Biol 59 869–880, Denbow et al. (1998) Poult Sci 77 878–881, Brinch-Pedersen et al. (2000) Mol Breed 6 195–206).

[0959] In particular embodiments, the value-added trait is related to the envisaged health benefits of the compounds present in the plant. For instance, in particular embodiments, the

value-added crop is obtained by applying the methods of the invention to ensure the modification of or induce/increase the synthesis of one or more of the following compounds:

- Carotenoids, such as α-Carotene present in carrots which Neutralizes free radicals that may cause damage to cells or β-Carotene present in various fruits and vegetables which neutralizes free radicals;
- Lutein present in green vegetables which contributes to maintenance of healthy vision;
- Lycopene present in tomato and tomato products, which is believed to reduce the risk of prostate cancer;
- Zeaxanthin, present in citrus and maize, which contributes to mainteance of healthy vision
- Dietary fiber such as insoluble fiber present in wheat bran which may reduce the risk of breast and/or colon cancer and β-Glucan present in oat, soluble fiber present in Psylium and whole cereal grains which may reduce the risk of cardiovascular disease (CVD);
- Fatty acids, such as ω-3 fatty acids which may reduce the risk of CVD and improve mental and visual functions, Conjugated linoleic acid, which may improve body composition, may decrease risk of certain cancers and GLA which may reduce inflammation risk of cancer and CVD, may improve body composition ;
- Flavonoids such as Hydroxycinnamates, present in wheat which have Antioxidant-like activities, may reduce risk of degenerative diseases, flavonols, catechins and tannins present in fruits and vegetables which neutralize free radicals and may reduce risk of cancer ;
- Glucosinolates, indoles, isothiocyanates, such as Sulforaphane, present in Cruciferous vegetables (broccoli, kale), horseradish, which neutralize free radicals, may reduce risk of cancer;
- Phenolics, such as stilbenes present in grape which May reduce risk of degenerative diseases, heart disease, and cancer, may have longevity effect and caffeic acid and ferulic acid present in vegetables and citrus which have Antioxidant-like activities, may reduce risk of degenerative diseases, heart disease, and eye disease, and epicatechin present in cacao which has Antioxidant-like activities, may reduce risk of degenerative diseases and heart disease;
- Plant stanols/sterols present in maize, soy, wheat and wooden oils which May reduce risk of coronary heart disease by lowering blood cholesterol levels;

- Fructans, inulins, fructo-oligosaccharides present in Jerusalem artichoke, shallot, onion powder which may improve gastrointestinal health;
- Saponins present in soybean, which may lower LDL cholesterol ;
- Soybean protein present in soybean which may reduce risk of heart disease;
- Phytoestrogens such as isoflavones present in soybean which May reduce menopause symptoms, such as hot flashes, may reduce osteoporosis and CVD and lignans present in flax, rye and vegetables, which May protect against heart disease and some cancers, may lower LDL cholesterol, total cholesterol.;
- Sulfides and thiols such as diallyl sulphide present in onion, garlic, olive, leek and scallon and Allyl methyl trisulfide, dithiolthiones present in cruciferous vegetables which may lower LDL cholesterol, helps to maintain healthy immune system; and
- Tannins, such as proanthocyanidins, present in cranberry, cocoa, which may improve urinary tract health, may reduce risk of CVD and high blood pressure.

[0960] In addition, the methods of the present invention also envisage modifying protein/starch functionality, shelf life, taste/aesthetics, fiber quality, and allergen, antinutrient, and toxin reduction traits.

[0961] Accordingly, the invention encompasses methods for producing plants with nutritional added value, said methods comprising introducing into a plant cell a gene encoding an enzyme involved in the production of a component of added nutritional value using the AD-functionalized CRISPR system as described herein and regenerating a plant from said plant cell, said plant characterized in an increase expression of said component of added nutritional value. In particular embodiments, the AD-functionalized CRISPR system is used to modify the endogenous synthesis of these compounds indirectly, e.g. by modifying one or more transcription factors that controls the metabolism of this compound. Methods for introducing a gene of interest into a plant cell and/or modifying an endogenous gene using the AD-functionalized CRISPR system are described herein above.

[0962] Some specific examples of modifications in plants that have been modified to confer value-added traits are: plants with modified fatty acid metabolism, for example, by transforming a plant with an antisense gene of stearyl-ACP desaturase to increase stearic acid content of the plant. See Knultzon et al., Proc. Natl. Acad. Sci. U.S.A. 89:2624 (1992). Another example involves decreasing phytate content, for example by cloning and then reintroducing DNA associated with the single allele which may be responsible for maize mutants characterized by low levels of phytic acid. See Raboy et al, Maydica 35:383 (1990).

[0963] Similarly, expression of the maize (Zea mays) Tfs C1 and R, which regulate the production of flavonoids in maize aleurone layers under the control of a strong promoter, resulted in a high accumulation rate of anthocyanins in Arabidopsis (Arabidopsis thaliana), presumably by activating the entire pathway (Bruce et al., 2000, Plant Cell 12:65–80). DellaPenna (Welsch et al., 2007 Annu Rev Plant Biol 57: 711–738) found that Tf RAP2.2 and its interacting partner SINAT2 increased carotenogenesis in Arabidopsis leaves. Expressing the Tf Dof1 induced the up-regulation of genes encoding enzymes for carbon skeleton production, a marked increase of amino acid content, and a reduction of the Glc level in transgenic Arabidopsis (Yanagisawa, 2004 Plant Cell Physiol 45: 386–391), and the DOF Tf AtDof1.1 (OBP2) up-regulated all steps in the glucosinolate biosynthetic pathway in Arabidopsis (Skirycz et al., 2006 Plant J 47: 10–24).

Reducing allergen in plants

[0964] In particular embodiments the methods provided herein are used to generate plants with a reduced level of allergens, making them safer for the consumer. In particular embodiments, the methods comprise modifying expression of one or more genes responsible for the production of plant allergens. For instance, in particular embodiments, the methods comprise down-regulating expression of a Lol p5 gene in a plant cell, such as a ryegrass plant cell and regenerating a plant therefrom so as to reduce allergenicity of the pollen of said plant (Bhalla et al. 1999, Proc. Natl. Acad. Sci. USA Vol. 96: 11676–11680).

[0965] Peanut allergies and allergies to legumes generally are a real and serious health concern. The AD-functionalized CRISPR system of the present invention can be used to identify and then mutate genes encoding allergenic proteins of such legumes. Without limitation as to such genes and proteins, Nicolaou et al. identifies allergenic proteins in peanuts, soybeans, lentils, peas, lupin, green beans, and mung beans. See, Nicolaou et al., Current Opinion in Allergy and Clinical Immunology 2011;11(3):222).

Screening methods for endogenous genes of interest

[0966] The methods provided herein further allow the identification of genes of value encoding enzymes involved in the production of a component of added nutritional value or generally genes affecting agronomic traits of interest, across species, phyla, and plant kingdom. By selectively targeting e.g. genes encoding enzymes of metabolic pathways in plants using the AD-functionalized CRISPR system as described herein, the genes responsible for certain nutritional aspects of a plant can be identified. Similarly, by selectively targeting genes which may affect a desirable agronomic trait, the relevant genes can be identified. Accordingly, the

present invention encompasses screening methods for genes encoding enzymes involved in the production of compounds with a particular nutritional value and/or agronomic traits.

Use of AD-functionalized CRISPR system in biofuel production

[0967] The term "biofuel" as used herein is an alternative fuel made from plant and plant-derived resources. Renewable biofuels can be extracted from organic matter whose energy has been obtained through a process of carbon fixation or are made through the use or conversion of biomass. This biomass can be used directly for biofuels or can be converted to convenient energy containing substances by thermal conversion, chemical conversion, and biochemical conversion. This biomass conversion can result in fuel in solid, liquid, or gas form. There are two types of biofuels: bioethanol and biodiesel. Bioethanol is mainly produced by the sugar fermentation process of cellulose (starch), which is mostly derived from maize and sugar cane. Biodiesel on the other hand is mainly produced from oil crops such as rapeseed, palm, and soybean. Biofuels are used mainly for transportation.

Enhancing plant properties for biofuel production

[0968] In particular embodiments, the methods using the AD-functionalized CRISPR system as described herein are used to alter the properties of the cell wall in order to facilitate access by key hydrolysing agents for a more efficient release of sugars for fermentation. In particular embodiments, the biosynthesis of cellulose and/or lignin are modified. Cellulose is the major component of the cell wall. The biosynthesis of cellulose and lignin are co-regulated. By reducing the proportion of lignin in a plant the proportion of cellulose can be increased. In particular embodiments, the methods described herein are used to downregulate lignin biosynthesis in the plant so as to increase fermentable carbohydrates. More particularly, the methods described herein are used to downregulate at least a first lignin biosynthesis gene selected from the group consisting of 4-coumarate 3-hydroxylase (C3H), phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H), hydroxycinnamoyl transferase (HCT), caffeic acid O-methyltransferase (COMT), caffeoyl CoA 3-O-methyltransferase (CCoAOMT), ferulate 5- hydroxylase (F5H), cinnamyl alcohol dehydrogenase (CAD), cinnamoyl CoA-reductase (CCR), 4- coumarate-CoA ligase (4CL), monolignol-lignin-specific glycosyltransferase, and aldehyde dehydrogenase (ALDH) as disclosed in WO 2008064289 A2.

[0969] In particular embodiments, the methods described herein are used to produce plant mass that produces lower levels of acetic acid during fermentation (see also WO 2010096488). More particularly, the methods disclosed herein are used to generate mutations in homologs to CaslL to reduce polysaccharide acetylation.

Modifying yeast for Biofuel production

[0970] In particular embodiments, the AD-functionalized CRISPR system provided herein is used for bioethanol production by recombinant micro-organisms. For instance, the AD-functionalized CRISPR system can be used to engineer micro-organisms, such as yeast, to generate biofuel or biopolymers from fermentable sugars and optionally to be able to degrade plant-derived lignocellulose derived from agricultural waste as a source of fermentable sugars. In some embodiments, the AD-functionalized CRISPR system is used to modify endogenous metabolic pathways which compete with the biofuel production pathway.

[0971] Accordingly, in more particular embodiments, the methods described herein are used to modify a micro-organism as follows: to modify at least one nucleic acid encoding for an enzyme in a metabolic pathway in said host cell, wherein said pathway produces a metabolite other than acetaldehyde from pyruvate or ethanol from acetaldehyde, and wherein said modification results in a reduced production of said metabolite, or to introduce at least one nucleic acid encoding for an inhibitor of said enzyme.

Modifying Algae and plants for production of vegetable oils or biofuels

[0972] Transgenic algae or other plants such as rape may be particularly useful in the production of vegetable oils or biofuels such as alcohols (especially methanol and ethanol), for instance. These may be engineered to express or overexpress high levels of oil or alcohols for use in the oil or biofuel industries.

[0973] According to particular embodiments of the invention, the AD-functionalized CRISPR system is used to generate lipid-rich diatoms which are useful in biofuel production.

[0974] In particular embodiments it is envisaged to specifically modify genes that are involved in the modification of the quantity of lipids and/or the quality of the lipids produced by the algal cell. Examples of genes encoding enzymes involved in the pathways of fatty acid synthesis can encode proteins having for instance acetyl-CoA carboxylase, fatty acid synthase, 3-ketoacyl_acyl- carrier protein synthase III, glycerol-3-phospate deshydrogenase (G3PDH), Enoyl-acyl carrier protein reductase (Enoyl-ACP-reductase), glycerol-3-phosphate acyltransferase, lysophosphatidic acyl transferase or diacylglycerol acyltransferase, phospholipid:diacylglycerol acyltransferase, phoshatidate phosphatase, fatty acid thioesterase such as palmitoyi protein thioesterase, or malic enzyme activities. In further embodiments it is envisaged to generate diatoms that have increased lipid accumulation. This can be achieved by targeting genes that decrease lipid catabolisation. Of particular interest for use in the methods of the present invention are genes involved in the activation of both triacylglycerol and free fatty acids, as well as genes directly involved in β-oxidation of fatty acids, such as acyl-CoA

synthetase, 3-ketoacyl-CoA thiolase, acyl-CoA oxidase activity and phosphoglucomutase. The AD-functionalized CRISPR system and methods described herein can be used to specifically activate such genes in diatoms as to increase their lipid content.

[0975] Organisms such as microalgae are widely used for synthetic biology. Stovicek et al. (Metab. Eng. Comm., 2015; 2:13 describes genome editing of industrial yeast, for example, Saccharomyces cerevisae, to efficiently produce robust strains for industrial production. Stovicek used a CRISPR-Cas9 system codon-optimized for yeast to simultaneously disrupt both alleles of an endogenous gene and knock in a heterologous gene. Cas9 and guide RNA were expressed from genomic or episomal 2μ-based vector locations. The authors also showed that gene disruption efficiency could be improved by optimization of the levels of Cas9 and guide RNA expression. Hlavová et al. (Biotechnol. Adv. 2015) discusses development of species or strains of microalgae using techniques such as CRISPR to target nuclear and chloroplast genes for insertional mutagenesis and screening.

[0976] US 8945839 describes a method for engineering Micro-Algae (Chlamydomonas reinhardtii cells) species) using Cas9 . Using similar tools, the methods of the AD-functionalized CRISPR system described herein can be applied on Chlamydomonas species and other algae. In particular embodiments, a CRISPR-Cas protein (e.g., Cas13), adenosine deaminase (which may be fused to the CRISPR-Cas protein or an aptamer-binding adaptor protein), and guide RNA are introduced in algae expressed using a vector that expresses the CRISPR-Cas protein and optionally the adenosine deaminase under the control of a constitutive promoter such as Hsp70A-Rbc S2 or Beta2 -tubulin. Guide RNA will be delivered using a vector containing T7 promoter. Alternatively, mRNA and in vitro transcribed guide RNA can be delivered to algal cells. Electroporation protocol follows standard recommended protocol from the GeneArt Chlamydomonas Engineering kit.

The use of AD-functionalized Compositions in the generation of micro-organisms capable of fatty acid production

[0977] In particular embodiments, the methods of the invention are used for the generation of genetically engineered micro-organisms capable of the production of fatty esters, such as fatty acid methyl esters ("FAME") and fatty acid ethyl esters ("FAEE"),

[0978] Typically, host cells can be engineered to produce fatty esters from a carbon source, such as an alcohol, present in the medium, by expression or overexpression of a gene encoding a thioesterase, a gene encoding an acyl-CoA synthase, and a gene encoding an ester synthase. Accordingly, the methods provided herein are used to modify a micro-organisms so as to overexpress or introduce a thioesterase gene, a gene encloding an acyl-CoA synthase, and a

gene encoding an ester synthase. In particular embodiments, the thioesterase gene is selected from tesA, 'tesA, tesB,fatB, fatB2,fatB3,fatAl, or fatA. In particular embodiments, the gene encoding an acyl-CoA synthase is selected from fadDJadK, BH3103, pfl-4354, EAV15023, fadDl, fadD2, RPC_4074,fadDD35, fadDD22, faa39, or an identified gene encoding an enzyme having the same properties. In particular embodiments, the gene encoding an ester synthase is a gene encoding a synthase/acyl-CoA:diacylglycerl acyltransferase from Simmondsia chinensis, Acinetobacter sp. ADP , Alcanivorax borkumensis, Pseudomonas aeruginosa, Fundibacter jadensis, Arabidopsis thaliana, or Alkaligenes eutrophus, or a variant thereof. Additionally or alternatively, the methods provided herein are used to decrease expression in said micro-organism of of at least one of a gene encoding an acyl-CoA dehydrogenase, a gene encoding an outer membrane protein receptor, and a gene encoding a transcriptional regulator of fatty acid biosynthesis. In particular embodiments one or more of these genes is inactivated, such as by introduction of a mutation. In particular embodiments, the gene encoding an acyl-CoA dehydrogenase is fadE. In particular embodiments, the gene encoding a transcriptional regulator of fatty acid biosynthesis encodes a DNA transcription repressor, for example, fabR.

[0979] Additionally or alternatively, said micro-organism is modified to reduce expression of at least one of a gene encoding a pyruvate formate lyase, a gene encoding a lactate dehydrogenase, or both. In particular embodiments, the gene encoding a pyruvate formate lyase is pflB. In particular embodiments, the gene encoding a lactate dehydrogenase is IdhA. In particular embodiments one or more of these genes is inactivated, such as by introduction of a mutation therein.

[0980] In particular embodiments, the micro-organism is selected from the genus *Escherichia, Bacillus, Lactobacillus, Rhodococcus, Synechococcus, Synechoystis, Pseudomonas, Aspergillus, Trichoderma, Neurospora, Fusarium, Humicola, Rhizomucor, Kluyveromyces, Pichia, Mucor, Myceliophtora, Penicillium, Phanerochaete, Pleurotus, Trametes, Chrysosporium, Saccharomyces, Stenotrophamonas, Schizosaccharomyces, Yarrowia, or Streptomyces.*

The use of AD-functionalized CRISPR system in the generation of micro-organisms capable of organic acid production

[0981] The methods provided herein are further used to engineer micro-organisms capable of organic acid production, more particularly from pentose or hexose sugars. In particular embodiments, the methods comprise introducing into a micro-organism an exogenous LDH gene. In particular embodiments, the organic acid production in said micro-organisms is

additionally or alternatively increased by inactivating endogenous genes encoding proteins involved in an endogenous metabolic pathway which produces a metabolite other than the organic acid of interest and/or wherein the endogenous metabolic pathway consumes the organic acid. In particular embodiments, the modification ensures that the production of the metabolite other than the organic acid of interest is reduced. According to particular embodiments, the methods are used to introduce at least one engineered gene deletion and/or inactivation of an endogenous pathway in which the organic acid is consumed or a gene encoding a product involved in an endogenous pathway which produces a metabolite other than the organic acid of interest. In particular embodiments, the at least one engineered gene deletion or inactivation is in one or more gene encoding an enzyme selected from the group consisting of pyruvate decarboxylase (pdc), fumarate reductase, alcohol dehydrogenase (adh), acetaldehyde dehydrogenase, phosphoenolpyruvate carboxylase (ppc), D-lactate dehydrogenase (d-ldh), L-lactate dehydrogenase (l-ldh), lactate 2-monooxygenase. In further embodiments the at least one engineered gene deletion and/or inactivation is in an endogenous gene encoding pyruvate decarboxylase (pdc).

[0982] In further embodiments, the micro-organism is engineered to produce lactic acid and the at least one engineered gene deletion and/or inactivation is in an endogenous gene encoding lactate dehydrogenase. Additionally or alternatively, the micro-organism comprises at least one engineered gene deletion or inactivation of an endogenous gene encoding a cytochrome-dependent lactate dehydrogenase, such as a cytochrome B2-dependent L-lactate dehydrogenase.

The use of AD-functionalized CRISPR system in the generation of improved xylose or cellobiose utilizing yeasts strains

[0983] In particular embodiments, the AD-functionalized CRISPR system may be applied to select for improved xylose or cellobiose utilizing yeast strains. Error-prone PCR can be used to amplify one (or more) genes involved in the xylose utilization or cellobiose utilization pathways. Examples of genes involved in xylose utilization pathways and cellobiose utilization pathways may include, without limitation, those described in Ha, S.J., et al. (2011) Proc. Natl. Acad. Sci. USA 108(2):504-9 and Galazka, J.M., et al. (2010) Science 330(6000):84-6. Resulting libraries of double-stranded DNA molecules, each comprising a random mutation in such a selected gene could be co-transformed with the components of the AD-functionalized CRISPR system into a yeast strain (for instance S288C) and strains can be selected with enhanced xylose or cellobiose utilization capacity, as described in WO2015138855.

The use of AD-functionalized CRISPR system in the generation of improved yeasts strains for use in isoprenoid biosynthesis

[0984] Tadas Jakočiūnas et al. described the successful application of a multiplex CRISPR-Cas9 system for genome engineering of up to 5 different genomic loci in one transformation step in baker's yeast Saccharomyces cerevisiae (Metabolic Engineering Volume 28, March 2015, Pages 213–222) resulting in strains with high mevalonate production, a key intermediate for the industrially important isoprenoid biosynthesis pathway. In particular embodiments, the AD-functionalized CRISPR system may be applied in a multiplex genome engineering method as described herein for identifying additional high producing yeast strains for use in isoprenoid synthesis.

Improved plants and yeast cells

[0985] The present invention also provides plants and yeast cells obtainable and obtained by the methods provided herein. The improved plants obtained by the methods described herein may be useful in food or feed production through expression of genes which, for instance ensure tolerance to plant pests, herbicides, drought, low or high temperatures, excessive water, etc.

[0986] The improved plants obtained by the methods described herein, especially crops and algae may be useful in food or feed production through expression of, for instance, higher protein, carbohydrate, nutrient or vitamin levels than would normally be seen in the wildtype. In this regard, improved plants, especially pulses and tubers are preferred.

[0987] Improved algae or other plants such as rape may be particularly useful in the production of vegetable oils or biofuels such as alcohols (especially methanol and ethanol), for instance. These may be engineered to express or overexpress high levels of oil or alcohols for use in the oil or biofuel industries.

[0988] The invention also provides for improved parts of a plant. Plant parts include, but are not limited to, leaves, stems, roots, tubers, seeds, endosperm, ovule, and pollen. Plant parts as envisaged herein may be viable, nonviable, regeneratable, and/or non- regeneratable.

[0989] It is also encompassed herein to provide plant cells and plants generated according to the methods of the invention. Gametes, seeds, embryos, either zygotic or somatic, progeny or hybrids of plants comprising the genetic modification, which are produced by traditional breeding methods, are also included within the scope of the present invention. Such plants may contain a heterologous or foreign DNA sequence inserted at or instead of a target sequence. Alternatively, such plants may contain only an alteration (mutation, deletion, insertion,

substitution) in one or more nucleotides. As such, such plants will only be different from their progenitor plants by the presence of the particular modification.

[0990] Thus, the invention provides a plant, animal or cell, produced by the present methods, or a progeny thereof. The progeny may be a clone of the produced plant or animal, or may result from sexual reproduction by crossing with other individuals of the same species to introgress further desirable traits into their offspring. The cell may be in vivo or ex vivo in the cases of multicellular organisms, particularly animals or plants.

[0991] The methods for genome editing using the AD-functionalized CRISPR system as described herein can be used to confer desired traits on essentially any plant, algae, fungus, yeast, etc. A wide variety of plants, algae, fungus, yeast, etc and plant algae, fungus, yeast cell or tissue systems may be engineered for the desired physiological and agronomic characteristics described herein using the nucleic acid constructs of the present disclosure and the various transformation methods mentioned above.

[0992] In particular embodiments, the methods described herein are used to modify endogenous genes or to modify their expression without the permanent introduction into the genome of the plant, algae, fungus, yeast, etc of any foreign gene, including those encoding CRISPR components, so as to avoid the presence of foreign DNA in the genome of the plant. This can be of interest as the regulatory requirements for non-transgenic plants are less rigorous.

[0993] The methods described herein generally result in the generation of "improved plants, algae, fungi, yeast, etc" in that they have one or more desirable traits compared to the wildtype plant. In particular embodiments, non-transgenic genetically modified plants, algae, fungi, yeast, etc., parts or cells are obtained, in that no exogenous DNA sequence is incorporated into the genome of any of the cells of the plant. In such embodiments, the improved plants, algae, fungi, yeast, etc. are non-transgenic. Where only the modification of an endogenous gene is ensured and no foreign genes are introduced or maintained in the plant, algae, fungi, yeast, etc. genome, the resulting genetically modified crops contain no foreign genes and can thus basically be considered non-transgenic. The different applications of the AD-functionalized CRISPR system for plant, algae, fungi, yeast, etc. genome editing include, but are not limited to: editing of endogenous genes to confer an agricultural trait of interest. Examplary genes conferring agronomic traits include, but are not limited to genes that confer resistance to pests or diseases; genes involved in plant diseases, such as those listed in WO 2013046247; genes that confer resistance to herbicides, fungicides, or the like; genes involved in (abiotic) stress tolerance. Other aspects of the use of the CRISPR-Cas system include, but

are not limited to: create (male) sterile plants; increasing the fertility stage in plants/algae etc; generate genetic variation in a crop of interest; affect fruit-ripening; increasing storage life of plants/algae etc; reducing allergen in plants/algae etc; ensure a value added trait (e.g. nutritional improvement); Sscreening methods for endogenous genes of interest; biofuel, fatty acid, organic acid, etc production.

AD-Functionalized Compositions Can Be Used In Non-Human Organisms

[0994] In an aspect, the invention provides a non-human eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. In other aspects, the invention provides a eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. The organism in some embodiments of these aspects may be an animal; for example a mammal. Also, the organism may be an arthropod such as an insect. The present invention may also be extended to other agricultural applications such as, for example, farm and production animals. For example, pigs have many features that make them attractive as biomedical models, especially in regenerative medicine. In particular, pigs with severe combined immunodeficiency (SCID) may provide useful models for regenerative medicine, xenotransplantation (discussed also elsewhere herein), and tumor development and will aid in developing therapies for human SCID patients. Lee et al., (Proc Natl Acad Sci U S A. 2014 May 20;111(20):7260-5) utilized a reporter-guided transcription activator-like effector nuclease (TALEN) system to generated targeted modifications of recombination activating gene (RAG) 2 in somatic cells at high efficiency, including some that affected both alleles. The AD-functionalized CRISPR system may be applied to a similar system.

[0995] The methods of Lee et al., (Proc Natl Acad Sci U S A. 2014 May 20;111(20):7260-5) may be applied to the present invention analogously as follows. Mutated pigs are produced by targeted modification of RAG2 in fetal fibroblast cells followed by SCNT and embryo transfer. Constructs coding for CRISPR Cas and a reporter are electroporated into fetal-derived fibroblast cells. After 48 h, transfected cells expressing the green fluorescent protein are sorted into individual wells of a 96-well plate at an estimated dilution of a single cell per well. Targeted modification of RAG2 are screened by amplifying a genomic DNA fragment flanking any CRISPR Cas cutting sites followed by sequencing the PCR products. After screening and ensuring lack of off-site mutations, cells carrying targeted modification of RAG2 are used for SCNT. The polar body, along with a portion of the adjacent cytoplasm of oocyte, presumably containing the metaphase II plate, are removed, and a donor cell are placed in the perivitelline. The reconstructed embryos are then electrically porated to fuse the donor cell with the oocyte

and then chemically activated. The activated embryos are incubated in Porcine Zygote Medium 3 (PZM3) with 0.5 μM Scriptaid (S7817; Sigma-Aldrich) for 14–16 h. Embryos are then washed to remove the Scriptaid and cultured in PZM3 until they were transferred into the oviducts of surrogate pigs.

[0996] The present invention is used to create a platform to model a disease or disorder of an animal, in some embodiments a mammal, in some embodiments a human. In certain embodiments, such models and platforms are rodent based, in non-limiting examples rat or mouse. Such models and platforms can take advantage of distinctions among and comparisons between inbred rodent strains. In certain embodiments, such models and platforms primate, horse, cattle, sheep, goat, swine, dog, cat or bird-based, for example to directly model diseases and disorders of such animals or to create modified and/or improved lines of such animals. Advantageously, in certain embodiments, an animal based platform or model is created to mimic a human disease or disorder. For example, the similarities of swine to humans make swine an ideal platform for modeling human diseases. Compared to rodent models, development of swine models has been costly and time intensive. On the other hand, swine and other animals are much more similar to humans genetically, anatomically, physiologically and pathophysiologically. The present invention provides a high efficiency platform for targeted gene and genome editing, gene and genome modification and gene and genome regulation to be used in such animal platforms and models. Though ethical standards block development of human models and in many cases models based on non-human primates, the present invention is used with *in vitro* systems, including but not limited to cell culture systems, three dimensional models and systems, and organoids to mimic, model, and investigate genetics, anatomy, physiology and pathophysiology of structures, organs, and systems of humans. The platforms and models provide manipulation of single or multiple targets.

[0997] In certain embodiments, the present invention is applicable to disease models like that of Schomberg et al. (FASEB Journal, April 2016; 30(1):Suppl 571.1). To model the inherited disease neurofibromatosis type 1 (NF-1) Schomberg used CRISPR-Cas9 to introduce mutations in the swine neurofibromin 1 gene by cytosolic microinjection of CRISPR/Cas9 components into swine embryos. CRISPR guide RNAs (gRNA) were created for regions targeting sites both upstream and downstream of an exon within the gene for targeted cleavage by Cas9 and repair was mediated by a specific single-stranded oligodeoxynucleotide (ssODN) template to introduce a 2500 bp deletion. The CRISPR-Cas system was also used to engineer swine with specific NF-1 mutations or clusters of mutations, and futher can be used to engineer mutations that are specific to or representative of a given human individual. The invention is

similarly used to develop animal models, including but not limited to swine models, of human multigenic diseases. According to the invention, multiple genetic loci in one gene or in multiple genes are simultaneously targeted using multiplexed guides and optionally one or multiple templates.

[0998] The present invention is also applicable to modifying SNPs of other animals, such as cows. Tan et al. (Proc Natl Acad Sci U S A. 2013 Oct 8; 110(41): 16526–16531) expanded the livestock gene editing toolbox to include transcription activator-like (TAL) effector nuclease (TALEN)- and clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9- stimulated homology-directed repair (HDR) using plasmid, rAAV, and oligonucleotide templates. Gene specific guide RNA sequences were cloned into the Church lab guide RNA vector (Addgene ID: 41824) according to their methods (Mali P, et al. (2013) RNA-Guided Human Genome Engineering via Cas9. Science 339(6121):823-826). The Cas9 nuclease was provided either by co-transfection of the hCas9 plasmid (Addgene ID: 41815) or mRNA synthesized from RCIScript-hCas9. This RCIScript-hCas9 was constructed by sub-cloning the XbaI-AgeI fragment from the hCas9 plasmid (encompassing the hCas9 cDNA) into the RCIScript plasmid.

[0999] Heo et al. (Stem Cells Dev. 2015 Feb 1;24(3):393-402. doi: 10.1089/scd.2014.0278. Epub 2014 Nov 3) reported highly efficient gene targeting in the bovine genome using bovine pluripotent cells and clustered regularly interspaced short palindromic repeat (CRISPR)/Cas9 nuclease. First, Heo et al. generate induced pluripotent stem cells (iPSCs) from bovine somatic fibroblasts by the ectopic expression of yamanaka factors and GSK3β and MEK inhibitor (2i) treatment. Heo et al. observed that these bovine iPSCs are highly similar to naïve pluripotent stem cells with regard to gene expression and developmental potential in teratomas. Moreover, CRISPR-Cas9 nuclease, which was specific for the bovine NANOG locus, showed highly efficient editing of the bovine genome in bovine iPSCs and embryos.

[01000] Igenity® provides a profile analysis of animals, such as cows, to perform and transmit traits of economic traits of economic importance, such as carcass composition, carcass quality, maternal and reproductive traits and average daily gain. The analysis of a comprehensive Igenity® profile begins with the discovery of DNA markers (most often single nucleotide polymorphisms or SNPs). All the markers behind the Igenity® profile were discovered by independent scientists at research institutions, including universities, research organizations, and government entities such as USDA. Markers are then analyzed at Igenity® in validation populations. Igenity® uses multiple resource populations that represent various production environments and biological types, often working with industry partners from the

seedstock, cow-calf, feedlot and/or packing segments of the beef industry to collect phenotypes that are not commonly available. Cattle genome databases are widely available, see, e.g., the NAGRP Cattle Genome Coordination Program (http://www.animalgenome.org/cattle/maps/db.html). Thus, the present invention maybe applied to target bovine SNPs. One of skill in the art may utilize the above protocols for targeting SNPs and apply them to bovine SNPs as described, for example, by Tan et al. or Heo et al.

[01001] Qingjian Zou et al. (Journal of Molecular Cell Biology Advance Access published October 12, 2015) demonstrated increased muscle mass in dogs by targeting targeting the first exon of the dog Myostatin (MSTN) gene (a negative regulator of skeletal muscle mass). First, the efficiency of the sgRNA was validated, using cotransfection of the the sgRNA targeting MSTN with a Cas9 vector into canine embryonic fibroblasts (CEFs). Thereafter, MSTN KO dogs were generated by micro-injecting embryos with normal morphology with a mixture of Cas9 mRNA and MSTN sgRNA and auto-transplantation of the zygotes into the oviduct of the same female dog. The knock-out puppies displayed an obvious muscular phenotype on thighs compared with its wild-type littermate sister. This can also be performed using the AD-functionalized CRISPR systems provided herein.

Livestock - Pigs

[01002] Viral targets in livestock may include, in some embodiments, porcine CD163, for example on porcine macrophages. CD163 is associated with infection (thought to be through viral cell entry) by PRRSv (Porcine Reproductive and Respiratory Syndrome virus, an arterivirus). Infection by PRRSv, especially of porcine alveolar macrophages (found in the lung), results in a previously incurable porcine syndrome ("Mystery swine disease" or "blue ear disease") that causes suffering, including reproductive failure, weight loss and high mortality rates in domestic pigs. Opportunistic infections, such as enzootic pneumonia, meningitis and ear oedema, are often seen due to immune deficiency through loss of macrophage activity. It also has significant economic and environmental repercussions due to increased antibiotic use and financial loss (an estimated $660m per year).

[01003] As reported by Kristin M Whitworth and Dr Randall Prather et al. (Nature Biotech 3434 published online 07 December 2015) at the University of Missouri and in collaboration with Genus Plc, CD163 was targeted using CRISPR-Cas9 and the offspring of edited pigs were resistant when exposed to PRRSv. One founder male and one founder female, both of whom had mutations in exon 7 of CD163, were bred to produce offspring. The founder male possessed an 11-bp deletion in exon 7 on one allele, which results in a frameshift mutation and missense

translation at amino acid 45 in domain 5 and a subsequent premature stop codon at amino acid 64. The other allele had a 2-bp addition in exon 7 and a 377-bp deletion in the preceding intron, which were predicted to result in the expression of the first 49 amino acids of domain 5, followed by a premature stop code at amino acid 85. The sow had a 7 bp addition in one allele that when translated was predicted to express the first 48 amino acids of domain 5, followed by a premature stop codon at amino acid 70. The sow's other allele was unamplifiable. Selected offspring were predicted to be a null animal (CD163–/–), i.e. a CD163 knock out.

[01004] Accordingly, in some embodiments, porcine alveolar macrophages may be targeted by the CRISPR protein. In some embodiments, porcine CD163 may be targeted by the CRISPR protein. In some embodiments, porcine CD163 may be knocked out through induction of a DSB or through insertions or deletions, for example targeting deletion or modification of exon 7, including one or more of those described above, or in other regions of the gene, for example deletion or modification of exon 5.

[01005] An edited pig and its progeny are also envisaged, for example a CD163 knock out pig. This may be for livestock, breeding or modelling purposes (i.e. a porcine model). Semen comprising the gene knock out is also provided.

[01006] CD163 is a member of the scavenger receptor cysteine-rich (SRCR) superfamily. Based on in vitro studies SRCR domain 5 of the protein is the domain responsible for unpackaging and release of the viral genome. As such, other members of the SRCR superfamily may also be targeted in order to assess resistance to other viruses. PRRSV is also a member of the mammalian arterivirus group, which also includes murine lactate dehydrogenase-elevating virus, simian hemorrhagic fever virus and equine arteritis virus. The arteriviruses share important pathogenesis properties, including macrophage tropism and the capacity to cause both severe disease and persistent infection. Accordingly, arteriviruses, and in particular murine lactate dehydrogenase-elevating virus, simian hemorrhagic fever virus and equine arteritis virus, may be targeted, for example through porcine CD163 or homologues thereof in other species, and murine, simian and equine models and knockout also provided.

[01007] Indeed, this approach may be extended to viruses or bacteria that cause other livestock diseases that may be transmitted to humans, such as Swine Influenza Virus (SIV) strains which include influenza C and the subtypes of influenza A known as H1N1, H1N2, H2N1, H3N1, H3N2, and H2N3, as well as pneumonia, meningitis and oedema mentioned above.

[01008] In some embodiments, the AD-functionalized CRISPR system described herein can be used to genetically modify a pig genome to inactivate one or more porcine endogenous

retrovirus (PERVs) loci to facilitate clinical application of porcine-to-human xenotransplantation. See Yang et al., Science 350(6264):1101-1104 (2015), which is incorporated herein by reference in its entirey. In some embodiments, the AD-functionalized CRISPR system described herein can be used to produce a genetically modified pig that does not comprise any active porcine endogenous retrovirus (PERVs) locus.

Therapeutic Targeting with AD-Functionalized Compositions

As will be apparent, it is envisaged that AD-functionalized CRISPR system can be used to target any polynucleotide sequence of interest. The invention provides a non-naturally occurring or engineered composition, or one or more polynucleotides encoding components of said composition, or vector or delivery systems comprising one or more polynucleotides encoding components of said composition for use in a modifying a target cell *in vivo*, *ex vivo* or *in vitro* and, may be conducted in a manner alters the cell such that once modified the progeny or cell line of the CRISPR modified cell retains the altered phenotype. The modified cells and progeny may be part of a multi-cellular organism such as a plant or animal with *ex vivo* or *in vivo* application of CRISPR system to desired cell types. The CRISPR invention may be a therapeutic method of treatment. The therapeutic method of treatment may comprise gene or genome editing, or gene therapy. Additional diseases that may be treated using the compositions and methods of the present invention are are further disclosed in ClinVar database (Landrum et al., Nucleic Acids Res. 2016 Jan 4;44(D1):D862-8; Landrum et al., Nucleic Acids Res. 2014 Jan 1;42(1):D980-5; http://www.ncbi.nlm.nih.gov/books/NBK174587/).

Adoptive Cell Therapies

[01009] The present invention also contemplates use of the AD-functionalized CRISPR system described herein tomodify cells for adoptive therapies. Aspects of the invention accordingly involve the adoptive transfer of immune system cells, such as T cells, specific for selected antigens, such as tumor associated antigens (see Maus et al., 2014, Adoptive Immunotherapy for Cancer or Viruses, Annual Review of Immunology, Vol. 32: 189-225; Rosenberg and Restifo, 2015, Adoptive cell transfer as personalized immunotherapy for human cancer, Science Vol. 348 no. 6230 pp. 62-68; and, Restifo et al., 2015, Adoptive immunotherapy for cancer: harnessing the T cell response. Nat. Rev. Immunol. 12(4): 269-281; and Jenson and Riddell, 2014, Design and implementation of adoptive therapy with chimeric antigen receptor-modified T cells. Immunol Rev. 257(1): 127–144). Various strategies may for example be employed to genetically modify T cells by altering the specificity of the T cell receptor (TCR) for example by introducing new TCR α and β chains with selected

peptide specificity (see U.S. Patent No. 8,697,854; PCT Patent Publications: WO2003020763, WO2004033685, WO2004044004, WO2005114215, WO2006000830, WO2008038002, WO2008039818, WO2004074322, WO2005113595, WO2006125962, WO2013166321, WO2013039889, WO2014018863, WO2014083173; U.S. Patent No. 8,088,379).

[01010] As an alternative to, or addition to, TCR modifications, chimeric antigen receptors (CARs) may be used in order to generate immunoresponsive cells, such as T cells, specific for selected targets, such as malignant cells, with a wide variety of receptor chimera constructs having been described (see U.S. Patent Nos. 5,843,728; 5,851,828; 5,912,170; 6,004,811; 6,284,240; 6,392,013; 6,410,014; 6,753,162; 8,211,422; and, PCT Publication WO9215322). Alternative CAR constructs may be characterized as belonging to successive generations. First-generation CARs typically consist of a single-chain variable fragment of an antibody specific for an antigen, for example comprising a VL linked to a VH of a specific antibody, linked by a flexible linker, for example by a CD8α hinge domain and a CD8α transmembrane domain, to the transmembrane and intracellular signaling domains of either CD3ζ or FcRγ (scFv-CD3ζ or scFv-FcRγ; see U.S. Patent No. 7,741,465; U.S. Patent No. 5,912,172; U.S. Patent No. 5,906,936). Second-generation CARs incorporate the intracellular domains of one or more costimulatory molecules, such as CD28, OX40 (CD134), or 4-1BB (CD137) within the endodomain (for example scFv-CD28/OX40/4-1BB-CD3ζ; see U.S. Patent Nos. 8,911,993; 8,916,381; 8,975,071; 9,101,584; 9,102,760; 9,102,761). Third-generation CARs include a combination of costimulatory endodomains, such a CD3ζ-chain, CD97, GDI la-CD18, CD2, ICOS, CD27, CD154, CDS, OX40, 4-1BB, or CD28 signaling domains (for example scFv-CD28-4-1BB-CD3ζ or scFv-CD28-OX40-CD3ζ; see U.S. Patent No. 8,906,682; U.S. Patent No. 8,399,645; U.S. Pat. No. 5,686,281; PCT Publication No. WO2014134165; PCT Publication No. WO2012079000). Alternatively, costimulation may be orchestrated by expressing CARs in antigen-specific T cells, chosen so as to be activated and expanded following engagement of their native αβTCR, for example by antigen on professional antigen-presenting cells, with attendant costimulation. In addition, additional engineered receptors may be provided on the immunoresponsive cells, for example to improve targeting of a T-cell attack and/or minimize side effects.

[01011] Alternative techniques may be used to transform target immunoresponsive cells, such as protoplast fusion, lipofection, transfection or electroporation. A wide variety of vectors may be used, such as retroviral vectors, lentiviral vectors, adenoviral vectors, adeno-associated viral vectors, plasmids or transposons, such as a Sleeping Beauty transposon (see U.S. Patent Nos. 6,489,458; 7,148,203; 7,160,682; 7,985,739; 8,227,432), may be used to introduce CARs,

for example using 2nd generation antigen-specific CARs signaling through CD3ζ and either CD28 or CD137. Viral vectors may for example include vectors based on HIV, SV40, EBV, HSV or BPV.

[01012] Cells that are targeted for transformation may for example include T cells, Natural Killer (NK) cells, cytotoxic T lymphocytes (CTL), regulatory T cells, human embryonic stem cells, tumor-infiltrating lymphocytes (TIL) or a pluripotent stem cell from which lymphoid cells may be differentiated. T cells expressing a desired CAR may for example be selected through co-culture with γ-irradiated activating and propagating cells (AaPC), which co-express the cancer antigen and co-stimulatory molecules. The engineered CAR T-cells may be expanded, for example by co-culture on AaPC in presence of soluble factors, such as IL-2 and IL-21. This expansion may for example be carried out so as to provide memory CAR+ T cells (which may for example be assayed by non-enzymatic digital array and/or multi-panel flow cytometry). In this way, CAR T cells may be provided that have specific cytotoxic activity against antigen-bearing tumors (optionally in conjunction with production of desired chemokines such as interferon-γ). CAR T cells of this kind may for example be used in animal models, for example to threat tumor xenografts.

[01013] Approaches such as the foregoing may be adapted to provide methods of treating and/or increasing survival of a subject having a disease, such as a neoplasia, for example by administering an effective amount of an immunoresponsive cell comprising an antigen recognizing receptor that binds a selected antigen, wherein the binding activates the immunoreponsive cell, thereby treating or preventing the disease (such as a neoplasia, a pathogen infection, an autoimmune disorder, or an allogeneic transplant reaction). Dosing in CAR T cell therapies may for example involve administration of from 106 to 109 cells/kg, with or without a course of lymphodepletion, for example with cyclophosphamide.

[01014] In one embodiment, the treatment can be administrated into patients undergoing an immunosuppressive treatment. The cells or population of cells, may be made resistant to at least one immunosuppressive agent due to the inactivation of a gene encoding a receptor for such immunosuppressive agent. Not being bound by a theory, the immunosuppressive treatment should help the selection and expansion of the immunoresponsive or T cells according to the invention within the patient.

[01015] The administration of the cells or population of cells according to the present invention may be carried out in any convenient manner, including by aerosol inhalation, injection, ingestion, transfusion, implantation or transplantation. The cells or population of cells may be administered to a patient subcutaneously, intradermally, intratumorally,

intranodally, intramedullary, intramuscularly, by intravenous or intralymphatic injection, or intraperitoneally. In one embodiment, the cell compositions of the present invention are preferably administered by intravenous injection.

[01016] The administration of the cells or population of cells can consist of the administration of 104- 109 cells per kg body weight, preferably 105 to 106 cells/kg body weight including all integer values of cell numbers within those ranges. Dosing in CAR T cell therapies may for example involve administration of from 106 to 109 cells/kg, with or without a course of lymphodepletion, for example with cyclophosphamide. The cells or population of cells can be administrated in one or more doses. In another embodiment, the effective amount of cells are administrated as a single dose. In another embodiment, the effective amount of cells are administrated as more than one dose over a period time. Timing of administration is within the judgment of managing physician and depends on the clinical condition of the patient. The cells or population of cells may be obtained from any source, such as a blood bank or a donor. While individual needs vary, determination of optimal ranges of effective amounts of a given cell type for a particular disease or conditions are within the skill of one in the art. An effective amount means an amount which provides a therapeutic or prophylactic benefit. The dosage administrated will be dependent upon the age, health and weight of the recipient, kind of concurrent treatment, if any, frequency of treatment and the nature of the effect desired.

[01017] In another embodiment, the effective amount of cells or composition comprising those cells are administrated parenterally. The administration can be an intravenous administration. The administration can be directly done by injection within a tumor.

[01018] To guard against possible adverse reactions, engineered immunoresponsive cells may be equipped with a transgenic safety switch, in the form of a transgene that renders the cells vulnerable to exposure to a specific signal. For example, the herpes simplex viral thymidine kinase (TK) gene may be used in this way, for example by introduction into allogeneic T lymphocytes used as donor lymphocyte infusions following stem cell transplantation (Greco, et al., Improving the safety of cell therapy with the TK-suicide gene. Front. Pharmacol. 2015; 6: 95). In such cells, administration of a nucleoside prodrug such as ganciclovir or acyclovir causes cell death. Alternative safety switch constructs include inducible caspase 9, for example triggered by administration of a small-molecule dimerizer that brings together two nonfunctional icasp9 molecules to form the active enzyme. A wide variety of alternative approaches to implementing cellular proliferation controls have been described (see U.S. Patent Publication No. 20130071414; PCT Patent Publication WO2011146862; PCT Patent Publication WO2014011987; PCT Patent Publication

WO2013040371; Zhou et al. BLOOD, 2014, 123/25:3895 – 3905; Di Stasi et al., The New England Journal of Medicine 2011; 365:1673-1683; Sadelain M, The New England Journal of Medicine 2011; 365:1735-173; Ramos et al., Stem Cells 28(6):1107-15 (2010)).

[01019] In a further refinement of adoptive therapies, genome editing with a AD-functionalized CRISPR-Cas system as described herein may be used to tailor immunoresponsive cells to alternative implementations, for example providing edited CAR T cells (see Poirot et al., 2015, Multiplex genome edited T-cell manufacturing platform for "off-the-shelf" adoptive T-cell immunotherapies, Cancer Res 75 (18): 3853). For example, immunoresponsive cells may be edited to delete expression of some or all of the class of HLA type II and/or type I molecules, or to knockout selected genes that may inhibit the desired immune response, such as the PD1 gene.

[01020] Cells may be edited using a AD-functionalized CRISPR system as described herein. AD-functionalized CRISPR systems may be delivered to an immune cell by any method described herein. In preferred embodiments, cells are edited ex vivo and transferred to a subject in need thereof. Immunoresponsive cells, CAR-T cells or any cells used for adoptive cell transfer may be edited. Editing may be performed to eliminate potential alloreactive T-cell receptors (TCR), disrupt the target of a chemotherapeutic agent, block an immune checkpoint, activate a T cell, and/or increase the differentiation and/or proliferation of functionally exhausted or dysfunctional CD8+ T-cells (see PCT Patent Publications: WO2013176915, WO2014059173, WO2014172606, WO2014184744, and WO2014191128). Editing may result in inactivation of a gene.

[01021] T cell receptors (TCR) are cell surface receptors that participate in the activation of T cells in response to the presentation of antigen. The TCR is generally made from two chains, α and β, which assemble to form a heterodimer and associates with the CD3-transducing subunits to form the T cell receptor complex present on the cell surface. Each α and β chain of the TCR consists of an immunoglobulin-like N-terminal variable (V) and constant (C) region, a hydrophobic transmembrane domain, and a short cytoplasmic region. As for immunoglobulin molecules, the variable region of the α and β chains are generated by V(D)J recombination, creating a large diversity of antigen specificities within the population of T cells. However, in contrast to immunoglobulins that recognize intact antigen, T cells are activated by processed peptide fragments in association with an MHC molecule, introducing an extra dimension to antigen recognition by T cells, known as MHC restriction. Recognition of MHC disparities between the donor and recipient through the T cell receptor leads to T cell proliferation and the potential development of graft versus host disease (GVHD). The inactivation of TCRα or TCRβ

can result in the elimination of the TCR from the surface of T cells preventing recognition of alloantigen and thus GVHD. However, TCR disruption generally results in the elimination of the CD3 signaling component and alters the means of further T cell expansion.

[01022] Allogeneic cells are rapidly rejected by the host immune system. It has been demonstrated that, allogeneic leukocytes present in non-irradiated blood products will persist for no more than 5 to 6 days (Boni, Muranski et al. 2008 Blood 1;112(12):4746-54). Thus, to prevent rejection of allogeneic cells, the host's immune system usually has to be suppressed to some extent. However, in the case of adoptive cell transfer the use of immunosuppressive drugs also have a detrimental effect on the introduced therapeutic T cells. Therefore, to effectively use an adoptive immunotherapy approach in these conditions, the introduced cells would need to be resistant to the immunosuppressive treatment. Thus, in a particular embodiment, the present invention further comprises a step of modifying T cells to make them resistant to an immunosuppressive agent, preferably by inactivating at least one gene encoding a target for an immunosuppressive agent. An immunosuppressive agent is an agent that suppresses immune function by one of several mechanisms of action. An immunosuppressive agent can be, but is not limited to a calcineurin inhibitor, a target of rapamycin, an interleukin-2 receptor α-chain blocker, an inhibitor of inosine monophosphate dehydrogenase, an inhibitor of dihydrofolic acid reductase, a corticosteroid or an immunosuppressive antimetabolite. The present invention allows conferring immunosuppressive resistance to T cells for immunotherapy by inactivating the target of the immunosuppressive agent in T cells. As non-limiting examples, targets for an immunosuppressive agent can be a receptor for an immunosuppressive agent such as: CD52, glucocorticoid receptor (GR), a FKBP family gene member and a cyclophilin family gene member.

[01023] Immune checkpoints are inhibitory pathways that slow down or stop immune reactions and prevent excessive tissue damage from uncontrolled activity of immune cells. In certain embodiments, the immune checkpoint targeted is the programmed death-1 (PD-1 or CD279) gene (PDCD1). In other embodiments, the immune checkpoint targeted is cytotoxic T-lymphocyte-associated antigen (CTLA-4). In additional embodiments, the immune checkpoint targeted is another member of the CD28 and CTLA4 Ig superfamily such as BTLA, LAG3, ICOS, PDL1 or KIR. In further additional embodiments, the immune checkpoint targeted is a member of the TNFR superfamily such as CD40, OX40, CD137, GITR, CD27 or TIM-3.

[01024] Additional immune checkpoints include Src homology 2 domain-containing protein tyrosine phosphatase 1 (SHP-1) (Watson HA, et al., SHP-1: the next checkpoint target for

cancer immunotherapy? Biochem Soc Trans. 2016 Apr 15;44(2):356-62). SHP-1 is a widely expressed inhibitory protein tyrosine phosphatase (PTP). In T-cells, it is a negative regulator of antigen-dependent activation and proliferation. It is a cytosolic protein, and therefore not amenable to antibody-mediated therapies, but its role in activation and proliferation makes it an attractive target for genetic manipulation in adoptive transfer strategies, such as chimeric antigen receptor (CAR) T cells. Immune checkpoints may also include T cell immunoreceptor with Ig and ITIM domains (TIGIT/Vstm3/WUCAM/VSIG9) and VISTA (Le Mercier I, et al., (2015) Beyond CTLA-4 and PD-1, the generation Z of negative checkpoint regulators. Front. Immunol. 6:418).

[01025] WO2014172606 relates to the use of MT1 and/or MT1 inhibitors to increase proliferation and/or activity of exhausted CD8+ T-cells and to decrease CD8+ T-cell exhaustion (e.g., decrease functionally exhausted or unresponsive CD8+ immune cells). In certain embodiments, metallothioneins are targeted by gene editing in adoptively transferred T cells.

[01026] In certain embodiments, targets of gene editing may be at least one targeted locus involved in the expression of an immune checkpoint protein. Such targets may include, but are not limited to CTLA4, PPP2CA, PPP2CB, PTPN6, PTPN22, PDCD1, ICOS (CD278), PDL1, KIR, LAG3, HAVCR2, BTLA, CD160, TIGIT, CD96, CRTAM, LAIR1, SIGLEC7, SIGLEC9, CD244 (2B4), TNFRSF10B, TNFRSF10A, CASP8, CASP10, CASP3, CASP6, CASP7, FADD, FAS, TGFBRII, TGFRBRI, SMAD2, SMAD3, SMAD4, SMAD10, SKI, SKIL, TGIF1, IL10RA, IL10RB, HMOX2, IL6R, IL6ST, EIF2AK4, CSK, PAG1, SIT1, FOXP3, PRDM1, BATF, VISTA, GUCY1A2, GUCY1A3, GUCY1B2, GUCY1B3, MT1, MT2, CD40, OX40, CD137, GITR, CD27, SHP-1 or TIM-3. In preferred embodiments, the gene locus involved in the expression of PD-1 or CTLA-4 genes is targeted. In other preferred embodiments, combinations of genes are targeted, such as but not limited to PD-1 and TIGIT.

[01027] In other embodiments, at least two genes are edited. Pairs of genes may include, but are not limited to PD1 and TCRα, PD1 and TCRβ, CTLA-4 and TCRα, CTLA-4 and TCRβ, LAG3 and TCRα, LAG3 and TCRβ, Tim3 and TCRα, Tim3 and TCRβ, BTLA and TCRα, BTLA and TCRβ, BY55 and TCRα, BY55 and TCRβ, TIGIT and TCRα, TIGIT and TCRβ, B7H5 and TCRα, B7H5 and TCRβ, LAIR1 and TCRα, LAIR1 and TCRβ, SIGLEC10 and TCRα, SIGLEC10 and TCRβ, 2B4 and TCRα, 2B4 and TCRβ.

[01028] Whether prior to or after genetic modification of the T cells, the T cells can be activated and expanded generally using methods as described, for example, in U.S. Patents 6,352,694; 6,534,055; 6,905,680; 5,858,358; 6,887,466; 6,905,681; 7,144,575; 7,232,566;

7,175,843; 5,883,223; 6,905,874; 6,797,514; 6,867,041; and 7,572,631. T cells can be expanded in vitro or in vivo.

[01029] The practice of the present invention employs techniques known in the field of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA, which are within the skill of the art. See MOLECULAR CLONING: A LABORATORY MANUAL, 2nd edition (1989) (Sambrook, Fritsch and Maniatis); MOLECULAR CLONING: A LABORATORY MANUAL, 4th edition (2012) (Green and Sambrook); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (1987) (F. M. Ausubel, et al. eds.); the series METHODS IN ENZYMOLOGY (Academic Press, Inc.); PCR 2: A PRACTICAL APPROACH (1995) (M.J. MacPherson, B.D. Hames and G.R. Taylor eds.); ANTIBODIES, A LABORATORY MANUAL (1988) (Harlow and Lane, eds.); ANTIBODIES A LABORATORY MANUAL, 2nd edition (2013) (E.A. Greenfield ed.); and ANIMAL CELL CULTURE (1987) (R.I. Freshney, ed.).

Screens/Diagnostics/Treatments using CRISPR systems

Cancer

[01030] The methods and compositions of the invention can be used to identify cell states, components, and mechanisms associated with drug-tolerance and persistence of disease cells. Terai et al. (Cancer Research, 19-Dec-2017, doi: 10.1158/0008-5472.CAN-17-1904) reported a genome-wide CRISPR/Cas9 enhancer/suppressor screen in EGFR-dependent lung cancer PC9 cells treated with erlotinib + THZ1 (CDK7/12 inhibitor) combination therapy to identify multiple genes that enhanced erlotinib/THZ1 synergy, as well as components and pathways that suppress synergy. Wang et al. (Cell Rep. 2017 Feb 7;18(6):1543-1557. doi: 10.1016/j.celrep.2017.01.031.; Krall et al., Elife. 2017 Feb 1;6. pii: e18970. doi: 10.7554/eLife.18970) reported the use of genome-wide CRISPR loss-of-function screens to identify mediator of resistance to MAPK inhibitors. Donovan et al. (PLoS One. 2017 Jan 24;12(1):e0170445. doi: 10.1371/journal.pone.0170445. eCollection 2017) used a CRISPR-mediated approach to mutagenesis to identify novel gain-of-function and drug resistant alleles of the MAPK signaling pathway genes. Wang et al. (Cell. 2017 Feb 23;168(5):890-903.e15. doi: 10.1016/j.cell.2017.01.013. Epub 2017 Feb 2) used genome-wide CRISPR screens to identify gene networks and synthetic lethal interactions with oncogenic Ras. Chow et al. (Nat Neurosci. 2017 Oct;20(10):1329-1341. doi: 10.1038/nn.4620. Epub 2017 Aug 14) developed an adeno-associated virus-mediated, autochthonous genetic CRISPR screen in glioblastoma to identify functional suppressors in glioblastoma. Xue et al. (Nature. 2014 Oct

16;514(7522):380-4. doi: 10.1038/nature13589. Epub 2014 Aug 6) employed CRISPR-mediated direct mutation of cancer genes in the mouse liver.

[01031] Chen et al. (J Clin Invest. 2017 Dec 4. pii: 90793. doi: 10.1172/JCI90793. [Epub ahead of print]) used a CRISPR-based screen to identify MYCN-amplified neuroblastoma dependency on EZH2. Support testing of EZH2 inhibitors in patients with MYCN-amplified neuroblastoma.

[01032] Vijai et al. (Cancer Discov. 2016 Nov;6(11):1267-1275. Epub 2016 Sep 21) reported use of CRISPR to generate heterozygous mutations in the mammary epithelial cell line to assess risk for breast cancer.

[01033] Chakraborty et al. (Sci Transl Med. 2017 Jul 12;9(398). pii: eaal5272. doi: 10.1126/scitranslmed.aal5272) used a CRISPR-based screen to identify EZH1 as potential target to treat clear cell renal cell carcinoma

Metabolic disease

[01034] The methods and compositions of the invention provide advantages over conventional gene therapy methods in the treatment of inherited metabolic diseases of the liver, including but not limited to familial hypercholesterolemia, hemophilia, ornithine transcarbamylase deficiency, hereditary tyrosinemia type 1, and alpha-1 antitrypsin deficiency. See, Bryson et al., Yale J. Biol. Med. 90(4):553-566, 19-Dec-2017.

[01035] Bompada et al. (Int J Biochem Cell Biol. 2016 Dec;81(Pt A):82-91. doi: 10.1016/j.biocel.2016.10.022. Epub 2016 Oct 29) described the use of CRISPR to knockout histone acetyltransferase in pancreatic beta cells to demonstrate that histone acetylation serves as a key regulator of glucose-induced increase in TXNIP gene expression and thereby glucotoxicity-induced apoptosis.

Muscle

[0002] Provenzano et al. (Mol Ther Nucleic Acids. 9:337-348. 15-Dec-2017;. doi: 10.1016/j.omtn.2017.10.006. Epub 2017 Oct 14) reported CRISPR/Cas9-mediated deletion of CTG expansions and permanent reversion to a normal phenotype in myogenic cells from myotonic dystrophy 1 patients. The methods and compositions of the instant invention are similarly applicable to nucleotide repeat disorders, not limited to CTG expansions. Tabebordbar et al. (2016 Jan 22;351(6271):407-411. doi: 10.1126/science.aad5177. Epub 2015 Dec 31) reports the use of CRISPR to edit the Dmd exon 23 locus to correct disruptive mutations in DMD. Tabebordbar shows that programmable CRISPR complexes can be delivered locally and systemically to terminally differentiated skeletal muscle fibers and cardiomyocytes, as well as muscle satellite cells, in neonatal and adult mice, where they

mediate targeted gene modification, restore dystrophin expression and partially recover functional deficiencies of dystrophic muscle. See also Nelson et al., (Science. 2016 Jan 22;351(6271):403-7. doi: 10.1126/science.aad5143. Epub 2015 Dec 31).

Infectious disease

[0003] Sidik et al. (Cell. 2016 Sep 8;166(6):1423-1435.e12. doi: 10.1016/j.cell.2016.08.019. Epub 2016 Sep 2) and Patel et al. (Nature. 2017 Aug 31;548(7669):537-542. doi: 10.1038/nature23477. Epub 2017 Aug 7) describe a CRISPR screen in Toxoplasma and expansion of antiparasitic interventions.

[0004] There are several reports of genome-wide CRISPR screens to identify components and processes underlying host-pathogen interactions. Examples include Blondel et al. (Cell Host Microbe. 2016 Aug 10;20(2):226-37. doi: 10.1016/j.chom.2016.06.010. Epub 2016 Jul 21), Shapiro et al. (Nat Microbiol. 2018 Jan;3(1):73-82. doi: 10.1038/s41564-017-0043-0. Epub 2017 Oct 23) and Park et al. (Nat Genet. 2017 Feb;49(2):193-203. doi: 10.1038/ng.3741. Epub 2016 Dec 19).

[0005] Ma et al. (Cell Host Microbe. 2017 May 10;21(5):580-591.e7. doi: 10.1016/j.chom.2017.04.005) employed genome-wide CRISPR loss-of-function screens to identify viral transformation-driven synthetic lethal targets for therapeutic intervention.

Cardiovascular diseases

[0006] CRISPR systems can be used as to tool to identify genes or genetic variant associated with vascular disease. This is useful for identifying potential treatment or preventative targets. Xu et al. (Atherosclerosis, 2017 Sep. 21 pii: S0021-9150(17)31265-0. doi: 10.1016/j.atherosclerosis.2017.08.031. [Epub ahead of print]) reports the use of CRISPR to knockout the ANGPTL3 gene to confirm the role of ANGPTL3 in regulating plasma level of LDL-C. Gupta et al., (Cell. 2017 Jul 27;170(3):522-533.e15. doi: 10.1016/j.cell.2017.06.049) reports the use of CRISPR to edit stem cell-derived enthothelial cells to identify genetic variant associated with vascular diseases. Beaudoin et al., (Arterioscler Thromb Vasc Biol. 2015 Jun;35(6):1472-1479. doi: 10.1161/ATVBAHA.115.305534. Epub 2015 Apr 2), reports the use of CRISPR genome editing to disrupt binding of the transcription factors MEF2 at the locus. This sets the stage for exploring how PHACTR1 functions in the vascular endothelium influence coronary artery disease. Pashos et al. (Cell Stem Cell. 2017 Apr 6;20(4):558-570.e10. doi: 10.1016/j.stem.2017.03.017.) reports on using CRISPR technology to target pluripotent stem cells and hepatocyte-like cells to identify functional variants and lipid-functional genes.

Neurological Diseases

[0007] The invention provides methods and compositions for investigating and treating neurological diseases and disorders. Nakayama et al., (Am J Hum Genet. 2015 May 7;96(5):709-19. doi: 10.1016/j.ajhg.2015.03.003. Epub 2015 Apr 9) report use of CRISPR to study the role of PYCR2 in human CNS development and to identify potential target for microcephaly and hypomyelination. Swiech et al. (Nat Biotechnol. 2015 Jan;33(1):102-6. doi: 10.1038/nbt.3055. Epub 2014 Oct 19) report use of CRISPR to target single (Mecp2) as well as multiple genes (Dnmt1, Dnmt3a and Dnmt3b) in the adult mouse brain *in vivo*. Shin et al. (Hum Mol Genet. 2016 Oct 15;25(20):4566-4576. doi: 10.1093/hmg/ddw286) describes the use of CRISPR to inactivate Huntingon's disease mutation. Platt et al. (Cell Rep. 2017 Apr 11;19(2):335-350. doi: 10.1016/j.celrep.2017.03.052) report use of CRISPR knockin mice to identify Chd8's role in autism spectrum disorder. Seo et al. (J Neurosci. 2017 Oct 11;37(41):9917-9924. doi: 10.1523/JNEUROSCI.0621-17.2017. Epub 2017 Sep 14) describe use of CRISPR to generate models of neurodegenerative disorders. Petersen et al. (Neuron. 2017 Dec 6;96(5):1003-1012.e7. doi: 10.1016/j.neuron.2017.10.008. Epub 2017 Nov 2) demonstrate CRISPR knockout of activin A receptor type I in oligodendrocyte progenitor cells to identify potential targets for diseases with remyelination failure. The methods and compositions of the instant invention are similarly applicable.

Other applications of CRISPR technology

[0008] Renneville et al (Blood. 2015 Oct 15;126(16):1930-9. doi: 10.1182/blood-2015-06-649087. Epub 2015 Aug 28) report use of CRISPR to study the roles of EHMT1 and EMHT2 in fetal hemoglobin expression and to identify novel therapeutic target for SCD.

[0009] Tothova et al. (Cell Stem Cell. 2017 Oct 5;21(4):547-555.e8. doi: 10.1016/j.stem.2017.07.015) reported the use of CRISPR in hematopoietic stem and progenitor cells for generating models of human myeloid diseases.

[0010] Giani et al. (Cell Stem Cell. 2016 Jan 7;18(1):73-78. doi: 10.1016/j.stem.2015.09.015. Epub 2015 Oct 22) report that inactivation of SH2B3 by CRISPR/Cas9 genome editing in human pluripotent stem cells allowed enhanced erythroid cell expansion with preserved differentiation.

[0011] Wakabayashi et al. (Proc Natl Acad Sci U S A. 2016 Apr 19;113(16):4434-9. doi: 10.1073/pnas.1521754113. Epub 2016 Apr 4) employed CRISPR to gain insight into GATA1 transcriptional activity and to investigate the pathogenicity of noncoding variants in human erythroid disorders.

[0012] Mandal et al. (Cell Stem Cell. 2014 Nov 6;15(5):643-52. doi: 10.1016/j.stem.2014.10.004. Epub 2014 Nov 6) describe CRISPR/Cas9 targeting of two

clinically relevant genes, B2M and CCR5, in primary human CD4+ T cells and CD34+ hematopoietic stem and progenitor cells (HSPCs)

[0013] Polfus et al. (Am J Hum Genet. 2016 Sep 1;99(3):785. doi: 10.1016/j.ajhg.2016.08.002. Epub 2016 Sep 1) used CRISPR to edit hematopoietic cell lines and follow-up targeted knockdown experiments in primary human hematopoietic stem and progenitor cells and investigate the role of GFI1B variants in human hematopoiesis.

[0014] Najm et al. (Nat Biotechnol. 2017 Dec 18. doi: 10.1038/nbt.4048. [Epub ahead of print]) reports the use of CRISPR complex having a pair SaCas9 and SpCas9 to achieve dual targeting to generate high-complexity pooled dual-knockout libraries to identify synthetic lethal and buffering gene pairs across multiple cell types, including MAPK pathway genes and apoptotic genes.

[0015] Manguso et al. (Nature. 2017 Jul 27;547(7664):413-418. doi: 10.1038/nature23270. Epub 2017 Jul 19.) reports the use of CRISPR screens to identify and/or confirm new immunotherapy targets. See also Roland et al. (Proc Natl Acad Sci U S A. 2017 Jun 20;114(25):6581-6586. doi: 10.1073/pnas.1701263114. Epub 2017 Jun 12.); Erb et al. (Nature. 2017 Mar 9;543(7644):270-274. doi: 10.1038/nature21688. Epub 2017 Mar 1.); Hong et al., (Nat Commun. 2016 Jun 22;7:11987. doi: 10.1038/ncomms11987); Fei et al., (Proc Natl Acad Sci U S A. 2017 Jun 27;114(26):E5207-E5215. doi: 10.1073/pnas.1617467114. Epub 2017 Jun 13.); Zhang et al., (Cancer Discov. 2017 Sep 29. doi: 10.1158/2159-8290.CD-17-0532. [Epub ahead of print]).

[0016] Joung et al. (Nature. 2017 Aug 17;548(7667):343-346. doi: 10.1038/nature23451. Epub 2017 Aug 9.) reports the use of genome-wide screens to analyze long non-coding RNAs (lncRNA); see also Zhu et al., (Nat Biotechnol. 2016 Dec;34(12):1279-1286. doi: 10.1038/nbt.3715. Epub 2016 Oct 31); Sanjana et al., (Science. 2016 Sep 30;353(6307):1545-1549).

[0017] Barrow et al. (Mol Cell. 2016 Oct 6;64(1):163-175. doi: 10.1016/j.molcel.2016.08.023. Epub 2016 Sep 22.) reports the use of genome-wide CRISPR screens to search for therapeutic targets for mitochondrial diseases. See also Vafai et al., (PLoS One. 2016 Sep 13;11(9):e0162686. doi: 10.1371/journal.pone.0162686. eCollection 2016).

[0018] Guo et al. (Elife. 2017 Dec 5;6. pii: e29329. doi: 10.7554/eLife.29329) reports the use of CRISPR to target human chondrocytes to elucidate biological mechanisms for human growth.

[0019] Ramanan et al. (Sci Rep. 2015 Jun 2;5:10833. doi: 10.1038/srep10833) reports the use of CRISPR to target and cleave conserved regions in the HBV genome.

Correction of Disease-Associated Mutations and Pathogenic SNPs

[01036] In one aspect, the invention described herein provides methods for modifying an adenosine residue at a target locus with the aim of remedying and/or preventing a diseased condition that is or is likely to be caused by a G-to-A or C-to-T point mutation or a pathogenic single nucleotide polymorphism (SNP).

Diseases Affecting the Brain and Central Nervous System

[01037] Pathogenic G-to-A or C-to-T mutations/SNPs associated with various diseases affecting the brain and central nervous system are reported in the ClinVar database and disclosed in Table A, including but not limited to Alzheimer's Disease, Parkinson's Disease, Autism, Amyotrophyic lateral sclerosis (ALS), Schizophrenia, Adrenoleukodystrophy, Aicardi Goutieres syndrome, Fabry disease, Lesch-Nyhan syndrome, and Menkes Disease. Accordingly, an aspect of the invention relates to a method for correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with any of these diseases, as discussed below.

[01038] Nakayama et al., (Am J Hum Genet. 2015 May 7;96(5):709-19. doi: 10.1016/j.ajhg.2015.03.003. Epub 2015 Apr 9) report use of CRISPR to study the role of PYCR2 in human CNS development and to identify potential target for microcephaly and hypomyelination. Swiech et al. (Nat Biotechnol. 2015 Jan;33(1):102-6. doi: 10.1038/nbt.3055. Epub 2014 Oct 19) report use of CRISPR to target single (Mecp2) as well as multiple genes (Dnmt1, Dnmt3a and Dnmt3b) in the adult mouse brain *in vivo*. Shin et al. (Hum Mol Genet. 2016 Oct 15;25(20):4566-4576. doi: 10.1093/hmg/ddw286) describes the use of CRISPR to inactivate Huntingon's disease mutation.

Alzheimer's Disease

[01039] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Alzheimer's Disease. In some embodiments, the pathogenic mutations/SNPs are present in at least one gene selected from PSEN1, PSEN2, and APP, including at least the followings:

NM_000021.3(PSEN1):c.796G>A (p.Gly266Ser)

NM_000484.3(APP):c.2017G>A (p.Ala673Thr)

NM_000484.3(APP):c.2149G>A (p.Val717Ile)

NM_000484.3(APP):c.2137G>A (p.Ala713Thr)

NM_000484.3(APP):c.2143G>A (p.Val715Met)

NM_000484.3(APP):c.2141C>T (p.Thr714Ile)

NM_000021.3(PSEN1):c.438G>A (p.Met146Ile)

NM_000021.3(PSEN1):c.1229G>A (p.Cys410Tyr)

NM_000021.3(PSEN1):c.487C>T (p.His163Tyr)

NM_000021.3(PSEN1):c.799C>T (p.Pro267Ser)

NM_000021.3(PSEN1):c.236C>T (p.Ala79Val)

NM_000021.3(PSEN1):c.509C>T (p.Ser170Phe)

NM_000447.2(PSEN2):c.1289C>T (p.Thr430Met)

NM_000447.2(PSEN2):c.717G>A (p.Met239Ile)

NM_000447.2(PSEN2):c.254C>T (p.Ala85Val)

NM_000021.3(PSEN1):c.806G>A (p.Arg269His)

NM_000484.3(APP):c.2018C>T (p.Ala673Val).

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Alzheimer's Disease by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in at least one gene selected from PSEN1, PSEN2, and APP, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Parkinson's Disease

[01040] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Parkinson's Disease. In some embodiments, In some embodiment, the pathogenic mutations/SNPs are present in at least one gene selected from SNCA, PLA2G6, FBXO7, VPS35, EIF4G1, DNAJC6, PRKN, SYNJ1, CHCHD2, PINK1, PARK7, LRRK2, ATP13A2, and GBA, including at least the followings:

NM_000345.3(SNCA):c.157G>A (p.Ala53Thr)

NM_000345.3(SNCA):c.152G>A (p.Gly51Asp)

NM_003560.3(PLA2G6):c.2222G>A (p.Arg741Gln)

NM_003560.3(PLA2G6):c.2239C>T (p.Arg747Trp)

NM_003560.3(PLA2G6):c.1904G>A (p.Arg635Gln)

NM_003560.3(PLA2G6):c.1354C>T (p.Gln452Ter)

NM_012179.3(FBXO7):c.1492C>T (p.Arg498Ter)

NM_012179.3(FBXO7):c.65C>T (p.Thr22Met)

NM_018206.5(VPS35):c.1858G>A (p.Asp620Asn)

NM_198241.2(EIF4G1):c.3614G>A (p.Arg1205His)

NM_198241.2(EIF4G1):c.1505C>T (p.Ala502Val)

NM_001256865.1(DNAJC6):c.2200C>T (p.Gln734Ter)

NM_001256865.1(DNAJC6):c.2326C>T (p.Gln776Ter)

NM_004562.2(PRKN):c.931C>T (p.Gln311Ter)

NM_004562.2(PRKN):c.1358G>A (p.Trp453Ter)

NM_004562.2(PRKN):c.635G>A (p.Cys212Tyr)

NM_203446.2(SYNJ1):c.773G>A (p.Arg258Gln)

NM_001320327.1(CHCHD2):c.182C>T (p.Thr61Ile)

NM_001320327.1(CHCHD2):c.434G>A (p.Arg145Gln)

NM_001320327.1(CHCHD2):c.300+5G>A

NM_032409.2(PINK1):c.926G>A (p.Gly309Asp)

NM_032409.2(PINK1):c.1311G>A (p.Trp437Ter)

NM_032409.2(PINK1):c.736C>T (p.Arg246Ter)

NM_032409.2(PINK1):c.836G>A (p.Arg279His)

NM_032409.2(PINK1):c.938C>T (p.Thr313Met)

NM_032409.2(PINK1):c.1366C>T (p.Gln456Ter)

NM_007262.4(PARK7):c.78G>A (p.Met26Ile)

NM_198578.3(LRRK2):c.4321C>T (p.Arg1441Cys)

NM_198578.3(LRRK2):c.4322G>A (p.Arg1441His)

NM_198578.3(LRRK2):c.1256C>T (p.Ala419Val)

NM_198578.3(LRRK2):c.6055G>A (p.Gly2019Ser)

NM_022089.3(ATP13A2):c.1306+5G>A

NM_022089.3(ATP13A2):c.2629G>A (p.Gly877Arg)

NM_022089.3(ATP13A2):c.490C>T (p.Arg164Trp)

NM_001005741.2(GBA):c.1444G>A (p.Asp482Asn)

m.15950G>A.

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Parkinson's Disease by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs in at least one gene selected from SNCA, PLA2G6, FBXO7, VPS35, EIF4G1, DNAJC6, PRKN, SYNJ1, CHCHD2, PINK1, PARK7, LRRK2, ATP13A2, and GBA, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Autism

[01041] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Autism. In some embodiments, the pathogenic mutations/SNPs are present in at least one

gene selected from MECP2, NLGN3, SLC9A9, EHMT1, CHD8, NLGN4X , GSPT2, and PTEN, including at least the followings:

NM_001110792.1(MECP2):c.916C>T (p.Arg306Ter)

NM_004992.3(MECP2):c.473C>T (p.Thr158Met)

NM_018977.3(NLGN3):c.1351C>T (p.Arg451Cys)

NM_173653.3(SLC9A9):c.1267C>T (p.Arg423Ter)

NM_024757.4(EHMT1):c.3413G>A (p.Trp1138Ter)

NM_020920.3(CHD8):c.2875C>T (p.Gln959Ter)

NM_020920.3(CHD8):c.3172C>T (p.Arg1058Ter)

NM_181332.2(NLGN4X):c.301C>T (p.Arg101Ter)

NM_018094.4(GSPT2):c.1021G>A (p.Val341Ile)

NM_000314.6(PTEN):c.392C>T (p.Thr131Ile)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Autism by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in at least one gene selected from MECP2, NLGN3, SLC9A9, EHMT1, CHD8, NLGN4X , GSPT2, and PTEN, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Amyotrophyic lateral sclerosis (ALS)

[01042] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with ALS. In some embodiments, the pathogenic mutations/SNPs are present in at least one gene selected from SOD1, VCP, UBQLN2, ERBB4, HNRNPA1, TUBA4A, SOD1, TARDBP, FIG4, OPTN, SETX, SPG11, FUS, VAPB, ANG, CHCHD10, SQSTM1, and TBK1, including at least the followings:

NM_000454.4(SOD1):c.289G>A (p.Asp97Asn)

NM_007126.3(VCP):c.1774G>A (p.Asp592Asn)

NM_007126.3(VCP):c.464G>A (p.Arg155His)

NM_007126.3(VCP):c.572G>A (p.Arg191Gln)

NM_013444.3(UBQLN2):c.1489C>T (p.Pro497Ser)

NM_013444.3(UBQLN2):c.1525C>T (p.Pro509Ser)

NM_013444.3(UBQLN2):c.1573C>T (p.Pro525Ser)

NM_013444.3(UBQLN2):c.1490C>T (p.Pro497Leu)

NM_005235.2(ERBB4):c.2780G>A (p.Arg927Gln)

NM_005235.2(ERBB4):c.3823C>T (p.Arg1275Trp)

NM_031157.3(HNRNPA1):c.940G>A (p.Asp314Asn)

NM_006000.2(TUBA4A):c.643C>T (p.Arg215Cys)

NM_006000.2(TUBA4A):c.958C>T (p.Arg320Cys)

NM_006000.2(TUBA4A):c.959G>A (p.Arg320His)

NM_006000.2(TUBA4A):c.1220G>A (p.Trp407Ter)

NM_006000.2(TUBA4A):c.1147G>A (p.Ala383Thr)

NM_000454.4(SOD1):c.112G>A (p.Gly38Arg)

NM_000454.4(SOD1):c.124G>A (p.Gly42Ser)

NM_000454.4(SOD1):c.125G>A (p.Gly42Asp)

NM_000454.4(SOD1):c.14C>T (p.Ala5Val)

NM_000454.4(SOD1):c.13G>A (p.Ala5Thr)

NM_000454.4(SOD1):c.436G>A (p.Ala146Thr)

NM_000454.4(SOD1):c.64G>A (p.Glu22Lys)

NM_000454.4(SOD1):c.404G>A (p.Ser135Asn)

NM_000454.4(SOD1):c.49G>A (p.Gly17Ser)

NM_000454.4(SOD1):c.217G>A (p.Gly73Ser)

NM_007375.3(TARDBP):c.892G>A (p.Gly298Ser)

NM_007375.3(TARDBP):c.943G>A (p.Ala315Thr)

NM_007375.3(TARDBP):c.883G>A (p.Gly295Ser)

NM_007375.3(TARDBP):c.*697G>A

NM_007375.3(TARDBP):c.1144G>A (p.Ala382Thr)

NM_007375.3(TARDBP):c.859G>A (p.Gly287Ser)

NM_014845.5(FIG4):c.547C>T (p.Arg183Ter)

NM_001008211.1(OPTN):c.1192C>T (p.Gln398Ter)

NM_015046.5(SETX):c.6407G>A (p.Arg2136His)

NM_015046.5(SETX):c.8C>T (p.Thr3Ile)

NM_025137.3(SPG11):c.118C>T (p.Gln40Ter)

NM_025137.3(SPG11):c.267G>A (p.Trp89Ter)

NM_025137.3(SPG11):c.5974C>T (p.Arg1992Ter)

NM_004960.3(FUS):c.1553G>A (p.Arg518Lys)

NM_004960.3(FUS):c.1561C>T (p.Arg521Cys)

NM_004960.3(FUS):c.1562G>A (p.Arg521His)

NM_004960.3(FUS):c.1520G>A (p.Gly507Asp)

NM_004960.3(FUS):c.1483C>T (p.Arg495Ter)

NM_004960.3(FUS):c.616G>A (p.Gly206Ser)

NM_004960.3(FUS):c.646C>T (p.Arg216Cys)

NM_004738.4(VAPB):c.166C>T (p.Pro56Ser)

NM_004738.4(VAPB):c.137C>T (p.Thr46Ile)

NM_001145.4(ANG):c.164G>A (p.Arg55Lys)

NM_001145.4(ANG):c.155G>A (p.Ser52Asn)

NM_001145.4(ANG):c.407C>T (p.Pro136Leu)

NM_001145.4(ANG):c.409G>A (p.Val137Ile)

NM_001301339.1(CHCHD10):c.239C>T (p.Pro80Leu)

NM_001301339.1(CHCHD10):c.176C>T (p.Ser59Leu)

NM_001142298.1(SQSTM1):c.-47-1924C>T

NM_003900.4(SQSTM1):c.1160C>T (p.Pro387Leu)

NM_003900.4(SQSTM1):c.1175C>T (p.Pro392Leu)

NM_013254.3(TBK1):c.1340+1G>A

NM_013254.3(TBK1):c.2086G>A (p.Glu696Lys)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing ALS by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in at least one gene selected from SOD1, VCP, UBQLN2, ERBB4, HNRNPA1, TUBA4A, SOD1, TARDBP, FIG4, OPTN, SETX, SPG11, FUS, VAPB, ANG, CHCHD10, SQSTM1, and TBK1,, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Schizophrenia

[01043] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Schizophrenia. In some embodiment, the pathogenic mutations/SNPs are present in at least one gene selected from PRODH, SETD1A, and SHANK3, including at least the followings:

NM_016335.4(PRODH):c.1292G>A (p.Arg431His)

NM_016335.4(PRODH):c.1397C>T (p.Thr466Met)

NM_014712.2(SETD1A):c.2209C>T (p.Gln737Ter)

NM_033517.1(SHANK3):c.3349C>T (p.Arg1117Ter)

NM_033517.1(SHANK3):c.1606C>T (p.Arg536Trp)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or

preventing Schizophrenia by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in at least one gene selected from PRODH, SETD1A, and SHANK3, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Adrenoleukodystrophy

[01044] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Adrenoleukodystrophy. In some embodiment, the pathogenic mutations/SNPs are present in at least the ABCD1 gene, including at least the followings:

NM_000033.3(ABCD1):c.421G>A (p.Ala141Thr)
NM_000033.3(ABCD1):c.796G>A (p.Gly266Arg)
NM_000033.3(ABCD1):c.1252C>T (p.Arg418Trp)
NM_000033.3(ABCD1):c.1552C>T (p.Arg518Trp)
NM_000033.3(ABCD1):c.1850G>A (p.Arg617His)
NM_000033.3(ABCD1):c.1396C>T (p.Gln466Ter)
NM_000033.3(ABCD1):c.1553G>A (p.Arg518Gln)
NM_000033.3(ABCD1):c.1679C>T (p.Pro560Leu)
NM_000033.3(ABCD1):c.1771C>T (p.Arg591Trp)
NM_000033.3(ABCD1):c.1802G>A (p.Trp601Ter)
NM_000033.3(ABCD1):c.346G>A (p.Gly116Arg)
NM_000033.3(ABCD1):c.406C>T (p.Gln136Ter)
NM_000033.3(ABCD1):c.1661G>A (p.Arg554His)
NM_000033.3(ABCD1):c.1825G>A (p.Glu609Lys)
NM_000033.3(ABCD1):c.1288C>T (p.Gln430Ter)
NM_000033.3(ABCD1):c.1781-1G>A
NM_000033.3(ABCD1):c.529C>T (p.Gln177Ter)
NM_000033.3(ABCD1):c.1866-10G>A

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Adrenoleukodystrophy by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in at least the ABCD1 gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Aicardi Goutieres syndrome

[01045] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Aicardi Goutieres syndrome. In some embodiment, the pathogenic mutations/SNPs are present in at least one gene selected from TREX1, RNASEH2C, ADAR, and IFIH1, including at least the followings:

NM_016381.5(TREX1):c.794G>A (p.Trp265Ter)

NM_033629.4(TREX1):c.52G>A (p.Asp18Asn)

NM_033629.4(TREX1):c.490C>T (p.Arg164Ter)

NM_032193.3(RNASEH2C):c.205C>T (p.Arg69Trp)

NM_001111.4(ADAR):c.3019G>A (p.Gly1007Arg)

NM_022168.3(IFIH1):c.2336G>A (p.Arg779His)

NM_022168.3(IFIH1):c.2335C>T (p.Arg779Cys)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Aicardi Goutieres syndrome by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in at least one gene selected from TREX1, RNASEH2C, ADAR, and IFIH1, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Fabry disease

[01046] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Fabry disease. In some embodiment, the pathogenic mutations/SNPs are present in at least the GLA gene, including at least the followings:

NM_000169.2(GLA):c.1024C>T (p.Arg342Ter)

NM_000169.2(GLA):c.1066C>T (p.Arg356Trp)

NM_000169.2(GLA):c.1025G>A (p.Arg342Gln)

NM_000169.2(GLA):c.281G>A (p.Cys94Tyr)

NM_000169.2(GLA):c.677G>A (p.Trp226Ter)

NM_000169.2(GLA):c.734G>A (p.Trp245Ter)

NM_000169.2(GLA):c.748C>T (p.Gln250Ter)

NM_000169.2(GLA):c.658C>T (p.Arg220Ter)

NM_000169.2(GLA):c.730G>A (p.Asp244Asn)

NM_000169.2(GLA):c.369+1G>A

NM_000169.2(GLA):c.335G>A (p.Arg112His)

NM_000169.2(GLA):c.485G>A (p.Trp162Ter)

NM_000169.2(GLA):c.661C>T (p.Gln221Ter)

NM_000169.2(GLA):c.916C>T (p.Gln306Ter)

NM_000169.2(GLA):c.1072G>A (p.Glu358Lys)

NM_000169.2(GLA):c.1087C>T (p.Arg363Cys)

NM_000169.2(GLA):c.1088G>A (p.Arg363His)

NM_000169.2(GLA):c.605G>A (p.Cys202Tyr)

NM_000169.2(GLA):c.830G>A (p.Trp277Ter)

NM_000169.2(GLA):c.979C>T (p.Gln327Ter)

NM_000169.2(GLA):c.422C>T (p.Thr141Ile)

NM_000169.2(GLA):c.285G>A (p.Trp95Ter)

NM_000169.2(GLA):c.735G>A (p.Trp245Ter)

NM_000169.2(GLA):c.639+919G>A

NM_000169.2(GLA):c.680G>A (p.Arg227Gln)

NM_000169.2(GLA):c.679C>T (p.Arg227Ter)

NM_000169.2(GLA):c.242G>A (p.Trp81Ter)

NM_000169.2(GLA):c.901C>T (p.Arg301Ter)

NM_000169.2(GLA):c.974G>A (p.Gly325Asp)

NM_000169.2(GLA):c.847C>T (p.Gln283Ter)

NM_000169.2(GLA):c.469C>T (p.Gln157Ter)

NM_000169.2(GLA):c.1118G>A (p.Gly373Asp)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Fabry disease by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in at least the GLA gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Lesch-Nyhan syndrome

[01047] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Lesch-Nyhan syndrome. In some embodiment, the pathogenic mutations/SNPs are present in at least the HPRT1 gene, including at least the followings:

NM_000194.2(HPRT1):c.151C>T (p.Arg51Ter)

NM_000194.2(HPRT1):c.384+1G>A

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Lesch-Nyhan syndrome by correcting one or more pathogenic G-to-A or C-to-T

mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in at least the HPRT1 gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Menkes Disease

[01048] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Menkes Disease. In some embodiment, the pathogenic mutations/SNPs are present in at least the ATP7A gene, including at least the followings:

NM_000052.6(ATP7A):c.601C>T (p.Arg201Ter)

NM_000052.6(ATP7A):c.2938C>T (p.Arg980Ter)

NM_000052.6(ATP7A):c.3056G>A (p.Gly1019Asp)

NM_000052.6(ATP7A):c.598C>T (p.Gln200Ter)

NM_000052.6(ATP7A):c.1225C>T (p.Arg409Ter)

NM_000052.6(ATP7A):c.1544-1G>A

NM_000052.6(ATP7A):c.1639C>T (p.Arg547Ter)

NM_000052.6(ATP7A):c.1933C>T (p.Arg645Ter)

NM_000052.6(ATP7A):c.1946+5G>A

NM_000052.6(ATP7A):c.1950G>A (p.Trp650Ter)

NM_000052.6(ATP7A):c.2179G>A (p.Gly727Arg)

NM_000052.6(ATP7A):c.2187G>A (p.Trp729Ter)

NM_000052.6(ATP7A):c.2383C>T (p.Arg795Ter)

NM_000052.6(ATP7A):c.2499-1G>A

NM_000052.6(ATP7A):c.2555C>T (p.Pro852Leu)

NM_000052.6(ATP7A):c.2956C>T (p.Arg986Ter)

NM_000052.6(ATP7A):c.3112-1G>A

NM_000052.6(ATP7A):c.3466C>T (p.Gln1156Ter)

NM_000052.6(ATP7A):c.3502C>T (p.Gln1168Ter)

NM_000052.6(ATP7A):c.3764G>A (p.Gly1255Glu)

NM_000052.6(ATP7A):c.3943G>A (p.Gly1315Arg)

NM_000052.6(ATP7A):c.4123+1G>A

NM_000052.6(ATP7A):c.4226+5G>A

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Menkes Disease by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs

present in at least the ATP7A gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Eye Diseases

[01049] The invention provides efficient treatment of inherited and acquired ocular diseases of the retina. Holmgaard et al. (Mol. Ther. Nucleic Acids 9:89-99, 15-Dec-2017 doi: 10.1016/j.omtn.2017.08.016. Epub 2017 Sep 21) reported indel formation at high frequencies when SpCas9 was delivered by lentiviral vectors (LVs) encoding SpCas9 targeted to Vegfa and there was a significant reduction of of VEGFA in transduced cells. Duan et al. (J Biol Chem. 2016 Jul 29;291(31):16339-47. doi: 10.1074/jbc.M116.729467. Epub 2016 May 31) describe use of CRISPR to target MDM2 genomic locus in human primary retinal pigment epithelial cells

[01050] The methods and compositions of the instant invention are similarly applicable to treatment of ocular diseases, including age-related macular degeneration.

[01051] Huang et al. (Nat Commun. 2017 Jul 24;8(1):112. doi: 10.1038/s41467-017-00140-3 employed CRISPR to edit VEGFR2 to treat angiogenesis-associated diseases.

[01052] Pathogenic G-to-A or C-to-T mutations/SNPs associated with various eye diseases are reported in the ClinVar database and disclosed in Table A, including but not limited to Stargardt Disease, Bardet-Biedl Syndrome, Cone-rod dystrophy, Congenital Stationary Night Blindness, Usher Syndrome, Leber Congenital Amaurosis, Retinitis Pigmentosa, and Achromatopsia. Accordingly, an aspect of the invention relates to a method for correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with any of these diseases, as discussed below.

Stargardt Disease

[01053] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Stargardt Disease. In some embodiment, the pathogenic mutations/SNPs are present in the ABCA4 gene, including at least the followings:

NM_000350.2(ABCA4):c.4429C>T (p.Gln1477Ter)

NM_000350.2(ABCA4):c.6647C>T (p.Ala2216Val)

NM_000350.2(ABCA4):c.5312+1G>A

NM_000350.2(ABCA4):c.5189G>A (p.Trp1730Ter)

NM_000350.2(ABCA4):c.4352+1G>A

NM_000350.2(ABCA4):c.4253+5G>A

NM_000350.2(ABCA4):c.3871C>T (p.Gln1291Ter)

NM_000350.2(ABCA4):c.3813G>A (p.Glu1271=)
NM_000350.2(ABCA4):c.1293G>A (p.Trp431Ter)
NM_000350.2(ABCA4):c.206G>A (p.Trp69Ter)
NM_000350.2(ABCA4):c.3322C>T (p.Arg1108Cys)
NM_000350.2(ABCA4):c.1804C>T (p.Arg602Trp)
NM_000350.2(ABCA4):c.1937+1G>A
NM_000350.2(ABCA4):c.2564G>A (p.Trp855Ter)
NM_000350.2(ABCA4):c.4234C>T (p.Gln1412Ter)
NM_000350.2(ABCA4):c.4457C>T (p.Pro1486Leu)
NM_000350.2(ABCA4):c.4594G>A (p.Asp1532Asn)
NM_000350.2(ABCA4):c.4919G>A (p.Arg1640Gln)
NM_000350.2(ABCA4):c.5196+1G>A
NM_000350.2(ABCA4):c.6316C>T (p.Arg2106Cys)
NM_000350.2(ABCA4):c.3056C>T (p.Thr1019Met)
NM_000350.2(ABCA4):c.52C>T (p.Arg18Trp)
NM_000350.2(ABCA4):c.122G>A (p.Trp41Ter)
NM_000350.2(ABCA4):c.1903C>T (p.Gln635Ter)
NM_000350.2(ABCA4):c.194G>A (p.Gly65Glu)
NM_000350.2(ABCA4):c.3085C>T (p.Gln1029Ter)
NM_000350.2(ABCA4):c.4195G>A (p.Glu1399Lys)
NM_000350.2(ABCA4):c.454C>T (p.Arg152Ter)
NM_000350.2(ABCA4):c.45G>A (p.Trp15Ter)
NM_000350.2(ABCA4):c.4610C>T (p.Thr1537Met)
NM_000350.2(ABCA4):c.6112C>T (p.Arg2038Trp)
NM_000350.2(ABCA4):c.6118C>T (p.Arg2040Ter)
NM_000350.2(ABCA4):c.6342G>A (p.Val2114=)
NM_000350.2(ABCA4):c.6658C>T (p.Gln2220Ter)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Stargardt Disease by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in the ABCA4 gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Bardet-Biedl Syndrome

[01054] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Bardet-Biedl Syndrome. In some embodiment, the pathogenic mutations/SNPs are present in at least one gene selected from BBS1, BBS2, BBS7, BBS9, BBS10, BBS12, LZTFL1, and TRIM32, including at least the followings:

NM_024649.4(BBS1):c.416G>A (p.Trp139Ter)

NM_024649.4(BBS1):c.871C>T (p.Gln291Ter)

NM_198428.2(BBS9):c.263+1G>A

NM_001178007.1(BBS12):c.1704G>A (p.Trp568Ter)

NM_001276378.1(LZTFL1):c.271C>T (p.Arg91Ter)

NM_031885.3(BBS2):c.1864C>T (p.Arg622Ter)

NM_198428.2(BBS9):c.1759C>T (p.Arg587Ter)

NM_198428.2(BBS9):c.1789+1G>A

NM_024649.4(BBS1):c.432+1G>A

NM_176824.2(BBS7):c.632C>T (p.Thr211Ile)

NM_012210.3(TRIM32):c.388C>T (p.Pro130Ser)

NM_031885.3(BBS2):c.823C>T (p.Arg275Ter)

NM_024685.3(BBS10):c.145C>T (p.Arg49Trp)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Bardet-Biedl Syndrome by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in at least one gene selected from BBS1, BBS2, BBS7, BBS9, BBS10, BBS12, LZTFL1, and TRIM32, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Cone-rod dystrophy

[01055] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Cone-rod dystrophy. In some embodiment, the pathogenic mutations/SNPs are present in at least one gene selected from RPGRIP1, DRAM2, ABCA4, ADAM9, and CACNA1F, including at least the followings:

NM_020366.3(RPGRIP1):c.154C>T (p.Arg52Ter)

NM_178454.5(DRAM2):c.494G>A (p.Trp165Ter)

NM_178454.5(DRAM2):c.131G>A (p.Ser44Asn)

NM_000350.2(ABCA4):c.161G>A (p.Cys54Tyr)

NM_000350.2(ABCA4):c.5714+5G>A

NM_000350.2(ABCA4):c.880C>T (p.Gln294Ter)

NM_000350.2(ABCA4):c.6079C>T (p.Leu2027Phe)

NM_000350.2(ABCA4):c.3113C>T (p.Ala1038Val)

NM_000350.2(ABCA4):c.634C>T (p.Arg212Cys)

NM_003816.2(ADAM9):c.490C>T (p.Arg164Ter)

NM_005183.3(CACNA1F):c.244C>T (p.Arg82Ter)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Cone-rod dystrophy by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in at least one gene selected from RPGRIP1, DRAM2, ABCA4, ADAM9, and CACNA1F, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Congenital Stationary Night Blindness

[01056] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Congenital Stationary Night Blindness. In some embodiment, the pathogenic mutations/SNPs are present in at least one gene selected from GRM6, TRPM1, GPR179, and CACNA1F, including at least the followings:

NM_000843.3(GRM6):c.1462C>T (p.Gln488Ter)

NM_002420.5(TRPM1):c.2998C>T (p.Arg1000Ter)

NM_001004334.3(GPR179):c.673C>T (p.Gln225Ter)

NM_005183.3(CACNA1F):c.2576+1G>A

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Congenital Stationary Night Blindness by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in at least one gene selected from GRM6, TRPM1, GPR179, and CACNA1F, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Usher Syndrome

[01057] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Usher Syndrome. In some embodiment, the pathogenic mutations/SNPs are present in at least

one gene selected from MYO7A, USH1C, CDH23, PCDH15, USH2A, ADGRV1, WHRN, and CLRN1, including at least the followings:

NM_000260.3(MYO7A):c.640G>A (p.Gly214Arg)

NM_000260.3(MYO7A):c.1200+1G>A

NM_000260.3(MYO7A):c.141G>A (p.Trp47Ter)

NM_000260.3(MYO7A):c.1556G>A (p.Gly519Asp)

NM_000260.3(MYO7A):c.1900C>T (p.Arg634Ter)

NM_000260.3(MYO7A):c.1963C>T (p.Gln655Ter)

NM_000260.3(MYO7A):c.2094+1G>A

NM_000260.3(MYO7A):c.4293G>A (p.Trp1431Ter)

NM_000260.3(MYO7A):c.5101C>T (p.Arg1701Ter)

NM_000260.3(MYO7A):c.5617C>T (p.Arg1873Trp)

NM_000260.3(MYO7A):c.5660C>T (p.Pro1887Leu)

NM_000260.3(MYO7A):c.6070C>T (p.Arg2024Ter)

NM_000260.3(MYO7A):c.470+1G>A

NM_000260.3(MYO7A):c.5968C>T (p.Gln1990Ter)

NM_000260.3(MYO7A):c.3719G>A (p.Arg1240Gln)

NM_000260.3(MYO7A):c.494C>T (p.Thr165Met)

NM_000260.3(MYO7A):c.5392C>T (p.Gln1798Ter)

NM_000260.3(MYO7A):c.5648G>A (p.Arg1883Gln)

NM_000260.3(MYO7A):c.448C>T (p.Arg150Ter)

NM_000260.3(MYO7A):c.700C>T (p.Gln234Ter)

NM_000260.3(MYO7A):c.635G>A (p.Arg212His)

NM_000260.3(MYO7A):c.1996C>T (p.Arg666Ter)

NM_005709.3(USH1C):c.216G>A (p.Val72=)

NM_022124.5(CDH23):c.7362+5G>A

NM_022124.5(CDH23):c.3481C>T (p.Arg1161Ter)

NM_022124.5(CDH23):c.3628C>T (p.Gln1210Ter)

NM_022124.5(CDH23):c.5272C>T (p.Gln1758Ter)

NM_022124.5(CDH23):c.5712+1G>A

NM_022124.5(CDH23):c.5712G>A (p.Thr1904=)

NM_022124.5(CDH23):c.5923+1G>A

NM_022124.5(CDH23):c.6049+1G>A

NM_022124.5(CDH23):c.7776G>A (p.Trp2592Ter)

NM_022124.5(CDH23):c.9556C>T (p.Arg3186Ter)
NM_022124.5(CDH23):c.3706C>T (p.Arg1236Ter)
NM_022124.5(CDH23):c.4309C>T (p.Arg1437Ter)
NM_022124.5(CDH23):c.6050-9G>A
NM_033056.3(PCDH15):c.3316C>T (p.Arg1106Ter)
NM_033056.3(PCDH15):c.7C>T (p.Arg3Ter)
NM_033056.3(PCDH15):c.1927C>T (p.Arg643Ter)
NM_001142772.1(PCDH15):c.400C>T (p.Arg134Ter)
NM_033056.3(PCDH15):c.3358C>T (p.Arg1120Ter)
NM_206933.2(USH2A):c.11048-1G>A
NM_206933.2(USH2A):c.1143+1G>A
NM_206933.2(USH2A):c.11954G>A (p.Trp3985Ter)
NM_206933.2(USH2A):c.12868C>T (p.Gln4290Ter)
NM_206933.2(USH2A):c.14180G>A (p.Trp4727Ter)
NM_206933.2(USH2A):c.14911C>T (p.Arg4971Ter)
NM_206933.2(USH2A):c.5788C>T (p.Arg1930Ter)
NM_206933.2(USH2A):c.5858-1G>A
NM_206933.2(USH2A):c.6224G>A (p.Trp2075Ter)
NM_206933.2(USH2A):c.820C>T (p.Arg274Ter)
NM_206933.2(USH2A):c.8981G>A (p.Trp2994Ter)
NM_206933.2(USH2A):c.9304C>T (p.Gln3102Ter)
NM_206933.2(USH2A):c.13010C>T (p.Thr4337Met)
NM_206933.2(USH2A):c.14248C>T (p.Gln4750Ter)
NM_206933.2(USH2A):c.6398G>A (p.Trp2133Ter)
NM_206933.2(USH2A):c.632G>A (p.Trp211Ter)
NM_206933.2(USH2A):c.6601C>T (p.Gln2201Ter)
NM_206933.2(USH2A):c.13316C>T (p.Thr4439Ile)
NM_206933.2(USH2A):c.4405C>T (p.Gln1469Ter)
NM_206933.2(USH2A):c.9570+1G>A
NM_206933.2(USH2A):c.8740C>T (p.Arg2914Ter)
NM_206933.2(USH2A):c.8681+1G>A
NM_206933.2(USH2A):c.1000C>T (p.Arg334Trp)
NM_206933.2(USH2A):c.14175G>A (p.Trp4725Ter)
NM_206933.2(USH2A):c.9390G>A (p.Trp3130Ter)

NM_206933.2(USH2A):c.908G>A (p.Arg303His)

NM_206933.2(USH2A):c.5776+1G>A

NM_206933.2(USH2A):c.11156G>A (p.Arg3719His)

NM_032119.3(ADGRV1):c.2398C>T (p.Arg800Ter)

NM_032119.3(ADGRV1):c.7406G>A (p.Trp2469Ter)

NM_032119.3(ADGRV1):c.12631C>T (p.Arg4211Ter)

NM_032119.3(ADGRV1):c.7129C>T (p.Arg2377Ter)

NM_032119.3(ADGRV1):c.14885G>A (p.Trp4962Ter)

NM_015404.3(WHRN):c.1267C>T (p.Arg423Ter)

NM_174878.2(CLRN1):c.619C>T (p.Arg207Ter)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Enhanced Usher Syndrome by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in at least one gene selected from MYO7A, USH1C, CDH23, PCDH15, USH2A, ADGRV1, WHRN, and CLRN1, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Leber Congenital Amaurosis

[01058] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Leber Congenital Amaurosis. In some embodiment, the pathogenic mutations/SNPs are present in at least one gene selected from TULP1, RPE65, SPATA7, AIPL1, CRB1, NMNAT1, and PEX1, including at least the followings:

NM_003322.5(TULP1):c.1495+1G>A

NM_000329.2(RPE65):c.11+5G>A

NM_018418.4(SPATA7):c.322C>T (p.Arg108Ter)

NM_014336.4(AIPL1):c.784G>A (p.Gly262Ser)

NM_201253.2(CRB1):c.1576C>T (p.Arg526Ter)

NM_201253.2(CRB1):c.3307G>A (p.Gly1103Arg)

NM_201253.2(CRB1):c.2843G>A (p.Cys948Tyr)

NM_022787.3(NMNAT1):c.769G>A (p.Glu257Lys)

NM_000466.2(PEX1):c.2528G>A (p.Gly843Asp)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Leber Congenital Amaurosis by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs

present in at least one gene selected from TULP1, RPE65, SPATA7, AIPL1, CRB1, NMNAT1, and PEX1, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Retinitis Pigmentosa

[01059] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Retinitis Pigmentosa. In some embodiment, the pathogenic mutations/SNPs are present in at least one gene selected from CRB1, IFT140, RP1, IMPDH1, PRPF31, RPGR, ABCA4, RPE65, EYS, NRL, FAM161A, NR2E3, USH2A, RHO, PDE6B, KLHL7, PDE6A, CNGB1, BEST1, C2orf71, PRPH2, CA4, CERKL, RPE65, PDE6B, and ADGRV1, including at least the followings:

NM_001257965.1(CRB1):c.2711G>A (p.Cys904Tyr)

NM_014714.3(IFT140):c.3827G>A (p.Gly1276Glu)

NM_006269.1(RP1):c.2029C>T (p.Arg677Ter)

NM_000883.3(IMPDH1):c.931G>A (p.Asp311Asn)

NM_015629.3(PRPF31):c.1273C>T (p.Gln425Ter)

NM_015629.3(PRPF31):c.1073+1G>A

NM_000328.2(RPGR):c.1387C>T (p.Gln463Ter)

NM_000350.2(ABCA4):c.4577C>T (p.Thr1526Met)

NM_000350.2(ABCA4):c.6229C>T (p.Arg2077Trp)

NM_000329.2(RPE65):c.271C>T (p.Arg91Trp)

NM_001142800.1(EYS):c.2194C>T (p.Gln732Ter)

NM_001142800.1(EYS):c.490C>T (p.Arg164Ter)

NM_006177.3(NRL):c.151C>T (p.Pro51Ser)

NM_001201543.1(FAM161A):c.1567C>T (p.Arg523Ter)

NM_014249.3(NR2E3):c.166G>A (p.Gly56Arg)

NM_206933.2(USH2A):c.2209C>T (p.Arg737Ter)

NM_206933.2(USH2A):c.14803C>T (p.Arg4935Ter)

NM_206933.2(USH2A):c.10073G>A (p.Cys3358Tyr)

NM_000539.3(RHO):c.541G>A (p.Glu181Lys)

NM_000283.3(PDE6B):c.892C>T (p.Gln298Ter)

NM_001031710.2(KLHL7):c.458C>T (p.Ala153Val)

NM_000440.2(PDE6A):c.1926+1G>A

NM_001297.4(CNGB1):c.2128C>T (p.Gln710Ter)

NM_001297.4(CNGB1):c.952C>T (p.Gln318Ter)

NM_004183.3(BEST1):c.682G>A (p.Asp228Asn)

NM_001029883.2(C2orf71):c.1828C>T (p.Gln610Ter)

NM_000322.4(PRPH2):c.647C>T (p.Pro216Leu)

NM_000717.4(CA4):c.40C>T (p.Arg14Trp)

NM_201548.4(CERKL):c.769C>T (p.Arg257Ter)

NM_000329.2(RPE65):c.118G>A (p.Gly40Ser)

NM_000322.4(PRPH2):c.499G>A (p.Gly167Ser)

NM_000539.3(RHO):c.403C>T (p.Arg135Trp)

NM_000283.3(PDE6B):c.2193+1G>A

NM_032119.3(ADGRV1):c.6901C>T (p.Gln2301Ter)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Retinitis Pigmentosa by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in at least one gene selected from CRB1, IFT140, RP1, IMPDH1, PRPF31, RPGR, ABCA4, RPE65, EYS, NRL, FAM161A, NR2E3, USH2A, RHO, PDE6B, KLHL7, PDE6A, CNGB1, BEST1, C2orf71, PRPH2, CA4, CERKL, RPE65, PDE6B, and ADGRV1, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Achromatopsia

[01060] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Achromatopsia. In some embodiment, the pathogenic mutations/SNPs are present in at least one gene selected from CNGA3, CNGB3, and ATF6, including at least the followings:

NM_001298.2(CNGA3):c.847C>T (p.Arg283Trp)

NM_001298.2(CNGA3):c.101+1G>A

NM_001298.2(CNGA3):c.1585G>A (p.Val529Met)

NM_019098.4(CNGB3):c.1578+1G>A

NM_019098.4(CNGB3):c.607C>T (p.Arg203Ter)

NM_019098.4(CNGB3):c.1119G>A (p.Trp373Ter)

NM_007348.3(ATF6):c.970C>T (p.Arg324Cys)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Achromatopsia by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs

present in at least one gene selected from CNGA3, CNGB3, and ATF6, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Diseases Affecting Hearing

[01061] Pathogenic G-to-A or C-to-T mutations/SNPs associated with various diseases affecting hearing are reported in the ClinVar database and disclosed in Table A, including but not limited to deafness and Nonsyndromic hearing loss. Accordingly, an aspect of the invention relates to a method for correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with any of these diseases, as discussed below.

Deafness

[01062] Gao et al. (Nature. 2017 Dec 20. doi: 10.1038/nature25164. [Epub ahead of print]) reported genome editing using CRISPR-Cas9 to target Tmc1 gene in mice and reduce progressive hearing loss and deafness In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with deafness. In some embodiment, the pathogenic mutations/SNPs are present in at least one gene selected from FGF3, MYO7A, STRC, ACTG1, SLC17A8, TMC1, GJB2, MYH14, COCH, CDH23, USH1C, GJB2, MYO7A, PCDH15, MYO15A, MYO3A, WHRN, DFNB59, TMC1, LOXHD1, TMPRSS3, OTOGL, OTOF, JAG1, and MARVELD2, including at least the followings:

NM_005247.2(FGF3):c.283C>T (p.Arg95Trp)

NM_000260.3(MYO7A):c.652G>A (p.Asp218Asn)

NM_000260.3(MYO7A):c.689C>T (p.Ala230Val)

NM_153700.2(STRC):c.4057C>T (p.Gln1353Ter)

NM_001614.3(ACTG1):c.721G>A (p.Glu241Lys)

NM_139319.2(SLC17A8):c.632C>T (p.Ala211Val)

NM_138691.2(TMC1):c.1714G>A (p.Asp572Asn)

NM_004004.5(GJB2):c.598G>A (p.Gly200Arg)

NM_004004.5(GJB2):c.71G>A (p.Trp24Ter)

NM_004004.5(GJB2):c.416G>A (p.Ser139Asn)

NM_004004.5(GJB2):c.224G>A (p.Arg75Gln)

NM_004004.5(GJB2):c.95G>A (p.Arg32His)

NM_004004.5(GJB2):c.250G>A (p.Val84Met)

NM_004004.5(GJB2):c.428G>A (p.Arg143Gln)

NM_004004.5(GJB2):c.551G>A (p.Arg184Gln)

NM_004004.5(GJB2):c.223C>T (p.Arg75Trp)

NM_024729.3(MYH14):c.359C>T (p.Ser120Leu)

NM_004086.2(COCH):c.151C>T (p.Pro51Ser)

NM_022124.5(CDH23):c.4021G>A (p.Asp1341Asn)

NM_153700.2(STRC):c.4701+1G>A

NM_153676.3(USH1C):c.496+1G>A

NM_004004.5(GJB2):c.131G>A (p.Trp44Ter)

NM_004004.5(GJB2):c.283G>A (p.Val95Met)

NM_004004.5(GJB2):c.298C>T (p.His100Tyr)

NM_004004.5(GJB2):c.427C>T (p.Arg143Trp)

NM_004004.5(GJB2):c.109G>A (p.Val37Ile)

NM_004004.5(GJB2):c.-23+1G>A

NM_004004.5(GJB2):c.148G>A (p.Asp50Asn)

NM_004004.5(GJB2):c.134G>A (p.Gly45Glu)

NM_004004.5(GJB2):c.370C>T (p.Gln124Ter)

NM_004004.5(GJB2):c.230G>A (p.Trp77Ter)

NM_004004.5(GJB2):c.231G>A (p.Trp77Ter)

NM_000260.3(MYO7A):c.5899C>T (p.Arg1967Ter)

NM_000260.3(MYO7A):c.2005C>T (p.Arg669Ter)

NM_033056.3(PCDH15):c.733C>T (p.Arg245Ter)

NM_016239.3(MYO15A):c.3866+1G>A

NM_016239.3(MYO15A):c.6178-1G>A

NM_016239.3(MYO15A):c.8714-1G>A

NM_017433.4(MYO3A):c.2506-1G>A

NM_015404.3(WHRN):c.1417-1G>A

NM_001042702.3(DFNB59):c.499C>T (p.Arg167Ter)

NM_138691.2(TMC1):c.100C>T (p.Arg34Ter)

NM_138691.2(TMC1):c.1165C>T (p.Arg389Ter)

NM_144612.6(LOXHD1):c.2008C>T (p.Arg670Ter)

NM_144612.6(LOXHD1):c.4714C>T (p.Arg1572Ter)

NM_144612.6(LOXHD1):c.4480C>T (p.Arg1494Ter)

NM_024022.2(TMPRSS3):c.325C>T (p.Arg109Trp)

NM_173591.3(OTOGL):c.3076C>T (p.Gln1026Ter)

NM_194248.2(OTOF):c.4483C>T (p.Arg1495Ter)

NM_194248.2(OTOF):c.2122C>T (p.Arg708Ter)

NM_194248.2(OTOF):c.2485C>T (p.Gln829Ter)

NM_001038603.2(MARVELD2):c.1498C>T (p.Arg500Ter)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing deafness by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in at least one gene selected from FGF3, MYO7A, STRC, ACTG1, SLC17A8, TMC1, GJB2, MYH14, COCH, CDH23, USH1C, GJB2, MYO7A, PCDH15, MYO15A, MYO3A, WHRN, DFNB59, TMC1, LOXHD1, TMPRSS3, OTOGL, OTOF, JAG1, and MARVELD2, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Nonsyndromic hearing loss

[01063] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Nonsyndromic hearing loss. In some embodiment, the pathogenic mutations/SNPs are present in at least one gene selected from GJB2, POU3F4, MYO15A, TMPRSS3, LOXHD1, OTOF, MYO6, OTOA, STRC, TRIOBP, MARVELD2, TMC1, TECTA, OTOGL, and GIPC3, including at least the followings:

NM_004004.5(GJB2):c.169C>T (p.Gln57Ter)

NM_000307.4(POU3F4):c.499C>T (p.Arg167Ter)

NM_016239.3(MYO15A):c.8767C>T (p.Arg2923Ter)

NM_024022.2(TMPRSS3):c.323-6G>A

NM_024022.2(TMPRSS3):c.916G>A (p.Ala306Thr)

NM_144612.6(LOXHD1):c.2497C>T (p.Arg833Ter)

NM_194248.2(OTOF):c.2153G>A (p.Trp718Ter)

NM_194248.2(OTOF):c.2818C>T (p.Gln940Ter)

NM_194248.2(OTOF):c.4799+1G>A

NM_004999.3(MYO6):c.826C>T (p.Arg276Ter)

NM_144672.3(OTOA):c.1880+1G>A

NM_153700.2(STRC):c.5188C>T (p.Arg1730Ter)

NM_153700.2(STRC):c.3670C>T (p.Arg1224Ter)

NM_153700.2(STRC):c.4402C>T (p.Arg1468Ter)

NM_024022.2(TMPRSS3):c.1192C>T (p.Gln398Ter)

NM_001039141.2(TRIOBP):c.6598C>T (p.Arg2200Ter)

NM_016239.3(MYO15A):c.7893+1G>A

NM_016239.3(MYO15A):c.5531+1G>A

NM_016239.3(MYO15A):c.6046+1G>A

NM_144612.6(LOXHD1):c.3169C>T (p.Arg1057Ter)

NM_001038603.2(MARVELD2):c.1331+1G>A

NM_138691.2(TMC1):c.1676G>A (p.Trp559Ter)

NM_138691.2(TMC1):c.1677G>A (p.Trp559Ter)

NM_005422.2(TECTA):c.5977C>T (p.Arg1993Ter)

NM_173591.3(OTOGL):c.4987C>T (p.Arg1663Ter)

NM_153700.2(STRC):c.3493C>T (p.Gln1165Ter)

NM_153700.2(STRC):c.3217C>T (p.Arg1073Ter)

NM_016239.3(MYO15A):c.5896C>T (p.Arg1966Ter)

NM_133261.2(GIPC3):c.411+1G>A

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Nonsyndromic hearing loss by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in at least one gene selected from GJB2, POU3F4, MYO15A, TMPRSS3, LOXHD1, OTOF, MYO6, OTOA, STRC, TRIOBP, MARVELD2, TMC1, TECTA, OTOGL, and GIPC3, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Blood Disorders

[01064] Pathogenic G-to-A or C-to-T mutations/SNPs associated with various blood disorders are reported in the ClinVar database and disclosed in Table A, including but not limited to Beta-thalassemia, Hemophilia A, Hemophilia B, Hemophilia C, and Wiskott-Aldrich syndrome. Accordingly, an aspect of the invention relates to a method for correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with any of these diseases, as discussed below.

Beta-thalassemia

[01065] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Beta-thalassemia. In some embodiment, the pathogenic mutations/SNPs are present in at least the HBB gene, including at least the followings:

NM_000518.4(HBB):c.-137C>T

NM_000518.4(HBB):c.-50-88C>T

NM_000518.4(HBB):c.-140C>T

NM_000518.4(HBB):c.316-197C>T

NM_000518.4(HBB):c.93-21G>A

NM_000518.4(HBB):c.114G>A (p.Trp38Ter)

NM_000518.4(HBB):c.118C>T (p.Gln40Ter)

NM_000518.4(HBB):c.92+1G>A

NM_000518.4(HBB):c.315+1G>A

NM_000518.4(HBB):c.92+5G>A

NM_000518.4(HBB):c.-50-101C>T

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Beta-thalassemia by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in the HBB gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Hemophilia A

[01066] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Hemophilia A. In some embodiment, the pathogenic mutations/SNPs are present in at least the F8 gene, including at least the followings:

NM_000132.3(F8):c.3169G>A (p.Glu1057Lys)

NM_000132.3(F8):c.902G>A (p.Arg301His)

NM_000132.3(F8):c.1834C>T (p.Arg612Cys)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Hemophilia A by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in the F8 gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Factor V Leiden

[01067] In some embodiments, the methods, systems, and compositions described herein are used to correct Factor V Leiden mutations. This disease-causing single point mutation (G1746 -> A) represents the most abundant genetic risk factor in heritable multifactorial thrombophilia in the Caucasian population. Due to the point mutation, a single amino acid substitution (R534[RIGHTWARDS ARROW]Q) appears at the Protein C dependent proteolytic cleavage site (R533R534) of the blood coagulation factor F5. Whereas the heterozygous defect is accompanied by an only minor increase in thrombosis risk (ca. 8-fold), the homozygous defect has a much more pronounced effect (>80-fold increased risk).19

Directed RNA editing has the potential to compensate for this genetic defect by its repair at the RNA level.

Hemophilia B

[01068] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Hemophilia B. In some embodiment, the pathogenic mutations/SNPs are present in at least the F9 gene, including at least the followings:

NM_000133.3(F9):c.835G>A (p.Ala279Thr)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Hemophilia B by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in the F9 gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Hemophilia C

[01069] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Hemophilia C. In some embodiment, the pathogenic mutations/SNPs are present in at least the F11 gene, including at least the followings:

NM_000128.3(F11):c.400C>T (p.Gln134Ter)

NM_000128.3(F11):c.1432G>A (p.Gly478Arg)

NM_000128.3(F11):c.1288G>A (p.Ala430Thr)

NM_000128.3(F11):c.326-1G>A

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Hemophilia C by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in the F11 gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Wiskott-Aldrich syndrome

[01070] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Wiskott-Aldrich syndrome. In some embodiment, the pathogenic mutations/SNPs are present in at least the WAS gene, including at least the followings:

NM_000377.2(WAS):c.37C>T (p.Arg13Ter)

NM_000377.2(WAS):c.257G>A (p.Arg86His)

NM_000377.2(WAS):c.777+1G>A

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Wiskott-Aldrich syndrome by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in the WAS gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Liver Diseases

[01071] Pathogenic G-to-A or C-to-T mutations/SNPs associated with various liver diseases are reported in the ClinVar database and disclosed in Table A, including but not limited to Transthyretin amyloidosis, Alpha-1-antitrypsin deficiency, Wilson's disease, and Phenylketonuria. Accordingly, an aspect of the invention relates to a method for correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with any of these diseases, as discussed below.

Transthyretin amyloidosis

[01072] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Transthyretin amyloidosis. In some embodiment, the pathogenic mutations/SNPs are present in at least the TTR gene, including at least the followings:

NM_000371.3(TTR):c.424G>A (p.Val142Ile)

NM_000371.3(TTR):c.148G>A (p.Val50Met)

NM_000371.3(TTR):c.118G>A (p.Val40Ile)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Transthyretin amyloidosis by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in the TTR gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Alpha-1-antitrypsin deficiency

[01073] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Alpha-1-antitrypsin deficiency. In some embodiment, the pathogenic mutations/SNPs are present in at least the **SERPINA1** gene, including at least the followings:

NM_000295.4(SERPINA1):c.538C>T (p.Gln180Ter)

NM_001127701.1(SERPINA1):c.1178C>T (p.Pro393Leu)

NM_001127701.1(SERPINA1):c.230C>T (p.Ser77Phe)

NM_001127701.1(SERPINA1):c.1096G>A (p.Glu366Lys)

NM_000295.4(SERPINA1):c.1177C>T (p.Pro393Ser)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Alpha-1-antitrypsin deficiency by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in the **SERPINA1** gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Wilson's disease

[01074] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Wilson's disease. In some embodiment, the pathogenic mutations/SNPs are present in at least the **ATP7B** gene, including at least the followings:

NM_000053.3(ATP7B):c.2293G>A (p.Asp765Asn)

NM_000053.3(ATP7B):c.3955C>T (p.Arg1319Ter)

NM_000053.3(ATP7B):c.2865+1G>A

NM_000053.3(ATP7B):c.3796G>A (p.Gly1266Arg)

NM_000053.3(ATP7B):c.2621C>T (p.Ala874Val)

NM_000053.3(ATP7B):c.2071G>A (p.Gly691Arg)

NM_000053.3(ATP7B):c.2128G>A (p.Gly710Ser)

NM_000053.3(ATP7B):c.2336G>A (p.Trp779Ter)

NM_000053.3(ATP7B):c.4021G>A (p.Gly1341Ser)

NM_000053.3(ATP7B):c.3182G>A (p.Gly1061Glu)

NM_000053.3(ATP7B):c.4114C>T (p.Gln1372Ter)

NM_000053.3(ATP7B):c.1708-1G>A

NM_000053.3(ATP7B):c.865C>T (p.Gln289Ter)

NM_000053.3(ATP7B):c.2930C>T (p.Thr977Met)

NM_000053.3(ATP7B):c.3659C>T (p.Thr1220Met)

NM_000053.3(ATP7B):c.2605G>A (p.Gly869Arg)

NM_000053.3(ATP7B):c.2975C>T (p.Pro992Leu)

NM_000053.3(ATP7B):c.2519C>T (p.Pro840Leu)

NM_000053.3(ATP7B):c.2906G>A (p.Arg969Gln)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Wilson's disease by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs

present in the ATP7B gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Phenylketonuria

[01075] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Phenylketonuria. In some embodiment, the pathogenic mutations/SNPs are present in at least the **PAH** gene, including at least the followings:

NM_000277.1(PAH):c.1315+1G>A

NM_000277.1(PAH):c.1222C>T (p.Arg408Trp)

NM_000277.1(PAH):c.838G>A (p.Glu280Lys)

NM_000277.1(PAH):c.331C>T (p.Arg111Ter)

NM_000277.1(PAH):c.782G>A (p.Arg261Gln)

NM_000277.1(PAH):c.754C>T (p.Arg252Trp)

NM_000277.1(PAH):c.473G>A (p.Arg158Gln)

NM_000277.1(PAH):c.727C>T (p.Arg243Ter)

NM_000277.1(PAH):c.842C>T (p.Pro281Leu)

NM_000277.1(PAH):c.728G>A (p.Arg243Gln)

NM_000277.1(PAH):c.1066-11G>A

NM_000277.1(PAH):c.781C>T (p.Arg261Ter)

NM_000277.1(PAH):c.1223G>A (p.Arg408Gln)

NM_000277.1(PAH):c.1162G>A (p.Val388Met)

NM_000277.1(PAH):c.1066-3C>T

NM_000277.1(PAH):c.1208C>T (p.Ala403Val)

NM_000277.1(PAH):c.890G>A (p.Arg297His)

NM_000277.1(PAH):c.926C>T (p.Ala309Val)

NM_000277.1(PAH):c.441+1G>A

NM_000277.1(PAH):c.526C>T (p.Arg176Ter)

NM_000277.1(PAH):c.688G>A (p.Val230Ile)

NM_000277.1(PAH):c.721C>T (p.Arg241Cys)

NM_000277.1(PAH):c.745C>T (p.Leu249Phe)

NM_000277.1(PAH):c.442-1G>A

NM_000277.1(PAH):c.842+1G>A

NM_000277.1(PAH):c.776C>T (p.Ala259Val)

NM_000277.1(PAH):c.1200-1G>A

NM_000277.1(PAH):c.912+1G>A

NM_000277.1(PAH):c.1065+1G>A

NM_000277.1(PAH):c.472C>T (p.Arg158Trp)

NM_000277.1(PAH):c.755G>A (p.Arg252Gln)

NM_000277.1(PAH):c.809G>A (p.Arg270Lys)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Phenylketonuria by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in the PAH gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Kidney Diseases

[01076] Pathogenic G-to-A or C-to-T mutations/SNPs associated with various kidney diseases are reported in the ClinVar database and disclosed in Table A, including but not limited to Autosomal recessive polycystic kidney disease and Renal carnitine transport defect. Accordingly, an aspect of the invention relates to a method for correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with any of these diseases, as discussed below.

Autosomal recessive polycystic kidney disease

[01077] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Autosomal recessive polycystic kidney disease. In some embodiment, the pathogenic mutations/SNPs are present in at least the PKHD1 gene, including at least the followings:

NM_138694.3(PKHD1):c.10444C>T (p.Arg3482Cys)

NM_138694.3(PKHD1):c.9319C>T (p.Arg3107Ter)

NM_138694.3(PKHD1):c.1480C>T (p.Arg494Ter)

NM_138694.3(PKHD1):c.707+1G>A

NM_138694.3(PKHD1):c.1486C>T (p.Arg496Ter)

NM_138694.3(PKHD1):c.8303-1G>A

NM_138694.3(PKHD1):c.2854G>A (p.Gly952Arg)

NM_138694.3(PKHD1):c.7194G>A (p.Trp2398Ter)

NM_138694.3(PKHD1):c.10219C>T (p.Gln3407Ter)

NM_138694.3(PKHD1):c.107C>T (p.Thr36Met)

NM_138694.3(PKHD1):c.8824C>T (p.Arg2942Ter)

NM_138694.3(PKHD1):c.982C>T (p.Arg328Ter)

NM_138694.3(PKHD1):c.4870C>T (p.Arg1624Trp)

NM_138694.3(PKHD1):c.1602+1G>A

NM_138694.3(PKHD1):c.1694-1G>A

NM_138694.3(PKHD1):c.2341C>T (p.Arg781Ter)

NM_138694.3(PKHD1):c.2407+1G>A

NM_138694.3(PKHD1):c.2452C>T (p.Gln818Ter)

NM_138694.3(PKHD1):c.5236+1G>A

NM_138694.3(PKHD1):c.6499C>T (p.Gln2167Ter)

NM_138694.3(PKHD1):c.2725C>T (p.Arg909Ter)

NM_138694.3(PKHD1):c.370C>T (p.Arg124Ter)

NM_138694.3(PKHD1):c.2810G>A (p.Trp937Ter)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Autosomal recessive polycystic kidney disease by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in the PKHD1 gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Renal carnitine transport defect

[01078] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Renal carnitine transport defect. In some embodiment, the pathogenic mutations/SNPs are present in at least the SLC22A5 gene, including at least the followings:

NM_003060.3(SLC22A5):c.760C>T (p.Arg254Ter)

NM_003060.3(SLC22A5):c.396G>A (p.Trp132Ter)

NM_003060.3(SLC22A5):c.844C>T (p.Arg282Ter)

NM_003060.3(SLC22A5):c.505C>T (p.Arg169Trp)

NM_003060.3(SLC22A5):c.1319C>T (p.Thr440Met)

NM_003060.3(SLC22A5):c.1195C>T (p.Arg399Trp)

NM_003060.3(SLC22A5):c.695C>T (p.Thr232Met)

NM_003060.3(SLC22A5):c.845G>A (p.Arg282Gln)

NM_003060.3(SLC22A5):c.1193C>T (p.Pro398Leu)

NM_003060.3(SLC22A5):c.1463G>A (p.Arg488His)

NM_003060.3(SLC22A5):c.338G>A (p.Cys113Tyr)

NM_003060.3(SLC22A5):c.136C>T (p.Pro46Ser)

NM_003060.3(SLC22A5):c.506G>A (p.Arg169Gln)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Renal carnitine transport defect by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in the SLC22A5 gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Cardiovascular Disease

[01079] The embodiments disclosed herein can directly be used to treat or prevent cardiovascular diseases for known targets. Khera et al. (Nat Rev Genet. 2017 Jun;18(6):331-344. doi: 10.1038/nrg.2016.160. Epub 2017 Mar 13) described common variant association studies linking approximately 60 genetic loci to coronary risk used to facilitate a better understanding of causal risk factors, underlying biology development of new therapeutics. Khera explains, for example that inactivating mutations in PCSK9 decreased levels of circulating LDL cholesterol and reduced risk of CAD leading to intense interest in development of PCSK9 inhibitors. Further, antisense oligonucleotides designed to mimic protective mutations in APOC3 or LPA demonstrated a ~70% reduction in triglyceride levels and 80% reduction in circulating lipoprotein(a) levels, respectively. In addition, Wang et al., (Arterioscler Thromb Vasc Biol. 2016 May;36(5):783-6. doi: 10.1161/ATVBAHA.116.307227. Epub 2016 Mar 3) and Ding et al. (Circ Res. 2014 Aug 15;115(5):488-92. doi: 10.1161/CIRCRESAHA.115.304351. Epub 2014 Jun 10.) report the use of CRISPR to target the gene Pcsk9 for the prevention of cardiovascular disease.

Muscle Diseases

[01080] Pathogenic G-to-A or C-to-T mutations/SNPs associated with various muscle diseases are reported in the ClinVar database and disclosed in Table A, including but not limited to Duchenne muscular dystrophy, Becker muscular dystrophy, Limb-girdle muscular dystrophy, Emery-Dreifuss muscular dystrophy, and Facioscapulohumeral muscular dystrophy. Accordingly, an aspect of the invention relates to a method for correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with any of these diseases, as discussed below.

Duchenne muscular dystrophy

[01081] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Duchenne muscular dystrophy. In some embodiment, the pathogenic mutations/SNPs are present in at least the DMD gene, including at least the followings:

NM_004006.2(DMD):c.2797C>T (p.Gln933Ter)

NM_004006.2(DMD):c.4870C>T (p.Gln1624Ter)
NM_004006.2(DMD):c.5551C>T (p.Gln1851Ter)
NM_004006.2(DMD):c.3188G>A (p.Trp1063Ter)
NM_004006.2(DMD):c.8357G>A (p.Trp2786Ter)
NM_004006.2(DMD):c.7817G>A (p.Trp2606Ter)
NM_004006.2(DMD):c.7755G>A (p.Trp2585Ter)
NM_004006.2(DMD):c.5917C>T (p.Gln1973Ter)
NM_004006.2(DMD):c.5641C>T (p.Gln1881Ter)
NM_004006.2(DMD):c.5131C>T (p.Gln1711Ter)
NM_004006.2(DMD):c.4240C>T (p.Gln1414Ter)
NM_004006.2(DMD):c.3427C>T (p.Gln1143Ter)
NM_004006.2(DMD):c.2407C>T (p.Gln803Ter)
NM_004006.2(DMD):c.2368C>T (p.Gln790Ter)
NM_004006.2(DMD):c.1683G>A (p.Trp561Ter)
NM_004006.2(DMD):c.1663C>T (p.Gln555Ter)
NM_004006.2(DMD):c.1388G>A (p.Trp463Ter)
NM_004006.2(DMD):c.1331+1G>A
NM_004006.2(DMD):c.1324C>T (p.Gln442Ter)
NM_004006.2(DMD):c.355C>T (p.Gln119Ter)
NM_004006.2(DMD):c.94-1G>A
NM_004006.2(DMD):c.5506C>T (p.Gln1836Ter)
NM_004006.2(DMD):c.1504C>T (p.Gln502Ter)
NM_004006.2(DMD):c.5032C>T (p.Gln1678Ter)
NM_004006.2(DMD):c.457C>T (p.Gln153Ter)
NM_004006.2(DMD):c.1594C>T (p.Gln532Ter)
NM_004006.2(DMD):c.1150-1G>A
NM_004006.2(DMD):c.6223C>T (p.Gln2075Ter)
NM_004006.2(DMD):c.3747G>A (p.Trp1249Ter)
NM_004006.2(DMD):c.2861G>A (p.Trp954Ter)
NM_004006.2(DMD):c.9563+1G>A
NM_004006.2(DMD):c.4483C>T (p.Gln1495Ter)
NM_004006.2(DMD):c.4312C>T (p.Gln1438Ter)
NM_004006.2(DMD):c.8209C>T (p.Gln2737Ter)
NM_004006.2(DMD):c.4071+1G>A

NM_004006.2(DMD):c.2665C>T (p.Arg889Ter)

NM_004006.2(DMD):c.2202G>A (p.Trp734Ter)

NM_004006.2(DMD):c.2077C>T (p.Gln693Ter)

NM_004006.2(DMD):c.1653G>A (p.Trp551Ter)

NM_004006.2(DMD):c.1061G>A (p.Trp354Ter)

NM_004006.2(DMD):c.8914C>T (p.Gln2972Ter)

NM_004006.2(DMD):c.6118-1G>A

NM_004006.2(DMD):c.4729C>T (p.Arg1577Ter)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Duchenne muscular dystrophy by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in the DMD gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Becker muscular dystrophy

[01082] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Becker muscular dystrophy. In some embodiment, the pathogenic mutations/SNPs are present in at least the DMD gene, including at least the followings:

NM_004006.2(DMD):c.3413G>A (p.Trp1138Ter)

NM_004006.2(DMD):c.358-1G>A

NM_004006.2(DMD):c.10108C>T (p.Arg3370Ter)

NM_004006.2(DMD):c.6373C>T (p.Gln2125Ter)

NM_004006.2(DMD):c.9568C>T (p.Arg3190Ter)

NM_004006.2(DMD):c.8713C>T (p.Arg2905Ter)

NM_004006.2(DMD):c.1615C>T (p.Arg539Ter)

NM_004006.2(DMD):c.3151C>T (p.Arg1051Ter)

NM_004006.2(DMD):c.3432+1G>A

NM_004006.2(DMD):c.5287C>T (p.Arg1763Ter)

NM_004006.2(DMD):c.5530C>T (p.Arg1844Ter)

NM_004006.2(DMD):c.8608C>T (p.Arg2870Ter)

NM_004006.2(DMD):c.8656C>T (p.Gln2886Ter)

NM_004006.2(DMD):c.8944C>T (p.Arg2982Ter)

NM_004006.2(DMD):c.5899C>T (p.Arg1967Ter)

NM_004006.2(DMD):c.10033C>T (p.Arg3345Ter)

NM_004006.2(DMD):c.10086+1G>A

NM_004019.2(DMD):c.1020G>A (p.Thr340=)

NM_004006.2(DMD):c.1261C>T (p.Gln421Ter)

NM_004006.2(DMD):c.1465C>T (p.Gln489Ter)

NM_004006.2(DMD):c.1990C>T (p.Gln664Ter)

NM_004006.2(DMD):c.2032C>T (p.Gln678Ter)

NM_004006.2(DMD):c.2332C>T (p.Gln778Ter)

NM_004006.2(DMD):c.2419C>T (p.Gln807Ter)

NM_004006.2(DMD):c.2650C>T (p.Gln884Ter)

NM_004006.2(DMD):c.2804-1G>A

NM_004006.2(DMD):c.3276+1G>A

NM_004006.2(DMD):c.3295C>T (p.Gln1099Ter)

NM_004006.2(DMD):c.336G>A (p.Trp112Ter)

NM_004006.2(DMD):c.3580C>T (p.Gln1194Ter)

NM_004006.2(DMD):c.4117C>T (p.Gln1373Ter)

NM_004006.2(DMD):c.649+1G>A

NM_004006.2(DMD):c.6906G>A (p.Trp2302Ter)

NM_004006.2(DMD):c.7189C>T (p.Gln2397Ter)

NM_004006.2(DMD):c.7309+1G>A

NM_004006.2(DMD):c.7657C>T (p.Arg2553Ter)

NM_004006.2(DMD):c.7682G>A (p.Trp2561Ter)

NM_004006.2(DMD):c.7683G>A (p.Trp2561Ter)

NM_004006.2(DMD):c.7894C>T (p.Gln2632Ter)

NM_004006.2(DMD):c.9361+1G>A

NM_004006.2(DMD):c.9564-1G>A

NM_004006.2(DMD):c.2956C>T (p.Gln986Ter)

NM_004006.2(DMD):c.883C>T (p.Arg295Ter)

NM_004006.2(DMD):c.31+36947G>A

NM_004006.2(DMD):c.10279C>T (p.Gln3427Ter)

NM_004006.2(DMD):c.433C>T (p.Arg145Ter)

NM_004006.2(DMD):c.9G>A (p.Trp3Ter)

NM_004006.2(DMD):c.10171C>T (p.Arg3391Ter)

NM_004006.2(DMD):c.583C>T (p.Arg195Ter)

NM_004006.2(DMD):c.9337C>T (p.Arg3113Ter)

NM_004006.2(DMD):c.8038C>T (p.Arg2680Ter)
NM_004006.2(DMD):c.1812+1G>A
NM_004006.2(DMD):c.1093C>T (p.Gln365Ter)
NM_004006.2(DMD):c.1704+1G>A
NM_004006.2(DMD):c.1912C>T (p.Gln638Ter)
NM_004006.2(DMD):c.133C>T (p.Gln45Ter)
NM_004006.2(DMD):c.5868G>A (p.Trp1956Ter)
NM_004006.2(DMD):c.565C>T (p.Gln189Ter)
NM_004006.2(DMD):c.5089C>T (p.Gln1697Ter)
NM_004006.2(DMD):c.2512C>T (p.Gln838Ter)
NM_004006.2(DMD):c.10477C>T (p.Gln3493Ter)
NM_004006.2(DMD):c.93+1G>A
NM_004006.2(DMD):c.4174C>T (p.Gln1392Ter)
NM_004006.2(DMD):c.3940C>T (p.Arg1314Ter)See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Becker muscular dystrophy by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in the DMD gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Limb-girdle muscular dystrophy

[01083] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Limb-girdle muscular dystrophy. In some embodiment, the pathogenic mutations/SNPs are present in at least one gene selected from SGCB, MYOT, LMNA, CAPN3, DYSF, SGCA, TTN, ANO5, TRAPPC11, LMNA, POMT1, and FKRP, including at least the followings:
NM_000232.4(SGCB):c.31C>T (p.Gln11Ter)
NM_006790.2(MYOT):c.164C>T (p.Ser55Phe)
NM_006790.2(MYOT):c.170C>T (p.Thr57Ile)
NM_170707.3(LMNA):c.1488+1G>A
NM_170707.3(LMNA):c.1609-1G>A
NM_000070.2(CAPN3):c.1715G>A (p.Arg572Gln)
NM_000070.2(CAPN3):c.2243G>A (p.Arg748Gln)
NM_000070.2(CAPN3):c.145C>T (p.Arg49Cys)
NM_000070.2(CAPN3):c.1319G>A (p.Arg440Gln)
NM_000070.2(CAPN3):c.1343G>A (p.Arg448His)

NM_000070.2(CAPN3):c.1465C>T (p.Arg489Trp)

NM_000070.2(CAPN3):c.1714C>T (p.Arg572Trp)

NM_000070.2(CAPN3):c.2306G>A (p.Arg769Gln)

NM_000070.2(CAPN3):c.133G>A (p.Ala45Thr)

NM_000070.2(CAPN3):c.499-1G>A

NM_000070.2(CAPN3):c.439C>T (p.Arg147Ter)

NM_000070.2(CAPN3):c.1063C>T (p.Arg355Trp)

NM_000070.2(CAPN3):c.1250C>T (p.Thr417Met)

NM_000070.2(CAPN3):c.245C>T (p.Pro82Leu)

NM_000070.2(CAPN3):c.2242C>T (p.Arg748Ter)

NM_000070.2(CAPN3):c.1318C>T (p.Arg440Trp)

NM_000070.2(CAPN3):c.1333G>A (p.Gly445Arg)

NM_000070.2(CAPN3):c.1957C>T (p.Gln653Ter)

NM_000070.2(CAPN3):c.1801-1G>A

NM_000070.2(CAPN3):c.2263+1G>A

NM_000070.2(CAPN3):c.956C>T (p.Pro319Leu)

NM_000070.2(CAPN3):c.1468C>T (p.Arg490Trp)

NM_000070.2(CAPN3):c.802-9G>A

NM_000070.2(CAPN3):c.1342C>T (p.Arg448Cys)

NM_000070.2(CAPN3):c.1303G>A (p.Glu435Lys)

NM_000070.2(CAPN3):c.1993-1G>A

NM_003494.3(DYSF):c.3113G>A (p.Arg1038Gln)

NM_001130987.1(DYSF):c.5174+1G>A

NM_001130987.1(DYSF):c.159G>A (p.Trp53Ter)

NM_001130987.1(DYSF):c.2929C>T (p.Arg977Trp)

NM_001130987.1(DYSF):c.4282C>T (p.Gln1428Ter)

NM_001130987.1(DYSF):c.1577-1G>A

NM_003494.3(DYSF):c.5529G>A (p.Trp1843Ter)

NM_001130987.1(DYSF):c.1576+1G>A

NM_001130987.1(DYSF):c.4462C>T (p.Gln1488Ter)

NM_003494.3(DYSF):c.5429G>A (p.Arg1810Lys)

NM_003494.3(DYSF):c.5077C>T (p.Arg1693Trp)

NM_001130978.1(DYSF):c.1813C>T (p.Gln605Ter)

NM_003494.3(DYSF):c.3230G>A (p.Trp1077Ter)

NM_003494.3(DYSF):c.265C>T (p.Arg89Ter)

NM_003494.3(DYSF):c.4434G>A (p.Trp1478Ter)

NM_003494.3(DYSF):c.3478C>T (p.Gln1160Ter)

NM_001130987.1(DYSF):c.1372G>A (p.Gly458Arg)

NM_003494.3(DYSF):c.4090C>T (p.Gln1364Ter)

NM_001130987.1(DYSF):c.2409+1G>A

NM_003494.3(DYSF):c.1708C>T (p.Gln570Ter)

NM_003494.3(DYSF):c.1956G>A (p.Trp652Ter)

NM_001130987.1(DYSF):c.5004-1G>A

NM_003494.3(DYSF):c.331C>T (p.Gln111Ter)

NM_001130978.1(DYSF):c.5776C>T (p.Arg1926Ter)

NM_003494.3(DYSF):c.6124C>T (p.Arg2042Cys)

NM_003494.3(DYSF):c.2643+1G>A

NM_003494.3(DYSF):c.4253G>A (p.Gly1418Asp)

NM_003494.3(DYSF):c.610C>T (p.Arg204Ter)

NM_003494.3(DYSF):c.1834C>T (p.Gln612Ter)

NM_003494.3(DYSF):c.5668-7G>A

NM_001130978.1(DYSF):c.3137G>A (p.Arg1046His)

NM_003494.3(DYSF):c.1053+1G>A

NM_003494.3(DYSF):c.1398-1G>A

NM_003494.3(DYSF):c.1481-1G>A

NM_003494.3(DYSF):c.2311C>T (p.Gln771Ter)

NM_003494.3(DYSF):c.2869C>T (p.Gln957Ter)

NM_003494.3(DYSF):c.4756C>T (p.Arg1586Ter)

NM_003494.3(DYSF):c.5509G>A (p.Asp1837Asn)

NM_003494.3(DYSF):c.5644C>T (p.Gln1882Ter)

NM_003494.3(DYSF):c.5946+1G>A

NM_003494.3(DYSF):c.937+1G>A

NM_003494.3(DYSF):c.5266C>T (p.Gln1756Ter)

NM_003494.3(DYSF):c.3832C>T (p.Gln1278Ter)

NM_003494.3(DYSF):c.5525+1G>A

NM_003494.3(DYSF):c.3112C>T (p.Arg1038Ter)

NM_000023.3(SGCA):c.293G>A (p.Arg98His)

NM_000023.3(SGCA):c.850C>T (p.Arg284Cys)

NM_000023.3(SGCA):c.403C>T (p.Gln135Ter)

NM_000023.3(SGCA):c.409G>A (p.Glu137Lys)

NM_000023.3(SGCA):c.747+1G>A

NM_000023.3(SGCA):c.229C>T (p.Arg77Cys)

NM_000023.3(SGCA):c.101G>A (p.Arg34His)

NM_000023.3(SGCA):c.739G>A (p.Val247Met)

NM_001256850.1(TTN):c.87394C>T (p.Arg29132Ter)

NM_213599.2(ANO5):c.762+1G>A

NM_213599.2(ANO5):c.1213C>T (p.Gln405Ter)

NM_213599.2(ANO5):c.1639C>T (p.Arg547Ter)

NM_213599.2(ANO5):c.1406G>A (p.Trp469Ter)

NM_213599.2(ANO5):c.1210C>T (p.Arg404Ter)

NM_213599.2(ANO5):c.2272C>T (p.Arg758Cys)

NM_213599.2(ANO5):c.41-1G>A

NM_213599.2(ANO5):c.172C>T (p.Arg58Trp)

NM_213599.2(ANO5):c.1898+1G>A

NM_021942.5(TRAPPC11):c.1287+5G>A

NM_170707.3(LMNA):c.1608+1G>A

NM_007171.3(POMT1):c.1864C>T (p.Arg622Ter)

NM_024301.4(FKRP):c.313C>T (p.Gln105Ter)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Limb-girdle muscular dystrophy by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in at least one gene selected from SGCB, MYOT, LMNA, CAPN3, DYSF, SGCA, TTN, ANO5, TRAPPC11, LMNA, POMT1, and FKRP, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Emery-Dreifuss muscular dystrophy

[01084] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Emery-Dreifuss muscular dystrophy. In some embodiment, the pathogenic mutations/SNPs are present in at least the EMD or SYNE1 gene, including at least the followings:

NM_000117.2(EMD):c.3G>A (p.Met1Ile)

NM_033071.3(SYNE1):c.11908C>T (p.Arg3970Ter)

NM_033071.3(SYNE1):c.21721C>T (p.Gln7241Ter)

NM_000117.2(EMD):c.130C>T (p.Gln44Ter)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Emery-Dreifuss muscular dystrophy by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in the EMD or SYNE1 gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Facioscapulohumeral muscular dystrophy

[01085] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Facioscapulohumeral muscular dystrophy. In some embodiment, the pathogenic mutations/SNPs are present in at least the SMCHD1 gene, including at least the followings:

NM_015295.2(SMCHD1):c.3801+1G>A

NM_015295.2(SMCHD1):c.1843-1G>A

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Facioscapulohumeral muscular dystrophy by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in the SMCHD1 gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Inborn Errors of Metabolism (IEM)

[01086] Pathogenic G-to-A or C-to-T mutations/SNPs associated with various IEMs are reported in the ClinVar database and disclosed in Table A, including but not limited to Primary hyperoxaluria type 1, Argininosuccinate lyase deficiency, Ornithine carbamoyltransferase deficiency, and Maple syrup urine disease. Accordingly, an aspect of the invention relates to a method for correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with any of these diseases, as discussed below.

***Primary hyperoxaluria type* 1**

[01087] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Primary hyperoxaluria type 1. In some embodiment, the pathogenic mutations/SNPs are present in at least the AGXT gene, including at least the followings:

NM_000030.2(AGXT):c.245G>A (p.Gly82Glu)

NM_000030.2(AGXT):c.698G>A (p.Arg233His)

NM_000030.2(AGXT):c.466G>A (p.Gly156Arg)

NM_000030.2(AGXT):c.106C>T (p.Arg36Cys)

NM_000030.2(AGXT):c.346G>A (p.Gly116Arg)

NM_000030.2(AGXT):c.568G>A (p.Gly190Arg)

NM_000030.2(AGXT):c.653C>T (p.Ser218Leu)

NM_000030.2(AGXT):c.737G>A (p.Trp246Ter)

NM_000030.2(AGXT):c.1049G>A (p.Gly350Asp)

NM_000030.2(AGXT):c.473C>T (p.Ser158Leu)

NM_000030.2(AGXT):c.907C>T (p.Gln303Ter)

NM_000030.2(AGXT):c.996G>A (p.Trp332Ter)

NM_000030.2(AGXT):c.508G>A (p.Gly170Arg)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Primary hyperoxaluria type 1 by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in the AGXT gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Argininosuccinate lyase deficiency

[01088] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Argininosuccinate lyase deficiency. In some embodiment, the pathogenic mutations/SNPs are present in at least the ASL gene, including at least the followings:

NM_001024943.1(ASL):c.1153C>T (p.Arg385Cys)

NM_000048.3(ASL):c.532G>A (p.Val178Met)

NM_000048.3(ASL):c.545G>A (p.Arg182Gln)

NM_000048.3(ASL):c.175G>A (p.Glu59Lys)

NM_000048.3(ASL):c.718+5G>A

NM_000048.3(ASL):c.889C>T (p.Arg297Trp)

NM_000048.3(ASL):c.1360C>T (p.Gln454Ter)

NM_000048.3(ASL):c.1060C>T (p.Gln354Ter)

NM_000048.3(ASL):c.35G>A (p.Arg12Gln)

NM_000048.3(ASL):c.446+1G>A

NM_000048.3(ASL):c.544C>T (p.Arg182Ter)

NM_000048.3(ASL):c.1135C>T (p.Arg379Cys)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Argininosuccinate lyase deficiency by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T

mutations/SNPs present in the ASL gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Ornithine carbamoyltransferase deficiency

[01089] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Ornithine carbamoyltransferase deficiency. In some embodiment, the pathogenic mutations/SNPs are present in at least the OTC gene, including at least the followings:

NM_000531.5(OTC):c.119G>A (p.Arg40His)

NM_000531.5(OTC):c.422G>A (p.Arg141Gln)

NM_000531.5(OTC):c.829C>T (p.Arg277Trp)

NM_000531.5(OTC):c.674C>T (p.Pro225Leu)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Ornithine carbamoyltransferase deficiency by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in the OTC gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Maple syrup urine disease

[01090] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Maple syrup urine disease. In some embodiment, the pathogenic mutations/SNPs are present in at least one gene selected from BCKDHA, BCKDHB, DBT, and DLD, including at least the followings:

NM_000709.3(BCKDHA):c.476G>A (p.Arg159Gln)

NM_183050.3(BCKDHB):c.3G>A (p.Met1Ile)

NM_183050.3(BCKDHB):c.554C>T (p.Pro185Leu)

NM_001918.3(DBT):c.1033G>A (p.Gly345Arg)

NM_000709.3(BCKDHA):c.940C>T (p.Arg314Ter)

NM_000709.3(BCKDHA):c.793C>T (p.Arg265Trp)

NM_000709.3(BCKDHA):c.868G>A (p.Gly290Arg)

NM_000108.4(DLD):c.1123G>A (p.Glu375Lys)

NM_000709.3(BCKDHA):c.1234G>A (p.Val412Met)

NM_000709.3(BCKDHA):c.288+1G>A

NM_000709.3(BCKDHA):c.979G>A (p.Glu327Lys)

NM_001918.3(DBT):c.901C>T (p.Arg301Cys)

NM_183050.3(BCKDHB):c.509G>A (p.Arg170His)

NM_183050.3(BCKDHB):c.799C>T (p.Gln267Ter)

NM_183050.3(BCKDHB):c.853C>T (p.Arg285Ter)

NM_183050.3(BCKDHB):c.970C>T (p.Arg324Ter)

NM_183050.3(BCKDHB):c.832G>A (p.Gly278Ser)

NM_000709.3(BCKDHA):c.1036C>T (p.Arg346Cys)

NM_000709.3(BCKDHA):c.288+9C>T

NM_000709.3(BCKDHA):c.632C>T (p.Thr211Met)

NM_000709.3(BCKDHA):c.659C>T (p.Ala220Val)

NM_000709.3(BCKDHA):c.964C>T (p.Gln322Ter)

NM_001918.3(DBT):c.1291C>T (p.Arg431Ter)

NM_001918.3(DBT):c.251G>A (p.Trp84Ter)

NM_001918.3(DBT):c.871C>T (p.Arg291Ter)

NM_000056.4(BCKDHB):c.1016C>T (p.Ser339Leu)

NM_000056.4(BCKDHB):c.344-1G>A

NM_000056.4(BCKDHB):c.633+1G>A

NM_000056.4(BCKDHB):c.952-1G>A

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Maple syrup urine disease by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in at least one gene selected from BCKDHA, BCKDHB, DBT, and DLD, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Cancer-Related Diseases

[01091] Pathogenic G-to-A or C-to-T mutations/SNPs associated with various cancers and cancer-related diseases are reported in the ClinVar database and disclosed in Table A, including but not limited to Breast-Ovarian Cancer and Lynch syndrome. Accordingly, an aspect of the invention relates to a method for correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with any of these diseases, as discussed below.

Breast-Ovarian Cancer

[01092] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Breast-Ovarian Cancer. In some embodiment, the pathogenic mutations/SNPs are present in at least the BRCA1or BRCA2 gene, including at least the followings:

NM_007294.3(BRCA1):c.5095C>T (p.Arg1699Trp)

NM_000059.3(BRCA2):c.7558C>T (p.Arg2520Ter)
NM_007294.3(BRCA1):c.2572C>T (p.Gln858Ter)
NM_007294.3(BRCA1):c.3607C>T (p.Arg1203Ter)
NM_007294.3(BRCA1):c.5503C>T (p.Arg1835Ter)
NM_007294.3(BRCA1):c.2059C>T (p.Gln687Ter)
NM_007294.3(BRCA1):c.4675+1G>A
NM_007294.3(BRCA1):c.5251C>T (p.Arg1751Ter)
NM_007294.3(BRCA1):c.5444G>A (p.Trp1815Ter)
NM_000059.3(BRCA2):c.9318G>A (p.Trp3106Ter)
NM_000059.3(BRCA2):c.9382C>T (p.Arg3128Ter)
NM_000059.3(BRCA2):c.274C>T (p.Gln92Ter)
NM_000059.3(BRCA2):c.6952C>T (p.Arg2318Ter)
NM_007294.3(BRCA1):c.1687C>T (p.Gln563Ter)
NM_007294.3(BRCA1):c.2599C>T (p.Gln867Ter)
NM_007294.3(BRCA1):c.784C>T (p.Gln262Ter)
NM_007294.3(BRCA1):c.280C>T (p.Gln94Ter)
NM_007294.3(BRCA1):c.5542C>T (p.Gln1848Ter)
NM_007294.3(BRCA1):c.5161C>T (p.Gln1721Ter)
NM_007294.3(BRCA1):c.4573C>T (p.Gln1525Ter)
NM_007294.3(BRCA1):c.4270C>T (p.Gln1424Ter)
NM_007294.3(BRCA1):c.4225C>T (p.Gln1409Ter)
NM_007294.3(BRCA1):c.4066C>T (p.Gln1356Ter)
NM_007294.3(BRCA1):c.3679C>T (p.Gln1227Ter)
NM_007294.3(BRCA1):c.1918C>T (p.Gln640Ter)
NM_007294.3(BRCA1):c.963G>A (p.Trp321Ter)
NM_007294.3(BRCA1):c.718C>T (p.Gln240Ter)
NM_000059.3(BRCA2):c.9196C>T (p.Gln3066Ter)
NM_000059.3(BRCA2):c.9154C>T (p.Arg3052Trp)
NM_007294.3(BRCA1):c.3991C>T (p.Gln1331Ter)
NM_007294.3(BRCA1):c.4097-1G>A
NM_007294.3(BRCA1):c.1059G>A (p.Trp353Ter)
NM_007294.3(BRCA1):c.1115G>A (p.Trp372Ter)
NM_007294.3(BRCA1):c.1138C>T (p.Gln380Ter)
NM_007294.3(BRCA1):c.1612C>T (p.Gln538Ter)

NM_007294.3(BRCA1):c.1621C>T (p.Gln541Ter)
NM_007294.3(BRCA1):c.1630C>T (p.Gln544Ter)
NM_007294.3(BRCA1):c.178C>T (p.Gln60Ter)
NM_007294.3(BRCA1):c.1969C>T (p.Gln657Ter)
NM_007294.3(BRCA1):c.2275C>T (p.Gln759Ter)
NM_007294.3(BRCA1):c.2410C>T (p.Gln804Ter)
NM_007294.3(BRCA1):c.2869C>T (p.Gln957Ter)
NM_007294.3(BRCA1):c.2923C>T (p.Gln975Ter)
NM_007294.3(BRCA1):c.3268C>T (p.Gln1090Ter)
NM_007294.3(BRCA1):c.3430C>T (p.Gln1144Ter)
NM_007294.3(BRCA1):c.3544C>T (p.Gln1182Ter)
NM_007294.3(BRCA1):c.4075C>T (p.Gln1359Ter)
NM_007294.3(BRCA1):c.4201C>T (p.Gln1401Ter)
NM_007294.3(BRCA1):c.4399C>T (p.Gln1467Ter)
NM_007294.3(BRCA1):c.4552C>T (p.Gln1518Ter)
NM_007294.3(BRCA1):c.5054C>T (p.Thr1685Ile)
NM_007294.3(BRCA1):c.514C>T (p.Gln172Ter)
NM_007294.3(BRCA1):c.5239C>T (p.Gln1747Ter)
NM_007294.3(BRCA1):c.5266C>T (p.Gln1756Ter)
NM_007294.3(BRCA1):c.5335C>T (p.Gln1779Ter)
NM_007294.3(BRCA1):c.5345G>A (p.Trp1782Ter)
NM_007294.3(BRCA1):c.5511G>A (p.Trp1837Ter)
NM_007294.3(BRCA1):c.5536C>T (p.Gln1846Ter)
NM_007294.3(BRCA1):c.55C>T (p.Gln19Ter)
NM_007294.3(BRCA1):c.949C>T (p.Gln317Ter)
NM_007294.3(BRCA1):c.928C>T (p.Gln310Ter)
NM_007294.3(BRCA1):c.5117G>A (p.Gly1706Glu)
NM_007294.3(BRCA1):c.5136G>A (p.Trp1712Ter)
NM_007294.3(BRCA1):c.4327C>T (p.Arg1443Ter)
NM_007294.3(BRCA1):c.1471C>T (p.Gln491Ter)
NM_007294.3(BRCA1):c.1576C>T (p.Gln526Ter)
NM_007294.3(BRCA1):c.160C>T (p.Gln54Ter)
NM_007294.3(BRCA1):c.2683C>T (p.Gln895Ter)
NM_007294.3(BRCA1):c.2761C>T (p.Gln921Ter)

NM_007294.3(BRCA1):c.3895C>T (p.Gln1299Ter)
NM_007294.3(BRCA1):c.4339C>T (p.Gln1447Ter)
NM_007294.3(BRCA1):c.4372C>T (p.Gln1458Ter)
NM_007294.3(BRCA1):c.5153G>A (p.Trp1718Ter)
NM_007294.3(BRCA1):c.5445G>A (p.Trp1815Ter)
NM_007294.3(BRCA1):c.5510G>A (p.Trp1837Ter)
NM_007294.3(BRCA1):c.5346G>A (p.Trp1782Ter)
NM_007294.3(BRCA1):c.1116G>A (p.Trp372Ter)
NM_007294.3(BRCA1):c.1999C>T (p.Gln667Ter)
NM_007294.3(BRCA1):c.4183C>T (p.Gln1395Ter)
NM_007294.3(BRCA1):c.4810C>T (p.Gln1604Ter)
NM_007294.3(BRCA1):c.850C>T (p.Gln284Ter)
NM_007294.3(BRCA1):c.1058G>A (p.Trp353Ter)
NM_007294.3(BRCA1):c.131G>A (p.Cys44Tyr)
NM_007294.3(BRCA1):c.1600C>T (p.Gln534Ter)
NM_007294.3(BRCA1):c.3286C>T (p.Gln1096Ter)
NM_007294.3(BRCA1):c.3403C>T (p.Gln1135Ter)
NM_007294.3(BRCA1):c.34C>T (p.Gln12Ter)
NM_007294.3(BRCA1):c.4258C>T (p.Gln1420Ter)
NM_007294.3(BRCA1):c.4609C>T (p.Gln1537Ter)
NM_007294.3(BRCA1):c.5154G>A (p.Trp1718Ter)
NM_007294.3(BRCA1):c.5431C>T (p.Gln1811Ter)
NM_007294.3(BRCA1):c.241C>T (p.Gln81Ter)
NM_007294.3(BRCA1):c.3331C>T (p.Gln1111Ter)
NM_007294.3(BRCA1):c.3967C>T (p.Gln1323Ter)
NM_007294.3(BRCA1):c.415C>T (p.Gln139Ter)
NM_007294.3(BRCA1):c.505C>T (p.Gln169Ter)
NM_007294.3(BRCA1):c.5194-12G>A
NM_007294.3(BRCA1):c.5212G>A (p.Gly1738Arg)
NM_007294.3(BRCA1):c.5332+1G>A
NM_007294.3(BRCA1):c.1480C>T (p.Gln494Ter)
NM_007294.3(BRCA1):c.2563C>T (p.Gln855Ter)
NM_007294.3(BRCA1):c.1066C>T (p.Gln356Ter)
NM_007294.3(BRCA1):c.3718C>T (p.Gln1240Ter)

NM_007294.3(BRCA1):c.3817C>T (p.Gln1273Ter)

NM_007294.3(BRCA1):c.3937C>T (p.Gln1313Ter)

NM_007294.3(BRCA1):c.4357+1G>A

NM_007294.3(BRCA1):c.5074+1G>A

NM_007294.3(BRCA1):c.5277+1G>A

NM_007294.3(BRCA1):c.2338C>T (p.Gln780Ter)

NM_007294.3(BRCA1):c.3598C>T (p.Gln1200Ter)

NM_007294.3(BRCA1):c.3841C>T (p.Gln1281Ter)

NM_007294.3(BRCA1):c.4222C>T (p.Gln1408Ter)

NM_007294.3(BRCA1):c.4524G>A (p.Trp1508Ter)

NM_007294.3(BRCA1):c.5353C>T (p.Gln1785Ter)

NM_007294.3(BRCA1):c.962G>A (p.Trp321Ter)

NM_007294.3(BRCA1):c.220C>T (p.Gln74Ter)

NM_007294.3(BRCA1):c.2713C>T (p.Gln905Ter)

NM_007294.3(BRCA1):c.2800C>T (p.Gln934Ter)

NM_007294.3(BRCA1):c.4612C>T (p.Gln1538Ter)

NM_007294.3(BRCA1):c.3352C>T (p.Gln1118Ter)

NM_007294.3(BRCA1):c.4834C>T (p.Gln1612Ter)

NM_007294.3(BRCA1):c.4523G>A (p.Trp1508Ter)

NM_007294.3(BRCA1):c.5135G>A (p.Trp1712Ter)

NM_007294.3(BRCA1):c.1155G>A (p.Trp385Ter)

NM_007294.3(BRCA1):c.4987-1G>A

NM_000059.3(BRCA2):c.9573G>A (p.Trp3191Ter)

NM_000059.3(BRCA2):c.1945C>T (p.Gln649Ter)

NM_000059.3(BRCA2):c.217C>T (p.Gln73Ter)

NM_000059.3(BRCA2):c.523C>T (p.Gln175Ter)

NM_000059.3(BRCA2):c.2548C>T (p.Gln850Ter)

NM_000059.3(BRCA2):c.2905C>T (p.Gln969Ter)

NM_000059.3(BRCA2):c.4689G>A (p.Trp1563Ter)

NM_000059.3(BRCA2):c.4972C>T (p.Gln1658Ter)

NM_000059.3(BRCA2):c.1184G>A (p.Trp395Ter)

NM_000059.3(BRCA2):c.2137C>T (p.Gln713Ter)

NM_000059.3(BRCA2):c.3217C>T (p.Gln1073Ter)

NM_000059.3(BRCA2):c.3523C>T (p.Gln1175Ter)

NM_000059.3(BRCA2):c.4783C>T (p.Gln1595Ter)
NM_000059.3(BRCA2):c.5800C>T (p.Gln1934Ter)
NM_000059.3(BRCA2):c.6478C>T (p.Gln2160Ter)
NM_000059.3(BRCA2):c.7033C>T (p.Gln2345Ter)
NM_000059.3(BRCA2):c.7495C>T (p.Gln2499Ter)
NM_000059.3(BRCA2):c.7501C>T (p.Gln2501Ter)
NM_000059.3(BRCA2):c.7887G>A (p.Trp2629Ter)
NM_000059.3(BRCA2):c.8910G>A (p.Trp2970Ter)
NM_000059.3(BRCA2):c.9139C>T (p.Gln3047Ter)
NM_000059.3(BRCA2):c.9739C>T (p.Gln3247Ter)
NM_000059.3(BRCA2):c.582G>A (p.Trp194Ter)
NM_000059.3(BRCA2):c.7963C>T (p.Gln2655Ter)
NM_000059.3(BRCA2):c.8695C>T (p.Gln2899Ter)
NM_000059.3(BRCA2):c.8869C>T (p.Gln2957Ter)
NM_000059.3(BRCA2):c.1117C>T (p.Gln373Ter)
NM_000059.3(BRCA2):c.1825C>T (p.Gln609Ter)
NM_000059.3(BRCA2):c.2455C>T (p.Gln819Ter)
NM_000059.3(BRCA2):c.2881C>T (p.Gln961Ter)
NM_000059.3(BRCA2):c.3265C>T (p.Gln1089Ter)
NM_000059.3(BRCA2):c.3283C>T (p.Gln1095Ter)
NM_000059.3(BRCA2):c.3442C>T (p.Gln1148Ter)
NM_000059.3(BRCA2):c.3871C>T (p.Gln1291Ter)
NM_000059.3(BRCA2):c.439C>T (p.Gln147Ter)
NM_000059.3(BRCA2):c.4525C>T (p.Gln1509Ter)
NM_000059.3(BRCA2):c.475+1G>A
NM_000059.3(BRCA2):c.5344C>T (p.Gln1782Ter)
NM_000059.3(BRCA2):c.5404C>T (p.Gln1802Ter)
NM_000059.3(BRCA2):c.5773C>T (p.Gln1925Ter)
NM_000059.3(BRCA2):c.5992C>T (p.Gln1998Ter)
NM_000059.3(BRCA2):c.6469C>T (p.Gln2157Ter)
NM_000059.3(BRCA2):c.7261C>T (p.Gln2421Ter)
NM_000059.3(BRCA2):c.7303C>T (p.Gln2435Ter)
NM_000059.3(BRCA2):c.7471C>T (p.Gln2491Ter)
NM_000059.3(BRCA2):c.7681C>T (p.Gln2561Ter)

NM_000059.3(BRCA2):c.7738C>T (p.Gln2580Ter)
NM_000059.3(BRCA2):c.7886G>A (p.Trp2629Ter)
NM_000059.3(BRCA2):c.8140C>T (p.Gln2714Ter)
NM_000059.3(BRCA2):c.8363G>A (p.Trp2788Ter)
NM_000059.3(BRCA2):c.8572C>T (p.Gln2858Ter)
NM_000059.3(BRCA2):c.8773C>T (p.Gln2925Ter)
NM_000059.3(BRCA2):c.8821C>T (p.Gln2941Ter)
NM_000059.3(BRCA2):c.9109C>T (p.Gln3037Ter)
NM_000059.3(BRCA2):c.9317G>A (p.Trp3106Ter)
NM_000059.3(BRCA2):c.9466C>T (p.Gln3156Ter)
NM_000059.3(BRCA2):c.9572G>A (p.Trp3191Ter)
NM_000059.3(BRCA2):c.8490G>A (p.Trp2830Ter)
NM_000059.3(BRCA2):c.5980C>T (p.Gln1994Ter)
NM_000059.3(BRCA2):c.7721G>A (p.Trp2574Ter)
NM_000059.3(BRCA2):c.196C>T (p.Gln66Ter)
NM_000059.3(BRCA2):c.7618-1G>A
NM_000059.3(BRCA2):c.8489G>A (p.Trp2830Ter)
NM_000059.3(BRCA2):c.7857G>A (p.Trp2619Ter)
NM_000059.3(BRCA2):c.1261C>T (p.Gln421Ter)
NM_000059.3(BRCA2):c.1456C>T (p.Gln486Ter)
NM_000059.3(BRCA2):c.3319C>T (p.Gln1107Ter)
NM_000059.3(BRCA2):c.5791C>T (p.Gln1931Ter)
NM_000059.3(BRCA2):c.6070C>T (p.Gln2024Ter)
NM_000059.3(BRCA2):c.7024C>T (p.Gln2342Ter)
NM_000059.3(BRCA2):c.961C>T (p.Gln321Ter)
NM_000059.3(BRCA2):c.9380G>A (p.Trp3127Ter)
NM_000059.3(BRCA2):c.8364G>A (p.Trp2788Ter)
NM_000059.3(BRCA2):c.7758G>A (p.Trp2586Ter)
NM_000059.3(BRCA2):c.2224C>T (p.Gln742Ter)
NM_000059.3(BRCA2):c.5101C>T (p.Gln1701Ter)
NM_000059.3(BRCA2):c.5959C>T (p.Gln1987Ter)
NM_000059.3(BRCA2):c.7060C>T (p.Gln2354Ter)
NM_000059.3(BRCA2):c.9100C>T (p.Gln3034Ter)
NM_000059.3(BRCA2):c.9148C>T (p.Gln3050Ter)

NM_000059.3(BRCA2):c.9883C>T (p.Gln3295Ter)
NM_000059.3(BRCA2):c.1414C>T (p.Gln472Ter)
NM_000059.3(BRCA2):c.1689G>A (p.Trp563Ter)
NM_000059.3(BRCA2):c.581G>A (p.Trp194Ter)
NM_000059.3(BRCA2):c.6490C>T (p.Gln2164Ter)
NM_000059.3(BRCA2):c.7856G>A (p.Trp2619Ter)
NM_000059.3(BRCA2):c.8970G>A (p.Trp2990Ter)
NM_000059.3(BRCA2):c.92G>A (p.Trp31Ter)
NM_000059.3(BRCA2):c.9376C>T (p.Gln3126Ter)
NM_000059.3(BRCA2):c.93G>A (p.Trp31Ter)
NM_000059.3(BRCA2):c.1189C>T (p.Gln397Ter)
NM_000059.3(BRCA2):c.2818C>T (p.Gln940Ter)
NM_000059.3(BRCA2):c.2979G>A (p.Trp993Ter)
NM_000059.3(BRCA2):c.3166C>T (p.Gln1056Ter)
NM_000059.3(BRCA2):c.4285C>T (p.Gln1429Ter)
NM_000059.3(BRCA2):c.6025C>T (p.Gln2009Ter)
NM_000059.3(BRCA2):c.772C>T (p.Gln258Ter)
NM_000059.3(BRCA2):c.7877G>A (p.Trp2626Ter)
NM_000059.3(BRCA2):c.3109C>T (p.Gln1037Ter)
NM_000059.3(BRCA2):c.4222C>T (p.Gln1408Ter)
NM_000059.3(BRCA2):c.7480C>T (p.Arg2494Ter)
NM_000059.3(BRCA2):c.7878G>A (p.Trp2626Ter)
NM_000059.3(BRCA2):c.9076C>T (p.Gln3026Ter)
NM_000059.3(BRCA2):c.1855C>T (p.Gln619Ter)
NM_000059.3(BRCA2):c.4111C>T (p.Gln1371Ter)
NM_000059.3(BRCA2):c.5656C>T (p.Gln1886Ter)
NM_000059.3(BRCA2):c.7757G>A (p.Trp2586Ter)
NM_000059.3(BRCA2):c.8243G>A (p.Gly2748Asp)
NM_000059.3(BRCA2):c.8878C>T (p.Gln2960Ter)
NM_000059.3(BRCA2):c.8487+1G>A
NM_000059.3(BRCA2):c.8677C>T (p.Gln2893Ter)
NM_000059.3(BRCA2):c.250C>T (p.Gln84Ter)
NM_000059.3(BRCA2):c.6124C>T (p.Gln2042Ter)
NM_000059.3(BRCA2):c.7617+1G>A

NM_000059.3(BRCA2):c.8575C>T (p.Gln2859Ter)
NM_000059.3(BRCA2):c.8174G>A (p.Trp2725Ter)
NM_000059.3(BRCA2):c.3187C>T (p.Gln1063Ter)
NM_000059.3(BRCA2):c.9381G>A (p.Trp3127Ter)
NM_000059.3(BRCA2):c.2095C>T (p.Gln699Ter)
NM_000059.3(BRCA2):c.1642C>T (p.Gln548Ter)
NM_000059.3(BRCA2):c.8608C>T (p.Gln2870Ter)
NM_000059.3(BRCA2):c.3412C>T (p.Gln1138Ter)
NM_000059.3(BRCA2):c.4246C>T (p.Gln1416Ter)
NM_000059.3(BRCA2):c.6475C>T (p.Gln2159Ter)
NM_000059.3(BRCA2):c.7366C>T (p.Gln2456Ter)
NM_000059.3(BRCA2):c.7516C>T (p.Gln2506Ter)
NM_000059.3(BRCA2):c.8969G>A (p.Trp2990Ter)
NM_000059.3(BRCA2):c.6487C>T (p.Gln2163Ter)
NM_000059.3(BRCA2):c.2978G>A (p.Trp993Ter)
NM_000059.3(BRCA2):c.7615C>T (p.Gln2539Ter)
NM_000059.3(BRCA2):c.9106C>T (p.Gln3036Ter)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Breast-Ovarian Cancer by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in the BRCA1or BRCA2 gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Lynch syndrome

[01093] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Lynch syndrome. In some embodiment, the pathogenic mutations/SNPs are present in at least one gene selected from MSH6, MSH2, EPCAM, PMS2, and MLH1, including at least the followings:

NM_000179.2(MSH6):c.1045C>T (p.Gln349Ter)
NM_000251.2(MSH2):c.1384C>T (p.Gln462Ter)
NM_002354.2(EPCAM):c.133C>T (p.Gln45Ter)
NM_002354.2(EPCAM):c.429G>A (p.Trp143Ter)
NM_002354.2(EPCAM):c.523C>T (p.Gln175Ter)
NM_000179.2(MSH6):c.2680C>T (p.Gln894Ter)

NM_000251.2(MSH2):c.350G>A (p.Trp117Ter)

NM_000179.2(MSH6):c.2735G>A (p.Trp912Ter)

NM_000179.2(MSH6):c.3556+1G>A

NM_000251.2(MSH2):c.388C>T (p.Gln130Ter)

NM_000535.6(PMS2):c.1912C>T (p.Gln638Ter)

NM_000535.6(PMS2):c.1891C>T (p.Gln631Ter)

NM_000249.3(MLH1):c.454-1G>A

NM_000251.2(MSH2):c.1030C>T (p.Gln344Ter)

NM_000179.2(MSH6):c.2330G>A (p.Trp777Ter)

NM_000179.2(MSH6):c.2191C>T (p.Gln731Ter)

NM_000179.2(MSH6):c.2764C>T (p.Arg922Ter)

NM_000179.2(MSH6):c.2815C>T (p.Gln939Ter)

NM_000179.2(MSH6):c.3020G>A (p.Trp1007Ter)

NM_000179.2(MSH6):c.3436C>T (p.Gln1146Ter)

NM_000179.2(MSH6):c.3647-1G>A

NM_000179.2(MSH6):c.3772C>T (p.Gln1258Ter)

NM_000179.2(MSH6):c.3838C>T (p.Gln1280Ter)

NM_000179.2(MSH6):c.706C>T (p.Gln236Ter)

NM_000179.2(MSH6):c.730C>T (p.Gln244Ter)

NM_000249.3(MLH1):c.1171C>T (p.Gln391Ter)

NM_000249.3(MLH1):c.1192C>T (p.Gln398Ter)

NM_000249.3(MLH1):c.1225C>T (p.Gln409Ter)

NM_000249.3(MLH1):c.1276C>T (p.Gln426Ter)

NM_000249.3(MLH1):c.1528C>T (p.Gln510Ter)

NM_000249.3(MLH1):c.1609C>T (p.Gln537Ter)

NM_000249.3(MLH1):c.1613G>A (p.Trp538Ter)

NM_000249.3(MLH1):c.1614G>A (p.Trp538Ter)

NM_000249.3(MLH1):c.1624C>T (p.Gln542Ter)

NM_000249.3(MLH1):c.1684C>T (p.Gln562Ter)

NM_000249.3(MLH1):c.1731+1G>A

NM_000249.3(MLH1):c.1731+5G>A

NM_000249.3(MLH1):c.1732-1G>A

NM_000249.3(MLH1):c.1896G>A (p.Glu632=)

NM_000249.3(MLH1):c.1989+1G>A

NM_000249.3(MLH1):c.1990-1G>A

NM_000249.3(MLH1):c.1998G>A (p.Trp666Ter)

NM_000249.3(MLH1):c.208-1G>A

NM_000249.3(MLH1):c.2101C>T (p.Gln701Ter)

NM_000249.3(MLH1):c.2136G>A (p.Trp712Ter)

NM_000249.3(MLH1):c.2224C>T (p.Gln742Ter)

NM_000249.3(MLH1):c.230G>A (p.Cys77Tyr)

NM_000249.3(MLH1):c.256C>T (p.Gln86Ter)

NM_000249.3(MLH1):c.436C>T (p.Gln146Ter)

NM_000249.3(MLH1):c.445C>T (p.Gln149Ter)

NM_000249.3(MLH1):c.545G>A (p.Arg182Lys)

NM_000249.3(MLH1):c.731G>A (p.Gly244Asp)

NM_000249.3(MLH1):c.76C>T (p.Gln26Ter)

NM_000249.3(MLH1):c.842C>T (p.Ala281Val)

NM_000249.3(MLH1):c.882C>T (p.Leu294=)

NM_000249.3(MLH1):c.901C>T (p.Gln301Ter)

NM_000251.2(MSH2):c.1013G>A (p.Gly338Glu)

NM_000251.2(MSH2):c.1034G>A (p.Trp345Ter)

NM_000251.2(MSH2):c.1129C>T (p.Gln377Ter)

NM_000251.2(MSH2):c.1183C>T (p.Gln395Ter)

NM_000251.2(MSH2):c.1189C>T (p.Gln397Ter)

NM_000251.2(MSH2):c.1204C>T (p.Gln402Ter)

NM_000251.2(MSH2):c.1276+1G>A

NM_000251.2(MSH2):c.1528C>T (p.Gln510Ter)

NM_000251.2(MSH2):c.1552C>T (p.Gln518Ter)

NM_000251.2(MSH2):c.1720C>T (p.Gln574Ter)

NM_000251.2(MSH2):c.1777C>T (p.Gln593Ter)

NM_000251.2(MSH2):c.1885C>T (p.Gln629Ter)

NM_000251.2(MSH2):c.2087C>T (p.Pro696Leu)

NM_000251.2(MSH2):c.2251G>A (p.Gly751Arg)

NM_000251.2(MSH2):c.2291G>A (p.Trp764Ter)

NM_000251.2(MSH2):c.2292G>A (p.Trp764Ter)

NM_000251.2(MSH2):c.2446C>T (p.Gln816Ter)

NM_000251.2(MSH2):c.2470C>T (p.Gln824Ter)

NM_000251.2(MSH2):c.2536C>T (p.Gln846Ter)
NM_000251.2(MSH2):c.2581C>T (p.Gln861Ter)
NM_000251.2(MSH2):c.2634G>A (p.Glu878=)
NM_000251.2(MSH2):c.2635C>T (p.Gln879Ter)
NM_000251.2(MSH2):c.28C>T (p.Gln10Ter)
NM_000251.2(MSH2):c.472C>T (p.Gln158Ter)
NM_000251.2(MSH2):c.478C>T (p.Gln160Ter)
NM_000251.2(MSH2):c.484G>A (p.Gly162Arg)
NM_000251.2(MSH2):c.490G>A (p.Gly164Arg)
NM_000251.2(MSH2):c.547C>T (p.Gln183Ter)
NM_000251.2(MSH2):c.577C>T (p.Gln193Ter)
NM_000251.2(MSH2):c.643C>T (p.Gln215Ter)
NM_000251.2(MSH2):c.645+1G>A
NM_000251.2(MSH2):c.652C>T (p.Gln218Ter)
NM_000251.2(MSH2):c.754C>T (p.Gln252Ter)
NM_000251.2(MSH2):c.792+1G>A
NM_000251.2(MSH2):c.942G>A (p.Gln314=)
NM_000535.6(PMS2):c.949C>T (p.Gln317Ter)
NM_000249.3(MLH1):c.306+1G>A
NM_000249.3(MLH1):c.62C>T (p.Ala21Val)
NM_000251.2(MSH2):c.1865C>T (p.Pro622Leu)
NM_000179.2(MSH6):c.426G>A (p.Trp142Ter)
NM_000251.2(MSH2):c.715C>T (p.Gln239Ter)
NM_000249.3(MLH1):c.350C>T (p.Thr117Met)
NM_000251.2(MSH2):c.1915C>T (p.His639Tyr)
NM_000251.2(MSH2):c.289C>T (p.Gln97Ter)
NM_000251.2(MSH2):c.2785C>T (p.Arg929Ter)
NM_000249.3(MLH1):c.131C>T (p.Ser44Phe)
NM_000249.3(MLH1):c.1219C>T (p.Gln407Ter)
NM_000249.3(MLH1):c.306+5G>A
NM_000251.2(MSH2):c.1801C>T (p.Gln601Ter)
NM_000535.6(PMS2):c.1144+1G>A
NM_000251.2(MSH2):c.1984C>T (p.Gln662Ter)
NM_000249.3(MLH1):c.381-1G>A

NM_000535.6(PMS2):c.631C>T (p.Arg211Ter)

NM_000251.2(MSH2):c.790C>T (p.Gln264Ter)

NM_000251.2(MSH2):c.366+1G>A

NM_000249.3(MLH1):c.298C>T (p.Arg100Ter)

NM_000179.2(MSH6):c.3013C>T (p.Arg1005Ter)

NM_000179.2(MSH6):c.694C>T (p.Gln232Ter)

NM_000179.2(MSH6):c.742C>T (p.Arg248Ter)

NM_000249.3(MLH1):c.1039-1G>A

NM_000249.3(MLH1):c.142C>T (p.Gln48Ter)

NM_000249.3(MLH1):c.1790G>A (p.Trp597Ter)

NM_000249.3(MLH1):c.1961C>T (p.Pro654Leu)

NM_000249.3(MLH1):c.2103+1G>A

NM_000249.3(MLH1):c.2135G>A (p.Trp712Ter)

NM_000249.3(MLH1):c.588+5G>A

NM_000249.3(MLH1):c.790+1G>A

NM_000251.2(MSH2):c.1035G>A (p.Trp345Ter)

NM_000251.2(MSH2):c.1255C>T (p.Gln419Ter)

NM_000251.2(MSH2):c.1861C>T (p.Arg621Ter)

NM_000251.2(MSH2):c.226C>T (p.Gln76Ter)

NM_000251.2(MSH2):c.2653C>T (p.Gln885Ter)

NM_000251.2(MSH2):c.508C>T (p.Gln170Ter)

NM_000251.2(MSH2):c.862C>T (p.Gln288Ter)

NM_000251.2(MSH2):c.892C>T (p.Gln298Ter)

NM_000251.2(MSH2):c.970C>T (p.Gln324Ter)

NM_000179.2(MSH6):c.4001G>A (p.Arg1334Gln)

NM_000251.2(MSH2):c.1662-1G>A

NM_000535.6(PMS2):c.1882C>T (p.Arg628Ter)

NM_000535.6(PMS2):c.2174+1G>A

NM_000535.6(PMS2):c.2404C>T (p.Arg802Ter)

NM_000179.2(MSH6):c.3991C>T (p.Arg1331Ter)

NM_000179.2(MSH6):c.2503C>T (p.Gln835Ter)

NM_000179.2(MSH6):c.718C>T (p.Arg240Ter)

NM_000249.3(MLH1):c.1038G>A (p.Gln346=)

NM_000249.3(MLH1):c.245C>T (p.Thr82Ile)

NM_000249.3(MLH1):c.83C>T (p.Pro28Leu)
NM_000249.3(MLH1):c.884G>A (p.Ser295Asn)
NM_000249.3(MLH1):c.982C>T (p.Gln328Ter)
NM_000251.2(MSH2):c.1046C>T (p.Pro349Leu)
NM_000251.2(MSH2):c.1120C>T (p.Gln374Ter)
NM_000251.2(MSH2):c.1285C>T (p.Gln429Ter)
NM_000251.2(MSH2):c.1477C>T (p.Gln493Ter)
NM_000251.2(MSH2):c.2152C>T (p.Gln718Ter)
NM_000535.6(PMS2):c.703C>T (p.Gln235Ter)
NM_000249.3(MLH1):c.2141G>A (p.Trp714Ter)
NM_000251.2(MSH2):c.1009C>T (p.Gln337Ter)
NM_000251.2(MSH2):c.1216C>T (p.Arg406Ter)
NM_000179.2(MSH6):c.3202C>T (p.Arg1068Ter)
NM_000251.2(MSH2):c.1165C>T (p.Arg389Ter)
NM_000249.3(MLH1):c.1943C>T (p.Pro648Leu)
NM_000249.3(MLH1):c.200G>A (p.Gly67Glu)
NM_000249.3(MLH1):c.793C>T (p.Arg265Cys)
NM_000249.3(MLH1):c.2059C>T (p.Arg687Trp)
NM_000249.3(MLH1):c.677G>A (p.Arg226Gln)
NM_000249.3(MLH1):c.2041G>A (p.Ala681Thr)
NM_000249.3(MLH1):c.1942C>T (p.Pro648Ser)
NM_000249.3(MLH1):c.676C>T (p.Arg226Ter)
NM_000251.2(MSH2):c.2038C>T (p.Arg680Ter)
NM_000179.2(MSH6):c.1483C>T (p.Arg495Ter)
NM_000179.2(MSH6):c.2194C>T (p.Arg732Ter)
NM_000179.2(MSH6):c.3103C>T (p.Arg1035Ter)
NM_000179.2(MSH6):c.892C>T (p.Arg298Ter)
NM_000249.3(MLH1):c.1459C>T (p.Arg487Ter)
NM_000249.3(MLH1):c.1731G>A (p.Ser577=)
NM_000249.3(MLH1):c.184C>T (p.Gln62Ter)
NM_000249.3(MLH1):c.1975C>T (p.Arg659Ter)
NM_000249.3(MLH1):c.199G>A (p.Gly67Arg)
NM_000251.2(MSH2):c.1076+1G>A
NM_000251.2(MSH2):c.1147C>T (p.Arg383Ter)

NM_000251.2(MSH2):c.181C>T (p.Gln61Ter)

NM_000251.2(MSH2):c.212-1G>A

NM_000251.2(MSH2):c.2131C>T (p.Arg711Ter)

NM_000535.6(PMS2):c.697C>T (p.Gln233Ter)

NM_000535.6(PMS2):c.1261C>T (p.Arg421Ter)

NM_000251.2(MSH2):c.2047G>A (p.Gly683Arg)

NM_000535.6(PMS2):c.400C>T (p.Arg134Ter)

NM_000535.6(PMS2):c.1927C>T (p.Gln643Ter)

NM_000179.2(MSH6):c.1444C>T (p.Arg482Ter)

NM_000179.2(MSH6):c.2731C>T (p.Arg911Ter)

NM_000535.6(PMS2):c.943C>T (p.Arg315Ter)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Lynch syndrome by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in at least one gene selected from BCKDHA, BCKDHB, DBT, and DLD, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Other Genetic Diseases

[01094] Pathogenic G-to-A or C-to-T mutations/SNPs associated with additional genetic diseases are also reported in the ClinVar database and disclosed in Table A, including but not limited to Marfan syndrome, Hurler syndrome, Glycogen Storage Disease, and Cystic Fibrosis. Accordingly, an aspect of the invention relates to a method for correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with any of these diseases, as discussed below.

Marfan syndrome

[01095] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Marfan syndrome. In some embodiment, the pathogenic mutations/SNPs are present in at least the FBN1 gene, including at least the followings:

NM_000138.4(FBN1):c.1879C>T (p.Arg627Cys)

NM_000138.4(FBN1):c.1051C>T (p.Gln351Ter)

NM_000138.4(FBN1):c.184C>T (p.Arg62Cys)

NM_000138.4(FBN1):c.2855-1G>A

NM_000138.4(FBN1):c.3164G>A (p.Cys1055Tyr)

NM_000138.4(FBN1):c.368G>A (p.Cys123Tyr)

NM_000138.4(FBN1):c.4955G>A (p.Cys1652Tyr)
NM_000138.4(FBN1):c.7180C>T (p.Arg2394Ter)
NM_000138.4(FBN1):c.8267G>A (p.Trp2756Ter)
NM_000138.4(FBN1):c.1496G>A (p.Cys499Tyr)
NM_000138.4(FBN1):c.6886C>T (p.Gln2296Ter)
NM_000138.4(FBN1):c.3373C>T (p.Arg1125Ter)
NM_000138.4(FBN1):c.640G>A (p.Gly214Ser)
NM_000138.4(FBN1):c.5038C>T (p.Gln1680Ter)
NM_000138.4(FBN1):c.434G>A (p.Cys145Tyr)
NM_000138.4(FBN1):c.2563C>T (p.Gln855Ter)
NM_000138.4(FBN1):c.7466G>A (p.Cys2489Tyr)
NM_000138.4(FBN1):c.2089C>T (p.Gln697Ter)
NM_000138.4(FBN1):c.592C>T (p.Gln198Ter)
NM_000138.4(FBN1):c.6695G>A (p.Cys2232Tyr)
NM_000138.4(FBN1):c.6164-1G>A
NM_000138.4(FBN1):c.5627G>A (p.Cys1876Tyr)
NM_000138.4(FBN1):c.4061G>A (p.Trp1354Ter)
NM_000138.4(FBN1):c.1982G>A (p.Cys661Tyr)
NM_000138.4(FBN1):c.6784C>T (p.Gln2262Ter)
NM_000138.4(FBN1):c.409C>T (p.Gln137Ter)
NM_000138.4(FBN1):c.364C>T (p.Arg122Cys)
NM_000138.4(FBN1):c.3217G>A (p.Glu1073Lys)
NM_000138.4(FBN1):c.4460-8G>A
NM_000138.4(FBN1):c.4786C>T (p.Arg1596Ter)
NM_000138.4(FBN1):c.7806G>A (p.Trp2602Ter)
NM_000138.4(FBN1):c.247+1G>A
NM_000138.4(FBN1):c.2495G>A (p.Cys832Tyr)
NM_000138.4(FBN1):c.493C>T (p.Arg165Ter)
NM_000138.4(FBN1):c.5504G>A (p.Cys1835Tyr)
NM_000138.4(FBN1):c.5863C>T (p.Gln1955Ter)
NM_000138.4(FBN1):c.6658C>T (p.Arg2220Ter)
NM_000138.4(FBN1):c.7606G>A (p.Gly2536Arg)
NM_000138.4(FBN1):c.7955G>A (p.Cys2652Tyr)
NM_000138.4(FBN1):c.3037G>A (p.Gly1013Arg)

NM_000138.4(FBN1):c.8080C>T (p.Arg2694Ter)

NM_000138.4(FBN1):c.1633C>T (p.Arg545Cys)

NM_000138.4(FBN1):c.7205-1G>A

NM_000138.4(FBN1):c.4621C>T (p.Arg1541Ter)

NM_000138.4(FBN1):c.1090C>T (p.Arg364Ter)

NM_000138.4(FBN1):c.1585C>T (p.Arg529Ter)

NM_000138.4(FBN1):c.4781G>A (p.Gly1594Asp)

NM_000138.4(FBN1):c.643C>T (p.Arg215Ter)

NM_000138.4(FBN1):c.3668G>A (p.Cys1223Tyr)

NM_000138.4(FBN1):c.8326C>T (p.Arg2776Ter)

NM_000138.4(FBN1):c.6354C>T (p.Ile2118=)

NM_000138.4(FBN1):c.1468+5G>A

NM_000138.4(FBN1):c.1546C>T (p.Arg516Ter)

NM_000138.4(FBN1):c.4615C>T (p.Arg1539Ter)

NM_000138.4(FBN1):c.5368C>T (p.Arg1790Ter)

NM_000138.4(FBN1):c.1285C>T (p.Arg429Ter)

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Marfan syndrome by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in the FBN1 gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Hurler syndrome

[01096] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Hurler syndrome. In some embodiment, the pathogenic mutations/SNPs are present in at least the IDUA gene, including at least the followings:

NM_000203.4(IDUA):c.972+1G>A

NM_000203.4(IDUA):c.1855C>T (p.Arg619Ter)

NM_000203.4(IDUA):c.152G>A (p.Gly51Asp)

NM_000203.4(IDUA):c.1205G>A (p.Trp402Ter)

NM_000203.4(IDUA):c.208C>T (p.Gln70Ter)

NM_000203.4(IDUA):c.1045G>A (p.Asp349Asn)

NM_000203.4(IDUA):c.1650+5G>A

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Hurler syndrome by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in the IDUA gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Glycogen Storage Disease

[01097] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Glycogen Storage Disease. In some embodiment, the pathogenic mutations/SNPs are present in at least one gene selected from GAA, AGL, PHKB, PRKAG2, G6PC, PGAM2, GBE1, PYGM, and PFKM, including at least the followings:

NM_000152.4(GAA):c.1927G>A (p.Gly643Arg)

NM_000152.4(GAA):c.2173C>T (p.Arg725Trp)

NM_000642.2(AGL):c.3980G>A (p.Trp1327Ter)

NM_000642.2(AGL):c.16C>T (p.Gln6Ter)

NM_000642.2(AGL):c.2039G>A (p.Trp680Ter)

NM_000293.2(PHKB):c.1546C>T (p.Gln516Ter)

NM_016203.3(PRKAG2):c.1592G>A (p.Arg531Gln)

NM_000151.3(G6PC):c.248G>A (p.Arg83His)

NM_000151.3(G6PC):c.724C>T (p.Gln242Ter)

NM_000151.3(G6PC):c.883C>T (p.Arg295Cys)

NM_000151.3(G6PC):c.247C>T (p.Arg83Cys)

NM_000151.3(G6PC):c.1039C>T (p.Gln347Ter)

NM_000152.4(GAA):c.1561G>A (p.Glu521Lys)

NM_000642.2(AGL):c.2590C>T (p.Arg864Ter)

NM_000642.2(AGL):c.3682C>T (p.Arg1228Ter)

NM_000642.2(AGL):c.118C>T (p.Gln40Ter)

NM_000642.2(AGL):c.256C>T (p.Gln86Ter)

NM_000642.2(AGL):c.2681+1G>A

NM_000642.2(AGL):c.2158-1G>A

NM_000290.3(PGAM2):c.233G>A (p.Trp78Ter)

NM_000152.4(GAA):c.1548G>A (p.Trp516Ter)

NM_000152.4(GAA):c.2014C>T (p.Arg672Trp)

NM_000152.4(GAA):c.546G>A (p.Thr182=)

NM_000152.4(GAA):c.1802C>T (p.Ser601Leu)

NM_000152.4(GAA):c.1754+1G>A

NM_000152.4(GAA):c.1082C>T (p.Pro361Leu)

NM_000152.4(GAA):c.2560C>T (p.Arg854Ter)

NM_000152.4(GAA):c.655G>A (p.Gly219Arg)

NM_000152.4(GAA):c.1933G>A (p.Asp645Asn)

NM_000152.4(GAA):c.1979G>A (p.Arg660His)

NM_000152.4(GAA):c.1465G>A (p.Asp489Asn)

NM_000152.4(GAA):c.2512C>T (p.Gln838Ter)

NM_000158.3(GBE1):c.1543C>T (p.Arg515Cys)

NM_005609.3(PYGM):c.1726C>T (p.Arg576Ter)

NM_005609.3(PYGM):c.1827G>A (p.Lys609=)

NM_005609.3(PYGM):c.148C>T (p.Arg50Ter)

NM_005609.3(PYGM):c.613G>A (p.Gly205Ser)

NM_005609.3(PYGM):c.1366G>A (p.Val456Met)

NM_005609.3(PYGM):c.1768+1G>A

NM_001166686.1(PFKM):c.450+1G>A

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Glycogen Storage Disease by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in at least one gene selected from GAA, AGL, PHKB, PRKAG2, G6PC, PGAM2, GBE1, PYGM, and PFKM, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

Cystic Fibrosis

[01098] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic G-to-A or C-to-T mutations/SNPs associated with Cystic Fibrosis. In some embodiment, the pathogenic mutations/SNPs are present in the CFTR gene, including at least the followings:

NM_000492.3(CFTR):c.3712C>T (p.Gln1238Ter)

NM_000492.3(CFTR):c.3484C>T (p.Arg1162Ter)

NM_000492.3(CFTR):c.1766+1G>A

NM_000492.3(CFTR):c.1477C>T (p.Gln493Ter)

NM_000492.3(CFTR):c.2538G>A (p.Trp846Ter)

NM_000492.3(CFTR):c.2551C>T (p.Arg851Ter)

NM_000492.3(CFTR):c.3472C>T (p.Arg1158Ter)

NM_000492.3(CFTR):c.1475C>T (p.Ser492Phe)

NM_000492.3(CFTR):c.1679G>A (p.Arg560Lys)

NM_000492.3(CFTR):c.3197G>A (p.Arg1066His)

NM_000492.3(CFTR):c.3873+1G>A

NM_000492.3(CFTR):c.3196C>T (p.Arg1066Cys)

NM_000492.3(CFTR):c.2490+1G>A

NM_000492.3(CFTR):c.3718-1G>A

NM_000492.3(CFTR):c.171G>A (p.Trp57Ter)

NM_000492.3(CFTR):c.3937C>T (p.Gln1313Ter)

NM_000492.3(CFTR):c.274G>A (p.Glu92Lys)

NM_000492.3(CFTR):c.1013C>T (p.Thr338Ile)

NM_000492.3(CFTR):c.3266G>A (p.Trp1089Ter)

NM_000492.3(CFTR):c.1055G>A (p.Arg352Gln)

NM_000492.3(CFTR):c.1654C>T (p.Gln552Ter)

NM_000492.3(CFTR):c.2668C>T (p.Gln890Ter)

NM_000492.3(CFTR):c.3611G>A (p.Trp1204Ter)

NM_000492.3(CFTR):c.1585-8G>A

NM_000492.3(CFTR):c.223C>T (p.Arg75Ter)

NM_000492.3(CFTR):c.1680-1G>A

NM_000492.3(CFTR):c.349C>T (p.Arg117Cys)

NM_000492.3(CFTR):c.1203G>A (p.Trp401Ter)

NM_000492.3(CFTR):c.1240C>T (p.Gln414Ter)

NM_000492.3(CFTR):c.1202G>A (p.Trp401Ter)

NM_000492.3(CFTR):c.1209+1G>A

NM_000492.3(CFTR):c.115C>T (p.Gln39Ter)

NM_000492.3(CFTR):c.1116+1G>A

NM_000492.3(CFTR):c.1393-1G>A

NM_000492.3(CFTR):c.1573C>T (p.Gln525Ter)

NM_000492.3(CFTR):c.164+1G>A

NM_000492.3(CFTR):c.166G>A (p.Glu56Lys)

NM_000492.3(CFTR):c.170G>A (p.Trp57Ter)

NM_000492.3(CFTR):c.2053C>T (p.Gln685Ter)

NM_000492.3(CFTR):c.2125C>T (p.Arg709Ter)

NM_000492.3(CFTR):c.2290C>T (p.Arg764Ter)
NM_000492.3(CFTR):c.2353C>T (p.Arg785Ter)
NM_000492.3(CFTR):c.2374C>T (p.Arg792Ter)
NM_000492.3(CFTR):c.2537G>A (p.Trp846Ter)
NM_000492.3(CFTR):c.292C>T (p.Gln98Ter)
NM_000492.3(CFTR):c.2989-1G>A
NM_000492.3(CFTR):c.3293G>A (p.Trp1098Ter)
NM_000492.3(CFTR):c.4144C>T (p.Gln1382Ter)
NM_000492.3(CFTR):c.4231C>T (p.Gln1411Ter)
NM_000492.3(CFTR):c.4234C>T (p.Gln1412Ter)
NM_000492.3(CFTR):c.579+5G>A
NM_000492.3(CFTR):c.595C>T (p.His199Tyr)
NM_000492.3(CFTR):c.613C>T (p.Pro205Ser)
NM_000492.3(CFTR):c.658C>T (p.Gln220Ter)
NM_000492.3(CFTR):c.1117-1G>A
NM_000492.3(CFTR):c.3294G>A (p.Trp1098Ter)
NM_000492.3(CFTR):c.1865G>A (p.Gly622Asp)
NM_000492.3(CFTR):c.743+1G>A
NM_000492.3(CFTR):c.1679+1G>A
NM_000492.3(CFTR):c.1657C>T (p.Arg553Ter)
NM_000492.3(CFTR):c.1675G>A (p.Ala559Thr)
NM_000492.3(CFTR):c.165-1G>A
NM_000492.3(CFTR):c.200C>T (p.Pro67Leu)
NM_000492.3(CFTR):c.2834C>T (p.Ser945Leu)
NM_000492.3(CFTR):c.3846G>A (p.Trp1282Ter)
NM_000492.3(CFTR):c.1652G>A (p.Gly551Asp)
NM_000492.3(CFTR):c.4426C>T (p.Gln1476Ter)
NM_000492.3:c.3718-2477C>T
NM_000492.3(CFTR):c.2988+1G>A
NM_000492.3(CFTR):c.2657+5G>A
NM_000492.3(CFTR):c.2988G>A (p.Gln996=)
NM_000492.3(CFTR):c.274-1G>A
NM_000492.3(CFTR):c.3612G>A (p.Trp1204Ter)
NM_000492.3(CFTR):c.1646G>A (p.Ser549Asn)

NM_000492.3(CFTR):c.3752G>A (p.Ser1251Asn)

NM_000492.3(CFTR):c.4046G>A (p.Gly1349Asp)

NM_000492.3(CFTR):c.532G>A (p.Gly178Arg)

NM_000492.3(CFTR):c.3731G>A (p.Gly1244Glu)

NM_000492.3(CFTR):c.1651G>A (p.Gly551Ser)

NM_000492.3(CFTR):c.1585-1G>A

NM_000492.3(CFTR):c.1000C>T (p.Arg334Trp)

NM_000492.3(CFTR):c.254G>A (p.Gly85Glu)

NM_000492.3(CFTR):c.1040G>A (p.Arg347His)

NM_000492.3(CFTR):c.273+1G>A

See Table A. Accordingly, an aspect of the invention relates to a method for treating or preventing Cystic Fibrosis by correcting one or more pathogenic G-to-A or C-to-T mutations/SNPs, particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs present in the CFTR gene, and more particularly one or more pathogenic G-to-A or C-to-T mutations/SNPs described above.

[01099] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with familial 2 breast-ovarian cancer, wherein the pathogenic A>G mutation or SNP is located in the BRCA2 gene (HGVS: U43746.1:n.7829+1G>A). Accordingly, an additional aspect of the invention relates to a method for treating or preventing familial 2 breast-ovarian cancer by correcting the aforementioned pathogenic A>G mutation or SNP.

[01100] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with hereditary factor IX deficiency, wherein the pathogenic A>G mutation or SNP is located at GRCh38: ChrX: 139537145 in the F9 gene, which results in an Arg to Gln substitution. Accordingly, an additional aspect of the invention relates to a method for treating or preventing hereditary factor IX deficiency by correcting the aforementioned pathogenic A>G mutation or SNP.

[01101] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with beta-plus-thalassemia, beta thalassemia, and beta thalassemia major, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr11: 5226820 in the HBB gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing with beta-

plus-thalassemia, beta thalassemia, and beta thalassemia major by correcting the aforementioned pathogenic A>G mutation or SNP.

[01102] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with Marfan syndrome, wherein the pathogenic A>G mutation or SNP is located in the FBN1 gene (IVS2DS, G-A, +1), as reported by Yamamoto et al. J Hum Genet. 2000;45(2):115-8. Accordingly, an additional aspect of the invention relates to a method for treating or preventing Marfan syndrome by correcting the aforementioned pathogenic A>G mutation or SNP.

[01103] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with Wiskott-Aldrich syndrome, wherein the pathogenic A>G mutation or SNP is located at position -1 of intro 6 of the WAS gene (IVS6AS, G-A, -1), as reported by Kwan et al. (1995). Accordingly, an additional aspect of the invention relates to a method for treating or preventing Wiskott-Aldrich syndrome by correcting the aforementioned pathogenic A>G mutation or SNP.

[01104] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with cystic fibrosis, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr7:117590440 in the CFTR gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing cystic fibrosis by correcting the aforementioned pathogenic A>G mutation or SNP.

[01105] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with cystic fibrosis and hereditary pancreatitis, wherein the pathogenic A>G mutation or SNP is located GRCh38: Chr7:117606754 in the CFTR gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing cystic fibrosis and hereditary pancreatitis by correcting the aforementioned pathogenic A>G mutation or SNP.

[01106] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with cystic fibrosis, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr7: 117587738 in the CFTR gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing cystic fibrosis by correcting the aforementioned pathogenic A>G mutation or SNP.

[01107] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with Turcot syndrome and Lynch syndrome, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr2:47470964 in the MSH2 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing Turcot syndrome and Lynch syndrome by correcting the aforementioned pathogenic A>G mutation or SNP.

[01108] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with cystic fibrosis, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr7: 117642437 in the CFTR gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing cystic fibrosis by correcting the aforementioned pathogenic A>G mutation or SNP.

[01109] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with Lynch syndrome II and Lynch syndrome, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr3:37001058 in the MLH1 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing Lynch syndrome II and Lynch syndrome by correcting the aforementioned pathogenic A>G mutation or SNP.

[01110] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with cystic fibrosis, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr7: 117642594 in the CFTR gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing cystic fibrosis by correcting the aforementioned pathogenic A>G mutation or SNP.

[01111] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with cystic fibrosis, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr7: 117592658 in the CFTR gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing cystic fibrosis by correcting the aforementioned pathogenic A>G mutation or SNP.

[01112] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with familial 1 breast-ovarian cancer, hereditary breast and ovarian cancer syndrome, and hereditary cancer-predisposing syndrome, wherein the pathogenic A>G mutation or SNP is located at

GRCh38: Chr17:43057051 in the BRCA1 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing familial 1 breast-ovarian cancer, hereditary breast and ovarian cancer syndrome, and hereditary cancer-predisposing syndrome by correcting the aforementioned pathogenic A>G mutation or SNP.

[01113] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with dihydropyrimidine dehydrogenase deficiency, Hirschsprung disease 1, fluorouracil response, pyrimidine analogues response - toxicity/ADR, capecitabine response - toxicity/ADR, fluorouracil response - toxicity/ADR, tegafur response - toxicity/ADR, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr1:97450058 in the DPYD gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing dihydropyrimidine dehydrogenase deficiency, Hirschsprung disease 1, fluorouracil response, pyrimidine analogues response - toxicity/ADR, capecitabine response - toxicity/ADR, fluorouracil response - toxicity/ADR, tegafur response - toxicity/ADR by correcting the aforementioned pathogenic A>G mutation or SNP.

[01114] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with Lynch syndrome, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr2:47478520 in the MSH2 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing the Lynch syndrome by correcting the aforementioned pathogenic A>G mutation or SNP.

[01115] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with Lynch syndrome, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr3:37011819 in the MLH1 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing the Lynch syndrome by correcting the aforementioned pathogenic A>G mutation or SNP.

[01116] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with Lynch syndrome, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr3: 37014545 in the MLH1 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing the Lynch syndrome by correcting the aforementioned pathogenic A>G mutation or SNP.

[01117] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with Lynch syndrome, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr3: 37011867 in the MLH1 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing the Lynch syndrome by correcting the aforementioned pathogenic A>G mutation or SNP.

[01118] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with Lynch syndrome, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr3: 37025636 in the MLH1 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing the Lynch syndrome by correcting the aforementioned pathogenic A>G mutation or SNP.

[01119] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with Lynch syndrome, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr3: 37004475 in the MLH1 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing the Lynch syndrome by correcting the aforementioned pathogenic A>G mutation or SNP.

[01120] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with Lynch syndrome and hereditary cancer-predisposing syndrome, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr2:47416430 in the MSH2 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing Lynch syndrome and hereditary cancer-predisposing syndrome by correcting the aforementioned pathogenic A>G mutation or SNP.

[01121] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with Lynch syndrome and hereditary cancer-predisposing syndrome, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr2: 47408400 in the MSH2 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing Lynch syndrome and hereditary cancer-predisposing syndrome by correcting the aforementioned pathogenic A>G mutation or SNP.

[01122] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with

Lynch syndrome and hereditary cancer-predisposing syndrome, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr3:36996710 in the MLH1 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing Lynch syndrome and hereditary cancer-predisposing syndrome by correcting the aforementioned pathogenic A>G mutation or SNP.

[01123] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with familial 1 breast-ovarian cancer, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr17:43067696 in the BRCA1 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing familial 1 breast-ovarian cancer by correcting the aforementioned pathogenic A>G mutation or SNP.

[01124] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with familial 2 breast-ovarian cancer and hereditary breast and ovarian cancer syndrome, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr13:32356610 in the BRCA2 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing familial 2 breast-ovarian cancer and hereditary breast and ovarian cancer syndrome by correcting the aforementioned pathogenic A>G mutation or SNP.

[01125] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with primary dilated cardiomyopathy and primary familial hypertrophic cardiomyopathy, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr14:23419993 in the MYH7 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing primary dilated cardiomyopathy and primary familial hypertrophic cardiomyopathy by correcting the aforementioned pathogenic A>G mutation or SNP.

[01126] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with primary familial hypertrophic cardiomyopathy, camptocormism, and hypertrophic cardiomyopathy, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr14:23415225 in the MYH7 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing primary familial hypertrophic cardiomyopathy, camptocormism, and hypertrophic cardiomyopathy by correcting the aforementioned pathogenic A>G mutation or SNP.

[01127] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with familial cancer of breast, familial 2 breast-ovarian cancer, hereditary breast and ovarian cancer syndrome, and hereditary cancer-predisposing syndrome, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr13:32357741 in the BRCA2 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing the familial cancer of breast, familial 2 breast-ovarian cancer, hereditary breast and ovarian cancer syndrome, and hereditary cancer-predisposing syndrome by correcting the aforementioned pathogenic A>G mutation or SNP.

[01128] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with primary dilated cardiomyopathy, hypertrophic cardiomyopathy, cardiomyopathy, and left ventricular noncompaction, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr14:23431584 in the MYH7 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing primary dilated cardiomyopathy, hypertrophic cardiomyopathy, cardiomyopathy, and left ventricular noncompaction by correcting the aforementioned pathogenic A>G mutation or SNP.

[01129] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with familial 1 breast-ovarian cancer, hereditary breast and ovarian cancer syndrome, and hereditary cancer-predisposing syndrome, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr17:43067607 in the BRCA1 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing familial 1 breast-ovarian cancer, hereditary breast and ovarian cancer syndrome, and hereditary cancer-predisposing syndrome by correcting the aforementioned pathogenic A>G mutation or SNP.

[01130] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with familial 1 breast-ovarian cancer, hereditary breast and ovarian cancer syndrome, hereditary cancer-predisposing syndrome, and breast cancer, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr17:43047666 in the BRCA1 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing familial 1 breast-ovarian cancer, hereditary breast and ovarian cancer syndrome, hereditary cancer-predisposing syndrome, and breast cancer by correcting the aforementioned pathogenic A>G mutation or SNP.

[01131] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with familial 2 breast-ovarian cancer, hereditary breast and ovarian cancer syndrome, and hereditary cancer-predisposing syndrome, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr13:32370558 in the BRCA2 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing familial 2 breast-ovarian cancer, hereditary breast and ovarian cancer syndrome, and hereditary cancer-predisposing syndrome by correcting the aforementioned pathogenic A>G mutation or SNP.

[01132] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with familial 1 breast-ovarian cancer, hereditary breast and ovarian cancer syndrome, hereditary cancer-predisposing syndrome, and breast cancer, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr17:43074330 in the BRCA1 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing familial 1 breast-ovarian cancer, hereditary breast and ovarian cancer syndrome, hereditary cancer-predisposing syndrome, and breast cancer by correcting the aforementioned pathogenic A>G mutation or SNP.

[01133] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic A-to-G (A>G) mutation or SNP believed to be associated with familial 1 breast-ovarian cancer, hereditary breast and ovarian cancer syndrome, and hereditary cancer-predisposing syndrome, wherein the pathogenic A>G mutation or SNP is located at GRCh38: Chr17: 43082403 in the BRCA1 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing familial 1 breast-ovarian cancer, hereditary breast and ovarian cancer syndrome, and hereditary cancer-predisposing syndrome by correcting the aforementioned pathogenic A>G mutation or SNP.

[01134] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic C-to-T (C>T) mutation or SNP believed to be associated with cystic fibrosis and hereditary pancreatitis, wherein the pathogenic C>T mutation or SNP is located at GRCh38: Chr7:117639961 in the CFTR gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing the cystic fibrosis and hereditary pancreatitis by correcting the aforementioned pathogenic C>T mutation or SNP.

[01135] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic C-to-T (C>T) mutation or SNP believed to be associated with familial 2 breast-ovarian cancer, wherein the pathogenic C>T mutation or SNP is located at

GRCh38: Chr13:32336492 in the BRCA2 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing the familial 2 breast-ovarian cancer by correcting the aforementioned pathogenic C>T mutation or SNP.

[01136] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic C-to-T (C>T) mutation or SNP believed to be associated with familial 1 breast-ovarian cancer, wherein the pathogenic C>T mutation or SNP is located at GRCh38: Chr17:43063365 in the BRCA1 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing the familial 1 breast-ovarian cancer by correcting the aforementioned pathogenic C>T mutation or SNP.

[01137] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic C-to-T (C>T) mutation or SNP believed to be associated with familial 1 breast-ovarian cancer, wherein the pathogenic C>T mutation or SNP is located at GRCh38: Chr17:43093613 in the BRCA1 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing the familial 1 breast-ovarian cancer by correcting the aforementioned pathogenic C>T mutation or SNP.

[01138] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic C-to-T (C>T) mutation or SNP believed to be associated with familial cancer of breast, and familial 1 breast-ovarian cancer, wherein the pathogenic C>T mutation or SNP is located at at GRCh38: Chr17:43093931 of the BRCA1 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing familial cancer of breast, and familial 1 breast-ovarian cancer by correcting the aforementioned pathogenic C>T mutation or SNP.

[01139] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic C-to-T (C>T) mutation or SNP believed to be associated with familial hypertrophic cardiomyopathy 1, primary familial hypertrophic cardiomyopathy, and hypertrophic cardiomyopathy, wherein the pathogenic C>T mutation or SNP is located at GRCh38: Chr14:23429279 of the MYH7 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing familial hypertrophic cardiomyopathy 1, primary familial hypertrophic cardiomyopathy, and hypertrophic cardiomyopathy by correcting the aforementioned pathogenic C>T mutation or SNP.

[01140] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic C-to-T (C>T) mutation or SNP believed to be associated with familial 2 breast-ovarian cancer, hereditary breast and ovarian cancer syndrome, and hereditary cancer-predisposing syndrome, wherein the pathogenic C>T mutation or SNP is located at

GRCh38: Chr13:32356472 of the BRCA2 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing familial 2 breast-ovarian cancer, hereditary breast and ovarian cancer syndrome, and hereditary cancer-predisposing syndrome by correcting the aforementioned pathogenic C>T mutation or SNP.

[01141] In some embodiments, the methods, systems, and compositions described herein are used to correct a pathogenic C-to-T (C>T) mutation or SNP believed to be associated with familial hypertrophic cardiomyopathy 1, primary familial hypertrophic cardiomyopathy, familial restrictive cardiomyopathy, and hypertrophic cardiomyopathy, wherein the pathogenic C>T mutation or SNP is located at GRCh38: Chr14:23429005 in the MYH7 gene. Accordingly, an additional aspect of the invention relates to a method for treating or preventing familial hypertrophic cardiomyopathy 1, primary familial hypertrophic cardiomyopathy, familial restrictive cardiomyopathy, and hypertrophic cardiomyopathy by correcting the aforementioned pathogenic C>T mutation or SNP.

[01142] Additional pathogenic A>G mutations and SNPs are found in the ClinVar database Accordingly, an additional aspect of the present disclosure relates to correction of a pathogenic A>G mutation or SNP listed in ClinVar using the methods, systems, and compositions described herein to treat or prevent a disease or condition associated therewith.

[01143] Additional pathogenic C>T mutations and SNPs are also found in the ClinVar database. Accordingly, an additional aspect of the present disclosure relates to correction of a pathogenic C>T mutation or SNP listed in ClinVar using the methods, systems, and compositions described herein to treat or prevent a disease or condition associated therewith. Other T mutations or SNPS that may be addressed using the embodiments disclosed herein are listed in a table found in the ASCII text filed entitled “Clin_var_pathogenic_SNPS_TC_txt” filed herewith.

Modification of phosphorylation sites and other post-translational modifications

[01144] The present invention also contemplates use of the AD-functionalized CRISPR system described herein to modify phosphorylation sites and other post-translational modifications (PTMs). The AD-functionalized CRISPR system described herein can edit residues associated with post-translational modifications (FIG. 140A and 140B). Protein phosphorylations are involved in multiple cellular processes and are relatively easy to target (Humprey et al. Trends Endocrinol Metab 2015, 26(12):676-687). Current technologies to target phosphorylations sites or other PTMs include whole protein knockdown or knockout, base editing, and small molecule. These methods, however, all have certain drawbacks. Protein target knockdown or knockout will remove whole protein instead of just the PTMs, base editing

is permanent, whereas small molecules are also hard to develop and may have unknown targets. Using the AD-functionalized CRISPR system described herein to remove phosphorylations site may allow study of the function of phosphorylations, for example, it can be used for screening kinase targets to determine relative contributions to phenotype, or for transcriptome-wide screening for potential small molecules. Targeting PTMs using AD-functionalized CRISPR system can also have therapeutic potential in cancer, inflammation, metabolism, and differentiation.

[01145] In certain embodiments, the AD-functionalized CRISPR system described herein can be used to target Stat3 and/or IRF-5 phosphorylation to reduce inflammation. The target sites can be selected from the group consisting of Stat3 Tyr705, IRF-5 Thr10, Ser158, Ser309, Ser317, Ser451 and Ser462, all of which are involved in interleukin signaling and/or autoimmunity (Sadreev et al. PLOS One 2014, 9(10): e110913). Accordingly, an additional aspect of the invention relates to a method for treating or preventing autoimmune disease by targeting the aforementioned phosphorylation sites.

[01146] In certain embodiments, the AD-functionalized CRISPR system described herein can be used to target Insulin receptor substrate (IRS) phosphorylation. The target sites can be selected from the group consisting of Ser-265, Ser-302, Ser-325, Ser-336, Ser-358, Ser-407, and Ser-408 of IRS-1. The phosphorylation of these sites reduces insulin sensitivity (Copps and White Diabetologia 2012, Oct;55(10):2565-2582), and reducing inhibitory serine phosphorylation at these sites can rescue insulin sensitivity. Accordingly, an additional aspect of the invention relates to a method for treating or preventing diabetes by targeting the aforementioned phosphorylation sites.

Making hypomorphic mutations

[01147] In certain embodiments, the AD-functionalized CRISPR system described herein can be used to make hypomorphic mutations. Engineering hypomorphic mutations can lead to significant downregulation of essential genes without lethality, which allows for straightforward creation of models for diseases that involve hypomorphic mutations and decreasing levels of certain proteins in a fine-tuned manner for therapeutic applications. PolyA track insertion is an existing technology to create hypomorphic mutants. Using the AD-functionalized CRISPR system for introducing hypomorphic mutations is minimally disruptive, precise, and can be fine-tuned.

[01148] In certain embodiments, the AD-functionalized CRISPR system can be used for targeted editing of immune checkpoint proteins. Immune checkpoint blockade is used in cancer therapy to enhance anti-tumor immunity by promoting T-cell activation and proliferation,

which includes anti-CTLA4 and anti-PD-1 therapies (Byun et al., Nat Reviews Endocrinology 2017). Using the AD-functionalized CRISPR system can improve efficacy over existing CTLA-4, PD-1/PD-L1 inhibitor therapies. The AD-functionalized CRISPR system can also be employed to inhibit other suppressive immune checkpoints (such as TIM-3, KIRs, and LAG-3) and to introduce hypomorphic mutations to immune activating checkpoints such as 4-IBB and GITR. In particular embodiments, the AD-functionalized CRISPR system can be used for targeted editing of CTLA-4/B7-1 interaction surface ^{99}MYPPPY^{104} stem loop (Stamper et. al., Nature 2001, Mar 29;410(6828):608-11), for example, the C-to-U editing can convert proline to serine or leucine, whereas the A-to-I editing can convert tyrosine to cysteine and methionine to valine. In particular embodiments, the AD-functionalized CRISPR system can be used for targeted editing of CTLA-4/B7-2 interface at E33, R35, T53, and E97 (Schwartz et. al., Nature 2001, Mar 29;410(6828):604-8; Peach et. al., Cell (1994)), for example, the C-to-U editing can convert arginine to cysteine, stop codon, or tryptophan, whereas the A-to-I editing can convert glutamic acid to glycine, and arginine to glycine. Accordingly, an additional aspect of the invention relates to a method for treating or preventing cancer by editing the aforementioned residues involved in immune checkpoint protein interactions.

Modulating protein stability

[01149] In certain embodiments, the AD-functionalized CRISPR system described herein can be used to modulate protein stability. In particular embodiments, the AD-functionalized CRISPR system can be used for general degron targeting. A degron is a portion of a protein that is important in regulation of protein degradation rates. Known degrons include short amino acid sequences, structural motifs and exposed amino acids (often Lysine or Arginine) located anywhere in the protein. Some proteins can contain multiple degrons. While there are many types of different degrons, and a high degree of variability even within these groups, degrons are all similar for their involvement in regulating the rate of a protein degradation and can be categorized as "Ubiquitin-dependent" or "Ubiquitin-independent".

[01150] In certain example embodiments, the AD- functionalized CRISPR system can be used for targeted editing of the degron present in SMN2, a protein involved in spinal muscular atrophy (SMA). SMA is caused by homozygous survival of motor neurons 1(SMN1) gene deletions, leaving a duplicate gene, SMN2, as the sole source of SMN protein. SMA disease severity correlates to the amount of functional protein. For example, severe SMA (type I) patients typically have one or two SMN2 copies, intermediate severity SMA (type II) patients usually have three SMN2 copies, and patients with mild SMA (type III) mostly have three or four SMN2 copies. Most of the mRNA produced from SMN2 pre-mRNA is exon 7-skipped

(about 80%), resulting in a highly unstable and almost undetectable protein (SMNDelta7). This splicing defect creates a degradation signal (degron; SMNDelta7-DEG) at SMNDelta7's C-terminal 15 amino acids. The S270A mutation inactivates SMNDelta7-DEG, generating a stable SMNDelta7 that rescues viability of SMN-deleted cells. (Cho and Dreyfuss, Genes and Dev., 2010, Mar 1;24(5):438-42). The AD- functionalized CRISPR system can be used for targeted editing of S270, thereby disrupts the degron present in SMN2. Accordingly, an additional aspect of the invention relates to a method for treating or preventing SMA by editing the aforementioned residues involved in regulating SMN stability.

[01151] In certain embodiments, the AD- functionalized CRISPR system can be used for disrupting the D-box degrons, resulting in the conversion of Arg to Gly, or Leu to The. In other embodiments, the AD- functionalized CRISPR system can be used for disrupting the KEN-box degrons, resulting in the conversion of Lys to Arg/Glu, Glu to Gly, or Asn to Ser/Asp.

[01152] The N-degrons were first characterized in yeast to the PEST sequence of mouse ornithine decarboxylase. A PEST sequence is a peptide sequence that is rich in proline (P), glutamic acid (E), serine (S), and threonine (T). This sequence is associated with proteins that have a short intracellular half-life; hence, it is hypothesized that the PEST sequence acts as a signal peptide for protein degradation. The AD-functionalized CRISPR system can be used for targeted editing of the PEST sequence, hence regulating protein stability. In particular example embodiments, the AD-functionalized CRISPR system can be used for targeting the PEST sequence or a regulated, ubiquitin-independent degron in IκBα (Fortmann et al, JMB Molecular Bio 2015, Aug 28; 427(17): 2748-2756). In particular embodiments, the AD-functionalized CRISPR system can be used for editing a PEST sequence in NANOG to promote embryonic stem cell (ESC) pluripotency. In particular embodiments, the AD-functionalized CRISPR system can be used for editing a PEST sequence in Cdc25A phosphatase. In other embodiments, the AD-functionalized CRISPR system can also be employed to facilitate protein degradation, for example, by mutating the residues to enhance the degree of degradation or by mutating the N-terminal methionine.

Targeting ion channels for therapy

[01153] In certain embodiments, the AD-functionalized CRISPR system described herein can be used to target ion channels. Ions regulate many physiological processes, including heart contractility, nervous system signal transduction, and control of pulmonary vasculature pressure. Small molecules that affect ion channels, such as Digoxin and Lidocaine are widely used in clinical medicine. These small molecules, however, have toxicity issues and only act

on shorter time scales whereas the diseases being treated such as heart failure or arrhythmias, are often chronic. Knockdown approach is also not desirable as it may affect other biological roles played by the ion channels.

[01154] In certain embodiments, the AD-functionalized CRISPR system can be used to make stop codons to block ion channels. In certain embodiments, the AD-functionalized CRISPR system can be used to make stop codons to skip exons. The ion channels can be sodium or potassium ion channels. In particular embodiments, the AD-functionalized CRISPR system can be used to make mutations selected from the group consisting of V36I, F216S, S241T, R277X, Y328X, N395K, S459X, E693X, I767X, R830X, I848T, L858H, L858H, L858F, A863P, W897X, R996C, F1200LfsX33, I1235LfsX2, V1298F, V1298D, V1299F, F1449V, c.4336-7_10delGTTTX, I1461T, F1462V, T1464I, R1488X, M1267K, K1659X, W1689X in the sodium-channel subunit Nav1.7 (Drenth and Waxman, JCI, 2007, Dec;117(12):3603-9). In certain embodiments, the AD-functionalized CRISPR system can be used to edit RNA in neurons. The resulting ion channel activity change can be assessed via patch-clamping and pain sensitivity can be examined using existing mouse models (Gao et al., J Neurosci. 2009 Apr 1;29(13):4096-108). Accordingly, an additional aspect of the invention relates to a method for treating or preventing heart failure or arrhythmia by editing the aforementioned residues involved in ion channel activities.

TGFbeta modulation to prevent cardiac remodeling

[01155] In certain embodiments, the AD-functionalized CRISPR system can be used to modulate TGFbeta signaling to prevent cardiac remodeling. After myocardial infarction, TGFbeta signaling promotes cardiac fibrosis and cardiomyocyte apoptosis and blocks the inflammatory response that can heal the cardiac tissue. Therefore negative heart remodeling can be prevented by blocking TGFbeta signaling. The type II TGFbeta receptor requires autophosphorylation at Ser213 and Ser409 as well as Thr259, 336, and 424 for activity. The AD-functionalized CRISPR system can be used to mutate the serines to Leu or Phe, or tyrosines to Cys, which can prevent autophosphorylation and TGFbeta activation in fibroblasts and cardiomyocytes.

[01156] In certain embodiments, the AD-functionalized CRISPR system can be used to mutate the Smad transcription factors downstream of the TGFbeta receptor to prevent their activation via phosphorylation. The AD-functionalized CRISPR system can mutate the phosphorylation sites selected from the group consisting of Thr8, Thr179, Ser208, and Ser213 of Smad3 and Ser245, Ser250, Ser255, and Thr8 of Smad2. The AD-functionalized CRISPR system can be used to mutate the serines to Leu or Phe, or threonines to Ile or Met. Accordingly,

an additional aspect of the invention relates to a method for preventing cardiac remodeling by editing the aforementioned residues involved in TGFbeta signaling.

Other applications

[01157] In certain embodiments, the AD-functionalized CRISPR system can be used in lineage tracing. In certain embodiments, the AD-functionalized CRISPR system can be used for sensing with REPAIR system. Different orthologs can be induced and editing can be focused on synthetic transcripts. In certain embodiments, the AD-functionalized CRISPR system can be used for saturation mutagenesis on specific proteins to identify functional domains. In certain embodiments, the AD-functionalized CRISPR system can be used to identify RNA binding protein interactions. The AD- functionalized CRISPR system can be used to map protein-protein binding interfaces. Saturation mutagenesis on be performed on one protein followed by FRET and cell sorting to determine which guide RNA disrupts protein-protein interactions.

[01158] In certain embodiments, the AD-functionalized CRISPR system can be used for transient inactivation or activation of proteins, generating heterozygous protective mutations, pre or pro-protein cleavage sites, generation of neoantigens, creating conditional fusion proteins, editing of poly-A signals, RNA targeting to introduce other epitranscriptomic modifications, for identification or modification of RNA binding protein sites, mapping RNA-RNA contacts, or editing co-localized RNPs.

[01159] In some embodiments, the AD-functionalized CRISPR system can be used for modification ubiquitination or acetylation sites, tissue regeneration, cell differentiation, creating motifs recognized by ubiquitin ligases, single cell barcoding, creating splice sites, or altering antigen receptors.

WORKING EXAMPLES

EXAMPLE 1

[01160] Adenine deaminases (ADs) is capable of deaminating adenines at specific sites in double stranded RNA.

[01161] The facts that some ADs can effect adenine deamination on DNA-RNAn RNA duplexes (e.g. Zheng et al., Nucleic Acids Research 2017) presents a unique opportunity to develop an RNA guided AD by taking advantage of the RNA duplex formed between the guide RNA and its complementary DNA target in the R-loop formed during RNA-guided

DNA binding by inactive Cas13. By using inactive Cas13 to recruit an AD, the AD enzyme will then act on the adenine in the RNA-DNAn RNA duplex.

[01162] In one embodiment, an inactive Cas13, such as Cas13b is obtained using the following mutations: R116A, H121A, R1177A and H1182A. To increase the efficiency of editing by AD, a mutated ADAR is used such as the mutated hADAR2d comprising mutation E488Q.

Designs for the recruitment of AD to a specific locus:

[01163] 1. NLS-tagged inactive Cas13 is fused to AD on either the N- or C-terminal end. A variety of linkers are used including flexible linkers such as GSG5 or less flexible linkers such as LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR (SEQ ID No. 11).

[01164] 2. The guide RNA scaffold is modified with aptamers such as MS2 binding sites (e.g. Konermann et al., Nature 2015). NLS-tagged AD-MS2 binding protein fusions is co-introduced into target cells along with (NLS-tagged inactive or Cas13b) and corresponding guide RNA.

[01165] 3. AD is inserted into an internal loop of NLS-tagged inactive or nickase Cas13.

Designs for the RNA guide:

[01166] 1. Guide sequences of a length corresponding to that of a natural guide sequence of the Cas13 protein are designed to target the RNA of interest.

[01167] 2. RNA guide with longer than canonical length is used to form RNA duplexes outside of the protein-guide RNA-target DNA complex.

[01168] For each of these RNA guide designs, the base on the RNA that is opposite of the adenine on the target RNA strand would be specified as a C as opposed to U.

Choice and Designs of ADs:

[01169] A number of ADs are used, and each will have varying levels of activity. These ADs

[01170] 1. Human ADARs (hADAR1, hADAR2, hADAR3)

[01171] 2. Squid Octopus vulgaris ADARs

[01172] 3. Squid Sepia ADARS; Doryteusthis opalescens ADARS

[01173] ADATs (human ADAT, Drosophila ADAT)

[01174] Mutations can also be used to increase the activity of ADAR reacting against a DNA-RNAn RNA duplex. For example, for the human ADAR genes, the hADAR1d(E1008Q) or hADAR2d(E488Q) mutation is used to increase their activity against a DNA-RNA target.

[01175] Each ADAR has varying levels of sequence context requirement. For example, for hADAR1d (E1008Q), tAg and aAg sites are efficiently deaminated, whereas aAt and cAc are less efficiently edited, and gAa and gAc are even less edited. However, the context requirement will vary for different ADARs.

[01176] A schematic showing of one version of the system is provided in Figure 1. The amino acid sequences of example AD proteins are provided in Figures 4.

EXAMPLE 2

Cluc/Gluc tiling for Cas13a/Cas13b interference

[01177] To compare knockdown efficiency between Cas13a and Cas13b, Cypridina and Gaussia luciferase genes were tiled with 24 or 96 guides, respectively (Fig. 10). Guides were matched for Cas13a and Cas13b, and show increased knockdown efficiency for Cas13b, with all but one guide for each gene showing higher efficiency for Cas13b.

ADAR editing quantification by NGS

[01178] Cas13b-ADAR2 RNA editing efficiency was tested by designing a luciferase reporter with a premature stop codon UAG, which prevents expression of the luciferase (Fig. 11A). 7 guides of varying length were designed and positioned relative to the UAG stop codon that all contained a C mismatch to the A in the UAG. The C mismatch is known to create a bubble at the site of editing which is favored by the ADAR catalytic domain. RNA editing by Cas13b-ADAR2 would convert the UAG to a UIG (UGG), which introduces a tryptophan instead of the stop codon, and allows translation to proceed. Expression of the guides and Cas13b12-ADAR2 fusion in HEK293FT cells restored luciferase expression to varying levels with the greatest restoration occurring for guide 5 (Fig. 111B). In general, there is increasing levels of editing from guides 1-5 as the editing site is moved further away from the 3' end of the crRNA where the direct repeat is and thus where the protein binds. This likely indicates that the part of the crRNA:target duplex that is bound by the protein is inaccessible to the ADAR catalytic domain. Guides 5, 6, and 7 show the greatest amount of activity because the editing site is on the far end of the guide away from the DR/protein binding area and because their guides are much longer, generating a longer RNA duplex that is favored by ADAR. ADAR activity is optimal when the editing site is in the middle of a RNA duplex. The relative expression of luciferase activity is normalized to the non-targeting guide condition.

[01179] These samples were sequenced to precisely quantitate the RNA editing efficiency (Fig. 11C). The editing efficiency is listed in parentheses next to the guide label.

Overall, the percent editing identified by sequencing matched the relative levels of luciferase expression restoration seen in Fig. 11B. Guide 5 showed the most RNA editing with a rate of 45% conversion to G at the on-target A. In some instances, there is a small amount of off-target A-G editing in the region. These may be reduced by introducing G mismatches in the guide sequence, which are disfavored by the ADAR catalytic domain.

[01180] In addition to editing the luciferase reporter transcript, guides were designed to edit out-of-frame UAG sites in the KRAS and PPIB transcripts, with two guides targeting each transcript (Fig. 12). The guides were designed with the same principles as guide 5 above (a 45nt spacer with the editing site 27nt away from the 3' DR and a C mismatch to the editing site adenosine). The KRAS guides were able to achieve 6.5% and 13.7% editing at the on-target adenosine and the PPIB guides were able to achieve 7.7% and 9.2% editing. There are also some off-targets present for some of these guides which can be reduced by designing G mismatches in the spacers against possible off-target adenosines that are nearby. It does seem that off-targets seem to happen with the duplex region 3' of the target adenosine.

Cas13a/b + shRNA specificity from RNA Seq

[01181] To determine the specificity of the Cas13b12 knockdown, RNA sequencing was performed on all mRNAs across the transcriptome (Fig. 13A). The knockdown of guides targeting Gluc and KRAS was compared against non-targeting guides and found that Cas13a2 and Cas13b12 had specific knockdown of the target transcript (red dot in Fig. 13A) while the shRNAs had many off-targets as evidenced by the greater variance in the distribution. The number of significant off-targets for each of these conditions is shown in Fig. 13B. Significant off-targets are measured by a t-test with FDR correction ($p < 0.01$) for any off-target transcripts that are changed by greater than 2 fold or less than 0.8 fold. The Cas13a and Cas13b conditions had very few off targets compared to the hundreds of off-targets found for the shRNA conditions. The knockdown efficiency for each of the conditions is shown in Fig. 13C.

Mismatch specificity to reduce off targets (A:A or A:G)

[01182] To reduce off targets at adenosines near the target adenosine editing site, guides were designed that have G or A mismatches to the potential off-target adenosines (Fig. 14 and Table below). Mismatches with G or A are not favored for activity by the ADAR catalytic domain.

Name	Guide
Luciferase guide WT with C mismatch (SEQ ID No. 162)	catagaatgttctaaaCCAtcctgcggcctctactctgc attcaa
Luciferase guide WT with C mismatch with 1 G MM (SEQ ID No. 163)	catagaatgttcGaaaCCAtcctgcggcctctactctg cattcaa
Luciferase guide WT with C mismatch with 2 G MM (SEQ ID No. 164)	catagaatgGtcGaaaCCAtcctgcggcctctactct gcattcaa
Luciferase guide WT with C mismatch with 1 G MM (SEQ ID No. 165)	catagaatgttcAaaaCCAtcctgcggcctctactctg cattcaa
Luciferase guide WT with C mismatch with 2 G MM (SEQ ID No. 166)	catagaatgAtcAaaaCCAtcctgcggcctctactct gcattcaa
KRAS guide WT with Cmismatch (SEQ ID No. 167)	ggtttctccatcaattacCacttgcttcctgtaggaatcct ctatt
KRAS guide with C mismatch with 1 G MM (SEQ ID No. 168)	ggtttctccatcaatGacCacttgcttcctgtaggaatcc tctatt
KRAS guide with C mismatch with 2 G MM (SEQ ID No. 169)	ggtttctccatcaaGGacCacttgcttcctgtaggaatc ctctatt
KRAS guide with C mismatch with 1 A MM (SEQ ID No. 170)	ggtttctccatcaatAacCacttgcttcctgtaggaatcc tctatt
KRAS guide with C mismatch with 2 A MM (SEQ ID No. 171)	ggtttctccatcaaAAacCacttgcttcctgtaggaatc ctctatt
PPIB guide WT with C mismatch (SEQ ID No. 172)	gcctttctctcctgtagcCaaggccacaaaattatccact gttttt
PPIB guide WT with C mismatch with 1 G MM (SEQ ID No. 173)	gcctttctctcctgGagcCaaggccacaaaattatccac tgttttt
PPIB guide WT with C mismatch with 1 A MM (SEQ ID No. 174)	gcctttctctcctgAagcCaaggccacaaaattatccac tgttttt

[01183] The guides in the Table above were designed to have a C mismatch against the on-target adenosine to be edited and G or A mismatches against known off-target sites (based off of the RNA sequencing from above). Mismatches in the spacer sequence are capitalized.

Mismatch for on-target activity

[01184] Prior research on the catalytic domain of ADAR2 has demonstrated that different bases opposite the target A can influence the amount of inosine editing (Zheng et al. (2017), Nucleic Acid Research, 45(6):3369-3377). Specifically, U and C are found opposite natural ADAR-edited A's, whereas G and A are not. To test whether or not A and G mismatches with the edited A can be used to suppress ADAR activity a guide known to be active with a C mismatch is tested with all other 3 possible bases on the luciferase reporter assay (Fig. 16). Relative activities are quantified by assessing luciferase activity. Guide sequences are provided in the Table below.

Mismatch	Guide sequence
Mismatch-C (SEQ ID No. 175)	GcatagaatgttctaaaCCAtcctgcggcctctactct gcattcaa
Mismatch-G (SEQ ID No. 176)	GcatagaatgttctaaaCGAtcctgcggcctctactct gcattcaa
Mismatch-T (SEQ ID No. 177)	GcatagaatgttctaaaCTAtcctgcggcctctactct gcattcaa
Mismatch-A (SEQ ID No. 178)	GcatagaatgttctaaaCAAtcctgcggcctctactct gcattcaa

Improvement of editing and reduction of off-target modification by chemical modification of gRNAs

[01185] gRNAs which are chemically modified as exemplified in Vogel et al. (2014), Angew Chem Int Ed, 53:6267-6271, doi:10.1002/anie.201402634) to reduce off-target activity and to improve on-target efficiency. 2′-O-methyl and phosphothioate modified guide RNAs in general improve editing efficiency in cells.

Motif Preference

[01186] ADAR has been known to demonstrate a preference for neighboring nucleotides on either side of the edited A (www.nature.com/nsmb/journal/v23/n5/full/nsmb.3203.html, Matthews et al. (2017), Nature Structural Mol Biol, 23(5): 426-433). The preference is systematically tested by targeting Cypridina luciferase transcripts with variable bases surrounding the targeted A (Fig. 17).

Larger bubbles to enhance RNA editing efficiency

[01187] To enhance RNA editing efficiency on non-preferred 5' or 3' neighboring bases, intentional mismatches in neighboring bases are introduced, which has been demonstrated in vitro to allow for editing of non-preferred motifs (https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gku272; Schneider et al (2014), Nucleic Acid Res, 42(10):e87); Fukuda et al. (2017), Scienticic Reports, 7, doi:10.1038/srep41478). Additional mismatches are tested, such as guanosine substitutions, to see if they reduce natural preferences (Fig. 18).

Editing of multiple A's in a transcript

[01188] Results suggest that As opposite Cs in the targeting window of the ADAR deaminase domain are preferentially edited over other bases. Additionally, As base-paired with Us within a few bases of the targeted base show low levels of editing by Cas13b-ADAR fusions, suggesting that there is flexibility for the enzyme to edit multiple As (Fig. 19). These two observations suggest that multiple As in the activity window of Cas13b-ADAR fusions could be specified for editing by mismatching all As to be edited with Cs. To test this the most promising guides from the optimization experiment are taken and multiple A:C mismatches in the activity window are designed to test the possibility of creating multiple A:I edits. The editing rates for this experiment is be quanitifed using NGS. To suppress potential off-target editing in the activity window, non-target As are paired with As or Gs (depending on the results from the base preference experiment).

Guide length titration for RNA editing

[01189] ADAR naturally works on inter- or intra-molecular RNA duplexes of >20 bp in length (see also Nishikura et al. (2010), Annu Rev Biochem, 79:321-349). The results demonstrated that longer crRNAs, resulting in longer duplexes, had higher levels of activity. To systematically compare the activity of guides of different lengths for RNA editing activity we have designed guides of 30, 50, 70 and 84 bases to correct the stop codon in our luciferase reporter assay (Fig. 20 and Table below). We have designed these guides such that the position of the edited A is present at all possible even distances within the mRNA:crRNA duplex with respect to the 3' end of the specificity determining region of the crRNA (i.e. +2, +4 etc.).

Sequence Name	Sequence

Guide_Cas13bC-luc_30_ADAR0Top (SEQ ID No. 179)	GCATCCTGCGGCCTCTACTCTGCATTCAATT
Guide_Cas13bC-luc_30_ADAR1Top(SEQ ID No. 180)	GACCATCCTGCGGCCTCTACTCTGCATTCAA
Guide_Cas13bC-luc_30_ADAR2Top(SEQ ID No. 181)	GAAACCATCCTGCGGCCTCTACTCTGCATTC
Guide_Cas13bC-luc_30_ADAR3Top(SEQ ID No. 182)	GCTAAACCATCCTGCGGCCTCTACTCTGCAT
Guide_Cas13bC-luc_30_ADAR4Top(SEQ ID No. 183)	GTTCTAAACCATCCTGCGGCCTCTACTCTGC
Guide_Cas13bC-luc_30_ADAR5Top(SEQ ID No. 184)	GTGTTCTAAACCATCCTGCGGCCTCTACTCT
Guide_Cas13bC-luc_30_ADAR6Top(SEQ ID No. 185)	GAATGTTCTAAACCATCCTGCGGCCTCTACT
Guide_Cas13bC-luc_30_ADAR7Top(SEQ ID No. 186)	GAGAATGTTCTAAACCATCCTGCGGCCTCTA
Guide_Cas13bC-luc_30_ADAR8Top(SEQ ID No. 187)	GATAGAATGTTCTAAACCATCCTGCGGCCTC
Guide_Cas13bC-luc_30_ADAR9Top(SEQ ID No. 188)	GCCATAGAATGTTCTAAACCATCCTGCGGCC
Guide_Cas13bC-luc_30_ADAR10Top(SEQ ID No. 189)	GTTCCATAGAATGTTCTAAACCATCCTGCGG
Guide_Cas13bC-luc_30_ADAR11Top(SEQ ID No. 190)	GCTTTCCATAGAATGTTCTAAACCATCCTGC
Guide_Cas13bC-luc_30_ADAR12Top(SEQ ID No. 191)	GCTCTTTCCATAGAATGTTCTAAACCATCCT
Guide_Cas13bC-luc_30_ADAR13Top(SEQ ID No. 192)	GATCTCTTTCCATAGAATGTTCTAAACCATC
Guide_Cas13bC-luc_30_ADAR14Top(SEQ ID No. 193)	GGAATCTCTTTCCATAGAATGTTCTAAACCA
Guide_Cas13bC-luc_50_ADAR0Top(SEQ ID No. 194)	GCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCA
Guide_Cas13bC-luc_50_ADAR1Top(SEQ ID No. 195)	GACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGG
Guide_Cas13bC-luc_50_ADAR2Top(SEQ ID No. 196)	GAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTC
Guide_Cas13bC-luc_50_ADAR3Top(SEQ ID No. 197)	GCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACAT
Guide_Cas13bC-luc_50_ADAR4Top(SEQ ID No. 198)	GTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACAC
Guide_Cas13bC-luc_50_ADAR5Top(SEQ ID No. 199)	GTGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGAC

Guide_Cas13bC-luc_50_ADAR6Top(SEQ ID No. 200)	GAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTG
Guide_Cas13bC-luc_50_ADAR7Top(SEQ ID No. 201)	GAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATAC
Guide_Cas13bC-luc_50_ADAR8Top(SEQ ID No. 202)	GATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACAT
Guide_Cas13bC-luc_50_ADAR9Top(SEQ ID No. 203)	GCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTAC
Guide_Cas13bC-luc_50_ADAR10Top(SEQ ID No. 204)	GTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATT
Guide_Cas13bC-luc_50_ADAR11Top(SEQ ID No. 205)	GCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAA
Guide_Cas13bC-luc_50_ADAR12Top(SEQ ID No. 206)	GCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTC
Guide_Cas13bC-luc_50_ADAR13Top(SEQ ID No. 207)	GATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCAT
Guide_Cas13bC-luc_50_ADAR14Top(SEQ ID No. 208)	GGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGC
Guide_Cas13bC-luc_50_ADAR15Top(SEQ ID No. 209)	GTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCT
Guide_Cas13bC-luc_50_ADAR16Top(SEQ ID No. 210)	GACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACT
Guide_Cas13bC-luc_50_ADAR17Top(SEQ ID No. 211)	GGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTA
Guide_Cas13bC-luc_50_ADAR18Top(SEQ ID No. 212)	GTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTC
Guide_Cas13bC-luc_50_ADAR19Top(SEQ ID No. 213)	GCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCC
Guide_Cas13bC-luc_50_ADAR20Top(SEQ ID No. 214)	GTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGG
Guide_Cas13bC-luc_50_ADAR21Top(SEQ ID No. 215)	GGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGC
Guide_Cas13bC-luc_50_ADAR22Top(SEQ ID No. 216)	GCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCT
Guide_Cas13bC-luc_50_ADAR23Top(SEQ ID No. 217)	GACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATC

Guide_Cas13bC-luc_50_ADAR24Top(SEQ ID No. 218)	GGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCA
Guide_Cas13bC-luc_70_ADAR0Top(SEQ ID No. 219)	GCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTTCCTGGTTTAT
Guide_Cas13bC-luc_70_ADAR1Top(SEQ ID No. 220)	GACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTTCCTGGTTT
Guide_Cas13bC-luc_70_ADAR2Top(SEQ ID No. 221)	GAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTTCCTGGT
Guide_Cas13bC-luc_70_ADAR3Top(SEQ ID No. 222)	GCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTTCCTG
Guide_Cas13bC-luc_70_ADAR4Top(SEQ ID No. 223)	GTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTTCC
Guide_Cas13bC-luc_70_ADAR5Top(SEQ ID No. 224)	GTGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTT
Guide_Cas13bC-luc_70_ADAR6Top(SEQ ID No. 225)	GAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTT
Guide_Cas13bC-luc_70_ADAR7Top(SEQ ID No. 226)	GAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGT
Guide_Cas13bC-luc_70_ADAR8Top(SEQ ID No. 227)	GATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACAT
Guide_Cas13bC-luc_70_ADAR9Top(SEQ ID No. 228)	GCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAAC
Guide_Cas13bC-luc_70_ADAR10Top(SEQ ID No. 229)	GTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCA
Guide_Cas13bC-luc_70_ADAR11Top(SEQ ID No. 230)	GCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGG
Guide_Cas13bC-luc_70_ADAR12Top(SEQ ID No. 231)	GCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTC
Guide_Cas13bC-luc_70_ADAR13Top(SEQ ID No. 232)	GATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACAT
Guide_Cas13bC-luc_70_ADAR14Top(SEQ ID No. 233)	GGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACAC

Guide_Cas13bC-luc_70_ADAR15Top(SEQ ID No. 234)	GTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGAC
Guide_Cas13bC-luc_70_ADAR16Top(SEQ ID No. 235)	GACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTG
Guide_Cas13bC-luc_70_ADAR17Top(SEQ ID No. 236)	GGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATAC
Guide_Cas13bC-luc_70_ADAR18Top(SEQ ID No. 237)	GTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACAT
Guide_Cas13bC-luc_70_ADAR19Top(SEQ ID No. 238)	GCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTAC
Guide_Cas13bC-luc_70_ADAR20Top(SEQ ID No. 239)	GTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATT
Guide_Cas13bC-luc_70_ADAR21Top(SEQ ID No. 240)	GGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAA
Guide_Cas13bC-luc_70_ADAR22Top(SEQ ID No. 241)	GCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTC
Guide_Cas13bC-luc_70_ADAR23Top(SEQ ID No. 242)	GACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCAT
Guide_Cas13bC-luc_70_ADAR24Top(SEQ ID No. 243)	GGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGC
Guide_Cas13bC-luc_70_ADAR25Top(SEQ ID No. 244)	GATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCT
Guide_Cas13bC-luc_70_ADAR26Top(SEQ ID No. 245)	GGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACT
Guide_Cas13bC-luc_70_ADAR27Top(SEQ ID No. 246)	GACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTA
Guide_Cas13bC-luc_70_ADAR28Top(SEQ ID No. 247)	GACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTC
Guide_Cas13bC-luc_70_ADAR29Top(SEQ ID No. 248)	GCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCC
Guide_Cas13bC-luc_70_ADAR30Top(SEQ ID No. 249)	GCCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGG

Guide_Cas13bC-luc_70_ADAR31Top(SEQ ID No. 250)	GGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGC
Guide_Cas13bC-luc_70_ADAR32Top(SEQ ID No. 251)	GTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCT
Guide_Cas13bC-luc_70_ADAR33Top(SEQ ID No. 252)	GCCTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATC
Guide_Cas13bC-luc_70_ADAR34Top(SEQ ID No. 253)	GTTCCTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCA
Guide_Cas13bC-luc_84_ADAR0Top(SEQ ID No. 254)	GCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTTCCTGGTTTATTTTCACACAGTCCA
Guide_Cas13bC-luc_84_ADAR1Top(SEQ ID No. 255)	GACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTTCCTGGTTTATTTTCACACAGTC
Guide_Cas13bC-luc_84_ADAR2Top(SEQ ID No. 256)	GAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTTCCTGGTTTATTTTCACACAG
Guide_Cas13bC-luc_84_ADAR3Top(SEQ ID No. 257)	GCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTTCCTGGTTTATTTTCACAC
Guide_Cas13bC-luc_84_ADAR4Top(SEQ ID No. 258)	GTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTTCCTGGTTTATTTTCAC
Guide_Cas13bC-luc_84_ADAR5Top(SEQ ID No. 259)	GTGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTTCCTGGTTTATTTTC
Guide_Cas13bC-luc_84_ADAR6Top(SEQ ID No. 260)	GAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTTCCTGGTTTATTT
Guide_Cas13bC-luc_84_ADAR7Top(SEQ ID No. 261)	GAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTTCCTGGTTTAT
Guide_Cas13bC-luc_84_ADAR8Top(SEQ ID No. 262)	GATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTCCTGGTTT

Guide_Cas13bC-luc_84_ADAR9Top(SEQ ID No. 263)	GCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTTCCTGGT
Guide_Cas13bC-luc_84_ADAR10Top(SEQ ID No. 264)	GTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTTCCTG
Guide_Cas13bC-luc_84_ADAR11Top(SEQ ID No. 265)	GCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTTCC
Guide_Cas13bC-luc_84_ADAR12Top(SEQ ID No. 266)	GCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTT
Guide_Cas13bC-luc_84_ADAR13Top(SEQ ID No. 267)	GATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTT
Guide_Cas13bC-luc_84_ADAR14Top(SEQ ID No. 268)	GGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGT
Guide_Cas13bC-luc_84_ADAR15Top(SEQ ID No. 269)	GTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACAT
Guide_Cas13bC-luc_84_ADAR16Top(SEQ ID No. 270)	GACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAAC
Guide_Cas13bC-luc_84_ADAR17Top(SEQ ID No. 271)	GGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCA
Guide_Cas13bC-luc_84_ADAR18Top(SEQ ID No. 272)	GTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGG
Guide_Cas13bC-luc_84_ADAR19Top(SEQ ID No. 273)	GCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTC
Guide_Cas13bC-luc_84_ADAR20Top(SEQ ID No. 274)	GTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACAT

Guide_Cas13bC-luc_84_ADAR21Top(SEQ ID No. 275)	GGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACAC
Guide_Cas13bC-luc_84_ADAR22Top(SEQ ID No. 276)	GCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGAC
Guide_Cas13bC-luc_84_ADAR23Top(SEQ ID No. 277)	GACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTG
Guide_Cas13bC-luc_84_ADAR24Top(SEQ ID No. 278)	GGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATAC
Guide_Cas13bC-luc_84_ADAR25Top(SEQ ID No. 279)	GATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACAT
Guide_Cas13bC-luc_84_ADAR26Top(SEQ ID No. 280)	GGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTAC
Guide_Cas13bC-luc_84_ADAR27Top(SEQ ID No. 281)	GACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATT
Guide_Cas13bC-luc_84_ADAR28Top(SEQ ID No. 282)	GACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAA
Guide_Cas13bC-luc_84_ADAR29Top(SEQ ID No. 283)	GCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTC
Guide_Cas13bC-luc_84_ADAR30Top(SEQ ID No. 284)	GCCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCAT
Guide_Cas13bC-luc_84_ADAR31Top(SEQ ID No. 285)	GGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGC
Guide_Cas13bC-luc_84_ADAR32Top(SEQ ID No. 286)	GTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCT

Guide_Cas13bC-luc_84_ADAR33Top(SEQ ID No. 287)	GCCTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACT
Guide_Cas13bC-luc_84_ADAR34Top(SEQ ID No. 288)	GTTCCTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTA
Guide_Cas13bC-luc_84_ADAR35Top(SEQ ID No. 289)	GGGTTCCTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTC
Guide_Cas13bC-luc_84_ADAR36Top(SEQ ID No. 290)	GTTGGTTCCTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCC
Guide_Cas13bC-luc_84_ADAR37Top(SEQ ID No. 291)	GCCTTGGTTCCTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGG
Guide_Cas13bC-luc_84_ADAR38Top(SEQ ID No. 292)	GGCCCTTGGTTCCTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGC
Guide_Cas13bC-luc_84_ADAR39Top(SEQ ID No. 293)	GCCGCCCTTGGTTCCTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCT
Guide_Cas13bC-luc_84_ADAR40Top(SEQ ID No. 294)	GCGCCGCCCTTGGTTCCTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATC
Guide_Cas13bC-luc_84_ADAR41Top(SEQ ID No. 295)	GGTCGCCGCCCTTGGTTCCTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCA

Reversing causal disease mutations

[01190] The three genes in the Table below are synthesised with the pathogenic G>A mutation that introduces a pre-termination stop site into the gene and integrate them into a non-human cell line. The ability of Cas13b12-ADAR2 to correct the transcripts by changing the stop codon UAG to UIG (UGG) and thus restore protein translation is tested.

Protein length (aa)	**Full length candidates**	**Gene**	**Disease**

498	NM_004992.3(MECP2):c.311G>A (p.Trp104Ter)	MECP2	Rett syndrome
414	NM_007375.3(TARDBP):c.943G>A (p.Ala315Thr)	TARDBP	Amyotrophic lateral sclerosis type 10
393	NM_000546.5(TP53):c.273G>A (p.Trp91Ter)	TP53	Li-Fraumeni syndrome|Hereditary cancer-predisposing syndrome

[01191] Forty-eight more pathogenic G>A mutations are shown in Table 5 below along with the accompanying disease. 200bp fragments around these mutations are synthesised, rather than the entire gene, and cloned in front of a GFP. When the pre-termination site is restored, that will allow translation of the GFP and correction can be measured by fluorescence in high-throughput - in addition to RNA sequencing.

Table 5

	Candidate	**Gene**	**Disease**
1	NM_004006.2(DMD):c.3747G>A (p.Trp1249Ter)	DMD	Duchenne muscular dystrophy
2	M_000344.3(SMN1):c.305G>A (p.Trp102Ter)	SMN1	Spinal muscular atrophy, type II|Kugelberg-Welander disease
3	NM_000492.3(CFTR):c.3846G>A (p.Trp1282Ter)	CFTR	Cystic fibrosis|Hereditary pancreatitis|not provided|ataluren response - Efficacy
4	NM_004562.2(PRKN):c.1358G>A (p.Trp453Ter)	PRKN	Parkinson disease 2
5	NM_017651.4(AHI1):c.2174G>A (p.Trp725Ter)	AHI1	Joubert syndrome 3
6	NM_000238.3(KCNH2):c.3002G>A (p.Trp1001Ter)	KCNH2	Long QT syndrome|not provided
7	NM_000136.2(FANCC):c.1517G>A (p.Trp506Ter)	FANCC|C9orf3	Fanconi anemia, complementation group C
8	NM_001009944.2(PKD1):c.12420G>A (p.Trp4140Ter)	PKD1	Polycystic kidney disease, adult type
9	NM_177965.3(C8orf37):c.555G>A (p.Trp185Ter)	C8orf37	Retinitis pigmentosa 64
10	NM_000833.4(GRIN2A):c.3813G>A (p.Trp1271Ter)	GRIN2A	Epilepsy, focal, with speech disorder and with or without mental retardation

11	NM_000548.4(TSC2):c.2108G>A (p.Trp703Ter)	TSC2	Tuberous sclerosis 2|Tuberous sclerosis syndrome
12	NM_000267.3(NF1):c.7044G>A (p.Trp2348Ter)	NF1	Neurofibromatosis, type 1
13	NM_000520.5(HEXA):c.1454G>A (p.Trp485Ter)	HEXA	Tay-Sachs disease
14	NM_130838.1(UBE3A):c.2304G>A (p.Trp768Ter)	UBE3A	Angelman syndrome
15	NM_000543.4(SMPD1):c.168G>A (p.Trp56Ter)	SMPD1	Niemann-Pick disease, type A
16	NM_000218.2(KCNQ1):c.1175G>A (p.Trp392Ter)	KCNQ1	Long QT syndrome
17	NM_000256.3(MYBPC3):c.3293G>A (p.Trp1098Ter)	MYBPC3	Primary familial hypertrophic cardiomyopathy
18	NM_000038.5(APC):c.1262G>A (p.Trp421Ter)	APC	Familial adenomatous polyposis 1
19	NM_000249.3(MLH1):c.1998G>A (p.Trp666Ter)	MLH1	Lynch syndrome
20	NM_000054.4(AVPR2):c.878G>A (p.Trp293Ter)	AVPR2	Nephrogenic diabetes insipidus, X-linked
21	NM_001204.6(BMPR2):c.893G>A (p.W298*)	BMPR2	Primary pulmonary hypertension
22	NM_004560.3(ROR2):c.2247G>A (p.Trp749Ter)	ROR2	Brachydactyly type B1
23	NM_000518.4(HBB):c.114G>A (p.Trp38Ter)	HBB	beta^0^ Thalassemia|beta Thalassemia
24	NM_024577.3(SH3TC2):c.920G>A (p.Trp307Ter)	SH3TC2	Charcot-Marie-Tooth disease, type 4C
25	NM_206933.2(USH2A):c.9390G>A (p.Trp3130Ter)	USH2A	Usher syndrome, type 2A
26	NM_000179.2(MSH6):c.3020G>A (p.Trp1007Ter)	MSH6	Lynch syndrome
27	NM_002977.3(SCN9A):c.2691G>A (p.Trp897Ter)	SCN9A|LOC101929680	Indifference to pain, congenital, autosomal recessive
28	NM_000090.3(COL3A1):c.30G>A (p.Trp10Ter)	COL3A1	Ehlers-Danlos syndrome, type 4
29	NM_000551.3(VHL):c.263G>A (p.Trp88Ter)	VHL	Von Hippel-Lindau syndrome|not provided
30	NM_015627.2(LDLRAP1):c.65G>A (p.Trp22Ter)	LDLRAP1	Hypercholesterolemia, autosomal recessive

31	NM_000132.3(F8):c.3144G>A (p.Trp1048Ter)	F8	Hereditary factor VIII deficiency disease
32	NM_002185.4(IL7R):c.651G>A (p.Trp217Ter)	IL7R	Severe combined immunodeficiency, autosomal recessive, T cell-negative, B cell-positive, NK cell-positive
33	NM_000527.4(LDLR):c.1449G>A (p.Trp483Ter)	LDLR	Familial hypercholesterolemia
34	NM_002294.2(LAMP2):c.962G>A (p.Trp321Ter)	LAMP2	Danon disease
35	NM_000271.4(NPC1):c.1142G>A (p.Trp381Ter)	NPC1	Niemann-Pick disease type C1
36	NM_000267.3(NF1):c.1713G>A (p.Trp571Ter)	NF1	Neurofibromatosis, type 1
37	NM_000035.3(ALDOB):c.888G>A (p.Trp296Ter)	ALDOB	Hereditary fructosuria
38	NM_000090.3(COL3A1):c.3833G>A (p.Trp1278Ter)	COL3A1	Ehlers-Danlos syndrome, type 4
39	NM_001369.2(DNAH5):c.8465G>A (p.Trp2822Ter)	DNAH5	Primary ciliary dyskinesia
40	NM_178443.2(FERMT3):c.48G>A (p.Trp16Ter)	FERMT3	Leukocyte adhesion deficiency, type III
41	NM_005359.5(SMAD4):c.906G>A (p.Trp302Ter)	SMAD4	Juvenile polyposis syndrome
42	NM_032119.3(ADGRV1):c.7406G>A (p.Trp2469Ter)	ADGRV1	Usher syndrome, type 2C
43	NM_000206.2(IL2RG):c.710G>A (p.Trp237Ter)	IL2RG	X-linked severe combined immunodeficiency
44	NM_007294.3(BRCA1):c.5511G>A (p.Trp1837Ter)	BRCA1	Familial cancer of breast|Breast-ovarian cancer, familial 1
45	NM_130799.2(MEN1):c.1269G>A (p.Trp423Ter)	MEN1	Hereditary cancer-predisposing syndrome
46	NM_000071.2(CBS):c.162G>A (p.Trp54Ter)	CBS	Homocystinuria due to CBS deficiency
47	NM_000059.3(BRCA2):c.582G>A (p.Trp194Ter)	BRCA2	Familial cancer of breast|Breast-ovarian cancer, familial 2
48	NM_000053.3(ATP7B):c.2336G>A (p.Trp779Ter)	ATP7B	Wilson disease

EXAMPLE 3

[01192] Efficient and precise nucleic acid editing holds great promise for treating genetic disease, particularly at the level of RNA, where disease-relevant transcripts can be rescued to yield functional protein products. Type VI CRISPR-Cas systems contain the programmable single-effector RNA-guided RNases Cas13. Here, we profile the diversity of Type VI systems to engineer a Cas13 ortholog capable of robust knockdown and demonstrate RNA editing by using catalytically-inactive Cas13 (dCas13) to direct adenosine deaminase activity to transcripts in mammalian cells. By fusing the ADAR2 deaminase domain to dCas13 and engineering guide RNAs to create an optimal RNA duplex substrate, we achieve targeted editing of specific single adenosines to inosines (which is read out as guanosine during translation) with efficiencies routinely ranging from 20-40% and up to 89%. This system, referred to as RNA Editing for Programmable A to I Replacement (REPAIR), can be further engineered to achieve high specificity. An engineered variant, REPAIRv2, displays greater than 170-fold increase in specificity while maintaining robust on-target A to I editing as well as minimize the system to ease viral delivery. We use REPAIRv2 to edit full-length transcripts containing known pathogenic mutations and create functional truncated versions suitable for packaging in adeno-associated viral (AAV) vectors. REPAIR presents a promising RNA editing platform with broad applicability for research, therapeutics, and biotechnology. Precise nucleic acid editing technologies are valuable for studying cellular function and as novel therapeutics. Although current editing tools, such as the Cas9 nuclease, can achieve programmable modification of genomic loci, edits are often heterogenous due to insertions or deletions or require a donor template for precise editing. Base editors, such as dCas9-APOBEC fusions, allow for editing without generating a double stranded break, but may lack precision due to the nature of cytidine deaminase activity, which edits any cytidine in a target window. Furthermore, the requirement for a protospacer adjacent motif (PAM) limits the number of possible editing sites. Here, we describe the development of a precise and flexible RNA base editing tool using the RNA-guided RNA targeting Cas13 enzyme from type VI prokaryotic clustered regularly interspaced short palindromic repeats (CRISPR) adaptive immune system.

[01193] Precise nucleic acid editing technologies are valuable for studying cellular function and as novel therapeutics. Current editing tools, based on programmable nucleases such as the prokaryotic clustered regularly interspaced short palindromic repeats (CRISPR)-associated nucleases Cas9 (1-4) or Cpf1(5), have been widely adopted for mediating targeted DNA cleavage which in turn drives targeted gene disruption through non-homologous end

joining (NHEJ) or precise gene editing through template-dependent homology-directed repair (HDR)(6). NHEJ utilizes host machineries that are active in both dividing and post-mitotic cells and provides efficient gene disruption by generating a mixture of insertion or deletion (indel) mutations that can lead to frame shifts in protein coding genes. HDR, in contrast, is mediated by host machineries whose expression is largely limited to replicating cells. As such, the development of gene-editing capabilities in post-mitotic cells remains a major challenge. Recently, DNA base editors, such as the use of catalytically inactive Cas9 (dCas9) to target cytidine deaminase activity to specific genome targets to effect cytosine to thymine conversions within a target window, allow for editing without generating a DNA double strand break and significantly reduces the formation of indels(7, 8). However the targeting range of DNA base editors is limited due to the requirement of Cas9 for a protospacer adjacent motif (PAM) at the editing site(9). Here, we describe the development of a precise and flexible RNA base editing technology using the type VI CRISPR-associated RNA-guided RNase Cas13(10-13).

[01194] Cas13 enzymes have two Higher Eukaryotes and Prokaryotes Nucleotide-binding (HEPN) endoRNase domains that mediate precise RNA cleavage(10, 11). Three Cas13 protein families have been identified to date: Cas13a (previously known as C2c2), Cas13b, and Cas13c(12, 13). We recently reported Cas13a enzymes can be adapted as tools for nucleic acid detection(14) as well as mammalian and plant cell RNA knockdown and transcript tracking(15). The RNA-guided nature of Cas13 enzymes makes them attractive tool for RNA binding and perturbation applications.

[01195] The adenosine deaminase acting on RNA (ADAR) family of enzymes mediates endogenous editing of transcripts via hydrolytic deamination of adenosine to inosine, a nucleobase that is functionally equivalent to guanosine in translation and splicing(16). There are two functional human ADAR orthologs, ADAR1 and ADAR2, which consist of N-terminal double stranded RNA-binding domains and a C-terminal catalytic deamination domain. Endogenous target sites of ADAR1 and ADAR2 contain substantial double stranded identity, and the catalytic domains require duplexed regions for efficient editing in vitro and in vivo (18, 19). Importantly, the ADAR catalytic domain is capable of deaminating target adenosines without any protein co-factors in vitro (20). ADAR1 has been found to target mainly repetitive regions whereas ADAR2 mainly targets non-repetitive coding regions (17).Although ADAR proteins have preferred motifs for editing that could restrict the potential flexibility of targeting, hyperactive mutants, such as

ADAR(E488Q)(21), relax sequence constraints and improve adenosine to inosine editing rates. ADARs preferentially deaminate adenosines opposite cytidine bases in RNA duplexes(22), providing a promising opportunity for precise base editing. Although previous approaches have engineered targeted ADAR fusions via RNA guides (23-26), the specificity of these approaches has not been reported and their respective targeting mechanisms rely on RNA-RNA hybridization without the assistance of protein partners that may enhance target recognition and stringency.

[01196] Here we assay the entire family of Cas13 enzymes for RNA knockdown activity in mammalian cells and identify the Cas13b ortholog from Prevotella sp. P5-125 (PspCas13b) as the most efficient and specific for mammalian cell applications. We then fuse the ADAR2 deaminase domain (ADARDD) to catalytically inactive PspCas13b and demonstrate RNA editing for programmable A to I (G) replacement (REPAIR) of reporter and endogenous transcripts as well as disease-relevant mutations. Lastly, we employ a rational mutagenesis scheme to improve the specificity of dCas13b-ADAR2DD fusions to generate REPAIRv2 with more than 170 fold increase in specificity..

Methods

Design and cloning of bacterial constructs

[01197] Mammalian codon optimized Cas13b constructs were cloned into the chloramphenicol resistant pACYC184 vector under control of the Lac promoter. Two corresponding direct-repeat (DR) sequences separated by BsaI restriction sites were then inserted downstream of Cas13b, under control of the pJ23119 promoter. Last, oligos for targeting spacers were phosphorylated using T4 PNK (New England Biolabs), annealed and ligated into BsaI digested vectors using T7 ligase (Enzymatics) to generate targeting Cas13b vectors. Guide sequences used are in Table 11.

Bacterial PFS screens

[01198] Ampicillin resistance plasmids for PFS screens were cloned by inserting PCR products containing Cas13b targets with 2 5' randomized nucleotides and 4 3' randomized nucleotides separated by a target site immediately downstream of the start codon of the ampicillin resistance gene bla using NEB Gibson Assembly (New England Biolabs). 100 ng of ampicillin-resistant target plasmids were then electroporated with 65-100 ng chloramphenicol-resistant Cas13b bacterial targeting plasmids into Endura Electrocompetent Cells. Plasmids were added to cells, incubated 15 minutes on ice, electroporated using the manufacturer's protocol, and then 950 uL of recovery media was

added to cells before a one hour outgrowth at 37C. The outgrowth was plated onto chloramphenicol and ampicillin double selection plates. Serial dilutions of the outgrowth were used to estimate the cfu/ng DNA. 16 hours post plating, cells were scraped off plates and surviving plasmid DNA harvested using the Qiagen Plasmid Plus Maxi Kit (Qiagen). Surviving Cas13b target sequences and their flanking regions were amplified by PCR and sequenced using an Illumina NextSeq. To assess PFS preferences, the positions containing randomized nucleotides in the original library were extracted, and sequences depleted relative to the vector only condition that were present in both bioreplicates were extracted using custom python scripts. The -log2 of the ratio of PFS abundance in the Cas13b condition compared to the vector only control was then used to calculate preferred motifs. Specifically, all sequences having -log2(sample/vector) depletion ratios above a specific threshold were used to generate weblogos of sequence motifs (weblogo.berkeley.edu). The specific depletion ratio values used to generate weblogos for each Cas13b ortholog are listed in Table 9.

Design and cloning of mammalian constructs for RNA interference

[01199] To generate vectors for testing Cas13 orthologs in mammalian cells, mammalian codon optimized Cas13a, Cas13b, and Cas13c genes were PCR amplified and golden-gate cloned into a mammalian expression vector containing dual NLS sequences and a C-terminal msfGFP, under control of the EF1alpha promoter. For further optimization Cas13 orthologs were golden gate cloned into destination vectors containing different C-terminal localization tags under control of the EF1alpha promoter.

[01200] The dual luciferase reporter was cloned by PCR amplifying Gaussia and Cypridinia luciferase coding DNA, the EF1alpha and CMV promoters and assembly using the NEB Gibson Assembly (New England Biolabs).

[01201] For expression of mammalian guide RNA for Cas13a, Cas13b, or Cas13c orthologs, the corresponding direct repeat sequences were synthesized with golden-gate acceptor sites and cloned under U6 expression via restriction digest cloning. Individual guides were then cloned into the corresponding expression backbones for each ortholog by golden gate cloning. All Cas13 plasmids are listed in Supplementary Table 10. All Cas13 guide sequences for knockdown experiments are listed in Supplementary Tables 11-13.

Measurement of Cas13 expression in mammalian cells

[01202] Dual-NLS Cas13-msfGFP constructs were transfected into HEK293FT cells with targeting and non-targeting guides. GFP fluorescence was measured 48 hours post transfection in the non-targeting guide condition using a plate reader.

Cloning of pooled mismatch libraries for Cas13 interference specificity

[01203] Pooled mismatch library target sites were created by PCR. Oligos containing semi-degenerate target sequences in G-luciferase containing a mixture of 94% of the correct base and 2% of each incorrect base at each position within the target were used as one primer, and an oligo corresponding to a non-targeted region of G-luciferase was used as the second primer in the PCR reaction. The mismatch library target was then cloned into the dual luciferase reporter in place of the wildtype G-luciferase using NEB Gibson assembly (New England Biolabs).

Design and cloning of mammalian constructs for RNA editing

[01204] PspCas13b was made catalytically inactive (dPspCas13b) via two histidine to alanine mutations (H133A/H1058A) at the catalytic site of the HEPN domains. The deaminase domains of human ADAR1 and ADAR2 were synthesized and PCR amplified for gibson cloning into pcDNA-CMV vector backbones and were fused to dPspCas13b at the C-terminus via GS or GSGGGGS (SEQ ID No. 296) linkers. For the experiment in which we tested different linkers we cloned the following additional linkers between dPspCas13b and ADAR2dd: GGGGSGGGGSGGGGS, EAAAK (SEQ ID No. 297), GGSGGSGGSGGSGGSGGS (SEQ ID No. 298), and SGSETPGTSESATPES (SEQ ID No. 299)(XTEN). Specificity mutants were generated by gibson cloning the appropriate mutants into the dPspCas13b-GSGGGGS backbone.

[01205] The luciferase reporter vector for measuring RNA editing activity was generated by creating a W85X mutation (TGG>TAG) in the luciferase reporter vector used for knockdown experiments. This reporter vector expresses functional Gluc as a normalization control, but a defective Cluc due to the addition of a pretermination site. To test ADAR editing motif preferences, we cloned every possible motif around the adenosine at codon 85 (XAX) of Cluc. All plasmids are listed in Supplementary Table 10.

[01206] For testing PFS preference of REPAIR, we cloned a pooled plasmid library containing a 6 basepair degenerate PFS sequence upstream of a target region and adenosine editing site. The library was synthesized as an ultramer from Integrated DNA Technologies (IDT) and was made double stranded via annealing a primer and Klenow fragment of DNA polymerase I (New England Biolabs) fill in of the sequence. This dsDNA fragment

containing the degenerate sequence was then gibson cloned into the digested reporter vector and this was then isopropanol precipitated and purified. The cloned library was then electroporated into Endura competent E. coli cells (Lucigen) and plated on 245mm x 245mm square bioassay plates (Nunc). After 16 hours, colonies were harvested and midiprepped using endotoxin-free MACHEREY-NAGEL midiprep kits. Cloned libraries were verified by next generation sequencing.

[01207] For cloning disease-relevant mutations for testing REPAIR activity, 34 G>A mutations related to disease pathogenesis as defined in ClinVar were selected and 200bp regions surrounding these mutations were golden gate cloned between mScarlett and EGFP under a CMV promoter. Two additional G>A mutations in AVPR2 and FANCC were selected for Gibson cloning the whole gene sequence under expression of EF1alpha.

[01208] For expression of mammalian guide RNA for REPAIR, the PspCas13b direct repeat sequences were synthesized with golden-gate acceptor sites and cloned under U6 expression via restriction digest cloning. Individual guides were then cloned into this expression backbones by golden gate cloning. Guide sequences for REPAIR experiments are listed in Supplementary Table 14.

Mammalian cell culture

[01209] Mammalian cell culture experiments were performed in the HEK293FT line (American Type Culture Collection (ATCC)), which was grown in Dulbecco's Modified Eagle Medium with high glucose, sodium pyruvate, and GlutaMAX (Thermo Fisher Scientific), additionally supplemented with 1× penicillin–streptomycin (Thermo Fisher Scientific) and 10% fetal bovine serum (VWR Seradigm). Cells were maintained at confluency below 80%.

[01210] Unless otherwise noted, all transfections were performed with Lipofectamine 2000 (Thermo Fisher Scientific) in 96-well plates coated with poly-D-lysine (BD Biocoat). Cells were plated at approximately 20,000 cells/well sixteen hours prior to transfection to ensure 90% confluency at the time of transfection. For each well on the plate, transfection plasmids were combined with Opti-MEM I Reduced Serum Medium (Thermo Fisher) to a total of 25 µl. Separately, 24.5 ul of Opti-MEM was combined with 0.5 ul of Lipofectamine 2000. Plasmid and Lipofectamine solutions were then combined and incubated for 5 minutes, after which they were pipetted onto cells. The U2OS transfections were performed using Lipofectamine 3000 according to the manufacturer's protocol.

RNA knockdown mammalian cell assays

[01211] To assess RNA targeting in mammalian cells with reporter constructs, 150 ng of Cas13 construct was co-transfected with 300 ng of guide expression plasmid and 12.5 ng of the knockdown reporter construct. 48 hours post-transfection, media containing secreted luciferase was removed from cells, diluted 1:5 in PBS, and measured for activity with BioLux Cypridinia and Biolux Gaussia luciferase assay kits (New England Biolabs) on a plate reader (Biotek Synergy Neo2) with an injection protocol. All replicates performed are biological replicates.

[01212] For targeting of endogenous genes, 150 ng of Cas13 construct was co-transfected with 300 ng of guide expression plasmid. 48 hours post-transfection, cells were lysed and RNA was harvested and reverse transcribed using a previously described (33) modification of the Cells-to-Ct kit (Thermo Fisher Scientific). cDNA expression was measured via qPCR using TaqMan qPCR probes for the KRAS transcript (Thermo Fisher Scientific), GAPDH control probes (Thermo Fisher Scientific), and Fast Advanced Master Mix (Thermo Fisher Scientific). qPCR reactions were read out on a LightCycler 480 Instrument II (Roche), with four 5 ul technical replicates in 384-well format..

Evaluation of RNA specificity using pooled library of mismatched targets

[01213] The ability of Cas13 to interfere with the mismatched target library was tested using HEK293FT cells seeded in 6 well plates. ~70% confluent cells were transfected using 2400 ng Cas13 vector, 4800 ng of guide and 240 ng of mismatched target library. 48 hours post transfection, cells were harvested and RNA extracted using the QIAshredder (Qiagen) and the Qiagen RNeasy Mini Kit. 1ug of extracted RNA was reverse transcribed using the qScript Flex cDNA synthesis kit (Quantabio) following the manufacturer's gene-specific priming protocol and a Gluc specific RT primer. cDNA was then amplified and sequenced on an Illumina NextSeq.

[01214] The sequencing was analyzed by counting reads per sequence and depletion scores were calculated by determining the log2(-read count ratio) value, where read count ratio is the ratio of read counts in the targeting guide condition versus the non-targeting guide condition. This score value represents the level of Cas13 activity on the sequence, with higher values representing stronger depletion and thus higher Cas13 cleavage activity. Separate distributions for the single mismatch and double mismatch sequences were determined and plotted as heatmaps with a depletion score for each mismatch identity. For double mismatch sequences the average of all possible double mismatches at a given position were plotted.

Transcriptome-wide profiling of Cas13 in mammalian cells by RNA sequencing

[01215] For measurement of transcriptome-wide specificity, 150 ng of Cas13 construct, 300 ng of guide expression plasmid and 15 ng of the knockdown reporter construct were co-transfected; for shRNA conditions, 300 ng of shRNA targeting plasmid, 15 ng of the knockdown reporter construct, and 150 ng of EF1-alpha driven mCherry (to balance reporter load) were co-transfected. 48 hours after transfection, RNA was purified with the RNeasy Plus Mini kit (Qiagen), mRNA was selected for using NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs) and prepared for sequencing with the NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs). RNA sequencing libraries were then sequenced on a NextSeq (Illumina).

[01216] To analyze transcriptome-wide sequencing data, reads were aligned RefSeq GRCh38 assembly using Bowtie and RSEM version 1.2.31 with default parameters (34): accurate transcript quantification from RNA-Seq data with or without a reference genome]. Transcript expression was quantified as log2(TPM + 1), genes were filtered for log2(TPM + 1) >2.5 For selection of differentially expressed genes, only genes with differential changes of >2 or <.75 were considered. Statistical significance of differential expression was evaluated Student's T-test on three targeting replicates versus non-targeting replicates, and filtered for a false discovery rate of <0.01% by Benjamini-Hochberg procedure.

ADAR RNA editing in mammalian cells transfections

[01217] To assess REPAIR activity in mammalian cells, we transfected 150ng of REPAIR vector, 300ng of guide expression plasmid, and 40ng of the RNA editing reporter. After 48 hours, RNA from cells were harvested and reverse transcribed using a method previously described (33) with a gene specific reverse transcription primer. The extracted cDNA was then subjected to two rounds of PCR to add Illumina adaptors and sample barcodes using NEBNext High-Fidelity 2X PCR Master Mix (New England Biolabs). The library was then subjected to next generation sequencing on an Illumina NextSeq or MiSeq. RNA editing rates were then evaluated at all adenosine within the sequencing window.

[01218] In experiments where the luciferase reporter was targeted for RNA editing, we also harvested the media with secreted luciferase prior to RNA harvest. In this case, because the corrected Cluc might be at low levels, we did not dilute the media. We measured luciferase activity with BioLux Cypridinia and Biolux Gaussia luciferase assay kits (New England Biolabs) on a plate reader (Biotek Synergy Neo2) with an injection protocol. All replicates performed are biological replicates.

PFS binding mammalian screen

[01219] To determine the contribution of the PFS to editing efficiency, 625 ng of PFS target library, 4.7 ug of guide, and 2.35 ug of REPAIR were co-transfected on HEK293FT cells plated in 225 cm2 flasks. Plasmids were mixed with 33 ul of PLUS reagent (Thermo Fisher Scientific), brought to 533 ul with Opti-MEM, incubated for 5 minutes, combined with 30 ul of Lipofectamine 2000 and 500 ul of Opti-MEM, incubated for an additional 5 minutes, and then pipetted onto cells. 48 hours post-transfection, RNA was harvested with the RNeasy Plus Mini kit (Qiagen), reverse transcribed with qScript Flex (Quantabio) using a gene specific primer, and amplified with two rounds of PCR using NEBNext High-Fidelity 2X PCR Master Mix (New England Biolabs) to add Illumina adaptors and sample barcodes. The library was sequenced on an Illumina NextSeq, and RNA editing rates at the target adenosine were mapped to PFS identity. To increase coverage, the PFS was computationally collapsed to 4 nucleotides. REPAIR editing rates were calculated for each PFS, averaged over biological replicates with non-targeting rates for the corresponding PFS subtracted.

Whole-transcriptome sequencing to evaluate ADAR editing specificity

[01220] For analyzing off-target RNA editing sites across the transcriptome, we harvested total RNA from cells 48 hours post transfection using the RNeasy Plus Miniprep kit (Qiagen). The mRNA fraction is then enriched using a NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB) and this RNA is then prepared for sequencing using NEBNext Ultra RNA Library Prep Kit for Illumina (NEB). The libraries were then sequenced on an Illumina NextSeq and loaded such that there was at least 5 million reads per sample.

RNA editing analysis for targeted and transcriptome wide experiments

[01221] Analysis of the transcriptome-wide editing RNA sequencing data was performed on the FireCloud computational framework (https://software.broadinstitute.org/firecloud/) using a custom workflow we developed: https://portal.firecloud.org/#methods/m/rna_editing_final_workflow/rna_editing_final_workflow/1. For analysis, unless otherwise denoted, sequence files were randomly downsampled to 5 million reads. An index was generated using the RefSeq GRCh38 assembly with Gluc and Cluc sequences added and reads were aligned and quantified using Bowtie/RSEM version 1.3.0. Alignment BAMs were then sorted and analyzed for RNA editing sites using REDitools (35,36) with the following parameters: -t 8 -e -d -l -U [AG or TC] -p -u -m20 -T6-0 -W -v 1 -n 0.0. Any significant edits found in untransfected or EGFP-

transfected conditions were considered to be SNPs or artifacts of the transfection and filtered out from the analysis of off-targets. Off-targets were considered significant if the Fisher's exact test yielded a p-value less than 0.05 after multiple hypothesis correction by Benjamini Hochberg correction and at least 2 of 3 biological replicates identified the edit site. Overlap of edits between samples was calculated relative to the maximum possible overlap, equivalent to the fewer number of edits between the two samples. The percentage of overlapping edit sites was calculated as the number of shared edit sites divided by minimum number of edits of the two samples, multiplied by 100. For the high-coverage sequencing analysis, an additional layer of filtering for known SNP positions was performed using the Kaviar (*37*) method for identifying SNPs.

[01222] For analyzing the predicted variant effects of each off-target, the list of off-target edit sites was analyzed using the variant annotation integrator (https://genome.ucsc.edu/cgi-bin/hgVai) as part of the UCSC genome browser suite of tools using the SIFT and PolyPhen-2 annotations. To declare whether the off-target genes are oncogenic, a database of oncogenic annotations from the COSMIC catalogue of somatic mutations in cancer (cancer.sanger.ac.uk).

[01223] For analyzing whether the REPAIR constructs perturbed RNA levels, the transcript per million (TPM) values output from the RSEM analysis were used for expression counts and transformed to log-space by taking the log2(TPM+1). To find differentially regulated genes, a Student's t-test was performed on three targeting guide replicates versus three non-targeting guide replicates. The statistical analysis was only performed on genes with log2(TPM+1) values greater than 2.5 and genes were only considered differentially regulated if they had a fold change greater than 2 or less than 0.8. Genes were reported if they had a false discovery rate of less than 0.01.

Results

Comprehensive Characterization of Cas13 Family Members in Mammalian Cells

[01224] We previously developed LwaCas13a for mammalian knockdown applications, but it required an msfGFP stabilization domain for efficient knockdown and, although the specificity was high, knockdown efficiencies were not consistently below 50%(*15*). We sought to identify a more robust RNA-targeting CRISPR system by characterizing a genetically diverse set of Cas13 family members to assess their RNA knockdown activity in mammalian cells (Fig. 49A). We cloned 21 Cas13a, 15 Cas13b, and 7 Cas13c mammalian codon-optimized orthologs (Table 6) into an expression vector with N- and C-terminal

nuclear export signal (NES) sequences and a C-terminal msfGFP to enhance protein stability. To assay interference in mammalian cells, we designed a dual reporter construct expressing the orthogonal *Gaussia* (Gluc) and *Cypridinia* (Cluc) luciferases under separate promoters, which allows one luciferase to function as a measure of Cas13 interference activity and the other to serve as an internal control. For each ortholog, we designed PFS-compatible guide RNAs, using the Cas13b PFS motifs derived from an ampicillin interference assay (Fig 55; Table 7) and the 3' H PFS (not G) from previous reports of Cas13a activity(*10*).

[01225] We transfected HEK293FT cells with Cas13 expression, guide RNA and reporter plasmids and quantified levels of the targeted Gluc 48 hours later (Figure 49B, 69A). Testing two guide RNAs for each Cas13 ortholog revealed a range of activity levels, including five Cas13b orthologs with similar or increased interference across both guide RNAs relative to the recently characterized LwaCas13a (Figure 49B) , and we observed only a weak correlation between Cas13 expression and interference activity (Fig. 69B-D).. We selected these five Cas13b orthologs, as well as the top two Cas13a orthologs for further engineering.

[01226] We next tested for Cas13-mediated knockdown of Gluc without msfGFP, in order to select orthologs that do not require stabilization domains for robust activity. We hypothesized that, in addition to msfGFP, Cas13 activity could be affected by subcellular localization, as previously reported for optimization of LwaCas13a(*15*). Therefore, we tested the interference activity of the seven selected Cas13 orthologs C-terminally fused to one of six different localization tags without msfGFP. Using the luciferase reporter assay, we found that PspCas13b and PguCas13b C-terminally fused to the HIV Rev gene NES and RanCas13b C-terminally fused to the MAPK NES had the highest levels of interference activity (fig. 56A). To further distinguish activity levels of the top orthologs, we compared the three optimized Cas13b constructs to the optimal LwaCas13a-msfGFP fusion and shRNA for their ability to knockdown the *KRAS* transcript using position-matched guides (fig. 56B). We observed the highest levels interference for PspCas13b (average knockdown 62.9%) and thus selected this for further comparison to LwaCas13a.

[01227] To more rigorously define the activity level of PspCas13b and LwaCas13a we designed position matched guides tiling along both Gluc and Cluc and assayed their activity using our luciferase reporter assay. We tested 93 and 20 position matched guides targeting Gluc and Cluc, respectively, and found that PspCas13b had consistently increased levels of

knockdown relative to LwaCas13a (average of 92.3% for PspCas13b vs. 40.1% knockdown for LwaCas13a) (Fig. 49C,D).

Specificity of Cas13 mammalian interference activity

[01228] To characterize the interference specificities of PspCas13b and LwaCas13a we designed a plasmid library of luciferase targets containing single mismatches and double mismatches throughout the target sequence and the three flanking 5' and 3' base pairs (fig. 56C). We transfected HEK293FT cells with either LwaCas13a or PspCas13b, a fixed guide RNA targeting the unmodified target sequence, and the mismatched target library corresponding to the appropriate system. We then performed targeted RNA sequencing of uncleaved transcripts to quantify depletion of mismatched target sequences. We found that LwaCas13a and PspCas13b had a central region that was relatively intolerant to single mismatches, extending from base pairs 12-26 for the PspCas13b target and 13-24 for the LwaCas13a target (fig. 56D). Double mismatches were even less tolerated than single mutations, with little knockdown activity observed over a larger window, extending from base pairs 12-29 for PspCas13b and 8-27 for LwaCas13a in their respective targets (fig. 56E). Additionally, because there are mismatches included in the three nucleotides flanking the 5' and 3' ends of the target sequence, we could assess PFS constraints on Cas13 knockdown activity. Sequencing showed that almost all PFS combinations allowed robust knockdown, indicating that a PFS constraint for interference in mammalian cells likely does not exist for either enzyme tested. These results indicate that Cas13a and Cas13b display similar sequence constraints and sensitivities against mismatches.

[01229] We next characterized the interference specificity of PspCas13b and LwaCas13a across the mRNA fraction of the transcriptome. We performed transcriptome-wide mRNA sequencing to detect significant differentially expressed genes. LwaCas13a and PspCas13b demonstrated robust knockdown of Gluc (Fig. 49E,F) and were highly specific compared to a position-matched shRNA, which showed hundreds of off-targets (Fig. 49G), consistent with our previous characterization of LwaCas13a specificity in mammalian cells (*15*).

Cas13-ADAR fusions enable targeted RNA editing

[01230] Given that PspCas13b achieved consistent, robust, and specific knockdown of mRNA in mammalian cells, we envisioned that it could be adapted as an RNA binding platform to recruit the deaminase domain of ADARs ($ADAR_{DD}$) for programmable RNA editing. To engineer a PspCas13b lacking nuclease activity (dPspCas13b, referred to as

dCas13b from here), we mutated conserved catalytic residues in the HEPN domains and observed loss of luciferase RNA knockdown activity (fig. 57A). We hypothesized that a dCas13b-ADAR$_{DD}$ fusion could be recruited by a guide RNA to target adenosines, with the hybridized RNA creating the required duplex substrate for ADAR activity (Fig. 50A). To enhance target adenosine deamination rates we introduced two additional modifications to our initial RNA editing design: we introduced a mismatched cytidine opposite the target adenosine, which has been previously reported to increase deamination frequency, and fused dCas13b with the deaminase domains of human ADAR1 or ADAR2 containing hyperactivating mutations to enhance catalytic activity (ADAR1$_{DD}$(E1008Q)(27) or ADAR2$_{DD}$(E488Q)(21)).

[01231] To test the activity of dCas13b-ADAR$_{DD}$ we generated an RNA-editing reporter on Cluc by introducing a nonsense mutation (W85X (UGG->UAG)), which could functionally be repaired to the wildtype codon through A->I editing (Fig. 50B) and then be detected as restoration of Cluc luminescence. We evenly tiled guides with spacers 30, 50, 70 or 84 nucleotides in length across the target adenosine to determine the optimal guide placement and design (Fig. 50C). We found that dCas13b-ADAR1$_{DD}$ required longer guides to repair the Cluc reporter, while dCas13b-ADAR2$_{DD}$ was functional with all guide lengths tested (Fig. 50C). We also found that the hyperactive E488Q mutation improved editing efficiency, as luciferase restoration with the wildtype ADAR2$_{DD}$ was reduced (fig. 57B). From this demonstration of activity, we chose dCas13b-ADAR2$_{DD}$(E488Q) for further characterization and designated this approach as RNA Editing for Programmable A to I Replacement version 1 (REPAIRv1).

[01232] To validate that restoration of luciferase activity was due to *bona fide* editing events, we measured editing of Cluc transcripts subject to REPAIRv1 directly via reverse transcription and targeted next-generation sequencing. We tested 30- and 50-nt spacers around the target site and found that both guide lengths resulted in the expected A to I edit, with 50-nt spacers achieving higher editing percentages (Fig. 50D,E, fig. 57C). We also observed that 50-nt spacers had an increased propensity for editing at non-targeted adenosines, likely due to increased regions of duplex RNA (Fig. 50E, Fig. 57C).

[01233] We next targeted an endogenous gene, *PPIB*. We designed 50-nt spacers tiling *PPIB* and found that we could edit the *PPIB* transcript with up to 28% editing efficiency (Fig. 57D). To test if REPAIR could be further optimized, we modified the linker between dCas13b and ADAR2$_{DD}$(E488Q) (fig. 57E, Table 8) and found that linker choice

modestly affected luciferase activity restoration. Additionally, we tested the ability of dCas13b and guide alone to mediate editing events, finding that the ADAR deaminase domain is required for editing (Fig. 70A-D).

Defining the sequence parameters for RNA editing

[01234] Given that we could achieve precise RNA editing at a test site, we wanted to characterize the sequence constraints for programming the system against any RNA target in the transcriptome. Sequence constraints could arise from dCas13b targeting limitations, such as the PFS, or from ADAR sequence preferences(*26*). To investigate PFS constraints on REPAIRv1, we designed a plasmid library carrying a series of four randomized nucleotides at the 5' end of a target site on the Cluc transcript (Fig. 51A). We targeted the center adenosine within either a UAG or AAC motif and found that for both motifs, all PFSs demonstrated detectable levels of RNA editing, with a majority of the PFSs having greater than 50% editing at the target site (Fig. 51B). Next, we sought to determine if the $ADAR2_{DD}$ in REPAIRv1 had any sequence constraints immediately flanking the targeted base, as has been reported previously for $ADAR2_{DD}$(*26*).We tested every possible combination of 5' and 3' flanking nucleotides directly surrounding the target adenosine (Fig. 51C), and found that REPAIRv1 was capable of editing all motifs (Fig. 51D). Lastly, we analyzed whether the identity of the base opposite the target A in the spacer sequence affected editing efficiency and found that an A-C mismatch had the highest luciferase restoration with A-G, A-U, and A-A having drastically reduced REPAIRv1 activity (fig. 57F, 70E).

Correction of disease-relevant human mutations using REPAIRv1

[01235] To demonstrate the broad applicability of the REPAIRv1 system for RNA editing in mammalian cells, we designed REPAIRv1 guides against two disease relevant mutations: 878G>A (*AVPR2* W293X) in X-linked Nephrogenic diabetes insipidus and 1517G>A (*FANCC* W506X) in Fanconi anemia. We transfected expression constructs for cDNA of genes carrying these mutations into HEK293FT cells and tested whether REPAIRv1 could correct the mutations. Using guide RNAs containing 50-nt spacers, we were able to achieve 35% correction of *AVPR2* and 23% correction of *FANCC* (Fig. 52A-D). We then tested the ability of REPAIRv1 to correct 34 different disease-relevant G>A mutations (Table 9) and found that we were able to achieve significant editing at 33 sites with up to 28% editing efficiency (Fig. 52E). The mutations we chose are only a fraction of the pathogenic G to A mutations (5,739) in the ClinVar database, which also includes an additional 11,943 G to A variants (Fig. 52F and fig. 58). Because there are no sequence

constraints, REPAIRv1 is capable of potentially editing all these disease relevant mutations, especially given that we observed significant editing regardless of the target motif (Fig. 51C and Fig. 52G).

[01236] Delivering the REPAIRv1 system to diseased cells is a prerequisite for therapeutic use, and we therefore sought to design REPAIRv1 constructs that could be packaged into therapeutically relevant viral vectors, such as adeno-associated viral (AAV) vectors. AAV vectors have a packaging limit of 4.7kb, which cannot accommodate the large size of dCas13b-$ADAR_{DD}$ (4473 bp) along with promoter and expression regulatory elements. To reduce the size, we tested a variety of N-terminal and C-terminal truncations of dCas13 fused to $ADAR2_{DD}$(E488Q) for RNA editing activity. We found that all C-terminal truncations tested were still functional and able to restore luciferase signal (fig. 59), and the largest truncation, C-terminal Δ984-1090 (total size of the fusion protein 4,152bp) was small enough to fit within the packaging limit of AAV vectors.

Transcriptome-wide specificity of REPAIRv1

[01237] Although RNA knockdown with PspCas13b was highly specific, in our luciferase tiling experiments, we observed off-target adenosine editing within the guide:target duplex (Fig. 50E). To see if this was a widespread phenomenon, we tiled an endogenous transcript, *KRAS*, and measured the degree of off-target editing near the target adenosine (Fig. 53A). We found that for *KRAS*, while the on-target editing rate was 23%, there were many sites around the target site that also had detectable A to G edits (Fig. 53B).

[01238] Because of the observed off-target editing within the guide:target duplex, we evaluated all possible transcriptome off-targets by performing RNA sequencing on all mRNAs. RNA sequencing revealed that there was a significant number A to G off-target events, with 1,732 off-targets in the targeting condition and 925 off-targets in the non-targeting condition, with 828 off-targets overlapping (Fig. 53C,D). Of all the editing sites across the transcriptome, the on-target editing site had the highest editing rate, with 89% A to G conversion.

[01239] Given the high specificity of Cas13 targeting, we reasoned that the off-targets may arise from ADAR. We repeated the Cluc targeting experiment, this time comparing transcriptome changes for REPAIRv1 with a targeting guide, REPAIRv1 with a non-targeting guide, REPAIRv1 alone, or ADARDD(E488Q) alone (fig. 71). We found differentially expressed genes and off-target editing events in each condition (fig. 71,C). Interestingly, there was a high degree of overlap in the off-target editing events between

ADARDD(E488Q) and all REPAIRv1 off-target edits, supporting the hypothesis that REPAIR off-target edits are driven by dCas13b-independent ADARDD(E488Q) editing events (fig. 71).

[01240] Two RNA-guided ADAR systems have been described previously (fig. 60A). The first utilizes a fusion of $ADAR2_{DD}$ to the small viral protein lambda N (ʎN), which binds to the BoxB-ʎ RNA hairpin(*22*). A guide RNA with double BoxB-ʎ hairpins guides $ADAR2_{DD}$ to edit sites encoded in the guide RNA(*23*). The second design utilizes full length ADAR2 (ADAR2) and a guide RNA with a hairpin that the double strand RNA binding domains (dsRBDs) of ADAR2 recognize(*21, 24*). We analyzed the editing efficiency of these two systems compared to REPAIRv1 and found that the BoxB-ADAR2 and ADAR2 systems demonstrated 63% and 36% editing rates, respectively, compared to the 89% editing rate achieved by REPAIRv1 (fig. 60B-E). Additionally, the BoxB and ADAR2 systems created 2018 and 174 observed off targets, respectively, in the targeting guide conditions, compared to the 1,229 off targets in the REPAIRv1 targeting guide condition. Notably, all the conditions with the two $ADAR2_{DD}$-based systems (REPAIRv1 and BoxB) showed a high percentage of overlap in their off-targets while the ADAR2 system had a largely distinct set of off-targets (fig. 60F). The overlap in off-targets between the targeting and non-targeting conditions and between REPAIRv1 and BoxB conditions suggest $ADAR2_{DD}$ drove off-targets independent of dCas13 targeting (fig. 60F).

Improving specificity of REPAIRv1 through rational protein engineering

[01241] To improve the specificity of REPAIR, we employed structure-guided protein engineering of $ADAR2_{DD}$(E488Q). Because of the guide-independent nature of off-targets, we hypothesized that destabilizing $ADAR2_{DD}$(E488Q)-RNA binding would selectively decrease off-target editing, but maintain on-target editing due to increased local concentration from dCas13b tethering of $ADAR2_{DD}$(E488Q) to the target site. We mutagenized $ADAR2_{DD}$(E488Q) residues previously determined to contact the duplex region of the target RNA (Fig. 54A)(*18*) on the $ADAR2_{DD}$(E488Q) background. To assess efficiency and specificity, we tested 17 single mutants with both targeting and non-targeting guides, under the assumption that background luciferase restoration in the non-targeting condition detected would be indicative of broader off-target activity. We found that mutations at the selected residues had significant effects on the luciferase activity for targeting and non-targeting guides (Fig. 54A,B, fig. 61A). A majority of mutants either significantly improved the luciferase activity for the targeting guide or increased the ratio of

targeting to non-targeting guide activity, which we termed the specificity score (Fig. 54A,B). We selected a subset of these mutants (Fig. 54B) for transcriptome-wide specificity profiling by next generation sequencing. As expected, off-targets measured from transcriptome-wide sequencing correlated with our specificity score (fig. 61B) for mutants. We found that with the exception of $ADAR2_{DD}$(E488Q/R455E), all sequenced REPAIRv1 mutants could effectively edit the reporter transcript (Fig. 54C), with many mutants showing reduction in the number of off-targets (fig 61C, 62). We further explored the surrounding motifs of off-targets for specificity mutants, and found that REPAIRv1 and most of the engineered mutants exhibited a strong 3' G preference for their edits, in agreement with the characterized ADAR2 motif (fig. 63A)(*28*). We selected the mutant $ADAR2_{DD}$(E488Q/T375G) for future experiments, as it had the highest percent editing of the four mutants with the lowest numbers of transcriptome-wide off targets and termed it REPAIRv2. Compared to REPAIRv1, REPAIRv2 exhibited increased specificity, with a reduction from 18,385 to 20 transcriptome-wide off-targets by high-coverage sequencing (125X coverage, 10ng DNA transfection) (Fig. 54D). In the region surrounding the targeted adenosine in Cluc, REPAIRv2 had reduced off-target editing, visible in sequencing traces (Fig. 54E). In motifs derived from next-generation sequencing, REPAIRv1 presented a strong preference towards 3' G, but showed off-targeting edits for all motifs (fig. 63B); by contrast, REPAIRv2 only edited the strongest off-target motifs (fig. 63C). The distribution of edits on transcripts was heavily skewed, with highly-edited genes having over 60 edits (fig. 64A,B), whereas REPAIRv2 only edited one transcript (*EEF1A1*) multiple times (fig. 64D-F). REPAIRv1 off-target edits were predicted to result in numerous variants, including 1000 missense mutations (fig. 64C) with 93 oncogenic events (fig. 64D). In contrast, REPAIRv2 only had 6 missense mutations (fig. 64E), none of which had oncogenic consequences (fig. 64F). This reduction in predicted off-target effects distinguishes REPAIRv2 from other RNA editing approaches. Experiments with different dosages of guide RNA suggests that dose response may reduce off target activity (Fig. 68).

[01242] Analysis of the sequence surrounding off-target edits for REPAIRv1 or v2 did not reveal homology to guide sequences, suggesting that off-targets are likely dCas13b-independent (fig. 72), consistent with the high overlap of off-targets between REPAIRv1 and the ADAR deaminase domain (fig. 71D). To directly compare REPAIRv2 against other programmable ADAR systems, we repeated our Cluc targeting experiments with all systems at two different dosages of ADAR vector, finding that REPAIRv2 had comparable on-target

editing to BoxB and ADAR2 but with significantly fewer off-target editing events at both dosages (fig 73). REPAIRv2 had enhanced specificity compared to REPAIRv1 at both dosages (fig. 73B), a finding that also extended to two guides targeting distinct sites on PPIB (fig. 74A-D). It is also worth noting that, in general, the lower dosage condition (10 ng) had fewer off-targets than the higher dosage condition (150 ng) (fig. 70).

[01243] To assess editing specificity with greater sensitivity, we sequenced the low dosage condition (10 ng of transfected DNA) of REPAIRv1 and v2 at significantly higher sequencing depth (125X coverage of the transcriptome). Increased numbers of off-targets were found at higher sequencing depths corresponding to detection of rarer off-target events (fig. 75). Furthermore, we speculated that different transcriptome states could also potentially alter the number of off-targeting events. Therefore, we tested REPAIRv2 activity in the osteosarcoma U2OS cell line, observing 6 and 7 off-targets for the targeting and non-targeting guide, respectively (fig. 76).

[01244] Applicant targeted REPAIRv2 to endogenous genes to test if the specificity-enhancing mutations reduced nearby edits in target transcripts while maintaining high-efficiency on-target editing. For guides targeting either *KRAS* or *PPIB*, Applicant found that REPAIRv2 had no detectable off-target edits and could effectively edit the on-target adenosine at 27.1% and 13%, respectively (Fig. 54F,G). This specificity extended to additional target sites, including regions that demonstrate high-levels of background in non-targeting conditions for REPAIRv1, such as other *KRAS* or *PPIB* target sites (Fig. 65). Overall, REPAIRv2 eliminated off-targets in duplexed regions around the edited adenosine and showed dramatically enhanced transcriptome-wide specificity.

Conclusion

[01245] Applicant has shown here that the RNA-guided RNA-targeting type VI-B effector Cas13b is capable of highly efficient and specific RNA knockdown, providing the basis for improved tools for interrogating essential genes and non-coding RNA as well as controlling cellular processes at the transcriptomic level. Catalytically inactive Cas13b (dCas13b) retains programmable RNA binding capability, which we leveraged here by fusing dCas13b to the adenosine deaminase ADAR2 to achieve precise A to I edits, a system we term REPAIRv1 (RNA Editing for Programmable A to I Replacement version 1). Further engineering of the system produced REPAIRv2, a method with comparable or increased activity relative to current editing platforms with dramatically improved specificity than

previously described RNA editing platforms (*25, 29*) while maintaining high levels of on-target efficacy.

[01246] Although Cas13b exhibits high fidelity, Applicant's initial results with dCas13b-ADAR2$_{DD}$ fusions revealed thousands of off-targets. To address this, Applicant employed a rational mutagenesis strategy to vary the ADAR2$_{DD}$ residues that contact the RNA duplex, identifying a variant, ADAR2$_{DD}$(E488Q/T375G), capable of precise, efficient, and highly specific editing when fused to dCas13b. Editing efficiency with this variant was comparable to or better than that achieved with two currently available systems, BoxB-ADAR$_{DD}$ or ADAR2 editing. Moreover, the REPAIRv2 system created only 10 observable off-targets in the whole transcriptome, at least an order of magnitude better than both alternative editing technologies. While it is possible that ADAR could deaminate adenosine bases on the DNA strand in RNA-DNA heteroduplexes (*20*), it is unlikely to do so in this case as Cas13b does not bind DNA efficiently and that REPAIR is cytoplasmically localized. Additionally, the lack of homology of off-target sites to the guide sequence and the strong overlap of off-targets with the ADAR$_{DD}$(E488Q)-only condition suggest that off-targets are not mediated by off-target guide binding. Deeper sequencing and novel inosine enrichment methods could further refine our understanding of REPAIR specificity in the future.

[01247] The REPAIR system offers many advantages compared to other nucleic acid editing tools. First, the exact target site can be encoded in the guide by placing a cytidine within the guide extension across from the desired adenosine to create a favorable A-C mismatch ideal for ADAR editing activity. Second, Cas13 has no targeting sequence constraints, such as a PFS or PAM, and no motif preference surrounding the target adenosine, allowing any adenosine in the transcriptome to be potentially targeted with the REPAIR system. We do note, however, that DNA base editors can target either the sense or anti-sense strand, while the REPAIR system is limited to transcribed sequences, thereby constraining the total number of possible editing sites we could target. However, due to the more flexible nature of targeting with REPAIR, this system can effect more edits within ClinVar (Fig. 52C) than Cas9-DNA base editors. Third, the REPAIR system directly deaminates target adenosines to inosines and does not rely on endogenous repair pathways, such as base-excision or mismatch repair, to generate desired editing outcomes. Thus, REPAIR should be possible in non-dividing cells that cannot support other forms of editing, such as in post-mitotic cells, such as in neurons. Fourth, RNA editing can be transient, allowing the potential for temporal control over editing outcomes. This property will likely

be useful for treating diseases caused by temporary changes in cell state, such as local inflammation.

[01248] The REPAIR system provides multiple opportunities for additional engineering. Cas13b possesses pre-crRNA processing activity(*13*), allowing for multiplex editing of multiple variants, which alone might not alter disease risk, but together might have additive effects and disease-modifying potential. Extension of our rational design approach, such as combining promising mutations, could further increase the specificity and efficiency of the system, while unbiased screening approaches could identify additional residues for improving REPAIR activity and specificity.

[01249] Currently, the base conversions achievable by REPAIR are limited to generating inosine from adenosine; additional fusions of dCas13 with other catalytic RNA editing domains, such as APOBEC, could enable cytidine to uridine editing. Additionally, mutagenesis of ADAR could relax the substrate preference to target cytidine, allowing for the enhanced specificity conferred by the duplexed RNA substrate requirement to be exploited by C->U editors. Adenosine to inosine editing on DNA substrates may also be possible with catalytically inactive DNA-targeting CRISPR effectors, such as dCas9 or dCpf1, either through formation of DNA-RNA heteroduplex targets(*20*) or mutagenesis of the ADAR domain.

[01250] REPAIR could be applied towards a range of therapeutic indications where A to I (A to G) editing can reverse or slow disease progression (fig. 66). First, expression of REPAIR for targeting causal, Mendelian G to A mutations in disease-relevant tissues could be used to revert deleterious mutations and treat disease. For example, stable REPAIR expression via AAV in brain tissue could be used to correct the *GRIN2A* missense mutation c.2191G>A (Asp731Asn) that causes focal epilepsy(*28*) or the *APP* missense mutation c.2149G>A (Val717Ile) causing early-onset Alzheimer's disease(*29*). Second, REPAIR could be used to treat disease by modifying the function of proteins involved in disease-related signal transduction. For instance, REPAIR editing would allow the re-coding of some serine, threonine and tyrosine residues that are the targets of kinases (fig. 66). Phosphorylation of these residues in disease-relevant proteins affects disease progression for many disorders including Alzheimer's disease and multiple neurodegenerative conditions(*30*). Third, REPAIR could be used to change the sequence of expressed, risk-modifying G to A variants to pre-emptively decrease the chance of entering a disease state for patients. The most intriguing case are the 'protective' risk-modifying alleles, which

dramatically decrease the chance of entering a disease state, and in some cases, confer additional health benefits. For instance, REPAIR could be used to functionally mimic A to G alleles of *PCSK9* and *IFIH1* that protect against cardiovascular disease and psoriatic arthritis(*31, 39*), respectively. Last, REPAIR can be used to therapeutically modify splice acceptor and donor sites for exon modulation therapies. REPAIR can change AU to IU or AA to AI, the functional equivalent of the consensus 5' splice donor or 3' splice acceptor sites respectively, creating new splice junctions. Additionally, REPAIR editing can mutate the consensus 3' splice acceptor site from AG->IG to promote skipping of the adjacent downstream exon, a therapeutic strategy that has received significant interest for the treatment of DMD. Modulation of splice sites could have broad applications in diseases where anti-sense oligos have had some success, such as for modulation of SMN2 splicing for treatment of spinal muscular atrophy(*32*).

[01251] We have demonstrated the use of the PspCas13b enzyme as both an RNA knockdown and RNA editing tool. The dCas13b platform for programmable RNA binding has many applications, including live transcript imaging, splicing modification, targeted localization of transcripts, pull down of RNA-binding proteins, and epitranscriptomic modifications. Here, we used dCas13 to create REPAIR, adding to the existing suite of nucleic acid editing technologies. REPAIR provides a new approach for treating genetic disease or mimicking protective alleles, and establishes RNA editing as a useful tool for modifying genetic function.

Table 6 Cas13 Orthologs used in this study

Cas13 ID	Cas13 abbreviation	Host Organism	Protein Accession
Cas13a1	LshCas13a	Leptotrichia shahii	WP_018451595.1
Cas13a2	LwaCas13a	Leptotrichia wadei (Lw2)	WP_021746774.1
Cas13a3	LseCas13a	Listeria seeligeri	WP_012985477.1
Cas13a4	LbmCas13a	Lachnospiraceae bacterium MA2020	WP_044921188.1
Cas13a5	LbnCas13a	Lachnospiraceae bacterium NK4A179	WP_022785443.1
Cas13a6	CamCas13a	[Clostridium] aminophilum DSM 10710	WP_031473346.1
Cas13a7	CgaCas13a	Carnobacterium gallinarum DSM 4847	WP_034560163.1
Cas13a8	Cga2Cas13a	Carnobacterium gallinarum DSM 4847	WP_034563842.1
Cas13a9	Pprcas13a	Paludibacter propionicigenes WB4	WP_013443710.1

Cas13a10	LweCas13a	Listeria weihenstephanensis FSL R9-0317	WP_036059185.1
Cas13a11	LbfCas13a	Listeriaceae bacterium FSL M6-0635	WP_036091002.1
Cas13a12	Lwa2cas13a	Leptotrichia wadei F0279	WP_021746774.1
Cas13a13	RcsCas13a	Rhodobacter capsulatus SB 1003	WP_013067728.1
Cas13a14	RcrCas13a	Rhodobacter capsulatus R121	WP_023911507.1
Cas13a15	RcdCas13a	Rhodobacter capsulatus DE442	WP_023911507.1
Cas13a16	LbuCas13a	Leptotrichia buccalis C-1013-b	WP_015770004.1
Cas13a17	HheCas13a	Herbinix hemicellulosilytica	CRZ35554.1
Cas13a18	EreCas13a	[Eubacterium] rectale	WP_055061018.1
Cas13a19	EbaCas13a	Eubacteriaceae bacterium CHKCI004	WP_090127496.1
Cas13a20	BmaCas13a	Blautia sp. Marseille-P2398	WP_062808098.1
Cas13a21	LspCas13a	Leptotrichia sp. oral taxon 879 str. F0557	WP_021744063.1
Cas13b1	BzoCas13b	Bergeyella zoohelcum	WP_002664492
Cas13b2	PinCas13b	Prevotella intermedia	WP_036860899
Cas13b3	PbuCas13b	Prevotella buccae	WP_004343973
Cas13b4	AspCas13b	Alistipes sp. ZOR0009	WP_047447901
Cas13b5	PsmCas13b	Prevotella sp. MA2016	WP_036929175
Cas13b6	RanCas13b	Riemerella anatipestifer	WP_004919755
Cas13b7	PauCas13b	Prevotella aurantiaca	WP_025000926
Cas13b8	PsaCas13b	Prevotella saccharolytica	WP_051522484
Cas13b9	Pin2Cas13b	Prevotella intermedia	WP_061868553
Cas13b10	CcaCas13b	Capnocytophaga canimorsus	WP_013997271
Cas13b11	PguCas13b	Porphyromonas gulae	WP_039434803
Cas13b12	PspCas13b	Prevotella sp. P5-125	WP_044065294
Cas13b13	FbrCas13b	Flavobacterium branchiophilum	WP_014084666
Cas13b14	PgiCas13b	Porphyromonas gingivalis	WP_053444417
Cas13b15	Pin3Cas13b	Prevotella intermedia	WP_050955369
Cas13c1	FnsCas13c	Fusobacterium necrophorum subsp. funduliforme ATCC 51357 contig00003	WP_005959231.1
Cas13c2	FndCas13c	Fusobacterium necrophorum DJ-2 contig0065, whole genome shotgun sequence	WP_035906563.1
Cas13c3	FnbCas13c	Fusobacterium necrophorum BFTR-1 contig0068	WP_035935671.1
Cas13c4	FnfCas13c	Fusobacterium necrophorum subsp. funduliforme 1_1_36S cont1.14	EHO19081.1

Cas13c5	FpeCas13c	Fusobacterium perfoetens ATCC 29250 T364DRAFT_scaffold00009.9_C	WP_027128616.1
Cas13c6	FulCas13c	Fusobacterium ulcerans ATCC 49185 cont2.38	WP_040490876.1
Cas13c7	AspCas13c	Anaerosalibacter sp. ND1 genome assembly Anaerosalibacter massiliensis ND1	WP_042678931.1

Table 7 PFS cutoffs in bacterial screens

Cas13b ortholog	Key	-Log_2 depletion score used to generate PFS motif
Bergeyella zoohelcum	1	2
Prevotella intermedia locus 1	2	1
Prevotella buccae	3	3
Alistipes sp. ZOR0009	4	1
Prevotella sp. MA2016	5	2
Riemerella anatipestifer	6	4
Prevotella aurantiaca	7	1
Prevotella saccharolytica	8	0
Prevotella intermedia locus 2	9	0
Capnocytophaga canimorsus	10	3
Porphyromonas gulae	11	4
Prevotella sp. P5-125	12	2.1
Flavobacterium branchiophilum	13	1
Porphyromonas gingivalis	14	3
Prevotella intermedia locus 2	15	4

Table 8 dCas13b-ADAR linker sequences used in this study for RNA editing in mammalian cells.

Figure	linker
50C	GSGGGGS
50E	GS
57B	GSGGGGS
57C	GS
57D	GS
57E: GS	GS
57E: GSGGGGS	GSGGGGS

57E: (GGGS)3	GGGGSGGGGSGGGGS
57E: Rigid	EAAAK
57E: (GGS)6	GGSGGSGGSGGSGGSGGS
57E: XTEN	SGSETPGTSESATPES
51B	GS
57F	GS
51C	GS
52B	GS
52D	GS
52E	GS
51A: Δ984-1090, Δ1026-1090, Δ1053-1090	GS
51A: Δ1-125, Δ1-88, Δ1-72	GSGGGGS
53B	GS
53C	GS
53D	GS
60A	GS
60C	GS
60D	GS
61D	GS
54A	GS
62A	GS
54B	GS
62B	GS
62C	GS
63A	GS
63B	GS
54C	GS
54D	GS
54E	GS
54F	GS
66A	GS
66A	GS

Table 9 Disease information for disease-relevant mutations

Full length candidates	**Gene**	**Disease**
NM_000054.4(AVPR2):c.878G>A (p.Trp293Ter)	AVPR2	Nephrogenic diabetes insipidus, X-linked
NM_000136.2(FANCC):c.1517G>A (p.Trp506Ter)	FANCC	Fanconi anemia, complementation group C
Additional simulated candiates		
Candidate	**Gene**	**Disease**

NM_000206.2(IL2RG):c.710G>A (p.Trp237Ter)	IL2RG	X-linked severe combined immunodeficiency
NM_000132.3(F8):c.3144G>A (p.Trp1048Ter)	F8	Hereditary factor VIII deficiency disease
NM_000527.4(LDLR):c.1449G>A (p.Trp483Ter)	LDLR	Familial hypercholesterolemia
NM_000071.2(CBS):c.162G>A (p.Trp54Ter)	CBS	Homocystinuria due to CBS deficiency
NM_000518.4(HBB):c.114G>A (p.Trp38Ter)	HBB	beta^0^ Thalassemia|beta Thalassemia
NM_000035.3(ALDOB):c.888G>A (p.Trp296Ter)	ALDOB	Hereditary fructosuria
NM_004006.2(DMD):c.3747G>A (p.Trp1249Ter)	DMD	Duchenne muscular dystrophy
NM_005359.5(SMAD4):c.906G>A (p.Trp302Ter)	SMAD4	Juvenile polyposis syndrome
NM_000059.3(BRCA2):c.582G>A (p.Trp194Ter)	BRCA2	Familial cancer of breast|Breast-ovarian cancer, familial 2
NM_000833.4(GRIN2A):c.3813G>A (p.Trp1271Ter)	GRIN2A	Epilepsy, focal, with speech disorder and with or without mental retardation
NM_002977.3(SCN9A):c.2691G>A (p.Trp897Ter)	SCN9A	Indifference to pain, congenital, autosomal recessive
NM_007375.3(TARDBP):c.943G>A (p.Ala315Thr)	TARDBP	Amyotrophic lateral sclerosis type 10
NM_000492.3(CFTR):c.3846G>A (p.Trp1282Ter)	CFTR	Cystic fibrosis|Hereditary pancreatitis|not provided|ataluren response - Efficacy
NM_130838.1(UBE3A):c.2304G>A (p.Trp768Ter)	UBE3A	Angelman syndrome
NM_000543.4(SMPD1):c.168G>A (p.Trp56Ter)	SMPD1	Niemann-Pick disease, type A
NM_206933.2(USH2A):c.9390G>A (p.Trp3130Ter)	USH2A	Usher syndrome, type 2A
NM_130799.2(MEN1):c.1269G>A (p.Trp423Ter)	MEN1	Hereditary cancer-predisposing syndrome
NM_177965.3(C8orf37):c.555G>A (p.Trp185Ter)	C8orf37	Retinitis pigmentosa 64
NM_000249.3(MLH1):c.1998G>A (p.Trp666Ter)	MLH1	Lynch syndrome
NM_000548.4(TSC2):c.2108G>A (p.Trp703Ter)	TSC2	Tuberous sclerosis 2|Tuberous sclerosis syndrome
NM_000267.3(NF1):c.7044G>A (p.Trp2348Ter)	NF1	Neurofibromatosis, type 1

NM_000179.2(MSH6):c.3020G>A (p.Trp1007Ter)	MSH6	Lynch syndrome
NM_000344.3(SMN1):c.305G>A (p.Trp102Ter)	SMN1	Spinal muscular atrophy, type II|Kugelberg-Welander disease
NM_024577.3(SH3TC2):c.920G>A (p.Trp307Ter)	SH3TC2	Charcot-Marie-Tooth disease, type 4C
NM_001369.2(DNAH5):c.8465G>A (p.Trp2822Ter)	DNAH5	Primary ciliary dyskinesia
NM_004992.3(MECP2):c.311G>A (p.Trp104Ter)	MECP2	Rett syndrome
NM_032119.3(ADGRV1):c.7406G>A (p.Trp2469Ter)	ADGRV1	Usher syndrome, type 2C
NM_017651.4(AHI1):c.2174G>A (p.Trp725Ter)	AHI1	Joubert syndrome 3
NM_004562.2(PRKN):c.1358G>A (p.Trp453Ter)	PRKN	Parkinson disease 2
NM_000090.3(COL3A1):c.3833G>A (p.Trp1278Ter)	COL3A1	Ehlers-Danlos syndrome, type 4
NM_007294.3(BRCA1):c.5511G>A (p.Trp1837Ter)	BRCA1	Familial cancer of breast|Breast-ovarian cancer, familial 1
NM_000256.3(MYBPC3):c.3293G>A (p.Trp1098Ter)	MYBPC3	Primary familial hypertrophic cardiomyopathy
NM_000038.5(APC):c.1262G>A (p.Trp421Ter)	APC	Familial adenomatous polyposis 1
NM_001204.6(BMPR2):c.893G>A (p.W298*)	BMPR2	Primary pulmonary hypertension

Table 10: Key plasmids used in this study

Plasmid name	**Description**
pAB0006	CMV-Cluciferase-polyA EF1a-G-luciferase-polyA
pAB0040	CMV-Cluciferase(STOP85)-polyA EF1a-G-luciferase-polyA
pAB0048	pCDNA-ADAR2-N-terminal B12-HIV NES
pAB0050	pAB0001-A02 (crRNA backbone)
pAB0051	pAB0001-B06 (crRNA backbone)
pAB0052	pAB0001-B11 (crRNA backbone)
pAB0053	pAB0001-B12 (crRNA backbone)
pAB0014.B6	EF1a-BsiWI-Cas13b6-NES-mapk
pAB0013.B11	EF1a-BsiWI-Cas13b11-NES-HIV
pAB0013.B12	EF1a-BsiWI-Cas13b12-NES-HIV
pAB0051	pAB0001-B06 (crRNA backbone)
pAB0052	pAB0001-B11 (crRNA backbone)
pAB0053	pAB0001-B12 (crRNA backbone)

pAB0079	pCDNA-ADAR1hu-EQmutant-N-terminal destination vector
pAB0085	pCDNA-ADAR2 (E488Q)hu-EQmutant-N-terminal destination vector
pAB0095	EF1a-BsiWI-Cas13-B12-NES-HIV, with double H HEPN mutant
pAB0114	pCDNA-wtADAR2hu-EQmutant-N-terminal destination vector
pAB0120	Luciferase ADAR guide optimal (guide 24 from wC0054)
pAB0122	pAB0001-B12 NT guide for ADAR experiments
pAB0151	pCDNA-ADAR2hu-EQmutant-N-terminal destination vector C-term delta 984-1090
pAB0180	T375G specificity mutant
pAB0181	T375G Cas13b C-term delta 984-1090
pAB0440	CMV-dCas13b6-HIVNES-GS-dADAR2

Table 11: Guide/shRNA sequences used in this study for knockdown in mammalian cells

Name	Spacer sequence	Interference Mechanism	Notes	First figure
Bacterial PFS guide	GCCAGCUUUCCGGGCAUUGGCUUCCAUC (SEQ ID No. 300)	Cas13b	Used for all orthologs	
Cas13a-Gluc guide 1	GCCAGCUUUCCGGGCAUUGGCUUCCAUC (SEQ ID No. 301)	Cas13a	Used for all Cas13a orthologs	Figure 49B
Cas13a-Gluc guide 2	ACCCAGGAAUCUCAGGAAUGUCGACGAU (SEQ ID No. 302)	Cas13a	Used for all Cas13a orthologs	Figure 49B
Cas13a-non-targeting guide (LacZ)	AGGGUUUUCCCAGUCACGACGUUGUAAA (SEQ ID No. 303)	Cas13a	Used for all Cas13a orthologs	Figure 49B
Cas13b-Gluc guide 1.1	GGGCAUUGGCUUCCAUCUCUUUGAGCACCU (SEQ ID No. 304)	Cas13b	Used for orthologs 1-3, 6, 7, 10, 11, 12, 14, 15	Figure 49B
Cas13b-Gluc guide 1.2	GUGCAGCCAGCUUUCCGGGCAUUGGCUUCC (SEQ ID No. 305)	Cas13b	Used for ortholog 4	Figure 49B
Cas13b-Gluc guide 1.3	GCAGCCAGCUUUCCGGGCAUUGGCUUCCAU (SEQ ID No. 306)	Cas13b	Used for ortholog 5	Figure 49B
Cas13b-Gluc guide 1.4	GGCUUCCAUCUCUUUGAGCACCUCCAGCGG (SEQ ID No. 307)	Cas13b	Used for ortholog 8, 9	Figure 49B

Cas13b-Gluc guide 1.5	GGAAUGUCGACGAUCGCCUCGCCUAUGCCG (SEQ ID No. 308)	Cas13b	Used for ortholog 13	Figure 49B
Cas13b-Gluc guide 2.1	GAAUGUCGACGAUCGCCUCGCCUAUGCCGC (SEQ ID No. 309)	Cas13b	Used for orthologs 1-3, 6, 7, 10, 11, 14, 15	Figure 49B
Cas13b-Gluc guide 2.2	GACCUGUGCGAUGAACUGCUCCAUGGGCUC (SEQ ID No. 310)	Cas13b	Used for ortholog 12	Figure 49B
Cas13b-Gluc guide 2.2	GUGUGGCAGCGUCCUGGGAUGAACUUCUUC (SEQ ID No. 311)	Cas13b	Used for ortholog 4	Figure 49B
Cas13b-Gluc guide 2.3	GUGGCAGCGUCCUGGGAUGAACUUCUUCAU (SEQ ID No. 312)	Cas13b	Used for ortholog 5	Figure 49B
Cas13b-Gluc guide 2.4	GCUUCUUGCCGGGCAACUUCCCGCGGUCAG (SEQ ID No. 313)	Cas13b	Used for ortholog 8, 9	Figure 49B
Cas13b-Gluc guide 2.6	GCAGGGUUUUCCCAGUCACGACGUUGUAAAA (SEQ ID No. 314)	Cas13b	Used for ortholog 13	Figure 49B
Cas13b-non targeting guide	GCAGGGUUUUCCCAGUCACGACGUUGUAAAA (SEQ ID No. 315)	Cas13b	Used for all orthologs	Figure 49B
Cas13a-Gluc guide-RNASeq	ACCCAGGAAUCUCAGGAAUGUCGACGAU (SEQ ID No. 316)	Cas13a		Figure 49E
shRNA-Gluc guide	CAGCUUUCCGGGCAUUGGCUU (SEQ ID No. 317)	shRNA		Figure 49F
Cas13b-Gluc guide-RNASeq	CCGCUGGAGGUGCUCAAAGAGAUGGAAGCC (SEQ ID No. 318)	Cas13b		Figure 49F
Cas13a-Gluc-guide-1	GCCAGCUUUCCGGGCAUUGGCUUCCAUC (SEQ ID No. 319)	Cas13a		Figure 56A
Cas13a-Gluc-guide-2	ACCCAGGAAUCUCAGGAAUGUCGACGAU (SEQ ID No. 320)	Cas13a		Figure 56A
Cas13b-Gluc-opt-guide-1	GGGCAUUGGCUUCCAUCUCUUUGAGCACCU (SEQ ID No. 321)	Cas13b		Figure 56A

Cas13b-Gluc-opt-guide-2	GAAUGUCGACGAUCGCCUCGCCUAUGCCGC (SEQ ID No. 322)	Cas13b		Figure 56A
Cas13a KRAS guide 1	CAAGGCACUCUUGCCUACGCCACCAGCU (SEQ ID No. 323)	Cas13a		Figure 56B
Cas13a KRAS guide 2	UCAUAUUCGUCCACAAAAUGAUUCUGAA (SEQ ID No. 324)	Cas13a		Figure 56B
Cas13a KRAS guide 3	AUUAUUUAUGGCAAAUACACAAAGAAAG (SEQ ID No. 325)	Cas13a		Figure 56B
Cas13a KRAS guide 4	GAAUAUCUUCAAAUGAUUUAGUAUUAUU (SEQ ID No. 326)	Cas13a		Figure 56B
Cas13a KRAS guide 5	ACCAUAGGUACAUCUUCAGAGUCCUUAA (SEQ ID No. 327)	Cas13a		Figure 56B
Cas13b KRAS guide 1	GUCAAGGCACUCUUGCCUACGCCACCAGCU (SEQ ID No. 328)	Cas13b		Figure 56B
Cas13b KRAS guide 2	GAUCAUAUUCGUCCACAAAAUGAUUCUGAA (SEQ ID No. 329)	Cas13b		Figure 56B
Cas13b KRAS guide 3	GUAUUAUUUAUGGCAAAUACACAAAGAAAG (SEQ ID No. 330)	Cas13b		Figure 56B
Cas13b KRAS guide 4	GUGAAUAUCUUCAAAUGAUUUAGUAUUAUU (SEQ ID No. 331)	Cas13b		Figure 56B
Cas13b KRAS guide 5	GGACCAUAGGUACAUCUUCAGAGUCCUUAA (SEQ ID No. 332)	Cas13b		Figure 56B
shRNA KRAS guide 1	aagagtgccttgacgatacagcCUCGAGgctgtatcgtcaaggcactctt (SEQ ID No. 333)	shRNA		Figure 56B
shRNA KRAS guide 2	aatcattttgtggacgaatatCUCGAGatattcgtccacaaaatgatt (SEQ ID No. 334)	shRNA		Figure 56B
shRNA KRAS guide 3	aaataatactaaatcatttgaCUCGAGtcaaatgatttagtattattt (SEQ ID No. 335)	shRNA		Figure 56B
shRNA KRAS guide 4	aataatactaaatcatttgaaCUCGAGttcaaatgatttagtattatt (SEQ ID No. 336)	shRNA		Figure 56B
shRNA KRAS guide 5	aaggactctgaagatgtacctCUCGAGaggtacatcttcagagtcctt (SEQ ID No. 337)	shRNA		Figure 56B

Table 12: Guide sequences used for Gluc knockdown

Name	Spacer sequence	Position	Notes	First figure
Gluc tiling guide 1	GAGAUCAGGGCAAACAGAACUUUGACUCCC (SEQ ID No. 338)	2	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 2	GGAUGCAGAUCAGGGCAAACAGAACUUUGA (SEQ ID No. 339)	7	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 3	GCACAGCGAUGCAGAUCAGGGCAAACAGAA (SEQ ID No. 340)	13	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 4	GCUCGGCCACAGCGAUGCAGAUCAGGGCAA (SEQ ID No. 341)	19	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 5	GGGGCUUGGCCUCGGCCACAGCGAUGCAGA (SEQ ID No. 342)	28	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 6	GUGGGCUUGGCCUCGGCCACAGCGAUGCAG (SEQ ID No. 343)	29	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 7	GUCUCGGUGGGCUUGGCCUCGGCCACAGCG (SEQ ID No. 344)	35	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 8	GUUCGUUGUUCUCGGUGGGCUUGGCCUCGG (SEQ ID No. 345)	43	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 9	GGAAGUCUUCGUUGUUCUCGGUGGGCUUGG (SEQ ID No. 346)	49	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 10	GAUGUUGAAGUCUUCGUUGUUCUCGGUGGG (SEQ ID No. 347)	54	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 11	GCGGCCACGAUGUUGAAGUCUUCGUUGUUC (SEQ ID No. 348)	62	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C

Gluc tiling guide 12	GUGGCCACGGCCAC GAUGUUGAAGUCUU CG (SEQ ID No. 349)	68	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 13	GGUUGCUGGCCACG GCCACGAUGUUGAA GU (SEQ ID No. 350)	73	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 14	GUCGCGAAGUUGCU GGCCACGGCCACGA UG (SEQ ID No. 351)	80	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 15	GCCGUGGUCGCGAA GUUGCUGGCCACGG CC (SEQ ID No. 352)	86	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 16	GCGAGAUCCGUGGU CGCGAAGUUGCUGG CC (SEQ ID No. 353)	92	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 17	GCAGCAUCGAGAUC CGUGGUCGCGAAGU UG (SEQ ID No. 354)	98	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 18	GGGUCAGCAUCGAG AUCCGUGGUCGCGA AG (SEQ ID No. 355)	101	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 19	GCUUCCCGCGGUCA GCAUCGAGAUCCGU GG (SEQ ID No. 356)	109	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 20	GGGGCAACUUCCCG CGGUCAGCAUCGAG AU (SEQ ID No. 357)	115	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 21	GUCUUGCCGGGCAA CUUCCCGCGGUCAG CA (SEQ ID No. 358)	122	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 22	GGCAGCUUCUUGCC GGGCAACUUCCCGC GG (SEQ ID No. 359)	128	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 23	GCCAGCGGCAGCUU CUUGCCGGGCAACU UC (SEQ ID No. 360)	134	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C

Gluc tiling guide 24	GCACCUCCAGCGGC AGCUUCUUGCCGGG CA (SEQ ID No. 361)	139	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 25	GCUUUGAGCACCUC CAGCGGCAGCUUCU UG (SEQ ID No. 362)	146	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 26	GCAUCUCUUUGAGC ACCUCCAGCGGCAG CU (SEQ ID No. 363)	151	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 27	GUCCAUCUCUUUGA GCACCUCCAGCGGC AG (SEQ ID No. 364)	153	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 28	GGGCAUUGGCUUCC AUCUCUUUGAGCAC CU (SEQ ID No. 365)	163	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 29	GUCCGGGCAUUGGC UUCCAUCUCUUUGA GC (SEQ ID No. 366)	167	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 30	GGCCAGCUUUCCGG GCAUUGGCUUCCAU CU (SEQ ID No. 367)	175	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 31	GGGUGCAGCCAGCU UUCCGGGCAUUGGC UU (SEQ ID No. 368)	181	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 32	GAGCCCCUGGUGCA GCCAGCUUUCCGGG CA (SEQ ID No. 369)	188	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 33	GAUCAGACAGCCCC UGGUGCAGCCAGCU UU (SEQ ID No. 370)	195	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 34	GGCAGAUCAGACAG CCCCUGGUGCAGCC AG (SEQ ID No. 371)	199	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 35	GACAGGCAGAUCAG ACAGCCCCUGGUGC AG (SEQ ID No. 372)	203	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C

Gluc tiling guide 36	GUGAUGUGGGACAGGCAGAUCAGACAGCCC (SEQ ID No. 373)	212	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 37	GACUUGAUGUGGGACAGGCAGAUCAGACAG (SEQ ID No. 374)	215	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 38	GGGGCGUGCACUUGAUGUGGGACAGGCAGA (SEQ ID No. 375)	223	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 39	GCUUCAUCUUGGGCGUGCACUUGAUGUGGG (SEQ ID No. 376)	232	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 40	GUGAACUUCUUCAUCUUGGGCGUGCACUUG (SEQ ID No. 377)	239	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 41	GGGAUGAACUUCUUCAUCUUGGGCGUGCAC (SEQ ID No. 378)	242	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 42	GUGGGAUGAACUUCUUCAUCUUGGGCGUGC (SEQ ID No. 379)	244	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 43	GGGCAGCGUCCUGGGAUGAACUUCUUCAUC (SEQ ID No. 380)	254	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 44	GGGUGUGGCAGCGUCCUGGGAUGAACUUCU (SEQ ID No. 381)	259	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 45	GUUCGUAGGUGUGGCAGCGUCCUGGGAUGA (SEQ ID No. 382)	265	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 46	GCGCCUUCGUAGGUGUGGCAGCGUCCUGGG (SEQ ID No. 383)	269	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 47	GUCUUUGUCGCCUUCGUAGGUGUGGCAGCG (SEQ ID No. 384)	276	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C

Gluc tiling guide 48	GCUUUGUCGCCUUC GUAGGUGUGGCAGC GU (SEQ ID No. 385)	275	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 49	GUGCCGCCCUGUGC GGACUCUUUGUCGC CU (SEQ ID No. 386)	293	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 50	GUAUGCCGCCCUGU GCGGACUCUUUGUC GC (SEQ ID No. 387)	295	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 51	GCCUCGCCUAUGCC GCCCUGUGCGGACU CU (SEQ ID No. 388)	302	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 52	GGAUCGCCUCGCCU AUGCCGCCCUGUGC GG (SEQ ID No. 389)	307	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 53	GAUGUCGACGAUCG CCUCGCCUAUGCCG CC (SEQ ID No. 390)	315	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 54	GCAGGAAUGUCGAC GAUCGCCUCGCCUA UG (SEQ ID No. 391)	320	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 55	GAAUCUCAGGAAUG UCGACGAUCGCCUC GC (SEQ ID No. 392)	325	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 56	GCCCAGGAAUCUCA GGAAUGUCGACGAU CG (SEQ ID No. 393)	331	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 57	GCCUUGAACCCAGG AAUCUCAGGAAUGU CG (SEQ ID No. 394)	338	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 58	GCCAAGUCCUUGAA CCCAGGAAUCUCAG GA (SEQ ID No. 395)	344	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 59	GUGGGCUCCAAGUC CUUGAACCCAGGAA UC (SEQ ID No. 396)	350	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C

Gluc tiling guide 60	GCCAUGGGCUCCAAGUCCUUGAACCCAGGA (SEQ ID No. 397)	353	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 61	GGAACUGCUCCAUGGGCUCCAAGUCCUUGA (SEQ ID No. 398)	361	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 62	GUGCGAUGAACUGCUCCAUGGGCUCCAAGU (SEQ ID No. 399)	367	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 63	GGACCUGUGCGAUGAACUGCUCCAUGGGCU (SEQ ID No. 400)	373	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 64	GACAGAUCGACCUGUGCGAUGAACUGCUCC (SEQ ID No. 401)	380	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 65	GACACACAGAUCGACCUGUGCGAUGAACUG (SEQ ID No. 402)	384	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 66	GUGCAGUCCACACACAGAUCGACCUGUGCG (SEQ ID No. 403)	392	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 67	GCCAGUUGUGCAGUCCACACACAGAUCGAC (SEQ ID No. 404)	399	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 68	GGGCAGCCAGUUGUGCAGUCCACACACAGA (SEQ ID No. 405)	404	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 69	GUUUGAGGCAGCCAGUUGUGCAGUCCACAC (SEQ ID No. 406)	409	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 70	GAAGCCCUUUGAGGCAGCCAGUUGUGCAGU (SEQ ID No. 407)	415	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 71	GCACGUUGGCAAGCCCUUUGAGGCAGCCAG (SEQ ID No. 408)	424	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C

Gluc tiling guide 72	GACUGCACGUUGGCAAGCCCUUUGAGGCAG (SEQ ID No. 409)	428	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 73	GGGUCAGAACACUGCACGUUGGCAAGCCCU (SEQ ID No. 410)	437	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 74	GCAGGUCAGAACACUGCACGUUGGCAAGCC (SEQ ID No. 411)	439	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 75	GAGCAGGUCAGAACACUGCACGUUGGCAAG (SEQ ID No. 412)	441	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 76	GGCCACUUCUUGAGCAGGUCAGAACACUGC (SEQ ID No. 413)	452	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 77	GCGGCAGCCACUUCUUGAGCAGGUCAGAAC (SEQ ID No. 414)	457	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 78	GUGCGGCAGCCACUUCUUGAGCAGGUCAGA (SEQ ID No. 415)	459	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 79	GAGCGUUGCGGCAGCCACUUCUUGAGCAGG (SEQ ID No. 416)	464	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 80	GAAAGGUCGCACAGCGUUGCGGCAGCCACU (SEQ ID No. 417)	475	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 81	GCUGGCAAAGGUCGCACAGCGUUGCGGCAG (SEQ ID No. 418)	480	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 82	GGGCAAAGGUCGCACAGCGUUGCGGCAGCC (SEQ ID No. 419)	478	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 83	GUGGAUCUUGCUGGCAAAGGUCGCACAGCG (SEQ ID No. 420)	489	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C

Gluc tiling guide 84	GCACCUGGCCCUGGAUCUUGCUGGCAAAGG (SEQ ID No. 421)	499	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 85	GUGGCCCUGGAUCUUGCUGGCAAAGGUCGC (SEQ ID No. 422)	495	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 86	GUGAUCUUGUCCACCUGGCCCUGGAUCUUG (SEQ ID No. 423)	509	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 87	GCCCCUUGAUCUUGUCCACCUGGCCCUGGA (SEQ ID No. 424)	514	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 88	GCCCUUGAUCUUGUCCACCUGGCCCUGGAU (SEQ ID No. 425)	513	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 89	GCCUUGAUCUUGUCCACCUGGCCCUGGAUC (SEQ ID No. 426)	512	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 90	GGCAAAGGUCGCACAGCGUUGCGGCAGCCA (SEQ ID No. 427)	477	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 91	GCAAAGGUCGCACAGCGUUGCGGCAGCCAC (SEQ ID No. 428)	476	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 92	GAAGGUCGCACAGCGUUGCGGCAGCCACUU (SEQ ID No. 429)	474	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Gluc tiling guide 93	GAGGUCGCACAGCGUUGCGGCAGCCACUUC (SEQ ID No. 430)	473	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Non-targeting guide 1	GGUAAUGCCUGGCUUGUCGACGCAUAGUCUG (SEQ ID No. 431)	N/A	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C
Non-targeting guide 2	GGGAACCUUGGCCGUUAUAAAGUCUGACCAG (SEQ ID No. 432)	N/A	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C

Non-targeting guide 3	GGAGGGUGAGAAUUUAGAACCAAGAUUGUUG (SEQ ID No. 433)	N/A	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49C

Table 13: Guide sequences used for Cluc knockdown

Name	Spacer sequence	Position	Notes	First figure
Cluc tiling guide 1	GAGUCCUGGCAAUGAACAGUGGCGCAGUAG (SEQ ID No. 434)	32	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D
Cluc tiling guide 2	GGGUGCCACAGCUGCUAUCAAUACAUUCUC (SEQ ID No. 435)	118	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D
Cluc tiling guide 3	GUUACAUACUGACACAUUCGGCAACAUGUU (SEQ ID No. 436)	197	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D
Cluc tiling guide 4	GUAUGUACCAGGUUCCUGGAACUGGAAUCU (SEQ ID No. 437)	276	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D
Cluc tiling guide 5	GCCUUGGUUCCAUCCAGGUUCUCCAGGGUG (SEQ ID No. 438)	350	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D
Cluc tiling guide 6	GCAGUGAUGGGAUUCUCAGUAGCUUGAGCG (SEQ ID No. 439)	431	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D
Cluc tiling guide 7	GAGCCUGGCAUCUCAACAACAGCGAUGGUG (SEQ ID No. 440)	512	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D
Cluc tiling guide 8	GUGUCUGGGGCGAUUCUUACAGAUCUUCCU (SEQ ID No. 441)	593	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D
Cluc tiling guide 9	GCUGGAUCUGAAGUGAAGUCUGUAUCUUCC (SEQ ID No. 442)	671	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D
Cluc tiling guide 10	GGCAACGUCAUCAGGAUUUCCAUAGAGUGG (SEQ ID No. 443)	747	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D

Cluc tiling guide 11	GAGGCGCAGGAGAUG GUGUAGUAGUAGAAG (SEQ ID No. 444)	830	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D
Cluc tiling guide 13	GAGGGACCCUGGAAU UGGUAUCUUGCUUUG (SEQ ID No. 445)	986	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D
Cluc tiling guide 14	GGUAAGAGUCAACAU UCCUGUGUGAAACCU (SEQ ID No. 446)	1066	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D
Cluc tiling guide 15	GACCAGAAUCUGUUU UCCAUCAACAAUGAG (SEQ ID No. 447)	1143	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D
Cluc tiling guide 16	GAUGGCUGUAGUCAG UAUGUCACCAUCUUG (SEQ ID No. 448)	1227	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D
Cluc tiling guide 17	GUACCAUCGAAUGGA UCUCUAAUAUGUACG (SEQ ID No. 449)	1304	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D
Cluc tiling guide 18	GAGAUCACAGGCUCC UUCAGCAUCAAAAGA (SEQ ID No. 450)	1380	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D
Cluc tiling guide 19	GCUUUGACCGGCGAA GAGACUAUUGCAGAG (SEQ ID No. 451)	1461	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D
Cluc tiling guide 20	GCCCCUCAGGCAAUA CUCGUACAUGCAUCG (SEQ ID No. 452)	1539	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D
Cluc tiling guide 21	GCUGGUACUUCUAGG GUGUCUCCAUGCUUU (SEQ ID No. 453)	1619	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D
Non-targeting guide 1	GGUAAUGCCUGGCUU GUCGACGCAUAGUCU G (SEQ ID No. 454)	N/A	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D
Non-targeting guide 2	GGGAACCUUGGCCGU UAUAAAGUCUGACCA G (SEQ ID No. 455)	N/A	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D

Non-targeting guide 3	GGAGGGUGAGAAUUUAGAACCAAGAUUGUUG (SEQ ID No. 456)	N/A	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	49D

Table 14: Guide sequences used in this study for RNA editing in mammalian cells. Mismatched base flips are capitalized

Name	Spacer sequence	Notes	First figure
Tiling 30 nt 30 mismatch distance	gCauccugcggccucuacucugcauucaauu (SEQ ID No. 457)	Has a 5' G for U6 expression	50C
Tiling 30 nt 28 mismatch distance	gacCauccugcggccucuacucugcauucaa (SEQ ID No. 458)	Has a 5' G for U6 expression	50C
Tiling 30 nt 26 mismatch distance	gaaacCauccugcggccucuacucugcauuc (SEQ ID No. 459)	Has a 5' G for U6 expression	50C
Tiling 30 nt 24 mismatch distance	gcuaaacCauccugcggccucuacucugcau (SEQ ID No. 460)	Has a 5' G for U6 expression	50C
Tiling 30 nt 22 mismatch distance	guucuaaacCauccugcggccucuacucugc (SEQ ID No. 461)	Has a 5' G for U6 expression	50C
Tiling 30 nt 20 mismatch distance	guguucuaaacCauccugcggccucuacucu (SEQ ID No. 462)	Has a 5' G for U6 expression	50C
Tiling 30 nt 18 mismatch distance	gaauguucuaaacCauccugcggccucuacu (SEQ ID No. 463)	Has a 5' G for U6 expression	50C
Tiling 30 nt 16 mismatch distance	gagaauguucuaaacCauccugcggccucua (SEQ ID No. 464)	Has a 5' G for U6 expression	50C
Tiling 30 nt 14 mismatch distance	gauagaauguucuaaacCauccugcggccuc (SEQ ID No. 465)	Has a 5' G for U6 expression	50C
Tiling 30 nt 12 mismatch distance	gccauagaauguucuaaacCauccugcggcc (SEQ ID No. 466)	Has a 5' G for U6 expression	50C
Tiling 30 nt 10 mismatch distance	guuccauagaauguucuaaacCauccugcgg (SEQ ID No. 467)	Has a 5' G for U6 expression	50C
Tiling 30 nt 8 mismatch distance	gcuuuccauagaauguucuaaacCauccugc (SEQ ID No. 468)	Has a 5' G for U6 expression	50C
Tiling 30 nt 6 mismatch distance	gcucuuuccauagaauguucuaaacCauccu (SEQ ID No. 469)	Has a 5' G for U6 expression	50C

Tiling 30 nt 4 mismatch distance	gaucucuuuccauagaauguucuaaa cCauc (SEQ ID No. 470)	Has a 5' G for U6 expression	50C
Tiling 30 nt 2 mismatch distance	ggaaucucuuuccauagaauguucua aacCa (SEQ ID No. 471)	Has a 5' G for U6 expression	50C
Tiling 50 nt 50 mismatch distance	gCauccugcggccucuacucugcauu caauuacauacugacacauucggca (SEQ ID No. 472)	Has a 5' G for U6 expression	50C
Tiling 50 nt 48 mismatch distance	gacCauccugcggccucuacucugca uucaauuacauacugacacauucgg (SEQ ID No. 473)	Has a 5' G for U6 expression	50C
Tiling 50 nt 46 mismatch distance	gaaacCauccugcggccucuacucug cauucaauuacauacugacacauuc (SEQ ID No. 474)	Has a 5' G for U6 expression	50C
Tiling 50 nt 44 mismatch distance	gcuaaacCauccugcggccucuacuc ugcauucaauuacauacugacacau (SEQ ID No. 475)	Has a 5' G for U6 expression	50C
Tiling 50 nt 42 mismatch distance	guucuaaacCauccugcggccucuac ucugcauucaauuacauacugacac (SEQ ID No. 476)	Has a 5' G for U6 expression	50C
Tiling 50 nt 40 mismatch distance	guguucuaaacCauccugcggccucu acucugcauucaauuacauacugac (SEQ ID No. 477)	Has a 5' G for U6 expression	50C
Tiling 50 nt 38 mismatch distance	gaauguucuaaacCauccugcggccu cuacucugcauucaauuacauacug (SEQ ID No. 478)	Has a 5' G for U6 expression	50C
Tiling 50 nt 36 mismatch distance	gagaauguucuaaacCauccugcggc cucuacucugcauucaauuacauac (SEQ ID No. 479)	Has a 5' G for U6 expression	50C
Tiling 50 nt 34 mismatch distance	gauagaauguucuaaacCauccugcg gccucuacucugcauucaauuacau (SEQ ID No. 480)	Has a 5' G for U6 expression	50C
Tiling 50 nt 32 mismatch distance	gccauagaauguucuaaacCauccug cggccucuacucugcauucaauuac (SEQ ID No. 481)	Has a 5' G for U6 expression	50C
Tiling 50 nt 30 mismatch distance	guuccauagaauguucuaaacCaucc ugcggccucuacucugcauucaauu (SEQ ID No. 482)	Has a 5' G for U6 expression	50C
Tiling 50 nt 28 mismatch distance	gcuuuccauagaauguucuaaacCau ccugcggccucuacucugcauucaa (SEQ ID No. 483)	Has a 5' G for U6 expression	50C
Tiling 50 nt 26 mismatch distance	gcucuuuccauagaauguucuaaacC auccugcggccucuacucugcauuc (SEQ ID No. 484)	Has a 5' G for U6 expression	50C
Tiling 50 nt 24 mismatch distance	gaucucuuuccauagaauguucuaaa cCauccugcggccucuacucugcau (SEQ ID No. 485)	Has a 5' G for U6 expression	50C

Tiling 50 nt 22 mismatch distance	ggaaucucuuuccauagaauguucua aacCauccugcggccucuacucugc (SEQ ID No. 486)	Has a 5' G for U6 expression	50C
Tiling 50 nt 20 mismatch distance	guggaaucucuuuccauagaauguuc uaaacCauccugcggccucuacucu (SEQ ID No. 487)	Has a 5' G for U6 expression	50C
Tiling 50 nt 18 mismatch distance	gacuggaaucucuuuccauagaaugu ucuaaacCauccugcggccucuacu (SEQ ID No. 488)	Has a 5' G for U6 expression	50C
Tiling 50 nt 16 mismatch distance	ggaacuggaaucucuuuccauagaau guucuaaacCauccugcggccucua (SEQ ID No. 489)	Has a 5' G for U6 expression	50C
Tiling 50 nt 14 mismatch distance	guggaacuggaaucucuuuccauaga auguucuaaacCauccugcggccu c(SEQ ID No. 490)	Has a 5' G for U6 expression	50C
Tiling 50 nt 12 mismatch distance	gccuggaacuggaaucucuuuccaua gaauguucuaaacCauccugcggcc (SEQ ID No. 491)	Has a 5' G for U6 expression	50C
Tiling 50 nt 10 mismatch distance	guuccuggaacuggaaucucuuucca uagaauguucuaaacCauccugcgg (SEQ ID No. 492)	Has a 5' G for U6 expression	50C
Tiling 50 nt 8 mismatch distance	ggguuccuggaacuggaaucucuuuc cauagaauguucuaaacCauccugc (SEQ ID No. 493)	Has a 5' G for U6 expression	50C
Tiling 50 nt 6 mismatch distance	gcagguuccuggaacuggaaucucuu uccauagaauguucuaaacCauccu (SEQ ID No. 494)	Has a 5' G for U6 expression	50C
Tiling 50 nt 4 mismatch distance	gaccagguuccuggaacuggaaucuc uuuccauagaauguucuaaacCauc (SEQ ID No. 495)	Has a 5' G for U6 expression	50C
Tiling 50 nt 2 mismatch distance	gguaccagguuccuggaacuggaauc ucuuuccauagaauguucuaaacCa (SEQ ID No. 496)	Has a 5' G for U6 expression	50C
Tiling 70 nt 70 mismatch distance	gCauccugcggccucuacucugcauu caauuacauacugacacauucggcaac auguuuuuccugguuuau (SEQ ID No. 497)	Has a 5' G for U6 expression	50C
Tiling 70 nt 68 mismatch distance	gacCauccugcggccucuacucugca uucaauuacauacugacacauucggca acauguuuuuccugguuu (SEQ ID No. 498)	Has a 5' G for U6 expression	50C
Tiling 70 nt 66 mismatch distance	gaaacCauccugcggccucuacucug cauucaauuacauacugacacauucgg caacauguuuuuccuggu (SEQ ID No. 499)	Has a 5' G for U6 expression	50C
Tiling 70 nt 64 mismatch distance	gcuaaacCauccugcggccucuacuc ugcauucaauuacauacugacacauuc ggcaacauguuuuuccug (SEQ ID No. 500)	Has a 5' G for U6 expression	50C

Tiling 70 nt 62 mismatch distance	guucuaaacCauccugcggccucuac ucugcauucaauuacauacugacacau ucggcaacauguuuuucc (SEQ ID No. 501)	Has a 5' G for U6 expression	50C
Tiling 70 nt 60 mismatch distance	guguucuaaacCauccugcggccucu acucugcauucaauuacauacugacac auucggcaacauguuuuu (SEQ ID No. 502)	Has a 5' G for U6 expression	50C
Tiling 70 nt 58 mismatch distance	gaauguucuaaacCauccugcggccu cuacucugcauucaauuacauacugac acauucggcaacauguuu (SEQ ID No. 503)	Has a 5' G for U6 expression	50C
Tiling 70 nt 56 mismatch distance	gagaauguucuaaacCauccugcggc cucuacucugcauucaauuacauacug acacauucggcaacaugu (SEQ ID No. 504)	Has a 5' G for U6 expression	50C
Tiling 70 nt 54 mismatch distance	gauagaauguucuaaacCauccugcg gccucuacucugcauucaauuacauac ugacacauucggcaacau (SEQ ID No. 505)	Has a 5' G for U6 expression	50C
Tiling 70 nt 52 mismatch distance	gccauagaauguucuaaacCauccug cggccucuacucugcauucaauuacau acugacacauucggcaac (SEQ ID No. 506)	Has a 5' G for U6 expression	50C
Tiling 70 nt 50 mismatch distance	guuccauagaauguucuaaacCaucc ugcggccucuacucugcauucaauua cauacugacacauucggca (SEQ ID No. 507)	Has a 5' G for U6 expression	50C
Tiling 70 nt 48 mismatch distance	gcuuuccauagaauguucuaaacCau ccugcggccucuacucugcauucaau uacauacugacacauucgg (SEQ ID No. 508)	Has a 5' G for U6 expression	50C
Tiling 70 nt 46 mismatch distance	gcucuuuccauagaauguucuaaacC auccugcggccucuacucugcauucaa uuacauacugacacauuc (SEQ ID No. 509)	Has a 5' G for U6 expression	50C
Tiling 70 nt 44 mismatch distance	gaucucuuuccauagaauguucuaaa cCauccugcggccucuacucugcauu caauuacauacugacacau (SEQ ID No. 510)	Has a 5' G for U6 expression	50C
Tiling 70 nt 42 mismatch distance	ggaaucucuuuccauagaauguucua aacCauccugcggccucuacucugca uucaauuacauacugacac (SEQ ID No. 511)	Has a 5' G for U6 expression	50C
Tiling 70 nt 40 mismatch distance	guggaaucucuuuccauagaauguuc uaaacCauccugcggccucuacucug cauucaauuacauacugac (SEQ ID No. 512)	Has a 5' G for U6 expression	50C

Tiling 70 nt 38 mismatch distance	gacuggaaucucuuuccauagaaugu ucuaaacCauccugcggccucuacuc ugcauucaauuacauacug (SEQ ID No. 513)	Has a 5' G for U6 expression	50C
Tiling 70 nt 36 mismatch distance	ggaacuggaaucucuuuccauagaau guucuaaacCauccugcggccucuac ucugcauucaauuacauac (SEQ ID No. 514)	Has a 5' G for U6 expression	50C
Tiling 70 nt 34 mismatch distance	guggaacuggaaucucuuuccauaga auguucuaaacCauccugcggccucu acucugcauucaauuacau (SEQ ID No. 515)	Has a 5' G for U6 expression	50C
Tiling 70 nt 32 mismatch distance	gccuggaacuggaaucucuuuccaua gaauguucuaaacCauccugcggccu cuacucugcauucaauuac (SEQ ID No. 516)	Has a 5' G for U6 expression	50C
Tiling 70 nt 30 mismatch distance	guuccuggaacuggaaucucuuucca uagaauguucuaaacCauccugcggc cucuacucugcauucaauu (SEQ ID No. 517)	Has a 5' G for U6 expression	50C
Tiling 70 nt 28 mismatch distance	ggguuccuggaacuggaaucucuuuc cauagaauguucuaaacCauccugcg gccucuacucugcauucaa (SEQ ID No. 518)	Has a 5' G for U6 expression	50C
Tiling 70 nt 26 mismatch distance	gcagguuccuggaacuggaaucucuu uccauagaauguucuaaacCauccug cggccucuacucugcauuc (SEQ ID No. 519)	Has a 5' G for U6 expression	50C
Tiling 70 nt 24 mismatch distance	gaccagguuccuggaacuggaaucuc uuuccauagaauguucuaaacCaucc ugcggccucuacucugcau (SEQ ID No. 520)	Has a 5' G for U6 expression	50C
Tiling 70 nt 22 mismatch distance	gguaccagguuccuggaacuggaauc ucuuuccauagaauguucuaaacCau ccugcggccucuacucugc (SEQ ID No. 521)	Has a 5' G for U6 expression	50C
Tiling 70 nt 20 mismatch distance	gauguaccagguuccuggaacuggaa ucucuuuccauagaauguucuaaacC auccugcggccucuacucu (SEQ ID No. 522)	Has a 5' G for U6 expression	50C
Tiling 70 nt 18 mismatch distance	gguauguaccagguuccuggaacugg aaucucuuuccauagaauguucuaaac Cauccugcggccucuacu (SEQ ID No. 523)	Has a 5' G for U6 expression	50C
Tiling 70 nt 16 mismatch distance	gacguauguaccagguuccuggaacu ggaaucucuuuccauagaauguucua aacCauccugcggccucua (SEQ ID No. 524)	Has a 5' G for U6 expression	50C

Tiling 70 nt 14 mismatch distance	gacacguauguaccagguuccuggaacuggaaucucuuuccauagaauguucuaaacCauccugcggccuc (SEQ ID No. 525)	Has a 5' G for U6 expression	50C
Tiling 70 nt 12 mismatch distance	gcaacacguauguaccagguuccuggaacuggaaucucuuuccauagaauguucuaaacCauccugcggcc (SEQ ID No. 526)	Has a 5' G for U6 expression	50C
Tiling 70 nt 10 mismatch distance	gcccaacacguauguaccagguuccuggaacuggaaucucuuuccauagaauguucuaaacCauccugcgg (SEQ ID No. 527)	Has a 5' G for U6 expression	50C
Tiling 70 nt 8 mismatch distance	ggacccaacacguauguaccagguuccuggaacuggaaucucuuuccauagaauguucuaaacCauccugc (SEQ ID No. 528)	Has a 5' G for U6 expression	50C
Tiling 70 nt 6 mismatch distance	guugacccaacacguauguaccagguuccuggaacuggaaucucuuuccauagaauguucuaaacCauccu (SEQ ID No. 529)	Has a 5' G for U6 expression	50C
Tiling 70 nt 4 mismatch distance	gccuugacccaacacguauguaccagguuccuggaacuggaaucucuuuccauagaauguucuaaacCauc (SEQ ID No. 530)	Has a 5' G for U6 expression	50C
Tiling 70 nt 2 mismatch distance	guuccuugacccaacacguauguaccaggguuccuggaacuggaaucucuuuccauagaauguucuaaacCa (SEQ ID No. 531)	Has a 5' G for U6 expression	50C
Tiling 84 nt 84 mismatch distance	gCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacauguuuuuccugguuuauuuucacacagucca (SEQ ID No. 532)	Has a 5' G for U6 expression	50C
Tiling 84 nt 82 mismatch distance	gacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacauguuuuuccugguuuauuuucacacaguc (SEQ ID No. 533)	Has a 5' G for U6 expression	50C
Tiling 84 nt 80 mismatch distance	gaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacauguuuuuccugguuuauuuucacacag (SEQ ID No. 534)	Has a 5' G for U6 expression	50C
Tiling 84 nt 78 mismatch distance	gcuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacauguuuuuccugguuuauuucacac (SEQ ID No. 535)	Has a 5' G for U6 expression	50C
Tiling 84 nt 76 mismatch distance	guucuaaacCauccugcggccucuacucugcauucaauuacauacugacacaucggcaacauguuuuuccugguuuauuucac (SEQ ID No. 536)	Has a 5' G for U6 expression	50C

Tiling 84 nt 74 mismatch distance	guguucuaaacCauccugcggccucu acucugcauucaauuacauacugacac auucggcaacauguuuuuccuggguuu auuuuc (SEQ ID No. 537)	Has a 5' G for U6 expression	50C
Tiling 84 nt 72 mismatch distance	gaauguucuaaacCauccugcggccu cuacucugcauucaauuacauacugac acauucggcaacauguuuuuccuggu uuauuu (SEQ ID No. 538)	Has a 5' G for U6 expression	50C
Tiling 84 nt 70 mismatch distance	gagaauguucuaaacCauccugcggc cucuacucugcauucaauuacauacug acacauucggcaacauguuuuuccug guuuau (SEQ ID No. 539)	Has a 5' G for U6 expression	50C
Tiling 84 nt 68 mismatch distance	gauagaauguucuaaacCauccugcg gccucuacucugcauucaauuacauac ugacacauucggcaacauguuuuucc ugguuu (SEQ ID No. 540)	Has a 5' G for U6 expression	50C
Tiling 84 nt 66 mismatch distance	gccauagaauguucuaaacCauccug cggccucuacucugcauucaauuacau acugacacauucggcaacauguuuuu ccuggu (SEQ ID No. 541)	Has a 5' G for U6 expression	50C
Tiling 84 nt 64 mismatch distance	guuccauagaauguucuaaacCaucc ugcggccucuacucugcauucaauua cauacugacacauucggcaacauguuu uuccug (SEQ ID No. 542)	Has a 5' G for U6 expression	50C
Tiling 84 nt 62 mismatch distance	gcuuuccauagaauguucuaaacCau ccugcggccucuacucugcauucaau uacauacugacacauucggcaacaugu uuuucc (SEQ ID No. 543)	Has a 5' G for U6 expression	50C
Tiling 84 nt 60 mismatch distance	gcucuuuccauagaauguucuaaacC auccugcggccucuacucugcauucaa uuacauacugacacauucggcaacaug uuuuu (SEQ ID No. 544)	Has a 5' G for U6 expression	50C
Tiling 84 nt 58 mismatch distance	gaucucuuuccauagaauguucuaaa cCauccugcggccucuacucugcauu caauuacauacugacacauucggcaac auguuu (SEQ ID No. 545)	Has a 5' G for U6 expression	50C
Tiling 84 nt 56 mismatch distance	ggaaucucuuuccauagaauguucua aacCauccugcggccucuacucugca uucaauuacauacugacacauucggca acaugu (SEQ ID No. 546)	Has a 5' G for U6 expression	50C
Tiling 84 nt 54 mismatch distance	guggaaucucuuuccauagaauguuc uaaacCauccugcggccucuacucug cauucaauuacauacugacacauucgg caacau (SEQ ID No. 547)	Has a 5' G for U6 expression	50C
Tiling 84 nt 52 mismatch distance	gacuggaaucucuuuccauagaaugu ucuaaacCauccugcggccucuacuc ugcauucaauuacauacugacacauuc ggcaac (SEQ ID No. 548)	Has a 5' G for U6 expression	50C

Tiling 84 nt 50 mismatch distance	ggaacuggaaucucuuuccauagaau guucuaaacCauccugcggccucuac ucugcauucaauuacauacugacacau ucggca (SEQ ID No. 549)	Has a 5' G for U6 expression	50C
Tiling 84 nt 48 mismatch distance	guggaacuggaaucucuuuccauaga auguucuaaacCauccugcggccucu acucugcauucaauuacauacugacac auucgg (SEQ ID No. 550)	Has a 5' G for U6 expression	50C
Tiling 84 nt 46 mismatch distance	gccuggaacuggaaucucuuuccaua gaauguucuaaacCauccugcggccu cuacucugcauucaauuacauacugac acauuc (SEQ ID No. 551)	Has a 5' G for U6 expression	50C
Tiling 84 nt 44 mismatch distance	guuccuggaacuggaaucucuuucca uagaauguucuaaacCauccugcggc cucuacucugcauucaauuacauacug acacau (SEQ ID No. 552)	Has a 5' G for U6 expression	50C
Tiling 84 nt 42 mismatch distance	ggguuccuggaacuggaaucucuuuc cauagaauguucuaaacCauccugcg gccucuacucugcauucaauuacauac ugacac (SEQ ID No. 553)	Has a 5' G for U6 expression	50C
Tiling 84 nt 40 mismatch distance	gcagguuccuggaacuggaaucucuu uccauagaauguucuaaacCauccug cggccucuacucugcauucaauuacau acugac (SEQ ID No. 554)	Has a 5' G for U6 expression	50C
Tiling 84 nt 38 mismatch distance	gaccagguuccuggaacuggaaucuc uuuccauagaauguucuaaacCaucc ugcggccucuacucugcauucaauua cauacug (SEQ ID No. 555)	Has a 5' G for U6 expression	50C
Tiling 84 nt 36 mismatch distance	gguaccagguuccuggaacuggaauc ucuuuccauagaauguucuaaacCau ccugcggccucuacucugcauucaau uacauac (SEQ ID No. 556)	Has a 5' G for U6 expression	50C
Tiling 84 nt 34 mismatch distance	gauguaccagguuccuggaacuggaa ucucuuuccauagaauguucuaaacC auccugcggccucuacucugcauucaa uuacau (SEQ ID No. 557)	Has a 5' G for U6 expression	50C
Tiling 84 nt 32 mismatch distance	gguauguaccagguuccuggaacugg aaucucuuuccauagaauguucuaaac Cauccugcggccucuacucugcauuc aauuac (SEQ ID No. 558)	Has a 5' G for U6 expression	50C
Tiling 84 nt 30 mismatch distance	gacguauguaccagguuccuggaacu ggaaucucuuuccauagaauguucua aacCauccugcggccucuacucugca uucaauu (SEQ ID No. 559)	Has a 5' G for U6 expression	50C
Tiling 84 nt 28 mismatch distance	gacacguauguaccagguuccuggaa cuggaaucucuuuccauagaauguuc uaaacCauccugcggccucuacucug cauucaa (SEQ ID No. 560)	Has a 5' G for U6 expression	50C

Tiling 84 nt 26 mismatch distance	gcaacacguauguaccagguuccugg aacuggaaucucuuuccauagaaugu ucuaaacCauccugcggccucuacuc ugcauuc (SEQ ID No. 561)	Has a 5' G for U6 expression	50C
Tiling 84 nt 24 mismatch distance	gcccaacacguauguaccagguuccug gaacuggaaucucuuuccauagaaug uucuaaacCauccugcggccucuacu cugcau (SEQ ID No. 562)	Has a 5' G for U6 expression	50C
Tiling 84 nt 22 mismatch distance	ggacccaacacguauguaccagguucc uggaacuggaaucucuuuccauagaa uguucuaaacCauccugcggccucua cucugc (SEQ ID No. 563)	Has a 5' G for U6 expression	50C
Tiling 84 nt 20 mismatch distance	guugacccaacacguauguaccaggu uccuggaacuggaaucucuuuccaua gaauguucuaaacCauccugcggccu cuacucu (SEQ ID No. 564)	Has a 5' G for U6 expression	50C
Tiling 84 nt 18 mismatch distance	gccuugacccaacacguauguaccagg uuccuggaacuggaaucucuuuccau agaauguucuaaacCauccugcggcc ucuacu (SEQ ID No. 565)	Has a 5' G for U6 expression	50C
Tiling 84 nt 16 mismatch distance	guuccuugacccaacacguauguacca gguuccuggaacuggaaucucuuucc auagaauguucuaaacCauccugcgg ccucua (SEQ ID No. 566)	Has a 5' G for U6 expression	50C
Tiling 84 nt 14 mismatch distance	ggguuccuugacccaacacguaugua ccagguuccuggaacuggaaucucuu uccauagaauguucuaaacCauccug cggccuc (SEQ ID No. 567)	Has a 5' G for U6 expression	50C
Tiling 84 nt 12 mismatch distance	guugguuccuugacccaacacguaug uaccagguuccuggaacuggaaucuc uuuccauagaauguucuaaacCaucc ugcggcc (SEQ ID No. 568)	Has a 5' G for U6 expression	50C
Tiling 84 nt 10 mismatch distance	gccuugguuccuugacccaacacgua uguaccagguuccuggaacuggaauc ucuuuccauagaauguucuaaacCau ccugcgg (SEQ ID No. 569)	Has a 5' G for U6 expression	50C
Tiling 84 nt 8 mismatch distance	ggcccuugguuccuugacccaacacg uauguaccagguuccuggaacuggaa ucucuuuccauagaauguucuaaacC auccugc (SEQ ID No. 570)	Has a 5' G for U6 expression	50C
Tiling 84 nt 6 mismatch distance	gccgcccuugguuccuugacccaacac guauguaccagguuccuggaacugga aucucuuuccauagaauguucuaaac Cauccu (SEQ ID No. 571)	Has a 5' G for U6 expression	50C
Tiling 84 nt 4 mismatch distance	gcgccgcccuugguuccuugacccaac acguauguaccagguuccuggaacug gaaucucuuuccauagaauguucuaa acCauc (SEQ ID No. 572)	Has a 5' G for U6 expression	50C

Tiling 84 nt 2 mismatch distance	ggucgccgcccuugguuccuugaccc aacacguauguaccagguuccuggaac uggaaucucuuuccauagaauguucu aaacCa (SEQ ID No. 573)	Has a 5' G for U6 expression	50C
ADAR non-targeting guide	GUAAUGCCUGGCUUGUC GACGCAUAGUCUG (SEQ ID No. 574)	Has a 5' G for U6 expression	50C
PFS binding screen guide for TAG motif	gaaaacgcagguuccucCaguuucgg gagcagcgcacgucucccuguaguc (SEQ ID No. 575)	Has a 5' G for U6 expression	51B
PFS binding screen guide for AAC motif	gacgcagguuccucuagCuucgggag cagcgcacgucucccuguagucaag (SEQ ID No. 576)	Has a 5' G for U6 expression	51B
PFS binding screen non-targeting	GUAAUGCCUGGCUUGUC GACGCAUAGUCUG (SEQ ID No. 577)	Has a 5' G for U6 expression	51B
Motif preference targeting guide	gauagaauguucuaaacCauccugcg gccucuacucugcauucaauuacau (SEQ ID No. 578)	Has a 5' G for U6 expression	51C
Motif preference non-targeting guide	GUAAUGCCUGGCUUGUC GACGCAUAGUCUG (SEQ ID No. 579)	Has a 5' G for U6 expression	51C
PPIB tiling guide 50 mismatch distance	gCaaggccacaaaauuauccacuguu uuuggaacagucuuuccgaagagac (SEQ ID No. 580)	Has a 5' G for U6 expression	57D
PPIB tiling guide 42 mismatch distance	gccuguagcCaaggccacaaaauuau ccacuguuuuuggaacagucuuucc (SEQ ID No. 581)	Has a 5' G for U6 expression	57D
PPIB tiling guide 34 mismatch distance	gcuuucucuccuguagcCaaggccac aaaauuauccacuguuuuuggaaca (SEQ ID No. 582)	Has a 5' G for U6 expression	57D
PPIB tiling guide 26 mismatch distance	ggccaaauccuuucucuccuguagcC aaggccacaaaauuauccacuguuu (SEQ ID No. 583)	Has a 5' G for U6 expression	57D
PPIB tiling guide 18 mismatch distance	guuuuuguagccaaauccuuucucuc cuguagcCaaggccacaaaauuauc (SEQ ID No. 584)	Has a 5' G for U6 expression	57D
PPIB tiling guide 10 mismatch distance	gauuugcuguuuuuguagccaaaucc uuucucuccuguagcCaaggccaca (SEQ ID No. 585)	Has a 5' G for U6 expression	57D
PPIB tiling guide 2 mismatch distance	gacgauggaauuugcuguuuuuguag ccaaauccuuucucuccuguagcCa (SEQ ID No. 586)	Has a 5' G for U6 expression	57D
Targeting guide, opposite base G	gauagaauguucuaaacGauccugcg gccucuacucugcauucaauuacau (SEQ ID No. 587)	Has a 5' G for U6 expression	57D
Targeting guide, opposite base A	gauagaauguucuaaacAauccugcg gccucuacucugcauucaauuacau (SEQ ID No. 588)	Has a 5' G for U6 expression	57D

Targeting guide, opposite base C	gauagaauguucuaaacUauccugcg gccucuacucugcauucaauuacau (SEQ ID No. 589)	Has a 5' G for U6 expression	57D
AVPR2 guide 37 mismatch distance	ggucccacgcggccCacagcugcacc aggaagaagggugcccagcacagca (SEQ ID No. 590)	Has a 5' G for U6 expression	52A
AVPR2 guide 35 mismatch distance	gggguccccacgcggccCacagcugca ccaggaagaagggugcccagcacag (SEQ ID No. 591)	Has a 5' G for U6 expression	52A
AVPR2 guide 33 mismatch distance	gccggguccccacgcggccCacagcug caccaggaagaagggugcccagcac (SEQ ID No. 592)	Has a 5' G for U6 expression	52A
FANCC guide 37 mismatch distance	gggugaugacauccCaggcgaucgug uggccuccaggagcccagagcagga (SEQ ID No. 593)	Has a 5' G for U6 expression	52B
FANCC guide 35 mismatch distance	gagggugaugacauccCaggcgaucg uguggccuccaggagcccagagcag (SEQ ID No. 594)	Has a 5' G for U6 expression	52B
FANCC guide 32 mismatch distance	gaucagggugaugacauccCaggcga ucguguggccuccaggagcccagag (SEQ ID No. 595)	Has a 5' G for U6 expression	52B
Synthetic disease gene target IL2RG	gguggcuccauucacucCaaugcuga gcacuuccacagagugggguuaaagc (SEQ ID No. 596)	Has a 5' G for U6 expression	52E
Synthetic disease gene target F8	guuucuaauauauuuugCcagacuga uggacuauucucaauuaauaaugau (SEQ ID No. 597)	Has a 5' G for U6 expression	52E
Synthetic disease gene target LDLR	gagauguugcuguggauCcaguccac agccagcccgucgggggccuggaug (SEQ ID No. 598)	Has a 5' G for U6 expression	52E
Synthetic disease gene target CBS	gcaggccggcccagcugCcaggugca ccugcucggagcaucgggccggauc (SEQ ID No. 599)	Has a 5' G for U6 expression	52E
Synthetic disease gene target HBB	gcaaagaaccucuggguCcaagggua gaccaccagcagccugcccagggcc (SEQ ID No. 600)	Has a 5' G for U6 expression	52E
Synthetic disease gene target ALDOB	gaagagaaacuuaguuuCcagggcuu ugguagagggcaaagguugauagca (SEQ ID No. 601)	Has a 5' G for U6 expression	52E
Synthetic disease gene target DMD	gucagccuagugcagagCcacuggua guuggugguuagaguuucaaguucc (SEQ ID No. 602)	Has a 5' G for U6 expression	52E
Synthetic disease gene target SMAD4	ggcucauugugaacaggCcaguaaug uccgggaugggggcggcauaggcggg (SEQ ID No. 603)	Has a 5' G for U6 expression	52E
Synthetic disease gene target BRCA2	guagcuaaagaacuugaCcaagacau aucaggauccaccucagcuccuaga (SEQ ID No. 604)	Has a 5' G for U6 expression	52E

Synthetic disease gene target GRIN2A	ggggcauuguucugugcCcaguccu gcugguagaccugcucccggguggcu (SEQ ID No. 605)	Has a 5' G for U6 expression	52E
Synthetic disease gene target SCN9A	gagaagucguucaugugCcaccgugg gagcguacagucaucauugaucuug (SEQ ID No. 606)	Has a 5' G for U6 expression	52E
Synthetic disease gene target TARDBP	gggauuaaugcugaacgCaccaaagu ucaucccaccacccauauuacuacc (SEQ ID No. 607)	Has a 5' G for U6 expression	52E
Synthetic disease gene target CFTR	gcuccaaaggcuuuccuCcacuguug caaaguuauugaaucccaagacaca (SEQ ID No. 608)	Has a 5' G for U6 expression	52E
Synthetic disease gene target UBE3A	gaugaaugaacgauuucCcagaacuc ccuaaucagaacagagucccuggua (SEQ ID No. 609)	Has a 5' G for U6 expression	52E
Synthetic disease gene target SMPD1	ggagccucugccggagcCcagagaac ccgagagucagacagagccagcgcc (SEQ ID No. 610)	Has a 5' G for U6 expression	52E
Synthetic disease gene target USH2A	ggcuuccguggagacacCcaaucaau uugaagagaucuugaagugaugcca (SEQ ID No. 611)	Has a 5' G for U6 expression	52E
Synthetic disease gene target MEN1	gugggacugcccuccucCcauuugca gaugccgucguagaaucgcagcagg (SEQ ID No. 612)	Has a 5' G for U6 expression	52E
Synthetic disease gene target C8orf37	gcuucuucaauaguucuCcagcuaca cuggcaggcauaugcccguguuccu (SEQ ID No. 613)	Has a 5' G for U6 expression	52E
Synthetic disease gene target MLH1	gauuccuuuucuucgucCcaauucac cucaguggcuagucgaagaaugaag (SEQ ID No. 614)	Has a 5' G for U6 expression	52E
Synthetic disease gene target TSC2	gcagcuucagcaccuucCagucagac uccugcuucaagcacugcagcagga (SEQ ID No. 615)	Has a 5' G for U6 expression	52E
Synthetic disease gene target NF1	gccauuugcuugcagugCcacuccag aggauuccggauugccauaaauacu (SEQ ID No. 616)	Has a 5' G for U6 expression	52E
Synthetic disease gene target MSH6	guucaauaguuuuggucCaguaucg uuuacagcccuucuugguagauuuca (SEQ ID No. 617)	Has a 5' G for U6 expression	52E
Synthetic disease gene target SMN1	ggcaaccgucuucugacCaaauggca gaacauuuguccccaacuuuccacu (SEQ ID No. 618)	Has a 5' G for U6 expression	52E
Synthetic disease gene target SH3TC2	gcgacuuuccaaugaacCacugaagc ccagguaugacaaagccgaugaucu (SEQ ID No. 619)	Has a 5' G for U6 expression	52E
Synthetic disease gene target DNAH5	guuuacacucaugcuucCacagcuuu aacagaucauuugguuccuugauga (SEQ ID No. 620)	Has a 5' G for U6 expression	52E

Synthetic disease gene target MECP2	gcuuaagcuuccgugucCagccuuca ggcaggguggggucaucauacaugg (SEQ ID No. 621)	Has a 5' G for U6 expression	52E
Synthetic disease gene target ADGRV1	ggacagcugggcugaucCaugauguc auccagaaacacuggggacccucag (SEQ ID No. 622)	Has a 5' G for U6 expression	52E
Synthetic disease gene target AHI1	gucucaucucaacuuucCauauccgu aucauggaaucauagcauccuguaa (SEQ ID No. 623)	Has a 5' G for U6 expression	52E
Synthetic disease gene target PRKN	gcaugcagacgcgguucCacucgcag ccacaguuccagcaccacucgagcc (SEQ ID No. 624)	Has a 5' G for U6 expression	52E
Synthetic disease gene target COL3A1	guugguuagggucaaccCaguauucu ccacucuugaguucaggauggcaga (SEQ ID No. 625)	Has a 5' G for U6 expression	52E
Synthetic disease gene target BRCA1	gcuacacuguccaacacCcacucucg ggucaccacaggugccucacacauc (SEQ ID No. 626)	Has a 5' G for U6 expression	52E
Synthetic disease gene target MYBPC3	gcugcacuguguaccccCagagcucc guguugccgacauccugggguggcu (SEQ ID No. 627)	Has a 5' G for U6 expression	52E
Synthetic disease gene target APC	gagcuuccugccacuccCaacagguu ucacaguaagcgcguaucuguucca (SEQ ID No. 628)	Has a 5' G for U6 expression	52E
Synthetic disease gene target BMPR2	gacggcaagagcuuaccCagucacuu guguggagacuuaaauacuugcaua (SEQ ID No. 629)	Has a 5' G for U6 expression	52E
KRAS tiling guide 50 mismatch distance	gCaaggccacaaaauuauccacuguu uuuggaacagucuuuccgaagagac (SEQ ID No. 630)	Has a 5' G for U6 expression	53A
KRAS tiling guide 42 mismatch distance	gccuguagcCaaggccacaaaauuau ccacuguuuuuggaacagucuuucc (SEQ ID No. 631)	Has a 5' G for U6 expression	53A
KRAS tiling guide 34 mismatch distance	gcuuucucuccuguagcCaaggccac aaaauuauccacuguuuuuggaaca (SEQ ID No. 632)	Has a 5' G for U6 expression	53A
KRAS tiling guide 26 mismatch distance	ggccaaauccuuucucuccuguagcC aaggccacaaaauuauccacuguuu (SEQ ID No. 633)	Has a 5' G for U6 expression	53A
KRAS tiling guide 18 mismatch distance	guuuuuguagccaaauccuuucucuc cuguagcCaaggccacaaaauuauc (SEQ ID No. 634)	Has a 5' G for U6 expression	53A

KRAS tiling guide 10 mismatch distance	gauuugcuguuuuuguagccaaaucc uuucucuccuguagcCaaggccaca (SEQ ID No. 635)	Has a 5' G for U6 expression	53A
KRAS tiling guide 2 mismatch distance	gacgauggaauuugcuguuuuuguag ccaaauccuuucucuccuguagcCa (SEQ ID No. 636)	Has a 5' G for U6 expression	53A
KRAS tiling non-targeting guide	GUAAUGCCUGGCUUGUC GACGCAUAGUCUG (SEQ ID No. 637)	Has a 5' G for U6 expression	53A
Luciferase W85X targeting guide for transcriptome specificity	gauagaauguucuaaacCauccugcg gccucuacucugcauucaauuacau (SEQ ID No. 638)	Has a 5' G for U6 expression	53B
Non-targeting guide for transcriptome specificity	GCAGGGUUUUCCCAGUC ACGACGUUGUAAAGUUG (SEQ ID No. 639)	Has a 5' G for U6 expression	53C
endogenous KRAS guide 2	gucaaggcacucuugccCacgccacc agcuccaacuaccacaaguuuauau (SEQ ID No. 640)	Has a 5' G for U6 expression	54F
endogenous PPIB guide 1	gcaaagaucacccggccCacaucuuca ucuccaauucguaggucaaaauac (SEQ ID No. 641)	Has a 5' G for U6 expression	54G
endogenous KRAS guide 1	GcgccaccagcuccaacCaccacaag uuuauauucagucauuuucagcagg (SEQ ID No. 642)	Has a 5' G for U6 expression	54F
endogenous KRAS guide 3	GuuucuccaucaauuacCacuugcu uccuguaggaauccucuauuGUugg a (SEQ ID No. 643)	Has a 5' G for U6 expression	54F
endogenous PPIB guide 2	GcuuucucuccuguagcCaaggccac aaaauuauccacuguuuuuggaaca (SEQ ID No. 644)	Has a 5' G for U6 expression	54G
endogenous non-targeting guide	GUAAUGCCUGGCUUGUC GACGCAUAGUCUG (SEQ ID No. 645)	Has a 5' G for U6 expression	54F
BoxB Cluc guide	ucuuuccauaGGCCCUGAAAA AGGGCCuguucuaaacCauccug cggccucuacucGGCCCUGAAA AAGGGCCauucaauuac (SEQ ID No. 646)	Has a 5' G for U6 expression	62B
BoxB non-targeting guide	cagcuggcgaGGCCCUGAAAA AGGGCCggggaugugcCgcaagg cgauuaaguuggGGCCCUGAA AAAGGGCCacgccagggu (SEQ ID No. 647)	Has a 5' G for U6 expression	62B
Stafforst full length ADAR2 guide 1	GUGGAAUAGUAUAACAA UAUGCUAAAUGUUGUUA UAGUAUCCCACucuaaaCCA	Has a 5' G for U6 expression	62C

	uccugcgGGGCCCUCUUCAG GGCCC (SEQ ID No. 648)		
Stafforst full length ADAR2 non-targeting guide	GUGGAAUAGUAUAACAA UAUGCUAAAUGUUGUUA UAGUAUCCCACacccuggcgu uacccaGGGCCCUCUUCAGG GCCC (SEQ ID No. 649)	Has a 5' G for U6 expression	62C

EXAMPLE 4 – REPAIRv3 Search

[01252] To identifying additional ADAR mutants with increased efficiency and specificity, Cas13b-ADAR fusions with various mutations in the ADAR deaminase domain were assayed on the luciferase target.

[01253] As shown in FIG. 77, R455H and S458F mutants each exhibited increased efficiency and specificity compared to REPAIRv1 (dCas13b-ADAR2$_{DD}$(E488Q)).

[01254] As shown in FIG. 79, H460P mutant exhibited increased efficiency compared to REPAIRv2 (dCas13b-ADAR2$_{DD}$(E488Q/T375G)) and increased specificity compared to REPAIRv1.H460I and A476E mutants each exhibited increased efficiency and specificity compared to REPAIRv1, and increased efficiency compared to REPAIRv2.

[01255] As shown in FIG. 81, V351Y, V351M and V351T mutants each exhibited increased specificity compared to REPAIRv1 at similar efficiency, and increased efficiency compared to REPAIRv2 at similar specificity.

[01256] As shown in FIG. 82, T375H, T375C and T375Q mutants each exhibited increased specificity compared to REPAIRv1 at similar efficiency, and increased efficiency compared to REPAIRv2 at similar specificity.

[01257] As shown in FIG. 83, R455H mutant exhibited increased specificity compared to REPAIRv1 at similar efficiency, and increased efficiency compared to REPAIRv2 at similar specificity.

[01258] As shown in FIG. 84, V351Y, V351M, V351T, T375H, T375C, T375Q, G478R, S485F, and H460I mutants each exhibited increased specificity compared to REPAIRv1, and increased efficiency compared to REPAIRv2. pMAX was used as a GFP control.

[01259] As shown in FIG. 86, V351Y, V351M, T375H, T375Q, and H460P mutants each exhibited increased specificity compared to REPAIRv1, and increased efficiency compared to REPAIRv2.

[01260] As shown in FIGs. 89-90, a number of combination mutants exhibited increased specificity compared to REPAIRv1, and increased efficiency compared to

REPAIRv2. Among them, T375S/S458F combination mutant exhibited both increased efficiency and increased specificity compared to REPAIRv1, as well as increased efficiency compared to REPAIRv2.

EXAMPLE 5 – ADAR Mutants Having C-to-U Deamination Activity

[01261] To identifying ADAR mutants having C-to-U deamination activity, Cas13b-ADAR fusions with various mutations in the ADAR deaminase domain were assayed on the luciferase target.

[01262] As shown in FIGs. 96-97, a number of V351, T375 and R455 mutants were tested for their ability to catalyze C-to-U deamination activity, and certain V351 mutants having C-to-U activity were further validated. Guide sequence used in contruct guide-paring showin FIG. 95 are provided below

		Guide sequence (SEQ ID Nos. 774-781)	Guide seq + G (SEQ ID Nos. 782-789)	Revcom (SEQ ID Nos. 790-797)	Top order (SEQ ID Nos. 709-805)	Bottom order (SEQ ID Nos. 806-813)	
GlucA	50	ttcatcttgggcgtgcActtgatgtgggacaggcagatcagacagcccct	GttcatcttgggcgtgcActtgatgtgggacaggcagatcagacagcccct	aggggctgtctgatctgcctgtcccacatcaagtgcacgcccaagatgaac	caccGttcatcttgggcgtgcActtgatgtgggacaggcagatcagacagcccct	caacaggggctgtctgatctgcctgtcccacatcaagtgcacgcccaagatgaac	
GlucC	50	ttcatcttgggcgtgcCcttgatgtgggacaggcagatcagacagcccct	GttcatcttgggcgtgcCcttgatgtgggacaggcagatcagacagcccct	aggggctgtctgatctgcctgtcccacatcaagggcacgcccaagatgaac	caccGttcatcttgggcgtgcCcttgatgtgggacaggcagatcagacagcccct	caacaggggctgtctgatctgcctgtcccacatcaaggcacgcccaagatgaac	
GlucT	50	ttcatcttgggcgtgcTcttgatgtgggacaggcagatcagacagcccct	GttcatcttgggcgtgcTcttgatgtgggacaggcagatcagacagcccct	aggggctgtctgatctgcctgtcccacatcaagagcacgcccaagatgaac	caccGttcatcttgggcgtgcTcttgatgtgggacaggcagatcagacagcccct	caacaggggctgtctgatctgcctgtcccacatcaagagcacgcccaagatgaac	
GlucG	50	ttcatcttgggcgtgcGcttgatgtgggacaggcagatcagacagcccct	GttcatcttgggcgtgcGcttgatgtgggacaggcagatcagacagcccct	aggggctgtctgatctgcctgtcccacatcaagcgcacgcccaagatgaac	caccGttcatcttgggcgtgcGcttgatgtgggacaggcagatcagacagcccct	caacaggggctgtctgatctgcctgtcccacatcaagcgcacgcccaagatgaac	
GlucA	30	gcActtgatgtgggacaggcagatcagaca	GgcActtgatgtggacaggcagatcagaca	tgtctgatctgcctgtccacatcaagtgcc	caccGgcActtgatgtgggacaggcagatcagaca	caactgtctgatctgcctgtcccacatcaagtgcc	
GlucC	30	gcCcttgatgtgggacaggcagatcagaca	GgcCcttgatgtggacaggcagatcagaca	tgtctgatctgcctgtccacatcaagggcc	caccGgcCcttgatgtgggacaggcagatcagaca	caactgtctgatctgcctgtcccacatcaagggcc	
GlucT	30	gcTcttgatgtgggacaggcagatcagaca	GgcTcttgatgtggacaggcagatcagaca	tgtctgatctgcctgtccacatcaagagcc	caccGgcTcttgatgtgggacaggcagatcagaca	caactgtctgatctgcctgtcccacatcaagagcc	
GlucG	30	gcGcttgatgtgggacaggcagatcagaca	GgcGcttgatgtggacaggcagatcagaca	tgtctgatctgcctgtccacatcaagcgcc	caccGgcGcttgatgtgggacaggcagatcagaca	caactgtctgatctgcctgtcccacatcaagcgcc	

PC0715	caccGttcatcttgggcgtgcActtgatgtgggacaggcagatcagacagcccct (SEQ ID No. 814)	Guide targeting pAB0212 50bp A flip F v2

PC0716	caccGttcatcttgggcgtgcCcttgatgtgggacaggcagatcagacagcccct (SEQ ID No. 815)	Guide targeting pAB0212 50bp C flip F v2
PC0717	caccGttcatcttgggcgtgcTcttgatgtgggacaggcagatcagacagcccct (SEQ ID No. 816)	Guide targeting pAB0212 50bp T flip F v2
PC0718	caccGttcatcttgggcgtgcGcttgatgtgggacaggcagatcagacagcccct (SEQ ID No. 817)	Guide targeting pAB0212 50bp G flip F v2
PC0719	caccGgcActtgatgtgggacaggcagatcagaca(SEQ ID No. 818)	Guide targeting pAB0212 30bp A flip F v2
PC0720	caccGgcCcttgatgtgggacaggcagatcagaca(SEQ ID No. 819)	Guide targeting pAB0212 30bp C flip F v2
PC0721	caccGgcTcttgatgtgggacaggcagatcagaca(SEQ ID No. 820)	Guide targeting pAB0212 30bp T flip F v2
PC0722	caccGgcGcttgatgtgggacaggcagatcagaca(SEQ ID No. 821)	Guide targeting pAB0212 30bp G flip F v2
PC0723	Caacaggggctgtctgatctgcctgtcccacatcaagtgcacgcccaagatgaac (SEQ ID No. 822)	Guide targeting pAB0212 50bp A flip R v2
PC0724	Caacaggggctgtctgatctgcctgtcccacatcaagggcacgcccaagatgaac (SEQ ID No. 823)	Guide targeting pAB0212 50bp C flip R v2
PC0725	Caacaggggctgtctgatctgcctgtcccacatcaagagcacgcccaagatgaac (SEQ ID No. 824)	Guide targeting pAB0212 50bp T flip R v2
PC0726	Caacaggggctgtctgatctgcctgtcccacatcaagcgcacgcccaagatgaac (SEQ ID No. 825)	Guide targeting pAB0212 50bp G flip R v2
PC0727	Caactgtctgatctgcctgtcccacatcaagtgcc (SEQ ID No. 826)	Guide targeting pAB0212 30bp A flip R v2
PC0728	Caactgtctgatctgcctgtcccacatcaagggcc (SEQ ID No. 827)	Guide targeting pAB0212 30bp C flip R v2
PC0729	Caactgtctgatctgcctgtcccacatcaagagcc (SEQ ID No. 828)	Guide targeting pAB0212 30bp T flip R v2
PC0730	caactgtctgatctgcctgtcccacatcaagcgcc(SEQ ID No. 829)	Guide targeting pAB0212 30bp G flip R v2

[01263] Guide sequences used in contruct guide-paring show in FIG. 99 are provided below.

Position	Guide sequence (SEQ ID Nos. 830-845)	Guide seq + G (SEQ ID Nos. 846-861)	Revcom (SEQ ID Nos 862 – 877)	Top order (SEQ ID Nos. 878-893)	Bottom order (SEQ ID Nos. 894-909)
0	CTCTTTGTCGCCTTCGTAGGTGTGGCAGCG	GCTCTTTGTCGCCTTCGTAGGTGTGGCAGCG	cgctgccacacctacgaaggcgacaaagagc	caccGCTCTTTGTCGCCTTCGTAGGTGTGGCAGCG	caaccgctgccacacctacgaaggcgacaaagagc
6	GTCGCCTTCGTAGGTGTGGCAGCGTCCTGG	GGTCGCCTTCGTAGGTGTGGCAGCGTCCTGG	ccaggacgctgccacacctacgaaggcgacc	caccGGTCGCCTTCGTAGGTGTGGCAGCGTCCTGG	caacccaggacgctgccacacctacgaaggcgacc
12	TTCGTAGGTGTGGCAGCGTCCTGGGATGAA	GTTCGTAGGTGTGGCAGCGTCCTGGGATGAA	ttcatcccaggacgctgccacacctacgaac	caccGTTCGTAGGTGTGGCAGCGTCCTGGGATGAA	caacttcatcccaggacgctgccacacctacgaac
18	GGTGTGGCAGCGTCCTGGGATGAACTTCTT	GGGTGTGGCAGCGTCCTGGGATGAACTTCTT	aagaagttcatcccaggacgctgccacaccc	caccGGGTGTGGCAGCGTCCTGGGATGAACTTCTT	caacaagaagttcatcccaggacgctgccacacc
24	GCAGCGTCCTGGGATGAACTTCTTCATCTT	GGCAGCGTCCTGGGATGAACTTCTTCATCTT	aagatgaagaagttcatcccaggacgctgcc	caccGGCAGCGTCCTGGGATGAACTTCTTCATCTT	caacaagatgaagaagttcatcccaggacgctgcc
30	TCCTGGGATGAACTTCTTCATCTTGGCGT	GTCCTGGGATGAACTTCTTCATCTTGGGCGT	acgcccaagatgaagaagttcatcccaggac	caccGTCCTGGGATGAACTTCTTCATCTTGGGCGT	caacacgcccaagatgaagaagttcatcccaggac
36	GATGAACTTCTTCATCTTGGGCGTGCGCTT	GGATGAACTTCTTCATCTTGGGCGTGCGCTT	aagcgcacgcccaagatgaagaagttcatcc	caccGGATGAACTTCTTCATCTTGGGCGTGCGCTT	caacaagcgcacgcccaagatgaagaagttcatcc

42	CTTCTTCATCTTGGGCGTGCGCTTGATGTG	GCTTCTTCATCTTGGGCGTGCGCTTGATGTG	cacatcaagcgcacgcccaagatgaagaagc	caccGCTTCTTCATCTTGGGCGTGCGCTTGATGTG	caaccacatcaagcgcacgcccaagatgaagaagc
48	CATCTTGGGCGTGCGCTTGATGTGGGACAG	GCATCTTGGGCGTGCGCTTGATGTGGGACAG	ctgtcccacatcaagcgcacgcccaagatgc	caccGCATCTTGGGCGTGCGCTTGATGTGGGACAG	caacctgtcccacatcaagcgcacgcccaagatgc
54	GGGCGTGCGCTTGATGTGGGACAGGCAGAT	GGGGCGTGCGCTTGATGTGGGACAGGCAGAT	atctgcctgtcccacatcaagcgcaccgcccc	caccGGGGCGTGCGCTTGATGTGGGACAGGCAGAT	caacatctgcctgtcccacatcaagcgcacgccc
60	GCGCTTGATGTGGGACAGGCAGATCAGACA	GGCGCTTGATGTGGGACAGGCAGATCAGACA	tgtctgatctgcctgtcccacatcaagcgcc	caccGGCGCTTGATGTGGGACAGGCAGATCAGACA	caactgtctgatctgcctgtcccacatcaagcgcc
66	GATGTGGGACAGGCAGATCAGACAGCCCCT	GGATGTGGGACAGGCAGATCAGACAGCCCCT	aggggctgtctgatctgcctgtcccacatcc	caccGGATGTGGGACAGGCAGATCAGACAGCCCCT	caacaggggctgtctgatctgcctgtcccacatcc
72	GGACAGGCAGATCAGACAGCCCCTGGTGCA	GGGACAGGCAGATCAGACAGCCCCTGGTGCA	tgcaccaggggctgtctgatctgcctgtccc	caccGGGACAGGCAGATCAGACAGCCCCTGGTGCA	caactgcaccaggggctgtctgatctgcctgtccc
78	GCAGATCAGACAGCCCCTGGTGCAGCCAGC	GGCAGATCAGACAGCCCCTGGTGCAGCCAGC	gctggctgcaccaggggctgtctgatctgcc	caccGGCAGATCAGACAGCCCCTGGTGCAGCCAGC	caacgctggctgcaccaggggctgtctgatctgcc
84	CAGACAGCCCCTGGTGCAGCCAGCTTTCCG	GCAGACAGCCCCTGGTGCAGCCAGCTTTCCG	cggaaagctggctgcaccaggggctgtctgc	caccGCAGACAGCCCCTGGTGCAGCCAGCTTTCCG	caaccggaaagctggctgcaccaggggctgtctgc
NT	GTAATGCCTGGCTTGTCGACGCATAGTCTG	GGTAATGCCTGGCTTGTCGACGCATAGTCTG	cagactatgcgtcgacaagccaggcattacc	caccGGTAATGCCTGGCTTGTCGACGCATAGTCTG	caaccagactatgcgtcgacaagccaggcattacc

References

1. P. D. Hsu, E. S. Lander, F. Zhang, Development and applications of CRISPR-Cas9 for genome engineering. Cell 157, 1262-1278 (2014).
2. A. C. Komor, A. H. Badran, D. R. Liu, CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell 168, 20-36 (2017).
3. L. Cong et al., Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823 (2013).
4. P. Mali et al., RNA-guided human genome engineering via Cas9. Science 339, 823-826 (2013).
5. B. Zetsche et al., Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163, 759-771 (2015).
6. H. Kim, J. S. Kim, A guide to genome engineering with programmable nucleases. Nat Rev Genet 15, 321-334 (2014).
7. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420-424 (2016).
8. K. Nishida et al., Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science 353, (2016).
9. Y. B. Kim et al., Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nat Biotechnol 35, 371-376 (2017).
10. O. O. Abudayyeh et al., C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science 353, aaf5573 (2016).
11. S. Shmakov et al., Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Mol Cell 60, 385-397 (2015).

12. S. Shmakov et al., Diversity and evolution of class 2 CRISPR-Cas systems. Nat Rev Microbiol 15, 169-182 (2017).
13. A. A. Smargon et al., Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. Mol Cell 65, 618-630 e617 (2017).
14. J. S. Gootenberg et al., Nucleic acid detection with CRISPR-Cas13a/C2c2. Science 356, 438-442 (2017).
15. O. O. Abudayyeh et al., RNA targeting with CRISPR-Cas13a. Nature in press, (2017).
16. K. Nishikura, Functions and regulation of RNA editing by ADAR deaminases. Annu Rev Biochem 79, 321-349 (2010).
17. M. H. Tan *et al.*, Dynamic landscape and regulation of RNA editing in mammals. *Nature* **550**, 249-254 (2017)
18. B. L. Bass, H. Weintraub, An unwinding activity that covalently modifies its double-stranded RNA substrate. Cell 55, 1089-1098 (1988).
19. M. M. Matthews et al., Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. Nat Struct Mol Biol 23, 426-433 (2016).
20. Y. Zheng, C. Lorenzo, P. A. Beal, DNA editing in DNA/RNA hybrids by adenosine deaminases that act on RNA. Nucleic Acids Res 45, 3369-3377 (2017).
21. A. Kuttan, B. L. Bass, Mechanistic insights into editing-site specificity of ADARs. Proc Natl Acad Sci U S A 109, E3295-3304 (2012).
22. S. K. Wong, S. Sato, D. W. Lazinski, Substrate recognition by ADAR1 and ADAR2. RNA 7, 846-858 (2001).
23. M. Fukuda et al., Construction of a guide-RNA for site-directed RNA mutagenesis utilising intracellular A-to-I RNA editing. Sci Rep 7, 41478 (2017).
24. M. F. Montiel-Gonzalez, I. Vallecillo-Viejo, G. A. Yudowski, J. J. Rosenthal, Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. Proc Natl Acad Sci U S A 110, 18285-18290 (2013).
25. M. F. Montiel-Gonzalez, I. C. Vallecillo-Viejo, J. J. Rosenthal, An efficient system for selectively altering genetic information within mRNAs. Nucleic Acids Res 44, e157 (2016).
26. J. Wettengel, P. Reautschnig, S. Geisler, P. J. Kahle, T. Stafforst, Harnessing human ADAR2 for RNA repair - Recoding a PINK1 mutation rescues mitophagy. Nucleic Acids Res 45, 2797-2808 (2017).
27. Y. Wang, J. Havel, P. A. Beal, A Phenotypic Screen for Functional Mutants of Human Adenosine Deaminase Acting on RNA 1. ACS Chem Biol 10, 2512-2519 (2015).
28. K. A. Lehmann, B. L. Bass, Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. Biochemistry 39, 12875-12884 (2000).
29. T. Stafforst, M. F. Schneider, An RNA-deaminase conjugate selectively repairs point mutations. *Angew Chem Int Ed Engl* **51**, 11166-11169 (2012).
30. C. Ballatore, V. M. Lee, J. Q. Trojanowski, Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. Nat Rev Neurosci 8, 663-672 (2007).
31. Y. Li et al., Carriers of rare missense variants in IFIH1 are protected from psoriasis. J Invest Dermatol 130, 2768-2772 (2010).
32. R. S. Finkel et al., Treatment of infantile-onset spinal muscular atrophy with nusinersen: a phase 2, open-label, dose-escalation study. Lancet 388, 3017-3026 (2016).
33. J. Joung *et al.*, Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat Protoc* **12**, 828-863 (2017).

34. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
35. E. Picardi, A. M. D'Erchia, A. Montalvo, G. Pesole, Using REDItools to Detect RNA Editing Events in NGS Datasets. *Curr Protoc Bioinformatics* **49**, 12 12 11-15 (2015).
36. E. Picardi, G. Pesole, REDItools: high-throughput RNA editing detection made easy. *Bioinformatics* **29**, 1813-1814 (2013).
37. G. Glusman, J. Caballero, D. E. Mauldin, L. Hood, J. C. Roach, Kaviar: an accessible system for testing SNV novelty. *Bioinformatics* **27**, 3216-3217 (2011).
38. J. D. Watson, *Molecular biology of the gene*. (Pearson, Boston, ed. Seventh edition, 2014), pp. xxxiv, 872 pages.
39. R. C. Ferreira *et al.*, Association of IFIH1 and other autoimmunity risk alleles with selective IgA deficiency. *Nat Genet* **42**, 777-780 (2010).

[01264] The invention further relates to the following aspects, which are described hereinbelow as numbered statements:

1. A method of modifying an Adenine in a target RNA sequence of interest, comprising delivering to said target RNA:

(a) a catalytically inactive (dead) Cas13 protein;

(b) a guide molecule which comprises a guide sequence linked to a direct repeat sequence; and

(c) an adenosine deaminase protein or catalytic domain thereof;

wherein said adenosine deaminase protein or catalytic domain thereof is covalently or non-covalently linked to said dead Cas13 protein or said guide molecule or is adapted to link thereto after delivery;

wherein guide molecule forms a complex with said dead Cas13 protein and directs said complex to bind said target RNA sequence of interest, wherein said guide sequence is capable of hybridizing with a target sequence comprising said Adenine to form an RNA duplex, wherein said guide sequence comprises a non-pairing Cytosine at a position corresponding to said Adenine resulting in an A-C mismatch in the RNA duplex formed;

wherein said adenosine deaminase protein or catalytic domain thereof deaminates said Adenine in said RNA duplex.

2. The method of statement 1, wherein said Cas13 protein is Cas13a, Cas13b or Cas 13c.

3. The method of statement 1, wherein said adenosine deaminase protein or catalytic domain thereof is fused to N- or C-terminus of said dead Cas13 protein.

4. The method of statement 3, wherein said adenosine deaminase protein or catalytic domain thereof is fused to said dead Cas13 protein by a linker.

5. The method of statement 4, wherein said linker is $(GGGGS)_{3-11}$, GSG_5 or LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR, or wherein said linker is an XTEN linker.

6. The method of statement 1, wherein said adenosine deaminase protein or catalytic domain thereof is linked to an adaptor protein and said guide molecule or said dead Cas13 protein comprises an aptamer sequence capable of binding to said adaptor protein.

7. The method of statement 6, wherein said adaptor sequence is selected from MS2, PP7, Qβ, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕCb5, ϕCb8r, ϕCb12r, ϕCb23r, 7s and PRR1.

8. The method of statement 1, wherein said adenosine deaminase protein or catalytic domain thereof is inserted into an internal loop of said dead Cas13 protein.

9. The method of statement 8, wherein said Cas13 protein is a Cas13a protein and said Cas13a comprises one or more mutations the two HEPN domains, particularly at postion R474 and R1046 of Cas 13a protein originating from *Leptotrichia wadei* or amino acid positions corresponding thereto of a Cas13a ortholog.

10. The method of statement 9, wherein said Cas13 protein is a Cas13b protein and said Cas13b comprises a mutation in one or more of positions R116, H121, R1177, H1182 of Cas13b protein originating from *Bergeyella zoohelcum* ATCC 43767 or amino acid positions corresponding thereto of a Cas13b ortholog.

11. The method of statement, wherein said mutation is one or more of R116A, H121A, R1177A, H1182A of Cas13b protein originating from *Bergeyella zoohelcum* ATCC 43767 or amino acid positions corresponding thereto of a Cas13b ortholog.

12. The method of any of statement 1, wherein said guide sequence has a length of about 20-53 nt, preferably 25-53 nt, more preferably 29-53 nt capable of forming said RNA duplex with said target sequence.

13. The method of any of statement 12, wherein said guide sequence has a length of about 40-50 nt capable of forming said RNA duplex with said target sequence.

14. The method of statement 1, wherein the distance between said non-pairing C and the 5' end of said guide sequence is 20-30 nucleotides.

15. The method of any of the preceding statements, wherein said adenosine deaminase protein or catalytic domain thereof is a human, cephalopod, or Drosophila adenosine deaminase protein or catalytic domain thereof.

16. The method of statement 1, wherein said adenosine deaminase protein or catalytic domain thereof has been modified to comprise a mutation at glutamic acid^{488} of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein.

17. The method of statement 16, wherein said glutamic acid residue at position 488 or a corresponding position in a homologous ADAR protein is replaced by a glutamine residue (E488Q).

18. The method of statement 16 or 17, wherein said adenosine deaminase protein or catalytic domain thereof is a mutated hADAR2d comprising mutation E488Q or a mutated hADAR1d comprising mutation E1008Q.

19. The method of any of the preceding statements wherein the guide sequence comprises more than one mismatch corresponding to different adenosine sites in the target RNA sequence or wherein two guide molecules are used, each comprising a mismatch corresponding to a different adenosine sites in the target RNA sequence.

20. The method of any of the preceding statements, wherein said Cas13 protein and optionally said adenosine deaminase protein or catalytic domain thereof comprise one or more heterologous nuclear localization signal(s) (NLS(s)).

21. The method of any of the preceding statements, wherein said method comprises, determining said target sequence of interest and selecting said adenosine deaminase protein or catalytic domain thereof which most efficiently deaminates said Adenine present in said target sequence.

22. The method of any of the preceding statements, wherein said catalytically inactive Cas13 protein is obtained from a Cas13 nuclease derived from a bacterial species selected from the group consisting of the bacterial species listed in any of Tables 1, 2, 3, or 4.

23. The method of any of the preceding statements, wherein said target RNA sequence of interest is within a cell.

24. The method of statement 23, wherein said cell is a eukaryotic cell.

25. The method of statement 24, wherein said cell is a non-human animal cell.

26. The method of statement 24, wherein said cell is a human cell.

27. The method of statement 24, wherein said cell is a plant cell.

28. The method of any of the preceding statements, wherein said target locus of interest is within an animal.

29. The method of any of the preceding statements, wherein said target locus of interest is within a plant.

30. The method of any of the preceding statements, wherein said target RNA sequence of interest is comprised in an RNA polynucleotide *in vitro*.

31. The method of any of the preceding statements, wherein said components (a), (b) and (c) are delivered to said cell as a ribonucleoprotein complex.

32. The method of any of the preceding statements, wherein said components (a), (b) and (c) are delivered to said cell as one or more polynucleotide molecules.

33. The method of statement 32, wherein said one or more polynucleotide molecules comprise one or more mRNA molecules encoding components (a) and/or (c).

34. The method of statement 33, wherein said one or more polynucleotide molecules are comprised within one or more vectors.

35. The method of statement 34, wherein said one or more polynucleotide molecules comprise one or more regulatory elements operably configured to express said Cas13 protein, said guide molecule, and said adenosine deaminase protein or catalytic domain thereof, optionally wherein said one or more regulatory elements comprise inducible promoters.

36. The method of statement 32, wherein said one or more polynucleotide molecules or said ribonucleoprotein complex are delivered via particles, vesicles, or one or more viral vectors.

37. The method of statement 36, wherein said particles comprise a lipid, a sugar, a metal or a protein.

38. The method of statement 37, wherein said particles comprise lipid nanoparticles.

39. The method of statement 36, wherein said vesicles comprise exosomes or liposomes.

40. The method of statement 34, wherein said one or more viral vectors comprise one or more of adenovirus, one or more lentivirus or one or more adeno-associated virus.

41. The method of any of the preceding statements, where said method modifies a cell, a cell line or an organism by manipulation of one or more target RNA sequences.

42. The method of statement 41, wherein said deamination of said Adenine in said target RNA of interest remedies a disease caused by transcripts containing a pathogenic G→A or C→T point mutation.

43. The method of statement 42, wherein said disease is selected from Meier-Gorlin syndrome, Seckel syndrome 4, Joubert syndrome 5, Leber congenital amaurosis 10; Charcot-Marie-Tooth disease, type 2; Charcot-Marie-Tooth disease, type 2; Usher syndrome, type 2C; Spinocerebellar ataxia 28; Spinocerebellar ataxia 28; Spinocerebellar ataxia 28; Long QT syndrome 2; Sjögren-Larsson syndrome; Hereditary fructosuria; Hereditary fructosuria; Neuroblastoma; Neuroblastoma; Kallmann syndrome 1; Kallmann syndrome 1; Kallmann syndrome 1; Metachromatic leukodystrophy, Rett syndrome, Amyotrophic lateral sclerosis type 10, Li-Fraumeni syndrome, or a disease listed in Table 5.

44. The method of statement 42, wherein said disease is a premature termination disease.

45. The method of statement 41, wherein said modification affects the fertility of an organism.

46. The method of statement 41, wherein said modification affects splicing of said target RNA sequence.

47. The method of statement 41, wherein said modification introduces a mutation in a transcript introducing an amino acid change and causing expression of a new antigen in a cancer cell.

48. The method of statement 41, wherein said target RNA is comprised within a microRNA.

49. The method of statement 41, wherein said deamination of said Adenine in said target RNA of interest causes a gain of function or a loss of function of a gene.

50. The method of statement 49, wherein said gene is a gene expressed by a cancer cell.

51. A modified cell obtained from the method of any of the preceding statements, or progeny of said modified cell, wherein said cell comprises a hypoxanthine or a guanine in replace of said Adenine in said target RNA of interest compared to a corresponding cell not subjected to the method.

52. The modified cell or progeny thereof of statement 51, wherein said cell is a eukaryotic cell.

53. The modified cell or progeny thereof of statement 51, wherein said cell is an animal cell.

54. The modified cell or progeny thereof of statement 51, wherein said cell is a human cell.

55. The modified cell or progeny thereof of statement 51, wherein said cell is a therapeutic T cell.

56. The modified cell or progeny thereof of statement 51, wherein said cell is an antibody-producing B cell.

57. The modified cell or progeny thereof of statement 51, wherein said cell is a plant cell.

58. A non-human animal comprising said modified cell of statement 51.

59. A plant comprising said modified cell of statement 58.

60. A method for cell therapy, comprising administering to a patient in need thereof said modified cell of any of statements 51-55, wherein presence of said modified cell remedies a disease in the patient.

61. An engineered, non-naturally occurring system suitable for modifying an Adenine in a target locus of interest, comprising

a) a guide molecule which comprises a guide sequence linked to a direct repeat sequence, or a nucleotide sequence encoding said guide molecule;
b) a catalytically inactive Cas13 protein, or a nucleotide sequence encoding said catalytically inactive Cas13 protein;
c) an adenosine deaminase protein or catalytic domain thereof, or a nucleotide sequence encoding said adenosine deaminase protein or catalytic domain thereof;

wherein said adenosine deaminase protein or catalytic domain thereof is covalently or non-covalently linked to said Cas13 protein or said guide molecule or is adapted to link thereto after delivery;

wherein said guide sequence is capable of hybridizing with a target RNA sequence comprising an Adenine to form an RNA duplex, wherein said guide sequence comprises a non-pairing Cytosine at a position corresponding to said Adenine resulting in an A-C mismatch in the RNA duplex formed.

62. An engineered, non-naturally occurring vector system suitable for modifying an Adenine in a target locus of interest, comprising the nucleotide sequences of a), b) and c) of statement 61.

63. The engineered, non-naturally occurring vector system of statement 62, comprising one or more vectors comprising:

a) a first regulatory element operably linked to a nucleotide sequence encoding said guide molecule which comprises said guide sequence,

b) a second regulatory element operably linked to a nucleotide sequence encoding said catalytically inactive Cas13 protein; and

c) a nucleotide sequence encoding an adenosine deaminase protein or catalytic domain thereof which is under control of said first or second regulatory element or operably linked to a third regulatory element;

wherein, if said nucleotide sequence encoding an adenosine deaminase protein or catalytic domain thereof is operably linked to a third regulatory element, said adenosine deaminase protein or catalytic domain thereof is adapted to link to said guide molecule or said Cas13 protein after expression;

wherein components (a), (b) and (c) are located on the same or different vectors of the system.

64. An *in vitro* or *ex vivo* host cell or progeny thereof or cell line or progeny thereof comprising the system of any of statements 61-63.

65. The host cell or progeny thereof or cell line or progeny thereof of statement 64, wherein said cell is a eukaryotic cell.

66. The host cell or progeny thereof or cell line or progeny thereof of statement 64, wherein said cell is an animal cell.

67. The host cell or progeny thereof or cell line or progeny thereof of statement 64, wherein said cell is a human cell.

68. The host cell or progeny thereof or cell line or progeny thereof of statement 64, wherein said cell is a plant cell.

69. The method according to statement 1, wherein said cytosine is not 5' flanked by guanosine.

70. The method according to statement 1, wherein said adenosine deaminase is ADAR, optionally huADAR, optionally (hu)ADAR1 or (hu)ADAR2, preferably huADAR2.

71. The method according to statement 1, wherein said Cas13, preferably Cas13b, is truncated, preferably C-terminally truncated, preferably wherein said Cas 13 is a truncated functional variant of the corresponding wild type Cas13.

72. The method according to statement 1, wherein said adenosine deaminase is capable of deaminating adenosine in RNA or is an RNA specific adenosine deaminase.

73. The method according to statement 1 or 16, wherein said adenosine deaminase protein or catalytic domain thereof has been modified to comprise one or more mutation of the ADAR, preferably a mutation as described herein, for instance a mutation as provided in any of Figures 43-47, or a corresponding mutation in an ADAR homologue or orthologue.

***

[01265] The present disclosure is not to be limited in terms of the particular embodiments described in this application. Many modifications and variations can be made without departing from its spirit and scope, as will be apparent to those skilled in the art. Functionally equivalent methods and compositions within the scope of the disclosure, in addition to those enumerated herein, will be apparent to those skilled in the art from the foregoing descriptions. Such modifications and variations are intended to fall within the scope of the appended claims. The present disclosure is to be limited only by the terms of the appended claims, along with the full scope of equivalents to which such claims are entitled. It is to be understood that this disclosure is not limited to particular methods, reagents, compounds compositions or biological systems, which can of course vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting.

2018290843 17 Sep 2024

[01266] In addition, where features or aspects of the disclosure are described in terms of Markush groups, those skilled in the art will recognize that the disclosure is also thereby described in terms of any individual member or subgroup of members of the Markush group.

[01267] Other embodiments are set forth in the following claims.

[01268] The discussion of documents, acts, materials, devices, articles and the like is included in this specification solely for the purpose of providing a context for the present invention. It is not suggested or represented that any or all of these matters formed part of the prior art base or were common general knowledge in the field relevant to the present invention as it existed before the priority date of each claim of this application.

[01269] Unless the context requires otherwise, where the terms "comprise", "comprises", "comprised" or "comprising" are used in this specification (including the claims) they are to be interpreted as specifying the presence of the stated features, integers, steps or components, but not precluding the presence of one or more other features, integers, steps or components, or group thereof.

2018290843 07 Apr 2025

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. An engineered composition for site directed base editing comprising a Cas13 and an adenosine deaminase, or catalytic domain thereof, wherein the adenosine deaminase comprises one or more mutations that changes the functionality of the adenosine deaminase relative to wild type, wherein the one or more mutations comprise:

(i) one or more mutations at one or more amino acid residues selected from T375, a S486, and a V351 in a human adenosine deaminase acting on RNA 2 (ADAR2) or deaminase domain thereof comprising an amino acid sequence SEQ ID NO: 650 or in a corresponding position thereto in a structural or functional homolog or ortholog ADAR or adenosine deaminase domain thereof,

or

(ii) an E448Q mutation and one or more mutations at one or more amino acid residues selected from T375, a S486, and a V351 in a human ADAR2 or deaminase domain thereof comprising an amino acid sequence SEQ ID NO: 650 or in a corresponding position thereto in a structural or functional homolog or ortholog ADAR or adenosine deaminase domain thereof, and

the mutations in (i) or (ii) changes the ability of the adenosine deaminase to deaminate cytosine.

2. The composition of claim 1, wherein the adenosine deaminase, or catalytic domain thereof, comprises one or more mutations that increase activity or specificity of the adenosine deaminase relative to wild type.

3. The composition of claim 1 or claim 2, wherein (i) the one or more mutations comprise one or more of a T375S mutation, a S486A mutation, a V351G mutation in a human ADAR2 or deaminase domain thereof comprising an amino acid sequence SEQ ID NO: 650 or in a corresponding position thereto in a structural or functional homolog or ortholog ADAR or adenosine deaminase domain thereof or, (ii) the ADAR or adenosine deaminase domain thereof comprises an E448Q mutation and the one or more mutations comprise one or more of T375S mutation, a S486A mutation, a V351G mutation or any combination thereof in a human ADAR2 or deaminase domain thereof comprising an amino acid sequence SEQ ID NO: 650 or in a corresponding position thereto in a structural or functional homolog or ortholog ADAR or adenosine deaminase domain thereof.

2018290843 07 Apr 2025

4. The composition of claim 3, wherein the CRISPR system is catalytically inactive.

5. The composition of claim 3 or 4, wherein the the Cas13 protein is Cas13a, Cas13b or Cas13c, preferably wherein said Cas13 a Cas13 listed in any of Tables 1, 2, 3, 4, or 6 or is from a bacterial species listed in any of Tables 1, 2, 3, 4, or 6, preferably wherein said Cas13 protein is Prevotella sp.P5-125 Cas13b, Porphyromas gulae Cas13b, or Riemerella anatipestifer Cas13b; preferably Prevotella sp.P5-125 Cas13b.

6. The composition of claim 3, 4 or 5, wherein said guide molecule comprises a guide sequence is capable of hybridizing with a target RNA sequence comprising an Adenine to form an RNA duplex, wherein said guide sequence comprises a non-pairing Cytosine at a position corresponding to said Adenine resulting in an A-C mismatch in the RNA duplex formed.

7. The composition of claim 5, wherein said Cas13 protein is a Cas13a protein and said Cas13a comprises one or more mutations the two HEPN domains, particularly at position R474 and R1046 of Cas13a protein originating from *Leptotrichia wadei* or amino acid positions corresponding thereto of a Cas13a ortholog, or wherein said Cas13 protein is a Cas13b protein and said Cas13b comprises a mutation in one or more of positions R116, H121, R1177, H1182, preferably R116A, H121A, R1177A, H1182A of Cas13b protein originating from *Bergeyella zoohelcum* ATCC 43767 or amino acid positions corresponding thereto of a Cas13b ortholog, or wherein said Cas13 protein is a Cas13b protein and said Cas13b comprises a mutation in one or more of positions R128, H133, R1053, H1058, preferably H133 and H1058, preferably H133A and H1058A, of a Cas13b protein originating from Prevotella sp. P5-125 or amino acid positions corresponding thereto of a Cas13b ortholog.

8. The composition of claim 5, wherein said Cas13, preferably Cas13b, is truncated, preferably C-terminally truncated, preferably wherein said Cas13 is a truncated functional variant of the corresponding wild type Cas13, optionally wherein said truncated Cas13b is encoded by nt 1-984 of Prevotella sp.P5-125 Cas13b or the corresponding nt of a Cas13b orthologue or homologue.

9. The composition of claim 5, wherein said Cas13 is a catalytically inactive Cas13, preferably Cas13b6.

2018290843 07 Apr 2025

10. The composition of claim 8, wherein said guide sequence has a length of about 20-53 nt, preferably 25-53 nt, more preferably 29-53 nt or 40-50 nt capable of forming said RNA duplex with said target sequence, and/or wherein the distance between said non-pairing C and the 5' end of said guide sequence is 20-30 nucleotides..

11. The composition of claims 10, wherein the guide sequence comprises more than one mismatch corresponding to different adenosine sites in the target RNA sequence or wherein two guide molecules are used, each comprising a mismatch corresponding to a different adenosine sites in the target RNA sequence.

12. The composition of any one of claims 1 to 11, wherein adenosine deaminase protein or catalytic domain thereof is fused to a N- or C-terminus of said oligonucleotide targeting protein, optionally by a linker, preferably where said linker is $(GGGGS)_{3-11}$, GSG_5 or LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR, or wherein said linker is an XTEN linker.

13. The composition of any one of claims 5 to 11, wherein said adenosine deaminase protein or catalytic domain thereof is inserted into an internal loop of said dead Cas13 protein.

14. The composition of any one of claims 5 to 11, wherein said adenosine deaminase protein or catalytic domain thereof is linked to an adaptor protein and said guide molecule or said dead Cas13 protein comprises an aptamer sequence capable of binding to said adaptor protein, preferably wherein said adaptor sequence is selected from MS2, PP7, Qβ, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕCb5, ϕCb8r, ϕCb12r, ϕCb23r, 7s and PRR1.

15. The composition of any one of claims 1 to 14, wherein said adenosine deaminase protein or catalytic domain thereof capable of deaminating adenosine or cytosine in RNA or is an RNA specific adenosine deaminase and/or is a bacterial, human, cephalopod, or Drosophila adenosine deaminase protein or catalytic domain thereof, preferably TadA, more preferably ADAR, optionally huADAR, optionally (hu)ADAR1 or (hu)ADAR2, preferably huADAR2 or catalytic domain thereof.

16. The composition of claim 15, wherein the ADAR protein is a mutated hADAR2d comprising mutation E488Q or a mutated hADAR1d comprising mutation E1008Q.

2018290843 07 Apr 2025

17. The composition of any one of claims 1 to 16, wherein said targeting domain and optionally said adenosine protein or catalytic domain thereof comprise one or more heterologous nuclear export signal(s) (NES(s)) or nuclear localization signal(s) (NLS(s)), preferably an HIV Rev NES or MAPK NES, preferably C-terminal.

18. The composition of any one of claims 1 to 17, wherein said target RNA sequence of interest is within a cell, preferably a eukaryotic cell, most preferably a human or non-human animal cell, or a plant cell.

19. The composition of any one of claims 1 to 18 for use in prophylactic or therapeutic treatment, preferably wherein said target locus of interest is within a human or animal.

20. A method of modifying an Adenine or Cytidine in a target RNA sequence of interest, comprising delivering to said target RNA, the composition according to any one of claims 1 to 19.

21. The method of claim 20, wherein the targeting domain comprises the CRISPR system of any one of claims 3 to 5, wherein said guide molecule forms a complex with said CRISPR effector protein and directs said complex to bind said target RNA sequence of interest, wherein said guide sequence is capable of hybridizing with a target sequence comprising said Adenine or Cytosine to form an RNA duplex; wherein said adenosine deaminase protein or catalytic domain thereof deaminates said Adenine or Cytosine in said RNA duplex.

22. The method of claim 20, wherein the CRISPR system comprises the Cas13 of any one of claims 5 to 19.

23. The method of claims 20 or 21, wherein the CRISPR system and the adenonsine deaminase, or catalytic domain thereof, are delivered as one or more polynucleotide molecules, as a ribonucleoprotein complex, optionally via particles, vesicles, or one or more viral vectors.

24. The method of anyone of claims 20 to 22 or the composition of any one of claims 1 to 19 is for use in the treatment or prevention of a disease caused by transcripts containing a pathogenic G→A or C→T point mutation.

25. An isolated cell obtained from the method of any one of claims 20 to 23 and/or comprising the composition of any one of claims 1-19, or progeny of said modified cell, preferably wherein said cell comprises a hypoxanthine or a guanine in replace of said

2018290843 07 Apr 2025

Adenine in said target RNA of interest compared to a corresponding cell not subjected to the method.

26. The cell or progeny thereof of claim 25, wherein said cell is a eukaryotic cell, preferably a human or non-human animal cell, optionally a therapeutic T cell or an antibody-producing B-cell or wherein said cell is a plant cell.

27. A non-human animal comprising said modified cell or progeny thereof of claim 25 or 26.

28. A plant comprising said modified cell or progeny thereof of claim 25.

29. A modified cell according to claim 25 or 26 for use in therapy, preferably cell therapy.

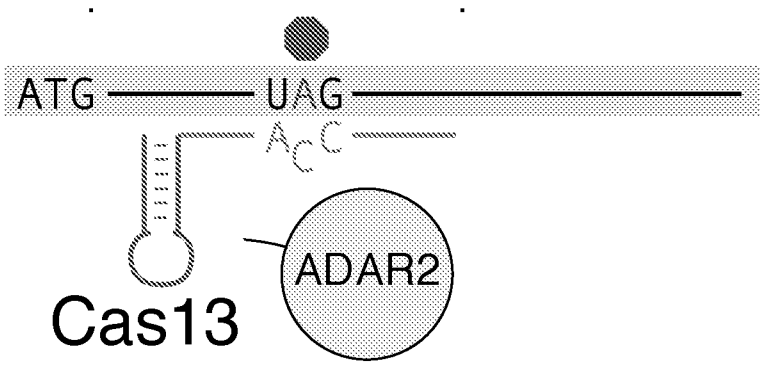


FIG. 1

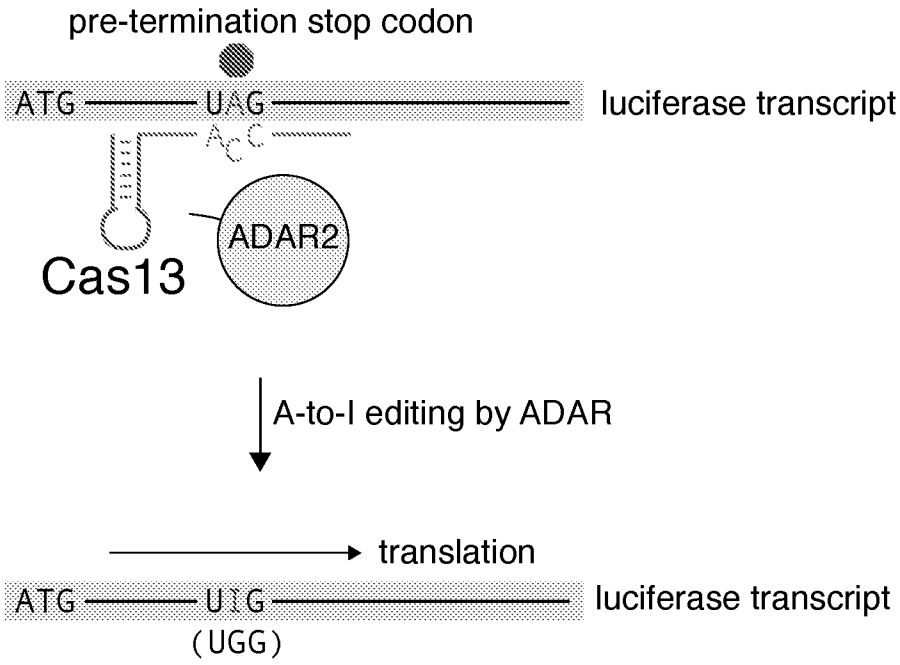

pre-termination stop codon
ATG
UAG
luciferase transcript
Cas13
ADAR2
A-to-I editing by ADAR
translation
ATG
UIG
(UGG)
luciferase transcript

FIG. 2

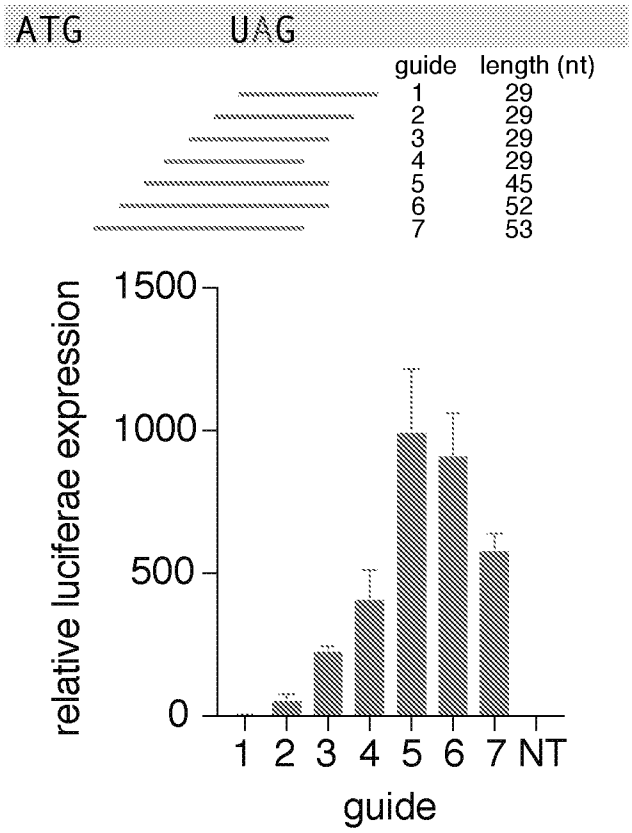

ATG
UAG
guide
length (nt)
1
29
2
29
3
29
4
29
5
45
6
52
7
53
relative luciferae expression
1500
1000
500
0
1 2 3 4 5 6 7 NT
guide

FIG. 3

ADAR Name	Protein Sequence
Homo sapiens_ADAR2_E_Q_Mutant (SEQ ID NO :650)	QLHLPQVLADAVSRLVLGKFGDLTDNFSSPHARRKVLAGVVMT TGTDVKDAKVISVSTGTKCINGEYMSDRGLALNDCHAEIISRRSL LRFLYTQLELYLNNKDDQKRSIFQKSERGGFRLKENVQFHLYIST SPCGDARIFSPHEPILEEPADRHPNRKARGQLRTKIESGQGTIPVR SNASIQTWDGVLQGERLLTMSCSDKIARWNVVGIQGSLLSIFVEP IYFSSIILGSLYHGDHLSRAMYQRISNIEDLPPLYTLNKPLLSGISN AEARQPGKAPNFSVNWTVGDSAIEVINATTGKDELGRASRLCKH ALYCRWMRVHGKVPSHLLRSKITKPNVYHESKLAAKEYQAAKA RLFTAFIKAGLGAWVEKPTEQDQFSLT*
Homo sapiens_ADAR1_E_Q_Mutant (SEQ ID NO :651)	SLGTGNRCVKGDSLSLKGETVNDCHAEIISRRGFIRFLYSELMKY NSQTAKDSIFEPAKGGEKLQIKKTVSFHLYISTAPCGDGALFDKS CSDRAMESTESRHYPVFENPKQGKLRTKVENGQGTIPVESSDIVP TWDGIRLGERLRTMSCSDKILRWNVLGLQGALLTHFLQPIYLKS VTLGYLFSQGHLTRAICCRVTRDGSAFEDGLRHPFIVNHPKVGR VSIYDSKRQSGKTKETSVNWCLADGYDLEILDGTRGTVDGPRNE LSRVSKKNIFLLFKKLCSFRYRRDLLRLSYGEAKKAARDYETAK NYFKKGLKDMGYGNWISKPQEEKNF
Octopus vulgaris_ADAR1_E_Q_Mutant (SEQ ID NO :652)	SVGTGNRCLTGDHLSLEGNSVNDSHAEMITRRGFLRYLYRHLLE YDAEVPNDLFEKGERSICRIKTNITFHLYISTAPCGDGALFSPRDT DSSNAKMEEENKHIHNPTFSSSVQGLLRTKVEGGQGTIPIDADFT EQTWDGIQRGERLRTMSCSDKICRWNVVGLQGALLSHFIEPIYL DSLTLGYLYDHGHLARAVCCRIERGEASVNQLLPEGYRLNHPW LGRVTACDPPRETQKTKSLSINWCYDDEKSEVLDGTAGICYTAIE KNLFSRLTKHNLYEEFKRVCRKFDRNDLLTAPSYNKAKMMATP FQTAKNVMLKKLKENNCGTWVSKPIEEEMF
Sepia_ADAR1_E_Q_Mutant (SEQ ID NO :653)	SVGTGNRCLTGDRLSLEGNSVNDSHAEMVTRRGFLRYLYKHLL EYDPEKPHDLFEKGERSLCRIKTNITFHLYISTAPCGDGALFSPRD TDSSNVKVDEENKHVHNPTFSSSVQGLLRTKVEGGQGTIPIDADF TEQTWDGIQRGERLRTMSCSDKICRWNVVGLQGALLSHFVEPIY LESLTLGYLYDHGHLARAVCCRIERGEASVNQLLPEGYRLNHPW LGRVTACDPPRETQKTKSLSINWCYDDEKSEVLDGTAGICYTAIE KNLFSRLTKHSLYEEFKKVCQKFEREDLLNVTSYNKAKMMAIPF QTAKNVMLKKLKENNCGTWVSKPIEEEMF

FIG. 4

Octopus vulgaris_ADAR2_E_Q_Mutant (SEQ ID NO :654)	GIGTGTKCINGEHMSDRGFGVNDCHAEIIARRCFLRYIYDQLELH LSDNSDVRNSSIFELRDKGGYQLKENIQFHLYISTAPCGDARIFSP HGQDVETGDRHPNRKARGQLRTKIESGQGTIPVRTSGVIQTWDG VLEGERLLTMSCSDKIARWNVLGIQGSLLSHFMNPIYLESIILGSL YHSDHLSRAMYSRISIIENLPEPFHLNRPFLSGISSPESRQPGKAPN FGINWRKEDETFEVINAMTGRVEGGSVSRICKQALFGRFMSLYG KLSSLTGQSVTTRPTHYSDAKAAVMEYQLAKQCVFQAFQKAGL GNWVQKPIEQDQF
Sepia_ADAR2_E_Q_Mutant(SEQ ID NO :655)	GIGTGTKCINGEYMNDRGFAVNDCHAEIIARRCFLRFIYDQLEMH LSEDPEVRGQSVFELRDGGGYKLKPNIHFHLYISTAPCGDARIFSP HGQDVETGDRHPNRKARGQLRTKIESGQGTIPVRSSGFIQTWDG VLEGERLLTMSCSDKIARWNVLGIQGALLCHFMHPIYLESIILGSL YHSDHLSRAVYCRIASIENLPDLFQLNRPFLSGISSPESRQPGKAP NFGINWRRNDDTFEVINAMTGRVEGGNMSRICKQALFDRFMNL YGRLSSLTGQSVTTRPTLYSEAKAAVMEYQLAKQCVFQAFQKA GLGNWVQKPIEQDQF
Doryteusthis opalescens_ADAR2_E_Q_Mutant(SEQ ID NO :656)	GIGTGTKCINGEYMNDRGFAVNDCHAEIIARRCFLRFIYDQLELH LSDNAEVRGQSIFELRDAGGYKLKPNIQFHLYISTAPCGDARIFSP HGQDVETGDRHPNRKARGQLRTKIESGQGTIPVRSSGFIQTWDG VLEGERLLTMSCSDKIARWNVLGVQGALLCHFMHPIYLESIILGS LYHSDHLSRAVYCRIAAIENLPDLFRLNRPFLSGISSPESRQPGKA PNFGINWRRNDDSFEVINAMTGRVEGGSMSRICKQALFDRFMNL YGKLSSLTGQSVTTRPALYSEAKATVMEYQLAKQCVFQAFQKA GLGNWVQKPIEQDQF

FIG. 4
(continued)

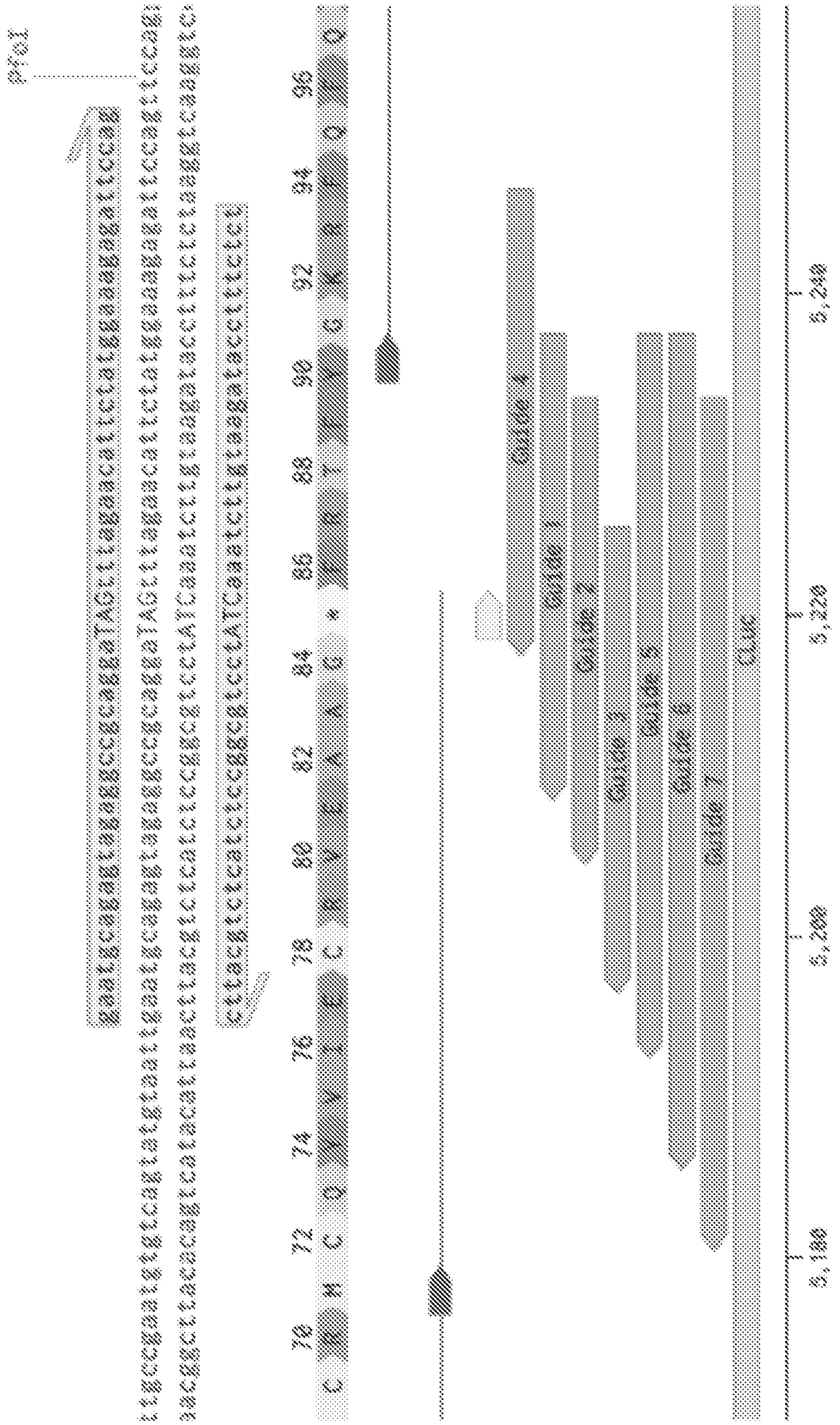

PfoI
Guide 4
Guide 1
Guide 2
Guide 3
Guide 5
Guide 6
Guide 7
CLuc
5,180
5,200
5,220
5,240

FIG. 5

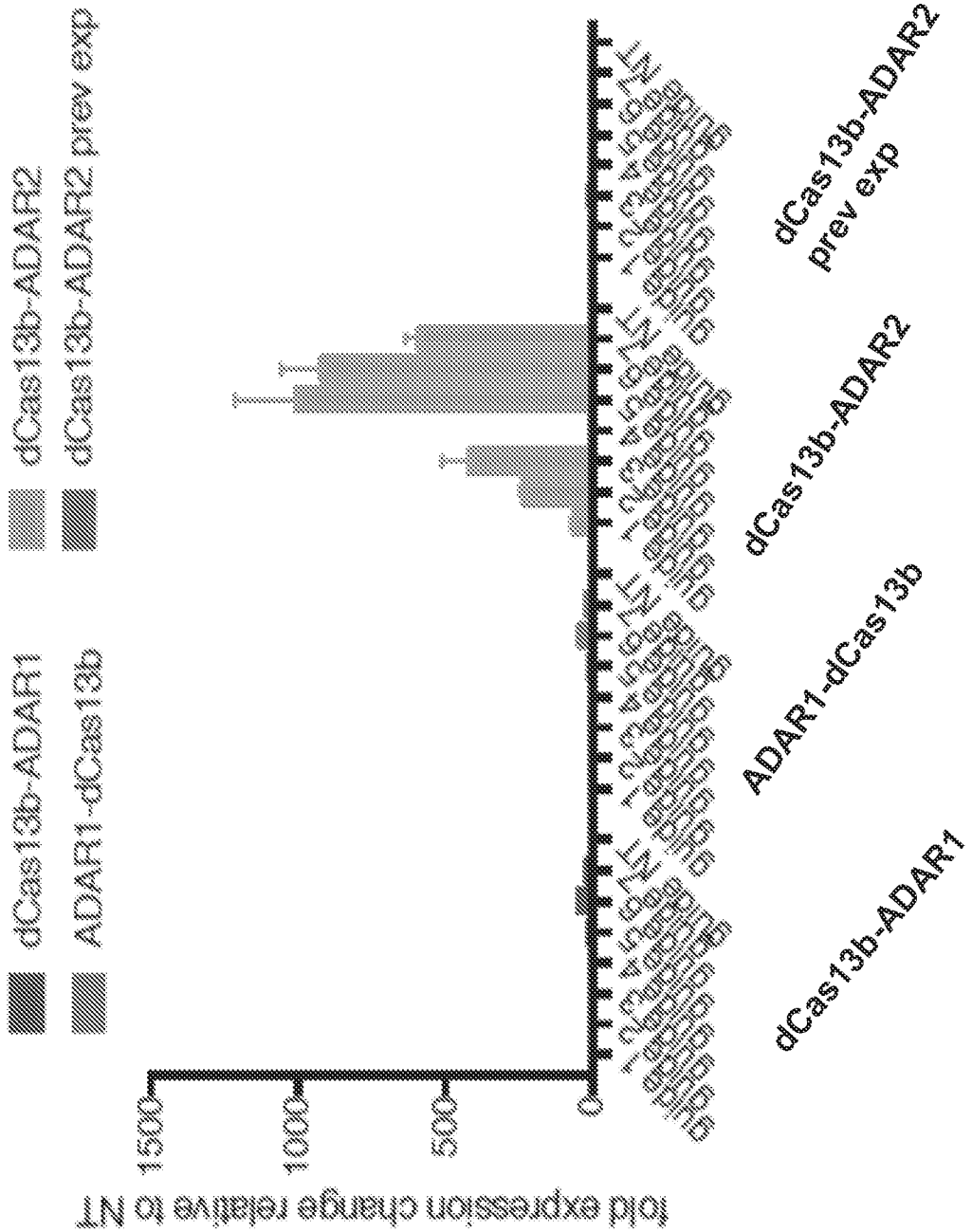

dCas13b-ADAR1
ADAR1-dCas13b
dCas13b-ADAR2
dCas13b-ADAR2 prev exp
fold expression change relative to NT
1500
1000
500
0
dCas13b-ADAR1
ADAR1-dCas13b
dCas13b-ADAR2
dCas13b-ADAR2 prev exp

FIG. 6

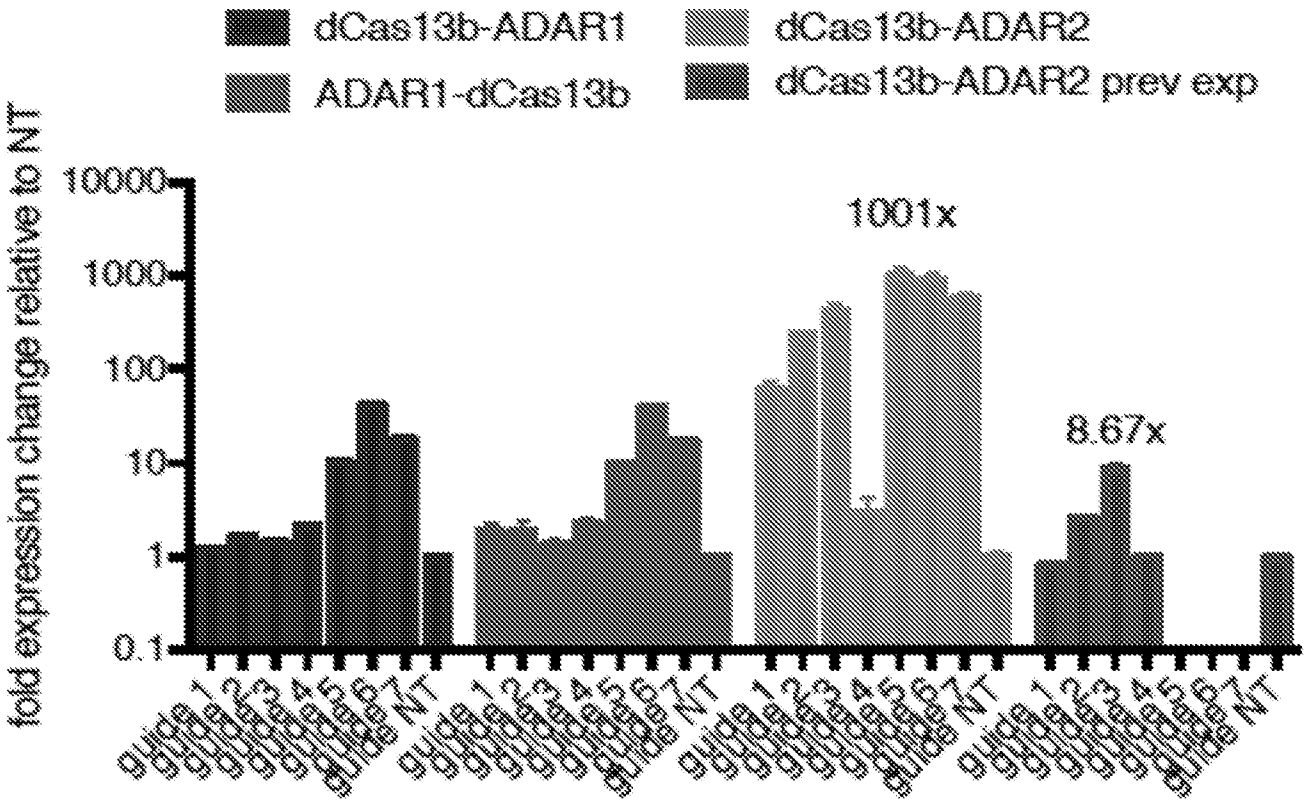

dCas13b-ADAR1
ADAR1-dCas13b
dCas13b-ADAR2
dCas13b-ADAR2 prev exp
fold expression change relative to NT
10000
1000
100
10
1
0.1
1001x
8.67x
guide 1
guide 2
guide 3
guide 4
guide 5
guide 6
guide 7
guide 8
guide NT

FIG. 7

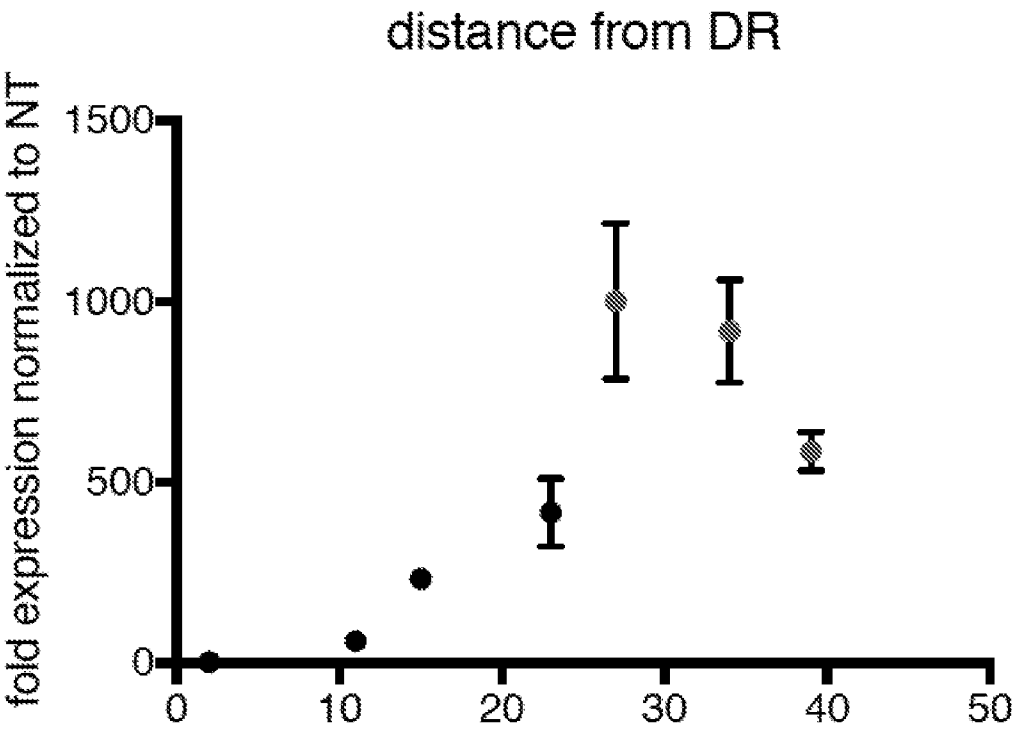

distance from DR
fold expression normalized to NT
1500
1000
500
0
0
10
20
30
40
50

FIG. 8

A

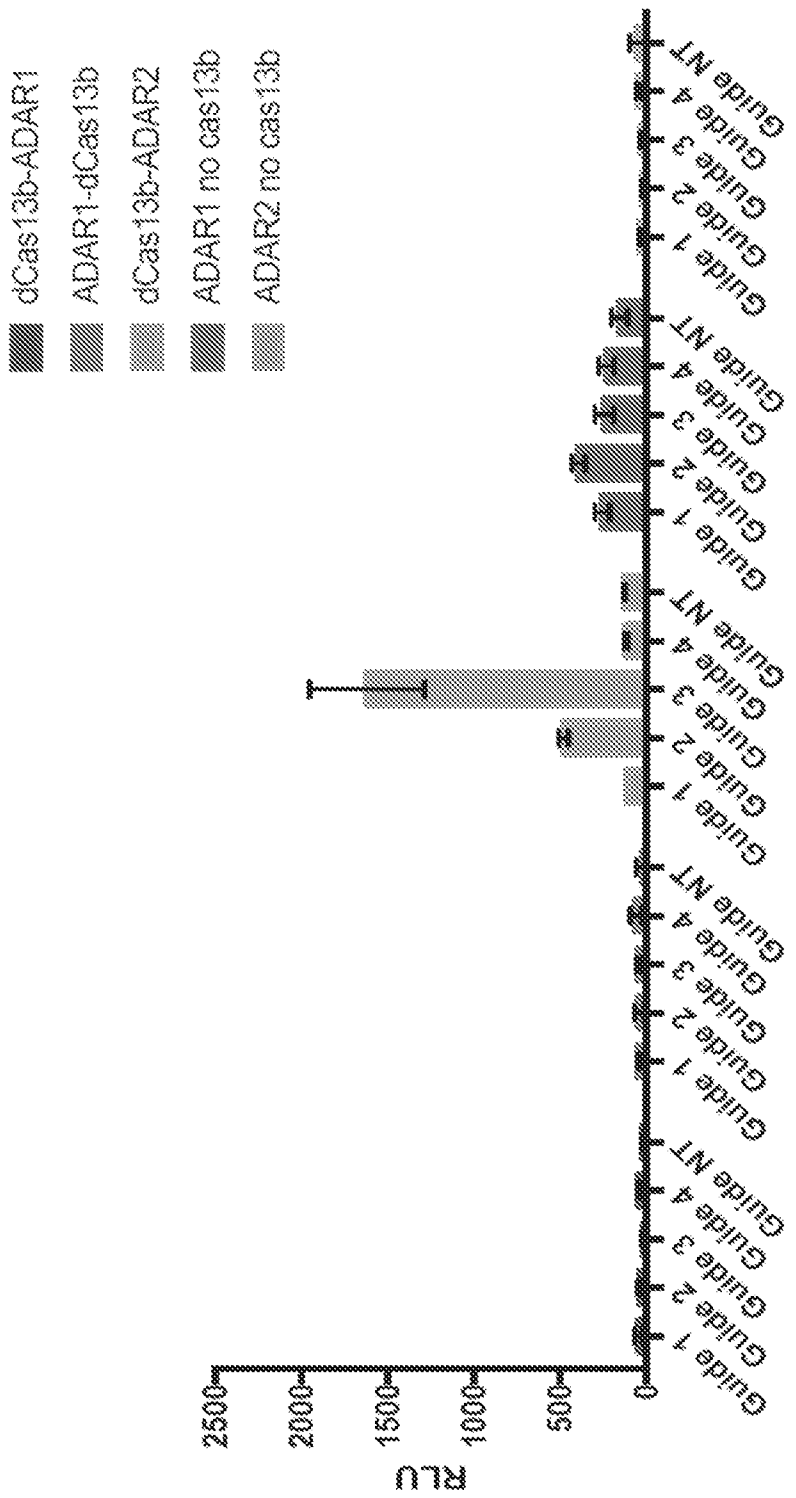

dCas13b-ADAR1
ADAR1-dCas13b
dCas13b-ADAR2
ADAR1 no cas13b
ADAR2 no cas13b
RLU
2500
2000
1500
1000
500
0
Guide 1
Guide 2
Guide 3
Guide 4
Guide NT

FIG. 9A

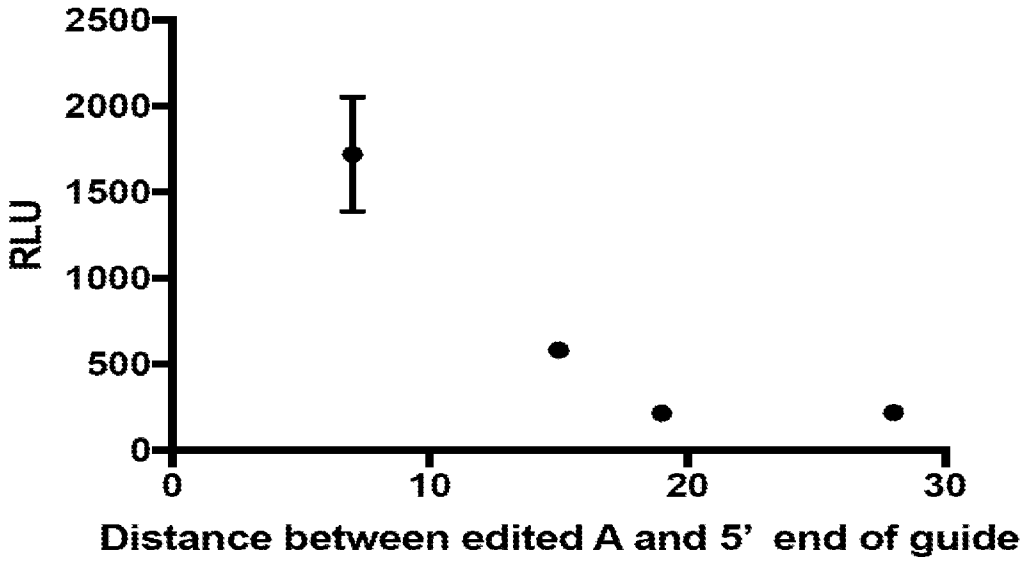

2500
2000
1500
1000
500
0
RLU
0
10
20
30
Distance between edited A and 5' end of guide

FIG. 9B

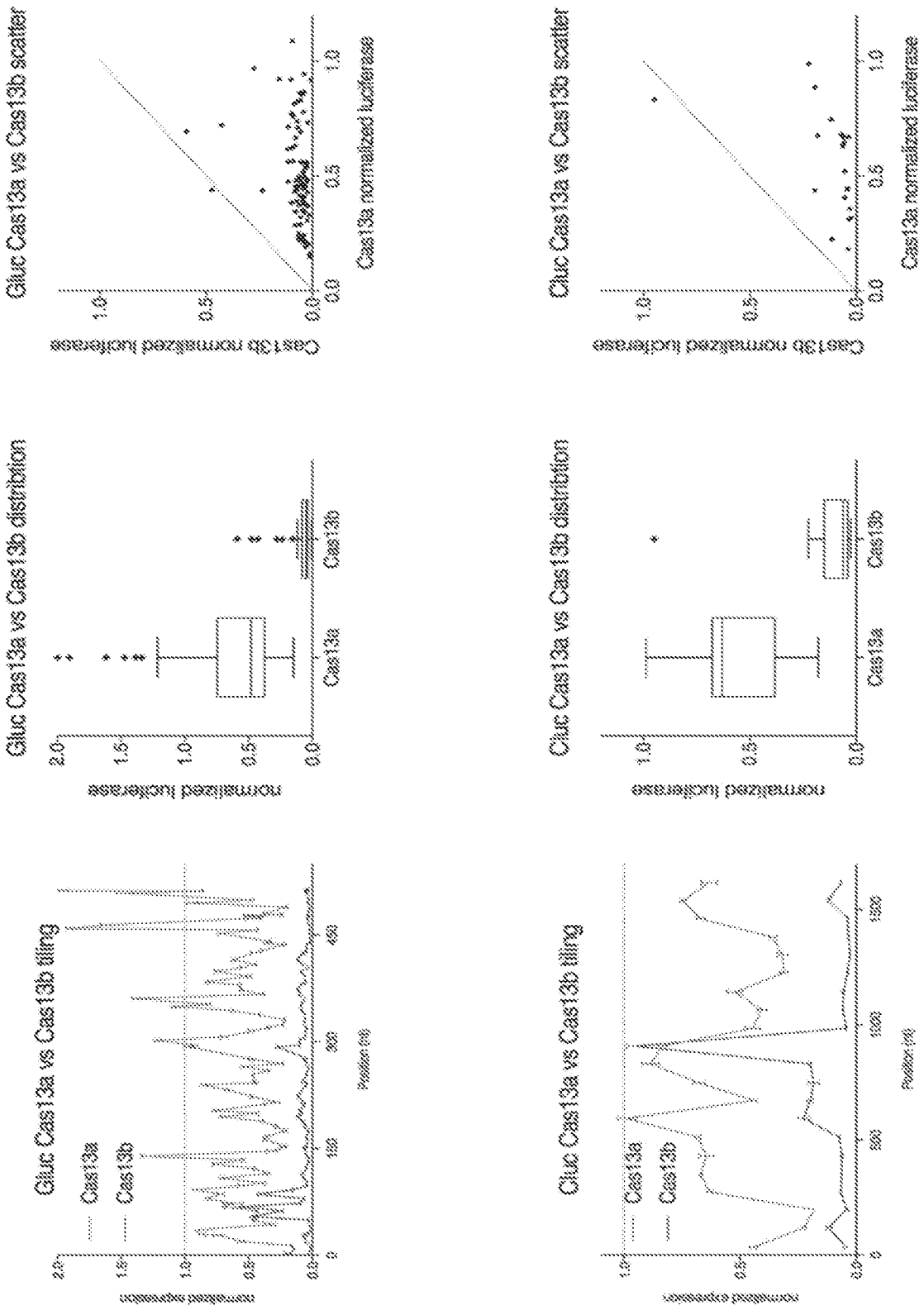
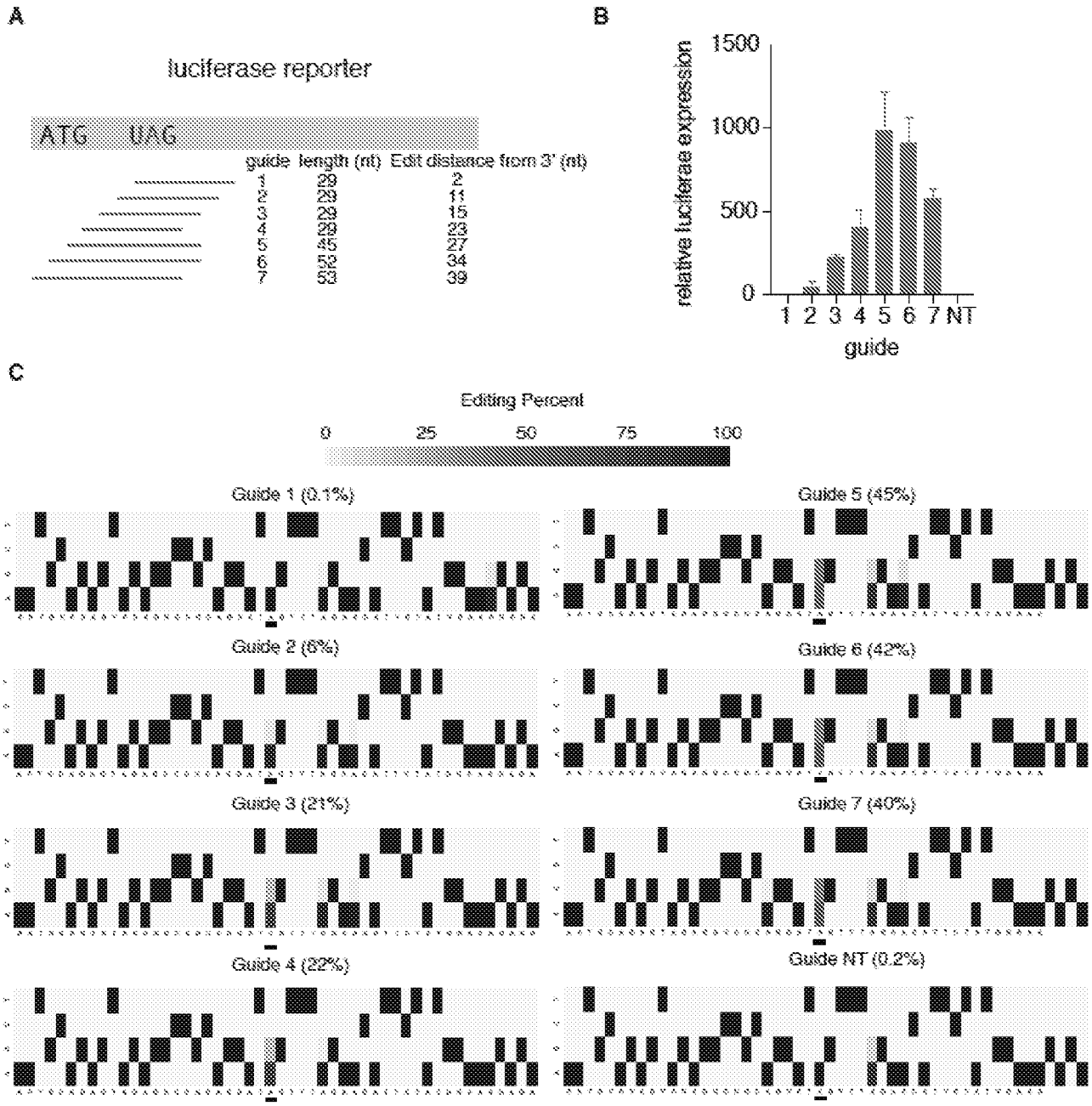
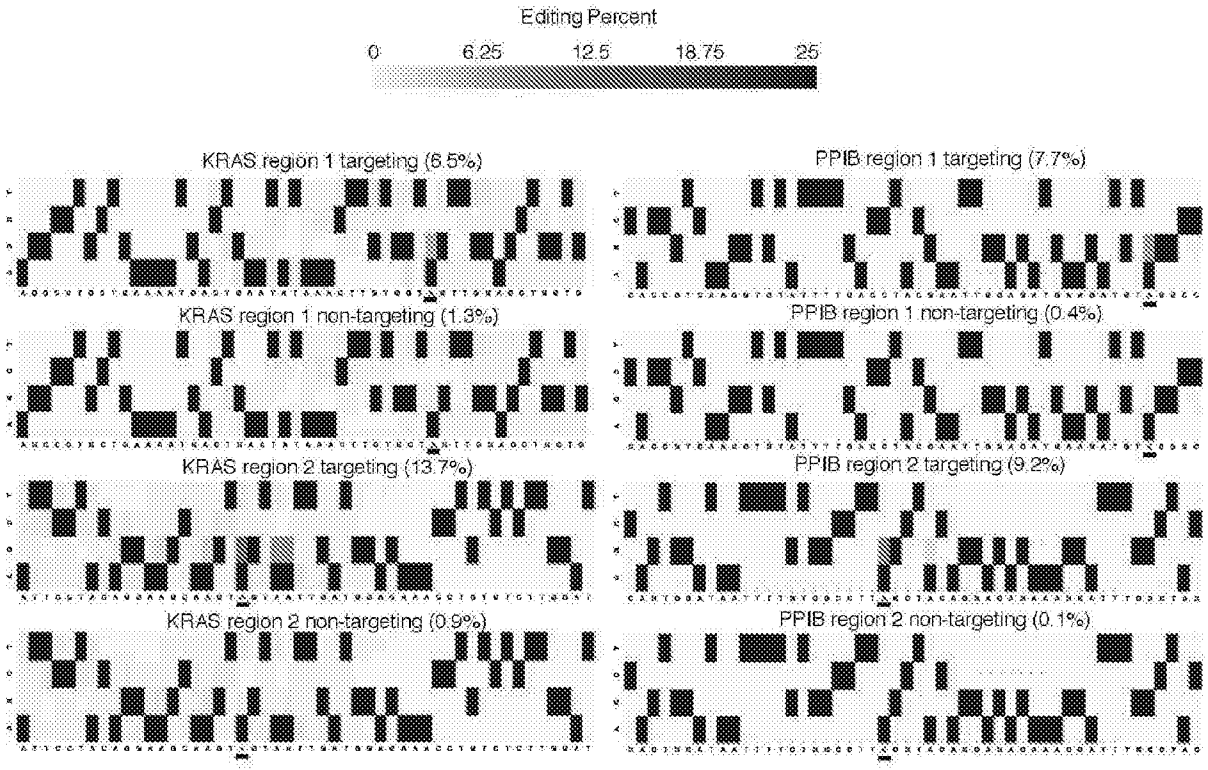

Gluc Cas13a vs Cas13b tiling
Cas13a
Cas13b
normalized expression
Position (nt)
Gluc Cas13a vs Cas13b distribution
normalized luciferase
Cas13a
Cas13b
Gluc Cas13a vs Cas13b scatter
Cas13b normalized luciferase
Cas13a normalized luciferase

FIG. 10



Targeted A is shown in red and underlined. On-target percent editing is in parentheses.

FIG. 11



Targeted A is shown in red and underlined. On-target percent editing is in parentheses.

FIG. 12

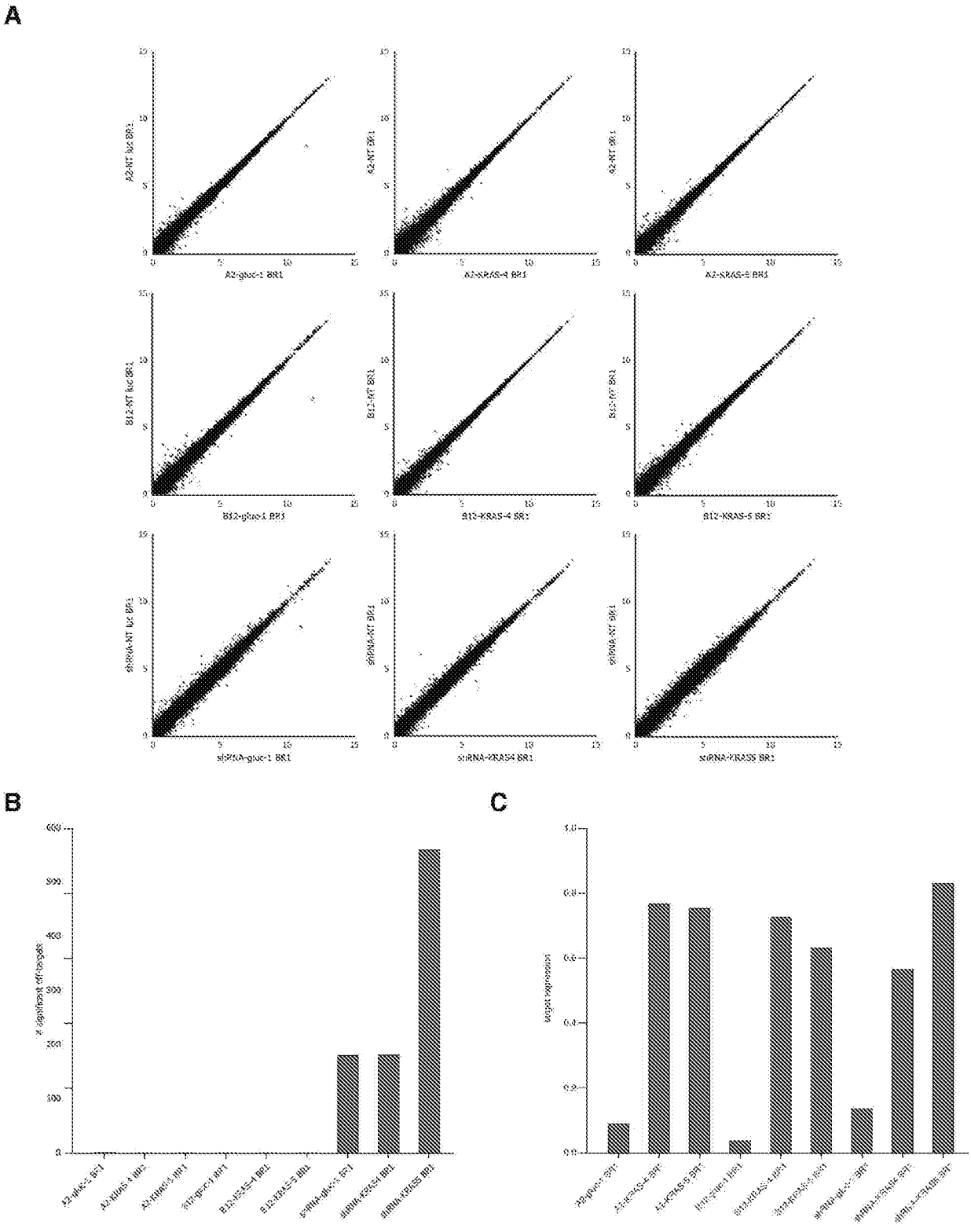

A
A2-NT luc BR1
A2-gluc-1 BR1
A2-NT BR1
A2-KRAS-4 BR1
A2-KRAS-5 BR1
B12-NT luc BR1
B12-gluc-1 BR1
B12-NT BR1
B12-KRAS-4 BR1
B12-KRAS-5 BR1
shRNA-NT luc BR1
shRNA-gluc-1 BR1
shRNA-NT BR1
shRNA-KRAS4 BR1
shRNA-KRAS5 BR1
B
significant off-targets
A2-gluc-1 BR1
A2-KRAS-4 BR1
A2-KRAS-5 BR1
B12-gluc-1 BR1
B12-KRAS-4 BR1
B12-KRAS-5 BR1
shRNA-gluc-1 BR1
shRNA-KRAS4 BR1
shRNA-KRAS5 BR1
C
Target expression

FIG. 13

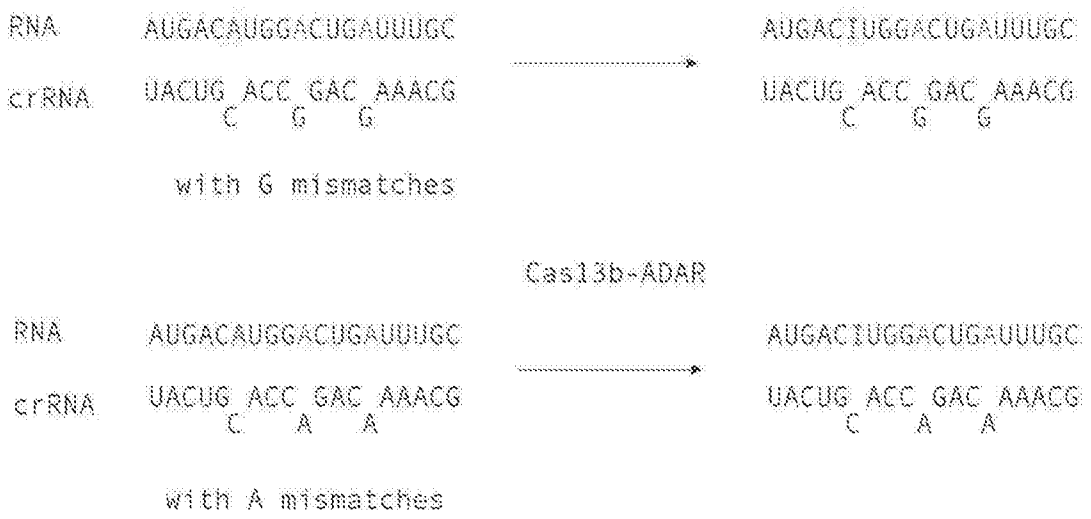

RNA
AUGACAUGGACUGAUUUGC
AUGACIUGGACUGAUUUGC
crRNA
UACUG ACC GAC AAACG
C G G
UACUG ACC GAC AAACG
C G G
with G mismatches
Cas13b-ADAR
RNA
AUGACAUGGACUGAUUUGC
AUGACIUGGACUGAUUUGC
crRNA
UACUG ACC GAC AAACG
C A A
UACUG ACC GAC AAACG
C A A
with A mismatches

FIG. 14

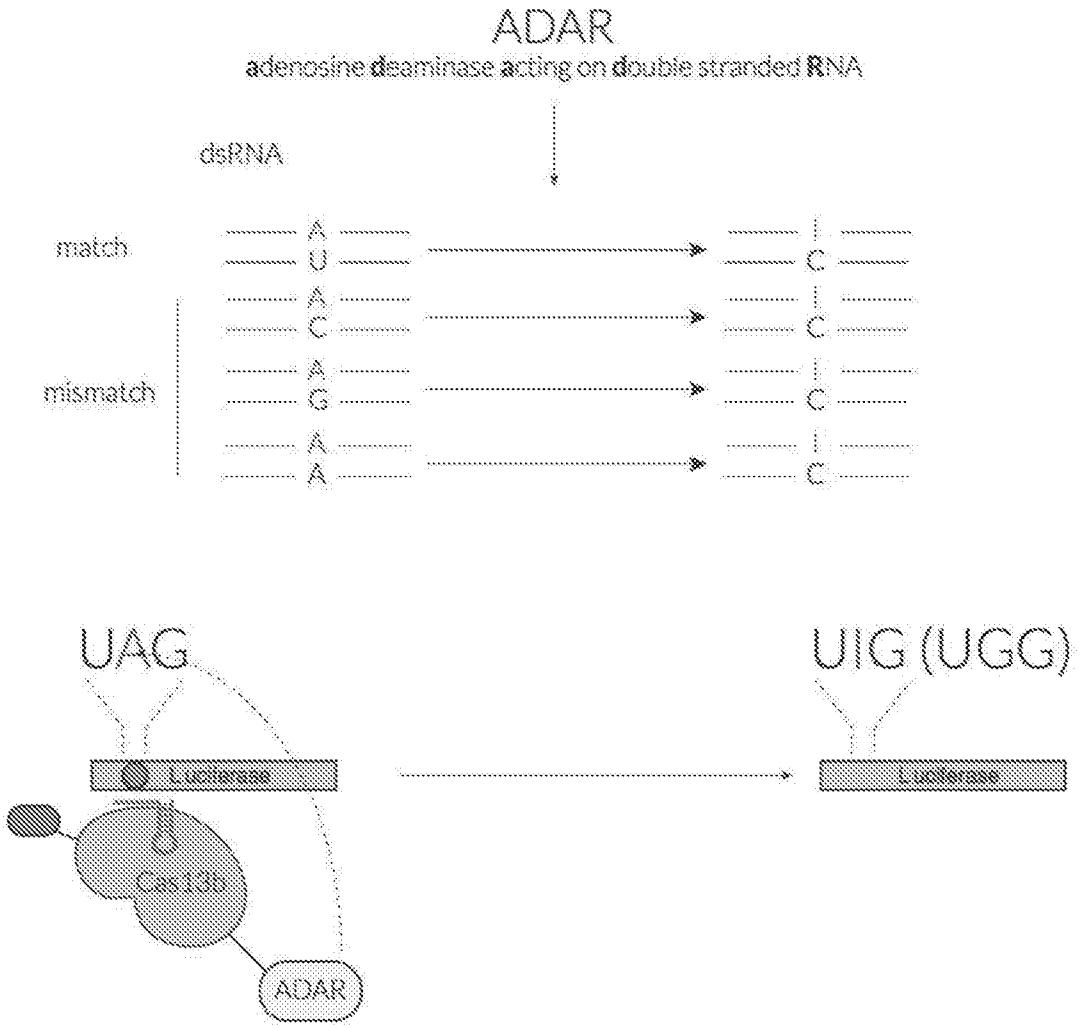

ADAR
adenosine deaminase acting on double stranded RNA
dsRNA
match
mismatch
A
U
A
C
A
G
A
A
I
C
UAG
UIG (UGG)
Luciferase
Luciferase
Cas13b
ADAR

FIG. 15

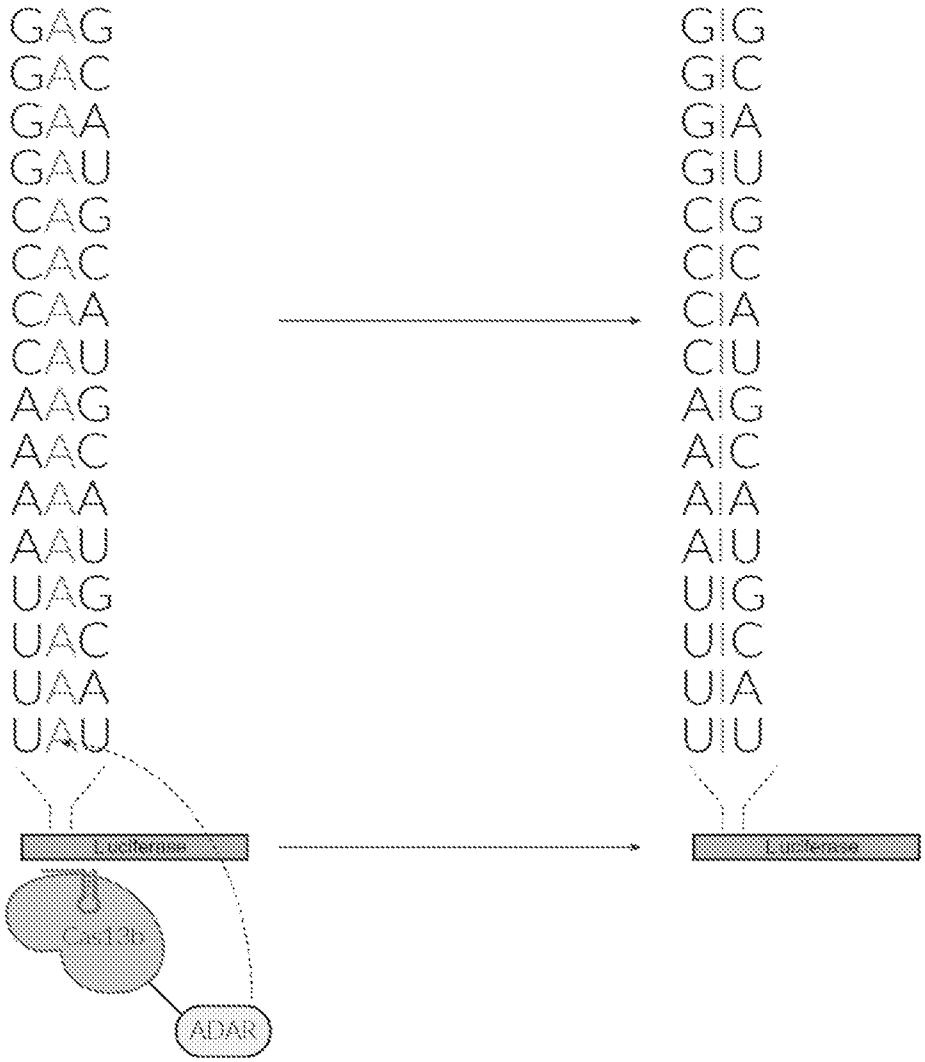

GAG
GAC
GAA
GAU
CAG
CAC
CAA
CAU
AAG
AAC
AAA
AAU
UAG
UAC
UAA
UAU
G|G
G|C
G|A
G|U
C|G
C|C
C|A
C|U
A|G
A|C
A|A
A|U
U|G
U|C
U|A
U|U
ADAR

FIG. 16

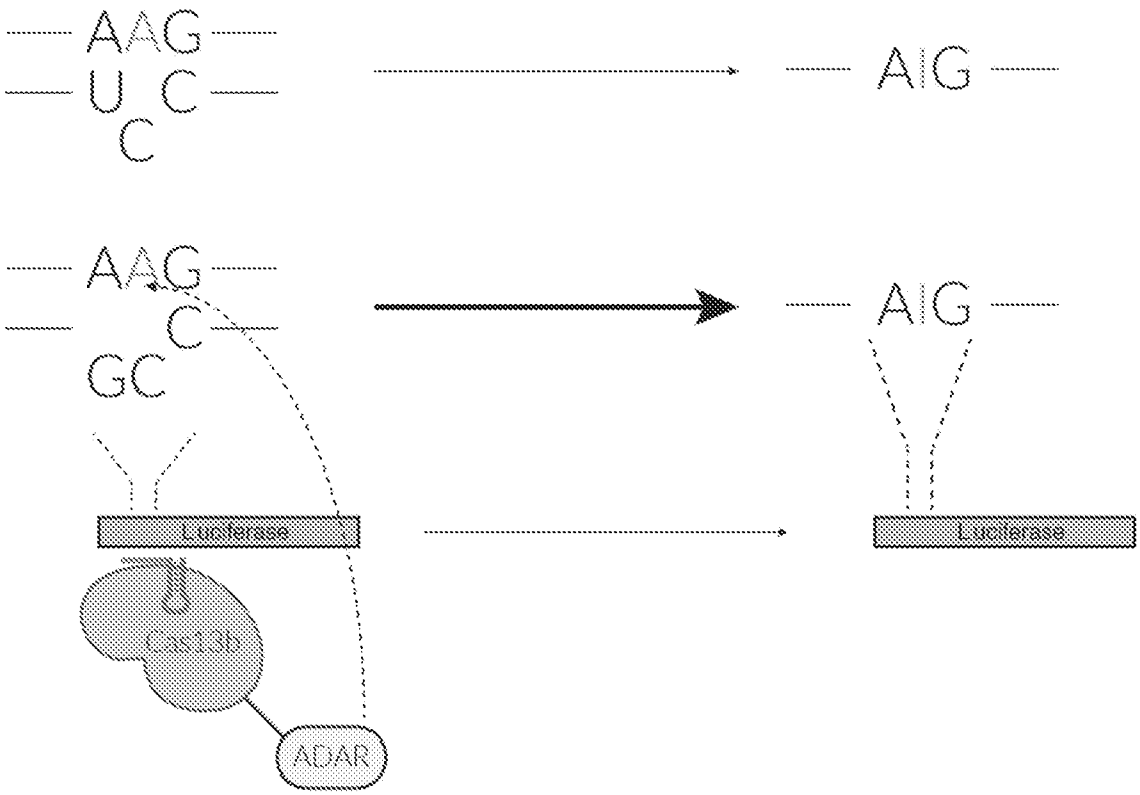


FIG. 17

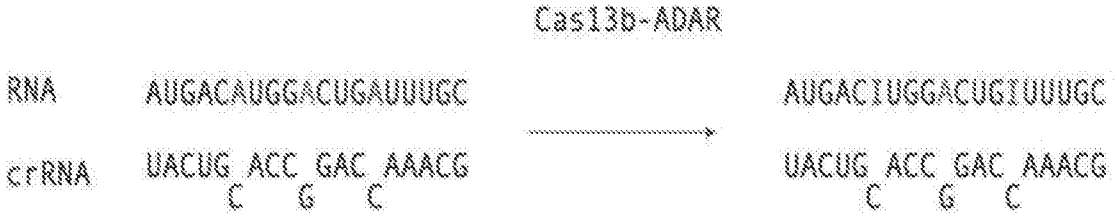

Cas13b-ADAR
RNA
AUGACAUGGACUGAUUUGC
AUGACIUGGACUGIUUUGC
crRNA
UACUG ACC GAC AAACG
C G C
UACUG ACC GAC AAACG
C G C

FIG. 18

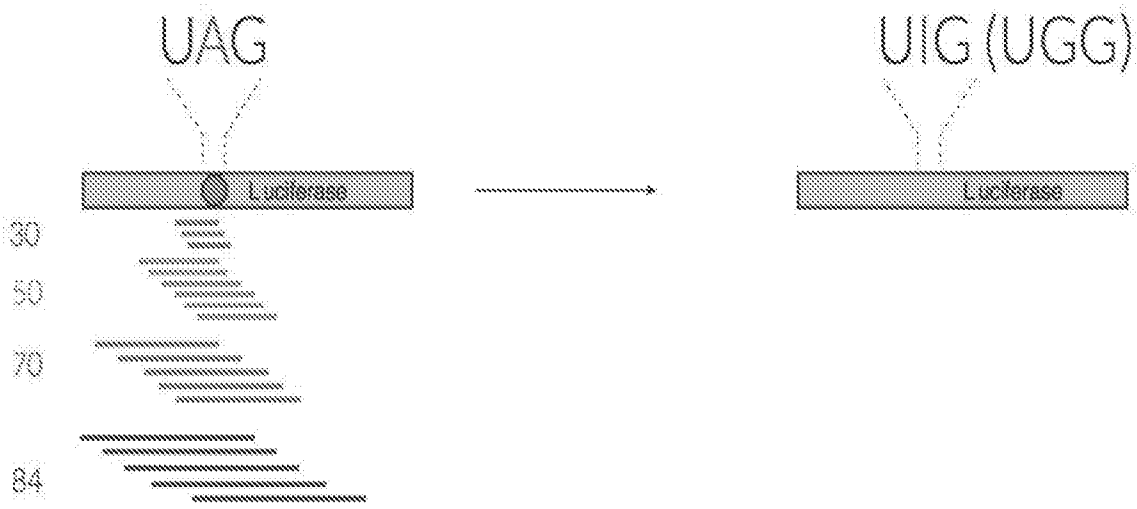

UAG
UIG (UGG)
Luciferase
Luciferase
30
50
70
84

FIG. 19

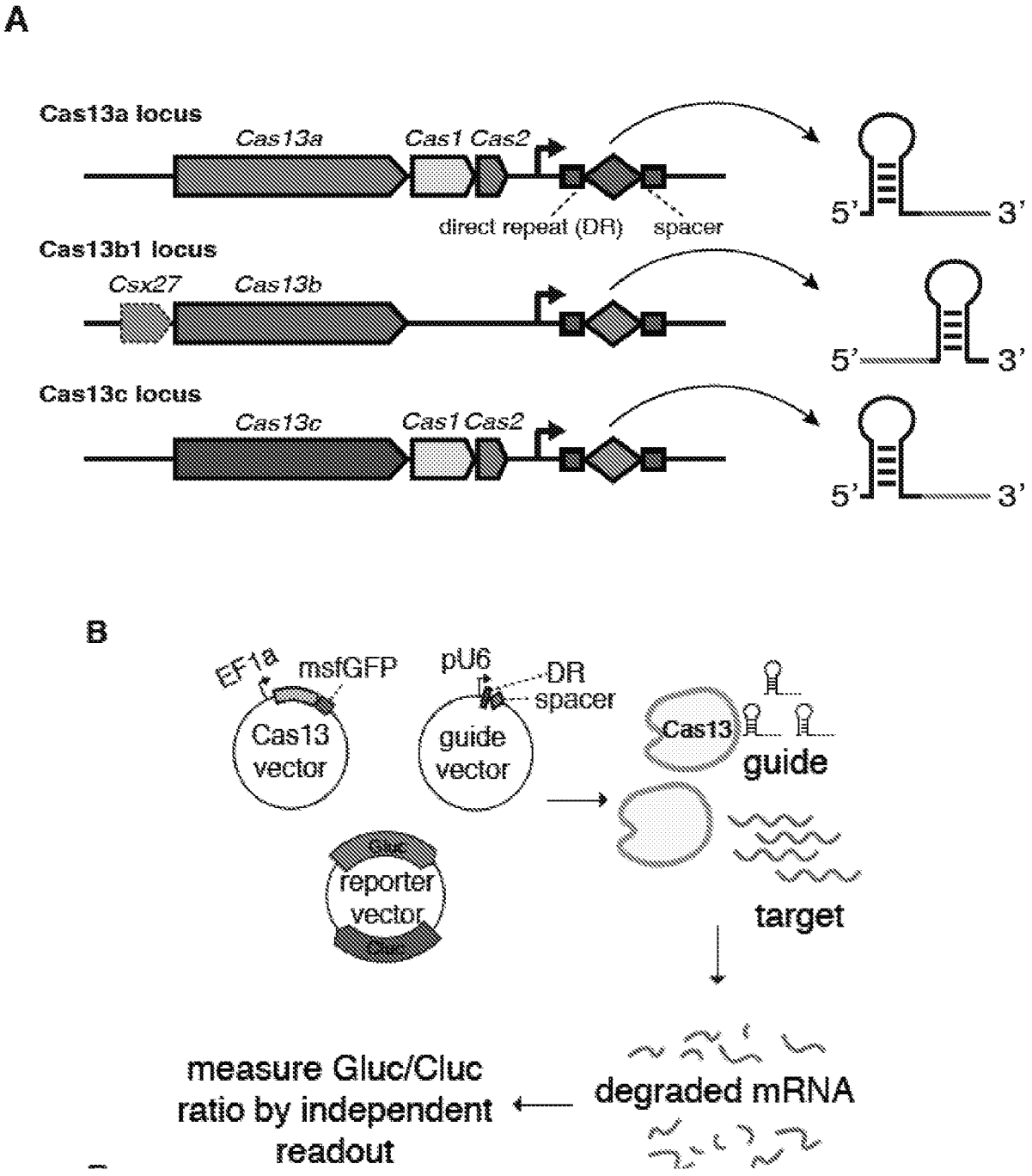
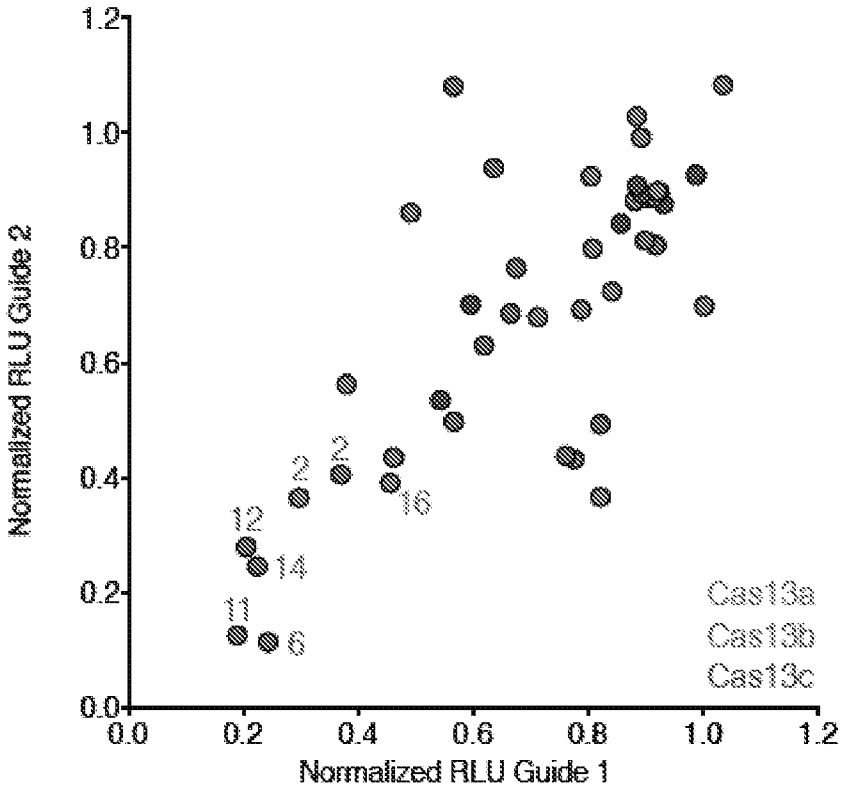

A
Cas13a locus
Cas13a
Cas1 Cas2
direct repeat (DR)
spacer
5'
3'
Cas13b1 locus
Csx27
Cas13b
5'
3'
Cas13c locus
Cas13c
Cas1 Cas2
5'
3'
B
EF1a
msfGFP
Cas13 vector
pU6
DR
spacer
guide vector
Cas13
guide
reporter vector
target
degraded mRNA
measure Gluc/Cluc ratio by independent readout

FIG. 20

C



D

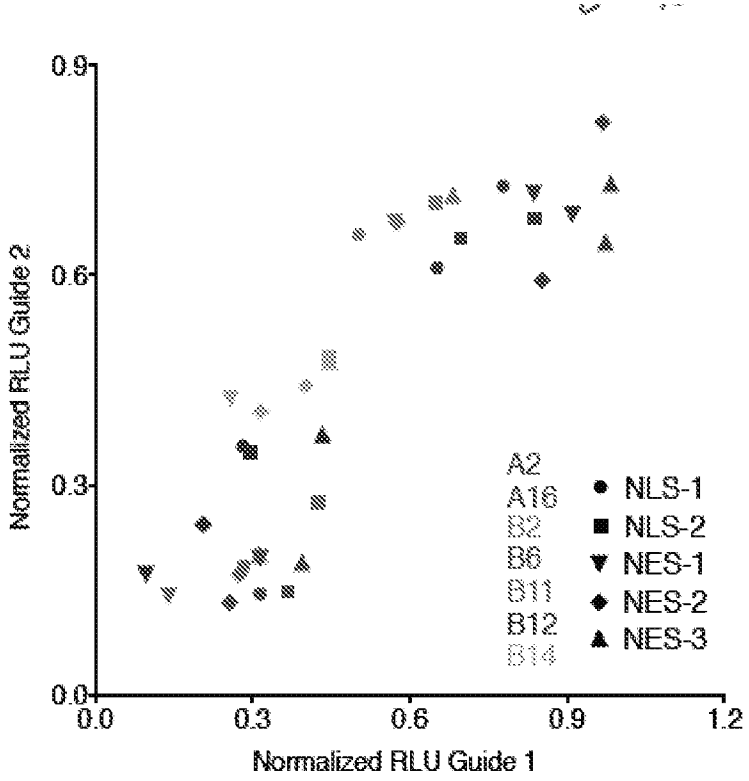


FIG. 20 (continued)

E

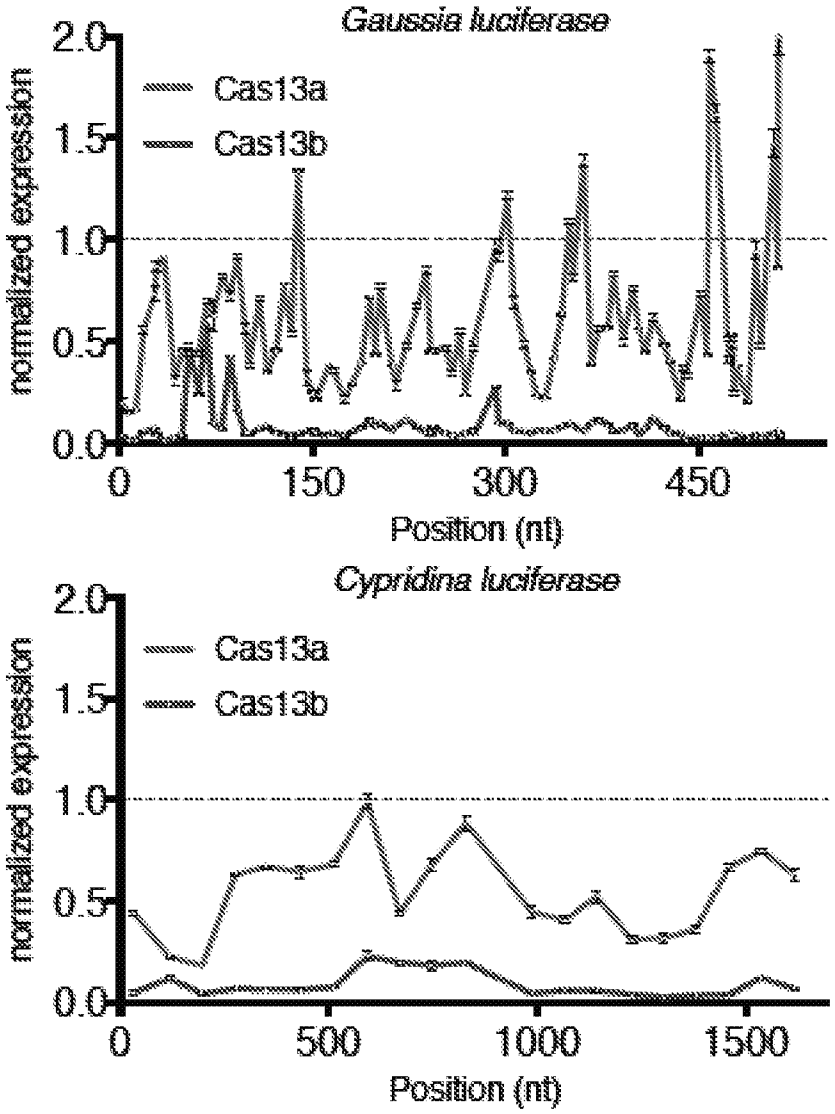


FIG. 20 (continued)

F

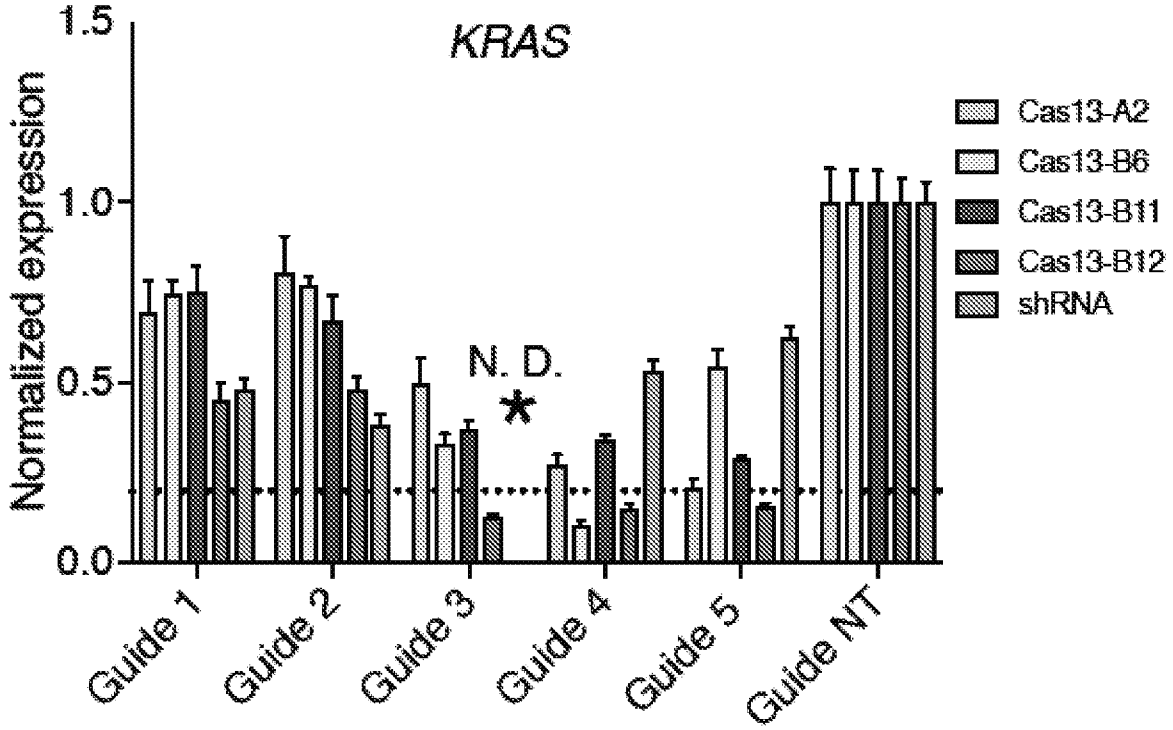


FIG. 20 (continued)

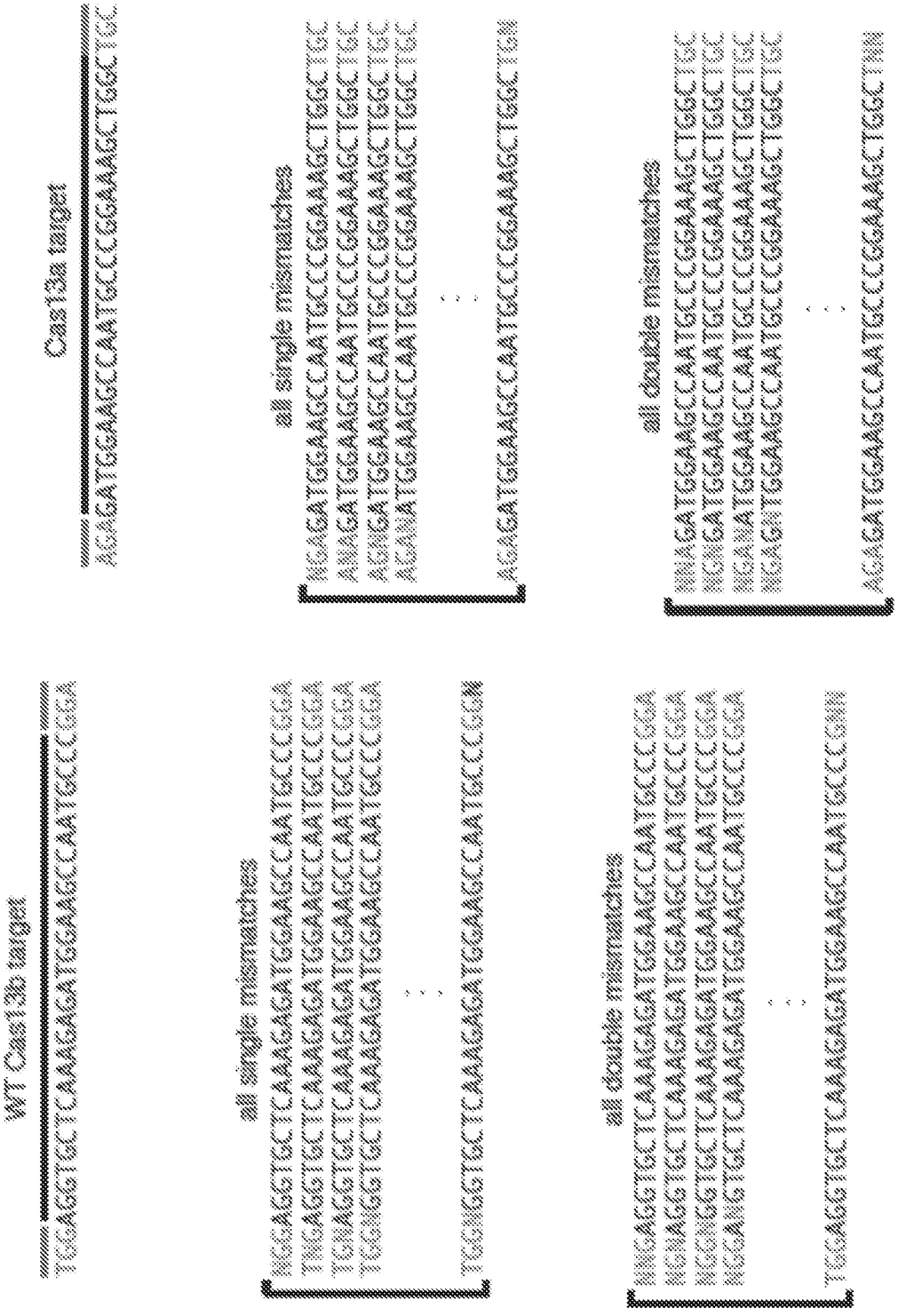

WT Cas13b target
TCGAGGTGCTCAAAGAGATGGAAGCCAATGCCCCGGA
all single mismatches
NCGAGGTGCTCAAAGAGATGGAAGCCAATGCCCCGGA
TNGAGGTGCTCAAAGAGATGGAAGCCAATGCCCCGGA
TGNAGGTGCTCAAAGAGATGGAAGCCAATGCCCCGGA
TCGNGGTGCTCAAAGAGATGGAAGCCAATGCCCCGGA
. . .
TCGAGGTGCTCAAAGAGATGGAAGCCAATGCCCCGGM
all double mismatches
NNCAGGTGCTCAAAGAGATGGAAGCCAATGCCCCGGA
NCNAGGTGCTCAAAGAGATGGAAGCCAATGCCCCGGA
NCGNGGTGCTCAAAGAGATGGAAGCCAATGCCCCGGA
NCGANGTGCTCAAAGAGATGGAAGCCAATGCCCCGGA
. . .
TCGAGGTGCTCAAAGAGATGGAAGCCAATGCCCCGMM
Cas13a target
AGAGATGGAAGCCAATGCCCCGGAAAGCTGGCTGC
all single mismatches
NGAGATGGAAGCCAATGCCCCGGAAAGCTGGCTGC
ANAGATGGAAGCCAATGCCCCGGAAAGCTGGCTGC
AGNGATGGAAGCCAATGCCCCGGAAAGCTGGCTGC
ACANATGGAAGCCAATGCCCCGGAAAGCTGGCTGC
. . .
AGAGATGGAAGCCAATGCCCCGGAAAGCTGGCTGM
all double mismatches
NNAGATGGAAGCCAATGCCCCGGAAAGCTGGCTGC
NCNGATGGAAGCCAATGCCCCGGAAAGCTGGCTGC
NCANATGGAAGCCAATGCCCCGGAAAGCTGGCTGC
NCAGNTGGAAGCCAATGCCCCGGAAAGCTGGCTGC
. . .
AGAGATGGAAGCCAATGCCCCGGAAAGCTGGCTMM

FIG. 21

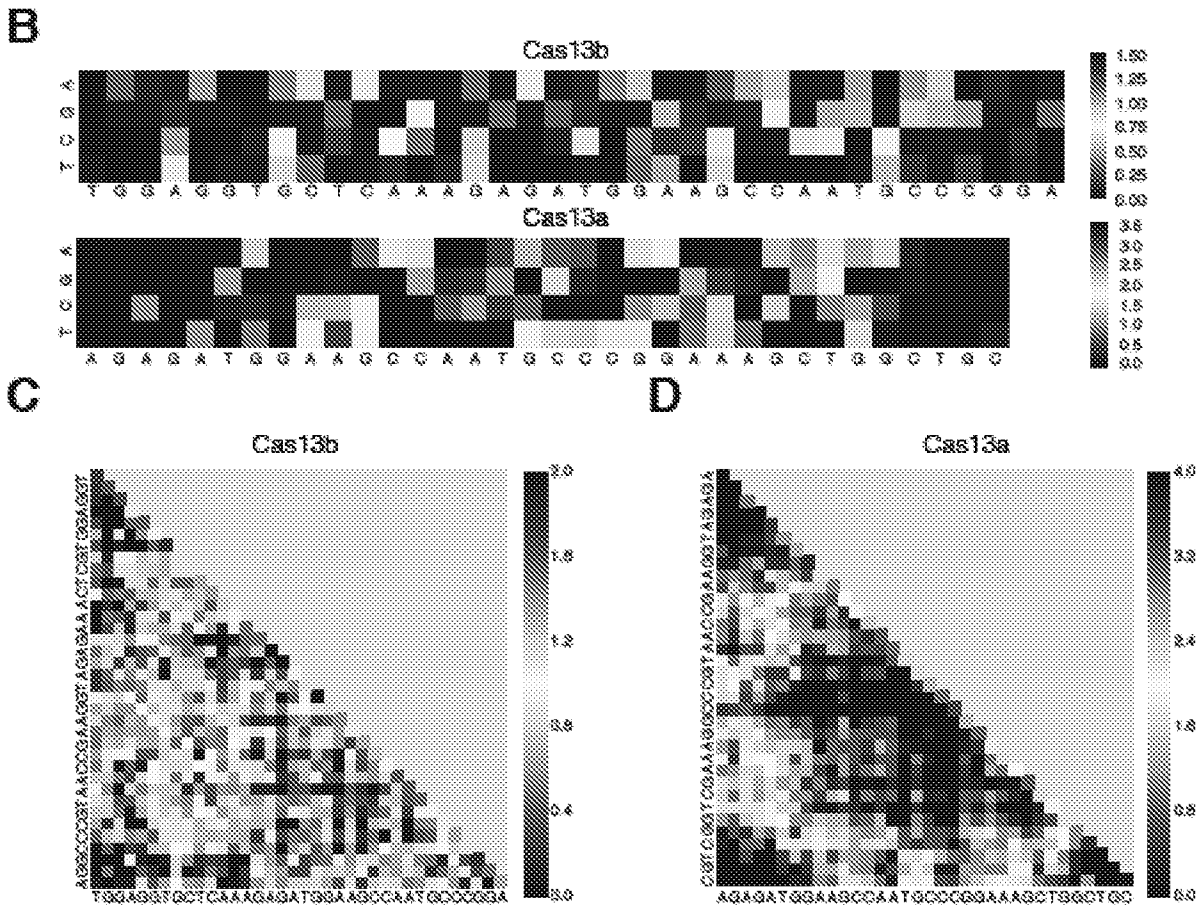


FIG. 21 (continued)

E

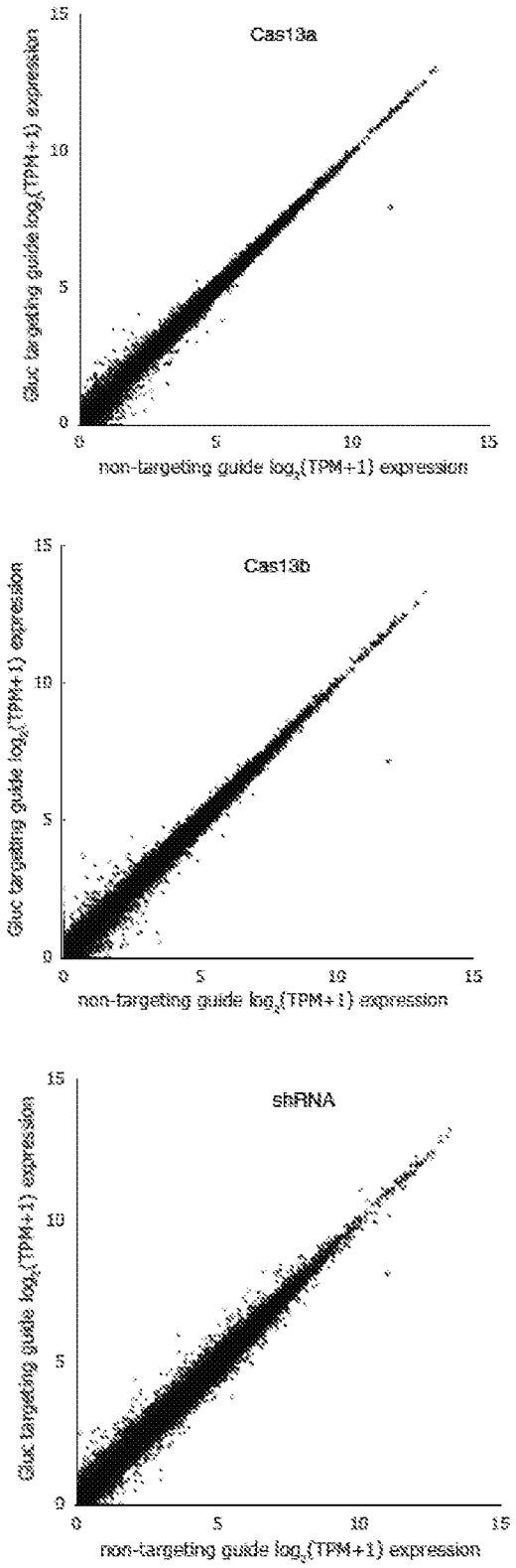
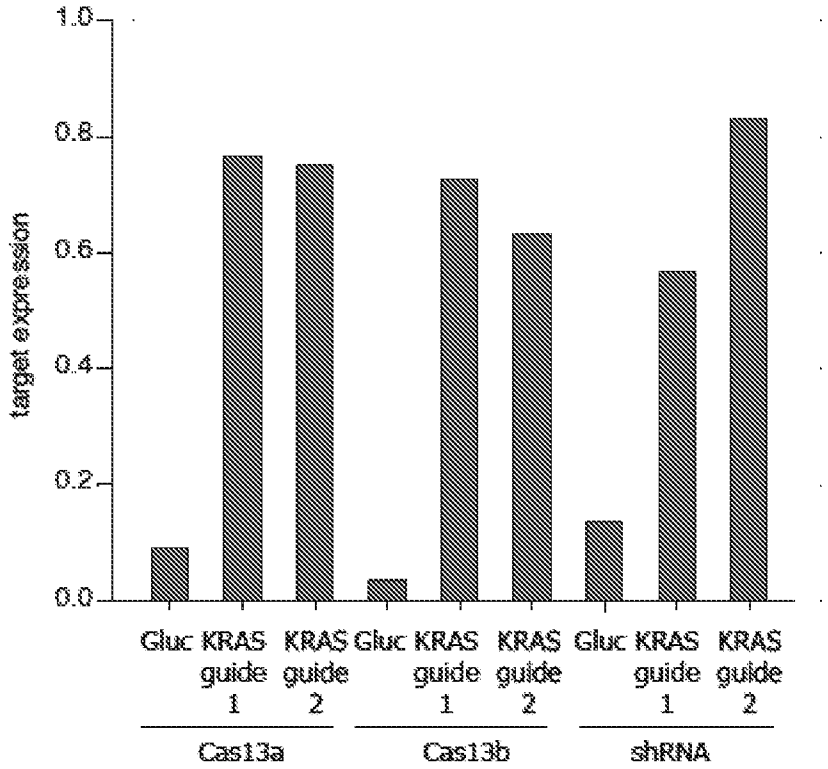


FIG. 21 (continued)

F



G

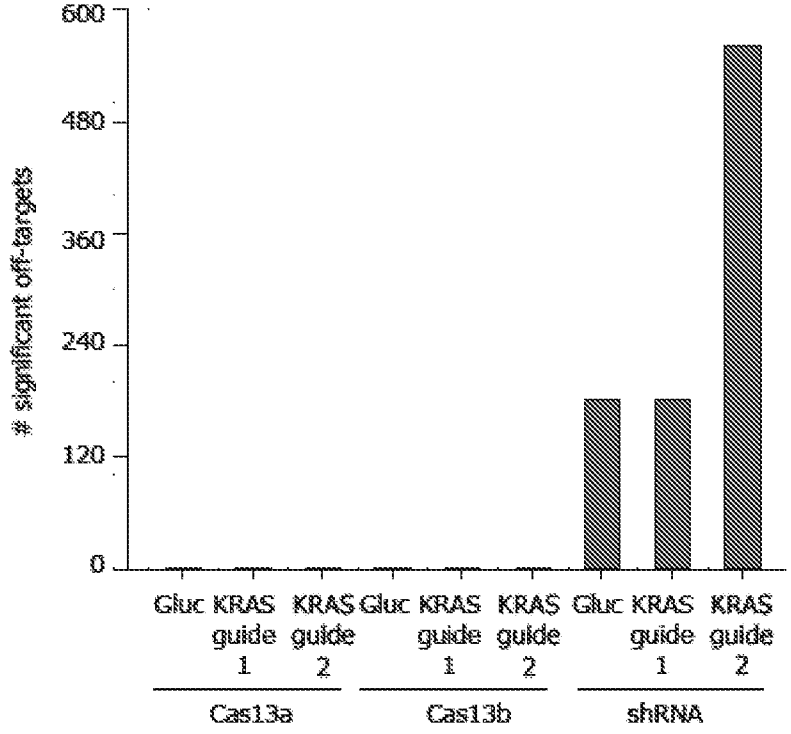
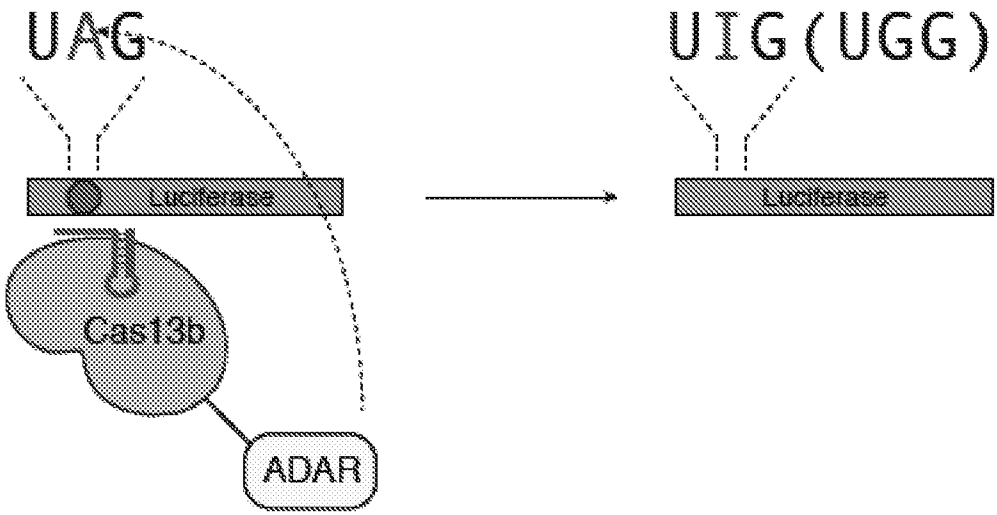


FIG. 21 (continued)

A


UAG
UIG(UGG)
Luciferase
Luciferase
Cas13b
ADAR

B

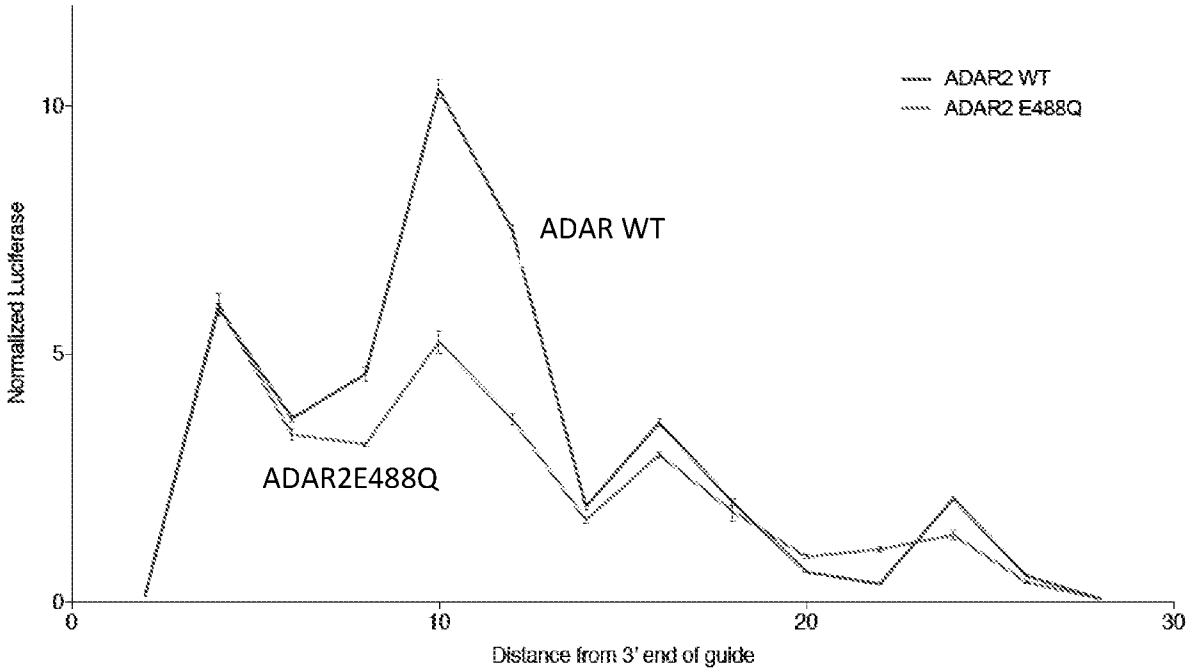

ADAR2 WT
ADAR2 E488Q
ADAR WT
ADAR2E488Q
Normalized Luciferase
Distance from 3' end of guide
0
5
10
0
10
20
30

FIG. 22

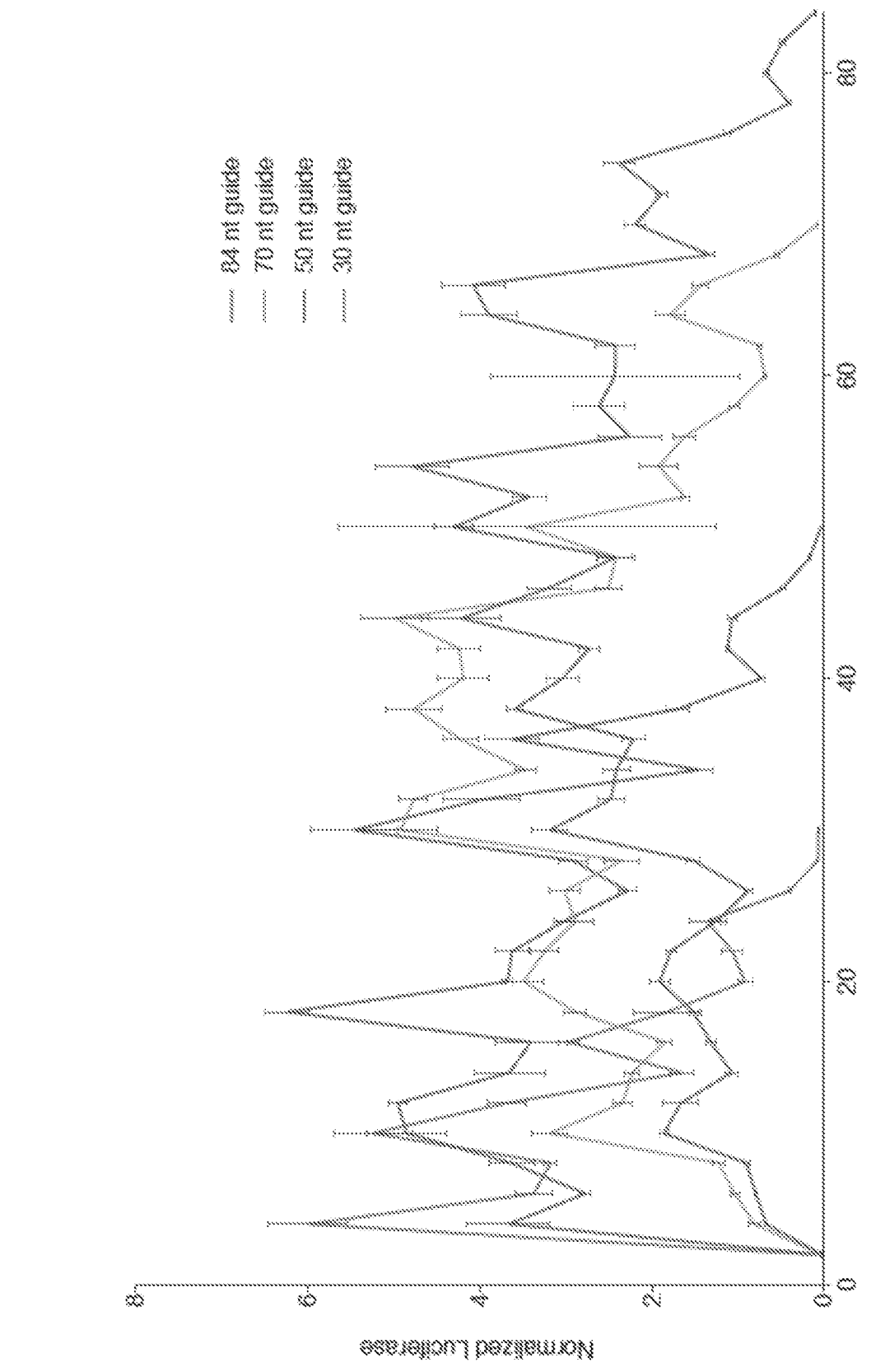

84 nt guide
70 nt guide
50 nt guide
30 nt guide
Normalized Luciferase
Distance from 3' end of guide
0
2
4
6
8
0
20
40
60
80

FIG. 22

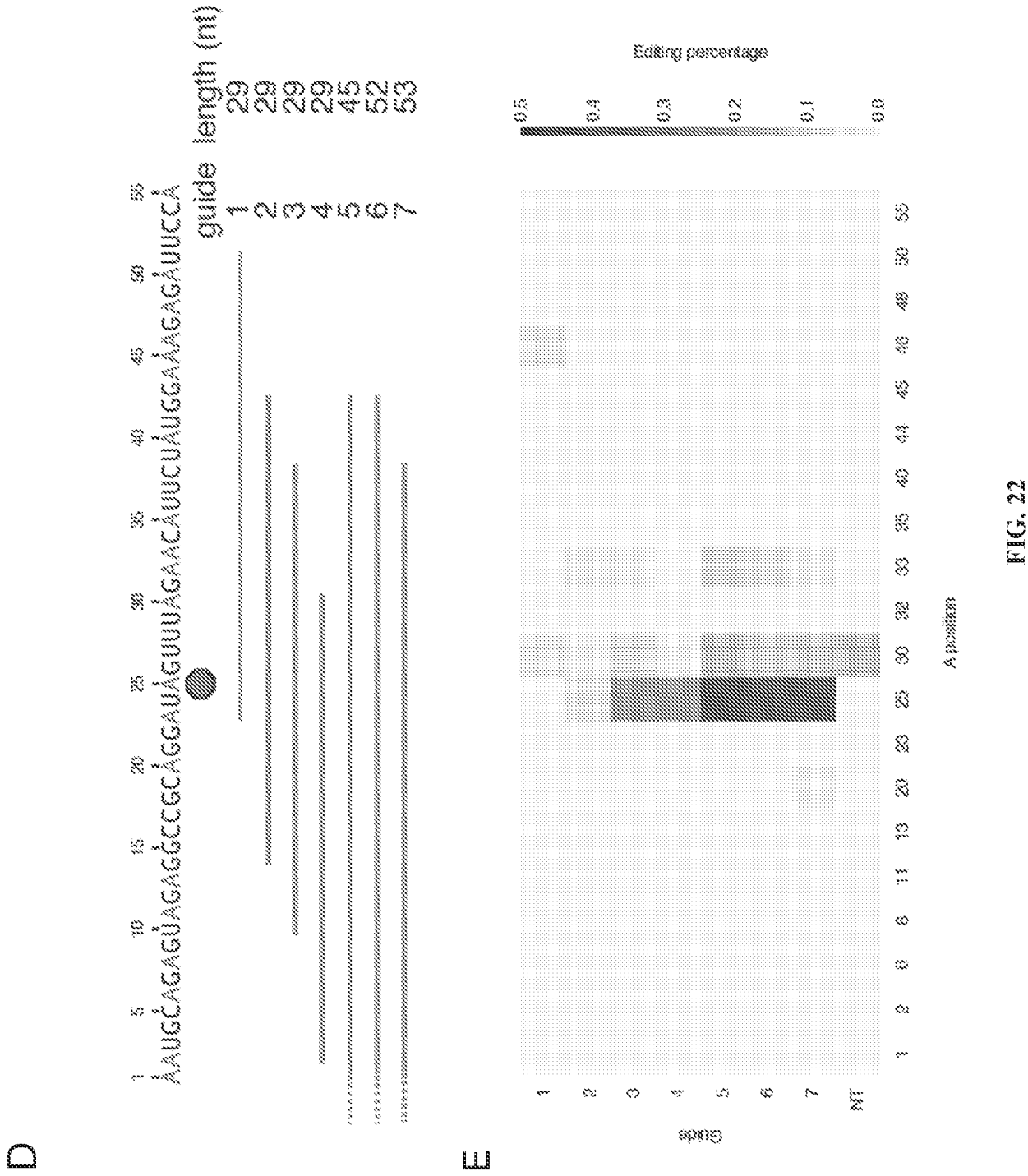

D
guide length (nt)
1 29
2 29
3 29
4 29
5 45
6 52
7 53
E
Editing percentage
0.5
0.4
0.3
0.2
0.1
0.0
guide
MT
A position

FIG. 22

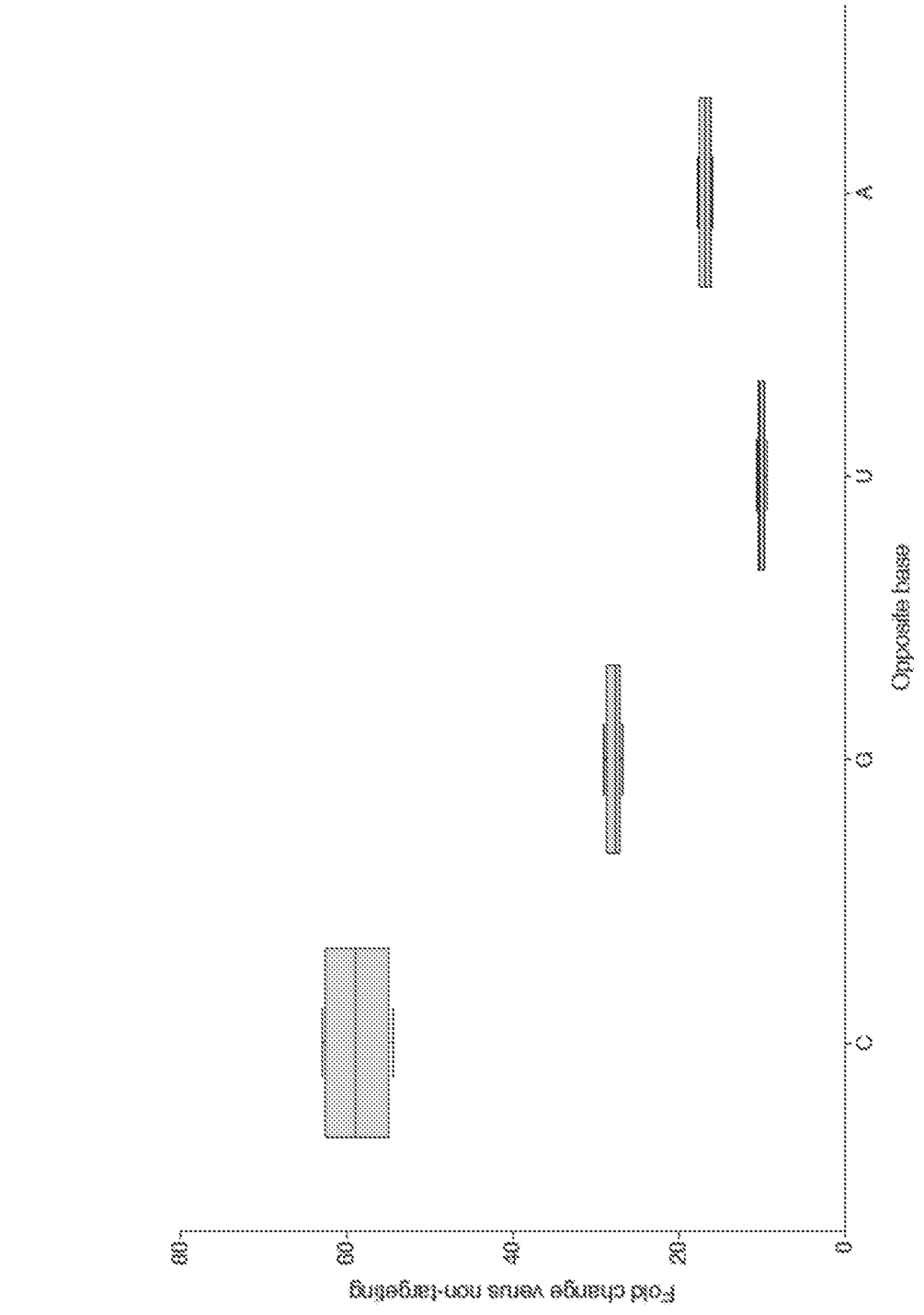

Fold change versus non-targeting
0
20
40
60
C
G
U
A
Oligonucleotide base

FIG. 22

A

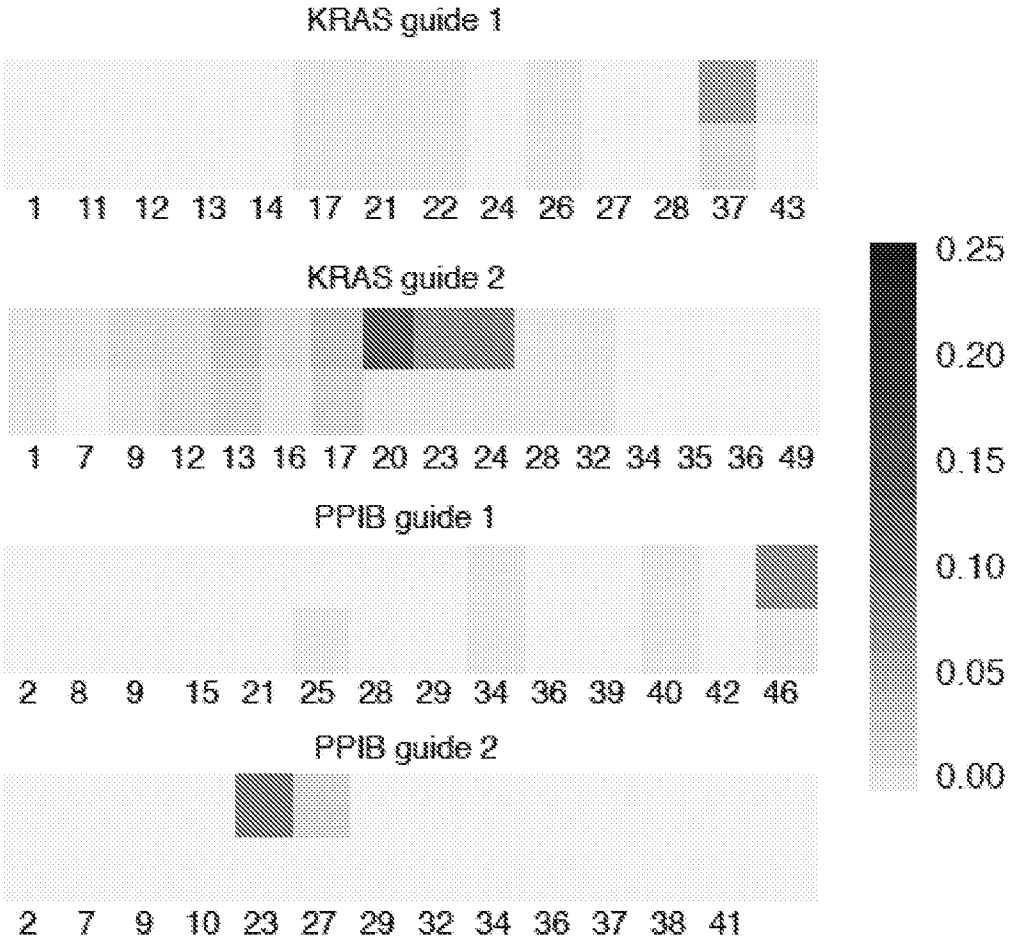

KRAS guide 1
1 11 12 13 14 17 21 22 24 26 27 28 37 43
KRAS guide 2
1 7 9 12 13 16 17 20 23 24 28 32 34 35 36 49
PPIB guide 1
2 8 9 15 21 25 28 29 34 36 39 40 42 46
PPIB guide 2
2 7 9 10 23 27 29 32 34 36 37 38 41
0.25
0.20
0.15
0.10
0.05
0.00

FIG. 23



FIG. 24

A

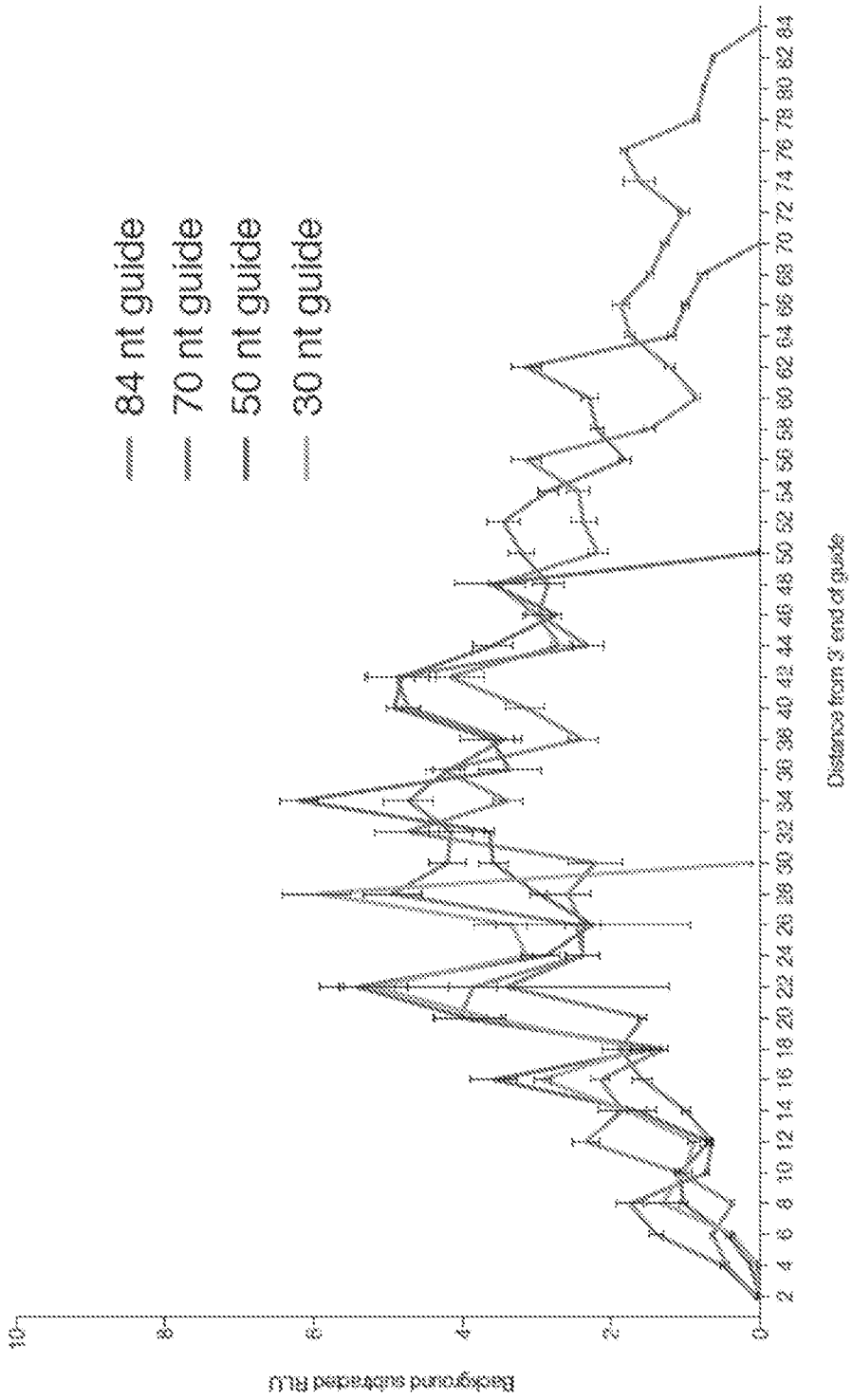

84 nt guide
70 nt guide
50 nt guide
30 nt guide
Background subtracted RLU
Distance from 3' end of guide

FIG. 25

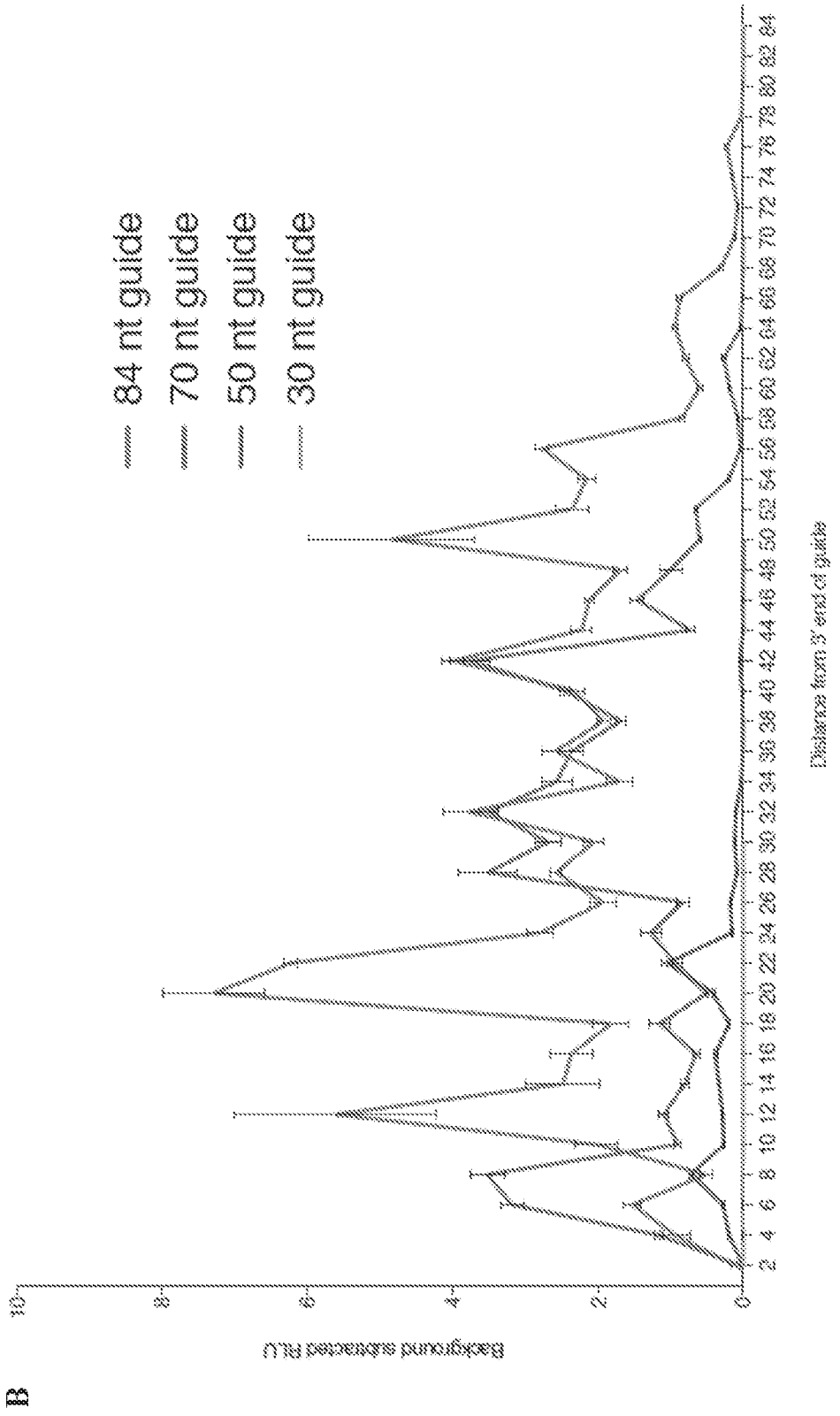

B
Background subtracted RLU
0
2
4
6
8
10
2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56 58 60 62 64 66 68 70 72 74 76 78 80 82 84
Distance from 5' end of guide
84 nt guide
70 nt guide
50 nt guide
30 nt guide

FIG. 25

C

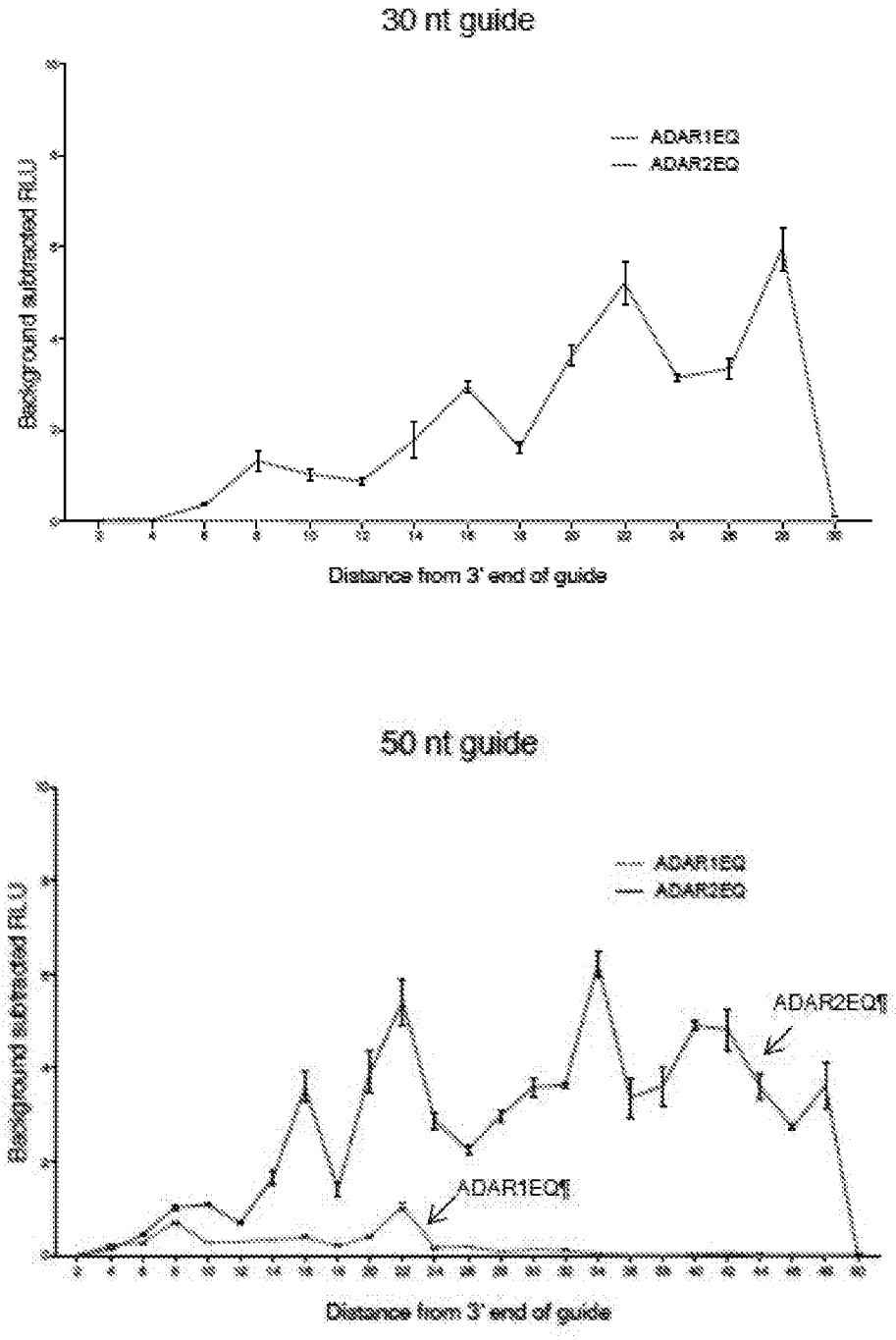


FIG. 25 (continued)

C

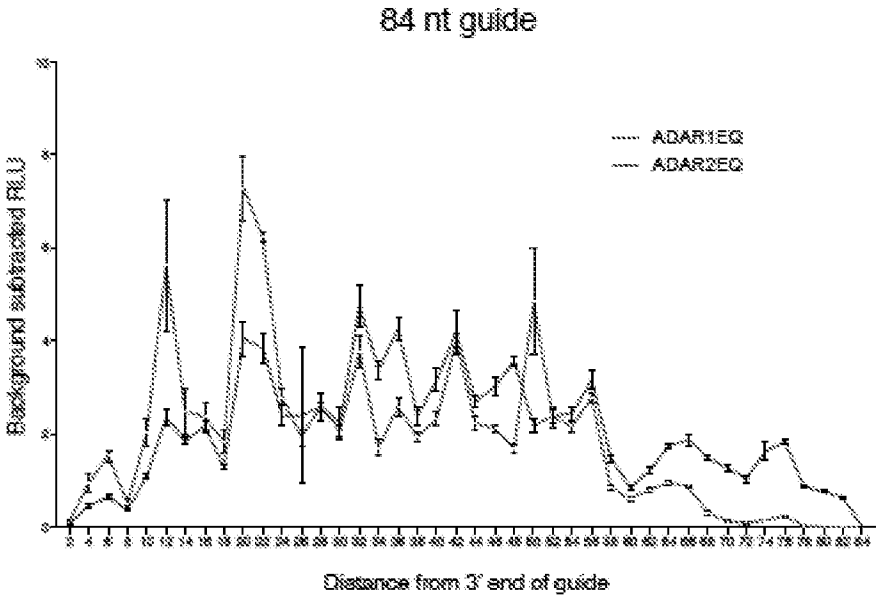


FIG. 25 (continued)

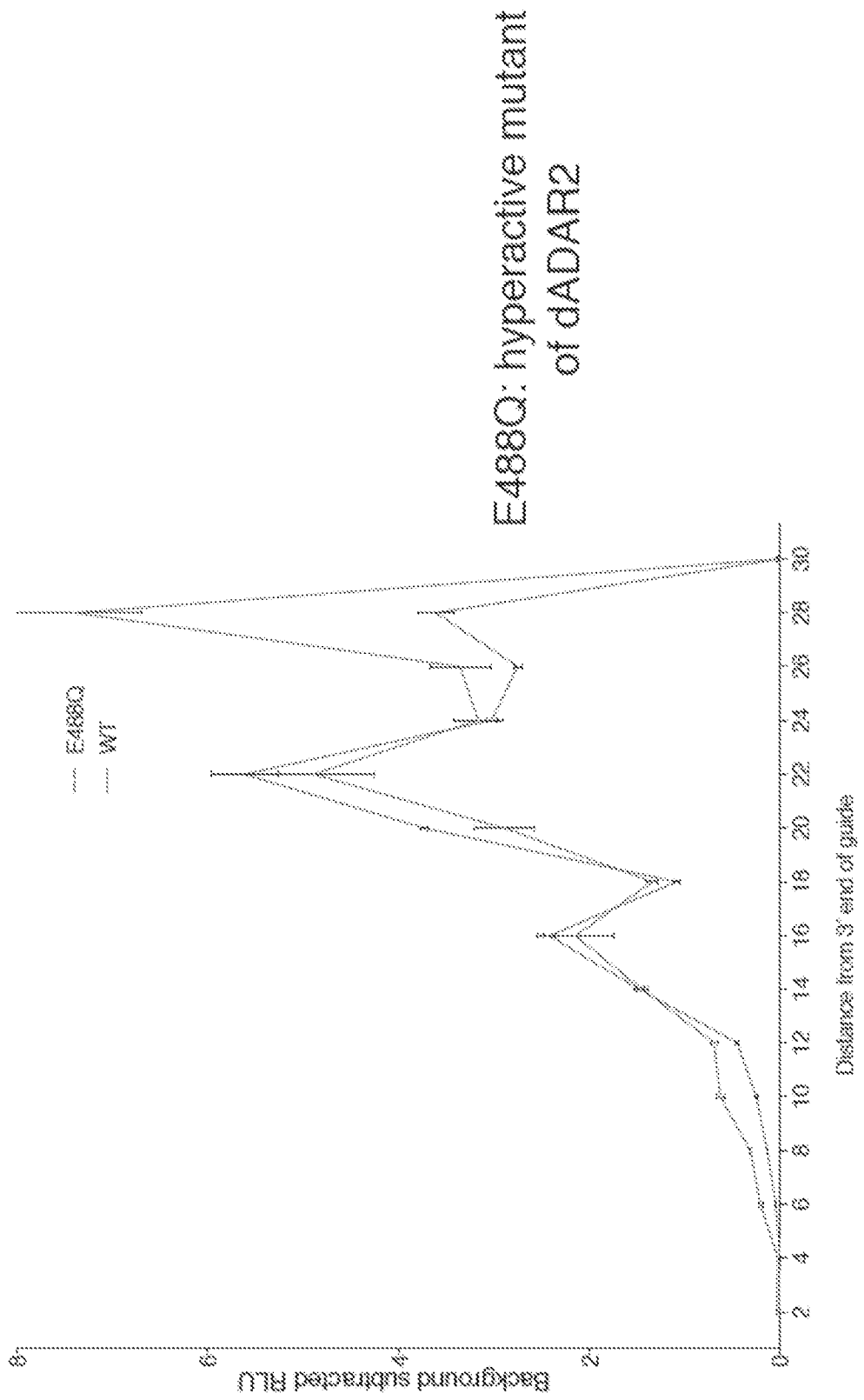

E488Q: hyperactive mutant of dADAR2
E488Q
WT
Background subtracted RLU
0
2
4
6
8
Distance from 3' end of guide
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30

FIG. 26

A

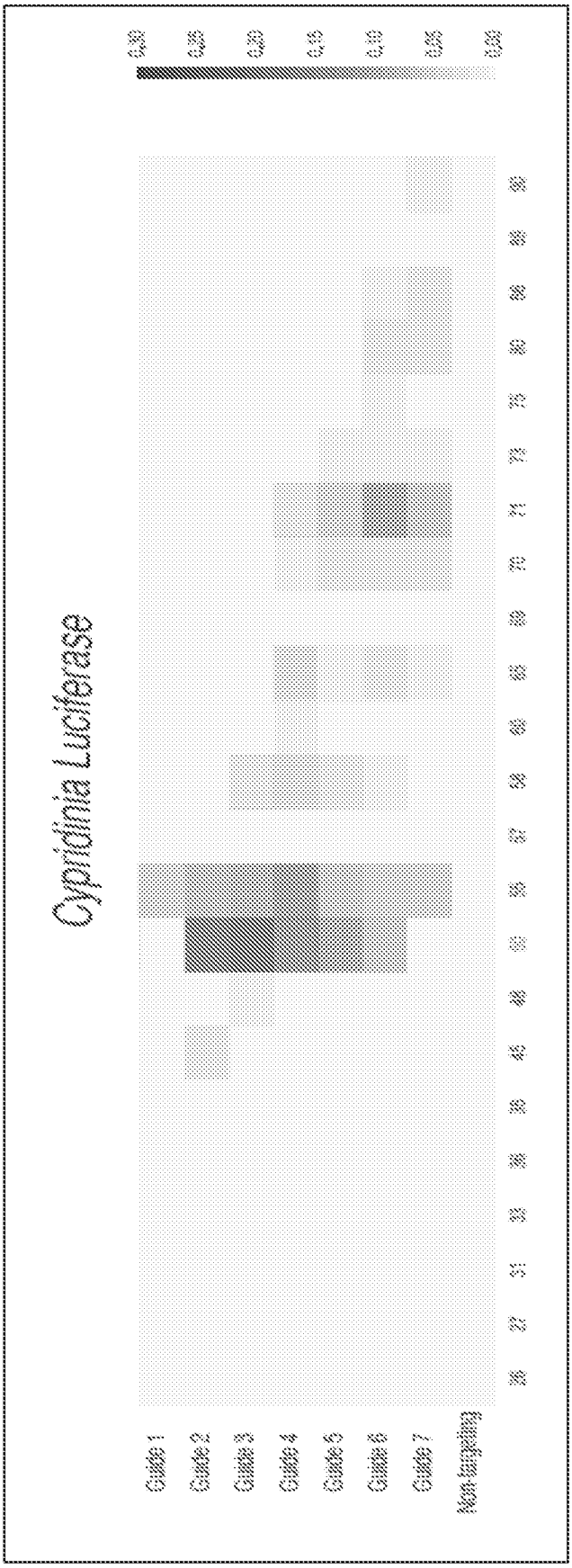

Cypridina Luciferase
Guide 1
Guide 2
Guide 3
Guide 4
Guide 5
Guide 6
Guide 7
Non-targeting

FIG. 27

B

Guide	3' start pos.	5' start pos.	A-C pos.
1	1	50	50
2	9	59	50
3	17	67	50
4	25	75	50
5	33	83	50
6	41	91	50
7	49	99	50
NT			

FIG. 27 Cont

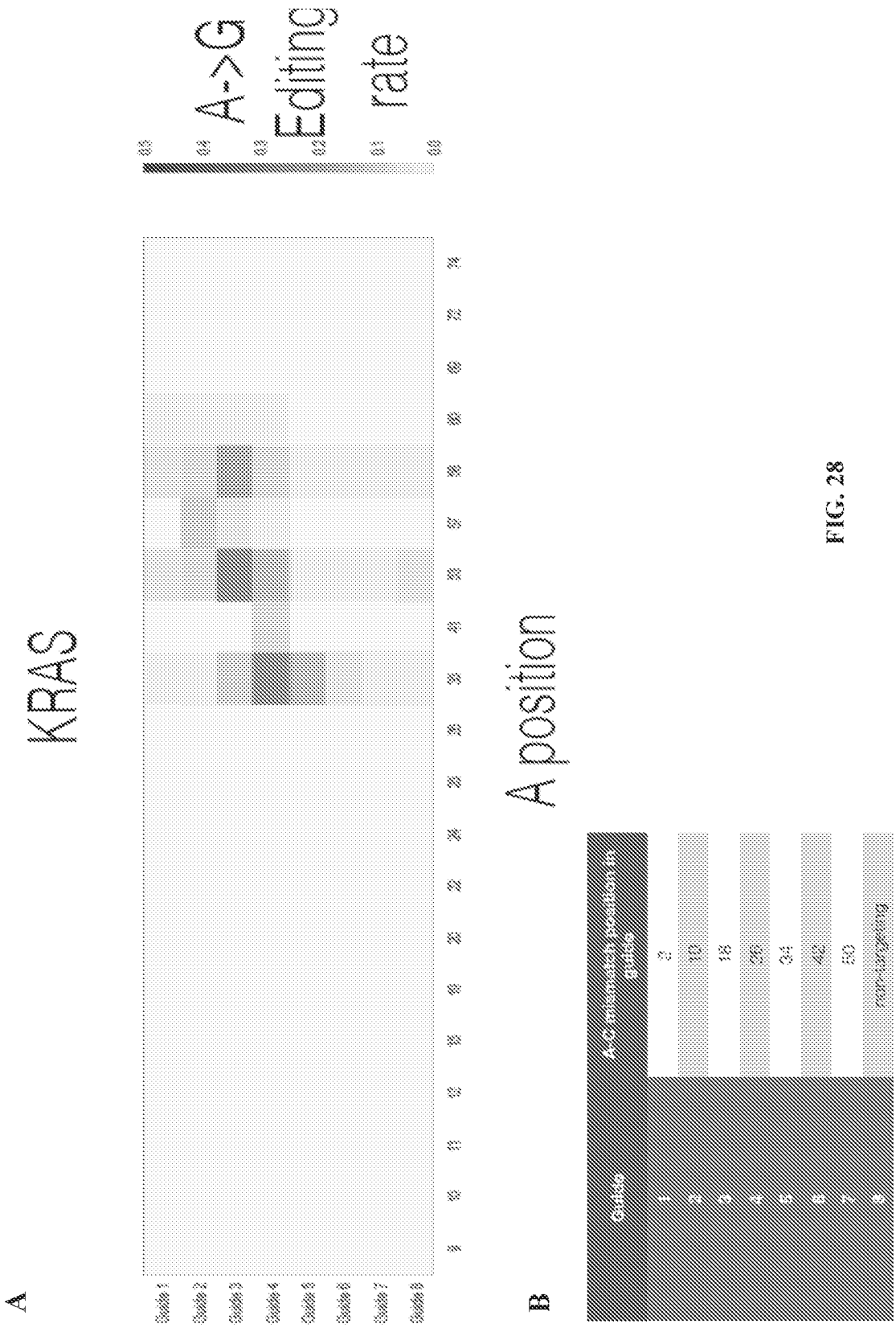

A
KRAS
A-->G Editing rate
A position
B
A-G mismatch positions in gRNA
gRNA
20
12
18
26
34
42
50
non-targeting

FIG. 28

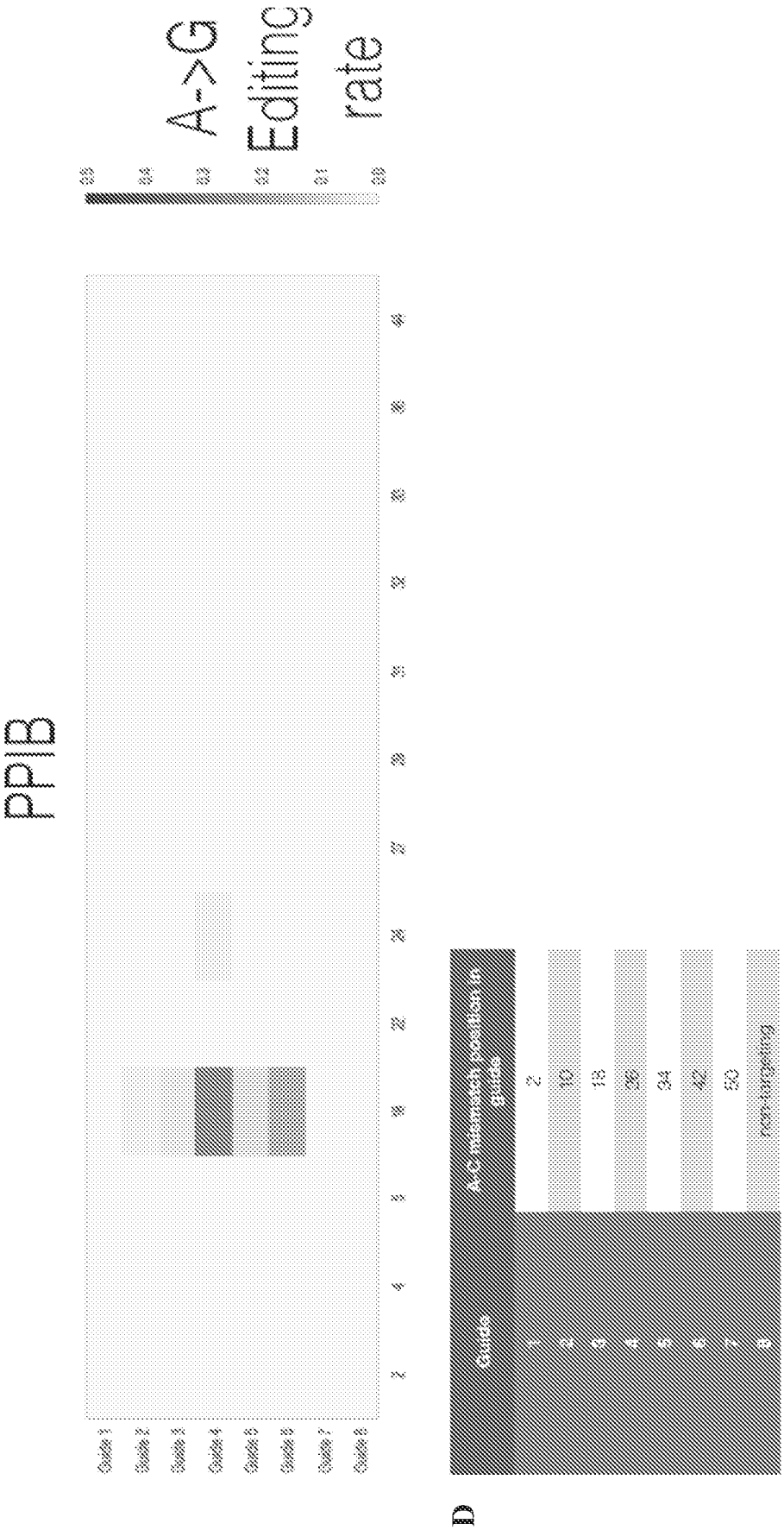
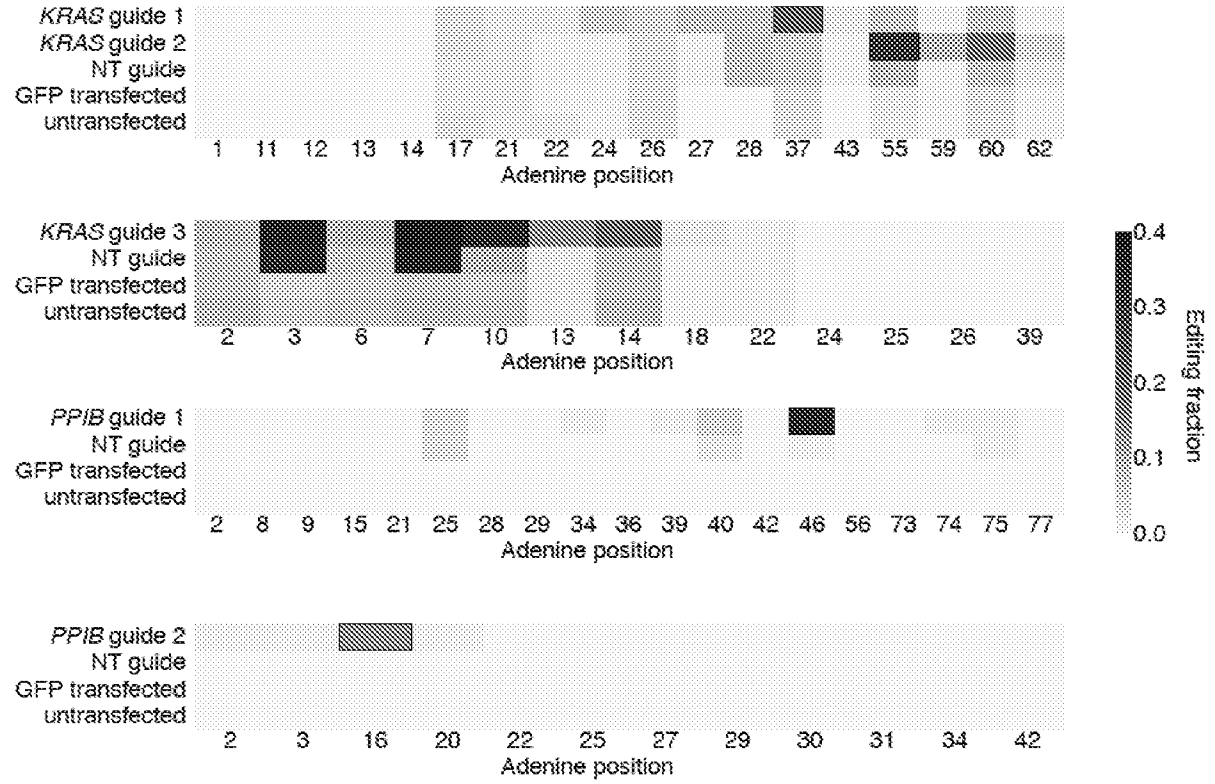

C
PPIB
A->G Editing rate
Guide 1
Guide 2
Guide 3
Guide 4
Guide 5
Guide 6
Guide 7
Guide 8
D
A>G mismatch position in guide
Guide
2
10
18
26
34
42
50
non-targeting

FIG. 28 cont.


KRAS guide 1
KRAS guide 2
NT guide
GFP transfected
untransfected
1 11 12 13 14 17 21 22 24 26 27 28 37 43 55 59 60 62
Adenine position
KRAS guide 3
NT guide
GFP transfected
untransfected
2 3 6 7 10 13 14 18 22 24 25 26 39
Adenine position
PPIB guide 1
NT guide
GFP transfected
untransfected
2 8 9 15 21 25 28 29 34 36 39 40 42 46 56 73 74 75 77
Adenine position
PPIB guide 2
NT guide
GFP transfected
untransfected
2 3 16 20 22 25 27 29 30 31 34 42
Adenine position
0.4
0.3
0.2
0.1
0.0
Editing fraction

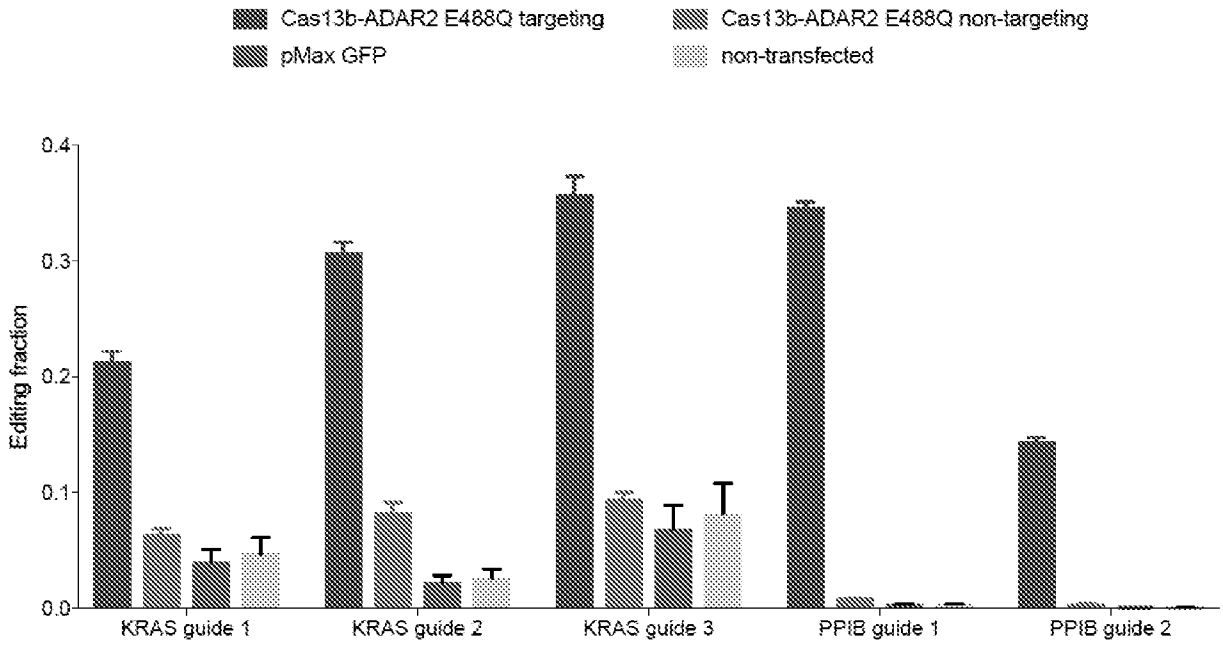

Cas13b-ADAR2 E488Q targeting
Cas13b-ADAR2 E488Q non-targeting
pMax GFP
non-transfected
0.4
0.3
0.2
0.1
0.0
Editing fraction
KRAS guide 1
KRAS guide 2
KRAS guide 3
PPIB guide 1
PPIB guide 2

FIG. 29

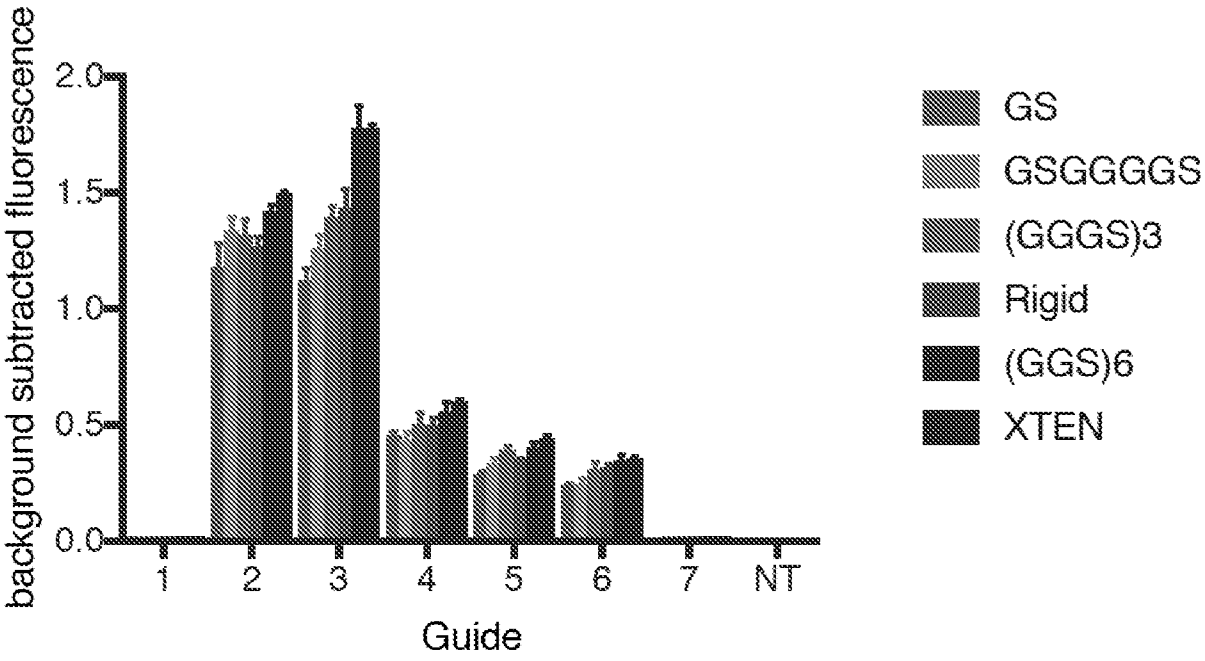

background subtracted fluorescence
2.0
1.5
1.0
0.5
0.0
1
2
3
4
5
6
7
NT
Guide
GS
GSGGGGS
(GGGS)3
Rigid
(GGS)6
XTEN

FIG. 30

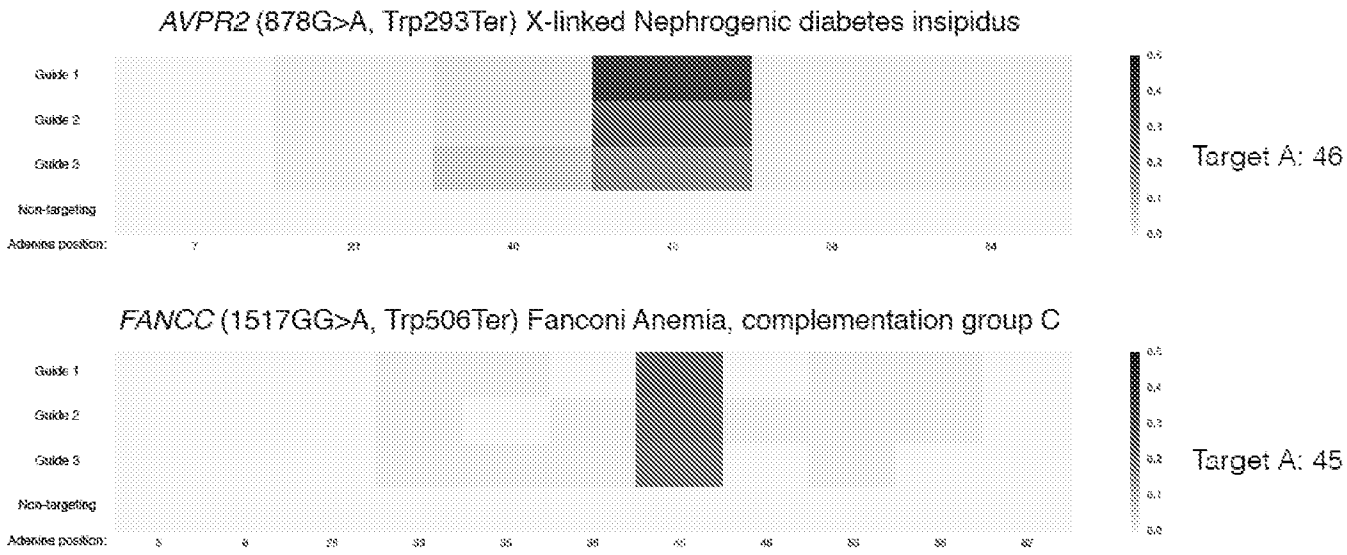

AVPR2 (878G>A, Trp293Ter) X-linked Nephrogenic diabetes insipidus
Guide 1
Guide 2
Guide 3
Non-targeting
Adenine position:
Target A: 46
FANCC (1517GG>A, Trp506Ter) Fanconi Anemia, complementation group C
Guide 1
Guide 2
Guide 3
Non-targeting
Adenine position:
Target A: 45

FIG. 31

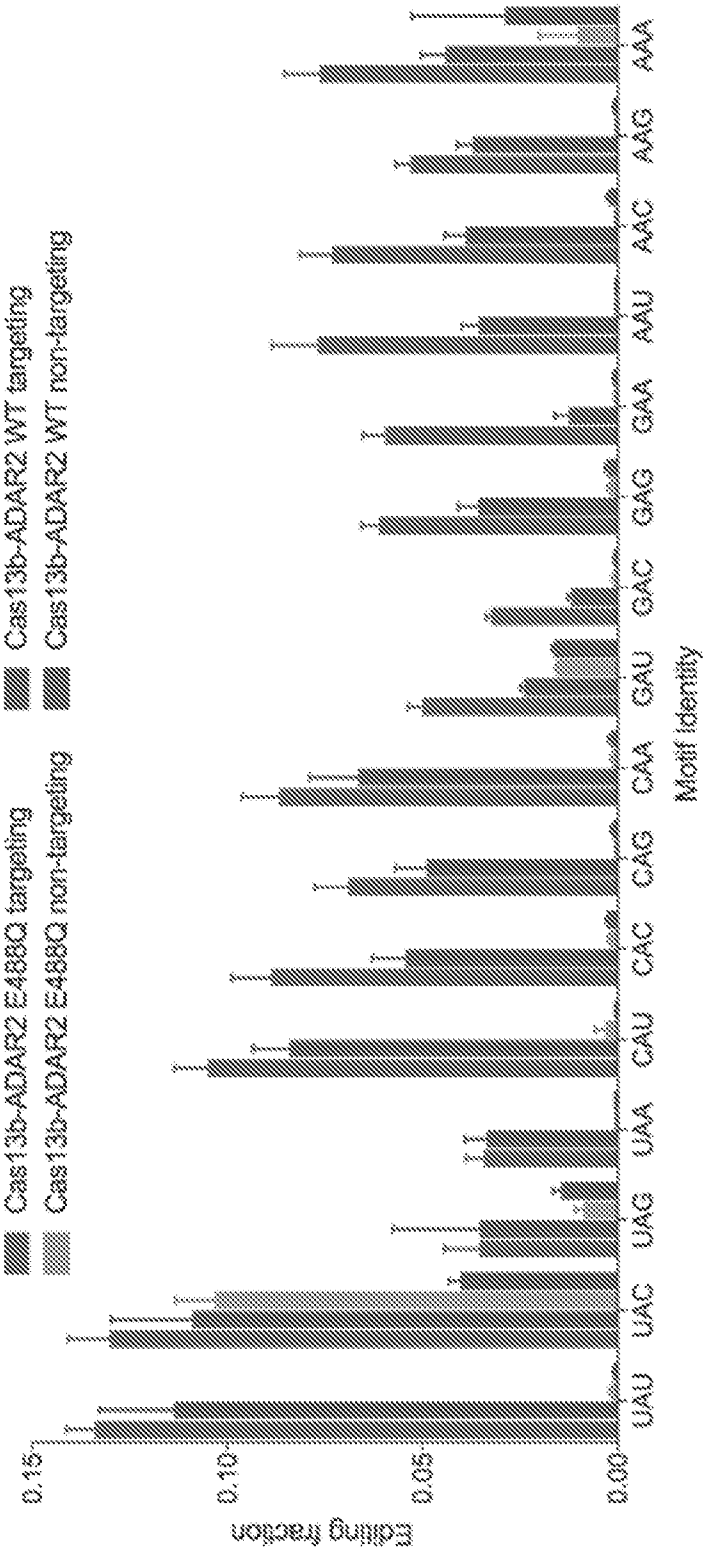
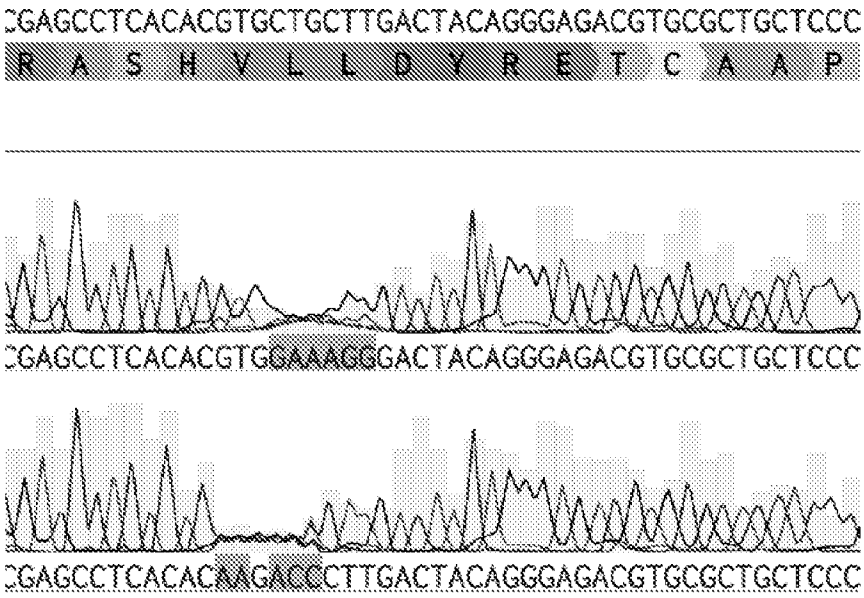

Cas13b-ADAR2 E488Q targeting
Cas13b-ADAR2 E488Q non-targeting
Cas13b-ADAR2 WT targeting
Cas13b-ADAR2 WT non-targeting
Editing fraction
0.00
0.05
0.10
0.15
UAU
UAC
UAG
UAA
CAU
CAC
CAG
CAA
GAU
GAC
GAG
GAA
AAU
AAC
AAG
AAA
Motif identity

FIG. 32

A


CGAGCCTCACACGTGCTGCTTGACTACAGGGAGACGTGCGCTGCTCCC
R A S H V L L D Y R E T C A A P
CGAGCCTCACACGTGGAAACGGACTACAGGGAGACGTGCGCTGCTCCC
CGAGCCTCACACAAGACCCTTGACTACAGGGAGACGTGCGCTGCTCCC

B

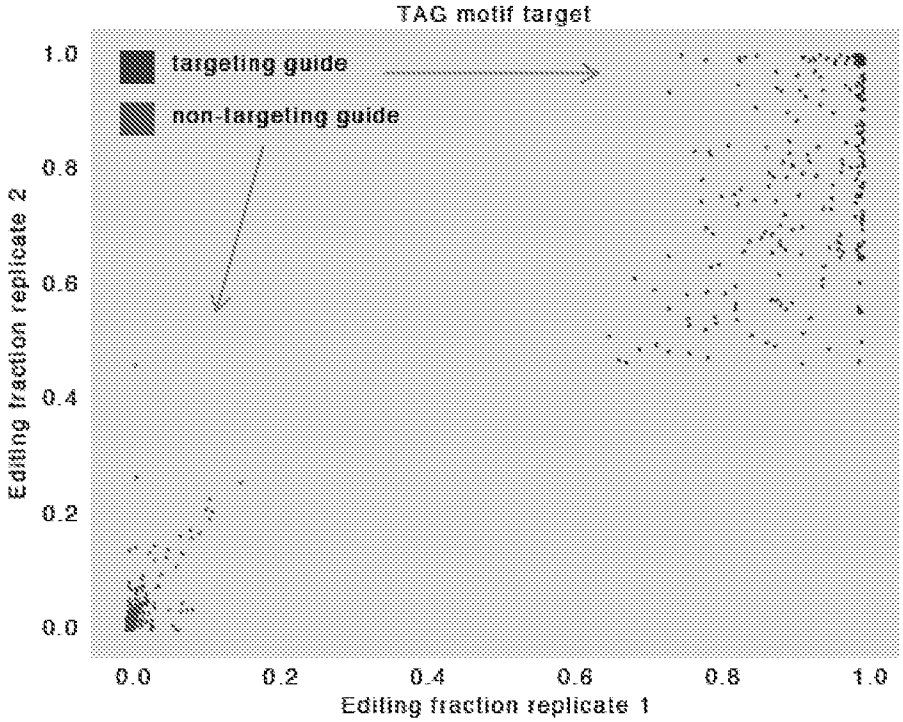
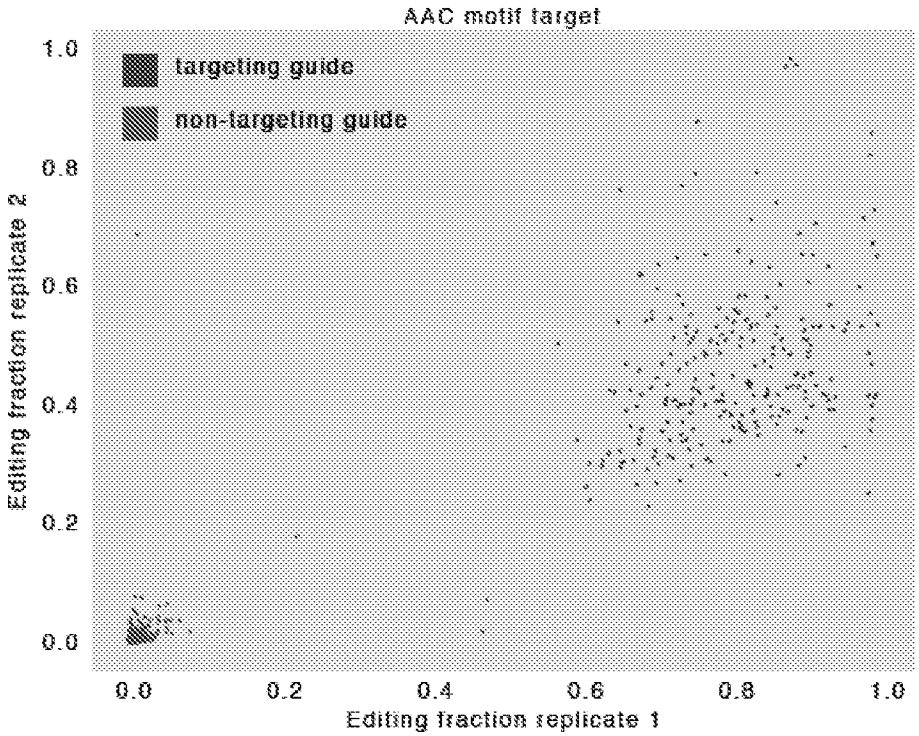

TAG motif target
targeting guide
non-targeting guide
Editing fraction replicate 2
Editing fraction replicate 1
0.0
0.2
0.4
0.6
0.8
1.0

FIG. 33

C



D

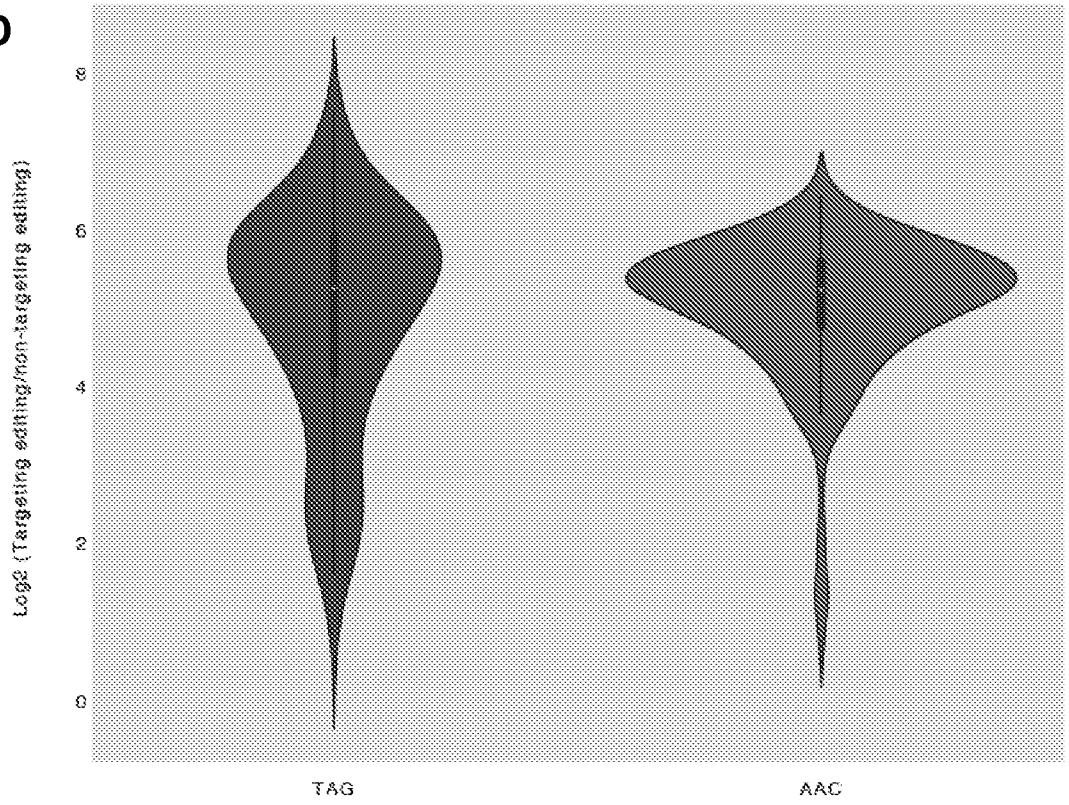


FIG. 33 (continued)

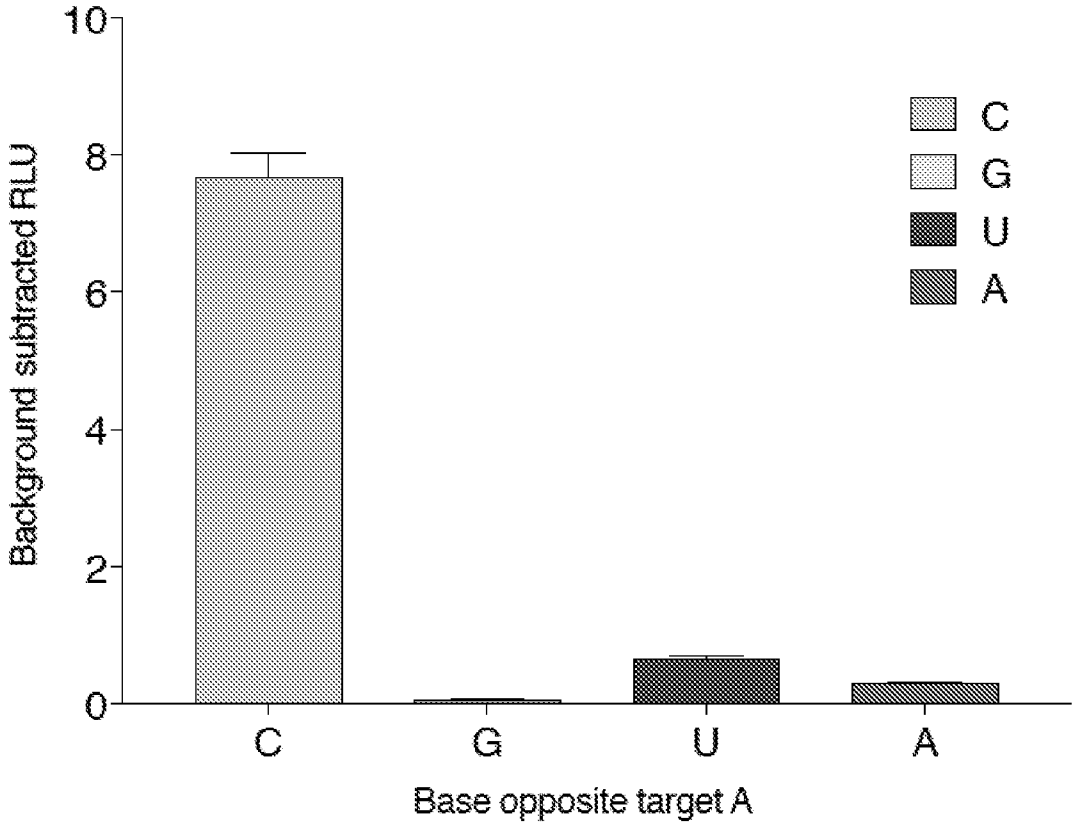

10
8
6
4
2
0
Background subtracted RLU
C
G
U
A
C
G
U
A
Base opposite target A

FIG. 34

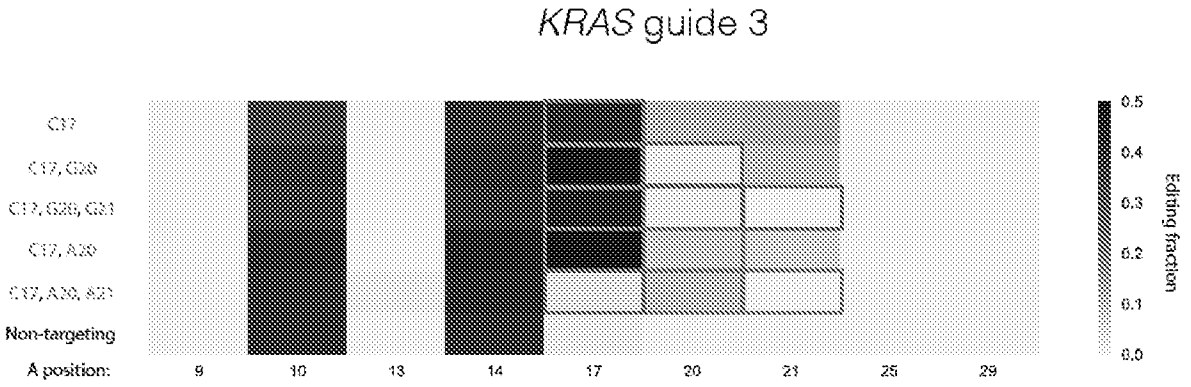
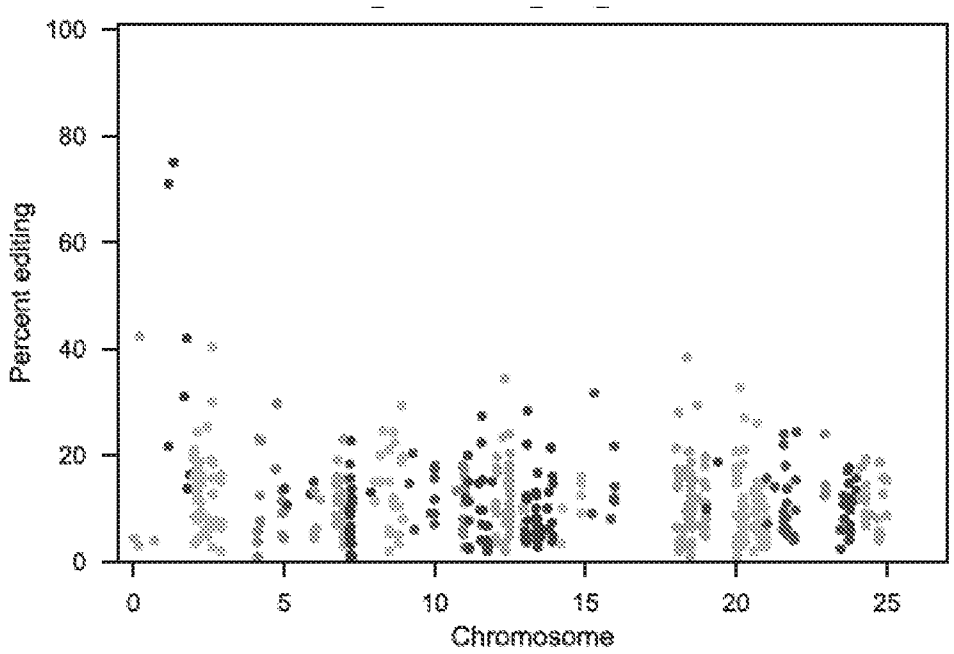

KRAS guide 3
C17
C17, G20
C17, G20, G21
C17, A20
C17, A20, A21
Non-targeting
A position:
9
10
13
14
17
20
21
25
29
0.5
0.4
0.3
0.2
0.1
0.0
Editing fraction

FIG. 35

A


100
80
60
40
20
0
Percent editing
0
5
10
15
20
25
Chromosome

B

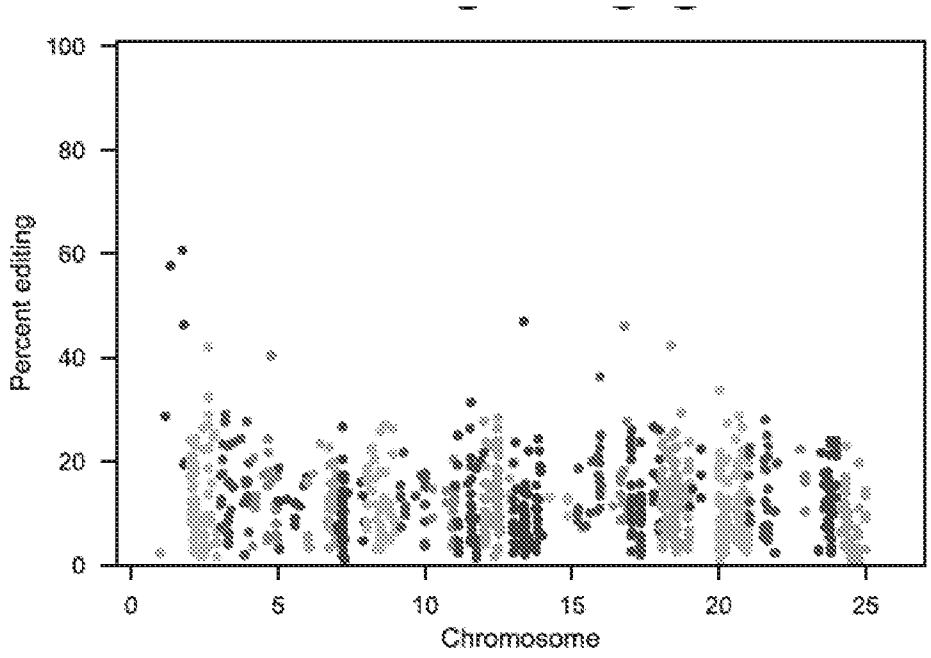
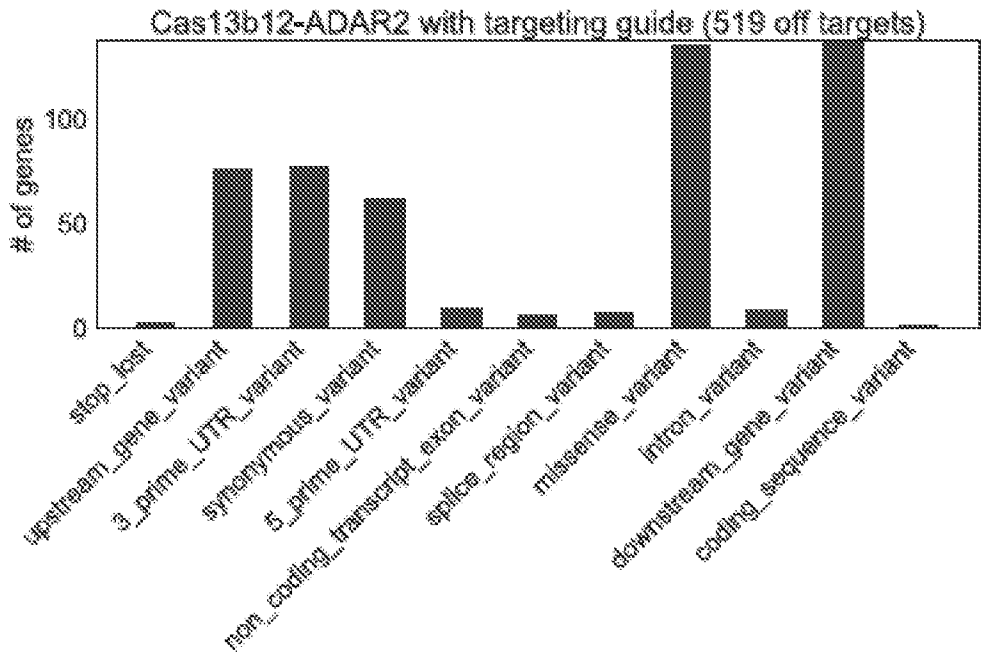

100
80
60
40
20
0
Percent editing
0
5
10
15
20
25
Chromosome

FIG. 36

A


Cas13b12-ADAR2 with targeting guide (519 off targets)
of genes
100
50
0
stop_lost
upstream_gene_variant
3_prime_UTR_variant
synonymous_variant
5_prime_UTR_variant
non_coding_transcript_exon_variant
splice_region_variant
missense_variant
intron_variant
downstream_gene_variant
coding_sequence_variant

B

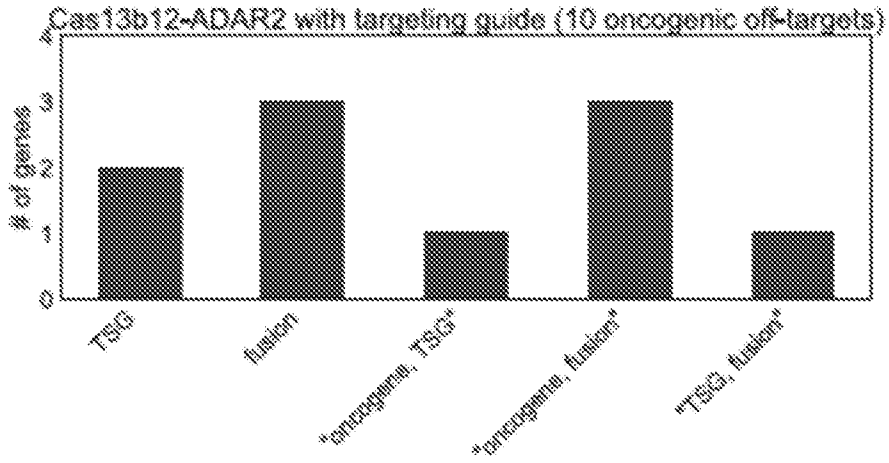
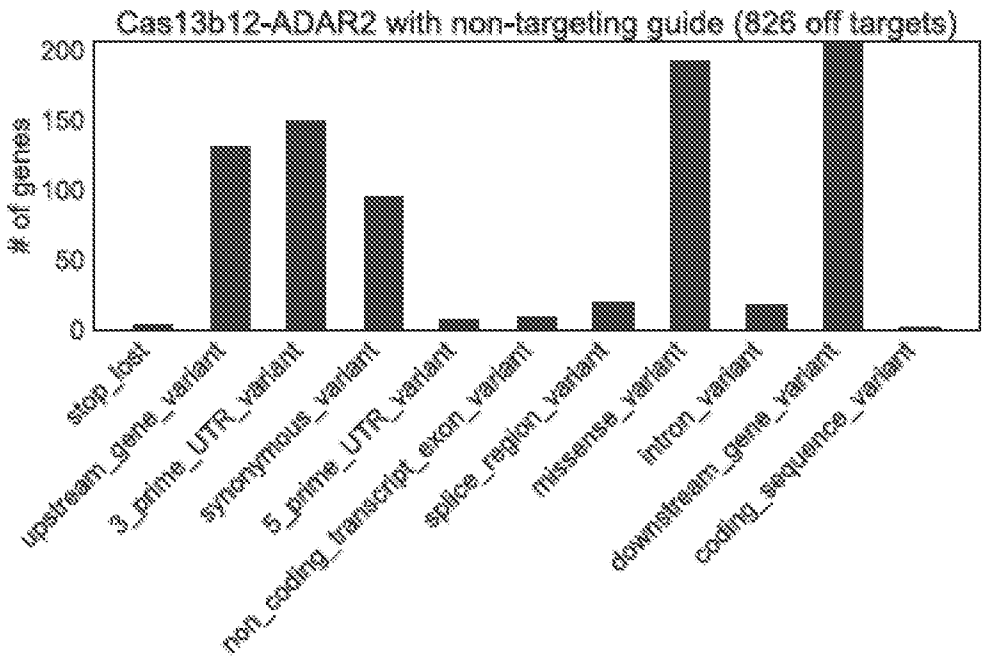

Cas13b12-ADAR2 with targeting guide (10 oncogenic off-targets)
of genes
3
2
1
0
TSG
fusion
*oncogene, TSG"
*oncogene, fusion"
*TSG, fusion"

FIG. 37

C



D

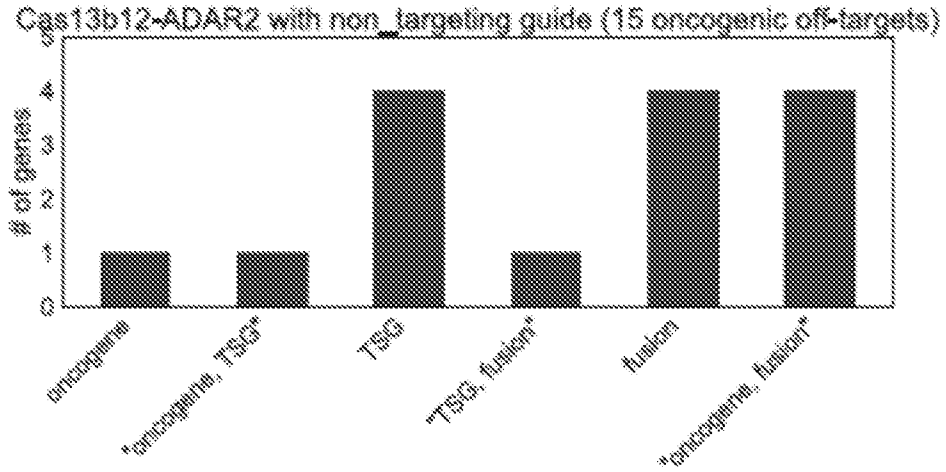


FIG. 37 (continued)

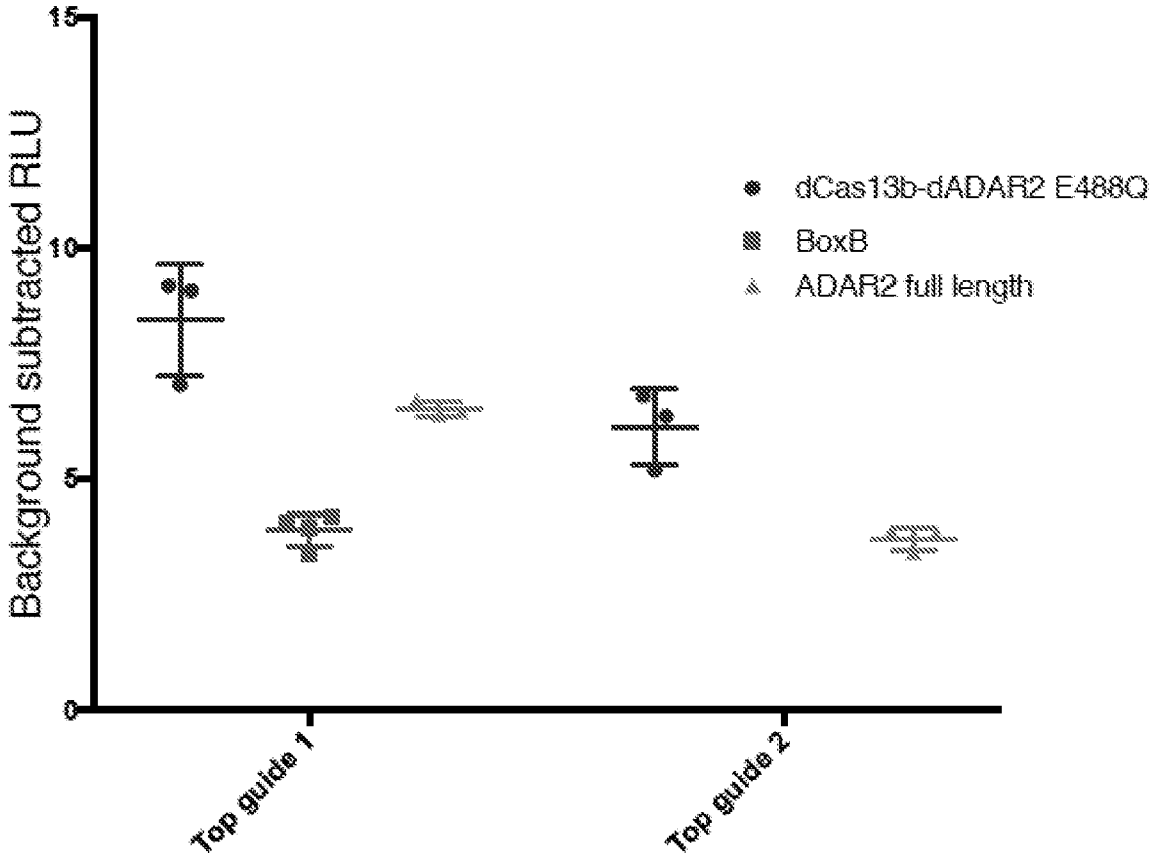
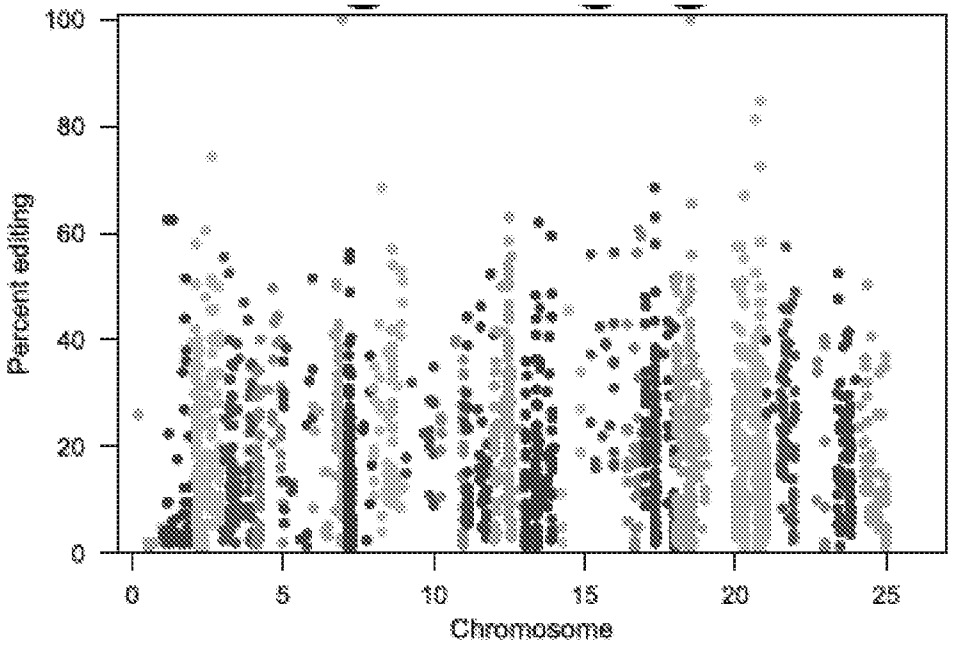
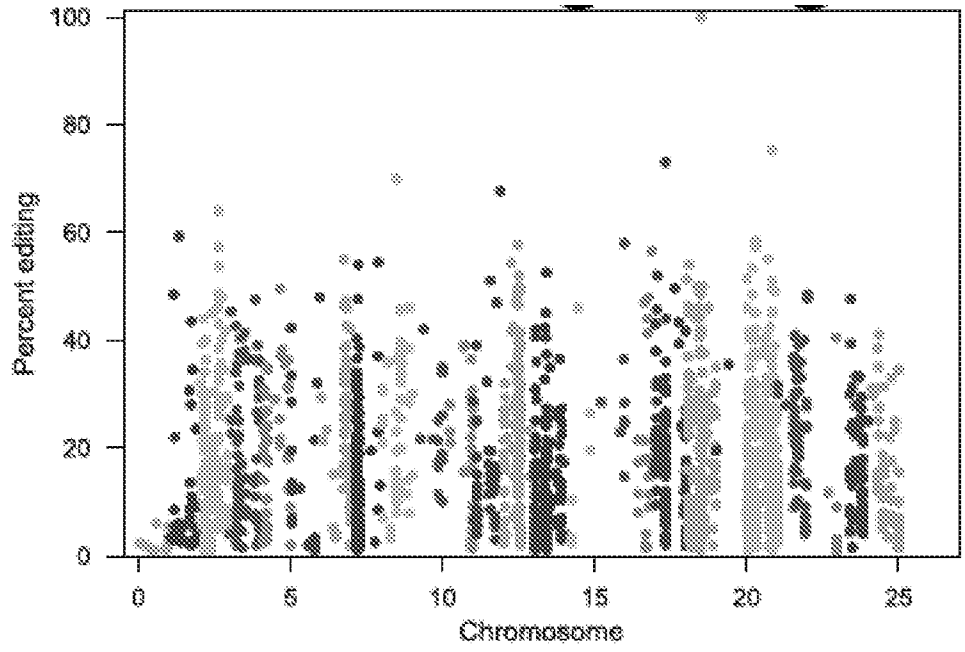

15
10
5
0
Background subtracted RLU
dCas13b-dADAR2 E488Q
BoxB
ADAR2 full length
Top guide 1
Top guide 2

FIG. 38

A


Percent editing
100
80
60
40
20
0
0
5
10
15
20
25
Chromosome

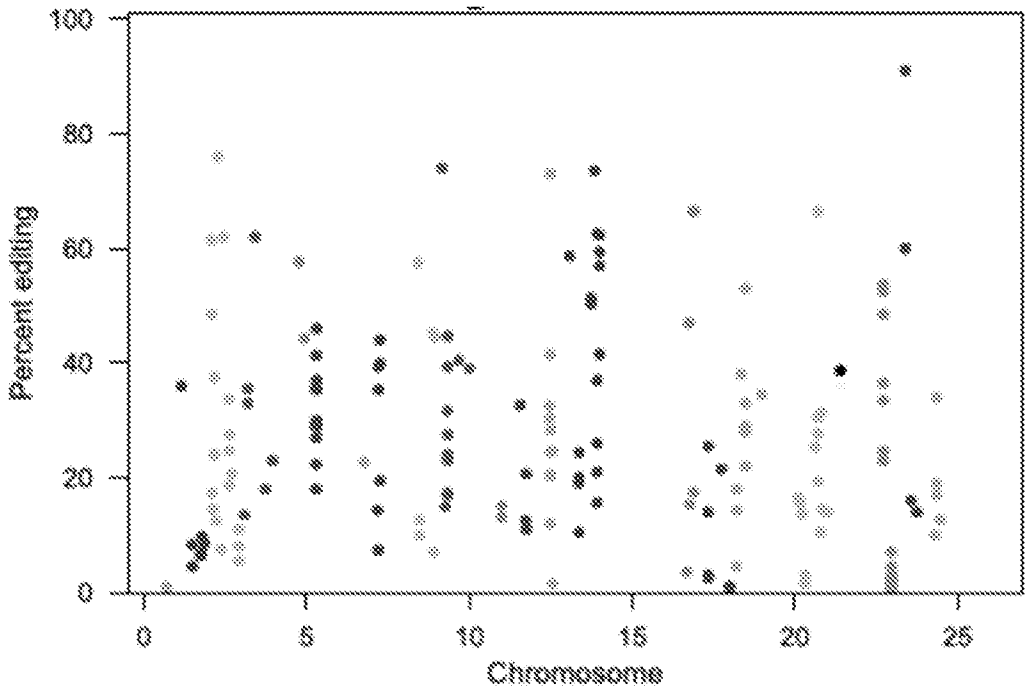
B


Percent editing
100
80
60
40
20
0
0
5
10
15
20
25
Chromosome

39

FIG.

C



D

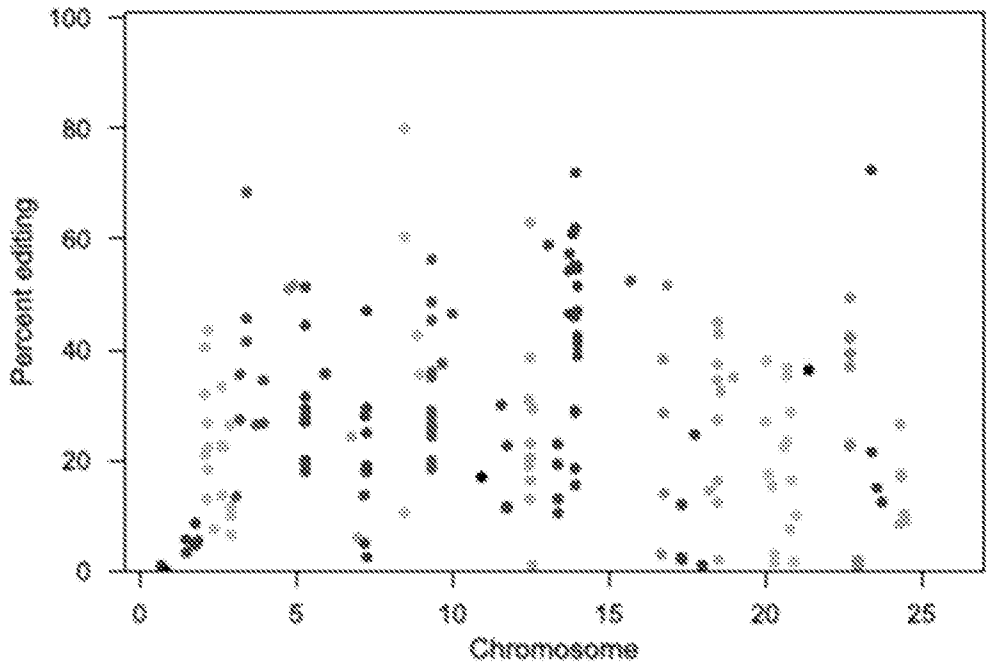


FIG. 39 (continued)

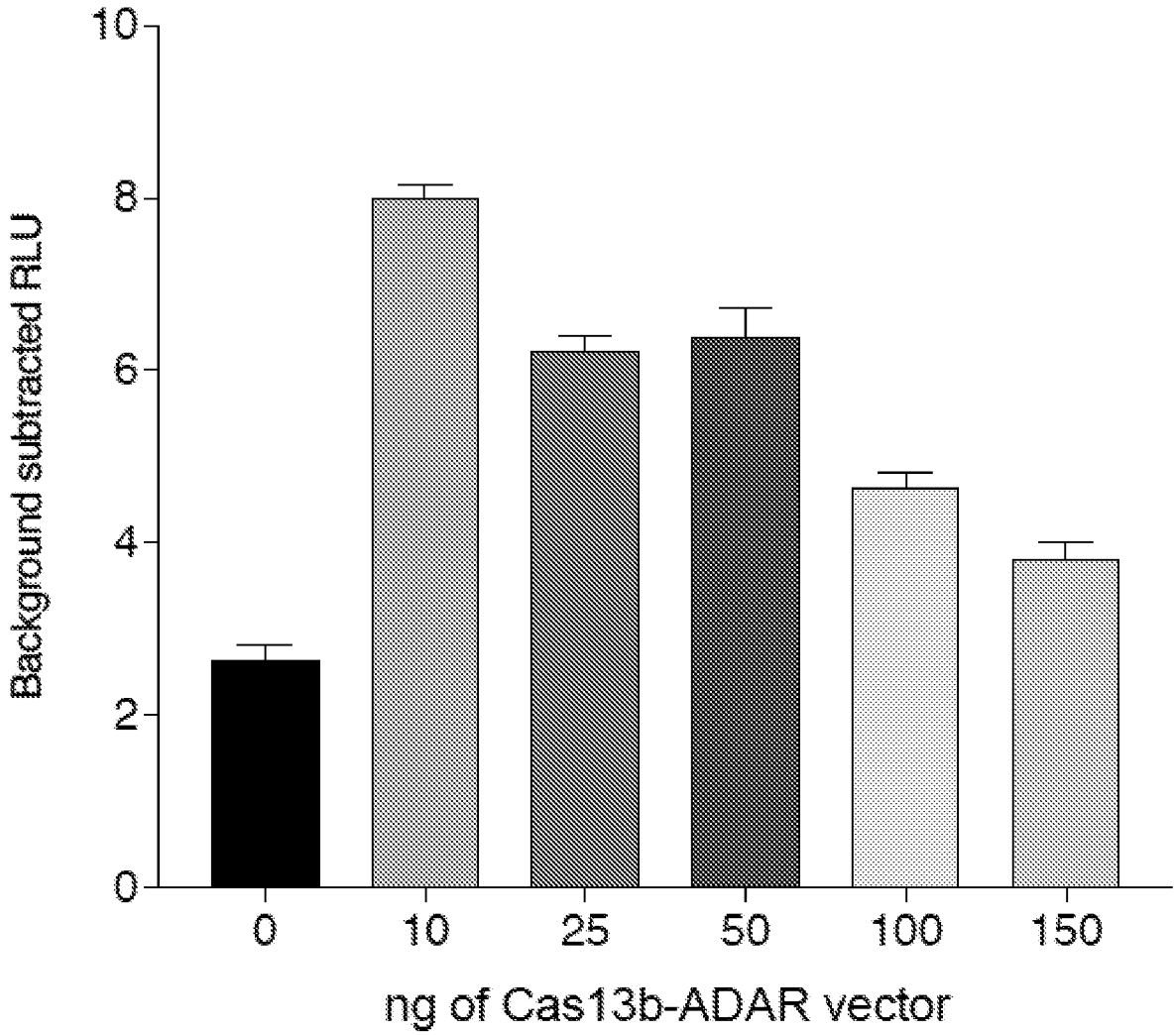
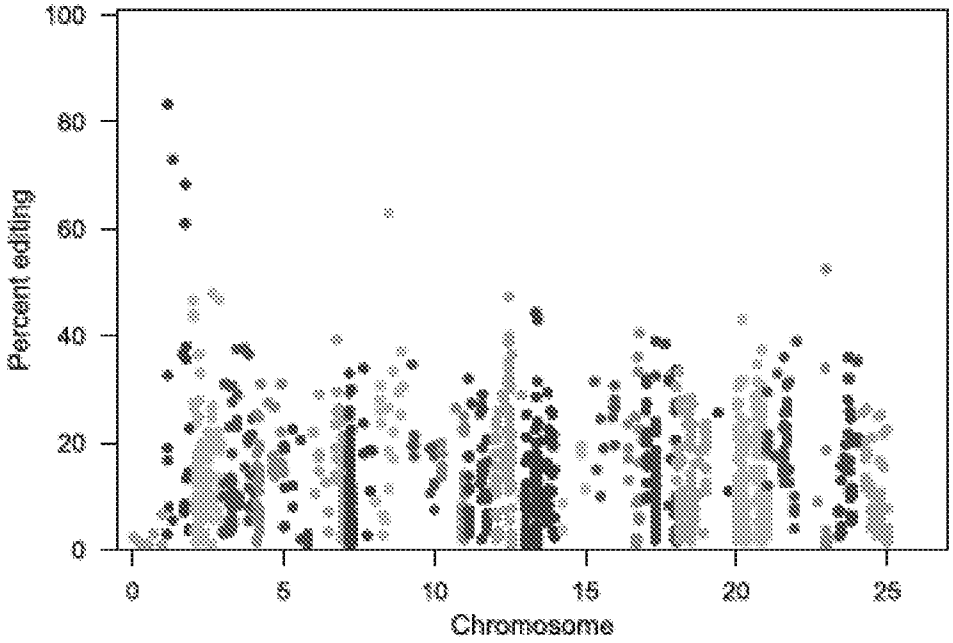

Background subtracted RLU
10
8
6
4
2
0
0
10
25
50
100
150
ng of Cas13b-ADAR vector

FIG. 40

A


100
80
60
40
20
0
Percent editing
0
5
10
15
20
25
Chromosome

B

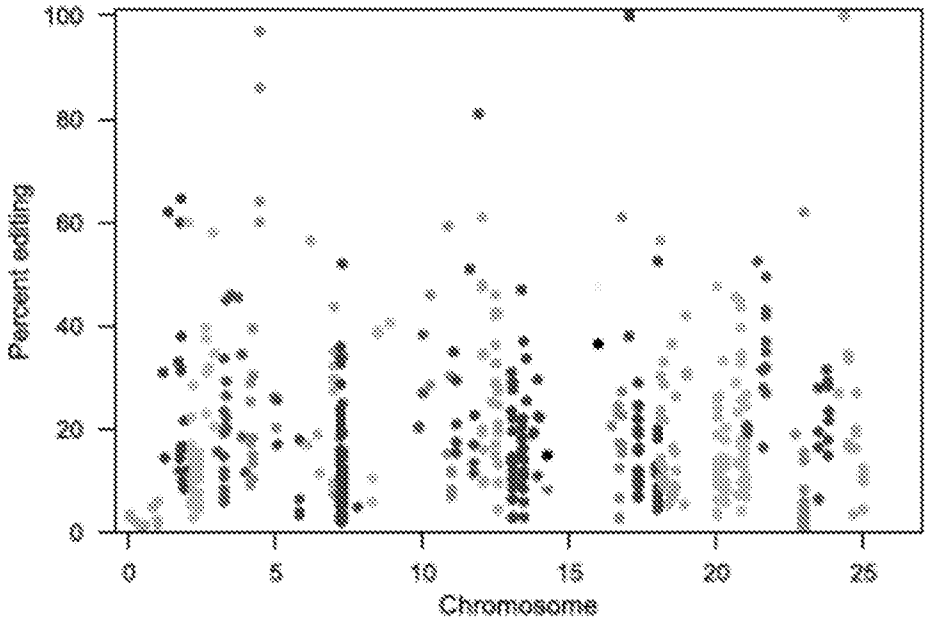
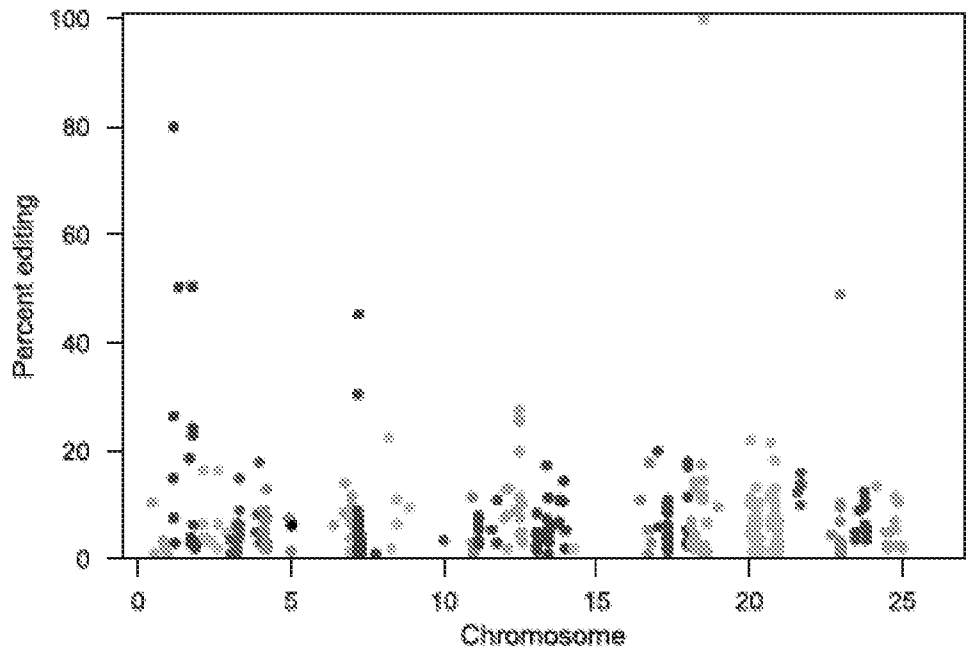

100
80
60
40
20
0
Percent editing
0
5
10
15
20
25
Chromosome

FIG. 41

C



D

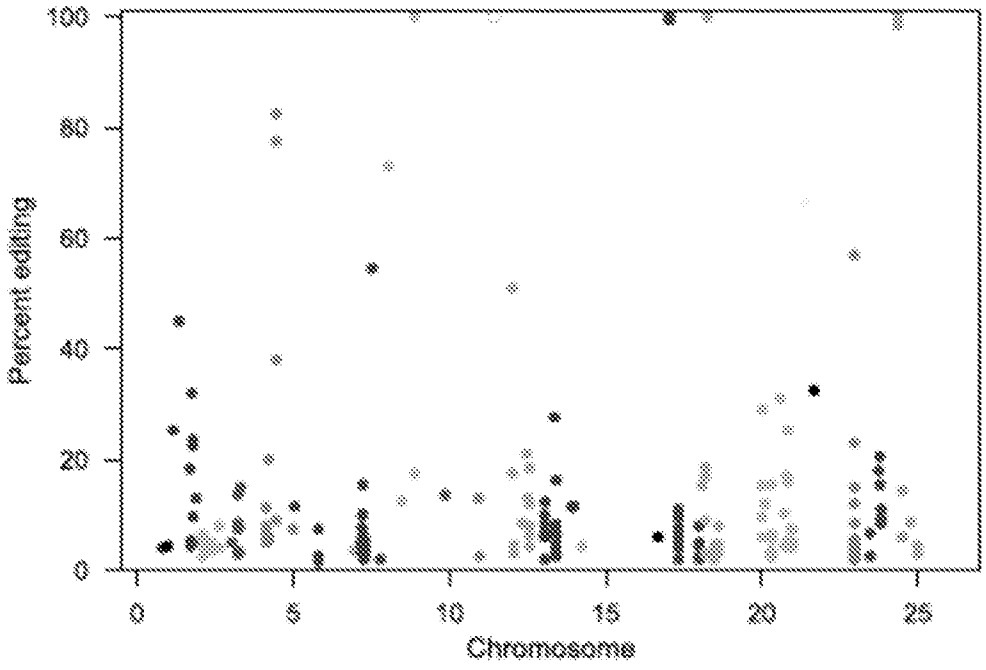
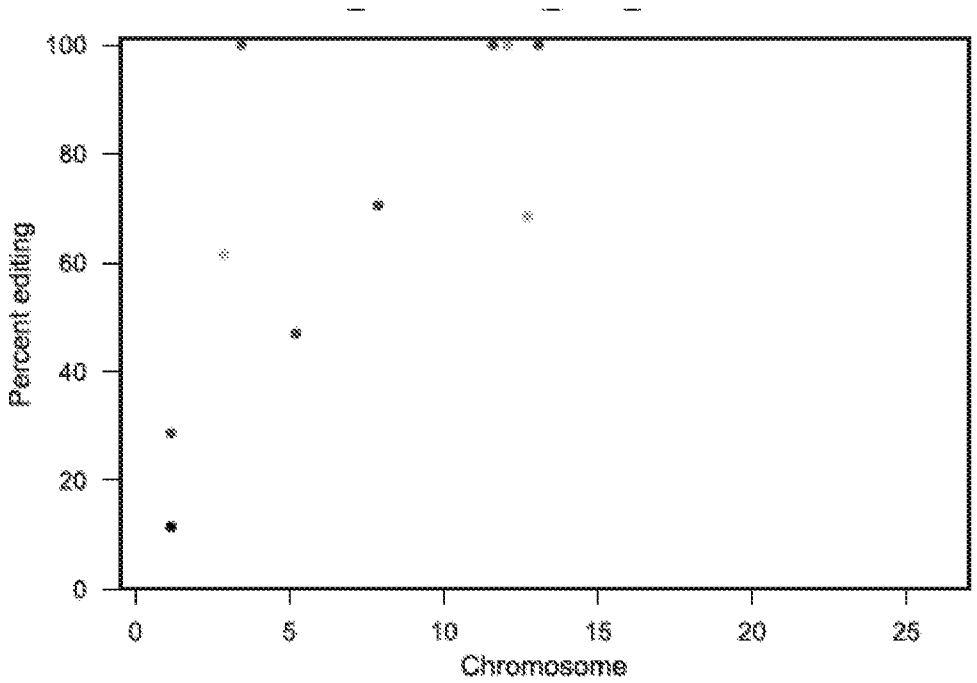


FIG. 41 (continued)

A


Percent editing
100
80
60
40
20
0
0
5
10
15
20
25
Chromosome

B

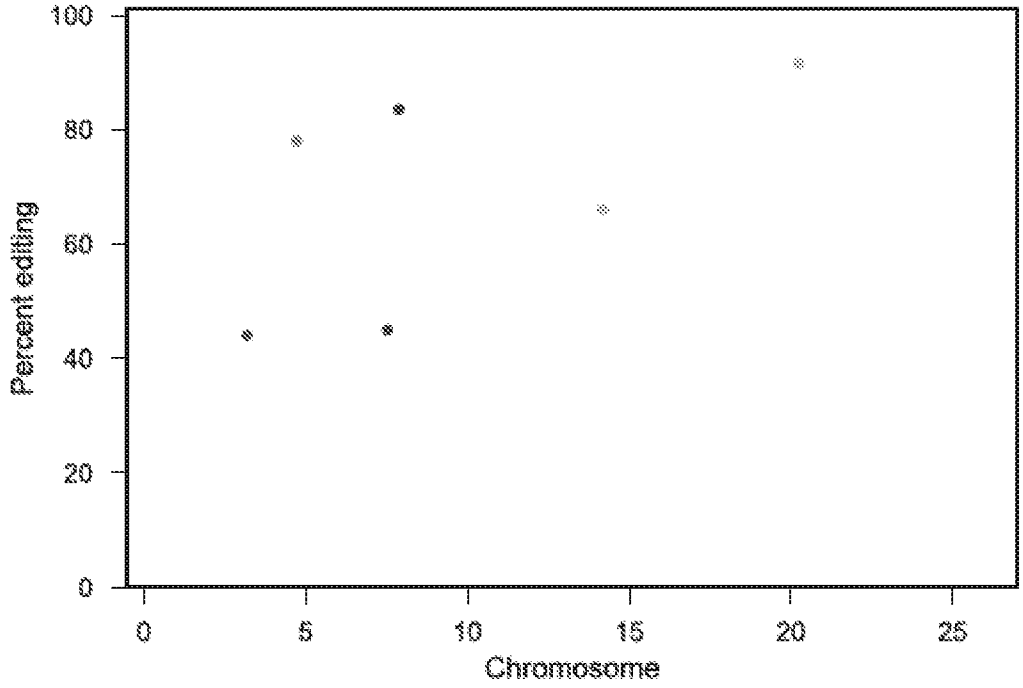

Percent editing
100
80
60
40
20
0
0
5
10
15
20
25
Chromosome

FIG. 42

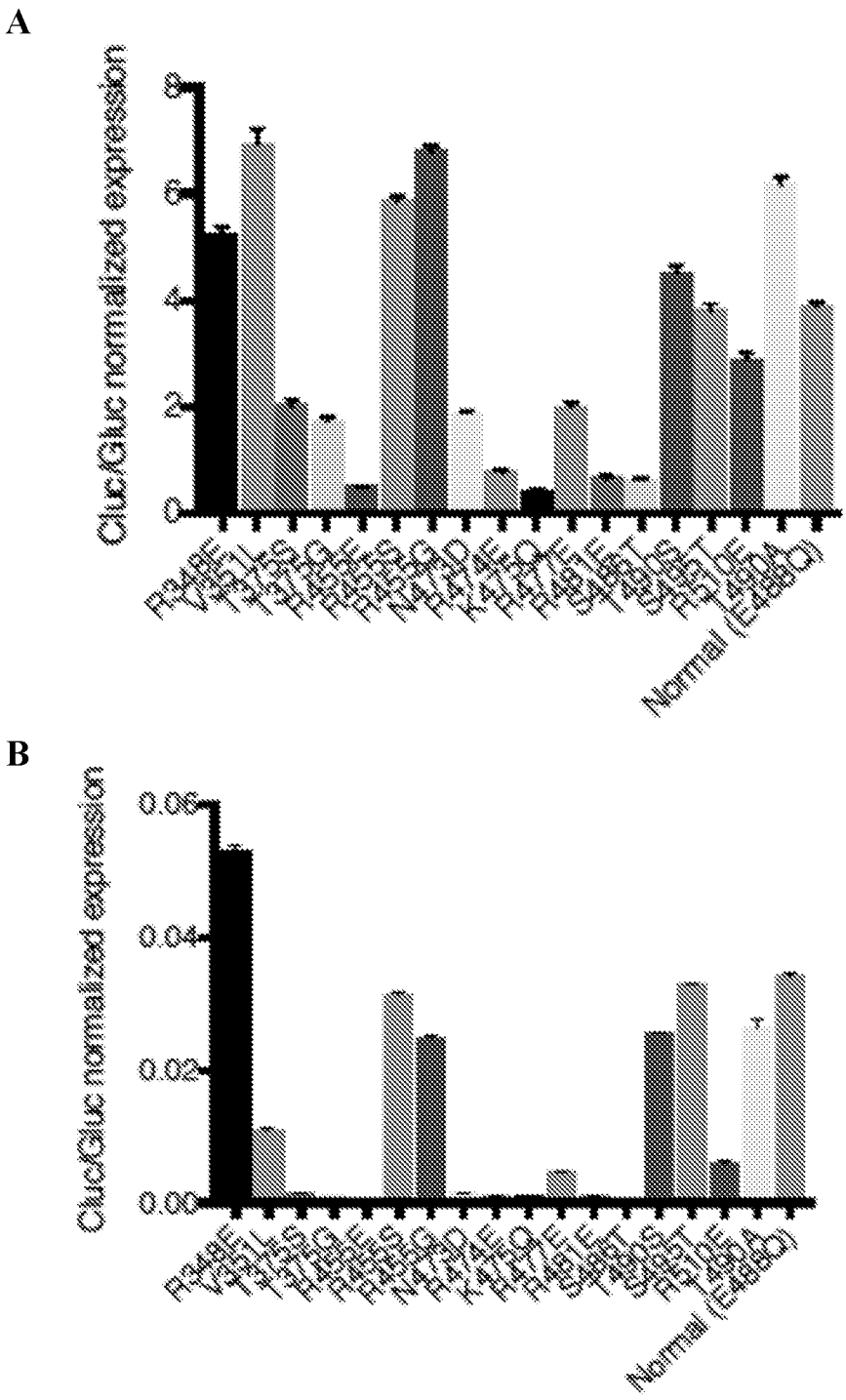
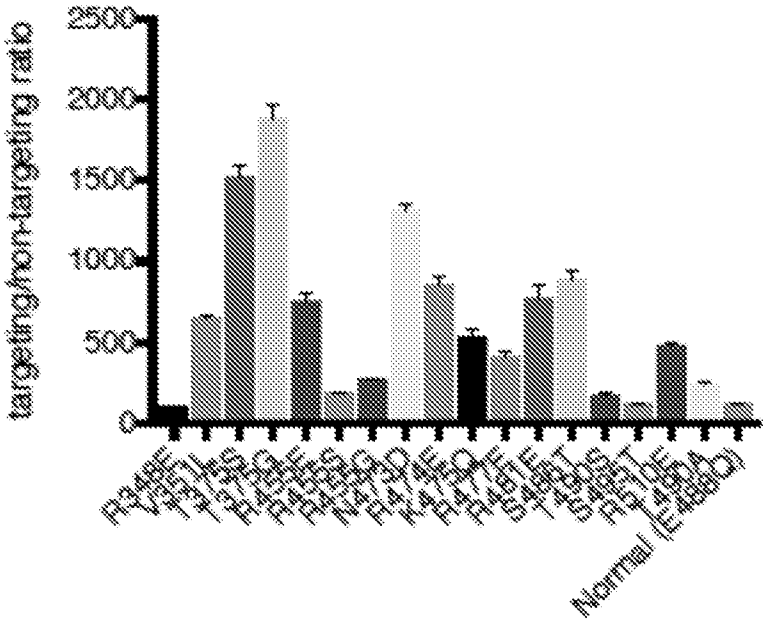

A
Cluc/Gluc normalized expression
0
2
4
6
8
Normal (E4B6)
B
Cluc/Gluc normalized expression
0.00
0.02
0.04
0.06
Normal (E4B6)

FIG. 43

C



D

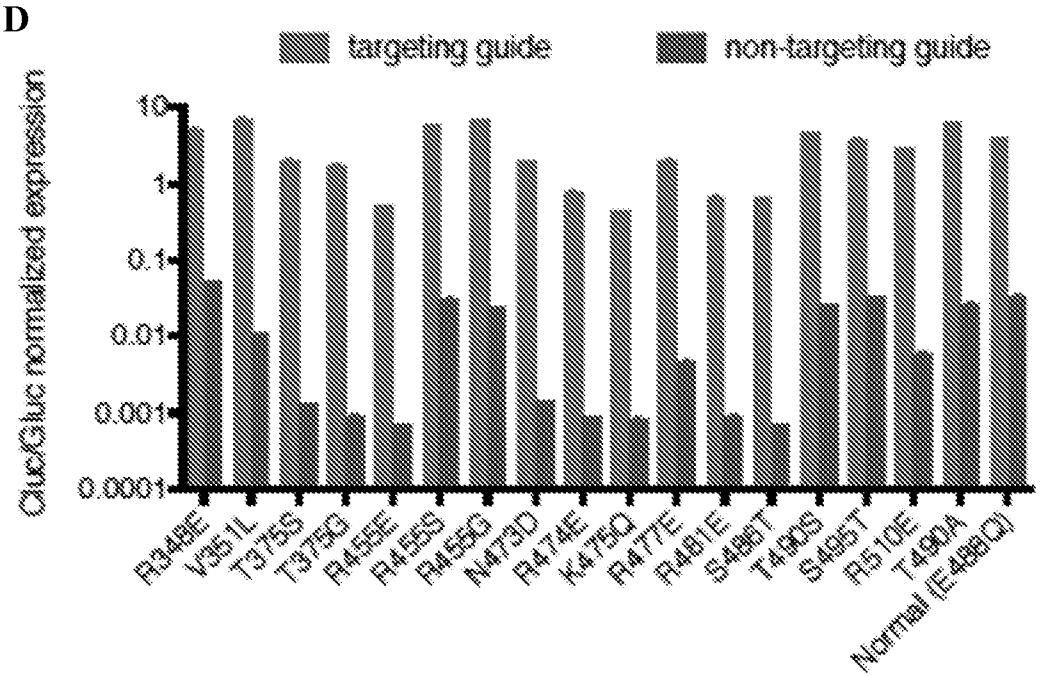


FIG. 43 (continued)

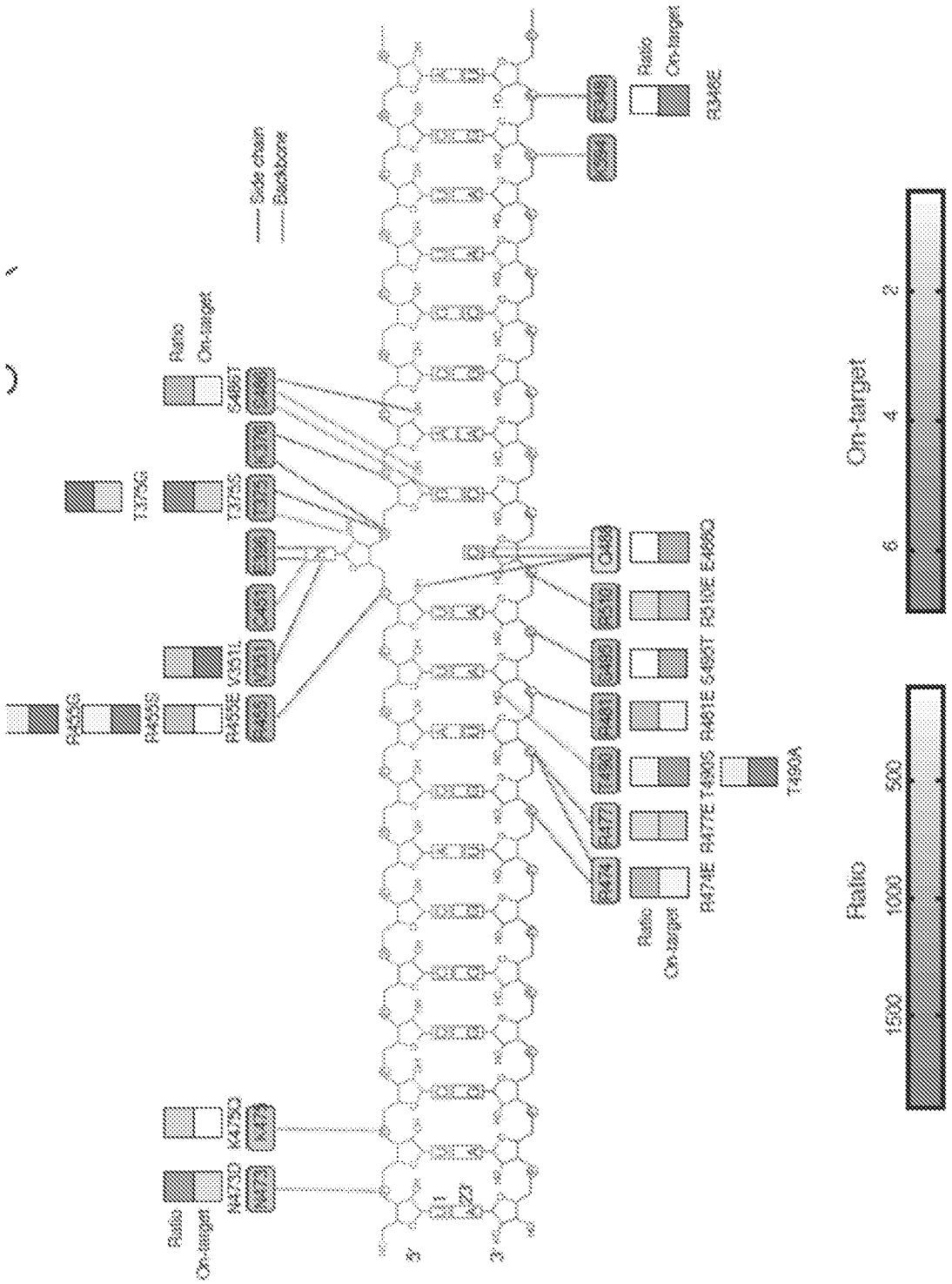

Side chain
Backbone
Ratio
On-target
On-target
Ratio
2
4
6
500
1000
1500

FIG. 44

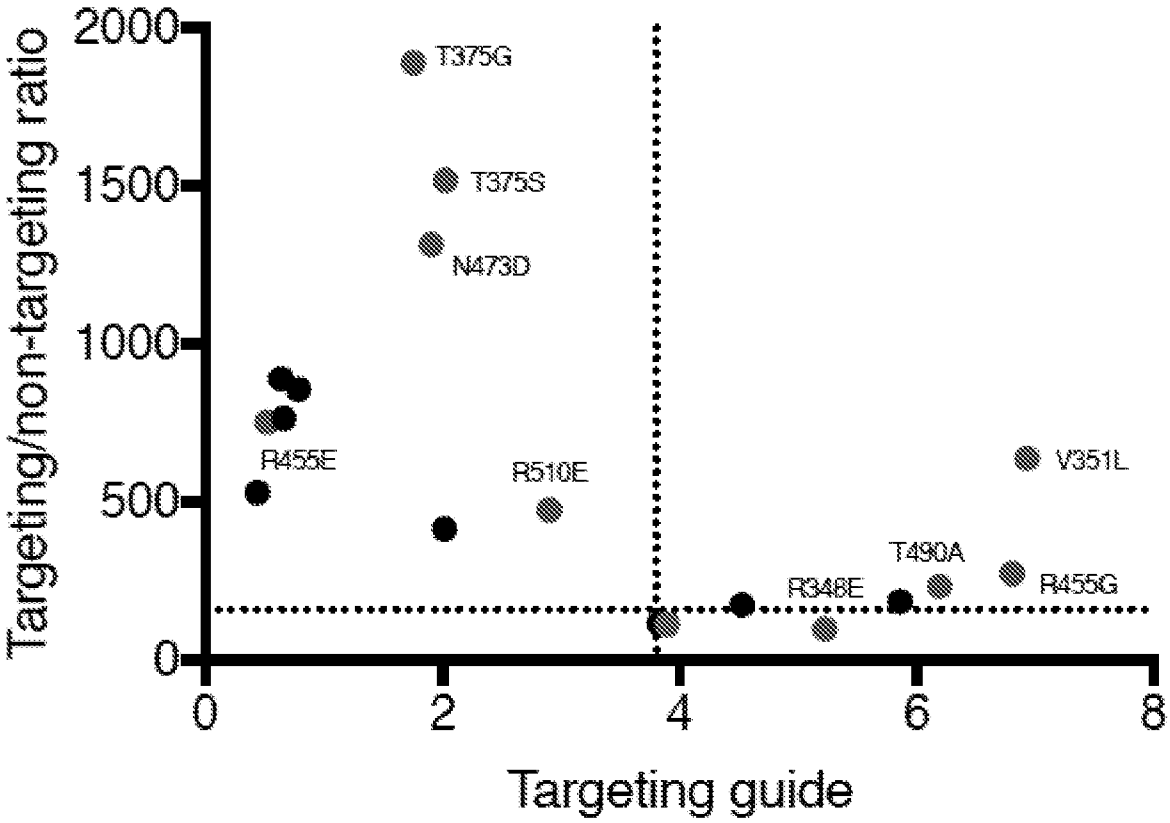

Targeting/non-targeting ratio
2000
1500
1000
500
0
T375G
T375S
N473D
R455E
R510E
V351L
T490A
R348E
R455G
0
2
4
6
8
Targeting guide

FIG. 45

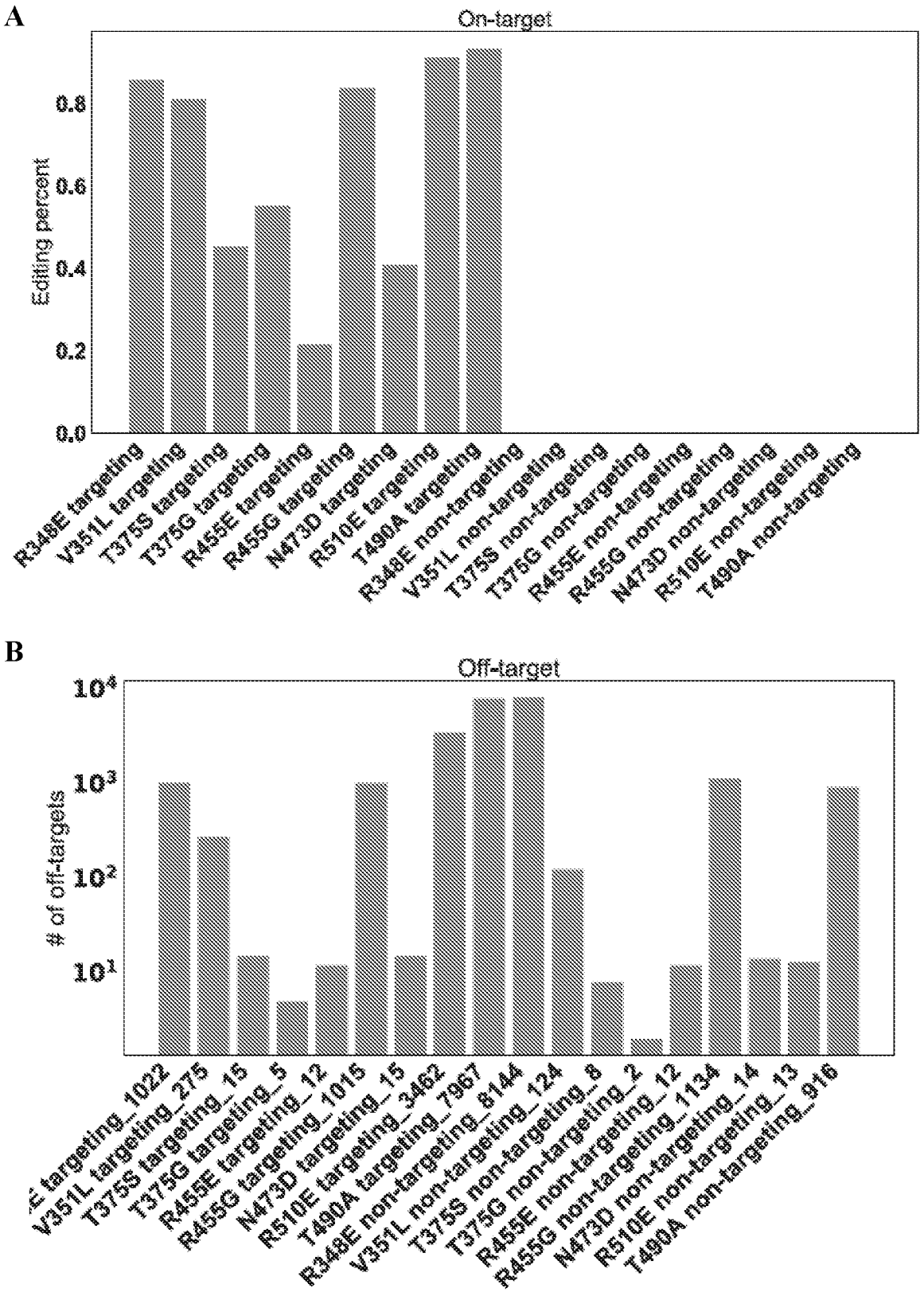

A
On-target
Editing percent
0.0
0.2
0.4
0.6
0.8
R348E targeting
V351L targeting
T375S targeting
T375G targeting
R455E targeting
R455G targeting
N473D targeting
R510E targeting
T490A targeting
R348E non-targeting
V351L non-targeting
T375S non-targeting
T375G non-targeting
R455E non-targeting
R455G non-targeting
N473D non-targeting
R510E non-targeting
T490A non-targeting
B
Off-target
of off-targets
10^1
10^2
10^3
10^4
V351L targeting_275
T375S targeting_15
T375G targeting_5
R455E targeting_12
R455G targeting_1015
N473D targeting_15
R510E targeting_3462
T490A targeting_7967
R348E non-targeting_8144
V351L non-targeting_124
T375S non-targeting_8
T375G non-targeting_2
R455E non-targeting_12
R455G non-targeting_1134
N473D non-targeting_14
R510E non-targeting_13
T490A non-targeting_916

FIG. 46

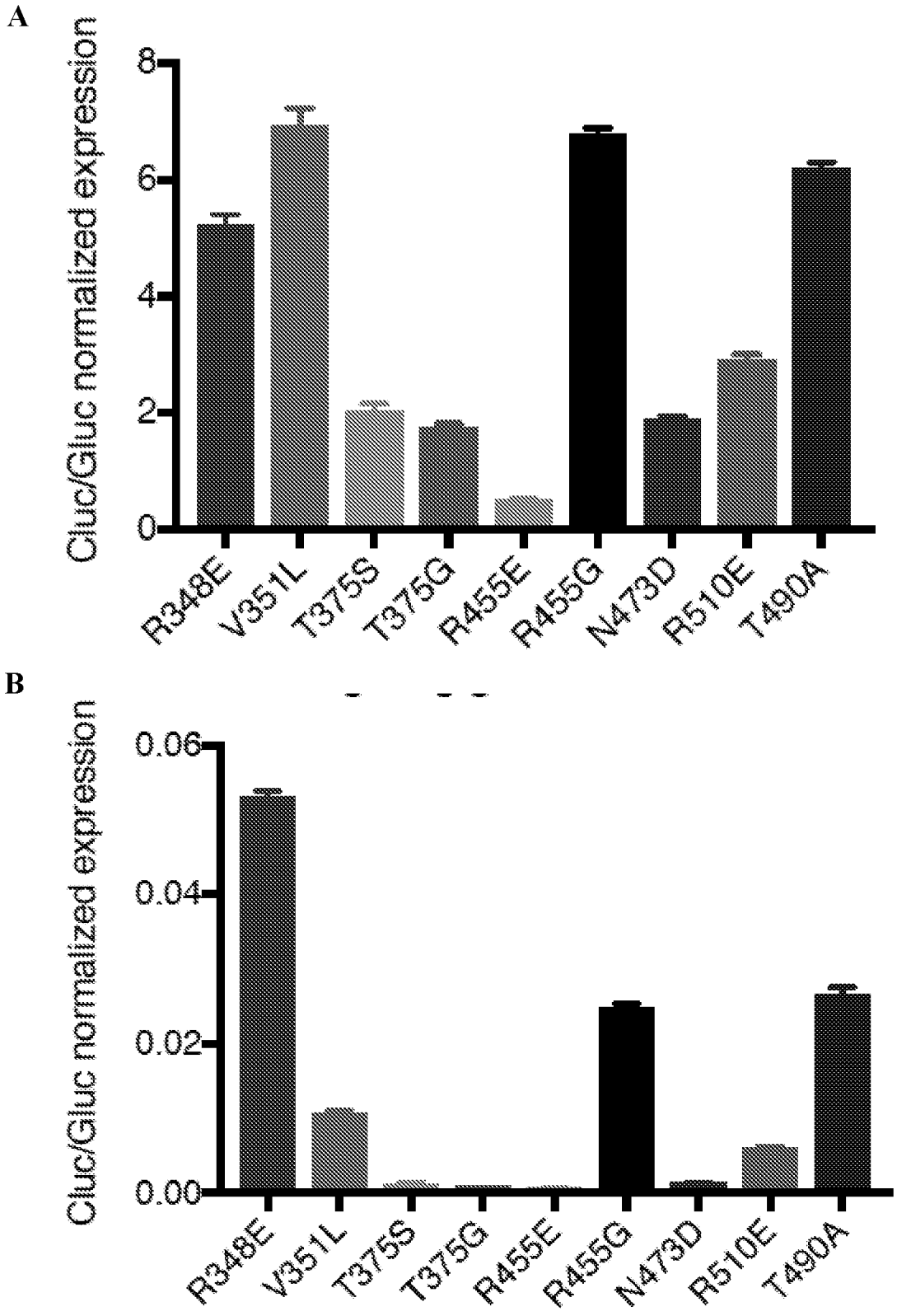

A
Cluc/Gluc normalized expression
0
2
4
6
8
R348E
V351L
T375S
T375G
R455E
R455G
N473D
R510E
T490A
B
Cluc/Gluc normalized expression
0.00
0.02
0.04
0.06
R348E
V351L
T375S
T375G
R455E
R455G
N473D
R510E
T490A

FIG. 47

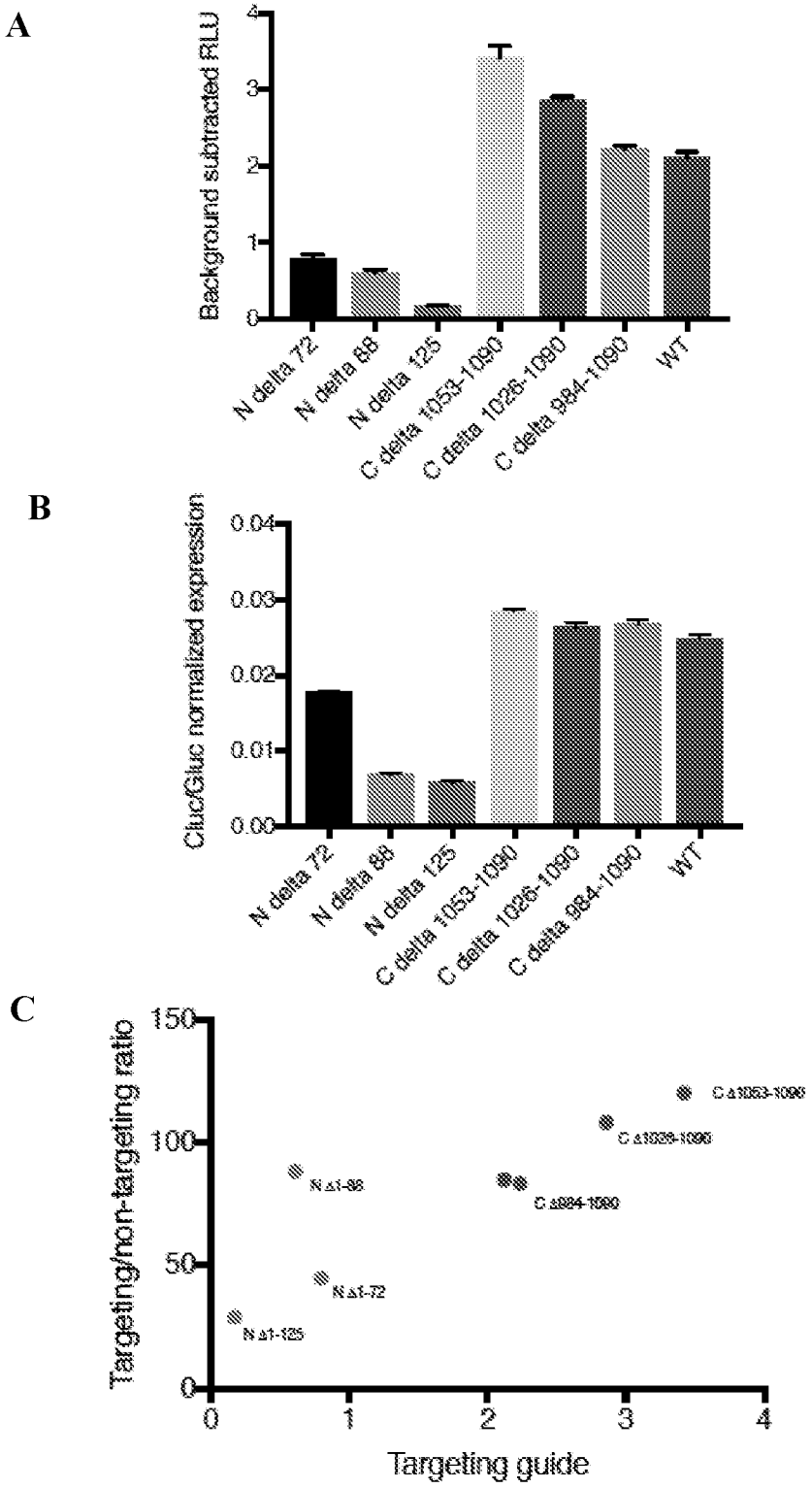

A
Background subtracted RLU
0
1
2
3
4
N delta 72
N delta 88
N delta 125
C delta 1053-1090
C delta 1026-1090
C delta 984-1090
WT
B
Cluc/Gluc normalized expression
0.00
0.01
0.02
0.03
0.04
N delta 72
N delta 88
N delta 125
C delta 1053-1090
C delta 1026-1090
C delta 984-1090
WT
C
Targeting/non-targeting ratio
0
50
100
150
0
1
2
3
4
Targeting guide
N Δ1-88
N Δ1-72
N Δ1-125
C Δ984-1090
C Δ1026-1090
C Δ1053-1090

FIG. 48

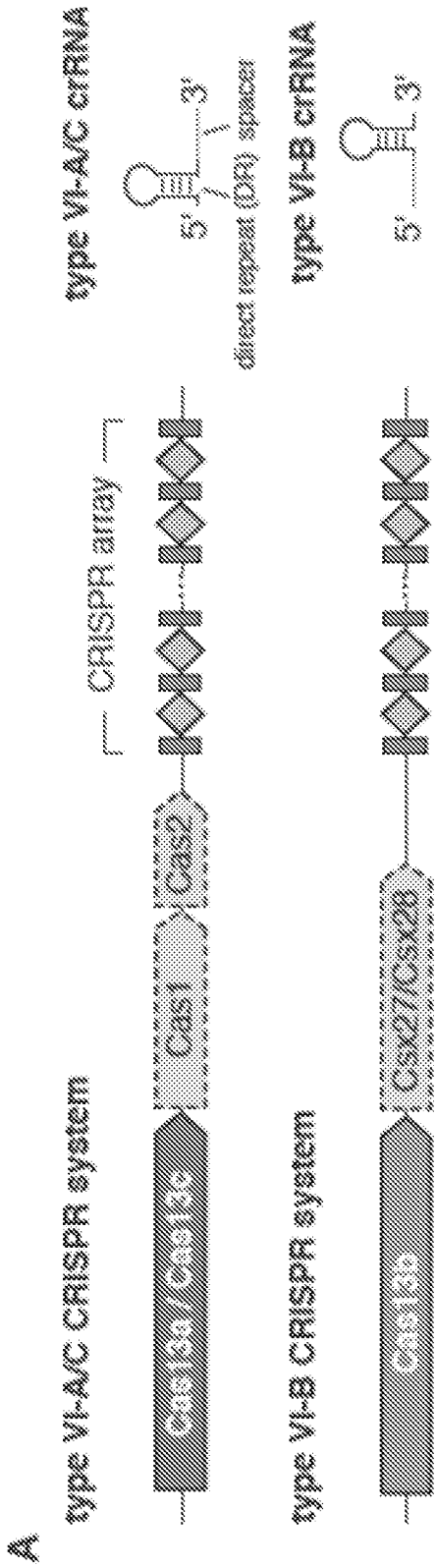

A
type VI-A/C CRISPR system
CRISPR array
Cas1
Cas2
Cas13a/C2c2/C2c6c
type VI-A/C crRNA
5'
3'
direct repeat (DR)
spacer
type VI-B CRISPR system
Cas13b
Csx27/Csx28
type VI-B crRNA
5'
3'

FIG. 49A

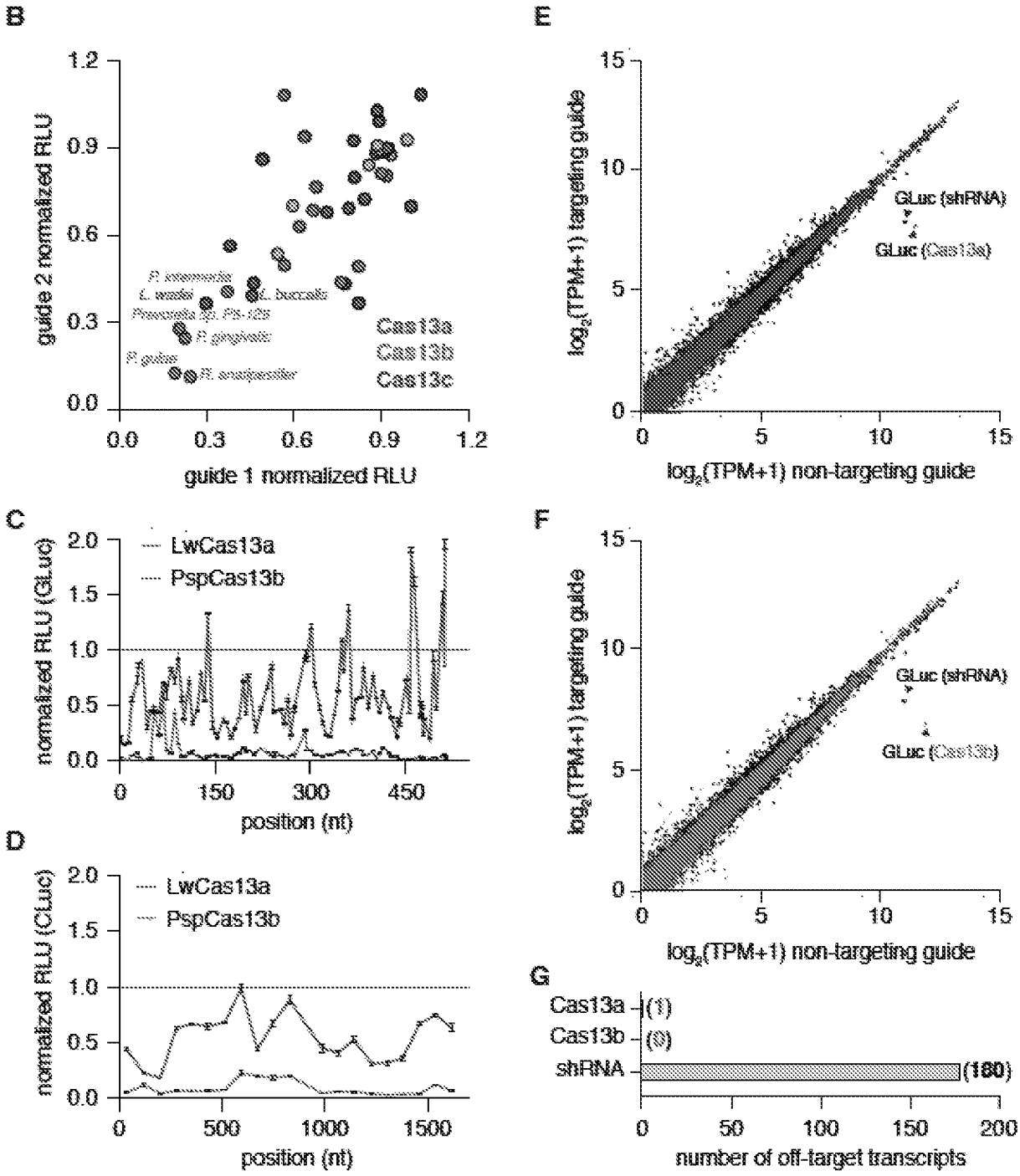


FIG. 49 (continued)

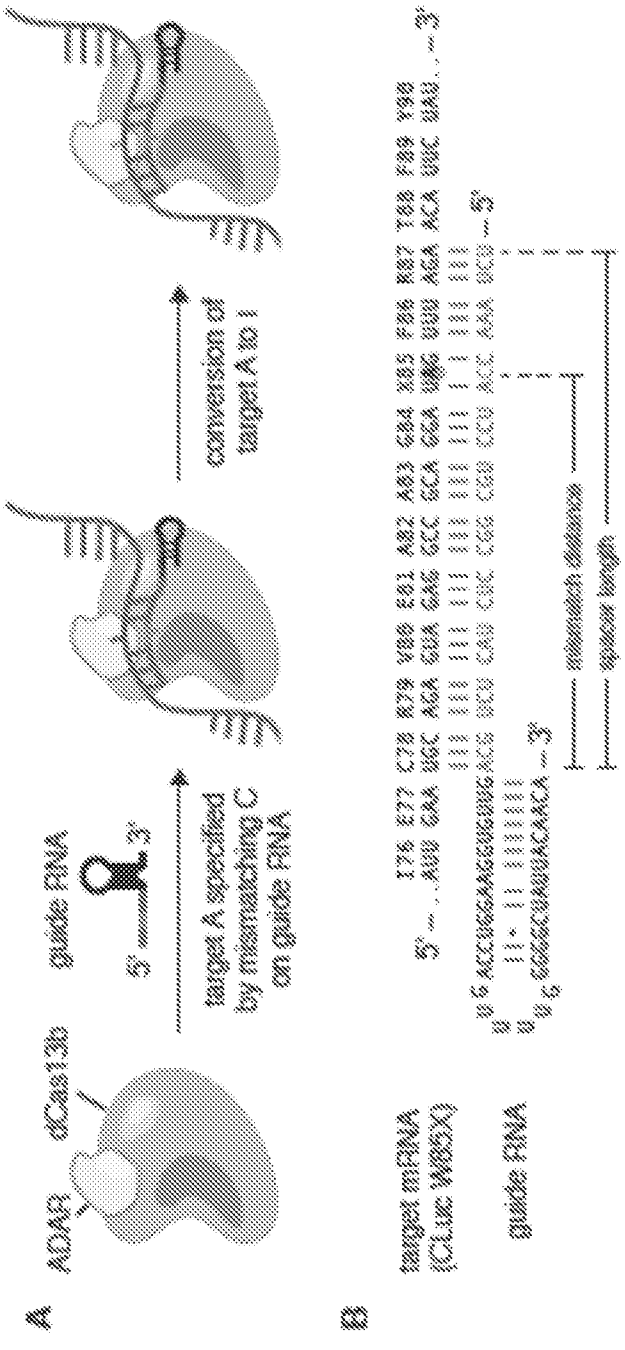

A
ADAR
dCas13b
guide RNA
5'
3'
targeted A specified by mismatching C on guide RNA
conversion of targeted A to I
B
target mRNA (CLuc W85X)
guide RNA
5'
3'

FIG. 50

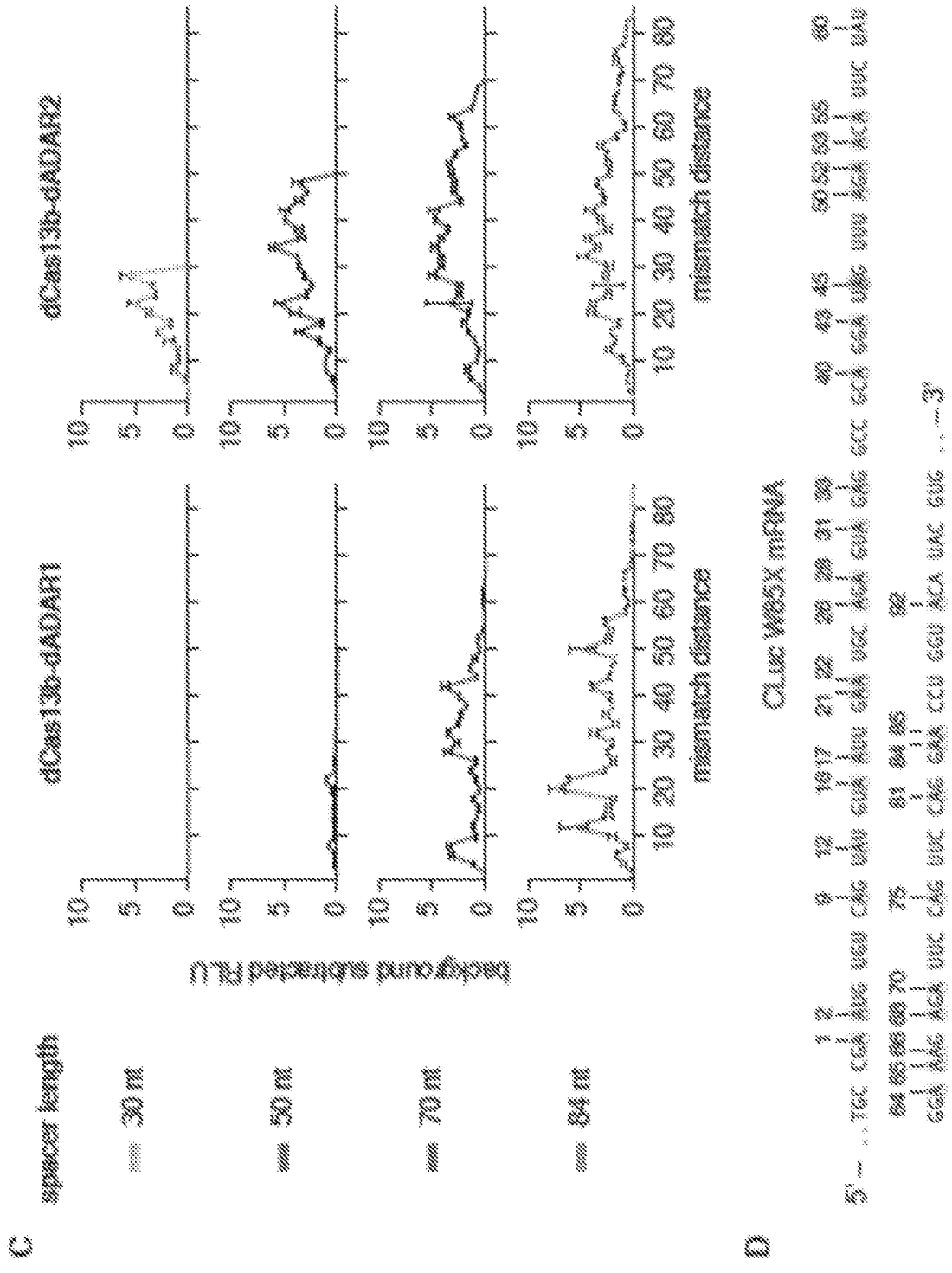

C
spacer length
30 nt
50 nt
70 nt
84 nt
dCas13b-dADAR1
dCas13b-dADAR2
background subtracted RLU
mismatch distance
D
CLuc W85X mRNA
5' ... tcc cca ttc ucu cAc uuc cAc uuc uuc cAc cAc cAA AAu cAA ucc cAA uuc cAA cAu cAu ccc cca cAA cAc uAc uuc uAc uuc ... -3'

FIG. 50

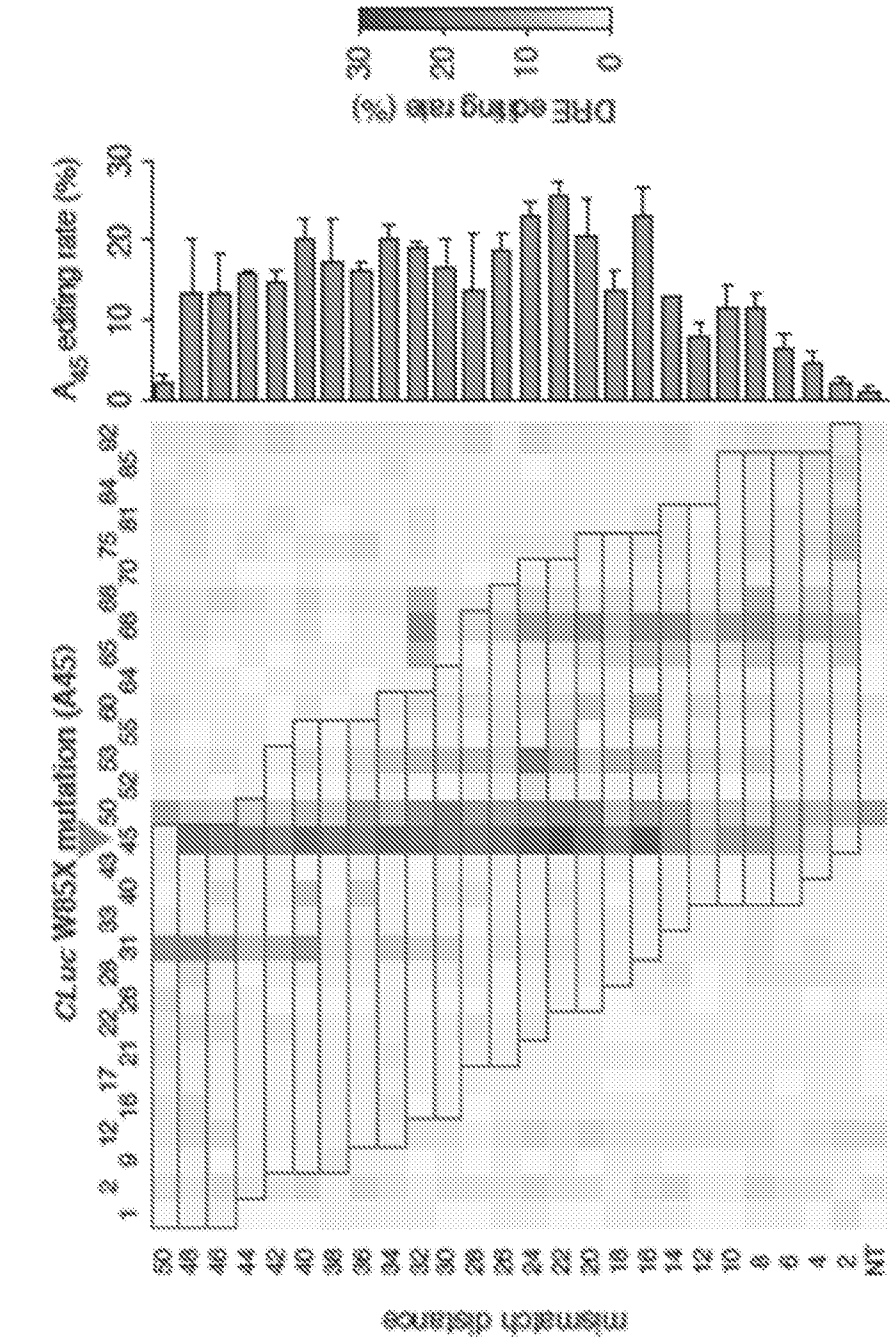

DNE editing rate (%)
30
20
10
0
A45 editing rate (%)
30
20
10
0
CLuc W85X mutation (A45)
1
2
6
10
17
18
21
22
26
28
31
33
40
45
49
50
52
58
60
64
65
66
70
75
81
84
85
92
mismatch distance
3'
5'

FIG. 50

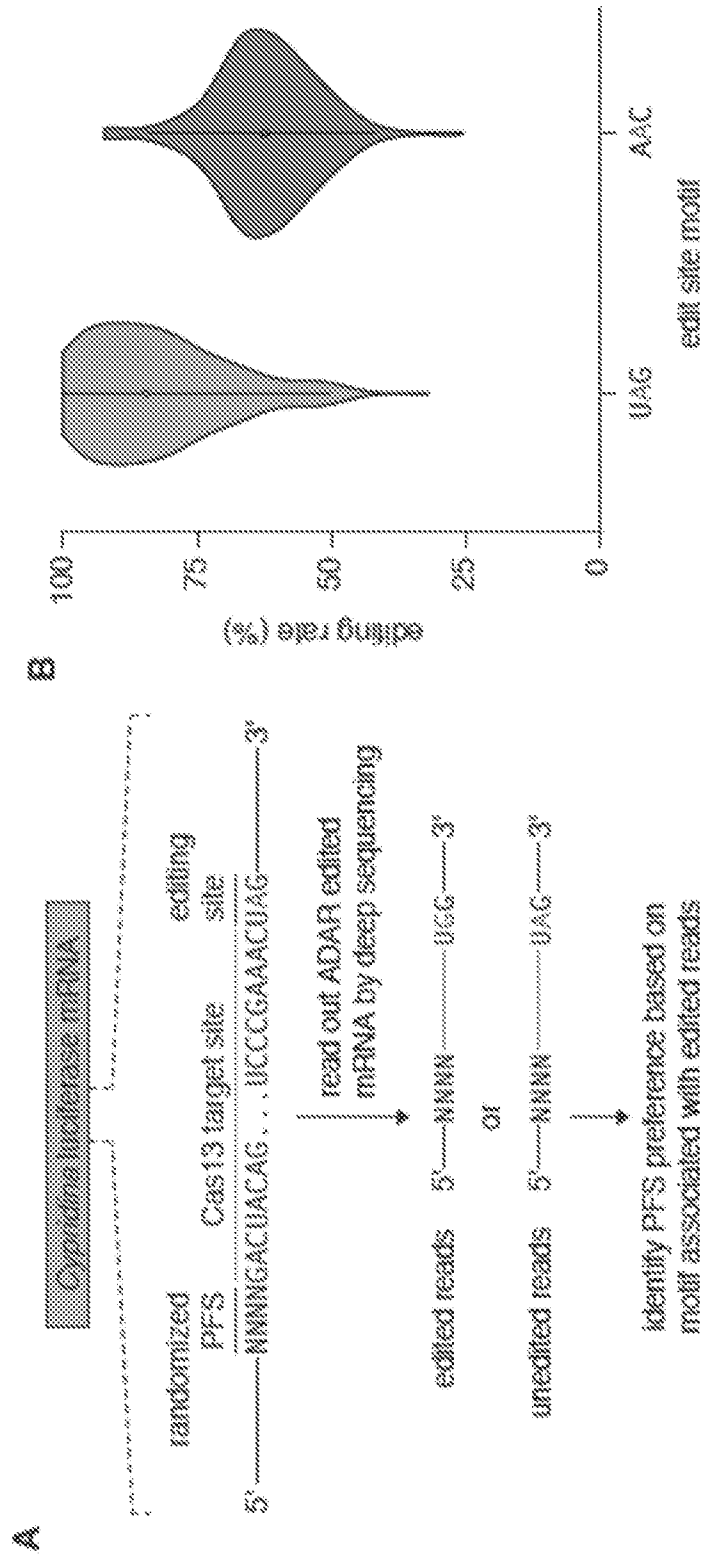

A
Construct library of mRNA
randomized PFS
editing site
Cas13 target site
5'
NNNNCACUACAG . . . UCCCGAAACUAG
3'
read out ADAR edited mRNA by deep sequencing
edited reads
5'
NNNN
UGG
3'
or
unedited reads
5'
NNNN
UAG
3'
identify PFS preference based on motif associated with edited reads
B
editing rate (%)
100
75
50
25
0
UAG
AAC
edit site motif

FIG. 51

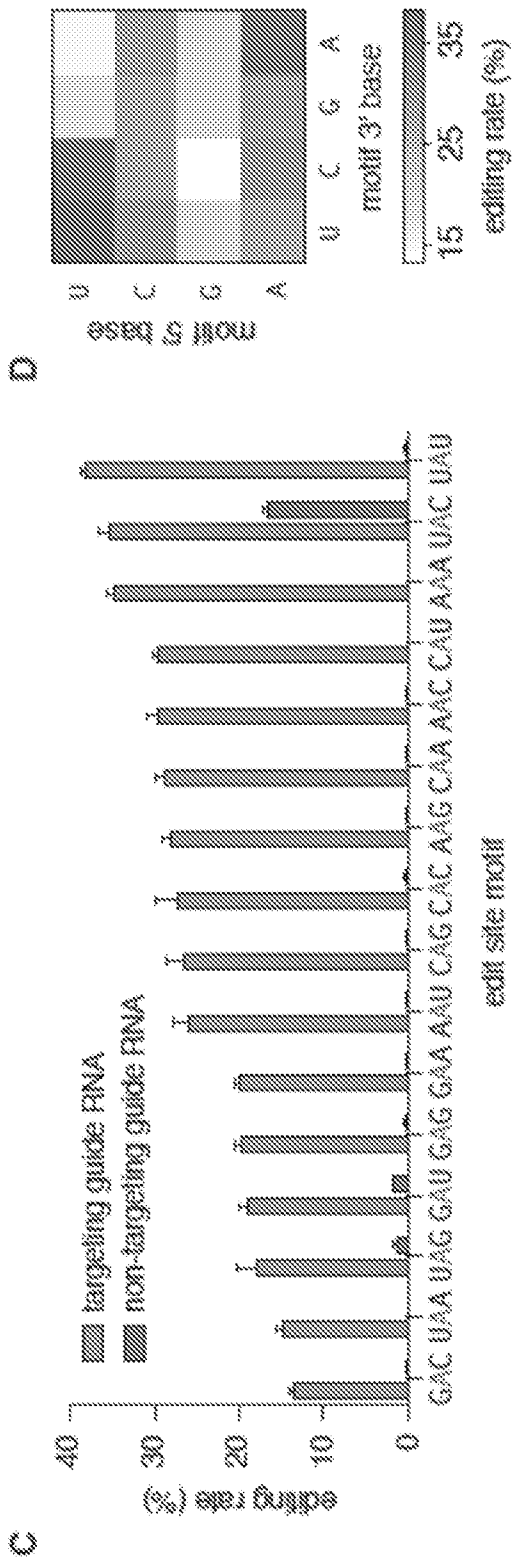


FIG. 51 (continued)

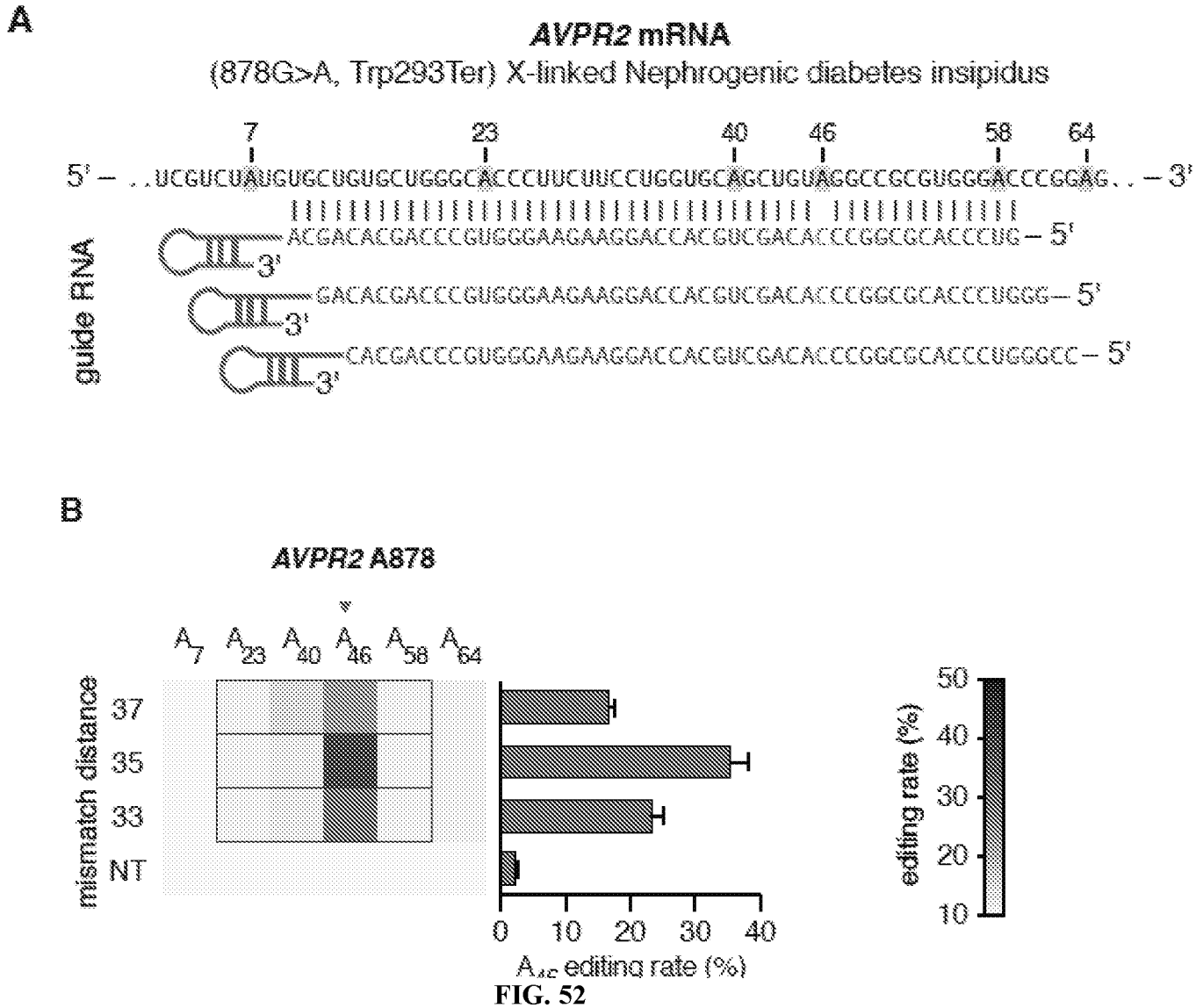

A
AVPR2 mRNA
(878G>A, Trp293Ter) X-linked Nephrogenic diabetes insipidus
7
23
40
46
58
64
5' – ..UCGUCUAUGUGCUGUGCUGGGCACCCUUCUUCCUGGUGCAGCUGUAGGCCGCGUGGGACCCGGAG.. – 3'
guide RNA
ACGACACGACCCGUGGGAAGAAGGACCACGUCGACACCCGGCGCACCCUG – 5'
3'
GACACGACCCGUGGGAAGAAGGACCACGUCGACACCCGGCGCACCCUGGG – 5'
3'
CACGACCCGUGGGAAGAAGGACCACGUCGACACCCGGCGCACCCUGGGCC – 5'
3'
B
AVPR2 A878
A7 A23 A40 A46 A58 A64
mismatch distance
37
35
33
NT
0 10 20 30 40
editing rate (%)
50
40
30
20
10
editing rate (%)

FIG. 52

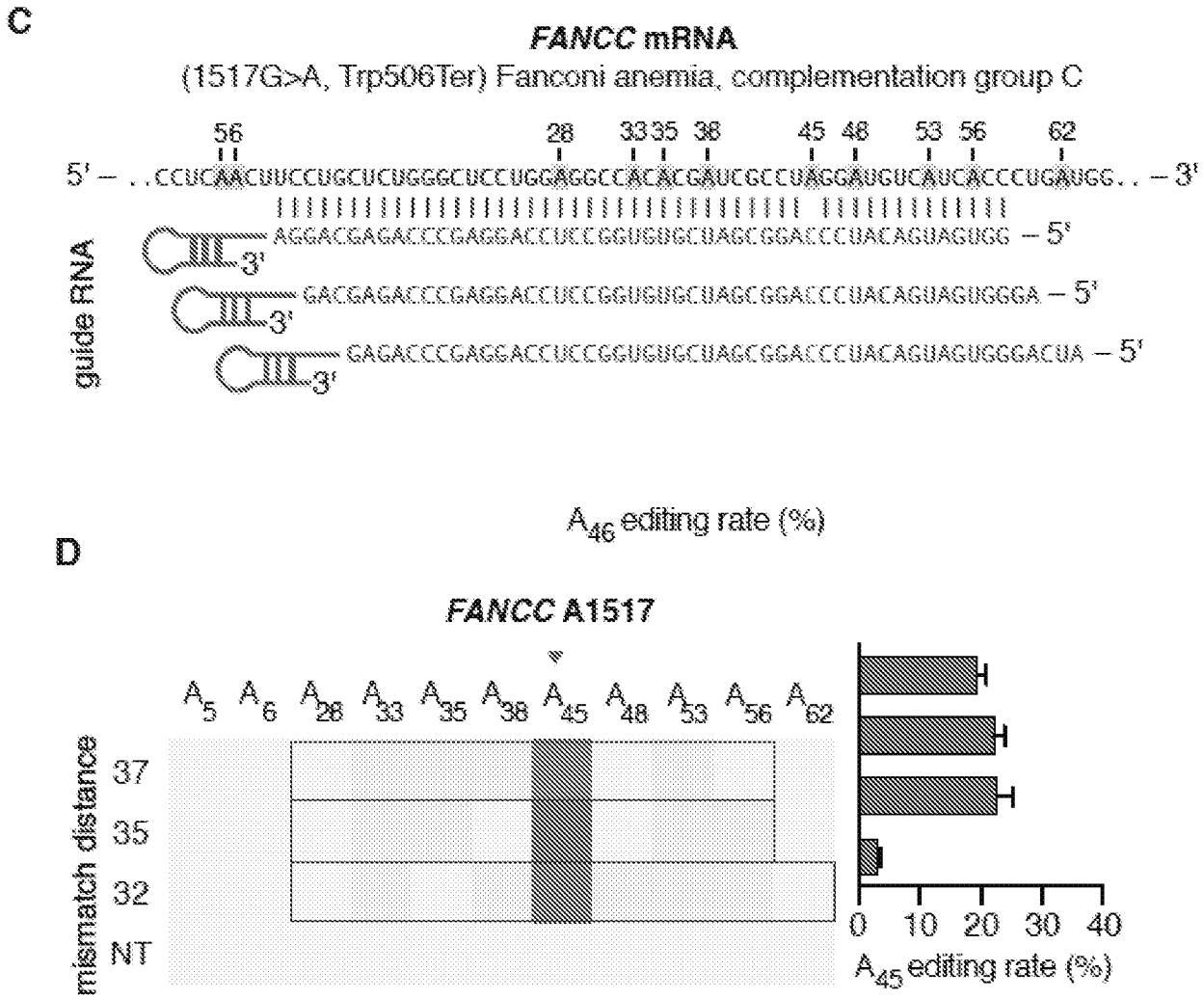
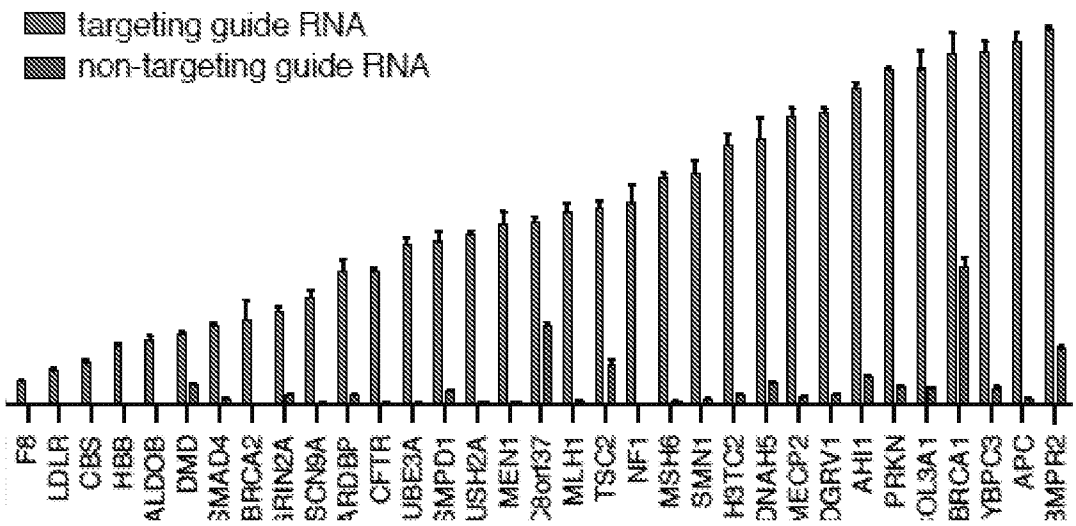


FIG. 52 (continued)

E



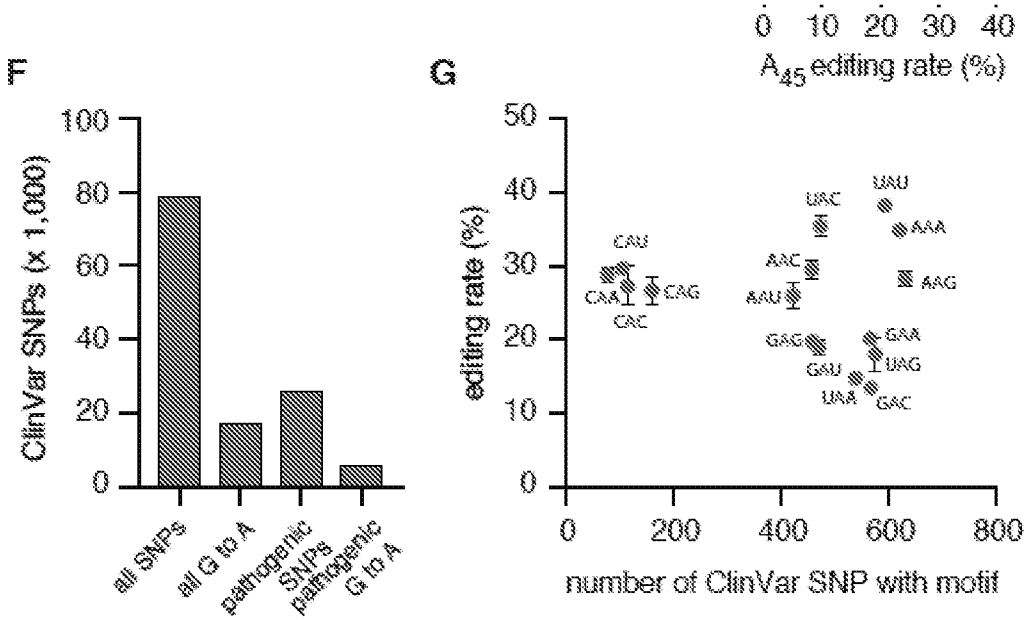


FIG. 52 (continued)

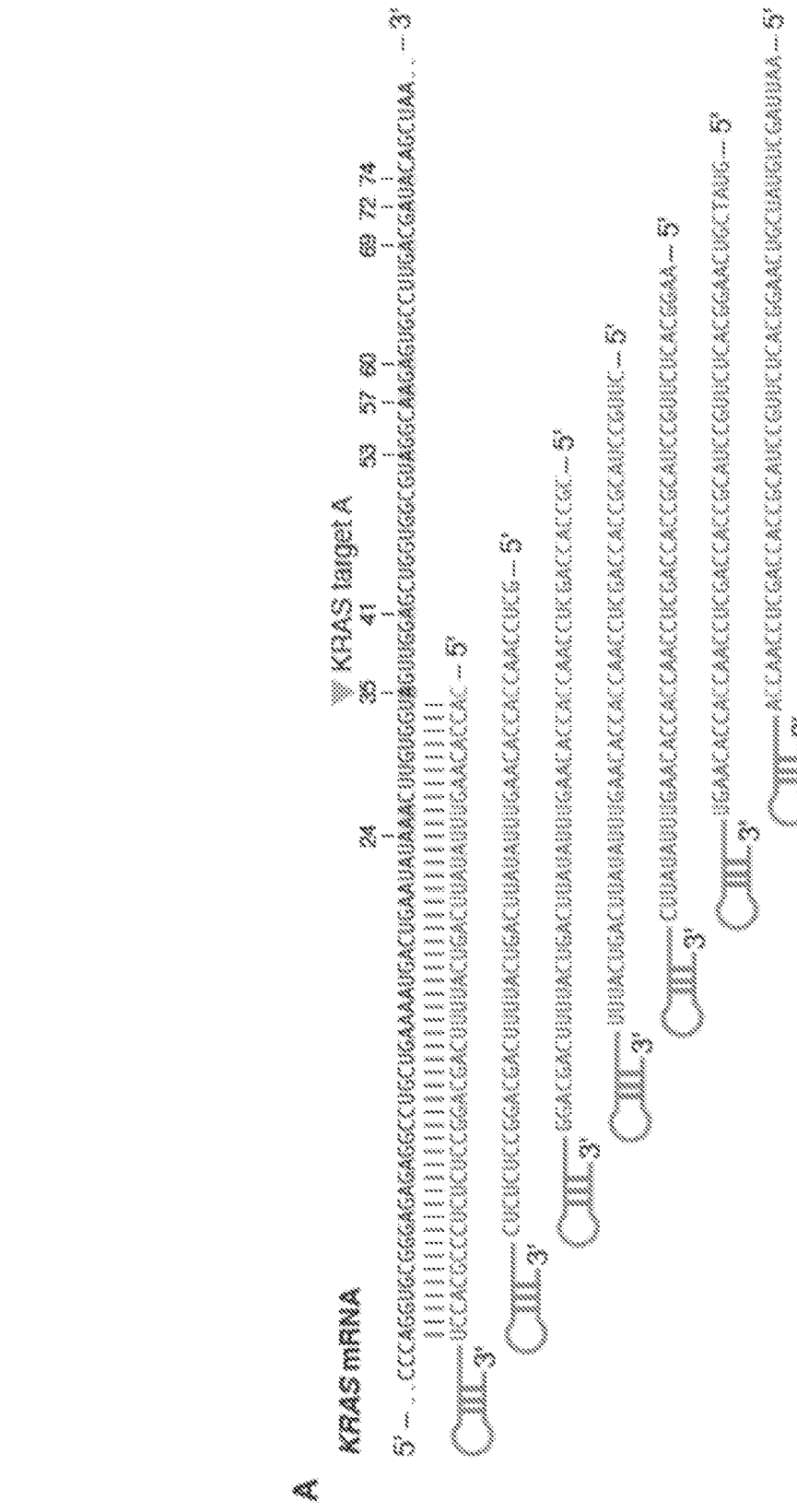

A
KRAS mRNA
KRAS target A
24
35
41
53
57
60
69
72
74
5'
3'
CIL
CIL
CIL
CIL
CIL
CIL

FIG. 53

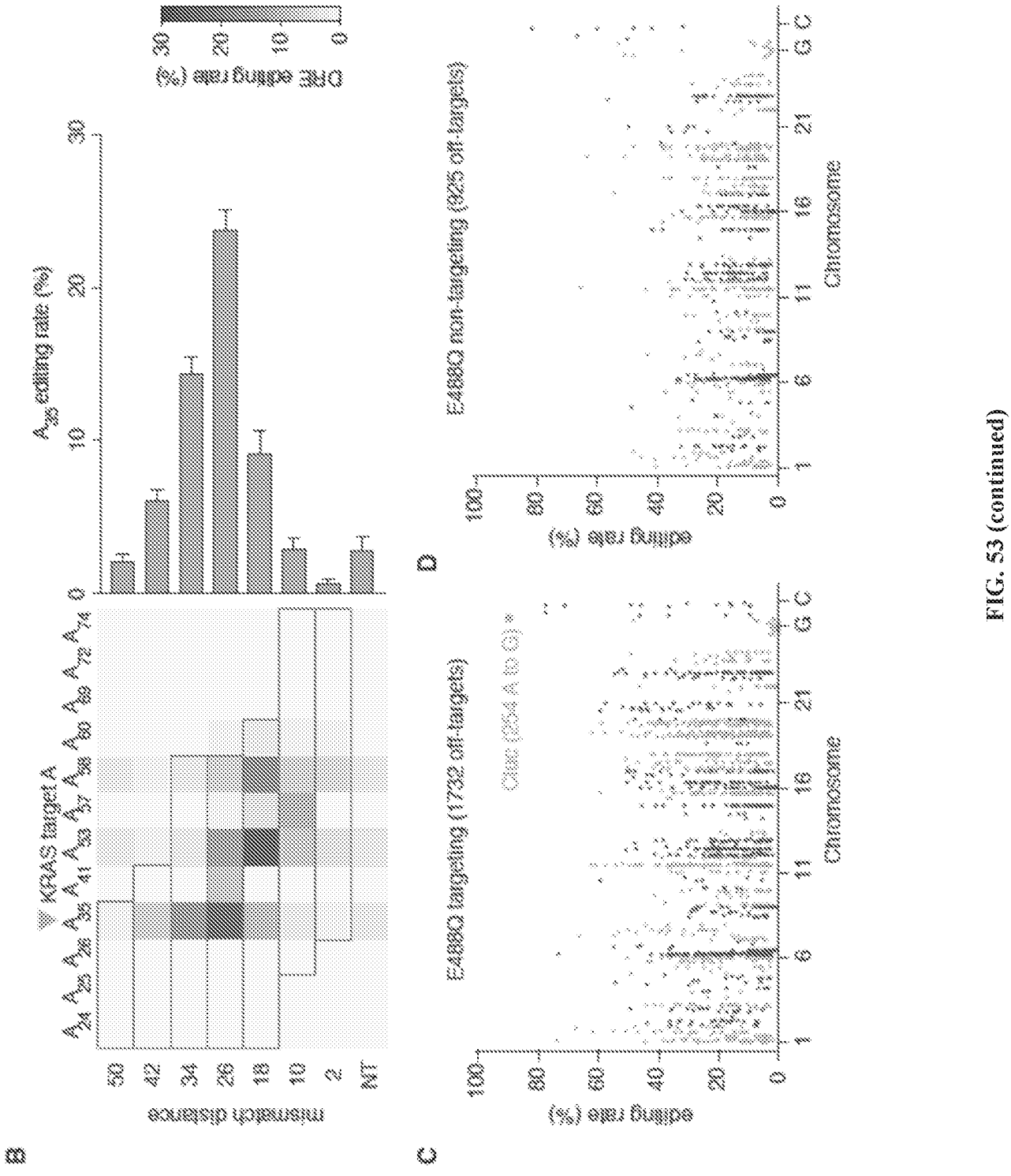


FIG. 53 (continued)

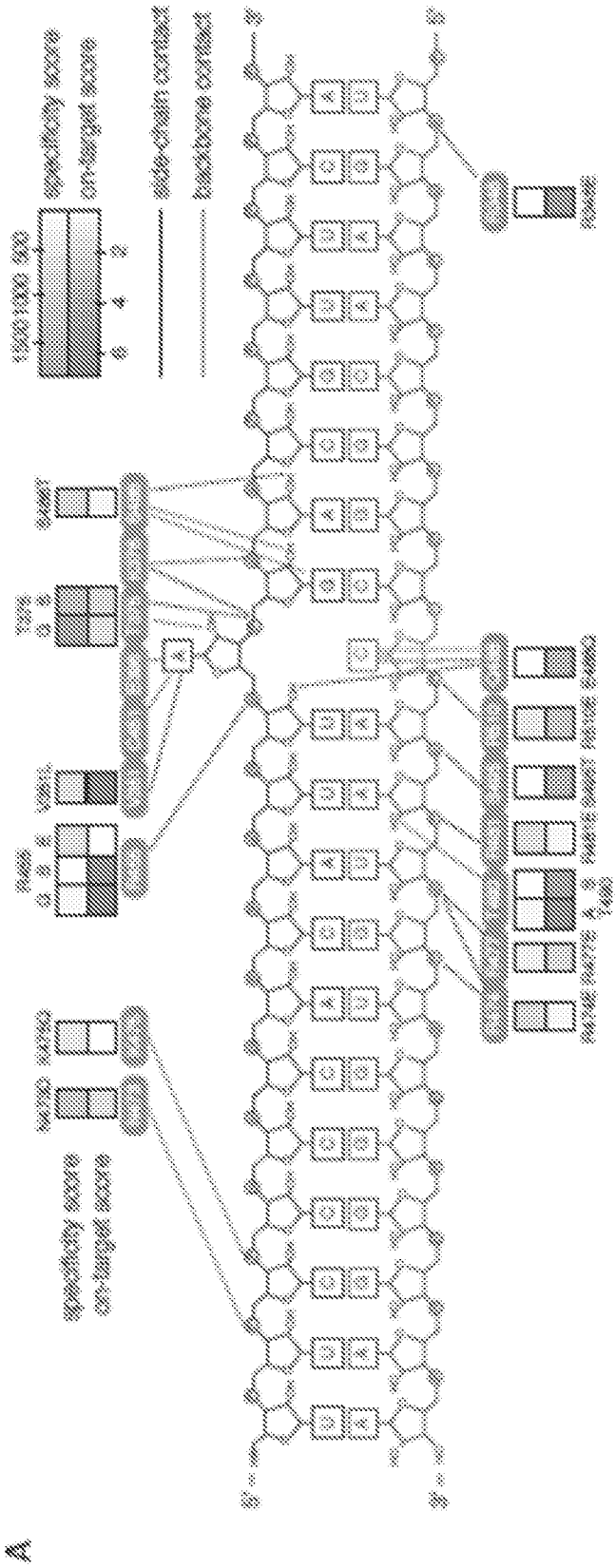

activity score
context score
side-chain contact
backbone contact
activity sum
context sum

FIG. 54

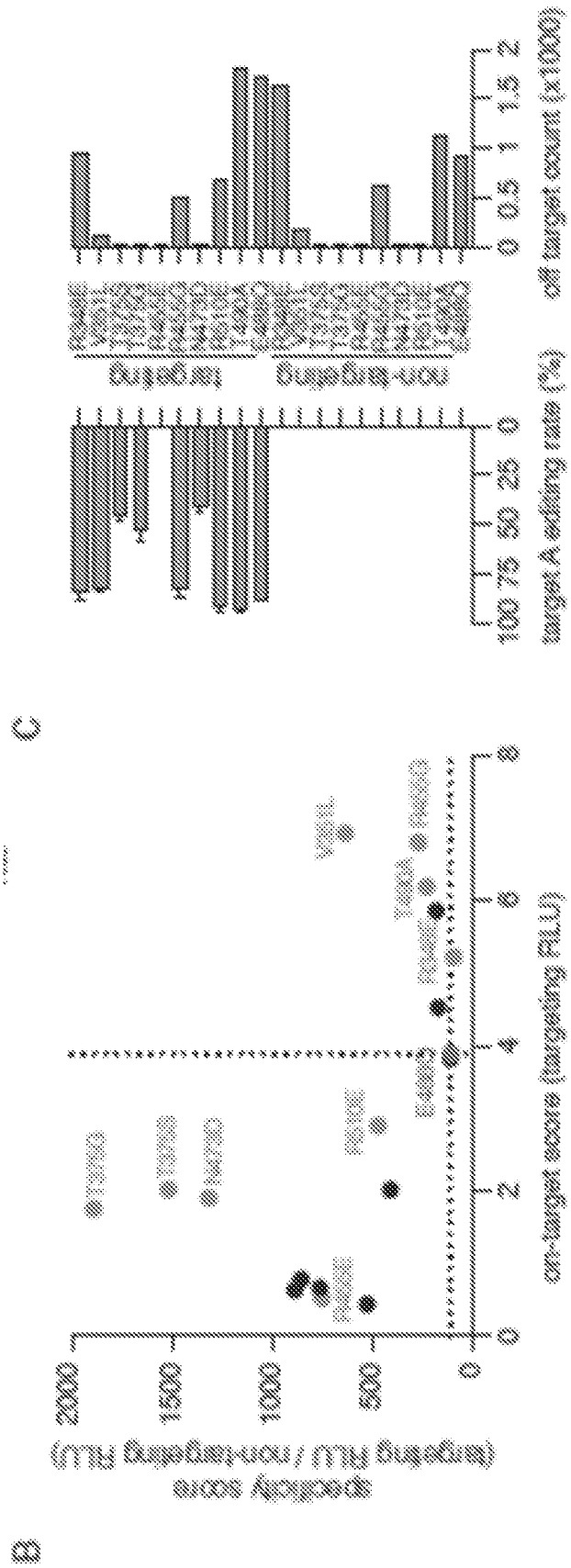


FIG. 54 (continued)

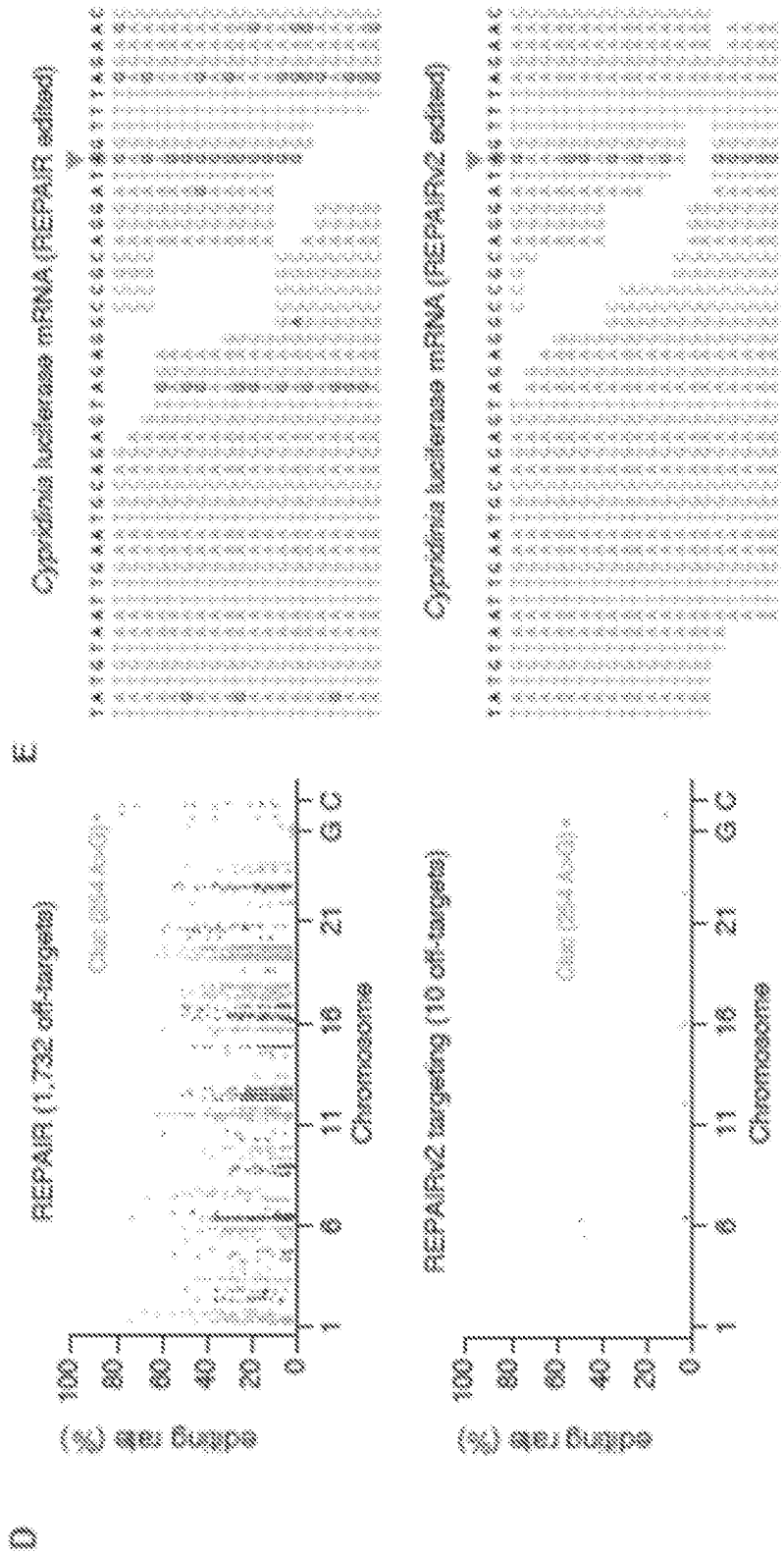
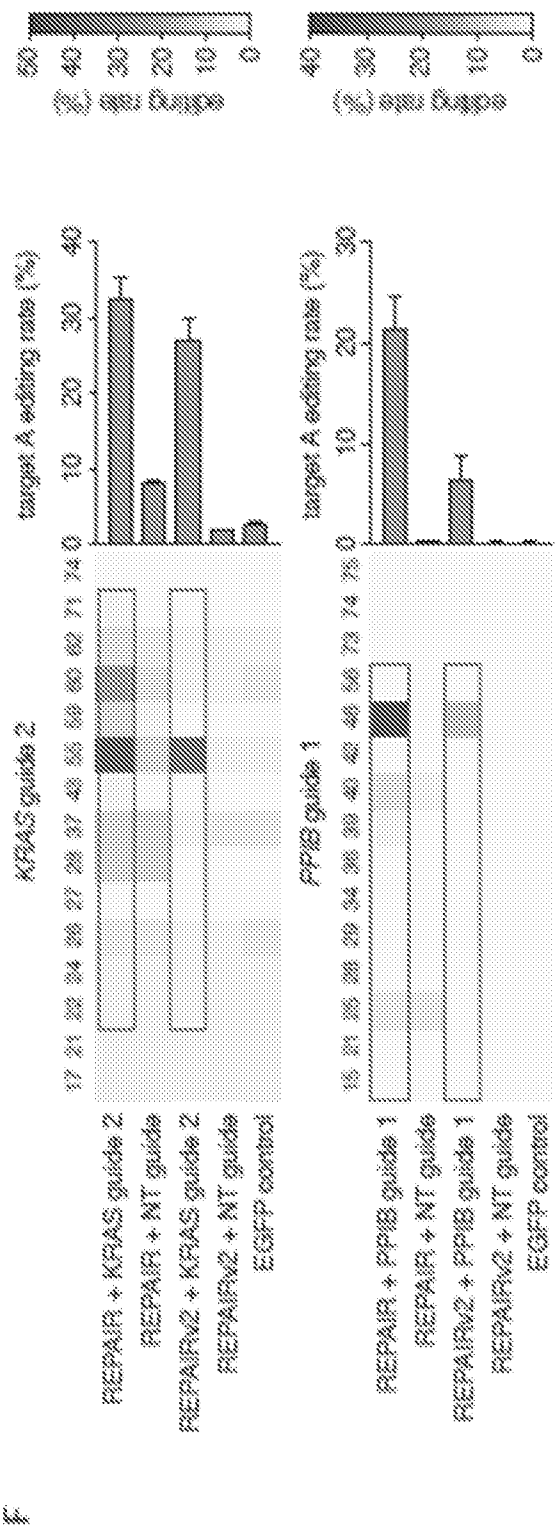


FIG. 54 (continued)



F

FIG. 54 (continued)

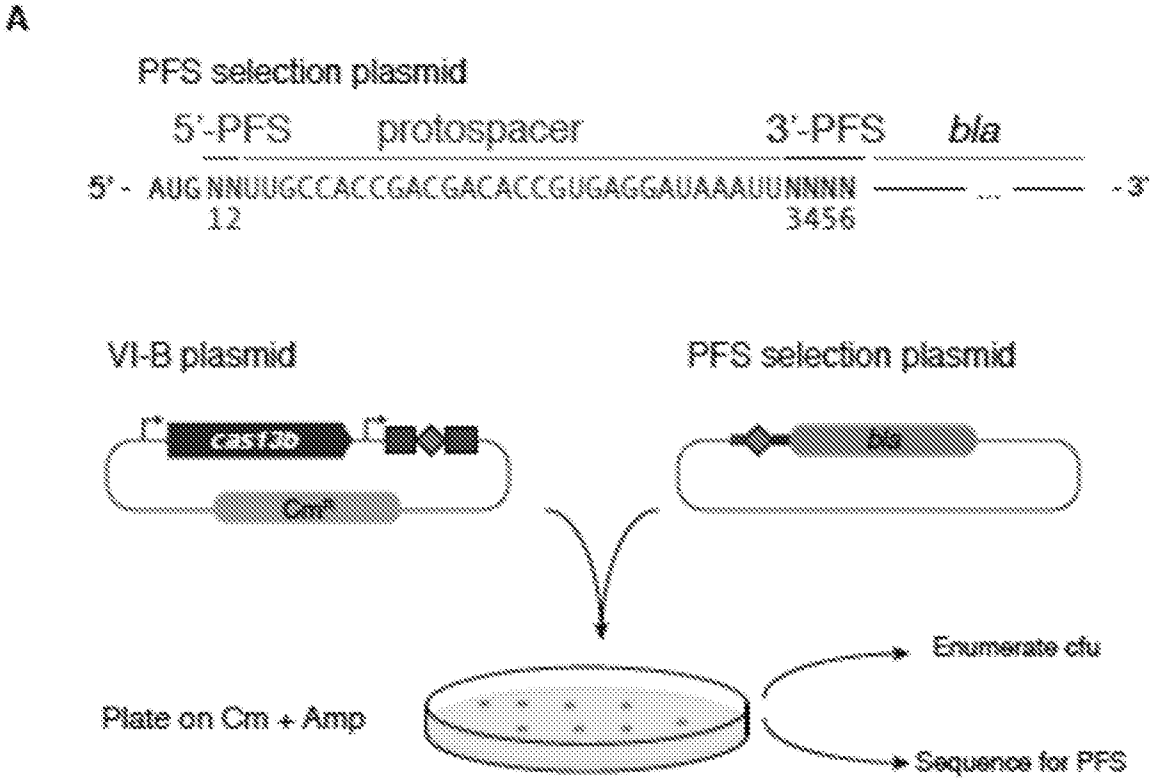

A
PFS selection plasmid
5'-PFS
protospacer
3'-PFS
bla
5' - AUG NNUUGCCACCGACGACACCGUGAGGAUAAAUUNNNN - 3'
1 2
3 4 5 6
VI-B plasmid
PFS selection plasmid
Cm
Plate on Cm + Amp
Enumerate cfu
Sequence for PFS

FIG. 55

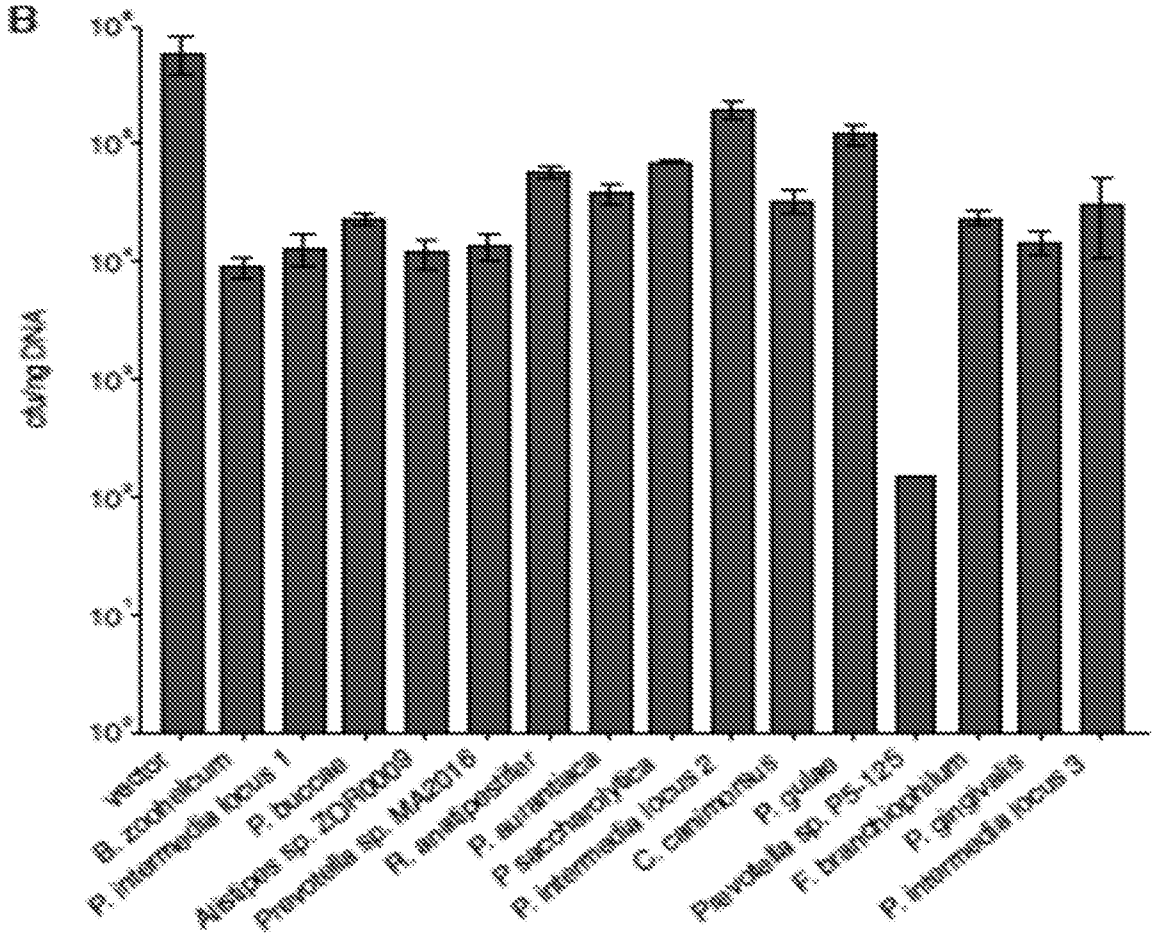


FIG. 55 (continued)

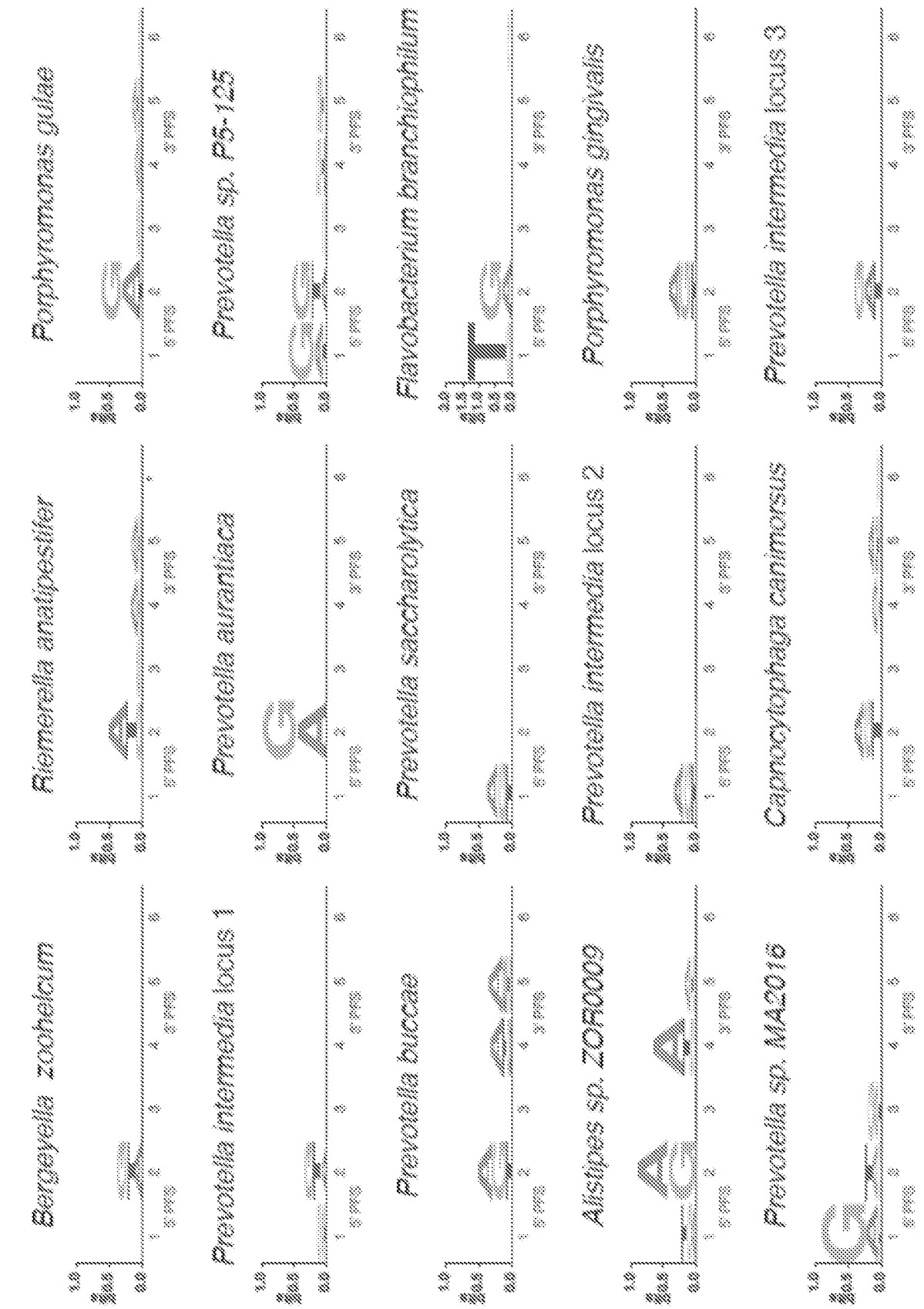
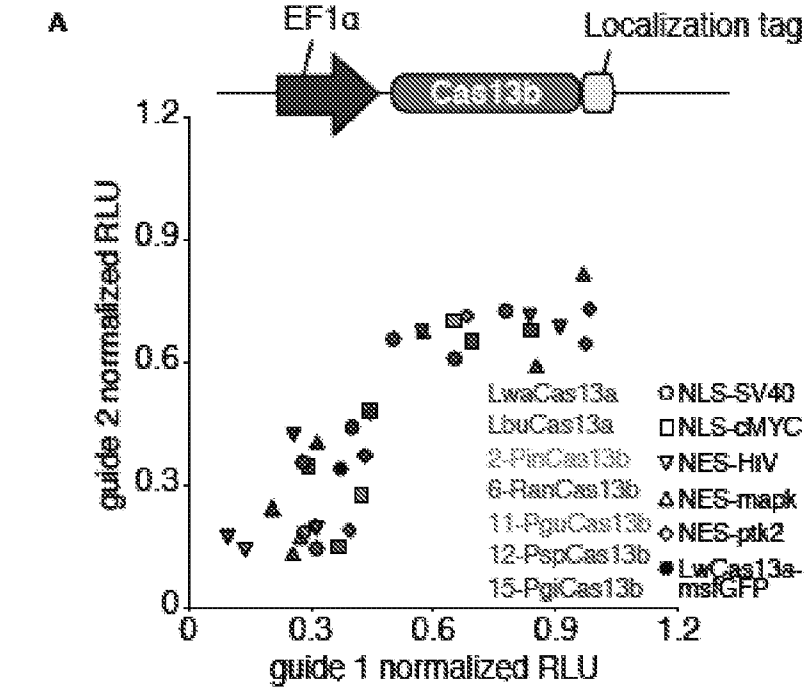

C
Bergeyella zoohelcum
Alemerella anatipestifer
Porphyromonas gulae
Prevotella intermedia locus 1
Prevotella aurantiaca
Prevotella sp. P5-125
Prevotella buccae
Prevotella saccharolytica
Flavobacterium branchiophilum
Alistipes sp. ZOR0009
Prevotella intermedia locus 2
Porphyromonas gingivalis
Prevotella sp. MA2016
Capnocytophaga canimorsus
Prevotella intermedia locus 3

FIG. 55 cont.


A
EF1α
Localization tag
Cas13b
guide 2 normalized RLU
guide 1 normalized RLU
0
0.3
0.6
0.9
1.2
LwaCas13a
LbuCas13a
2-PinCas13b
6-RanCas13b
11-PguCas13b
12-PspCas13b
15-PgiCas13b
NLS-SV40
NLS-cMYC
NES-HIV
NES-mapk
NES-ptk2
LwaCas13a-msfGFP

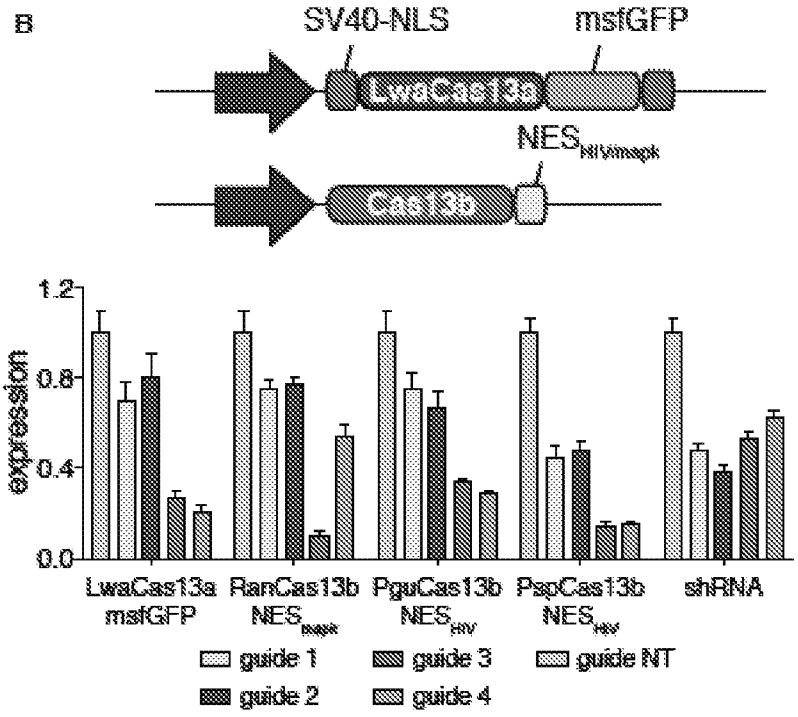

B
SV40-NLS
msfGFP
LwaCas13b
NES
Cas13b
expression
0.0
0.4
0.8
1.2
LwaCas13a msfGFP
RanCas13b NES
PguCas13b NES
PspCas13b NES
shRNA
guide 1
guide 2
guide 3
guide 4
guide NT

FIG. 56

C

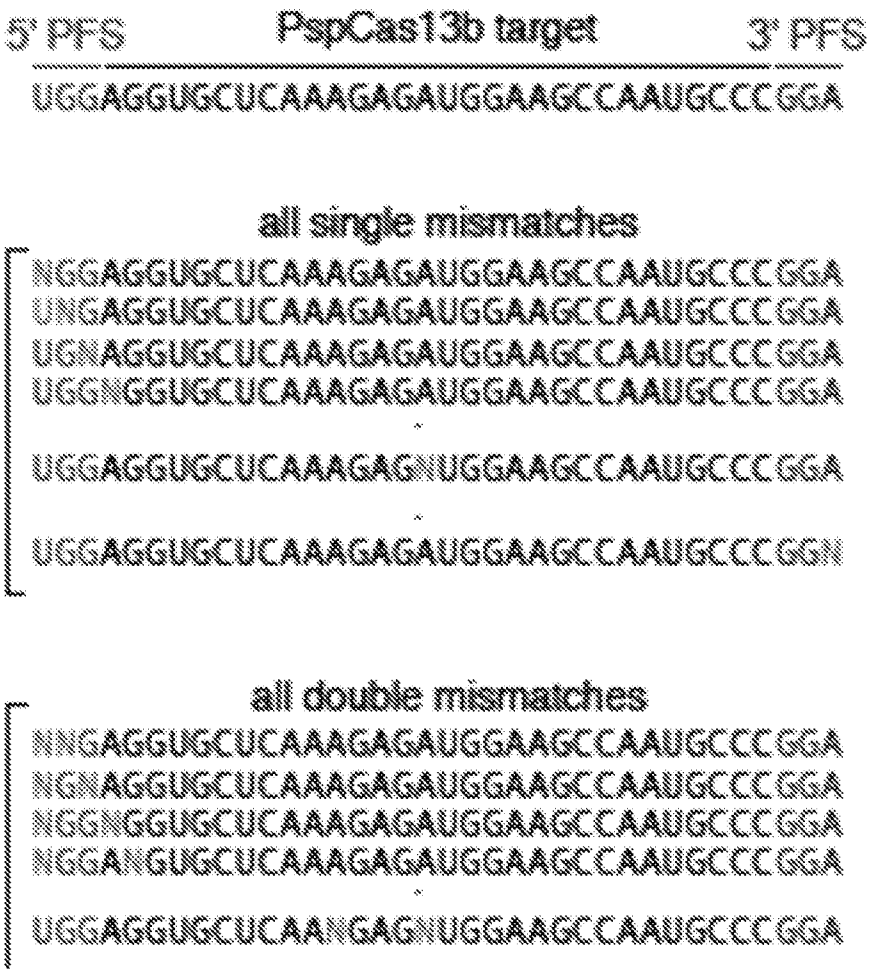


FIG. 56 (continued)

D

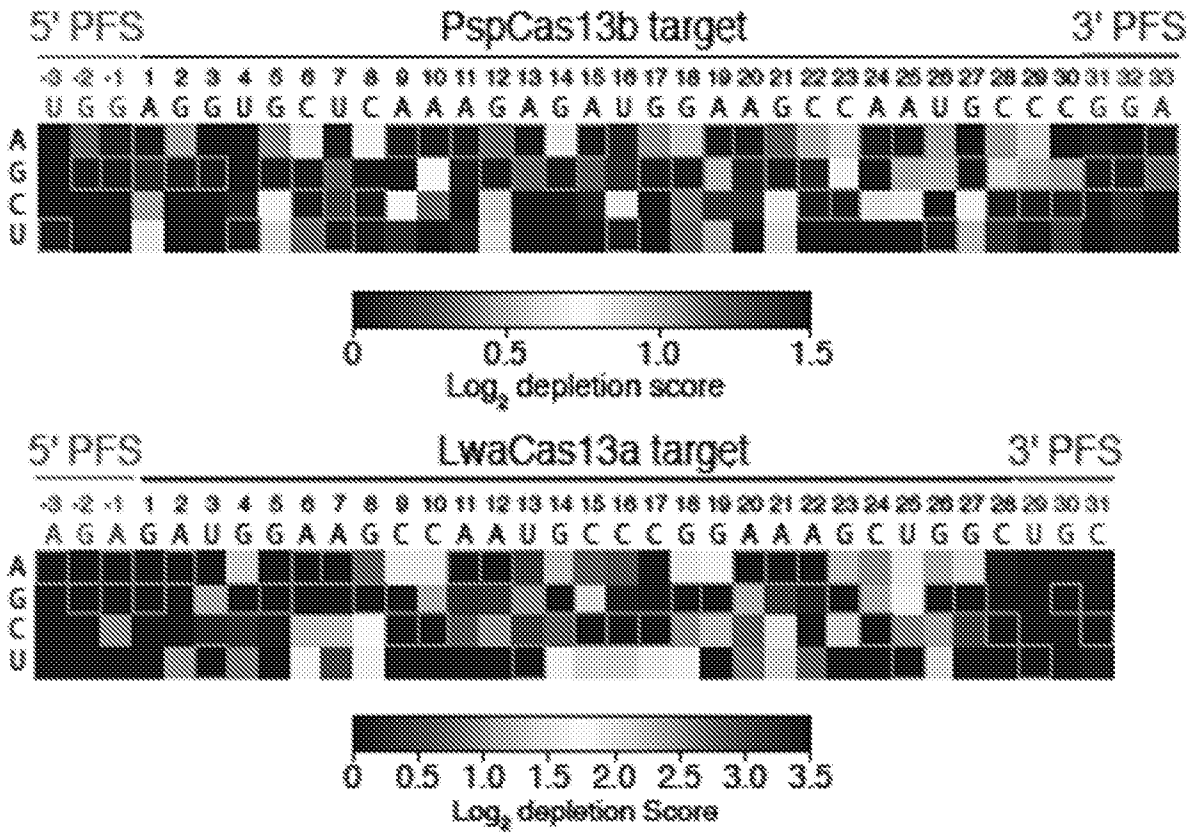


FIG. 56 (continued)

E

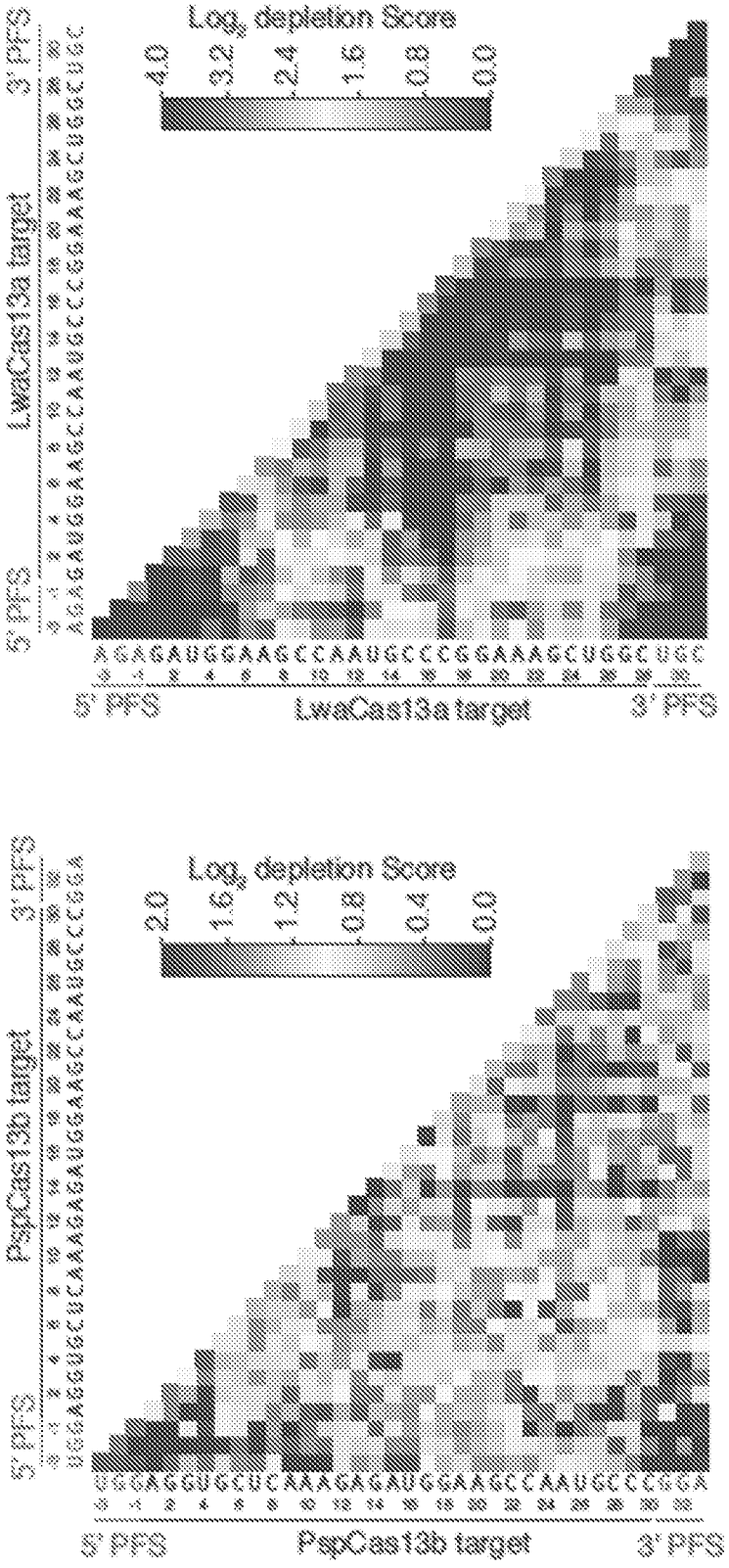

Log2 depletion Score
4.0
3.2
2.4
1.6
0.8
0.0
5' PFS
LwaCas13a target
3' PFS
Log2 depletion Score
2.0
1.6
1.2
0.8
0.4
0.0
5' PFS
PspCas13b target
3' PFS

FIG. 56 cont.

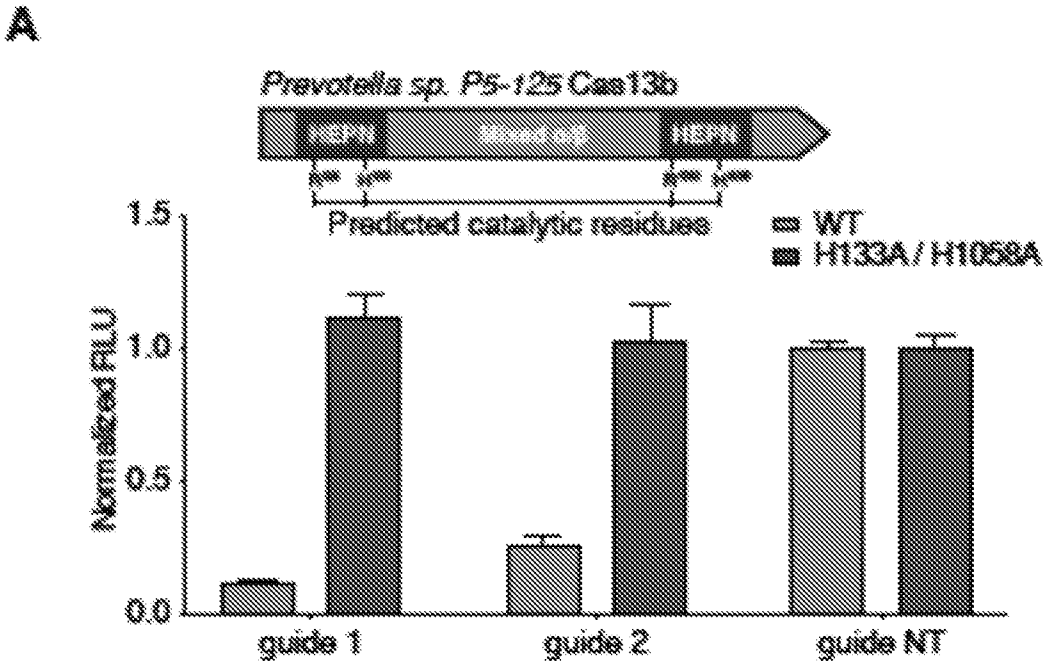

A
Prevotella sp. P5-125 Cas13b
NCPH
Predicted catalytic residues
WT
H133A / H1058A
Normalized FLU
1.5
1.0
0.5
0.0
guide 1
guide 2
guide NT

FIG. 57

B

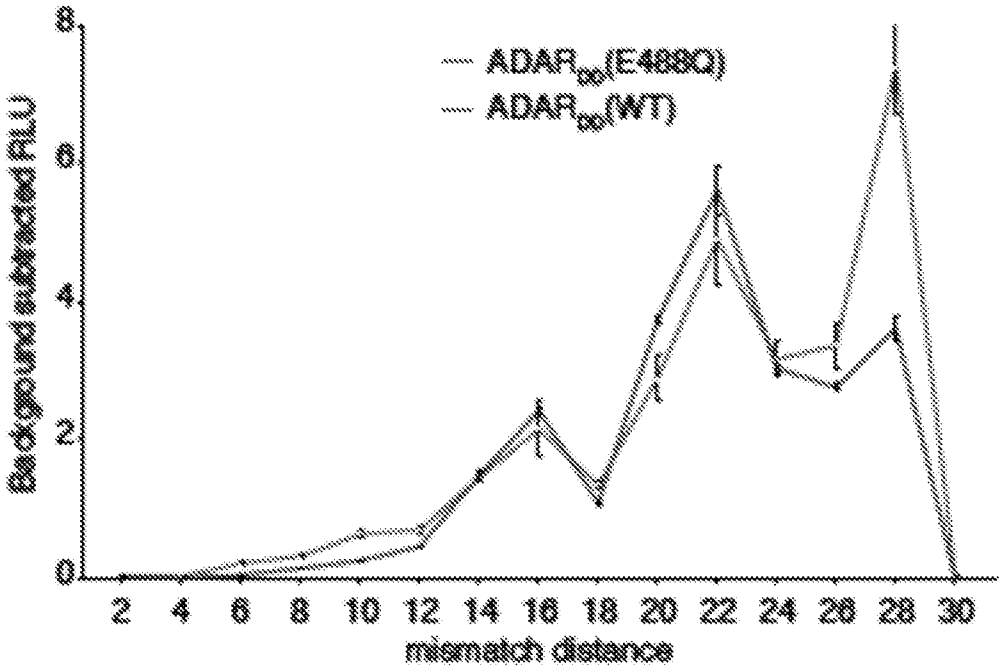


FIG. 57 (continued)

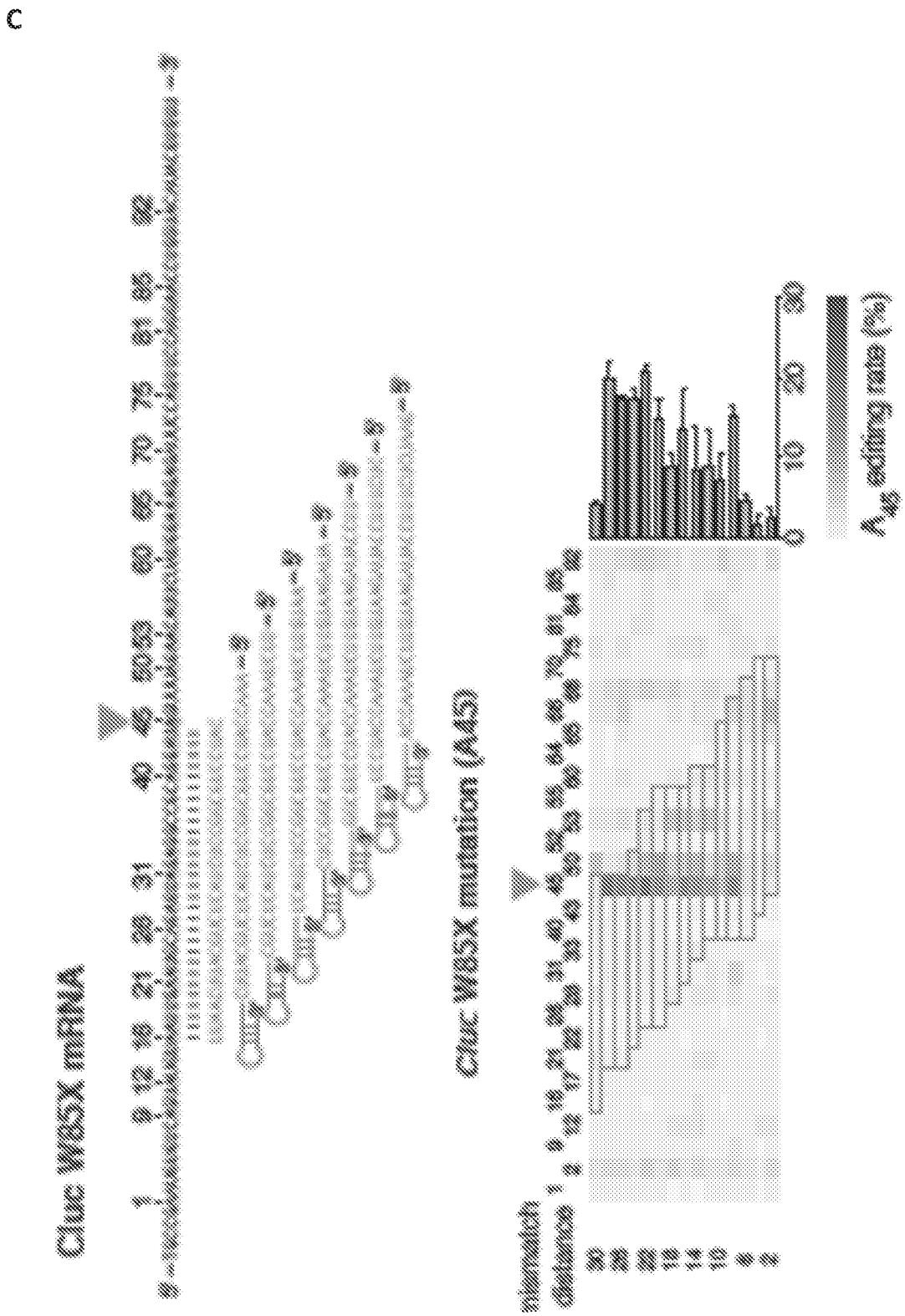


FIG. 57 (continued)

D

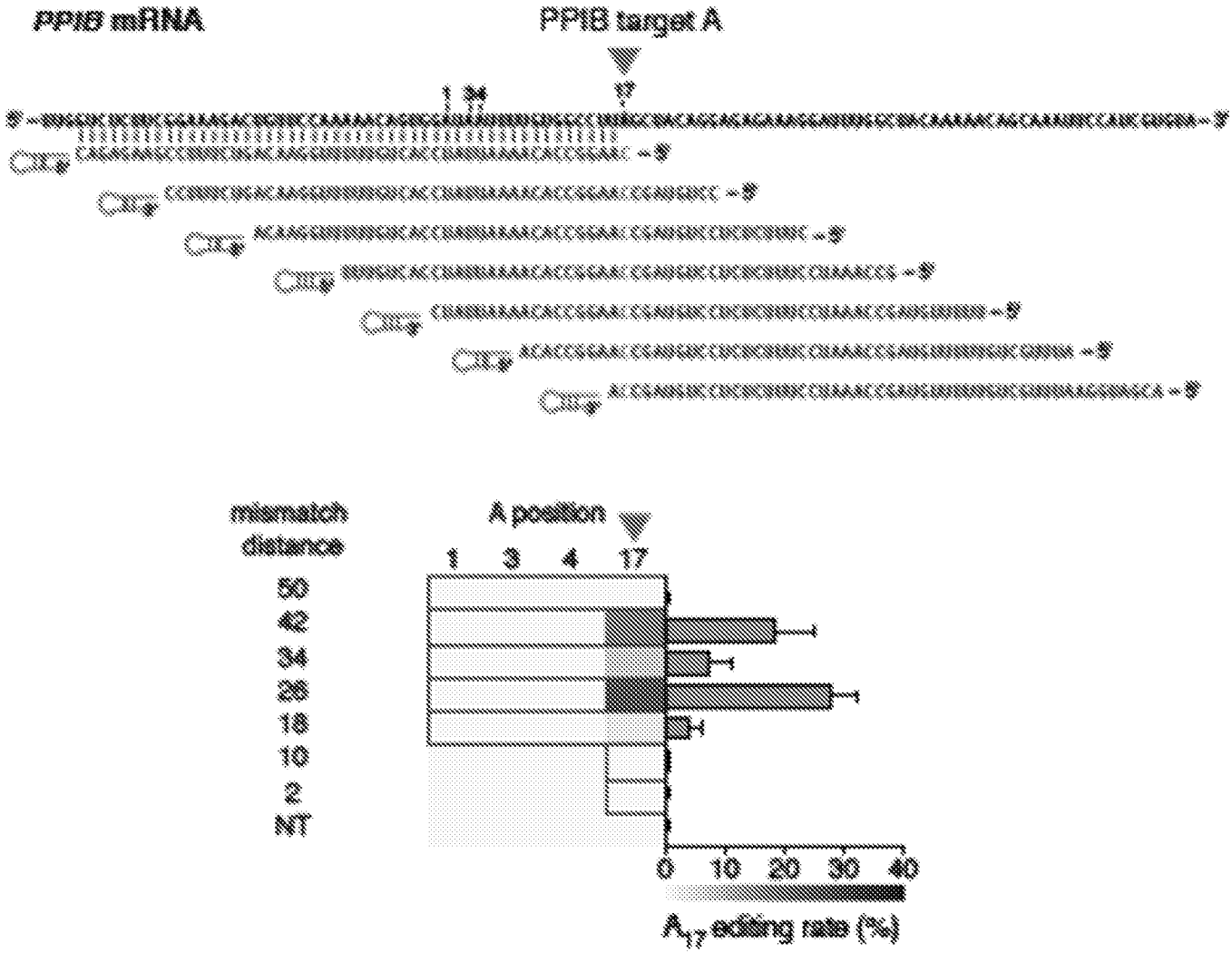


FIG. 57 (continued)

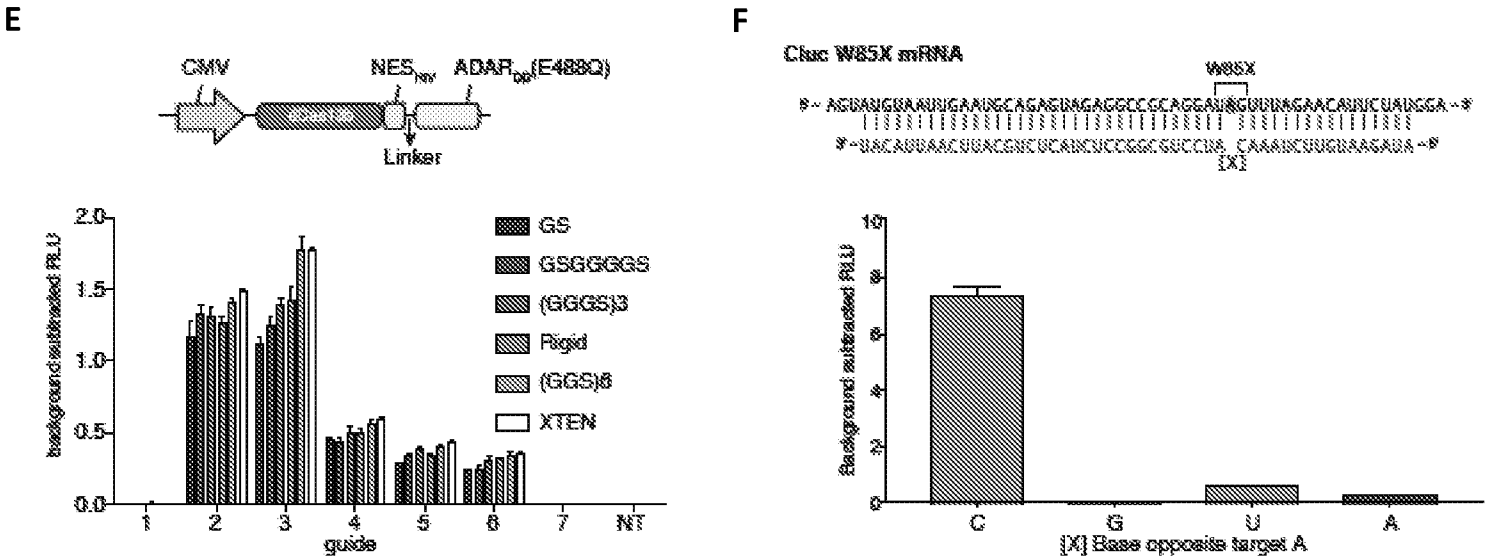


FIG. 57 (continued)

A

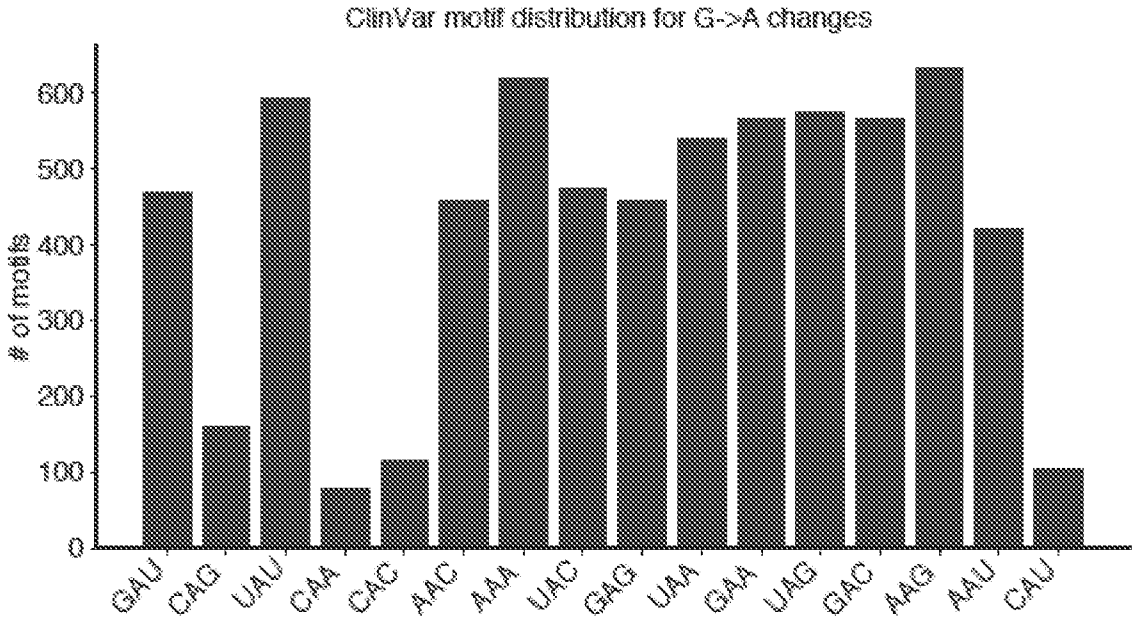

ClinVar motif distribution for G->A changes
of motifs
0
100
200
300
400
500
600
GAU
CAG
UAU
CAA
CAC
AAC
AAA
UAC
GAG
UAA
GAA
UAG
GAC
AAG
AAU
CAU

FIG. 58

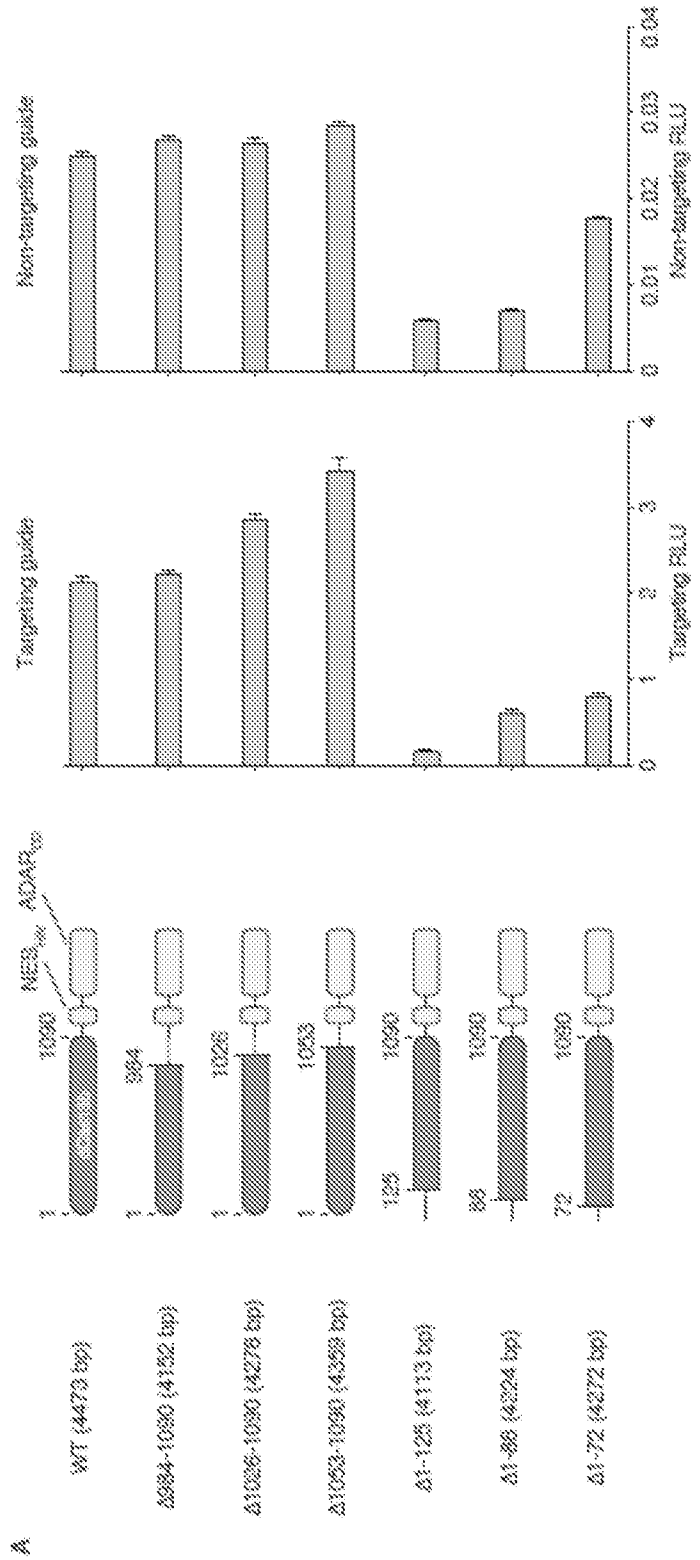

A
WT (4473 bp)
Δ984-1030 (4132 bp)
Δ1026-1030 (4278 bp)
Δ1053-1060 (4359 bp)
Δ1-125 (4113 bp)
Δ1-86 (4224 bp)
Δ1-72 (4272 bp)
MEB$_{WT}$
ADAR$_{CD}$
1030
984
1026
1053
1060
125
86
72
1
Targeting guide
Non-targeting guide
0
1
2
3
4
Targeting RLU
0
0.01
0.02
0.03
0.04
Non-targeting RLU

FIG. 59

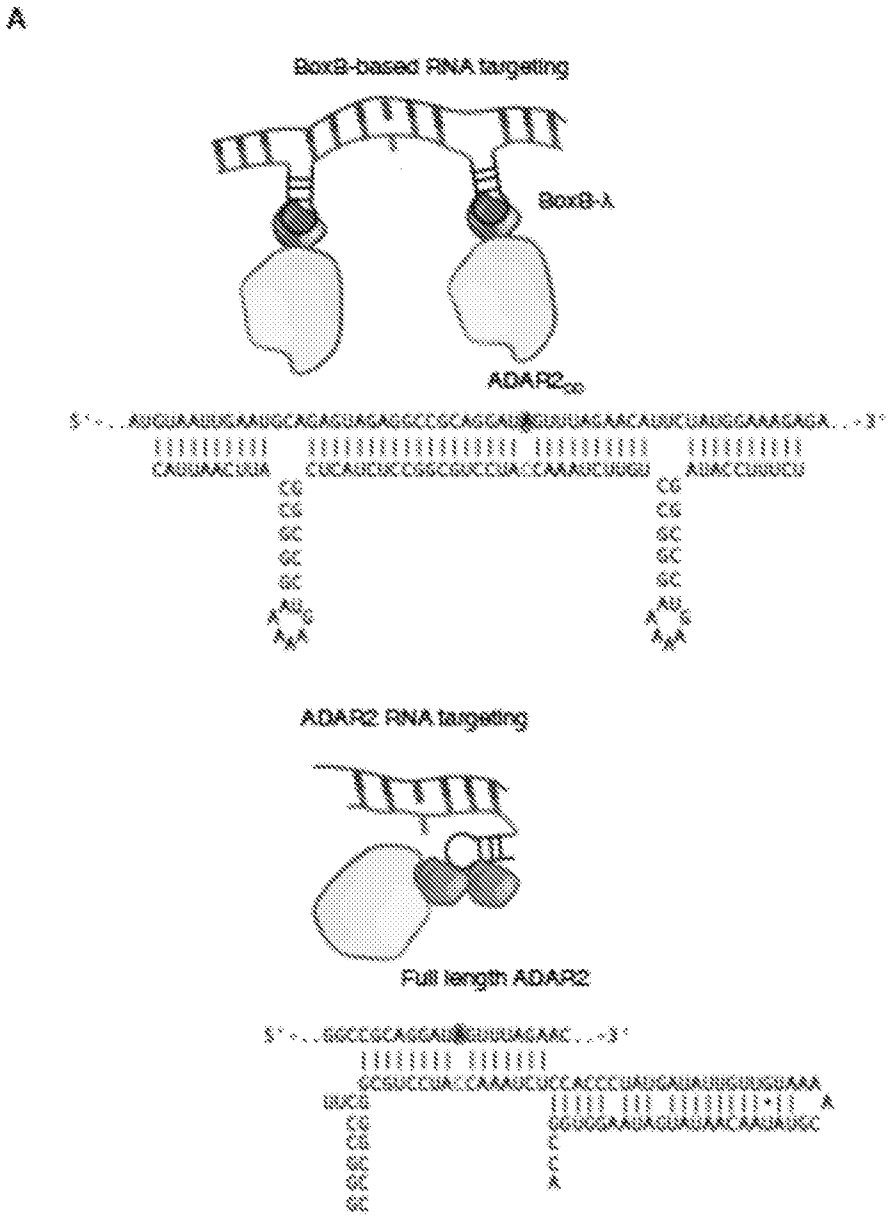

A
BoxB-based RNA targeting
BoxB-λ
ADAR2dd
ADAR2 RNA targeting
Full length ADAR2

FIG. 60

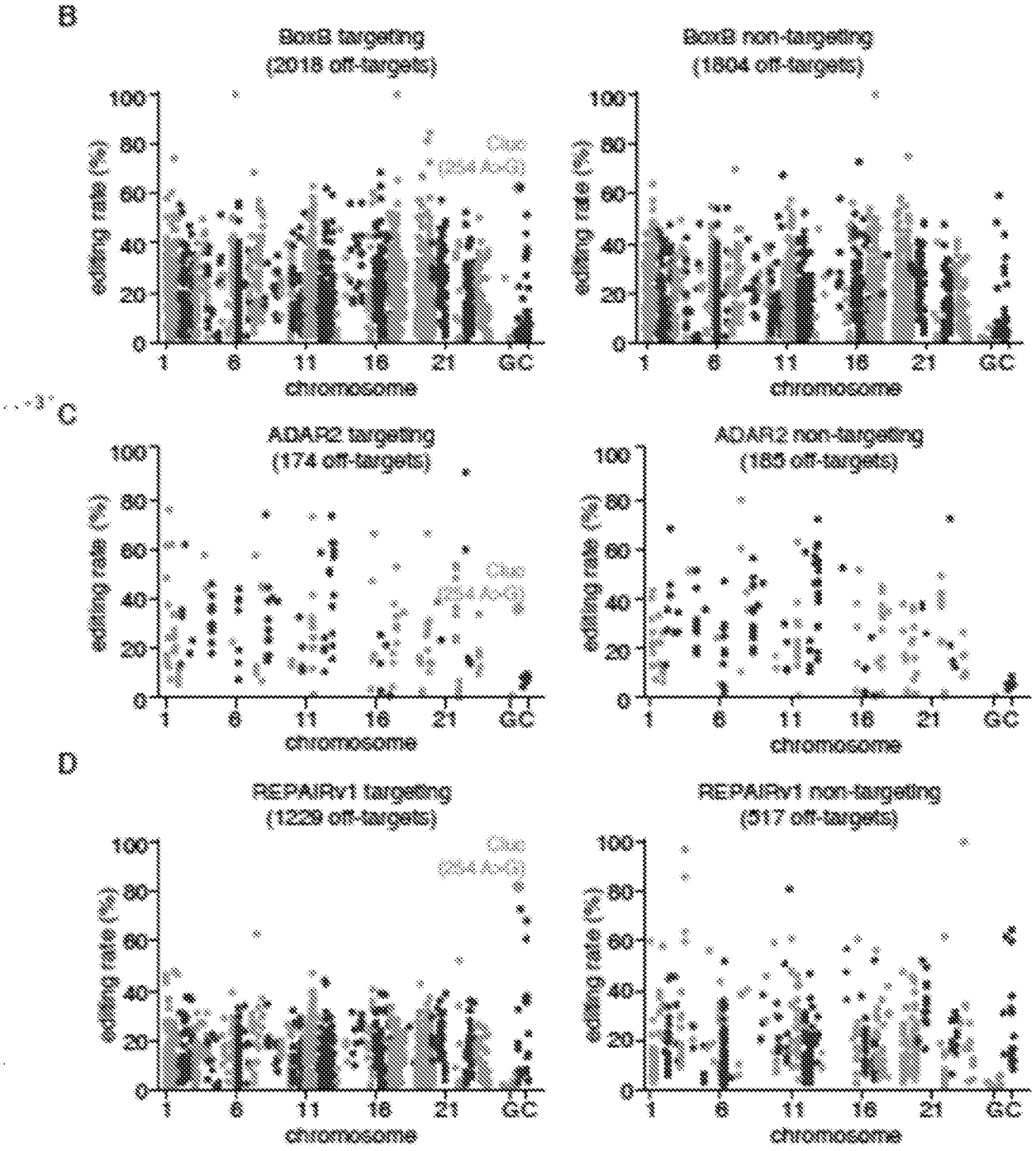


FIG. 60 (continued)

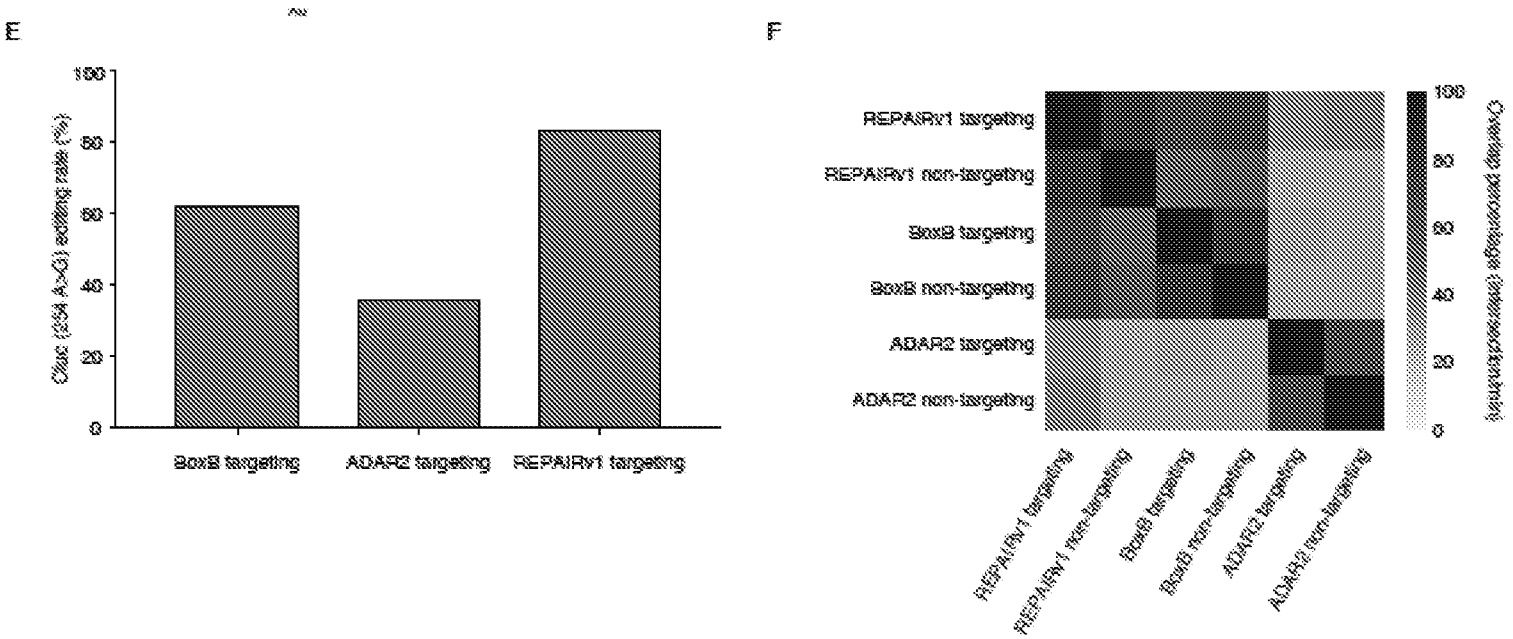


FIG. 60 (continued)

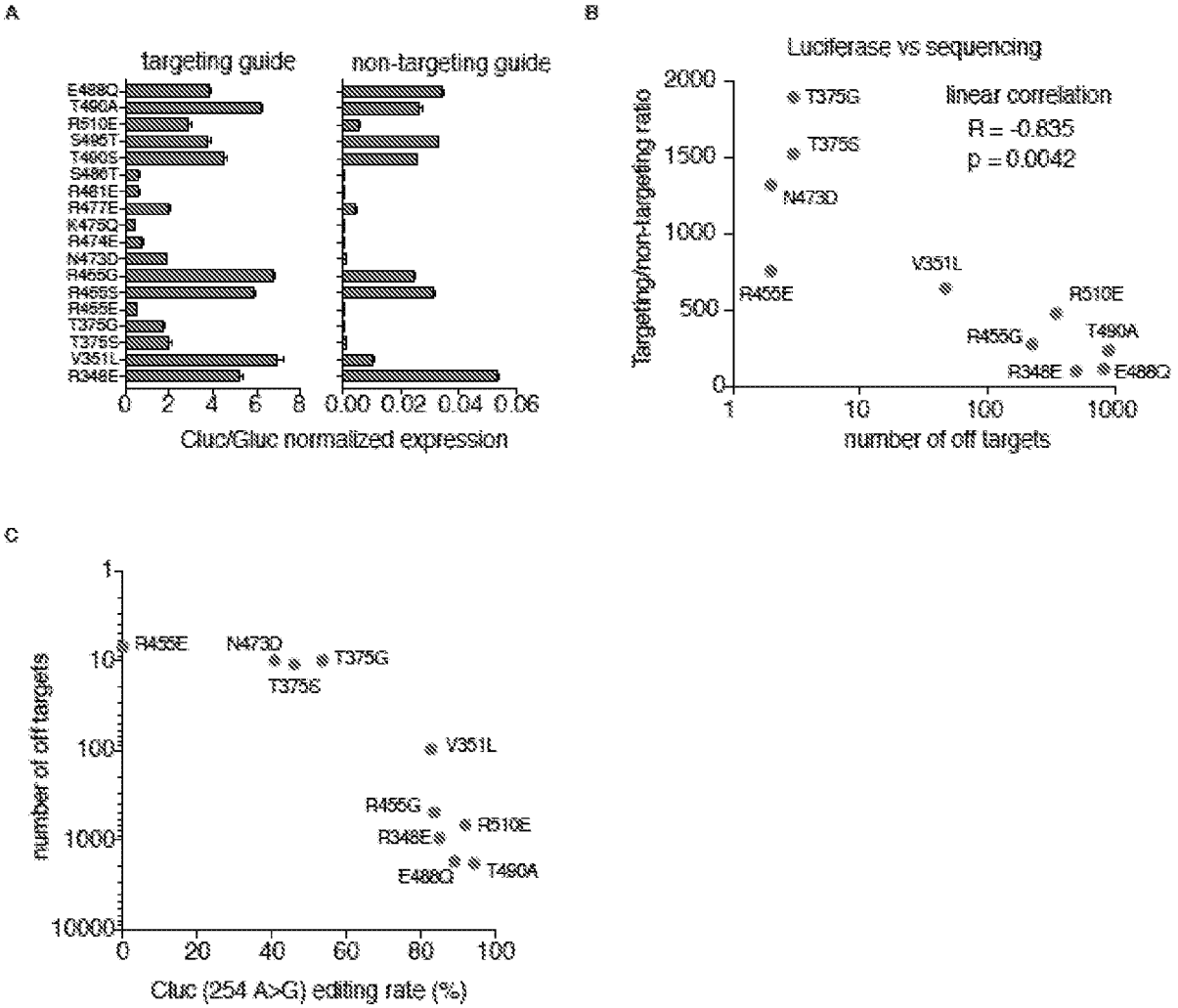

A
targeting guide
non-targeting guide
E488Q
T490A
R510E
S485T
T490S
S486T
R481E
R477E
K475Q
R474E
N473D
R455G
R455S
R455E
T375G
T375S
V351L
R348E
0 2 4 6 8
0.00 0.02 0.04 0.06
Cluc/Gluc normalized expression
B
Luciferase vs sequencing
linear correlation
R = -0.835
p = 0.0042
targeting/non-targeting guide ratio
number of off targets
C
number of off targets
Cluc (254 A>G) editing rate (%)

FIG. 61

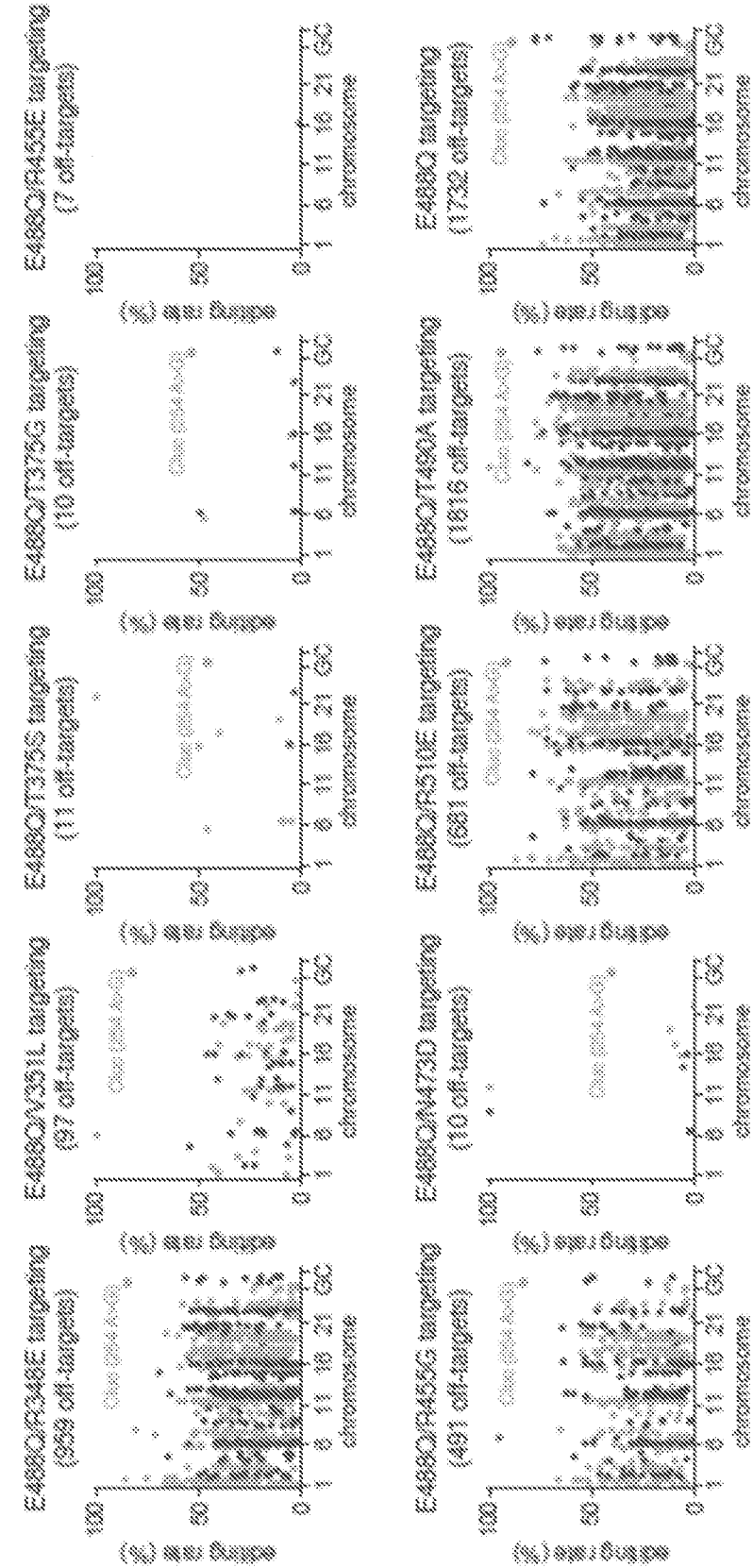

A
E488OR348E targeting (369 off-targets)
E488ON731L targeting (97 off-targets)
E488OT3735 targeting (11 off-targets)
E488OT373G targeting (10 off-targets)
E488OR453E targeting (7 off-targets)
E488OR455G targeting (491 off-targets)
E488ON473D targeting (10 off-targets)
E488OR510E targeting (681 off-targets)
E488OT490A targeting (1816 off-targets)
E488O targeting (1732 off-targets)
on-target
off-target
editing rate (%)
chromosome
1 6 11 16 21 GC

FIG. 62

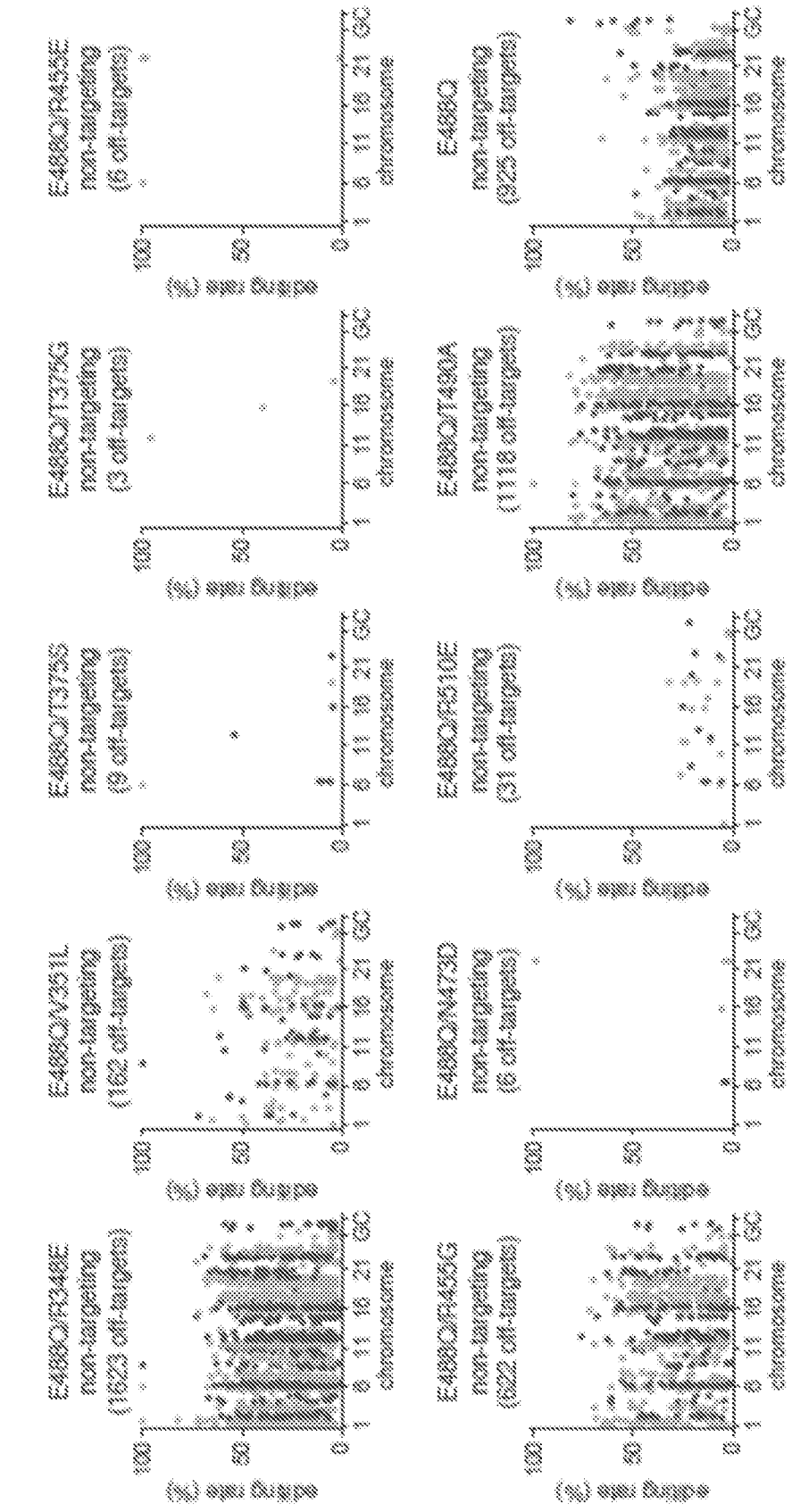

E488OR349E
non-targeting
(1623 off-targets)
E488OV351L
non-targeting
(162 off-targets)
E488OT375S
non-targeting
(9 off-targets)
E488OT377G
non-targeting
(3 off-targets)
E488OR433E
non-targeting
(6 off-targets)
E488OR450G
non-targeting
(622 off-targets)
E488ON473D
non-targeting
(6 off-targets)
E488OR510E
non-targeting
(31 off-targets)
E488OT490A
non-targeting
(1118 off-targets)
E488Q
non-targeting
(925 off-targets)
(%) edits @ pos
chromosome

FIG. 62 cont.

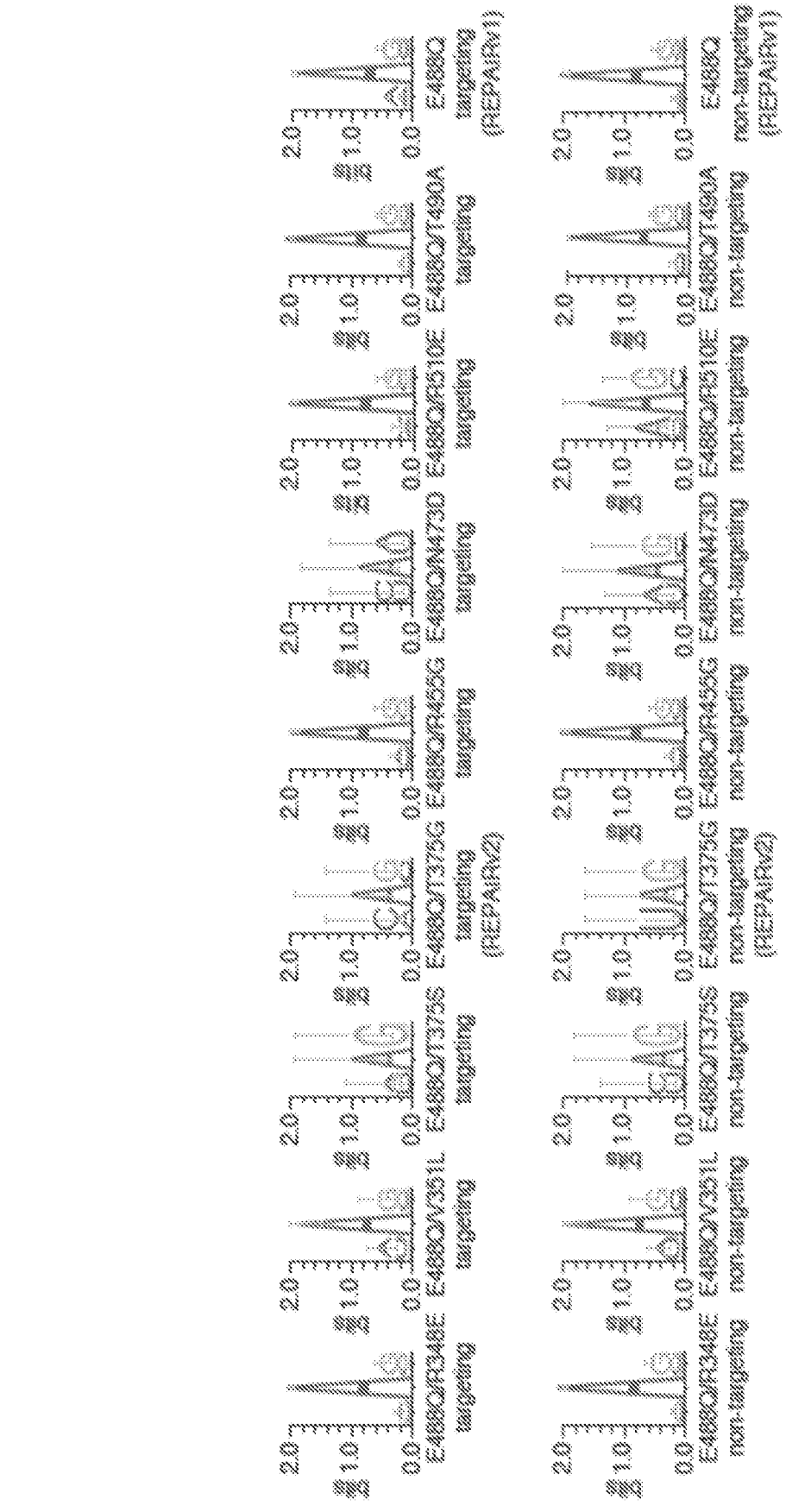

A
E488QR348E
E488QV351L
E488QT375S
E488QT375G
E488QR455G
E488QM473I
E488QR510E
E488QT490A
E488Q
targeting
non-targeting
(REPAIRv2)
(REPAIRv1)
%
2.0
1.0
0.0

FIG. 63

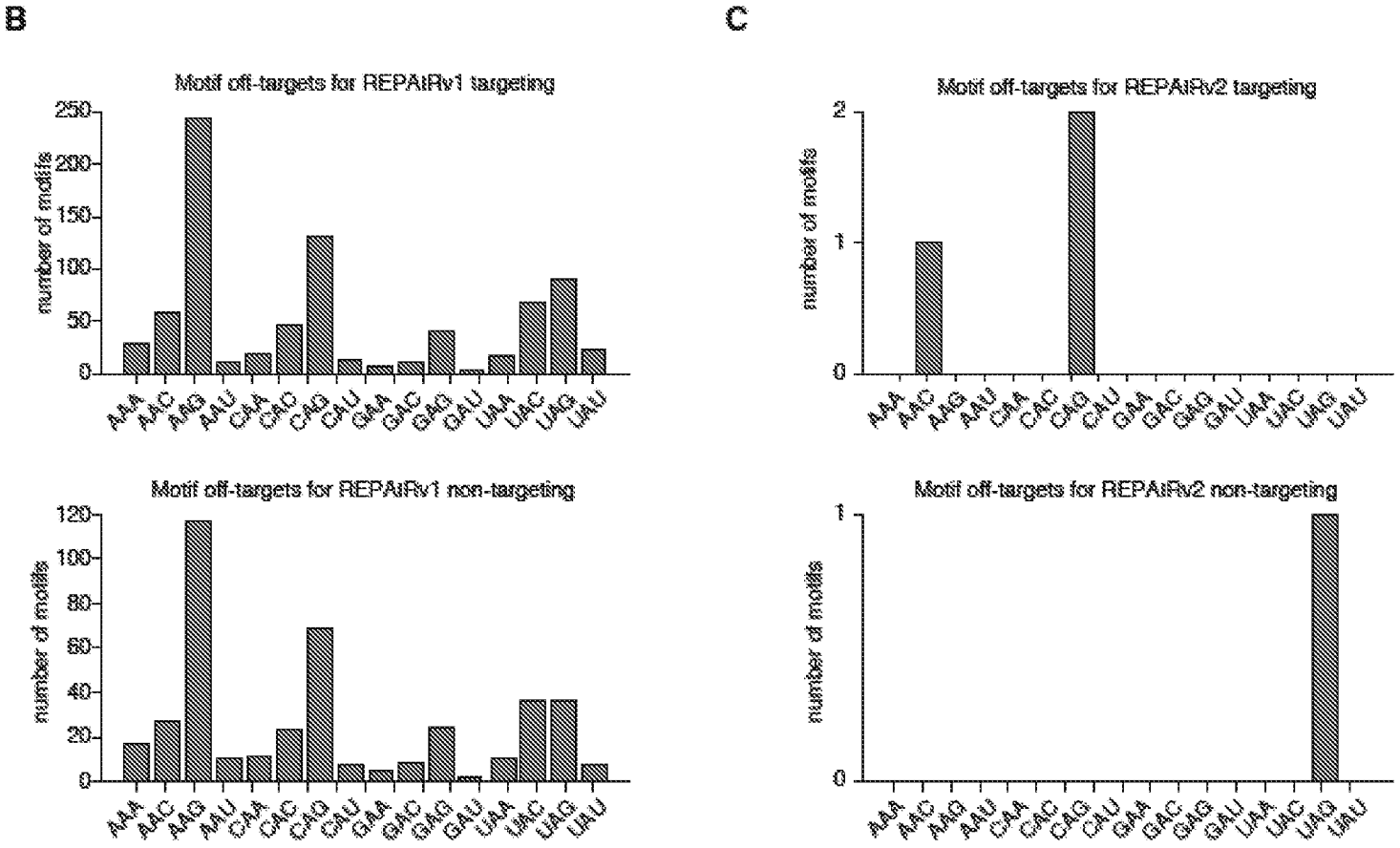


FIG. 63 (continued)

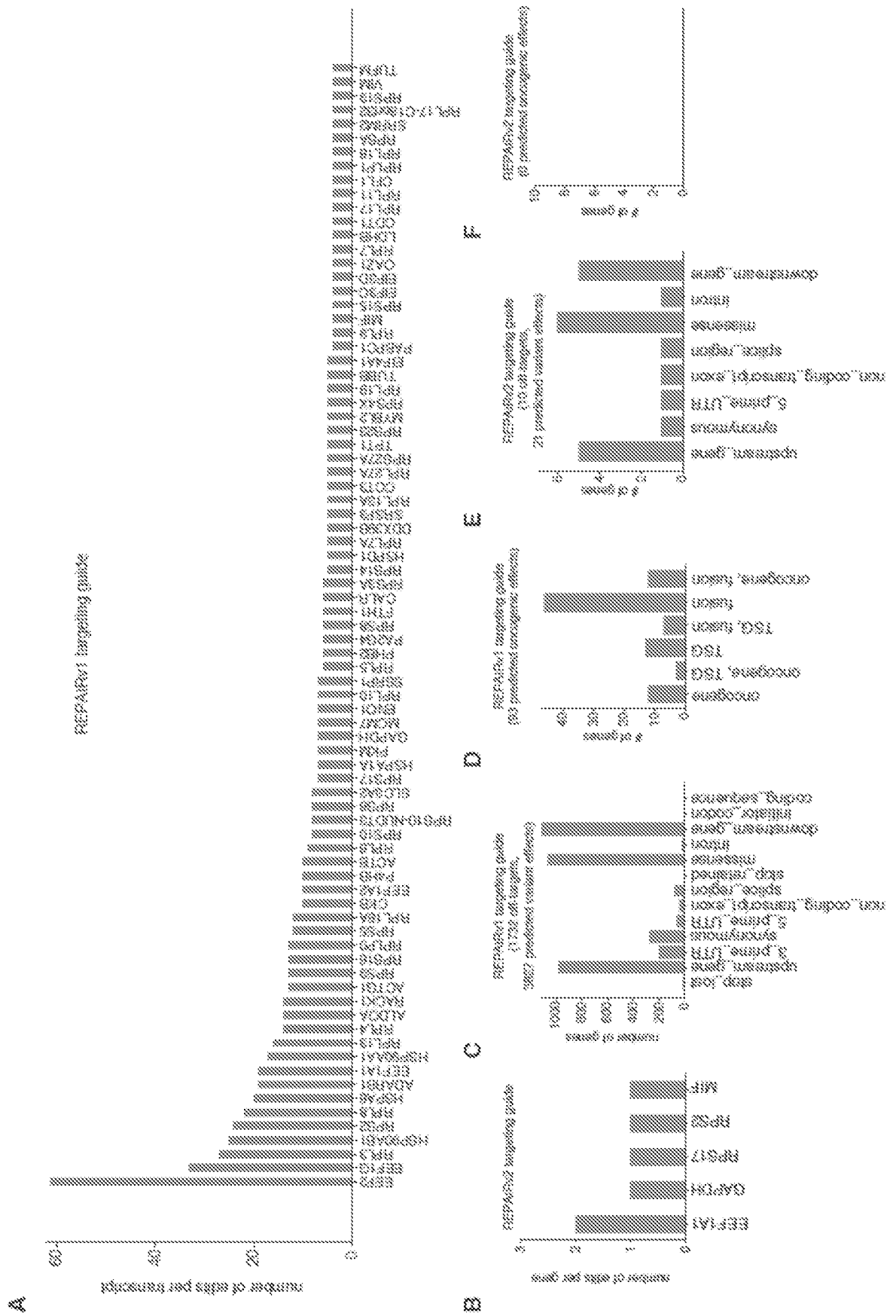

A
REPAIRv1 targeting guide
number of edits per transcript
B
REPAIRv2 targeting guide
number of edits per gene
C
REPAIRv1 targeting guide
(1,732 edits)
(2962 predicted nonsense effects)
number of genes
D
REPAIRv1 targeting guide
(60 predicted nonsense effects)
of genes
oncogene
oncogene, TSG
TSG
TSG, fusion
fusion
oncogene, fusion
E
REPAIRv2 targeting guide
(2,732 edits)
(2 predicted nonsense effects)
of genes
upstream_gene
synonymous
5_prime_UTR
non_coding_transcript_exon
splice_region
missense
intron
downstream_gene
F
REPAIRv2 targeting guide
(0 predicted nonsense effects)
of genes

FIG. 64

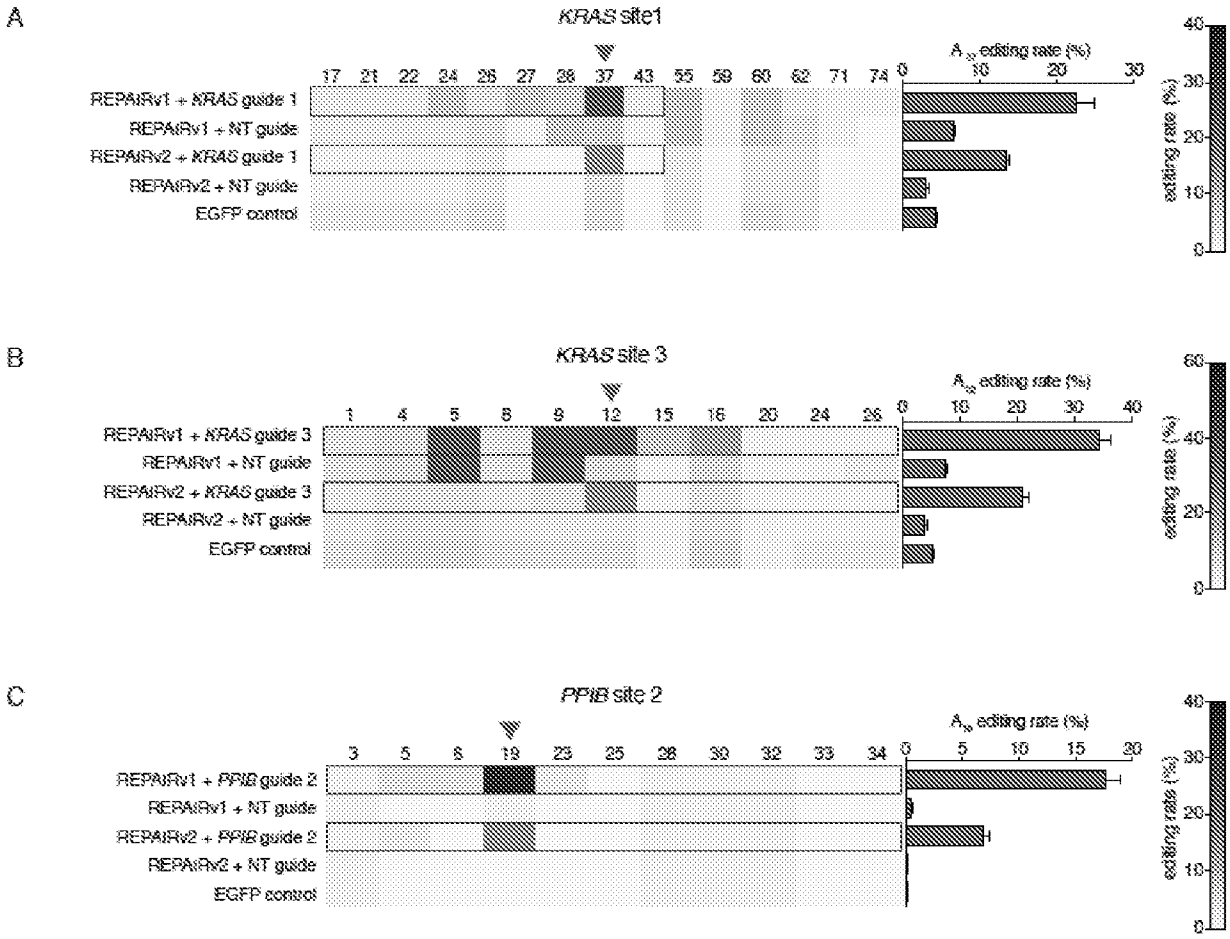

A
KRAS site1
REPAIRv1 + KRAS guide 1
REPAIRv1 + NT guide
REPAIRv2 + KRAS guide 1
REPAIRv2 + NT guide
EGFP control
editing rate (%)
B
KRAS site 3
REPAIRv1 + KRAS guide 3
REPAIRv1 + NT guide
REPAIRv2 + KRAS guide 3
REPAIRv2 + NT guide
EGFP control
editing rate (%)
C
PPIB site 2
REPAIRv1 + PPIB guide 2
REPAIRv1 + NT guide
REPAIRv2 + PPIB guide 2
REPAIRv2 + NT guide
EGFP control
editing rate (%)

FIG. 65

A

Initial Codon		Edited Codon
Tyr	→	Cys
stop	→	Trp
His	→	Arg
Gln	→	Arg
Asn	→	Ser
		Asp
Lys	→	Arg
		Glu
Ser	→	Gly
Arg	→	Gly
Thr	→	Ala
Ile	→	Val
Met	→	Val

B

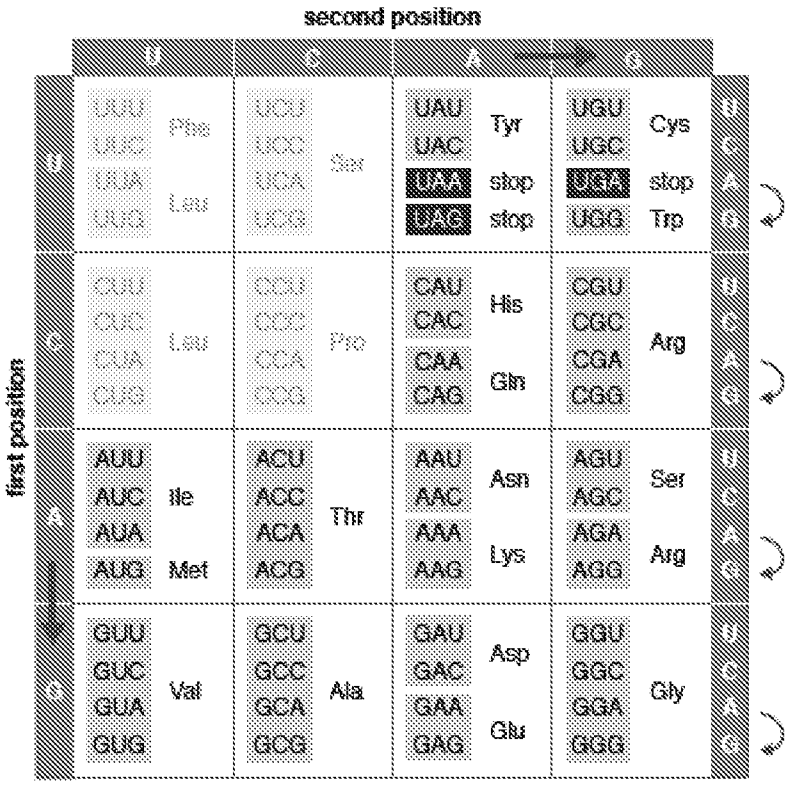


FIG. 66

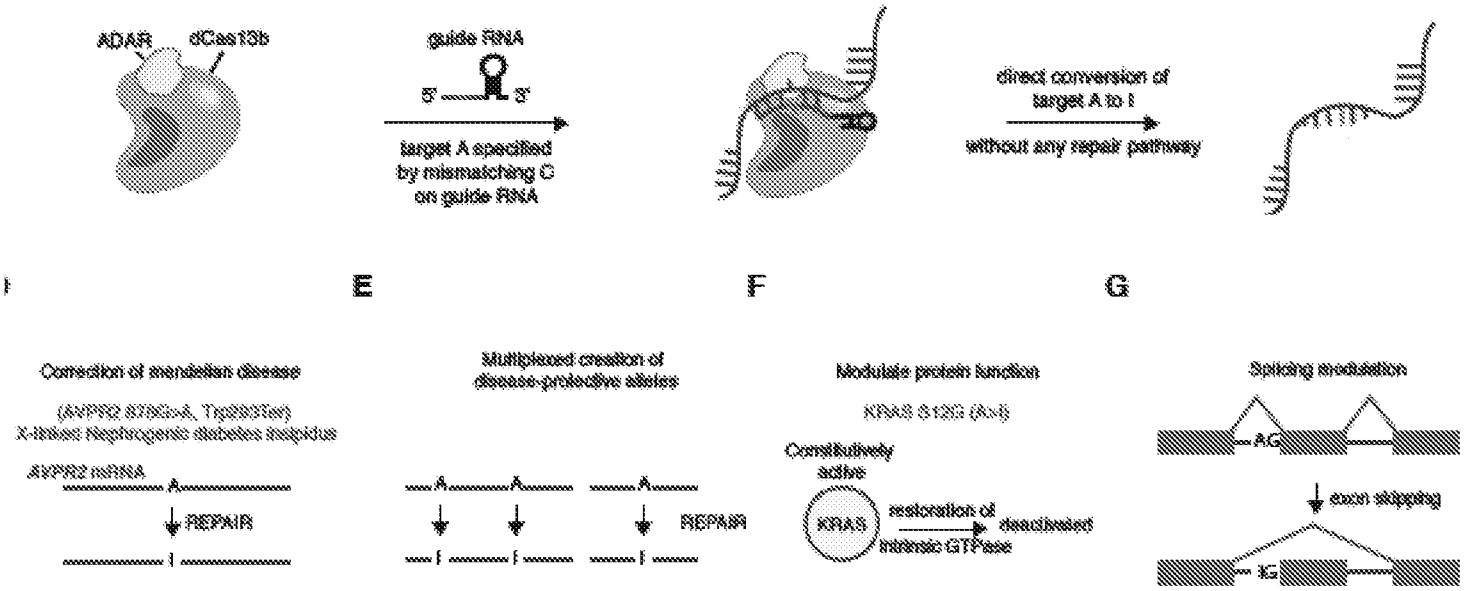


FIG. 66
(continued)

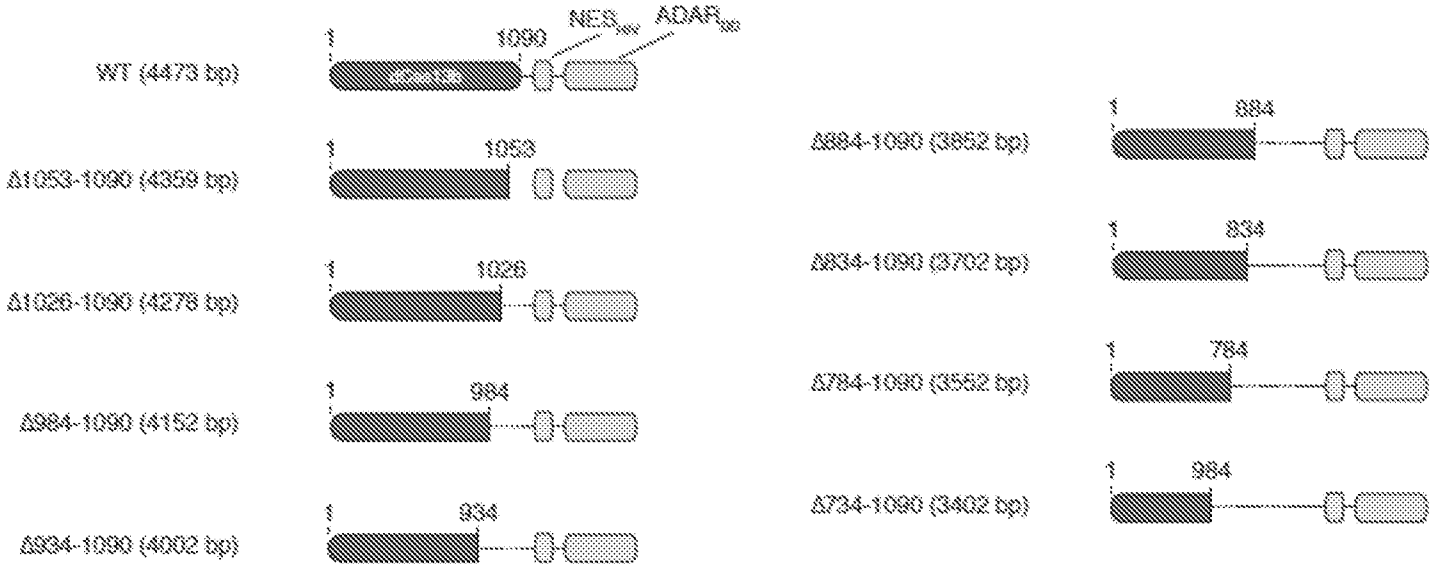

WT (4473 bp)
1
1090
NES
ADAR
Δ1053-1090 (4359 bp)
1053
Δ1026-1090 (4278 bp)
1026
Δ984-1090 (4152 bp)
984
Δ934-1090 (4002 bp)
934
Δ884-1090 (3852 bp)
884
Δ834-1090 (3702 bp)
834
Δ784-1090 (3552 bp)
784
Δ734-1090 (3402 bp)
734

FIG. 67

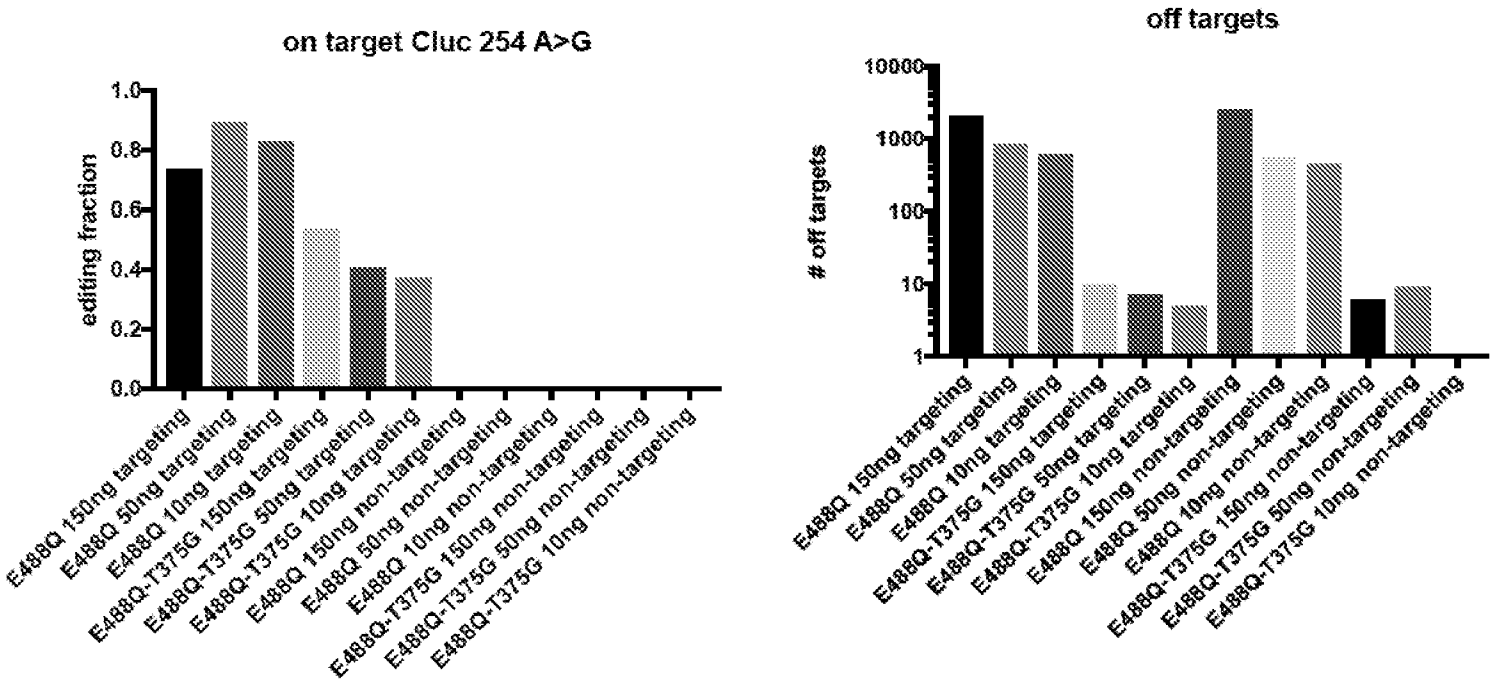

on target Cluc 254 A>G
editing fraction
1.0
0.8
0.6
0.4
0.2
0.0
E488Q 150ng targeting
E488Q 50ng targeting
E488Q 10ng targeting
E488Q-T375G 150ng targeting
E488Q-T375G 50ng targeting
E488Q-T375G 10ng targeting
E488Q 150ng non-targeting
E488Q 50ng non-targeting
E488Q 10ng non-targeting
E488Q-T375G 150ng non-targeting
E488Q-T375G 50ng non-targeting
E488Q-T375G 10ng non-targeting
off targets
off targets
10000
1000
100
10
1
E488Q 150ng targeting
E488Q 50ng targeting
E488Q 10ng targeting
E488Q-T375G 150ng targeting
E488Q-T375G 50ng targeting
E488Q-T375G 10ng targeting
E488Q 150ng non-targeting
E488Q 50ng non-targeting
E488Q 10ng non-targeting
E488Q-T375G 150ng non-targeting
E488Q-T375G 50ng non-targeting
E488Q-T375G 10ng non-targeting

FIG. 68

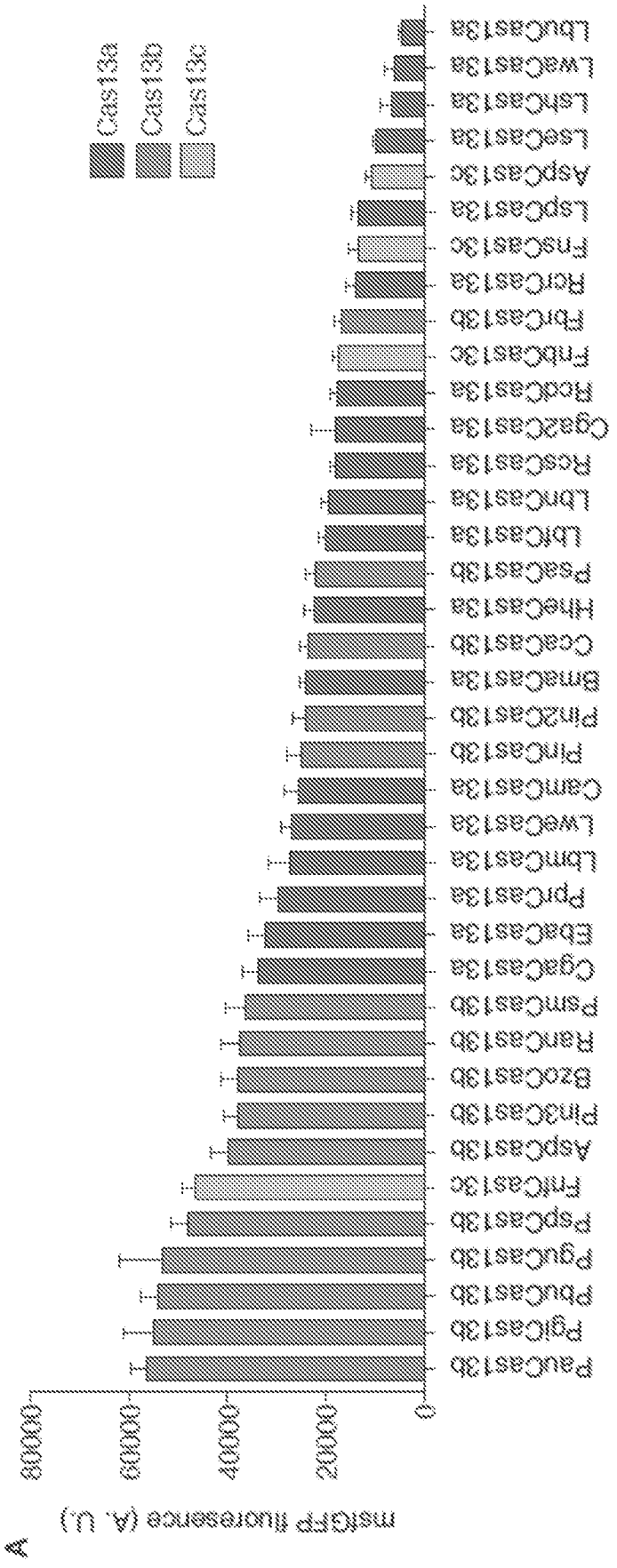

A
msfGFP fluorescence (A. U.)
80000
60000
40000
20000
0
Cas13a
Cas13b
Cas13c
PauCas13b
PglCas13b
PbuCas13b
PguCas13b
PspCas13b
FmCas13c
AspCas13b
Pin3Cas13b
BzoCas13b
RanCas13b
PsmCas13b
CgaCas13a
EbaCas13a
PprCas13a
LbmCas13a
LweCas13a
CamCas13a
PinCas13b
Pin2Cas13b
BmaCas13a
CcaCas13b
HheCas13a
PsaCas13b
LbiCas13a
LbnCas13a
RcsCas13a
Cga2Cas13a
RcdCas13a
FnbCas13c
FbrCas13b
RcrCas13a
FnsCas13c
LspCas13a
AspCas13c
LseCas13a
LshCas13a
LwaCas13a
LbuCas13a

FIG. 69

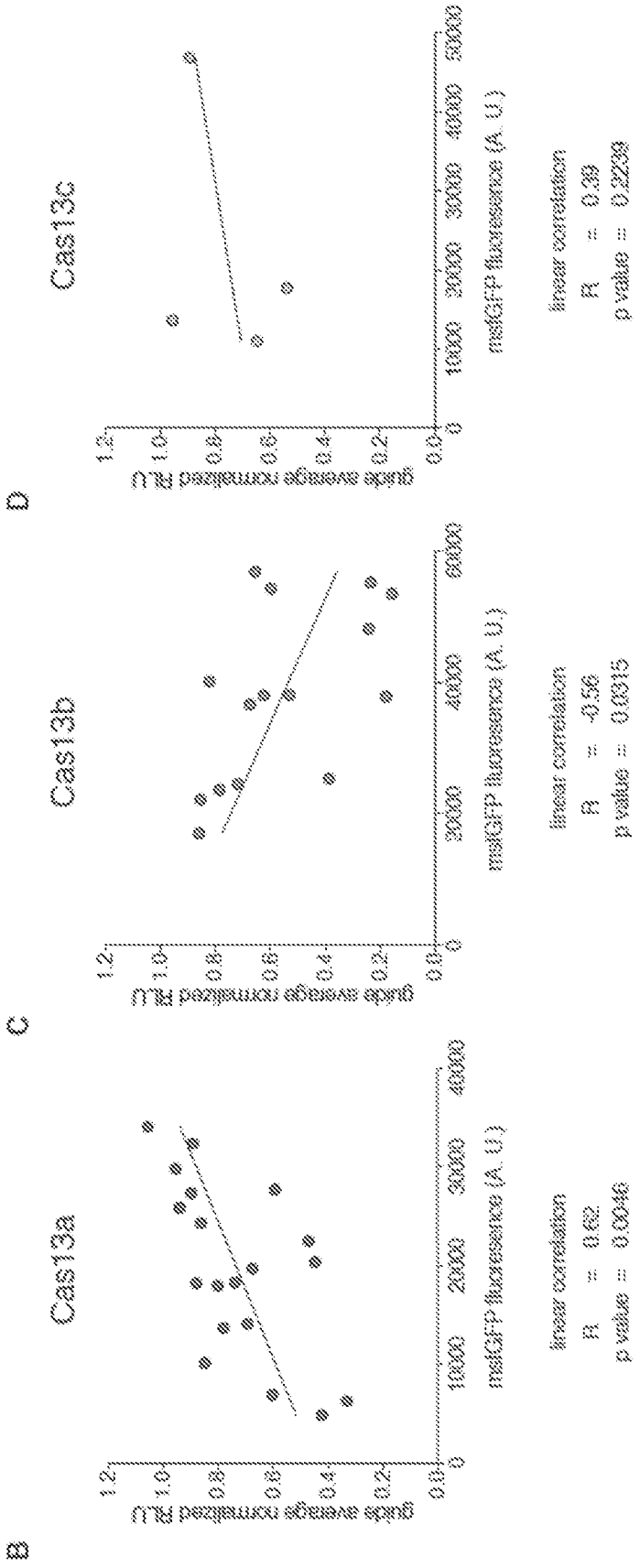


FIG. 69 (continued)

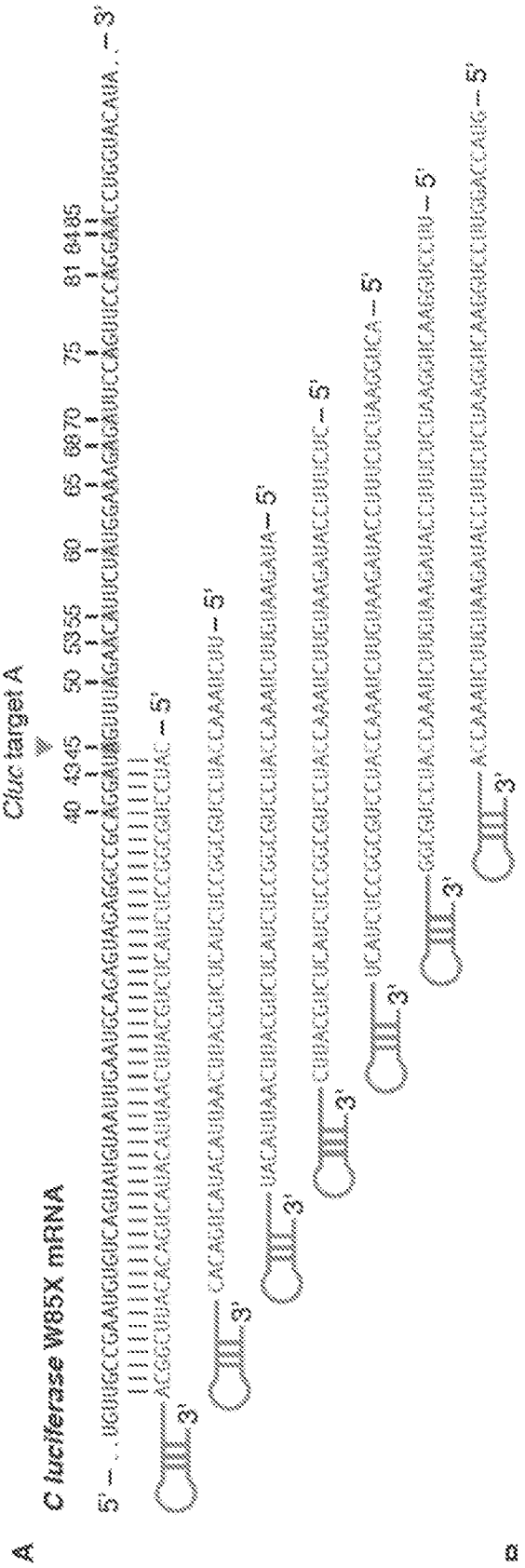

A
C luciferase W85X mRNA
Cluc target A
40 43 45 50 53 55 60 65 68 70 75 81 84 85
5' – UGUUGCCGAAUGUGCAGUAUGAAUGCAAUGCAGACCGCACAGGCAUGCUUAGAGAUUCAAUCGAUUGUGAAGAGAGAUUCCAGUUCCAGGACCUGGUACAUA – 3'
3' ACGGCUUACACAGUCAUACAUUACGUUACGUCUAUGCGGCGUGUCCGGGGGAGGUCCUAU – 5'
3' CAUAUUACAUUAACUUACUUGAGUCUUAUCAAAUCUU – 5'
3' UAUAUUACAUUAACUUACUUGAGUCUUAUCAAAUCUUGUAAGAUA – 5'
3' CUUACUUCUUAUCUUGAGUCUUAUCAAAUCUUGUAAGAUACUUUCUC – 5'
3' UCAUCUUCGGAUCUUACAAAUCUUGUAAGAUACUUUCUCUAAGGUCA – 5'
3' GUAGUCUUACUAAAUCUUUGUAAGAUACUUUCUCUAAGUUCAAGUCCUU – 5'
3' AUCAAAUCUUGUAAGAUACUUUCUCUAAGUUCAAGUCCUUGGACCAUG – 5'
B

FIG. 70

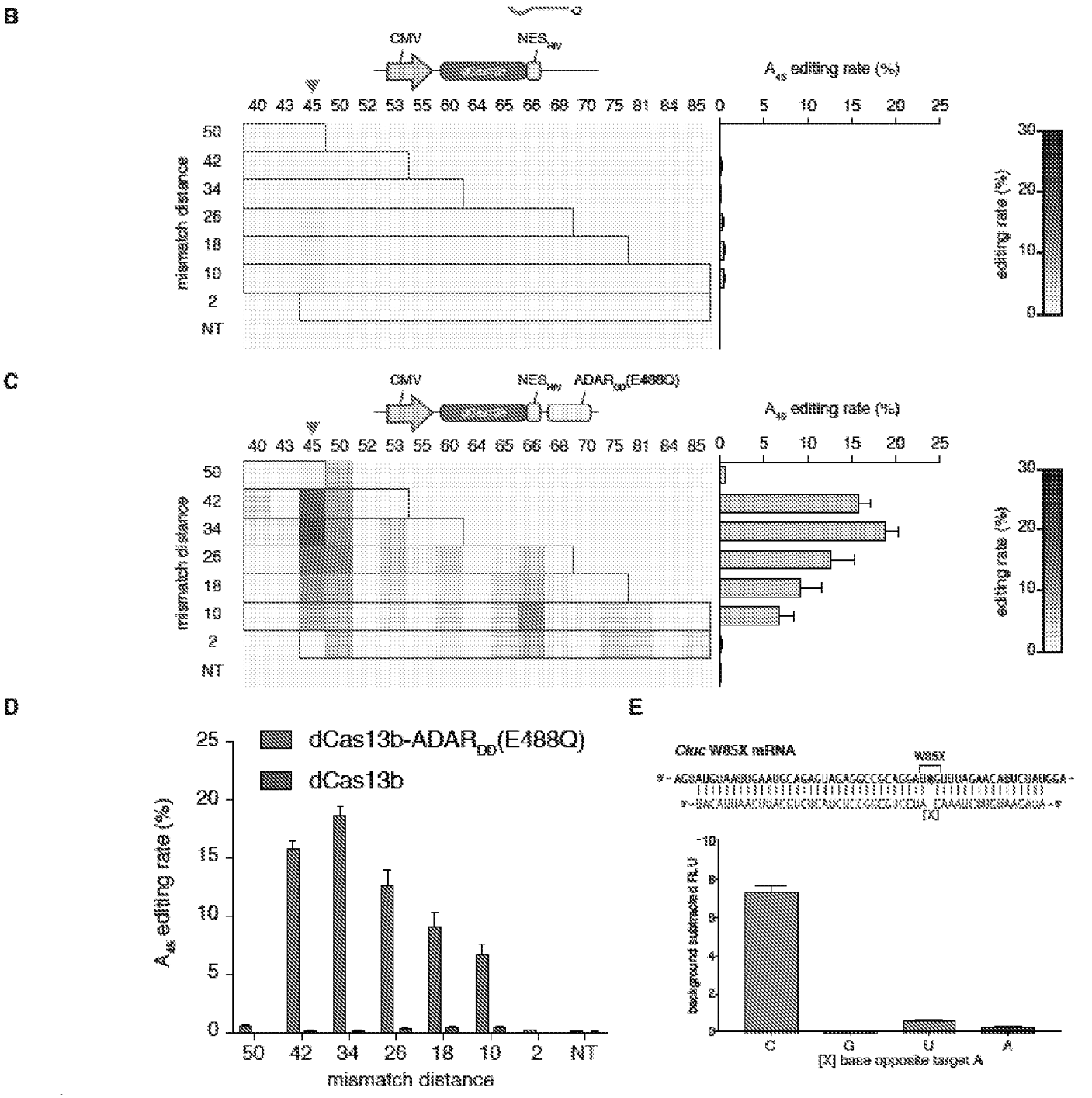


FIG. 70 (continued)

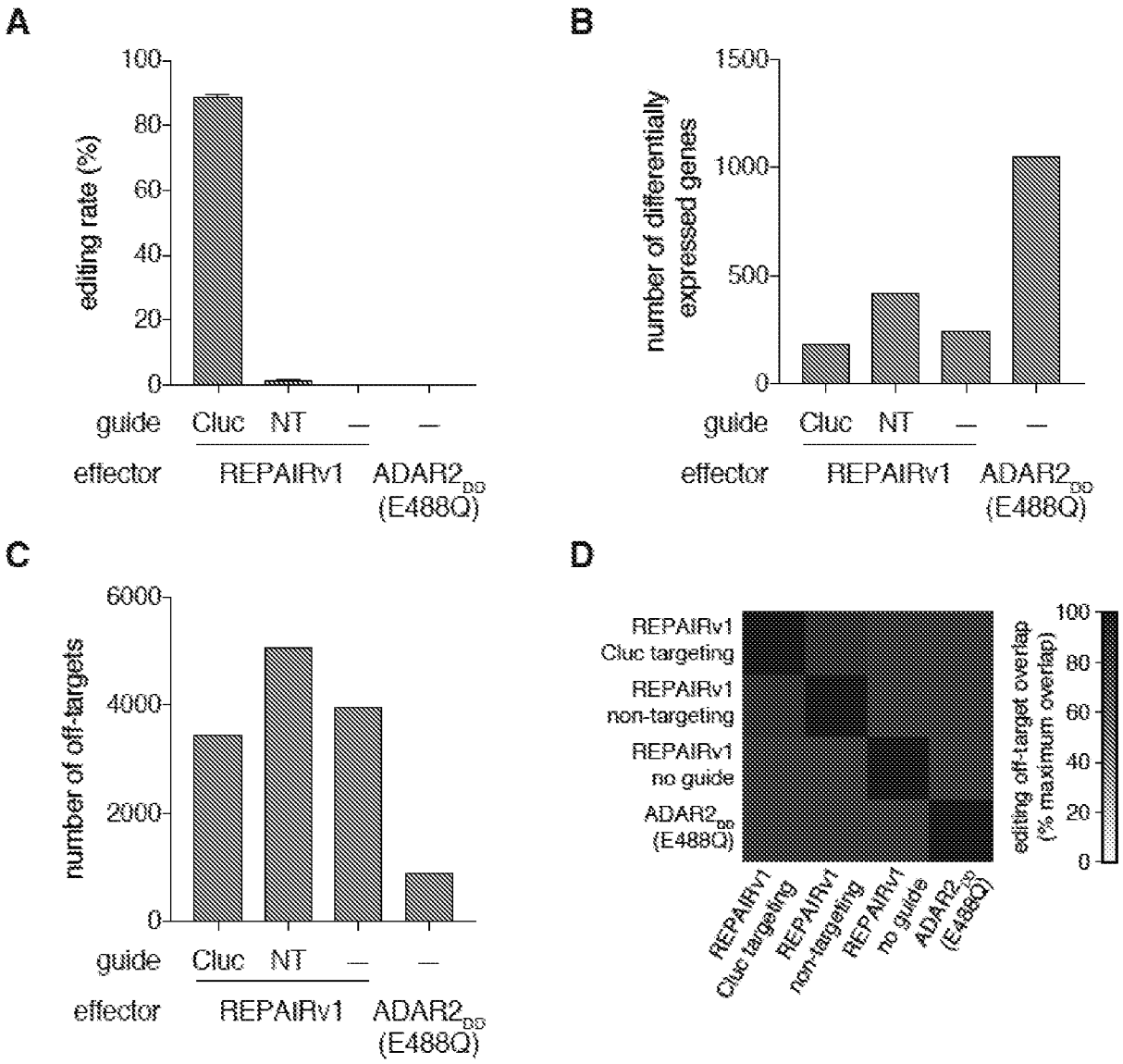

A
editing rate (%)
100
80
60
40
20
0
guide
Cluc
NT
—
—
effector
REPAIRv1
ADAR2DD (E488Q)
B
number of differentially expressed genes
1500
1000
500
0
guide
Cluc
NT
—
—
effector
REPAIRv1
ADAR2DD (E488Q)
C
number of off-targets
6000
4000
2000
0
guide
Cluc
NT
—
—
effector
REPAIRv1
ADAR2DD (E488Q)
D
REPAIRv1 Cluc targeting
REPAIRv1 non-targeting
REPAIRv1 no guide
ADAR2DD (E488Q)
editing off-target overlap (% maximum overlap)
100
80
60
40
20
0

FIG. 71

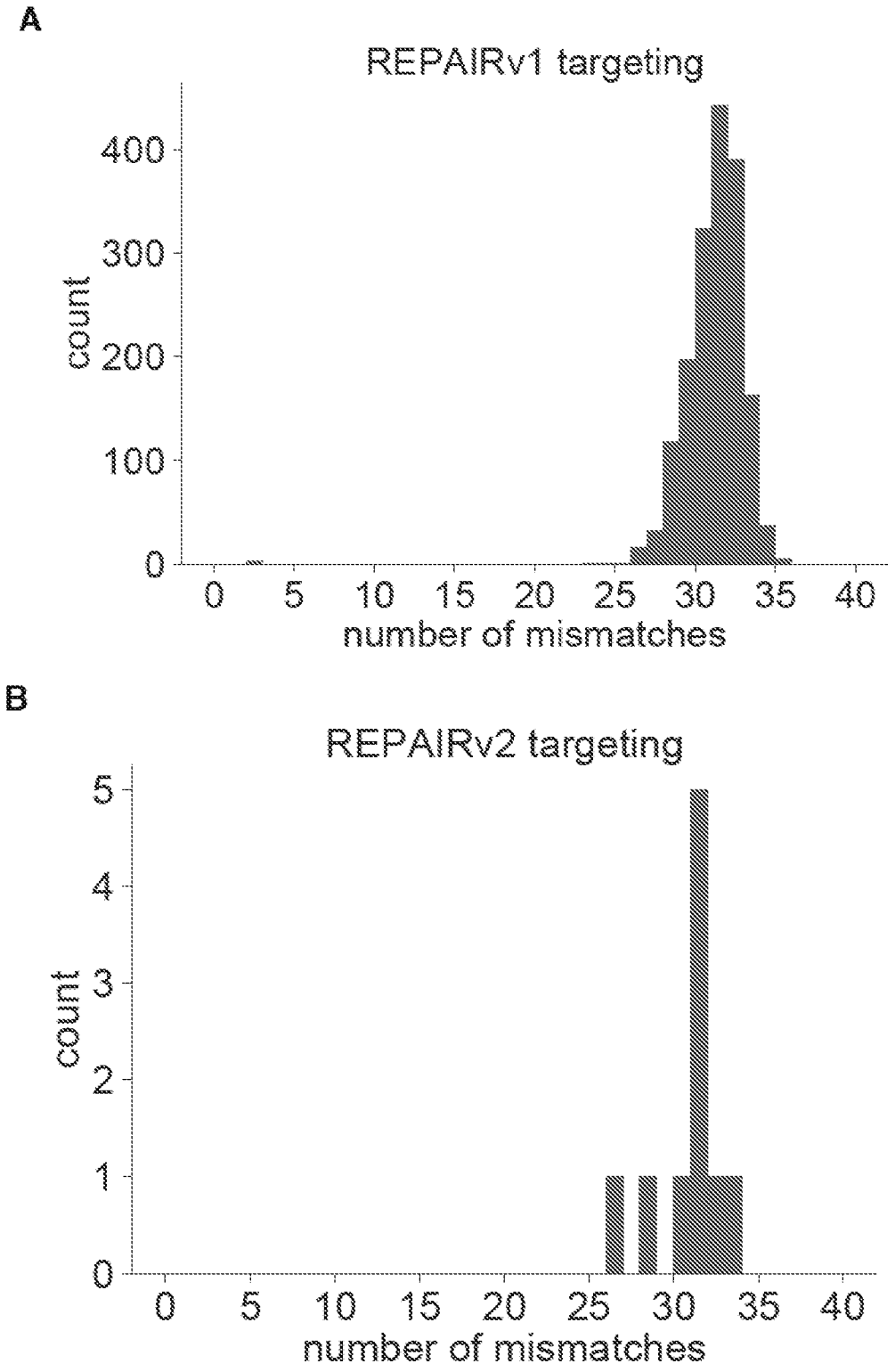

A
REPAIRv1 targeting
count
400
300
200
100
0
0 5 10 15 20 25 30 35 40
number of mismatches
B
REPAIRv2 targeting
count
5
4
3
2
1
0
0 5 10 15 20 25 30 35 40
number of mismatches

FIG. 72

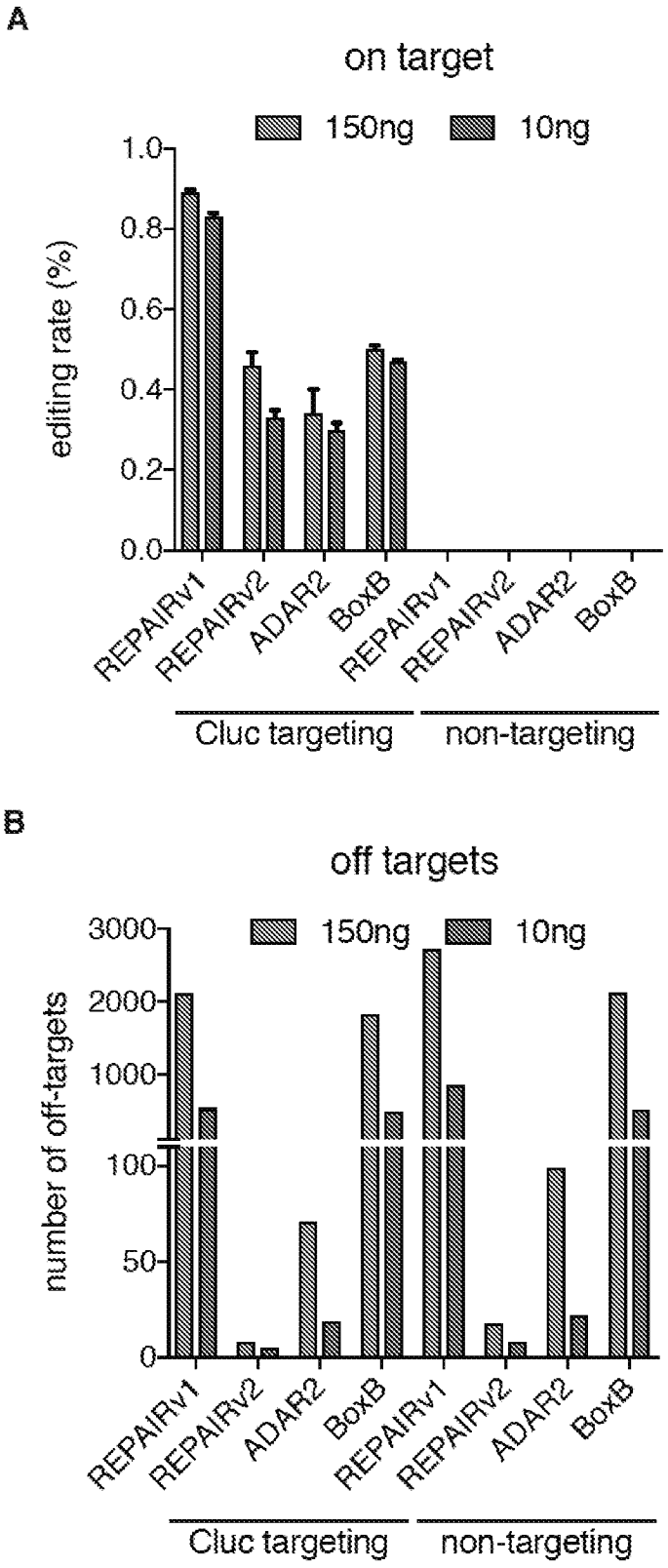

A
on target
150ng
10ng
editing rate (%)
1.0
0.8
0.6
0.4
0.2
0.0
REPAIRv1
REPAIRv2
ADAR2
BoxB
REPAIRv1
REPAIRv2
ADAR2
BoxB
Cluc targeting
non-targeting
B
off targets
150ng
10ng
number of off-targets
3000
2000
1000
100
50
0
REPAIRv1
REPAIRv2
ADAR2
BoxB
REPAIRv1
REPAIRv2
ADAR2
BoxB
Cluc targeting
non-targeting

FIG. 73

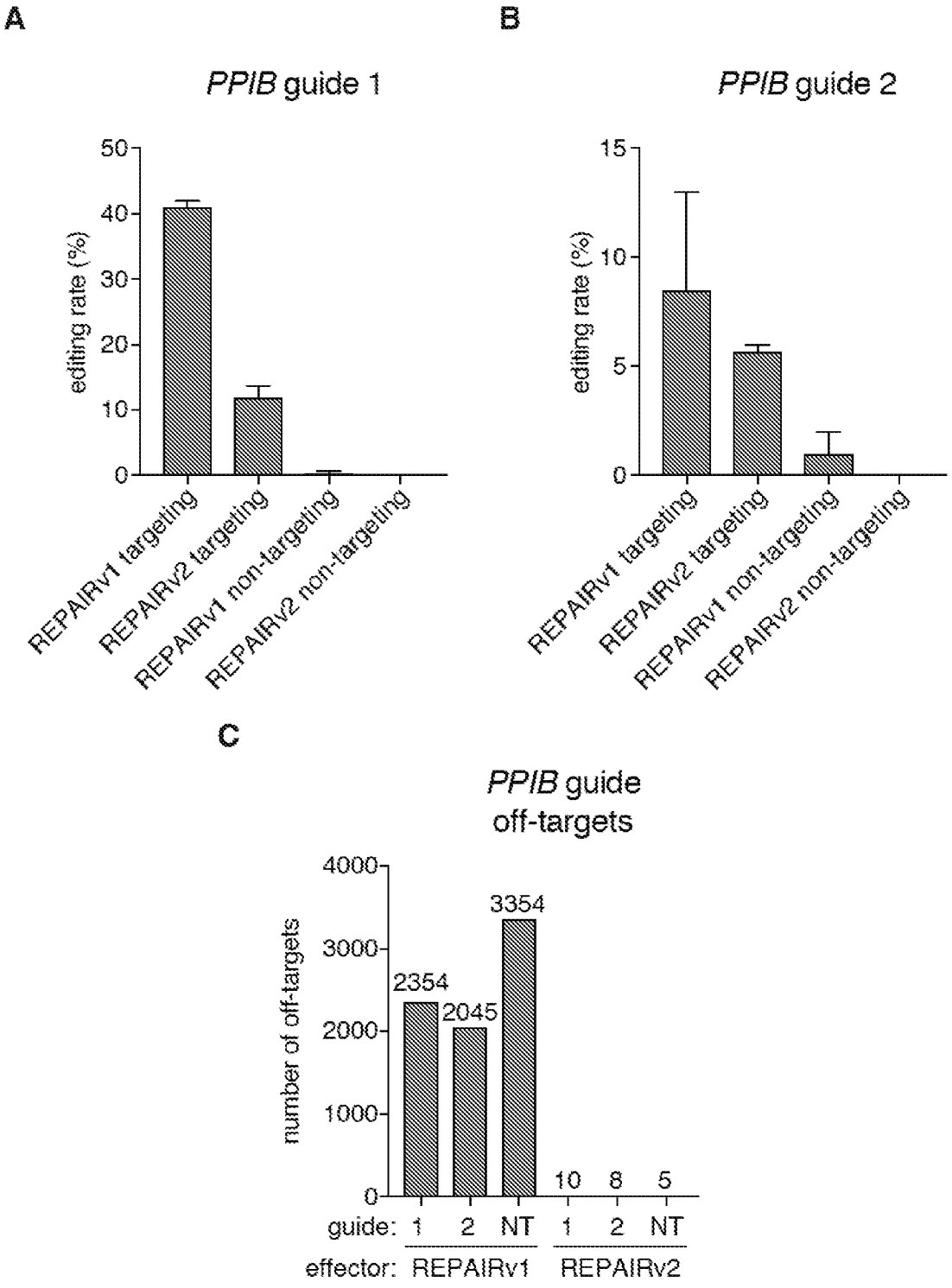

A
PPIB guide 1
editing rate (%)
50
40
30
20
10
0
REPAIRv1 targeting
REPAIRv2 targeting
REPAIRv1 non-targeting
REPAIRv2 non-targeting
B
PPIB guide 2
editing rate (%)
15
10
5
0
REPAIRv1 targeting
REPAIRv2 targeting
REPAIRv1 non-targeting
REPAIRv2 non-targeting
C
PPIB guide
off-targets
number of off-targets
4000
3000
2000
1000
0
2354
2045
3354
10
8
5
guide: 1 2 NT 1 2 NT
effector: REPAIRv1 REPAIRv2

FIG. 74

D

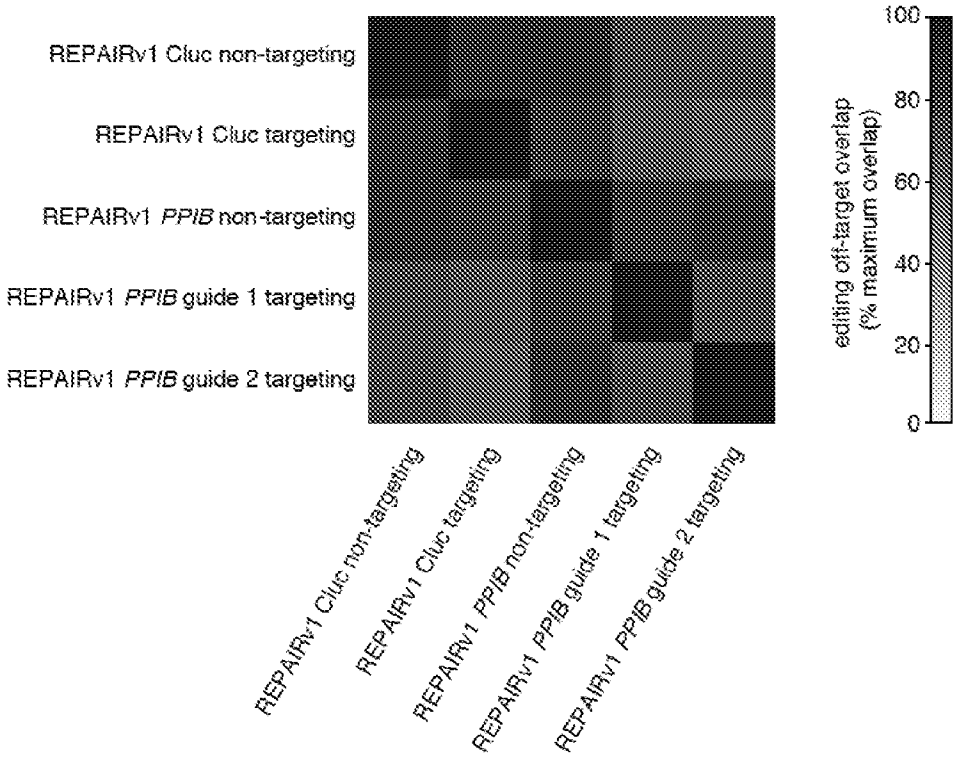


FIG. 74 (continued)

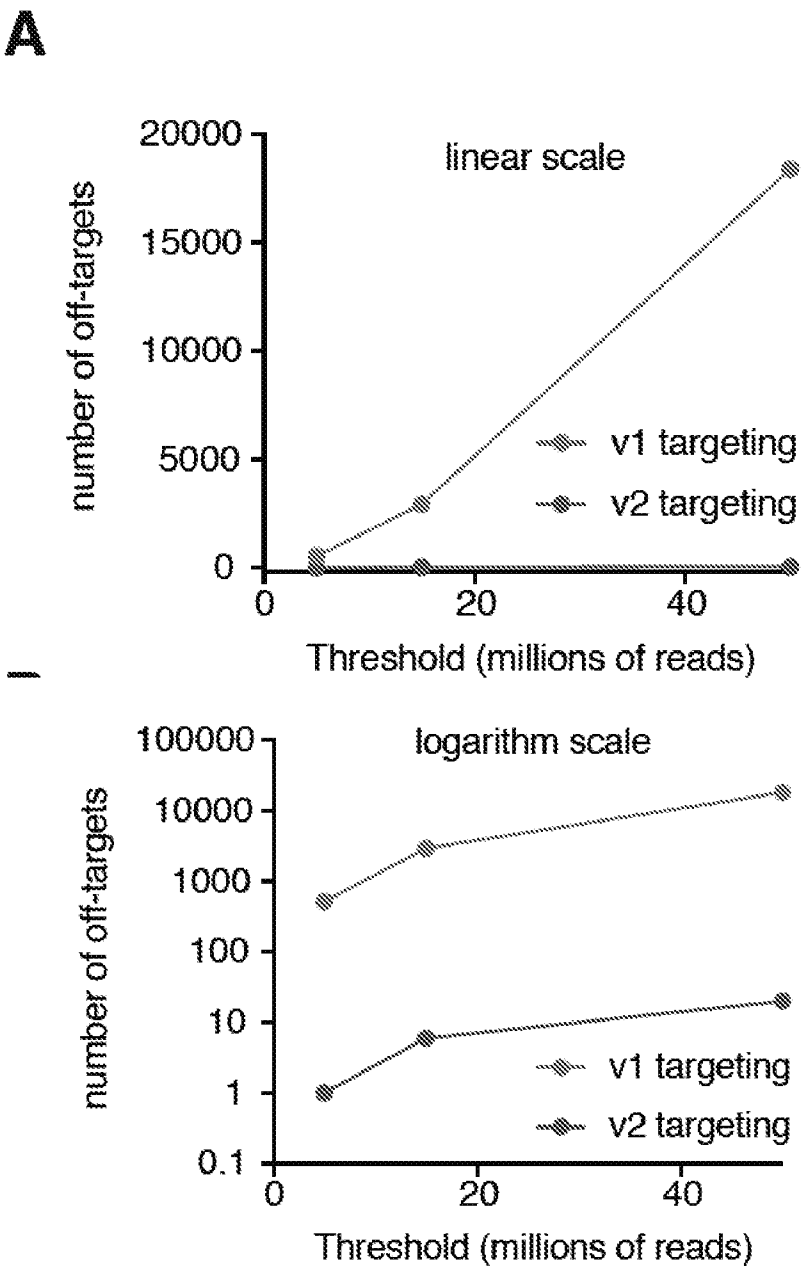

A
linear scale
number of off-targets
20000
15000
10000
5000
0
0
20
40
Threshold (millions of reads)
v1 targeting
v2 targeting
logarithm scale
number of off-targets
100000
10000
1000
100
10
1
0.1
0
20
40
Threshold (millions of reads)
v1 targeting
v2 targeting

FIG. 75

FIG. 75 (continued)

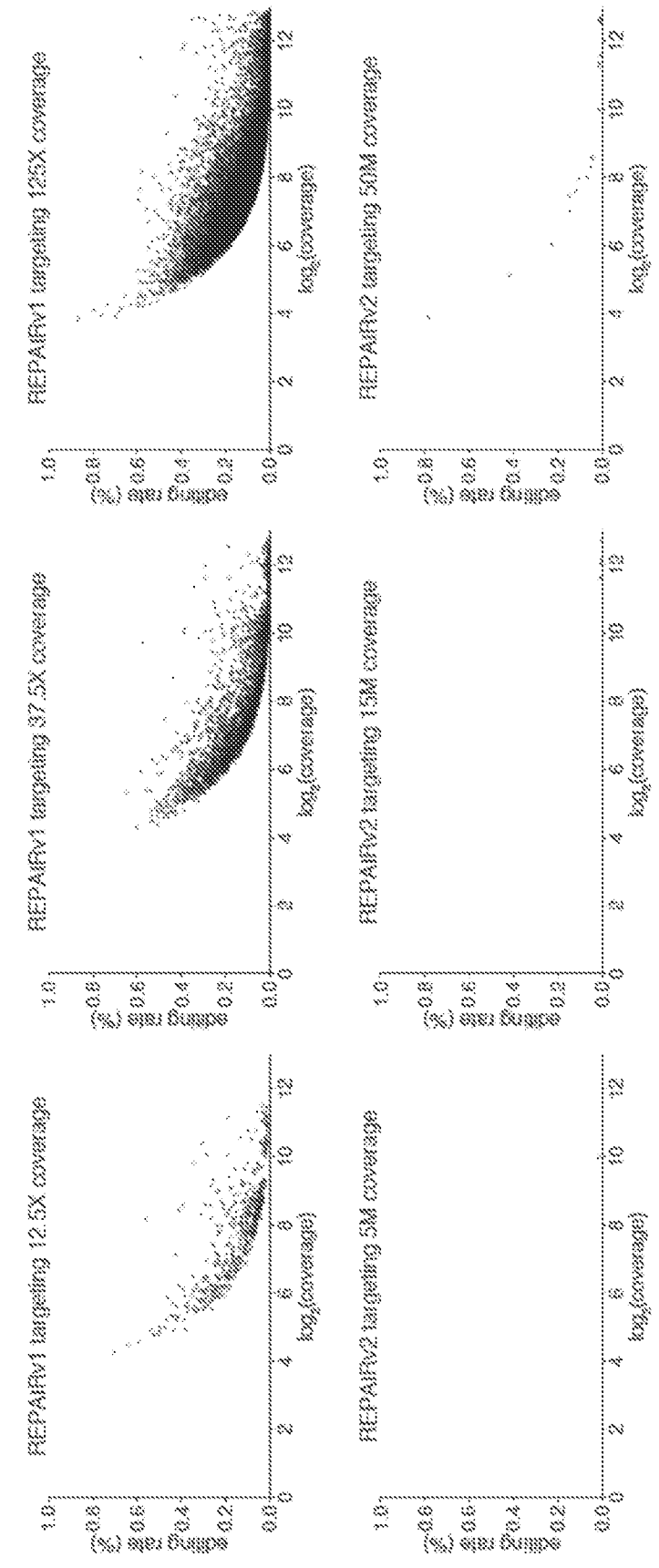


FIG. 75 (continued)

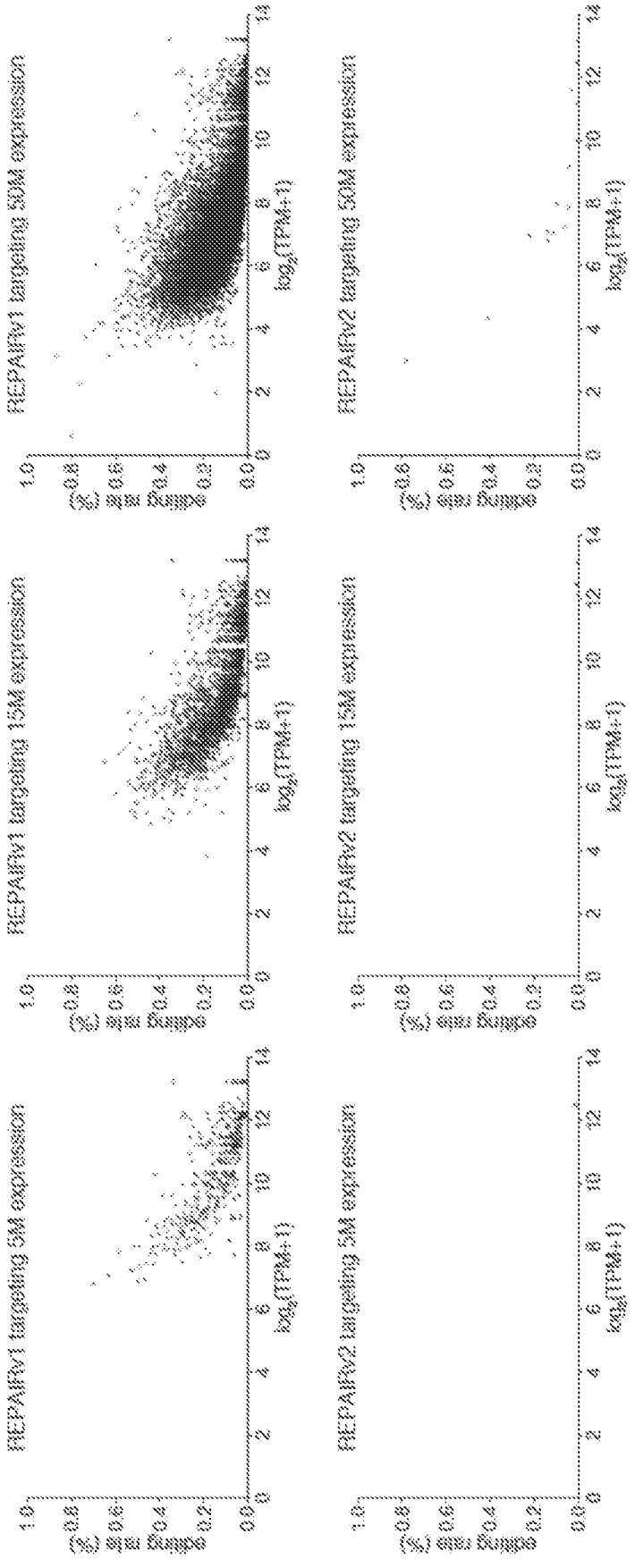


FIG. 75 (continued)

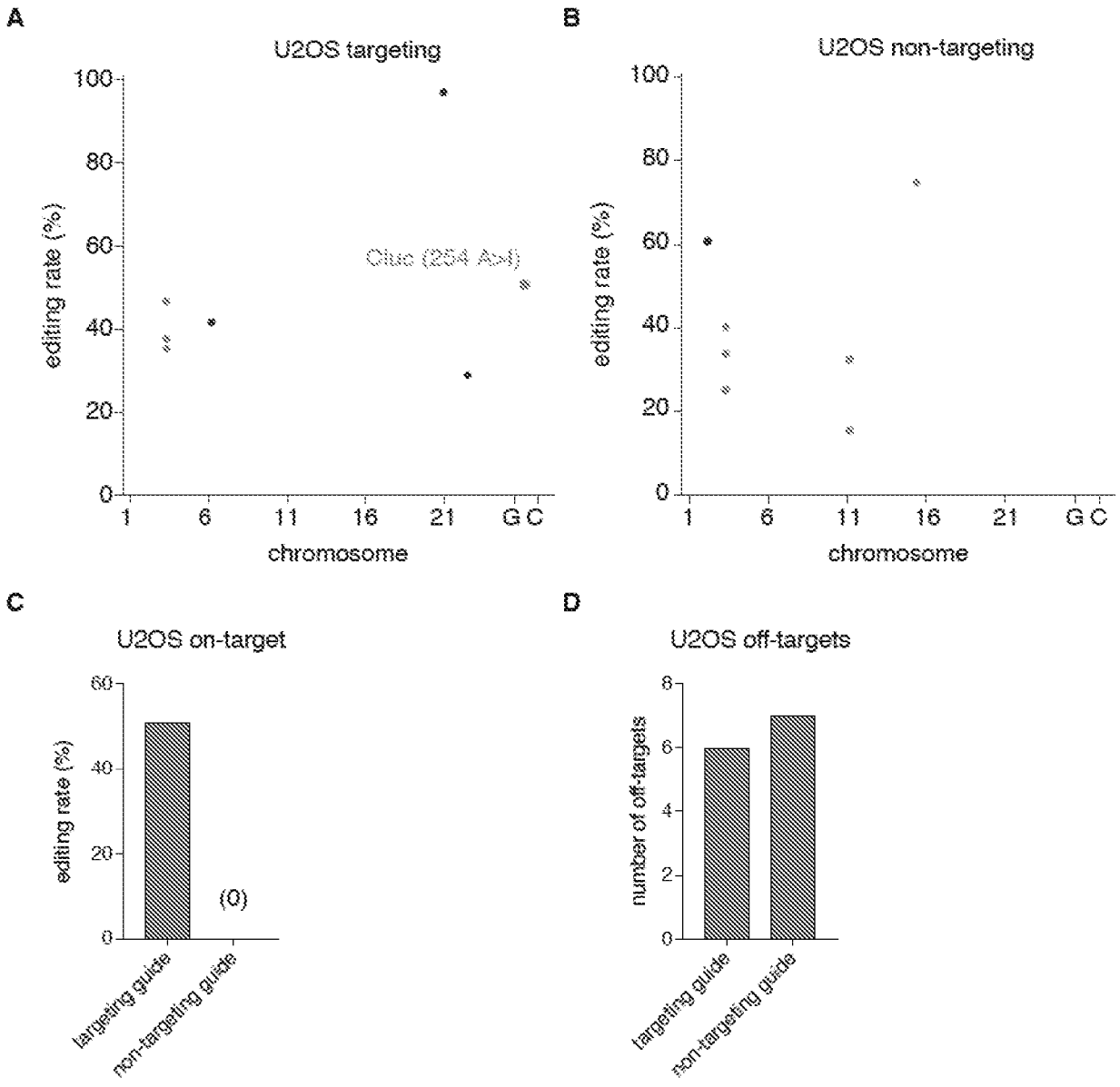

A
U2OS targeting
editing rate (%)
100
80
60
40
20
0
Cluc (254 A>I)
1
6
11
16
21
G C
chromosome
B
U2OS non-targeting
editing rate (%)
100
80
60
40
20
0
1
6
11
16
21
G C
chromosome
C
U2OS on-target
editing rate (%)
60
40
20
0
(0)
targeting guide
non-targeting guide
D
U2OS off-targets
number of off-targets
8
6
4
2
0
targeting guide
non-targeting guide

FIG. 76

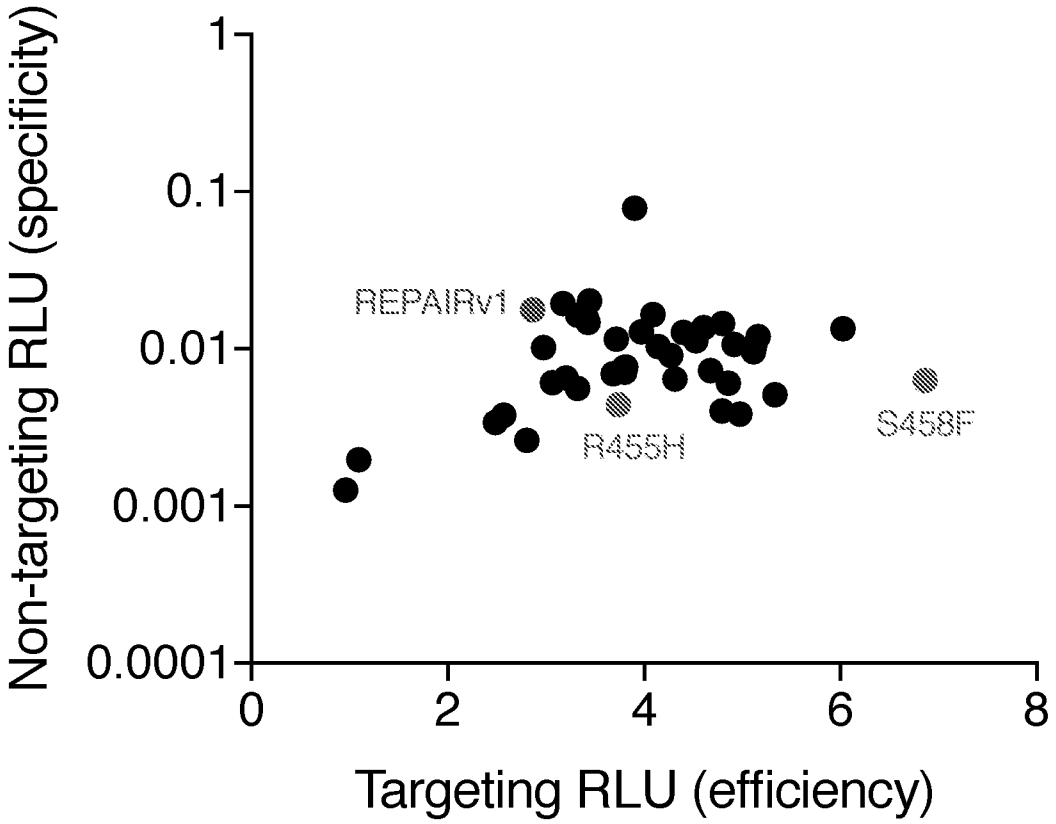

Non-targeting RLU (specificity)
1
0.1
0.01
0.001
0.0001
0
2
4
6
8
Targeting RLU (efficiency)
REPAIRv1
R455H
S458F

FIG. 77

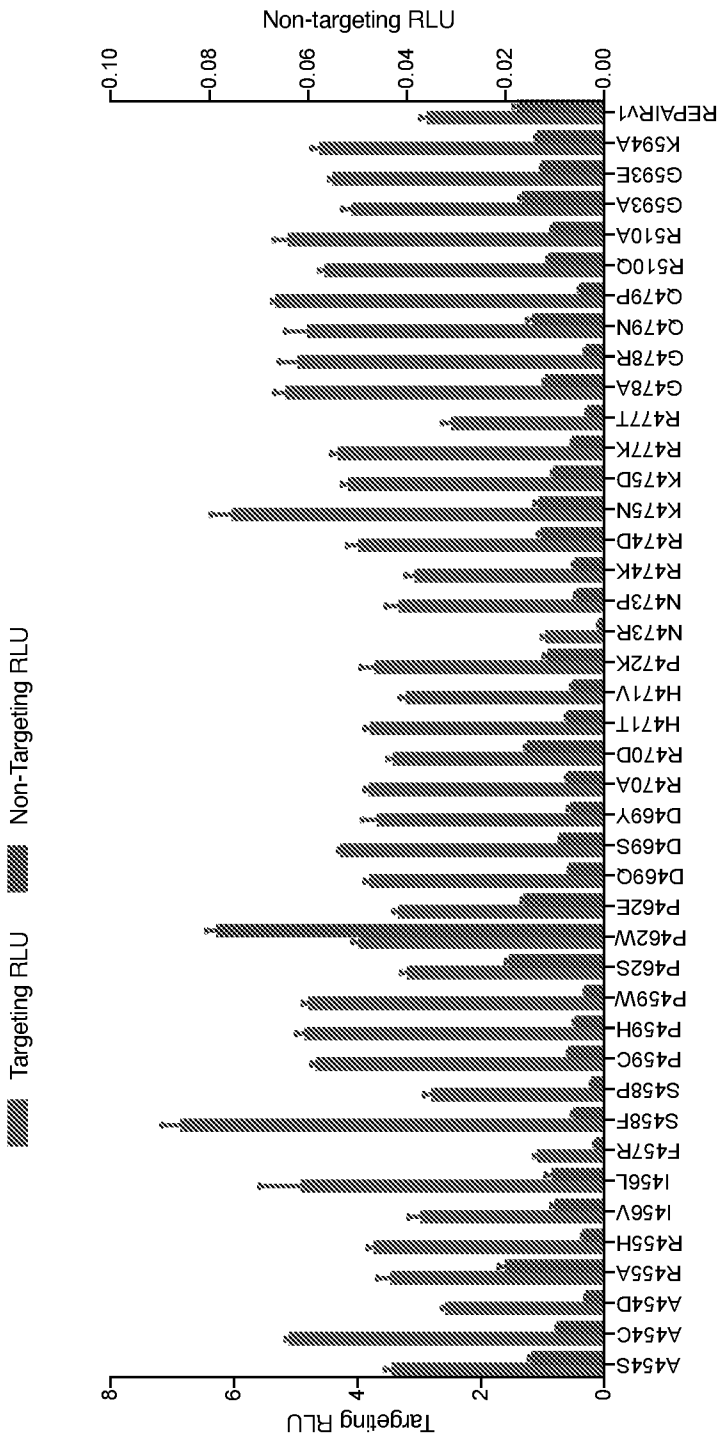

Non-targeting RLU
0.10
0.08
0.06
0.04
0.02
0.00
Targeting RLU
8
6
4
2
0
Targeting RLU
Non-Targeting RLU
ADAR Mutant
A454S
A454C
A454D
R455A
R455H
I456V
I456L
F457R
S458F
S458P
P459C
P459H
P459W
P462S
P462W
P462E
D469Q
D469S
D469Y
R470A
R470D
H471T
H471V
P472K
N473R
N473P
R474K
R474D
K475N
K475D
R477K
R477T
G478A
G478R
Q479N
Q479P
R510Q
R510A
G593A
G593E
K594A
REPAIRv1

FIG. 78

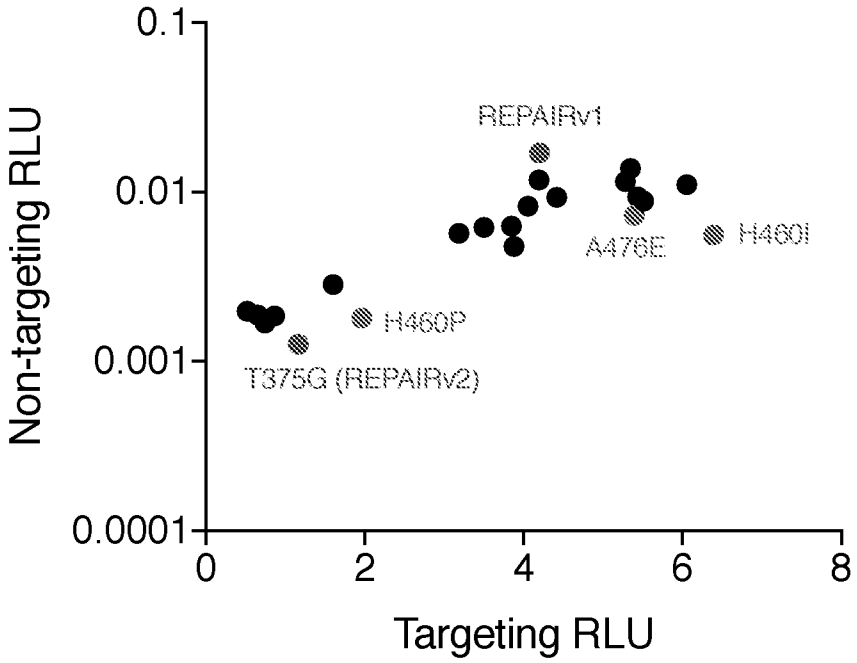

0.1
0.01
0.001
0.0001
Non-targeting RLU
0
2
4
6
8
Targeting RLU
REPAIRv1
A476E
H460I
H460P
T375G (REPAIRv2)

FIG. 79

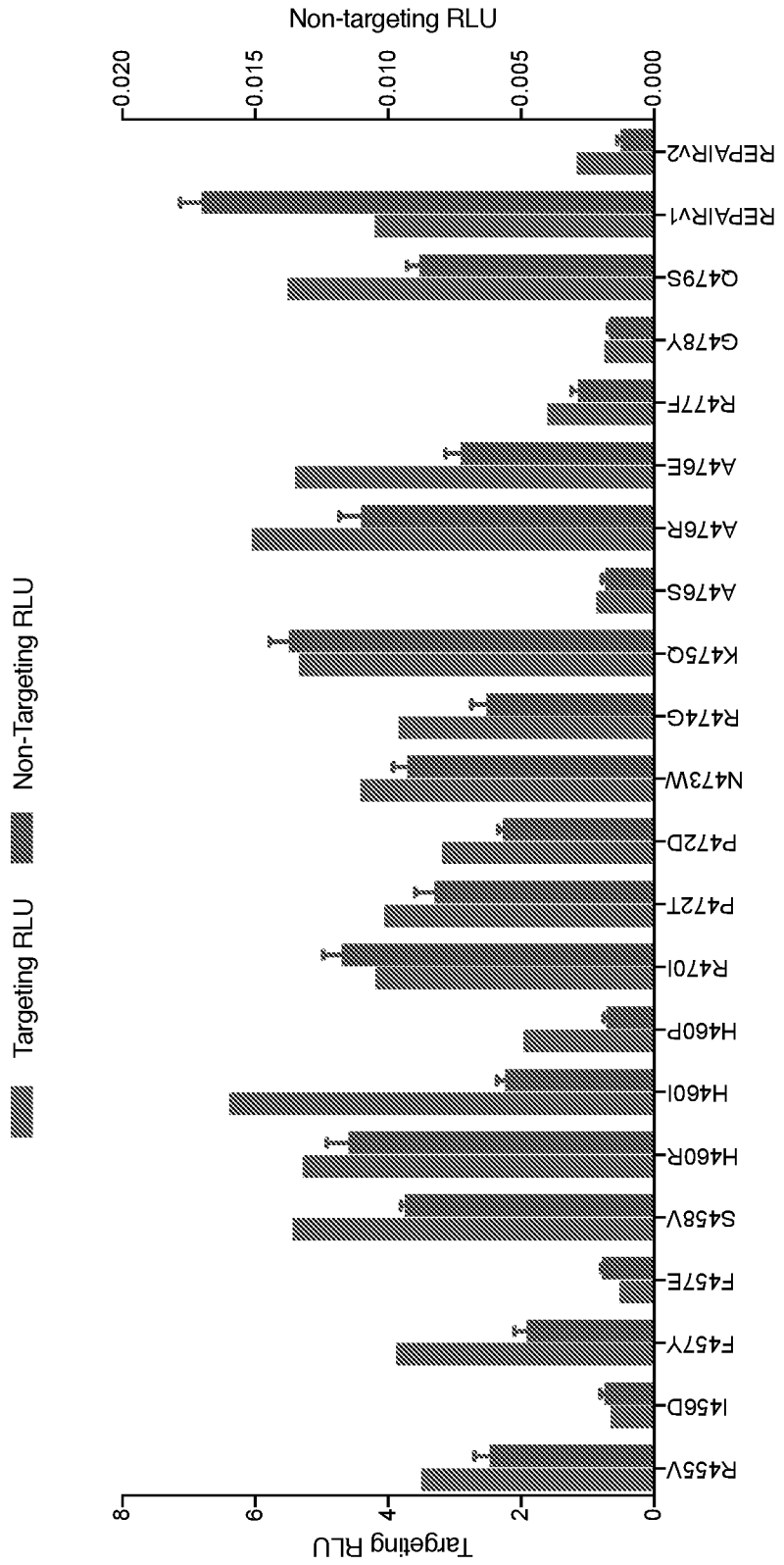

Non-targeting RLU
0.020
0.015
0.010
0.005
0.000
Targeting RLU
8
6
4
2
0
Targeting RLU
Non-Targeting RLU
R455V
I456D
F457Y
F457E
S458V
H460R
H460I
H460P
R470I
P472T
P472D
N473W
R474G
K475Q
A476S
A476R
A476E
R477F
G478Y
Q479S
REPAIRv1
REPAIRv2
ADAR Mutant

FIG. 80

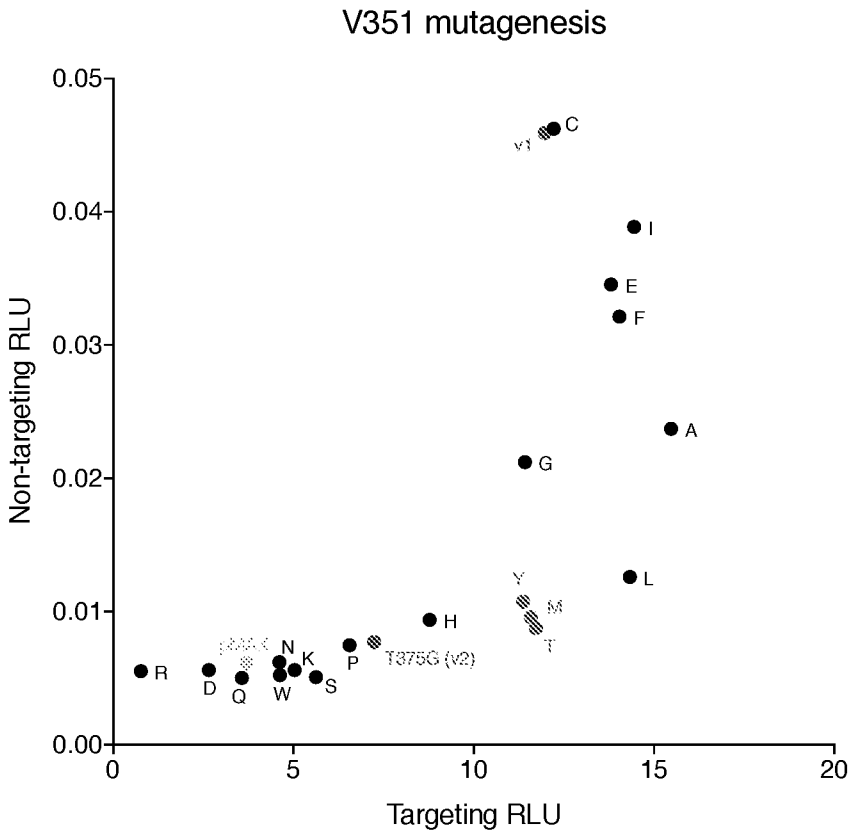

V351 mutagenesis
Non-targeting RLU
Targeting RLU
0.05
0.04
0.03
0.02
0.01
0.00
0
5
10
15
20
C
I
E
F
A
G
L
Y
M
T
H
N
K
P
T375G (v2)
R
D
Q
W
S

FIG. 81

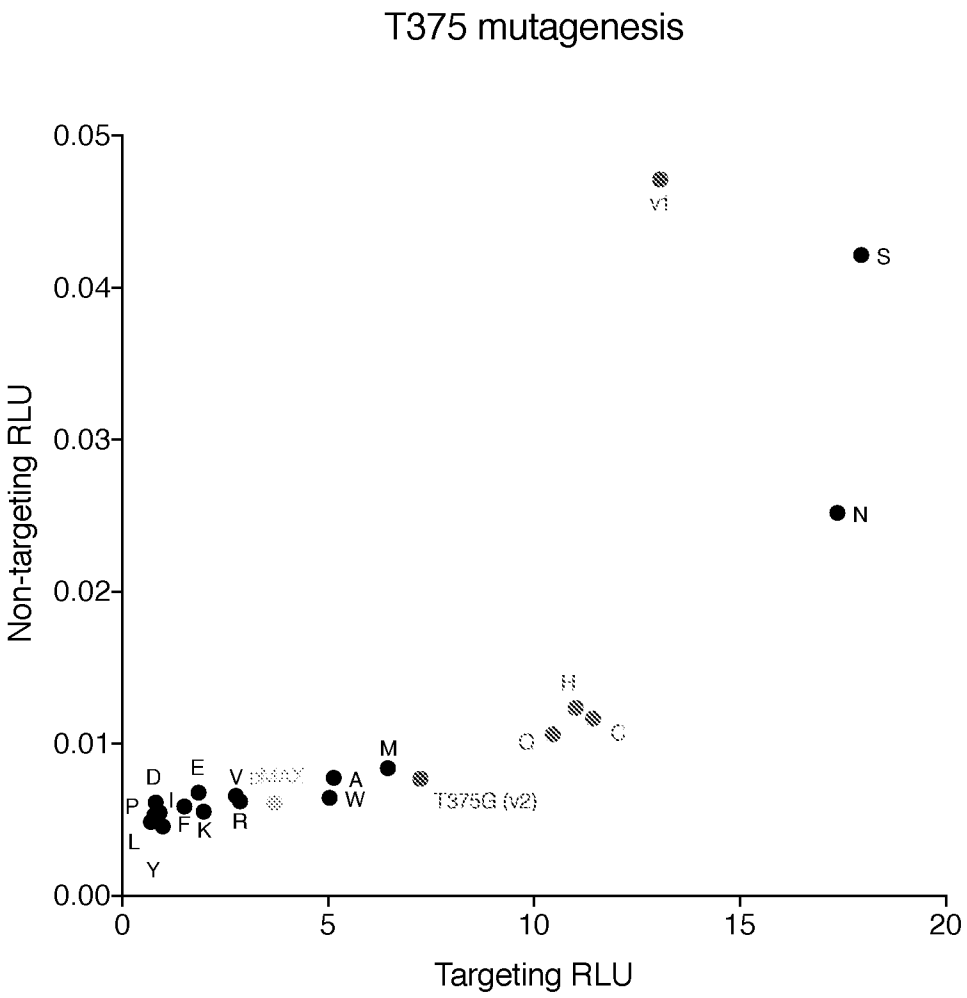

T375 mutagenesis
Non-targeting RLU
Targeting RLU
0.05
0.04
0.03
0.02
0.01
0.00
0
5
10
15
20
v1
S
N
H
Q
C
M
D
E
V
A
W
P
I
F
K
R
L
Y
T375G (v2)

FIG. 82

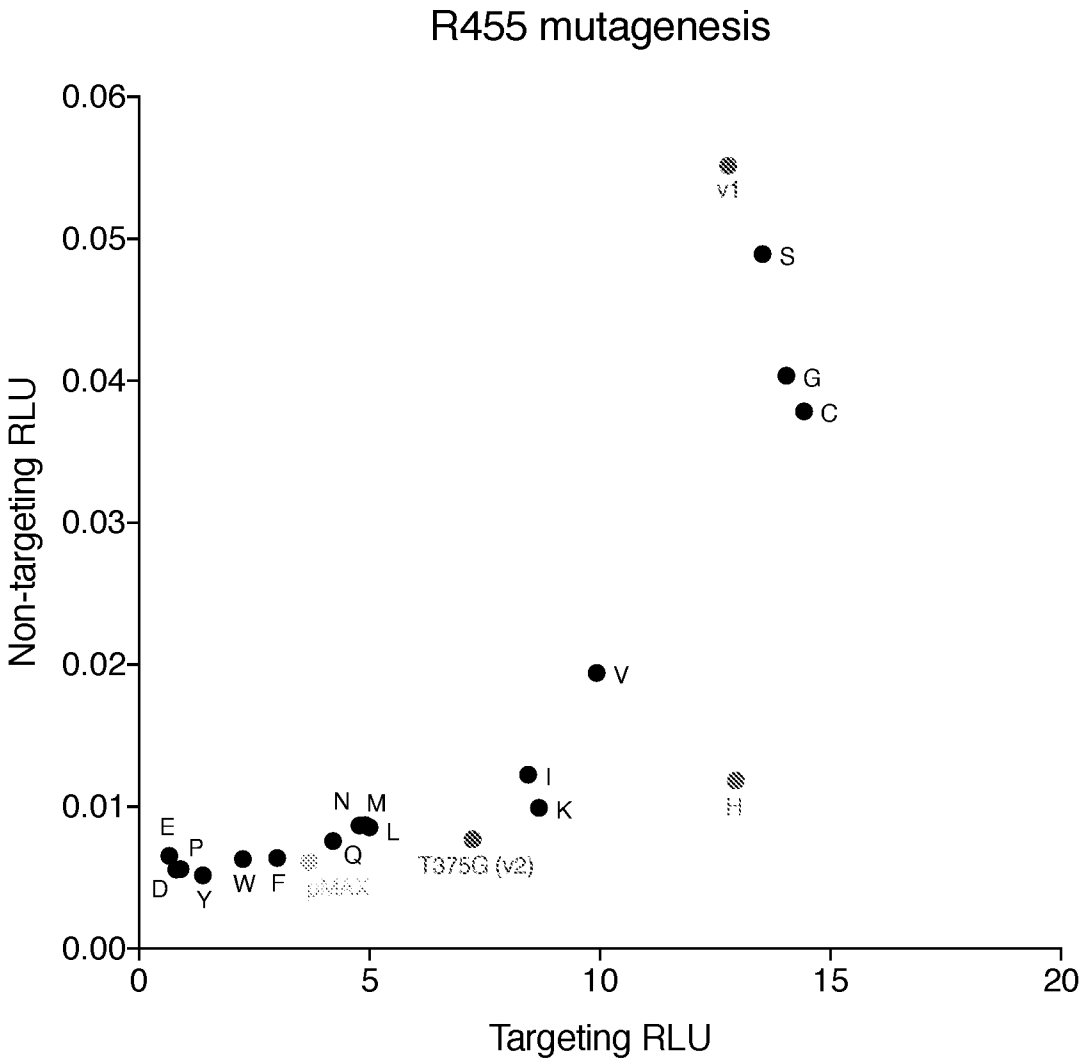

R455 mutagenesis
Non-targeting RLU
Targeting RLU
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0
5
10
15
20
v1
S
G
C
V
I
K
H
N
M
L
Q
E
P
D
Y
W
F
pMAX
T375G (v2)

FIG. 83

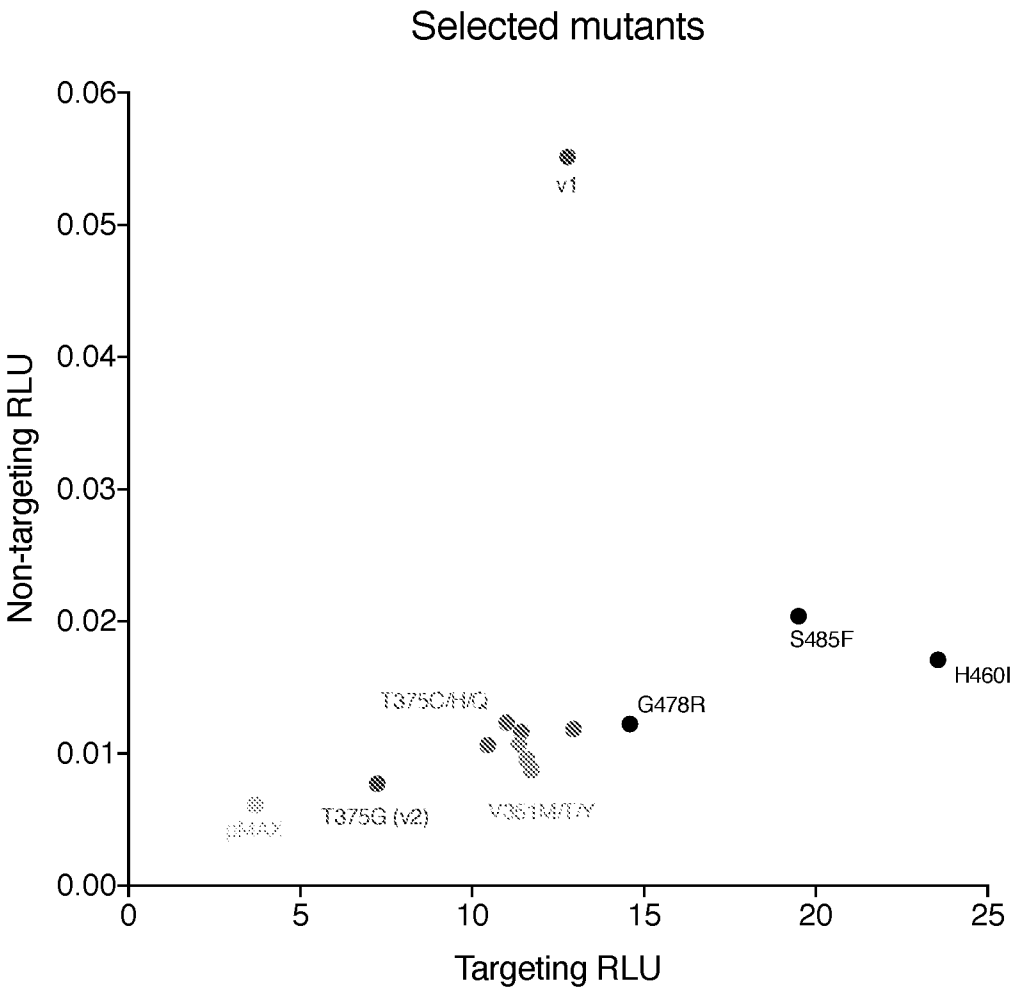

Selected mutants
Non-targeting RLU
Targeting RLU
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0
5
10
15
20
25
v1
S485F
H460I
G478R
T375C/H/Q
T375G (v2)
V351M/T/Y

FIG. 84

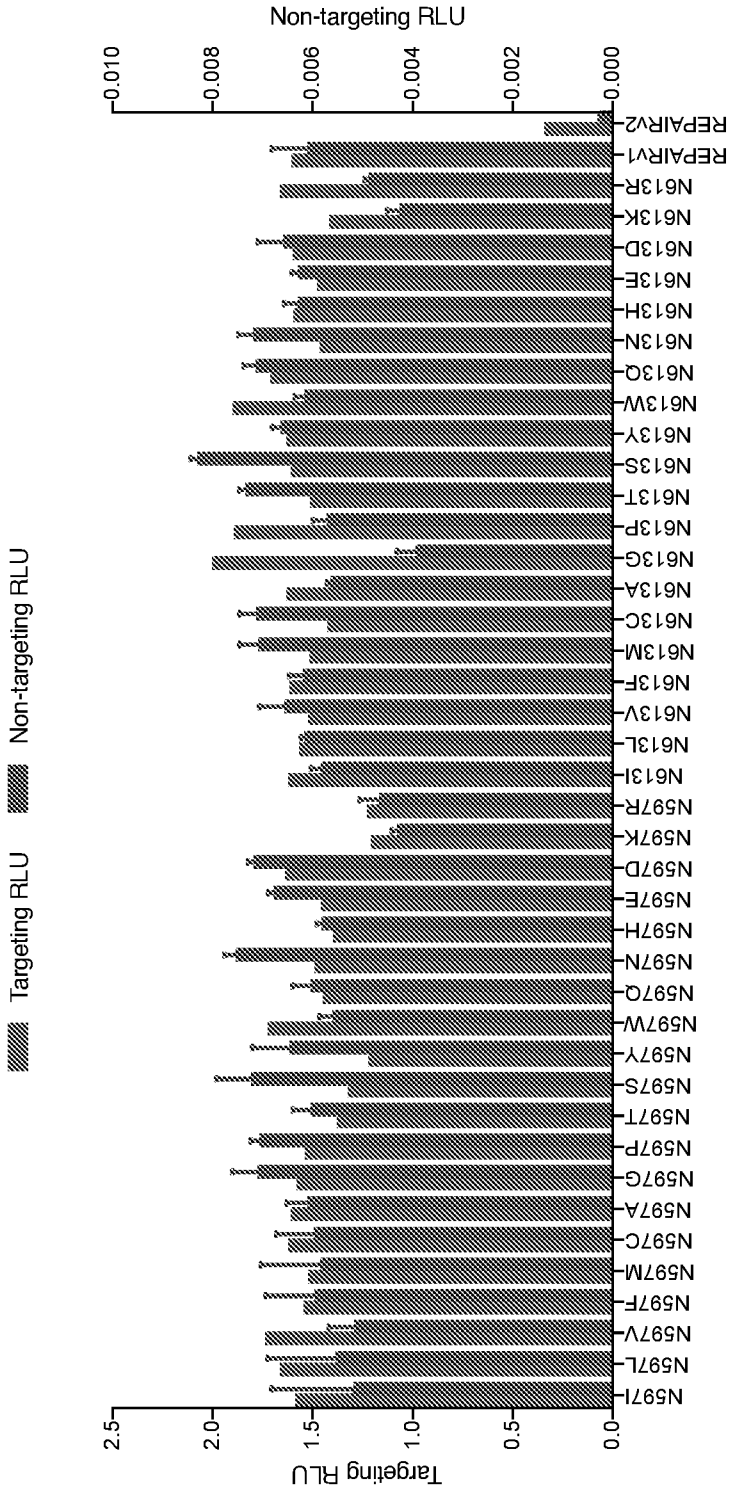

Non-targeting RLU
0.010
0.008
0.006
0.004
0.002
0.000
REPAIRv2
REPAIRv1
N613R
N613K
N613D
N613E
N613H
N613N
N613Q
N613W
N613Y
N613S
N613T
N613P
N613G
N613A
N613C
N613M
N613F
N613V
N613L
N613I
N597R
N597K
N597D
N597E
N597H
N597N
N597Q
N597W
N597Y
N597S
N597T
N597P
N597G
N597A
N597C
N597M
N597F
N597V
N597L
N597I
2.5
2.0
1.5
1.0
0.5
0.0
Targeting RLU
Targeting RLU
Non-targeting RLU
ADAR Mutant

FIG. 85

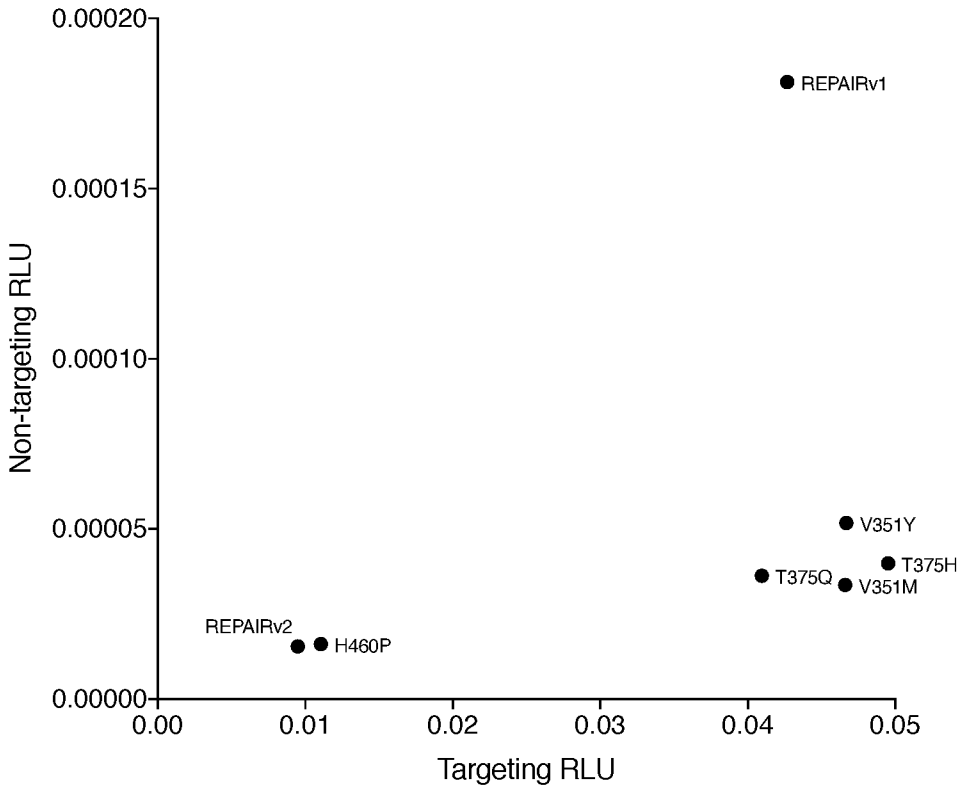

0.00020
0.00015
0.00010
0.00005
0.00000
Non-targeting RLU
0.00
0.01
0.02
0.03
0.04
0.05
Targeting RLU
REPAIRv1
V351Y
T375H
T375Q
V351M
REPAIRv2
H460P

FIG. 86

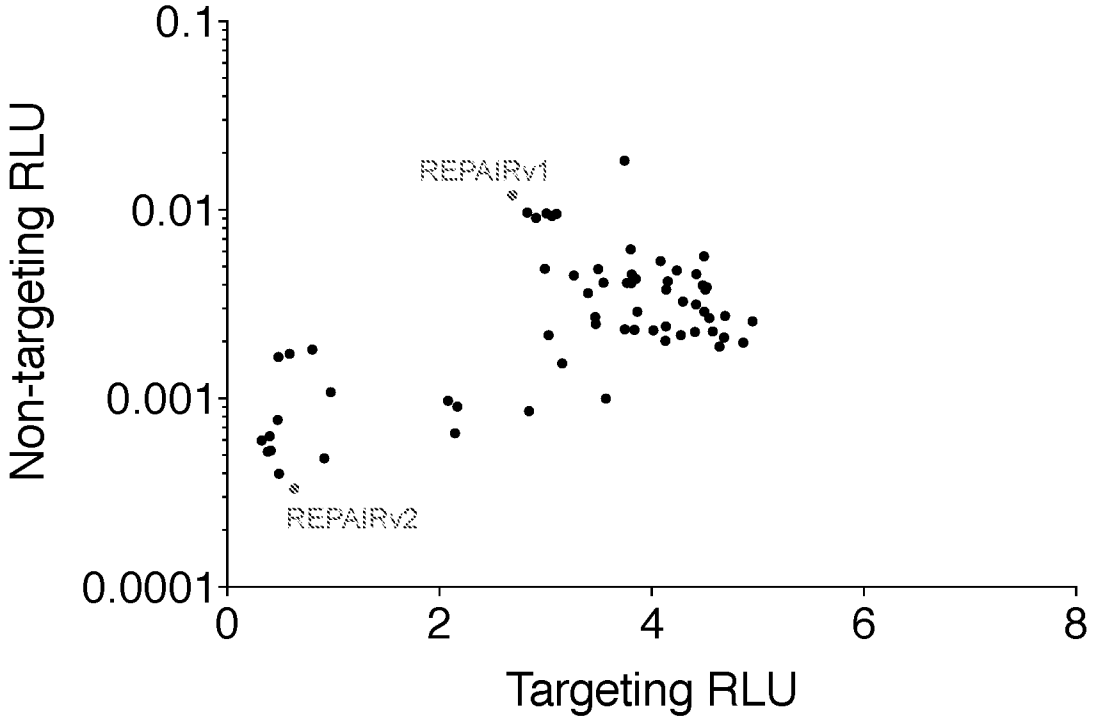

0.1
0.01
0.001
0.0001
Non-targeting RLU
REPAIRv1
REPAIRv2
0
2
4
6
8
Targeting RLU

FIG. 87

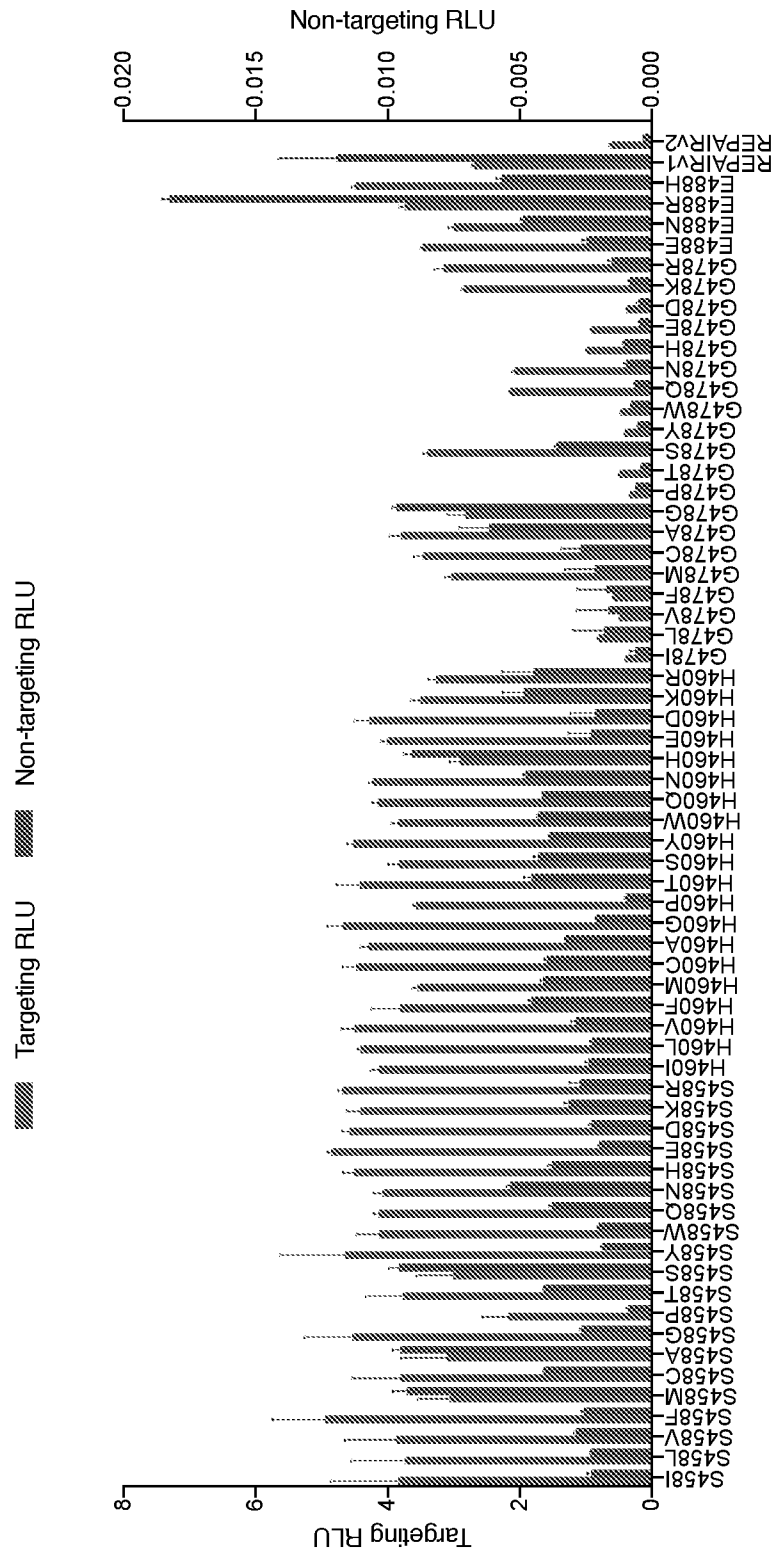

Non-targeting RLU
0.020
0.015
0.010
0.005
0.000
Targeting RLU
8
6
4
2
0
Targeting RLU
Non-targeting RLU
ADAR Mutant
REPAIRv2
REPAIRv1
E488H
E488R
E488N
E488E
G478R
G478K
G478D
G478E
G478H
G478N
G478Q
G478W
G478Y
G478S
G478T
G478P
G478G
G478A
G478C
G478M
G478F
G478V
G478L
G478I
H460R
H460K
H460D
H460E
H460H
H460N
H460Q
H460W
H460Y
H460S
H460T
H460P
H460G
H460A
H460C
H460M
H460F
H460V
H460L
H460I
S458R
S458K
S458D
S458E
S458H
S458N
S458Q
S458W
S458Y
S458S
S458T
S458P
S458G
S458A
S458C
S458M
S458F
S458V
S458L
S458I

FIG. 88

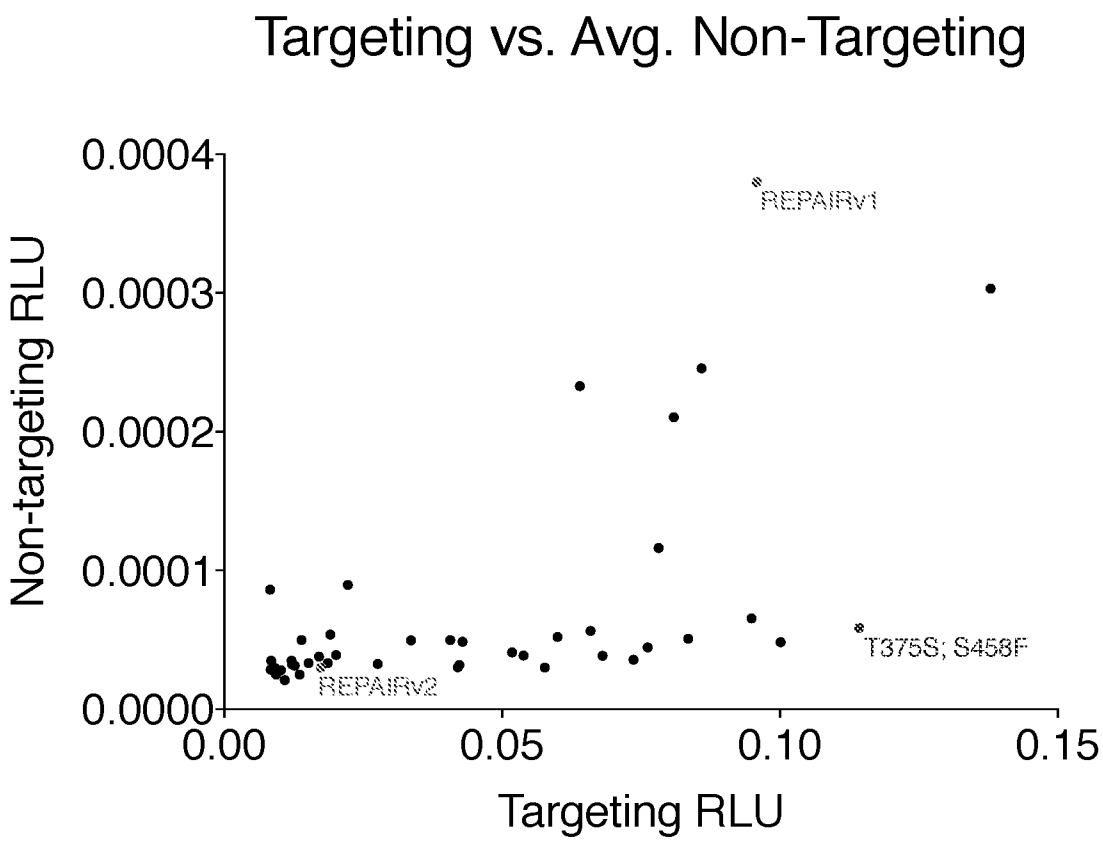

Targeting vs. Avg. Non-Targeting
Non-targeting RLU
0.0004
0.0003
0.0002
0.0001
0.0000
0.00
0.05
0.10
0.15
Targeting RLU
REPAIRv1
T375S; S458F
REPAIRv2

FIG. 89

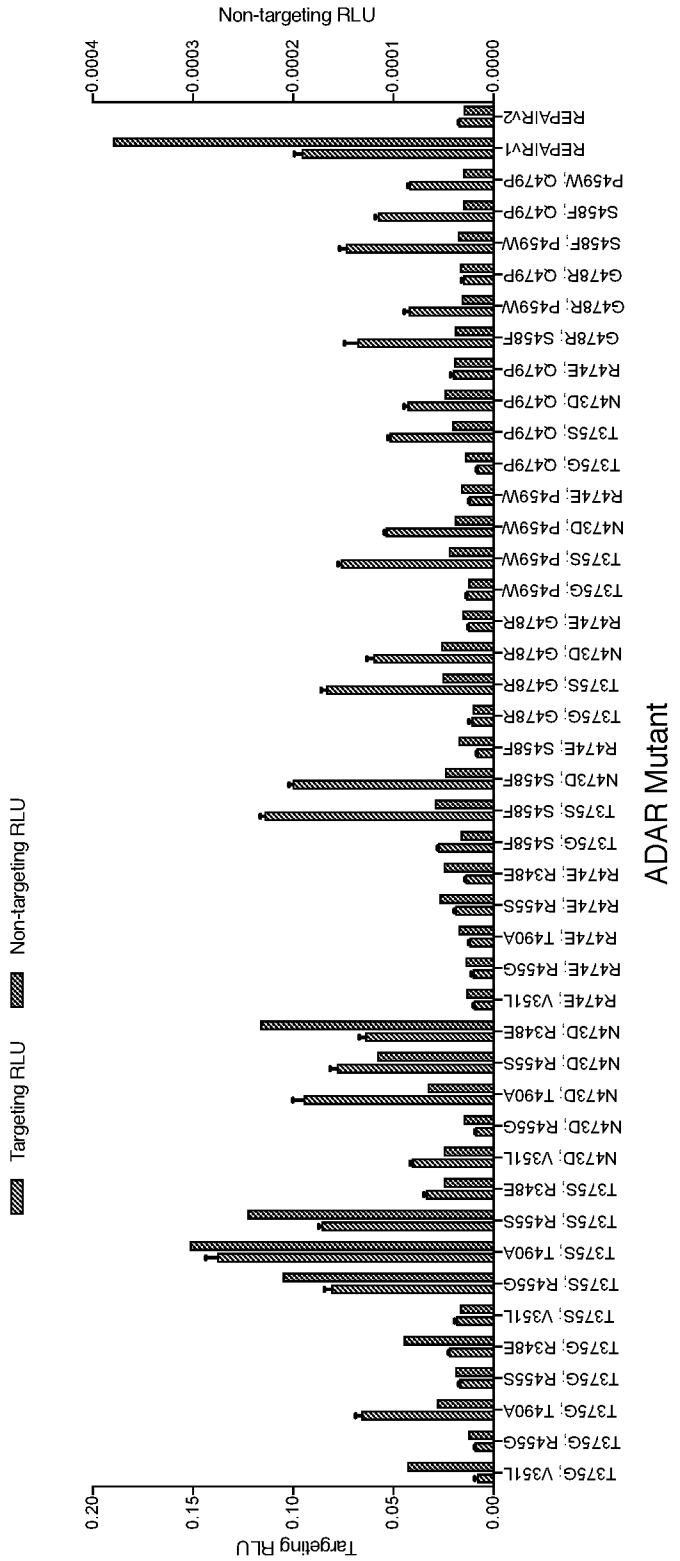

Non-targeting RLU
0.0004
0.0003
0.0002
0.0001
0.0000
REPAIRv2
REPAIRv1
P459W, Q479P
S458F, Q479P
S458F, P459W
G478R, Q479P
G478R, P459W
G478R, S458F
R474E, Q479P
N473D, Q479P
T375S, Q479P
T375G, Q479P
R474E, P459W
N473D, P459W
T375S, P459W
T375G, P459W
R474E, G478R
N473D, G478R
T375S, G478R
T375G, G478R
R474E, S458F
N473D, S458F
T375S, S458F
T375G, S458F
R474E, R348E
R474E, R455S
R474E, T490A
R474E, R455G
R474E, V351L
N473D, R348E
N473D, R455S
N473D, T490A
N473D, R455G
N473D, V351L
T375S, R348E
T375S, R455S
T375S, T490A
T375S, R455G
T375S, V351L
T375G, R348E
T375G, R455S
T375G, T490A
T375G, R455G
T375G, V351L
0.20
0.15
0.10
0.05
0.00
Targeting RLU
Targeting RLU
Non-targeting RLU
ADAR Mutant

FIG. 90

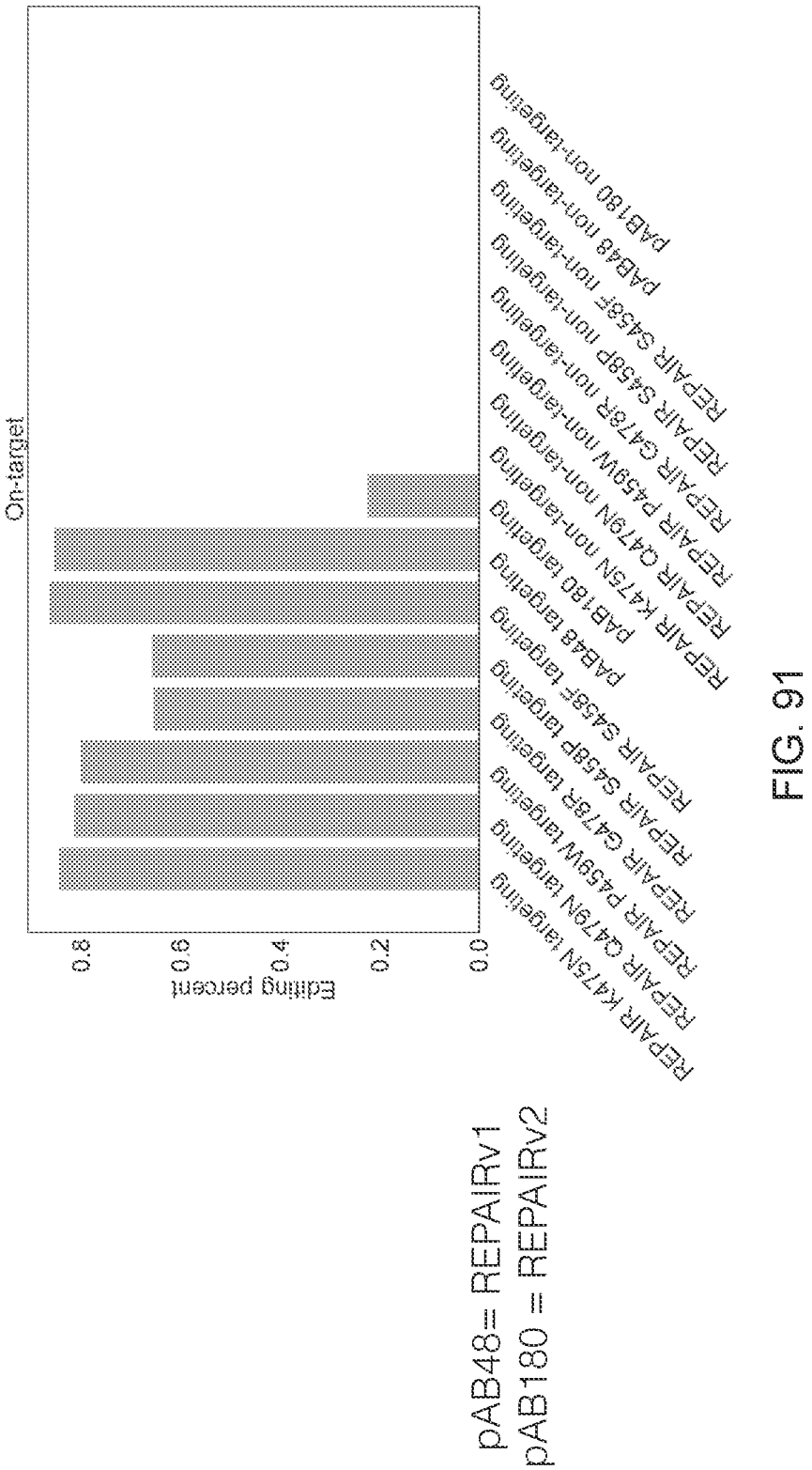


FIG. 91

pAB48= REPAIRv1
pAB180 = REPAIRv2

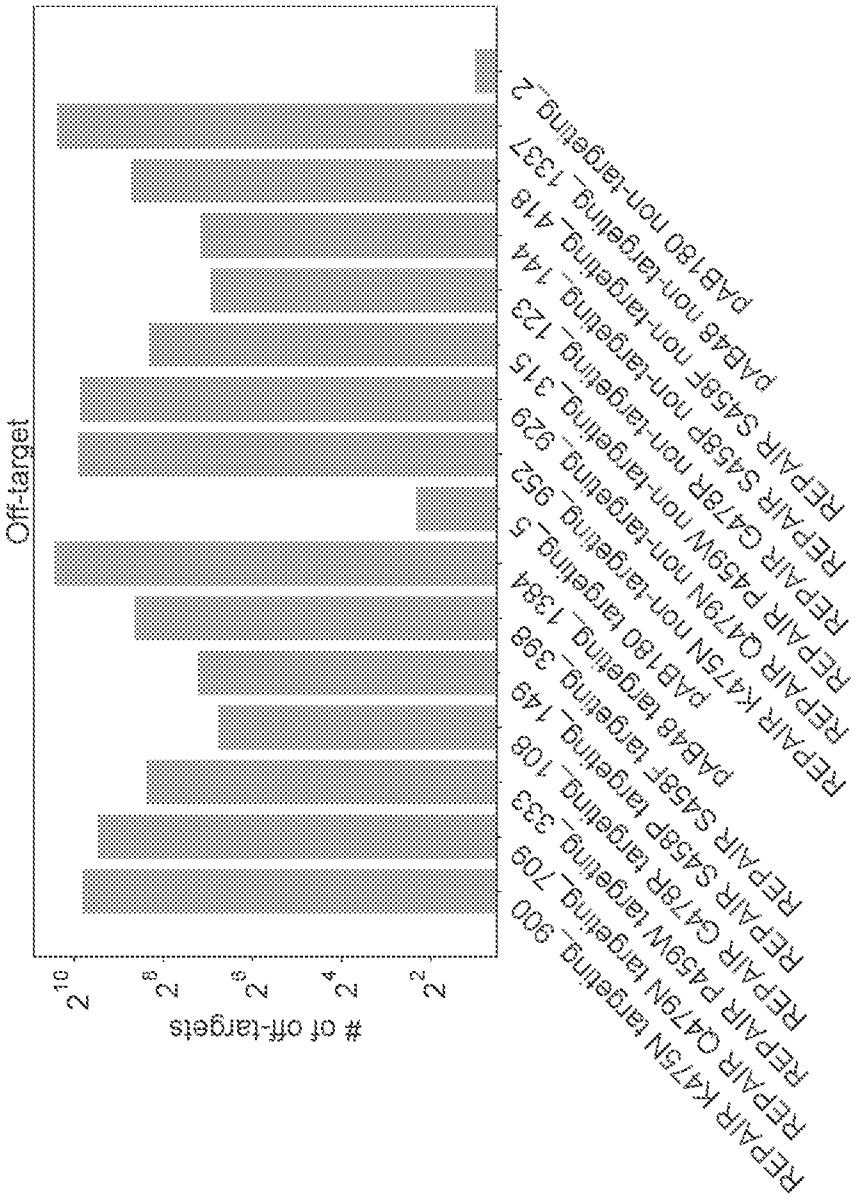

Off-target
of off-targets
2^{10}
2^{8}
2^{6}
2^{4}
2^{2}
REPAIR K475N targeting_900
REPAIR Q479N targeting_709
REPAIR P459W targeting_333
REPAIR Q478R targeting_108
REPAIR S458P targeting_149
pAB48 targeting_398
pAB180 targeting_1384
REPAIR K475N non-targeting_5
REPAIR Q479N non-targeting_952
REPAIR P459W non-targeting_629
REPAIR Q478R non-targeting_315
REPAIR S458P non-targeting_123
REPAIR S458E non-targeting_144
pAB48 non-targeting_418
pAB180 non-targeting_1337
REPAIR S458E non-targeting_2

FIG. 92

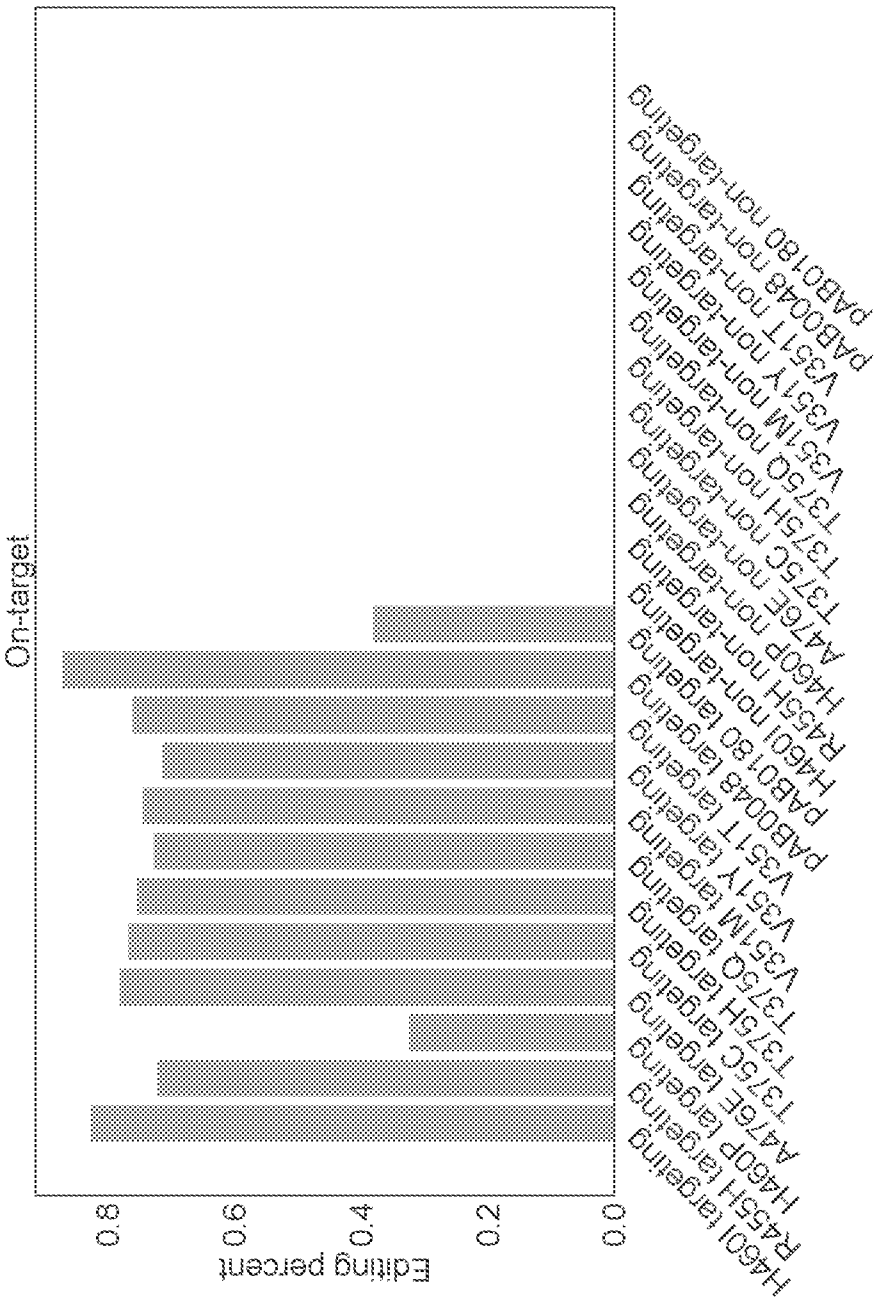

On-target
Editing percent
0.0
0.2
0.4
0.6
0.8
H460I targeting
R455H targeting
H460P targeting
A476E targeting
T375C targeting
T375H targeting
V351O targeting
V351M targeting
V351Y targeting
pAB0051 targeting
pAB0048 targeting
H460I non-targeting
R455H non-targeting
H460P non-targeting
A476E non-targeting
T375C non-targeting
T375H non-targeting
V351O non-targeting
V351M non-targeting
V351Y non-targeting
pAB0051 non-targeting
pAB0048 non-targeting
pAB0180 non-targeting

FIG. 93

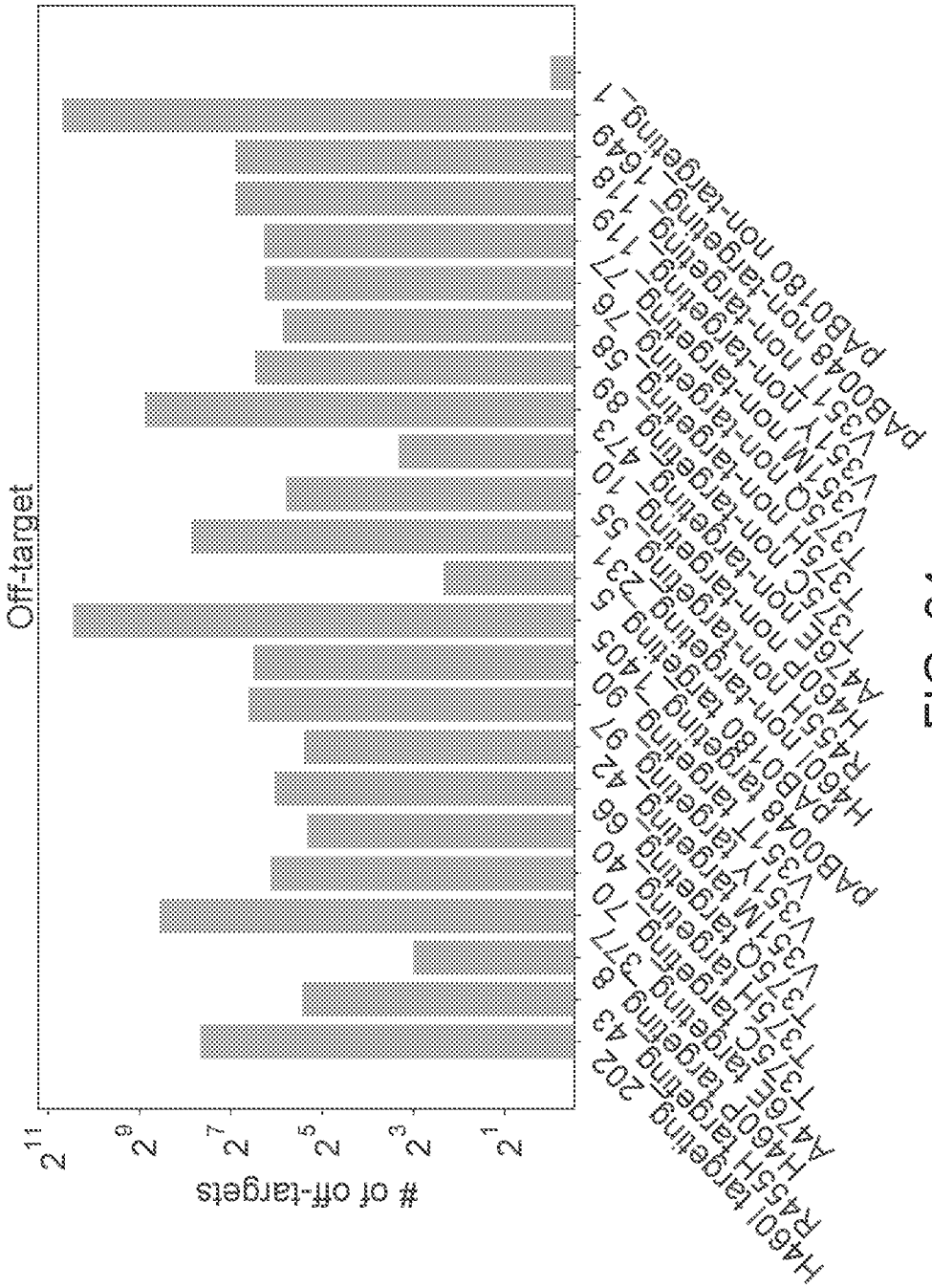

Off-target
of off-targets
2^11
2^9
2^7
2^5
2^3
2^1
H46Or targeting 202 A3 G77 targeting 37T Q0 A0 G6 A2 G7 Q0 targeting 405 S3 targeting 23A S5 targeting 10 A73 G8 G8 targeting 17G
R45SH targeting 37S Q targeting V35M targeting V35Y targeting 37T targeting 405 targeting 180 targeting 405 S3 targeting 23A S5 targeting 10 A73 G8 G8 targeting 17G
A476P targeting 37S Q targeting V35Y targeting 37T targeting 405 targeting 180 targeting 405 S3 targeting 23A S5 targeting 10 A73 G8 G8 targeting 17G
pAB004 targeting 37T targeting 405 targeting 180 targeting 405 S3 targeting 23A S5 targeting 10 A73 G8 G8 targeting 17G
H46Or R45SH targeting 37S Q targeting V35M targeting V35Y targeting 37T targeting 405 targeting 180 targeting 405 S3 targeting 23A S5 targeting 10 A73 G8 G8 targeting 17G
R45SH A476P targeting 37S Q targeting V35Y targeting 37T targeting 405 targeting 180 targeting 405 S3 targeting 23A S5 targeting 10 A73 G8 G8 targeting 17G
A476P T37SC targeting 37S Q targeting V35Y targeting 37T targeting 405 targeting 180 targeting 405 S3 targeting 23A S5 targeting 10 A73 G8 G8 targeting 17G
T37SC T37SH targeting 37S Q targeting V35Y targeting 37T targeting 405 targeting 180 targeting 405 S3 targeting 23A S5 targeting 10 A73 G8 G8 targeting 17G
T37SH T37SQ targeting 37S Q targeting V35Y targeting 37T targeting 405 targeting 180 targeting 405 S3 targeting 23A S5 targeting 10 A73 G8 G8 targeting 17G
T37SQ V35M targeting 37S Q targeting V35Y targeting 37T targeting 405 targeting 180 targeting 405 S3 targeting 23A S5 targeting 10 A73 G8 G8 targeting 17G
V35M V35Y targeting 37S Q targeting V35Y targeting 37T targeting 405 targeting 180 targeting 405 S3 targeting 23A S5 targeting 10 A73 G8 G8 targeting 17G
V35Y pAB004 targeting 37S Q targeting V35Y targeting 37T targeting 405 targeting 180 targeting 405 S3 targeting 23A S5 targeting 10 A73 G8 G8 targeting 17G
pAB004 H46Or targeting 37S Q targeting V35Y targeting 37T targeting 405 targeting 180 targeting 405 S3 targeting 23A S5 targeting 10 A73 G8 G8 targeting 17G
H46Or R45SH targeting 37S Q targeting V35Y targeting 37T targeting 405 targeting 180 targeting 405 S3 targeting 23A S5 targeting 10 A73 G8 G8 targeting 17G
R45SH A476P targeting 37S Q targeting V35Y targeting 37T targeting 405 targeting 180 targeting 405 S3 targeting 23A S5 targeting 10 A73 G8 G8 targeting 17G
A476P T37SC targeting 37S Q targeting V35Y targeting 37T targeting 405 targeting 180 targeting 405 S3 targeting 23A S5 targeting 10 A73 G8 G8 targeting 17G
T37SC T37SH targeting 37S Q targeting V35Y targeting 37T targeting 405 targeting 180 targeting 405 S3 targeting 23A S5 targeting 10 A73 G8 G8 targeting 17G
T37SH T37SQ targeting 37S Q targeting V35Y targeting 37T targeting 405 targeting 180 targeting 405 S3 targeting 23A S5 targeting 10 A73 G8 G8 targeting 17G
T37SQ V35M targeting 37S Q targeting V35Y targeting 37T targeting 405 targeting 180 targeting 405 S3 targeting 23A S5 targeting 10 A73 G8 G8 targeting 17G
V35M V35Y targeting 37S Q targeting V35Y targeting 37T targeting 405 targeting 180 targeting 405 S3 targeting 23A S5 targeting 10 A73 G8 G8 targeting 17G
V35Y pAB004 targeting 37S Q targeting V35Y targeting 37T targeting 405 targeting 180 targeting 405 S3 targeting 23A S5 targeting 10 A73 G8 G8 targeting 17G
pAB004 non-targeting_1

FIG. 94

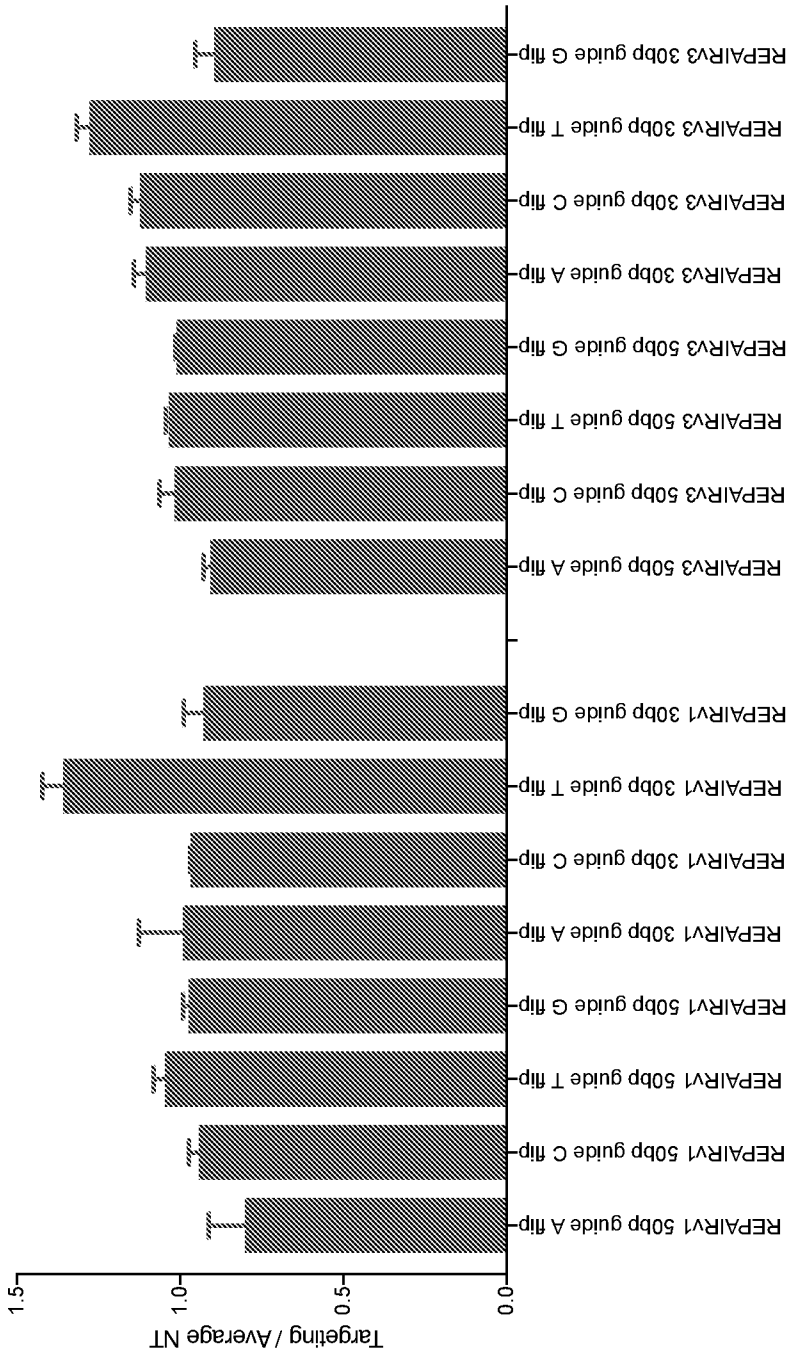

Targeting / Average NT
0.0
0.5
1.0
1.5
REPAIRv1 50bp guide A flip
REPAIRv1 50bp guide C flip
REPAIRv1 50bp guide T flip
REPAIRv1 50bp guide G flip
REPAIRv1 30bp guide A flip
REPAIRv1 30bp guide C flip
REPAIRv1 30bp guide T flip
REPAIRv1 30bp guide G flip
REPAIRv3 50bp guide A flip
REPAIRv3 50bp guide C flip
REPAIRv3 50bp guide T flip
REPAIRv3 50bp guide G flip
REPAIRv3 30bp guide A flip
REPAIRv3 30bp guide C flip
REPAIRv3 30bp guide T flip
REPAIRv3 30bp guide G flip
Construct-Guide Pairing

FIG. 95

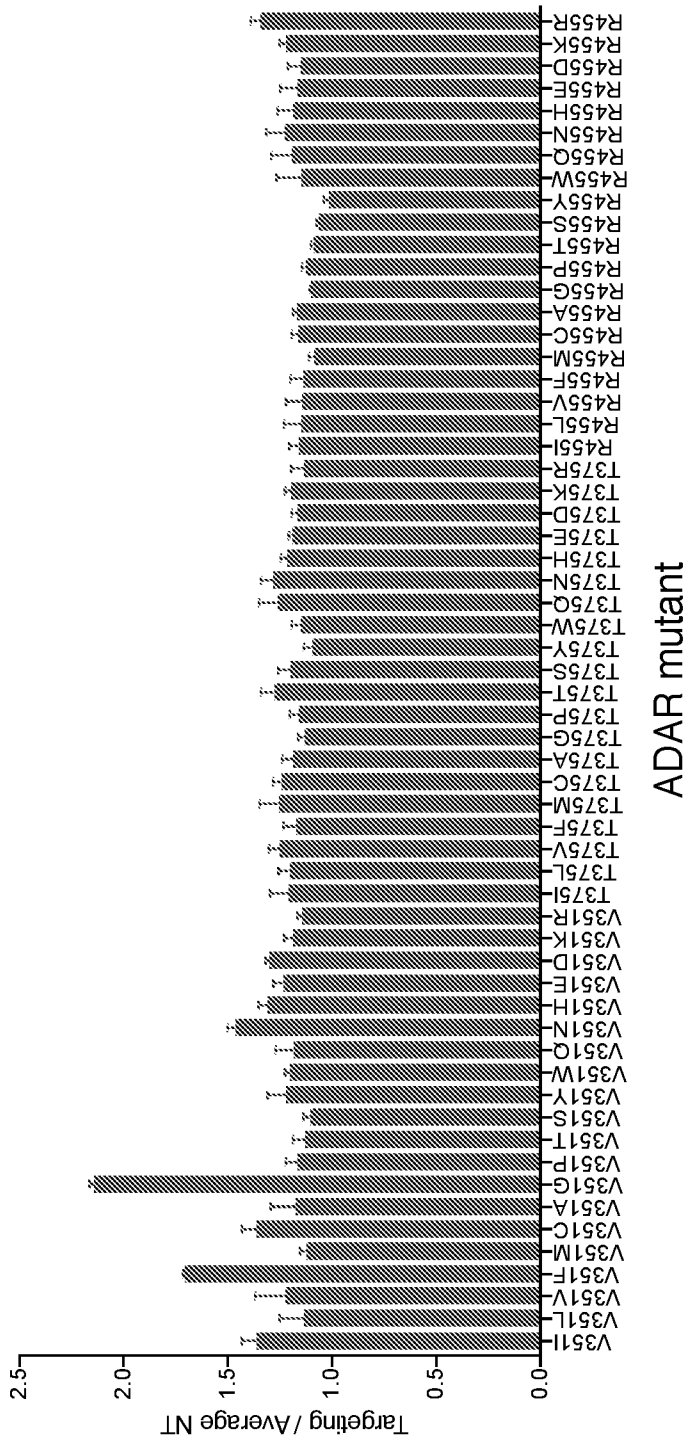

Targeting / Average NT
2.5
2.0
1.5
1.0
0.5
0.0
V351I
V351L
V351V
V351F
V351M
V351C
V351A
V351G
V351P
V351T
V351S
V351Y
V351W
V351Q
V351N
V351H
V351E
V351D
V351K
V351R
T375I
T375L
T375V
T375F
T375M
T375C
T375A
T375G
T375P
T375T
T375S
T375Y
T375W
T375Q
T375N
T375H
T375E
T375D
T375K
T375R
R455I
R455L
R455V
R455F
R455M
R455C
R455A
R455G
R455P
R455T
R455S
R455Y
R455W
R455Q
R455N
R455H
R455E
R455D
R455K
R455R
ADAR mutant

FIG. 96

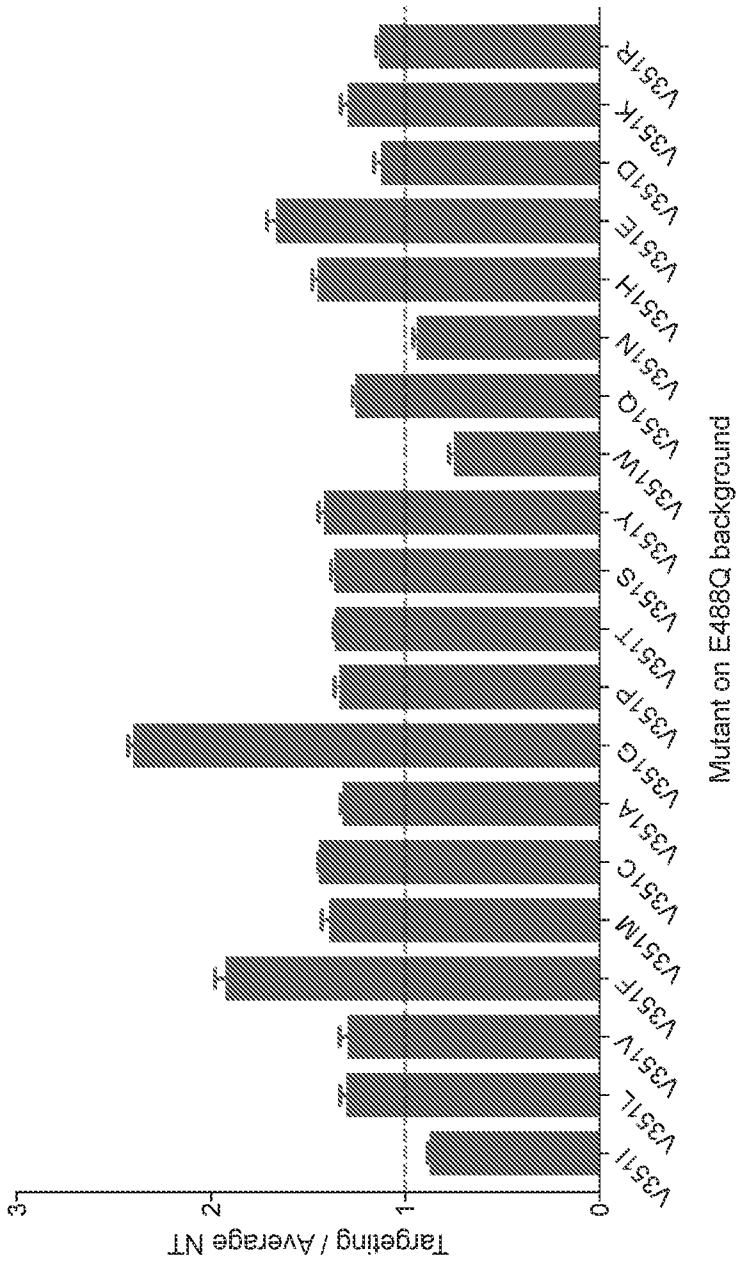

Targeting / Average NT
0
1
2
3
V351I
V351L
V351V
V351F
V351M
V351C
V351A
V351G
V351P
V351T
V351S
V351Y
V351W
V351Q
V351N
V351H
V351E
V351D
V351K
V351R
Mutant on E488Q background

FIG. 97

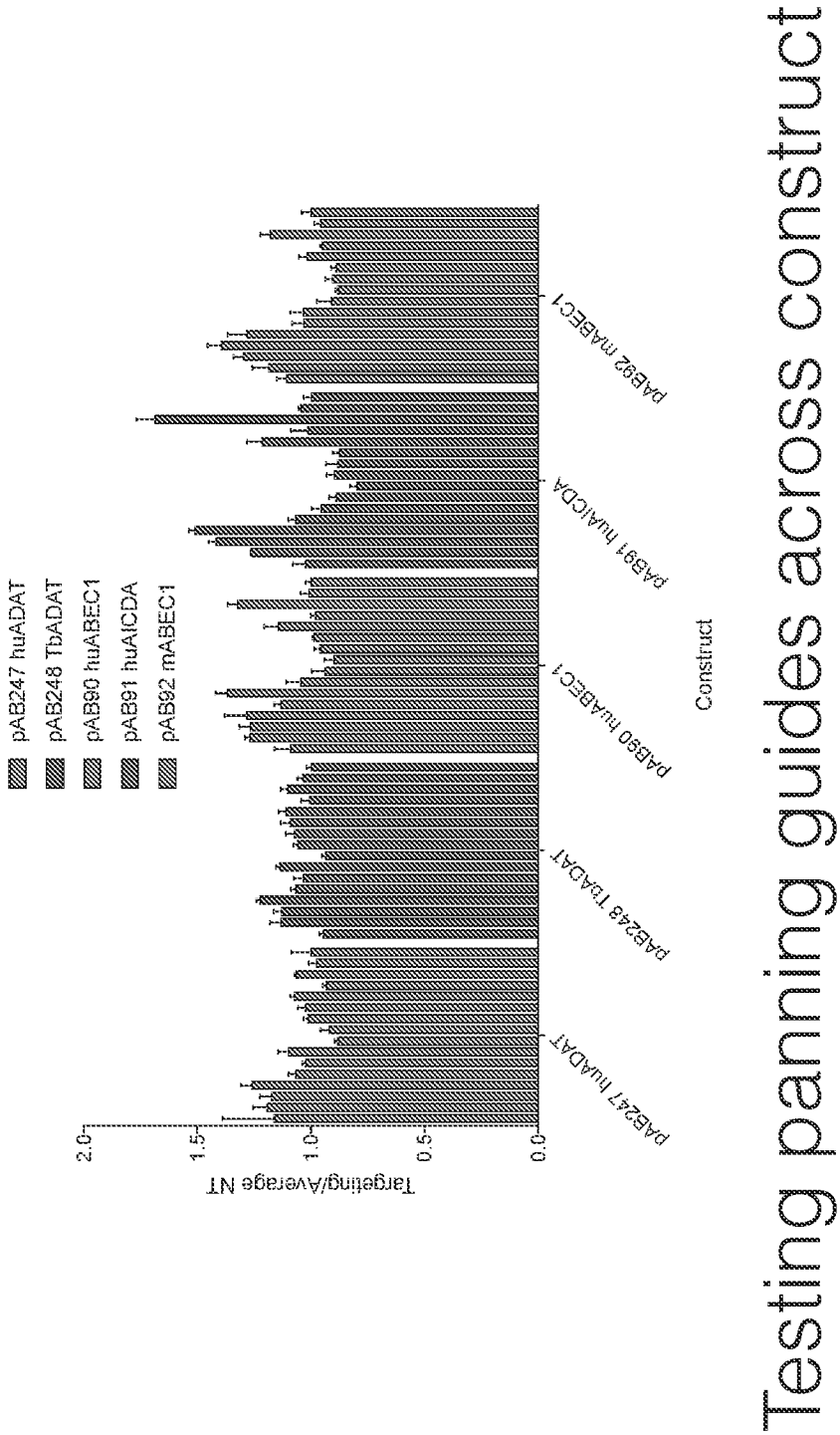

Testing panning guides across construct
pAB247 huADAT
pAB248 TbADAT
pAB90 huABEC1
pAB91 huAICDA
pAB92 mABEC1
Targeting/Average NT
2.0
1.5
1.0
0.5
0.0
pAB247 huADAT
pAB248 TbADAT
pAB90 huABEC1
pAB91 huAICDA
pAB92 mABEC1
Construct

FIG. 98

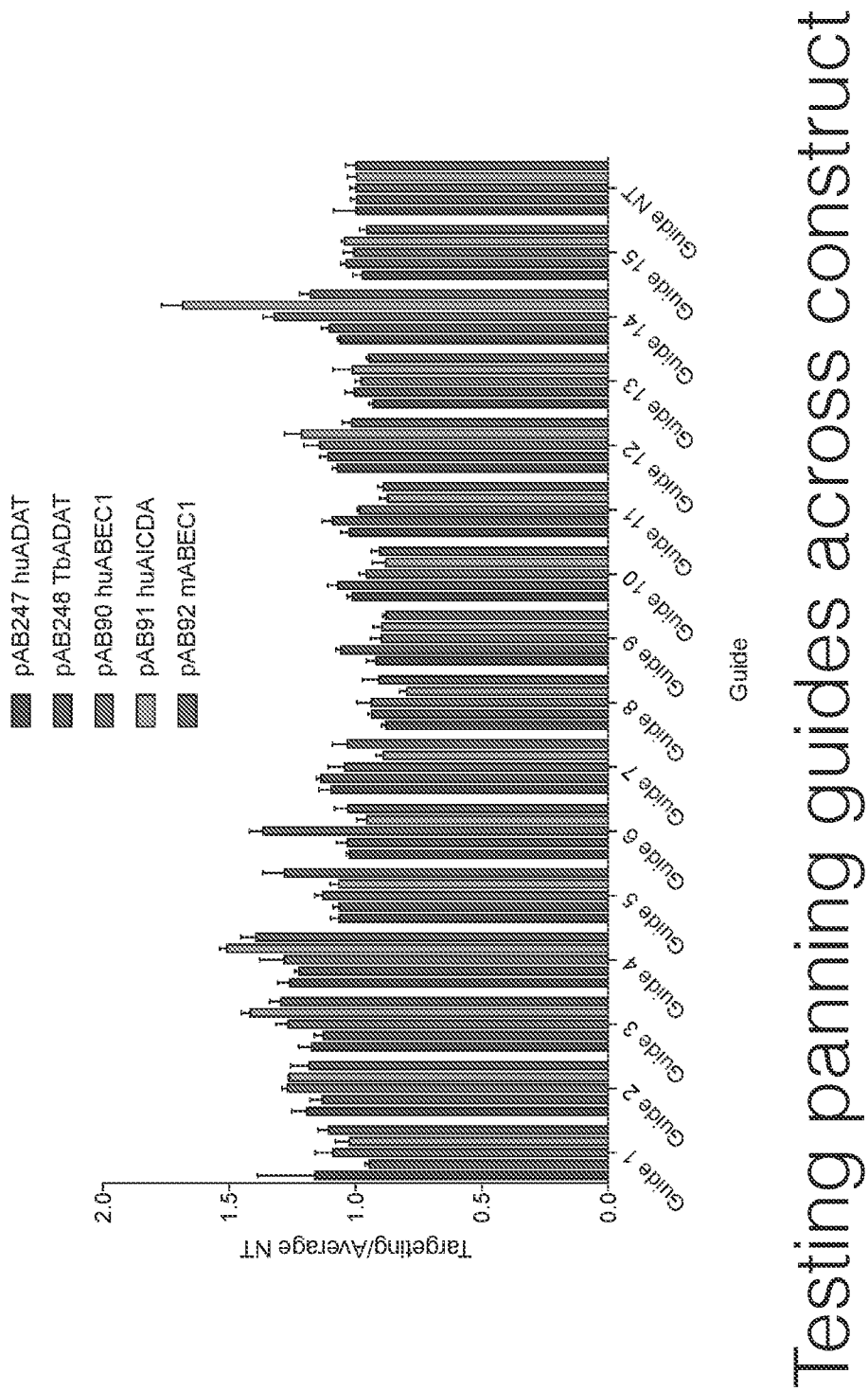

Testing panning guides across construct
pAB247 huADAT
pAB248 TbADAT
pAB90 huABEC1
pAB91 huAICDA
pAB92 mABEC1
Targeting/Average NT
2.0
1.5
1.0
0.5
0.0
Guide 1
Guide 2
Guide 3
Guide 4
Guide 5
Guide 6
Guide 7
Guide 8
Guide 9
Guide 10
Guide 11
Guide 12
Guide 13
Guide 14
Guide 15
Guide NT
Guide

FIG. 99

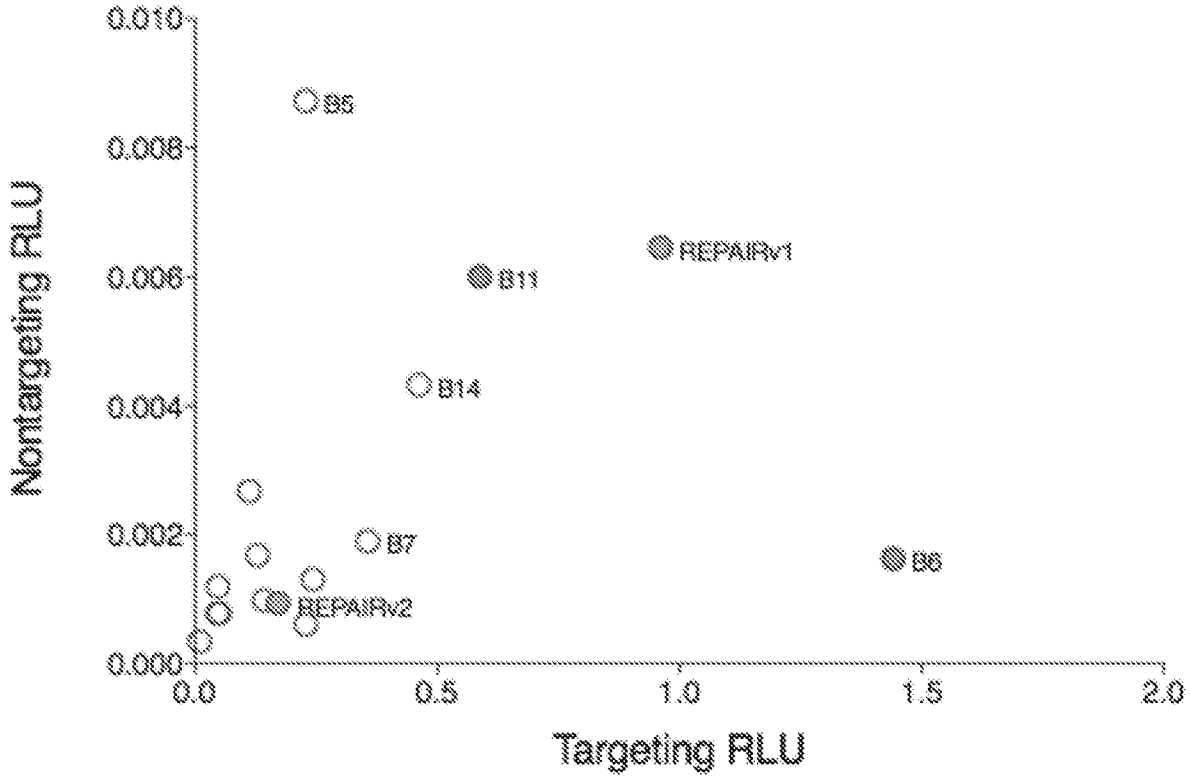

0.010
0.008
0.006
0.004
0.002
0.000
Nontargeting RLU
B5
REPAIRv1
B11
B14
B7
B6
REPAIRv2
0.0
0.5
1.0
1.5
2.0
Targeting RLU

FIG. 100

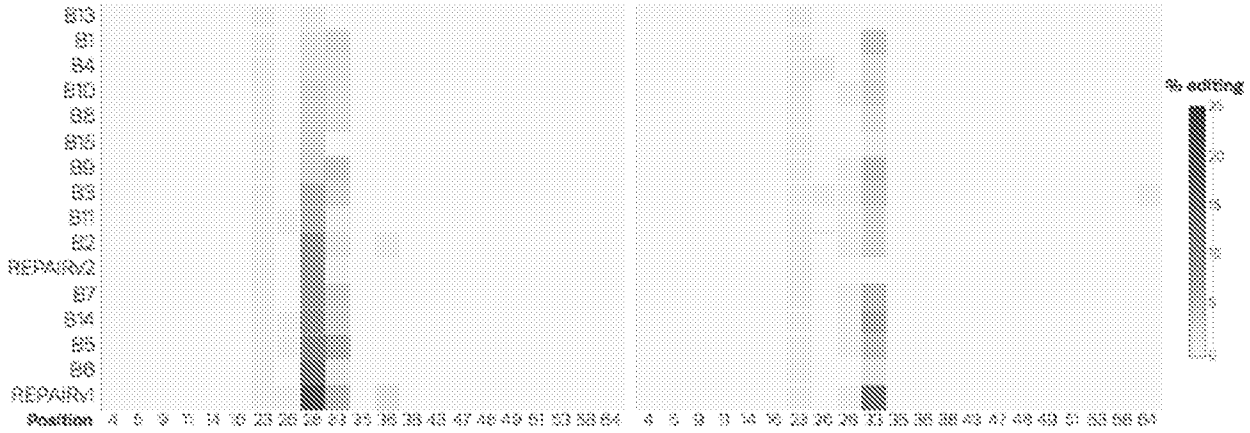

B13
B1
B4
B10
B8
B15
B9
B3
B11
B2
REPAIRv2
B7
B14
B5
B6
REPAIRv1
Position
% editing

FIG. 101

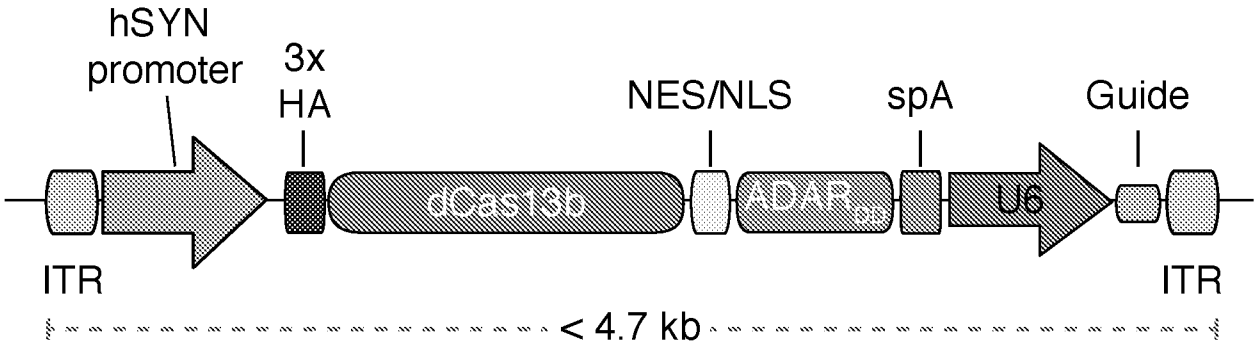

hSYN
promoter
3x
HA
NES/NLS
spA
Guide
dCas13b
ADAR
U6
ITR
ITR
< 4.7 kb

FIG. 102

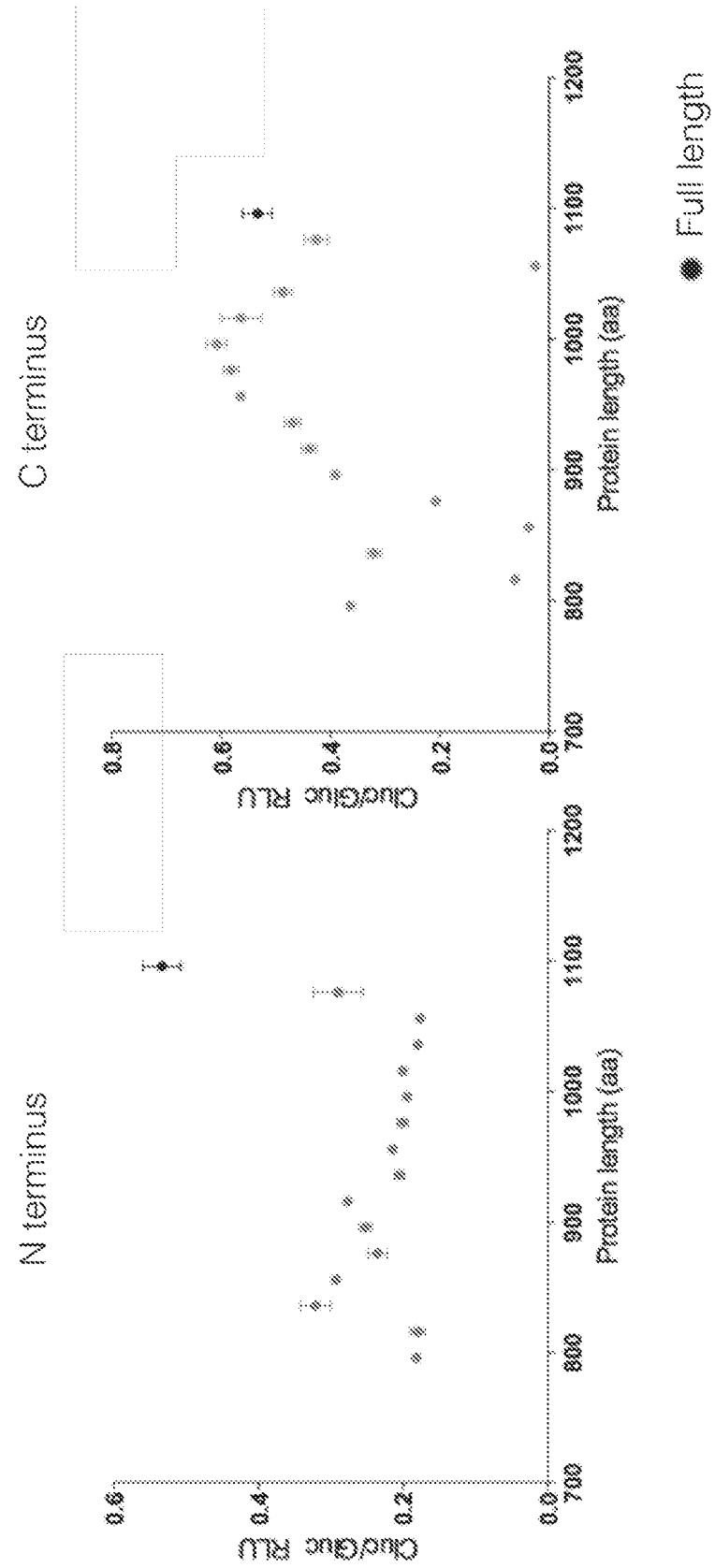


FIG. 103B

FIG. 103A

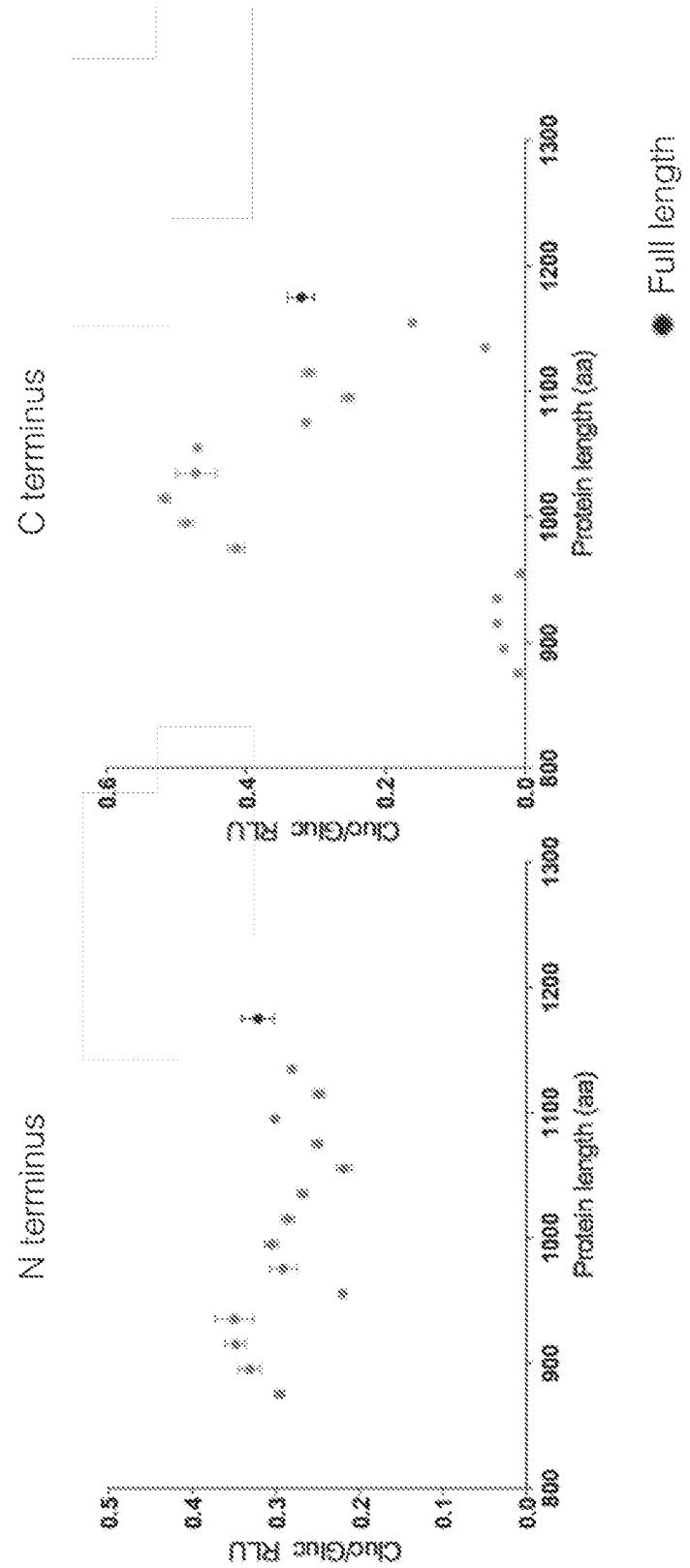


FIG. 104A

FIG. 104B

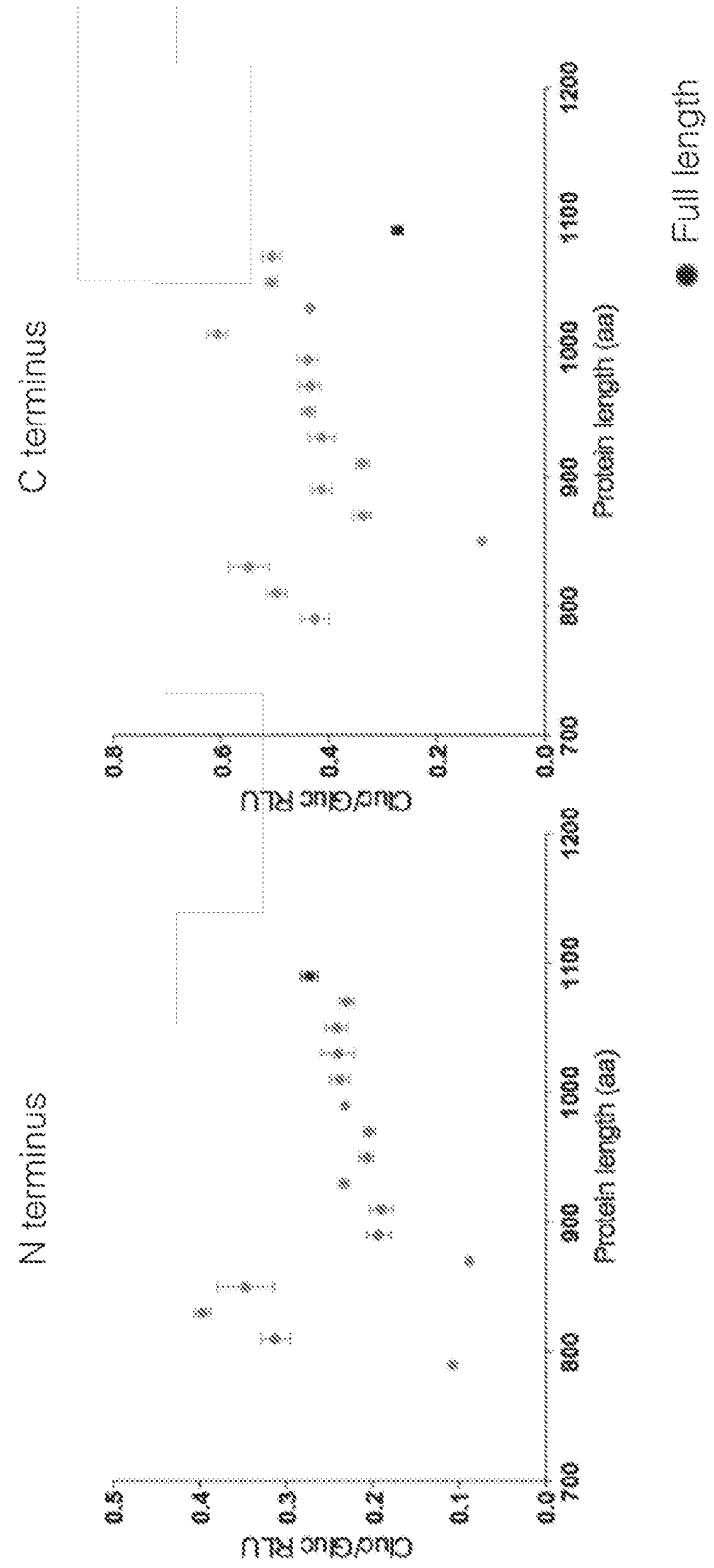


FIG. 105A

FIG. 105B

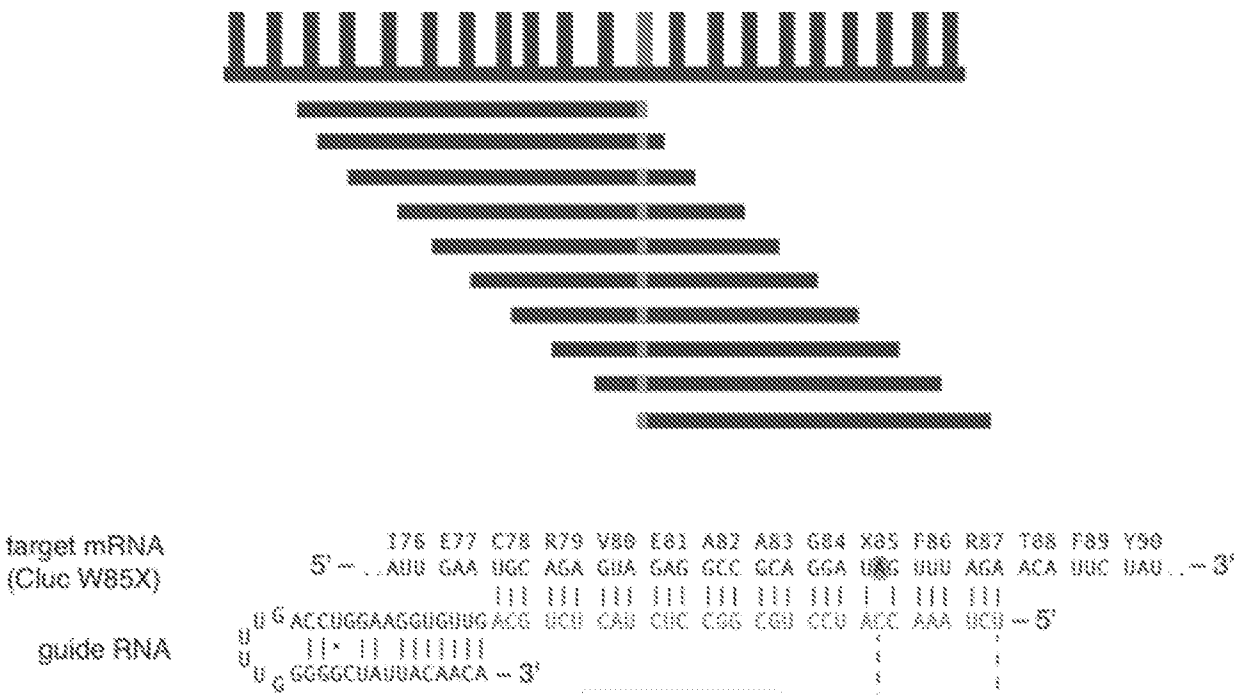


FIG. 106

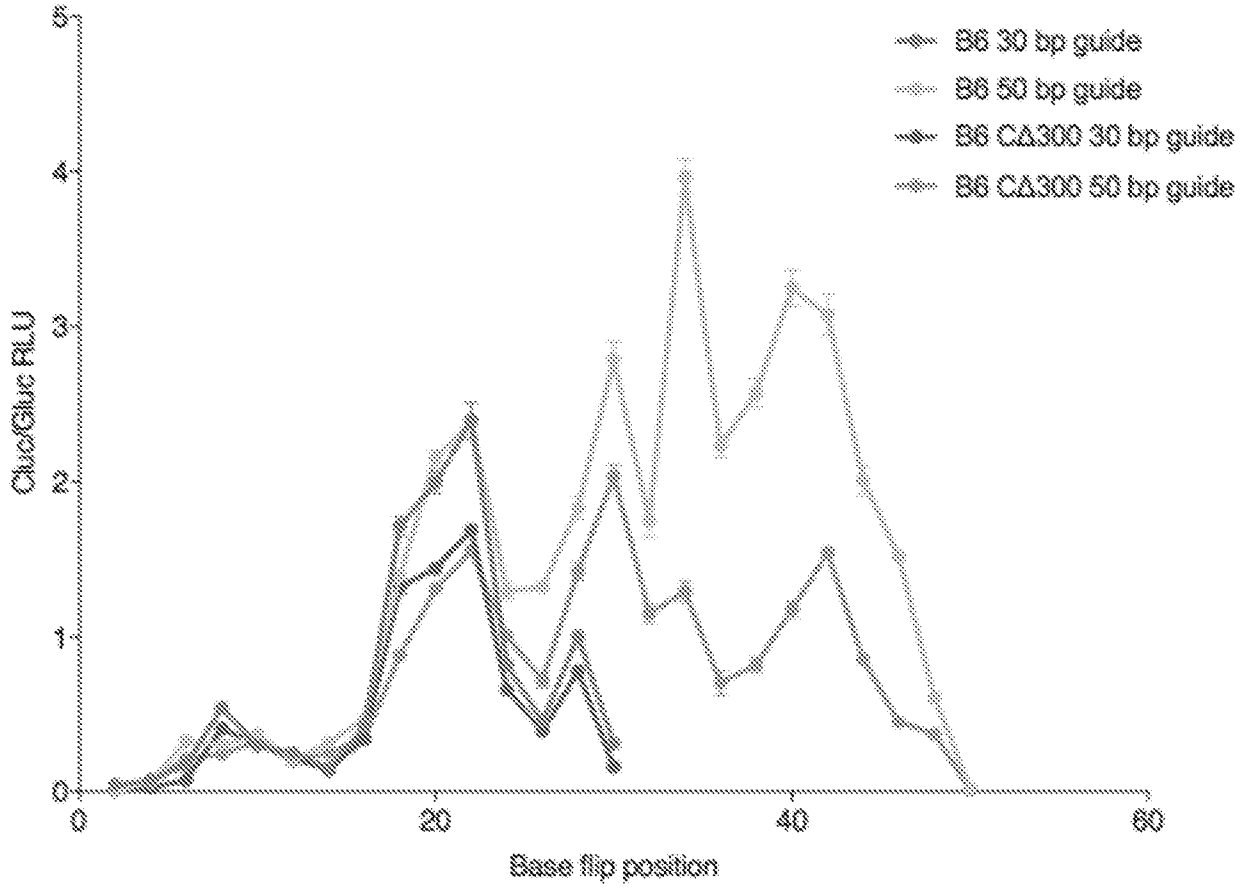

B6 30 bp guide
B6 50 bp guide
B6 CΔ300 30 bp guide
B6 CΔ300 50 bp guide
Cluc/Gluc RLU
0
1
2
3
4
5
0
20
40
60
Base flip position

FIG. 107

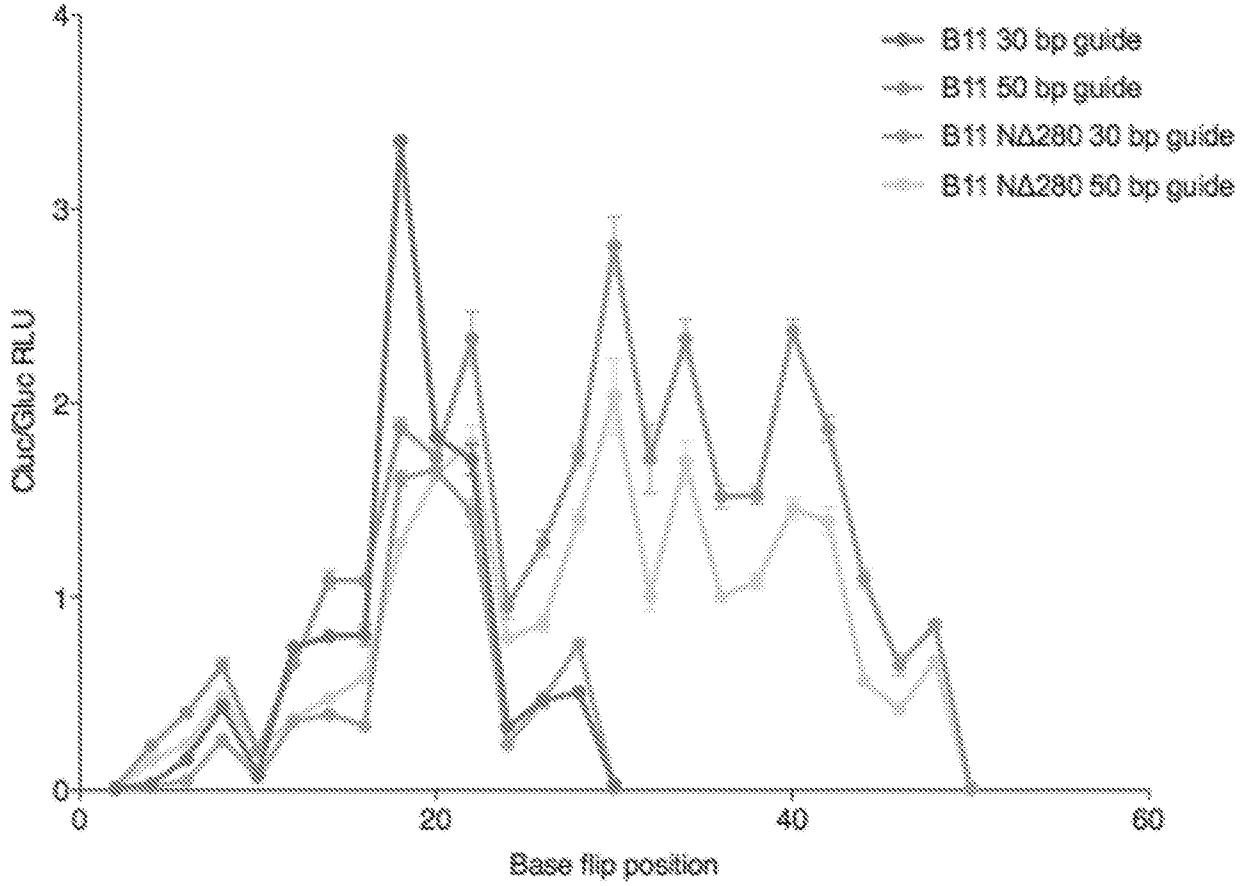

B11 30 bp guide
B11 50 bp guide
B11 NΔ280 30 bp guide
B11 NΔ280 50 bp guide
Cluc/Gluc RLU
0
1
2
3
4
0
20
40
60
Base flip position

FIG. 108

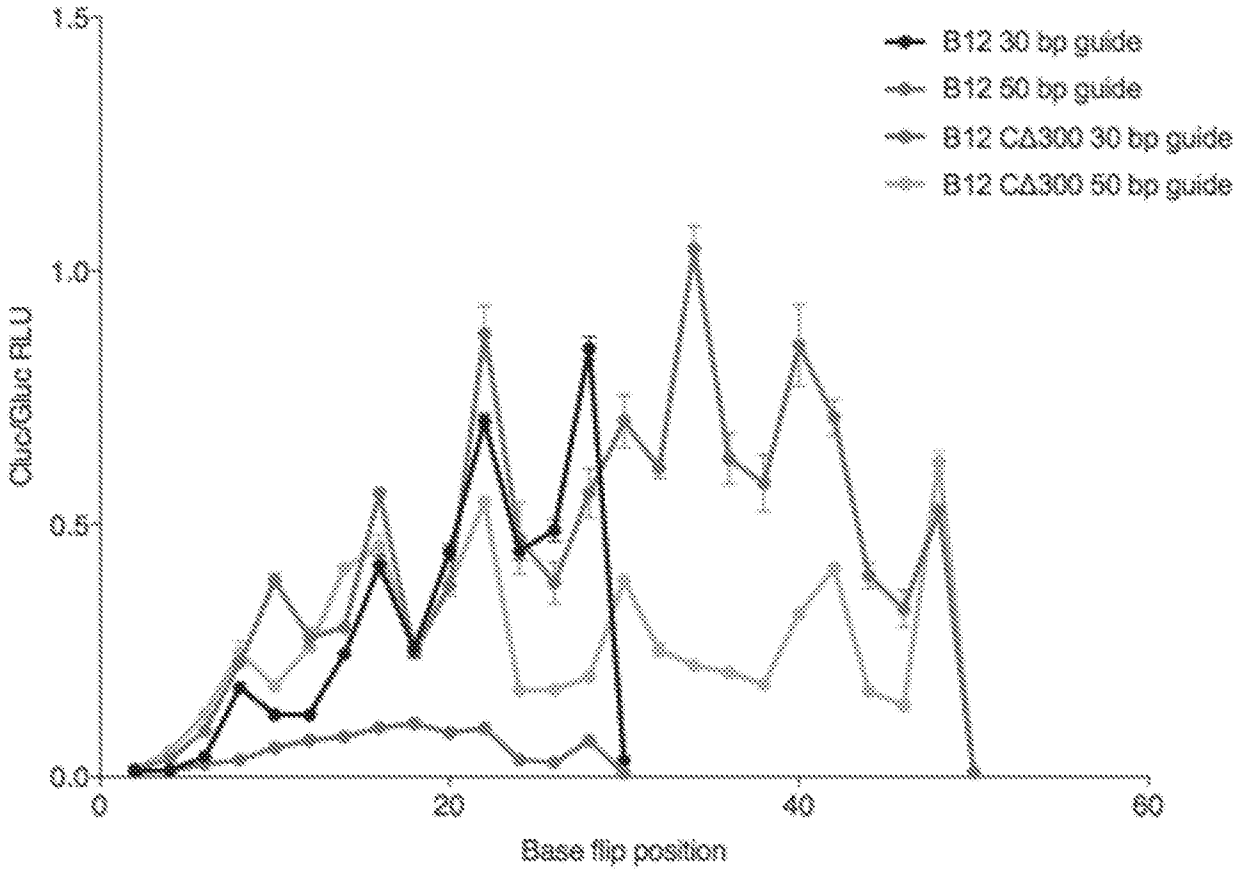

B12 30 bp guide
B12 50 bp guide
B12 CΔ300 30 bp guide
B12 CΔ300 50 bp guide
Cluc/Gluc RLU
1.5
1.0
0.5
0.0
0
20
40
60
Base flip position

FIG. 109

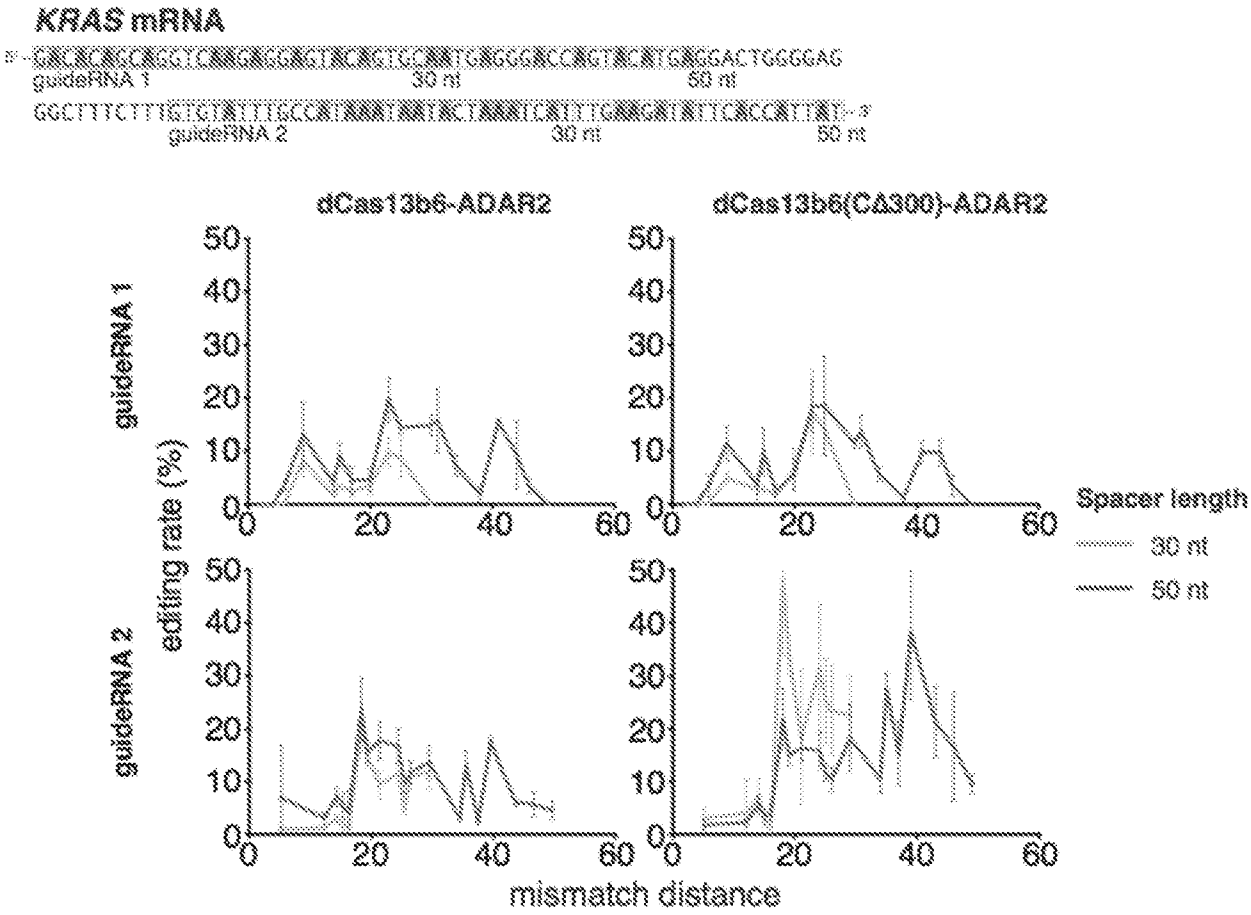

KRAS mRNA
guideRNA 1
30 nt
50 nt
guideRNA 2
30 nt
50 nt
dCas13b6-ADAR2
dCas13b6(CΔ300)-ADAR2
guideRNA 1
guideRNA 2
editing rate (%)
mismatch distance
Spacer length
30 nt
50 nt

FIG. 110

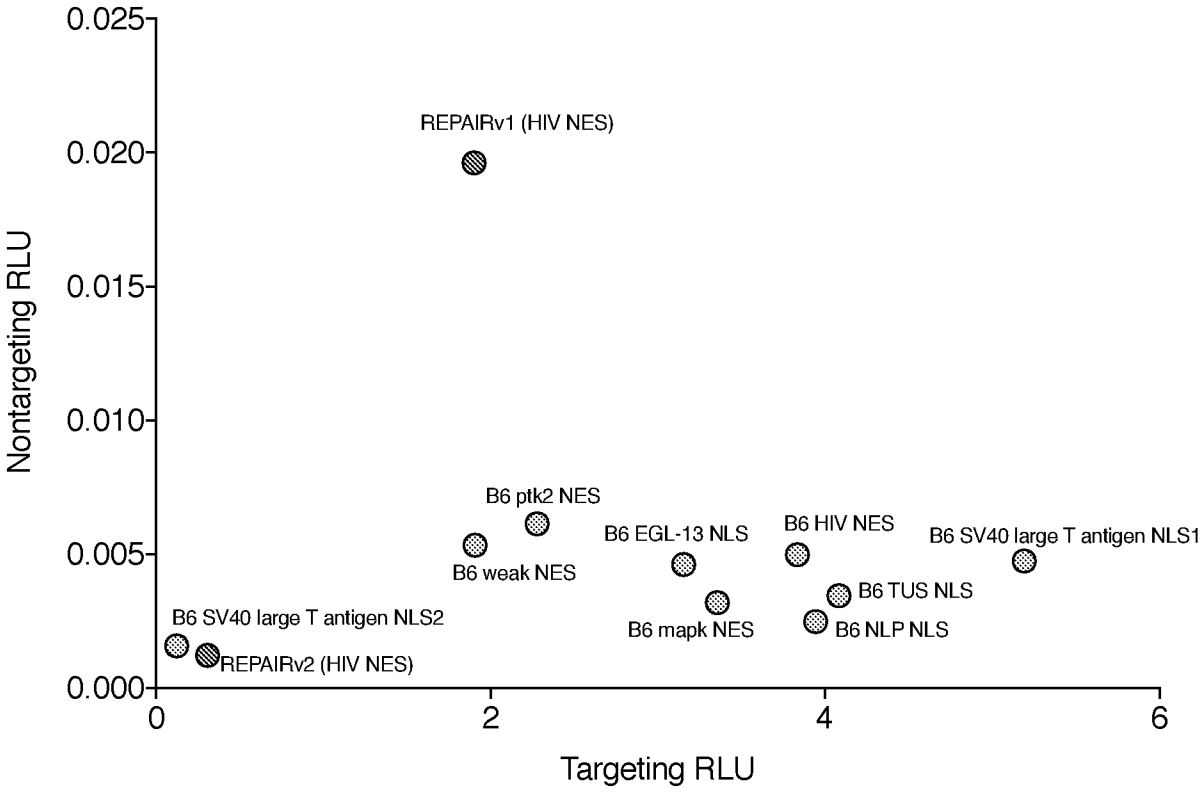

0.025
0.020
0.015
0.010
0.005
0.000
Nontargeting RLU
0
2
4
6
Targeting RLU
REPAIRv1 (HIV NES)
B6 ptk2 NES
B6 weak NES
B6 EGL-13 NLS
B6 HIV NES
B6 SV40 large T antigen NLS1
B6 TUS NLS
B6 mapk NES
B6 NLP NLS
B6 SV40 large T antigen NLS2
REPAIRv2 (HIV NES)

FIG. 111

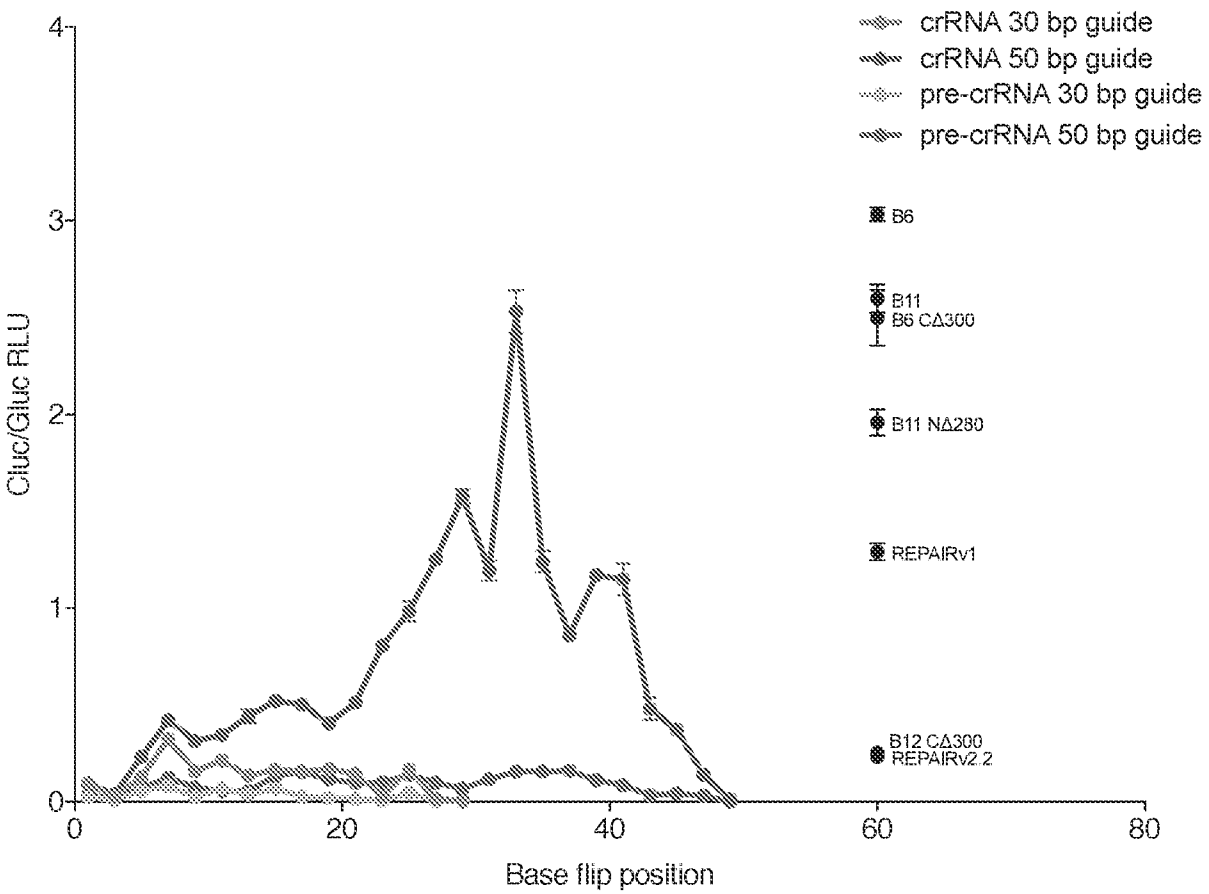

crRNA 30 bp guide
crRNA 50 bp guide
pre-crRNA 30 bp guide
pre-crRNA 50 bp guide
B6
B11
B6 CΔ300
B11 NΔ280
REPAIRv1
B12 CΔ300
REPAIRv2.2
Cluc/Gluc RLU
Base flip position
0
1
2
3
4
0
20
40
60
80

FIG. 112

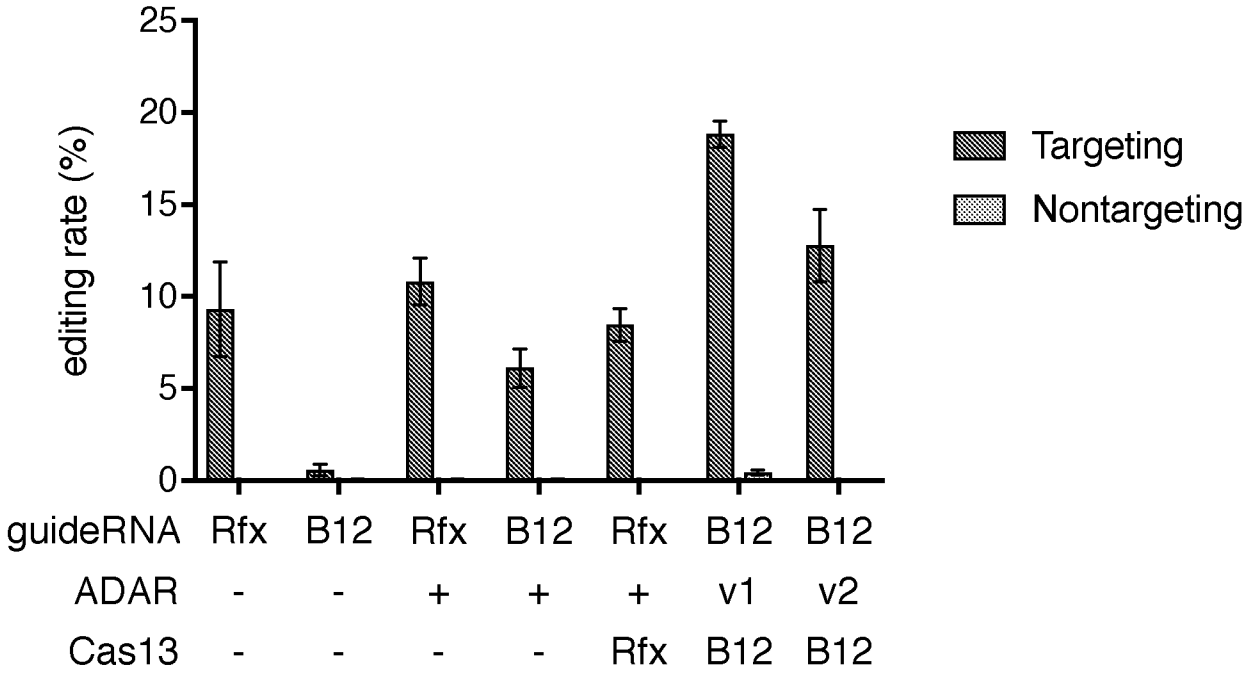

25
20
15
10
5
0
editing rate (%)
Targeting
Nontargeting
guideRNA Rfx B12 Rfx B12 Rfx B12 B12
ADAR - - + + + v1 v2
Cas13 - - - - Rfx B12 B12

FIG. 113

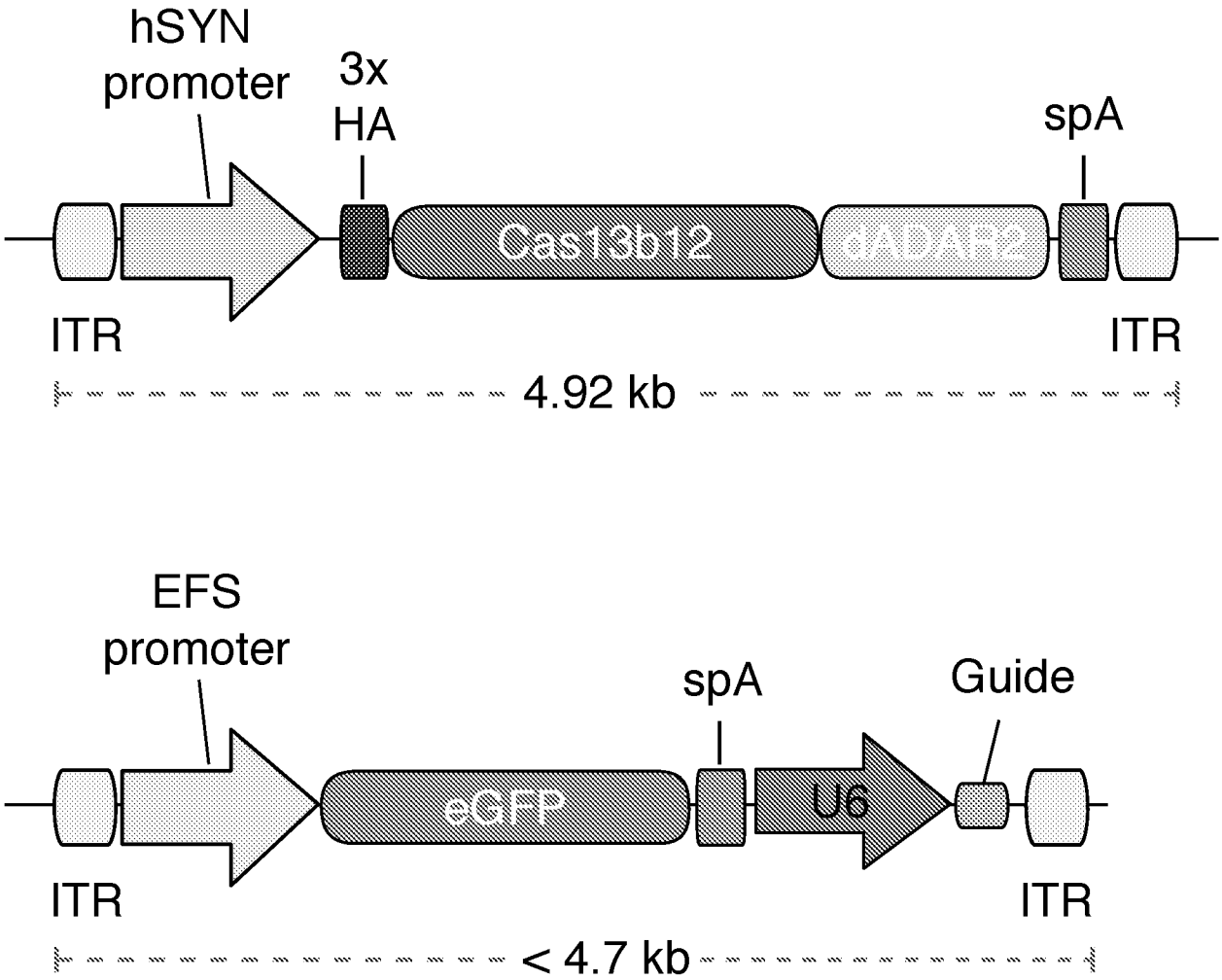

hSYN
promoter
3x
HA
spA
Cas13b12
dADAR2
ITR
ITR
4.92 kb
EFS
promoter
spA
Guide
eGFP
U6
ITR
ITR
< 4.7 kb

FIG. 114

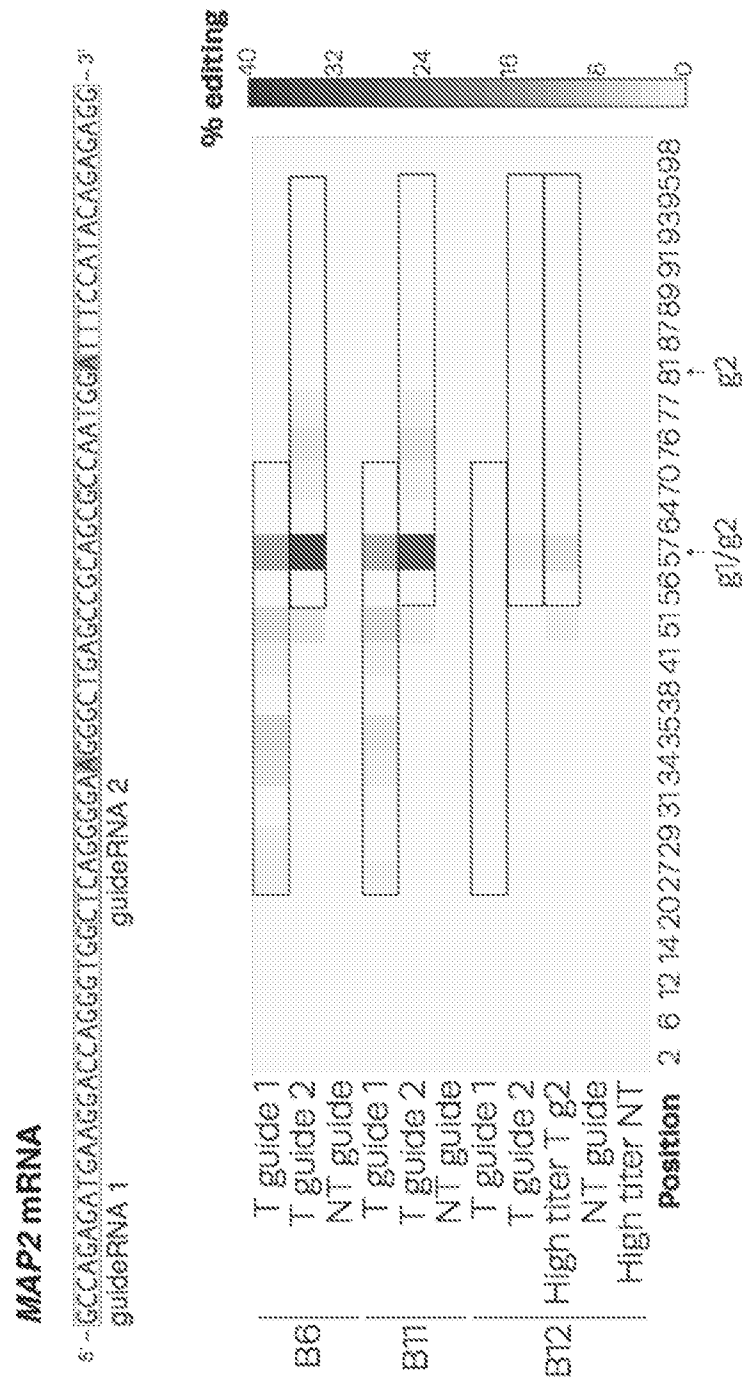

MAP2 mRNA
5' GCCAGAGATGAAGGACCAGGGTGGCTCAGGGGATGGGCTGAGCCGCAGCCAACGCCAATGGTTTCCATACAGAGGG 3'
guideRNA 1
guideRNA 2
% editing
40
32
24
16
8
0
B6
B11
B12
T guide 1
T guide 2
NT guide
T guide 1
T guide 2
NT guide
T guide 1
T guide 2
High titer T g2
NT guide
High titer NT
Position 2 6 12 14 20 27 29 31 34 35 38 41 51 56 57 64 70 76 77 81 87 89 91 93 95 98
g1/g2
g2

FIG. 115

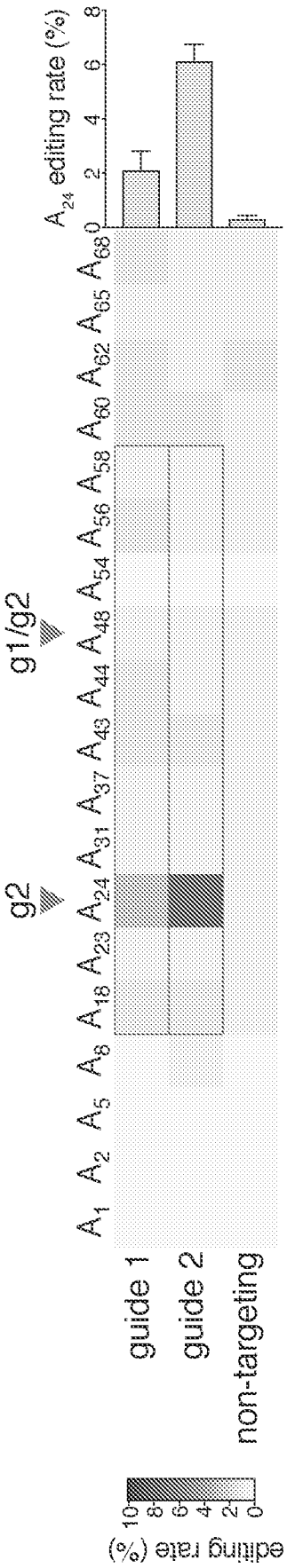

A24 editing rate (%)
0
2
4
6
8
A1 A2 A5 A8 A18 A23 A24 A31 A37 A43 A44 A48 A54 A56 A58 A60 A62 A65 A68
g1/g2
g2
guide 1
guide 2
non-targeting
editing rate (%)
10 8 6 4 2 0

FIG. 116

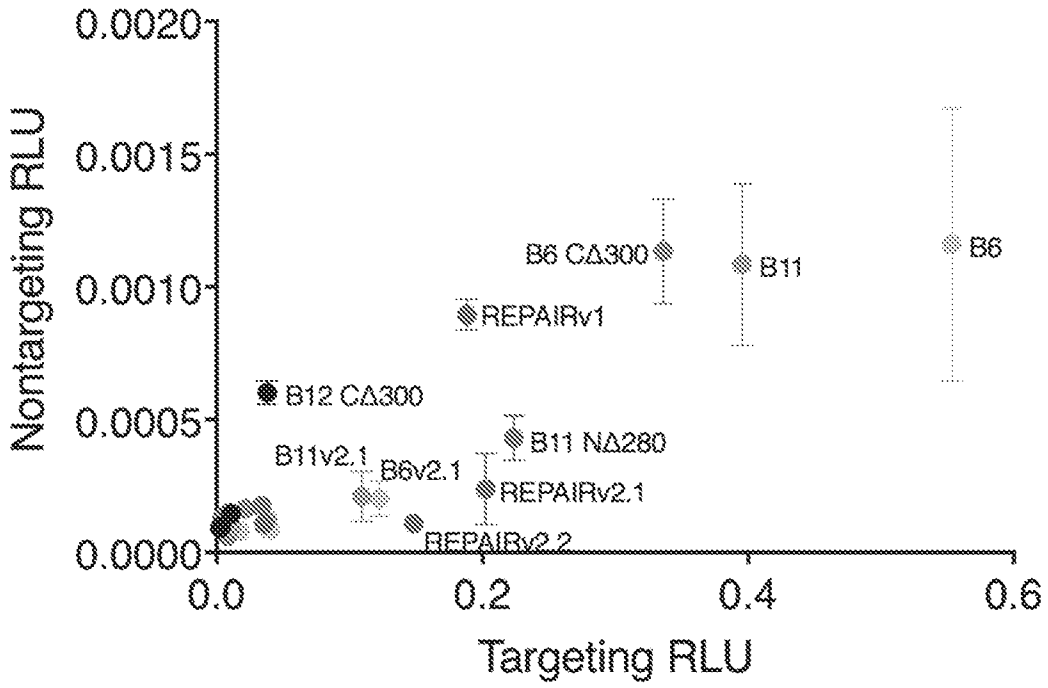

0.0020
0.0015
0.0010
0.0005
0.0000
Nontargeting RLU
0.0
0.2
0.4
0.6
Targeting RLU
B6 CΔ300
B11
B6
REPAIRv1
B12 CΔ300
B11 NΔ280
B11v2.1
B6v2.1
REPAIRv2.1
REPAIRv2.2

FIG. 117

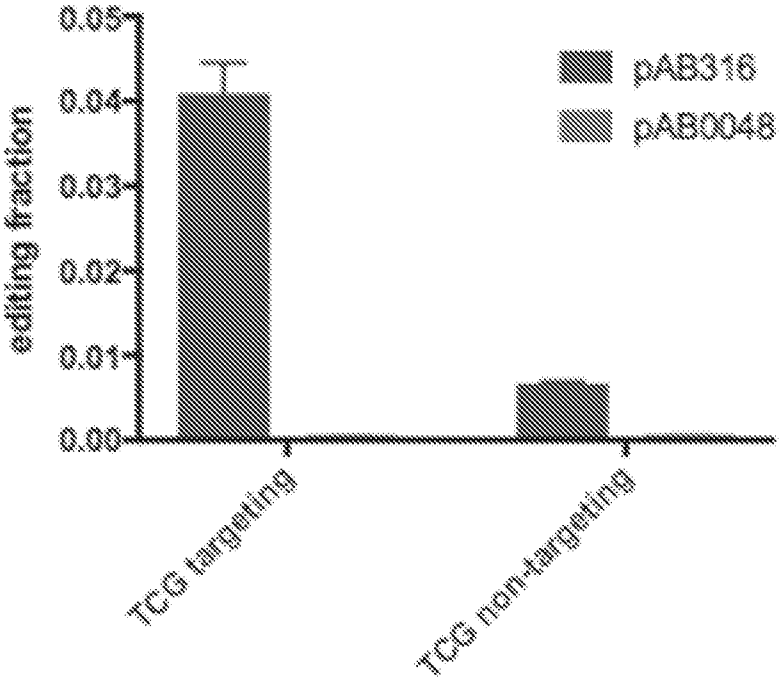

0.05
0.04
0.03
0.02
0.01
0.00
editing fraction
pAB316
pAB0048
TCG targeting
TCG non-targeting

FIG. 118

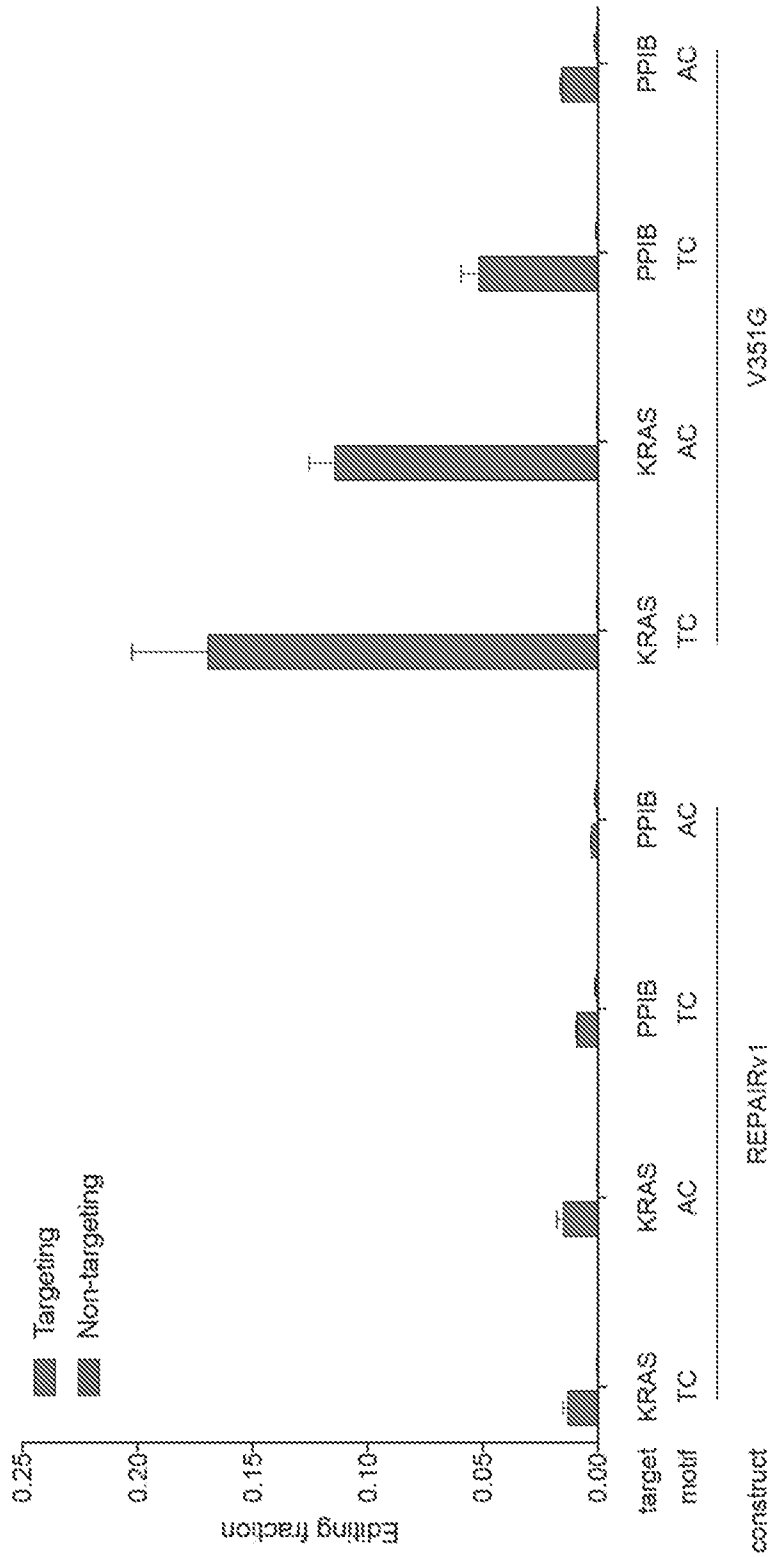

Targeting
Non-targeting
Editing fraction
0.25
0.20
0.15
0.10
0.05
0.00
target
motif
construct
KRAS
TC
KRAS
AC
PPIB
TC
PPIB
AC
KRAS
TC
KRAS
AC
PPIB
TC
PPIB
AC
REPAIRv1
V351G

FIG. 119

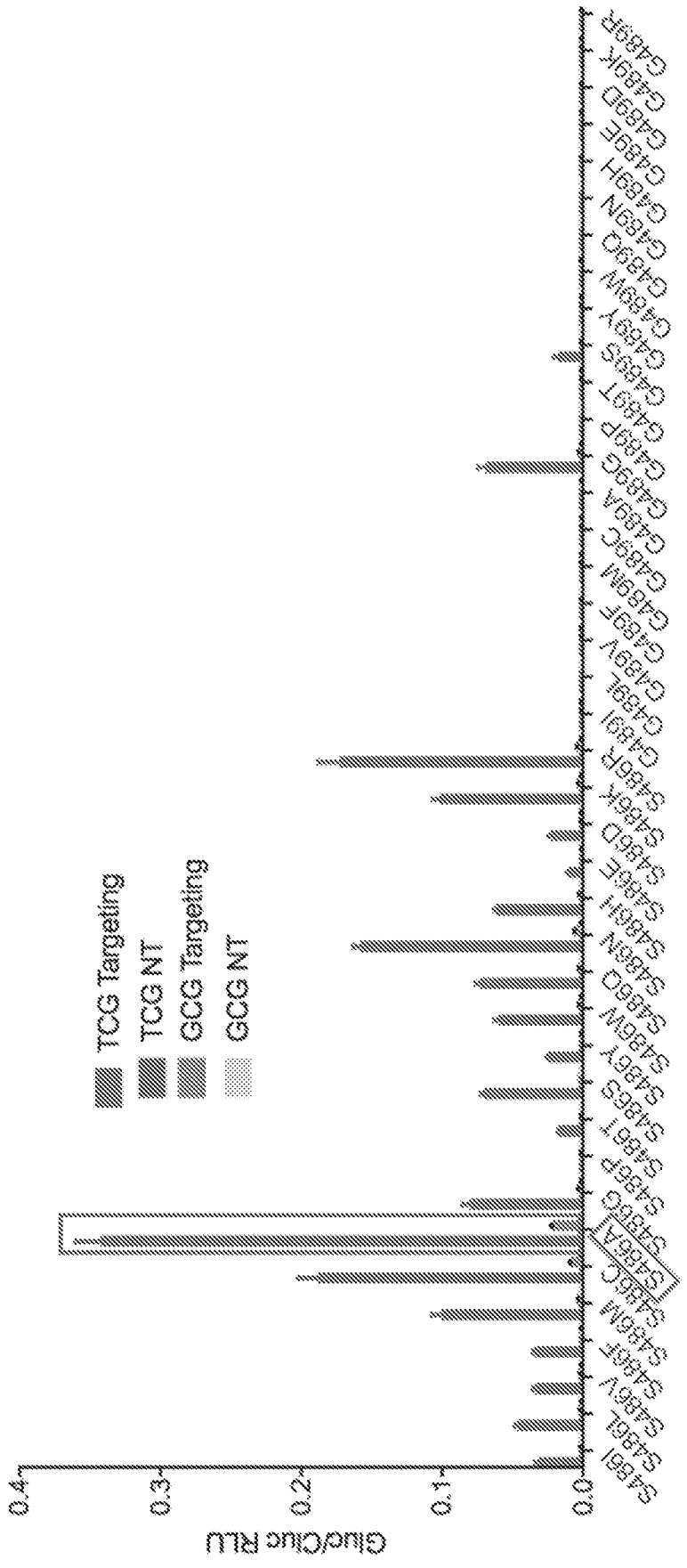

TCG Targeting
TCG NT
GCG Targeting
GCG NT
Gluc/Cluc RLU
0.4
0.3
0.2
0.1
0.0
Mutant

FIG. 120

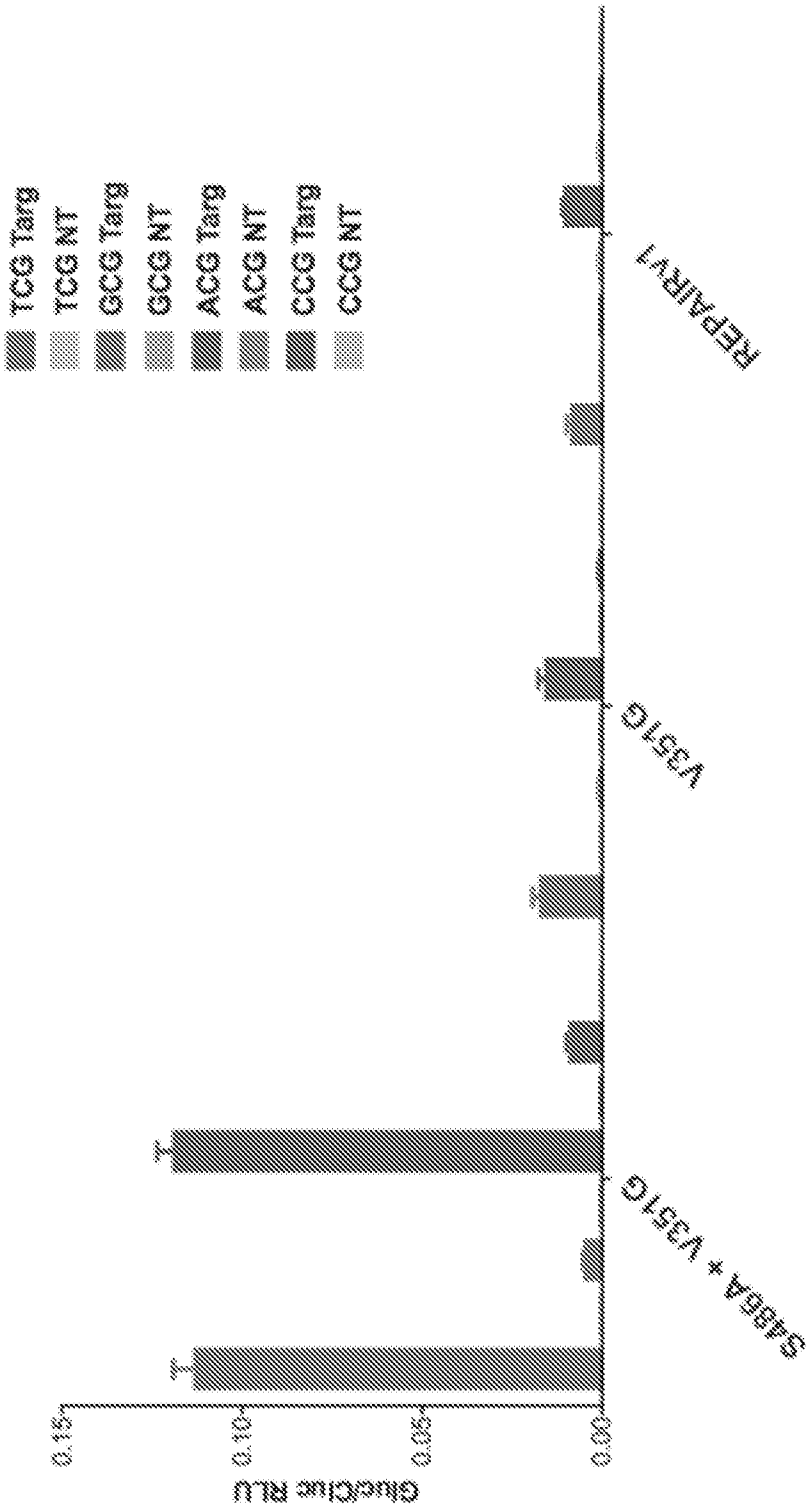

TCG Targ
TCG NT
GCG Targ
GCG NT
ACG Targ
ACG NT
CCG Targ
CCG NT
GlucClue RLU
0.15
0.10
0.05
0.00
S486A + V351G
V351G
REPAIRV1
ADAR Mutant

FIG. 121

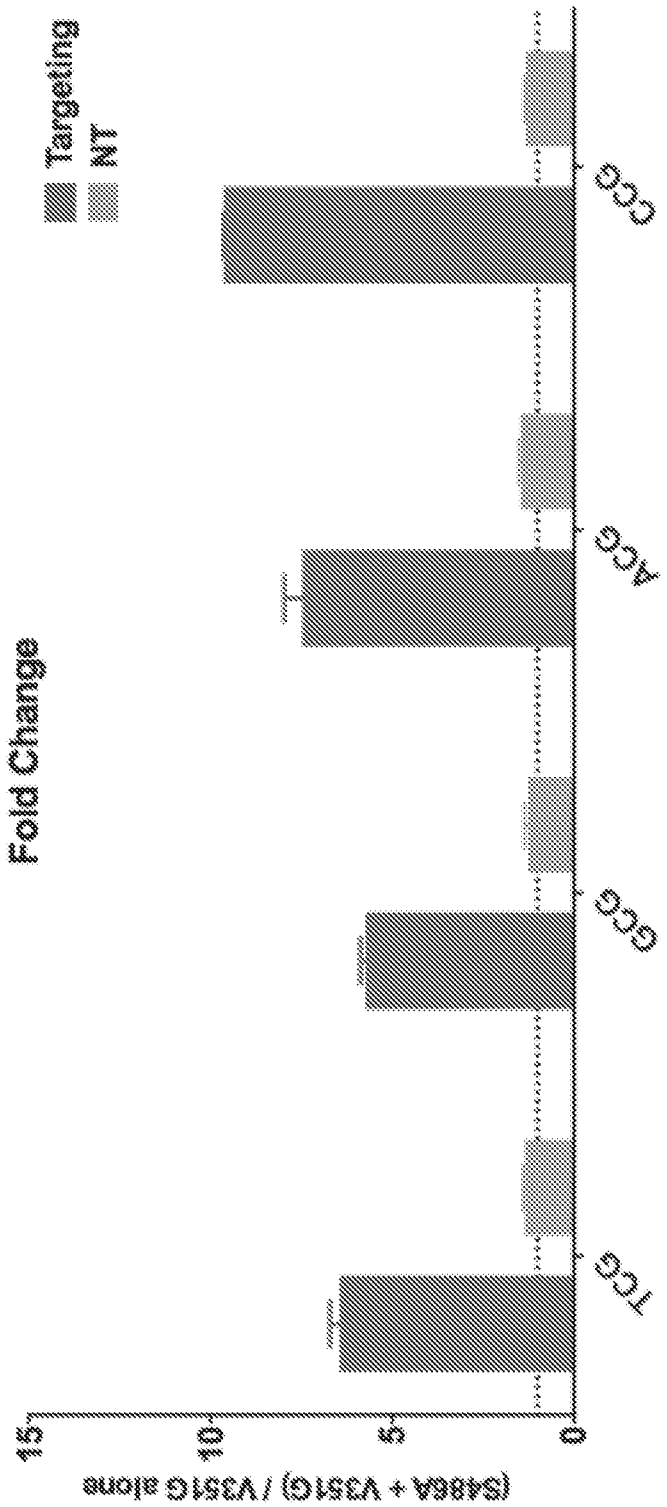

Targeting
NT
Fold Change
(S484A + V351G) / V351G alone
15
10
5
0
TCC
GCC
ACC
GCC
Motif

FIG. 122

FIG. 123A

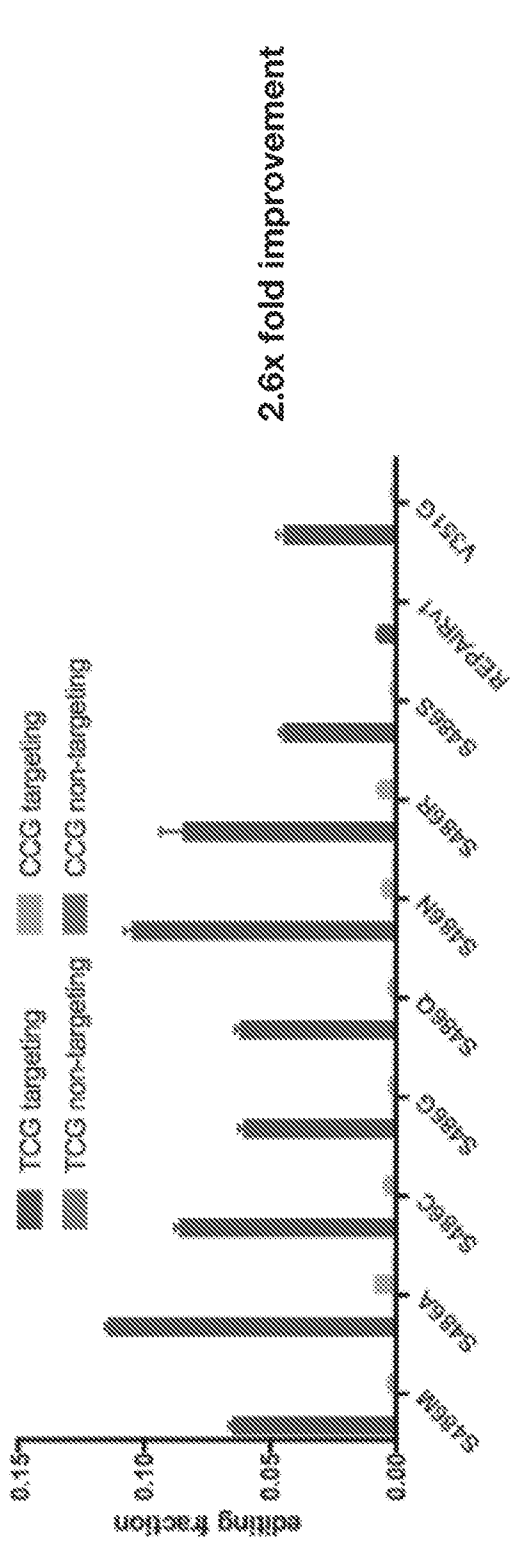
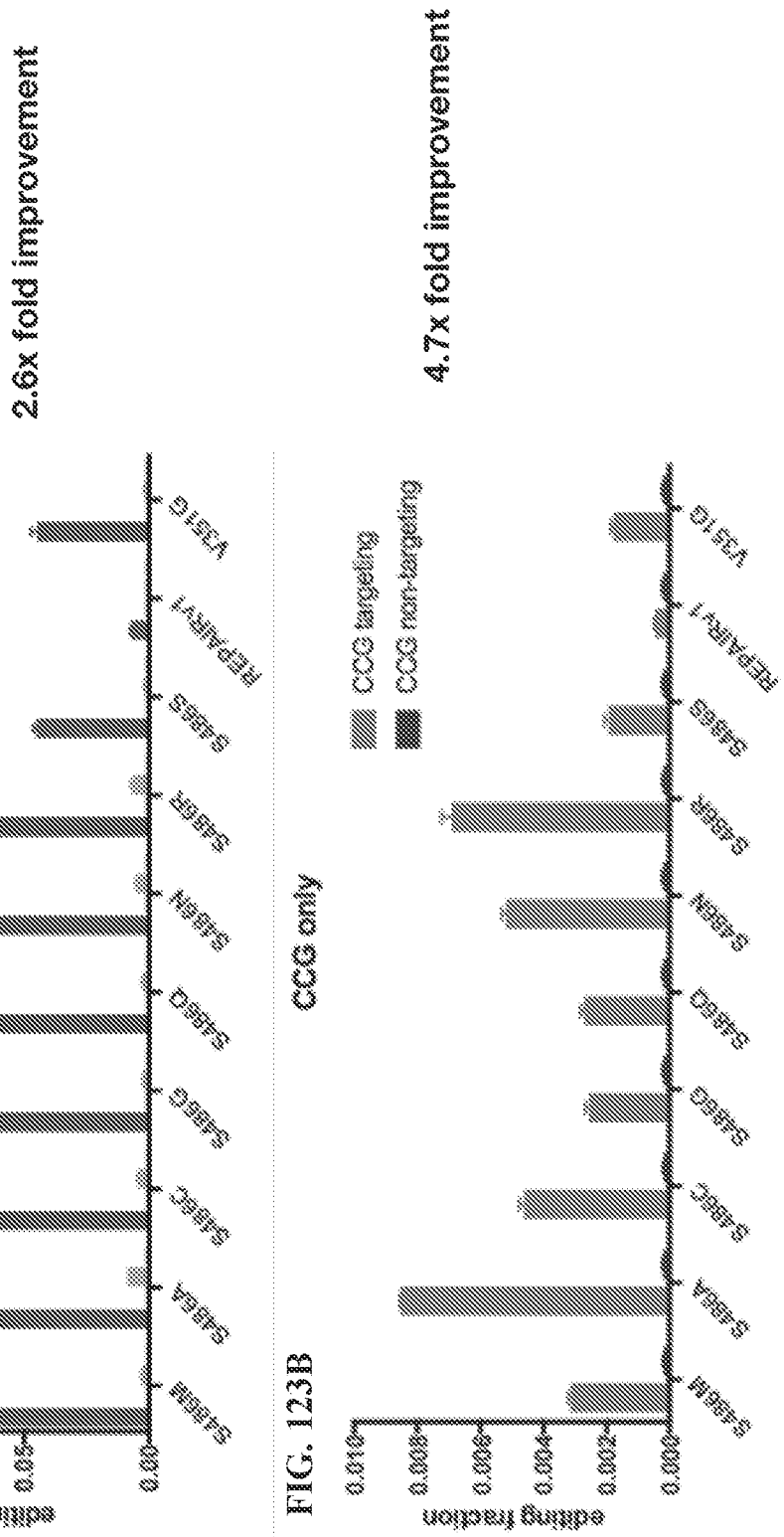


FIG. 123B



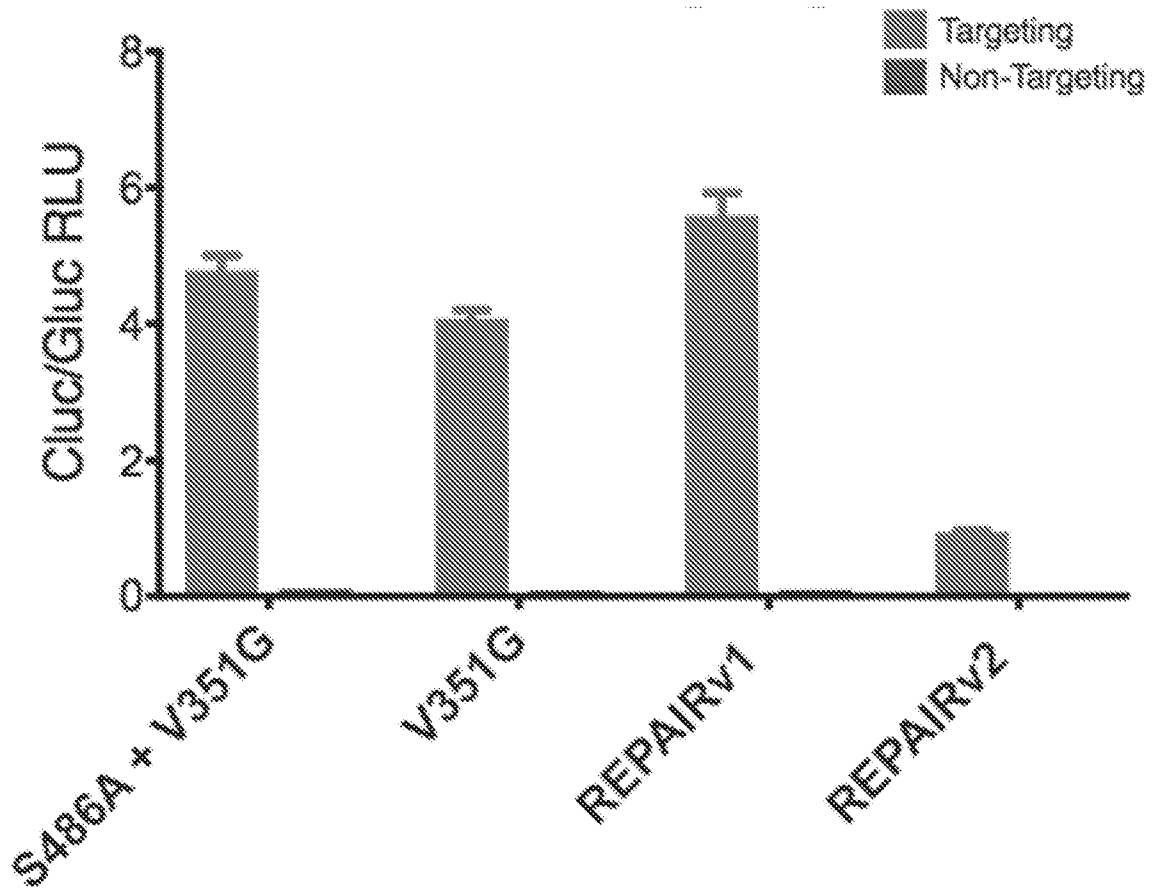

Targeting
Non-Targeting
Cluc/Gluc RLU
0
2
4
6
8
S486A + V351G
V351G
REPAIRv1
REPAIRv2

FIG. 124

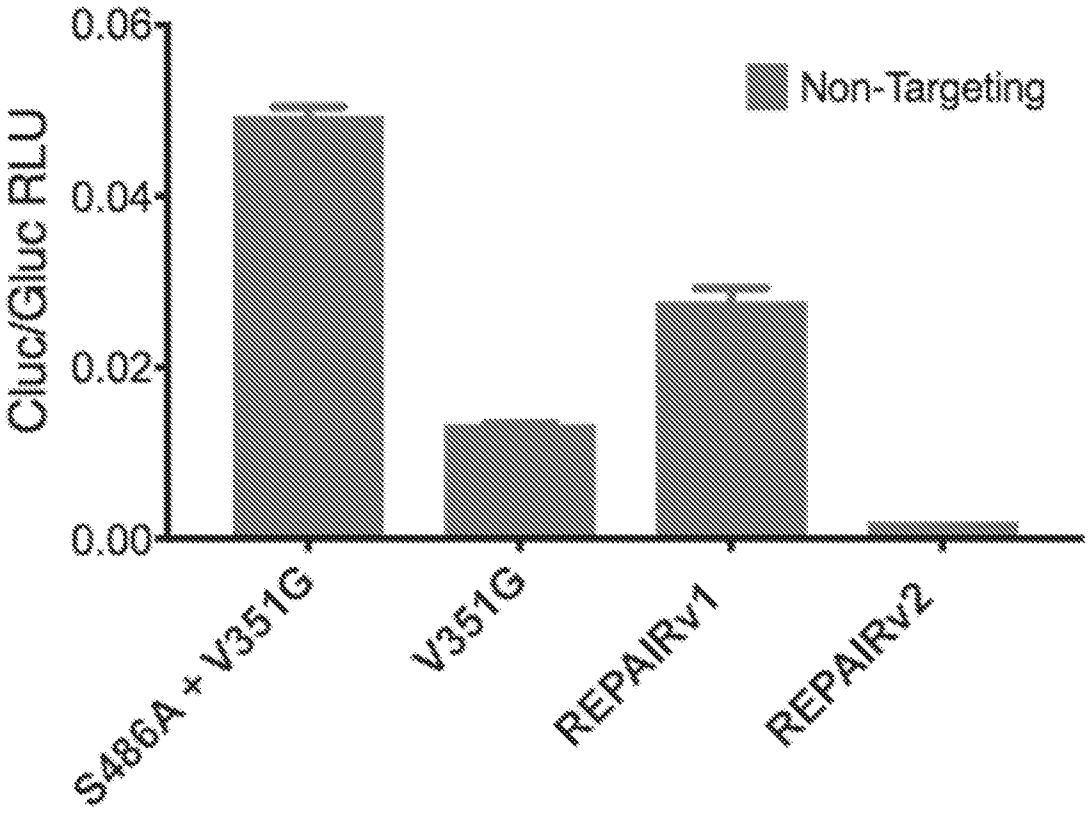

0.06
0.04
0.02
0.00
Cluc/Gluc RLU
Non-Targeting
S486A + V351G
V351G
REPAIRv1
REPAIRv2

FIG. 125

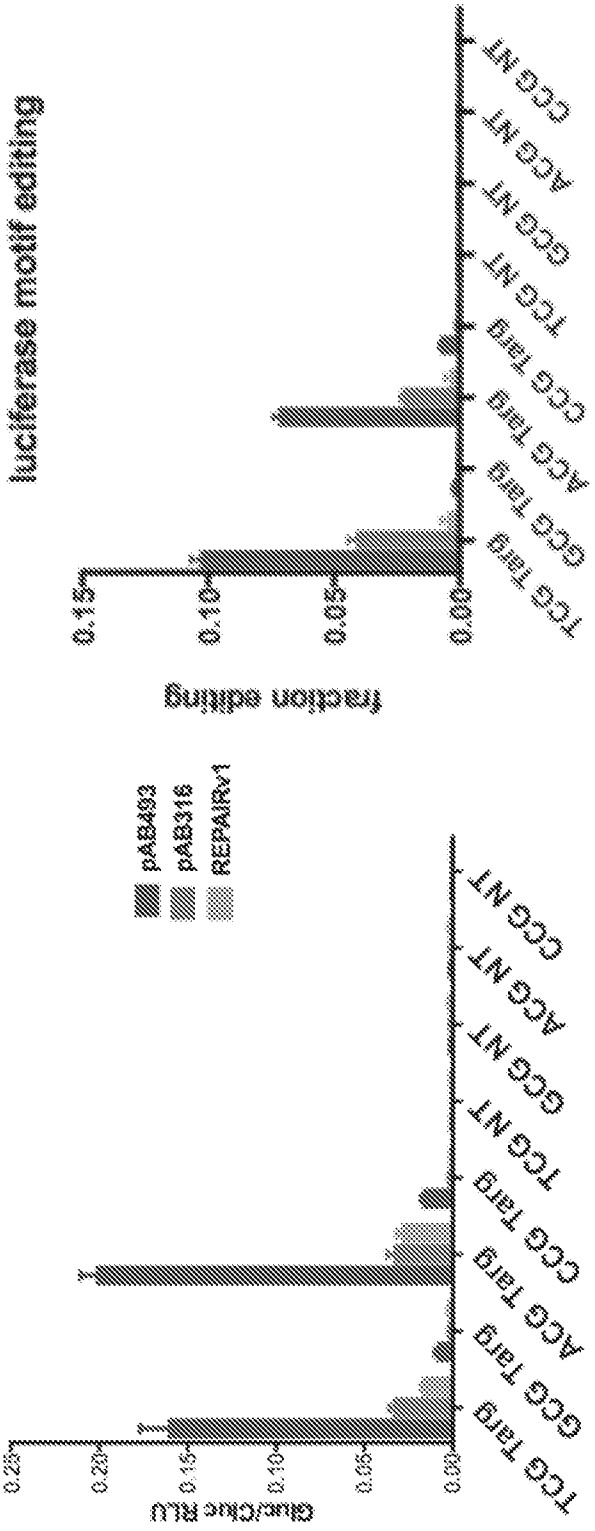


FIG. 126A

FIG. 126B

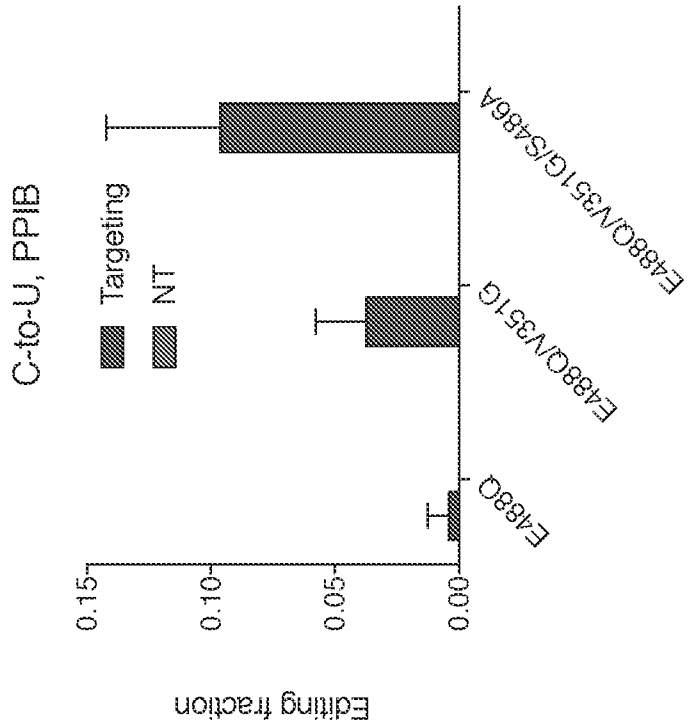


FIG. 127B

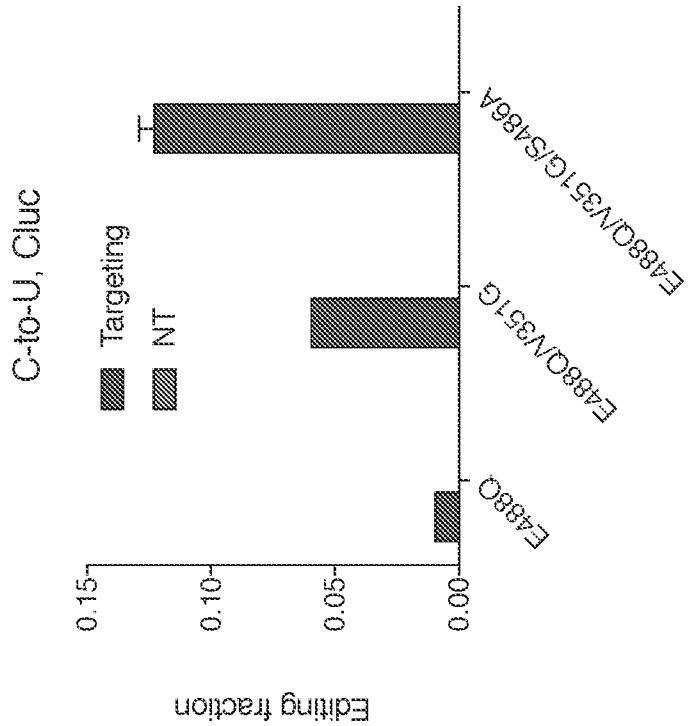


FIG. 127A

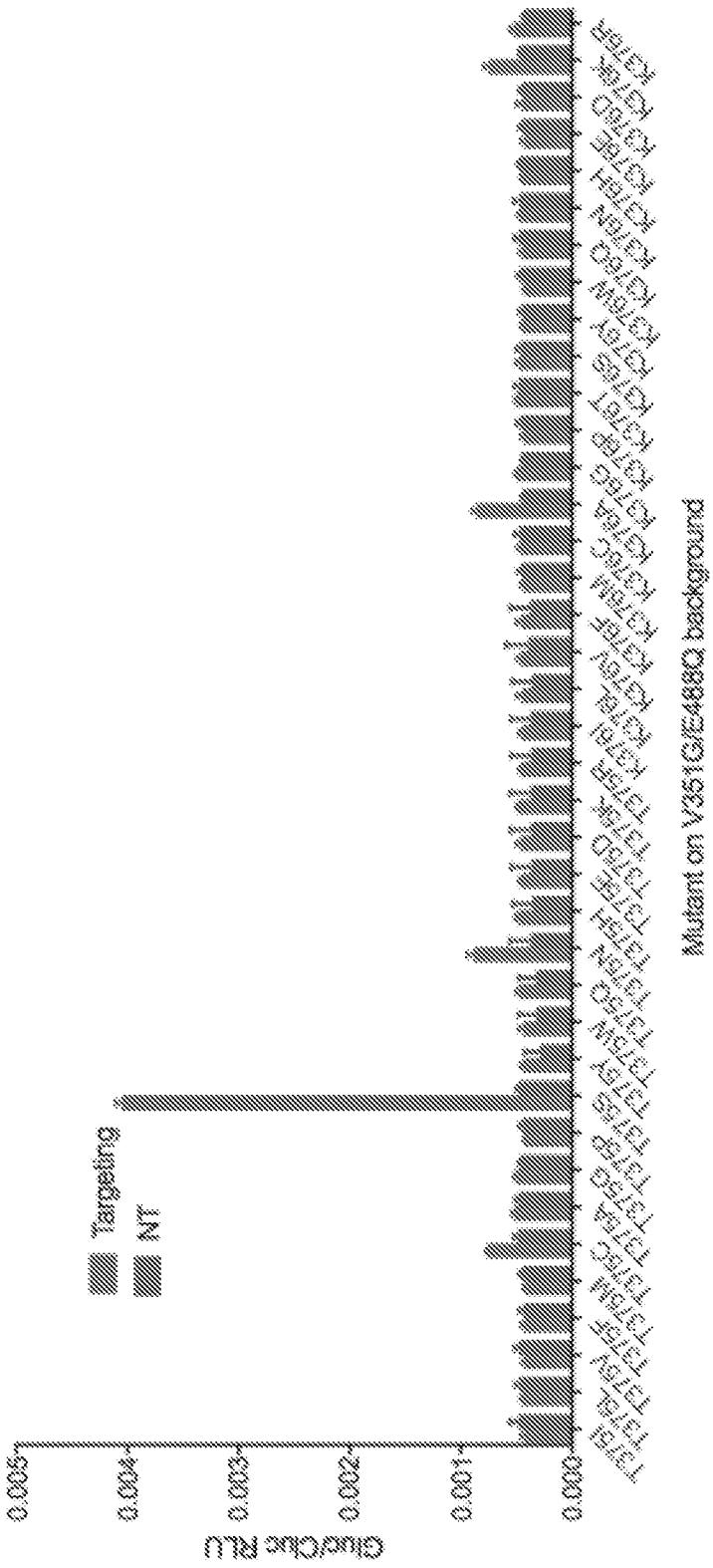

Targeting
NT
Glo/Cluc RLU
0.005
0.004
0.003
0.002
0.001
0.000
Mutant on V351G/E488Q background

FIG. 128

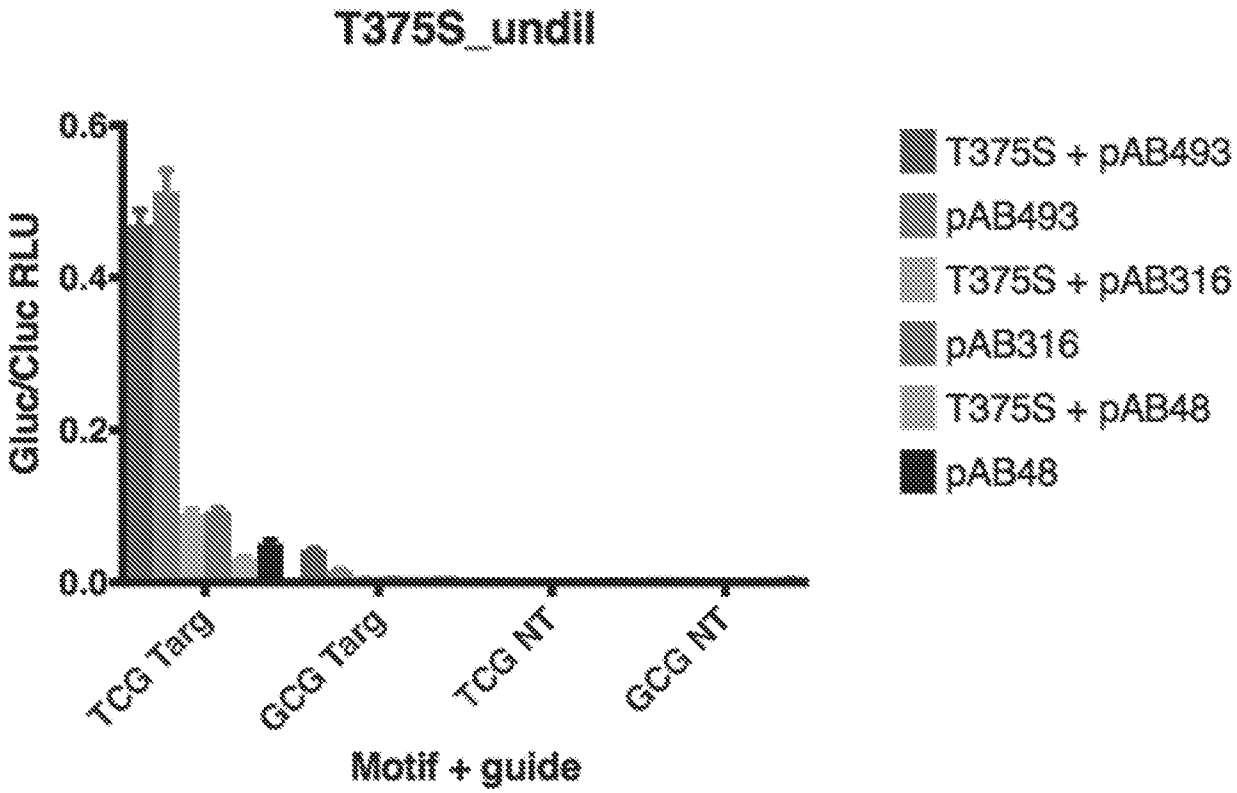

T375S_undil
Gluc/Cluc RLU
0.6
0.4
0.2
0.0
TCG Targ
GCG Targ
TCG NT
GCG NT
Motif + guide
T375S + pAB493
pAB493
T375S + pAB316
pAB316
T375S + pAB48
pAB48

FIG. 129

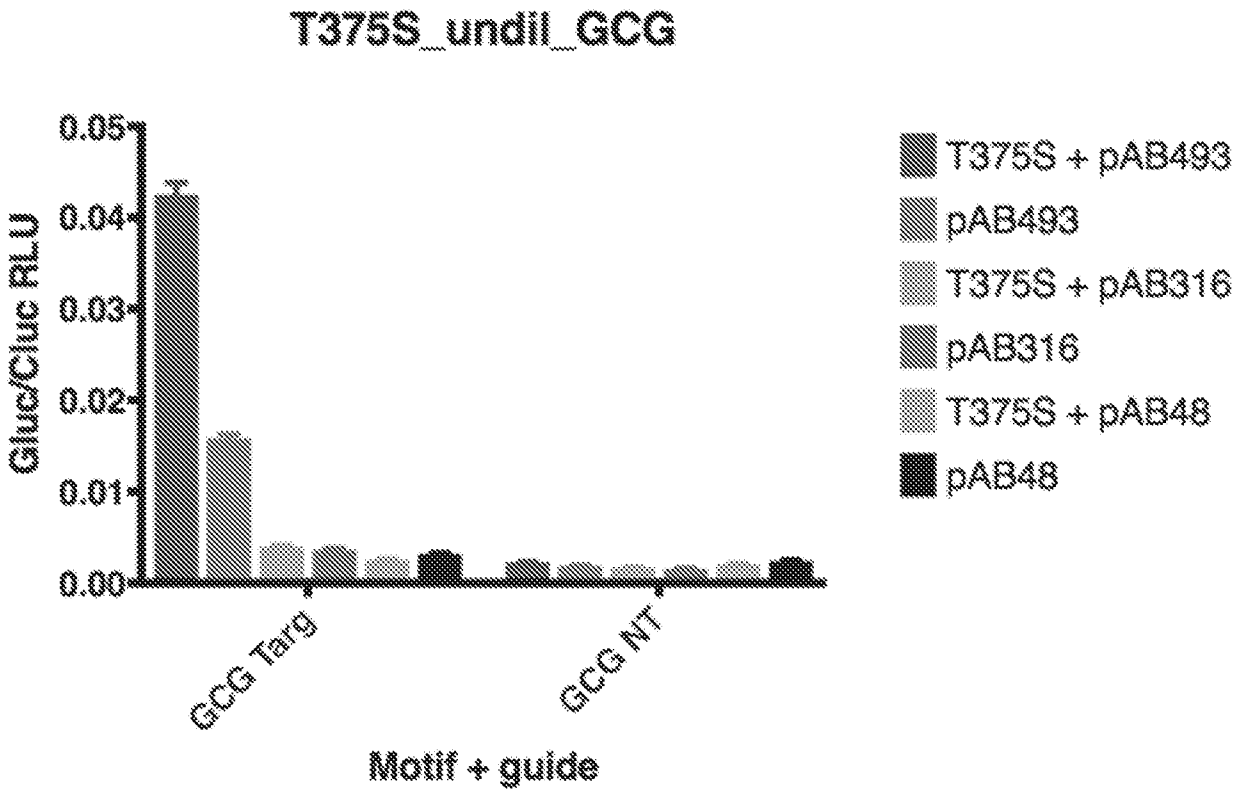

T375S_undil_GCG
Gluc/Cluc RLU
0.05
0.04
0.03
0.02
0.01
0.00
GCG Targ
GCG NT
Motif + guide
T375S + pAB493
pAB493
T375S + pAB316
pAB316
T375S + pAB48
pAB48

FIG. 130

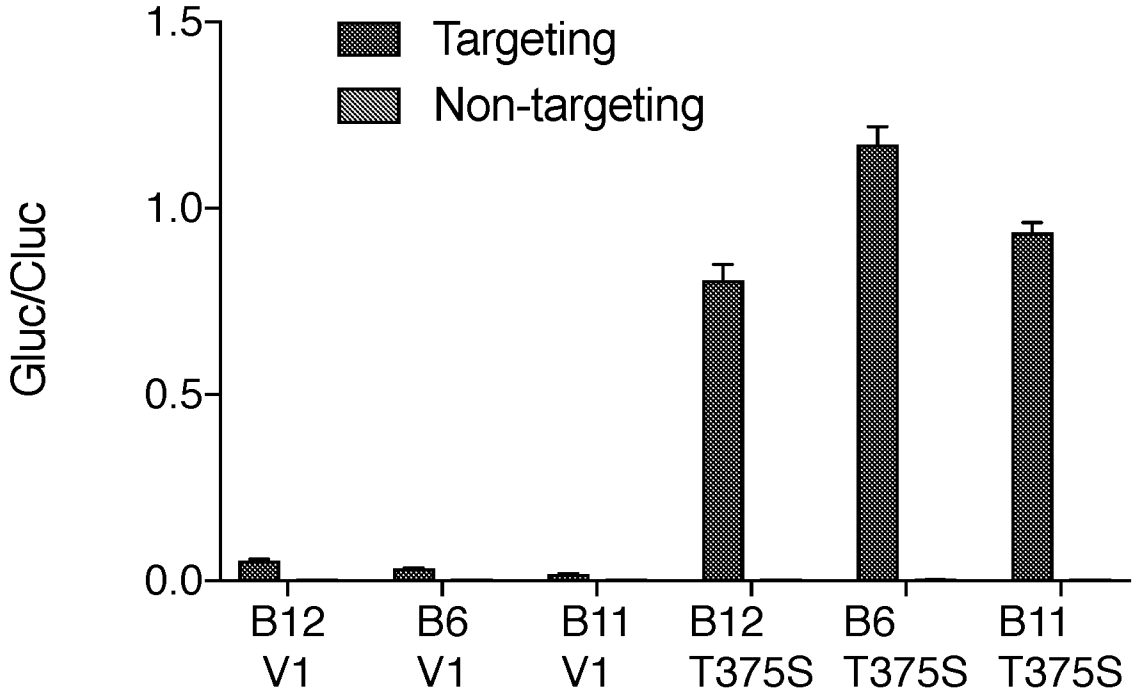

1.5
1.0
0.5
0.0
Gluc/Cluc
Targeting
Non-targeting
B12 V1
B6 V1
B11 V1
B12 T375S
B6 T375S
B11 T375S

FIG. 131

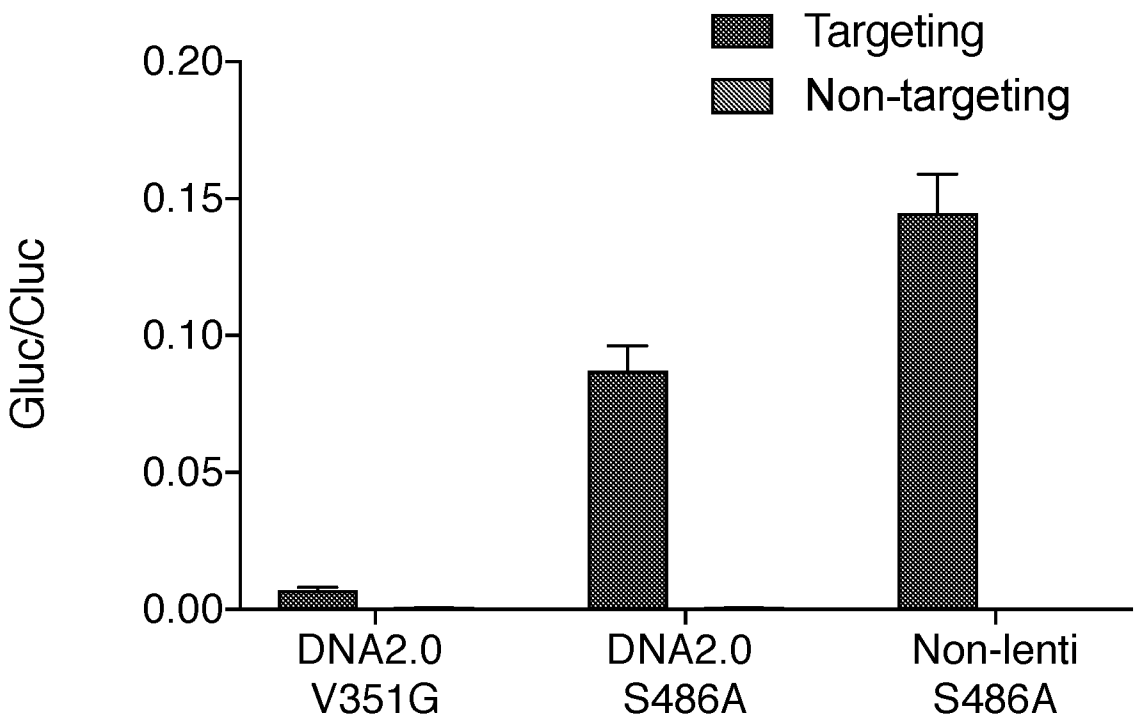

Targeting
Non-targeting
Gluc/Cluc
0.20
0.15
0.10
0.05
0.00
DNA2.0 V351G
DNA2.0 S486A
Non-lenti S486A

FIG. 132

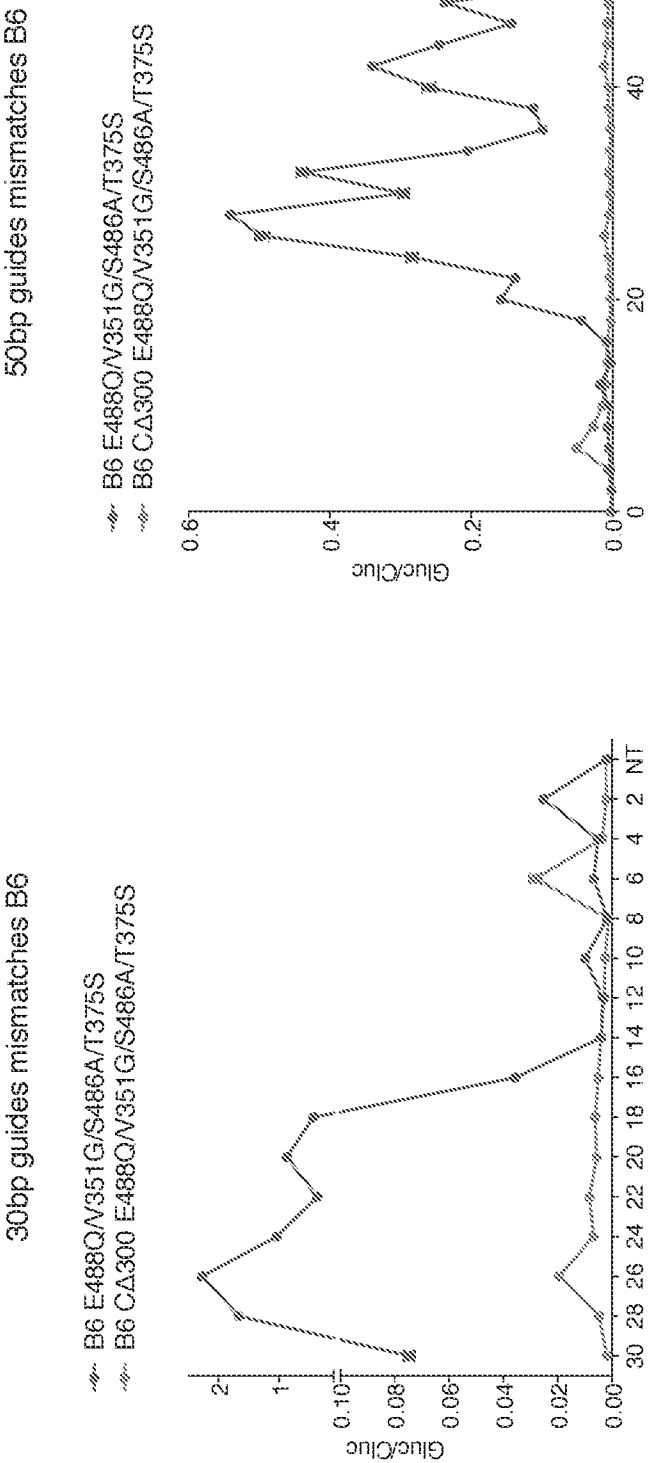


FIG. 133A

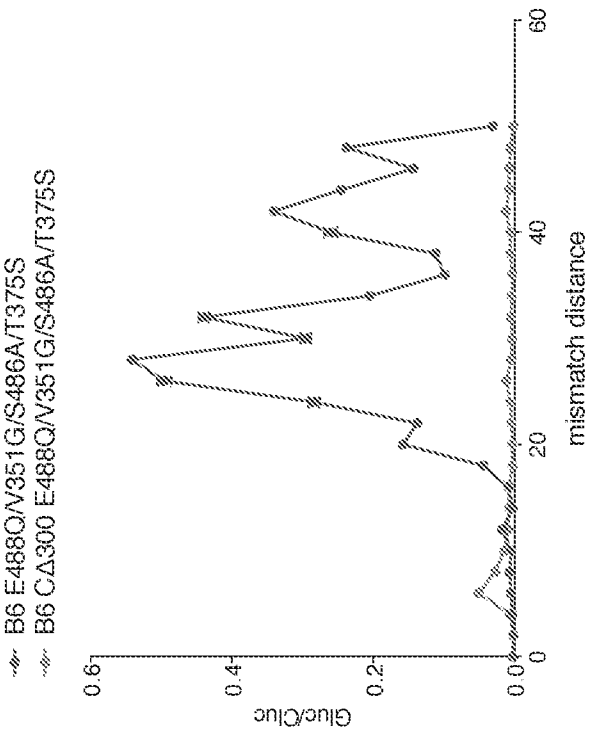


FIG. 133B

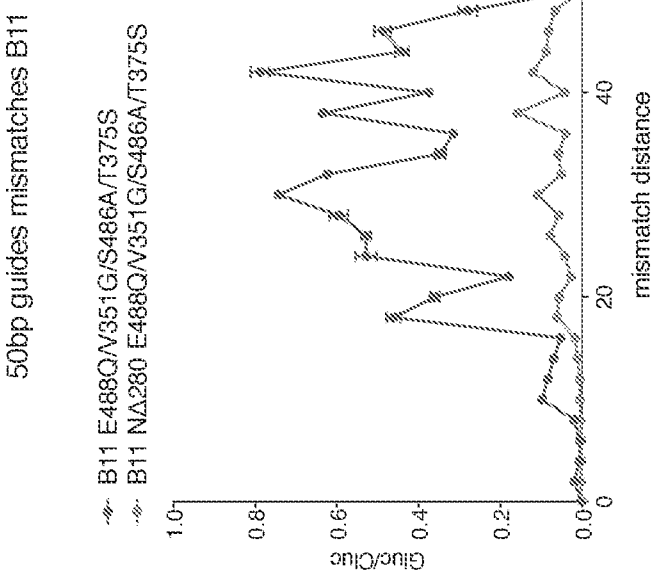


FIG. 134B

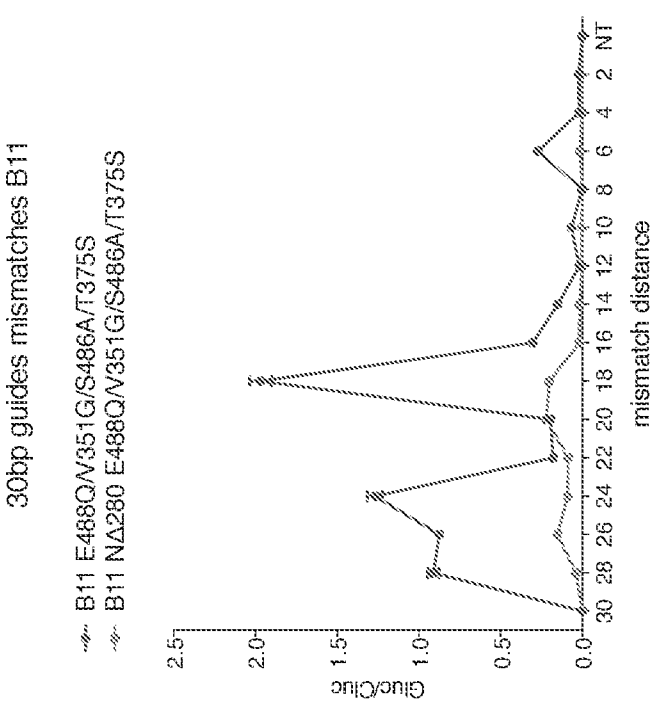


FIG. 134A

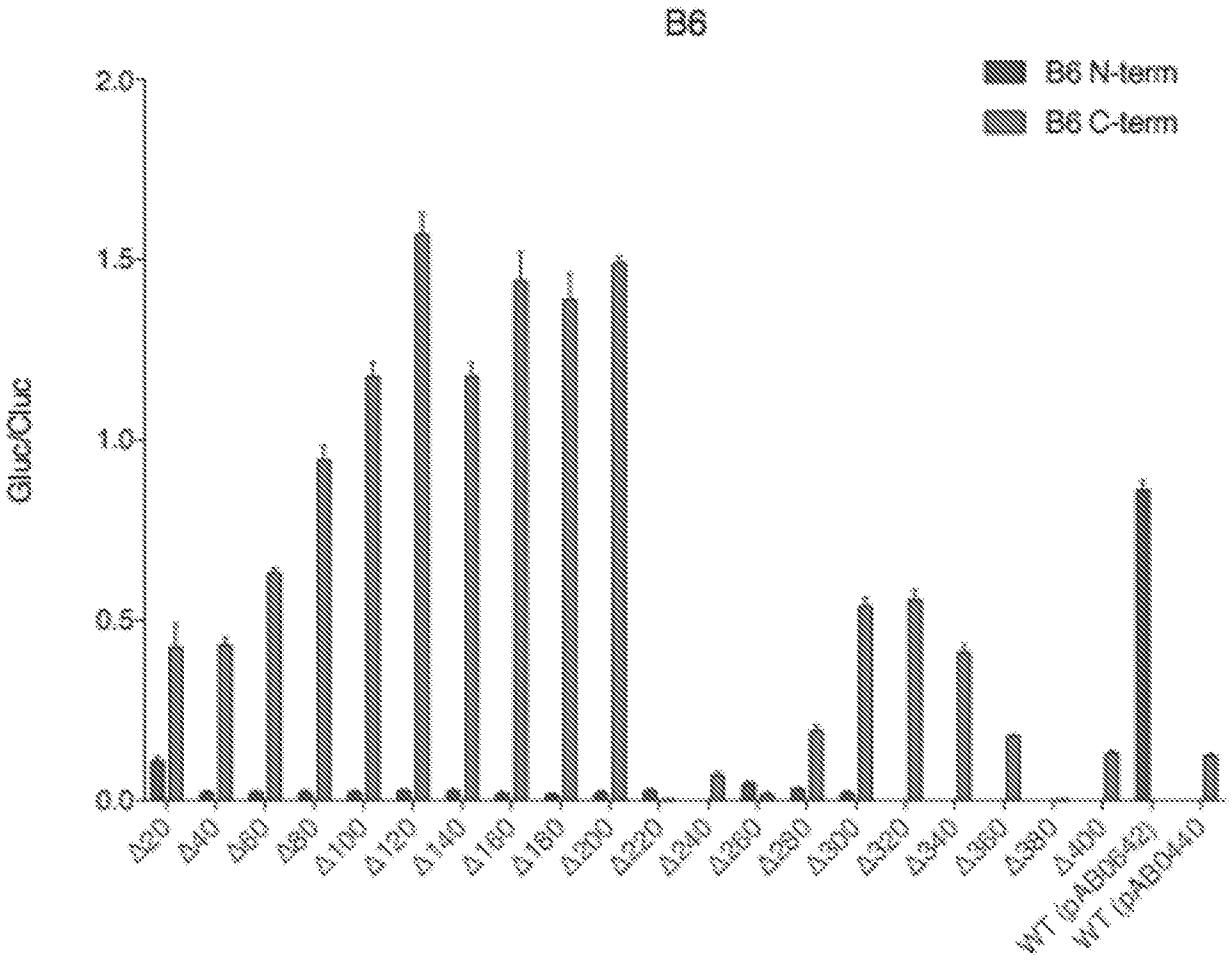

B6
B6 N-term
B6 C-term
Gluc/Gluc
2.0
1.5
1.0
0.5
0.0
Δ20
Δ40
Δ60
Δ80
Δ100
Δ120
Δ140
Δ160
Δ180
Δ200
Δ220
Δ240
Δ260
Δ280
Δ300
Δ320
Δ340
Δ360
Δ380
Δ400
WT (pAB06-42)
WT (pAB0440

FIG. 135

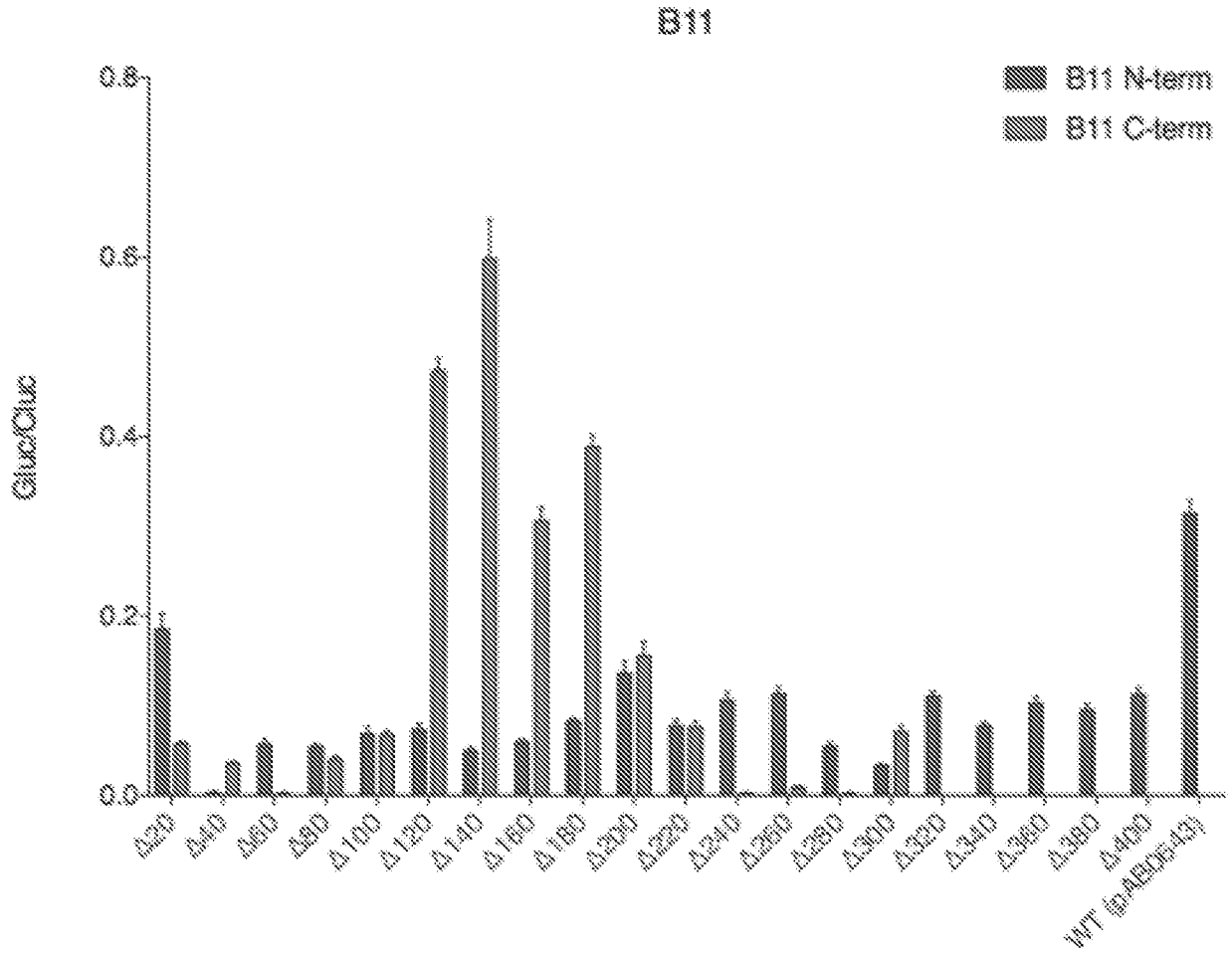

B11
B11 N-term
B11 C-term
Gluc/Gluc
0.0
0.2
0.4
0.6
0.8
Δ20
Δ40
Δ60
Δ80
Δ100
Δ120
Δ140
Δ160
Δ180
Δ200
Δ220
Δ240
Δ260
Δ280
Δ300
Δ320
Δ340
Δ360
Δ380
Δ400
WT (pAB06-43)

FIG. 136

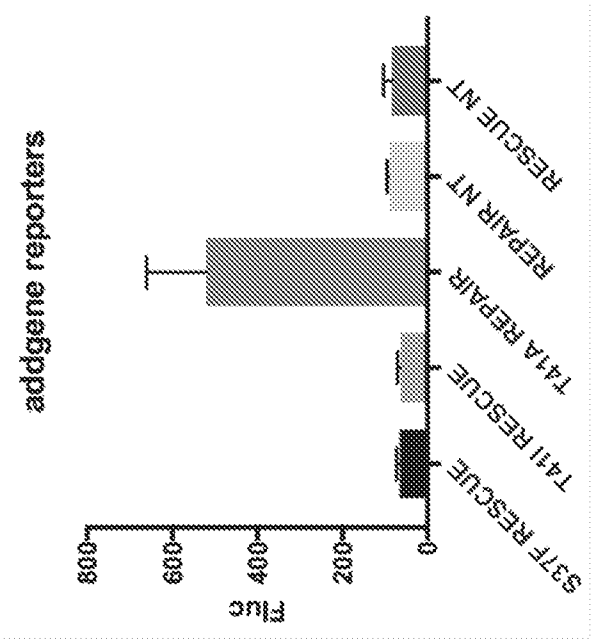


FIG. 137B

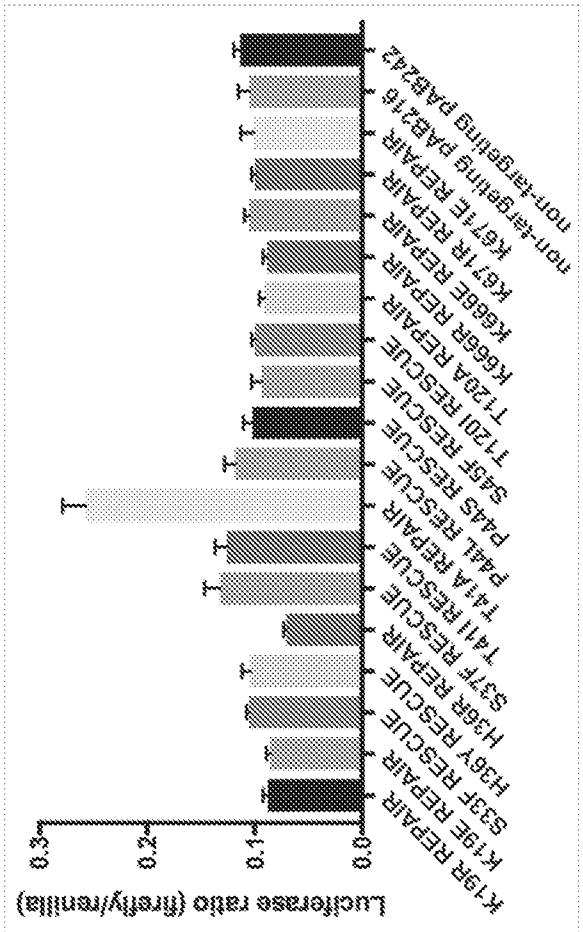


FIG. 137A

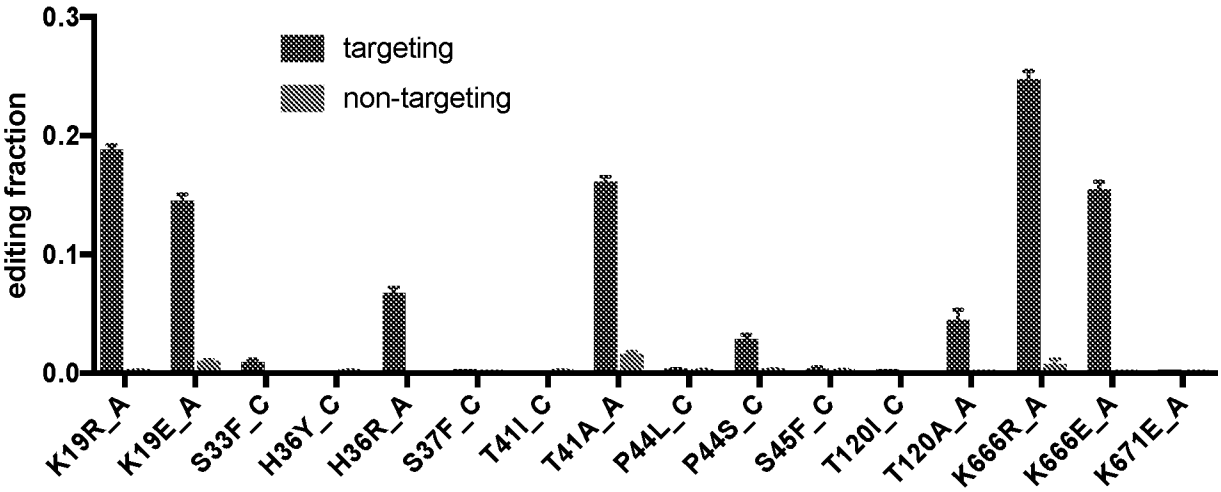

0.3
0.2
0.1
0.0
editing fraction
targeting
non-targeting
K19R_A
K19E_A
S33F_C
H36Y_C
H36R_A
S37F_C
T41I_C
T41A_A
P44L_C
P44S_C
S45F_C
T120I_C
T120A_A
K666R_A
K666E_A
K671E_A

FIG. 138

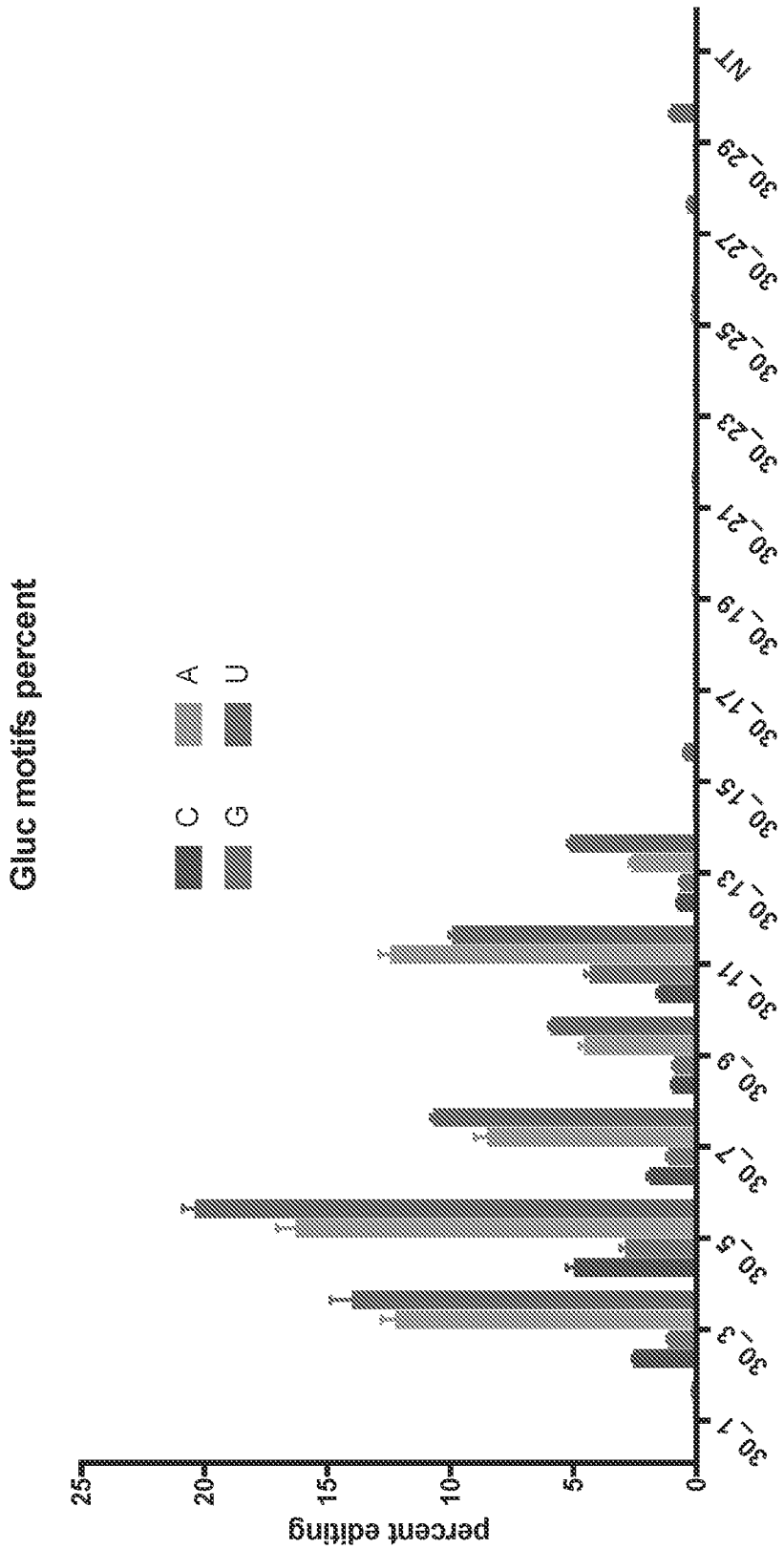

Gluc motifs percent
C
A
G
U
percent editing
0
5
10
15
20
25
30_1
30_3
30_5
30_7
30_9
30_11
30_13
30_15
30_17
30_19
30_21
30_23
30_25
30_27
30_29
NT

FIG. 139

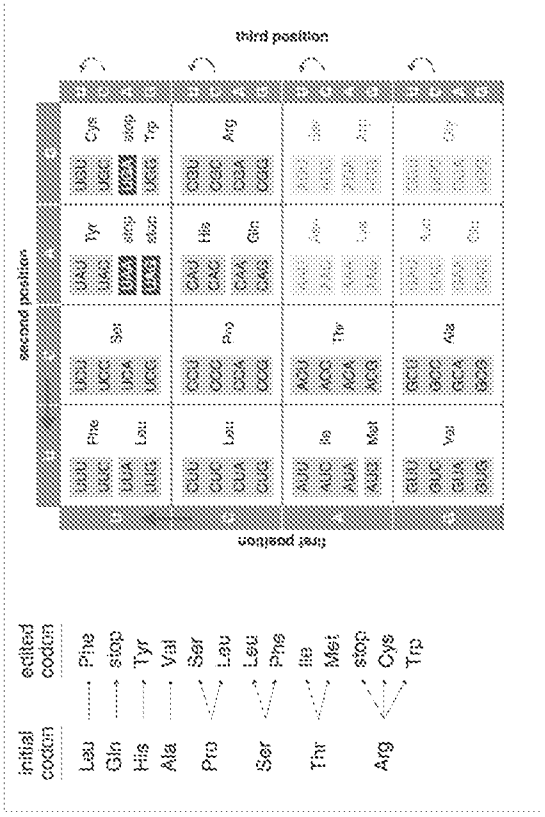


FIG. 140B

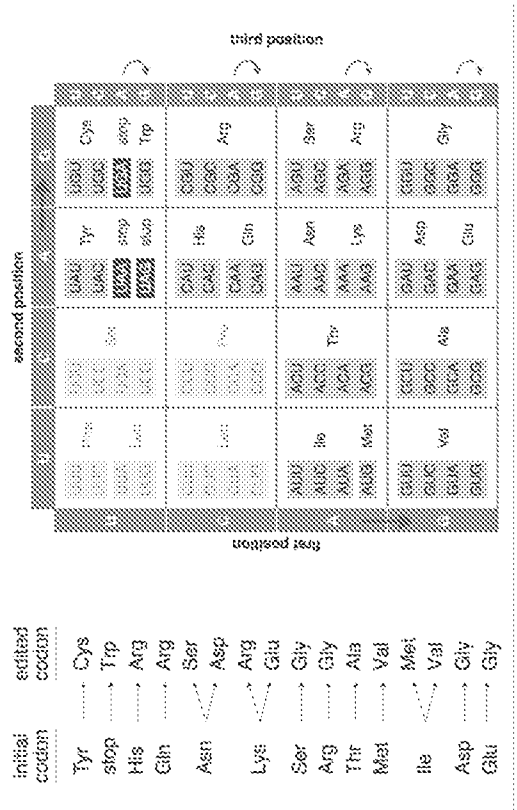


FIG. 140A

SEQUENCE LISTING

<110> The BROAD Institute, Inc.
Massachusetts Institute of Technology
The President and Fellows of Harvard College
Cox, David
Kaanan, Soumya
Gootenberg, Jonathan
Abudayyeh, Omar
Zhang, Feng

<120> Systems, Methods, and Compositions for Targeted Nucleic Acid Editing

<130> BROD-2174WP

<150> US 62/525,181
<151> 2017-06-26

<150> US 62/528,391
<151> 2017-07-03

<150> US 62/534,016
<151> 2017-07-18

<150> US 62/561,638
<151> 2017-09-21

<150> US 62/568,304
<151> 2017-10-04

<150> US 62/574,158
<151> 2017-10-18

<150> US 62/591,187
<151> 2017-11-27

<150> US 62/610,105
<151> 2017-12-22

<160> 918

<170> PatentIn version 3.5

<210> 1
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 1

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
1 5 10 15

<210> 2
<211> 20
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 2

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
1 5 10 15

Gly Gly Gly Ser
20

<210> 3
<211> 25
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 3

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Gly Ser
20 25

<210> 4
<211> 30
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 4

```
Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
1               5                   10                  15


Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
            20                  25                  30


<210>  5
<211>  35
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Synthetic or Unknown

<400>  5

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
1               5                   10                  15


Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
            20                  25                  30


Gly Gly Ser
        35


<210>  6
<211>  40
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Synthetic or Unknown

<400>  6

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
1               5                   10                  15


Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
            20                  25                  30


Gly Gly Ser Gly Gly Gly Gly Ser
        35                  40
```

<210> 7
<211> 45
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 7

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
20 25 30

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
35 40 45

<210> 8
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 8

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
1 5 10 15

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
20 25 30

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly
35 40 45

Gly Ser
50

<210> 9
<211> 55
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 9

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
1   5   10   15

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
            20   25   30

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly
        35   40   45

Gly Ser Gly Gly Gly Gly Ser
    50   55

<210> 10
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 10

Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly
1   5   10   15

<210> 11
<211> 32
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 11

Leu Glu Pro Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser
1   5   10   15

Phe Ser Gln Ser Gly Ala Leu Thr Arg His Gln Arg Thr His Thr Arg
            20   25   30

<210> 12
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 12

Gly Gly Ser Gly Gly Ser Gly Gly Ser
1   5

<210> 13
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 13

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
1   5   10   15

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
            20   25   30

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly
        35   40   45

Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
    50   55   60

<210> 14
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 14

Gly Gly Gly Gly Ser
1 5

<210> 15
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 15

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
1 5 10

<210> 16
<211> 20
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 16

Pro Lys Lys Lys Arg Lys Val Glu Ala Ser Ser Pro Lys Lys Arg Lys
1 5 10 15

Val Glu Ala Ser
20

<210> 17
<211> 7
<212> PRT
<213> SV40

<400> 17

Pro Lys Lys Lys Arg Lys Val
1 5

<210> 18
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 18

Pro Lys Lys Lys Arg Lys Val Glu Ala Ser
1   5   10

<210> 19
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 19

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys
1   5   10   15

<210> 20
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 20

Pro Ala Ala Lys Arg Val Lys Leu Asp
1   5

<210> 21
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 21

Arg Gln Arg Arg Asn Glu Leu Lys Arg Ser Pro
1   5   10

<210> 22

<211> 38
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 22

Asn Gln Ser Ser Asn Phe Gly Pro Met Lys Gly Gly Asn Phe Gly Gly
1 5 10 15

Arg Ser Ser Gly Pro Tyr Gly Gly Gly Gly Gln Tyr Phe Ala Lys Pro
20 25 30

Arg Asn Gln Gly Gly Tyr
35

<210> 23
<211> 42
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 23

Arg Met Arg Ile Glx Phe Lys Asn Lys Gly Lys Asp Thr Ala Glu Leu
1 5 10 15

Arg Arg Arg Arg Val Glu Val Ser Val Glu Leu Arg Lys Ala Lys Lys
20 25 30

Asp Glu Gln Ile Leu Lys Arg Arg Asn Val
35 40

<210> 24
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 24

Val Ser Arg Lys Arg Pro Arg Pro
1 5

<210> 25
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 25

Pro Pro Lys Lys Ala Arg Glu Asp
1 5

<210> 26
<211> 8
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (2)..(2)
<223> O at position 2 represents pyrrolysine

<400> 26

Pro Xaa Pro Lys Lys Lys Pro Leu
1 5

<210> 27
<211> 12
<212> PRT
<213> Mus

<400> 27

Ser Ala Leu Ile Lys Lys Lys Lys Lys Met Ala Pro
1 5 10

<210> 28
<211> 5
<212> PRT
<213> Influenza virus

<400> 28

Asp Arg Leu Arg Arg
1 5

<210> 29
<211> 7
<212> PRT
<213> Influenza virus

<400> 29

Pro Lys Gln Lys Lys Arg Lys
1 5

<210> 30
<211> 10
<212> PRT
<213> Hepatitis virus

<400> 30

Arg Lys Leu Lys Lys Lys Ile Lys Lys Leu
1 5 10

<210> 31
<211> 10
<212> PRT
<213> Mus

<400> 31

Arg Glu Lys Lys Lys Phe Leu Lys Arg Arg
1 5 10

<210> 32
<211> 20
<212> PRT
<213> Homo sapiens

<400> 32

Lys Arg Lys Gly Asp Glu Val Asp Gly Val Asp Glu Val Ala Lys Lys
1 5 10 15

Lys Ser Lys Lys
20

<210> 33
<211> 17
<212> PRT
<213> Homo sapiens

<400> 33

Arg Lys Cys Leu Gln Ala Gly Met Asn Leu Glu Ala Arg Lys Thr Lys
1 5 10 15

Lys

<210> 34
<211> 1340
<212> PRT
<213> Lachnospiracea sp

<400> 34

Met Gln Ile Ser Lys Val Asn His Lys His Val Ala Val Gly Gln Lys
1 5 10 15

Asp Arg Glu Arg Ile Thr Gly Phe Ile Tyr Asn Asp Pro Val Gly Asp
20 25 30

Glu Lys Ser Leu Glu Asp Val Val Ala Lys Arg Ala Asn Asp Thr Lys
35 40 45

Val Leu Phe Asn Val Phe Asn Thr Lys Asp Leu Tyr Asp Ser Gln Glu
50 55 60

Ser Asp Lys Ser Glu Lys Asp Lys Glu Ile Ile Ser Lys Gly Ala Lys
65 70 75 80

Phe Val Ala Lys Ser Phe Asn Ser Ala Ile Thr Ile Leu Lys Lys Gln
85 90 95

Asn Lys Ile Tyr Ser Thr Leu Thr Ser Gln Gln Val Ile Lys Glu Leu
100 105 110

```
Lys Asp Lys Phe Gly Gly Ala Arg Ile Tyr Asp Asp Asp Ile Glu Glu
        115                 120                 125


Ala Leu Thr Glu Thr Leu Lys Lys Ser Phe Arg Lys Glu Asn Val Arg
    130                 135                 140


Asn Ser Ile Lys Val Leu Ile Glu Asn Ala Ala Gly Ile Arg Ser Ser
145                 150                 155                 160


Leu Ser Lys Asp Glu Glu Glu Leu Ile Gln Glu Tyr Phe Val Lys Gln
                165                 170                 175


Leu Val Glu Glu Tyr Thr Lys Thr Lys Leu Gln Lys Asn Val Val Lys
            180                 185                 190


Ser Ile Lys Asn Gln Asn Met Val Ile Gln Pro Asp Ser Asp Ser Gln
        195                 200                 205


Val Leu Ser Leu Ser Glu Ser Arg Arg Glu Lys Gln Ser Ser Ala Val
    210                 215                 220


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


Phe Leu Thr Asp Tyr Ala Val Leu Asp Glu Asp Glu Arg Asn Ser Leu
                245                 250                 255


Leu Trp Lys Leu Arg Asn Leu Val Asn Leu Tyr Phe Tyr Gly Ser Glu
            260                 265                 270


Ser Ile Arg Asp Tyr Ser Tyr Thr Lys Glu Lys Ser Val Trp Lys Glu
        275                 280                 285


His Asp Glu Gln Lys Ala Asn Lys Thr Leu Phe Ile Asp Glu Ile Cys
    290                 295                 300


His Ile Thr Lys Ile Gly Lys Asn Gly Lys Glu Gln Lys Val Leu Asp
305                 310                 315                 320
```

```
Tyr Glu Glu Asn Arg Ser Arg Cys Arg Lys Gln Asn Ile Asn Tyr Tyr
                325                 330                 335

Arg Ser Ala Leu Asn Tyr Ala Lys Asn Asn Thr Ser Gly Ile Phe Glu
            340                 345                 350

Asn Glu Asp Ser Asn His Phe Trp Ile His Leu Ile Glu Asn Glu Val
        355                 360                 365

Glu Arg Leu Tyr Asn Gly Ile Glu Asn Gly Glu Glu Phe Lys Phe Glu
    370                 375                 380

Thr Gly Tyr Ile Ser Glu Lys Val Trp Lys Ala Val Ile Asn His Leu
385                 390                 395                 400

Ser Ile Lys Tyr Ile Ala Leu Gly Lys Ala Val Tyr Asn Tyr Ala Met
                405                 410                 415

Lys Glu Leu Ser Ser Pro Gly Asp Ile Glu Pro Gly Lys Ile Asp Asp
            420                 425                 430

Ser Tyr Ile Asn Gly Ile Thr Ser Phe Asp Tyr Glu Ile Ile Lys Ala
        435                 440                 445

Glu Glu Ser Leu Gln Arg Asp Ile Ser Met Asn Val Val Phe Ala Thr
    450                 455                 460

Asn Tyr Leu Ala Cys Ala Thr Val Asp Thr Asp Lys Asp Phe Leu Leu
465                 470                 475                 480

Phe Ser Lys Glu Asp Ile Arg Ser Cys Thr Lys Lys Asp Gly Asn Leu
                485                 490                 495

Cys Lys Asn Ile Met Gln Phe Trp Gly Gly Tyr Ser Thr Trp Lys Asn
            500                 505                 510

Phe Cys Glu Glu Tyr Leu Lys Asp Asp Lys Asp Ala Leu Glu Leu Leu
        515                 520                 525
```

```
Tyr Ser Leu Lys Ser Met Leu Tyr Ser Met Arg Asn Ser Ser Phe His
    530                 535                 540

Phe Ser Thr Glu Asn Val Asp Asn Gly Ser Trp Asp Thr Glu Leu Ile
545                 550                 555                 560

Gly Lys Leu Phe Glu Glu Asp Cys Asn Arg Ala Ala Arg Ile Glu Lys
                565                 570                 575

Glu Lys Phe Tyr Asn Asn Asn Leu His Met Phe Tyr Ser Ser Ser Leu
            580                 585                 590

Leu Glu Lys Val Leu Glu Arg Leu Tyr Ser Ser His His Glu Arg Ala
        595                 600                 605

Ser Gln Val Pro Ser Phe Asn Arg Val Phe Val Arg Lys Asn Phe Pro
    610                 615                 620

Ser Ser Leu Ser Glu Gln Arg Ile Thr Pro Lys Phe Thr Asp Ser Lys
625                 630                 635                 640

Asp Glu Gln Ile Trp Gln Ser Ala Val Tyr Tyr Leu Cys Lys Glu Ile
                645                 650                 655

Tyr Tyr Asn Asp Phe Leu Gln Ser Lys Glu Ala Tyr Lys Leu Phe Arg
            660                 665                 670

Glu Gly Val Lys Asn Leu Asp Lys Asn Asp Ile Asn Asn Gln Lys Ala
        675                 680                 685

Ala Asp Ser Phe Lys Gln Ala Val Val Tyr Tyr Gly Lys Ala Ile Gly
    690                 695                 700

Asn Ala Thr Leu Ser Gln Val Cys Gln Ala Ile Met Thr Glu Tyr Asn
705                 710                 715                 720

Arg Gln Asn Asn Asp Gly Leu Lys Lys Lys Ser Ala Tyr Ala Glu Lys
                725                 730                 735
```

```
Gln Asn Ser Asn Lys Tyr Lys His Tyr Pro Leu Phe Leu Lys Gln Val
            740                 745                 750

Leu Gln Ser Ala Phe Trp Glu Tyr Leu Asp Glu Asn Lys Glu Ile Tyr
        755                 760                 765

Gly Phe Ile Ser Ala Gln Ile His Lys Ser Asn Val Glu Ile Lys Ala
    770                 775                 780

Glu Asp Phe Ile Ala Asn Tyr Ser Ser Gln Gln Tyr Lys Lys Leu Val
785                 790                 795                 800

Asp Lys Val Lys Lys Thr Pro Glu Leu Gln Lys Trp Tyr Thr Leu Gly
                805                 810                 815

Arg Leu Ile Asn Pro Arg Gln Ala Asn Gln Phe Leu Gly Ser Ile Arg
            820                 825                 830

Asn Tyr Val Gln Phe Val Lys Asp Ile Gln Arg Arg Ala Lys Glu Asn
        835                 840                 845

Gly Asn Pro Ile Arg Asn Tyr Tyr Glu Val Leu Glu Ser Asp Ser Ile
    850                 855                 860

Ile Lys Ile Leu Glu Met Cys Thr Lys Leu Asn Gly Thr Thr Ser Asn
865                 870                 875                 880

Asp Ile His Asp Tyr Phe Arg Asp Glu Asp Glu Tyr Ala Glu Tyr Ile
                885                 890                 895

Ser Gln Phe Val Asn Phe Gly Asp Val His Ser Gly Ala Ala Leu Asn
            900                 905                 910

Ala Phe Cys Asn Ser Glu Ser Glu Gly Lys Lys Asn Gly Ile Tyr Tyr
        915                 920                 925

Asp Gly Ile Asn Pro Ile Val Asn Arg Asn Trp Val Leu Cys Lys Leu
    930                 935                 940
```

```
Tyr Gly Ser Pro Asp Leu Ile Ser Lys Ile Ile Ser Arg Val Asn Glu
945                 950                 955                 960


Asn Met Ile His Asp Phe His Lys Gln Glu Asp Leu Ile Arg Glu Tyr
                965                 970                 975


Gln Ile Lys Gly Ile Cys Ser Asn Lys Lys Glu Gln Gln Asp Leu Arg
            980                 985                 990


Thr Phe Gln Val Leu Lys Asn Arg  Val Glu Leu Arg Asp  Ile Val Glu
        995                 1000                1005


Tyr Ser  Glu Ile Ile Asn Glu  Leu Tyr Gly Gln Leu  Ile Lys Trp
    1010                1015                1020


Cys Tyr  Leu Arg Glu Arg Asp  Leu Met Tyr Phe Gln  Leu Gly Phe
    1025                1030                1035


His Tyr  Leu Cys Leu Asn Asn  Ala Ser Ser Lys Glu  Ala Asp Tyr
    1040                1045                1050


Ile Lys  Ile Asn Val Asp Asp  Arg Asn Ile Ser Gly  Ala Ile Leu
    1055                1060                1065


Tyr Gln  Ile Ala Ala Met Tyr  Ile Asn Gly Leu Pro  Val Tyr Tyr
    1070                1075                1080


Lys Lys  Asp Asp Met Tyr Val  Ala Leu Lys Ser Gly  Lys Lys Ala
    1085                1090                1095


Ser Asp  Glu Leu Asn Ser Asn  Glu Gln Thr Ser Lys  Lys Ile Asn
    1100                1105                1110


Tyr Phe  Leu Lys Tyr Gly Asn  Asn Ile Leu Gly Asp  Lys Lys Asp
    1115                1120                1125


Gln Leu  Tyr Leu Ala Gly Leu  Glu Leu Phe Glu Asn  Val Ala Glu
    1130                1135                1140
```

```
His Glu  Asn Ile Ile Ile Phe  Arg Asn Glu Ile Asp  His Phe His
    1145                 1150                 1155


Tyr Phe  Tyr Asp Arg Asp Arg  Ser Met Leu Asp Leu  Tyr Ser Glu
    1160                 1165                 1170


Val Phe  Asp Arg Phe Phe Thr  Tyr Asp Met Lys Leu  Arg Lys Asn
    1175                 1180                 1185


Val Val  Asn Met Leu Tyr Asn  Ile Leu Leu Asp His  Asn Ile Val
    1190                 1195                 1200


Ser Ser  Phe Val Phe Glu Thr  Gly Glu Lys Lys Val  Gly Arg Gly
    1205                 1210                 1215


Asp Ser  Glu Val Ile Lys Pro  Ser Ala Lys Ile Arg  Leu Arg Ala
    1220                 1225                 1230


Asn Asn  Gly Val Ser Ser Asp  Val Phe Thr Tyr Lys  Val Gly Ser
    1235                 1240                 1245


Lys Asp  Glu Leu Lys Ile Ala  Thr Leu Pro Ala Lys  Asn Glu Glu
    1250                 1255                 1260


Phe Leu  Leu Asn Val Ala Arg  Leu Ile Tyr Tyr Pro  Asp Met Glu
    1265                 1270                 1275


Ala Val  Ser Glu Asn Met Val  Arg Glu Gly Val Val  Lys Val Glu
    1280                 1285                 1290


Lys Ser  Asn Asp Lys Lys Gly  Lys Ile Ser Arg Gly  Ser Asn Thr
    1295                 1300                 1305


Arg Ser  Ser Asn Gln Ser Lys  Tyr Asn Asn Lys Ser  Lys Asn Arg
    1310                 1315                 1320


Met Asn  Tyr Ser Met Gly Ser  Ile Phe Glu Lys Met  Asp Leu Lys
    1325                 1330                 1335
```

Phe Asp
1340

<210> 35
<211> 1437
<212> PRT
<213> Lachnospiracea NK4A179

<400> 35

Met Lys Ile Ser Lys Val Arg Glu Glu Asn Arg Gly Ala Lys Leu Thr
1 5 10 15

Val Asn Ala Lys Thr Ala Val Val Ser Glu Asn Arg Ser Gln Glu Gly
20 25 30

Ile Leu Tyr Asn Asp Pro Ser Arg Tyr Gly Lys Ser Arg Lys Asn Asp
35 40 45

Glu Asp Arg Asp Arg Tyr Ile Glu Ser Arg Leu Lys Ser Ser Gly Lys
50 55 60

Leu Tyr Arg Ile Phe Asn Glu Asp Lys Asn Lys Arg Glu Thr Asp Glu
65 70 75 80

Leu Gln Trp Phe Leu Ser Glu Ile Val Lys Lys Ile Asn Arg Arg Asn
85 90 95

Gly Leu Val Leu Ser Asp Met Leu Ser Val Asp Asp Arg Ala Phe Glu
100 105 110

Lys Ala Phe Glu Lys Tyr Ala Glu Leu Ser Tyr Thr Asn Arg Arg Asn
115 120 125

Lys Val Ser Gly Ser Pro Ala Phe Glu Thr Cys Gly Val Asp Ala Ala
130 135 140

Thr Ala Glu Arg Leu Lys Gly Ile Ile Ser Glu Thr Asn Phe Ile Asn
145 150 155 160

Arg Ile Lys Asn Asn Ile Asp Asn Lys Val Ser Glu Asp Ile Ile Asp

```
                165                 170                 175

Arg Ile Ile Ala Lys Tyr Leu Lys Lys Ser Leu Cys Arg Glu Arg Val
            180                 185                 190

Lys Arg Gly Leu Lys Lys Leu Leu Met Asn Ala Phe Asp Leu Pro Tyr
        195                 200                 205

Ser Asp Pro Asp Ile Asp Val Gln Arg Asp Phe Ile Asp Tyr Val Leu
    210                 215                 220

Glu Asp Phe Tyr His Val Arg Ala Lys Ser Gln Val Ser Arg Ser Ile
225                 230                 235                 240

Lys Asn Met Asn Met Pro Val Gln Pro Glu Gly Asp Gly Lys Phe Ala
                245                 250                 255

Ile Thr Val Ser Lys Gly Gly Thr Glu Ser Gly Asn Lys Arg Ser Ala
            260                 265                 270

Glu Lys Glu Ala Phe Lys Lys Phe Leu Ser Asp Tyr Ala Ser Leu Asp
        275                 280                 285

Glu Arg Val Arg Asp Asp Met Leu Arg Arg Met Arg Arg Leu Val Val
    290                 295                 300

Leu Tyr Phe Tyr Gly Ser Asp Asp Ser Lys Leu Ser Asp Val Asn Glu
305                 310                 315                 320

Lys Phe Asp Val Trp Glu Asp His Ala Ala Arg Arg Val Asp Asn Arg
                325                 330                 335

Glu Phe Ile Lys Leu Pro Leu Glu Asn Lys Leu Ala Asn Gly Lys Thr
            340                 345                 350

Asp Lys Asp Ala Glu Arg Ile Arg Lys Asn Thr Val Lys Glu Leu Tyr
        355                 360                 365

Arg Asn Gln Asn Ile Gly Cys Tyr Arg Gln Ala Val Lys Ala Val Glu
```

```
    370                 375                 380

Glu Asp Asn Asn Gly Arg Tyr Phe Asp Asp Lys Met Leu Asn Met Phe
385                 390                 395                 400

Phe Ile His Arg Ile Glu Tyr Gly Val Glu Lys Ile Tyr Ala Asn Leu
                405                 410                 415

Lys Gln Val Thr Glu Phe Lys Ala Arg Thr Gly Tyr Leu Ser Glu Lys
            420                 425                 430

Ile Trp Lys Asp Leu Ile Asn Tyr Ile Ser Ile Lys Tyr Ile Ala Met
        435                 440                 445

Gly Lys Ala Val Tyr Asn Tyr Ala Met Asp Glu Leu Asn Ala Ser Asp
    450                 455                 460

Lys Lys Glu Ile Glu Leu Gly Lys Ile Ser Glu Glu Tyr Leu Ser Gly
465                 470                 475                 480

Ile Ser Ser Phe Asp Tyr Glu Leu Ile Lys Ala Glu Glu Met Leu Gln
                485                 490                 495

Arg Glu Thr Ala Val Tyr Val Ala Phe Ala Ala Arg His Leu Ser Ser
            500                 505                 510

Gln Thr Val Glu Leu Asp Ser Glu Asn Ser Asp Phe Leu Leu Leu Lys
        515                 520                 525

Pro Lys Gly Thr Met Asp Lys Asn Asp Lys Asn Lys Leu Ala Ser Asn
    530                 535                 540

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

Leu Gln Tyr Phe Gly Gly His Ser Leu Trp Thr Asp Phe Pro Phe Asp
                565                 570                 575

Lys Tyr Leu Ala Gly Gly Lys Asp Asp Val Asp Phe Leu Thr Asp Leu
```

```
            580                 585                 590

Lys Asp Val Ile Tyr Ser Met Arg Asn Asp Ser Phe His Tyr Ala Thr
        595                 600                 605

Glu Asn His Asn Asn Gly Lys Trp Asn Lys Glu Leu Ile Ser Ala Met
    610                 615                 620

Phe Glu His Glu Thr Glu Arg Met Thr Val Val Met Lys Asp Lys Phe
625                 630                 635                 640

Tyr Ser Asn Asn Leu Pro Met Phe Tyr Lys Asn Asp Asp Leu Lys Lys
                645                 650                 655

Leu Leu Ile Asp Leu Tyr Lys Asp Asn Val Glu Arg Ala Ser Gln Val
            660                 665                 670

Pro Ser Phe Asn Lys Val Phe Val Arg Lys Asn Phe Pro Ala Leu Val
        675                 680                 685

Arg Asp Lys Asp Asn Leu Gly Ile Glu Leu Asp Leu Lys Ala Asp Ala
    690                 695                 700

Asp Lys Gly Glu Asn Glu Leu Lys Phe Tyr Asn Ala Leu Tyr Tyr Met
705                 710                 715                 720

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

Glu Arg Phe Ile Thr Lys Ala Thr Lys Val Ala Asp Asn Tyr Asp Arg
            740                 745                 750

Asn Lys Glu Arg Asn Leu Lys Asp Arg Ile Lys Ser Ala Gly Ser Asp
        755                 760                 765

Glu Lys Lys Lys Leu Arg Glu Gln Leu Gln Asn Tyr Ile Ala Glu Asn
    770                 775                 780

Asp Phe Gly Gln Arg Ile Lys Asn Ile Val Gln Val Asn Pro Asp Tyr
```

785 790 795 800

Thr Leu Ala Gln Ile Cys Gln Leu Ile Met Thr Glu Tyr Asn Gln Gln
805 810 815

Asn Asn Gly Cys Met Gln Lys Lys Ser Ala Ala Arg Lys Asp Ile Asn
820 825 830

Lys Asp Ser Tyr Gln His Tyr Lys Met Leu Leu Leu Val Asn Leu Arg
835 840 845

Lys Ala Phe Leu Glu Phe Ile Lys Glu Asn Tyr Ala Phe Val Leu Lys
850 855 860

Pro Tyr Lys His Asp Leu Cys Asp Lys Ala Asp Phe Val Pro Asp Phe
865 870 875 880

Ala Lys Tyr Val Lys Pro Tyr Ala Gly Leu Ile Ser Arg Val Ala Gly
885 890 895

Ser Ser Glu Leu Gln Lys Trp Tyr Ile Val Ser Arg Phe Leu Ser Pro
900 905 910

Ala Gln Ala Asn His Met Leu Gly Phe Leu His Ser Tyr Lys Gln Tyr
915 920 925

Val Trp Asp Ile Tyr Arg Arg Ala Ser Glu Thr Gly Thr Glu Ile Asn
930 935 940

His Ser Ile Ala Glu Asp Lys Ile Ala Gly Val Asp Ile Thr Asp Val
945 950 955 960

Asp Ala Val Ile Asp Leu Ser Val Lys Leu Cys Gly Thr Ile Ser Ser
965 970 975

Glu Ile Ser Asp Tyr Phe Lys Asp Asp Glu Val Tyr Ala Glu Tyr Ile
980 985 990

Ser Ser Tyr Leu Asp Phe Glu Tyr Asp Gly Gly Asn Tyr Lys Asp Ser

```
        995                 1000                1005

Leu Asn  Arg Phe Cys Asn Ser  Asp Ala Val Asn Asp  Gln Lys Val
    1010                1015                1020

Ala Leu  Tyr Tyr Asp Gly Glu  His Pro Lys Leu Asn  Arg Asn Ile
    1025                1030                1035

Ile Leu  Ser Lys Leu Tyr Gly  Glu Arg Arg Phe Leu  Glu Lys Ile
    1040                1045                1050

Thr Asp  Arg Val Ser Arg Ser  Asp Ile Val Glu Tyr  Tyr Lys Leu
    1055                1060                1065

Lys Lys  Glu Thr Ser Gln Tyr  Gln Thr Lys Gly Ile  Phe Asp Ser
    1070                1075                1080

Glu Asp  Glu Gln Lys Asn Ile  Lys Lys Phe Gln Glu  Met Lys Asn
    1085                1090                1095

Ile Val  Glu Phe Arg Asp Leu  Met Asp Tyr Ser Glu  Ile Ala Asp
    1100                1105                1110

Glu Leu  Gln Gly Gln Leu Ile  Asn Trp Ile Tyr Leu  Arg Glu Arg
    1115                1120                1125

Asp Leu  Met Asn Phe Gln Leu  Gly Tyr His Tyr Ala  Cys Leu Asn
    1130                1135                1140

Asn Asp  Ser Asn Lys Gln Ala  Thr Tyr Val Thr Leu  Asp Tyr Gln
    1145                1150                1155

Gly Lys  Lys Asn Arg Lys Ile  Asn Gly Ala Ile Leu  Tyr Gln Ile
    1160                1165                1170

Cys Ala  Met Tyr Ile Asn Gly  Leu Pro Leu Tyr Tyr  Val Asp Lys
    1175                1180                1185

Asp Ser  Ser Glu Trp Thr Val  Ser Asp Gly Lys Glu  Ser Thr Gly
```

```
    1190                1195                1200

Ala Lys  Ile Gly Glu Phe Tyr  Arg Tyr Ala Lys Ser  Phe Glu Asn
    1205                1210                1215

Thr Ser  Asp Cys Tyr Ala Ser  Gly Leu Glu Ile Phe  Glu Asn Ile
    1220                1225                1230

Ser Glu  His Asp Asn Ile Thr  Glu Leu Arg Asn Tyr  Ile Glu His
    1235                1240                1245

Phe Arg  Tyr Tyr Ser Ser Phe  Asp Arg Ser Phe Leu  Gly Ile Tyr
    1250                1255                1260

Ser Glu  Val Phe Asp Arg Phe  Phe Thr Tyr Asp Leu  Lys Tyr Arg
    1265                1270                1275

Lys Asn  Val Pro Thr Ile Leu  Tyr Asn Ile Leu Leu  Gln His Phe
    1280                1285                1290

Val Asn  Val Arg Phe Glu Phe  Val Ser Gly Lys Lys  Met Ile Gly
    1295                1300                1305

Ile Asp  Lys Lys Asp Arg Lys  Ile Ala Lys Glu Lys  Glu Cys Ala
    1310                1315                1320

Arg Ile  Thr Ile Arg Glu Lys  Asn Gly Val Tyr Ser  Glu Gln Phe
    1325                1330                1335

Thr Tyr  Lys Leu Lys Asn Gly  Thr Val Tyr Val Asp  Ala Arg Asp
    1340                1345                1350

Lys Arg  Tyr Leu Gln Ser Ile  Ile Arg Leu Leu Phe  Tyr Pro Glu
    1355                1360                1365

Lys Val  Asn Met Asp Glu Met  Ile Glu Val Lys Glu  Lys Lys Lys
    1370                1375                1380

Pro Ser  Asp Asn Asn Thr Gly  Lys Gly Tyr Ser Lys  Arg Asp Arg
```

1385 1390 1395

Gln Gln Asp Arg Lys Glu Tyr Asp Lys Tyr Lys Glu Lys Lys Lys
1400 1405 1410

Lys Glu Gly Asn Phe Leu Ser Gly Met Gly Gly Asn Ile Asn Trp
1415 1420 1425

Asp Glu Ile Asn Ala Gln Leu Lys Asn
1430 1435

<210> 36
<211> 1385
<212> PRT
<213> Clostridium aminophilum

<400> 36

Met Lys Phe Ser Lys Val Asp His Thr Arg Ser Ala Val Gly Ile Gln
1 5 10 15

Lys Ala Thr Asp Ser Val His Gly Met Leu Tyr Thr Asp Pro Lys Lys
20 25 30

Gln Glu Val Asn Asp Leu Asp Lys Arg Phe Asp Gln Leu Asn Val Lys
35 40 45

Ala Lys Arg Leu Tyr Asn Val Phe Asn Gln Ser Lys Ala Glu Glu Asp
50 55 60

Asp Asp Glu Lys Arg Phe Gly Lys Val Val Lys Lys Leu Asn Arg Glu
65 70 75 80

Leu Lys Asp Leu Leu Phe His Arg Glu Val Ser Arg Tyr Asn Ser Ile
85 90 95

Gly Asn Ala Lys Tyr Asn Tyr Tyr Gly Ile Lys Ser Asn Pro Glu Glu
100 105 110

Ile Val Ser Asn Leu Gly Met Val Glu Ser Leu Lys Gly Glu Arg Asp
115 120 125

```
Pro Gln Lys Val Ile Ser Lys Leu Leu Leu Tyr Tyr Leu Arg Lys Gly
    130                 135                 140

Leu Lys Pro Gly Thr Asp Gly Leu Arg Met Ile Leu Glu Ala Ser Cys
145                 150                 155                 160

Gly Leu Arg Lys Leu Ser Gly Asp Glu Lys Glu Leu Lys Val Phe Leu
                165                 170                 175

Gln Thr Leu Asp Glu Asp Phe Glu Lys Lys Thr Phe Lys Lys Asn Leu
            180                 185                 190

Ile Arg Ser Ile Glu Asn Gln Asn Met Ala Val Gln Pro Ser Asn Glu
        195                 200                 205

Gly Asp Pro Ile Ile Gly Ile Thr Gln Gly Arg Phe Asn Ser Gln Lys
    210                 215                 220

Asn Glu Glu Lys Ser Ala Ile Glu Arg Met Met Ser Met Tyr Ala Asp
225                 230                 235                 240

Leu Asn Glu Asp His Arg Glu Asp Val Leu Arg Lys Leu Arg Arg Leu
                245                 250                 255

Asn Val Leu Tyr Phe Asn Val Asp Thr Glu Lys Thr Glu Glu Pro Thr
            260                 265                 270

Leu Pro Gly Glu Val Asp Thr Asn Pro Val Phe Glu Val Trp His Asp
        275                 280                 285

His Glu Lys Gly Lys Glu Asn Asp Arg Gln Phe Ala Thr Phe Ala Lys
    290                 295                 300

Ile Leu Thr Glu Asp Arg Glu Thr Arg Lys Lys Glu Lys Leu Ala Val
305                 310                 315                 320

Lys Glu Ala Leu Asn Asp Leu Lys Ser Ala Ile Arg Asp His Asn Ile
                325                 330                 335
```

```
Met Ala Tyr Arg Cys Ser Ile Lys Val Thr Glu Gln Asp Lys Asp Gly
            340                 345                 350

Leu Phe Phe Glu Asp Gln Arg Ile Asn Arg Phe Trp Ile His His Ile
        355                 360                 365

Glu Ser Ala Val Glu Arg Ile Leu Ala Ser Ile Asn Pro Glu Lys Leu
    370                 375                 380

Tyr Lys Leu Arg Ile Gly Tyr Leu Gly Glu Lys Val Trp Lys Asp Leu
385                 390                 395                 400

Leu Asn Tyr Leu Ser Ile Lys Tyr Ile Ala Val Gly Lys Ala Val Phe
                405                 410                 415

His Phe Ala Met Glu Asp Leu Gly Lys Thr Gly Gln Asp Ile Glu Leu
            420                 425                 430

Gly Lys Leu Ser Asn Ser Val Ser Gly Gly Leu Thr Ser Phe Asp Tyr
        435                 440                 445

Glu Gln Ile Arg Ala Asp Glu Thr Leu Gln Arg Gln Leu Ser Val Glu
    450                 455                 460

Val Ala Phe Ala Ala Asn Asn Leu Phe Arg Ala Val Val Gly Gln Thr
465                 470                 475                 480

Gly Lys Lys Ile Glu Gln Ser Lys Ser Glu Glu Asn Glu Glu Asp Phe
                485                 490                 495

Leu Leu Trp Lys Ala Glu Lys Ile Ala Glu Ser Ile Lys Lys Glu Gly
            500                 505                 510

Glu Gly Asn Thr Leu Lys Ser Ile Leu Gln Phe Phe Gly Gly Ala Ser
        515                 520                 525

Ser Trp Asp Leu Asn His Phe Cys Ala Ala Tyr Gly Asn Glu Ser Ser
    530                 535                 540
```

```
Ala Leu Gly Tyr Glu Thr Lys Phe Ala Asp Asp Leu Arg Lys Ala Ile
545                 550                 555                 560

Tyr Ser Leu Arg Asn Glu Thr Phe His Phe Thr Thr Leu Asn Lys Gly
                565                 570                 575

Ser Phe Asp Trp Asn Ala Lys Leu Ile Gly Asp Met Phe Ser His Glu
            580                 585                 590

Ala Ala Thr Gly Ile Ala Val Glu Arg Thr Arg Phe Tyr Ser Asn Asn
        595                 600                 605

Leu Pro Met Phe Tyr Arg Glu Ser Asp Leu Lys Arg Ile Met Asp His
    610                 615                 620

Leu Tyr Asn Thr Tyr His Pro Arg Ala Ser Gln Val Pro Ser Phe Asn
625                 630                 635                 640

Ser Val Phe Val Arg Lys Asn Phe Arg Leu Phe Leu Ser Asn Thr Leu
                645                 650                 655

Asn Thr Asn Thr Ser Phe Asp Thr Glu Val Tyr Gln Lys Trp Glu Ser
            660                 665                 670

Gly Val Tyr Tyr Leu Phe Lys Glu Ile Tyr Tyr Asn Ser Phe Leu Pro
        675                 680                 685

Ser Gly Asp Ala His His Leu Phe Phe Glu Gly Leu Arg Arg Ile Arg
    690                 695                 700

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

Asn Ala Val Gln Asp Phe Gly Arg Arg Cys Asp Glu Leu Lys Asn Leu
                725                 730                 735

Ser Leu Ser Ala Ile Cys Gln Met Ile Met Thr Glu Tyr Asn Glu Gln
            740                 745                 750
```

```
Asn Asn Gly Asn Arg Lys Val Lys Ser Thr Arg Glu Asp Lys Arg Lys
        755                 760                 765

Pro Asp Ile Phe Gln His Tyr Lys Met Leu Leu Leu Arg Thr Leu Gln
    770                 775                 780

Glu Ala Phe Ala Ile Tyr Ile Arg Arg Glu Glu Phe Lys Phe Ile Phe
785                 790                 795                 800

Asp Leu Pro Lys Thr Leu Tyr Val Met Lys Pro Val Glu Glu Phe Leu
                805                 810                 815

Pro Asn Trp Lys Ser Gly Met Phe Asp Ser Leu Val Glu Arg Val Lys
            820                 825                 830

Gln Ser Pro Asp Leu Gln Arg Trp Tyr Val Leu Cys Lys Phe Leu Asn
        835                 840                 845

Gly Arg Leu Leu Asn Gln Leu Ser Gly Val Ile Arg Ser Tyr Ile Gln
    850                 855                 860

Phe Ala Gly Asp Ile Gln Arg Arg Ala Lys Ala Asn His Asn Arg Leu
865                 870                 875                 880

Tyr Met Asp Asn Thr Gln Arg Val Glu Tyr Tyr Ser Asn Val Leu Glu
                885                 890                 895

Val Val Asp Phe Cys Ile Lys Gly Thr Ser Arg Phe Ser Asn Val Phe
            900                 905                 910

Ser Asp Tyr Phe Arg Asp Glu Asp Ala Tyr Ala Asp Tyr Leu Asp Asn
        915                 920                 925

Tyr Leu Gln Phe Lys Asp Glu Lys Ile Ala Glu Val Ser Ser Phe Ala
    930                 935                 940

Ala Leu Lys Thr Phe Cys Asn Glu Glu Glu Val Lys Ala Gly Ile Tyr
945                 950                 955                 960
```

```
Met Asp Gly Glu Asn Pro Val Met Gln Arg Asn Ile Val Met Ala Lys
               965                 970                 975

Leu Phe Gly Pro Asp Glu Val Leu Lys Asn Val Val Pro Lys Val Thr
           980                 985                 990

Arg Glu Glu Ile Glu Glu Tyr Tyr  Gln Leu Glu Lys Gln  Ile Ala Pro
        995                 1000                1005

Tyr Arg  Gln Asn Gly Tyr Cys  Lys Ser Glu Glu Asp  Gln Lys Lys
    1010                1015                1020

Leu Leu  Arg Phe Gln Arg Ile  Lys Asn Arg Val Glu  Phe Gln Thr
    1025                1030                1035

Ile Thr  Glu Phe Ser Glu Ile  Ile Asn Glu Leu Leu  Gly Gln Leu
    1040                1045                1050

Ile Ser  Trp Ser Phe Leu Arg  Glu Arg Asp Leu Leu  Tyr Phe Gln
    1055                1060                1065

Leu Gly  Phe His Tyr Leu Cys  Leu His Asn Asp Thr  Glu Lys Pro
    1070                1075                1080

Ala Glu  Tyr Lys Glu Ile Ser  Arg Glu Asp Gly Thr  Val Ile Arg
    1085                1090                1095

Asn Ala  Ile Leu His Gln Val  Ala Ala Met Tyr Val  Gly Gly Leu
    1100                1105                1110

Pro Val  Tyr Thr Leu Ala Asp  Lys Lys Leu Ala Ala  Phe Glu Lys
    1115                1120                1125

Gly Glu  Ala Asp Cys Lys Leu  Ser Ile Ser Lys Asp  Thr Ala Gly
    1130                1135                1140

Ala Gly  Lys Lys Ile Lys Asp  Phe Phe Arg Tyr Ser  Lys Tyr Val
    1145                1150                1155
```

```
Leu Ile  Lys Asp Arg Met Leu  Thr Asp Gln Asn Gln  Lys Tyr Thr
    1160                1165                1170

Ile Tyr  Leu Ala Gly Leu Glu  Leu Phe Glu Asn Thr  Asp Glu His
    1175                1180                1185

Asp Asn  Ile Thr Asp Val Arg  Lys Tyr Val Asp His  Phe Lys Tyr
    1190                1195                1200

Tyr Ala  Thr Ser Asp Glu Asn  Ala Met Ser Ile Leu  Asp Leu Tyr
    1205                1210                1215

Ser Glu  Ile His Asp Arg Phe  Phe Thr Tyr Asp Met  Lys Tyr Gln
    1220                1225                1230

Lys Asn  Val Ala Asn Met Leu  Glu Asn Ile Leu Leu  Arg His Phe
    1235                1240                1245

Val Leu  Ile Arg Pro Glu Phe  Phe Thr Gly Ser Lys  Lys Val Gly
    1250                1255                1260

Glu Gly  Lys Lys Ile Thr Cys  Lys Ala Arg Ala Gln  Ile Glu Ile
    1265                1270                1275

Ala Glu  Asn Gly Met Arg Ser  Glu Asp Phe Thr Tyr  Lys Leu Ser
    1280                1285                1290

Asp Gly  Lys Lys Asn Ile Ser  Thr Cys Met Ile Ala  Ala Arg Asp
    1295                1300                1305

Gln Lys  Tyr Leu Asn Thr Val  Ala Arg Leu Leu Tyr  Tyr Pro His
    1310                1315                1320

Glu Ala  Lys Lys Ser Ile Val  Asp Thr Arg Glu Lys  Lys Asn Asn
    1325                1330                1335

Lys Lys  Thr Asn Arg Gly Asp  Gly Thr Phe Asn Lys  Gln Lys Gly
    1340                1345                1350
```

```
Thr Ala  Arg Lys Glu Lys Asp  Asn Gly Pro Arg Glu  Phe Asn Asp
    1355                1360                1365


Thr Gly  Phe Ser Asn Thr Pro  Phe Ala Gly Phe Asp  Pro Phe Arg
    1370                1375                1380


Asn Ser
    1385


<210>  37
<211>  1175
<212>  PRT
<213>  Carnobacterium gallinarum

<400>  37

Met Arg Ile Thr Lys Val Lys Ile Lys Leu Asp Asn Lys Leu Tyr Gln
1               5                   10                  15


Val Thr Met Gln Lys Glu Glu Lys Tyr Gly Thr Leu Lys Leu Asn Glu
            20                  25                  30


Glu Ser Arg Lys Ser Thr Ala Glu Ile Leu Arg Leu Lys Lys Ala Ser
        35                  40                  45


Phe Asn Lys Ser Phe His Ser Lys Thr Ile Asn Ser Gln Lys Glu Asn
    50                  55                  60


Lys Asn Ala Thr Ile Lys Lys Asn Gly Asp Tyr Ile Ser Gln Ile Phe
65                  70                  75                  80


Glu Lys Leu Val Gly Val Asp Thr Asn Lys Asn Ile Arg Lys Pro Lys
                85                  90                  95


Met Ser Leu Thr Asp Leu Lys Asp Leu Pro Lys Lys Asp Leu Ala Leu
            100                 105                 110


Phe Ile Lys Arg Lys Phe Lys Asn Asp Asp Ile Val Glu Ile Lys Asn
        115                 120                 125
```

```
Leu Asp Leu Ile Ser Leu Phe Tyr Asn Ala Leu Gln Lys Val Pro Gly
    130                 135                 140

Glu His Phe Thr Asp Glu Ser Trp Ala Asp Phe Cys Gln Glu Met Met
145                 150                 155                 160

Pro Tyr Arg Glu Tyr Lys Asn Lys Phe Ile Glu Arg Lys Ile Ile Leu
                165                 170                 175

Leu Ala Asn Ser Ile Glu Gln Asn Lys Gly Phe Ser Ile Asn Pro Glu
            180                 185                 190

Thr Phe Ser Lys Arg Lys Arg Val Leu His Gln Trp Ala Ile Glu Val
        195                 200                 205

Gln Glu Arg Gly Asp Phe Ser Ile Leu Asp Glu Lys Leu Ser Lys Leu
    210                 215                 220

Ala Glu Ile Tyr Asn Phe Lys Lys Met Cys Lys Arg Val Gln Asp Glu
225                 230                 235                 240

Leu Asn Asp Leu Glu Lys Ser Met Lys Lys Gly Lys Asn Pro Glu Lys
                245                 250                 255

Glu Lys Glu Ala Tyr Lys Lys Gln Lys Asn Phe Lys Ile Lys Thr Ile
            260                 265                 270

Trp Lys Asp Tyr Pro Tyr Lys Thr His Ile Gly Leu Ile Glu Lys Ile
        275                 280                 285

Lys Glu Asn Glu Glu Leu Asn Gln Phe Asn Ile Glu Ile Gly Lys Tyr
    290                 295                 300

Phe Glu His Tyr Phe Pro Ile Lys Lys Glu Arg Cys Thr Glu Asp Glu
305                 310                 315                 320

Pro Tyr Tyr Leu Asn Ser Glu Thr Ile Ala Thr Thr Val Asn Tyr Gln
                325                 330                 335
```

```
Leu Lys Asn Ala Leu Ile Ser Tyr Leu Met Gln Ile Gly Lys Tyr Lys
            340                 345                 350

Gln Phe Gly Leu Glu Asn Gln Val Leu Asp Ser Lys Lys Leu Gln Glu
        355                 360                 365

Ile Gly Ile Tyr Glu Gly Phe Gln Thr Lys Phe Met Asp Ala Cys Val
    370                 375                 380

Phe Ala Thr Ser Ser Leu Lys Asn Ile Ile Glu Pro Met Arg Ser Gly
385                 390                 395                 400

Asp Ile Leu Gly Lys Arg Glu Phe Lys Glu Ala Ile Ala Thr Ser Ser
                405                 410                 415

Phe Val Asn Tyr His His Phe Phe Pro Tyr Phe Pro Phe Glu Leu Lys
            420                 425                 430

Gly Met Lys Asp Arg Glu Ser Glu Leu Ile Pro Phe Gly Glu Gln Thr
        435                 440                 445

Glu Ala Lys Gln Met Gln Asn Ile Trp Ala Leu Arg Gly Ser Val Gln
    450                 455                 460

Gln Ile Arg Asn Glu Ile Phe His Ser Phe Asp Lys Asn Gln Lys Phe
465                 470                 475                 480

Asn Leu Pro Gln Leu Asp Lys Ser Asn Phe Glu Phe Asp Ala Ser Glu
                485                 490                 495

Asn Ser Thr Gly Lys Ser Gln Ser Tyr Ile Glu Thr Asp Tyr Lys Phe
            500                 505                 510

Leu Phe Glu Ala Glu Lys Asn Gln Leu Glu Gln Phe Phe Ile Glu Arg
        515                 520                 525

Ile Lys Ser Ser Gly Ala Leu Glu Tyr Tyr Pro Leu Lys Ser Leu Glu
    530                 535                 540
```

```
Lys Leu Phe Ala Lys Lys Glu Met Lys Phe Ser Leu Gly Ser Gln Val
545                 550                 555                 560


Val Ala Phe Ala Pro Ser Tyr Lys Lys Leu Val Lys Lys Gly His Ser
                565                 570                 575


Tyr Gln Thr Ala Thr Glu Gly Thr Ala Asn Tyr Leu Gly Leu Ser Tyr
            580                 585                 590


Tyr Asn Arg Tyr Glu Leu Lys Glu Glu Ser Phe Gln Ala Gln Tyr Tyr
        595                 600                 605


Leu Leu Lys Leu Ile Tyr Gln Tyr Val Phe Leu Pro Asn Phe Ser Gln
    610                 615                 620


Gly Asn Ser Pro Ala Phe Arg Glu Thr Val Lys Ala Ile Leu Arg Ile
625                 630                 635                 640


Asn Lys Asp Glu Ala Arg Lys Lys Met Lys Lys Asn Lys Lys Phe Leu
                645                 650                 655


Arg Lys Tyr Ala Phe Glu Gln Val Arg Glu Met Glu Phe Lys Glu Thr
            660                 665                 670


Pro Asp Gln Tyr Met Ser Tyr Leu Gln Ser Glu Met Arg Glu Glu Lys
        675                 680                 685


Val Arg Lys Ala Glu Lys Asn Asp Lys Gly Phe Glu Lys Asn Ile Thr
    690                 695                 700


Met Asn Phe Glu Lys Leu Leu Met Gln Ile Phe Val Lys Gly Phe Asp
705                 710                 715                 720


Val Phe Leu Thr Thr Phe Ala Gly Lys Glu Leu Leu Leu Ser Ser Glu
                725                 730                 735


Glu Lys Val Ile Lys Glu Thr Glu Ile Ser Leu Ser Lys Lys Ile Asn
            740                 745                 750
```

```
Glu Arg Glu Lys Thr Leu Lys Ala Ser Ile Gln Val Glu His Gln Leu
        755                 760                 765

Val Ala Thr Asn Ser Ala Ile Ser Tyr Trp Leu Phe Cys Lys Leu Leu
    770                 775                 780

Asp Ser Arg His Leu Asn Glu Leu Arg Asn Glu Met Ile Lys Phe Lys
785                 790                 795                 800

Gln Ser Arg Ile Lys Phe Asn His Thr Gln His Ala Glu Leu Ile Gln
                805                 810                 815

Asn Leu Leu Pro Ile Val Glu Leu Thr Ile Leu Ser Asn Asp Tyr Asp
            820                 825                 830

Glu Lys Asn Asp Ser Gln Asn Val Asp Val Ser Ala Tyr Phe Glu Asp
        835                 840                 845

Lys Ser Leu Tyr Glu Thr Ala Pro Tyr Val Gln Thr Asp Asp Arg Thr
    850                 855                 860

Arg Val Ser Phe Arg Pro Ile Leu Lys Leu Glu Lys Tyr His Thr Lys
865                 870                 875                 880

Ser Leu Ile Glu Ala Leu Leu Lys Asp Asn Pro Gln Phe Arg Val Ala
                885                 890                 895

Ala Thr Asp Ile Gln Glu Trp Met His Lys Arg Glu Glu Ile Gly Glu
            900                 905                 910

Leu Val Glu Lys Arg Lys Asn Leu His Thr Glu Trp Ala Glu Gly Gln
        915                 920                 925

Gln Thr Leu Gly Ala Glu Lys Arg Glu Glu Tyr Arg Asp Tyr Cys Lys
    930                 935                 940

Lys Ile Asp Arg Phe Asn Trp Lys Ala Asn Lys Val Thr Leu Thr Tyr
945                 950                 955                 960
```

```
Leu Ser Gln Leu His Tyr Leu Ile Thr Asp Leu Leu Gly Arg Met Val
               965                 970                 975


Gly Phe Ser Ala Leu Phe Glu Arg Asp Leu Val Tyr Phe Ser Arg Ser
           980                 985                 990


Phe Ser Glu Leu Gly Gly Glu Thr  Tyr His Ile Ser Asp  Tyr Lys Asn
        995                 1000                1005


Leu Ser  Gly Val Leu Arg Leu  Asn Ala Glu Val Lys  Pro Ile Lys
    1010                1015                1020


Ile Lys  Asn Ile Lys Val Ile  Asp Asn Glu Glu Asn  Pro Tyr Lys
    1025                1030                1035


Gly Asn  Glu Pro Glu Val Lys  Pro Phe Leu Asp Arg  Leu His Ala
    1040                1045                1050


Tyr Leu  Glu Asn Val Ile Gly  Ile Lys Ala Val His  Gly Lys Ile
    1055                1060                1065


Arg Asn  Gln Thr Ala His Leu  Ser Val Leu Gln Leu  Glu Leu Ser
    1070                1075                1080


Met Ile  Glu Ser Met Asn Asn  Leu Arg Asp Leu Met  Ala Tyr Asp
    1085                1090                1095


Arg Lys  Leu Lys Asn Ala Val  Thr Lys Ser Met Ile  Lys Ile Leu
    1100                1105                1110


Asp Lys  His Gly Met Ile Leu  Lys Leu Lys Ile Asp  Glu Asn His
    1115                1120                1125


Lys Asn  Phe Glu Ile Glu Ser  Leu Ile Pro Lys Glu  Ile Ile His
    1130                1135                1140


Leu Lys  Asp Lys Ala Ile Lys  Thr Asn Gln Val Ser  Glu Glu Tyr
    1145                1150                1155
```

```
Cys Gln  Leu Val Leu Ala Leu  Leu Thr Thr Asn Pro  Gly Asn Gln
    1160                1165                1170


Leu Asn
    1175


<210>  38
<211>  1164
<212>  PRT
<213>  Carnobacterium gallinarum

<400>  38

Met Arg Met Thr Lys Val Lys Ile Asn Gly Ser Pro Val Ser Met Asn
1               5                   10                  15


Arg Ser Lys Leu Asn Gly His Leu Val Trp Asn Gly Thr Thr Asn Thr
            20                  25                  30


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


Asn Lys Thr Leu Val Lys Ala Asp Gln Val Lys Gly Tyr Lys Val Leu
    50                  55                  60


Ala Glu Asn Ile Phe Ile Ile Phe Glu Gln Leu Glu Lys Ser Asn Ser
65                  70                  75                  80


Glu Lys Pro Ser Val Tyr Leu Asn Asn Ile Arg Arg Leu Lys Glu Ala
                85                  90                  95


Gly Leu Lys Arg Phe Phe Lys Ser Lys Tyr His Glu Glu Ile Lys Tyr
            100                 105                 110


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


Leu Phe Phe Asn Ala Val Asp Arg Ile Gln Glu Asp Lys Phe Asp Glu
    130                 135                 140
```

```
Lys Asn Trp Ser Tyr Phe Cys Lys Glu Met Ser Pro Tyr Leu Asp Tyr
145                 150                 155                 160


Lys Lys Ser Tyr Leu Asn Arg Lys Lys Glu Ile Leu Ala Asn Ser Ile
                165                 170                 175


Gln Gln Asn Arg Gly Phe Ser Met Pro Thr Ala Glu Glu Pro Asn Leu
            180                 185                 190


Leu Ser Lys Arg Lys Gln Leu Phe Gln Gln Trp Ala Met Lys Phe Gln
        195                 200                 205


Glu Ser Pro Leu Ile Gln Gln Asn Asn Phe Ala Val Glu Gln Phe Asn
    210                 215                 220


Lys Glu Phe Ala Asn Lys Ile Asn Glu Leu Ala Ala Val Tyr Asn Val
225                 230                 235                 240


Asp Glu Leu Cys Thr Ala Ile Thr Glu Lys Leu Met Asn Phe Asp Lys
                245                 250                 255


Asp Lys Ser Asn Lys Thr Arg Asn Phe Glu Ile Lys Lys Leu Trp Lys
            260                 265                 270


Gln His Pro His Asn Lys Asp Lys Ala Leu Ile Lys Leu Phe Asn Gln
        275                 280                 285


Glu Gly Asn Glu Ala Leu Asn Gln Phe Asn Ile Glu Leu Gly Lys Tyr
    290                 295                 300


Phe Glu His Tyr Phe Pro Lys Thr Gly Lys Lys Glu Ser Ala Glu Ser
305                 310                 315                 320


Tyr Tyr Leu Asn Pro Gln Thr Ile Ile Lys Thr Val Gly Tyr Gln Leu
                325                 330                 335


Arg Asn Ala Phe Val Gln Tyr Leu Leu Gln Val Gly Lys Leu His Gln
            340                 345                 350
```

```
Tyr Asn Lys Gly Val Leu Asp Ser Gln Thr Leu Gln Glu Ile Gly Met
        355                 360                 365


Tyr Glu Gly Phe Gln Thr Lys Phe Met Asp Ala Cys Val Phe Ala Ser
    370                 375                 380


Ser Ser Leu Arg Asn Ile Ile Gln Ala Thr Thr Asn Glu Asp Ile Leu
385                 390                 395                 400


Thr Arg Glu Lys Phe Lys Lys Glu Leu Glu Lys Asn Val Glu Leu Lys
                405                 410                 415


His Asp Leu Phe Phe Lys Thr Glu Ile Val Glu Glu Arg Asp Glu Asn
            420                 425                 430


Pro Ala Lys Lys Ile Ala Met Thr Pro Asn Glu Leu Asp Leu Trp Ala
        435                 440                 445


Ile Arg Gly Ala Val Gln Arg Val Arg Asn Gln Ile Phe His Gln Gln
    450                 455                 460


Ile Asn Lys Arg His Glu Pro Asn Gln Leu Lys Val Gly Ser Phe Glu
465                 470                 475                 480


Asn Gly Asp Leu Gly Asn Val Ser Tyr Gln Lys Thr Ile Tyr Gln Lys
                485                 490                 495


Leu Phe Asp Ala Glu Ile Lys Asp Ile Glu Ile Tyr Phe Ala Glu Lys
            500                 505                 510


Ile Lys Ser Ser Gly Ala Leu Glu Gln Tyr Ser Met Lys Asp Leu Glu
        515                 520                 525


Lys Leu Phe Ser Asn Lys Glu Leu Thr Leu Ser Leu Gly Gly Gln Val
    530                 535                 540


Val Ala Phe Ala Pro Ser Tyr Lys Lys Leu Tyr Lys Gln Gly Tyr Phe
545                 550                 555                 560
```

```
Tyr Gln Asn Glu Lys Thr Ile Glu Leu Glu Gln Phe Thr Asp Tyr Asp
                565                 570                 575

Phe Ser Asn Asp Val Phe Lys Ala Asn Tyr Tyr Leu Ile Lys Leu Ile
            580                 585                 590

Tyr His Tyr Val Phe Leu Pro Gln Phe Ser Gln Ala Asn Asn Lys Leu
        595                 600                 605

Phe Lys Asp Thr Val His Tyr Val Ile Gln Gln Asn Lys Glu Leu Asn
    610                 615                 620

Thr Thr Glu Lys Asp Lys Lys Asn Asn Lys Lys Ile Arg Lys Tyr Ala
625                 630                 635                 640

Phe Glu Gln Val Lys Leu Met Lys Asn Glu Ser Pro Glu Lys Tyr Met
                645                 650                 655

Gln Tyr Leu Gln Arg Glu Met Gln Glu Glu Arg Thr Ile Lys Glu Ala
            660                 665                 670

Lys Lys Thr Asn Glu Glu Lys Pro Asn Tyr Asn Phe Glu Lys Leu Leu
        675                 680                 685

Ile Gln Ile Phe Ile Lys Gly Phe Asp Thr Phe Leu Arg Asn Phe Asp
    690                 695                 700

Leu Asn Leu Asn Pro Ala Glu Glu Leu Val Gly Thr Val Lys Glu Lys
705                 710                 715                 720

Ala Glu Gly Leu Arg Lys Arg Lys Glu Arg Ile Ala Lys Ile Leu Asn
                725                 730                 735

Val Asp Glu Gln Ile Lys Thr Gly Asp Glu Glu Ile Ala Phe Trp Ile
            740                 745                 750

Phe Ala Lys Leu Leu Asp Ala Arg His Leu Ser Glu Leu Arg Asn Glu
        755                 760                 765
```

```
Met Ile Lys Phe Lys Gln Ser Ser Val Lys Lys Gly Leu Ile Lys Asn
    770                 775                 780

Gly Asp Leu Ile Glu Gln Met Gln Pro Ile Leu Glu Leu Cys Ile Leu
785                 790                 795                 800

Ser Asn Asp Ser Glu Ser Met Glu Lys Glu Ser Phe Asp Lys Ile Glu
                805                 810                 815

Val Phe Leu Glu Lys Val Glu Leu Ala Lys Asn Glu Pro Tyr Met Gln
            820                 825                 830

Glu Asp Lys Leu Thr Pro Val Lys Phe Arg Phe Met Lys Gln Leu Glu
        835                 840                 845

Lys Tyr Gln Thr Arg Asn Phe Ile Glu Asn Leu Val Ile Glu Asn Pro
    850                 855                 860

Glu Phe Lys Val Ser Glu Lys Ile Val Leu Asn Trp His Glu Glu Lys
865                 870                 875                 880

Glu Lys Ile Ala Asp Leu Val Asp Lys Arg Thr Lys Leu His Glu Glu
                885                 890                 895

Trp Ala Ser Lys Ala Arg Glu Ile Glu Glu Tyr Asn Glu Lys Ile Lys
            900                 905                 910

Lys Asn Lys Ser Lys Lys Leu Asp Lys Pro Ala Glu Phe Ala Lys Phe
        915                 920                 925

Ala Glu Tyr Lys Ile Ile Cys Glu Ala Ile Glu Asn Phe Asn Arg Leu
    930                 935                 940

Asp His Lys Val Arg Leu Thr Tyr Leu Lys Asn Leu His Tyr Leu Met
945                 950                 955                 960

Ile Asp Leu Met Gly Arg Met Val Gly Phe Ser Val Leu Phe Glu Arg
                965                 970                 975
```

```
Asp Phe Val Tyr Met Gly Arg Ser Tyr Ser Ala Leu Lys Lys Gln Ser
            980                 985                 990


Ile Tyr Leu Asn Asp Tyr Asp Thr  Phe Ala Asn Ile Arg  Asp Trp Glu
        995                 1000                1005


Val Asn  Glu Asn Lys His Leu  Phe Gly Thr Ser Ser  Ser Asp Leu
    1010                1015                1020


Thr Phe  Gln Glu Thr Ala Glu  Phe Lys Asn Leu Lys  Lys Pro Met
    1025                1030                1035


Glu Asn  Gln Leu Lys Ala Leu  Leu Gly Val Thr Asn  His Ser Phe
    1040                1045                1050


Glu Ile  Arg Asn Asn Ile Ala  His Leu His Val Leu  Arg Asn Asp
    1055                1060                1065


Gly Lys  Gly Glu Gly Val Ser  Leu Leu Ser Cys Met  Asn Asp Leu
    1070                1075                1080


Arg Lys  Leu Met Ser Tyr Asp  Arg Lys Leu Lys Asn  Ala Val Thr
    1085                1090                1095


Lys Ala  Ile Ile Lys Ile Leu  Asp Lys His Gly Met  Ile Leu Lys
    1100                1105                1110


Leu Thr  Asn Asn Asp His Thr  Lys Pro Phe Glu Ile  Glu Ser Leu
    1115                1120                1125


Lys Pro  Lys Lys Ile Ile His  Leu Glu Lys Ser Asn  His Ser Phe
    1130                1135                1140


Pro Met  Asp Gln Val Ser Gln  Glu Tyr Cys Asp Leu  Val Lys Lys
    1145                1150                1155


Met Leu  Val Phe Thr Asn
    1160
```

<210> 39
<211> 1154
<212> PRT
<213> Paludibacter propionicigenes

<400> 39

Met Arg Val Ser Lys Val Lys Val Lys Asp Gly Gly Lys Asp Lys Met
1 5 10 15

Val Leu Val His Arg Lys Thr Thr Gly Ala Gln Leu Val Tyr Ser Gly
20 25 30

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

Gln Ser Phe Asp Leu Ser Thr Leu Asn Lys Thr Ile Ile Lys Phe Asp
50 55 60

Thr Ala Lys Lys Gln Lys Leu Asn Val Asp Gln Tyr Lys Ile Val Glu
65 70 75 80

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

Glu Glu Ile Leu Pro Phe Leu Asn His Lys Phe Gln Glu Pro Val Lys
100 105 110

Tyr Trp Lys Asn Gly Lys Glu Glu Ser Phe Asn Leu Thr Leu Leu Ile
115 120 125

Val Glu Ala Val Gln Ala Gln Asp Lys Arg Lys Leu Gln Pro Tyr Tyr
130 135 140

Asp Trp Lys Thr Trp Tyr Ile Gln Thr Lys Ser Asp Leu Leu Lys Lys
145 150 155 160

Ser Ile Glu Asn Asn Arg Ile Asp Leu Thr Glu Asn Leu Ser Lys Arg
165 170 175

Lys Lys Ala Leu Leu Ala Trp Glu Thr Glu Phe Thr Ala Ser Gly Ser

180 185 190

Ile Asp Leu Thr His Tyr His Lys Val Tyr Met Thr Asp Val Leu Cys
195 200 205

Lys Met Leu Gln Asp Val Lys Pro Leu Thr Asp Asp Lys Gly Lys Ile
210 215 220

Asn Thr Asn Ala Tyr His Arg Gly Leu Lys Lys Ala Leu Gln Asn His
225 230 235 240

Gln Pro Ala Ile Phe Gly Thr Arg Glu Val Pro Asn Glu Ala Asn Arg
245 250 255

Ala Asp Asn Gln Leu Ser Ile Tyr His Leu Glu Val Val Lys Tyr Leu
260 265 270

Glu His Tyr Phe Pro Ile Lys Thr Ser Lys Arg Arg Asn Thr Ala Asp
275 280 285

Asp Ile Ala His Tyr Leu Lys Ala Gln Thr Leu Lys Thr Thr Ile Glu
290 295 300

Lys Gln Leu Val Asn Ala Ile Arg Ala Asn Ile Ile Gln Gln Gly Lys
305 310 315 320

Thr Asn His His Glu Leu Lys Ala Asp Thr Thr Ser Asn Asp Leu Ile
325 330 335

Arg Ile Lys Thr Asn Glu Ala Phe Val Leu Asn Leu Thr Gly Thr Cys
340 345 350

Ala Phe Ala Ala Asn Asn Ile Arg Asn Met Val Asp Asn Glu Gln Thr
355 360 365

Asn Asp Ile Leu Gly Lys Gly Asp Phe Ile Lys Ser Leu Leu Lys Asp
370 375 380

Asn Thr Asn Ser Gln Leu Tyr Ser Phe Phe Phe Gly Glu Gly Leu Ser

385 390 395 400

Thr Asn Lys Ala Glu Lys Glu Thr Gln Leu Trp Gly Ile Arg Gly Ala
405 410 415

Val Gln Gln Ile Arg Asn Asn Val Asn His Tyr Lys Lys Asp Ala Leu
420 425 430

Lys Thr Val Phe Asn Ile Ser Asn Phe Glu Asn Pro Thr Ile Thr Asp
435 440 445

Pro Lys Gln Gln Thr Asn Tyr Ala Asp Thr Ile Tyr Lys Ala Arg Phe
450 455 460

Ile Asn Glu Leu Glu Lys Ile Pro Glu Ala Phe Ala Gln Gln Leu Lys
465 470 475 480

Thr Gly Gly Ala Val Ser Tyr Tyr Thr Ile Glu Asn Leu Lys Ser Leu
485 490 495

Leu Thr Thr Phe Gln Phe Ser Leu Cys Arg Ser Thr Ile Pro Phe Ala
500 505 510

Pro Gly Phe Lys Lys Val Phe Asn Gly Gly Ile Asn Tyr Gln Asn Ala
515 520 525

Lys Gln Asp Glu Ser Phe Tyr Glu Leu Met Leu Glu Gln Tyr Leu Arg
530 535 540

Lys Glu Asn Phe Ala Glu Glu Ser Tyr Asn Ala Arg Tyr Phe Met Leu
545 550 555 560

Lys Leu Ile Tyr Asn Asn Leu Phe Leu Pro Gly Phe Thr Thr Asp Arg
565 570 575

Lys Ala Phe Ala Asp Ser Val Gly Phe Val Gln Met Gln Asn Lys Lys
580 585 590

Gln Ala Glu Lys Val Asn Pro Arg Lys Lys Glu Ala Tyr Ala Phe Glu

595 600 605

Ala Val Arg Pro Met Thr Ala Ala Asp Ser Ile Ala Asp Tyr Met Ala
610 615 620

Tyr Val Gln Ser Glu Leu Met Gln Glu Gln Asn Lys Lys Glu Glu Lys
625 630 635 640

Val Ala Glu Glu Thr Arg Ile Asn Phe Glu Lys Phe Val Leu Gln Val
645 650 655

Phe Ile Lys Gly Phe Asp Ser Phe Leu Arg Ala Lys Glu Phe Asp Phe
660 665 670

Val Gln Met Pro Gln Pro Gln Leu Thr Ala Thr Ala Ser Asn Gln Gln
675 680 685

Lys Ala Asp Lys Leu Asn Gln Leu Glu Ala Ser Ile Thr Ala Asp Cys
690 695 700

Lys Leu Thr Pro Gln Tyr Ala Lys Ala Asp Asp Ala Thr His Ile Ala
705 710 715 720

Phe Tyr Val Phe Cys Lys Leu Leu Asp Ala Ala His Leu Ser Asn Leu
725 730 735

Arg Asn Glu Leu Ile Lys Phe Arg Glu Ser Val Asn Glu Phe Lys Phe
740 745 750

His His Leu Leu Glu Ile Ile Glu Ile Cys Leu Leu Ser Ala Asp Val
755 760 765

Val Pro Thr Asp Tyr Arg Asp Leu Tyr Ser Ser Glu Ala Asp Cys Leu
770 775 780

Ala Arg Leu Arg Pro Phe Ile Glu Gln Gly Ala Asp Ile Thr Asn Trp
785 790 795 800

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

```
                805                 810                 815

Asn Ile Glu Leu Ser Val Lys Tyr Gly Thr Thr Lys Leu Leu Glu Gln
            820                 825                 830

Ile Ile Asn Lys Asp Thr Gln Phe Lys Thr Thr Glu Ala Asn Phe Thr
        835                 840                 845

Ala Trp Asn Thr Ala Gln Lys Ser Ile Glu Gln Leu Ile Lys Gln Arg
    850                 855                 860

Glu Asp His His Glu Gln Trp Val Lys Ala Lys Asn Ala Asp Asp Lys
865                 870                 875                 880

Glu Lys Gln Glu Arg Lys Arg Glu Lys Ser Asn Phe Ala Gln Lys Phe
                885                 890                 895

Ile Glu Lys His Gly Asp Asp Tyr Leu Asp Ile Cys Asp Tyr Ile Asn
            900                 905                 910

Thr Tyr Asn Trp Leu Asp Asn Lys Met His Phe Val His Leu Asn Arg
        915                 920                 925

Leu His Gly Leu Thr Ile Glu Leu Leu Gly Arg Met Ala Gly Phe Val
    930                 935                 940

Ala Leu Phe Asp Arg Asp Phe Gln Phe Phe Asp Glu Gln Gln Ile Ala
945                 950                 955                 960

Asp Glu Phe Lys Leu His Gly Phe Val Asn Leu His Ser Ile Asp Lys
                965                 970                 975

Lys Leu Asn Glu Val Pro Thr Lys Lys Ile Lys Glu Ile Tyr Asp Ile
            980                 985                 990

Arg Asn Lys Ile Ile Gln Ile Asn  Gly Asn Lys Ile Asn  Glu Ser Val
        995                 1000                1005

Arg Ala  Asn Leu Ile Gln Phe  Ile Ser Ser Lys Arg  Asn Tyr Tyr
```

1010 1015 1020

Asn Asn Ala Phe Leu His Val Ser Asn Asp Glu Ile Lys Glu Lys
1025 1030 1035

Gln Met Tyr Asp Ile Arg Asn His Ile Ala His Phe Asn Tyr Leu
1040 1045 1050

Thr Lys Asp Ala Ala Asp Phe Ser Leu Ile Asp Leu Ile Asn Glu
1055 1060 1065

Leu Arg Glu Leu Leu His Tyr Asp Arg Lys Leu Lys Asn Ala Val
1070 1075 1080

Ser Lys Ala Phe Ile Asp Leu Phe Asp Lys His Gly Met Ile Leu
1085 1090 1095

Lys Leu Lys Leu Asn Ala Asp His Lys Leu Lys Val Glu Ser Leu
1100 1105 1110

Glu Pro Lys Lys Ile Tyr His Leu Gly Ser Ser Ala Lys Asp Lys
1115 1120 1125

Pro Glu Tyr Gln Tyr Cys Thr Asn Gln Val Met Met Ala Tyr Cys
1130 1135 1140

Asn Met Cys Arg Ser Leu Leu Glu Met Lys Lys
1145 1150

<210> 40
<211> 970
<212> PRT
<213> Listeria weihenstephanensis

<400> 40

Met Leu Ala Leu Leu His Gln Glu Val Pro Ser Gln Lys Leu His Asn
1 5 10 15

Leu Lys Ser Leu Asn Thr Glu Ser Leu Thr Lys Leu Phe Lys Pro Lys
20 25 30

```
Phe Gln Asn Met Ile Ser Tyr Pro Pro Ser Lys Gly Ala Glu His Val
        35                  40                  45

Gln Phe Cys Leu Thr Asp Ile Ala Val Pro Ala Ile Arg Asp Leu Asp
    50                  55                  60

Glu Ile Lys Pro Asp Trp Gly Ile Phe Phe Glu Lys Leu Lys Pro Tyr
65                  70                  75                  80

Thr Asp Trp Ala Glu Ser Tyr Ile His Tyr Lys Gln Thr Thr Ile Gln
                85                  90                  95

Lys Ser Ile Glu Gln Asn Lys Ile Gln Ser Pro Asp Ser Pro Arg Lys
            100                 105                 110

Leu Val Leu Gln Lys Tyr Val Thr Ala Phe Leu Asn Gly Glu Pro Leu
        115                 120                 125

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

Ser Phe Lys Val Val Asp Leu Asn Glu Asp Lys Ser Ala Asn Tyr Lys
145                 150                 155                 160

Ile Lys Ala Cys Leu Gln Gln His Gln Arg Asn Ile Leu Asp Glu Leu
                165                 170                 175

Lys Glu Asp Pro Glu Leu Asn Gln Tyr Gly Ile Glu Val Lys Lys Tyr
            180                 185                 190

Ile Gln Arg Tyr Phe Pro Ile Lys Arg Ala Pro Asn Arg Ser Lys His
        195                 200                 205

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

Gln Gln Phe Lys Asn Ala Val Tyr His Tyr Val Leu Glu Gln Gly Lys
225                 230                 235                 240
```

```
Met Glu Ala Tyr Glu Leu Thr Asp Pro Lys Thr Lys Asp Leu Gln Asp
                245                 250                 255

Ile Arg Ser Gly Glu Ala Phe Ser Phe Lys Phe Ile Asn Ala Cys Ala
            260                 265                 270

Phe Ala Ser Asn Asn Leu Lys Met Ile Leu Asn Pro Glu Cys Glu Lys
        275                 280                 285

Asp Ile Leu Gly Lys Gly Asp Phe Lys Lys Asn Leu Pro Asn Ser Thr
    290                 295                 300

Thr Gln Ser Asp Val Val Lys Lys Met Ile Pro Phe Phe Ser Asp Glu
305                 310                 315                 320

Ile Gln Asn Val Asn Phe Asp Glu Ala Ile Trp Ala Ile Arg Gly Ser
                325                 330                 335

Ile Gln Gln Ile Arg Asn Glu Val Tyr His Cys Lys Lys His Ser Trp
            340                 345                 350

Lys Ser Ile Leu Lys Ile Lys Gly Phe Glu Phe Glu Pro Asn Asn Met
        355                 360                 365

Lys Tyr Thr Asp Ser Asp Met Gln Lys Leu Met Asp Lys Asp Ile Ala
    370                 375                 380

Lys Ile Pro Asp Phe Ile Glu Glu Lys Leu Lys Ser Ser Gly Ile Ile
385                 390                 395                 400

Arg Phe Tyr Ser His Asp Lys Leu Gln Ser Ile Trp Glu Met Lys Gln
                405                 410                 415

Gly Phe Ser Leu Leu Thr Thr Asn Ala Pro Phe Val Pro Ser Phe Lys
            420                 425                 430

Arg Val Tyr Ala Lys Gly His Asp Tyr Gln Thr Ser Lys Asn Arg Tyr
        435                 440                 445
```

```
Tyr Asp Leu Gly Leu Thr Thr Phe Asp Ile Leu Glu Tyr Gly Glu Glu
    450                 455                 460

Asp Phe Arg Ala Arg Tyr Phe Leu Thr Lys Leu Val Tyr Tyr Gln Gln
465                 470                 475                 480

Phe Met Pro Trp Phe Thr Ala Asp Asn Asn Ala Phe Arg Asp Ala Ala
                485                 490                 495

Asn Phe Val Leu Arg Leu Asn Lys Asn Arg Gln Gln Asp Ala Lys Ala
            500                 505                 510

Phe Ile Asn Ile Arg Glu Val Glu Glu Gly Glu Met Pro Arg Asp Tyr
        515                 520                 525

Met Gly Tyr Val Gln Gly Gln Ile Ala Ile His Glu Asp Ser Thr Glu
    530                 535                 540

Asp Thr Pro Asn His Phe Glu Lys Phe Ile Ser Gln Val Phe Ile Lys
545                 550                 555                 560

Gly Phe Asp Ser His Met Arg Ser Ala Asp Leu Lys Phe Ile Lys Asn
                565                 570                 575

Pro Arg Asn Gln Gly Leu Glu Gln Ser Glu Ile Glu Glu Met Ser Phe
            580                 585                 590

Asp Ile Lys Val Glu Pro Ser Phe Leu Lys Asn Lys Asp Asp Tyr Ile
        595                 600                 605

Ala Phe Trp Thr Phe Cys Lys Met Leu Asp Ala Arg His Leu Ser Glu
    610                 615                 620

Leu Arg Asn Glu Met Ile Lys Tyr Asp Gly His Leu Thr Gly Glu Gln
625                 630                 635                 640

Glu Ile Ile Gly Leu Ala Leu Leu Gly Val Asp Ser Arg Glu Asn Asp
                645                 650                 655
```

```
Trp Lys Gln Phe Phe Ser Ser Glu Arg Glu Tyr Glu Lys Ile Met Lys
            660                 665                 670

Gly Tyr Val Gly Glu Glu Leu Tyr Gln Arg Glu Pro Tyr Arg Gln Ser
        675                 680                 685

Asp Gly Lys Thr Pro Ile Leu Phe Arg Gly Val Glu Gln Ala Arg Lys
    690                 695                 700

Tyr Gly Thr Glu Thr Val Ile Gln Arg Leu Phe Asp Ala Ser Pro Glu
705                 710                 715                 720

Phe Lys Val Ser Lys Cys Asn Ile Thr Glu Trp Glu Arg Gln Lys Glu
                725                 730                 735

Thr Ile Glu Glu Thr Ile Glu Arg Arg Lys Glu Leu His Asn Glu Trp
            740                 745                 750

Glu Lys Asn Pro Lys Lys Pro Gln Asn Asn Ala Phe Phe Lys Glu Tyr
        755                 760                 765

Lys Glu Cys Cys Asp Ala Ile Asp Ala Tyr Asn Trp His Lys Asn Lys
    770                 775                 780

Thr Thr Leu Val Tyr Val Asn Glu Leu His His Leu Leu Ile Glu Ile
785                 790                 795                 800

Leu Gly Arg Tyr Val Gly Tyr Val Ala Ile Ala Asp Arg Asp Phe Gln
                805                 810                 815

Cys Met Ala Asn Gln Tyr Phe Lys His Ser Gly Ile Thr Glu Arg Val
            820                 825                 830

Glu Tyr Trp Gly Asp Asn Arg Leu Lys Ser Ile Lys Lys Leu Asp Thr
        835                 840                 845

Phe Leu Lys Lys Glu Gly Leu Phe Val Ser Glu Lys Asn Ala Arg Asn
    850                 855                 860
```

```
His Ile Ala His Leu Asn Tyr Leu Ser Leu Lys Ser Glu Cys Thr Leu
865                 870                 875                 880


Leu Tyr Leu Ser Glu Arg Leu Arg Glu Ile Phe Lys Tyr Asp Arg Lys
                885                 890                 895


Leu Lys Asn Ala Val Ser Lys Ser Leu Ile Asp Ile Leu Asp Arg His
            900                 905                 910


Gly Met Ser Val Val Phe Ala Asn Leu Lys Glu Asn Lys His Arg Leu
        915                 920                 925


Val Ile Lys Ser Leu Glu Pro Lys Lys Leu Arg His Leu Gly Glu Lys
    930                 935                 940


Lys Ile Asp Asn Gly Tyr Ile Glu Thr Asn Gln Val Ser Glu Glu Tyr
945                 950                 955                 960


Cys Gly Ile Val Lys Arg Leu Leu Glu Ile
                965                 970


<210>  41
<211>  1051
<212>  PRT
<213>  Listeriacea bacterium

<400>  41

Met Lys Ile Thr Lys Met Arg Val Asp Gly Arg Thr Ile Val Met Glu
1               5                   10                  15


Arg Thr Ser Lys Glu Gly Gln Leu Gly Tyr Glu Gly Ile Asp Gly Asn
            20                  25                  30


Lys Thr Thr Glu Ile Ile Phe Asp Lys Lys Lys Glu Ser Phe Tyr Lys
        35                  40                  45


Ser Ile Leu Asn Lys Thr Val Arg Lys Pro Asp Glu Lys Glu Lys Asn
    50                  55                  60
```

```
Arg Arg Lys Gln Ala Ile Asn Lys Ala Ile Asn Lys Glu Ile Thr Glu
65                  70                  75                  80

Leu Met Leu Ala Val Leu His Gln Glu Val Pro Ser Gln Lys Leu His
                85                  90                  95

Asn Leu Lys Ser Leu Asn Thr Glu Ser Leu Thr Lys Leu Phe Lys Pro
            100                 105                 110

Lys Phe Gln Asn Met Ile Ser Tyr Pro Pro Ser Lys Gly Ala Glu His
        115                 120                 125

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

Asp Glu Ile Lys Pro Asp Trp Gly Ile Phe Phe Glu Lys Leu Lys Pro
145                 150                 155                 160

Tyr Thr Asp Trp Ala Glu Ser Tyr Ile His Tyr Lys Gln Thr Thr Ile
                165                 170                 175

Gln Lys Ser Ile Glu Gln Asn Lys Ile Gln Ser Pro Asp Ser Pro Arg
            180                 185                 190

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

Leu Gly Leu Asp Leu Val Ala Lys Lys Tyr Lys Leu Ala Asp Leu Ala
    210                 215                 220

Glu Ser Phe Lys Leu Val Asp Leu Asn Glu Asp Lys Ser Ala Asn Tyr
225                 230                 235                 240

Lys Ile Lys Ala Cys Leu Gln Gln His Gln Arg Asn Ile Leu Asp Glu
                245                 250                 255

Leu Lys Glu Asp Pro Glu Leu Asn Gln Tyr Gly Ile Glu Val Lys Lys
            260                 265                 270
```

```
Tyr Ile Gln Arg Tyr Phe Pro Ile Lys Arg Ala Pro Asn Arg Ser Lys
        275                 280                 285

His Ala Arg Ala Asp Phe Leu Lys Lys Glu Leu Ile Glu Ser Thr Val
    290                 295                 300

Glu Gln Gln Phe Lys Asn Ala Val Tyr His Tyr Val Leu Glu Gln Gly
305                 310                 315                 320

Lys Met Glu Ala Tyr Glu Leu Thr Asp Pro Lys Thr Lys Asp Leu Gln
                325                 330                 335

Asp Ile Arg Ser Gly Glu Ala Phe Ser Phe Lys Phe Ile Asn Ala Cys
            340                 345                 350

Ala Phe Ala Ser Asn Asn Leu Lys Met Ile Leu Asn Pro Glu Cys Glu
        355                 360                 365

Lys Asp Ile Leu Gly Lys Gly Asn Phe Lys Lys Asn Leu Pro Asn Ser
    370                 375                 380

Thr Thr Arg Ser Asp Val Val Lys Lys Met Ile Pro Phe Phe Ser Asp
385                 390                 395                 400

Glu Leu Gln Asn Val Asn Phe Asp Glu Ala Ile Trp Ala Ile Arg Gly
                405                 410                 415

Ser Ile Gln Gln Ile Arg Asn Glu Val Tyr His Cys Lys Lys His Ser
            420                 425                 430

Trp Lys Ser Ile Leu Lys Ile Lys Gly Phe Glu Phe Glu Pro Asn Asn
        435                 440                 445

Met Lys Tyr Ala Asp Ser Asp Met Gln Lys Leu Met Asp Lys Asp Ile
    450                 455                 460

Ala Lys Ile Pro Glu Phe Ile Glu Glu Lys Leu Lys Ser Ser Gly Val
465                 470                 475                 480
```

```
Val Arg Phe Tyr Arg His Asp Glu Leu Gln Ser Ile Trp Glu Met Lys
                485                 490                 495

Gln Gly Phe Ser Leu Leu Thr Thr Asn Ala Pro Phe Val Pro Ser Phe
            500                 505                 510

Lys Arg Val Tyr Ala Lys Gly His Asp Tyr Gln Thr Ser Lys Asn Arg
        515                 520                 525

Tyr Tyr Asn Leu Asp Leu Thr Thr Phe Asp Ile Leu Glu Tyr Gly Glu
    530                 535                 540

Glu Asp Phe Arg Ala Arg Tyr Phe Leu Thr Lys Leu Val Tyr Tyr Gln
545                 550                 555                 560

Gln Phe Met Pro Trp Phe Thr Ala Asp Asn Asn Ala Phe Arg Asp Ala
                565                 570                 575

Ala Asn Phe Val Leu Arg Leu Asn Lys Asn Arg Gln Gln Asp Ala Lys
            580                 585                 590

Ala Phe Ile Asn Ile Arg Glu Val Glu Glu Gly Glu Met Pro Arg Asp
        595                 600                 605

Tyr Met Gly Tyr Val Gln Gly Gln Ile Ala Ile His Glu Asp Ser Ile
    610                 615                 620

Glu Asp Thr Pro Asn His Phe Glu Lys Phe Ile Ser Gln Val Phe Ile
625                 630                 635                 640

Lys Gly Phe Asp Arg His Met Arg Ser Ala Asn Leu Lys Phe Ile Lys
                645                 650                 655

Asn Pro Arg Asn Gln Gly Leu Glu Gln Ser Glu Ile Glu Glu Met Ser
            660                 665                 670

Phe Asp Ile Lys Val Glu Pro Ser Phe Leu Lys Asn Lys Asp Asp Tyr
        675                 680                 685
```

```
Ile Ala Phe Trp Ile Phe Cys Lys Met Leu Asp Ala Arg His Leu Ser
    690                 695                 700

Glu Leu Arg Asn Glu Met Ile Lys Tyr Asp Gly His Leu Thr Gly Glu
705                 710                 715                 720

Gln Glu Ile Ile Gly Leu Ala Leu Leu Gly Val Asp Ser Arg Glu Asn
                725                 730                 735

Asp Trp Lys Gln Phe Phe Ser Ser Glu Arg Glu Tyr Glu Lys Ile Met
            740                 745                 750

Lys Gly Tyr Val Val Glu Glu Leu Tyr Gln Arg Glu Pro Tyr Arg Gln
        755                 760                 765

Ser Asp Gly Lys Thr Pro Ile Leu Phe Arg Gly Val Glu Gln Ala Arg
    770                 775                 780

Lys Tyr Gly Thr Glu Thr Val Ile Gln Arg Leu Phe Asp Ala Asn Pro
785                 790                 795                 800

Glu Phe Lys Val Ser Lys Cys Asn Leu Ala Glu Trp Glu Arg Gln Lys
                805                 810                 815

Glu Thr Ile Glu Glu Thr Ile Lys Arg Arg Lys Glu Leu His Asn Glu
            820                 825                 830

Trp Ala Lys Asn Pro Lys Lys Pro Gln Asn Asn Ala Phe Phe Lys Glu
        835                 840                 845

Tyr Lys Glu Cys Cys Asp Ala Ile Asp Ala Tyr Asn Trp His Lys Asn
    850                 855                 860

Lys Thr Thr Leu Ala Tyr Val Asn Glu Leu His His Leu Leu Ile Glu
865                 870                 875                 880

Ile Leu Gly Arg Tyr Val Gly Tyr Val Ala Ile Ala Asp Arg Asp Phe
                885                 890                 895
```

```
Gln Cys Met Ala Asn Gln Tyr Phe Lys His Ser Gly Ile Thr Glu Arg
            900                 905                 910


Val Glu Tyr Trp Gly Asp Asn Arg Leu Lys Ser Ile Lys Lys Leu Asp
        915                 920                 925


Thr Phe Leu Lys Lys Glu Gly Leu Phe Val Ser Glu Lys Asn Ala Arg
    930                 935                 940


Asn His Ile Ala His Leu Asn Tyr Leu Ser Leu Lys Ser Glu Cys Thr
945                 950                 955                 960


Leu Leu Tyr Leu Ser Glu Arg Leu Arg Glu Ile Phe Lys Tyr Asp Arg
                965                 970                 975


Lys Leu Lys Asn Ala Val Ser Lys Ser Leu Ile Asp Ile Leu Asp Arg
            980                 985                 990


His Gly Met Ser Val Val Phe Ala  Asn Leu Lys Glu Asn  Lys His Arg
        995                 1000                1005


Leu Val  Ile Lys Ser Leu Glu  Pro Lys Lys Leu Arg  His Leu Gly
    1010                1015                1020


Gly Lys  Lys Ile Asp Gly Gly  Tyr Ile Glu Thr Asn  Gln Val Ser
    1025                1030                1035


Glu Glu  Tyr Cys Gly Ile Val  Lys Arg Leu Leu Glu  Met
    1040                1045                1050


<210>  42
<211>  1152
<212>  PRT
<213>  Leptotrichia wadei

<400>  42

Met Lys Val Thr Lys Val Asp Gly Ile Ser His Lys Lys Tyr Ile Glu
1               5                   10                  15
```

```
Glu Gly Lys Leu Val Lys Ser Thr Ser Glu Glu Asn Arg Thr Ser Glu
            20                  25                  30

Arg Leu Ser Glu Leu Leu Ser Ile Arg Leu Asp Ile Tyr Ile Lys Asn
        35                  40                  45

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

Lys Lys Phe Phe Ser Asn Lys Val Leu His Leu Lys Asp Ser Val Leu
65                  70                  75                  80

Tyr Leu Lys Asn Arg Lys Glu Lys Asn Ala Val Gln Asp Lys Asn Tyr
                85                  90                  95

Ser Glu Glu Asp Ile Ser Glu Tyr Asp Leu Lys Asn Lys Asn Ser Phe
            100                 105                 110

Ser Val Leu Lys Lys Ile Leu Leu Asn Glu Asp Val Asn Ser Glu Glu
        115                 120                 125

Leu Glu Ile Phe Arg Lys Asp Val Glu Ala Lys Leu Asn Lys Ile Asn
    130                 135                 140

Ser Leu Lys Tyr Ser Phe Glu Glu Asn Lys Ala Asn Tyr Gln Lys Ile
145                 150                 155                 160

Asn Glu Asn Asn Val Glu Lys Val Gly Gly Lys Ser Lys Arg Asn Ile
                165                 170                 175

Ile Tyr Asp Tyr Tyr Arg Glu Ser Ala Lys Arg Asn Asp Tyr Ile Asn
            180                 185                 190

Asn Val Gln Glu Ala Phe Asp Lys Leu Tyr Lys Lys Glu Asp Ile Glu
        195                 200                 205

Lys Leu Phe Phe Leu Ile Glu Asn Ser Lys Lys His Glu Lys Tyr Lys
    210                 215                 220
```

```
Ile Arg Glu Tyr Tyr His Lys Ile Ile Gly Arg Lys Asn Asp Lys Glu
225                 230                 235                 240


Asn Phe Ala Lys Ile Ile Tyr Glu Glu Ile Gln Asn Val Asn Asn Ile
                245                 250                 255


Lys Glu Leu Ile Glu Lys Ile Pro Asp Met Ser Glu Leu Lys Lys Ser
            260                 265                 270


Gln Val Phe Tyr Lys Tyr Tyr Leu Asp Lys Glu Glu Leu Asn Asp Lys
        275                 280                 285


Asn Ile Lys Tyr Ala Phe Cys His Phe Val Glu Ile Glu Met Ser Gln
    290                 295                 300


Leu Leu Lys Asn Tyr Val Tyr Lys Arg Leu Ser Asn Ile Ser Asn Asp
305                 310                 315                 320


Lys Ile Lys Arg Ile Phe Glu Tyr Gln Asn Leu Lys Lys Leu Ile Glu
                325                 330                 335


Asn Lys Leu Leu Asn Lys Leu Asp Thr Tyr Val Arg Asn Cys Gly Lys
            340                 345                 350


Tyr Asn Tyr Tyr Leu Gln Val Gly Glu Ile Ala Thr Ser Asp Phe Ile
        355                 360                 365


Ala Arg Asn Arg Gln Asn Glu Ala Phe Leu Arg Asn Ile Ile Gly Val
    370                 375                 380


Ser Ser Val Ala Tyr Phe Ser Leu Arg Asn Ile Leu Glu Thr Glu Asn
385                 390                 395                 400


Glu Asn Asp Ile Thr Gly Arg Met Arg Gly Lys Thr Val Lys Asn Asn
                405                 410                 415


Lys Gly Glu Glu Lys Tyr Val Ser Gly Glu Val Asp Lys Ile Tyr Asn
            420                 425                 430
```

```
Glu Asn Lys Gln Asn Glu Val Lys Glu Asn Leu Lys Met Phe Tyr Ser
        435                 440                 445


Tyr Asp Phe Asn Met Asp Asn Lys Asn Glu Ile Glu Asp Phe Phe Ala
    450                 455                 460


Asn Ile Asp Glu Ala Ile Ser Ser Ile Arg His Gly Ile Val His Phe
465                 470                 475                 480


Asn Leu Glu Leu Glu Gly Lys Asp Ile Phe Ala Phe Lys Asn Ile Ala
                485                 490                 495


Pro Ser Glu Ile Ser Lys Lys Met Phe Gln Asn Glu Ile Asn Glu Lys
            500                 505                 510


Lys Leu Lys Leu Lys Ile Phe Lys Gln Leu Asn Ser Ala Asn Val Phe
        515                 520                 525


Asn Tyr Tyr Glu Lys Asp Val Ile Ile Lys Tyr Leu Lys Asn Thr Lys
    530                 535                 540


Phe Asn Phe Val Asn Lys Asn Ile Pro Phe Val Pro Ser Phe Thr Lys
545                 550                 555                 560


Leu Tyr Asn Lys Ile Glu Asp Leu Arg Asn Thr Leu Lys Phe Phe Trp
                565                 570                 575


Ser Val Pro Lys Asp Lys Glu Glu Lys Asp Ala Gln Ile Tyr Leu Leu
            580                 585                 590


Lys Asn Ile Tyr Tyr Gly Glu Phe Leu Asn Lys Phe Val Lys Asn Ser
        595                 600                 605


Lys Val Phe Phe Lys Ile Thr Asn Glu Val Ile Lys Ile Asn Lys Gln
    610                 615                 620


Arg Asn Gln Lys Thr Gly His Tyr Lys Tyr Gln Lys Phe Glu Asn Ile
625                 630                 635                 640
```

```
Glu Lys Thr Val Pro Val Glu Tyr Leu Ala Ile Ile Gln Ser Arg Glu
                645                 650                 655


Met Ile Asn Asn Gln Asp Lys Glu Glu Lys Asn Thr Tyr Ile Asp Phe
            660                 665                 670


Ile Gln Gln Ile Phe Leu Lys Gly Phe Ile Asp Tyr Leu Asn Lys Asn
        675                 680                 685


Asn Leu Lys Tyr Ile Glu Ser Asn Asn Asn Asn Asp Asn Asn Asp Ile
    690                 695                 700


Phe Ser Lys Ile Lys Ile Lys Lys Asp Asn Lys Glu Lys Tyr Asp Lys
705                 710                 715                 720


Ile Leu Lys Asn Tyr Glu Lys His Asn Arg Asn Lys Glu Ile Pro His
                725                 730                 735


Glu Ile Asn Glu Phe Val Arg Glu Ile Lys Leu Gly Lys Ile Leu Lys
            740                 745                 750


Tyr Thr Glu Asn Leu Asn Met Phe Tyr Leu Ile Leu Lys Leu Leu Asn
        755                 760                 765


His Lys Glu Leu Thr Asn Leu Lys Gly Ser Leu Glu Lys Tyr Gln Ser
    770                 775                 780


Ala Asn Lys Glu Glu Thr Phe Ser Asp Glu Leu Glu Leu Ile Asn Leu
785                 790                 795                 800


Leu Asn Leu Asp Asn Asn Arg Val Thr Glu Asp Phe Glu Leu Glu Ala
                805                 810                 815


Asn Glu Ile Gly Lys Phe Leu Asp Phe Asn Glu Asn Lys Ile Lys Asp
            820                 825                 830


Arg Lys Glu Leu Lys Lys Phe Asp Thr Asn Lys Ile Tyr Phe Asp Gly
        835                 840                 845
```

```
Glu Asn Ile Ile Lys His Arg Ala Phe Tyr Asn Ile Lys Lys Tyr Gly
    850                 855                 860

Met Leu Asn Leu Leu Glu Lys Ile Ala Asp Lys Ala Lys Tyr Lys Ile
865                 870                 875                 880

Ser Leu Lys Glu Leu Lys Glu Tyr Ser Asn Lys Lys Asn Glu Ile Glu
                885                 890                 895

Lys Asn Tyr Thr Met Gln Gln Asn Leu His Arg Lys Tyr Ala Arg Pro
            900                 905                 910

Lys Lys Asp Glu Lys Phe Asn Asp Glu Asp Tyr Lys Glu Tyr Glu Lys
        915                 920                 925

Ala Ile Gly Asn Ile Gln Lys Tyr Thr His Leu Lys Asn Lys Val Glu
    930                 935                 940

Phe Asn Glu Leu Asn Leu Leu Gln Gly Leu Leu Leu Lys Ile Leu His
945                 950                 955                 960

Arg Leu Val Gly Tyr Thr Ser Ile Trp Glu Arg Asp Leu Arg Phe Arg
                965                 970                 975

Leu Lys Gly Glu Phe Pro Glu Asn His Tyr Ile Glu Glu Ile Phe Asn
            980                 985                 990

Phe Asp Asn Ser Lys Asn Val Lys  Tyr Lys Ser Gly Gln  Ile Val Glu
        995                 1000                1005

Lys Tyr  Ile Asn Phe Tyr Lys  Glu Leu Tyr Lys Asp  Asn Val Glu
    1010                1015                1020

Lys Arg  Ser Ile Tyr Ser Asp  Lys Lys Val Lys Lys  Leu Lys Gln
    1025                1030                1035

Glu Lys  Lys Asp Leu Tyr Ile  Arg Asn Tyr Ile Ala  His Phe Asn
    1040                1045                1050
```

```
Tyr Ile  Pro His Ala Glu Ile  Ser Leu Leu Glu Val  Leu Glu Asn
    1055                1060                1065


Leu Arg  Lys Leu Leu Ser Tyr  Asp Arg Lys Leu Lys  Asn Ala Ile
    1070                1075                1080


Met Lys  Ser Ile Val Asp Ile  Leu Lys Glu Tyr Gly  Phe Val Ala
    1085                1090                1095


Thr Phe  Lys Ile Gly Ala Asp  Lys Lys Ile Glu Ile  Gln Thr Leu
    1100                1105                1110


Glu Ser  Glu Lys Ile Val His  Leu Lys Asn Leu Lys  Lys Lys Lys
    1115                1120                1125


Leu Met  Thr Asp Arg Asn Ser  Glu Glu Leu Cys Glu  Leu Val Lys
    1130                1135                1140


Val Met  Phe Glu Tyr Lys Ala  Leu Glu
    1145                1150


<210>  43
<211>  1285
<212>  PRT
<213>  Rhodobacter capsulatus

<400>  43

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


Pro Ala Gly Gly Leu Lys Arg Lys Ile Ser Thr Asp Gly Lys Asn Arg
            20                  25                  30


Lys Glu Leu Pro Ala His Leu Ser Ser Asp Pro Lys Ala Leu Ile Gly
        35                  40                  45


Gln Trp Ile Ser Gly Ile Asp Lys Ile Tyr Arg Lys Pro Asp Ser Arg
    50                  55                  60


Lys Ser Asp Gly Lys Ala Ile His Ser Pro Thr Pro Ser Lys Met Gln
```

65 70 75 80

Phe Asp Ala Arg Asp Asp Leu Gly Glu Ala Phe Trp Lys Leu Val Ser
85 90 95

Glu Ala Gly Leu Ala Gln Asp Ser Asp Tyr Asp Gln Phe Lys Arg Arg
100 105 110

Leu His Pro Tyr Gly Asp Lys Phe Gln Pro Ala Asp Ser Gly Ala Lys
115 120 125

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

Trp Tyr Gly Ala Met Ser Lys Arg Gly Asn Asp Ala Lys Glu Leu Ala
145 150 155 160

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

Gln Pro Lys Arg Asn Pro Lys Thr Asp Lys Phe Ala Pro Gly Leu Val
180 185 190

Val Ala Arg Ala Leu Gly Ile Glu Ser Ser Val Leu Pro Arg Gly Met
195 200 205

Ala Arg Leu Ala Arg Asn Trp Gly Glu Glu Glu Ile Gln Thr Tyr Phe
210 215 220

Val Val Asp Val Ala Ala Ser Val Lys Glu Val Ala Lys Ala Ala Val
225 230 235 240

Ser Ala Ala Gln Ala Phe Asp Pro Pro Arg Gln Val Ser Gly Arg Ser
245 250 255

Leu Ser Pro Lys Val Gly Phe Ala Leu Ala Glu His Leu Glu Arg Val
260 265 270

Thr Gly Ser Lys Arg Cys Ser Phe Asp Pro Ala Ala Gly Pro Ser Val

```
        275                 280                 285

Leu Ala Leu His Asp Glu Val Lys Lys Thr Tyr Lys Arg Leu Cys Ala
    290                 295                 300

Arg Gly Lys Asn Ala Ala Arg Ala Phe Pro Ala Asp Lys Thr Glu Leu
305                 310                 315                 320

Leu Ala Leu Met Arg His Thr His Glu Asn Arg Val Arg Asn Gln Met
                325                 330                 335

Val Arg Met Gly Arg Val Ser Glu Tyr Arg Gly Gln Gln Ala Gly Asp
            340                 345                 350

Leu Ala Gln Ser His Tyr Trp Thr Ser Ala Gly Gln Thr Glu Ile Lys
        355                 360                 365

Glu Ser Glu Ile Phe Val Arg Leu Trp Val Gly Ala Phe Ala Leu Ala
    370                 375                 380

Gly Arg Ser Met Lys Ala Trp Ile Asp Pro Met Gly Lys Ile Val Asn
385                 390                 395                 400

Thr Glu Lys Asn Asp Arg Asp Leu Thr Ala Ala Val Asn Ile Arg Gln
                405                 410                 415

Val Ile Ser Asn Lys Glu Met Val Ala Glu Ala Met Ala Arg Arg Gly
            420                 425                 430

Ile Tyr Phe Gly Glu Thr Pro Glu Leu Asp Arg Leu Gly Ala Glu Gly
        435                 440                 445

Asn Glu Gly Phe Val Phe Ala Leu Leu Arg Tyr Leu Arg Gly Cys Arg
    450                 455                 460

Asn Gln Thr Phe His Leu Gly Ala Arg Ala Gly Phe Leu Lys Glu Ile
465                 470                 475                 480

Arg Lys Glu Leu Glu Lys Thr Arg Trp Gly Lys Ala Lys Glu Ala Glu
```

485 490 495

His Val Val Leu Thr Asp Lys Thr Val Ala Ala Ile Arg Ala Ile Ile
500 505 510

Asp Asn Asp Ala Lys Ala Leu Gly Ala Arg Leu Leu Ala Asp Leu Ser
515 520 525

Gly Ala Phe Val Ala His Tyr Ala Ser Lys Glu His Phe Ser Thr Leu
530 535 540

Tyr Ser Glu Ile Val Lys Ala Val Lys Asp Ala Pro Glu Val Ser Ser
545 550 555 560

Gly Leu Pro Arg Leu Lys Leu Leu Leu Lys Arg Ala Asp Gly Val Arg
565 570 575

Gly Tyr Val His Gly Leu Arg Asp Thr Arg Lys His Ala Phe Ala Thr
580 585 590

Lys Leu Pro Pro Pro Pro Ala Pro Arg Glu Leu Asp Asp Pro Ala Thr
595 600 605

Lys Ala Arg Tyr Ile Ala Leu Leu Arg Leu Tyr Asp Gly Pro Phe Arg
610 615 620

Ala Tyr Ala Ser Gly Ile Thr Gly Thr Ala Leu Ala Gly Pro Ala Ala
625 630 635 640

Arg Ala Lys Glu Ala Ala Thr Ala Leu Ala Gln Ser Val Asn Val Thr
645 650 655

Lys Ala Tyr Ser Asp Val Met Glu Gly Arg Thr Ser Arg Leu Arg Pro
660 665 670

Pro Asn Asp Gly Glu Thr Leu Arg Glu Tyr Leu Ser Ala Leu Thr Gly
675 680 685

Glu Thr Ala Thr Glu Phe Arg Val Gln Ile Gly Tyr Glu Ser Asp Ser

```
    690                 695                 700

Glu Asn Ala Arg Lys Gln Ala Glu Phe Ile Glu Asn Tyr Arg Arg Asp
705                 710                 715                 720

Met Leu Ala Phe Met Phe Glu Asp Tyr Ile Arg Ala Lys Gly Phe Asp
                725                 730                 735

Trp Ile Leu Lys Ile Glu Pro Gly Ala Thr Ala Met Thr Arg Ala Pro
            740                 745                 750

Val Leu Pro Glu Pro Ile Asp Thr Arg Gly Gln Tyr Glu His Trp Gln
        755                 760                 765

Ala Ala Leu Tyr Leu Val Met His Phe Val Pro Ala Ser Asp Val Ser
    770                 775                 780

Asn Leu Leu His Gln Leu Arg Lys Trp Glu Ala Leu Gln Gly Lys Tyr
785                 790                 795                 800

Glu Leu Val Gln Asp Gly Asp Ala Thr Asp Gln Ala Asp Ala Arg Arg
                805                 810                 815

Glu Ala Leu Asp Leu Val Lys Arg Phe Arg Asp Val Leu Val Leu Phe
            820                 825                 830

Leu Lys Thr Gly Glu Ala Arg Phe Glu Gly Arg Ala Ala Pro Phe Asp
        835                 840                 845

Leu Lys Pro Phe Arg Ala Leu Phe Ala Asn Pro Ala Thr Phe Asp Arg
    850                 855                 860

Leu Phe Met Ala Thr Pro Thr Thr Ala Arg Pro Ala Glu Asp Asp Pro
865                 870                 875                 880

Glu Gly Asp Gly Ala Ser Glu Pro Glu Leu Arg Val Ala Arg Thr Leu
                885                 890                 895

Arg Gly Leu Arg Gln Ile Ala Arg Tyr Asn His Met Ala Val Leu Ser
```

```
            900                 905                 910

Asp Leu Phe Ala Lys His Lys Val Arg Asp Glu Glu Val Ala Arg Leu
        915                 920                 925

Ala Glu Ile Glu Asp Glu Thr Gln Glu Lys Ser Gln Ile Val Ala Ala
    930                 935                 940

Gln Glu Leu Arg Thr Asp Leu His Asp Lys Val Met Lys Cys His Pro
945                 950                 955                 960

Lys Thr Ile Ser Pro Glu Glu Arg Gln Ser Tyr Ala Ala Ala Ile Lys
                965                 970                 975

Thr Ile Glu Glu His Arg Phe Leu Val Gly Arg Val Tyr Leu Gly Asp
            980                 985                 990

His Leu Arg Leu His Arg Leu Met  Met Asp Val Ile Gly  Arg Leu Ile
        995                 1000                1005

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

Asn Ala  Ser Lys Gln Leu Gly  Ala Gly Ala Asp Trp  Ala Val Thr
    1025                1030                1035

Ile Ala  Gly Ala Ala Asn Thr  Asp Ala Arg Thr Gln  Thr Arg Lys
    1040                1045                1050

Asp Leu  Ala His Phe Asn Val  Leu Asp Arg Ala Asp  Gly Thr Pro
    1055                1060                1065

Asp Leu  Thr Ala Leu Val Asn  Arg Ala Arg Glu Met  Met Ala Tyr
    1070                1075                1080

Asp Arg  Lys Arg Lys Asn Ala  Val Pro Arg Ser Ile  Leu Asp Met
    1085                1090                1095

Leu Ala  Arg Leu Gly Leu Thr  Leu Lys Trp Gln Met  Lys Asp His
```

```
    1100                1105                1110

Leu Leu  Gln Asp Ala Thr Ile  Thr Gln Ala Ala Ile  Lys His Leu
    1115                1120                1125

Asp Lys  Val Arg Leu Thr Val  Gly Gly Pro Ala Ala  Val Thr Glu
    1130                1135                1140

Ala Arg  Phe Ser Gln Asp Tyr  Leu Gln Met Val Ala  Ala Val Phe
    1145                1150                1155

Asn Gly  Ser Val Gln Asn Pro  Lys Pro Arg Arg Arg  Asp Asp Gly
    1160                1165                1170

Asp Ala  Trp His Lys Pro Pro  Lys Pro Ala Thr Ala  Gln Ser Gln
    1175                1180                1185

Pro Asp  Gln Lys Pro Pro Asn  Lys Ala Pro Ser Ala  Gly Ser Arg
    1190                1195                1200

Leu Pro  Pro Pro Gln Val Gly  Glu Val Tyr Glu Gly  Val Val Val
    1205                1210                1215

Lys Val  Ile Asp Thr Gly Ser  Leu Gly Phe Leu Ala  Val Glu Gly
    1220                1225                1230

Val Ala  Gly Asn Ile Gly Leu  His Ile Ser Arg Leu  Arg Arg Ile
    1235                1240                1245

Arg Glu  Asp Ala Ile Ile Val  Gly Arg Arg Tyr Arg  Phe Arg Val
    1250                1255                1260

Glu Ile  Tyr Val Pro Pro Lys  Ser Asn Thr Ser Lys  Leu Asn Ala
    1265                1270                1275

Ala Asp  Leu Val Arg Ile Asp
    1280                1285
```

<210> 44

<211> 1285
<212> PRT
<213> Rhodobacter capsulatus

<400> 44

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

Pro Ala Gly Gly Leu Lys Arg Lys Ile Ser Thr Asp Gly Lys Asn Arg
20 25 30

Lys Glu Leu Pro Ala His Leu Ser Ser Asp Pro Lys Ala Leu Ile Gly
35 40 45

Gln Trp Ile Ser Gly Ile Asp Lys Ile Tyr Arg Lys Pro Asp Ser Arg
50 55 60

Lys Ser Asp Gly Lys Ala Ile His Ser Pro Thr Pro Ser Lys Met Gln
65 70 75 80

Phe Asp Ala Arg Asp Asp Leu Gly Glu Ala Phe Trp Lys Leu Val Ser
85 90 95

Glu Ala Gly Leu Ala Gln Asp Ser Asp Tyr Asp Gln Phe Lys Arg Arg
100 105 110

Leu His Pro Tyr Gly Asp Lys Phe Gln Pro Ala Asp Ser Gly Ala Lys
115 120 125

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

Trp Tyr Gly Ala Met Ser Lys Arg Gly Asn Asp Ala Lys Glu Leu Ala
145 150 155 160

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

Gln Pro Lys Arg Asn Pro Lys Thr Asp Lys Phe Ala Pro Gly Leu Val
180 185 190

```
Val Ala Arg Ala Leu Gly Ile Glu Ser Ser Val Leu Pro Arg Gly Met
        195                 200                 205

Ala Arg Leu Ala Arg Asn Trp Gly Glu Glu Glu Ile Gln Thr Tyr Phe
    210                 215                 220

Val Val Asp Val Ala Ala Ser Val Lys Glu Val Ala Lys Ala Ala Val
225                 230                 235                 240

Ser Ala Ala Gln Ala Phe Asp Pro Pro Arg Gln Val Ser Gly Arg Ser
                245                 250                 255

Leu Ser Pro Lys Val Gly Phe Ala Leu Ala Glu His Leu Glu Arg Val
            260                 265                 270

Thr Gly Ser Lys Arg Cys Ser Phe Asp Pro Ala Ala Gly Pro Ser Val
        275                 280                 285

Leu Ala Leu His Asp Glu Val Lys Lys Thr Tyr Lys Arg Leu Cys Ala
    290                 295                 300

Arg Gly Lys Asn Ala Ala Arg Ala Phe Pro Ala Asp Lys Thr Glu Leu
305                 310                 315                 320

Leu Ala Leu Met Arg His Thr His Glu Asn Arg Val Arg Asn Gln Met
                325                 330                 335

Val Arg Met Gly Arg Val Ser Glu Tyr Arg Gly Gln Gln Ala Gly Asp
            340                 345                 350

Leu Ala Gln Ser His Tyr Trp Thr Ser Ala Gly Gln Thr Glu Ile Lys
        355                 360                 365

Glu Ser Glu Ile Phe Val Arg Leu Trp Val Gly Ala Phe Ala Leu Ala
    370                 375                 380

Gly Arg Ser Met Lys Ala Trp Ile Asp Pro Met Gly Lys Ile Val Asn
385                 390                 395                 400
```

```
Thr Glu Lys Asn Asp Arg Asp Leu Thr Ala Ala Val Asn Ile Arg Gln
                405                 410                 415

Val Ile Ser Asn Lys Glu Met Val Ala Glu Ala Met Ala Arg Arg Gly
            420                 425                 430

Ile Tyr Phe Gly Glu Thr Pro Glu Leu Asp Arg Leu Gly Ala Glu Gly
        435                 440                 445

Asn Glu Gly Phe Val Phe Ala Leu Leu Arg Tyr Leu Arg Gly Cys Arg
    450                 455                 460

Asn Gln Thr Phe His Leu Gly Ala Arg Ala Gly Phe Leu Lys Glu Ile
465                 470                 475                 480

Arg Lys Glu Leu Glu Lys Thr Arg Trp Gly Lys Ala Lys Glu Ala Glu
                485                 490                 495

His Val Val Leu Thr Asp Lys Thr Val Ala Ala Ile Arg Ala Ile Ile
            500                 505                 510

Asp Asn Asp Ala Lys Ala Leu Gly Ala Arg Leu Leu Ala Asp Leu Ser
        515                 520                 525

Gly Ala Phe Val Ala His Tyr Ala Ser Lys Glu His Phe Ser Thr Leu
    530                 535                 540

Tyr Ser Glu Ile Val Lys Ala Val Lys Asp Ala Pro Glu Val Ser Ser
545                 550                 555                 560

Gly Leu Pro Arg Leu Lys Leu Leu Leu Lys Arg Ala Asp Gly Val Arg
                565                 570                 575

Gly Tyr Val His Gly Leu Arg Asp Thr Arg Lys His Ala Phe Ala Thr
            580                 585                 590

Lys Leu Pro Pro Pro Pro Ala Pro Arg Glu Leu Asp Asp Pro Ala Thr
        595                 600                 605
```

```
Lys Ala Arg Tyr Ile Ala Leu Leu Arg Leu Tyr Asp Gly Pro Phe Arg
    610                 615                 620

Ala Tyr Ala Ser Gly Ile Thr Gly Thr Ala Leu Ala Gly Pro Ala Ala
625                 630                 635                 640

Arg Ala Lys Glu Ala Ala Thr Ala Leu Ala Gln Ser Val Asn Val Thr
                645                 650                 655

Lys Ala Tyr Ser Asp Val Met Glu Gly Arg Ser Ser Arg Leu Arg Pro
            660                 665                 670

Pro Asn Asp Gly Glu Thr Leu Arg Glu Tyr Leu Ser Ala Leu Thr Gly
        675                 680                 685

Glu Thr Ala Thr Glu Phe Arg Val Gln Ile Gly Tyr Glu Ser Asp Ser
    690                 695                 700

Glu Asn Ala Arg Lys Gln Ala Glu Phe Ile Glu Asn Tyr Arg Arg Asp
705                 710                 715                 720

Met Leu Ala Phe Met Phe Glu Asp Tyr Ile Arg Ala Lys Gly Phe Asp
                725                 730                 735

Trp Ile Leu Lys Ile Glu Pro Gly Ala Thr Ala Met Thr Arg Ala Pro
            740                 745                 750

Val Leu Pro Glu Pro Ile Asp Thr Arg Gly Gln Tyr Glu His Trp Gln
        755                 760                 765

Ala Ala Leu Tyr Leu Val Met His Phe Val Pro Ala Ser Asp Val Ser
    770                 775                 780

Asn Leu Leu His Gln Leu Arg Lys Trp Glu Ala Leu Gln Gly Lys Tyr
785                 790                 795                 800

Glu Leu Val Gln Asp Gly Asp Ala Thr Asp Gln Ala Asp Ala Arg Arg
                805                 810                 815
```

```
Glu Ala Leu Asp Leu Val Lys Arg Phe Arg Asp Val Leu Val Leu Phe
            820                 825                 830

Leu Lys Thr Gly Glu Ala Arg Phe Glu Gly Arg Ala Ala Pro Phe Asp
        835                 840                 845

Leu Lys Pro Phe Arg Ala Leu Phe Ala Asn Pro Ala Thr Phe Asp Arg
    850                 855                 860

Leu Phe Met Ala Thr Pro Thr Thr Ala Arg Pro Ala Glu Asp Asp Pro
865                 870                 875                 880

Glu Gly Asp Gly Ala Ser Glu Pro Glu Leu Arg Val Ala Arg Thr Leu
                885                 890                 895

Arg Gly Leu Arg Gln Ile Ala Arg Tyr Asn His Met Ala Val Leu Ser
            900                 905                 910

Asp Leu Phe Ala Lys His Lys Val Arg Asp Glu Glu Val Ala Arg Leu
        915                 920                 925

Ala Glu Ile Glu Asp Glu Thr Gln Glu Lys Ser Gln Ile Val Ala Ala
    930                 935                 940

Gln Glu Leu Arg Thr Asp Leu His Asp Lys Val Met Lys Cys His Pro
945                 950                 955                 960

Lys Thr Ile Ser Pro Glu Glu Arg Gln Ser Tyr Ala Ala Ala Ile Lys
                965                 970                 975

Thr Ile Glu Glu His Arg Phe Leu Val Gly Arg Val Tyr Leu Gly Asp
            980                 985                 990

His Leu Arg Leu His Arg Leu Met  Met Asp Val Ile Gly  Arg Leu Ile
        995                 1000                1005

Asp Tyr  Ala Gly Ala Tyr Glu  Arg Asp Thr Gly Thr  Phe Leu Ile
    1010                1015                1020
```

```
Asn Ala  Ser Lys Gln Leu Gly  Ala Gly Ala Asp Trp  Ala Val Thr
    1025                1030                1035

Ile Ala  Gly Ala Ala Asn Thr  Asp Ala Arg Thr Gln  Thr Arg Lys
    1040                1045                1050

Asp Leu  Ala His Phe Asn Val  Leu Asp Arg Ala Asp  Gly Thr Pro
    1055                1060                1065

Asp Leu  Thr Ala Leu Val Asn  Arg Ala Arg Glu Met  Met Ala Tyr
    1070                1075                1080

Asp Arg  Lys Arg Lys Asn Ala  Val Pro Arg Ser Ile  Leu Asp Met
    1085                1090                1095

Leu Ala  Arg Leu Gly Leu Thr  Leu Lys Trp Gln Met  Lys Asp His
    1100                1105                1110

Leu Leu  Gln Asp Ala Thr Ile  Thr Gln Ala Ala Ile  Lys His Leu
    1115                1120                1125

Asp Lys  Val Arg Leu Thr Val  Gly Gly Pro Ala Ala  Val Thr Glu
    1130                1135                1140

Ala Arg  Phe Ser Gln Asp Tyr  Leu Gln Met Val Ala  Ala Val Phe
    1145                1150                1155

Asn Gly  Ser Val Gln Asn Pro  Lys Pro Arg Arg Arg  Asp Asp Gly
    1160                1165                1170

Asp Ala  Trp His Lys Pro Pro  Lys Pro Ala Thr Ala  Gln Ser Gln
    1175                1180                1185

Pro Asp  Gln Lys Pro Pro Asn  Lys Ala Pro Ser Ala  Gly Ser Arg
    1190                1195                1200

Leu Pro  Pro Pro Gln Val Gly  Glu Val Tyr Glu Gly  Val Val Val
    1205                1210                1215
```

```
Lys Val  Ile Asp Thr Gly Ser  Leu Gly Phe Leu Ala  Val Glu Gly
    1220                1225                1230


Val Ala  Gly Asn Ile Gly Leu  His Ile Ser Arg Leu  Arg Arg Ile
    1235                1240                1245


Arg Glu  Asp Ala Ile Ile Val  Gly Arg Arg Tyr Arg  Phe Arg Val
    1250                1255                1260


Glu Ile  Tyr Val Pro Pro Lys  Ser Asn Thr Ser Lys  Leu Asn Ala
    1265                1270                1275


Ala Asp  Leu Val Arg Ile Asp
    1280                1285


<210>  45
<211>  1285
<212>  PRT
<213>  Rhodobacter capsulatus

<400>  45

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


Pro Ala Gly Gly Leu Lys Arg Lys Ile Ser Thr Asp Gly Lys Asn Arg
            20                  25                  30


Lys Glu Leu Pro Ala His Leu Ser Ser Asp Pro Lys Ala Leu Ile Gly
        35                  40                  45


Gln Trp Ile Ser Gly Ile Asp Lys Ile Tyr Arg Lys Pro Asp Ser Arg
    50                  55                  60


Lys Ser Asp Gly Lys Ala Ile His Ser Pro Thr Pro Ser Lys Met Gln
65                  70                  75                  80


Phe Asp Ala Arg Asp Asp Leu Gly Glu Ala Phe Trp Lys Leu Val Ser
                85                  90                  95
```

```
Glu Ala Gly Leu Ala Gln Asp Ser Asp Tyr Asp Gln Phe Lys Arg Arg
            100                 105                 110

Leu His Pro Tyr Gly Asp Lys Phe Gln Pro Ala Asp Ser Gly Ala Lys
        115                 120                 125

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

Trp Tyr Gly Ala Met Ser Lys Arg Gly Asn Asp Ala Lys Glu Leu Ala
145                 150                 155                 160

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

Gln Pro Lys Arg Asn Pro Lys Thr Asp Lys Phe Ala Pro Gly Leu Val
            180                 185                 190

Val Ala Arg Ala Leu Gly Ile Glu Ser Ser Val Leu Pro Arg Gly Met
        195                 200                 205

Ala Arg Leu Ala Arg Asn Trp Gly Glu Glu Glu Ile Gln Thr Tyr Phe
    210                 215                 220

Val Val Asp Val Ala Ala Ser Val Lys Glu Val Ala Lys Ala Ala Val
225                 230                 235                 240

Ser Ala Ala Gln Ala Phe Asp Pro Pro Arg Gln Val Ser Gly Arg Ser
                245                 250                 255

Leu Ser Pro Lys Val Gly Phe Ala Leu Ala Glu His Leu Glu Arg Val
            260                 265                 270

Thr Gly Ser Lys Arg Cys Ser Phe Asp Pro Ala Ala Gly Pro Ser Val
        275                 280                 285

Leu Ala Leu His Asp Glu Val Lys Lys Thr Tyr Lys Arg Leu Cys Ala
    290                 295                 300
```

```
Arg Gly Lys Asn Ala Ala Arg Ala Phe Pro Ala Asp Lys Thr Glu Leu
305                 310                 315                 320

Leu Ala Leu Met Arg His Thr His Glu Asn Arg Val Arg Asn Gln Met
                325                 330                 335

Val Arg Met Gly Arg Val Ser Glu Tyr Arg Gly Gln Gln Ala Gly Asp
            340                 345                 350

Leu Ala Gln Ser His Tyr Trp Thr Ser Ala Gly Gln Thr Glu Ile Lys
        355                 360                 365

Glu Ser Glu Ile Phe Val Arg Leu Trp Val Gly Ala Phe Ala Leu Ala
    370                 375                 380

Gly Arg Ser Met Lys Ala Trp Ile Asp Pro Met Gly Lys Ile Val Asn
385                 390                 395                 400

Thr Glu Lys Asn Asp Arg Asp Leu Thr Ala Ala Val Asn Ile Arg Gln
                405                 410                 415

Val Ile Ser Asn Lys Glu Met Val Ala Glu Ala Met Ala Arg Arg Gly
            420                 425                 430

Ile Tyr Phe Gly Glu Thr Pro Glu Leu Asp Arg Leu Gly Ala Glu Gly
        435                 440                 445

Asn Glu Gly Phe Val Phe Ala Leu Leu Arg Tyr Leu Arg Gly Cys Arg
    450                 455                 460

Asn Gln Thr Phe His Leu Gly Ala Arg Ala Gly Phe Leu Lys Glu Ile
465                 470                 475                 480

Arg Lys Glu Leu Glu Lys Thr Arg Trp Gly Lys Ala Lys Glu Ala Glu
                485                 490                 495

His Val Val Leu Thr Asp Lys Thr Val Ala Ala Ile Arg Ala Ile Ile
            500                 505                 510
```

```
Asp Asn Asp Ala Lys Ala Leu Gly Ala Arg Leu Leu Ala Asp Leu Ser
        515                 520                 525

Gly Ala Phe Val Ala His Tyr Ala Ser Lys Glu His Phe Ser Thr Leu
    530                 535                 540

Tyr Ser Glu Ile Val Lys Ala Val Lys Asp Ala Pro Glu Val Ser Ser
545                 550                 555                 560

Gly Leu Pro Arg Leu Lys Leu Leu Leu Lys Arg Ala Asp Gly Val Arg
                565                 570                 575

Gly Tyr Val His Gly Leu Arg Asp Thr Arg Lys His Ala Phe Ala Thr
            580                 585                 590

Lys Leu Pro Pro Pro Pro Ala Pro Arg Glu Leu Asp Asp Pro Ala Thr
        595                 600                 605

Lys Ala Arg Tyr Ile Ala Leu Leu Arg Leu Tyr Asp Gly Pro Phe Arg
    610                 615                 620

Ala Tyr Ala Ser Gly Ile Thr Gly Thr Ala Leu Ala Gly Pro Ala Ala
625                 630                 635                 640

Arg Ala Lys Glu Ala Ala Thr Ala Leu Ala Gln Ser Val Asn Val Thr
                645                 650                 655

Lys Ala Tyr Ser Asp Val Met Glu Gly Arg Ser Ser Arg Leu Arg Pro
            660                 665                 670

Pro Asn Asp Gly Glu Thr Leu Arg Glu Tyr Leu Ser Ala Leu Thr Gly
        675                 680                 685

Glu Thr Ala Thr Glu Phe Arg Val Gln Ile Gly Tyr Glu Ser Asp Ser
    690                 695                 700

Glu Asn Ala Arg Lys Gln Ala Glu Phe Ile Glu Asn Tyr Arg Arg Asp
705                 710                 715                 720
```

```
Met Leu Ala Phe Met Phe Glu Asp Tyr Ile Arg Ala Lys Gly Phe Asp
                725                 730                 735

Trp Ile Leu Lys Ile Glu Pro Gly Ala Thr Ala Met Thr Arg Ala Pro
            740                 745                 750

Val Leu Pro Glu Pro Ile Asp Thr Arg Gly Gln Tyr Glu His Trp Gln
        755                 760                 765

Ala Ala Leu Tyr Leu Val Met His Phe Val Pro Ala Ser Asp Val Ser
    770                 775                 780

Asn Leu Leu His Gln Leu Arg Lys Trp Glu Ala Leu Gln Gly Lys Tyr
785                 790                 795                 800

Glu Leu Val Gln Asp Gly Asp Ala Thr Asp Gln Ala Asp Ala Arg Arg
                805                 810                 815

Glu Ala Leu Asp Leu Val Lys Arg Phe Arg Asp Val Leu Val Leu Phe
            820                 825                 830

Leu Lys Thr Gly Glu Ala Arg Phe Glu Gly Arg Ala Ala Pro Phe Asp
        835                 840                 845

Leu Lys Pro Phe Arg Ala Leu Phe Ala Asn Pro Ala Thr Phe Asp Arg
    850                 855                 860

Leu Phe Met Ala Thr Pro Thr Thr Ala Arg Pro Ala Glu Asp Asp Pro
865                 870                 875                 880

Glu Gly Asp Gly Ala Ser Glu Pro Glu Leu Arg Val Ala Arg Thr Leu
                885                 890                 895

Arg Gly Leu Arg Gln Ile Ala Arg Tyr Asn His Met Ala Val Leu Ser
            900                 905                 910

Asp Leu Phe Ala Lys His Lys Val Arg Asp Glu Glu Val Ala Arg Leu
        915                 920                 925
```

```
Ala Glu Ile Glu Asp Glu Thr Gln Glu Lys Ser Gln Ile Val Ala Ala
    930                 935                 940

Gln Glu Leu Arg Thr Asp Leu His Asp Lys Val Met Lys Cys His Pro
945                 950                 955                 960

Lys Thr Ile Ser Pro Glu Glu Arg Gln Ser Tyr Ala Ala Ala Ile Lys
                965                 970                 975

Thr Ile Glu Glu His Arg Phe Leu Val Gly Arg Val Tyr Leu Gly Asp
            980                 985                 990

His Leu Arg Leu His Arg Leu Met  Met Asp Val Ile Gly  Arg Leu Ile
        995                 1000                1005

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

Asn Ala  Ser Lys Gln Leu Gly  Ala Gly Ala Asp Trp  Ala Val Thr
    1025                1030                1035

Ile Ala  Gly Ala Ala Asn Thr  Asp Ala Arg Thr Gln  Thr Arg Lys
    1040                1045                1050

Asp Leu  Ala His Phe Asn Val  Leu Asp Arg Ala Asp  Gly Thr Pro
    1055                1060                1065

Asp Leu  Thr Ala Leu Val Asn  Arg Ala Arg Glu Met  Met Ala Tyr
    1070                1075                1080

Asp Arg  Lys Arg Lys Asn Ala  Val Pro Arg Ser Ile  Leu Asp Met
    1085                1090                1095

Leu Ala  Arg Leu Gly Leu Thr  Leu Lys Trp Gln Met  Lys Asp His
    1100                1105                1110

Leu Leu  Gln Asp Ala Thr Ile  Thr Gln Ala Ala Ile  Lys His Leu
    1115                1120                1125
```

```
Asp Lys  Val Arg Leu Thr Val  Gly Gly Pro Ala Ala  Val Thr Glu
    1130                1135                1140


Ala Arg  Phe Ser Gln Asp Tyr  Leu Gln Met Val Ala  Ala Val Phe
    1145                1150                1155


Asn Gly  Ser Val Gln Asn Pro  Lys Pro Arg Arg Arg  Asp Asp Gly
    1160                1165                1170


Asp Ala  Trp His Lys Pro Pro  Lys Pro Ala Thr Ala  Gln Ser Gln
    1175                1180                1185


Pro Asp  Gln Lys Pro Pro Asn  Lys Ala Pro Ser Ala  Gly Ser Arg
    1190                1195                1200


Leu Pro  Pro Pro Gln Val Gly  Glu Val Tyr Glu Gly  Val Val Val
    1205                1210                1215


Lys Val  Ile Asp Thr Gly Ser  Leu Gly Phe Leu Ala  Val Glu Gly
    1220                1225                1230


Val Ala  Gly Asn Ile Gly Leu  His Ile Ser Arg Leu  Arg Arg Ile
    1235                1240                1245


Arg Glu  Asp Ala Ile Ile Val  Gly Arg Arg Tyr Arg  Phe Arg Val
    1250                1255                1260


Glu Ile  Tyr Val Pro Pro Lys  Ser Asn Thr Ser Lys  Leu Asn Ala
    1265                1270                1275


Ala Asp  Leu Val Arg Ile Asp
    1280                1285


<210>  46
<211>  1389
<212>  PRT
<213>  Unknown

<220>
<223>  Synthetic or Unknown
```

<400> 46

```
Met Gly Asn Leu Phe Gly His Lys Arg Trp Tyr Glu Val Arg Asp Lys
1               5                  10                  15

Lys Asp Phe Lys Ile Lys Arg Lys Val Lys Val Lys Arg Asn Tyr Asp
            20                  25                  30

Gly Asn Lys Tyr Ile Leu Asn Ile Asn Glu Asn Asn Asn Lys Glu Lys
        35                  40                  45

Ile Asp Asn Asn Lys Phe Ile Arg Lys Tyr Ile Asn Tyr Lys Lys Asn
    50                  55                  60

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

Leu Phe Lys Leu Lys Gly Lys Glu Gly Ile Ile Arg Ile Glu Asn Asn
                85                  90                  95

Asp Asp Phe Leu Glu Thr Glu Glu Val Val Leu Tyr Ile Glu Ala Tyr
            100                 105                 110

Gly Lys Ser Glu Lys Leu Lys Ala Leu Gly Ile Thr Lys Lys Lys Ile
        115                 120                 125

Ile Asp Glu Ala Ile Arg Gln Gly Ile Thr Lys Asp Asp Lys Lys Ile
    130                 135                 140

Glu Ile Lys Arg Gln Glu Asn Glu Glu Glu Ile Glu Ile Asp Ile Arg
145                 150                 155                 160

Asp Glu Tyr Thr Asn Lys Thr Leu Asn Asp Cys Ser Ile Ile Leu Arg
                165                 170                 175

Ile Ile Glu Asn Asp Glu Leu Glu Thr Lys Lys Ser Ile Tyr Glu Ile
            180                 185                 190

Phe Lys Asn Ile Asn Met Ser Leu Tyr Lys Ile Ile Glu Lys Ile Ile
```

```
        195                 200                 205

Glu Asn Glu Thr Glu Lys Val Phe Glu Asn Arg Tyr Tyr Glu Glu His
    210                 215                 220

Leu Arg Glu Lys Leu Leu Lys Asp Asp Lys Ile Asp Val Ile Leu Thr
225                 230                 235                 240

Asn Phe Met Glu Ile Arg Glu Lys Ile Lys Ser Asn Leu Glu Ile Leu
                245                 250                 255

Gly Phe Val Lys Phe Tyr Leu Asn Val Gly Gly Asp Lys Lys Lys Ser
            260                 265                 270

Lys Asn Lys Lys Met Leu Val Glu Lys Ile Leu Asn Ile Asn Val Asp
        275                 280                 285

Leu Thr Val Glu Asp Ile Ala Asp Phe Val Ile Lys Glu Leu Glu Phe
    290                 295                 300

Trp Asn Ile Thr Lys Arg Ile Glu Lys Val Lys Lys Val Asn Asn Glu
305                 310                 315                 320

Phe Leu Glu Lys Arg Arg Asn Arg Thr Tyr Ile Lys Ser Tyr Val Leu
                325                 330                 335

Leu Asp Lys His Glu Lys Phe Lys Ile Glu Arg Glu Asn Lys Lys Asp
            340                 345                 350

Lys Ile Val Lys Phe Phe Val Glu Asn Ile Lys Asn Asn Ser Ile Lys
        355                 360                 365

Glu Lys Ile Glu Lys Ile Leu Ala Glu Phe Lys Ile Asp Glu Leu Ile
    370                 375                 380

Lys Lys Leu Glu Lys Glu Leu Lys Lys Gly Asn Cys Asp Thr Glu Ile
385                 390                 395                 400

Phe Gly Ile Phe Lys Lys His Tyr Lys Val Asn Phe Asp Ser Lys Lys
```

405 410 415

Phe Ser Lys Lys Ser Asp Glu Glu Lys Glu Leu Tyr Lys Ile Ile Tyr
420 425 430

Arg Tyr Leu Lys Gly Arg Ile Glu Lys Ile Leu Val Asn Glu Gln Lys
435 440 445

Val Arg Leu Lys Lys Met Glu Lys Ile Glu Ile Glu Lys Ile Leu Asn
450 455 460

Glu Ser Ile Leu Ser Glu Lys Ile Leu Lys Arg Val Lys Gln Tyr Thr
465 470 475 480

Leu Glu His Ile Met Tyr Leu Gly Lys Leu Arg His Asn Asp Ile Asp
485 490 495

Met Thr Thr Val Asn Thr Asp Asp Phe Ser Arg Leu His Ala Lys Glu
500 505 510

Glu Leu Asp Leu Glu Leu Ile Thr Phe Phe Ala Ser Thr Asn Met Glu
515 520 525

Leu Asn Lys Ile Phe Ser Arg Glu Asn Ile Asn Asn Asp Glu Asn Ile
530 535 540

Asp Phe Phe Gly Gly Asp Arg Glu Lys Asn Tyr Val Leu Asp Lys Lys
545 550 555 560

Ile Leu Asn Ser Lys Ile Lys Ile Ile Arg Asp Leu Asp Phe Ile Asp
565 570 575

Asn Lys Asn Asn Ile Thr Asn Asn Phe Ile Arg Lys Phe Thr Lys Ile
580 585 590

Gly Thr Asn Glu Arg Asn Arg Ile Leu His Ala Ile Ser Lys Glu Arg
595 600 605

Asp Leu Gln Gly Thr Gln Asp Asp Tyr Asn Lys Val Ile Asn Ile Ile

610 615 620

Gln Asn Leu Lys Ile Ser Asp Glu Glu Val Ser Lys Ala Leu Asn Leu
625 630 635 640

Asp Val Val Phe Lys Asp Lys Lys Asn Ile Ile Thr Lys Ile Asn Asp
645 650 655

Ile Lys Ile Ser Glu Glu Asn Asn Asn Asp Ile Lys Tyr Leu Pro Ser
660 665 670

Phe Ser Lys Val Leu Pro Glu Ile Leu Asn Leu Tyr Arg Asn Asn Pro
675 680 685

Lys Asn Glu Pro Phe Asp Thr Ile Glu Thr Glu Lys Ile Val Leu Asn
690 695 700

Ala Leu Ile Tyr Val Asn Lys Glu Leu Tyr Lys Lys Leu Ile Leu Glu
705 710 715 720

Asp Asp Leu Glu Glu Asn Glu Ser Lys Asn Ile Phe Leu Gln Glu Leu
725 730 735

Lys Lys Thr Leu Gly Asn Ile Asp Glu Ile Asp Glu Asn Ile Ile Glu
740 745 750

Asn Tyr Tyr Lys Asn Ala Gln Ile Ser Ala Ser Lys Gly Asn Asn Lys
755 760 765

Ala Ile Lys Lys Tyr Gln Lys Lys Val Ile Glu Cys Tyr Ile Gly Tyr
770 775 780

Leu Arg Lys Asn Tyr Glu Glu Leu Phe Asp Phe Ser Asp Phe Lys Met
785 790 795 800

Asn Ile Gln Glu Ile Lys Lys Gln Ile Lys Asp Ile Asn Asp Asn Lys
805 810 815

Thr Tyr Glu Arg Ile Thr Val Lys Thr Ser Asp Lys Thr Ile Val Ile

820 825 830

Asn Asp Asp Phe Glu Tyr Ile Ile Ser Ile Phe Ala Leu Leu Asn Ser
835 840 845

Asn Ala Val Ile Asn Lys Ile Arg Asn Arg Phe Phe Ala Thr Ser Val
850 855 860

Trp Leu Asn Thr Ser Glu Tyr Gln Asn Ile Ile Asp Ile Leu Asp Glu
865 870 875 880

Ile Met Gln Leu Asn Thr Leu Arg Asn Glu Cys Ile Thr Glu Asn Trp
885 890 895

Asn Leu Asn Leu Glu Glu Phe Ile Gln Lys Met Lys Glu Ile Glu Lys
900 905 910

Asp Phe Asp Asp Phe Lys Ile Gln Thr Lys Lys Glu Ile Phe Asn Asn
915 920 925

Tyr Tyr Glu Asp Ile Lys Asn Asn Ile Leu Thr Glu Phe Lys Asp Asp
930 935 940

Ile Asn Gly Cys Asp Val Leu Glu Lys Lys Leu Glu Lys Ile Val Ile
945 950 955 960

Phe Asp Asp Glu Thr Lys Phe Glu Ile Asp Lys Lys Ser Asn Ile Leu
965 970 975

Gln Asp Glu Gln Arg Lys Leu Ser Asn Ile Asn Lys Lys Asp Leu Lys
980 985 990

Lys Lys Val Asp Gln Tyr Ile Lys Asp Lys Asp Gln Glu Ile Lys Ser
995 1000 1005

Lys Ile Leu Cys Arg Ile Ile Phe Asn Ser Asp Phe Leu Lys Lys
1010 1015 1020

Tyr Lys Lys Glu Ile Asp Asn Leu Ile Glu Asp Met Glu Ser Glu

```
    1025                1030                1035

Asn Glu  Asn Lys Phe Gln Glu  Ile Tyr Tyr Pro Lys  Glu Arg Lys
    1040                1045                1050

Asn Glu  Leu Tyr Ile Tyr Lys  Lys Asn Leu Phe Leu  Asn Ile Gly
    1055                1060                1065

Asn Pro  Asn Phe Asp Lys Ile  Tyr Gly Leu Ile Ser  Asn Asp Ile
    1070                1075                1080

Lys Met  Ala Asp Ala Lys Phe  Leu Phe Asn Ile Asp  Gly Lys Asn
    1085                1090                1095

Ile Arg  Lys Asn Lys Ile Ser  Glu Ile Asp Ala Ile  Leu Lys Asn
    1100                1105                1110

Leu Asn  Asp Lys Leu Asn Gly  Tyr Ser Lys Glu Tyr  Lys Glu Lys
    1115                1120                1125

Tyr Ile  Lys Lys Leu Lys Glu  Asn Asp Asp Phe Phe  Ala Lys Asn
    1130                1135                1140

Ile Gln  Asn Lys Asn Tyr Lys  Ser Phe Glu Lys Asp  Tyr Asn Arg
    1145                1150                1155

Val Ser  Glu Tyr Lys Lys Ile  Arg Asp Leu Val Glu  Phe Asn Tyr
    1160                1165                1170

Leu Asn  Lys Ile Glu Ser Tyr  Leu Ile Asp Ile Asn  Trp Lys Leu
    1175                1180                1185

Ala Ile  Gln Met Ala Arg Phe  Glu Arg Asp Met His  Tyr Ile Val
    1190                1195                1200

Asn Gly  Leu Arg Glu Leu Gly  Ile Ile Lys Leu Ser  Gly Tyr Asn
    1205                1210                1215

Thr Gly  Ile Ser Arg Ala Tyr  Pro Lys Arg Asn Gly  Ser Asp Gly
```

```
    1220                1225                1230

Phe Tyr  Thr Thr Thr Ala Tyr  Tyr Lys Phe Phe Asp  Glu Glu Ser
    1235                1240                1245

Tyr Lys  Lys Phe Glu Lys Ile  Cys Tyr Gly Phe Gly  Ile Asp Leu
    1250                1255                1260

Ser Glu  Asn Ser Glu Ile Asn  Lys Pro Glu Asn Glu  Ser Ile Arg
    1265                1270                1275

Asn Tyr  Ile Ser His Phe Tyr  Ile Val Arg Asn Pro  Phe Ala Asp
    1280                1285                1290

Tyr Ser  Ile Ala Glu Gln Ile  Asp Arg Val Ser Asn  Leu Leu Ser
    1295                1300                1305

Tyr Ser  Thr Arg Tyr Asn Asn  Ser Thr Tyr Ala Ser  Val Phe Glu
    1310                1315                1320

Val Phe  Lys Lys Asp Val Asn  Leu Asp Tyr Asp Glu  Leu Lys Lys
    1325                1330                1335

Lys Phe  Lys Leu Ile Gly Asn  Asn Asp Ile Leu Glu  Arg Leu Met
    1340                1345                1350

Lys Pro  Lys Lys Val Ser Val  Leu Glu Leu Glu Ser  Tyr Asn Ser
    1355                1360                1365

Asp Tyr  Ile Lys Asn Leu Ile  Ile Glu Leu Leu Thr  Lys Ile Glu
    1370                1375                1380

Asn Thr  Asn Asp Thr Leu
    1385
```

<210> 47
<211> 1197
<212> PRT
<213> L. wadei

<400> 47

```
Met Lys Val Thr Lys Val Asp Gly Ile Ser His Lys Lys Tyr Ile Glu
1               5                  10                  15

Glu Gly Lys Leu Val Lys Ser Thr Ser Glu Glu Asn Arg Thr Ser Glu
            20                  25                  30

Arg Leu Ser Glu Leu Leu Ser Ile Arg Leu Asp Ile Tyr Ile Lys Asn
        35                  40                  45

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

Lys Lys Phe Phe Ser Asn Lys Val Leu His Leu Lys Asp Ser Val Leu
65                  70                  75                  80

Tyr Leu Lys Asn Arg Lys Glu Lys Asn Ala Val Gln Asp Lys Asn Tyr
                85                  90                  95

Ser Glu Glu Asp Ile Ser Glu Tyr Asp Leu Lys Asn Lys Asn Ser Phe
            100                 105                 110

Ser Val Leu Lys Lys Ile Leu Leu Asn Glu Asp Val Asn Ser Glu Glu
        115                 120                 125

Leu Glu Ile Phe Arg Lys Asp Val Glu Ala Lys Leu Asn Lys Ile Asn
    130                 135                 140

Ser Leu Lys Tyr Ser Phe Glu Glu Asn Lys Ala Asn Tyr Gln Lys Ile
145                 150                 155                 160

Asn Glu Asn Asn Val Glu Lys Val Gly Gly Lys Ser Lys Arg Asn Ile
                165                 170                 175

Ile Tyr Asp Tyr Tyr Arg Glu Ser Ala Lys Arg Asn Asp Tyr Ile Asn
            180                 185                 190

Asn Val Gln Glu Ala Phe Asp Lys Leu Tyr Lys Lys Glu Asp Ile Glu
        195                 200                 205
```

```
Lys Leu Phe Phe Leu Ile Glu Asn Ser Lys Lys His Glu Lys Tyr Lys
    210                 215                 220

Ile Arg Glu Tyr Tyr His Lys Ile Ile Gly Arg Lys Asn Asp Lys Glu
225                 230                 235                 240

Asn Phe Ala Lys Ile Ile Tyr Glu Glu Ile Gln Asn Val Asn Asn Ile
                245                 250                 255

Lys Glu Leu Ile Glu Lys Ile Pro Asp Met Ser Glu Leu Lys Lys Ser
            260                 265                 270

Gln Val Phe Tyr Lys Tyr Tyr Leu Asp Lys Glu Glu Leu Asn Asp Lys
        275                 280                 285

Asn Ile Lys Tyr Ala Phe Cys His Phe Val Glu Ile Glu Met Ser Gln
    290                 295                 300

Leu Leu Lys Asn Tyr Val Tyr Lys Arg Leu Ser Asn Ile Ser Asn Asp
305                 310                 315                 320

Lys Ile Lys Arg Ile Phe Glu Tyr Gln Asn Leu Lys Lys Leu Ile Glu
                325                 330                 335

Asn Lys Leu Leu Asn Lys Leu Asp Thr Tyr Val Arg Asn Cys Gly Lys
            340                 345                 350

Tyr Asn Tyr Tyr Leu Gln Val Gly Glu Ile Ala Thr Ser Asp Phe Ile
        355                 360                 365

Ala Arg Asn Arg Gln Asn Glu Ala Phe Leu Arg Asn Ile Ile Gly Val
    370                 375                 380

Ser Ser Val Ala Tyr Phe Ser Leu Arg Asn Ile Leu Glu Thr Glu Asn
385                 390                 395                 400

Glu Asn Asp Ile Thr Gly Arg Met Arg Gly Lys Thr Val Lys Asn Asn
                405                 410                 415
```

```
Lys Gly Glu Glu Lys Tyr Val Ser Gly Glu Val Asp Lys Ile Tyr Asn
            420                 425                 430

Glu Asn Lys Gln Asn Glu Val Lys Glu Asn Leu Lys Met Phe Tyr Ser
        435                 440                 445

Tyr Asp Phe Asn Met Asp Asn Lys Asn Glu Ile Glu Asp Phe Phe Ala
    450                 455                 460

Asn Ile Asp Glu Ala Ile Ser Ser Ile Arg His Gly Ile Val His Phe
465                 470                 475                 480

Asn Leu Glu Leu Glu Gly Lys Asp Ile Phe Ala Phe Lys Asn Ile Ala
                485                 490                 495

Pro Ser Glu Ile Ser Lys Lys Met Phe Gln Asn Glu Ile Asn Glu Lys
            500                 505                 510

Lys Leu Lys Leu Lys Ile Phe Lys Gln Leu Asn Ser Ala Asn Val Phe
        515                 520                 525

Asn Tyr Tyr Glu Lys Asp Val Ile Ile Lys Tyr Leu Lys Asn Thr Lys
    530                 535                 540

Phe Asn Phe Val Asn Lys Asn Ile Pro Phe Val Pro Ser Phe Thr Lys
545                 550                 555                 560

Leu Tyr Asn Lys Ile Glu Asp Leu Arg Asn Thr Leu Lys Phe Phe Trp
                565                 570                 575

Ser Val Pro Lys Asp Lys Glu Glu Lys Asp Ala Gln Ile Tyr Leu Leu
            580                 585                 590

Lys Asn Ile Tyr Tyr Gly Glu Phe Leu Asn Lys Phe Val Lys Asn Ser
        595                 600                 605

Lys Val Phe Phe Lys Ile Thr Asn Glu Val Ile Lys Ile Asn Lys Gln
    610                 615                 620
```

```
Arg Asn Gln Lys Thr Gly His Tyr Lys Tyr Gln Lys Phe Glu Asn Ile
625                 630                 635                 640

Glu Lys Thr Val Pro Val Glu Tyr Leu Ala Ile Ile Gln Ser Arg Glu
                645                 650                 655

Met Ile Asn Asn Gln Asp Lys Glu Glu Lys Asn Thr Tyr Ile Asp Phe
            660                 665                 670

Ile Gln Gln Ile Phe Leu Lys Gly Phe Ile Asp Tyr Leu Asn Lys Asn
        675                 680                 685

Asn Leu Lys Tyr Ile Glu Ser Asn Asn Asn Asn Asp Asn Asn Asp Ile
    690                 695                 700

Phe Ser Lys Ile Lys Ile Lys Lys Asp Asn Lys Glu Lys Tyr Asp Lys
705                 710                 715                 720

Ile Leu Lys Asn Tyr Glu Lys His Asn Arg Asn Lys Glu Ile Pro His
                725                 730                 735

Glu Ile Asn Glu Phe Val Arg Glu Ile Lys Leu Gly Lys Ile Leu Lys
            740                 745                 750

Tyr Thr Glu Asn Leu Asn Met Phe Tyr Leu Ile Leu Lys Leu Leu Asn
        755                 760                 765

His Lys Glu Leu Thr Asn Leu Lys Gly Ser Leu Glu Lys Tyr Gln Ser
    770                 775                 780

Ala Asn Lys Glu Glu Thr Phe Ser Asp Glu Leu Glu Leu Ile Asn Leu
785                 790                 795                 800

Leu Asn Leu Asp Asn Asn Arg Val Thr Glu Asp Phe Glu Leu Glu Ala
                805                 810                 815

Asn Glu Ile Gly Lys Phe Leu Asp Phe Asn Glu Asn Lys Ile Lys Asp
            820                 825                 830
```

```
Arg Lys Glu Leu Lys Lys Phe Asp Thr Asn Lys Ile Tyr Phe Asp Gly
        835                 840                 845

Glu Asn Ile Ile Lys His Arg Ala Phe Tyr Asn Ile Lys Lys Tyr Gly
    850                 855                 860

Met Leu Asn Leu Leu Glu Lys Ile Ala Asp Lys Ala Lys Tyr Lys Ile
865                 870                 875                 880

Ser Leu Lys Glu Leu Lys Glu Tyr Ser Asn Lys Lys Asn Glu Ile Glu
                885                 890                 895

Lys Asn Tyr Thr Met Gln Gln Asn Leu His Arg Lys Tyr Ala Arg Pro
            900                 905                 910

Lys Lys Asp Glu Lys Phe Asn Asp Glu Asp Tyr Lys Glu Tyr Glu Lys
        915                 920                 925

Ala Ile Gly Asn Ile Gln Lys Tyr Thr His Leu Lys Asn Lys Val Glu
    930                 935                 940

Phe Asn Glu Leu Asn Leu Leu Gln Gly Leu Leu Leu Lys Ile Leu His
945                 950                 955                 960

Arg Leu Val Gly Tyr Thr Ser Ile Trp Glu Arg Asp Leu Arg Phe Arg
                965                 970                 975

Leu Lys Gly Glu Phe Pro Glu Asn His Tyr Ile Glu Glu Ile Phe Asn
            980                 985                 990

Phe Asp Asn Ser Lys Asn Val Lys  Tyr Lys Ser Gly Gln  Ile Val Glu
        995                 1000                1005

Lys Tyr  Ile Asn Phe Tyr Lys  Glu Leu Tyr Lys Asp  Asn Val Glu
    1010                1015                1020

Lys Arg  Ser Ile Tyr Ser Asp  Lys Lys Val Lys Lys  Leu Lys Gln
    1025                1030                1035
```

```
Glu Lys  Lys Asp Leu Tyr Ile  Arg Asn Tyr Ile Ala  His Phe Asn
    1040                1045                1050

Tyr Ile  Pro His Ala Glu Ile  Ser Leu Leu Glu Val  Leu Glu Asn
    1055                1060                1065

Leu Arg  Lys Leu Leu Ser Tyr  Asp Arg Lys Leu Lys  Asn Ala Ile
    1070                1075                1080

Met Lys  Ser Ile Val Asp Ile  Leu Lys Glu Tyr Gly  Phe Val Ala
    1085                1090                1095

Thr Phe  Lys Ile Gly Ala Asp  Lys Lys Ile Glu Ile  Gln Thr Leu
    1100                1105                1110

Glu Ser  Glu Lys Ile Val His  Leu Lys Asn Leu Lys  Lys Lys Lys
    1115                1120                1125

Leu Met  Thr Asp Arg Asn Ser  Glu Glu Leu Cys Glu  Leu Val Lys
    1130                1135                1140

Val Met  Phe Glu Tyr Lys Ala  Leu Glu Lys Arg Pro  Ala Ala Thr
    1145                1150                1155

Lys Lys  Ala Gly Gln Ala Lys  Lys Lys Lys Gly Ser  Tyr Pro Tyr
    1160                1165                1170

Asp Val  Pro Asp Tyr Ala Tyr  Pro Tyr Asp Val Pro  Asp Tyr Ala
    1175                1180                1185

Tyr Pro  Tyr Asp Val Pro Asp  Tyr Ala
    1190                1195
```

<210> 48
<211> 1120
<212> PRT
<213> Listeria seeligeri

<400> 48

```
Met Trp Ile Ser Ile Lys Thr Leu Ile His His Leu Gly Val Leu Phe
1               5                  10                  15

Phe Cys Asp Tyr Met Tyr Asn Arg Arg Glu Lys Lys Ile Ile Glu Val
            20                  25                  30

Lys Thr Met Arg Ile Thr Lys Val Glu Val Asp Arg Lys Lys Val Leu
        35                  40                  45

Ile Ser Arg Asp Lys Asn Gly Gly Lys Leu Val Tyr Glu Asn Glu Met
    50                  55                  60

Gln Asp Asn Thr Glu Gln Ile Met His His Lys Lys Ser Ser Phe Tyr
65                  70                  75                  80

Lys Ser Val Val Asn Lys Thr Ile Cys Arg Pro Glu Gln Lys Gln Met
                85                  90                  95

Lys Lys Leu Val His Gly Leu Leu Gln Glu Asn Ser Gln Glu Lys Ile
            100                 105                 110

Lys Val Ser Asp Val Thr Lys Leu Asn Ile Ser Asn Phe Leu Asn His
        115                 120                 125

Arg Phe Lys Lys Ser Leu Tyr Tyr Phe Pro Glu Asn Ser Pro Asp Lys
    130                 135                 140

Ser Glu Glu Tyr Arg Ile Glu Ile Asn Leu Ser Gln Leu Leu Glu Asp
145                 150                 155                 160

Ser Leu Lys Lys Gln Gln Gly Thr Phe Ile Cys Trp Glu Ser Phe Ser
                165                 170                 175

Lys Asp Met Glu Leu Tyr Ile Asn Trp Ala Glu Asn Tyr Ile Ser Ser
            180                 185                 190

Lys Thr Lys Leu Ile Lys Lys Ser Ile Arg Asn Asn Arg Ile Gln Ser
        195                 200                 205
```

```
Thr Glu Ser Arg Ser Gly Gln Leu Met Asp Arg Tyr Met Lys Asp Ile
    210                 215                 220

Leu Asn Lys Asn Lys Pro Phe Asp Ile Gln Ser Val Ser Glu Lys Tyr
225                 230                 235                 240

Gln Leu Glu Lys Leu Thr Ser Ala Leu Lys Ala Thr Phe Lys Glu Ala
                245                 250                 255

Lys Lys Asn Asp Lys Glu Ile Asn Tyr Lys Leu Lys Ser Thr Leu Gln
            260                 265                 270

Asn His Glu Arg Gln Ile Ile Glu Glu Leu Lys Glu Asn Ser Glu Leu
        275                 280                 285

Asn Gln Phe Asn Ile Glu Ile Arg Lys His Leu Glu Thr Tyr Phe Pro
    290                 295                 300

Ile Lys Lys Thr Asn Arg Lys Val Gly Asp Ile Arg Asn Leu Glu Ile
305                 310                 315                 320

Gly Glu Ile Gln Lys Ile Val Asn His Arg Leu Lys Asn Lys Ile Val
                325                 330                 335

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

Thr Val Asn Ser Asn Ser Leu Gln Lys Ile Lys Ile Glu Glu Ala Phe
        355                 360                 365

Ala Leu Lys Phe Ile Asn Ala Cys Leu Phe Ala Ser Asn Asn Leu Arg
    370                 375                 380

Asn Met Val Tyr Pro Val Cys Lys Lys Asp Ile Leu Met Ile Gly Glu
385                 390                 395                 400

Phe Lys Asn Ser Phe Lys Glu Ile Lys His Lys Lys Phe Ile Arg Gln
                405                 410                 415
```

```
Trp Ser Gln Phe Phe Ser Gln Glu Ile Thr Val Asp Asp Ile Glu Leu
            420                 425                 430

Ala Ser Trp Gly Leu Arg Gly Ala Ile Ala Pro Ile Arg Asn Glu Ile
        435                 440                 445

Ile His Leu Lys Lys His Ser Trp Lys Lys Phe Phe Asn Asn Pro Thr
    450                 455                 460

Phe Lys Val Lys Lys Ser Lys Ile Ile Asn Gly Lys Thr Lys Asp Val
465                 470                 475                 480

Thr Ser Glu Phe Leu Tyr Lys Glu Thr Leu Phe Lys Asp Tyr Phe Tyr
                485                 490                 495

Ser Glu Leu Asp Ser Val Pro Glu Leu Ile Ile Asn Lys Met Glu Ser
            500                 505                 510

Ser Lys Ile Leu Asp Tyr Tyr Ser Ser Asp Gln Leu Asn Gln Val Phe
        515                 520                 525

Thr Ile Pro Asn Phe Glu Leu Ser Leu Leu Thr Ser Ala Val Pro Phe
    530                 535                 540

Ala Pro Ser Phe Lys Arg Val Tyr Leu Lys Gly Phe Asp Tyr Gln Asn
545                 550                 555                 560

Gln Asp Glu Ala Gln Pro Asp Tyr Asn Leu Lys Leu Asn Ile Tyr Asn
                565                 570                 575

Glu Lys Ala Phe Asn Ser Glu Ala Phe Gln Ala Gln Tyr Ser Leu Phe
            580                 585                 590

Lys Met Val Tyr Tyr Gln Val Phe Leu Pro Gln Phe Thr Thr Asn Asn
        595                 600                 605

Asp Leu Phe Lys Ser Ser Val Asp Phe Ile Leu Thr Leu Asn Lys Glu
    610                 615                 620
```

```
Arg Lys Gly Tyr Ala Lys Ala Phe Gln Asp Ile Arg Lys Met Asn Lys
625                 630                 635                 640


Asp Glu Lys Pro Ser Glu Tyr Met Ser Tyr Ile Gln Ser Gln Leu Met
                645                 650                 655


Leu Tyr Gln Lys Lys Gln Glu Glu Lys Glu Lys Ile Asn His Phe Glu
            660                 665                 670


Lys Phe Ile Asn Gln Val Phe Ile Lys Gly Phe Asn Ser Phe Ile Glu
        675                 680                 685


Lys Asn Arg Leu Thr Tyr Ile Cys His Pro Thr Lys Asn Thr Val Pro
    690                 695                 700


Glu Asn Asp Asn Ile Glu Ile Pro Phe His Thr Asp Met Asp Asp Ser
705                 710                 715                 720


Asn Ile Ala Phe Trp Leu Met Cys Lys Leu Leu Asp Ala Lys Gln Leu
                725                 730                 735


Ser Glu Leu Arg Asn Glu Met Ile Lys Phe Ser Cys Ser Leu Gln Ser
            740                 745                 750


Thr Glu Glu Ile Ser Thr Phe Thr Lys Ala Arg Glu Val Ile Gly Leu
        755                 760                 765


Ala Leu Leu Asn Gly Glu Lys Gly Cys Asn Asp Trp Lys Glu Leu Phe
    770                 775                 780


Asp Asp Lys Glu Ala Trp Lys Lys Asn Met Ser Leu Tyr Val Ser Glu
785                 790                 795                 800


Glu Leu Leu Gln Ser Leu Pro Tyr Thr Gln Glu Asp Gly Gln Thr Pro
                805                 810                 815


Val Ile Asn Arg Ser Ile Asp Leu Val Lys Lys Tyr Gly Thr Glu Thr
            820                 825                 830
```

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


Lys Asp Ile Ala Lys Leu His Glu Tyr Asp Val Thr Glu Lys Ile Ala
    850                 855                 860


Gln Gln Glu Ser Leu His Lys Gln Trp Ile Glu Lys Pro Gly Leu Ala
865                 870                 875                 880


Arg Asp Ser Ala Trp Thr Lys Lys Tyr Gln Asn Val Ile Asn Asp Ile
                885                 890                 895


Ser Asn Tyr Gln Trp Ala Lys Thr Lys Val Glu Leu Thr Gln Val Arg
            900                 905                 910


His Leu His Gln Leu Thr Ile Asp Leu Leu Ser Arg Leu Ala Gly Tyr
        915                 920                 925


Met Ser Ile Ala Asp Arg Asp Phe Gln Phe Ser Ser Asn Tyr Ile Leu
    930                 935                 940


Glu Arg Glu Asn Ser Glu Tyr Arg Val Thr Ser Trp Ile Leu Leu Ser
945                 950                 955                 960


Glu Asn Lys Asn Lys Asn Lys Tyr Asn Asp Tyr Glu Leu Tyr Asn Leu
                965                 970                 975


Lys Asn Ala Ser Ile Lys Val Ser Ser Lys Asn Asp Pro Gln Leu Lys
            980                 985                 990


Val Asp Leu Lys Gln Leu Arg Leu  Thr Leu Glu Tyr Leu  Glu Leu Phe
        995                 1000                1005


Asp Asn  Arg Leu Lys Glu Lys  Arg Asn Asn Ile Ser  His Phe Asn
    1010                1015                1020


Tyr Leu  Asn Gly Gln Leu Gly  Asn Ser Ile Leu Glu  Leu Phe Asp
    1025                1030                1035
```

```
Asp Ala  Arg Asp Val Leu Ser  Tyr Asp Arg Lys Leu  Lys Asn Ala
    1040                1045                1050


Val Ser  Lys Ser Leu Lys Glu  Ile Leu Ser Ser His  Gly Met Glu
    1055                1060                1065


Val Thr  Phe Lys Pro Leu Tyr  Gln Thr Asn His His  Leu Lys Ile
    1070                1075                1080


Asp Lys  Leu Gln Pro Lys Lys  Ile His His Leu Gly  Glu Lys Ser
    1085                1090                1095


Thr Val  Ser Ser Asn Gln Val  Ser Asn Glu Tyr Cys  Gln Leu Val
    1100                1105                1110


Arg Thr  Leu Leu Thr Met Lys
    1115                1120


<210>  49
<211>  1159
<212>  PRT
<213>  Leptotrichia buccalis

<400>  49

Met Lys Val Thr Lys Val Gly Gly Ile Ser His Lys Lys Tyr Thr Ser
1               5                   10                  15


Glu Gly Arg Leu Val Lys Ser Glu Ser Glu Glu Asn Arg Thr Asp Glu
            20                  25                  30


Arg Leu Ser Ala Leu Leu Asn Met Arg Leu Asp Met Tyr Ile Lys Asn
        35                  40                  45


Pro Ser Ser Thr Glu Thr Lys Glu Asn Gln Lys Arg Ile Gly Lys Leu
    50                  55                  60


Lys Lys Phe Phe Ser Asn Lys Met Val Tyr Leu Lys Asp Asn Thr Leu
65                  70                  75                  80
```

```
Ser Leu Lys Asn Gly Lys Lys Glu Asn Ile Asp Arg Glu Tyr Ser Glu
                85                  90                  95

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

Leu Lys Lys Ile Tyr Leu Asn Glu Asn Val Asn Ser Glu Glu Leu Glu
        115                 120                 125

Val Phe Arg Asn Asp Ile Lys Lys Lys Leu Asn Lys Ile Asn Ser Leu
    130                 135                 140

Lys Tyr Ser Phe Glu Lys Asn Lys Ala Asn Tyr Gln Lys Ile Asn Glu
145                 150                 155                 160

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

Asp Tyr Tyr Arg Glu Ser Ala Lys Arg Asp Ala Tyr Val Ser Asn Val
            180                 185                 190

Lys Glu Ala Phe Asp Lys Leu Tyr Lys Glu Glu Asp Ile Ala Lys Leu
        195                 200                 205

Val Leu Glu Ile Glu Asn Leu Thr Lys Leu Glu Lys Tyr Lys Ile Arg
    210                 215                 220

Glu Phe Tyr His Glu Ile Ile Gly Arg Lys Asn Asp Lys Glu Asn Phe
225                 230                 235                 240

Ala Lys Ile Ile Tyr Glu Glu Ile Gln Asn Val Asn Asn Met Lys Glu
                245                 250                 255

Leu Ile Glu Lys Val Pro Asp Met Ser Glu Leu Lys Lys Ser Gln Val
            260                 265                 270

Phe Tyr Lys Tyr Tyr Leu Asp Lys Glu Glu Leu Asn Asp Lys Asn Ile
        275                 280                 285
```

```
Lys Tyr Ala Phe Cys His Phe Val Glu Ile Glu Met Ser Gln Leu Leu
    290                 295                 300

Lys Asn Tyr Val Tyr Lys Arg Leu Ser Asn Ile Ser Asn Asp Lys Ile
305                 310                 315                 320

Lys Arg Ile Phe Glu Tyr Gln Asn Leu Lys Lys Leu Ile Glu Asn Lys
                325                 330                 335

Leu Leu Asn Lys Leu Asp Thr Tyr Val Arg Asn Cys Gly Lys Tyr Asn
            340                 345                 350

Tyr Tyr Leu Gln Asp Gly Glu Ile Ala Thr Ser Asp Phe Ile Ala Arg
        355                 360                 365

Asn Arg Gln Asn Glu Ala Phe Leu Arg Asn Ile Ile Gly Val Ser Ser
    370                 375                 380

Val Ala Tyr Phe Ser Leu Arg Asn Ile Leu Glu Thr Glu Asn Glu Asn
385                 390                 395                 400

Asp Ile Thr Gly Arg Met Arg Gly Lys Thr Val Lys Asn Asn Lys Gly
                405                 410                 415

Glu Glu Lys Tyr Val Ser Gly Glu Val Asp Lys Ile Tyr Asn Glu Asn
            420                 425                 430

Lys Lys Asn Glu Val Lys Glu Asn Leu Lys Met Phe Tyr Ser Tyr Asp
        435                 440                 445

Phe Asn Met Asp Asn Lys Asn Glu Ile Glu Asp Phe Phe Ala Asn Ile
    450                 455                 460

Asp Glu Ala Ile Ser Ser Ile Arg His Gly Ile Val His Phe Asn Leu
465                 470                 475                 480

Glu Leu Glu Gly Lys Asp Ile Phe Ala Phe Lys Asn Ile Ala Pro Ser
                485                 490                 495
```

```
Glu Ile Ser Lys Lys Met Phe Gln Asn Glu Ile Asn Glu Lys Lys Leu
            500                 505                 510


Lys Leu Lys Ile Phe Arg Gln Leu Asn Ser Ala Asn Val Phe Arg Tyr
        515                 520                 525


Leu Glu Lys Tyr Lys Ile Leu Asn Tyr Leu Lys Arg Thr Arg Phe Glu
    530                 535                 540


Phe Val Asn Lys Asn Ile Pro Phe Val Pro Ser Phe Thr Lys Leu Tyr
545                 550                 555                 560


Ser Arg Ile Asp Asp Leu Lys Asn Ser Leu Gly Ile Tyr Trp Lys Thr
                565                 570                 575


Pro Lys Thr Asn Asp Asp Asn Lys Thr Lys Glu Ile Ile Asp Ala Gln
            580                 585                 590


Ile Tyr Leu Leu Lys Asn Ile Tyr Tyr Gly Glu Phe Leu Asn Tyr Phe
        595                 600                 605


Met Ser Asn Asn Gly Asn Phe Phe Glu Ile Ser Lys Glu Ile Ile Glu
    610                 615                 620


Leu Asn Lys Asn Asp Lys Arg Asn Leu Lys Thr Gly Phe Tyr Lys Leu
625                 630                 635                 640


Gln Lys Phe Glu Asp Ile Gln Glu Lys Ile Pro Lys Glu Tyr Leu Ala
                645                 650                 655


Asn Ile Gln Ser Leu Tyr Met Ile Asn Ala Gly Asn Gln Asp Glu Glu
            660                 665                 670


Glu Lys Asp Thr Tyr Ile Asp Phe Ile Gln Lys Ile Phe Leu Lys Gly
        675                 680                 685


Phe Met Thr Tyr Leu Ala Asn Asn Gly Arg Leu Ser Leu Ile Tyr Ile
    690                 695                 700
```

```
Gly Ser Asp Glu Glu Thr Asn Thr Ser Leu Ala Glu Lys Lys Gln Glu
705                 710                 715                 720


Phe Asp Lys Phe Leu Lys Lys Tyr Glu Gln Asn Asn Asn Ile Lys Ile
                725                 730                 735


Pro Tyr Glu Ile Asn Glu Phe Leu Arg Glu Ile Lys Leu Gly Asn Ile
            740                 745                 750


Leu Lys Tyr Thr Glu Arg Leu Asn Met Phe Tyr Leu Ile Leu Lys Leu
        755                 760                 765


Leu Asn His Lys Glu Leu Thr Asn Leu Lys Gly Ser Leu Glu Lys Tyr
    770                 775                 780


Gln Ser Ala Asn Lys Glu Glu Ala Phe Ser Asp Gln Leu Glu Leu Ile
785                 790                 795                 800


Asn Leu Leu Asn Leu Asp Asn Asn Arg Val Thr Glu Asp Phe Glu Leu
                805                 810                 815


Glu Ala Asp Glu Ile Gly Lys Phe Leu Asp Phe Asn Gly Asn Lys Val
            820                 825                 830


Lys Asp Asn Lys Glu Leu Lys Lys Phe Asp Thr Asn Lys Ile Tyr Phe
        835                 840                 845


Asp Gly Glu Asn Ile Ile Lys His Arg Ala Phe Tyr Asn Ile Lys Lys
    850                 855                 860


Tyr Gly Met Leu Asn Leu Leu Glu Lys Ile Ala Asp Lys Ala Gly Tyr
865                 870                 875                 880


Lys Ile Ser Ile Glu Glu Leu Lys Lys Tyr Ser Asn Lys Lys Asn Glu
                885                 890                 895


Ile Glu Lys Asn His Lys Met Gln Glu Asn Leu His Arg Lys Tyr Ala
            900                 905                 910
```

```
Arg Pro Arg Lys Asp Glu Lys Phe Thr Asp Glu Asp Tyr Glu Ser Tyr
        915                 920                 925


Lys Gln Ala Ile Glu Asn Ile Glu Glu Tyr Thr His Leu Lys Asn Lys
    930                 935                 940


Val Glu Phe Asn Glu Leu Asn Leu Leu Gln Gly Leu Leu Leu Arg Ile
945                 950                 955                 960


Leu His Arg Leu Val Gly Tyr Thr Ser Ile Trp Glu Arg Asp Leu Arg
                965                 970                 975


Phe Arg Leu Lys Gly Glu Phe Pro Glu Asn Gln Tyr Ile Glu Glu Ile
            980                 985                 990


Phe Asn Phe Glu Asn Lys Lys Asn  Val Lys Tyr Lys Gly  Gly Gln Ile
        995                 1000                1005


Val Glu  Lys Tyr Ile Lys Phe  Tyr Lys Glu Leu His  Gln Asn Asp
    1010                1015                1020


Glu Val  Lys Ile Asn Lys Tyr  Ser Ser Ala Asn Ile  Lys Val Leu
    1025                1030                1035


Lys Gln  Glu Lys Lys Asp Leu  Tyr Ile Arg Asn Tyr  Ile Ala His
    1040                1045                1050


Phe Asn  Tyr Ile Pro His Ala  Glu Ile Ser Leu Leu  Glu Val Leu
    1055                1060                1065


Glu Asn  Leu Arg Lys Leu Leu  Ser Tyr Asp Arg Lys  Leu Lys Asn
    1070                1075                1080


Ala Val  Met Lys Ser Val Val  Asp Ile Leu Lys Glu  Tyr Gly Phe
    1085                1090                1095


Val Ala  Thr Phe Lys Ile Gly  Ala Asp Lys Lys Ile  Gly Ile Gln
    1100                1105                1110
```

```
Thr Leu  Glu Ser Glu Lys Ile  Val His Leu Lys Asn  Leu Lys Lys
    1115                1120                1125


Lys Lys  Leu Met Thr Asp Arg  Asn Ser Glu Glu Leu  Cys Lys Leu
    1130                1135                1140


Val Lys  Ile Met Phe Glu Tyr  Lys Met Glu Glu Lys  Lys Ser Glu
    1145                1150                1155


Asn



<210>  50
<211>  1285
<212>  PRT
<213>  Herbinix hemicellulosilytica

<400>  50

Met Lys Leu Thr Arg Arg Arg Ile Ser Gly Asn Ser Val Asp Gln Lys
1               5                   10                  15


Ile Thr Ala Ala Phe Tyr Arg Asp Met Ser Gln Gly Leu Leu Tyr Tyr
            20                  25                  30


Asp Ser Glu Asp Asn Asp Cys Thr Asp Lys Val Ile Glu Ser Met Asp
        35                  40                  45


Phe Glu Arg Ser Trp Arg Gly Arg Ile Leu Lys Asn Gly Glu Asp Asp
    50                  55                  60


Lys Asn Pro Phe Tyr Met Phe Val Lys Gly Leu Val Gly Ser Asn Asp
65                  70                  75                  80


Lys Ile Val Cys Glu Pro Ile Asp Val Asp Ser Asp Pro Asp Asn Leu
                85                  90                  95


Asp Ile Leu Ile Asn Lys Asn Leu Thr Gly Phe Gly Arg Asn Leu Lys
            100                 105                 110


Ala Pro Asp Ser Asn Asp Thr Leu Glu Asn Leu Ile Arg Lys Ile Gln
```

```
        115                 120                 125

Ala Gly Ile Pro Glu Glu Glu Val Leu Pro Glu Leu Lys Lys Ile Lys
    130                 135                 140

Glu Met Ile Gln Lys Asp Ile Val Asn Arg Lys Glu Gln Leu Leu Lys
145                 150                 155                 160

Ser Ile Lys Asn Asn Arg Ile Pro Phe Ser Leu Glu Gly Ser Lys Leu
                165                 170                 175

Val Pro Ser Thr Lys Lys Met Lys Trp Leu Phe Lys Leu Ile Asp Val
            180                 185                 190

Pro Asn Lys Thr Phe Asn Glu Lys Met Leu Glu Lys Tyr Trp Glu Ile
        195                 200                 205

Tyr Asp Tyr Asp Lys Leu Lys Ala Asn Ile Thr Asn Arg Leu Asp Lys
    210                 215                 220

Thr Asp Lys Lys Ala Arg Ser Ile Ser Arg Ala Val Ser Glu Glu Leu
225                 230                 235                 240

Arg Glu Tyr His Lys Asn Leu Arg Thr Asn Tyr Asn Arg Phe Val Ser
                245                 250                 255

Gly Asp Arg Pro Ala Ala Gly Leu Asp Asn Gly Gly Ser Ala Lys Tyr
            260                 265                 270

Asn Pro Asp Lys Glu Glu Phe Leu Leu Phe Leu Lys Glu Val Glu Gln
        275                 280                 285

Tyr Phe Lys Lys Tyr Phe Pro Val Lys Ser Lys His Ser Asn Lys Ser
    290                 295                 300

Lys Asp Lys Ser Leu Val Asp Lys Tyr Lys Asn Tyr Cys Ser Tyr Lys
305                 310                 315                 320

Val Val Lys Lys Glu Val Asn Arg Ser Ile Ile Asn Gln Leu Val Ala
```

325 330 335

Gly Leu Ile Gln Gln Gly Lys Leu Leu Tyr Tyr Phe Tyr Tyr Asn Asp
340 345 350

Thr Trp Gln Glu Asp Phe Leu Asn Ser Tyr Gly Leu Ser Tyr Ile Gln
355 360 365

Val Glu Glu Ala Phe Lys Lys Ser Val Met Thr Ser Leu Ser Trp Gly
370 375 380

Ile Asn Arg Leu Thr Ser Phe Phe Ile Asp Asp Ser Asn Thr Val Lys
385 390 395 400

Phe Asp Asp Ile Thr Thr Lys Lys Ala Lys Glu Ala Ile Glu Ser Asn
405 410 415

Tyr Phe Asn Lys Leu Arg Thr Cys Ser Arg Met Gln Asp His Phe Lys
420 425 430

Glu Lys Leu Ala Phe Phe Tyr Pro Val Tyr Val Lys Asp Lys Lys Asp
435 440 445

Arg Pro Asp Asp Asp Ile Glu Asn Leu Ile Val Leu Val Lys Asn Ala
450 455 460

Ile Glu Ser Val Ser Tyr Leu Arg Asn Arg Thr Phe His Phe Lys Glu
465 470 475 480

Ser Ser Leu Leu Glu Leu Leu Lys Glu Leu Asp Asp Lys Asn Ser Gly
485 490 495

Gln Asn Lys Ile Asp Tyr Ser Val Ala Ala Glu Phe Ile Lys Arg Asp
500 505 510

Ile Glu Asn Leu Tyr Asp Val Phe Arg Glu Gln Ile Arg Ser Leu Gly
515 520 525

Ile Ala Glu Tyr Tyr Lys Ala Asp Met Ile Ser Asp Cys Phe Lys Thr

```
    530                 535                 540

Cys Gly Leu Glu Phe Ala Leu Tyr Ser Pro Lys Asn Ser Leu Met Pro
545                 550                 555                 560

Ala Phe Lys Asn Val Tyr Lys Arg Gly Ala Asn Leu Asn Lys Ala Tyr
                565                 570                 575

Ile Arg Asp Lys Gly Pro Lys Glu Thr Gly Asp Gln Gly Gln Asn Ser
            580                 585                 590

Tyr Lys Ala Leu Glu Glu Tyr Arg Glu Leu Thr Trp Tyr Ile Glu Val
        595                 600                 605

Lys Asn Asn Asp Gln Ser Tyr Asn Ala Tyr Lys Asn Leu Leu Gln Leu
    610                 615                 620

Ile Tyr Tyr His Ala Phe Leu Pro Glu Val Arg Glu Asn Glu Ala Leu
625                 630                 635                 640

Ile Thr Asp Phe Ile Asn Arg Thr Lys Glu Trp Asn Arg Lys Glu Thr
                645                 650                 655

Glu Glu Arg Leu Asn Thr Lys Asn Asn Lys Lys His Lys Asn Phe Asp
            660                 665                 670

Glu Asn Asp Asp Ile Thr Val Asn Thr Tyr Arg Tyr Glu Ser Ile Pro
        675                 680                 685

Asp Tyr Gln Gly Glu Ser Leu Asp Asp Tyr Leu Lys Val Leu Gln Arg
    690                 695                 700

Lys Gln Met Ala Arg Ala Lys Glu Val Asn Glu Lys Glu Glu Gly Asn
705                 710                 715                 720

Asn Asn Tyr Ile Gln Phe Ile Arg Asp Val Val Val Trp Ala Phe Gly
                725                 730                 735

Ala Tyr Leu Glu Asn Lys Leu Lys Asn Tyr Lys Asn Glu Leu Gln Pro
```

```
            740                 745                 750

Pro Leu Ser Lys Glu Asn Ile Gly Leu Asn Asp Thr Leu Lys Glu Leu
        755                 760                 765

Phe Pro Glu Glu Lys Val Lys Ser Pro Phe Asn Ile Lys Cys Arg Phe
    770                 775                 780

Ser Ile Ser Thr Phe Ile Asp Asn Lys Gly Lys Ser Thr Asp Asn Thr
785                 790                 795                 800

Ser Ala Glu Ala Val Lys Thr Asp Gly Lys Glu Asp Glu Lys Asp Lys
                805                 810                 815

Lys Asn Ile Lys Arg Lys Asp Leu Leu Cys Phe Tyr Leu Phe Leu Arg
            820                 825                 830

Leu Leu Asp Glu Asn Glu Ile Cys Lys Leu Gln His Gln Phe Ile Lys
        835                 840                 845

Tyr Arg Cys Ser Leu Lys Glu Arg Arg Phe Pro Gly Asn Arg Thr Lys
    850                 855                 860

Leu Glu Lys Glu Thr Glu Leu Leu Ala Glu Leu Glu Glu Leu Met Glu
865                 870                 875                 880

Leu Val Arg Phe Thr Met Pro Ser Ile Pro Glu Ile Ser Ala Lys Ala
                885                 890                 895

Glu Ser Gly Tyr Asp Thr Met Ile Lys Lys Tyr Phe Lys Asp Phe Ile
            900                 905                 910

Glu Lys Lys Val Phe Lys Asn Pro Lys Thr Ser Asn Leu Tyr Tyr His
        915                 920                 925

Ser Asp Ser Lys Thr Pro Val Thr Arg Lys Tyr Met Ala Leu Leu Met
    930                 935                 940

Arg Ser Ala Pro Leu His Leu Tyr Lys Asp Ile Phe Lys Gly Tyr Tyr
```

```
945                 950                 955                 960

Leu Ile Thr Lys Lys Glu Cys Leu Glu Tyr Ile Lys Leu Ser Asn Ile
                965                 970                 975

Ile Lys Asp Tyr Gln Asn Ser Leu Asn Glu Leu His Glu Gln Leu Glu
            980                 985                 990

Arg Ile Lys Leu Lys Ser Glu Lys  Gln Asn Gly Lys Asp  Ser Leu Tyr
        995                 1000                1005

Leu Asp  Lys Lys Asp Phe Tyr  Lys Val Lys Glu Tyr  Val Glu Asn
    1010                1015                1020

Leu Glu  Gln Val Ala Arg Tyr  Lys His Leu Gln His  Lys Ile Asn
    1025                1030                1035

Phe Glu  Ser Leu Tyr Arg Ile  Phe Arg Ile His Val  Asp Ile Ala
    1040                1045                1050

Ala Arg  Met Val Gly Tyr Thr  Gln Asp Trp Glu Arg  Asp Met His
    1055                1060                1065

Phe Leu  Phe Lys Ala Leu Val  Tyr Asn Gly Val Leu  Glu Glu Arg
    1070                1075                1080

Arg Phe  Glu Ala Ile Phe Asn  Asn Asn Asp Asp Asn  Asn Asp Gly
    1085                1090                1095

Arg Ile  Val Lys Lys Ile Gln  Asn Asn Leu Asn Asn  Lys Asn Arg
    1100                1105                1110

Glu Leu  Val Ser Met Leu Cys  Trp Asn Lys Lys Leu  Asn Lys Asn
    1115                1120                1125

Glu Phe  Gly Ala Ile Ile Trp  Lys Arg Asn Pro Ile  Ala His Leu
    1130                1135                1140

Asn His  Phe Thr Gln Thr Glu  Gln Asn Ser Lys Ser  Ser Leu Glu
```

1145 1150 1155

Ser Leu Ile Asn Ser Leu Arg Ile Leu Leu Ala Tyr Asp Arg Lys
1160 1165 1170

Arg Gln Asn Ala Val Thr Lys Thr Ile Asn Asp Leu Leu Leu Asn
1175 1180 1185

Asp Tyr His Ile Arg Ile Lys Trp Glu Gly Arg Val Asp Glu Gly
1190 1195 1200

Gln Ile Tyr Phe Asn Ile Lys Glu Lys Glu Asp Ile Glu Asn Glu
1205 1210 1215

Pro Ile Ile His Leu Lys His Leu His Lys Lys Asp Cys Tyr Ile
1220 1225 1230

Tyr Lys Asn Ser Tyr Met Phe Asp Lys Gln Lys Glu Trp Ile Cys
1235 1240 1245

Asn Gly Ile Lys Glu Glu Val Tyr Asp Lys Ser Ile Leu Lys Cys
1250 1255 1260

Ile Gly Asn Leu Phe Lys Phe Asp Tyr Glu Asp Lys Asn Lys Ser
1265 1270 1275

Ser Ala Asn Pro Lys His Thr
1280 1285

<210> 51
<211> 1344
<212> PRT
<213> Eubacterium rectale

<400> 51

Met Leu Arg Arg Asp Lys Glu Val Lys Lys Leu Tyr Asn Val Phe Asn
1 5 10 15

Gln Ile Gln Val Gly Thr Lys Pro Lys Lys Trp Asn Asn Asp Glu Lys
20 25 30

```
Leu Ser Pro Glu Glu Asn Glu Arg Arg Ala Gln Gln Lys Asn Ile Lys
        35                  40                  45

Met Lys Asn Tyr Lys Trp Arg Glu Ala Cys Ser Lys Tyr Val Glu Ser
    50                  55                  60

Ser Gln Arg Ile Ile Asn Asp Val Ile Phe Tyr Ser Tyr Arg Lys Ala
65                  70                  75                  80

Lys Asn Lys Leu Arg Tyr Met Arg Lys Asn Glu Asp Ile Leu Lys Lys
                85                  90                  95

Met Gln Glu Ala Glu Lys Leu Ser Lys Phe Ser Gly Gly Lys Leu Glu
            100                 105                 110

Asp Phe Val Ala Tyr Thr Leu Arg Lys Ser Leu Val Val Ser Lys Tyr
        115                 120                 125

Asp Thr Gln Glu Phe Asp Ser Leu Ala Ala Met Val Val Phe Leu Glu
    130                 135                 140

Cys Ile Gly Lys Asn Asn Ile Ser Asp His Glu Arg Glu Ile Val Cys
145                 150                 155                 160

Lys Leu Leu Glu Leu Ile Arg Lys Asp Phe Ser Lys Leu Asp Pro Asn
                165                 170                 175

Val Lys Gly Ser Gln Gly Ala Asn Ile Val Arg Ser Val Arg Asn Gln
            180                 185                 190

Asn Met Ile Val Gln Pro Gln Gly Asp Arg Phe Leu Phe Pro Gln Val
        195                 200                 205

Tyr Ala Lys Glu Asn Glu Thr Val Thr Asn Lys Asn Val Glu Lys Glu
    210                 215                 220

Gly Leu Asn Glu Phe Leu Leu Asn Tyr Ala Asn Leu Asp Asp Glu Lys
225                 230                 235                 240
```

```
Arg Ala Glu Ser Leu Arg Lys Leu Arg Arg Ile Leu Asp Val Tyr Phe
                245                 250                 255

Ser Ala Pro Asn His Tyr Glu Lys Asp Met Asp Ile Thr Leu Ser Asp
            260                 265                 270

Asn Ile Glu Lys Glu Lys Phe Asn Val Trp Glu Lys His Glu Cys Gly
        275                 280                 285

Lys Lys Glu Thr Gly Leu Phe Val Asp Ile Pro Asp Val Leu Met Glu
    290                 295                 300

Ala Glu Ala Glu Asn Ile Lys Leu Asp Ala Val Val Glu Lys Arg Glu
305                 310                 315                 320

Arg Lys Val Leu Asn Asp Arg Val Arg Lys Gln Asn Ile Ile Cys Tyr
                325                 330                 335

Arg Tyr Thr Arg Ala Val Val Glu Lys Tyr Asn Ser Asn Glu Pro Leu
            340                 345                 350

Phe Phe Glu Asn Asn Ala Ile Asn Gln Tyr Trp Ile His His Ile Glu
        355                 360                 365

Asn Ala Val Glu Arg Ile Leu Lys Asn Cys Lys Ala Gly Lys Leu Phe
    370                 375                 380

Lys Leu Arg Lys Gly Tyr Leu Ala Glu Lys Val Trp Lys Asp Ala Ile
385                 390                 395                 400

Asn Leu Ile Ser Ile Lys Tyr Ile Ala Leu Gly Lys Ala Val Tyr Asn
                405                 410                 415

Phe Ala Leu Asp Asp Ile Trp Lys Asp Lys Lys Asn Lys Glu Leu Gly
            420                 425                 430

Ile Val Asp Glu Arg Ile Arg Asn Gly Ile Thr Ser Phe Asp Tyr Glu
        435                 440                 445
```

```
Met Ile Lys Ala His Glu Asn Leu Gln Arg Glu Leu Ala Val Asp Ile
    450                 455                 460

Ala Phe Ser Val Asn Asn Leu Ala Arg Ala Val Cys Asp Met Ser Asn
465                 470                 475                 480

Leu Gly Asn Lys Glu Ser Asp Phe Leu Leu Trp Lys Arg Asn Asp Ile
                485                 490                 495

Ala Asp Lys Leu Lys Asn Lys Asp Asp Met Ala Ser Val Ser Ala Val
            500                 505                 510

Leu Gln Phe Phe Gly Gly Lys Ser Ser Trp Asp Ile Asn Ile Phe Lys
        515                 520                 525

Asp Ala Tyr Lys Gly Lys Lys Lys Tyr Asn Tyr Glu Val Arg Phe Ile
    530                 535                 540

Asp Asp Leu Arg Lys Ala Ile Tyr Cys Ala Arg Asn Glu Asn Phe His
545                 550                 555                 560

Phe Lys Thr Ala Leu Val Asn Asp Glu Lys Trp Asn Thr Glu Leu Phe
                565                 570                 575

Gly Lys Ile Phe Glu Arg Glu Thr Glu Phe Cys Leu Asn Val Glu Lys
            580                 585                 590

Asp Arg Phe Tyr Ser Asn Asn Leu Tyr Met Phe Tyr Gln Val Ser Glu
        595                 600                 605

Leu Arg Asn Met Leu Asp His Leu Tyr Ser Arg Ser Val Ser Arg Ala
    610                 615                 620

Ala Gln Val Pro Ser Tyr Asn Ser Val Ile Val Arg Thr Ala Phe Pro
625                 630                 635                 640

Glu Tyr Ile Thr Asn Val Leu Gly Tyr Gln Lys Pro Ser Tyr Asp Ala
                645                 650                 655
```

```
Asp Thr Leu Gly Lys Trp Tyr Ser Ala Cys Tyr Tyr Leu Leu Lys Glu
            660                 665                 670

Ile Tyr Tyr Asn Ser Phe Leu Gln Ser Asp Arg Ala Leu Gln Leu Phe
        675                 680                 685

Glu Lys Ser Val Lys Thr Leu Ser Trp Asp Asp Lys Lys Gln Gln Arg
    690                 695                 700

Ala Val Asp Asn Phe Lys Asp His Phe Ser Asp Ile Lys Ser Ala Cys
705                 710                 715                 720

Thr Ser Leu Ala Gln Val Cys Gln Ile Tyr Met Thr Glu Tyr Asn Gln
                725                 730                 735

Gln Asn Asn Gln Ile Lys Lys Val Arg Ser Ser Asn Asp Ser Ile Phe
            740                 745                 750

Asp Gln Pro Val Tyr Gln His Tyr Lys Val Leu Leu Lys Lys Ala Ile
        755                 760                 765

Ala Asn Ala Phe Ala Asp Tyr Leu Lys Asn Asn Lys Asp Leu Phe Gly
    770                 775                 780

Phe Ile Gly Lys Pro Phe Lys Ala Asn Glu Ile Arg Glu Ile Asp Lys
785                 790                 795                 800

Glu Gln Phe Leu Pro Asp Trp Thr Ser Arg Lys Tyr Glu Ala Leu Cys
                805                 810                 815

Ile Glu Val Ser Gly Ser Gln Glu Leu Gln Lys Trp Tyr Ile Val Gly
            820                 825                 830

Lys Phe Leu Asn Ala Arg Ser Leu Asn Leu Met Val Gly Ser Met Arg
        835                 840                 845

Ser Tyr Ile Gln Tyr Val Thr Asp Ile Lys Arg Arg Ala Ala Ser Ile
    850                 855                 860
```

```
Gly Asn Glu Leu His Val Ser Val His Asp Val Glu Lys Val Glu Lys
865                 870                 875                 880


Trp Val Gln Val Ile Glu Val Cys Ser Leu Leu Ala Ser Arg Thr Ser
                885                 890                 895


Asn Gln Phe Glu Asp Tyr Phe Asn Asp Lys Asp Asp Tyr Ala Arg Tyr
            900                 905                 910


Leu Lys Ser Tyr Val Asp Phe Ser Asn Val Asp Met Pro Ser Glu Tyr
        915                 920                 925


Ser Ala Leu Val Asp Phe Ser Asn Glu Glu Gln Ser Asp Leu Tyr Val
    930                 935                 940


Asp Pro Lys Asn Pro Lys Val Asn Arg Asn Ile Val His Ser Lys Leu
945                 950                 955                 960


Phe Ala Ala Asp His Ile Leu Arg Asp Ile Val Glu Pro Val Ser Lys
                965                 970                 975


Asp Asn Ile Glu Glu Phe Tyr Ser Gln Lys Ala Glu Ile Ala Tyr Cys
            980                 985                 990


Lys Ile Lys Gly Lys Glu Ile Thr  Ala Glu Glu Gln Lys  Ala Val Leu
        995                 1000                1005


Lys Tyr  Gln Lys Leu Lys Asn  Arg Val Glu Leu Arg  Asp Ile Val
    1010                1015                1020


Glu Tyr  Gly Glu Ile Ile Asn  Glu Leu Leu Gly Gln  Leu Ile Asn
    1025                1030                1035


Trp Ser  Phe Met Arg Glu Arg  Asp Leu Leu Tyr Phe  Gln Leu Gly
    1040                1045                1050


Phe His  Tyr Asp Cys Leu Arg  Asn Asp Ser Lys Lys  Pro Glu Gly
    1055                1060                1065
```

```
Tyr Lys  Asn Ile Lys Val Asp  Glu Asn Ser Ile Lys  Asp Ala Ile
    1070                1075                1080

Leu Tyr  Gln Ile Ile Gly Met  Tyr Val Asn Gly Val  Thr Val Tyr
    1085                1090                1095

Ala Pro  Glu Lys Asp Gly Asp  Lys Leu Lys Glu Gln  Cys Val Lys
    1100                1105                1110

Gly Gly  Val Gly Val Lys Val  Ser Ala Phe His Arg  Tyr Ser Lys
    1115                1120                1125

Tyr Leu  Gly Leu Asn Glu Lys  Thr Leu Tyr Asn Ala  Gly Leu Glu
    1130                1135                1140

Ile Phe  Glu Val Val Ala Glu  His Glu Asp Ile Ile  Asn Leu Arg
    1145                1150                1155

Asn Gly  Ile Asp His Phe Lys  Tyr Tyr Leu Gly Asp  Tyr Arg Ser
    1160                1165                1170

Met Leu  Ser Ile Tyr Ser Glu  Val Phe Asp Arg Phe  Phe Thr Tyr
    1175                1180                1185

Asp Ile  Lys Tyr Gln Lys Asn  Val Leu Asn Leu Leu  Gln Asn Ile
    1190                1195                1200

Leu Leu  Arg His Asn Val Ile  Val Glu Pro Ile Leu  Glu Ser Gly
    1205                1210                1215

Phe Lys  Thr Ile Gly Glu Gln  Thr Lys Pro Gly Ala  Lys Leu Ser
    1220                1225                1230

Ile Arg  Ser Ile Lys Ser Asp  Thr Phe Gln Tyr Lys  Val Lys Gly
    1235                1240                1245

Gly Thr  Leu Ile Thr Asp Ala  Lys Asp Glu Arg Tyr  Leu Glu Thr
    1250                1255                1260
```

```
Ile Arg  Lys Ile Leu Tyr Tyr  Ala Glu Asn Glu Glu  Asp Asn Leu
    1265                1270                1275


Lys Lys  Ser Val Val Val Thr  Asn Ala Asp Lys Tyr  Glu Lys Asn
    1280                1285                1290


Lys Glu  Ser Asp Asp Gln Asn  Lys Gln Lys Glu Lys  Lys Asn Lys
    1295                1300                1305


Asp Asn  Lys Gly Lys Lys Asn  Glu Glu Thr Lys Ser  Asp Ala Glu
    1310                1315                1320


Lys Asn  Asn Asn Glu Arg Leu  Ser Tyr Asn Pro Phe  Ala Asn Leu
    1325                1330                1335


Asn Phe  Lys Leu Ser Asn
    1340


<210>  52
<211>  1373
<212>  PRT
<213>  Eubacteriacea bacterium

<400>  52

Met Lys Ile Ser Lys Glu Ser His Lys Arg Thr Ala Val Ala Val Met
1               5                   10                  15


Glu Asp Arg Val Gly Gly Val Val Tyr Val Pro Gly Gly Ser Gly Ile
            20                  25                  30


Asp Leu Ser Asn Asn Leu Lys Lys Arg Ser Met Asp Thr Lys Ser Leu
        35                  40                  45


Tyr Asn Val Phe Asn Gln Ile Gln Ala Gly Thr Ala Pro Ser Glu Tyr
    50                  55                  60


Glu Trp Lys Asp Tyr Leu Ser Glu Ala Glu Asn Lys Lys Arg Glu Ala
65                  70                  75                  80
```

```
Gln Lys Met Ile Gln Lys Ala Asn Tyr Glu Leu Arg Arg Glu Cys Glu
                85                  90                  95

Asp Tyr Ala Lys Lys Ala Asn Leu Ala Val Ser Arg Ile Ile Phe Ser
            100                 105                 110

Lys Lys Pro Lys Lys Ile Phe Ser Asp Asp Asp Ile Ile Ser His Met
        115                 120                 125

Lys Lys Gln Arg Leu Ser Lys Phe Lys Gly Arg Met Glu Asp Phe Val
    130                 135                 140

Leu Ile Ala Leu Arg Lys Ser Leu Val Val Ser Thr Tyr Asn Gln Glu
145                 150                 155                 160

Val Phe Asp Ser Arg Lys Ala Ala Thr Val Phe Leu Lys Asn Ile Gly
                165                 170                 175

Lys Lys Asn Ile Ser Ala Asp Asp Glu Arg Gln Ile Lys Gln Leu Met
            180                 185                 190

Ala Leu Ile Arg Glu Asp Tyr Asp Lys Trp Asn Pro Asp Lys Asp Ser
        195                 200                 205

Ser Asp Lys Lys Glu Ser Ser Gly Thr Lys Val Ile Arg Ser Ile Glu
    210                 215                 220

His Gln Asn Met Val Ile Gln Pro Glu Lys Asn Lys Leu Ser Leu Ser
225                 230                 235                 240

Lys Ile Ser Asn Val Gly Lys Lys Thr Lys Thr Lys Gln Lys Glu Lys
                245                 250                 255

Ala Gly Leu Asp Ala Phe Leu Lys Glu Tyr Ala Gln Ile Asp Glu Asn
            260                 265                 270

Ser Arg Met Glu Tyr Leu Lys Lys Leu Arg Arg Leu Leu Asp Thr Tyr
        275                 280                 285
```

```
Phe Ala Ala Pro Ser Ser Tyr Ile Lys Gly Ala Ala Val Ser Leu Pro
    290                 295                 300


Glu Asn Ile Asn Phe Ser Ser Glu Leu Asn Val Trp Glu Arg His Glu
305                 310                 315                 320


Ala Ala Lys Lys Val Asn Ile Asn Phe Val Glu Ile Pro Glu Ser Leu
                325                 330                 335


Leu Asn Ala Glu Gln Asn Asn Asn Lys Ile Asn Lys Val Glu Gln Glu
            340                 345                 350


His Ser Leu Glu Gln Leu Arg Thr Asp Ile Arg Arg Arg Asn Ile Thr
        355                 360                 365


Cys Tyr His Phe Ala Asn Ala Leu Ala Ala Asp Glu Arg Tyr His Thr
    370                 375                 380


Leu Phe Phe Glu Asn Met Ala Met Asn Gln Phe Trp Ile His His Met
385                 390                 395                 400


Glu Asn Ala Val Glu Arg Ile Leu Lys Lys Cys Asn Val Gly Thr Leu
                405                 410                 415


Phe Lys Leu Arg Ile Gly Tyr Leu Ser Glu Lys Val Trp Lys Asp Met
            420                 425                 430


Leu Asn Leu Leu Ser Ile Lys Tyr Ile Ala Leu Gly Lys Ala Val Tyr
        435                 440                 445


His Phe Ala Leu Asp Asp Ile Trp Lys Ala Asp Ile Trp Lys Asp Ala
    450                 455                 460


Ser Asp Lys Asn Ser Gly Lys Ile Asn Asp Leu Thr Leu Lys Gly Ile
465                 470                 475                 480


Ser Ser Phe Asp Tyr Glu Met Val Lys Ala Gln Glu Asp Leu Gln Arg
                485                 490                 495
```

```
Glu Met Ala Val Gly Val Ala Phe Ser Thr Asn Asn Leu Ala Arg Val
            500                 505                 510

Thr Cys Lys Met Asp Asp Leu Ser Asp Ala Glu Ser Asp Phe Leu Leu
        515                 520                 525

Trp Asn Lys Glu Ala Ile Arg Arg His Val Lys Tyr Thr Glu Lys Gly
    530                 535                 540

Glu Ile Leu Ser Ala Ile Leu Gln Phe Phe Gly Gly Arg Ser Leu Trp
545                 550                 555                 560

Asp Glu Ser Leu Phe Glu Lys Ala Tyr Ser Asp Ser Asn Tyr Glu Leu
                565                 570                 575

Lys Phe Leu Asp Asp Leu Lys Arg Ala Ile Tyr Ala Ala Arg Asn Glu
            580                 585                 590

Thr Phe His Phe Lys Thr Ala Ala Ile Asp Gly Gly Ser Trp Asn Thr
        595                 600                 605

Arg Leu Phe Gly Ser Leu Phe Glu Lys Glu Ala Gly Leu Cys Leu Asn
    610                 615                 620

Val Glu Lys Asn Lys Phe Tyr Ser Asn Asn Leu Val Leu Phe Tyr Lys
625                 630                 635                 640

Gln Glu Asp Leu Arg Val Phe Leu Asp Lys Leu Tyr Gly Lys Glu Cys
                645                 650                 655

Ser Arg Ala Ala Gln Ile Pro Ser Tyr Asn Thr Ile Leu Pro Arg Lys
            660                 665                 670

Ser Phe Ser Asp Phe Met Lys Gln Leu Leu Gly Leu Lys Glu Pro Val
        675                 680                 685

Tyr Gly Ser Ala Ile Leu Asp Gln Trp Tyr Ser Ala Cys Tyr Tyr Leu
    690                 695                 700
```

```
Phe Lys Glu Val Tyr Tyr Asn Leu Phe Leu Gln Asp Ser Ser Ala Lys
705                 710                 715                 720

Ala Leu Phe Glu Lys Ala Val Lys Ala Leu Lys Gly Ala Asp Lys Lys
                725                 730                 735

Gln Glu Lys Ala Val Glu Ser Phe Arg Lys Arg Tyr Trp Glu Ile Ser
            740                 745                 750

Lys Asn Ala Ser Leu Ala Glu Ile Cys Gln Ser Tyr Ile Thr Glu Tyr
        755                 760                 765

Asn Gln Gln Asn Asn Lys Glu Arg Lys Val Arg Ser Ala Asn Asp Gly
    770                 775                 780

Met Phe Asn Glu Pro Ile Tyr Gln His Tyr Lys Met Leu Leu Lys Glu
785                 790                 795                 800

Ala Leu Lys Met Ala Phe Ala Ser Tyr Ile Lys Asn Asp Lys Glu Leu
                805                 810                 815

Lys Phe Val Tyr Lys Pro Thr Glu Lys Leu Phe Glu Val Ser Gln Asp
            820                 825                 830

Asn Phe Leu Pro Asn Trp Asn Ser Glu Lys Tyr Asn Thr Leu Ile Ser
        835                 840                 845

Glu Val Lys Asn Ser Pro Asp Leu Gln Lys Trp Tyr Ile Val Gly Lys
    850                 855                 860

Phe Met Asn Ala Arg Met Leu Asn Leu Leu Leu Gly Ser Met Arg Ser
865                 870                 875                 880

Tyr Leu Gln Tyr Val Ser Asp Ile Gln Lys Arg Ala Ala Gly Leu Gly
                885                 890                 895

Glu Asn Gln Leu His Leu Ser Ala Glu Asn Val Gly Gln Val Lys Lys
            900                 905                 910
```

```
Trp Ile Gln Val Leu Glu Val Cys Leu Leu Leu Ser Val Arg Ile Ser
        915                 920                 925


Asp Lys Phe Thr Asp Tyr Phe Lys Asp Glu Glu Glu Tyr Ala Ser Tyr
    930                 935                 940


Leu Lys Glu Tyr Val Asp Phe Glu Asp Ser Ala Met Pro Ser Asp Tyr
945                 950                 955                 960


Ser Ala Leu Leu Ala Phe Ser Asn Glu Gly Lys Ile Asp Leu Tyr Val
                965                 970                 975


Asp Ala Ser Asn Pro Lys Val Asn Arg Asn Ile Ile Gln Ala Lys Leu
            980                 985                 990


Tyr Ala Pro Asp Met Val Leu Lys  Lys Val Val Lys Lys  Ile Ser Gln
        995                 1000                1005


Asp Glu  Cys Lys Glu Phe Asn  Glu Lys Lys Glu Gln  Ile Met Gln
    1010                1015                1020


Phe Lys  Asn Lys Gly Asp Glu  Val Ser Trp Glu Glu  Gln Gln Lys
    1025                1030                1035


Ile Leu  Glu Tyr Gln Lys Leu  Lys Asn Arg Val Glu  Leu Arg Asp
    1040                1045                1050


Leu Ser  Glu Tyr Gly Glu Leu  Ile Asn Glu Leu Leu  Gly Gln Leu
    1055                1060                1065


Ile Asn  Trp Ser Tyr Leu Arg  Glu Arg Asp Leu Leu  Tyr Phe Gln
    1070                1075                1080


Leu Gly  Phe His Tyr Ser Cys  Leu Met Asn Glu Ser  Lys Lys Pro
    1085                1090                1095


Asp Ala  Tyr Lys Thr Ile Arg  Arg Gly Thr Val Ser  Ile Glu Asn
    1100                1105                1110
```

```
Ala Val  Leu Tyr Gln Ile Ile  Ala Met Tyr Ile Asn  Gly Phe Pro
    1115                1120                1125

Val Tyr  Ala Pro Glu Lys Gly  Glu Leu Lys Pro Gln  Cys Lys Thr
    1130                1135                1140

Gly Ser  Ala Gly Gln Lys Ile  Arg Ala Phe Cys Gln  Trp Ala Ser
    1145                1150                1155

Met Val  Glu Lys Lys Lys Tyr  Glu Leu Tyr Asn Ala  Gly Leu Glu
    1160                1165                1170

Leu Phe  Glu Val Val Lys Glu  His Asp Asn Ile Ile  Asp Leu Arg
    1175                1180                1185

Asn Lys  Ile Asp His Phe Lys  Tyr Tyr Gln Gly Asn  Asp Ser Ile
    1190                1195                1200

Leu Ala  Leu Tyr Gly Glu Ile  Phe Asp Arg Phe Phe  Thr Tyr Asp
    1205                1210                1215

Met Lys  Tyr Arg Asn Asn Val  Leu Asn His Leu Gln  Asn Ile Leu
    1220                1225                1230

Leu Arg  His Asn Val Ile Ile  Lys Pro Ile Ile Ser  Lys Asp Lys
    1235                1240                1245

Lys Glu  Val Gly Arg Gly Lys  Met Lys Asp Arg Ala  Ala Phe Leu
    1250                1255                1260

Leu Glu  Glu Val Ser Ser Asp  Arg Phe Thr Tyr Lys  Val Lys Glu
    1265                1270                1275

Gly Glu  Arg Lys Ile Asp Ala  Lys Asn Arg Leu Tyr  Leu Glu Thr
    1280                1285                1290

Val Arg  Asp Ile Leu Tyr Phe  Pro Asn Arg Ala Val  Asn Asp Lys
    1295                1300                1305
```

```
Gly Glu  Asp Val Ile Ile Cys  Ser Lys Lys Ala Gln  Asp Leu Asn
    1310                1315                1320


Glu Lys  Lys Ala Asp Arg Asp  Lys Asn His Asp Lys  Ser Lys Asp
    1325                1330                1335


Thr Asn  Gln Lys Lys Glu Gly  Lys Asn Gln Glu Glu  Lys Ser Glu
    1340                1345                1350


Asn Lys  Glu Pro Tyr Ser Asp  Arg Met Thr Trp Lys  Pro Phe Ala
    1355                1360                1365


Gly Ile  Lys Leu Glu
    1370


<210>  53
<211>  1276
<212>  PRT
<213>  Blautia sp.

<400>  53

Met Lys Ile Ser Lys Val Asp His Val Lys Ser Gly Ile Asp Gln Lys
1               5                   10                  15


Leu Ser Ser Gln Arg Gly Met Leu Tyr Lys Gln Pro Gln Lys Lys Tyr
            20                  25                  30


Glu Gly Lys Gln Leu Glu Glu His Val Arg Asn Leu Ser Arg Lys Ala
        35                  40                  45


Lys Ala Leu Tyr Gln Val Phe Pro Val Ser Gly Asn Ser Lys Met Glu
    50                  55                  60


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


Leu Asp Ser Gly Lys Thr Ser Glu Glu Ile Val Gly Tyr Ile Asn Thr
                85                  90                  95
```

```
Tyr Ser Val Ala Ser Gln Ile Ser Gly Asp His Ile Gln Glu Leu Val
            100                 105                 110


Asp Gln His Leu Lys Glu Ser Leu Arg Lys Tyr Thr Cys Val Gly Asp
        115                 120                 125


Lys Arg Ile Tyr Val Pro Asp Ile Ile Val Ala Leu Leu Lys Ser Lys
    130                 135                 140


Phe Asn Ser Glu Thr Leu Gln Tyr Asp Asn Ser Glu Leu Lys Ile Leu
145                 150                 155                 160


Ile Asp Phe Ile Arg Glu Asp Tyr Leu Lys Glu Lys Gln Ile Lys Gln
                165                 170                 175


Ile Val His Ser Ile Glu Asn Asn Ser Thr Pro Leu Arg Ile Ala Glu
            180                 185                 190


Ile Asn Gly Gln Lys Arg Leu Ile Pro Ala Asn Val Asp Asn Pro Lys
        195                 200                 205


Lys Ser Tyr Ile Phe Glu Phe Leu Lys Glu Tyr Ala Gln Ser Asp Pro
    210                 215                 220


Lys Gly Gln Glu Ser Leu Leu Gln His Met Arg Tyr Leu Ile Leu Leu
225                 230                 235                 240


Tyr Leu Tyr Gly Pro Asp Lys Ile Thr Asp Asp Tyr Cys Glu Glu Ile
                245                 250                 255


Glu Ala Trp Asn Phe Gly Ser Ile Val Met Asp Asn Glu Gln Leu Phe
            260                 265                 270


Ser Glu Glu Ala Ser Met Leu Ile Gln Asp Arg Ile Tyr Val Asn Gln
        275                 280                 285


Gln Ile Glu Glu Gly Arg Gln Ser Lys Asp Thr Ala Lys Val Lys Lys
    290                 295                 300
```

```
Asn Lys Ser Lys Tyr Arg Met Leu Gly Asp Lys Ile Glu His Ser Ile
305                 310                 315                 320


Asn Glu Ser Val Val Lys His Tyr Gln Glu Ala Cys Lys Ala Val Glu
                325                 330                 335


Glu Lys Asp Ile Pro Trp Ile Lys Tyr Ile Ser Asp His Val Met Ser
            340                 345                 350


Val Tyr Ser Ser Lys Asn Arg Val Asp Leu Asp Lys Leu Ser Leu Pro
        355                 360                 365


Tyr Leu Ala Lys Asn Thr Trp Asn Thr Trp Ile Ser Phe Ile Ala Met
    370                 375                 380


Lys Tyr Val Asp Met Gly Lys Gly Val Tyr His Phe Ala Met Ser Asp
385                 390                 395                 400


Val Asp Lys Val Gly Lys Gln Asp Asn Leu Ile Ile Gly Gln Ile Asp
                405                 410                 415


Pro Lys Phe Ser Asp Gly Ile Ser Ser Phe Asp Tyr Glu Arg Ile Lys
            420                 425                 430


Ala Glu Asp Asp Leu His Arg Ser Met Ser Gly Tyr Ile Ala Phe Ala
        435                 440                 445


Val Asn Asn Phe Ala Arg Ala Ile Cys Ser Asp Glu Phe Arg Lys Lys
    450                 455                 460


Asn Arg Lys Glu Asp Val Leu Thr Val Gly Leu Asp Glu Ile Pro Leu
465                 470                 475                 480


Tyr Asp Asn Val Lys Arg Lys Leu Leu Gln Tyr Phe Gly Gly Ala Ser
                485                 490                 495


Asn Trp Asp Asp Ser Ile Ile Asp Ile Ile Asp Asp Lys Asp Leu Val
            500                 505                 510
```

```
Ala Cys Ile Lys Glu Asn Leu Tyr Val Ala Arg Asn Val Asn Phe His
        515                 520                 525


Phe Ala Gly Ser Glu Lys Val Gln Lys Lys Gln Asp Asp Ile Leu Glu
    530                 535                 540


Glu Ile Val Arg Lys Glu Thr Arg Asp Ile Gly Lys His Tyr Arg Lys
545                 550                 555                 560


Val Phe Tyr Ser Asn Asn Val Ala Val Phe Tyr Cys Asp Glu Asp Ile
                565                 570                 575


Ile Lys Leu Met Asn His Leu Tyr Gln Arg Glu Lys Pro Tyr Gln Ala
            580                 585                 590


Gln Ile Pro Ser Tyr Asn Lys Val Ile Ser Lys Thr Tyr Leu Pro Asp
        595                 600                 605


Leu Ile Phe Met Leu Leu Lys Gly Lys Asn Arg Thr Lys Ile Ser Asp
    610                 615                 620


Pro Ser Ile Met Asn Met Phe Arg Gly Thr Phe Tyr Phe Leu Leu Lys
625                 630                 635                 640


Glu Ile Tyr Tyr Asn Asp Phe Leu Gln Ala Ser Asn Leu Lys Glu Met
                645                 650                 655


Phe Cys Glu Gly Leu Lys Asn Asn Val Lys Asn Lys Lys Ser Glu Lys
            660                 665                 670


Pro Tyr Gln Asn Phe Met Arg Arg Phe Glu Glu Leu Glu Asn Met Gly
        675                 680                 685


Met Asp Phe Gly Glu Ile Cys Gln Gln Ile Met Thr Asp Tyr Glu Gln
    690                 695                 700


Gln Asn Lys Gln Lys Lys Lys Thr Ala Thr Ala Val Met Ser Glu Lys
705                 710                 715                 720
```

```
Asp Lys Lys Ile Arg Thr Leu Asp Asn Asp Thr Gln Lys Tyr Lys His
                725                 730                 735

Phe Arg Thr Leu Leu Tyr Ile Gly Leu Arg Glu Ala Phe Ile Ile Tyr
            740                 745                 750

Leu Lys Asp Glu Lys Asn Lys Glu Trp Tyr Glu Phe Leu Arg Glu Pro
        755                 760                 765

Val Lys Arg Glu Gln Pro Glu Glu Lys Glu Phe Val Asn Lys Trp Lys
    770                 775                 780

Leu Asn Gln Tyr Ser Asp Cys Ser Glu Leu Ile Leu Lys Asp Ser Leu
785                 790                 795                 800

Ala Ala Ala Trp Tyr Val Val Ala His Phe Ile Asn Gln Ala Gln Leu
                805                 810                 815

Asn His Leu Ile Gly Asp Ile Lys Asn Tyr Ile Gln Phe Ile Ser Asp
            820                 825                 830

Ile Asp Arg Arg Ala Lys Ser Thr Gly Asn Pro Val Ser Glu Ser Thr
        835                 840                 845

Glu Ile Gln Ile Glu Arg Tyr Arg Lys Ile Leu Arg Val Leu Glu Phe
    850                 855                 860

Ala Lys Phe Phe Cys Gly Gln Ile Thr Asn Val Leu Thr Asp Tyr Tyr
865                 870                 875                 880

Gln Asp Glu Asn Asp Phe Ser Thr His Val Gly His Tyr Val Lys Phe
                885                 890                 895

Glu Lys Lys Asn Met Glu Pro Ala His Ala Leu Gln Ala Phe Ser Asn
            900                 905                 910

Ser Leu Tyr Ala Cys Gly Lys Glu Lys Lys Lys Ala Gly Phe Tyr Tyr
        915                 920                 925
```

```
Asp Gly Met Asn Pro Ile Val Asn Arg Asn Ile Thr Leu Ala Ser Met
    930                 935                 940

Tyr Gly Asn Lys Lys Leu Leu Glu Asn Ala Met Asn Pro Val Thr Glu
945                 950                 955                 960

Gln Asp Ile Arg Lys Tyr Tyr Ser Leu Met Ala Glu Leu Asp Ser Val
                965                 970                 975

Leu Lys Asn Gly Ala Val Cys Lys Ser Glu Asp Glu Gln Lys Asn Leu
            980                 985                 990

Arg His Phe Gln Asn Leu Lys Asn  Arg Ile Glu Leu Val  Asp Val Leu
        995                 1000                1005

Thr Leu  Ser Glu Leu Val Asn  Asp Leu Val Ala Gln  Leu Ile Gly
    1010                1015                1020

Trp Val  Tyr Ile Arg Glu Arg  Asp Met Met Tyr Leu  Gln Leu Gly
    1025                1030                1035

Leu His  Tyr Ile Lys Leu Tyr  Phe Thr Asp Ser Val  Ala Glu Asp
    1040                1045                1050

Ser Tyr  Leu Arg Thr Leu Asp  Leu Glu Glu Gly Ser  Ile Ala Asp
    1055                1060                1065

Gly Ala  Val Leu Tyr Gln Ile  Ala Ser Leu Tyr Ser  Phe Asn Leu
    1070                1075                1080

Pro Met  Tyr Val Lys Pro Asn  Lys Ser Ser Val Tyr  Cys Lys Lys
    1085                1090                1095

His Val  Asn Ser Val Ala Thr  Lys Phe Asp Ile Phe  Glu Lys Glu
    1100                1105                1110

Tyr Cys  Asn Gly Asp Glu Thr  Val Ile Glu Asn Gly  Leu Arg Leu
    1115                1120                1125
```

```
Phe Glu  Asn Ile Asn Leu His  Lys Asp Met Val Lys  Phe Arg Asp
    1130                1135                1140


Tyr Leu  Ala His Phe Lys Tyr  Phe Ala Lys Leu Asp  Glu Ser Ile
    1145                1150                1155


Leu Glu  Leu Tyr Ser Lys Ala  Tyr Asp Phe Phe Phe  Ser Tyr Asn
    1160                1165                1170


Ile Lys  Leu Lys Lys Ser Val  Ser Tyr Val Leu Thr  Asn Val Leu
    1175                1180                1185


Leu Ser  Tyr Phe Ile Asn Ala  Lys Leu Ser Phe Ser  Thr Tyr Lys
    1190                1195                1200


Ser Ser  Gly Asn Lys Thr Val  Gln His Arg Thr Thr  Lys Ile Ser
    1205                1210                1215


Val Val  Ala Gln Thr Asp Tyr  Phe Thr Tyr Lys Leu  Arg Ser Ile
    1220                1225                1230


Val Lys  Asn Lys Asn Gly Val  Glu Ser Ile Glu Asn  Asp Asp Arg
    1235                1240                1245


Arg Cys  Glu Val Val Asn Ile  Ala Ala Arg Asp Lys  Glu Phe Val
    1250                1255                1260


Asp Glu  Val Cys Asn Val Ile  Asn Tyr Asn Ser Asp  Lys
    1265                1270                1275
```

<210> 54
<211> 1385
<212> PRT
<213> Leptotrichia sp.

<400> 54

```
Met Gly Asn Leu Phe Gly His Lys Arg Trp Tyr Glu Val Arg Asp Lys
1               5                   10                  15


Lys Asp Phe Lys Ile Lys Arg Lys Val Lys Val Lys Arg Asn Tyr Asp
```

```
            20                  25                  30

Gly Asn Lys Tyr Ile Leu Asn Ile Asn Glu Asn Asn Asn Lys Glu Lys
        35                  40                  45

Ile Asp Asn Asn Lys Phe Ile Gly Glu Phe Val Asn Tyr Lys Lys Asn
    50                  55                  60

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

Leu Phe Lys Leu Lys Gly Lys Glu Glu Ile Ile Arg Ile Glu Asn Asn
                85                  90                  95

Asp Asp Phe Leu Glu Thr Glu Glu Val Val Leu Tyr Ile Glu Val Tyr
            100                 105                 110

Gly Lys Ser Glu Lys Leu Lys Ala Leu Glu Ile Thr Lys Lys Lys Ile
        115                 120                 125

Ile Asp Glu Ala Ile Arg Gln Gly Ile Thr Lys Asp Asp Lys Lys Ile
    130                 135                 140

Glu Ile Lys Arg Gln Glu Asn Glu Glu Glu Ile Glu Ile Asp Ile Arg
145                 150                 155                 160

Asp Glu Tyr Thr Asn Lys Thr Leu Asn Asp Cys Ser Ile Ile Leu Arg
                165                 170                 175

Ile Ile Glu Asn Asp Glu Leu Glu Thr Lys Lys Ser Ile Tyr Glu Ile
            180                 185                 190

Phe Lys Asn Ile Asn Met Ser Leu Tyr Lys Ile Ile Glu Lys Ile Ile
        195                 200                 205

Glu Asn Glu Thr Glu Lys Val Phe Glu Asn Arg Tyr Tyr Glu Glu His
    210                 215                 220

Leu Arg Glu Lys Leu Leu Lys Asp Asn Lys Ile Asp Val Ile Leu Thr
```

225 230 235 240

Asn Phe Met Glu Ile Arg Glu Lys Ile Lys Ser Asn Leu Glu Ile Met
245 250 255

Gly Phe Val Lys Phe Tyr Leu Asn Val Ser Gly Asp Lys Lys Lys Ser
260 265 270

Glu Asn Lys Lys Met Phe Val Glu Lys Ile Leu Asn Thr Asn Val Asp
275 280 285

Leu Thr Val Glu Asp Ile Val Asp Phe Ile Val Lys Glu Leu Lys Phe
290 295 300

Trp Asn Ile Thr Lys Arg Ile Glu Lys Val Lys Lys Phe Asn Asn Glu
305 310 315 320

Phe Leu Glu Asn Arg Arg Asn Arg Thr Tyr Ile Lys Ser Tyr Val Leu
325 330 335

Leu Asp Lys His Glu Lys Phe Lys Ile Glu Arg Glu Asn Lys Lys Asp
340 345 350

Lys Ile Val Lys Phe Phe Val Glu Asn Ile Lys Asn Asn Ser Ile Lys
355 360 365

Glu Lys Ile Glu Lys Ile Leu Ala Glu Phe Lys Ile Asn Glu Leu Ile
370 375 380

Lys Lys Leu Glu Lys Glu Leu Lys Lys Gly Asn Cys Asp Thr Glu Ile
385 390 395 400

Phe Gly Ile Phe Lys Lys His Tyr Lys Val Asn Phe Asp Ser Lys Lys
405 410 415

Phe Ser Asn Lys Ser Asp Glu Glu Lys Glu Leu Tyr Lys Ile Ile Tyr
420 425 430

Arg Tyr Leu Lys Gly Arg Ile Glu Lys Ile Leu Val Asn Glu Gln Lys

```
        435                 440                 445

Val Arg Leu Lys Lys Met Glu Lys Ile Glu Ile Glu Lys Ile Leu Asn
    450                 455                 460

Glu Ser Ile Leu Ser Glu Lys Ile Leu Lys Arg Val Lys Gln Tyr Thr
465                 470                 475                 480

Leu Glu His Ile Met Tyr Leu Gly Lys Leu Arg His Asn Asp Ile Val
                485                 490                 495

Lys Met Thr Val Asn Thr Asp Asp Phe Ser Arg Leu His Ala Lys Glu
            500                 505                 510

Glu Leu Asp Leu Glu Leu Ile Thr Phe Phe Ala Ser Thr Asn Met Glu
        515                 520                 525

Leu Asn Lys Ile Phe Asn Gly Lys Glu Lys Val Thr Asp Phe Phe Gly
    530                 535                 540

Phe Asn Leu Asn Gly Gln Lys Ile Thr Leu Lys Glu Lys Val Pro Ser
545                 550                 555                 560

Phe Lys Leu Asn Ile Leu Lys Lys Leu Asn Phe Ile Asn Asn Glu Asn
                565                 570                 575

Asn Ile Asp Glu Lys Leu Ser His Phe Tyr Ser Phe Gln Lys Glu Gly
            580                 585                 590

Tyr Leu Leu Arg Asn Lys Ile Leu His Asn Ser Tyr Gly Asn Ile Gln
        595                 600                 605

Glu Thr Lys Asn Leu Lys Gly Glu Tyr Glu Asn Val Glu Lys Leu Ile
    610                 615                 620

Lys Glu Leu Lys Val Ser Asp Glu Glu Ile Ser Lys Ser Leu Ser Leu
625                 630                 635                 640

Asp Val Ile Phe Glu Gly Lys Val Asp Ile Ile Asn Lys Ile Asn Ser
```

645 650 655

Leu Lys Ile Gly Glu Tyr Lys Asp Lys Lys Tyr Leu Pro Ser Phe Ser
660 665 670

Lys Ile Val Leu Glu Ile Thr Arg Lys Phe Arg Glu Ile Asn Lys Asp
675 680 685

Lys Leu Phe Asp Ile Glu Ser Glu Lys Ile Ile Leu Asn Ala Val Lys
690 695 700

Tyr Val Asn Lys Ile Leu Tyr Glu Lys Ile Thr Ser Asn Glu Glu Asn
705 710 715 720

Glu Phe Leu Lys Thr Leu Pro Asp Lys Leu Val Lys Lys Ser Asn Asn
725 730 735

Lys Lys Glu Asn Lys Asn Leu Leu Ser Ile Glu Glu Tyr Tyr Lys Asn
740 745 750

Ala Gln Val Ser Ser Ser Lys Gly Asp Lys Lys Ala Ile Lys Lys Tyr
755 760 765

Gln Asn Lys Val Thr Asn Ala Tyr Leu Glu Tyr Leu Glu Asn Thr Phe
770 775 780

Thr Glu Ile Ile Asp Phe Ser Lys Phe Asn Leu Asn Tyr Asp Glu Ile
785 790 795 800

Lys Thr Lys Ile Glu Glu Arg Lys Asp Asn Lys Ser Lys Ile Ile Ile
805 810 815

Asp Ser Ile Ser Thr Asn Ile Asn Ile Thr Asn Asp Ile Glu Tyr Ile
820 825 830

Ile Ser Ile Phe Ala Leu Leu Asn Ser Asn Thr Tyr Ile Asn Lys Ile
835 840 845

Arg Asn Arg Phe Phe Ala Thr Ser Val Trp Leu Glu Lys Gln Asn Gly

```
    850                 855                 860

Thr Lys Glu Tyr Asp Tyr Glu Asn Ile Ile Ser Ile Leu Asp Glu Val
865                 870                 875                 880

Leu Leu Ile Asn Leu Leu Arg Glu Asn Asn Ile Thr Asp Ile Leu Asp
                885                 890                 895

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

Tyr Ile Lys Asn Tyr Ile Phe Glu Ser Asn Glu Glu Lys Leu Lys Lys
        915                 920                 925

Arg Leu Phe Cys Glu Glu Leu Val Asp Lys Glu Asp Ile Arg Lys Ile
    930                 935                 940

Phe Glu Asp Glu Asn Phe Lys Phe Lys Ser Phe Ile Lys Lys Asn Glu
945                 950                 955                 960

Ile Gly Asn Phe Lys Ile Asn Phe Gly Ile Leu Ser Asn Leu Glu Cys
                965                 970                 975

Asn Ser Glu Val Glu Ala Lys Lys Ile Ile Gly Lys Asn Ser Lys Lys
            980                 985                 990

Leu Glu Ser Phe Ile Gln Asn Ile  Ile Asp Glu Tyr Lys  Ser Asn Ile
        995                 1000                1005

Arg Thr  Leu Phe Ser Ser Glu  Phe Leu Glu Lys Tyr  Lys Glu Glu
    1010                1015                1020

Ile Asp  Asn Leu Val Glu Asp  Thr Glu Ser Glu Asn  Lys Asn Lys
    1025                1030                1035

Phe Glu  Lys Ile Tyr Tyr Pro  Lys Glu His Lys Asn  Glu Leu Tyr
    1040                1045                1050

Ile Tyr  Lys Lys Asn Leu Phe  Leu Asn Ile Gly Asn  Pro Asn Phe
```

```
    1055                1060                1065

Asp Lys  Ile Tyr Gly Leu Ile  Ser Lys Asp Ile Lys  Asn Val Asp
    1070                1075                1080

Thr Lys  Ile Leu Phe Asp Asp  Asp Ile Lys Lys Asn  Lys Ile Ser
    1085                1090                1095

Glu Ile  Asp Ala Ile Leu Lys  Asn Leu Asn Asp Lys  Leu Asn Gly
    1100                1105                1110

Tyr Ser  Asn Asp Tyr Lys Ala  Lys Tyr Val Asn Lys  Leu Lys Glu
    1115                1120                1125

Asn Asp  Asp Phe Phe Ala Lys  Asn Ile Gln Asn Glu  Asn Tyr Ser
    1130                1135                1140

Ser Phe  Gly Glu Phe Glu Lys  Asp Tyr Asn Lys Val  Ser Glu Tyr
    1145                1150                1155

Lys Lys  Ile Arg Asp Leu Val  Glu Phe Asn Tyr Leu  Asn Lys Ile
    1160                1165                1170

Glu Ser  Tyr Leu Ile Asp Ile  Asn Trp Lys Leu Ala  Ile Gln Met
    1175                1180                1185

Ala Arg  Phe Glu Arg Asp Met  His Tyr Ile Val Asn  Gly Leu Arg
    1190                1195                1200

Glu Leu  Gly Ile Ile Lys Leu  Ser Gly Tyr Asn Thr  Gly Ile Ser
    1205                1210                1215

Arg Ala  Tyr Pro Lys Arg Asn  Gly Ser Asp Gly Phe  Tyr Thr Thr
    1220                1225                1230

Thr Ala  Tyr Tyr Lys Phe Phe  Asp Glu Glu Ser Tyr  Lys Lys Phe
    1235                1240                1245

Glu Lys  Ile Cys Tyr Gly Phe  Gly Ile Asp Leu Ser  Glu Asn Ser
```

1250 1255 1260

Glu Ile Asn Lys Pro Glu Asn Glu Ser Ile Arg Asn Tyr Ile Ser
1265 1270 1275

His Phe Tyr Ile Val Arg Asn Pro Phe Ala Asp Tyr Ser Ile Ala
1280 1285 1290

Glu Gln Ile Asp Arg Val Ser Asn Leu Leu Ser Tyr Ser Thr Arg
1295 1300 1305

Tyr Asn Asn Ser Thr Tyr Ala Ser Val Phe Glu Val Phe Lys Lys
1310 1315 1320

Asp Val Asn Leu Asp Tyr Asp Glu Leu Lys Lys Lys Phe Arg Leu
1325 1330 1335

Ile Gly Asn Asn Asp Ile Leu Glu Arg Leu Met Lys Pro Lys Lys
1340 1345 1350

Val Ser Val Leu Glu Leu Glu Ser Tyr Asn Ser Asp Tyr Ile Lys
1355 1360 1365

Asn Leu Ile Ile Glu Leu Leu Thr Lys Ile Glu Asn Thr Asn Asp
1370 1375 1380

Thr Leu
1385

<210> 55
<211> 1334
<212> PRT
<213> Lachnospiraceae bacteroi,

<400> 55

Met Lys Ile Ser Lys Val Asp His Thr Arg Met Ala Val Ala Lys Gly
1 5 10 15

Asn Gln His Arg Arg Asp Glu Ile Ser Gly Ile Leu Tyr Lys Asp Pro
20 25 30

```
Thr Lys Thr Gly Ser Ile Asp Phe Asp Glu Arg Phe Lys Lys Leu Asn
        35                  40                  45

Cys Ser Ala Lys Ile Leu Tyr His Val Phe Asn Gly Ile Ala Glu Gly
    50                  55                  60

Ser Asn Lys Tyr Lys Asn Ile Val Asp Lys Val Asn Asn Asn Leu Asp
65                  70                  75                  80

Arg Val Leu Phe Thr Gly Lys Ser Tyr Asp Arg Lys Ser Ile Ile Asp
                85                  90                  95

Ile Asp Thr Val Leu Arg Asn Val Glu Lys Ile Asn Ala Phe Asp Arg
            100                 105                 110

Ile Ser Thr Glu Glu Arg Glu Gln Ile Ile Asp Asp Leu Leu Glu Ile
        115                 120                 125

Gln Leu Arg Lys Gly Leu Arg Lys Gly Lys Ala Gly Leu Arg Glu Val
    130                 135                 140

Leu Leu Ile Gly Ala Gly Val Ile Val Arg Thr Asp Lys Lys Gln Glu
145                 150                 155                 160

Ile Ala Asp Phe Leu Glu Ile Leu Asp Glu Asp Phe Asn Lys Thr Asn
                165                 170                 175

Gln Ala Lys Asn Ile Lys Leu Ser Ile Glu Asn Gln Gly Leu Val Val
            180                 185                 190

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

Gln Lys Gly Lys Ser Ser Lys Lys Ala Gln Glu Lys Glu Ala Leu Ser
    210                 215                 220

Ala Phe Leu Leu Asp Tyr Ala Asp Leu Asp Lys Asn Val Arg Phe Glu
225                 230                 235                 240
```

```
Tyr Leu Arg Lys Ile Arg Arg Leu Ile Asn Leu Tyr Phe Tyr Val Lys
                245                 250                 255

Asn Asp Asp Val Met Ser Leu Thr Glu Ile Pro Ala Glu Val Asn Leu
            260                 265                 270

Glu Lys Asp Phe Asp Ile Trp Arg Asp His Glu Gln Arg Lys Glu Glu
        275                 280                 285

Asn Gly Asp Phe Val Gly Cys Pro Asp Ile Leu Leu Ala Asp Arg Asp
    290                 295                 300

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

Arg Glu Ser Ile Arg Glu Lys Asn Ile Lys Arg Tyr Arg Phe Ser Ile
                325                 330                 335

Lys Thr Ile Glu Lys Asp Asp Gly Thr Tyr Phe Phe Ala Asn Lys Gln
            340                 345                 350

Ile Ser Val Phe Trp Ile His Arg Ile Glu Asn Ala Val Glu Arg Ile
        355                 360                 365

Leu Gly Ser Ile Asn Asp Lys Lys Leu Tyr Arg Leu Arg Leu Gly Tyr
    370                 375                 380

Leu Gly Glu Lys Val Trp Lys Asp Ile Leu Asn Phe Leu Ser Ile Lys
385                 390                 395                 400

Tyr Ile Ala Val Gly Lys Ala Val Phe Asn Phe Ala Met Asp Asp Leu
                405                 410                 415

Gln Glu Lys Asp Arg Asp Ile Glu Pro Gly Lys Ile Ser Glu Asn Ala
            420                 425                 430

Val Asn Gly Leu Thr Ser Phe Asp Tyr Glu Gln Ile Lys Ala Asp Glu
        435                 440                 445
```

```
Met Leu Gln Arg Glu Val Ala Val Asn Val Ala Phe Ala Ala Asn Asn
    450                 455                 460

Leu Ala Arg Val Thr Val Asp Ile Pro Gln Asn Gly Glu Lys Glu Asp
465                 470                 475                 480

Ile Leu Leu Trp Asn Lys Ser Asp Ile Lys Lys Tyr Lys Lys Asn Ser
                485                 490                 495

Lys Lys Gly Ile Leu Lys Ser Ile Leu Gln Phe Phe Gly Gly Ala Ser
            500                 505                 510

Thr Trp Asn Met Lys Met Phe Glu Ile Ala Tyr His Asp Gln Pro Gly
        515                 520                 525

Asp Tyr Glu Glu Asn Tyr Leu Tyr Asp Ile Ile Gln Ile Ile Tyr Ser
    530                 535                 540

Leu Arg Asn Lys Ser Phe His Phe Lys Thr Tyr Asp His Gly Asp Lys
545                 550                 555                 560

Asn Trp Asn Arg Glu Leu Ile Gly Lys Met Ile Glu His Asp Ala Glu
                565                 570                 575

Arg Val Ile Ser Val Glu Arg Glu Lys Phe His Ser Asn Asn Leu Pro
            580                 585                 590

Met Phe Tyr Lys Asp Ala Asp Leu Lys Lys Ile Leu Asp Leu Leu Tyr
        595                 600                 605

Ser Asp Tyr Ala Gly Arg Ala Ser Gln Val Pro Ala Phe Asn Thr Val
    610                 615                 620

Leu Val Arg Lys Asn Phe Pro Glu Phe Leu Arg Lys Asp Met Gly Tyr
625                 630                 635                 640

Lys Val His Phe Asn Asn Pro Glu Val Glu Asn Gln Trp His Ser Ala
                645                 650                 655
```

```
Val Tyr Tyr Leu Tyr Lys Glu Ile Tyr Tyr Asn Leu Phe Leu Arg Asp
            660                 665                 670

Lys Glu Val Lys Asn Leu Phe Tyr Thr Ser Leu Lys Asn Ile Arg Ser
        675                 680                 685

Glu Val Ser Asp Lys Lys Gln Lys Leu Ala Ser Asp Asp Phe Ala Ser
    690                 695                 700

Arg Cys Glu Glu Ile Glu Asp Arg Ser Leu Pro Glu Ile Cys Gln Ile
705                 710                 715                 720

Ile Met Thr Glu Tyr Asn Ala Gln Asn Phe Gly Asn Arg Lys Val Lys
                725                 730                 735

Ser Gln Arg Val Ile Glu Lys Asn Lys Asp Ile Phe Arg His Tyr Lys
            740                 745                 750

Met Leu Leu Ile Lys Thr Leu Ala Gly Ala Phe Ser Leu Tyr Leu Lys
        755                 760                 765

Gln Glu Arg Phe Ala Phe Ile Gly Lys Ala Thr Pro Ile Pro Tyr Glu
    770                 775                 780

Thr Thr Asp Val Lys Asn Phe Leu Pro Glu Trp Lys Ser Gly Met Tyr
785                 790                 795                 800

Ala Ser Phe Val Glu Glu Ile Lys Asn Asn Leu Asp Leu Gln Glu Trp
                805                 810                 815

Tyr Ile Val Gly Arg Phe Leu Asn Gly Arg Met Leu Asn Gln Leu Ala
            820                 825                 830

Gly Ser Leu Arg Ser Tyr Ile Gln Tyr Ala Glu Asp Ile Glu Arg Arg
        835                 840                 845

Ala Ala Glu Asn Arg Asn Lys Leu Phe Ser Lys Pro Asp Glu Lys Ile
    850                 855                 860
```

```
Glu Ala Cys Lys Lys Ala Val Arg Val Leu Asp Leu Cys Ile Lys Ile
865                 870                 875                 880

Ser Thr Arg Ile Ser Ala Glu Phe Thr Asp Tyr Phe Asp Ser Glu Asp
                885                 890                 895

Asp Tyr Ala Asp Tyr Leu Glu Lys Tyr Leu Lys Tyr Gln Asp Asp Ala
            900                 905                 910

Ile Lys Glu Leu Ser Gly Ser Ser Tyr Ala Ala Leu Asp His Phe Cys
        915                 920                 925

Asn Lys Asp Asp Leu Lys Phe Asp Ile Tyr Val Asn Ala Gly Gln Lys
    930                 935                 940

Pro Ile Leu Gln Arg Asn Ile Val Met Ala Lys Leu Phe Gly Pro Asp
945                 950                 955                 960

Asn Ile Leu Ser Glu Val Met Glu Lys Val Thr Glu Ser Ala Ile Arg
                965                 970                 975

Glu Tyr Tyr Asp Tyr Leu Lys Lys Val Ser Gly Tyr Arg Val Arg Gly
            980                 985                 990

Lys Cys Ser Thr Glu Lys Glu Gln  Glu Asp Leu Leu Lys  Phe Gln Arg
        995                 1000                1005

Leu Lys  Asn Ala Val Glu Phe  Arg Asp Val Thr Glu  Tyr Ala Glu
    1010                1015                1020

Val Ile  Asn Glu Leu Leu Gly  Gln Leu Ile Ser Trp  Ser Tyr Leu
    1025                1030                1035

Arg Glu  Arg Asp Leu Leu Tyr  Phe Gln Leu Gly Phe  His Tyr Met
    1040                1045                1050

Cys Leu  Lys Asn Lys Ser Phe  Lys Pro Ala Glu Tyr  Val Asp Ile
    1055                1060                1065
```

```
Arg Arg  Asn Asn Gly Thr Ile  Ile His Asn Ala Ile  Leu Tyr Gln
    1070                1075                1080

Ile Val  Ser Met Tyr Ile Asn  Gly Leu Asp Phe Tyr  Ser Cys Asp
    1085                1090                1095

Lys Glu  Gly Lys Thr Leu Lys  Pro Ile Glu Thr Gly  Lys Gly Val
    1100                1105                1110

Gly Ser  Lys Ile Gly Gln Phe  Ile Lys Tyr Ser Gln  Tyr Leu Tyr
    1115                1120                1125

Asn Asp  Pro Ser Tyr Lys Leu  Glu Ile Tyr Asn Ala  Gly Leu Glu
    1130                1135                1140

Val Phe  Glu Asn Ile Asp Glu  His Asp Asn Ile Thr  Asp Leu Arg
    1145                1150                1155

Lys Tyr  Val Asp His Phe Lys  Tyr Tyr Ala Tyr Gly  Asn Lys Met
    1160                1165                1170

Ser Leu  Leu Asp Leu Tyr Ser  Glu Phe Phe Asp Arg  Phe Phe Thr
    1175                1180                1185

Tyr Asp  Met Lys Tyr Gln Lys  Asn Val Val Asn Val  Leu Glu Asn
    1190                1195                1200

Ile Leu  Leu Arg His Phe Val  Ile Phe Tyr Pro Lys  Phe Gly Ser
    1205                1210                1215

Gly Lys  Lys Asp Val Gly Ile  Arg Asp Cys Lys Lys  Glu Arg Ala
    1220                1225                1230

Gln Ile  Glu Ile Ser Glu Gln  Ser Leu Thr Ser Glu  Asp Phe Met
    1235                1240                1245

Phe Lys  Leu Asp Asp Lys Ala  Gly Glu Glu Ala Lys  Lys Phe Pro
    1250                1255                1260
```

```
Ala Arg  Asp Glu Arg Tyr Leu  Gln Thr Ile Ala Lys  Leu Leu Tyr
    1265                1270                1275


Tyr Pro  Asn Glu Ile Glu Asp  Met Asn Arg Phe Met  Lys Lys Gly
    1280                1285                1290


Glu Thr  Ile Asn Lys Lys Val  Gln Phe Asn Arg Lys  Lys Lys Ile
    1295                1300                1305


Thr Arg  Lys Gln Lys Asn Asn  Ser Ser Asn Glu Val  Leu Ser Ser
    1310                1315                1320


Thr Met  Gly Tyr Leu Phe Lys  Asn Ile Lys Leu
    1325                1330


<210>  56
<211>  500
<212>  PRT
<213>  Chloroflexus aggregans

<400>  56

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


Asp Thr Pro Leu Ala Ala Ala Gln Thr Pro Ala Ala Asp Ala Ala Val
            20                  25                  30


Ala Ala Thr Pro Ala Pro Ala Glu Ala Val Ala Pro Thr Pro Glu Gln
        35                  40                  45


Ala Val Asp Gln Pro Ala Thr Thr Gly Glu Ser Glu Ala Pro Val Thr
    50                  55                  60


Thr Ala Gln Ala Ala Ala His Glu Ala Glu Pro Ala Glu Ala Thr Gly
65                  70                  75                  80


Ala Ser Phe Thr Pro Val Ser Glu Gln Gln Pro Gln Lys Pro Arg Arg
                85                  90                  95
```

```
Leu Lys Asp Leu Gln Pro Gly Met Glu Leu Glu Gly Lys Val Thr Ser
            100                 105                 110


Ile Ala Leu Tyr Gly Ile Phe Val Asp Val Gly Val Gly Arg Asp Gly
        115                 120                 125


Leu Val His Ile Ser Glu Met Ser Asp Arg Arg Ile Asp Thr Pro Ser
    130                 135                 140


Glu Leu Val Gln Ile Gly Asp Thr Val Lys Val Trp Val Lys Ser Val
145                 150                 155                 160


Asp Leu Asp Ala Arg Arg Ile Ser Leu Thr Met Leu Asn Pro Ser Arg
                165                 170                 175


Gly Glu Lys Pro Arg Arg Ser Arg Gln Ser Gln Pro Ala Gln Pro Gln
            180                 185                 190


Pro Arg Arg Gln Glu Val Asp Arg Glu Lys Leu Ala Ser Leu Lys Val
        195                 200                 205


Gly Glu Ile Val Glu Gly Val Ile Thr Gly Phe Ala Pro Phe Gly Ala
    210                 215                 220


Phe Ala Asp Ile Gly Val Gly Lys Asp Gly Leu Ile His Ile Ser Glu
225                 230                 235                 240


Leu Ser Glu Gly Arg Val Glu Lys Pro Glu Asp Ala Val Lys Val Gly
                245                 250                 255


Glu Arg Tyr Gln Phe Lys Val Leu Glu Ile Asp Gly Glu Gly Thr Arg
            260                 265                 270


Ile Ser Leu Ser Leu Arg Arg Ala Gln Arg Thr Gln Arg Met Gln Gln
        275                 280                 285


Leu Glu Pro Gly Gln Ile Ile Glu Gly Thr Val Ser Gly Ile Ala Thr
    290                 295                 300
```

```
Phe Gly Ala Phe Val Asp Ile Gly Val Gly Arg Asp Gly Leu Val His
305                 310                 315                 320


Ile Ser Ala Leu Ala Pro His Arg Val Ala Lys Val Glu Asp Val Val
                325                 330                 335


Lys Val Gly Asp Lys Val Lys Val Lys Val Leu Gly Val Asp Pro Gln
            340                 345                 350


Ser Lys Arg Ile Ser Leu Thr Met Arg Leu Glu Glu Glu Gln Pro Ala
        355                 360                 365


Thr Thr Ala Gly Asp Glu Ala Ala Glu Pro Ala Glu Glu Val Thr Pro
    370                 375                 380


Thr Arg Arg Gly Asn Leu Glu Arg Phe Ala Ala Ala Ala Gln Thr Ala
385                 390                 395                 400


Arg Glu Arg Ser Glu Arg Gly Glu Arg Ser Glu Arg Gly Glu Arg Arg
                405                 410                 415


Glu Arg Arg Glu Arg Arg Pro Ala Gln Ser Ser Pro Asp Thr Tyr Ile
            420                 425                 430


Val Gly Glu Asp Asp Asp Glu Ser Phe Glu Gly Asn Ala Thr Ile Glu
        435                 440                 445


Asp Leu Leu Thr Lys Phe Gly Gly Ser Ser Ser Arg Arg Asp Arg Asp
    450                 455                 460


Arg Arg Arg Arg His Glu Asp Asp Asp Asp Glu Glu Met Glu Arg Pro
465                 470                 475                 480


Ser Asn Arg Arg Gln Arg Glu Ala Ile Arg Arg Thr Leu Gln Gln Ile
                485                 490                 495


Gly Tyr Asp Glu
            500
```

<210> 57
<211> 249
<212> PRT
<213> Demequina aurantiaca

<400> 57

```
Met Asp Leu Thr Trp His Ala Leu Leu Ile Leu Phe Ile Val Ala Leu
1               5                  10                  15


Leu Ala Gly Phe Leu Asp Thr Leu Ala Gly Gly Gly Gly Leu Leu Thr
            20                  25                  30


Val Pro Ala Leu Leu Leu Thr Gly Ile Pro Pro Leu Gln Ala Leu Gly
        35                  40                  45


Thr Asn Lys Leu Gln Ser Ser Phe Gly Thr Gly Met Ala Thr Tyr Gln
    50                  55                  60


Val Ile Arg Lys Lys Arg Val His Trp Arg Asp Val Arg Trp Pro Met
65                  70                  75                  80


Val Trp Ala Phe Leu Gly Ser Ala Ala Gly Ala Val Ala Val Gln Phe
                85                  90                  95


Ile Asp Thr Asp Ala Leu Leu Ile Ile Ile Pro Val Val Leu Ala Leu
            100                 105                 110


Val Ala Ala Tyr Phe Leu Phe Val Pro Lys Ser His Leu Pro Pro Pro
        115                 120                 125


Glu Pro Arg Met Ser Asp Pro Ala Tyr Glu Ala Thr Leu Val Pro Ile
    130                 135                 140


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


Ala Leu Ser Gly Val Ala Leu Arg Ala Lys Thr Leu Val Gln Ser Thr
                165                 170                 175
```

```
Ala Ile Ala Lys Thr Leu Asn Phe Ala Thr Asn Phe Ala Ala Leu Leu
            180                 185                 190


Val Phe Ala Phe Ala Gly His Met Leu Trp Thr Val Gly Ala Val Met
        195                 200                 205


Ile Ala Gly Gln Leu Ile Gly Ala Tyr Ala Gly Ser His Met Leu Phe
    210                 215                 220


Arg Val Asn Pro Leu Val Leu Arg Val Leu Ile Val Val Met Ser Leu
225                 230                 235                 240


Gly Met Leu Ile Arg Val Leu Leu Asp
                245
```

<210> 58
<211> 1235
<212> PRT
<213> Thalassospira sp.

<400> 58

```
Met Arg Ile Ile Lys Pro Tyr Gly Arg Ser His Val Glu Gly Val Ala
1               5                   10                  15


Thr Gln Glu Pro Arg Arg Lys Leu Arg Leu Asn Ser Ser Pro Asp Ile
            20                  25                  30


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


Ala Gln Trp Ile Ser Ala Ile Asp Lys Ile Ala Thr Lys Pro Lys Pro
    50                  55                  60


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


Asp Ala Ala Trp Gln His Val Met Ala Glu Asn Leu Leu Pro Ala Ala
                85                  90                  95


Thr Asp Pro Ala Ile Arg Glu Lys Leu His Leu Ile Trp Gln Ser Lys
```

```
            100                 105                 110

Ile Ala Pro Trp Gly Thr Ala Arg Pro Gln Ala Glu Lys Asp Gly Lys
        115                 120                 125

Pro Thr Pro Lys Gly Gly Trp Tyr Glu Arg Phe Cys Gly Val Leu Ser
    130                 135                 140

Pro Glu Ala Ile Thr Gln Asn Val Ala Arg Gln Ile Ala Lys Asp Ile
145                 150                 155                 160

Tyr Asp His Leu His Val Ala Ala Lys Arg Lys Gly Arg Glu Pro Ala
                165                 170                 175

Lys Gln Gly Glu Ser Ser Asn Lys Pro Gly Lys Phe Lys Pro Asp Arg
            180                 185                 190

Lys Arg Gly Leu Ile Glu Glu Arg Ala Glu Ser Ile Ala Lys Asn Ala
        195                 200                 205

Leu Arg Pro Gly Ser His Ala Pro Cys Pro Trp Gly Pro Asp Asp Gln
    210                 215                 220

Ala Thr Tyr Glu Gln Ala Gly Asp Val Ala Gly Gln Ile Tyr Ala Ala
225                 230                 235                 240

Ala Arg Asp Cys Leu Glu Glu Lys Lys Arg Arg Ser Gly Asn Arg Asn
                245                 250                 255

Thr Ser Ser Val Gln Tyr Leu Pro Arg Asp Leu Ala Ala Lys Ile Leu
            260                 265                 270

Tyr Ala Gln Tyr Gly Arg Val Phe Gly Pro Asp Thr Thr Ile Lys Ala
        275                 280                 285

Ala Leu Asp Glu Gln Pro Ser Leu Phe Ala Leu His Lys Ala Ile Lys
    290                 295                 300

Asp Cys Tyr His Arg Leu Ile Asn Asp Ala Arg Lys Arg Asp Ile Leu
```

305 310 315 320

Arg Ile Leu Pro Arg Asn Met Ala Ala Leu Phe Arg Leu Val Arg Ala
325 330 335

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

Ile His Tyr His Ala Ser Glu Gln Gly Lys Ser Glu His His Gly Ile
355 360 365

Arg Asp Tyr Trp Pro Ser Gln Gln Asp Ile Gln Asn Ser Arg Phe Trp
370 375 380

Gly Ser Asp Gly Gln Ala Asp Ile Lys Arg His Glu Ala Phe Ser Arg
385 390 395 400

Ile Trp Arg His Ile Ile Ala Leu Ala Ser Arg Thr Leu His Asp Trp
405 410 415

Ala Asp Pro His Ser Gln Lys Phe Ser Gly Glu Asn Asp Asp Ile Leu
420 425 430

Leu Leu Ala Lys Asp Ala Ile Glu Asp Asp Val Phe Lys Ala Gly His
435 440 445

Tyr Glu Arg Lys Cys Asp Val Leu Phe Gly Ala Gln Ala Ser Leu Phe
450 455 460

Cys Gly Ala Glu Asp Phe Glu Lys Ala Ile Leu Lys Gln Ala Ile Thr
465 470 475 480

Gly Thr Gly Asn Leu Arg Asn Ala Thr Phe His Phe Lys Gly Lys Val
485 490 495

Arg Phe Glu Lys Glu Leu Gln Glu Leu Thr Lys Asp Val Pro Val Glu
500 505 510

Val Gln Ser Ala Ile Ala Ala Leu Trp Gln Lys Asp Ala Glu Gly Arg

```
        515                 520                 525

Thr Arg Gln Ile Ala Glu Thr Leu Gln Ala Val Leu Ala Gly His Phe
    530                 535                 540

Leu Thr Glu Glu Gln Asn Arg His Ile Phe Ala Ala Leu Thr Ala Ala
545                 550                 555                 560

Met Ala Gln Pro Gly Asp Val Pro Leu Pro Arg Leu Arg Arg Val Leu
                565                 570                 575

Ala Arg His Asp Ser Ile Cys Gln Arg Gly Arg Ile Leu Pro Leu Ser
            580                 585                 590

Pro Cys Pro Asp Arg Ala Lys Leu Glu Glu Ser Pro Ala Leu Thr Cys
        595                 600                 605

Gln Tyr Thr Val Leu Lys Met Leu Tyr Asp Gly Pro Phe Arg Ala Trp
    610                 615                 620

Leu Ala Gln Gln Asn Ser Thr Ile Leu Asn His Tyr Ile Asp Ser Thr
625                 630                 635                 640

Ile Ala Arg Thr Asp Lys Ala Ala Arg Asp Met Asn Gly Arg Lys Leu
                645                 650                 655

Ala Gln Ala Glu Lys Asp Leu Ile Thr Ser Arg Ala Ala Asp Leu Pro
            660                 665                 670

Arg Leu Ser Val Asp Glu Lys Met Gly Asp Phe Leu Ala Arg Leu Thr
        675                 680                 685

Ala Ala Thr Ala Thr Glu Met Arg Val Gln Arg Gly Tyr Gln Ser Asp
    690                 695                 700

Gly Glu Asn Ala Gln Lys Gln Ala Ala Phe Ile Gly Gln Phe Glu Cys
705                 710                 715                 720

Asp Val Ile Gly Arg Ala Phe Ala Asp Phe Leu Asn Gln Ser Gly Phe
```

725 730 735

Asp Phe Val Leu Lys Leu Lys Ala Asp Thr Pro Gln Pro Asp Ala Ala
740 745 750

Gln Cys Asp Val Thr Ala Leu Ile Ala Pro Asp Asp Ile Ser Val Ser
755 760 765

Pro Pro Gln Ala Trp Gln Gln Val Leu Tyr Phe Ile Leu His Leu Val
770 775 780

Pro Val Asp Asp Ala Ser His Leu Leu His Gln Ile Arg Lys Trp Gln
785 790 795 800

Val Leu Glu Gly Lys Glu Lys Pro Ala Gln Ile Ala His Asp Val Gln
805 810 815

Ser Val Leu Met Leu Tyr Leu Asp Met His Asp Ala Lys Phe Thr Gly
820 825 830

Gly Ala Ala Leu His Gly Ile Glu Lys Phe Ala Glu Phe Phe Ala His
835 840 845

Ala Ala Asp Phe Arg Ala Val Phe Pro Pro Gln Ser Leu Gln Asp Gln
850 855 860

Asp Arg Ser Ile Pro Arg Arg Gly Leu Arg Glu Ile Val Arg Phe Gly
865 870 875 880

His Leu Pro Leu Leu Gln His Met Ser Gly Thr Val Gln Ile Thr His
885 890 895

Asp Asn Val Val Ala Trp Gln Ala Ala Arg Thr Ala Gly Ala Thr Gly
900 905 910

Met Ser Pro Ile Ala Arg Arg Gln Lys Gln Arg Glu Glu Leu His Ala
915 920 925

Leu Ala Val Glu Arg Thr Ala Arg Phe Arg Asn Ala Asp Leu Gln Asn

930 935 940

Tyr Met His Ala Leu Val Asp Val Ile Lys His Arg Gln Leu Ser Ala
945 950 955 960

Gln Val Thr Leu Ser Asp Gln Val Arg Leu His Arg Leu Met Met Gly
965 970 975

Val Leu Gly Arg Leu Val Asp Tyr Ala Gly Leu Trp Glu Arg Asp Leu
980 985 990

Tyr Phe Val Val Leu Ala Leu Leu Tyr His His Gly Ala Thr Pro Asp
995 1000 1005

Asp Val Phe Lys Gly Gln Gly Lys Lys Asn Leu Ala Asp Gly Gln
1010 1015 1020

Val Val Ala Ala Leu Lys Pro Lys Asn Arg Lys Ala Ala Ala Pro
1025 1030 1035

Val Gly Val Phe Asp Asp Leu Asp His Tyr Gly Ile Tyr Gln Asp
1040 1045 1050

Asp Arg Gln Ser Ile Arg Asn Gly Leu Ser His Phe Asn Met Leu
1055 1060 1065

Arg Gly Gly Lys Ala Pro Asp Leu Ser His Trp Val Asn Gln Thr
1070 1075 1080

Arg Ser Leu Val Ala His Asp Arg Lys Leu Lys Asn Ala Val Ala
1085 1090 1095

Lys Ser Val Ile Glu Met Leu Ala Arg Glu Gly Phe Asp Leu Asp
1100 1105 1110

Trp Gly Ile Gln Thr Asp Arg Gly Gln His Ile Leu Ser His Gly
1115 1120 1125

Lys Ile Arg Thr Arg Gln Ala Gln His Phe Gln Lys Ser Arg Leu

1130 1135 1140

His Ile Val Lys Lys Ser Ala Lys Pro Asp Lys Asn Asp Thr Val
1145 1150 1155

Lys Ile Arg Glu Asn Leu His Gly Asp Ala Met Val Glu Arg Val
1160 1165 1170

Val Gln Leu Phe Ala Ala Gln Val Gln Lys Arg Tyr Asp Ile Thr
1175 1180 1185

Val Glu Lys Arg Leu Asp His Leu Phe Leu Lys Pro Gln Asp Gln
1190 1195 1200

Lys Gly Lys Asn Gly Ile His Thr His Asn Gly Trp Ser Lys Thr
1205 1210 1215

Glu Lys Lys Arg Arg Pro Ser Arg Glu Asn Arg Lys Gly Asn His
1220 1225 1230

Glu Asn
1235

<210> 59
<211> 1312
<212> PRT
<213> Pseudobutyrivibrio sp.

<400> 59

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

Glu Ser Asn Gly Ile Ile Gly Leu Leu Tyr Lys Asp Pro Leu Asn Glu
20 25 30

Lys Glu Lys Ile Glu Asp Val Val Asn Gln Arg Ala Asn Ser Thr Lys
35 40 45

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

```
Arg Ala Ser Lys Asp Leu Ala Lys Val Val Asn Lys Ala Ile Gly Asn
65                  70                  75                  80

Leu Lys Gly Asn Lys Lys Phe Asn Lys Lys Glu Gln Ile Thr Lys Gly
                85                  90                  95

Leu Asn Thr Lys Ile Ile Val Glu Glu Leu Lys Asn Val Leu Lys Asp
            100                 105                 110

Glu Lys Lys Leu Ile Val Asn Lys Asp Ile Ile Asp Glu Ala Cys Ser
        115                 120                 125

Arg Leu Leu Lys Thr Ser Phe Arg Thr Ala Lys Thr Lys Gln Ala Val
    130                 135                 140

Lys Met Ile Leu Thr Ala Val Leu Ile Glu Asn Thr Asn Leu Ser Lys
145                 150                 155                 160

Glu Asp Glu Ala Phe Val His Glu Tyr Phe Val Lys Lys Leu Val Asn
                165                 170                 175

Glu Tyr Asn Lys Thr Ser Val Lys Lys Gln Ile Pro Val Ala Leu Ser
            180                 185                 190

Asn Gln Asn Met Val Ile Gln Pro Asn Ser Val Asn Gly Thr Leu Glu
        195                 200                 205

Ile Ser Glu Thr Lys Lys Ser Lys Glu Thr Lys Thr Thr Glu Lys Asp
    210                 215                 220

Ala Phe Arg Ala Phe Leu Arg Asp Tyr Ala Thr Leu Asp Glu Asn Arg
225                 230                 235                 240

Arg His Lys Met Arg Leu Cys Leu Arg Asn Leu Val Asn Leu Tyr Phe
                245                 250                 255

Tyr Gly Glu Thr Ser Val Ser Lys Asp Asp Phe Asp Glu Trp Arg Asp
            260                 265                 270
```

```
His Glu Asp Lys Lys Gln Asn Asp Glu Leu Phe Val Lys Lys Ile Val
        275                 280                 285

Ser Ile Lys Thr Asp Arg Lys Gly Asn Val Lys Glu Val Leu Asp Val
    290                 295                 300

Asp Ala Thr Ile Asp Ala Ile Arg Thr Asn Asn Ile Ala Cys Tyr Arg
305                 310                 315                 320

Arg Ala Leu Ala Tyr Ala Asn Glu Asn Pro Asp Val Phe Phe Ser Asp
                325                 330                 335

Thr Met Leu Asn Lys Phe Trp Ile His His Val Glu Asn Glu Val Glu
            340                 345                 350

Arg Ile Tyr Gly His Ile Asn Asn Asn Thr Gly Asp Tyr Lys Tyr Gln
        355                 360                 365

Leu Gly Tyr Leu Ser Glu Lys Val Trp Lys Gly Ile Ile Asn Tyr Leu
    370                 375                 380

Ser Ile Lys Tyr Ile Ala Glu Gly Lys Ala Val Tyr Asn Tyr Ala Met
385                 390                 395                 400

Asn Ala Leu Ala Lys Asp Asn Asn Ser Asn Ala Phe Gly Lys Leu Asp
                405                 410                 415

Glu Lys Phe Val Asn Gly Ile Thr Ser Phe Glu Tyr Glu Arg Ile Lys
            420                 425                 430

Ala Glu Glu Thr Leu Gln Arg Glu Cys Ala Val Asn Ile Ala Phe Ala
        435                 440                 445

Ala Asn His Leu Ala Asn Ala Thr Val Asp Leu Asn Glu Lys Asp Ser
    450                 455                 460

Asp Phe Leu Leu Leu Lys His Glu Asp Asn Lys Asp Thr Leu Gly Ala
465                 470                 475                 480
```

```
Val Ala Arg Pro Asn Ile Leu Arg Asn Ile Leu Gln Phe Phe Gly Gly
                485                 490                 495

Lys Ser Arg Trp Asn Asp Phe Asp Phe Ser Gly Ile Asp Glu Ile Gln
            500                 505                 510

Leu Leu Asp Asp Leu Arg Lys Met Ile Tyr Ser Leu Arg Asn Ser Ser
        515                 520                 525

Phe His Phe Lys Thr Glu Asn Ile Asp Asn Asp Ser Trp Asn Thr Lys
    530                 535                 540

Leu Ile Gly Asp Met Phe Ala Tyr Asp Phe Asn Met Ala Gly Asn Val
545                 550                 555                 560

Gln Lys Asp Lys Met Tyr Ser Asn Asn Val Pro Met Phe Tyr Ser Thr
                565                 570                 575

Ser Asp Ile Glu Lys Met Leu Asp Arg Leu Tyr Ala Glu Val His Glu
            580                 585                 590

Arg Ala Ser Gln Val Pro Ser Phe Asn Ser Val Phe Val Arg Lys Asn
        595                 600                 605

Phe Pro Asp Tyr Leu Lys Asn Asp Leu Lys Ile Thr Ser Ala Phe Gly
    610                 615                 620

Val Asp Asp Ala Leu Lys Trp Gln Ser Ala Val Tyr Tyr Val Cys Lys
625                 630                 635                 640

Glu Ile Tyr Tyr Asn Asp Phe Leu Gln Asn Pro Glu Thr Phe Thr Met
                645                 650                 655

Leu Lys Asp Tyr Val Gln Cys Leu Pro Ile Asp Ile Asp Lys Ser Met
            660                 665                 670

Asp Gln Lys Leu Lys Ser Glu Arg Asn Ala His Lys Asn Phe Lys Glu
        675                 680                 685
```

```
Ala Phe Ala Thr Tyr Cys Lys Glu Cys Asp Ser Leu Ser Ala Ile Cys
    690                 695                 700

Gln Met Ile Met Thr Glu Tyr Asn Asn Gln Asn Lys Gly Asn Arg Lys
705                 710                 715                 720

Val Ile Ser Ala Arg Thr Lys Asp Gly Asp Lys Leu Ile Tyr Lys His
                725                 730                 735

Tyr Lys Met Ile Leu Phe Glu Ala Leu Lys Asn Val Phe Thr Ile Tyr
            740                 745                 750

Leu Glu Lys Asn Ile Asn Thr Tyr Gly Phe Leu Lys Lys Pro Lys Leu
        755                 760                 765

Ile Asn Asn Val Pro Ala Ile Glu Glu Phe Leu Pro Asn Tyr Asn Gly
    770                 775                 780

Arg Gln Tyr Glu Thr Leu Val Asn Arg Ile Thr Glu Glu Thr Glu Leu
785                 790                 795                 800

Gln Lys Trp Tyr Ile Val Gly Arg Leu Leu Asn Pro Lys Gln Val Asn
                805                 810                 815

Gln Leu Ile Gly Asn Phe Arg Ser Tyr Val Gln Tyr Val Asn Asp Val
            820                 825                 830

Ala Arg Arg Ala Lys Gln Thr Gly Asn Asn Leu Ser Asn Asp Asn Ile
        835                 840                 845

Ala Trp Asp Val Lys Asn Ile Ile Gln Ile Phe Asp Val Cys Thr Lys
    850                 855                 860

Leu Asn Gly Val Thr Ser Asn Ile Leu Glu Asp Tyr Phe Asp Asp Gly
865                 870                 875                 880

Asp Asp Tyr Ala Arg Tyr Leu Lys Asn Phe Val Asp Tyr Thr Asn Lys
                885                 890                 895
```

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

Ile Asp Gly Ile Lys Ile Gly Ile Tyr His Asp Gly Thr Asn Pro Ile
        915                 920                 925

Val Asn Arg Asn Ile Ile Gln Cys Lys Leu Tyr Gly Ala Thr Gly Ile
    930                 935                 940

Ile Ser Asp Leu Thr Lys Asp Gly Ser Ile Leu Ser Val Asp Tyr Glu
945                 950                 955                 960

Ile Ile Lys Lys Tyr Met Gln Met Gln Lys Glu Ile Lys Val Tyr Gln
                965                 970                 975

Gln Lys Gly Ile Cys Lys Thr Lys Glu Glu Gln Gln Asn Leu Lys Lys
            980                 985                 990

Tyr Gln Glu Leu Lys Asn Ile Val  Glu Leu Arg Asn Ile  Ile Asp Tyr
        995                 1000                1005

Ser Glu  Ile Leu Asp Glu Leu  Gln Gly Gln Leu Ile  Asn Trp Gly
    1010                1015                1020

Tyr Leu  Arg Glu Arg Asp Leu  Met Tyr Phe Gln Leu  Gly Phe His
    1025                1030                1035

Tyr Leu  Cys Leu His Asn Glu  Ser Lys Lys Pro Val  Gly Tyr Asn
    1040                1045                1050

Asn Ala  Gly Asp Ile Ser Gly  Ala Val Leu Tyr Gln  Ile Val Ala
    1055                1060                1065

Met Tyr  Thr Asn Gly Leu Ser  Leu Ile Asp Ala Asn  Gly Lys Ser
    1070                1075                1080

Lys Lys  Asn Ala Lys Ala Ser  Ala Gly Ala Lys Val  Gly Ser Phe
    1085                1090                1095
```

```
Cys Ser  Tyr Ser Lys Glu Ile  Arg Gly Val Asp Lys  Asp Thr Lys
    1100                1105                1110

Glu Asp  Asp Asp Pro Ile Tyr  Leu Ala Gly Val Glu  Leu Phe Glu
    1115                1120                1125

Asn Ile  Asn Glu His Gln Gln  Cys Ile Asn Leu Arg  Asn Tyr Ile
    1130                1135                1140

Glu His  Phe His Tyr Tyr Ala  Lys His Asp Arg Ser  Met Leu Asp
    1145                1150                1155

Leu Tyr  Ser Glu Val Phe Asp  Arg Phe Phe Thr Tyr  Asp Met Lys
    1160                1165                1170

Tyr Thr  Lys Asn Val Pro Asn  Met Met Tyr Asn Ile  Leu Leu Gln
    1175                1180                1185

His Leu  Val Val Pro Ala Phe  Glu Phe Gly Ser Ser  Glu Lys Arg
    1190                1195                1200

Leu Asp  Asp Asn Asp Glu Gln  Thr Lys Pro Arg Ala  Met Phe Thr
    1205                1210                1215

Leu Arg  Glu Lys Asn Gly Leu  Ser Ser Glu Gln Phe  Thr Tyr Arg
    1220                1225                1230

Leu Gly  Asp Gly Asn Ser Thr  Val Lys Leu Ser Ala  Arg Gly Asp
    1235                1240                1245

Asp Tyr  Leu Arg Ala Val Ala  Ser Leu Leu Tyr Tyr  Pro Asp Arg
    1250                1255                1260

Ala Pro  Glu Gly Leu Ile Arg  Asp Ala Glu Ala Glu  Asp Lys Phe
    1265                1270                1275

Ala Lys  Ile Asn His Ser Asn  Pro Lys Ser Asp Asn  Arg Asn Asn
    1280                1285                1290
```

```
Arg Gly  Asn Phe Lys Asn Pro  Lys Val Gln Trp Tyr  Asn Asn Lys
    1295                1300                1305


Thr Lys  Arg Lys
    1310


<210>  60
<211>  1343
<212>  PRT
<213>  Butyrivibrio sp.

<400>  60

Met Lys Ile Ser Lys Val Asp His Arg Lys Thr Ala Val Lys Ile Thr
1               5                   10                  15


Asp Asn Lys Gly Ala Glu Gly Phe Ile Tyr Gln Asp Pro Thr Arg Asp
            20                  25                  30


Ser Ser Thr Met Glu Gln Ile Ile Ser Asn Arg Ala Arg Ser Ser Lys
        35                  40                  45


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


Lys Tyr Thr Glu Ser Leu Ile Ile Tyr Val Asn Lys Ala Ile Lys Ser
65                  70                  75                  80


Leu Lys Gly Asp Lys Arg Asn Asn Lys Tyr Glu Glu Ile Thr Glu Ser
                85                  90                  95


Leu Lys Thr Glu Arg Val Leu Asn Ala Leu Ile Gln Ala Gly Asn Glu
            100                 105                 110


Phe Thr Cys Ser Glu Asn Asn Ile Glu Asp Ala Leu Asn Lys Tyr Leu
        115                 120                 125


Lys Lys Ser Phe Arg Val Gly Asn Thr Lys Ser Ala Leu Lys Lys Leu
    130                 135                 140
```

```
Leu Met Ala Ala Tyr Cys Gly Tyr Lys Leu Ser Ile Glu Glu Lys Glu
145                 150                 155                 160


Glu Ile Gln Asn Tyr Phe Val Asp Lys Leu Val Lys Glu Tyr Asn Lys
                165                 170                 175


Asp Thr Val Leu Lys Tyr Thr Ala Lys Ser Leu Lys His Gln Asn Met
            180                 185                 190


Val Val Gln Pro Asp Thr Asp Asn His Val Phe Leu Pro Ser Arg Ile
        195                 200                 205


Ala Gly Ala Thr Gln Asn Lys Met Ser Glu Lys Glu Ala Leu Thr Glu
    210                 215                 220


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


Arg Ile Ile Leu Arg Lys Leu Val Asn Leu Tyr Phe Tyr Glu Ser Pro
                245                 250                 255


Asp Phe Ile Tyr Pro Glu Asn Asn Glu Trp Lys Glu His Asp Asp Arg
            260                 265                 270


Lys Asn Lys Thr Glu Thr Phe Val Ser Pro Val Lys Val Asn Glu Glu
        275                 280                 285


Lys Asn Gly Lys Thr Phe Val Lys Ile Asp Val Pro Ala Thr Lys Asp
    290                 295                 300


Leu Ile Arg Leu Lys Asn Ile Glu Cys Tyr Arg Arg Ser Val Ala Glu
305                 310                 315                 320


Thr Ala Gly Asn Pro Ile Thr Tyr Phe Thr Asp His Asn Ile Ser Lys
                325                 330                 335


Phe Trp Ile His His Ile Glu Asn Glu Val Glu Lys Ile Phe Ala Leu
            340                 345                 350
```

```
Leu Lys Ser Asn Trp Lys Asp Tyr Gln Phe Ser Val Gly Tyr Ile Ser
        355                 360                 365


Glu Lys Val Trp Lys Glu Ile Ile Asn Tyr Leu Ser Ile Lys Tyr Ile
    370                 375                 380


Ala Ile Gly Lys Ala Val Tyr Asn Tyr Ala Leu Glu Asp Ile Lys Lys
385                 390                 395                 400


Asn Asp Gly Thr Leu Asn Phe Gly Val Ile Asp Pro Ser Phe Tyr Asp
                405                 410                 415


Gly Ile Asn Ser Phe Glu Tyr Glu Lys Ile Lys Ala Glu Glu Thr Phe
            420                 425                 430


Gln Arg Glu Val Ala Val Tyr Val Ser Phe Ala Val Asn His Leu Ser
        435                 440                 445


Ser Ala Thr Val Lys Leu Ser Glu Ala Gln Ser Asp Met Leu Val Leu
    450                 455                 460


Asn Lys Asn Asp Ile Glu Lys Ile Ala Tyr Gly Asn Thr Lys Arg Asn
465                 470                 475                 480


Ile Leu Gln Phe Phe Gly Gly Gln Ser Lys Trp Lys Glu Phe Asp Phe
                485                 490                 495


Asp Arg Tyr Ile Asn Pro Val Asn Tyr Thr Asp Ile Asp Phe Leu Phe
            500                 505                 510


Asp Ile Lys Lys Met Val Tyr Ser Leu Arg Asn Glu Ser Phe His Phe
        515                 520                 525


Thr Thr Thr Asp Thr Glu Ser Asp Trp Asn Lys Asn Leu Ile Ser Ala
    530                 535                 540


Met Phe Glu Tyr Glu Cys Arg Arg Ile Ser Thr Val Gln Lys Asn Lys
545                 550                 555                 560
```

```
Phe Phe Ser Asn Asn Leu Pro Leu Phe Tyr Gly Glu Asn Ser Leu Glu
                565                 570                 575

Arg Val Leu His Lys Leu Tyr Asp Asp Tyr Val Asp Arg Met Ser Gln
            580                 585                 590

Val Pro Ser Phe Gly Asn Val Phe Val Arg Lys Lys Phe Pro Asp Tyr
        595                 600                 605

Met Lys Glu Ile Gly Ile Lys His Asn Leu Ser Ser Glu Asp Asn Leu
    610                 615                 620

Lys Leu Gln Gly Ala Leu Tyr Phe Leu Tyr Lys Glu Ile Tyr Tyr Asn
625                 630                 635                 640

Ala Phe Ile Ser Ser Glu Lys Ala Met Lys Ile Phe Val Asp Leu Val
                645                 650                 655

Asn Lys Leu Asp Thr Asn Ala Arg Asp Asp Lys Gly Arg Ile Thr His
            660                 665                 670

Glu Ala Met Ala His Lys Asn Phe Lys Asp Ala Ile Ser His Tyr Met
        675                 680                 685

Thr His Asp Cys Ser Leu Ala Asp Ile Cys Gln Lys Ile Met Thr Glu
    690                 695                 700

Tyr Asn Gln Gln Asn Thr Gly His Arg Lys Lys Gln Thr Thr Tyr Ser
705                 710                 715                 720

Ser Glu Lys Asn Pro Glu Ile Phe Arg His Tyr Lys Met Ile Leu Phe
                725                 730                 735

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

Phe Asp Phe Ile Met Lys Pro Asn Ser Pro Lys Thr Asp Ile Lys Glu
        755                 760                 765
```

```
Glu Glu Phe Leu Pro Gln Tyr Lys Ser Cys Ala Tyr Asp Asn Leu Ile
    770                 775                 780

Lys Leu Ile Ala Asp Asn Val Glu Leu Gln Lys Trp Tyr Ile Thr Ala
785                 790                 795                 800

Arg Leu Leu Ser Pro Arg Glu Val Asn Gln Leu Ile Gly Ser Phe Arg
                805                 810                 815

Ser Tyr Lys Gln Phe Val Ser Asp Ile Glu Arg Arg Ala Lys Glu Thr
            820                 825                 830

Asn Asn Ser Leu Ser Lys Ser Gly Met Thr Val Asp Val Glu Asn Ile
        835                 840                 845

Thr Lys Val Leu Asp Leu Cys Thr Lys Leu Asn Gly Arg Phe Ser Asn
    850                 855                 860

Glu Leu Thr Asp Tyr Phe Asp Ser Lys Asp Asp Tyr Ala Val Tyr Val
865                 870                 875                 880

Ser Lys Phe Leu Asp Phe Gly Phe Lys Ile Asp Glu Lys Phe Pro Ala
                885                 890                 895

Ala Leu Leu Gly Glu Phe Cys Asn Lys Glu Glu Asn Gly Lys Lys Ile
            900                 905                 910

Gly Ile Tyr His Asn Gly Thr Glu Pro Ile Leu Asn Ser Asn Ile Ile
        915                 920                 925

Lys Ser Lys Leu Tyr Gly Ile Thr Asp Val Val Ser Arg Ala Val Lys
    930                 935                 940

Pro Val Ser Glu Lys Leu Ile Arg Glu Tyr Leu Gln Gln Glu Val Lys
945                 950                 955                 960

Ile Lys Pro Tyr Leu Glu Asn Gly Val Cys Lys Asn Lys Glu Glu Gln
                965                 970                 975
```

```
Ala Ala Leu Arg Lys Tyr Gln Glu Leu Lys Asn Arg Ile Glu Phe Arg
            980                 985                 990


Asp Ile Val Glu Tyr Ser Glu Ile  Ile Asn Glu Leu Met  Gly Gln Leu
        995                 1000                1005


Ile Asn  Phe Ser Tyr Leu Arg  Glu Arg Asp Leu Met  Tyr Phe Gln
    1010                1015                1020


Leu Gly  Phe His Tyr Leu Cys  Leu Asn Asn Tyr Gly  Ala Lys Pro
    1025                1030                1035


Glu Gly  Tyr Tyr Ser Ile Val  Asn Asp Lys Arg Thr  Ile Lys Gly
    1040                1045                1050


Ala Ile  Leu Tyr Gln Ile Val  Ala Met Tyr Thr Tyr  Gly Leu Pro
    1055                1060                1065


Ile Tyr  His Tyr Val Asp Gly  Thr Ile Ser Asp Arg  Arg Lys Asn
    1070                1075                1080


Lys Lys  Thr Val Leu Asp Thr  Leu Asn Ser Ser Glu  Thr Val Gly
    1085                1090                1095


Ala Lys  Ile Lys Tyr Phe Ile  Tyr Tyr Ser Asp Glu  Leu Phe Asn
    1100                1105                1110


Asp Ser  Leu Ile Leu Tyr Asn  Ala Gly Leu Glu Leu  Phe Glu Asn
    1115                1120                1125


Ile Asn  Glu His Glu Asn Ile  Val Asn Leu Arg Lys  Tyr Ile Asp
    1130                1135                1140


His Phe  Lys Tyr Tyr Val Ser  Gln Asp Arg Ser Leu  Leu Asp Ile
    1145                1150                1155


Tyr Ser  Glu Val Phe Asp Arg  Tyr Phe Thr Tyr Asp  Arg Lys Tyr
    1160                1165                1170
```

```
Lys Lys  Asn Val Met Asn Leu  Phe Ser Asn Ile Met  Leu Lys His
    1175                1180                1185


Phe Ile  Ile Thr Asp Phe Glu  Phe Ser Thr Gly Glu  Lys Thr Ile
    1190                1195                1200


Gly Glu  Lys Asn Thr Ala Lys  Lys Glu Cys Ala Lys  Val Arg Ile
    1205                1210                1215


Lys Arg  Gly Gly Leu Ser Ser  Asp Lys Phe Thr Tyr  Lys Phe Lys
    1220                1225                1230


Asp Ala  Lys Pro Ile Glu Leu  Ser Ala Lys Asn Thr  Glu Phe Leu
    1235                1240                1245


Asp Gly  Val Ala Arg Ile Leu  Tyr Tyr Pro Glu Asn  Val Val Leu
    1250                1255                1260


Thr Asp  Leu Val Arg Asn Ser  Glu Val Glu Asp Glu  Lys Arg Ile
    1265                1270                1275


Glu Lys  Tyr Asp Arg Asn His  Asn Ser Ser Pro Thr  Arg Lys Asp
    1280                1285                1290


Lys Thr  Tyr Lys Gln Asp Val  Lys Lys Asn Tyr Asn  Lys Lys Thr
    1295                1300                1305


Ser Lys  Ala Phe Asp Ser Ser  Lys Leu Asp Thr Lys  Ser Val Gly
    1310                1315                1320


Asn Asn  Leu Ser Asp Asn Pro  Val Leu Lys Gln Phe  Leu Ser Glu
    1325                1330                1335


Ser Lys  Lys Lys Arg
    1340


<210>  61
<211>  1276
<212>  PRT
```

<213> Blautia sp.

<400> 61

```
Met Lys Ile Ser Lys Val Asp His Val Lys Ser Gly Ile Asp Gln Lys
1               5                  10                  15


Leu Ser Ser Gln Arg Gly Met Leu Tyr Lys Gln Pro Gln Lys Lys Tyr
            20                  25                  30


Glu Gly Lys Gln Leu Glu Glu His Val Arg Asn Leu Ser Arg Lys Ala
        35                  40                  45


Lys Ala Leu Tyr Gln Val Phe Pro Val Ser Gly Asn Ser Lys Met Glu
    50                  55                  60


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


Leu Asp Ser Gly Lys Thr Ser Glu Glu Ile Val Gly Tyr Ile Asn Thr
                85                  90                  95


Tyr Ser Val Ala Ser Gln Ile Ser Gly Asp His Ile Gln Glu Leu Val
            100                 105                 110


Asp Gln His Leu Lys Glu Ser Leu Arg Lys Tyr Thr Cys Val Gly Asp
        115                 120                 125


Lys Arg Ile Tyr Val Pro Asp Ile Ile Val Ala Leu Leu Lys Ser Lys
    130                 135                 140


Phe Asn Ser Glu Thr Leu Gln Tyr Asp Asn Ser Glu Leu Lys Ile Leu
145                 150                 155                 160


Ile Asp Phe Ile Arg Glu Asp Tyr Leu Lys Glu Lys Gln Ile Lys Gln
                165                 170                 175


Ile Val His Ser Ile Glu Asn Asn Ser Thr Pro Leu Arg Ile Ala Glu
            180                 185                 190
```

```
Ile Asn Gly Gln Lys Arg Leu Ile Pro Ala Asn Val Asp Asn Pro Lys
        195                 200                 205

Lys Ser Tyr Ile Phe Glu Phe Leu Lys Glu Tyr Ala Gln Ser Asp Pro
    210                 215                 220

Lys Gly Gln Glu Ser Leu Leu Gln His Met Arg Tyr Leu Ile Leu Leu
225                 230                 235                 240

Tyr Leu Tyr Gly Pro Asp Lys Ile Thr Asp Asp Tyr Cys Glu Glu Ile
                245                 250                 255

Glu Ala Trp Asn Phe Gly Ser Ile Val Met Asp Asn Glu Gln Leu Phe
            260                 265                 270

Ser Glu Glu Ala Ser Met Leu Ile Gln Asp Arg Ile Tyr Val Asn Gln
        275                 280                 285

Gln Ile Glu Glu Gly Arg Gln Ser Lys Asp Thr Ala Lys Val Lys Lys
    290                 295                 300

Asn Lys Ser Lys Tyr Arg Met Leu Gly Asp Lys Ile Glu His Ser Ile
305                 310                 315                 320

Asn Glu Ser Val Val Lys His Tyr Gln Glu Ala Cys Lys Ala Val Glu
                325                 330                 335

Glu Lys Asp Ile Pro Trp Ile Lys Tyr Ile Ser Asp His Val Met Ser
            340                 345                 350

Val Tyr Ser Ser Lys Asn Arg Val Asp Leu Asp Lys Leu Ser Leu Pro
        355                 360                 365

Tyr Leu Ala Lys Asn Thr Trp Asn Thr Trp Ile Ser Phe Ile Ala Met
    370                 375                 380

Lys Tyr Val Asp Met Gly Lys Gly Val Tyr His Phe Ala Met Ser Asp
385                 390                 395                 400
```

```
Val Asp Lys Val Gly Lys Gln Asp Asn Leu Ile Ile Gly Gln Ile Asp
                405                 410                 415

Pro Lys Phe Ser Asp Gly Ile Ser Ser Phe Asp Tyr Glu Arg Ile Lys
            420                 425                 430

Ala Glu Asp Asp Leu His Arg Ser Met Ser Gly Tyr Ile Ala Phe Ala
        435                 440                 445

Val Asn Asn Phe Ala Arg Ala Ile Cys Ser Asp Glu Phe Arg Lys Lys
    450                 455                 460

Asn Arg Lys Glu Asp Val Leu Thr Val Gly Leu Asp Glu Ile Pro Leu
465                 470                 475                 480

Tyr Asp Asn Val Lys Arg Lys Leu Leu Gln Tyr Phe Gly Gly Ala Ser
                485                 490                 495

Asn Trp Asp Asp Ser Ile Ile Asp Ile Ile Asp Asp Lys Asp Leu Val
            500                 505                 510

Ala Cys Ile Lys Glu Asn Leu Tyr Val Ala Arg Asn Val Asn Phe His
        515                 520                 525

Phe Ala Gly Ser Glu Lys Val Gln Lys Lys Gln Asp Asp Ile Leu Glu
    530                 535                 540

Glu Ile Val Arg Lys Glu Thr Arg Asp Ile Gly Lys His Tyr Arg Lys
545                 550                 555                 560

Val Phe Tyr Ser Asn Asn Val Ala Val Phe Tyr Cys Asp Glu Asp Ile
                565                 570                 575

Ile Lys Leu Met Asn His Leu Tyr Gln Arg Glu Lys Pro Tyr Gln Ala
            580                 585                 590

Gln Ile Pro Ser Tyr Asn Lys Val Ile Ser Lys Thr Tyr Leu Pro Asp
        595                 600                 605
```

```
Leu Ile Phe Met Leu Leu Lys Gly Lys Asn Arg Thr Lys Ile Ser Asp
    610                 615                 620

Pro Ser Ile Met Asn Met Phe Arg Gly Thr Phe Tyr Phe Leu Leu Lys
625                 630                 635                 640

Glu Ile Tyr Tyr Asn Asp Phe Leu Gln Ala Ser Asn Leu Lys Glu Met
                645                 650                 655

Phe Cys Glu Gly Leu Lys Asn Asn Val Lys Asn Lys Lys Ser Glu Lys
            660                 665                 670

Pro Tyr Gln Asn Phe Met Arg Arg Phe Glu Glu Leu Glu Asn Met Gly
        675                 680                 685

Met Asp Phe Gly Glu Ile Cys Gln Gln Ile Met Thr Asp Tyr Glu Gln
    690                 695                 700

Gln Asn Lys Gln Lys Lys Lys Thr Ala Thr Ala Val Met Ser Glu Lys
705                 710                 715                 720

Asp Lys Lys Ile Arg Thr Leu Asp Asn Asp Thr Gln Lys Tyr Lys His
                725                 730                 735

Phe Arg Thr Leu Leu Tyr Ile Gly Leu Arg Glu Ala Phe Ile Ile Tyr
            740                 745                 750

Leu Lys Asp Glu Lys Asn Lys Glu Trp Tyr Glu Phe Leu Arg Glu Pro
        755                 760                 765

Val Lys Arg Glu Gln Pro Glu Glu Lys Glu Phe Val Asn Lys Trp Lys
    770                 775                 780

Leu Asn Gln Tyr Ser Asp Cys Ser Glu Leu Ile Leu Lys Asp Ser Leu
785                 790                 795                 800

Ala Ala Ala Trp Tyr Val Val Ala His Phe Ile Asn Gln Ala Gln Leu
                805                 810                 815
```

```
Asn His Leu Ile Gly Asp Ile Lys Asn Tyr Ile Gln Phe Ile Ser Asp
            820                 825                 830

Ile Asp Arg Arg Ala Lys Ser Thr Gly Asn Pro Val Ser Glu Ser Thr
        835                 840                 845

Glu Ile Gln Ile Glu Arg Tyr Arg Lys Ile Leu Arg Val Leu Glu Phe
    850                 855                 860

Ala Lys Phe Phe Cys Gly Gln Ile Thr Asn Val Leu Thr Asp Tyr Tyr
865                 870                 875                 880

Gln Asp Glu Asn Asp Phe Ser Thr His Val Gly His Tyr Val Lys Phe
                885                 890                 895

Glu Lys Lys Asn Met Glu Pro Ala His Ala Leu Gln Ala Phe Ser Asn
            900                 905                 910

Ser Leu Tyr Ala Cys Gly Lys Glu Lys Lys Lys Ala Gly Phe Tyr Tyr
        915                 920                 925

Asp Gly Met Asn Pro Ile Val Asn Arg Asn Ile Thr Leu Ala Ser Met
    930                 935                 940

Tyr Gly Asn Lys Lys Leu Leu Glu Asn Ala Met Asn Pro Val Thr Glu
945                 950                 955                 960

Gln Asp Ile Arg Lys Tyr Tyr Ser Leu Met Ala Glu Leu Asp Ser Val
                965                 970                 975

Leu Lys Asn Gly Ala Val Cys Lys Ser Glu Asp Glu Gln Lys Asn Leu
            980                 985                 990

Arg His Phe Gln Asn Leu Lys Asn  Arg Ile Glu Leu Val  Asp Val Leu
        995                 1000                1005

Thr Leu  Ser Glu Leu Val Asn  Asp Leu Val Ala Gln  Leu Ile Gly
    1010                1015                1020
```

```
Trp Val  Tyr Ile Arg Glu Arg  Asp Met Met Tyr Leu  Gln Leu Gly
    1025                1030                1035

Leu His  Tyr Ile Lys Leu Tyr  Phe Thr Asp Ser Val  Ala Glu Asp
    1040                1045                1050

Ser Tyr  Leu Arg Thr Leu Asp  Leu Glu Glu Gly Ser  Ile Ala Asp
    1055                1060                1065

Gly Ala  Val Leu Tyr Gln Ile  Ala Ser Leu Tyr Ser  Phe Asn Leu
    1070                1075                1080

Pro Met  Tyr Val Lys Pro Asn  Lys Ser Ser Val Tyr  Cys Lys Lys
    1085                1090                1095

His Val  Asn Ser Val Ala Thr  Lys Phe Asp Ile Phe  Glu Lys Glu
    1100                1105                1110

Tyr Cys  Asn Gly Asp Glu Thr  Val Ile Glu Asn Gly  Leu Arg Leu
    1115                1120                1125

Phe Glu  Asn Ile Asn Leu His  Lys Asp Met Val Lys  Phe Arg Asp
    1130                1135                1140

Tyr Leu  Ala His Phe Lys Tyr  Phe Ala Lys Leu Asp  Glu Ser Ile
    1145                1150                1155

Leu Glu  Leu Tyr Ser Lys Ala  Tyr Asp Phe Phe Phe  Ser Tyr Asn
    1160                1165                1170

Ile Lys  Leu Lys Lys Ser Val  Ser Tyr Val Leu Thr  Asn Val Leu
    1175                1180                1185

Leu Ser  Tyr Phe Ile Asn Ala  Lys Leu Ser Phe Ser  Thr Tyr Lys
    1190                1195                1200

Ser Ser  Gly Asn Lys Thr Val  Gln His Arg Thr Thr  Lys Ile Ser
    1205                1210                1215
```

```
Val Val  Ala Gln Thr Asp Tyr  Phe Thr Tyr Lys Leu  Arg Ser Ile
    1220                1225                1230


Val Lys  Asn Lys Asn Gly Val  Glu Ser Ile Glu Asn  Asp Asp Arg
    1235                1240                1245


Arg Cys  Glu Val Val Asn Ile  Ala Ala Arg Asp Lys  Glu Phe Val
    1250                1255                1260


Asp Glu  Val Cys Asn Val Ile  Asn Tyr Asn Ser Asp  Lys
    1265                1270                1275


<210>  62
<211>  1168
<212>  PRT
<213>  Leptotrichia marseille

<400>  62

Met Lys Ile Thr Lys Ile Asp Gly Ile Ser His Lys Lys Tyr Ile Lys
1               5                   10                  15


Glu Gly Lys Leu Val Lys Ser Thr Ser Glu Glu Asn Lys Thr Asp Glu
            20                  25                  30


Arg Leu Ser Glu Leu Leu Thr Ile Arg Leu Asp Thr Tyr Ile Lys Asn
        35                  40                  45


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


Lys Glu Phe Phe Ser Asn Lys Val Leu Tyr Leu Lys Asp Gly Ile Leu
65                  70                  75                  80


Tyr Leu Lys Asp Arg Arg Glu Lys Asn Gln Leu Gln Asn Lys Asn Tyr
                85                  90                  95


Ser Glu Glu Asp Ile Ser Glu Tyr Asp Leu Lys Asn Lys Asn Asn Phe
            100                 105                 110


Leu Val Leu Lys Lys Ile Leu Leu Asn Glu Asp Ile Asn Ser Glu Glu
```

```
        115                 120                 125

Leu Glu Ile Phe Arg Asn Asp Phe Glu Lys Lys Leu Asp Lys Ile Asn
    130                 135                 140

Ser Leu Lys Tyr Ser Leu Glu Glu Asn Lys Ala Asn Tyr Gln Lys Ile
145                 150                 155                 160

Asn Glu Asn Asn Ile Lys Lys Val Glu Gly Lys Ser Lys Arg Asn Ile
                165                 170                 175

Phe Tyr Asn Tyr Tyr Lys Asp Ser Ala Lys Arg Asn Asp Tyr Ile Asn
            180                 185                 190

Asn Ile Gln Glu Ala Phe Asp Lys Leu Tyr Lys Lys Glu Asp Ile Glu
        195                 200                 205

Asn Leu Phe Phe Leu Ile Glu Asn Ser Lys Lys His Glu Lys Tyr Lys
    210                 215                 220

Ile Arg Glu Cys Tyr His Lys Ile Ile Gly Arg Lys Asn Asp Lys Glu
225                 230                 235                 240

Asn Phe Ala Thr Ile Ile Tyr Glu Glu Ile Gln Asn Val Asn Asn Met
                245                 250                 255

Lys Glu Leu Ile Glu Lys Val Pro Asn Val Ser Glu Leu Lys Lys Ser
            260                 265                 270

Gln Val Phe Tyr Lys Tyr Tyr Leu Asn Lys Glu Lys Leu Asn Asp Glu
        275                 280                 285

Asn Ile Lys Tyr Val Phe Cys His Phe Val Glu Ile Glu Met Ser Lys
    290                 295                 300

Leu Leu Lys Asn Tyr Val Tyr Lys Lys Pro Ser Asn Ile Ser Asn Asp
305                 310                 315                 320

Lys Val Lys Arg Ile Phe Glu Tyr Gln Ser Leu Lys Lys Leu Ile Glu
```

```
                325                 330                 335

Asn Lys Leu Leu Asn Lys Leu Asp Thr Tyr Val Arg Asn Cys Gly Lys
            340                 345                 350

Tyr Ser Phe Tyr Leu Gln Asp Gly Glu Ile Ala Thr Ser Asp Phe Ile
        355                 360                 365

Val Gly Asn Arg Gln Asn Glu Ala Phe Leu Arg Asn Ile Ile Gly Val
    370                 375                 380

Ser Ser Thr Ala Tyr Phe Ser Leu Arg Asn Ile Leu Glu Thr Glu Asn
385                 390                 395                 400

Glu Asn Asp Ile Thr Gly Arg Met Arg Gly Lys Thr Val Lys Asn Asn
                405                 410                 415

Lys Gly Glu Glu Lys Tyr Ile Ser Gly Glu Ile Asp Lys Leu Tyr Asp
            420                 425                 430

Asn Asn Lys Gln Asn Glu Val Lys Lys Asn Leu Lys Met Phe Tyr Ser
        435                 440                 445

Tyr Asp Phe Asn Met Asn Ser Lys Lys Glu Ile Glu Asp Phe Phe Ser
    450                 455                 460

Asn Ile Asp Glu Ala Ile Ser Ser Ile Arg His Gly Ile Val His Phe
465                 470                 475                 480

Asn Leu Glu Leu Glu Gly Lys Asp Ile Phe Thr Phe Lys Asn Ile Val
                485                 490                 495

Pro Ser Gln Ile Ser Lys Lys Met Phe His Asp Glu Ile Asn Glu Lys
            500                 505                 510

Lys Leu Lys Leu Lys Ile Phe Lys Gln Leu Asn Ser Ala Asn Val Phe
        515                 520                 525

Arg Tyr Leu Glu Lys Tyr Lys Ile Leu Asn Tyr Leu Asn Arg Thr Arg
```

```
    530                 535                 540

Phe Glu Phe Val Asn Lys Asn Ile Pro Phe Val Pro Ser Phe Thr Lys
545                 550                 555                 560

Leu Tyr Ser Arg Ile Asp Asp Leu Lys Asn Ser Leu Gly Ile Tyr Trp
                565                 570                 575

Lys Thr Pro Lys Thr Asn Asp Asp Asn Lys Thr Lys Glu Ile Thr Asp
            580                 585                 590

Ala Gln Ile Tyr Leu Leu Lys Asn Ile Tyr Tyr Gly Glu Phe Leu Asn
        595                 600                 605

Tyr Phe Met Ser Asn Asn Gly Asn Phe Phe Glu Ile Thr Lys Glu Ile
    610                 615                 620

Ile Glu Leu Asn Lys Asn Asp Lys Arg Asn Leu Lys Thr Gly Phe Tyr
625                 630                 635                 640

Lys Leu Gln Lys Phe Glu Asn Leu Gln Glu Lys Thr Pro Lys Glu Tyr
                645                 650                 655

Leu Ala Asn Ile Gln Ser Leu Tyr Met Ile Asn Ala Gly Asn Gln Asp
            660                 665                 670

Glu Glu Glu Lys Asp Thr Tyr Ile Asp Phe Ile Gln Lys Ile Phe Leu
        675                 680                 685

Lys Gly Phe Met Thr Tyr Leu Ala Asn Asn Gly Arg Leu Ser Leu Ile
    690                 695                 700

Tyr Ile Gly Ser Asp Glu Glu Thr Asn Thr Ser Leu Ala Glu Lys Lys
705                 710                 715                 720

Gln Glu Phe Asp Lys Phe Leu Lys Lys Tyr Glu Gln Asn Asn Asn Ile
                725                 730                 735

Glu Ile Pro Tyr Glu Ile Asn Glu Phe Val Arg Glu Ile Lys Leu Gly
```

740 745 750

Lys Ile Leu Lys Tyr Thr Glu Arg Leu Asn Met Phe Tyr Leu Ile Leu
755 760 765

Lys Leu Leu Asn His Lys Glu Leu Thr Asn Leu Lys Gly Ser Leu Glu
770 775 780

Lys Tyr Gln Ser Ala Asn Lys Glu Glu Ala Phe Ser Asp Gln Leu Glu
785 790 795 800

Leu Ile Asn Leu Leu Asn Leu Asp Asn Asn Arg Val Thr Glu Asp Phe
805 810 815

Glu Leu Glu Ala Asp Glu Ile Gly Lys Phe Leu Asp Phe Asn Gly Asn
820 825 830

Lys Val Lys Asp Asn Lys Glu Leu Lys Lys Phe Asp Thr Asn Lys Ile
835 840 845

Tyr Phe Asp Gly Glu Asn Ile Ile Lys His Arg Ala Phe Tyr Asn Ile
850 855 860

Lys Lys Tyr Gly Met Leu Asn Leu Leu Glu Lys Ile Ser Asp Glu Ala
865 870 875 880

Lys Tyr Lys Ile Ser Ile Glu Glu Leu Lys Asn Tyr Ser Lys Lys Lys
885 890 895

Asn Glu Ile Glu Glu Asn His Thr Thr Gln Glu Asn Leu His Arg Lys
900 905 910

Tyr Ala Arg Pro Arg Lys Asp Glu Lys Phe Thr Asp Glu Asp Tyr Lys
915 920 925

Lys Tyr Glu Lys Ala Ile Arg Asn Ile Gln Gln Tyr Thr His Leu Lys
930 935 940

Asn Lys Val Glu Phe Asn Glu Leu Asn Leu Leu Gln Ser Leu Leu Leu

```
945                 950                 955                 960

Arg Ile Leu His Arg Leu Val Gly Tyr Thr Ser Ile Trp Glu Arg Asp
                965                 970                 975

Leu Arg Phe Arg Leu Lys Gly Glu Phe Pro Glu Asn Gln Tyr Ile Glu
            980                 985                 990

Glu Ile Phe Asn Phe Asp Asn Ser  Lys Asn Val Lys Tyr  Lys Asn Gly
        995                 1000                1005

Gln Ile  Val Glu Lys Tyr Ile  Asn Phe Tyr Lys Glu  Leu Tyr Lys
    1010                1015                1020

Asp Asp  Thr Glu Lys Ile Ser  Ile Tyr Ser Asp Lys  Lys Val Lys
    1025                1030                1035

Glu Leu  Lys Lys Glu Lys Lys  Asp Leu Tyr Ile Arg  Asn Tyr Ile
    1040                1045                1050

Ala His  Phe Asn Tyr Ile Pro  Asn Ala Glu Ile Ser  Leu Leu Glu
    1055                1060                1065

Met Leu  Glu Asn Leu Arg Lys  Leu Leu Ser Tyr Asp  Arg Lys Leu
    1070                1075                1080

Lys Asn  Ala Ile Met Lys Ser  Ile Val Asp Ile Leu  Lys Glu Tyr
    1085                1090                1095

Gly Phe  Val Val Thr Phe Lys  Ile Glu Lys Asp Lys  Lys Ile Arg
    1100                1105                1110

Ile Glu  Ser Leu Lys Ser Glu  Glu Val Val His Leu  Lys Lys Leu
    1115                1120                1125

Lys Leu  Lys Asp Asn Asp Lys  Lys Lys Glu Pro Ile  Lys Thr Tyr
    1130                1135                1140

Arg Asn  Ser Lys Glu Leu Cys  Lys Leu Val Lys Val  Met Phe Glu
```

1145 1150 1155

Tyr Lys Met Lys Glu Lys Lys Ser Glu Asn
1160 1165

<210> 63
<211> 1138
<212> PRT
<213> Bacterioides ihuae

<400> 63

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

Lys Met Val Leu Ile His Arg Lys Val Gly Glu Gly Thr Leu Val Leu
20 25 30

Asp Glu Asn Leu Ala Asp Leu Thr Ala Pro Ile Ile Asp Lys Tyr Lys
35 40 45

Asp Lys Ser Phe Glu Leu Ser Leu Leu Lys Gln Thr Leu Val Ser Glu
50 55 60

Lys Glu Met Asn Ile Pro Lys Cys Asp Lys Cys Thr Ala Lys Glu Arg
65 70 75 80

Cys Leu Ser Cys Lys Gln Arg Glu Lys Arg Leu Lys Glu Val Arg Gly
85 90 95

Ala Ile Glu Lys Thr Ile Gly Ala Val Ile Ala Gly Arg Asp Ile Ile
100 105 110

Pro Arg Leu Asn Ile Phe Asn Glu Asp Glu Ile Cys Trp Leu Ile Lys
115 120 125

Pro Lys Leu Arg Asn Glu Phe Thr Phe Lys Asp Val Asn Lys Gln Val
130 135 140

Val Lys Leu Asn Leu Pro Lys Val Leu Val Glu Tyr Ser Lys Lys Asn
145 150 155 160

```
Asp Pro Thr Leu Phe Leu Ala Tyr Gln Gln Trp Ile Ala Ala Tyr Leu
                165                 170                 175

Lys Asn Lys Lys Gly His Ile Lys Lys Ser Ile Leu Asn Asn Arg Val
            180                 185                 190

Val Ile Asp Tyr Ser Asp Glu Ser Lys Leu Ser Lys Arg Lys Gln Ala
        195                 200                 205

Leu Glu Leu Trp Gly Glu Glu Tyr Glu Thr Asn Gln Arg Ile Ala Leu
    210                 215                 220

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

Pro Asn Pro Glu Glu Tyr Val Ser Asp Lys Gly Glu Ile Arg Pro Ala
                245                 250                 255

Phe His Tyr Lys Leu Lys Asn Val Leu Gln Met His Gln Ser Thr Val
            260                 265                 270

Phe Gly Thr Asn Glu Ile Leu Cys Ile Asn Pro Ile Phe Asn Glu Asn
        275                 280                 285

Arg Ala Asn Ile Gln Leu Ser Ala Tyr Asn Leu Glu Val Val Lys Tyr
    290                 295                 300

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

Asn Gln Ala Ile Tyr Tyr Leu Lys Val Glu Thr Leu Lys Glu Arg Leu
                325                 330                 335

Ser Leu Gln Leu Glu Asn Ala Leu Arg Met Asn Leu Leu Gln Lys Gly
            340                 345                 350

Lys Ile Lys Lys His Glu Phe Asp Lys Asn Thr Cys Ser Asn Thr Leu
        355                 360                 365
```

```
Ser Gln Ile Lys Arg Asp Glu Phe Phe Val Leu Asn Leu Val Glu Met
    370                 375                 380

Cys Ala Phe Ala Ala Asn Asn Ile Arg Asn Ile Val Asp Lys Glu Gln
385                 390                 395                 400

Val Asn Glu Ile Leu Ser Lys Lys Asp Leu Cys Asn Ser Leu Ser Lys
                405                 410                 415

Asn Thr Ile Asp Lys Glu Leu Cys Thr Lys Phe Tyr Gly Ala Asp Phe
            420                 425                 430

Ser Gln Ile Pro Val Ala Ile Trp Ala Met Arg Gly Ser Val Gln Gln
        435                 440                 445

Ile Arg Asn Glu Ile Val His Tyr Lys Ala Glu Ala Ile Asp Lys Ile
    450                 455                 460

Phe Ala Leu Lys Thr Phe Glu Tyr Asp Asp Met Glu Lys Asp Tyr Ser
465                 470                 475                 480

Asp Thr Pro Phe Lys Gln Tyr Leu Glu Leu Ser Ile Glu Lys Ile Asp
                485                 490                 495

Ser Phe Phe Ile Glu Gln Leu Ser Ser Asn Asp Val Leu Asn Tyr Tyr
            500                 505                 510

Cys Thr Glu Asp Val Asn Lys Leu Leu Asn Lys Cys Lys Leu Ser Leu
        515                 520                 525

Arg Arg Thr Ser Ile Pro Phe Ala Pro Gly Phe Lys Thr Ile Tyr Glu
    530                 535                 540

Leu Gly Cys His Leu Gln Asp Ser Ser Asn Thr Tyr Arg Ile Gly His
545                 550                 555                 560

Tyr Leu Met Leu Ile Gly Gly Arg Val Ala Asn Ser Thr Val Thr Lys
                565                 570                 575
```

```
Ala Ser Lys Ala Tyr Pro Ala Tyr Arg Phe Met Leu Lys Leu Ile Tyr
            580                 585                 590

Asn His Leu Phe Leu Asn Lys Phe Leu Asp Asn His Asn Lys Arg Phe
        595                 600                 605

Phe Met Lys Ala Val Ala Phe Val Leu Lys Asp Asn Arg Glu Asn Ala
    610                 615                 620

Arg Asn Lys Phe Gln Tyr Ala Phe Lys Glu Ile Arg Met Met Asn Asn
625                 630                 635                 640

Asp Glu Ser Ile Ala Ser Tyr Met Ser Tyr Ile His Ser Leu Ser Val
                645                 650                 655

Gln Glu Gln Glu Lys Lys Gly Asp Lys Asn Asp Lys Val Arg Tyr Asn
            660                 665                 670

Thr Glu Lys Phe Ile Glu Lys Val Phe Val Lys Gly Phe Asp Asp Phe
        675                 680                 685

Leu Ser Trp Leu Gly Val Glu Phe Ile Leu Ser Pro Asn Gln Glu Glu
    690                 695                 700

Arg Asp Lys Thr Val Thr Arg Glu Glu Tyr Glu Asn Leu Met Ile Lys
705                 710                 715                 720

Asp Arg Val Glu His Ser Ile Asn Ser Asn Gln Glu Ser His Ile Ala
                725                 730                 735

Phe Phe Thr Phe Cys Lys Leu Leu Asp Ala Asn His Leu Ser Asp Leu
            740                 745                 750

Arg Asn Glu Trp Ile Lys Phe Arg Ser Ser Gly Asp Lys Glu Gly Phe
        755                 760                 765

Ser Tyr Asn Phe Ala Ile Asp Ile Ile Glu Leu Cys Leu Leu Thr Val
    770                 775                 780
```

```
Asp Arg Val Glu Gln Arg Arg Asp Gly Tyr Lys Glu Gln Thr Glu Leu
785                 790                 795                 800

Lys Glu Tyr Leu Ser Phe Phe Ile Lys Gly Asn Glu Ser Glu Asn Thr
                805                 810                 815

Val Trp Lys Gly Phe Tyr Phe Gln Gln Asp Asn Tyr Thr Pro Val Leu
            820                 825                 830

Tyr Ser Pro Ile Glu Leu Ile Arg Lys Tyr Gly Thr Leu Glu Leu Leu
        835                 840                 845

Lys Leu Ile Ile Val Asp Glu Asp Lys Ile Thr Gln Gly Glu Phe Glu
    850                 855                 860

Glu Trp Gln Thr Leu Lys Lys Val Val Glu Asp Lys Val Thr Arg Arg
865                 870                 875                 880

Asn Glu Leu His Gln Glu Trp Glu Asp Met Lys Asn Lys Ser Ser Phe
                885                 890                 895

Ser Gln Glu Lys Cys Ser Ile Tyr Gln Lys Leu Cys Arg Asp Ile Asp
            900                 905                 910

Arg Tyr Asn Trp Leu Asp Asn Lys Leu His Leu Val His Leu Arg Lys
        915                 920                 925

Leu His Asn Leu Val Ile Gln Ile Leu Ser Arg Met Ala Arg Phe Ile
    930                 935                 940

Ala Leu Trp Asp Arg Asp Phe Val Leu Leu Asp Ala Ser Arg Ala Asn
945                 950                 955                 960

Asp Asp Tyr Lys Leu Leu Ser Phe Phe Asn Phe Arg Asp Phe Ile Asn
                965                 970                 975

Ala Lys Lys Thr Lys Thr Asp Asp Glu Leu Leu Ala Glu Phe Gly Ser
            980                 985                 990
```

```
Lys Ile Glu Lys Lys Asn Ala Pro  Phe Ile Lys Ala Glu  Asp Val Pro
        995                 1000                1005


Leu Met  Val Glu Cys Ile Glu  Ala Lys Arg Ser Phe  Tyr Gln Lys
    1010                1015                1020


Val Phe  Phe Arg Asn Asn Leu  Gln Val Leu Ala Asp  Arg Asn Phe
    1025                1030                1035


Ile Ala  His Tyr Asn Tyr Ile  Ser Lys Thr Ala Lys  Cys Ser Leu
    1040                1045                1050


Phe Glu  Met Ile Ile Lys Leu  Arg Thr Leu Met Tyr  Tyr Asp Arg
    1055                1060                1065


Lys Leu  Arg Asn Ala Val Val  Lys Ser Ile Ala Asn  Val Phe Asp
    1070                1075                1080


Gln Asn  Gly Met Val Leu Gln  Leu Ser Leu Asp Asp  Ser His Glu
    1085                1090                1095


Leu Lys  Val Asp Lys Val Ile  Ser Lys Arg Ile Val  His Leu Lys
    1100                1105                1110


Asn Asn  Asn Ile Met Thr Asp  Gln Val Pro Glu Glu  Tyr Tyr Lys
    1115                1120                1125


Ile Cys  Arg Arg Leu Leu Glu  Met Lys Lys
    1130                1135


<210>  64
<211>  697
<212>  PRT
<213>  Porphyromonadaceae bacterium

<400>  64

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

```
Glu Lys Ala Pro Leu Cys Phe Ala Glu Lys Leu Ile Ser Gly Gly Val
            20                  25                  30

Phe Ser Tyr Tyr Pro Ser Glu Arg Leu Lys Glu Phe Val Gly Asn His
        35                  40                  45

Pro Phe Ser Leu Phe Arg Lys Thr Met Pro Phe Ser Pro Gly Phe Lys
    50                  55                  60

Arg Val Met Lys Ser Gly Gly Asn Tyr Gln Asn Ala Asn Arg Asp Gly
65                  70                  75                  80

Arg Phe Tyr Asp Leu Asp Ile Gly Val Tyr Leu Pro Lys Asp Gly Phe
                85                  90                  95

Gly Asp Glu Glu Trp Asn Ala Arg Tyr Phe Leu Met Lys Leu Ile Tyr
            100                 105                 110

Asn Gln Leu Phe Leu Pro Tyr Phe Ala Asp Ala Glu Asn His Leu Phe
        115                 120                 125

Arg Glu Cys Val Asp Phe Val Lys Arg Val Asn Arg Asp Tyr Asn Cys
    130                 135                 140

Lys Asn Asn Asn Ser Glu Glu Gln Ala Phe Ile Asp Ile Arg Ser Met
145                 150                 155                 160

Arg Glu Asp Glu Ser Ile Ala Asp Tyr Leu Ala Phe Ile Gln Ser Asn
                165                 170                 175

Ile Ile Ile Glu Glu Asn Lys Lys Lys Glu Thr Asn Lys Glu Gly Gln
            180                 185                 190

Ile Asn Phe Asn Lys Phe Leu Leu Gln Val Phe Val Lys Gly Phe Asp
        195                 200                 205

Ser Phe Leu Lys Asp Arg Thr Glu Leu Asn Phe Leu Gln Leu Pro Glu
    210                 215                 220
```

```
Leu Gln Gly Asp Gly Thr Arg Gly Asp Asp Leu Glu Ser Leu Asp Lys
225                 230                 235                 240


Leu Gly Ala Val Val Ala Val Asp Leu Lys Leu Asp Ala Thr Gly Ile
                245                 250                 255


Asp Ala Asp Leu Asn Glu Asn Ile Ser Phe Tyr Thr Phe Cys Lys Leu
            260                 265                 270


Leu Asp Ser Asn His Leu Ser Arg Leu Arg Asn Glu Ile Ile Lys Tyr
        275                 280                 285


Gln Ser Ala Asn Ser Asp Phe Ser His Asn Glu Asp Phe Asp Tyr Asp
    290                 295                 300


Arg Ile Ile Ser Ile Ile Glu Leu Cys Met Leu Ser Ala Asp His Val
305                 310                 315                 320


Ser Thr Asn Asp Asn Glu Ser Ile Phe Pro Asn Asn Asp Lys Asp Phe
                325                 330                 335


Ser Gly Ile Arg Pro Tyr Leu Ser Thr Asp Ala Lys Val Glu Thr Phe
            340                 345                 350


Glu Asp Leu Tyr Val His Ser Asp Ala Lys Thr Pro Ile Thr Asn Ala
        355                 360                 365


Thr Met Val Leu Asn Trp Lys Tyr Gly Thr Asp Lys Leu Phe Glu Arg
    370                 375                 380


Leu Met Ile Ser Asp Gln Asp Phe Leu Val Thr Glu Lys Asp Tyr Phe
385                 390                 395                 400


Val Trp Lys Glu Leu Lys Lys Asp Ile Glu Glu Lys Ile Lys Leu Arg
                405                 410                 415


Glu Glu Leu His Ser Leu Trp Val Asn Thr Pro Lys Gly Lys Lys Gly
            420                 425                 430
```

```
Ala Lys Lys Lys Asn Gly Arg Glu Thr Thr Gly Glu Phe Ser Glu Glu
        435                 440                 445

Asn Lys Lys Glu Tyr Leu Glu Val Cys Arg Glu Ile Asp Arg Tyr Val
    450                 455                 460

Asn Leu Asp Asn Lys Leu His Phe Val His Leu Lys Arg Met His Ser
465                 470                 475                 480

Leu Leu Ile Glu Leu Leu Gly Arg Phe Val Gly Phe Thr Tyr Leu Phe
                485                 490                 495

Glu Arg Asp Tyr Gln Tyr Tyr His Leu Glu Ile Arg Ser Arg Arg Asn
            500                 505                 510

Lys Asp Ala Gly Val Val Asp Lys Leu Glu Tyr Asn Lys Ile Lys Asp
        515                 520                 525

Gln Asn Lys Tyr Asp Lys Asp Asp Phe Phe Ala Cys Thr Phe Leu Tyr
    530                 535                 540

Glu Lys Ala Asn Lys Val Arg Asn Phe Ile Ala His Phe Asn Tyr Leu
545                 550                 555                 560

Thr Met Trp Asn Ser Pro Gln Glu Glu Glu His Asn Ser Asn Leu Ser
                565                 570                 575

Gly Ala Lys Asn Ser Ser Gly Arg Gln Asn Leu Lys Cys Ser Leu Thr
            580                 585                 590

Glu Leu Ile Asn Glu Leu Arg Glu Val Met Ser Tyr Asp Arg Lys Leu
        595                 600                 605

Lys Asn Ala Val Thr Lys Ala Val Ile Asp Leu Phe Asp Lys His Gly
    610                 615                 620

Met Val Ile Lys Phe Arg Ile Val Asn Asn Asn Asn Asn Asp Asn Lys
625                 630                 635                 640
```

```
Asn Lys His His Leu Glu Leu Asp Asp Ile Val Pro Lys Lys Ile Met
                645                 650                 655


His Leu Arg Gly Ile Lys Leu Lys Arg Gln Asp Gly Lys Pro Ile Pro
            660                 665                 670


Ile Gln Thr Asp Ser Val Asp Pro Leu Tyr Cys Arg Met Trp Lys Lys
        675                 680                 685


Leu Leu Asp Leu Lys Pro Thr Pro Phe
    690                 695


<210>  65
<211>  204
<212>  PRT
<213>  Listeria riparia

<400>  65

Met His Asp Ala Trp Ala Glu Asn Pro Lys Lys Pro Gln Ser Asp Ala
1               5                   10                  15


Phe Leu Lys Glu Tyr Lys Ala Cys Cys Glu Ala Ile Asp Thr Tyr Asn
            20                  25                  30


Trp His Lys Asn Lys Ala Thr Leu Val Tyr Val Asn Glu Leu His His
        35                  40                  45


Leu Leu Ile Asp Ile Leu Gly Arg Leu Val Gly Tyr Val Ala Ile Ala
    50                  55                  60


Asp Arg Asp Phe Gln Cys Met Ala Asn Gln Tyr Leu Lys Ser Ser Gly
65                  70                  75                  80


His Thr Glu Arg Val Asp Ser Trp Ile Asn Thr Ile Arg Lys Asn Arg
                85                  90                  95


Pro Asp Tyr Ile Glu Lys Leu Asp Ile Phe Met Asn Lys Ala Gly Leu
            100                 105                 110
```

```
Phe Val Ser Glu Lys Asn Gly Arg Asn Tyr Ile Ala His Leu Asn Tyr
        115                 120                 125


Leu Ser Pro Lys His Lys Tyr Ser Leu Leu Tyr Leu Phe Glu Lys Leu
    130                 135                 140


Arg Glu Met Leu Lys Tyr Asp Arg Lys Leu Lys Asn Ala Val Thr Lys
145                 150                 155                 160


Ser Leu Ile Asp Leu Leu Asp Lys His Gly Met Cys Val Val Phe Ala
                165                 170                 175


Asn Leu Lys Asn Asn Lys His Arg Leu Val Ile Ala Ser Leu Lys Pro
            180                 185                 190


Lys Lys Ile Glu Thr Phe Lys Trp Lys Lys Ile Lys
        195                 200


<210>  66
<211>  1082
<212>  PRT
<213>  Insolitispirillum peregrinum

<400>  66

Met Arg Ile Ile Arg Pro Tyr Gly Ser Ser Thr Val Ala Ser Pro Ser
1               5                   10                  15


Pro Gln Asp Ala Gln Pro Leu Arg Ser Leu Gln Arg Gln Asn Gly Thr
            20                  25                  30


Phe Asp Val Ala Glu Phe Ser Arg Arg His Pro Glu Leu Val Leu Ala
        35                  40                  45


Gln Trp Val Ala Met Leu Asp Lys Ile Ile Arg Lys Pro Ala Pro Gly
    50                  55                  60


Lys Asn Ser Thr Ala Leu Pro Arg Pro Thr Ala Glu Gln Arg Arg Leu
65                  70                  75                  80


Arg Gln Gln Val Gly Ala Ala Leu Trp Ala Glu Met Gln Arg His Thr
```

```
                85                  90                  95

Pro Val Pro Pro Glu Leu Lys Ala Val Trp Asp Ser Lys Val His Pro
            100                 105                 110

Tyr Ser Lys Asp Asn Ala Pro Ala Thr Ala Lys Thr Pro Ser His Arg
        115                 120                 125

Gly Arg Trp Tyr Asp Arg Phe Gly Asp Pro Glu Thr Ser Ala Ala Thr
    130                 135                 140

Val Ala Glu Gly Val Arg Arg His Leu Leu Asp Ser Ala Gln Pro Phe
145                 150                 155                 160

Arg Ala Asn Gly Gly Gln Pro Lys Gly Lys Gly Val Ile Glu His Arg
                165                 170                 175

Ala Leu Thr Ile Gln Asn Gly Thr Leu Leu His His His Gln Ser Glu
            180                 185                 190

Lys Ala Gly Pro Leu Pro Glu Asp Trp Ser Thr Tyr Arg Ala Asp Glu
        195                 200                 205

Leu Val Ser Thr Ile Gly Lys Asp Ala Arg Trp Ile Lys Val Ala Ala
    210                 215                 220

Ser Leu Tyr Gln His Tyr Gly Arg Ile Phe Gly Pro Thr Thr Pro Ile
225                 230                 235                 240

Ser Glu Ala Gln Thr Arg Pro Glu Phe Val Leu His Thr Ala Val Lys
                245                 250                 255

Ala Tyr Tyr Arg Arg Leu Phe Lys Glu Arg Lys Leu Pro Ala Glu Arg
            260                 265                 270

Leu Glu Arg Leu Leu Pro Arg Thr Gly Glu Ala Leu Arg His Ala Val
        275                 280                 285

Thr Val Gln His Gly Asn Arg Ser Leu Ala Asp Ala Val Arg Ile Gly
```

```
    290                 295                 300

Lys Ile Leu His Tyr Gly Trp Leu Gln Asn Gly Glu Pro Asp Pro Trp
305                 310                 315                 320

Pro Asp Asp Ala Ala Leu Tyr Ser Ser Arg Tyr Trp Gly Ser Asp Gly
                325                 330                 335

Gln Thr Asp Ile Lys His Ser Glu Ala Val Ser Arg Val Trp Arg Arg
            340                 345                 350

Ala Leu Thr Ala Ala Gln Arg Thr Leu Thr Ser Trp Leu Tyr Pro Ala
        355                 360                 365

Gly Thr Asp Ala Gly Asp Ile Leu Leu Ile Gly Gln Lys Pro Asp Ser
    370                 375                 380

Ile Asp Arg Asn Arg Leu Pro Leu Leu Tyr Gly Asp Ser Thr Arg His
385                 390                 395                 400

Trp Thr Arg Ser Pro Gly Asp Val Trp Leu Phe Leu Lys Gln Thr Leu
                405                 410                 415

Glu Asn Leu Arg Asn Ser Ser Phe His Phe Lys Thr Leu Ser Ala Phe
            420                 425                 430

Thr Ser His Leu Asp Gly Thr Cys Glu Ser Glu Pro Ala Glu Gln Gln
        435                 440                 445

Ala Ala Gln Ala Leu Trp Gln Asp Asp Arg Gln Gln Asp His Gln Gln
    450                 455                 460

Val Phe Leu Ser Leu Arg Ala Leu Asp Ala Thr Thr Tyr Leu Pro Thr
465                 470                 475                 480

Gly Pro Leu His Arg Ile Val Asn Ala Val Gln Ser Thr Asp Ala Thr
                485                 490                 495

Leu Pro Leu Pro Arg Phe Arg Arg Val Val Thr Arg Ala Ala Asn Thr
```

```
            500                 505                 510

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

Asp Asp Pro Leu Leu Arg Cys Arg Tyr Gly Val Leu Lys Leu Leu Tyr
    530                 535                 540

Glu Arg Gly Phe Arg Ala Trp Leu Glu Thr Arg Pro Ser Ile Ala Ser
545                 550                 555                 560

Cys Leu Asp Gln Ser Leu Lys Arg Ser Thr Lys Ala Ala Gln Thr Ile
                565                 570                 575

Asn Gly Lys Asn Ser Pro Gln Gly Val Glu Ile Leu Ser Arg Ala Thr
            580                 585                 590

Lys Leu Leu Gln Ala Glu Gly Gly Gly Gly His Gly Ile His Asp Leu
        595                 600                 605

Phe Asp Arg Leu Tyr Ala Ala Thr Ala Arg Glu Met Arg Val Gln Val
    610                 615                 620

Gly Tyr His His Asp Ala Glu Ala Ala Arg Gln Gln Ala Glu Phe Ile
625                 630                 635                 640

Glu Asp Leu Lys Cys Glu Val Val Ala Arg Ala Phe Cys Ala Tyr Leu
                645                 650                 655

Lys Thr Leu Gly Ile Gln Gly Asp Thr Phe Arg Arg Gln Pro Glu Pro
            660                 665                 670

Leu Pro Thr Trp Pro Asp Leu Pro Asp Leu Pro Ser Ser Thr Ile Gly
        675                 680                 685

Thr Ala Gln Ala Ala Leu Tyr Ser Val Leu His Leu Met Pro Val Glu
    690                 695                 700

Asp Val Gly Ser Leu Leu His Gln Leu Arg Arg Trp Leu Val Ala Leu
```

705 710 715 720

Gln Ala Arg Gly Gly Glu Asp Gly Thr Ala Ile Thr Ala Thr Ile Pro
725 730 735

Leu Leu Glu Leu Tyr Leu Asn Arg His Asp Ala Lys Phe Ser Gly Gly
740 745 750

Gly Ala Gly Thr Gly Leu Arg Trp Asp Asp Trp Gln Val Phe Phe Asp
755 760 765

Cys Gln Ala Thr Phe Asp Arg Val Phe Pro Pro Gly Pro Ala Leu Asp
770 775 780

Ser His Arg Leu Pro Leu Arg Gly Leu Arg Glu Val Leu Arg Phe Gly
785 790 795 800

Arg Val Asn Asp Leu Ala Ala Leu Ile Gly Gln Asp Lys Ile Thr Ala
805 810 815

Ala Glu Val Asp Arg Trp His Thr Ala Glu Gln Thr Ile Ala Ala Gln
820 825 830

Gln Gln Arg Arg Glu Ala Leu His Glu Gln Leu Ser Arg Lys Lys Gly
835 840 845

Thr Asp Ala Glu Val Asp Glu Tyr Arg Ala Leu Val Thr Ala Ile Ala
850 855 860

Asp His Arg His Leu Thr Ala His Val Thr Leu Ser Asn Val Val Arg
865 870 875 880

Leu His Arg Leu Met Thr Thr Val Leu Gly Arg Leu Val Asp Tyr Gly
885 890 895

Gly Leu Trp Glu Arg Asp Leu Thr Phe Val Thr Leu Tyr Glu Ala His
900 905 910

Arg Leu Gly Gly Leu Arg Asn Leu Leu Ser Glu Ser Arg Val Asn Lys

915 920 925

Phe Leu Asp Gly Gln Thr Pro Ala Ala Leu Ser Lys Lys Asn Asn Ala
930 935 940

Glu Glu Asn Gly Met Ile Ser Lys Val Leu Gly Asp Lys Ala Arg Arg
945 950 955 960

Gln Ile Arg Asn Asp Phe Ala His Phe Asn Met Leu Gln Gln Gly Lys
965 970 975

Lys Thr Ile Asn Leu Thr Asp Glu Ile Asn Asn Ala Arg Lys Leu Met
980 985 990

Ala His Asp Arg Lys Leu Lys Asn Ala Ile Thr Arg Ser Val Thr Thr
995 1000 1005

Leu Leu Gln Gln Asp Gly Leu Asp Ile Val Trp Thr Met Asp Ala
1010 1015 1020

Ser His Arg Leu Thr Asp Ala Lys Ile Asp Ser Arg Asn Ala Ile
1025 1030 1035

His Leu His Lys Thr His Asn Arg Ala Asn Ile Arg Glu Pro Leu
1040 1045 1050

His Gly Lys Ser Tyr Cys Arg Trp Val Ala Ala Leu Phe Gly Ala
1055 1060 1065

Thr Ser Thr Pro Ser Ala Thr Lys Lys Ser Asp Lys Ile Arg
1070 1075 1080

<210> 67
<211> 1224
<212> PRT
<213> Bergeyella zoohelcum

<400> 67

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

```
Lys Pro Gln Asp Lys Ser Tyr Phe Ala Gly Tyr Phe Asn Ala Ala Met
            20                  25                  30

Glu Asn Thr Asp Ser Val Phe Arg Glu Leu Gly Lys Arg Leu Lys Gly
        35                  40                  45

Lys Glu Tyr Thr Ser Glu Asn Phe Phe Asp Ala Ile Phe Lys Glu Asn
    50                  55                  60

Ile Ser Leu Val Glu Tyr Glu Arg Tyr Val Lys Leu Leu Ser Asp Tyr
65                  70                  75                  80

Phe Pro Met Ala Arg Leu Leu Asp Lys Lys Glu Val Pro Ile Lys Glu
                85                  90                  95

Arg Lys Glu Asn Phe Lys Lys Asn Phe Lys Gly Ile Ile Lys Ala Val
            100                 105                 110

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

Ile Thr Asp Glu Ile Phe Gly Val Leu Asp Glu Met Leu Lys Ser Thr
    130                 135                 140

Val Leu Thr Val Lys Lys Lys Lys Val Lys Thr Asp Lys Thr Lys Glu
145                 150                 155                 160

Ile Leu Lys Lys Ser Ile Glu Lys Gln Leu Asp Ile Leu Cys Gln Lys
                165                 170                 175

Lys Leu Glu Tyr Leu Arg Asp Thr Ala Arg Lys Ile Glu Glu Lys Arg
            180                 185                 190

Arg Asn Gln Arg Glu Arg Gly Glu Lys Glu Leu Val Ala Pro Phe Lys
        195                 200                 205

Tyr Ser Asp Lys Arg Asp Asp Leu Ile Ala Ala Ile Tyr Asn Asp Ala
    210                 215                 220
```

```
Phe Asp Val Tyr Ile Asp Lys Lys Lys Asp Ser Leu Lys Glu Ser Ser
225                 230                 235                 240

Lys Ala Lys Tyr Asn Thr Lys Ser Asp Pro Gln Gln Glu Glu Gly Asp
                245                 250                 255

Leu Lys Ile Pro Ile Ser Lys Asn Gly Val Val Phe Leu Leu Ser Leu
            260                 265                 270

Phe Leu Thr Lys Gln Glu Ile His Ala Phe Lys Ser Lys Ile Ala Gly
        275                 280                 285

Phe Lys Ala Thr Val Ile Asp Glu Ala Thr Val Ser Glu Ala Thr Val
    290                 295                 300

Ser His Gly Lys Asn Ser Ile Cys Phe Met Ala Thr His Glu Ile Phe
305                 310                 315                 320

Ser His Leu Ala Tyr Lys Lys Leu Lys Arg Lys Val Arg Thr Ala Glu
                325                 330                 335

Ile Asn Tyr Gly Glu Ala Glu Asn Ala Glu Gln Leu Ser Val Tyr Ala
            340                 345                 350

Lys Glu Thr Leu Met Met Gln Met Leu Asp Glu Leu Ser Lys Val Pro
        355                 360                 365

Asp Val Val Tyr Gln Asn Leu Ser Glu Asp Val Gln Lys Thr Phe Ile
    370                 375                 380

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

Met Glu Glu Glu Gln Val Ile His Pro Val Ile Arg Lys Arg Tyr Glu
                405                 410                 415

Asp Lys Phe Asn Tyr Phe Ala Ile Arg Phe Leu Asp Glu Phe Ala Gln
            420                 425                 430
```

```
Phe Pro Thr Leu Arg Phe Gln Val His Leu Gly Asn Tyr Leu His Asp
        435                 440                 445

Ser Arg Pro Lys Glu Asn Leu Ile Ser Asp Arg Arg Ile Lys Glu Lys
    450                 455                 460

Ile Thr Val Phe Gly Arg Leu Ser Glu Leu Glu His Lys Lys Ala Leu
465                 470                 475                 480

Phe Ile Lys Asn Thr Glu Thr Asn Glu Asp Arg Glu His Tyr Trp Glu
                485                 490                 495

Ile Phe Pro Asn Pro Asn Tyr Asp Phe Pro Lys Glu Asn Ile Ser Val
            500                 505                 510

Asn Asp Lys Asp Phe Pro Ile Ala Gly Ser Ile Leu Asp Arg Glu Lys
        515                 520                 525

Gln Pro Val Ala Gly Lys Ile Gly Ile Lys Val Lys Leu Leu Asn Gln
    530                 535                 540

Gln Tyr Val Ser Glu Val Asp Lys Ala Val Lys Ala His Gln Leu Lys
545                 550                 555                 560

Gln Arg Lys Ala Ser Lys Pro Ser Ile Gln Asn Ile Ile Glu Glu Ile
                565                 570                 575

Val Pro Ile Asn Glu Ser Asn Pro Lys Glu Ala Ile Val Phe Gly Gly
            580                 585                 590

Gln Pro Thr Ala Tyr Leu Ser Met Asn Asp Ile His Ser Ile Leu Tyr
        595                 600                 605

Glu Phe Phe Asp Lys Trp Glu Lys Lys Lys Glu Lys Leu Glu Lys Lys
    610                 615                 620

Gly Glu Lys Glu Leu Arg Lys Glu Ile Gly Lys Glu Leu Glu Lys Lys
625                 630                 635                 640
```

```
Ile Val Gly Lys Ile Gln Ala Gln Ile Gln Gln Ile Ile Asp Lys Asp
                645                 650                 655

Thr Asn Ala Lys Ile Leu Lys Pro Tyr Gln Asp Gly Asn Ser Thr Ala
            660                 665                 670

Ile Asp Lys Glu Lys Leu Ile Lys Asp Leu Lys Gln Glu Gln Asn Ile
        675                 680                 685

Leu Gln Lys Leu Lys Asp Glu Gln Thr Val Arg Glu Lys Glu Tyr Asn
    690                 695                 700

Asp Phe Ile Ala Tyr Gln Asp Lys Asn Arg Glu Ile Asn Lys Val Arg
705                 710                 715                 720

Asp Arg Asn His Lys Gln Tyr Leu Lys Asp Asn Leu Lys Arg Lys Tyr
                725                 730                 735

Pro Glu Ala Pro Ala Arg Lys Glu Val Leu Tyr Tyr Arg Glu Lys Gly
            740                 745                 750

Lys Val Ala Val Trp Leu Ala Asn Asp Ile Lys Arg Phe Met Pro Thr
        755                 760                 765

Asp Phe Lys Asn Glu Trp Lys Gly Glu Gln His Ser Leu Leu Gln Lys
    770                 775                 780

Ser Leu Ala Tyr Tyr Glu Gln Cys Lys Glu Glu Leu Lys Asn Leu Leu
785                 790                 795                 800

Pro Glu Lys Val Phe Gln His Leu Pro Phe Lys Leu Gly Gly Tyr Phe
                805                 810                 815

Gln Gln Lys Tyr Leu Tyr Gln Phe Tyr Thr Cys Tyr Leu Asp Lys Arg
            820                 825                 830

Leu Glu Tyr Ile Ser Gly Leu Val Gln Gln Ala Glu Asn Phe Lys Ser
        835                 840                 845
```

```
Glu Asn Lys Val Phe Lys Lys Val Glu Asn Glu Cys Phe Lys Phe Leu
    850                 855                 860

Lys Lys Gln Asn Tyr Thr His Lys Glu Leu Asp Ala Arg Val Gln Ser
865                 870                 875                 880

Ile Leu Gly Tyr Pro Ile Phe Leu Glu Arg Gly Phe Met Asp Glu Lys
                885                 890                 895

Pro Thr Ile Ile Lys Gly Lys Thr Phe Lys Gly Asn Glu Ala Leu Phe
            900                 905                 910

Ala Asp Trp Phe Arg Tyr Tyr Lys Glu Tyr Gln Asn Phe Gln Thr Phe
        915                 920                 925

Tyr Asp Thr Glu Asn Tyr Pro Leu Val Glu Leu Glu Lys Lys Gln Ala
    930                 935                 940

Asp Arg Lys Arg Lys Thr Lys Ile Tyr Gln Gln Lys Lys Asn Asp Val
945                 950                 955                 960

Phe Thr Leu Leu Met Ala Lys His Ile Phe Lys Ser Val Phe Lys Gln
                965                 970                 975

Asp Ser Ile Asp Gln Phe Ser Leu Glu Asp Leu Tyr Gln Ser Arg Glu
            980                 985                 990

Glu Arg Leu Gly Asn Gln Glu Arg  Ala Arg Gln Thr Gly  Glu Arg Asn
        995                 1000                1005

Thr Asn  Tyr Ile Trp Asn Lys  Thr Val Asp Leu Lys  Leu Cys Asp
    1010                1015                1020

Gly Lys  Ile Thr Val Glu Asn  Val Lys Leu Lys Asn  Val Gly Asp
    1025                1030                1035

Phe Ile  Lys Tyr Glu Tyr Asp  Gln Arg Val Gln Ala  Phe Leu Lys
    1040                1045                1050
```

```
Tyr Glu  Glu Asn Ile Glu Trp  Gln Ala Phe Leu Ile  Lys Glu Ser
    1055                1060                1065

Lys Glu  Glu Glu Asn Tyr Pro  Tyr Val Val Glu Arg  Glu Ile Glu
    1070                1075                1080

Gln Tyr  Glu Lys Val Arg Arg  Glu Glu Leu Leu Lys  Glu Val His
    1085                1090                1095

Leu Ile  Glu Glu Tyr Ile Leu  Glu Lys Val Lys Asp  Lys Glu Ile
    1100                1105                1110

Leu Lys  Lys Gly Asp Asn Gln  Asn Phe Lys Tyr Tyr  Ile Leu Asn
    1115                1120                1125

Gly Leu  Leu Lys Gln Leu Lys  Asn Glu Asp Val Glu  Ser Tyr Lys
    1130                1135                1140

Val Phe  Asn Leu Asn Thr Glu  Pro Glu Asp Val Asn  Ile Asn Gln
    1145                1150                1155

Leu Lys  Gln Glu Ala Thr Asp  Leu Glu Gln Lys Ala  Phe Val Leu
    1160                1165                1170

Thr Tyr  Ile Arg Asn Lys Phe  Ala His Asn Gln Leu  Pro Lys Lys
    1175                1180                1185

Glu Phe  Trp Asp Tyr Cys Gln  Glu Lys Tyr Gly Lys  Ile Glu Lys
    1190                1195                1200

Glu Lys  Thr Tyr Ala Glu Tyr  Phe Ala Glu Val Phe  Lys Lys Glu
    1205                1210                1215

Lys Glu  Ala Leu Ile Lys
    1220
```

<210> 68
<211> 1126

<212> PRT
<213> Prevotella intermedia

<400> 68

```
Met Glu Asp Asp Lys Lys Thr Thr Asp Ser Ile Arg Tyr Glu Leu Lys
1               5                  10                  15


Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn Val
            20                  25                  30


Tyr Ile Thr Val Asn His Ile Asn Lys Ile Leu Glu Glu Gly Glu Ile
        35                  40                  45


Asn Arg Asp Gly Tyr Glu Thr Thr Leu Lys Asn Thr Trp Asn Glu Ile
    50                  55                  60


Lys Asp Ile Asn Lys Lys Asp Arg Leu Ser Lys Leu Ile Ile Lys His
65                  70                  75                  80


Phe Pro Phe Leu Glu Ala Ala Thr Tyr Arg Leu Asn Pro Thr Asp Thr
                85                  90                  95


Thr Lys Gln Lys Glu Glu Lys Gln Ala Glu Ala Gln Ser Leu Glu Ser
            100                 105                 110


Leu Arg Lys Ser Phe Phe Val Phe Ile Tyr Lys Leu Arg Asp Leu Arg
        115                 120                 125


Asn His Tyr Ser His Tyr Lys His Ser Lys Ser Leu Glu Arg Pro Lys
    130                 135                 140


Phe Glu Glu Gly Leu Leu Glu Lys Met Tyr Asn Ile Phe Asn Ala Ser
145                 150                 155                 160


Ile Arg Leu Val Lys Glu Asp Tyr Gln Tyr Asn Lys Asp Ile Asn Pro
                165                 170                 175


Asp Glu Asp Phe Lys His Leu Asp Arg Thr Glu Glu Glu Phe Asn Tyr
            180                 185                 190
```

```
Tyr Phe Thr Lys Asp Asn Glu Gly Asn Ile Thr Glu Ser Gly Leu Leu
        195                 200                 205


Phe Phe Val Ser Leu Phe Leu Glu Lys Lys Asp Ala Ile Trp Met Gln
    210                 215                 220


Gln Lys Leu Arg Gly Phe Lys Asp Asn Arg Glu Asn Lys Lys Lys Met
225                 230                 235                 240


Thr Asn Glu Val Phe Cys Arg Ser Arg Met Leu Leu Pro Lys Leu Arg
                245                 250                 255


Leu Gln Ser Thr Gln Thr Gln Asp Trp Ile Leu Leu Asp Met Leu Asn
            260                 265                 270


Glu Leu Ile Arg Cys Pro Lys Ser Leu Tyr Glu Arg Leu Arg Glu Glu
        275                 280                 285


Asp Arg Glu Lys Phe Arg Val Pro Ile Glu Ile Ala Asp Glu Asp Tyr
    290                 295                 300


Asp Ala Glu Gln Glu Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp
305                 310                 315                 320


Arg Phe Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Tyr Asn Glu Ile Phe
                325                 330                 335


Thr Asn Leu Arg Phe Gln Ile Asp Leu Gly Thr Tyr His Phe Ser Ile
            340                 345                 350


Tyr Lys Lys Gln Ile Gly Asp Tyr Lys Glu Ser His His Leu Thr His
        355                 360                 365


Lys Leu Tyr Gly Phe Glu Arg Ile Gln Glu Phe Thr Lys Gln Asn Arg
    370                 375                 380


Pro Asp Glu Trp Arg Lys Phe Val Lys Thr Phe Asn Ser Phe Glu Thr
385                 390                 395                 400
```

```
Ser Lys Glu Pro Tyr Ile Pro Glu Thr Thr Pro His Tyr His Leu Glu
                405                 410                 415

Asn Gln Lys Ile Gly Ile Arg Phe Arg Asn Asp Asn Asp Lys Ile Trp
            420                 425                 430

Pro Ser Leu Lys Thr Asn Ser Glu Lys Asn Glu Lys Ser Lys Tyr Lys
        435                 440                 445

Leu Asp Lys Ser Phe Gln Ala Glu Ala Phe Leu Ser Val His Glu Leu
    450                 455                 460

Leu Pro Met Met Phe Tyr Tyr Leu Leu Leu Lys Thr Glu Asn Thr Asp
465                 470                 475                 480

Asn Asp Asn Glu Ile Glu Thr Lys Lys Lys Glu Asn Lys Asn Asp Lys
                485                 490                 495

Gln Glu Lys His Lys Ile Glu Glu Ile Ile Glu Asn Lys Ile Thr Glu
            500                 505                 510

Ile Tyr Ala Leu Tyr Asp Thr Phe Ala Asn Gly Glu Ile Lys Ser Ile
        515                 520                 525

Asp Glu Leu Glu Glu Tyr Cys Lys Gly Lys Asp Ile Glu Ile Gly His
    530                 535                 540

Leu Pro Lys Gln Met Ile Ala Ile Leu Lys Asp Glu His Lys Val Met
545                 550                 555                 560

Ala Thr Glu Ala Glu Arg Lys Gln Glu Glu Met Leu Val Asp Val Gln
                565                 570                 575

Lys Ser Leu Glu Ser Leu Asp Asn Gln Ile Asn Glu Glu Ile Glu Asn
            580                 585                 590

Val Glu Arg Lys Asn Ser Ser Leu Lys Ser Gly Lys Ile Ala Ser Trp
        595                 600                 605
```

```
Leu Val Asn Asp Met Met Arg Phe Gln Pro Val Gln Lys Asp Asn Glu
    610                 615                 620

Gly Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Gln Leu
625                 630                 635                 640

Leu Gln Arg Thr Leu Ala Phe Phe Gly Ser Glu His Glu Arg Leu Ala
                645                 650                 655

Pro Tyr Phe Lys Gln Thr Lys Leu Ile Glu Ser Ser Asn Pro His Pro
            660                 665                 670

Phe Leu Lys Asp Thr Glu Trp Glu Lys Cys Asn Asn Ile Leu Ser Phe
        675                 680                 685

Tyr Arg Ser Tyr Leu Glu Ala Lys Lys Asn Phe Leu Glu Ser Leu Lys
    690                 695                 700

Pro Glu Asp Trp Glu Lys Asn Gln Tyr Phe Leu Lys Leu Lys Glu Pro
705                 710                 715                 720

Lys Thr Lys Pro Lys Thr Leu Val Gln Gly Trp Lys Asn Gly Phe Asn
                725                 730                 735

Leu Pro Arg Gly Ile Phe Thr Glu Pro Ile Arg Lys Trp Phe Met Lys
            740                 745                 750

His Arg Glu Asn Ile Thr Val Ala Glu Leu Lys Arg Val Gly Leu Val
        755                 760                 765

Ala Lys Val Ile Pro Leu Phe Phe Ser Glu Glu Tyr Lys Asp Ser Val
    770                 775                 780

Gln Pro Phe Tyr Asn Tyr His Phe Asn Val Gly Asn Ile Asn Lys Pro
785                 790                 795                 800

Asp Glu Lys Asn Phe Leu Asn Cys Glu Glu Arg Arg Glu Leu Leu Arg
                805                 810                 815
```

```
Lys Lys Lys Asp Glu Phe Lys Lys Met Thr Asp Lys Glu Lys Glu Glu
            820                 825                 830

Asn Pro Ser Tyr Leu Glu Phe Lys Ser Trp Asn Lys Phe Glu Arg Glu
        835                 840                 845

Leu Arg Leu Val Arg Asn Gln Asp Ile Val Thr Trp Leu Leu Cys Met
    850                 855                 860

Glu Leu Phe Asn Lys Lys Lys Ile Lys Glu Leu Asn Val Glu Lys Ile
865                 870                 875                 880

Tyr Leu Lys Asn Ile Asn Thr Asn Thr Thr Lys Lys Glu Lys Asn Thr
                885                 890                 895

Glu Glu Lys Asn Gly Glu Glu Lys Asn Ile Lys Glu Lys Asn Asn Ile
            900                 905                 910

Leu Asn Arg Ile Met Pro Met Arg Leu Pro Ile Lys Val Tyr Gly Arg
        915                 920                 925

Glu Asn Phe Ser Lys Asn Lys Lys Lys Lys Ile Arg Arg Asn Thr Phe
    930                 935                 940

Phe Thr Val Tyr Ile Glu Glu Lys Gly Thr Lys Leu Leu Lys Gln Gly
945                 950                 955                 960

Asn Phe Lys Ala Leu Glu Arg Asp Arg Arg Leu Gly Gly Leu Phe Ser
                965                 970                 975

Phe Val Lys Thr Pro Ser Lys Ala Glu Ser Lys Ser Asn Thr Ile Ser
            980                 985                 990

Lys Leu Arg Val Glu Tyr Glu Leu  Gly Glu Tyr Gln Lys  Ala Arg Ile
        995                 1000                1005

Glu Ile  Ile Lys Asp Met Leu  Ala Leu Glu Lys Thr  Leu Ile Asp
    1010                1015                1020
```

```
Lys Tyr  Asn Ser Leu Asp Thr  Asp Asn Phe Asn Lys  Met Leu Thr
    1025                 1030                 1035


Asp Trp  Leu Glu Leu Lys Gly  Glu Pro Asp Lys Ala  Ser Phe Gln
    1040                 1045                 1050


Asn Asp  Val Asp Leu Leu Ile  Ala Val Arg Asn Ala  Phe Ser His
    1055                 1060                 1065


Asn Gln  Tyr Pro Met Arg Asn  Arg Ile Ala Phe Ala  Asn Ile Asn
    1070                 1075                 1080


Pro Phe  Ser Leu Ser Ser Ala  Asn Thr Ser Glu Glu  Lys Gly Leu
    1085                 1090                 1095


Gly Ile  Ala Asn Gln Leu Lys  Asp Lys Thr His Lys  Thr Ile Glu
    1100                 1105                 1110


Lys Ile  Ile Glu Ile Glu Lys  Pro Ile Glu Thr Lys  Glu
    1115                 1120                 1125


<210>  69
<211>  1127
<212>  PRT
<213>  Prevotella buccae

<400>  69

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


Phe Ala Phe Pro Lys Tyr Ile Thr Ile Met Glu Asn Lys Glu Lys Pro
            20                  25                  30


Glu Pro Ile Tyr Tyr Glu Leu Thr Asp Lys His Phe Trp Ala Ala Phe
        35                  40                  45


Leu Asn Leu Ala Arg His Asn Val Tyr Thr Thr Ile Asn His Ile Asn
    50                  55                  60
```

```
Arg Arg Leu Glu Ile Ala Glu Leu Lys Asp Asp Gly Tyr Met Met Gly
65                  70                  75                  80


Ile Lys Gly Ser Trp Asn Glu Gln Ala Lys Lys Leu Asp Lys Lys Val
                85                  90                  95


Arg Leu Arg Asp Leu Ile Met Lys His Phe Pro Phe Leu Glu Ala Ala
            100                 105                 110


Ala Tyr Glu Met Thr Asn Ser Lys Ser Pro Asn Asn Lys Glu Gln Arg
        115                 120                 125


Glu Lys Glu Gln Ser Glu Ala Leu Ser Leu Asn Asn Leu Lys Asn Val
    130                 135                 140


Leu Phe Ile Phe Leu Glu Lys Leu Gln Val Leu Arg Asn Tyr Tyr Ser
145                 150                 155                 160


His Tyr Lys Tyr Ser Glu Glu Ser Pro Lys Pro Ile Phe Glu Thr Ser
                165                 170                 175


Leu Leu Lys Asn Met Tyr Lys Val Phe Asp Ala Asn Val Arg Leu Val
            180                 185                 190


Lys Arg Asp Tyr Met His His Glu Asn Ile Asp Met Gln Arg Asp Phe
        195                 200                 205


Thr His Leu Asn Arg Lys Lys Gln Val Gly Arg Thr Lys Asn Ile Ile
    210                 215                 220


Asp Ser Pro Asn Phe His Tyr His Phe Ala Asp Lys Glu Gly Asn Met
225                 230                 235                 240


Thr Ile Ala Gly Leu Leu Phe Phe Val Ser Leu Phe Leu Asp Lys Lys
                245                 250                 255


Asp Ala Ile Trp Met Gln Lys Lys Leu Lys Gly Phe Lys Asp Gly Arg
            260                 265                 270
```

```
Asn Leu Arg Glu Gln Met Thr Asn Glu Val Phe Cys Arg Ser Arg Ile
        275                 280                 285

Ser Leu Pro Lys Leu Lys Leu Glu Asn Val Gln Thr Lys Asp Trp Met
    290                 295                 300

Gln Leu Asp Met Leu Asn Glu Leu Val Arg Cys Pro Lys Ser Leu Tyr
305                 310                 315                 320

Glu Arg Leu Arg Glu Lys Asp Arg Glu Ser Phe Lys Val Pro Phe Asp
                325                 330                 335

Ile Phe Ser Asp Asp Tyr Asn Ala Glu Glu Glu Pro Phe Lys Asn Thr
            340                 345                 350

Leu Val Arg His Gln Asp Arg Phe Pro Tyr Phe Val Leu Arg Tyr Phe
        355                 360                 365

Asp Leu Asn Glu Ile Phe Glu Gln Leu Arg Phe Gln Ile Asp Leu Gly
    370                 375                 380

Thr Tyr His Phe Ser Ile Tyr Asn Lys Arg Ile Gly Asp Glu Asp Glu
385                 390                 395                 400

Val Arg His Leu Thr His His Leu Tyr Gly Phe Ala Arg Ile Gln Asp
                405                 410                 415

Phe Ala Pro Gln Asn Gln Pro Glu Glu Trp Arg Lys Leu Val Lys Asp
            420                 425                 430

Leu Asp His Phe Glu Thr Ser Gln Glu Pro Tyr Ile Ser Lys Thr Ala
        435                 440                 445

Pro His Tyr His Leu Glu Asn Glu Lys Ile Gly Ile Lys Phe Cys Ser
    450                 455                 460

Ala His Asn Asn Leu Phe Pro Ser Leu Gln Thr Asp Lys Thr Cys Asn
465                 470                 475                 480
```

```
Gly Arg Ser Lys Phe Asn Leu Gly Thr Gln Phe Thr Ala Glu Ala Phe
                485                 490                 495

Leu Ser Val His Glu Leu Leu Pro Met Met Phe Tyr Tyr Leu Leu Leu
            500                 505                 510

Thr Lys Asp Tyr Ser Arg Lys Glu Ser Ala Asp Lys Val Glu Gly Ile
        515                 520                 525

Ile Arg Lys Glu Ile Ser Asn Ile Tyr Ala Ile Tyr Asp Ala Phe Ala
    530                 535                 540

Asn Asn Glu Ile Asn Ser Ile Ala Asp Leu Thr Arg Arg Leu Gln Asn
545                 550                 555                 560

Thr Asn Ile Leu Gln Gly His Leu Pro Lys Gln Met Ile Ser Ile Leu
                565                 570                 575

Lys Gly Arg Gln Lys Asp Met Gly Lys Glu Ala Glu Arg Lys Ile Gly
            580                 585                 590

Glu Met Ile Asp Asp Thr Gln Arg Arg Leu Asp Leu Leu Cys Lys Gln
        595                 600                 605

Thr Asn Gln Lys Ile Arg Ile Gly Lys Arg Asn Ala Gly Leu Leu Lys
    610                 615                 620

Ser Gly Lys Ile Ala Asp Trp Leu Val Asn Asp Met Met Arg Phe Gln
625                 630                 635                 640

Pro Val Gln Lys Asp Gln Asn Asn Ile Pro Ile Asn Asn Ser Lys Ala
                645                 650                 655

Asn Ser Thr Glu Tyr Arg Met Leu Gln Arg Ala Leu Ala Leu Phe Gly
            660                 665                 670

Ser Glu Asn Phe Arg Leu Lys Ala Tyr Phe Asn Gln Met Asn Leu Val
        675                 680                 685
```

```
Gly Asn Asp Asn Pro His Pro Phe Leu Ala Glu Thr Gln Trp Glu His
    690                 695                 700

Gln Thr Asn Ile Leu Ser Phe Tyr Arg Asn Tyr Leu Glu Ala Arg Lys
705                 710                 715                 720

Lys Tyr Leu Lys Gly Leu Lys Pro Gln Asn Trp Lys Gln Tyr Gln His
                725                 730                 735

Phe Leu Ile Leu Lys Val Gln Lys Thr Asn Arg Asn Thr Leu Val Thr
            740                 745                 750

Gly Trp Lys Asn Ser Phe Asn Leu Pro Arg Gly Ile Phe Thr Gln Pro
        755                 760                 765

Ile Arg Glu Trp Phe Glu Lys His Asn Asn Ser Lys Arg Ile Tyr Asp
    770                 775                 780

Gln Ile Leu Ser Phe Asp Arg Val Gly Phe Val Ala Lys Ala Ile Pro
785                 790                 795                 800

Leu Tyr Phe Ala Glu Glu Tyr Lys Asp Asn Val Gln Pro Phe Tyr Asp
                805                 810                 815

Tyr Pro Phe Asn Ile Gly Asn Arg Leu Lys Pro Lys Lys Arg Gln Phe
            820                 825                 830

Leu Asp Lys Lys Glu Arg Val Glu Leu Trp Gln Lys Asn Lys Glu Leu
        835                 840                 845

Phe Lys Asn Tyr Pro Ser Glu Lys Lys Lys Thr Asp Leu Ala Tyr Leu
    850                 855                 860

Asp Phe Leu Ser Trp Lys Lys Phe Glu Arg Glu Leu Arg Leu Ile Lys
865                 870                 875                 880

Asn Gln Asp Ile Val Thr Trp Leu Met Phe Lys Glu Leu Phe Asn Met
                885                 890                 895
```

```
Ala Thr Val Glu Gly Leu Lys Ile Gly Glu Ile His Leu Arg Asp Ile
            900                 905                 910

Asp Thr Asn Thr Ala Asn Glu Glu Ser Asn Asn Ile Leu Asn Arg Ile
        915                 920                 925

Met Pro Met Lys Leu Pro Val Lys Thr Tyr Glu Thr Asp Asn Lys Gly
    930                 935                 940

Asn Ile Leu Lys Glu Arg Pro Leu Ala Thr Phe Tyr Ile Glu Glu Thr
945                 950                 955                 960

Glu Thr Lys Val Leu Lys Gln Gly Asn Phe Lys Ala Leu Val Lys Asp
                965                 970                 975

Arg Arg Leu Asn Gly Leu Phe Ser Phe Ala Glu Thr Thr Asp Leu Asn
            980                 985                 990

Leu Glu Glu His Pro Ile Ser Lys  Leu Ser Val Asp Leu  Glu Leu Ile
        995                 1000                1005

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

Leu Glu  Lys Lys Leu Ile Asp  Lys Tyr Ser Thr Leu  Pro Thr Asp
    1025                1030                1035

Ser Phe  Arg Asn Met Leu Glu  Arg Trp Leu Gln Cys  Lys Ala Asn
    1040                1045                1050

Arg Pro  Glu Leu Lys Asn Tyr  Val Asn Ser Leu Ile  Ala Val Arg
    1055                1060                1065

Asn Ala  Phe Ser His Asn Gln  Tyr Pro Met Tyr Asp  Ala Thr Leu
    1070                1075                1080

Phe Ala  Glu Val Lys Lys Phe  Thr Leu Phe Pro Ser  Val Asp Thr
    1085                1090                1095
```

Lys Lys Ile Glu Leu Asn Ile Ala Pro Gln Leu Leu Glu Ile Val
1100 1105 1110

Gly Lys Ala Ile Lys Glu Ile Glu Lys Ser Glu Asn Lys Asn
1115 1120 1125

<210> 70
<211> 1135
<212> PRT
<213> Porphyrmonas gingivalis

<400> 70

Met Asn Thr Val Pro Ala Ser Glu Asn Lys Gly Gln Ser Arg Thr Val
1 5 10 15

Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu
20 25 30

Asn Leu Ile Glu Val Glu Ser His Val Arg Ile Lys Phe Gly Lys Lys
35 40 45

Lys Leu Asn Glu Glu Ser Leu Lys Gln Ser Leu Leu Cys Asp His Leu
50 55 60

Leu Ser Val Asp Arg Trp Thr Lys Val Tyr Gly His Ser Arg Arg Tyr
65 70 75 80

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

His Asp Ser Lys Thr Gly Val Asp Pro Asp Ser Ala Gln Arg Leu Ile
100 105 110

Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His
115 120 125

Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Glu Val Ser Pro Asp
130 135 140

Ile Ser Ser Phe Ile Thr Gly Thr Tyr Ser Leu Ala Cys Gly Arg Ala

```
145                 150                 155                 160

Gln Ser Arg Phe Ala Val Phe Phe Lys Pro Asp Asp Phe Val Leu Ala
                165                 170                 175

Lys Asn Arg Lys Glu Gln Leu Ile Ser Val Ala Asp Gly Lys Glu Cys
            180                 185                 190

Leu Thr Val Ser Gly Phe Ala Phe Phe Ile Cys Leu Phe Leu Asp Arg
        195                 200                 205

Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr
    210                 215                 220

Asp Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys
225                 230                 235                 240

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

Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu
            260                 265                 270

Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu
        275                 280                 285

Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asp Glu Glu
    290                 295                 300

Ser Arg Leu Leu Trp Asp Gly Ser Ser Asp Trp Ala Glu Ala Leu Thr
305                 310                 315                 320

Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe
                325                 330                 335

Ile Glu Glu Met Asp Leu Leu Lys Gly Ile Arg Phe Arg Val Asp Leu
            340                 345                 350

Gly Glu Ile Glu Leu Asp Ser Tyr Ser Lys Lys Val Gly Arg Asn Gly
```

```
        355                 360                 365

Glu Tyr Asp Arg Thr Ile Thr Asp His Ala Leu Ala Phe Gly Lys Leu
    370                 375                 380

Ser Asp Phe Gln Asn Glu Glu Glu Val Ser Arg Met Ile Ser Gly Glu
385                 390                 395                 400

Ala Ser Tyr Pro Val Arg Phe Ser Leu Phe Ala Pro Arg Tyr Ala Ile
                405                 410                 415

Tyr Asp Asn Lys Ile Gly Tyr Cys His Thr Ser Asp Pro Val Tyr Pro
            420                 425                 430

Lys Ser Lys Thr Gly Glu Lys Arg Ala Leu Ser Asn Pro Gln Ser Met
        435                 440                 445

Gly Phe Ile Ser Val His Asp Leu Arg Lys Leu Leu Leu Met Glu Leu
    450                 455                 460

Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Asp Phe Leu Arg Lys
465                 470                 475                 480

Ala Asn Arg Ile Leu Asp Glu Thr Ala Glu Gly Lys Leu Gln Phe Ser
                485                 490                 495

Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro
            500                 505                 510

Lys Ser Lys Asp Arg Arg Glu Lys Ala Glu Thr Thr Leu Glu Lys Tyr
        515                 520                 525

Lys Gln Glu Ile Lys Gly Arg Lys Asp Lys Leu Asn Ser Gln Leu Leu
    530                 535                 540

Ser Ala Phe Asp Met Asp Gln Arg Gln Leu Pro Ser Arg Leu Leu Asp
545                 550                 555                 560

Glu Trp Met Asn Ile Arg Pro Ala Ser His Ser Val Lys Leu Arg Thr
```

```
                565                 570                 575

Tyr Val Lys Gln Leu Asn Glu Asp Cys Arg Leu Arg Leu Arg Lys Phe
            580                 585                 590

Arg Lys Asp Gly Asp Gly Lys Ala Arg Ala Ile Pro Leu Val Gly Glu
        595                 600                 605

Met Ala Thr Phe Leu Ser Gln Asp Ile Val Arg Met Ile Ile Ser Glu
    610                 615                 620

Glu Thr Lys Lys Leu Ile Thr Ser Ala Tyr Tyr Asn Glu Met Gln Arg
625                 630                 635                 640

Ser Leu Ala Gln Tyr Ala Gly Glu Glu Asn Arg Arg Gln Phe Arg Ala
                645                 650                 655

Ile Val Ala Glu Leu Arg Leu Leu Asp Pro Ser Ser Gly His Pro Phe
            660                 665                 670

Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Gly Phe Tyr
        675                 680                 685

Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Ile Phe Tyr
    690                 695                 700

Arg Pro Glu Gln Asp Glu Asn Thr Lys Arg Arg Ile Ser Val Phe Phe
705                 710                 715                 720

Val Pro Asp Gly Glu Ala Arg Lys Leu Leu Pro Leu Leu Ile Arg Arg
                725                 730                 735

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

Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Val Met
        755                 760                 765

Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe
```

```
    770                 775                 780

Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr
785                 790                 795                 800

Gly Leu Arg Arg Glu Leu Asn Ile His Gly Lys Ser Val Ser Tyr Ile
                805                 810                 815

Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu
            820                 825                 830

Lys Thr Val Arg Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro
        835                 840                 845

Val Pro Pro Asp Leu Ala Ala Asp Ile Lys Arg Ser Phe His Arg Ala
    850                 855                 860

Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg
865                 870                 875                 880

Leu Met Leu Met Ala Ile Asn Lys Met Met Thr Asp Arg Glu Glu Asp
                885                 890                 895

Ile Leu Pro Gly Leu Lys Asn Ile Asp Ser Ile Leu Asp Glu Glu Asn
            900                 905                 910

Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu
        915                 920                 925

Gly Gly Asp Asn Ser Leu Ser Leu Val Pro Ala Thr Ile Glu Ile Lys
    930                 935                 940

Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg
945                 950                 955                 960

Arg Val Pro Gly Leu Met Ser His Phe Pro Glu His Lys Ala Thr Leu
                965                 970                 975

Asp Glu Val Lys Thr Leu Leu Gly Glu Tyr Asp Arg Cys Arg Ile Lys
```

980 985 990

Ile Phe Asp Trp Ala Phe Ala Leu Glu Gly Ala Ile Met Ser Asp Arg
995 1000 1005

Asp Leu Lys Pro Tyr Leu His Glu Ser Ser Ser Arg Glu Gly Lys
1010 1015 1020

Ser Gly Glu His Ser Thr Leu Val Lys Met Leu Val Glu Lys Lys
1025 1030 1035

Gly Cys Leu Thr Pro Asp Glu Ser Gln Tyr Leu Ile Leu Ile Arg
1040 1045 1050

Asn Lys Ala Ala His Asn Gln Phe Pro Cys Ala Ala Glu Met Pro
1055 1060 1065

Leu Ile Tyr Arg Asp Val Ser Ala Lys Val Gly Ser Ile Glu Gly
1070 1075 1080

Ser Ser Ala Lys Asp Leu Pro Glu Gly Ser Ser Leu Val Asp Ser
1085 1090 1095

Leu Trp Lys Lys Tyr Glu Met Ile Ile Arg Lys Ile Leu Pro Ile
1100 1105 1110

Leu Asp Pro Glu Asn Arg Phe Phe Gly Lys Leu Leu Asn Asn Met
1115 1120 1125

Ser Gln Pro Ile Asn Asp Leu
1130 1135

<210> 71
<211> 1115
<212> PRT
<213> Bacteroides pyrogenes

<400> 71

Met Glu Ser Ile Lys Asn Ser Gln Lys Ser Thr Gly Lys Thr Leu Gln
1 5 10 15

```
Lys Asp Pro Pro Tyr Phe Gly Leu Tyr Leu Asn Met Ala Leu Leu Asn
            20                  25                  30

Val Arg Lys Val Glu Asn His Ile Arg Lys Trp Leu Gly Asp Val Ala
        35                  40                  45

Leu Leu Pro Glu Lys Ser Gly Phe His Ser Leu Leu Thr Thr Asp Asn
    50                  55                  60

Leu Ser Ser Ala Lys Trp Thr Arg Phe Tyr Tyr Lys Ser Arg Lys Phe
65                  70                  75                  80

Leu Pro Phe Leu Glu Met Phe Asp Ser Asp Lys Lys Ser Tyr Glu Asn
                85                  90                  95

Arg Arg Glu Thr Ala Glu Cys Leu Asp Thr Ile Asp Arg Gln Lys Ile
            100                 105                 110

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

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

Leu Ile Ile Ser Ser Glu Met His Arg Phe Ile Glu Asn Ala Tyr Ser
145                 150                 155                 160

Phe Ala Leu Gln Lys Thr Arg Ala Arg Phe Thr Gly Val Phe Val Glu
                165                 170                 175

Thr Asp Phe Leu Gln Ala Glu Glu Lys Gly Asp Asn Lys Lys Phe Phe
            180                 185                 190

Ala Ile Gly Gly Asn Glu Gly Ile Lys Leu Lys Asp Asn Ala Leu Ile
        195                 200                 205

Phe Leu Ile Cys Leu Phe Leu Asp Arg Glu Glu Ala Phe Lys Phe Leu
    210                 215                 220
```

```
Ser Arg Ala Thr Gly Phe Lys Ser Thr Lys Glu Lys Gly Phe Leu Ala
225                 230                 235                 240

Val Arg Glu Thr Phe Cys Ala Leu Cys Cys Arg Gln Pro His Glu Arg
                245                 250                 255

Leu Leu Ser Val Asn Pro Arg Glu Ala Leu Leu Met Asp Met Leu Asn
            260                 265                 270

Glu Leu Asn Arg Cys Pro Asp Ile Leu Phe Glu Met Leu Asp Glu Lys
        275                 280                 285

Asp Gln Lys Ser Phe Leu Pro Leu Leu Gly Glu Glu Glu Gln Ala His
    290                 295                 300

Ile Leu Glu Asn Ser Leu Asn Asp Glu Leu Cys Glu Ala Ile Asp Asp
305                 310                 315                 320

Pro Phe Glu Met Ile Ala Ser Leu Ser Lys Arg Val Arg Tyr Lys Asn
                325                 330                 335

Arg Phe Pro Tyr Leu Met Leu Arg Tyr Ile Glu Glu Lys Asn Leu Leu
            340                 345                 350

Pro Phe Ile Arg Phe Arg Ile Asp Leu Gly Cys Leu Glu Leu Ala Ser
        355                 360                 365

Tyr Pro Lys Lys Met Gly Glu Glu Asn Asn Tyr Glu Arg Ser Val Thr
    370                 375                 380

Asp His Ala Met Ala Phe Gly Arg Leu Thr Asp Phe His Asn Glu Asp
385                 390                 395                 400

Ala Val Leu Gln Gln Ile Thr Lys Gly Ile Thr Asp Glu Val Arg Phe
                405                 410                 415

Ser Leu Tyr Ala Pro Arg Tyr Ala Ile Tyr Asn Asn Lys Ile Gly Phe
            420                 425                 430
```

```
Val Arg Thr Ser Gly Ser Asp Lys Ile Ser Phe Pro Thr Leu Lys Lys
        435                 440                 445

Lys Gly Gly Glu Gly His Cys Val Ala Tyr Thr Leu Gln Asn Thr Lys
    450                 455                 460

Ser Phe Gly Phe Ile Ser Ile Tyr Asp Leu Arg Lys Ile Leu Leu Leu
465                 470                 475                 480

Ser Phe Leu Asp Lys Asp Lys Ala Lys Asn Ile Val Ser Gly Leu Leu
                485                 490                 495

Glu Gln Cys Glu Lys His Trp Lys Asp Leu Ser Glu Asn Leu Phe Asp
            500                 505                 510

Ala Ile Arg Thr Glu Leu Gln Lys Glu Phe Pro Val Pro Leu Ile Arg
        515                 520                 525

Tyr Thr Leu Pro Arg Ser Lys Gly Gly Lys Leu Val Ser Ser Lys Leu
    530                 535                 540

Ala Asp Lys Gln Glu Lys Tyr Glu Ser Glu Phe Glu Arg Arg Lys Glu
545                 550                 555                 560

Lys Leu Thr Glu Ile Leu Ser Glu Lys Asp Phe Asp Leu Ser Gln Ile
                565                 570                 575

Pro Arg Arg Met Ile Asp Glu Trp Leu Asn Val Leu Pro Thr Ser Arg
            580                 585                 590

Glu Lys Lys Leu Lys Gly Tyr Val Glu Thr Leu Lys Leu Asp Cys Arg
        595                 600                 605

Glu Arg Leu Arg Val Phe Glu Lys Arg Glu Lys Gly Glu His Pro Leu
    610                 615                 620

Pro Pro Arg Ile Gly Glu Met Ala Thr Asp Leu Ala Lys Asp Ile Ile
625                 630                 635                 640
```

```
Arg Met Val Ile Asp Gln Gly Val Lys Gln Arg Ile Thr Ser Ala Tyr
                645                 650                 655

Tyr Ser Glu Ile Gln Arg Cys Leu Ala Gln Tyr Ala Gly Asp Asp Asn
            660                 665                 670

Arg Arg His Leu Asp Ser Ile Ile Arg Glu Leu Arg Leu Lys Asp Thr
        675                 680                 685

Lys Asn Gly His Pro Phe Leu Gly Lys Val Leu Arg Pro Gly Leu Gly
    690                 695                 700

His Thr Glu Lys Leu Tyr Gln Arg Tyr Phe Glu Glu Lys Lys Glu Trp
705                 710                 715                 720

Leu Glu Ala Thr Phe Tyr Pro Ala Ala Ser Pro Lys Arg Val Pro Arg
                725                 730                 735

Phe Val Asn Pro Pro Thr Gly Lys Gln Lys Glu Leu Pro Leu Ile Ile
            740                 745                 750

Arg Asn Leu Met Lys Glu Arg Pro Glu Trp Arg Asp Trp Lys Gln Arg
        755                 760                 765

Lys Asn Ser His Pro Ile Asp Leu Pro Ser Gln Leu Phe Glu Asn Glu
    770                 775                 780

Ile Cys Arg Leu Leu Lys Asp Lys Ile Gly Lys Glu Pro Ser Gly Lys
785                 790                 795                 800

Leu Lys Trp Asn Glu Met Phe Lys Leu Tyr Trp Asp Lys Glu Phe Pro
                805                 810                 815

Asn Gly Met Gln Arg Phe Tyr Arg Cys Lys Arg Arg Val Glu Val Phe
            820                 825                 830

Asp Lys Val Val Glu Tyr Glu Tyr Ser Glu Glu Gly Gly Asn Tyr Lys
        835                 840                 845
```

```
Lys Tyr Tyr Glu Ala Leu Ile Asp Glu Val Val Arg Gln Lys Ile Ser
    850                 855                 860

Ser Ser Lys Glu Lys Ser Lys Leu Gln Val Glu Asp Leu Thr Leu Ser
865                 870                 875                 880

Val Arg Arg Val Phe Lys Arg Ala Ile Asn Glu Lys Glu Tyr Gln Leu
                885                 890                 895

Arg Leu Leu Cys Glu Asp Asp Arg Leu Leu Phe Met Ala Val Arg Asp
            900                 905                 910

Leu Tyr Asp Trp Lys Glu Ala Gln Leu Asp Leu Asp Lys Ile Asp Asn
        915                 920                 925

Met Leu Gly Glu Pro Val Ser Val Ser Gln Val Ile Gln Leu Glu Gly
    930                 935                 940

Gly Gln Pro Asp Ala Val Ile Lys Ala Glu Cys Lys Leu Lys Asp Val
945                 950                 955                 960

Ser Lys Leu Met Arg Tyr Cys Tyr Asp Gly Arg Val Lys Gly Leu Met
                965                 970                 975

Pro Tyr Phe Ala Asn His Glu Ala Thr Gln Glu Gln Val Glu Met Glu
            980                 985                 990

Leu Arg His Tyr Glu Asp His Arg  Arg Arg Val Phe Asn  Trp Val Phe
        995                 1000                1005

Ala Leu  Glu Lys Ser Val Leu  Lys Asn Glu Lys Leu  Arg Arg Phe
    1010                1015                1020

Tyr Glu  Glu Ser Gln Gly Gly  Cys Glu His Arg Arg  Cys Ile Asp
    1025                1030                1035

Ala Leu  Arg Lys Ala Ser Leu  Val Ser Glu Glu Glu  Tyr Glu Phe
    1040                1045                1050
```

```
Leu Val  His Ile Arg Asn Lys  Ser Ala His Asn Gln  Phe Pro Asp
    1055                1060                1065


Leu Glu  Ile Gly Lys Leu Pro  Pro Asn Val Thr Ser  Gly Phe Cys
    1070                1075                1080


Glu Cys  Ile Trp Ser Lys Tyr  Lys Ala Ile Ile Cys  Arg Ile Ile
    1085                1090                1095


Pro Phe  Ile Asp Pro Glu Arg  Arg Phe Phe Gly Lys  Leu Leu Glu
    1100                1105                1110


Gln Lys
    1115


<210>  72
<211>  953
<212>  PRT
<213>  Alistipes sp.

<400>  72

Met Ser Asn Glu Ile Gly Ala Phe Arg Glu His Gln Phe Ala Tyr Ala
1               5                   10                  15


Pro Gly Asn Glu Lys Gln Glu Glu Ala Thr Phe Ala Thr Tyr Phe Asn
            20                  25                  30


Leu Ala Leu Ser Asn Val Glu Gly Met Met Phe Gly Glu Val Glu Ser
        35                  40                  45


Asn Pro Asp Lys Ile Glu Lys Ser Leu Asp Thr Leu Pro Pro Ala Ile
    50                  55                  60


Leu Arg Gln Ile Ala Ser Phe Ile Trp Leu Ser Lys Glu Asp His Pro
65                  70                  75                  80


Asp Lys Ala Tyr Ser Thr Glu Glu Val Lys Val Ile Val Thr Asp Leu
                85                  90                  95
```

```
Val Arg Arg Leu Cys Phe Tyr Arg Asn Tyr Phe Ser His Cys Phe Tyr
            100                 105                 110

Leu Asp Thr Gln Tyr Phe Tyr Ser Asp Glu Leu Val Asp Thr Thr Ala
        115                 120                 125

Ile Gly Glu Lys Leu Pro Tyr Asn Phe His His Phe Ile Thr Asn Arg
    130                 135                 140

Leu Phe Arg Tyr Ser Leu Pro Glu Ile Thr Leu Phe Arg Trp Asn Glu
145                 150                 155                 160

Gly Glu Arg Lys Tyr Glu Ile Leu Arg Asp Gly Leu Ile Phe Phe Cys
                165                 170                 175

Cys Leu Phe Leu Lys Arg Gly Gln Ala Glu Arg Phe Leu Asn Glu Leu
            180                 185                 190

Arg Phe Phe Lys Arg Thr Asp Glu Glu Gly Arg Ile Lys Arg Thr Ile
        195                 200                 205

Phe Thr Lys Tyr Cys Thr Arg Glu Ser His Lys His Ile Gly Ile Glu
    210                 215                 220

Glu Gln Asp Phe Leu Ile Phe Gln Asp Ile Ile Gly Asp Leu Asn Arg
225                 230                 235                 240

Val Pro Lys Val Cys Asp Gly Val Val Asp Leu Ser Lys Glu Asn Glu
                245                 250                 255

Arg Tyr Ile Lys Asn Arg Glu Thr Ser Asn Glu Ser Asp Glu Asn Lys
            260                 265                 270

Ala Arg Tyr Arg Leu Leu Ile Arg Glu Lys Asp Lys Phe Pro Tyr Tyr
        275                 280                 285

Leu Met Arg Tyr Ile Val Asp Phe Gly Val Leu Pro Cys Ile Thr Phe
    290                 295                 300
```

```
Lys Gln Asn Asp Tyr Ser Thr Lys Glu Gly Arg Gly Gln Phe His Tyr
305                 310                 315                 320

Gln Asp Ala Ala Val Ala Gln Glu Glu Arg Cys Tyr Asn Phe Val Val
                325                 330                 335

Arg Asn Gly Asn Val Tyr Tyr Ser Tyr Met Pro Gln Ala Gln Asn Val
            340                 345                 350

Val Arg Ile Ser Glu Leu Gln Gly Thr Ile Ser Val Glu Glu Leu Arg
        355                 360                 365

Asn Met Val Tyr Ala Ser Ile Asn Gly Lys Asp Val Asn Lys Ser Val
    370                 375                 380

Glu Gln Tyr Leu Tyr His Leu His Leu Leu Tyr Glu Lys Ile Leu Thr
385                 390                 395                 400

Ile Ser Gly Gln Thr Ile Lys Glu Gly Arg Val Asp Val Glu Asp Tyr
                405                 410                 415

Arg Pro Leu Leu Asp Lys Leu Leu Leu Arg Pro Ala Ser Asn Gly Glu
            420                 425                 430

Glu Leu Arg Arg Glu Leu Arg Lys Leu Leu Pro Lys Arg Val Cys Asp
        435                 440                 445

Leu Leu Ser Asn Arg Phe Asp Cys Ser Glu Gly Val Ser Ala Val Glu
    450                 455                 460

Lys Arg Leu Lys Ala Ile Leu Leu Arg His Glu Gln Leu Leu Leu Ser
465                 470                 475                 480

Gln Asn Pro Ala Leu His Ile Asp Lys Ile Lys Ser Val Ile Asp Tyr
                485                 490                 495

Leu Tyr Leu Phe Phe Ser Asp Asp Glu Lys Phe Arg Gln Gln Pro Thr
            500                 505                 510
```

```
Glu Lys Ala His Arg Gly Leu Lys Asp Glu Glu Phe Gln Met Tyr His
        515                 520                 525

Tyr Leu Val Gly Asp Tyr Asp Ser His Pro Leu Ala Leu Trp Lys Glu
    530                 535                 540

Leu Glu Ala Ser Gly Arg Leu Lys Pro Glu Met Arg Lys Leu Thr Ser
545                 550                 555                 560

Ala Thr Ser Leu His Gly Leu Tyr Met Leu Cys Leu Lys Gly Thr Val
                565                 570                 575

Glu Trp Cys Arg Lys Gln Leu Met Ser Ile Gly Lys Gly Thr Ala Lys
            580                 585                 590

Val Glu Ala Ile Ala Asp Arg Val Gly Leu Lys Leu Tyr Asp Lys Leu
        595                 600                 605

Lys Glu Tyr Thr Pro Glu Gln Leu Glu Arg Glu Val Lys Leu Val Val
    610                 615                 620

Met His Gly Tyr Ala Ala Ala Ala Thr Pro Lys Pro Lys Ala Gln Ala
625                 630                 635                 640

Ala Ile Pro Ser Lys Leu Thr Glu Leu Arg Phe Tyr Ser Phe Leu Gly
                645                 650                 655

Lys Arg Glu Met Ser Phe Ala Ala Phe Ile Arg Gln Asp Lys Lys Ala
            660                 665                 670

Gln Lys Leu Trp Leu Arg Asn Phe Tyr Thr Val Glu Asn Ile Lys Thr
        675                 680                 685

Leu Gln Lys Arg Gln Ala Ala Ala Asp Ala Ala Cys Lys Lys Leu Tyr
    690                 695                 700

Asn Leu Val Gly Glu Val Glu Arg Val His Thr Asn Asp Lys Val Leu
705                 710                 715                 720
```

```
Val Leu Val Ala Gln Arg Tyr Arg Glu Arg Leu Leu Asn Val Gly Ser
                725                 730                 735

Lys Cys Ala Val Thr Leu Asp Asn Pro Glu Arg Gln Gln Lys Leu Ala
            740                 745                 750

Asp Val Tyr Glu Val Gln Asn Ala Trp Leu Ser Ile Arg Phe Asp Asp
        755                 760                 765

Leu Asp Phe Thr Leu Thr His Val Asn Leu Ser Asn Leu Arg Lys Ala
    770                 775                 780

Tyr Asn Leu Ile Pro Arg Lys His Ile Leu Ala Phe Lys Glu Tyr Leu
785                 790                 795                 800

Asp Asn Arg Val Lys Gln Lys Leu Cys Glu Glu Cys Arg Asn Val Arg
                805                 810                 815

Arg Lys Glu Asp Leu Cys Thr Cys Cys Ser Pro Arg Tyr Ser Asn Leu
            820                 825                 830

Thr Ser Trp Leu Lys Glu Asn His Ser Glu Ser Ser Ile Glu Arg Glu
        835                 840                 845

Ala Ala Thr Met Met Leu Leu Asp Val Glu Arg Lys Leu Leu Ser Phe
    850                 855                 860

Leu Leu Asp Glu Arg Arg Lys Ala Ile Ile Glu Tyr Gly Lys Phe Ile
865                 870                 875                 880

Pro Phe Ser Ala Leu Val Lys Glu Cys Arg Leu Ala Asp Ala Gly Leu
                885                 890                 895

Cys Gly Ile Arg Asn Asp Val Leu His Asp Asn Val Ile Ser Tyr Ala
            900                 905                 910

Asp Ala Ile Gly Lys Leu Ser Ala Tyr Phe Pro Lys Glu Ala Ser Glu
        915                 920                 925
```

Ala Val Glu Tyr Ile Arg Arg Thr Lys Glu Val Arg Glu Gln Arg Arg
    930   935   940

Glu Glu Leu Met Ala Asn Ser Ser Gln
945   950

<210> 73
<211> 1322
<212> PRT
<213> Prevotella sp.

<400> 73

Met Ser Lys Glu Cys Lys Lys Gln Arg Gln Glu Lys Lys Arg Arg Leu
1   5   10   15

Gln Lys Ala Asn Phe Ser Ile Ser Leu Thr Gly Lys His Val Phe Gly
            20   25   30

Ala Tyr Phe Asn Met Ala Arg Thr Asn Phe Val Lys Thr Ile Asn Tyr
        35   40   45

Ile Leu Pro Ile Ala Gly Val Arg Gly Asn Tyr Ser Glu Asn Gln Ile
    50   55   60

Asn Lys Met Leu His Ala Leu Phe Leu Ile Gln Ala Gly Arg Asn Glu
65   70   75   80

Glu Leu Thr Thr Glu Gln Lys Gln Trp Glu Lys Lys Leu Arg Leu Asn
                85   90   95

Pro Glu Gln Gln Thr Lys Phe Gln Lys Leu Leu Phe Lys His Phe Pro
            100   105   110

Val Leu Gly Pro Met Met Ala Asp Val Ala Asp His Lys Ala Tyr Leu
        115   120   125

Asn Lys Lys Lys Ser Thr Val Gln Thr Glu Asp Glu Thr Phe Ala Met
    130   135   140

```
Leu Lys Gly Val Ser Leu Ala Asp Cys Leu Asp Ile Ile Cys Leu Met
145                 150                 155                 160


Ala Asp Thr Leu Thr Glu Cys Arg Asn Phe Tyr Thr His Lys Asp Pro
                165                 170                 175


Tyr Asn Lys Pro Ser Gln Leu Ala Asp Gln Tyr Leu His Gln Glu Met
            180                 185                 190


Ile Ala Lys Lys Leu Asp Lys Val Val Val Ala Ser Arg Arg Ile Leu
        195                 200                 205


Lys Asp Arg Glu Gly Leu Ser Val Asn Glu Val Glu Phe Leu Thr Gly
    210                 215                 220


Ile Asp His Leu His Gln Glu Val Leu Lys Asp Glu Phe Gly Asn Ala
225                 230                 235                 240


Lys Val Lys Asp Gly Lys Val Met Lys Thr Phe Val Glu Tyr Asp Asp
                245                 250                 255


Phe Tyr Phe Lys Ile Ser Gly Lys Arg Leu Val Asn Gly Tyr Thr Val
            260                 265                 270


Thr Thr Lys Asp Asp Lys Pro Val Asn Val Asn Thr Met Leu Pro Ala
        275                 280                 285


Leu Ser Asp Phe Gly Leu Leu Tyr Phe Cys Val Leu Phe Leu Ser Lys
    290                 295                 300


Pro Tyr Ala Lys Leu Phe Ile Asp Glu Val Arg Leu Phe Glu Tyr Ser
305                 310                 315                 320


Pro Phe Asp Asp Lys Glu Asn Met Ile Met Ser Glu Met Leu Ser Ile
                325                 330                 335


Tyr Arg Ile Arg Thr Pro Arg Leu His Lys Ile Asp Ser His Asp Ser
            340                 345                 350
```

```
Lys Ala Thr Leu Ala Met Asp Ile Phe Gly Glu Leu Arg Arg Cys Pro
        355                 360                 365


Met Glu Leu Tyr Asn Leu Leu Asp Lys Asn Ala Gly Gln Pro Phe Phe
    370                 375                 380


His Asp Glu Val Lys His Pro Asn Ser His Thr Pro Asp Val Ser Lys
385                 390                 395                 400


Arg Leu Arg Tyr Asp Asp Arg Phe Pro Thr Leu Ala Leu Arg Tyr Ile
                405                 410                 415


Asp Glu Thr Glu Leu Phe Lys Arg Ile Arg Phe Gln Leu Gln Leu Gly
            420                 425                 430


Ser Phe Arg Tyr Lys Phe Tyr Asp Lys Glu Asn Cys Ile Asp Gly Arg
        435                 440                 445


Val Arg Val Arg Arg Ile Gln Lys Glu Ile Asn Gly Tyr Gly Arg Met
    450                 455                 460


Gln Glu Val Ala Asp Lys Arg Met Asp Lys Trp Gly Asp Leu Ile Gln
465                 470                 475                 480


Lys Arg Glu Glu Arg Ser Val Lys Leu Glu His Glu Glu Leu Tyr Ile
                485                 490                 495


Asn Leu Asp Gln Phe Leu Glu Asp Thr Ala Asp Ser Thr Pro Tyr Val
            500                 505                 510


Thr Asp Arg Arg Pro Ala Tyr Asn Ile His Ala Asn Arg Ile Gly Leu
        515                 520                 525


Tyr Trp Glu Asp Ser Gln Asn Pro Lys Gln Tyr Lys Val Phe Asp Glu
    530                 535                 540


Asn Gly Met Tyr Ile Pro Glu Leu Val Val Thr Glu Asp Lys Lys Ala
545                 550                 555                 560
```

```
Pro Ile Lys Met Pro Ala Pro Arg Cys Ala Leu Ser Val Tyr Asp Leu
                565                 570                 575

Pro Ala Met Leu Phe Tyr Glu Tyr Leu Arg Glu Gln Gln Asp Asn Glu
            580                 585                 590

Phe Pro Ser Ala Glu Gln Val Ile Ile Glu Tyr Glu Asp Asp Tyr Arg
        595                 600                 605

Lys Phe Phe Lys Ala Val Ala Glu Gly Lys Leu Lys Pro Phe Lys Arg
    610                 615                 620

Pro Lys Glu Phe Arg Asp Phe Leu Lys Lys Glu Tyr Pro Lys Leu Arg
625                 630                 635                 640

Met Ala Asp Ile Pro Lys Lys Leu Gln Leu Phe Leu Cys Ser His Gly
                645                 650                 655

Leu Cys Tyr Asn Asn Lys Pro Glu Thr Val Tyr Glu Arg Leu Asp Arg
            660                 665                 670

Leu Thr Leu Gln His Leu Glu Glu Arg Glu Leu His Ile Gln Asn Arg
        675                 680                 685

Leu Glu His Tyr Gln Lys Asp Arg Asp Met Ile Gly Asn Lys Asp Asn
    690                 695                 700

Gln Tyr Gly Lys Lys Ser Phe Ser Asp Val Arg His Gly Ala Leu Ala
705                 710                 715                 720

Arg Tyr Leu Ala Gln Ser Met Met Glu Trp Gln Pro Thr Lys Leu Lys
                725                 730                 735

Asp Lys Glu Lys Gly His Asp Lys Leu Thr Gly Leu Asn Tyr Asn Val
            740                 745                 750

Leu Thr Ala Tyr Leu Ala Thr Tyr Gly His Pro Gln Val Pro Glu Glu
        755                 760                 765
```

```
Gly Phe Thr Pro Arg Thr Leu Glu Gln Val Leu Ile Asn Ala His Leu
    770                 775                 780

Ile Gly Gly Ser Asn Pro His Pro Phe Ile Asn Lys Val Leu Ala Leu
785                 790                 795                 800

Gly Asn Arg Asn Ile Glu Glu Leu Tyr Leu His Tyr Leu Glu Glu Glu
                805                 810                 815

Leu Lys His Ile Arg Ser Arg Ile Gln Ser Leu Ser Ser Asn Pro Ser
            820                 825                 830

Asp Lys Ala Leu Ser Ala Leu Pro Phe Ile His His Asp Arg Met Arg
        835                 840                 845

Tyr His Glu Arg Thr Ser Glu Glu Met Met Ala Leu Ala Ala Arg Tyr
    850                 855                 860

Thr Thr Ile Gln Leu Pro Asp Gly Leu Phe Thr Pro Tyr Ile Leu Glu
865                 870                 875                 880

Ile Leu Gln Lys His Tyr Thr Glu Asn Ser Asp Leu Gln Asn Ala Leu
                885                 890                 895

Ser Gln Asp Val Pro Val Lys Leu Asn Pro Thr Cys Asn Ala Ala Tyr
            900                 905                 910

Leu Ile Thr Leu Phe Tyr Gln Thr Val Leu Lys Asp Asn Ala Gln Pro
        915                 920                 925

Phe Tyr Leu Ser Asp Lys Thr Tyr Thr Arg Asn Lys Asp Gly Glu Lys
    930                 935                 940

Ala Glu Ser Phe Ser Phe Lys Arg Ala Tyr Glu Leu Phe Ser Val Leu
945                 950                 955                 960

Asn Asn Asn Lys Lys Asp Thr Phe Pro Phe Glu Met Ile Pro Leu Phe
                965                 970                 975
```

```
Leu Thr Ser Asp Glu Ile Gln Glu Arg Leu Ser Ala Lys Leu Leu Asp
            980                 985                 990


Gly Asp Gly Asn Pro Val Pro Glu  Val Gly Glu Lys Gly  Lys Pro Ala
        995                 1000                1005


Thr Asp  Ser Gln Gly Asn Thr  Ile Trp Lys Arg Arg  Ile Tyr Ser
    1010                1015                1020


Glu Val  Asp Asp Tyr Ala Glu  Lys Leu Thr Asp Arg  Asp Met Lys
    1025                1030                1035


Ile Ser  Phe Lys Gly Glu Trp  Glu Lys Leu Pro Arg  Trp Lys Gln
    1040                1045                1050


Asp Lys  Ile Ile Lys Arg Arg  Asp Glu Thr Arg Arg  Gln Met Arg
    1055                1060                1065


Asp Glu  Leu Leu Gln Arg Met  Pro Arg Tyr Ile Arg  Asp Ile Lys
    1070                1075                1080


Asp Asn  Glu Arg Thr Leu Arg  Arg Tyr Lys Thr Gln  Asp Met Val
    1085                1090                1095


Leu Phe  Leu Leu Ala Glu Lys  Met Phe Thr Asn Ile  Ile Ser Glu
    1100                1105                1110


Gln Ser  Ser Glu Phe Asn Trp  Lys Gln Met Arg Leu  Ser Lys Val
    1115                1120                1125


Cys Asn  Glu Ala Phe Leu Arg  Gln Thr Leu Thr Phe  Arg Val Pro
    1130                1135                1140


Val Thr  Val Gly Glu Thr Thr  Ile Tyr Val Glu Gln  Glu Asn Met
    1145                1150                1155


Ser Leu  Lys Asn Tyr Gly Glu  Phe Tyr Arg Phe Leu  Thr Asp Asp
    1160                1165                1170
```

```
Arg Leu  Met Ser Leu Leu Asn  Asn Ile Val Glu Thr  Leu Lys Pro
    1175                1180                1185


Asn Glu  Asn Gly Asp Leu Val  Ile Arg His Thr Asp  Leu Met Ser
    1190                1195                1200


Glu Leu  Ala Ala Tyr Asp Gln  Tyr Arg Ser Thr Ile  Phe Met Leu
    1205                1210                1215


Ile Gln  Ser Ile Glu Asn Leu  Ile Ile Thr Asn Asn  Ala Val Leu
    1220                1225                1230


Asp Asp  Pro Asp Ala Asp Gly  Phe Trp Val Arg Glu  Asp Leu Pro
    1235                1240                1245


Lys Arg  Asn Asn Phe Ala Ser  Leu Leu Glu Leu Ile  Asn Gln Leu
    1250                1255                1260


Asn Asn  Val Glu Leu Thr Asp  Asp Glu Arg Lys Leu  Leu Val Ala
    1265                1270                1275


Ile Arg  Asn Ala Phe Ser His  Asn Ser Tyr Asn Ile  Asp Phe Ser
    1280                1285                1290


Leu Ile  Lys Asp Val Lys His  Leu Pro Glu Val Ala  Lys Gly Ile
    1295                1300                1305


Leu Gln  His Leu Gln Ser Met  Leu Gly Val Glu Ile  Thr Lys
    1310                1315                1320


<210>  74
<211>  1322
<212>  PRT
<213>  Prevotella sp.

<400>  74

Met Ser Lys Glu Cys Lys Lys Gln Arg Gln Glu Lys Lys Arg Arg Leu
1               5                   10                  15


Gln Lys Ala Asn Phe Ser Ile Ser Leu Thr Gly Lys His Val Phe Gly
```

```
            20                  25                  30

Ala Tyr Phe Asn Met Ala Arg Thr Asn Phe Val Lys Thr Ile Asn Tyr
        35                  40                  45

Ile Leu Pro Ile Ala Gly Val Arg Gly Asn Tyr Ser Glu Asn Gln Ile
    50                  55                  60

Asn Lys Met Leu His Ala Leu Phe Leu Ile Gln Ala Gly Arg Asn Glu
65                  70                  75                  80

Glu Leu Thr Thr Glu Gln Lys Gln Trp Glu Lys Lys Leu Arg Leu Asn
                85                  90                  95

Pro Glu Gln Gln Thr Lys Phe Gln Lys Leu Leu Phe Lys His Phe Pro
            100                 105                 110

Val Leu Gly Pro Met Met Ala Asp Val Ala Asp His Lys Ala Tyr Leu
        115                 120                 125

Asn Lys Lys Lys Ser Thr Val Gln Thr Glu Asp Glu Thr Phe Ala Met
    130                 135                 140

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

Ala Asp Thr Leu Thr Glu Cys Arg Asn Phe Tyr Thr His Lys Asp Pro
                165                 170                 175

Tyr Asn Lys Pro Ser Gln Leu Ala Asp Gln Tyr Leu His Gln Glu Met
            180                 185                 190

Ile Ala Lys Lys Leu Asp Lys Val Val Val Ala Ser Arg Arg Ile Leu
        195                 200                 205

Lys Asp Arg Glu Gly Leu Ser Val Asn Glu Val Glu Phe Leu Thr Gly
    210                 215                 220

Ile Asp His Leu His Gln Glu Val Leu Lys Asp Glu Phe Gly Asn Ala
```

225 230 235 240

Lys Val Lys Asp Gly Lys Val Met Lys Thr Phe Val Glu Tyr Asp Asp
245 250 255

Phe Tyr Phe Lys Ile Ser Gly Lys Arg Leu Val Asn Gly Tyr Thr Val
260 265 270

Thr Thr Lys Asp Asp Lys Pro Val Asn Val Asn Thr Met Leu Pro Ala
275 280 285

Leu Ser Asp Phe Gly Leu Leu Tyr Phe Cys Val Leu Phe Leu Ser Lys
290 295 300

Pro Tyr Ala Lys Leu Phe Ile Asp Glu Val Arg Leu Phe Glu Tyr Ser
305 310 315 320

Pro Phe Asp Asp Lys Glu Asn Met Ile Met Ser Glu Met Leu Ser Ile
325 330 335

Tyr Arg Ile Arg Thr Pro Arg Leu His Lys Ile Asp Ser His Asp Ser
340 345 350

Lys Ala Thr Leu Ala Met Asp Ile Phe Gly Glu Leu Arg Arg Cys Pro
355 360 365

Met Glu Leu Tyr Asn Leu Leu Asp Lys Asn Ala Gly Gln Pro Phe Phe
370 375 380

His Asp Glu Val Lys His Pro Asn Ser His Thr Pro Asp Val Ser Lys
385 390 395 400

Arg Leu Arg Tyr Asp Asp Arg Phe Pro Thr Leu Ala Leu Arg Tyr Ile
405 410 415

Asp Glu Thr Glu Leu Phe Lys Arg Ile Arg Phe Gln Leu Gln Leu Gly
420 425 430

Ser Phe Arg Tyr Lys Phe Tyr Asp Lys Glu Asn Cys Ile Asp Gly Arg

```
        435                 440                 445

Val Arg Val Arg Arg Ile Gln Lys Glu Ile Asn Gly Tyr Gly Arg Met
    450                 455                 460

Gln Glu Val Ala Asp Lys Arg Met Asp Lys Trp Gly Asp Leu Ile Gln
465                 470                 475                 480

Lys Arg Glu Glu Arg Ser Val Lys Leu Glu His Glu Glu Leu Tyr Ile
                485                 490                 495

Asn Leu Asp Gln Phe Leu Glu Asp Thr Ala Asp Ser Thr Pro Tyr Val
            500                 505                 510

Thr Asp Arg Arg Pro Ala Tyr Asn Ile His Ala Asn Arg Ile Gly Leu
        515                 520                 525

Tyr Trp Glu Asp Ser Gln Asn Pro Lys Gln Tyr Lys Val Phe Asp Glu
    530                 535                 540

Asn Gly Met Tyr Ile Pro Glu Leu Val Val Thr Glu Asp Lys Lys Ala
545                 550                 555                 560

Pro Ile Lys Met Pro Ala Pro Arg Cys Ala Leu Ser Val Tyr Asp Leu
                565                 570                 575

Pro Ala Met Leu Phe Tyr Glu Tyr Leu Arg Glu Gln Gln Asp Asn Glu
            580                 585                 590

Phe Pro Ser Ala Glu Gln Val Ile Ile Glu Tyr Glu Asp Asp Tyr Arg
        595                 600                 605

Lys Phe Phe Lys Ala Val Ala Glu Gly Lys Leu Lys Pro Phe Lys Arg
    610                 615                 620

Pro Lys Glu Phe Arg Asp Phe Leu Lys Lys Glu Tyr Pro Lys Leu Arg
625                 630                 635                 640

Met Ala Asp Ile Pro Lys Lys Leu Gln Leu Phe Leu Cys Ser His Gly
```

645 650 655

Leu Cys Tyr Asn Asn Lys Pro Glu Thr Val Tyr Glu Arg Leu Asp Arg
660 665 670

Leu Thr Leu Gln His Leu Glu Glu Arg Glu Leu His Ile Gln Asn Arg
675 680 685

Leu Glu His Tyr Gln Lys Asp Arg Asp Met Ile Gly Asn Lys Asp Asn
690 695 700

Gln Tyr Gly Lys Lys Ser Phe Ser Asp Val Arg His Gly Ala Leu Ala
705 710 715 720

Arg Tyr Leu Ala Gln Ser Met Met Glu Trp Gln Pro Thr Lys Leu Lys
725 730 735

Asp Lys Glu Lys Gly His Asp Lys Leu Thr Gly Leu Asn Tyr Asn Val
740 745 750

Leu Thr Ala Tyr Leu Ala Thr Tyr Gly His Pro Gln Val Pro Glu Glu
755 760 765

Gly Phe Thr Pro Arg Thr Leu Glu Gln Val Leu Ile Asn Ala His Leu
770 775 780

Ile Gly Gly Ser Asn Pro His Pro Phe Ile Asn Lys Val Leu Ala Leu
785 790 795 800

Gly Asn Arg Asn Ile Glu Glu Leu Tyr Leu His Tyr Leu Glu Glu Glu
805 810 815

Leu Lys His Ile Arg Ser Arg Ile Gln Ser Leu Ser Ser Asn Pro Ser
820 825 830

Asp Lys Ala Leu Ser Ala Leu Pro Phe Ile His His Asp Arg Met Arg
835 840 845

Tyr His Glu Arg Thr Ser Glu Glu Met Met Ala Leu Ala Ala Arg Tyr

```
    850                 855                 860

Thr Thr Ile Gln Leu Pro Asp Gly Leu Phe Thr Pro Tyr Ile Leu Glu
865                 870                 875                 880

Ile Leu Gln Lys His Tyr Thr Glu Asn Ser Asp Leu Gln Asn Ala Leu
                885                 890                 895

Ser Gln Asp Val Pro Val Lys Leu Asn Pro Thr Cys Asn Ala Ala Tyr
            900                 905                 910

Leu Ile Thr Leu Phe Tyr Gln Thr Val Leu Lys Asp Asn Ala Gln Pro
        915                 920                 925

Phe Tyr Leu Ser Asp Lys Thr Tyr Thr Arg Asn Lys Asp Gly Glu Lys
    930                 935                 940

Ala Glu Ser Phe Ser Phe Lys Arg Ala Tyr Glu Leu Phe Ser Val Leu
945                 950                 955                 960

Asn Asn Asn Lys Lys Asp Thr Phe Pro Phe Glu Met Ile Pro Leu Phe
                965                 970                 975

Leu Thr Ser Asp Glu Ile Gln Glu Arg Leu Ser Ala Lys Leu Leu Asp
            980                 985                 990

Gly Asp Gly Asn Pro Val Pro Glu  Val Gly Glu Lys Gly  Lys Pro Ala
        995                 1000                1005

Thr Asp  Ser Gln Gly Asn Thr  Ile Trp Lys Arg Arg  Ile Tyr Ser
    1010                1015                1020

Glu Val  Asp Asp Tyr Ala Glu  Lys Leu Thr Asp Arg  Asp Met Lys
    1025                1030                1035

Ile Ser  Phe Lys Gly Glu Trp  Glu Lys Leu Pro Arg  Trp Lys Gln
    1040                1045                1050

Asp Lys  Ile Ile Lys Arg Arg  Asp Glu Thr Arg Arg  Gln Met Arg
```

```
    1055                1060                1065

Asp Glu  Leu Leu Gln Arg Met  Pro Arg Tyr Ile Arg  Asp Ile Lys
    1070                1075                1080

Asp Asn  Glu Arg Thr Leu Arg  Arg Tyr Lys Thr Gln  Asp Met Val
    1085                1090                1095

Leu Phe  Leu Leu Ala Glu Lys  Met Phe Thr Asn Ile  Ile Ser Glu
    1100                1105                1110

Gln Ser  Ser Glu Phe Asn Trp  Lys Gln Met Arg Leu  Ser Lys Val
    1115                1120                1125

Cys Asn  Glu Ala Phe Leu Arg  Gln Thr Leu Thr Phe  Arg Val Pro
    1130                1135                1140

Val Thr  Val Gly Glu Thr Thr  Ile Tyr Val Glu Gln  Glu Asn Met
    1145                1150                1155

Ser Leu  Lys Asn Tyr Gly Glu  Phe Tyr Arg Phe Leu  Thr Asp Asp
    1160                1165                1170

Arg Leu  Met Ser Leu Leu Asn  Asn Ile Val Glu Thr  Leu Lys Pro
    1175                1180                1185

Asn Glu  Asn Gly Asp Leu Val  Ile Arg His Thr Asp  Leu Met Ser
    1190                1195                1200

Glu Leu  Ala Ala Tyr Asp Gln  Tyr Arg Ser Thr Ile  Phe Met Leu
    1205                1210                1215

Ile Gln  Ser Ile Glu Asn Leu  Ile Ile Thr Asn Asn  Ala Val Leu
    1220                1225                1230

Asp Asp  Pro Asp Ala Asp Gly  Phe Trp Val Arg Glu  Asp Leu Pro
    1235                1240                1245

Lys Arg  Asn Asn Phe Ala Ser  Leu Leu Glu Leu Ile  Asn Gln Leu
```

1250 1255 1260

Asn Asn Val Glu Leu Thr Asp Asp Glu Arg Lys Leu Leu Val Ala
1265 1270 1275

Ile Arg Asn Ala Phe Ser His Asn Ser Tyr Asn Ile Asp Phe Ser
1280 1285 1290

Leu Ile Lys Asp Val Lys His Leu Pro Glu Val Ala Lys Gly Ile
1295 1300 1305

Leu Gln His Leu Gln Ser Met Leu Gly Val Glu Ile Thr Lys
1310 1315 1320

<210> 75
<211> 1095
<212> PRT
<213> Riemerella anatipestifer

<400> 75

Met Glu Lys Pro Leu Leu Pro Asn Val Tyr Thr Leu Lys His Lys Phe
1 5 10 15

Phe Trp Gly Ala Phe Leu Asn Ile Ala Arg His Asn Ala Phe Ile Thr
20 25 30

Ile Cys His Ile Asn Glu Gln Leu Gly Leu Lys Thr Pro Ser Asn Asp
35 40 45

Asp Lys Ile Val Asp Val Val Cys Glu Thr Trp Asn Asn Ile Leu Asn
50 55 60

Asn Asp His Asp Leu Leu Lys Lys Ser Gln Leu Thr Glu Leu Ile Leu
65 70 75 80

Lys His Phe Pro Phe Leu Thr Ala Met Cys Tyr His Pro Pro Lys Lys
85 90 95

Glu Gly Lys Lys Lys Gly His Gln Lys Glu Gln Gln Lys Glu Lys Glu
100 105 110

```
Ser Glu Ala Gln Ser Gln Ala Glu Ala Leu Asn Pro Ser Lys Leu Ile
        115                 120                 125

Glu Ala Leu Glu Ile Leu Val Asn Gln Leu His Ser Leu Arg Asn Tyr
    130                 135                 140

Tyr Ser His Tyr Lys His Lys Lys Pro Asp Ala Glu Lys Asp Ile Phe
145                 150                 155                 160

Lys His Leu Tyr Lys Ala Phe Asp Ala Ser Leu Arg Met Val Lys Glu
                165                 170                 175

Asp Tyr Lys Ala His Phe Thr Val Asn Leu Thr Arg Asp Phe Ala His
            180                 185                 190

Leu Asn Arg Lys Gly Lys Asn Lys Gln Asp Asn Pro Asp Phe Asn Arg
        195                 200                 205

Tyr Arg Phe Glu Lys Asp Gly Phe Phe Thr Glu Ser Gly Leu Leu Phe
    210                 215                 220

Phe Thr Asn Leu Phe Leu Asp Lys Arg Asp Ala Tyr Trp Met Leu Lys
225                 230                 235                 240

Lys Val Ser Gly Phe Lys Ala Ser His Lys Gln Arg Glu Lys Met Thr
                245                 250                 255

Thr Glu Val Phe Cys Arg Ser Arg Ile Leu Leu Pro Lys Leu Arg Leu
            260                 265                 270

Glu Ser Arg Tyr Asp His Asn Gln Met Leu Leu Asp Met Leu Ser Glu
        275                 280                 285

Leu Ser Arg Cys Pro Lys Leu Leu Tyr Glu Lys Leu Ser Glu Glu Asn
    290                 295                 300

Lys Lys His Phe Gln Val Glu Ala Asp Gly Phe Leu Asp Glu Ile Glu
305                 310                 315                 320
```

```
Glu Glu Gln Asn Pro Phe Lys Asp Thr Leu Ile Arg His Gln Asp Arg
                325                 330                 335

Phe Pro Tyr Phe Ala Leu Arg Tyr Leu Asp Leu Asn Glu Ser Phe Lys
            340                 345                 350

Ser Ile Arg Phe Gln Val Asp Leu Gly Thr Tyr His Tyr Cys Ile Tyr
        355                 360                 365

Asp Lys Lys Ile Gly Asp Glu Gln Glu Lys Arg His Leu Thr Arg Thr
    370                 375                 380

Leu Leu Ser Phe Gly Arg Leu Gln Asp Phe Thr Glu Ile Asn Arg Pro
385                 390                 395                 400

Gln Glu Trp Lys Ala Leu Thr Lys Asp Leu Asp Tyr Lys Glu Thr Ser
                405                 410                 415

Asn Gln Pro Phe Ile Ser Lys Thr Thr Pro His Tyr His Ile Thr Asp
            420                 425                 430

Asn Lys Ile Gly Phe Arg Leu Gly Thr Ser Lys Glu Leu Tyr Pro Ser
        435                 440                 445

Leu Glu Ile Lys Asp Gly Ala Asn Arg Ile Ala Lys Tyr Pro Tyr Asn
    450                 455                 460

Ser Gly Phe Val Ala His Ala Phe Ile Ser Val His Glu Leu Leu Pro
465                 470                 475                 480

Leu Met Phe Tyr Gln His Leu Thr Gly Lys Ser Glu Asp Leu Leu Lys
                485                 490                 495

Glu Thr Val Arg His Ile Gln Arg Ile Tyr Lys Asp Phe Glu Glu Glu
            500                 505                 510

Arg Ile Asn Thr Ile Glu Asp Leu Glu Lys Ala Asn Gln Gly Arg Leu
        515                 520                 525
```

```
Pro Leu Gly Ala Phe Pro Lys Gln Met Leu Gly Leu Leu Gln Asn Lys
    530                 535                 540

Gln Pro Asp Leu Ser Glu Lys Ala Lys Ile Lys Ile Glu Lys Leu Ile
545                 550                 555                 560

Ala Glu Thr Lys Leu Leu Ser His Arg Leu Asn Thr Lys Leu Lys Ser
                565                 570                 575

Ser Pro Lys Leu Gly Lys Arg Arg Glu Lys Leu Ile Lys Thr Gly Val
            580                 585                 590

Leu Ala Asp Trp Leu Val Lys Asp Phe Met Arg Phe Gln Pro Val Ala
        595                 600                 605

Tyr Asp Ala Gln Asn Gln Pro Ile Lys Ser Ser Lys Ala Asn Ser Thr
    610                 615                 620

Glu Phe Trp Phe Ile Arg Arg Ala Leu Ala Leu Tyr Gly Gly Glu Lys
625                 630                 635                 640

Asn Arg Leu Glu Gly Tyr Phe Lys Gln Thr Asn Leu Ile Gly Asn Thr
                645                 650                 655

Asn Pro His Pro Phe Leu Asn Lys Phe Asn Trp Lys Ala Cys Arg Asn
            660                 665                 670

Leu Val Asp Phe Tyr Gln Gln Tyr Leu Glu Gln Arg Glu Lys Phe Leu
        675                 680                 685

Glu Ala Ile Lys Asn Gln Pro Trp Glu Pro Tyr Gln Tyr Cys Leu Leu
    690                 695                 700

Leu Lys Ile Pro Lys Glu Asn Arg Lys Asn Leu Val Lys Gly Trp Glu
705                 710                 715                 720

Gln Gly Gly Ile Ser Leu Pro Arg Gly Leu Phe Thr Glu Ala Ile Arg
                725                 730                 735
```

```
Glu Thr Leu Ser Glu Asp Leu Met Leu Ser Lys Pro Ile Arg Lys Glu
            740                 745                 750

Ile Lys Lys His Gly Arg Val Gly Phe Ile Ser Arg Ala Ile Thr Leu
        755                 760                 765

Tyr Phe Lys Glu Lys Tyr Gln Asp Lys His Gln Ser Phe Tyr Asn Leu
    770                 775                 780

Ser Tyr Lys Leu Glu Ala Lys Ala Pro Leu Leu Lys Arg Glu Glu His
785                 790                 795                 800

Tyr Glu Tyr Trp Gln Gln Asn Lys Pro Gln Ser Pro Thr Glu Ser Gln
                805                 810                 815

Arg Leu Glu Leu His Thr Ser Asp Arg Trp Lys Asp Tyr Leu Leu Tyr
            820                 825                 830

Lys Arg Trp Gln His Leu Glu Lys Lys Leu Arg Leu Tyr Arg Asn Gln
        835                 840                 845

Asp Val Met Leu Trp Leu Met Thr Leu Glu Leu Thr Lys Asn His Phe
    850                 855                 860

Lys Glu Leu Asn Leu Asn Tyr His Gln Leu Lys Leu Glu Asn Leu Ala
865                 870                 875                 880

Val Asn Val Gln Glu Ala Asp Ala Lys Leu Asn Pro Leu Asn Gln Thr
                885                 890                 895

Leu Pro Met Val Leu Pro Val Lys Val Tyr Pro Ala Thr Ala Phe Gly
            900                 905                 910

Glu Val Gln Tyr His Lys Thr Pro Ile Arg Thr Val Tyr Ile Arg Glu
        915                 920                 925

Glu His Thr Lys Ala Leu Lys Met Gly Asn Phe Lys Ala Leu Val Lys
    930                 935                 940
```

```
Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe Ile Lys Glu Glu Asn Asp
945                 950                 955                 960


Thr Gln Lys His Pro Ile Ser Gln Leu Arg Leu Arg Arg Glu Leu Glu
                965                 970                 975


Ile Tyr Gln Ser Leu Arg Val Asp Ala Phe Lys Glu Thr Leu Ser Leu
            980                 985                 990


Glu Glu Lys Leu Leu Asn Lys His  Thr Ser Leu Ser Ser  Leu Glu Asn
        995                 1000                1005


Glu Phe  Arg Ala Leu Leu Glu  Glu Trp Lys Lys Glu  Tyr Ala Ala
    1010                1015                1020


Ser Ser  Met Val Thr Asp Glu  His Ile Ala Phe Ile  Ala Ser Val
    1025                1030                1035


Arg Asn  Ala Phe Cys His Asn  Gln Tyr Pro Phe Tyr  Lys Glu Ala
    1040                1045                1050


Leu His  Ala Pro Ile Pro Leu  Phe Thr Val Ala Gln  Pro Thr Thr
    1055                1060                1065


Glu Glu  Lys Asp Gly Leu Gly  Ile Ala Glu Ala Leu  Leu Lys Val
    1070                1075                1080


Leu Arg  Glu Tyr Cys Glu Ile  Val Lys Ser Gln Ile
    1085                1090                1095


<210>  76
<211>  1124
<212>  PRT
<213>  Prevotella aurantiaca

<400>  76

Met Glu Asp Asp Lys Lys Thr Thr Gly Ser Ile Ser Tyr Glu Leu Lys
1               5                   10                  15
```

```
Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn Val
            20                  25                  30

Tyr Ile Thr Ile Asn His Ile Asn Lys Leu Leu Glu Ile Arg Glu Ile
        35                  40                  45

Asp Asn Asp Glu Lys Val Leu Asp Ile Lys Thr Leu Trp Gln Lys Gly
    50                  55                  60

Asn Lys Asp Leu Asn Gln Lys Ala Arg Leu Arg Glu Leu Met Thr Lys
65                  70                  75                  80

His Phe Pro Phe Leu Glu Thr Ala Ile Tyr Thr Lys Asn Lys Glu Asp
                85                  90                  95

Lys Lys Glu Val Lys Gln Glu Lys Gln Ala Glu Ala Gln Ser Leu Glu
            100                 105                 110

Ser Leu Lys Asp Cys Leu Phe Leu Phe Leu Asp Lys Leu Gln Glu Ala
        115                 120                 125

Arg Asn Tyr Tyr Ser His Tyr Lys Tyr Ser Glu Phe Ser Lys Glu Pro
    130                 135                 140

Glu Phe Glu Glu Gly Leu Leu Glu Lys Met Tyr Asn Ile Phe Gly Asn
145                 150                 155                 160

Asn Ile Gln Leu Val Ile Asn Asp Tyr Gln His Asn Lys Asp Ile Asn
                165                 170                 175

Pro Asp Glu Asp Phe Lys His Leu Asp Arg Lys Gly Gln Phe Lys Tyr
            180                 185                 190

Ser Phe Ala Asp Asn Glu Gly Asn Ile Thr Glu Ser Gly Leu Leu Phe
        195                 200                 205

Phe Val Ser Leu Phe Leu Glu Lys Lys Asp Ala Ile Trp Met Gln Gln
    210                 215                 220
```

```
Lys Leu Asn Gly Phe Lys Asp Asn Leu Glu Asn Lys Lys Lys Met Thr
225                 230                 235                 240

His Glu Val Phe Cys Arg Ser Arg Ile Leu Met Pro Lys Leu Arg Leu
                245                 250                 255

Glu Ser Thr Gln Thr Gln Asp Trp Ile Leu Leu Asp Met Leu Asn Glu
            260                 265                 270

Leu Ile Arg Cys Pro Lys Ser Leu Tyr Glu Arg Leu Gln Gly Asp Asp
        275                 280                 285

Arg Glu Lys Phe Lys Val Pro Phe Asp Pro Ala Asp Glu Asp Tyr Asn
    290                 295                 300

Ala Glu Gln Glu Pro Phe Lys Asn Thr Leu Ile Arg His Gln Asp Arg
305                 310                 315                 320

Phe Pro Tyr Phe Val Leu Arg Tyr Phe Asp Tyr Asn Glu Ile Phe Lys
                325                 330                 335

Asn Leu Arg Phe Gln Ile Asp Leu Gly Thr Tyr His Phe Ser Ile Tyr
            340                 345                 350

Lys Lys Leu Ile Gly Gly Gln Lys Glu Asp Arg His Leu Thr His Lys
        355                 360                 365

Leu Tyr Gly Phe Glu Arg Ile Gln Glu Phe Ala Lys Gln Asn Arg Pro
    370                 375                 380

Asp Glu Trp Lys Ala Ile Val Lys Asp Leu Asp Thr Tyr Glu Thr Ser
385                 390                 395                 400

Asn Lys Arg Tyr Ile Ser Glu Thr Thr Pro His Tyr His Leu Glu Asn
                405                 410                 415

Gln Lys Ile Gly Ile Arg Phe Arg Asn Gly Asn Lys Glu Ile Trp Pro
            420                 425                 430
```

```
Ser Leu Lys Thr Asn Asp Glu Asn Asn Glu Lys Ser Lys Tyr Lys Leu
        435                 440                 445

Asp Lys Gln Tyr Gln Ala Glu Ala Phe Leu Ser Val His Glu Leu Leu
    450                 455                 460

Pro Met Met Phe Tyr Tyr Leu Leu Leu Lys Lys Glu Lys Pro Asn Asn
465                 470                 475                 480

Asp Glu Ile Asn Ala Ser Ile Val Glu Gly Phe Ile Lys Arg Glu Ile
                485                 490                 495

Arg Asn Ile Phe Lys Leu Tyr Asp Ala Phe Ala Asn Gly Glu Ile Asn
            500                 505                 510

Asn Ile Asp Asp Leu Glu Lys Tyr Cys Ala Asp Lys Gly Ile Pro Lys
        515                 520                 525

Arg His Leu Pro Lys Gln Met Val Ala Ile Leu Tyr Asp Glu His Lys
    530                 535                 540

Asp Met Val Lys Glu Ala Lys Arg Lys Gln Lys Glu Met Val Lys Asp
545                 550                 555                 560

Thr Lys Lys Leu Leu Ala Thr Leu Glu Lys Gln Thr Gln Lys Glu Lys
                565                 570                 575

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

Arg Trp Leu Val Asn Asp Met Met Arg Phe Gln Pro Val Gln Lys Asp
        595                 600                 605

Asn Glu Gly Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr
    610                 615                 620

Gln Met Leu Gln Arg Ser Leu Ala Leu Tyr Asn Asn Glu Glu Lys Pro
625                 630                 635                 640
```

```
Thr Arg Tyr Phe Arg Gln Val Asn Leu Ile Glu Ser Asn Asn Pro His
                645                 650                 655

Pro Phe Leu Lys Trp Thr Lys Trp Glu Glu Cys Asn Asn Ile Leu Thr
            660                 665                 670

Phe Tyr Tyr Ser Tyr Leu Thr Lys Lys Ile Glu Phe Leu Asn Lys Leu
        675                 680                 685

Lys Pro Glu Asp Trp Lys Lys Asn Gln Tyr Phe Leu Lys Leu Lys Glu
    690                 695                 700

Pro Lys Thr Asn Arg Glu Thr Leu Val Gln Gly Trp Lys Asn Gly Phe
705                 710                 715                 720

Asn Leu Pro Arg Gly Ile Phe Thr Glu Pro Ile Arg Glu Trp Phe Lys
                725                 730                 735

Arg His Gln Asn Asn Ser Lys Glu Tyr Glu Lys Val Glu Ala Leu Asp
            740                 745                 750

Arg Val Gly Leu Val Thr Lys Val Ile Pro Leu Phe Phe Lys Glu Glu
        755                 760                 765

Tyr Phe Lys Asp Lys Glu Glu Asn Phe Lys Glu Asp Thr Gln Lys Glu
    770                 775                 780

Ile Asn Asp Cys Val Gln Pro Phe Tyr Asn Phe Pro Tyr Asn Val Gly
785                 790                 795                 800

Asn Ile His Lys Pro Lys Glu Lys Asp Phe Leu His Arg Glu Glu Arg
                805                 810                 815

Ile Glu Leu Trp Asp Lys Lys Lys Asp Lys Phe Lys Gly Tyr Lys Glu
            820                 825                 830

Lys Ile Lys Ser Lys Lys Leu Thr Glu Lys Asp Lys Glu Glu Phe Arg
        835                 840                 845
```

```
Ser Tyr Leu Glu Phe Gln Ser Trp Asn Lys Phe Glu Arg Glu Leu Arg
    850                 855                 860

Leu Val Arg Asn Gln Asp Ile Val Thr Trp Leu Leu Cys Lys Glu Leu
865                 870                 875                 880

Ile Asp Lys Leu Lys Ile Asp Glu Leu Asn Ile Glu Glu Leu Lys Lys
                885                 890                 895

Leu Arg Leu Asn Asn Ile Asp Thr Asp Thr Ala Lys Lys Glu Lys Asn
            900                 905                 910

Asn Ile Leu Asn Arg Val Met Pro Met Glu Leu Pro Val Thr Val Tyr
        915                 920                 925

Glu Ile Asp Asp Ser His Lys Ile Val Lys Asp Lys Pro Leu His Thr
    930                 935                 940

Ile Tyr Ile Lys Glu Ala Glu Thr Lys Leu Leu Lys Gln Gly Asn Phe
945                 950                 955                 960

Lys Ala Leu Val Lys Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe Val
                965                 970                 975

Lys Thr Asn Ser Glu Ala Glu Ser Lys Arg Asn Pro Ile Ser Lys Leu
            980                 985                 990

Arg Val Glu Tyr Glu Leu Gly Glu  Tyr Gln Glu Ala Arg  Ile Glu Ile
        995                 1000                1005

Ile Gln  Asp Met Leu Ala Leu  Glu Glu Lys Leu Ile  Asn Lys Tyr
    1010                1015                1020

Lys Asp  Leu Pro Thr Asn Lys  Phe Ser Glu Met Leu  Asn Ser Trp
    1025                1030                1035

Leu Glu  Gly Lys Asp Glu Ala  Asp Lys Ala Arg Phe  Gln Asn Asp
    1040                1045                1050
```

```
Val Asp  Phe Leu Ile Ala Val  Arg Asn Ala Phe Ser  His Asn Gln
    1055                1060                1065


Tyr Pro  Met His Asn Lys Ile  Glu Phe Ala Asn Ile  Lys Pro Phe
    1070                1075                1080


Ser Leu  Tyr Thr Ala Asn Asn  Ser Glu Glu Lys Gly  Leu Gly Ile
    1085                1090                1095


Ala Asn  Gln Leu Lys Asp Lys  Thr Lys Glu Thr Thr  Asp Lys Ile
    1100                1105                1110


Lys Lys  Ile Glu Lys Pro Ile  Glu Thr Lys Glu
    1115                1120


<210>  77
<211>  1151
<212>  PRT
<213>  Prevotella saccharolytica

<400>  77

Met Glu Asp Lys Pro Phe Trp Ala Ala Phe Phe Asn Leu Ala Arg His
1               5                   10                  15


Asn Val Tyr Leu Thr Val Asn His Ile Asn Lys Leu Leu Asp Leu Glu
            20                  25                  30


Lys Leu Tyr Asp Glu Gly Lys His Lys Glu Ile Phe Glu Arg Glu Asp
        35                  40                  45


Ile Phe Asn Ile Ser Asp Asp Val Met Asn Asp Ala Asn Ser Asn Gly
    50                  55                  60


Lys Lys Arg Lys Leu Asp Ile Lys Lys Ile Trp Asp Asp Leu Asp Thr
65                  70                  75                  80


Asp Leu Thr Arg Lys Tyr Gln Leu Arg Glu Leu Ile Leu Lys His Phe
                85                  90                  95
```

```
Pro Phe Ile Gln Pro Ala Ile Ile Gly Ala Gln Thr Lys Glu Arg Thr
            100                 105                 110

Thr Ile Asp Lys Asp Lys Arg Ser Thr Ser Thr Ser Asn Asp Ser Leu
        115                 120                 125

Lys Gln Thr Gly Glu Gly Asp Ile Asn Asp Leu Leu Ser Leu Ser Asn
    130                 135                 140

Val Lys Ser Met Phe Phe Arg Leu Leu Gln Ile Leu Glu Gln Leu Arg
145                 150                 155                 160

Asn Tyr Tyr Ser His Val Lys His Ser Lys Ser Ala Thr Met Pro Asn
                165                 170                 175

Phe Asp Glu Asp Leu Leu Asn Trp Met Arg Tyr Ile Phe Ile Asp Ser
            180                 185                 190

Val Asn Lys Val Lys Glu Asp Tyr Ser Ser Asn Ser Val Ile Asp Pro
        195                 200                 205

Asn Thr Ser Phe Ser His Leu Ile Tyr Lys Asp Glu Gln Gly Lys Ile
    210                 215                 220

Lys Pro Cys Arg Tyr Pro Phe Thr Ser Lys Asp Gly Ser Ile Asn Ala
225                 230                 235                 240

Phe Gly Leu Leu Phe Phe Val Ser Leu Phe Leu Glu Lys Gln Asp Ser
                245                 250                 255

Ile Trp Met Gln Lys Lys Ile Pro Gly Phe Lys Lys Ala Ser Glu Asn
            260                 265                 270

Tyr Met Lys Met Thr Asn Glu Val Phe Cys Arg Asn His Ile Leu Leu
        275                 280                 285

Pro Lys Ile Arg Leu Glu Thr Val Tyr Asp Lys Asp Trp Met Leu Leu
    290                 295                 300
```

```
Asp Met Leu Asn Glu Val Val Arg Cys Pro Leu Ser Leu Tyr Lys Arg
305                 310                 315                 320


Leu Thr Pro Ala Ala Gln Asn Lys Phe Lys Val Pro Glu Lys Ser Ser
                325                 330                 335


Asp Asn Ala Asn Arg Gln Glu Asp Asp Asn Pro Phe Ser Arg Ile Leu
            340                 345                 350


Val Arg His Gln Asn Arg Phe Pro Tyr Phe Val Leu Arg Phe Phe Asp
        355                 360                 365


Leu Asn Glu Val Phe Thr Thr Leu Arg Phe Gln Ile Asn Leu Gly Cys
    370                 375                 380


Tyr His Phe Ala Ile Cys Lys Lys Gln Ile Gly Asp Lys Lys Glu Val
385                 390                 395                 400


His His Leu Ile Arg Thr Leu Tyr Gly Phe Ser Arg Leu Gln Asn Phe
                405                 410                 415


Thr Gln Asn Thr Arg Pro Glu Glu Trp Asn Thr Leu Val Lys Thr Thr
            420                 425                 430


Glu Pro Ser Ser Gly Asn Asp Gly Lys Thr Val Gln Gly Val Pro Leu
        435                 440                 445


Pro Tyr Ile Ser Tyr Thr Ile Pro His Tyr Gln Ile Glu Asn Glu Lys
    450                 455                 460


Ile Gly Ile Lys Ile Phe Asp Gly Asp Thr Ala Val Asp Thr Asp Ile
465                 470                 475                 480


Trp Pro Ser Val Ser Thr Glu Lys Gln Leu Asn Lys Pro Asp Lys Tyr
                485                 490                 495


Thr Leu Thr Pro Gly Phe Lys Ala Asp Val Phe Leu Ser Val His Glu
            500                 505                 510
```

```
Leu Leu Pro Met Met Phe Tyr Tyr Gln Leu Leu Leu Cys Glu Gly Met
        515                 520                 525


Leu Lys Thr Asp Ala Gly Asn Ala Val Glu Lys Val Leu Ile Asp Thr
    530                 535                 540


Arg Asn Ala Ile Phe Asn Leu Tyr Asp Ala Phe Val Gln Glu Lys Ile
545                 550                 555                 560


Asn Thr Ile Thr Asp Leu Glu Asn Tyr Leu Gln Asp Lys Pro Ile Leu
                565                 570                 575


Ile Gly His Leu Pro Lys Gln Met Ile Asp Leu Leu Lys Gly His Gln
            580                 585                 590


Arg Asp Met Leu Lys Ala Val Glu Gln Lys Lys Ala Met Leu Ile Lys
        595                 600                 605


Asp Thr Glu Arg Arg Leu Lys Leu Leu Asp Lys Gln Leu Lys Gln Glu
    610                 615                 620


Thr Asp Val Ala Ala Lys Asn Thr Gly Thr Leu Leu Lys Asn Gly Gln
625                 630                 635                 640


Ile Ala Asp Trp Leu Val Asn Asp Met Met Arg Phe Gln Pro Val Lys
                645                 650                 655


Arg Asp Lys Glu Gly Asn Pro Ile Asn Cys Ser Lys Ala Asn Ser Thr
            660                 665                 670


Glu Tyr Gln Met Leu Gln Arg Ala Phe Ala Phe Tyr Ala Thr Asp Ser
        675                 680                 685


Cys Arg Leu Ser Arg Tyr Phe Thr Gln Leu His Leu Ile His Ser Asp
    690                 695                 700


Asn Ser His Leu Phe Leu Ser Arg Phe Glu Tyr Asp Lys Gln Pro Asn
705                 710                 715                 720
```

```
Leu Ile Ala Phe Tyr Ala Ala Tyr Leu Lys Ala Lys Leu Glu Phe Leu
                725                 730                 735

Asn Glu Leu Gln Pro Gln Asn Trp Ala Ser Asp Asn Tyr Phe Leu Leu
            740                 745                 750

Leu Arg Ala Pro Lys Asn Asp Arg Gln Lys Leu Ala Glu Gly Trp Lys
        755                 760                 765

Asn Gly Phe Asn Leu Pro Arg Gly Leu Phe Thr Glu Lys Ile Lys Thr
    770                 775                 780

Trp Phe Asn Glu His Lys Thr Ile Val Asp Ile Ser Asp Cys Asp Ile
785                 790                 795                 800

Phe Lys Asn Arg Val Gly Gln Val Ala Arg Leu Ile Pro Val Phe Phe
                805                 810                 815

Asp Lys Lys Phe Lys Asp His Ser Gln Pro Phe Tyr Arg Tyr Asp Phe
            820                 825                 830

Asn Val Gly Asn Val Ser Lys Pro Thr Glu Ala Asn Tyr Leu Ser Lys
        835                 840                 845

Gly Lys Arg Glu Glu Leu Phe Lys Ser Tyr Gln Asn Lys Phe Lys Asn
    850                 855                 860

Asn Ile Pro Ala Glu Lys Thr Lys Glu Tyr Arg Glu Tyr Lys Asn Phe
865                 870                 875                 880

Ser Leu Trp Lys Lys Phe Glu Arg Glu Leu Arg Leu Ile Lys Asn Gln
                885                 890                 895

Asp Ile Leu Ile Trp Leu Met Cys Lys Asn Leu Phe Asp Glu Lys Ile
            900                 905                 910

Lys Pro Lys Lys Asp Ile Leu Glu Pro Arg Ile Ala Val Ser Tyr Ile
        915                 920                 925
```

```
Lys Leu Asp Ser Leu Gln Thr Asn Thr Ser Thr Ala Gly Ser Leu Asn
    930                 935                 940

Ala Leu Ala Lys Val Val Pro Met Thr Leu Ala Ile His Ile Asp Ser
945                 950                 955                 960

Pro Lys Pro Lys Gly Lys Ala Gly Asn Asn Glu Lys Glu Asn Lys Glu
                965                 970                 975

Phe Thr Val Tyr Ile Lys Glu Glu Gly Thr Lys Leu Leu Lys Trp Gly
            980                 985                 990

Asn Phe Lys Thr Leu Leu Ala Asp  Arg Arg Ile Lys Gly  Leu Phe Ser
        995                 1000                1005

Tyr Ile  Glu His Asp Asp Ile  Asp Leu Lys Gln His  Pro Leu Thr
    1010                1015                1020

Lys Arg  Arg Val Asp Leu Glu  Leu Asp Leu Tyr Gln  Thr Cys Arg
    1025                1030                1035

Ile Asp  Ile Phe Gln Gln Thr  Leu Gly Leu Glu Ala  Gln Leu Leu
    1040                1045                1050

Asp Lys  Tyr Ser Asp Leu Asn  Thr Asp Asn Phe Tyr  Gln Met Leu
    1055                1060                1065

Ile Gly  Trp Arg Lys Lys Glu  Gly Ile Pro Arg Asn  Ile Lys Glu
    1070                1075                1080

Asp Thr  Asp Phe Leu Lys Asp  Val Arg Asn Ala Phe  Ser His Asn
    1085                1090                1095

Gln Tyr  Pro Asp Ser Lys Lys  Ile Ala Phe Arg Arg  Ile Arg Lys
    1100                1105                1110

Phe Asn  Pro Lys Glu Leu Ile  Leu Glu Glu Glu Glu  Gly Leu Gly
    1115                1120                1125
```

```
Ile Ala  Thr Gln Met Tyr Lys  Glu Val Glu Lys Val  Val Asn Arg
    1130                1135                1140


Ile Lys  Arg Ile Glu Leu Phe  Asp
    1145                1150


<210>  78
<211>  1159
<212>  PRT
<213>  Myroides odoratimim

<400>  78

Met Lys Asp Ile Leu Thr Thr Asp Thr Thr Glu Lys Gln Asn Arg Phe
1               5                   10                  15


Tyr Ser His Lys Ile Ala Asp Lys Tyr Phe Phe Gly Gly Tyr Phe Asn
            20                  25                  30


Leu Ala Ser Asn Asn Ile Tyr Glu Val Phe Glu Glu Val Asn Lys Arg
        35                  40                  45


Asn Thr Phe Gly Lys Leu Ala Lys Arg Asp Asn Gly Asn Leu Lys Asn
    50                  55                  60


Tyr Ile Ile His Val Phe Lys Asp Glu Leu Ser Ile Ser Asp Phe Glu
65                  70                  75                  80


Lys Arg Val Ala Ile Phe Ala Ser Tyr Phe Pro Ile Leu Glu Thr Val
                85                  90                  95


Asp Lys Lys Ser Ile Lys Glu Arg Asn Arg Thr Ile Asp Leu Thr Leu
            100                 105                 110


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


Ala Val Asp Gln Leu Arg Asn Phe Tyr Thr His Tyr His His Ser Asp
    130                 135                 140


Ile Val Ile Glu Asn Lys Val Leu Asp Phe Leu Asn Ser Ser Phe Val
```

```
145                 150                 155                 160

Ser Thr Ala Leu His Val Lys Asp Lys Tyr Leu Lys Thr Asp Lys Thr
                165                 170                 175

Lys Glu Phe Leu Lys Glu Thr Ile Ala Ala Glu Leu Asp Ile Leu Ile
            180                 185                 190

Glu Ala Tyr Lys Lys Lys Gln Ile Glu Lys Lys Asn Thr Arg Phe Lys
        195                 200                 205

Ala Asn Lys Arg Glu Asp Ile Leu Asn Ala Ile Tyr Asn Glu Ala Phe
    210                 215                 220

Trp Ser Phe Ile Asn Asp Lys Asp Lys Asp Lys Asp Lys Glu Thr Val
225                 230                 235                 240

Val Ala Lys Gly Ala Asp Ala Tyr Phe Glu Lys Asn His His Lys Ser
                245                 250                 255

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

Leu Leu Ser Phe Phe Leu Thr Asn Lys Glu Met Asp Ser Leu Lys Ala
        275                 280                 285

Asn Leu Thr Gly Phe Lys Gly Lys Val Asp Arg Glu Ser Gly Asn Ser
    290                 295                 300

Ile Lys Tyr Met Ala Thr Gln Arg Ile Tyr Ser Phe His Thr Tyr Arg
305                 310                 315                 320

Gly Leu Lys Gln Lys Ile Arg Thr Ser Glu Glu Gly Val Lys Glu Thr
                325                 330                 335

Leu Leu Met Gln Met Ile Asp Glu Leu Ser Lys Val Pro Asn Val Val
            340                 345                 350

Tyr Gln His Leu Ser Thr Thr Gln Gln Asn Ser Phe Ile Glu Asp Trp
```

```
        355                 360                 365

Asn Glu Tyr Tyr Lys Asp Tyr Glu Asp Asp Val Glu Thr Asp Asp Leu
    370                 375                 380

Ser Arg Val Ile His Pro Val Ile Arg Lys Arg Tyr Glu Asp Arg Phe
385                 390                 395                 400

Asn Tyr Phe Ala Ile Arg Phe Leu Asp Glu Phe Phe Asp Phe Pro Thr
                405                 410                 415

Leu Arg Phe Gln Val His Leu Gly Asp Tyr Val His Asp Arg Arg Thr
            420                 425                 430

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

Thr Val Phe Ala Arg Leu Lys Asp Ile Asn Ser Ala Lys Ala Ser Tyr
    450                 455                 460

Phe His Ser Leu Glu Glu Gln Asp Lys Glu Glu Leu Asp Asn Lys Trp
465                 470                 475                 480

Thr Leu Phe Pro Asn Pro Ser Tyr Asp Phe Pro Lys Glu His Thr Leu
                485                 490                 495

Gln His Gln Gly Glu Gln Lys Asn Ala Gly Lys Ile Gly Ile Tyr Val
            500                 505                 510

Lys Leu Arg Asp Thr Gln Tyr Lys Glu Lys Ala Ala Leu Glu Glu Ala
        515                 520                 525

Arg Lys Ser Leu Asn Pro Lys Glu Arg Ser Ala Thr Lys Ala Ser Lys
    530                 535                 540

Tyr Asp Ile Ile Thr Gln Ile Ile Glu Ala Asn Asp Asn Val Lys Ser
545                 550                 555                 560

Glu Lys Pro Leu Val Phe Thr Gly Gln Pro Ile Ala Tyr Leu Ser Met
```

```
                565                 570                 575

Asn Asp Ile His Ser Met Leu Phe Ser Leu Leu Thr Asp Asn Ala Glu
            580                 585                 590

Leu Lys Lys Thr Pro Glu Glu Val Glu Ala Lys Leu Ile Asp Gln Ile
        595                 600                 605

Gly Lys Gln Ile Asn Glu Ile Leu Ser Lys Asp Thr Asp Thr Lys Ile
    610                 615                 620

Leu Lys Lys Tyr Lys Asp Asn Asp Leu Lys Glu Thr Asp Thr Asp Lys
625                 630                 635                 640

Ile Thr Arg Asp Leu Ala Arg Asp Lys Glu Glu Ile Glu Lys Leu Ile
                645                 650                 655

Leu Glu Gln Lys Gln Arg Ala Asp Asp Tyr Asn Tyr Thr Ser Ser Thr
            660                 665                 670

Lys Phe Asn Ile Asp Lys Ser Arg Lys Arg Lys His Leu Leu Phe Asn
        675                 680                 685

Ala Glu Lys Gly Lys Ile Gly Val Trp Leu Ala Asn Asp Ile Lys Arg
    690                 695                 700

Phe Met Phe Lys Glu Ser Lys Ser Lys Trp Lys Gly Tyr Gln His Thr
705                 710                 715                 720

Glu Leu Gln Lys Leu Phe Ala Tyr Phe Asp Thr Ser Lys Ser Asp Leu
                725                 730                 735

Glu Leu Ile Leu Ser Asn Met Val Met Val Lys Asp Tyr Pro Ile Glu
            740                 745                 750

Leu Ile Asp Leu Val Lys Lys Ser Arg Thr Leu Val Asp Phe Leu Asn
        755                 760                 765

Lys Tyr Leu Glu Ala Arg Leu Glu Tyr Ile Glu Asn Val Ile Thr Arg
```

```
    770                 775                 780

Val Lys Asn Ser Ile Gly Thr Pro Gln Phe Lys Thr Val Arg Lys Glu
785                 790                 795                 800

Cys Phe Thr Phe Leu Lys Lys Ser Asn Tyr Thr Val Val Ser Leu Asp
                805                 810                 815

Lys Gln Val Glu Arg Ile Leu Ser Met Pro Leu Phe Ile Glu Arg Gly
            820                 825                 830

Phe Met Asp Asp Lys Pro Thr Met Leu Glu Gly Lys Ser Tyr Lys Gln
        835                 840                 845

His Lys Glu Lys Phe Ala Asp Trp Phe Val His Tyr Lys Glu Asn Ser
    850                 855                 860

Asn Tyr Gln Asn Phe Tyr Asp Thr Glu Val Tyr Glu Ile Thr Thr Glu
865                 870                 875                 880

Asp Lys Arg Glu Lys Ala Lys Val Thr Lys Lys Ile Lys Gln Gln Gln
                885                 890                 895

Lys Asn Asp Val Phe Thr Leu Met Met Val Asn Tyr Met Leu Glu Glu
            900                 905                 910

Val Leu Lys Leu Ser Ser Asn Asp Arg Leu Ser Leu Asn Glu Leu Tyr
        915                 920                 925

Gln Thr Lys Glu Glu Arg Ile Val Asn Lys Gln Val Ala Lys Asp Thr
    930                 935                 940

Gln Glu Arg Asn Lys Asn Tyr Ile Trp Asn Lys Val Val Asp Leu Gln
945                 950                 955                 960

Leu Cys Asp Gly Leu Val His Ile Asp Asn Val Lys Leu Lys Asp Ile
                965                 970                 975

Gly Asn Phe Arg Lys Tyr Glu Asn Asp Ser Arg Val Lys Glu Phe Leu
```

```
            980                 985                 990

Thr Tyr Gln Ser Asp Ile Val Trp  Ser Ala Tyr Leu Ser  Asn Glu Val
        995                1000                1005

Asp Ser  Asn Lys Leu Tyr Val  Ile Glu Arg Gln Leu  Asp Asn Tyr
    1010                1015                1020

Glu Ser  Ile Arg Ser Lys Glu  Leu Leu Lys Glu Val  Gln Glu Ile
    1025                1030                1035

Glu Cys  Ser Val Tyr Asn Gln  Val Ala Asn Lys Glu  Ser Leu Lys
    1040                1045                1050

Gln Ser  Gly Asn Glu Asn Phe  Lys Gln Tyr Val Leu  Gln Gly Leu
    1055                1060                1065

Leu Pro  Ile Gly Met Asp Val  Arg Glu Met Leu Ile  Leu Ser Thr
    1070                1075                1080

Asp Val  Lys Phe Lys Lys Glu  Glu Ile Ile Gln Leu  Gly Gln Ala
    1085                1090                1095

Gly Glu  Val Glu Gln Asp Leu  Tyr Ser Leu Ile Tyr  Ile Arg Asn
    1100                1105                1110

Lys Phe  Ala His Asn Gln Leu  Pro Ile Lys Glu Phe  Phe Asp Phe
    1115                1120                1125

Cys Glu  Asn Asn Tyr Arg Ser  Ile Ser Asp Asn Glu  Tyr Tyr Ala
    1130                1135                1140

Glu Tyr  Tyr Met Glu Ile Phe  Arg Ser Ile Lys Glu  Lys Tyr Ala
    1145                1150                1155

Asn
```

<210> 79

<211> 1120
<212> PRT
<213> Prevotella intermedia

<400> 79

```
Met Glu Asp Asp Lys Lys Thr Thr Asp Ser Ile Arg Tyr Glu Leu Lys
1               5                  10                  15


Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn Val
            20                  25                  30


Tyr Ile Thr Val Asn His Ile Asn Lys Ile Leu Glu Glu Asp Glu Ile
        35                  40                  45


Asn Arg Asp Gly Tyr Glu Asn Thr Leu Glu Asn Ser Trp Asn Glu Ile
    50                  55                  60


Lys Asp Ile Asn Lys Lys Asp Arg Leu Ser Lys Leu Ile Ile Lys His
65                  70                  75                  80


Phe Pro Phe Leu Glu Ala Thr Thr Tyr Arg Gln Asn Pro Thr Asp Thr
                85                  90                  95


Thr Lys Gln Lys Glu Glu Lys Gln Ala Glu Ala Gln Ser Leu Glu Ser
            100                 105                 110


Leu Lys Lys Ser Phe Phe Val Phe Ile Tyr Lys Leu Arg Asp Leu Arg
        115                 120                 125


Asn His Tyr Ser His Tyr Lys His Ser Lys Ser Leu Glu Arg Pro Lys
    130                 135                 140


Phe Glu Glu Asp Leu Gln Asn Lys Met Tyr Asn Ile Phe Asp Val Ser
145                 150                 155                 160


Ile Gln Phe Val Lys Glu Asp Tyr Lys His Asn Thr Asp Ile Asn Pro
                165                 170                 175


Lys Lys Asp Phe Lys His Leu Asp Arg Lys Arg Lys Gly Lys Phe His
            180                 185                 190
```

```
Tyr Ser Phe Ala Asp Asn Glu Gly Asn Ile Thr Glu Ser Gly Leu Leu
        195                 200                 205

Phe Phe Val Ser Leu Phe Leu Glu Lys Lys Asp Ala Ile Trp Val Gln
    210                 215                 220

Lys Lys Leu Glu Gly Phe Lys Cys Ser Asn Lys Ser Tyr Gln Lys Met
225                 230                 235                 240

Thr Asn Glu Val Phe Cys Arg Ser Arg Met Leu Leu Pro Lys Leu Arg
                245                 250                 255

Leu Glu Ser Thr Gln Thr Gln Asp Trp Ile Leu Leu Asp Met Leu Asn
            260                 265                 270

Glu Leu Ile Arg Cys Pro Lys Ser Leu Tyr Glu Arg Leu Gln Gly Val
        275                 280                 285

Asn Arg Lys Lys Phe Tyr Val Ser Phe Asp Pro Ala Asp Glu Asp Tyr
    290                 295                 300

Asp Ala Glu Gln Glu Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp
305                 310                 315                 320

Arg Phe Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Tyr Asn Glu Val Phe
                325                 330                 335

Ala Asn Leu Arg Phe Gln Ile Asp Leu Gly Thr Tyr His Phe Ser Ile
            340                 345                 350

Tyr Lys Lys Leu Ile Gly Gly Gln Lys Glu Asp Arg His Leu Thr His
        355                 360                 365

Lys Leu Tyr Gly Phe Glu Arg Ile Gln Glu Phe Asp Lys Gln Asn Arg
    370                 375                 380

Pro Asp Glu Trp Lys Ala Ile Val Lys Asp Ser Asp Thr Phe Lys Lys
385                 390                 395                 400
```

```
Lys Glu Glu Lys Glu Glu Glu Lys Pro Tyr Ile Ser Glu Thr Thr Pro
                405                 410                 415

His Tyr His Leu Glu Asn Lys Lys Ile Gly Ile Ala Phe Lys Asn His
            420                 425                 430

Asn Ile Trp Pro Ser Thr Gln Thr Glu Leu Thr Asn Asn Lys Arg Lys
        435                 440                 445

Lys Tyr Asn Leu Gly Thr Ser Ile Lys Ala Glu Ala Phe Leu Ser Val
    450                 455                 460

His Glu Leu Leu Pro Met Met Phe Tyr Tyr Leu Leu Leu Lys Thr Glu
465                 470                 475                 480

Asn Thr Lys Asn Asp Asn Lys Val Gly Gly Lys Lys Glu Thr Lys Lys
                485                 490                 495

Gln Gly Lys His Lys Ile Glu Ala Ile Ile Glu Ser Lys Ile Lys Asp
            500                 505                 510

Ile Tyr Ala Leu Tyr Asp Ala Phe Ala Asn Gly Glu Ile Asn Ser Glu
        515                 520                 525

Asp Glu Leu Lys Glu Tyr Leu Lys Gly Lys Asp Ile Lys Ile Val His
    530                 535                 540

Leu Pro Lys Gln Met Ile Ala Ile Leu Lys Asn Glu His Lys Asp Met
545                 550                 555                 560

Ala Glu Lys Ala Glu Ala Lys Gln Glu Lys Met Lys Leu Ala Thr Glu
                565                 570                 575

Asn Arg Leu Lys Thr Leu Asp Lys Gln Leu Lys Gly Lys Ile Gln Asn
            580                 585                 590

Gly Lys Arg Tyr Asn Ser Ala Pro Lys Ser Gly Glu Ile Ala Ser Trp
        595                 600                 605
```

```
Leu Val Asn Asp Met Met Arg Phe Gln Pro Val Gln Lys Asp Glu Asn
    610                 615                 620

Gly Glu Ser Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Gln Leu
625                 630                 635                 640

Leu Gln Arg Thr Leu Ala Phe Phe Gly Ser Glu His Glu Arg Leu Ala
                645                 650                 655

Pro Tyr Phe Lys Gln Thr Lys Leu Ile Glu Ser Ser Asn Pro His Pro
            660                 665                 670

Phe Leu Asn Asp Thr Glu Trp Glu Lys Cys Ser Asn Ile Leu Ser Phe
        675                 680                 685

Tyr Arg Ser Tyr Leu Lys Ala Arg Lys Asn Phe Leu Glu Ser Leu Lys
    690                 695                 700

Pro Glu Asp Trp Glu Lys Asn Gln Tyr Phe Leu Met Leu Lys Glu Pro
705                 710                 715                 720

Lys Thr Asn Arg Glu Thr Leu Val Gln Gly Trp Lys Asn Gly Phe Asn
                725                 730                 735

Leu Pro Arg Gly Phe Phe Thr Glu Pro Ile Arg Lys Trp Phe Met Glu
            740                 745                 750

His Trp Lys Ser Ile Lys Val Asp Asp Leu Lys Arg Val Gly Leu Val
        755                 760                 765

Ala Lys Val Thr Pro Leu Phe Phe Ser Glu Lys Tyr Lys Asp Ser Val
    770                 775                 780

Gln Pro Phe Tyr Asn Tyr Pro Phe Asn Val Gly Asp Val Asn Lys Pro
785                 790                 795                 800

Lys Glu Glu Asp Phe Leu His Arg Glu Glu Arg Ile Glu Leu Trp Asp
                805                 810                 815
```

```
Lys Lys Lys Asp Lys Phe Lys Gly Tyr Lys Ala Lys Lys Lys Phe Lys
            820                 825                 830

Glu Met Thr Asp Lys Glu Lys Glu Glu His Arg Ser Tyr Leu Glu Phe
        835                 840                 845

Gln Ser Trp Asn Lys Phe Glu Arg Glu Leu Arg Leu Val Arg Asn Gln
    850                 855                 860

Asp Ile Val Thr Trp Leu Leu Cys Thr Glu Leu Ile Asp Lys Leu Lys
865                 870                 875                 880

Ile Asp Glu Leu Asn Ile Lys Glu Leu Lys Lys Leu Arg Leu Lys Asp
                885                 890                 895

Ile Asn Thr Asp Thr Ala Lys Lys Glu Lys Asn Asn Ile Leu Asn Arg
            900                 905                 910

Val Met Pro Met Glu Leu Pro Val Thr Val Tyr Lys Val Asn Lys Gly
        915                 920                 925

Gly Tyr Ile Ile Lys Asn Lys Pro Leu His Thr Ile Tyr Ile Lys Glu
    930                 935                 940

Ala Glu Thr Lys Leu Leu Lys Gln Gly Asn Phe Lys Ala Leu Val Lys
945                 950                 955                 960

Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe Val Lys Thr Pro Ser Glu
                965                 970                 975

Ala Glu Ser Glu Ser Asn Pro Ile Ser Lys Leu Arg Val Glu Tyr Glu
            980                 985                 990

Leu Gly Lys Tyr Gln Asn Ala Arg  Leu Asp Ile Ile Glu  Asp Met Leu
        995                 1000                1005

Ala Leu  Glu Lys Lys Leu Ile  Asp Lys Tyr Asn Ser  Leu Asp Thr
    1010                1015                1020
```

```
Asp Asn  Phe His Asn Met Leu  Thr Gly Trp Leu Glu  Leu Lys Gly
    1025                1030                1035


Glu Ala  Lys Lys Ala Arg Phe  Gln Asn Asp Val Lys  Leu Leu Thr
    1040                1045                1050


Ala Val  Arg Asn Ala Phe Ser  His Asn Gln Tyr Pro  Met Tyr Asp
    1055                1060                1065


Glu Asn  Leu Phe Gly Asn Ile  Glu Arg Phe Ser Leu  Ser Ser Ser
    1070                1075                1080


Asn Ile  Ile Glu Ser Lys Gly  Leu Asp Ile Ala Ala  Lys Leu Lys
    1085                1090                1095


Glu Glu  Val Ser Lys Ala Ala  Lys Lys Ile Gln Asn  Glu Glu Asp
    1100                1105                1110


Asn Lys  Lys Glu Lys Glu Thr
    1115                1120


<210>  80
<211>  1199
<212>  PRT
<213>  Capnocytophaga canimorsus

<400>  80

Met Lys Asn Ile Gln Arg Leu Gly Lys Gly Asn Glu Phe Ser Pro Phe
1               5                   10                  15


Lys Lys Glu Asp Lys Phe Tyr Phe Gly Gly Phe Leu Asn Leu Ala Asn
            20                  25                  30


Asn Asn Ile Glu Asp Phe Phe Lys Glu Ile Ile Thr Arg Phe Gly Ile
        35                  40                  45


Val Ile Thr Asp Glu Asn Lys Lys Pro Lys Glu Thr Phe Gly Glu Lys
    50                  55                  60
```

```
Ile Leu Asn Glu Ile Phe Lys Lys Asp Ile Ser Ile Val Asp Tyr Glu
65                  70                  75                  80


Lys Trp Val Asn Ile Phe Ala Asp Tyr Phe Pro Phe Thr Lys Tyr Leu
                85                  90                  95


Ser Leu Tyr Leu Glu Glu Met Gln Phe Lys Asn Arg Val Ile Cys Phe
            100                 105                 110


Arg Asp Val Met Lys Glu Leu Leu Lys Thr Val Glu Ala Leu Arg Asn
        115                 120                 125


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


Phe Tyr Phe Leu Asp Lys Val Leu Leu Asp Val Ser Leu Thr Val Lys
145                 150                 155                 160


Asn Lys Tyr Leu Lys Thr Asp Lys Thr Lys Glu Phe Leu Asn Gln His
                165                 170                 175


Ile Gly Glu Glu Leu Lys Glu Leu Cys Lys Gln Arg Lys Asp Tyr Leu
            180                 185                 190


Val Gly Lys Gly Lys Arg Ile Asp Lys Glu Ser Glu Ile Ile Asn Gly
        195                 200                 205


Ile Tyr Asn Asn Ala Phe Lys Asp Phe Ile Cys Lys Arg Glu Lys Gln
    210                 215                 220


Asp Asp Lys Glu Asn His Asn Ser Val Glu Lys Ile Leu Cys Asn Lys
225                 230                 235                 240


Glu Pro Gln Asn Lys Lys Gln Lys Ser Ser Ala Thr Val Trp Glu Leu
                245                 250                 255


Cys Ser Lys Ser Ser Ser Lys Tyr Thr Glu Lys Ser Phe Pro Asn Arg
            260                 265                 270
```

```
Glu Asn Asp Lys His Cys Leu Glu Val Pro Ile Ser Gln Lys Gly Ile
        275                 280                 285

Val Phe Leu Leu Ser Phe Phe Leu Asn Lys Gly Glu Ile Tyr Ala Leu
    290                 295                 300

Thr Ser Asn Ile Lys Gly Phe Lys Ala Lys Ile Thr Lys Glu Glu Pro
305                 310                 315                 320

Val Thr Tyr Asp Lys Asn Ser Ile Arg Tyr Met Ala Thr His Arg Met
                325                 330                 335

Phe Ser Phe Leu Ala Tyr Lys Gly Leu Lys Arg Lys Ile Arg Thr Ser
            340                 345                 350

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

Glu Thr Leu Met Leu Gln Met Leu Asp Glu Leu Asn Lys Val Pro Asp
    370                 375                 380

Val Val Tyr Gln Asn Leu Ser Glu Asp Val Gln Lys Thr Phe Ile Glu
385                 390                 395                 400

Asp Trp Asn Glu Tyr Leu Lys Glu Asn Asn Gly Asp Val Gly Thr Met
                405                 410                 415

Glu Glu Glu Gln Val Ile His Pro Val Ile Arg Lys Arg Tyr Glu Asp
            420                 425                 430

Lys Phe Asn Tyr Phe Ala Ile Arg Phe Leu Asp Glu Phe Ala Gln Phe
        435                 440                 445

Pro Thr Leu Arg Phe Gln Val His Leu Gly Asn Tyr Leu Cys Asp Lys
    450                 455                 460

Arg Thr Lys Gln Ile Cys Asp Thr Thr Thr Glu Arg Glu Val Lys Lys
465                 470                 475                 480
```

```
Lys Ile Thr Val Phe Gly Arg Leu Ser Glu Leu Glu Asn Lys Lys Ala
                485                 490                 495

Ile Phe Leu Asn Glu Arg Glu Glu Ile Lys Gly Trp Glu Val Phe Pro
            500                 505                 510

Asn Pro Ser Tyr Asp Phe Pro Lys Glu Asn Ile Ser Val Asn Tyr Lys
        515                 520                 525

Asp Phe Pro Ile Val Gly Ser Ile Leu Asp Arg Glu Lys Gln Pro Val
    530                 535                 540

Ser Asn Lys Ile Gly Ile Arg Val Lys Ile Ala Asp Glu Leu Gln Arg
545                 550                 555                 560

Glu Ile Asp Lys Ala Ile Lys Glu Lys Lys Leu Arg Asn Pro Lys Asn
                565                 570                 575

Arg Lys Ala Asn Gln Asp Glu Lys Gln Lys Glu Arg Leu Val Asn Glu
            580                 585                 590

Ile Val Ser Thr Asn Ser Asn Glu Gln Gly Glu Pro Val Val Phe Ile
        595                 600                 605

Gly Gln Pro Thr Ala Tyr Leu Ser Met Asn Asp Ile His Ser Val Leu
    610                 615                 620

Tyr Glu Phe Leu Ile Asn Lys Ile Ser Gly Glu Ala Leu Glu Thr Lys
625                 630                 635                 640

Ile Val Glu Lys Ile Glu Thr Gln Ile Lys Gln Ile Ile Gly Lys Asp
                645                 650                 655

Ala Thr Thr Lys Ile Leu Lys Pro Tyr Thr Asn Ala Asn Ser Asn Ser
            660                 665                 670

Ile Asn Arg Glu Lys Leu Leu Arg Asp Leu Glu Gln Glu Gln Gln Ile
        675                 680                 685
```

```
Leu Lys Thr Leu Leu Glu Glu Gln Gln Arg Glu Lys Asp Lys Lys
    690                 695                 700

Asp Lys Lys Ser Lys Arg Lys His Glu Leu Tyr Pro Ser Glu Lys Gly
705                 710                 715                 720

Lys Val Ala Val Trp Leu Ala Asn Asp Ile Lys Arg Phe Met Pro Lys
                725                 730                 735

Ala Phe Lys Glu Gln Trp Arg Gly Tyr His His Ser Leu Leu Gln Lys
            740                 745                 750

Tyr Leu Ala Tyr Tyr Glu Gln Ser Lys Glu Glu Leu Lys Asn Leu Leu
        755                 760                 765

Pro Lys Glu Val Phe Lys His Phe Pro Phe Lys Leu Lys Gly Tyr Phe
    770                 775                 780

Gln Gln Gln Tyr Leu Asn Gln Phe Tyr Thr Asp Tyr Leu Lys Arg Arg
785                 790                 795                 800

Leu Ser Tyr Val Asn Glu Leu Leu Leu Asn Ile Gln Asn Phe Lys Asn
                805                 810                 815

Asp Lys Asp Ala Leu Lys Ala Thr Glu Lys Glu Cys Phe Lys Phe Phe
            820                 825                 830

Arg Lys Gln Asn Tyr Ile Ile Asn Pro Ile Asn Ile Gln Ile Gln Ser
        835                 840                 845

Ile Leu Val Tyr Pro Ile Phe Leu Lys Arg Gly Phe Leu Asp Glu Lys
    850                 855                 860

Pro Thr Met Ile Asp Arg Glu Lys Phe Lys Glu Asn Lys Asp Thr Glu
865                 870                 875                 880

Leu Ala Asp Trp Phe Met His Tyr Lys Asn Tyr Lys Glu Asp Asn Tyr
                885                 890                 895
```

```
Gln Lys Phe Tyr Ala Tyr Pro Leu Glu Lys Val Glu Glu Lys Glu Lys
            900                 905                 910


Phe Lys Arg Asn Lys Gln Ile Asn Lys Gln Lys Lys Asn Asp Val Tyr
        915                 920                 925


Thr Leu Met Met Val Glu Tyr Ile Ile Gln Lys Ile Phe Gly Asp Lys
    930                 935                 940


Phe Val Glu Glu Asn Pro Leu Val Leu Lys Gly Ile Phe Gln Ser Lys
945                 950                 955                 960


Ala Glu Arg Gln Gln Asn Asn Thr His Ala Ala Thr Thr Gln Glu Arg
                965                 970                 975


Asn Leu Asn Gly Ile Leu Asn Gln Pro Lys Asp Ile Lys Ile Gln Gly
            980                 985                 990


Lys Ile Thr Val Lys Gly Val Lys  Leu Lys Asp Ile Gly  Asn Phe Arg
        995                 1000                1005


Lys Tyr  Glu Ile Asp Gln Arg  Val Asn Thr Phe Leu  Asp Tyr Glu
    1010                1015                1020


Pro Arg  Lys Glu Trp Met Ala  Tyr Leu Pro Asn Asp  Trp Lys Glu
    1025                1030                1035


Lys Glu  Lys Gln Gly Gln Leu  Pro Pro Asn Asn Val  Ile Asp Arg
    1040                1045                1050


Gln Ile  Ser Lys Tyr Glu Thr  Val Arg Ser Lys Ile  Leu Leu Lys
    1055                1060                1065


Asp Val  Gln Glu Leu Glu Lys  Ile Ile Ser Asp Glu  Ile Lys Glu
    1070                1075                1080


Glu His  Arg His Asp Leu Lys  Gln Gly Lys Tyr Tyr  Asn Phe Lys
    1085                1090                1095
```

```
Tyr Tyr  Ile Leu Asn Gly Leu  Leu Arg Gln Leu Lys  Asn Glu Asn
    1100                1105                1110


Val Glu  Asn Tyr Lys Val Phe  Lys Leu Asn Thr Asn  Pro Glu Lys
    1115                1120                1125


Val Asn  Ile Thr Gln Leu Lys  Gln Glu Ala Thr Asp  Leu Glu Gln
    1130                1135                1140


Lys Ala  Phe Val Leu Thr Tyr  Ile Arg Asn Lys Phe  Ala His Asn
    1145                1150                1155


Gln Leu  Pro Lys Lys Glu Phe  Trp Asp Tyr Cys Gln  Glu Lys Tyr
    1160                1165                1170


Gly Lys  Ile Glu Lys Glu Lys  Thr Tyr Ala Glu Tyr  Phe Ala Glu
    1175                1180                1185


Val Phe  Lys Arg Glu Lys Glu  Ala Leu Ile Lys
    1190                1195


<210>  81
<211>  1175
<212>  PRT
<213>  Porphyromonas gulae

<400>  81

Met Thr Glu Gln Ser Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1               5                   10                  15


Glu Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn
            20                  25                  30


Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
        35                  40                  45


Ala Asp Ile Thr Asn Asp Gln Asp Val Leu Ser Phe Lys Ala Leu Trp
    50                  55                  60
```

```
Lys Asn Phe Asp Asn Asp Leu Glu Arg Lys Ser Arg Leu Arg Ser Leu
65                  70                  75                  80


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


Leu Phe Glu Ser Lys Ser Ser Gly Asn Lys Ser Ser Lys Asn Lys Glu
            100                 105                 110


Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
        115                 120                 125


Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
    130                 135                 140


Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Ser Gly Ser Ser Glu Leu
145                 150                 155                 160


Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
                165                 170                 175


Val Ser Val Gln Arg Val Lys Ile Asp His Glu His Asn Asp Glu Val
            180                 185                 190


Asp Pro His Tyr His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
        195                 200                 205


Arg Tyr Gly His Asn Asp Asn Pro Ser Phe Lys His His Phe Val Asp
    210                 215                 220


Gly Glu Gly Met Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225                 230                 235                 240


Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
                245                 250                 255


Phe Lys Gly Gly Thr Glu Thr Tyr Gln Gln Met Thr Asn Glu Val Phe
            260                 265                 270
```

```
Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
        275                 280                 285

Met Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys
    290                 295                 300

Pro Lys Pro Leu Tyr Asp Arg Leu Arg Glu Asp Asp Arg Ala Cys Phe
305                 310                 315                 320

Arg Val Pro Val Asp Ile Leu Pro Asp Glu Asp Asp Thr Asp Gly Gly
                325                 330                 335

Gly Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
            340                 345                 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
        355                 360                 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
    370                 375                 380

Lys Met Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385                 390                 395                 400

Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
                405                 410                 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
            420                 425                 430

Lys Pro Tyr Ile Ser Gln Thr Ser Pro His Tyr His Ile Glu Lys Gly
        435                 440                 445

Lys Ile Gly Leu Arg Phe Met Pro Glu Gly Gln His Leu Trp Pro Ser
    450                 455                 460

Pro Glu Val Gly Thr Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465                 470                 475                 480
```

```
Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
                485                 490                 495


Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val
            500                 505                 510


Ser Ala Glu Arg Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
        515                 520                 525


Tyr Ala Val Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Arg Asp
    530                 535                 540


Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu
545                 550                 555                 560


Pro Arg Gln Met Ile Ala Ile Leu Ser Gln Glu His Lys Asp Met Glu
                565                 570                 575


Glu Lys Ile Arg Lys Lys Leu Gln Glu Met Met Ala Asp Thr Asp His
            580                 585                 590


Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly
        595                 600                 605


Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
    610                 615                 620


Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Ala Ser Gly
625                 630                 635                 640


Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
                645                 650                 655


Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro
            660                 665                 670


Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
        675                 680                 685
```

```
Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
    690                 695                 700

Arg Ser Tyr Leu Arg Ala Arg Lys Ala Phe Leu Glu Arg Ile Gly Arg
705                 710                 715                 720

Ser Asp Arg Val Glu Asn Arg Pro Phe Leu Leu Leu Lys Glu Pro Lys
                725                 730                 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Gly Glu Phe His Leu
            740                 745                 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
        755                 760                 765

Gly His Asp Glu Val Ala Ser Tyr Lys Glu Val Gly Phe Met Ala Lys
    770                 775                 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Glu Asp Arg Val Gln Pro
785                 790                 795                 800

Phe Tyr Asp Ser Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
                805                 810                 815

Gly Arg Phe Leu Ser Lys Glu Glu Arg Ala Glu Glu Trp Glu Arg Gly
            820                 825                 830

Lys Glu Arg Phe Arg Asp Leu Glu Ala Trp Ser Tyr Ser Ala Ala Arg
        835                 840                 845

Arg Ile Glu Asp Ala Phe Ala Gly Ile Glu Tyr Ala Ser Pro Gly Asn
    850                 855                 860

Lys Lys Lys Ile Glu Gln Leu Leu Arg Asp Leu Ser Leu Trp Glu Ala
865                 870                 875                 880

Phe Glu Ser Lys Leu Lys Val Arg Ala Asp Arg Ile Asn Leu Ala Lys
                885                 890                 895
```

```
Leu Lys Lys Glu Ile Leu Glu Ala Gln Glu His Pro Tyr His Asp Phe
            900                 905                 910

Lys Ser Trp Gln Lys Phe Glu Arg Glu Leu Arg Leu Val Lys Asn Gln
        915                 920                 925

Asp Ile Ile Thr Trp Met Met Cys Arg Asp Leu Met Glu Glu Asn Lys
    930                 935                 940

Val Glu Gly Leu Asp Thr Gly Thr Leu Tyr Leu Lys Asp Ile Arg Pro
945                 950                 955                 960

Asn Val Gln Glu Gln Gly Ser Leu Asn Val Leu Asn Arg Val Lys Pro
                965                 970                 975

Met Arg Leu Pro Val Val Val Tyr Arg Ala Asp Ser Arg Gly His Val
            980                 985                 990

His Lys Glu Glu Ala Pro Leu Ala  Thr Val Tyr Ile Glu  Glu Arg Asp
        995                 1000                1005

Thr Lys  Leu Leu Lys Gln Gly  Asn Phe Lys Ser Phe  Val Lys Asp
    1010                1015                1020

Arg Arg  Leu Asn Gly Leu Phe  Ser Phe Val Asp Thr  Gly Gly Leu
    1025                1030                1035

Ala Met  Glu Gln Tyr Pro Ile  Ser Lys Leu Arg Val  Glu Tyr Glu
    1040                1045                1050

Leu Ala  Lys Tyr Gln Thr Ala  Arg Val Cys Val Phe  Glu Leu Thr
    1055                1060                1065

Leu Arg  Leu Glu Glu Ser Leu  Leu Thr Arg Tyr Pro  His Leu Pro
    1070                1075                1080

Asp Glu  Ser Phe Arg Glu Met  Leu Glu Ser Trp Ser  Asp Pro Leu
    1085                1090                1095
```

```
Leu Ala  Lys Trp Pro Glu Leu  His Gly Lys Val Arg  Leu Leu Ile
    1100                1105                1110


Ala Val  Arg Asn Ala Phe Ser  His Asn Gln Tyr Pro  Met Tyr Asp
    1115                1120                1125


Glu Ala  Val Phe Ser Ser Ile  Arg Lys Tyr Asp Pro  Ser Ser Pro
    1130                1135                1140


Asp Ala  Ile Glu Glu Arg Met  Gly Leu Asn Ile Ala  His Arg Leu
    1145                1150                1155


Ser Glu  Glu Val Lys Gln Ala  Lys Glu Thr Val Glu  Arg Ile Ile
    1160                1165                1170


Gln Ala
    1175


<210>  82
<211>  1090
<212>  PRT
<213>  Prevotella sp.

<400>  82

Met Asn Ile Pro Ala Leu Val Glu Asn Gln Lys Lys Tyr Phe Gly Thr
1               5                  10                  15


Tyr Ser Val Met Ala Met Leu Asn Ala Gln Thr Val Leu Asp His Ile
            20                  25                  30


Gln Lys Val Ala Asp Ile Glu Gly Glu Gln Asn Glu Asn Asn Glu Asn
        35                  40                  45


Leu Trp Phe His Pro Val Met Ser His Leu Tyr Asn Ala Lys Asn Gly
    50                  55                  60


Tyr Asp Lys Gln Pro Glu Lys Thr Met Phe Ile Ile Glu Arg Leu Gln
65                  70                  75                  80


Ser Tyr Phe Pro Phe Leu Lys Ile Met Ala Glu Asn Gln Arg Glu Tyr
```

```
                85                  90                  95

Ser Asn Gly Lys Tyr Lys Gln Asn Arg Val Glu Val Asn Ser Asn Asp
            100                 105                 110

Ile Phe Glu Val Leu Lys Arg Ala Phe Gly Val Leu Lys Met Tyr Arg
        115                 120                 125

Asp Leu Thr Asn His Tyr Lys Thr Tyr Glu Glu Lys Leu Asn Asp Gly
    130                 135                 140

Cys Glu Phe Leu Thr Ser Thr Glu Gln Pro Leu Ser Gly Met Ile Asn
145                 150                 155                 160

Asn Tyr Tyr Thr Val Ala Leu Arg Asn Met Asn Glu Arg Tyr Gly Tyr
                165                 170                 175

Lys Thr Glu Asp Leu Ala Phe Ile Gln Asp Lys Arg Phe Lys Phe Val
            180                 185                 190

Lys Asp Ala Tyr Gly Lys Lys Lys Ser Gln Val Asn Thr Gly Phe Phe
        195                 200                 205

Leu Ser Leu Gln Asp Tyr Asn Gly Asp Thr Gln Lys Lys Leu His Leu
    210                 215                 220

Ser Gly Val Gly Ile Ala Leu Leu Ile Cys Leu Phe Leu Asp Lys Gln
225                 230                 235                 240

Tyr Ile Asn Ile Phe Leu Ser Arg Leu Pro Ile Phe Ser Ser Tyr Asn
                245                 250                 255

Ala Gln Ser Glu Glu Arg Arg Ile Ile Ile Arg Ser Phe Gly Ile Asn
            260                 265                 270

Ser Ile Lys Leu Pro Lys Asp Arg Ile His Ser Glu Lys Ser Asn Lys
        275                 280                 285

Ser Val Ala Met Asp Met Leu Asn Glu Val Lys Arg Cys Pro Asp Glu
```

```
    290                 295                 300

Leu Phe Thr Thr Leu Ser Ala Glu Lys Gln Ser Arg Phe Arg Ile Ile
305                 310                 315                 320

Ser Asp Asp His Asn Glu Val Leu Met Lys Arg Ser Ser Asp Arg Phe
                325                 330                 335

Val Pro Leu Leu Leu Gln Tyr Ile Asp Tyr Gly Lys Leu Phe Asp His
            340                 345                 350

Ile Arg Phe His Val Asn Met Gly Lys Leu Arg Tyr Leu Leu Lys Ala
        355                 360                 365

Asp Lys Thr Cys Ile Asp Gly Gln Thr Arg Val Arg Val Ile Glu Gln
    370                 375                 380

Pro Leu Asn Gly Phe Gly Arg Leu Glu Glu Ala Glu Thr Met Arg Lys
385                 390                 395                 400

Gln Glu Asn Gly Thr Phe Gly Asn Ser Gly Ile Arg Ile Arg Asp Phe
                405                 410                 415

Glu Asn Met Lys Arg Asp Asp Ala Asn Pro Ala Asn Tyr Pro Tyr Ile
            420                 425                 430

Val Asp Thr Tyr Thr His Tyr Ile Leu Glu Asn Asn Lys Val Glu Met
        435                 440                 445

Phe Ile Asn Asp Lys Glu Asp Ser Ala Pro Leu Leu Pro Val Ile Glu
    450                 455                 460

Asp Asp Arg Tyr Val Val Lys Thr Ile Pro Ser Cys Arg Met Ser Thr
465                 470                 475                 480

Leu Glu Ile Pro Ala Met Ala Phe His Met Phe Leu Phe Gly Ser Lys
                485                 490                 495

Lys Thr Glu Lys Leu Ile Val Asp Val His Asn Arg Tyr Lys Arg Leu
```

```
            500                 505                 510

Phe Gln Ala Met Gln Lys Glu Glu Val Thr Ala Glu Asn Ile Ala Ser
        515                 520                 525

Phe Gly Ile Ala Glu Ser Asp Leu Pro Gln Lys Ile Leu Asp Leu Ile
    530                 535                 540

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

Val Asp Asp Met Leu Thr Asp Thr Glu Arg Arg Ile Lys Arg Phe Lys
                565                 570                 575

Asp Asp Arg Lys Ser Ile Arg Ser Ala Asp Asn Lys Met Gly Lys Arg
            580                 585                 590

Gly Phe Lys Gln Ile Ser Thr Gly Lys Leu Ala Asp Phe Leu Ala Lys
        595                 600                 605

Asp Ile Val Leu Phe Gln Pro Ser Val Asn Asp Gly Glu Asn Lys Ile
    610                 615                 620

Thr Gly Leu Asn Tyr Arg Ile Met Gln Ser Ala Ile Ala Val Tyr Asp
625                 630                 635                 640

Ser Gly Asp Asp Tyr Glu Ala Lys Gln Gln Phe Lys Leu Met Phe Glu
                645                 650                 655

Lys Ala Arg Leu Ile Gly Lys Gly Thr Thr Glu Pro His Pro Phe Leu
            660                 665                 670

Tyr Lys Val Phe Ala Arg Ser Ile Pro Ala Asn Ala Val Glu Phe Tyr
        675                 680                 685

Glu Arg Tyr Leu Ile Glu Arg Lys Phe Tyr Leu Thr Gly Leu Ser Asn
    690                 695                 700

Glu Ile Lys Lys Gly Asn Arg Val Asp Val Pro Phe Ile Arg Arg Asp
```

```
705                 710                 715                 720

Gln Asn Lys Trp Lys Thr Pro Ala Met Lys Thr Leu Gly Arg Ile Tyr
                725                 730                 735

Ser Glu Asp Leu Pro Val Glu Leu Pro Arg Gln Met Phe Asp Asn Glu
            740                 745                 750

Ile Lys Ser His Leu Lys Ser Leu Pro Gln Met Glu Gly Ile Asp Phe
        755                 760                 765

Asn Asn Ala Asn Val Thr Tyr Leu Ile Ala Glu Tyr Met Lys Arg Val
    770                 775                 780

Leu Asp Asp Asp Phe Gln Thr Phe Tyr Gln Trp Asn Arg Asn Tyr Arg
785                 790                 795                 800

Tyr Met Asp Met Leu Lys Gly Glu Tyr Asp Arg Lys Gly Ser Leu Gln
                805                 810                 815

His Cys Phe Thr Ser Val Glu Glu Arg Glu Gly Leu Trp Lys Glu Arg
            820                 825                 830

Ala Ser Arg Thr Glu Arg Tyr Arg Lys Gln Ala Ser Asn Lys Ile Arg
        835                 840                 845

Ser Asn Arg Gln Met Arg Asn Ala Ser Ser Glu Glu Ile Glu Thr Ile
    850                 855                 860

Leu Asp Lys Arg Leu Ser Asn Ser Arg Asn Glu Tyr Gln Lys Ser Glu
865                 870                 875                 880

Lys Val Ile Arg Arg Tyr Arg Val Gln Asp Ala Leu Leu Phe Leu Leu
                885                 890                 895

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

Lys Leu Lys Glu Ile Met Pro Asp Ala Glu Lys Gly Ile Leu Ser Glu
```

915 920 925

Ile Met Pro Met Ser Phe Thr Phe Glu Lys Gly Gly Lys Lys Tyr Thr
930 935 940

Ile Thr Ser Glu Gly Met Lys Leu Lys Asn Tyr Gly Asp Phe Phe Val
945 950 955 960

Leu Ala Ser Asp Lys Arg Ile Gly Asn Leu Leu Glu Leu Val Gly Ser
965 970 975

Asp Ile Val Ser Lys Glu Asp Ile Met Glu Glu Phe Asn Lys Tyr Asp
980 985 990

Gln Cys Arg Pro Glu Ile Ser Ser Ile Val Phe Asn Leu Glu Lys Trp
995 1000 1005

Ala Phe Asp Thr Tyr Pro Glu Leu Ser Ala Arg Val Asp Arg Glu
1010 1015 1020

Glu Lys Val Asp Phe Lys Ser Ile Leu Lys Ile Leu Leu Asn Asn
1025 1030 1035

Lys Asn Ile Asn Lys Glu Gln Ser Asp Ile Leu Arg Lys Ile Arg
1040 1045 1050

Asn Ala Phe Asp His Asn Asn Tyr Pro Asp Lys Gly Val Val Glu
1055 1060 1065

Ile Lys Ala Leu Pro Glu Ile Ala Met Ser Ile Lys Lys Ala Phe
1070 1075 1080

Gly Glu Tyr Ala Ile Met Lys
1085 1090

<210> 83
<211> 1150
<212> PRT
<213> Flavobacterium branchiophilum

<400> 83

```
Met Glu Asn Leu Asn Lys Ile Leu Asp Lys Glu Asn Glu Ile Cys Ile
1               5                  10                  15

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

Ala Leu Asp Asn Ile Lys Ser Lys Gln Lys Asn Asp Leu Asn Lys Glu
        35                  40                  45

Ala Arg Leu His Tyr Phe Ser Ile Gly His Ser Phe Lys Gln Ile Asp
    50                  55                  60

Thr Lys Lys Val Phe Asp Tyr Val Leu Ile Glu Glu Leu Lys Asp Glu
65                  70                  75                  80

Lys Pro Leu Lys Phe Ile Thr Leu Gln Lys Asp Phe Phe Thr Lys Glu
                85                  90                  95

Phe Ser Ile Lys Leu Gln Lys Leu Ile Asn Ser Ile Arg Asn Ile Asn
            100                 105                 110

Asn His Tyr Val His Asn Phe Asn Asp Ile Asn Leu Asn Lys Ile Asp
        115                 120                 125

Ser Asn Val Phe His Phe Leu Lys Glu Ser Phe Glu Leu Ala Ile Ile
    130                 135                 140

Glu Lys Tyr Tyr Lys Val Asn Lys Lys Tyr Pro Leu Asp Asn Glu Ile
145                 150                 155                 160

Val Leu Phe Leu Lys Glu Leu Phe Ile Lys Asp Glu Asn Thr Ala Leu
                165                 170                 175

Leu Asn Tyr Phe Thr Asn Leu Ser Lys Asp Glu Ala Ile Glu Tyr Ile
            180                 185                 190

Leu Thr Phe Thr Ile Thr Glu Asn Lys Ile Trp Asn Ile Asn Asn Glu
        195                 200                 205
```

```
His Asn Ile Leu Asn Ile Glu Lys Gly Lys Tyr Leu Thr Phe Glu Ala
    210                 215                 220

Met Leu Phe Leu Ile Thr Ile Phe Leu Tyr Lys Asn Glu Ala Asn His
225                 230                 235                 240

Leu Leu Pro Lys Leu Tyr Asp Phe Lys Asn Asn Lys Ser Lys Gln Glu
                245                 250                 255

Leu Phe Thr Phe Phe Ser Lys Lys Phe Thr Ser Gln Asp Ile Asp Ala
            260                 265                 270

Glu Glu Gly His Leu Ile Lys Phe Arg Asp Met Ile Gln Tyr Leu Asn
        275                 280                 285

His Tyr Pro Thr Ala Trp Asn Asn Asp Leu Lys Leu Glu Ser Glu Asn
    290                 295                 300

Lys Asn Lys Ile Met Thr Thr Lys Leu Ile Asp Ser Ile Ile Glu Phe
305                 310                 315                 320

Glu Leu Asn Ser Asn Tyr Pro Ser Phe Ala Thr Asp Ile Gln Phe Lys
                325                 330                 335

Lys Glu Ala Lys Ala Phe Leu Phe Ala Ser Asn Lys Lys Arg Asn Gln
            340                 345                 350

Thr Ser Phe Ser Asn Lys Ser Tyr Asn Glu Glu Ile Arg His Asn Pro
        355                 360                 365

His Ile Lys Gln Tyr Arg Asp Glu Ile Ala Ser Ala Leu Thr Pro Ile
    370                 375                 380

Ser Phe Asn Val Lys Glu Asp Lys Phe Lys Ile Phe Val Lys Lys His
385                 390                 395                 400

Val Leu Glu Glu Tyr Phe Pro Asn Ser Ile Gly Tyr Glu Lys Phe Leu
                405                 410                 415
```

```
Glu Tyr Asn Asp Phe Thr Glu Lys Glu Lys Glu Asp Phe Gly Leu Lys
            420                 425                 430

Leu Tyr Ser Asn Pro Lys Thr Asn Lys Leu Ile Glu Arg Ile Asp Asn
        435                 440                 445

His Lys Leu Val Lys Ser His Gly Arg Asn Gln Asp Arg Phe Met Asp
    450                 455                 460

Phe Ser Met Arg Phe Leu Ala Glu Asn Asn Tyr Phe Gly Lys Asp Ala
465                 470                 475                 480

Phe Phe Lys Cys Tyr Lys Phe Tyr Asp Thr Gln Glu Gln Asp Glu Phe
                485                 490                 495

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

Val Thr Thr Tyr Ile Lys Tyr Glu Glu His Leu Lys Asn Tyr Ser Tyr
        515                 520                 525

Trp Asp Cys Pro Phe Val Glu Glu Asn Asn Ser Met Ser Val Lys Ile
    530                 535                 540

Ser Ile Gly Ser Glu Glu Lys Ile Leu Lys Ile Gln Arg Asn Leu Met
545                 550                 555                 560

Ile Tyr Phe Leu Glu Asn Ala Leu Tyr Asn Glu Asn Val Glu Asn Gln
                565                 570                 575

Gly Tyr Lys Leu Val Asn Asn Tyr Tyr Arg Glu Leu Lys Lys Asp Val
            580                 585                 590

Glu Glu Ser Ile Ala Ser Leu Asp Leu Ile Lys Ser Asn Pro Asp Phe
        595                 600                 605

Lys Ser Lys Tyr Lys Lys Ile Leu Pro Lys Arg Leu Leu His Asn Tyr
    610                 615                 620
```

```
Ala Pro Ala Lys Gln Asp Lys Ala Pro Glu Asn Ala Phe Glu Thr Leu
625                 630                 635                 640


Leu Lys Lys Ala Asp Phe Arg Glu Glu Gln Tyr Lys Lys Leu Leu Lys
                645                 650                 655


Lys Ala Glu His Glu Lys Asn Lys Glu Asp Phe Val Lys Arg Asn Lys
            660                 665                 670


Gly Lys Gln Phe Lys Leu His Phe Ile Arg Lys Ala Cys Gln Met Met
        675                 680                 685


Tyr Phe Lys Glu Lys Tyr Asn Thr Leu Lys Glu Gly Asn Ala Ala Phe
    690                 695                 700


Glu Lys Lys Asp Pro Val Ile Glu Lys Arg Lys Asn Lys Glu His Glu
705                 710                 715                 720


Phe Gly His His Lys Asn Leu Asn Ile Thr Arg Glu Glu Phe Asn Asp
                725                 730                 735


Tyr Cys Lys Trp Met Phe Ala Phe Asn Gly Asn Asp Ser Tyr Lys Lys
            740                 745                 750


Tyr Leu Arg Asp Leu Phe Ser Glu Lys His Phe Phe Asp Asn Gln Glu
        755                 760                 765


Tyr Lys Asn Leu Phe Glu Ser Ser Val Asn Leu Glu Ala Phe Tyr Ala
    770                 775                 780


Lys Thr Lys Glu Leu Phe Lys Lys Trp Ile Glu Thr Asn Lys Pro Thr
785                 790                 795                 800


Asn Asn Glu Asn Arg Tyr Thr Leu Glu Asn Tyr Lys Asn Leu Ile Leu
                805                 810                 815


Gln Lys Gln Val Phe Ile Asn Val Tyr His Phe Ser Lys Tyr Leu Ile
            820                 825                 830
```

```
Asp Lys Asn Leu Leu Asn Ser Glu Asn Asn Val Ile Gln Tyr Lys Ser
        835                 840                 845

Leu Glu Asn Val Glu Tyr Leu Ile Ser Asp Phe Tyr Phe Gln Ser Lys
    850                 855                 860

Leu Ser Ile Asp Gln Tyr Lys Thr Cys Gly Lys Leu Phe Asn Lys Leu
865                 870                 875                 880

Lys Ser Asn Lys Leu Glu Asp Cys Leu Leu Tyr Glu Ile Ala Tyr Asn
                885                 890                 895

Tyr Ile Asp Lys Lys Asn Val His Lys Ile Asp Ile Gln Lys Ile Leu
            900                 905                 910

Thr Ser Lys Ile Ile Leu Thr Ile Asn Asp Ala Asn Thr Pro Tyr Lys
        915                 920                 925

Ile Ser Val Pro Phe Asn Lys Leu Glu Arg Tyr Thr Glu Met Ile Ala
    930                 935                 940

Ile Lys Asn Gln Asn Asn Leu Lys Ala Arg Phe Leu Ile Asp Leu Pro
945                 950                 955                 960

Leu Tyr Leu Ser Lys Asn Lys Ile Lys Lys Gly Lys Asp Ser Ala Gly
                965                 970                 975

Tyr Glu Ile Ile Ile Lys Asn Asp Leu Glu Ile Glu Asp Ile Asn Thr
            980                 985                 990

Ile Asn Asn Lys Ile Ile Asn Asp  Ser Val Lys Phe Thr  Glu Val Leu
        995                 1000                1005

Met Glu  Leu Glu Lys Tyr Phe  Ile Leu Lys Asp Lys  Cys Ile Leu
    1010                1015                1020

Ser Lys  Asn Tyr Ile Asp Asn  Ser Glu Ile Pro Ser  Leu Lys Gln
    1025                1030                1035
```

```
Phe Ser  Lys Val Trp Ile Lys  Glu Asn Glu Asn Glu  Ile Ile Asn
    1040                1045                1050


Tyr Arg  Asn Ile Ala Cys His  Phe His Leu Pro Leu  Leu Glu Thr
    1055                1060                1065


Phe Asp  Asn Leu Leu Leu Asn  Val Glu Gln Lys Phe  Ile Lys Glu
    1070                1075                1080


Glu Leu  Gln Asn Val Ser Thr  Ile Asn Asp Leu Ser  Lys Pro Gln
    1085                1090                1095


Glu Tyr  Leu Ile Leu Leu Phe  Ile Lys Phe Lys His  Asn Asn Phe
    1100                1105                1110


Tyr Leu  Asn Leu Phe Asn Lys  Asn Glu Ser Lys Thr  Ile Lys Asn
    1115                1120                1125


Asp Lys  Glu Val Lys Lys Asn  Arg Val Leu Gln Lys  Phe Ile Asn
    1130                1135                1140


Gln Val  Ile Leu Lys Lys Lys
    1145                1150


<210>  84
<211>  1159
<212>  PRT
<213>  Myroides odoratimum

<400>  84

Met Lys Asp Ile Leu Thr Thr Asp Thr Thr Glu Lys Gln Asn Arg Phe
1               5                  10                  15


Tyr Ser His Lys Ile Ala Asp Lys Tyr Phe Phe Gly Gly Tyr Phe Asn
            20                  25                  30


Leu Ala Ser Asn Asn Ile Tyr Glu Val Phe Glu Glu Val Asn Lys Arg
        35                  40                  45
```

```
Asn Thr Phe Gly Lys Leu Ala Lys Arg Asp Asn Gly Asn Leu Lys Asn
    50                  55                  60

Tyr Ile Ile His Val Phe Lys Asp Glu Leu Ser Ile Ser Asp Phe Glu
65                  70                  75                  80

Lys Arg Val Ala Ile Phe Ala Ser Tyr Phe Pro Ile Leu Glu Thr Val
                85                  90                  95

Asp Lys Lys Ser Ile Lys Glu Arg Asn Arg Thr Ile Asp Leu Thr Leu
            100                 105                 110

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

Ala Val Asp Gln Leu Arg Asn Phe Tyr Thr His Tyr His His Ser Asp
    130                 135                 140

Ile Val Ile Glu Asn Lys Val Leu Asp Phe Leu Asn Ser Ser Phe Val
145                 150                 155                 160

Ser Thr Ala Leu His Val Lys Asp Lys Tyr Leu Lys Thr Asp Lys Thr
                165                 170                 175

Lys Glu Phe Leu Lys Glu Thr Ile Ala Ala Glu Leu Asp Ile Leu Ile
            180                 185                 190

Glu Ala Tyr Lys Lys Lys Gln Ile Glu Lys Lys Asn Thr Arg Phe Lys
        195                 200                 205

Ala Asn Lys Arg Glu Asp Ile Leu Asn Ala Ile Tyr Asn Glu Ala Phe
    210                 215                 220

Trp Ser Phe Ile Asn Asp Lys Asp Lys Asp Lys Asp Lys Glu Thr Val
225                 230                 235                 240

Val Ala Lys Gly Ala Asp Ala Tyr Phe Glu Lys Asn His His Lys Ser
                245                 250                 255
```

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

Leu Leu Ser Phe Phe Leu Thr Asn Lys Glu Met Asp Ser Leu Lys Ala
        275                 280                 285

Asn Leu Thr Gly Phe Lys Gly Lys Val Asp Arg Glu Ser Gly Asn Ser
    290                 295                 300

Ile Lys Tyr Met Ala Thr Gln Arg Ile Tyr Ser Phe His Thr Tyr Arg
305                 310                 315                 320

Gly Leu Lys Gln Lys Ile Arg Thr Ser Glu Glu Gly Val Lys Glu Thr
                325                 330                 335

Leu Leu Met Gln Met Ile Asp Glu Leu Ser Lys Val Pro Asn Val Val
            340                 345                 350

Tyr Gln His Leu Ser Thr Thr Gln Gln Asn Ser Phe Ile Glu Asp Trp
        355                 360                 365

Asn Glu Tyr Tyr Lys Asp Tyr Glu Asp Asp Val Glu Thr Asp Asp Leu
    370                 375                 380

Ser Arg Val Thr His Pro Val Ile Arg Lys Arg Tyr Glu Asp Arg Phe
385                 390                 395                 400

Asn Tyr Phe Ala Ile Arg Phe Leu Asp Glu Phe Phe Asp Phe Pro Thr
                405                 410                 415

Leu Arg Phe Gln Val His Leu Gly Asp Tyr Val His Asp Arg Arg Thr
            420                 425                 430

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

Thr Val Phe Ala Arg Leu Lys Asp Ile Asn Ser Ala Lys Ala Ser Tyr
    450                 455                 460
```

```
Phe His Ser Leu Glu Glu Gln Asp Lys Glu Glu Leu Asp Asn Lys Trp
465                 470                 475                 480


Thr Leu Phe Pro Asn Pro Ser Tyr Asp Phe Pro Lys Glu His Thr Leu
                485                 490                 495


Gln His Gln Gly Glu Gln Lys Asn Ala Gly Lys Ile Gly Ile Tyr Val
            500                 505                 510


Lys Leu Arg Asp Thr Gln Tyr Lys Glu Lys Ala Ala Leu Glu Glu Ala
        515                 520                 525


Arg Lys Ser Leu Asn Pro Lys Glu Arg Ser Ala Thr Lys Ala Ser Lys
    530                 535                 540


Tyr Asp Ile Ile Thr Gln Ile Ile Glu Ala Asn Asp Asn Val Lys Ser
545                 550                 555                 560


Glu Lys Pro Leu Val Phe Thr Gly Gln Pro Ile Ala Tyr Leu Ser Met
                565                 570                 575


Asn Asp Ile His Ser Met Leu Phe Ser Leu Leu Thr Asp Asn Ala Glu
            580                 585                 590


Leu Lys Lys Thr Pro Glu Glu Val Glu Ala Lys Leu Ile Asp Gln Ile
        595                 600                 605


Gly Lys Gln Ile Asn Glu Ile Leu Ser Lys Asp Thr Asp Thr Lys Ile
    610                 615                 620


Leu Lys Lys Tyr Lys Asp Asn Asp Leu Lys Glu Thr Asp Thr Asp Lys
625                 630                 635                 640


Ile Thr Arg Asp Leu Ala Arg Asp Lys Glu Glu Ile Glu Lys Leu Ile
                645                 650                 655


Leu Glu Gln Lys Gln Arg Ala Asp Asp Tyr Asn Tyr Thr Ser Ser Thr
            660                 665                 670
```

```
Lys Phe Asn Ile Asp Lys Ser Arg Lys Arg Lys His Leu Leu Phe Asn
        675                 680                 685

Ala Glu Lys Gly Lys Ile Gly Val Trp Leu Ala Asn Asp Ile Lys Arg
    690                 695                 700

Phe Met Phe Lys Glu Ser Lys Ser Lys Trp Lys Gly Tyr Gln His Ile
705                 710                 715                 720

Glu Leu Gln Lys Leu Phe Ala Tyr Phe Asp Thr Ser Lys Ser Asp Leu
                725                 730                 735

Glu Leu Ile Leu Ser Asn Met Val Met Val Lys Asp Tyr Pro Ile Glu
            740                 745                 750

Leu Ile Asp Leu Val Lys Lys Ser Arg Thr Leu Val Asp Phe Leu Asn
        755                 760                 765

Lys Tyr Leu Glu Ala Arg Leu Glu Tyr Ile Glu Asn Val Ile Thr Arg
    770                 775                 780

Val Lys Asn Ser Ile Gly Thr Pro Gln Phe Lys Thr Val Arg Lys Glu
785                 790                 795                 800

Cys Phe Thr Phe Leu Lys Lys Ser Asn Tyr Thr Val Val Ser Leu Asp
                805                 810                 815

Lys Gln Val Glu Arg Ile Leu Ser Met Pro Leu Phe Ile Glu Arg Gly
            820                 825                 830

Phe Met Asp Asp Lys Pro Thr Met Leu Glu Gly Lys Ser Tyr Lys Gln
        835                 840                 845

His Lys Glu Lys Phe Ala Asp Trp Phe Val His Tyr Lys Glu Asn Ser
    850                 855                 860

Asn Tyr Gln Asn Phe Tyr Asp Thr Glu Val Tyr Glu Ile Thr Thr Glu
865                 870                 875                 880
```

```
Asp Lys Arg Glu Lys Ala Lys Val Thr Lys Lys Ile Lys Gln Gln Gln
                885                 890                 895


Lys Asn Asp Val Phe Thr Leu Met Met Val Asn Tyr Met Leu Glu Glu
            900                 905                 910


Val Leu Lys Leu Ser Ser Asn Asp Arg Leu Ser Leu Asn Glu Leu Tyr
        915                 920                 925


Gln Thr Lys Glu Glu Arg Ile Val Asn Lys Gln Val Ala Lys Asp Thr
    930                 935                 940


Gln Glu Arg Asn Lys Asn Tyr Ile Trp Asn Lys Val Val Asp Leu Gln
945                 950                 955                 960


Leu Cys Asp Gly Leu Val His Ile Asp Asn Val Lys Leu Lys Asp Ile
                965                 970                 975


Gly Asn Phe Arg Lys Tyr Glu Asn Asp Ser Arg Val Lys Glu Phe Leu
            980                 985                 990


Thr Tyr Gln Ser Asp Ile Val Trp  Ser Ala Tyr Leu Ser  Asn Glu Val
        995                 1000                1005


Asp Ser  Asn Lys Leu Tyr Val  Ile Glu Arg Gln Leu  Asp Asn Tyr
    1010                1015                1020


Glu Ser  Ile Arg Ser Lys Glu  Leu Leu Lys Glu Val  Gln Glu Ile
    1025                1030                1035


Glu Cys  Ser Val Tyr Asn Gln  Val Ala Asn Lys Glu  Ser Leu Lys
    1040                1045                1050


Gln Ser  Gly Asn Glu Asn Phe  Lys Gln Tyr Val Leu  Gln Gly Leu
    1055                1060                1065


Leu Pro  Ile Gly Met Asp Val  Arg Glu Met Leu Ile  Leu Ser Thr
    1070                1075                1080
```

```
Asp Val  Lys Phe Lys Lys Glu  Glu Ile Ile Gln Leu  Gly Gln Ala
    1085                1090                1095


Gly Glu  Val Glu Gln Asp Leu  Tyr Ser Leu Ile Tyr  Ile Arg Asn
    1100                1105                1110


Lys Phe  Ala His Asn Gln Leu  Pro Ile Lys Glu Phe  Phe Asp Phe
    1115                1120                1125


Cys Glu  Asn Asn Tyr Arg Ser  Ile Ser Asp Asn Glu  Tyr Tyr Ala
    1130                1135                1140


Glu Tyr  Tyr Met Glu Ile Phe  Arg Ser Ile Lys Glu  Lys Tyr Ala
    1145                1150                1155


Asn



<210>  85
<211>  1214
<212>  PRT
<213>  Flavobacterium columnare

<400>  85

Met Ser Ser Lys Asn Glu Ser Tyr Asn Lys Gln Lys Thr Phe Asn His
1               5                   10                  15


Tyr Lys Gln Glu Asp Lys Tyr Phe Phe Gly Gly Phe Leu Asn Asn Ala
            20                  25                  30


Asp Asp Asn Leu Arg Gln Val Gly Lys Glu Phe Lys Thr Arg Ile Asn
        35                  40                  45


Phe Asn His Asn Asn Asn Glu Leu Ala Ser Val Phe Lys Asp Tyr Phe
    50                  55                  60


Asn Lys Glu Lys Ser Val Ala Lys Arg Glu His Ala Leu Asn Leu Leu
65                  70                  75                  80
```

```
Ser Asn Tyr Phe Pro Val Leu Glu Arg Ile Gln Lys His Thr Asn His
                85                  90                  95

Asn Phe Glu Gln Thr Arg Glu Ile Phe Glu Leu Leu Leu Asp Thr Ile
            100                 105                 110

Lys Lys Leu Arg Asp Tyr Tyr Thr His His Tyr His Lys Pro Ile Thr
        115                 120                 125

Ile Asn Pro Lys Ile Tyr Asp Phe Leu Asp Asp Thr Leu Leu Asp Val
    130                 135                 140

Leu Ile Thr Ile Lys Lys Lys Lys Val Lys Asn Asp Thr Ser Arg Glu
145                 150                 155                 160

Leu Leu Lys Glu Lys Leu Arg Pro Glu Leu Thr Gln Leu Lys Asn Gln
                165                 170                 175

Lys Arg Glu Glu Leu Ile Lys Lys Gly Lys Lys Leu Leu Glu Glu Asn
            180                 185                 190

Leu Glu Asn Ala Val Phe Asn His Cys Leu Ile Pro Phe Leu Glu Glu
        195                 200                 205

Asn Lys Thr Asp Asp Lys Gln Asn Lys Thr Val Ser Leu Arg Lys Tyr
    210                 215                 220

Arg Lys Ser Lys Pro Asn Glu Glu Thr Ser Ile Thr Leu Thr Gln Ser
225                 230                 235                 240

Gly Leu Val Phe Leu Met Ser Phe Phe Leu His Arg Lys Glu Phe Gln
                245                 250                 255

Val Phe Thr Ser Gly Leu Glu Arg Phe Lys Ala Lys Val Asn Thr Ile
            260                 265                 270

Lys Glu Glu Glu Ile Ser Leu Asn Lys Asn Asn Ile Val Tyr Met Ile
        275                 280                 285
```

```
Thr His Trp Ser Tyr Ser Tyr Tyr Asn Phe Lys Gly Leu Lys His Arg
    290                 295                 300

Ile Lys Thr Asp Gln Gly Val Ser Thr Leu Glu Gln Asn Asn Thr Thr
305                 310                 315                 320

His Ser Leu Thr Asn Thr Asn Thr Lys Glu Ala Leu Leu Thr Gln Ile
                325                 330                 335

Val Asp Tyr Leu Ser Lys Val Pro Asn Glu Ile Tyr Glu Thr Leu Ser
            340                 345                 350

Glu Lys Gln Gln Lys Glu Phe Glu Glu Asp Ile Asn Glu Tyr Met Arg
        355                 360                 365

Glu Asn Pro Glu Asn Glu Asp Ser Thr Phe Ser Ser Ile Val Ser His
    370                 375                 380

Lys Val Ile Arg Lys Arg Tyr Glu Asn Lys Phe Asn Tyr Phe Ala Met
385                 390                 395                 400

Arg Phe Leu Asp Glu Tyr Ala Glu Leu Pro Thr Leu Arg Phe Met Val
                405                 410                 415

Asn Phe Gly Asp Tyr Ile Lys Asp Arg Gln Lys Lys Ile Leu Glu Ser
            420                 425                 430

Ile Gln Phe Asp Ser Glu Arg Ile Ile Lys Lys Glu Ile His Leu Phe
        435                 440                 445

Glu Lys Leu Ser Leu Val Thr Glu Tyr Lys Lys Asn Val Tyr Leu Lys
    450                 455                 460

Glu Thr Ser Asn Ile Asp Leu Ser Arg Phe Pro Leu Phe Pro Asn Pro
465                 470                 475                 480

Ser Tyr Val Met Ala Asn Asn Asn Ile Pro Phe Tyr Ile Asp Ser Arg
                485                 490                 495
```

```
Ser Asn Asn Leu Asp Glu Tyr Leu Asn Gln Lys Lys Lys Ala Gln Ser
            500                 505                 510

Gln Asn Lys Lys Arg Asn Leu Thr Phe Glu Lys Tyr Asn Lys Glu Gln
        515                 520                 525

Ser Lys Asp Ala Ile Ile Ala Met Leu Gln Lys Glu Ile Gly Val Lys
    530                 535                 540

Asp Leu Gln Gln Arg Ser Thr Ile Gly Leu Leu Ser Cys Asn Glu Leu
545                 550                 555                 560

Pro Ser Met Leu Tyr Glu Val Ile Val Lys Asp Ile Lys Gly Ala Glu
                565                 570                 575

Leu Glu Asn Lys Ile Ala Gln Lys Ile Arg Glu Gln Tyr Gln Ser Ile
            580                 585                 590

Arg Asp Phe Thr Leu Asp Ser Pro Gln Lys Asp Asn Ile Pro Thr Thr
        595                 600                 605

Leu Ile Lys Thr Ile Asn Thr Asp Ser Ser Val Thr Phe Glu Asn Gln
    610                 615                 620

Pro Ile Asp Ile Pro Arg Leu Lys Asn Ala Leu Gln Lys Glu Leu Thr
625                 630                 635                 640

Leu Thr Gln Glu Lys Leu Leu Asn Val Lys Glu His Glu Ile Glu Val
                645                 650                 655

Asp Asn Tyr Asn Arg Asn Lys Asn Thr Tyr Lys Phe Lys Asn Gln Pro
            660                 665                 670

Lys Asn Lys Val Asp Asp Lys Lys Leu Gln Arg Lys Tyr Val Phe Tyr
        675                 680                 685

Arg Asn Glu Ile Arg Gln Glu Ala Asn Trp Leu Ala Ser Asp Leu Ile
    690                 695                 700
```

```
His Phe Met Lys Asn Lys Ser Leu Trp Lys Gly Tyr Met His Asn Glu
705                 710                 715                 720


Leu Gln Ser Phe Leu Ala Phe Phe Glu Asp Lys Lys Asn Asp Cys Ile
                725                 730                 735


Ala Leu Leu Glu Thr Val Phe Asn Leu Lys Glu Asp Cys Ile Leu Thr
            740                 745                 750


Lys Gly Leu Lys Asn Leu Phe Leu Lys His Gly Asn Phe Ile Asp Phe
        755                 760                 765


Tyr Lys Glu Tyr Leu Lys Leu Lys Glu Asp Phe Leu Ser Thr Glu Ser
    770                 775                 780


Thr Phe Leu Glu Asn Gly Phe Ile Gly Leu Pro Pro Lys Ile Leu Lys
785                 790                 795                 800


Lys Glu Leu Ser Lys Arg Leu Lys Tyr Ile Phe Ile Val Phe Gln Lys
                805                 810                 815


Arg Gln Phe Ile Ile Lys Glu Leu Glu Glu Lys Lys Asn Asn Leu Tyr
            820                 825                 830


Ala Asp Ala Ile Asn Leu Ser Arg Gly Ile Phe Asp Glu Lys Pro Thr
        835                 840                 845


Met Ile Pro Phe Lys Lys Pro Asn Pro Asp Glu Phe Ala Ser Trp Phe
    850                 855                 860


Val Ala Ser Tyr Gln Tyr Asn Asn Tyr Gln Ser Phe Tyr Glu Leu Thr
865                 870                 875                 880


Pro Asp Ile Val Glu Arg Asp Lys Lys Lys Lys Lys Tyr Lys Asn Leu Arg
                885                 890                 895


Ala Ile Asn Lys Val Lys Ile Gln Asp Tyr Tyr Leu Lys Leu Met Val
            900                 905                 910
```

```
Asp Thr Leu Tyr Gln Asp Leu Phe Asn Gln Pro Leu Asp Lys Ser Leu
        915                 920                 925


Ser Asp Phe Tyr Val Ser Lys Ala Glu Arg Glu Lys Ile Lys Ala Asp
    930                 935                 940


Ala Lys Ala Tyr Gln Lys Leu Asn Asp Ser Ser Leu Trp Asn Lys Val
945                 950                 955                 960


Ile His Leu Ser Leu Gln Asn Asn Arg Ile Thr Ala Asn Pro Lys Leu
                965                 970                 975


Lys Asp Ile Gly Lys Tyr Lys Arg Ala Leu Gln Asp Glu Lys Ile Ala
            980                 985                 990


Thr Leu Leu Thr Tyr Asp Ala Arg  Thr Trp Thr Tyr Ala  Leu Gln Lys
        995                 1000                1005


Pro Glu  Lys Glu Asn Glu Asn  Asp Tyr Lys Glu Leu  His Tyr Thr
    1010                1015                1020


Ala Leu  Asn Met Glu Leu Gln  Glu Tyr Glu Lys Val  Arg Ser Lys
    1025                1030                1035


Glu Leu  Leu Lys Gln Val Gln  Glu Leu Glu Lys Lys  Ile Leu Asp
    1040                1045                1050


Lys Phe  Tyr Asp Phe Ser Asn  Asn Ala Ser His Pro  Glu Asp Leu
    1055                1060                1065


Glu Ile  Glu Asp Lys Lys Gly  Lys Arg His Pro Asn  Phe Lys Leu
    1070                1075                1080


Tyr Ile  Thr Lys Ala Leu Leu  Lys Asn Glu Ser Glu  Ile Ile Asn
    1085                1090                1095


Leu Glu  Asn Ile Asp Ile Glu  Ile Leu Leu Lys Tyr  Tyr Asp Tyr
    1100                1105                1110
```

```
Asn Thr  Glu Glu Leu Lys Glu  Lys Ile Lys Asn Met  Asp Glu Asp
    1115                1120                1125


Glu Lys  Ala Lys Ile Ile Asn  Thr Lys Glu Asn Tyr  Asn Lys Ile
    1130                1135                1140


Thr Asn  Val Leu Ile Lys Lys  Ala Leu Val Leu Ile  Ile Ile Arg
    1145                1150                1155


Asn Lys  Met Ala His Asn Gln  Tyr Pro Pro Lys Phe  Ile Tyr Asp
    1160                1165                1170


Leu Ala  Asn Arg Phe Val Pro  Lys Lys Glu Glu Glu  Tyr Phe Ala
    1175                1180                1185


Thr Tyr  Phe Asn Arg Val Phe  Glu Thr Ile Thr Lys  Glu Leu Trp
    1190                1195                1200


Glu Asn  Lys Glu Lys Lys Asp  Lys Thr Gln Val
    1205                1210


<210>  86
<211>  1119
<212>  PRT
<213>  Porphymonas gingivalis

<400>  86

Met Thr Glu Gln Asn Glu Lys Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1               5                   10                  15


Glu Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn
            20                  25                  30


Ala Tyr Ile Thr Leu Ala His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
        35                  40                  45


Ala Asp Ile Thr Asn Asp Glu Asp Ile Leu Phe Phe Lys Gly Gln Trp
    50                  55                  60


Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ala Arg Leu Arg Ser Leu
```

65 70 75 80

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

Leu Phe Glu Ser Gln Ser Ser Gly Asn Lys Ser Ser Lys Lys Lys Glu
100 105 110

Leu Ser Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
115 120 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
130 135 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Pro Glu Ser Ser Glu Leu
145 150 155 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
165 170 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val
180 185 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
195 200 205

Lys Tyr Gly Asn Asn Asp Asn Pro Phe Phe Lys His His Phe Val Asp
210 215 220

Arg Glu Gly Thr Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225 230 235 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
245 250 255

Phe Lys Gly Gly Thr Glu Ala Tyr Gln Gln Met Thr Asn Glu Val Phe
260 265 270

Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg

```
        275                 280                 285

Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys
    290                 295                 300

Pro Lys Ser Leu Tyr Asp Arg Leu Arg Glu Glu Asp Arg Ala Arg Phe
305                 310                 315                 320

Arg Val Pro Val Asp Ile Leu Ser Asp Glu Asp Asp Thr Asp Gly Thr
                325                 330                 335

Glu Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
            340                 345                 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
        355                 360                 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
    370                 375                 380

Lys Asn Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385                 390                 395                 400

Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
                405                 410                 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
            420                 425                 430

Lys Pro Tyr Ile Thr Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly
        435                 440                 445

Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln His Leu Trp Pro Ser
    450                 455                 460

Pro Glu Val Gly Ala Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465                 470                 475                 480

Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
```

```
                485                 490                 495

Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val
            500                 505                 510

Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
        515                 520                 525

Tyr Ala Val Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Arg Asp
    530                 535                 540

Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu
545                 550                 555                 560

Pro Arg Gln Met Ile Ala Ile Leu Ser Gln Glu His Lys Asp Met Glu
                565                 570                 575

Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His
            580                 585                 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly
        595                 600                 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Val Ala Asp Trp Leu
    610                 615                 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly
625                 630                 635                 640

Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
                645                 650                 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro
            660                 665                 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
        675                 680                 685

Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
```

```
    690                 695                 700

Arg Ser Tyr Leu Glu Ala Arg Lys Ala Phe Leu Gln Ser Ile Gly Arg
705                 710                 715                 720

Ser Asp Arg Val Glu Asn His Arg Phe Leu Leu Leu Lys Glu Pro Lys
                725                 730                 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Gly Glu Phe His Leu
            740                 745                 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
        755                 760                 765

Gly Tyr Asp Glu Val Gly Ser Tyr Lys Glu Val Gly Phe Met Ala Lys
    770                 775                 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Ser Lys Asp Arg Val Gln Pro
785                 790                 795                 800

Phe Tyr Asp Tyr Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
                805                 810                 815

Gly Arg Phe Leu Ser Lys Glu Lys Arg Ala Glu Glu Trp Glu Ser Gly
            820                 825                 830

Lys Glu Arg Phe Arg Leu Ala Lys Leu Lys Lys Glu Ile Leu Glu Ala
        835                 840                 845

Lys Glu His Pro Tyr His Asp Phe Lys Ser Trp Gln Lys Phe Glu Arg
    850                 855                 860

Glu Leu Arg Leu Val Lys Asn Gln Asp Ile Ile Thr Trp Met Met Cys
865                 870                 875                 880

Arg Asp Leu Met Glu Glu Asn Lys Val Glu Gly Leu Asp Thr Gly Thr
                885                 890                 895

Leu Tyr Leu Lys Asp Ile Arg Thr Asp Val Gln Glu Gln Gly Ser Leu
```

```
            900                 905                 910

Asn Val Leu Asn Arg Val Lys Pro Met Arg Leu Pro Val Val Val Tyr
        915                 920                 925

Arg Ala Asp Ser Arg Gly His Val His Lys Glu Gln Ala Pro Leu Ala
    930                 935                 940

Thr Val Tyr Ile Glu Glu Arg Asp Thr Lys Leu Leu Lys Gln Gly Asn
945                 950                 955                 960

Phe Lys Ser Phe Val Lys Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe
                965                 970                 975

Val Asp Thr Gly Ala Leu Ala Met Glu Gln Tyr Pro Ile Ser Lys Leu
            980                 985                 990

Arg Val Glu Tyr Glu Leu Ala Lys  Tyr Gln Thr Ala Arg  Val Cys Ala
        995                 1000                1005

Phe Glu  Gln Thr Leu Glu Leu  Glu Glu Ser Leu Leu  Thr Arg Tyr
    1010                1015                1020

Pro His  Leu Pro Asp Lys Asn  Phe Arg Lys Met Leu  Glu Ser Trp
    1025                1030                1035

Ser Asp  Pro Leu Leu Asp Lys  Trp Pro Asp Leu His  Gly Asn Val
    1040                1045                1050

Arg Leu  Leu Ile Ala Val Arg  Asn Ala Phe Ser His  Asn Gln Tyr
    1055                1060                1065

Pro Met  Tyr Asp Glu Thr Leu  Phe Ser Ser Ile Arg  Lys Tyr Asp
    1070                1075                1080

Pro Ser  Ser Pro Asp Ala Ile  Glu Glu Arg Met Gly  Leu Asn Ile
    1085                1090                1095

Ala His  Arg Leu Ser Glu Glu  Val Lys Gln Ala Lys  Glu Met Val
```

1100 1105 1110

Glu Arg Ile Ile Gln Ala
1115

<210> 87
<211> 1175
<212> PRT
<213> Porphyromonas sp.

<400> 87

Met Thr Glu Gln Ser Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1 5 10 15

Glu Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn
20 25 30

Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
35 40 45

Ala Asp Ile Thr Asn Asp Gln Asp Val Leu Ser Phe Lys Ala Leu Trp
50 55 60

Lys Asn Phe Asp Asn Asp Leu Glu Arg Lys Ser Arg Leu Arg Ser Leu
65 70 75 80

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

Leu Phe Glu Ser Lys Ser Ser Gly Asn Lys Ser Ser Lys Asn Lys Glu
100 105 110

Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
115 120 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
130 135 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Ser Glu Ser Ser Glu Leu
145 150 155 160

```
Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
                165                 170                 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val
            180                 185                 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
        195                 200                 205

Arg Tyr Gly His Asn Asp Asn Pro Ser Phe Lys His His Phe Val Asp
    210                 215                 220

Ser Glu Gly Met Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225                 230                 235                 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
                245                 250                 255

Phe Lys Gly Gly Thr Glu Thr Tyr Gln Gln Met Thr Asn Glu Val Phe
            260                 265                 270

Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
        275                 280                 285

Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys
    290                 295                 300

Pro Lys Pro Leu Tyr Asp Arg Leu Arg Glu Asp Asp Arg Ala Cys Phe
305                 310                 315                 320

Arg Val Pro Val Asp Ile Leu Pro Asp Glu Asp Asp Thr Asp Gly Gly
                325                 330                 335

Gly Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
            340                 345                 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
        355                 360                 365
```

```
Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
    370                 375                 380

Lys Met Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385                 390                 395                 400

Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
                405                 410                 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
            420                 425                 430

Lys Pro Tyr Ile Ser Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly
        435                 440                 445

Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln His Leu Trp Pro Ser
    450                 455                 460

Pro Glu Val Gly Thr Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465                 470                 475                 480

Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
                485                 490                 495

Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val
            500                 505                 510

Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
        515                 520                 525

Tyr Ala Ile Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Leu Lys
    530                 535                 540

Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu
545                 550                 555                 560

Pro Lys Gln Met Ile Gly Ile Leu Ser Gln Glu Arg Lys Asp Met Glu
                565                 570                 575
```

```
Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His
            580                 585                 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly
        595                 600                 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
    610                 615                 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly
625                 630                 635                 640

Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
                645                 650                 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro
            660                 665                 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
        675                 680                 685

Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
    690                 695                 700

Arg Ser Tyr Leu Arg Ala Arg Lys Ala Phe Leu Glu Arg Ile Gly Arg
705                 710                 715                 720

Ser Asp Arg Val Glu Asn Cys Pro Phe Leu Leu Leu Lys Glu Pro Lys
                725                 730                 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Gly Glu Phe His Leu
            740                 745                 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
        755                 760                 765

Gly Tyr Asp Glu Val Gly Ser Tyr Arg Glu Val Gly Phe Met Ala Lys
    770                 775                 780
```

```
Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Glu Asp Arg Val Gln Pro
785                 790                 795                 800

Phe Tyr Asp Ser Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
                805                 810                 815

Gly Arg Phe Leu Ser Lys Glu Asp Arg Ala Glu Glu Trp Glu Arg Gly
            820                 825                 830

Lys Glu Arg Phe Arg Asp Leu Glu Ala Trp Ser His Ser Ala Ala Arg
        835                 840                 845

Arg Ile Lys Asp Ala Phe Ala Gly Ile Glu Tyr Ala Ser Pro Gly Asn
    850                 855                 860

Lys Lys Lys Ile Glu Gln Leu Leu Arg Asp Leu Ser Leu Trp Glu Ala
865                 870                 875                 880

Phe Glu Ser Lys Leu Lys Val Arg Ala Asp Lys Ile Asn Leu Ala Lys
                885                 890                 895

Leu Lys Lys Glu Ile Leu Glu Ala Gln Glu His Pro Tyr His Asp Phe
            900                 905                 910

Lys Ser Trp Gln Lys Phe Glu Arg Glu Leu Arg Leu Val Lys Asn Gln
        915                 920                 925

Asp Ile Ile Thr Trp Met Met Cys Arg Asp Leu Met Glu Glu Asn Lys
    930                 935                 940

Val Glu Gly Leu Asp Thr Gly Thr Leu Tyr Leu Lys Asp Ile Arg Pro
945                 950                 955                 960

Asn Val Gln Glu Gln Gly Ser Leu Asn Val Leu Asn Arg Val Lys Pro
                965                 970                 975

Met Arg Leu Pro Val Val Val Tyr Arg Ala Asp Ser Arg Gly His Val
            980                 985                 990
```

```
His Lys Glu Glu Ala Pro Leu Ala  Thr Val Tyr Ile Glu  Glu Arg Asp
        995                 1000                1005

Thr Lys  Leu Leu Lys Gln Gly  Asn Phe Lys Ser Phe  Val Lys Asp
    1010                1015                1020

Arg Arg  Leu Asn Gly Leu Phe  Ser Phe Val Asp Thr  Gly Gly Leu
    1025                1030                1035

Ala Met  Glu Gln Tyr Pro Ile  Ser Lys Leu Arg Val  Glu Tyr Glu
    1040                1045                1050

Leu Ala  Lys Tyr Gln Thr Ala  Arg Val Cys Val Phe  Glu Leu Thr
    1055                1060                1065

Leu Arg  Leu Glu Glu Ser Leu  Leu Ser Arg Tyr Pro  His Leu Pro
    1070                1075                1080

Asp Glu  Ser Phe Arg Glu Met  Leu Glu Ser Trp Ser  Asp Pro Leu
    1085                1090                1095

Leu Ala  Lys Trp Pro Glu Leu  His Gly Lys Val Arg  Leu Leu Ile
    1100                1105                1110

Ala Val  Arg Asn Ala Phe Ser  His Asn Gln Tyr Pro  Met Tyr Asp
    1115                1120                1125

Glu Ala  Val Phe Ser Ser Ile  Arg Lys Tyr Asp Pro  Ser Ser Pro
    1130                1135                1140

Asp Ala  Ile Glu Glu Arg Met  Gly Leu Asn Ile Ala  His Arg Leu
    1145                1150                1155

Ser Glu  Glu Val Lys Gln Ala  Lys Glu Thr Val Glu  Arg Ile Ile
    1160                1165                1170

Gln Ala
    1175
```

<210> 88
<211> 1132
<212> PRT
<213> Prevotella intermedia

<400> 88

```
Met Glu Asp Asp Lys Lys Thr Lys Glu Ser Thr Asn Met Leu Asp Asn
1               5                  10                  15


Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn Val Tyr
            20                  25                  30


Ile Thr Val Asn His Ile Asn Lys Val Leu Glu Leu Lys Asn Lys Lys
        35                  40                  45


Asp Gln Asp Ile Ile Ile Asp Asn Asp Gln Asp Ile Leu Ala Ile Lys
    50                  55                  60


Thr His Trp Glu Lys Val Asn Gly Asp Leu Asn Lys Thr Glu Arg Leu
65                  70                  75                  80


Arg Glu Leu Met Thr Lys His Phe Pro Phe Leu Glu Thr Ala Ile Tyr
                85                  90                  95


Thr Lys Asn Lys Glu Asp Lys Glu Glu Val Lys Gln Glu Lys Gln Ala
            100                 105                 110


Lys Ala Gln Ser Phe Asp Ser Leu Lys His Cys Leu Phe Leu Phe Leu
        115                 120                 125


Glu Lys Leu Gln Glu Ala Arg Asn Tyr Tyr Ser His Tyr Lys Tyr Ser
    130                 135                 140


Glu Ser Thr Lys Glu Pro Met Leu Glu Lys Glu Leu Leu Lys Lys Met
145                 150                 155                 160


Tyr Asn Ile Phe Asp Asp Asn Ile Gln Leu Val Ile Lys Asp Tyr Gln
                165                 170                 175
```

```
His Asn Lys Asp Ile Asn Pro Asp Glu Asp Phe Lys His Leu Asp Arg
            180                 185                 190

Thr Glu Glu Glu Phe Asn Tyr Tyr Phe Thr Thr Asn Lys Lys Gly Asn
        195                 200                 205

Ile Thr Ala Ser Gly Leu Leu Phe Phe Val Ser Leu Phe Leu Glu Lys
    210                 215                 220

Lys Asp Ala Ile Trp Met Gln Gln Lys Leu Arg Gly Phe Lys Asp Asn
225                 230                 235                 240

Arg Glu Ser Lys Lys Lys Met Thr His Glu Val Phe Cys Arg Ser Arg
                245                 250                 255

Met Leu Leu Pro Lys Leu Arg Leu Glu Ser Thr Gln Thr Gln Asp Trp
            260                 265                 270

Ile Leu Leu Asp Met Leu Asn Glu Leu Ile Arg Cys Pro Lys Ser Leu
        275                 280                 285

Tyr Glu Arg Leu Gln Gly Glu Tyr Arg Lys Lys Phe Asn Val Pro Phe
    290                 295                 300

Asp Ser Ala Asp Glu Asp Tyr Asp Ala Glu Gln Glu Pro Phe Lys Asn
305                 310                 315                 320

Thr Leu Val Arg His Gln Asp Arg Phe Pro Tyr Phe Ala Leu Arg Tyr
                325                 330                 335

Phe Asp Tyr Asn Glu Ile Phe Thr Asn Leu Arg Phe Gln Ile Asp Leu
            340                 345                 350

Gly Thr Tyr His Phe Ser Ile Tyr Lys Lys Leu Ile Gly Gly Gln Lys
        355                 360                 365

Glu Asp Arg His Leu Thr His Lys Leu Tyr Gly Phe Glu Arg Ile Gln
    370                 375                 380
```

```
Glu Phe Ala Lys Gln Asn Arg Thr Asp Glu Trp Lys Ala Ile Val Lys
385                 390                 395                 400

Asp Phe Asp Thr Tyr Glu Thr Ser Glu Glu Pro Tyr Ile Ser Glu Thr
                405                 410                 415

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

Asn Asp Asn Asp Glu Ile Trp Pro Ser Leu Lys Thr Asn Gly Glu Asn
        435                 440                 445

Asn Glu Lys Arg Lys Tyr Lys Leu Asp Lys Gln Tyr Gln Ala Glu Ala
    450                 455                 460

Phe Leu Ser Val His Glu Leu Leu Pro Met Met Phe Tyr Tyr Leu Leu
465                 470                 475                 480

Leu Lys Lys Glu Glu Pro Asn Asn Asp Lys Lys Asn Ala Ser Ile Val
                485                 490                 495

Glu Gly Phe Ile Lys Arg Glu Ile Arg Asp Ile Tyr Lys Leu Tyr Asp
            500                 505                 510

Ala Phe Ala Asn Gly Glu Ile Asn Asn Ile Asp Asp Leu Glu Lys Tyr
        515                 520                 525

Cys Glu Asp Lys Gly Ile Pro Lys Arg His Leu Pro Lys Gln Met Val
    530                 535                 540

Ala Ile Leu Tyr Asp Glu His Lys Asp Met Ala Glu Glu Ala Lys Arg
545                 550                 555                 560

Lys Gln Lys Glu Met Val Lys Asp Thr Lys Lys Leu Leu Ala Thr Leu
                565                 570                 575

Glu Lys Gln Thr Gln Gly Glu Ile Glu Asp Gly Gly Arg Asn Ile Arg
            580                 585                 590
```

```
Leu Leu Lys Ser Gly Glu Ile Ala Arg Trp Leu Val Asn Asp Met Met
        595                 600                 605


Arg Phe Gln Pro Val Gln Lys Asp Asn Glu Gly Asn Pro Leu Asn Asn
    610                 615                 620


Ser Lys Ala Asn Ser Thr Glu Tyr Gln Met Leu Gln Arg Ser Leu Ala
625                 630                 635                 640


Leu Tyr Asn Lys Glu Glu Lys Pro Thr Arg Tyr Phe Arg Gln Val Asn
                645                 650                 655


Leu Ile Asn Ser Ser Asn Pro His Pro Phe Leu Lys Trp Thr Lys Trp
            660                 665                 670


Glu Glu Cys Asn Asn Ile Leu Ser Phe Tyr Arg Ser Tyr Leu Thr Lys
        675                 680                 685


Lys Ile Glu Phe Leu Asn Lys Leu Lys Pro Glu Asp Trp Glu Lys Asn
    690                 695                 700


Gln Tyr Phe Leu Lys Leu Lys Glu Pro Lys Thr Asn Arg Glu Thr Leu
705                 710                 715                 720


Val Gln Gly Trp Lys Asn Gly Phe Asn Leu Pro Arg Gly Ile Phe Thr
                725                 730                 735


Glu Pro Ile Arg Glu Trp Phe Lys Arg His Gln Asn Asp Ser Glu Glu
            740                 745                 750


Tyr Glu Lys Val Glu Thr Leu Asp Arg Val Gly Leu Val Thr Lys Val
        755                 760                 765


Ile Pro Leu Phe Phe Lys Lys Glu Asp Ser Lys Asp Lys Glu Glu Tyr
    770                 775                 780


Leu Lys Lys Asp Ala Gln Lys Glu Ile Asn Asn Cys Val Gln Pro Phe
785                 790                 795                 800
```

```
Tyr Gly Phe Pro Tyr Asn Val Gly Asn Ile His Lys Pro Asp Glu Lys
                805                 810                 815

Asp Phe Leu Pro Ser Glu Glu Arg Lys Lys Leu Trp Gly Asp Lys Lys
            820                 825                 830

Tyr Lys Phe Lys Gly Tyr Lys Ala Lys Val Lys Ser Lys Lys Leu Thr
        835                 840                 845

Asp Lys Glu Lys Glu Glu Tyr Arg Ser Tyr Leu Glu Phe Gln Ser Trp
    850                 855                 860

Asn Lys Phe Glu Arg Glu Leu Arg Leu Val Arg Asn Gln Asp Ile Val
865                 870                 875                 880

Thr Trp Leu Leu Cys Thr Glu Leu Ile Asp Lys Leu Lys Val Glu Gly
                885                 890                 895

Leu Asn Val Glu Glu Leu Lys Lys Leu Arg Leu Lys Asp Ile Asp Thr
            900                 905                 910

Asp Thr Ala Lys Gln Glu Lys Asn Asn Ile Leu Asn Arg Val Met Pro
        915                 920                 925

Met Gln Leu Pro Val Thr Val Tyr Glu Ile Asp Asp Ser His Asn Ile
    930                 935                 940

Val Lys Asp Arg Pro Leu His Thr Val Tyr Ile Glu Glu Thr Lys Thr
945                 950                 955                 960

Lys Leu Leu Lys Gln Gly Asn Phe Lys Ala Leu Val Lys Asp Arg Arg
                965                 970                 975

Leu Asn Gly Leu Phe Ser Phe Val Asp Thr Ser Ser Glu Thr Glu Leu
            980                 985                 990

Lys Ser Asn Pro Ile Ser Lys Ser  Leu Val Glu Tyr Glu  Leu Gly Glu
        995                 1000                1005
```

```
Tyr Gln  Asn Ala Arg Ile Glu  Thr Ile Lys Asp Met  Leu Leu Leu
    1010                1015                1020


Glu Glu  Thr Leu Ile Glu Lys  Tyr Lys Thr Leu Pro  Thr Asp Asn
    1025                1030                1035


Phe Ser  Asp Met Leu Asn Gly  Trp Leu Glu Gly Lys  Asp Glu Ala
    1040                1045                1050


Asp Lys  Ala Arg Phe Gln Asn  Asp Val Lys Leu Leu  Val Ala Val
    1055                1060                1065


Arg Asn  Ala Phe Ser His Asn  Gln Tyr Pro Met Arg  Asn Arg Ile
    1070                1075                1080


Ala Phe  Ala Asn Ile Asn Pro  Phe Ser Leu Ser Ser  Ala Asp Thr
    1085                1090                1095


Ser Glu  Glu Lys Lys Leu Asp  Ile Ala Asn Gln Leu  Lys Asp Lys
    1100                1105                1110


Thr His  Lys Ile Ile Lys Arg  Ile Ile Glu Ile Glu  Lys Pro Ile
    1115                1120                1125


Glu Thr  Lys Glu
    1130


<210>  89
<211>  1134
<212>  PRT
<213>  Prevotella intermedia

<400>  89

Met Lys Met Glu Asp Asp Lys Lys Thr Lys Glu Ser Thr Asn Met Leu
1               5                  10                  15


Asp Asn Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn
            20                  25                  30
```

```
Val Tyr Ile Thr Val Asn His Ile Asn Lys Val Leu Glu Leu Lys Asn
        35                  40                  45

Lys Lys Asp Gln Asp Ile Ile Ile Asp Asn Asp Gln Asp Ile Leu Ala
    50                  55                  60

Ile Lys Thr His Trp Glu Lys Val Asn Gly Asp Leu Asn Lys Thr Glu
65                  70                  75                  80

Arg Leu Arg Glu Leu Met Thr Lys His Phe Pro Phe Leu Glu Thr Ala
                85                  90                  95

Ile Tyr Thr Lys Asn Lys Glu Asp Lys Glu Glu Val Lys Gln Glu Lys
            100                 105                 110

Gln Ala Lys Ala Gln Ser Phe Asp Ser Leu Lys His Cys Leu Phe Leu
        115                 120                 125

Phe Leu Glu Lys Leu Gln Glu Ala Arg Asn Tyr Tyr Ser His Tyr Lys
    130                 135                 140

Tyr Ser Glu Ser Thr Lys Glu Pro Met Leu Glu Lys Glu Leu Leu Lys
145                 150                 155                 160

Lys Met Tyr Asn Ile Phe Asp Asp Asn Ile Gln Leu Val Ile Lys Asp
                165                 170                 175

Tyr Gln His Asn Lys Asp Ile Asn Pro Asp Glu Asp Phe Lys His Leu
            180                 185                 190

Asp Arg Thr Glu Glu Glu Phe Asn Tyr Tyr Phe Thr Thr Asn Lys Lys
        195                 200                 205

Gly Asn Ile Thr Ala Ser Gly Leu Leu Phe Phe Val Ser Leu Phe Leu
    210                 215                 220

Glu Lys Lys Asp Ala Ile Trp Met Gln Gln Lys Leu Arg Gly Phe Lys
225                 230                 235                 240
```

```
Asp Asn Arg Glu Ser Lys Lys Lys Met Thr His Glu Val Phe Cys Arg
                245                 250                 255

Ser Arg Met Leu Leu Pro Lys Leu Arg Leu Glu Ser Thr Gln Thr Gln
            260                 265                 270

Asp Trp Ile Leu Leu Asp Met Leu Asn Glu Leu Ile Arg Cys Pro Lys
        275                 280                 285

Ser Leu Tyr Glu Arg Leu Gln Gly Glu Tyr Arg Lys Lys Phe Asn Val
    290                 295                 300

Pro Phe Asp Ser Ala Asp Glu Asp Tyr Asp Ala Glu Gln Glu Pro Phe
305                 310                 315                 320

Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe Pro Tyr Phe Ala Leu
                325                 330                 335

Arg Tyr Phe Asp Tyr Asn Glu Ile Phe Thr Asn Leu Arg Phe Gln Ile
            340                 345                 350

Asp Leu Gly Thr Tyr His Phe Ser Ile Tyr Lys Lys Leu Ile Gly Gly
        355                 360                 365

Gln Lys Glu Asp Arg His Leu Thr His Lys Leu Tyr Gly Phe Glu Arg
    370                 375                 380

Ile Gln Glu Phe Ala Lys Gln Asn Arg Thr Asp Glu Trp Lys Ala Ile
385                 390                 395                 400

Val Lys Asp Phe Asp Thr Tyr Glu Thr Ser Glu Glu Pro Tyr Ile Ser
                405                 410                 415

Glu Thr Ala Pro His Tyr His Leu Glu Asn Gln Lys Ile Gly Ile Arg
            420                 425                 430

Phe Arg Asn Asp Asn Asp Glu Ile Trp Pro Ser Leu Lys Thr Asn Gly
        435                 440                 445
```

```
Glu Asn Asn Glu Lys Arg Lys Tyr Lys Leu Asp Lys Gln Tyr Gln Ala
    450                 455                 460

Glu Ala Phe Leu Ser Val His Glu Leu Leu Pro Met Met Phe Tyr Tyr
465                 470                 475                 480

Leu Leu Leu Lys Lys Glu Glu Pro Asn Asn Asp Lys Lys Asn Ala Ser
                485                 490                 495

Ile Val Glu Gly Phe Ile Lys Arg Glu Ile Arg Asp Ile Tyr Lys Leu
            500                 505                 510

Tyr Asp Ala Phe Ala Asn Gly Glu Ile Asn Asn Ile Asp Asp Leu Glu
        515                 520                 525

Lys Tyr Cys Glu Asp Lys Gly Ile Pro Lys Arg His Leu Pro Lys Gln
    530                 535                 540

Met Val Ala Ile Leu Tyr Asp Glu His Lys Asp Met Ala Glu Glu Ala
545                 550                 555                 560

Lys Arg Lys Gln Lys Glu Met Val Lys Asp Thr Lys Lys Leu Leu Ala
                565                 570                 575

Thr Leu Glu Lys Gln Thr Gln Gly Glu Ile Glu Asp Gly Gly Arg Asn
            580                 585                 590

Ile Arg Leu Leu Lys Ser Gly Glu Ile Ala Arg Trp Leu Val Asn Asp
        595                 600                 605

Met Met Arg Phe Gln Pro Val Gln Lys Asp Asn Glu Gly Asn Pro Leu
    610                 615                 620

Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Gln Met Leu Gln Arg Ser
625                 630                 635                 640

Leu Ala Leu Tyr Asn Lys Glu Glu Lys Pro Thr Arg Tyr Phe Arg Gln
                645                 650                 655
```

```
Val Asn Leu Ile Asn Ser Ser Asn Pro His Pro Phe Leu Lys Trp Thr
            660                 665                 670

Lys Trp Glu Glu Cys Asn Asn Ile Leu Ser Phe Tyr Arg Ser Tyr Leu
        675                 680                 685

Thr Lys Lys Ile Glu Phe Leu Asn Lys Leu Lys Pro Glu Asp Trp Glu
    690                 695                 700

Lys Asn Gln Tyr Phe Leu Lys Leu Lys Glu Pro Lys Thr Asn Arg Glu
705                 710                 715                 720

Thr Leu Val Gln Gly Trp Lys Asn Gly Phe Asn Leu Pro Arg Gly Ile
                725                 730                 735

Phe Thr Glu Pro Ile Arg Glu Trp Phe Lys Arg His Gln Asn Asp Ser
            740                 745                 750

Glu Glu Tyr Glu Lys Val Glu Thr Leu Asp Arg Val Gly Leu Val Thr
        755                 760                 765

Lys Val Ile Pro Leu Phe Phe Lys Lys Glu Asp Ser Lys Asp Lys Glu
    770                 775                 780

Glu Tyr Leu Lys Lys Asp Ala Gln Lys Glu Ile Asn Asn Cys Val Gln
785                 790                 795                 800

Pro Phe Tyr Gly Phe Pro Tyr Asn Val Gly Asn Ile His Lys Pro Asp
                805                 810                 815

Glu Lys Asp Phe Leu Pro Ser Glu Glu Arg Lys Lys Leu Trp Gly Asp
            820                 825                 830

Lys Lys Tyr Lys Phe Lys Gly Tyr Lys Ala Lys Val Lys Ser Lys Lys
        835                 840                 845

Leu Thr Asp Lys Glu Lys Glu Glu Tyr Arg Ser Tyr Leu Glu Phe Gln
    850                 855                 860
```

```
Ser Trp Asn Lys Phe Glu Arg Glu Leu Arg Leu Val Arg Asn Gln Asp
865                 870                 875                 880


Ile Val Thr Trp Leu Leu Cys Thr Glu Leu Ile Asp Lys Leu Lys Val
                885                 890                 895


Glu Gly Leu Asn Val Glu Glu Leu Lys Lys Leu Arg Leu Lys Asp Ile
            900                 905                 910


Asp Thr Asp Thr Ala Lys Gln Glu Lys Asn Asn Ile Leu Asn Arg Val
        915                 920                 925


Met Pro Met Gln Leu Pro Val Thr Val Tyr Glu Ile Asp Asp Ser His
    930                 935                 940


Asn Ile Val Lys Asp Arg Pro Leu His Thr Val Tyr Ile Glu Glu Thr
945                 950                 955                 960


Lys Thr Lys Leu Leu Lys Gln Gly Asn Phe Lys Ala Leu Val Lys Asp
                965                 970                 975


Arg Arg Leu Asn Gly Leu Phe Ser Phe Val Asp Thr Ser Ser Glu Thr
            980                 985                 990


Glu Leu Lys Ser Asn Pro Ile Ser  Lys Ser Leu Val Glu  Tyr Glu Leu
        995                 1000                1005


Gly Glu  Tyr Gln Asn Ala Arg  Ile Glu Thr Ile Lys  Asp Met Leu
    1010                1015                1020


Leu Leu  Glu Glu Thr Leu Ile  Glu Lys Tyr Lys Thr  Leu Pro Thr
    1025                1030                1035


Asp Asn  Phe Ser Asp Met Leu  Asn Gly Trp Leu Glu  Gly Lys Asp
    1040                1045                1050


Glu Ala  Asp Lys Ala Arg Phe  Gln Asn Asp Val Lys  Leu Leu Val
    1055                1060                1065
```

```
Ala Val  Arg Asn Ala Phe Ser  His Asn Gln Tyr Pro  Met Arg Asn
    1070                1075                1080


Arg Ile  Ala Phe Ala Asn Ile  Asn Pro Phe Ser Leu  Ser Ser Ala
    1085                1090                1095


Asp Thr  Ser Glu Glu Lys Lys  Leu Asp Ile Ala Asn  Gln Leu Lys
    1100                1105                1110


Asp Lys  Thr His Lys Ile Ile  Lys Arg Ile Ile Glu  Ile Glu Lys
    1115                1120                1125


Pro Ile  Glu Thr Lys Glu
    1130


<210>  90
<211>  1133
<212>  PRT
<213>  Prevotella intermedia

<400>  90

Met Glu Asp Asp Lys Lys Thr Thr Asp Ser Ile Ser Tyr Glu Leu Lys
1               5                  10                  15


Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn Val
            20                  25                  30


Tyr Ile Thr Val Asn His Ile Asn Lys Val Leu Glu Leu Lys Asn Lys
        35                  40                  45


Lys Asp Gln Asp Ile Ile Ile Asp Asn Asp Gln Asp Ile Leu Ala Ile
    50                  55                  60


Lys Thr His Trp Glu Lys Val Asn Gly Asp Leu Asn Lys Thr Glu Arg
65                  70                  75                  80


Leu Arg Glu Leu Met Thr Lys His Phe Pro Phe Leu Glu Thr Ala Ile
                85                  90                  95


Tyr Ser Lys Asn Lys Glu Asp Lys Glu Glu Val Lys Gln Glu Lys Gln
```

```
            100                 105                 110

Ala Lys Ala Gln Ser Phe Asp Ser Leu Lys His Cys Leu Phe Leu Phe
        115                 120                 125

Leu Glu Lys Leu Gln Glu Thr Arg Asn Tyr Tyr Ser His Tyr Lys Tyr
    130                 135                 140

Ser Glu Ser Thr Lys Glu Pro Met Leu Glu Lys Glu Leu Leu Lys Lys
145                 150                 155                 160

Met Tyr Asn Ile Phe Asp Asp Asn Ile Gln Leu Val Ile Lys Asp Tyr
                165                 170                 175

Gln His Asn Lys Asp Ile Asn Pro Asp Glu Asp Phe Lys His Leu Asp
            180                 185                 190

Arg Thr Glu Glu Asp Phe Asn Tyr Tyr Phe Thr Arg Asn Lys Lys Gly
        195                 200                 205

Asn Ile Thr Glu Ser Gly Leu Leu Phe Phe Val Ser Leu Phe Leu Glu
    210                 215                 220

Lys Lys Asp Ala Ile Trp Met Gln Gln Lys Leu Arg Gly Phe Lys Asp
225                 230                 235                 240

Asn Arg Glu Ser Lys Lys Lys Met Thr His Glu Val Phe Cys Arg Ser
                245                 250                 255

Arg Met Leu Leu Pro Lys Leu Arg Leu Glu Ser Thr Gln Thr Gln Asp
            260                 265                 270

Trp Ile Leu Leu Asp Met Leu Asn Glu Leu Ile Arg Cys Pro Lys Ser
        275                 280                 285

Leu Tyr Glu Arg Leu Gln Gly Glu Asp Arg Glu Lys Phe Lys Val Pro
    290                 295                 300

Phe Asp Pro Ala Asp Glu Asp Tyr Asp Ala Glu Gln Glu Pro Phe Lys
```

305 310 315 320

Asn Thr Leu Val Arg His Gln Asp Arg Phe Pro Tyr Phe Ala Leu Arg
325 330 335

Tyr Phe Asp Tyr Asn Glu Ile Phe Thr Asn Leu Arg Phe Gln Ile Asp
340 345 350

Leu Gly Thr Phe His Phe Ser Ile Tyr Lys Lys Leu Ile Gly Gly Gln
355 360 365

Lys Glu Asp Arg His Leu Thr His Lys Leu Tyr Gly Phe Glu Arg Ile
370 375 380

Gln Glu Phe Ala Lys Gln Asn Arg Pro Asp Glu Trp Lys Ala Ile Val
385 390 395 400

Lys Asp Leu Asp Thr Tyr Glu Thr Ser Asn Glu Arg Tyr Ile Ser Glu
405 410 415

Thr Thr Pro His Tyr His Leu Glu Asn Gln Lys Ile Gly Ile Arg Phe
420 425 430

Arg Asn Asp Asn Asp Glu Ile Trp Pro Ser Leu Lys Thr Asn Gly Glu
435 440 445

Asn Asn Glu Lys Ser Lys Tyr Lys Leu Asp Lys Gln Tyr Gln Ala Glu
450 455 460

Ala Phe Leu Ser Val His Glu Leu Leu Pro Met Met Phe Tyr Tyr Leu
465 470 475 480

Leu Leu Lys Lys Glu Glu Pro Asn Asn Asp Lys Lys Asn Ala Ser Ile
485 490 495

Val Glu Gly Phe Ile Lys Arg Glu Ile Arg Asp Met Tyr Lys Leu Tyr
500 505 510

Asp Ala Phe Ala Asn Gly Glu Ile Asn Asn Ile Asp Asp Leu Glu Lys

```
        515                 520                 525

Tyr Cys Glu Asp Lys Gly Ile Pro Lys Arg His Leu Pro Lys Gln Met
    530                 535                 540

Val Ala Ile Leu Tyr Asp Glu His Lys Asp Met Val Lys Glu Ala Lys
545                 550                 555                 560

Arg Lys Gln Arg Lys Met Val Lys Asp Thr Glu Lys Leu Leu Ala Ala
                565                 570                 575

Leu Glu Lys Gln Thr Gln Glu Lys Thr Glu Asp Gly Gly Arg Asn Ile
            580                 585                 590

Arg Leu Leu Lys Ser Gly Glu Ile Ala Arg Trp Leu Val Asn Asp Met
        595                 600                 605

Met Arg Phe Gln Pro Val Gln Lys Asp Asn Glu Gly Asn Pro Leu Asn
    610                 615                 620

Asn Ser Lys Ala Asn Ser Thr Glu Tyr Gln Met Leu Gln Arg Ser Leu
625                 630                 635                 640

Ala Leu Tyr Asn Lys Glu Glu Lys Pro Thr Arg Tyr Phe Arg Gln Val
                645                 650                 655

Asn Leu Ile Asn Ser Ser Asn Pro His Pro Phe Leu Lys Trp Thr Lys
            660                 665                 670

Trp Glu Glu Cys Asn Asn Ile Leu Ser Phe Tyr Arg Ser Tyr Leu Thr
        675                 680                 685

Lys Lys Ile Glu Phe Leu Asn Lys Leu Lys Pro Glu Asp Trp Glu Lys
    690                 695                 700

Asn Gln Tyr Phe Leu Lys Leu Lys Glu Pro Lys Thr Asn Arg Glu Thr
705                 710                 715                 720

Leu Val Gln Gly Trp Lys Asn Gly Phe Asn Leu Pro Arg Gly Ile Phe
```

```
                725                 730                 735

Thr Glu Pro Ile Arg Glu Trp Phe Lys Arg His Gln Asn Asp Ser Lys
            740                 745                 750

Glu Tyr Glu Lys Val Glu Ala Leu Asp Arg Val Gly Leu Val Thr Lys
        755                 760                 765

Val Ile Pro Leu Phe Phe Lys Lys Glu Asp Ser Lys Asp Lys Glu Glu
    770                 775                 780

Asp Leu Lys Lys Asp Ala Gln Lys Glu Ile Asn Asn Cys Val Gln Pro
785                 790                 795                 800

Phe Tyr Ser Phe Pro Tyr Asn Val Gly Asn Ile His Lys Pro Asp Glu
                805                 810                 815

Lys Asp Phe Leu His Arg Glu Glu Arg Ile Glu Leu Trp Asp Lys Lys
            820                 825                 830

Lys Asp Lys Phe Lys Gly Tyr Lys Ala Lys Val Lys Ser Lys Lys Leu
        835                 840                 845

Thr Asp Lys Glu Lys Glu Glu Tyr Arg Ser Tyr Leu Glu Phe Gln Ser
    850                 855                 860

Trp Asn Lys Phe Glu Arg Glu Leu Arg Leu Val Arg Asn Gln Asp Ile
865                 870                 875                 880

Val Thr Trp Leu Leu Cys Thr Glu Leu Ile Asp Lys Leu Lys Val Glu
                885                 890                 895

Gly Leu Asn Val Glu Glu Leu Lys Lys Leu Arg Leu Lys Asp Ile Asp
            900                 905                 910

Thr Asp Thr Ala Lys Gln Glu Lys Asn Asn Ile Leu Asn Arg Val Met
        915                 920                 925

Pro Met Gln Leu Pro Val Thr Val Tyr Glu Ile Asp Asp Ser His Asn
```

```
    930                 935                 940

Ile Val Lys Asp Arg Pro Leu His Thr Val Tyr Ile Glu Glu Thr Lys
945                 950                 955                 960

Thr Lys Leu Leu Lys Gln Gly Asn Phe Lys Ala Leu Val Lys Asp Arg
                965                 970                 975

Arg Leu Asn Gly Leu Phe Ser Phe Val Asp Thr Ser Ser Glu Ala Glu
            980                 985                 990

Leu Lys Ser Asn Pro Ile Ser Lys  Ser Leu Val Glu Tyr  Glu Leu Gly
        995                 1000                1005

Glu Tyr  Gln Asn Ala Arg Ile  Glu Thr Ile Lys Asp  Met Leu Leu
    1010                1015                1020

Leu Glu  Glu Thr Leu Ile Glu  Lys Tyr Lys Asn Leu  Pro Thr Asp
    1025                1030                1035

Asn Phe  Ser Asp Met Leu Asn  Gly Trp Leu Glu Gly  Lys Asp Glu
    1040                1045                1050

Ala Asp  Lys Ala Arg Phe Gln  Asn Asp Val Lys Leu  Leu Val Ala
    1055                1060                1065

Val Arg  Asn Ala Phe Ser His  Asn Gln Tyr Pro Met  Arg Asn Arg
    1070                1075                1080

Ile Ala  Phe Ala Asn Ile Asn  Pro Phe Ser Leu Ser  Ser Ala Asp
    1085                1090                1095

Thr Ser  Glu Glu Lys Lys Leu  Asp Ile Ala Asn Gln  Leu Lys Asp
    1100                1105                1110

Lys Thr  His Lys Ile Ile Lys  Arg Ile Ile Glu Ile  Glu Lys Pro
    1115                1120                1125

Ile Glu  Thr Lys Glu
```

1130

<210> 91
<211> 1127
<212> PRT
<213> Prevotella buccae

<400> 91

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

Phe Ala Phe Pro Lys Tyr Ile Thr Ile Met Glu Asn Lys Glu Lys Pro
20 25 30

Glu Pro Ile Tyr Tyr Glu Leu Thr Asp Lys His Phe Trp Ala Ala Phe
35 40 45

Leu Asn Leu Ala Arg His Asn Val Tyr Thr Thr Ile Asn His Ile Asn
50 55 60

Arg Arg Leu Glu Ile Ala Glu Leu Lys Asp Asp Gly Tyr Met Met Gly
65 70 75 80

Ile Lys Gly Ser Trp Asn Glu Gln Ala Lys Lys Leu Asp Lys Lys Val
85 90 95

Arg Leu Arg Asp Leu Ile Met Lys His Phe Pro Phe Leu Glu Ala Ala
100 105 110

Ala Tyr Glu Met Thr Asn Ser Lys Ser Pro Asn Asn Lys Glu Gln Arg
115 120 125

Glu Lys Glu Gln Ser Glu Ala Leu Ser Leu Asn Asn Leu Lys Asn Val
130 135 140

Leu Phe Ile Phe Leu Glu Lys Leu Gln Val Leu Arg Asn Tyr Tyr Ser
145 150 155 160

His Tyr Lys Tyr Ser Glu Glu Ser Pro Lys Pro Ile Phe Glu Thr Ser
165 170 175

```
Leu Leu Lys Asn Met Tyr Lys Val Phe Asp Ala Asn Val Arg Leu Val
            180                 185                 190

Lys Arg Asp Tyr Met His His Glu Asn Ile Asp Met Gln Arg Asp Phe
        195                 200                 205

Thr His Leu Asn Arg Lys Lys Gln Val Gly Arg Thr Lys Asn Ile Ile
    210                 215                 220

Asp Ser Pro Asn Phe His Tyr His Phe Ala Asp Lys Glu Gly Asn Met
225                 230                 235                 240

Thr Ile Ala Gly Leu Leu Phe Phe Val Ser Leu Phe Leu Asp Lys Lys
                245                 250                 255

Asp Ala Ile Trp Met Gln Lys Lys Leu Lys Gly Phe Lys Asp Gly Arg
            260                 265                 270

Asn Leu Arg Glu Gln Met Thr Asn Glu Val Phe Cys Arg Ser Arg Ile
        275                 280                 285

Ser Leu Pro Lys Leu Lys Leu Glu Asn Val Gln Thr Lys Asp Trp Met
    290                 295                 300

Gln Leu Asp Met Leu Asn Glu Leu Val Arg Cys Pro Lys Ser Leu Tyr
305                 310                 315                 320

Glu Arg Leu Arg Glu Lys Asp Arg Glu Ser Phe Lys Val Pro Phe Asp
                325                 330                 335

Ile Phe Ser Asp Asp Tyr Asn Ala Glu Glu Glu Pro Phe Lys Asn Thr
            340                 345                 350

Leu Val Arg His Gln Asp Arg Phe Pro Tyr Phe Val Leu Arg Tyr Phe
        355                 360                 365

Asp Leu Asn Glu Ile Phe Glu Gln Leu Arg Phe Gln Ile Asp Leu Gly
    370                 375                 380
```

```
Thr Tyr His Phe Ser Ile Tyr Asn Lys Arg Ile Gly Asp Glu Asp Glu
385                 390                 395                 400

Val Arg His Leu Thr His His Leu Tyr Gly Phe Ala Arg Ile Gln Asp
                405                 410                 415

Phe Ala Pro Gln Asn Gln Pro Glu Glu Trp Arg Lys Leu Val Lys Asp
            420                 425                 430

Leu Asp His Phe Glu Thr Ser Gln Glu Pro Tyr Ile Ser Lys Thr Ala
        435                 440                 445

Pro His Tyr His Leu Glu Asn Glu Lys Ile Gly Ile Lys Phe Cys Ser
    450                 455                 460

Ala His Asn Asn Leu Phe Pro Ser Leu Gln Thr Asp Lys Thr Cys Asn
465                 470                 475                 480

Gly Arg Ser Lys Phe Asn Leu Gly Thr Gln Phe Thr Ala Glu Ala Phe
                485                 490                 495

Leu Ser Val His Glu Leu Leu Pro Met Met Phe Tyr Tyr Leu Leu Leu
            500                 505                 510

Thr Lys Asp Tyr Ser Arg Lys Glu Ser Ala Asp Lys Val Glu Gly Ile
        515                 520                 525

Ile Arg Lys Glu Ile Ser Asn Ile Tyr Ala Ile Tyr Asp Ala Phe Ala
    530                 535                 540

Asn Asn Glu Ile Asn Ser Ile Ala Asp Leu Thr Arg Arg Leu Gln Asn
545                 550                 555                 560

Thr Asn Ile Leu Gln Gly His Leu Pro Lys Gln Met Ile Ser Ile Leu
                565                 570                 575

Lys Gly Arg Gln Lys Asp Met Gly Lys Glu Ala Glu Arg Lys Ile Gly
            580                 585                 590
```

```
Glu Met Ile Asp Asp Thr Gln Arg Arg Leu Asp Leu Leu Cys Lys Gln
        595                 600                 605

Thr Asn Gln Lys Ile Arg Ile Gly Lys Arg Asn Ala Gly Leu Leu Lys
    610                 615                 620

Ser Gly Lys Ile Ala Asp Trp Leu Val Asn Asp Met Met Arg Phe Gln
625                 630                 635                 640

Pro Val Gln Lys Asp Gln Asn Asn Ile Pro Ile Asn Asn Ser Lys Ala
                645                 650                 655

Asn Ser Thr Glu Tyr Arg Met Leu Gln Arg Ala Leu Ala Leu Phe Gly
            660                 665                 670

Ser Glu Asn Phe Arg Leu Lys Ala Tyr Phe Asn Gln Met Asn Leu Val
        675                 680                 685

Gly Asn Asp Asn Pro His Pro Phe Leu Ala Glu Thr Gln Trp Glu His
    690                 695                 700

Gln Thr Asn Ile Leu Ser Phe Tyr Arg Asn Tyr Leu Glu Ala Arg Lys
705                 710                 715                 720

Lys Tyr Leu Lys Gly Leu Lys Pro Gln Asn Trp Lys Gln Tyr Gln His
                725                 730                 735

Phe Leu Ile Leu Lys Val Gln Lys Thr Asn Arg Asn Thr Leu Val Thr
            740                 745                 750

Gly Trp Lys Asn Ser Phe Asn Leu Pro Arg Gly Ile Phe Thr Gln Pro
        755                 760                 765

Ile Arg Glu Trp Phe Glu Lys His Asn Asn Ser Lys Arg Ile Tyr Asp
    770                 775                 780

Gln Ile Leu Ser Phe Asp Arg Val Gly Phe Val Ala Lys Ala Ile Pro
785                 790                 795                 800
```

```
Leu Tyr Phe Ala Glu Glu Tyr Lys Asp Asn Val Gln Pro Phe Tyr Asp
                805                 810                 815

Tyr Pro Phe Asn Ile Gly Asn Arg Leu Lys Pro Lys Lys Arg Gln Phe
            820                 825                 830

Leu Asp Lys Lys Glu Arg Val Glu Leu Trp Gln Lys Asn Lys Glu Leu
        835                 840                 845

Phe Lys Asn Tyr Pro Ser Glu Lys Lys Lys Thr Asp Leu Ala Tyr Leu
    850                 855                 860

Asp Phe Leu Ser Trp Lys Lys Phe Glu Arg Glu Leu Arg Leu Ile Lys
865                 870                 875                 880

Asn Gln Asp Ile Val Thr Trp Leu Met Phe Lys Glu Leu Phe Asn Met
                885                 890                 895

Ala Thr Val Glu Gly Leu Lys Ile Gly Glu Ile His Leu Arg Asp Ile
            900                 905                 910

Asp Thr Asn Thr Ala Asn Glu Glu Ser Asn Asn Ile Leu Asn Arg Ile
        915                 920                 925

Met Pro Met Lys Leu Pro Val Lys Thr Tyr Glu Thr Asp Asn Lys Gly
    930                 935                 940

Asn Ile Leu Lys Glu Arg Pro Leu Ala Thr Phe Tyr Ile Glu Glu Thr
945                 950                 955                 960

Glu Thr Lys Val Leu Lys Gln Gly Asn Phe Lys Ala Leu Val Lys Asp
                965                 970                 975

Arg Arg Leu Asn Gly Leu Phe Ser Phe Ala Glu Thr Thr Asp Leu Asn
            980                 985                 990

Leu Glu Glu His Pro Ile Ser Lys  Leu Ser Val Asp Leu  Glu Leu Ile
        995                 1000                1005
```

```
Lys Tyr  Gln Thr Thr Arg Ile  Ser Ile Phe Glu Met  Thr Leu Gly
    1010                1015                1020


Leu Glu  Lys Lys Leu Ile Asp  Lys Tyr Ser Thr Leu  Pro Thr Asp
    1025                1030                1035


Ser Phe  Arg Asn Met Leu Glu  Arg Trp Leu Gln Cys  Lys Ala Asn
    1040                1045                1050


Arg Pro  Glu Leu Lys Asn Tyr  Val Asn Ser Leu Ile  Ala Val Arg
    1055                1060                1065


Asn Ala  Phe Ser His Asn Gln  Tyr Pro Met Tyr Asp  Ala Thr Leu
    1070                1075                1080


Phe Ala  Glu Val Lys Lys Phe  Thr Leu Phe Pro Ser  Val Asp Thr
    1085                1090                1095


Lys Lys  Ile Glu Leu Asn Ile  Ala Pro Gln Leu Leu  Glu Ile Val
    1100                1105                1110


Gly Lys  Ala Ile Lys Glu Ile  Glu Lys Ser Glu Asn  Lys Asn
    1115                1120                1125


<210>  92
<211>  1125
<212>  PRT
<213>  Prevotella pallens

<400>  92

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


Lys Asp Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His
            20                  25                  30


Asn Val Tyr Ile Thr Val Asn His Ile Asn Lys Ile Leu Gly Glu Gly
        35                  40                  45
```

```
Glu Ile Asn Arg Asp Gly Tyr Glu Asn Thr Leu Glu Lys Ser Trp Asn
    50                  55                  60

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

Lys His Phe Pro Phe Leu Glu Val Thr Thr Tyr Gln Arg Asn Ser Ala
                85                  90                  95

Asp Thr Thr Lys Gln Lys Glu Glu Lys Gln Ala Glu Ala Gln Ser Leu
            100                 105                 110

Glu Ser Leu Lys Lys Ser Phe Phe Val Phe Ile Tyr Lys Leu Arg Asp
        115                 120                 125

Leu Arg Asn His Tyr Ser His Tyr Lys His Ser Lys Ser Leu Glu Arg
    130                 135                 140

Pro Lys Phe Glu Glu Asp Leu Gln Glu Lys Met Tyr Asn Ile Phe Asp
145                 150                 155                 160

Ala Ser Ile Gln Leu Val Lys Glu Asp Tyr Lys His Asn Thr Asp Ile
                165                 170                 175

Lys Thr Glu Glu Asp Phe Lys His Leu Asp Arg Lys Gly Gln Phe Lys
            180                 185                 190

Tyr Ser Phe Ala Asp Asn Glu Gly Asn Ile Thr Glu Ser Gly Leu Leu
        195                 200                 205

Phe Phe Val Ser Leu Phe Leu Glu Lys Lys Asp Ala Ile Trp Val Gln
    210                 215                 220

Lys Lys Leu Glu Gly Phe Lys Cys Ser Asn Glu Ser Tyr Gln Lys Met
225                 230                 235                 240

Thr Asn Glu Val Phe Cys Arg Ser Arg Met Leu Leu Pro Lys Leu Arg
                245                 250                 255
```

```
Leu Gln Ser Thr Gln Thr Gln Asp Trp Ile Leu Leu Asp Met Leu Asn
            260                 265                 270

Glu Leu Ile Arg Cys Pro Lys Ser Leu Tyr Glu Arg Leu Arg Glu Glu
        275                 280                 285

Asp Arg Lys Lys Phe Arg Val Pro Ile Glu Ile Ala Asp Glu Asp Tyr
    290                 295                 300

Asp Ala Glu Gln Glu Pro Phe Lys Asn Ala Leu Val Arg His Gln Asp
305                 310                 315                 320

Arg Phe Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Tyr Asn Glu Ile Phe
                325                 330                 335

Thr Asn Leu Arg Phe Gln Ile Asp Leu Gly Thr Tyr His Phe Ser Ile
            340                 345                 350

Tyr Lys Lys Gln Ile Gly Asp Tyr Lys Glu Ser His His Leu Thr His
        355                 360                 365

Lys Leu Tyr Gly Phe Glu Arg Ile Gln Glu Phe Thr Lys Gln Asn Arg
    370                 375                 380

Pro Asp Glu Trp Arg Lys Phe Val Lys Thr Phe Asn Ser Phe Glu Thr
385                 390                 395                 400

Ser Lys Glu Pro Tyr Ile Pro Glu Thr Thr Pro His Tyr His Leu Glu
                405                 410                 415

Asn Gln Lys Ile Gly Ile Arg Phe Arg Asn Asp Asn Asp Lys Ile Trp
            420                 425                 430

Pro Ser Leu Lys Thr Asn Ser Glu Lys Asn Glu Lys Ser Lys Tyr Lys
        435                 440                 445

Leu Asp Lys Ser Phe Gln Ala Glu Ala Phe Leu Ser Val His Glu Leu
    450                 455                 460
```

```
Leu Pro Met Met Phe Tyr Tyr Leu Leu Leu Lys Thr Glu Asn Thr Asp
465                 470                 475                 480


Asn Asp Asn Glu Ile Glu Thr Lys Lys Lys Glu Asn Lys Asn Asp Lys
                485                 490                 495


Gln Glu Lys His Lys Ile Glu Glu Ile Ile Glu Asn Lys Ile Thr Glu
            500                 505                 510


Ile Tyr Ala Leu Tyr Asp Ala Phe Ala Asn Gly Lys Ile Asn Ser Ile
        515                 520                 525


Asp Lys Leu Glu Glu Tyr Cys Lys Gly Lys Asp Ile Glu Ile Gly His
    530                 535                 540


Leu Pro Lys Gln Met Ile Ala Ile Leu Lys Ser Glu His Lys Asp Met
545                 550                 555                 560


Ala Thr Glu Ala Lys Arg Lys Gln Glu Glu Met Leu Ala Asp Val Gln
                565                 570                 575


Lys Ser Leu Glu Ser Leu Asp Asn Gln Ile Asn Glu Glu Ile Glu Asn
            580                 585                 590


Val Glu Arg Lys Asn Ser Ser Leu Lys Ser Gly Glu Ile Ala Ser Trp
        595                 600                 605


Leu Val Asn Asp Met Met Arg Phe Gln Pro Val Gln Lys Asp Asn Glu
    610                 615                 620


Gly Asn Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Gln Met
625                 630                 635                 640


Leu Gln Arg Ser Leu Ala Leu Tyr Asn Lys Glu Glu Lys Pro Thr Arg
                645                 650                 655


Tyr Phe Arg Gln Val Asn Leu Ile Glu Ser Ser Asn Pro His Pro Phe
            660                 665                 670
```

```
Leu Asn Asn Thr Glu Trp Glu Lys Cys Asn Asn Ile Leu Ser Phe Tyr
        675                 680                 685

Arg Ser Tyr Leu Glu Ala Lys Lys Asn Phe Leu Glu Ser Leu Lys Pro
    690                 695                 700

Glu Asp Trp Glu Lys Asn Gln Tyr Phe Leu Met Leu Lys Glu Pro Lys
705                 710                 715                 720

Thr Asn Cys Glu Thr Leu Val Gln Gly Trp Lys Asn Gly Phe Asn Leu
                725                 730                 735

Pro Arg Gly Ile Phe Thr Glu Pro Ile Arg Lys Trp Phe Met Glu His
            740                 745                 750

Arg Lys Asn Ile Thr Val Ala Glu Leu Lys Arg Val Gly Leu Val Ala
        755                 760                 765

Lys Val Ile Pro Leu Phe Phe Ser Glu Glu Tyr Lys Asp Ser Val Gln
    770                 775                 780

Pro Phe Tyr Asn Tyr Leu Phe Asn Val Gly Asn Ile Asn Lys Pro Asp
785                 790                 795                 800

Glu Lys Asn Phe Leu Asn Cys Glu Glu Arg Arg Glu Leu Leu Arg Lys
                805                 810                 815

Lys Lys Asp Glu Phe Lys Lys Met Thr Asp Lys Glu Lys Glu Glu Asn
            820                 825                 830

Pro Ser Tyr Leu Glu Phe Gln Ser Trp Asn Lys Phe Glu Arg Glu Leu
        835                 840                 845

Arg Leu Val Arg Asn Gln Asp Ile Val Thr Trp Leu Leu Cys Met Glu
    850                 855                 860

Leu Phe Asn Lys Lys Lys Ile Lys Glu Leu Asn Val Glu Lys Ile Tyr
865                 870                 875                 880
```

```
Leu Lys Asn Ile Asn Thr Asn Thr Thr Lys Lys Glu Lys Asn Thr Glu
                885                 890                 895


Glu Lys Asn Gly Glu Glu Lys Ile Ile Lys Glu Lys Asn Asn Ile Leu
            900                 905                 910


Asn Arg Ile Met Pro Met Arg Leu Pro Ile Lys Val Tyr Gly Arg Glu
        915                 920                 925


Asn Phe Ser Lys Asn Lys Lys Lys Lys Ile Arg Arg Asn Thr Phe Phe
    930                 935                 940


Thr Val Tyr Ile Glu Glu Lys Gly Thr Lys Leu Leu Lys Gln Gly Asn
945                 950                 955                 960


Phe Lys Ala Leu Glu Arg Asp Arg Arg Leu Gly Gly Leu Phe Ser Phe
                965                 970                 975


Val Lys Thr His Ser Lys Ala Glu Ser Lys Ser Asn Thr Ile Ser Lys
            980                 985                 990


Ser Arg Val Glu Tyr Glu Leu Gly  Glu Tyr Gln Lys Ala  Arg Ile Glu
        995                 1000                1005


Ile Ile  Lys Asp Met Leu Ala  Leu Glu Glu Thr Leu  Ile Asp Lys
    1010                1015                1020


Tyr Asn  Ser Leu Asp Thr Asp  Asn Phe His Asn Met  Leu Thr Gly
    1025                1030                1035


Trp Leu  Lys Leu Lys Asp Glu  Pro Asp Lys Ala Ser  Phe Gln Asn
    1040                1045                1050


Asp Val  Asp Leu Leu Ile Ala  Val Arg Asn Ala Phe  Ser His Asn
    1055                1060                1065


Gln Tyr  Pro Met Arg Asn Arg  Ile Ala Phe Ala Asn  Ile Asn Pro
    1070                1075                1080
```

```
Phe Ser  Leu Ser Ser Ala Asn  Thr Ser Glu Glu Lys  Gly Leu Gly
    1085                1090                1095


Ile Ala  Asn Gln Leu Lys Asp  Lys Thr His Lys Thr  Ile Glu Lys
    1100                1105                1110


Ile Ile  Glu Ile Glu Lys Pro  Ile Glu Thr Lys Glu
    1115                1120                1125


<210>  93
<211>  1157
<212>  PRT
<213>  Myroides odoratimum

<400>  93

Met Lys Asp Ile Leu Thr Thr Asp Thr Thr Glu Lys Gln Asn Arg Phe
1               5                   10                  15


Tyr Ser His Lys Ile Ala Asp Lys Tyr Phe Phe Gly Gly Tyr Phe Asn
            20                  25                  30


Leu Ala Ser Asn Asn Ile Tyr Glu Val Phe Glu Glu Val Asn Lys Arg
        35                  40                  45


Asn Thr Phe Gly Lys Leu Ala Lys Arg Asp Asn Gly Asn Leu Lys Asn
    50                  55                  60


Tyr Ile Ile His Val Phe Lys Asp Glu Leu Ser Ile Ser Asp Phe Glu
65                  70                  75                  80


Lys Arg Val Ala Ile Phe Ala Ser Tyr Phe Pro Ile Leu Glu Thr Val
                85                  90                  95


Asp Lys Lys Ser Ile Lys Glu Arg Asn Arg Thr Ile Asp Leu Thr Leu
            100                 105                 110


Ser Gln Arg Ile Arg Gln Phe Arg Glu Met Leu Ile Ser Leu Val Thr
        115                 120                 125
```

```
Ala Val Asp Gln Leu Arg Asn Phe Tyr Thr His Tyr His His Ser Glu
    130                 135                 140

Ile Val Ile Glu Asn Lys Val Leu Asp Phe Leu Asn Ser Ser Leu Val
145                 150                 155                 160

Ser Thr Ala Leu His Val Lys Asp Lys Tyr Leu Lys Thr Asp Lys Thr
                165                 170                 175

Lys Glu Phe Leu Lys Glu Thr Ile Ala Ala Glu Leu Asp Ile Leu Ile
            180                 185                 190

Glu Ala Tyr Lys Lys Lys Gln Ile Glu Lys Lys Asn Thr Arg Phe Lys
        195                 200                 205

Ala Asn Lys Arg Glu Asp Ile Leu Asn Ala Ile Tyr Asn Glu Ala Phe
    210                 215                 220

Trp Ser Phe Ile Asn Asp Lys Asp Lys Asp Lys Glu Thr Val Val Ala
225                 230                 235                 240

Lys Gly Ala Asp Ala Tyr Phe Glu Lys Asn His His Lys Ser Asn Asp
                245                 250                 255

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

Ser Phe Phe Leu Thr Asn Lys Glu Met Asp Ser Leu Lys Ala Asn Leu
        275                 280                 285

Thr Gly Phe Lys Gly Lys Val Asp Arg Glu Ser Gly Asn Ser Ile Lys
    290                 295                 300

Tyr Met Ala Thr Gln Arg Ile Tyr Ser Phe His Thr Tyr Arg Gly Leu
305                 310                 315                 320

Lys Gln Lys Ile Arg Thr Ser Glu Glu Gly Val Lys Glu Thr Leu Leu
                325                 330                 335
```

```
Met Gln Met Ile Asp Glu Leu Ser Lys Val Pro Asn Val Val Tyr Gln
            340                 345                 350

His Leu Ser Thr Thr Gln Gln Asn Ser Phe Ile Glu Asp Trp Asn Glu
        355                 360                 365

Tyr Tyr Lys Asp Tyr Glu Asp Asp Val Glu Thr Asp Asp Leu Ser Arg
    370                 375                 380

Val Ile His Pro Val Ile Arg Lys Arg Tyr Glu Asp Arg Phe Asn Tyr
385                 390                 395                 400

Phe Ala Ile Arg Phe Leu Asp Glu Phe Phe Asp Phe Pro Thr Leu Arg
                405                 410                 415

Phe Gln Val His Leu Gly Asp Tyr Val His Asp Arg Arg Thr Lys Gln
            420                 425                 430

Leu Gly Lys Val Glu Ser Asp Arg Ile Ile Lys Glu Lys Val Thr Val
        435                 440                 445

Phe Ala Arg Leu Lys Asp Ile Asn Ser Ala Lys Ala Asn Tyr Phe His
    450                 455                 460

Ser Leu Glu Glu Gln Asp Lys Glu Glu Leu Asp Asn Lys Trp Thr Leu
465                 470                 475                 480

Phe Pro Asn Pro Ser Tyr Asp Phe Pro Lys Glu His Thr Leu Gln His
                485                 490                 495

Gln Gly Glu Gln Lys Asn Ala Gly Lys Ile Gly Ile Tyr Val Lys Leu
            500                 505                 510

Arg Asp Thr Gln Tyr Lys Glu Lys Ala Ala Leu Glu Glu Ala Arg Lys
        515                 520                 525

Ser Leu Asn Pro Lys Glu Arg Ser Ala Thr Lys Ala Ser Lys Tyr Asp
    530                 535                 540
```

```
Ile Ile Thr Gln Ile Ile Glu Ala Asn Asp Asn Val Lys Ser Glu Lys
545                 550                 555                 560


Pro Leu Val Phe Thr Gly Gln Pro Ile Ala Tyr Leu Ser Met Asn Asp
                565                 570                 575


Ile His Ser Met Leu Phe Ser Leu Leu Thr Asp Asn Ala Glu Leu Lys
            580                 585                 590


Lys Thr Pro Glu Glu Val Glu Ala Lys Leu Ile Asp Gln Ile Gly Lys
        595                 600                 605


Gln Ile Asn Glu Ile Leu Ser Lys Asp Thr Asp Thr Lys Ile Leu Lys
    610                 615                 620


Lys Tyr Lys Asp Asn Asp Leu Lys Glu Thr Asp Thr Asp Lys Ile Thr
625                 630                 635                 640


Arg Asp Leu Ala Arg Asp Lys Glu Glu Ile Glu Lys Leu Ile Leu Glu
                645                 650                 655


Gln Lys Gln Arg Ala Asp Asp Tyr Asn Tyr Thr Ser Ser Thr Lys Phe
            660                 665                 670


Asn Ile Asp Lys Ser Arg Lys Arg Lys His Leu Leu Phe Asn Ala Glu
        675                 680                 685


Lys Gly Lys Ile Gly Val Trp Leu Ala Asn Asp Ile Lys Arg Phe Met
    690                 695                 700


Thr Glu Glu Phe Lys Ser Lys Trp Lys Gly Tyr Gln His Thr Glu Leu
705                 710                 715                 720


Gln Lys Leu Phe Ala Tyr Tyr Asp Thr Ser Lys Ser Asp Leu Asp Leu
                725                 730                 735


Ile Leu Ser Asp Met Val Met Val Lys Asp Tyr Pro Ile Glu Leu Ile
            740                 745                 750
```

```
Ala Leu Val Lys Lys Ser Arg Thr Leu Val Asp Phe Leu Asn Lys Tyr
        755                 760                 765


Leu Glu Ala Arg Leu Gly Tyr Met Glu Asn Val Ile Thr Arg Val Lys
    770                 775                 780


Asn Ser Ile Gly Thr Pro Gln Phe Lys Thr Val Arg Lys Glu Cys Phe
785                 790                 795                 800


Thr Phe Leu Lys Lys Ser Asn Tyr Thr Val Val Ser Leu Asp Lys Gln
                805                 810                 815


Val Glu Arg Ile Leu Ser Met Pro Leu Phe Ile Glu Arg Gly Phe Met
            820                 825                 830


Asp Asp Lys Pro Thr Met Leu Glu Gly Lys Ser Tyr Gln Gln His Lys
        835                 840                 845


Glu Lys Phe Ala Asp Trp Phe Val His Tyr Lys Glu Asn Ser Asn Tyr
    850                 855                 860


Gln Asn Phe Tyr Asp Thr Glu Val Tyr Glu Ile Thr Thr Glu Asp Lys
865                 870                 875                 880


Arg Glu Lys Ala Lys Val Thr Lys Lys Ile Lys Gln Gln Gln Lys Asn
                885                 890                 895


Asp Val Phe Thr Leu Met Met Val Asn Tyr Met Leu Glu Glu Val Leu
            900                 905                 910


Lys Leu Ser Ser Asn Asp Arg Leu Ser Leu Asn Glu Leu Tyr Gln Thr
        915                 920                 925


Lys Glu Glu Arg Ile Val Asn Lys Gln Val Ala Lys Asp Thr Gln Glu
    930                 935                 940


Arg Asn Lys Asn Tyr Ile Trp Asn Lys Val Val Asp Leu Gln Leu Cys
945                 950                 955                 960
```

```
Glu Gly Leu Val Arg Ile Asp Lys Val Lys Leu Lys Asp Ile Gly Asn
               965                 970                 975


Phe Arg Lys Tyr Glu Asn Asp Ser Arg Val Lys Glu Phe Leu Thr Tyr
            980                 985                 990


Gln Ser Asp Ile Val Trp Ser Ala  Tyr Leu Ser Asn Glu  Val Asp Ser
        995                 1000                1005


Asn Lys  Leu Tyr Val Ile Glu  Arg Gln Leu Asp Asn  Tyr Glu Ser
    1010                1015                1020


Ile Arg  Ser Lys Glu Leu Leu  Lys Glu Val Gln Glu  Ile Glu Cys
    1025                1030                1035


Ser Val  Tyr Asn Gln Val Ala  Asn Lys Glu Ser Leu  Lys Gln Ser
    1040                1045                1050


Gly Asn  Glu Asn Phe Lys Gln  Tyr Val Leu Gln Gly  Leu Val Pro
    1055                1060                1065


Ile Gly  Met Asp Val Arg Glu  Met Leu Ile Leu Ser  Thr Asp Val
    1070                1075                1080


Lys Phe  Ile Lys Glu Glu Ile  Ile Gln Leu Gly Gln  Ala Gly Glu
    1085                1090                1095


Val Glu  Gln Asp Leu Tyr Ser  Leu Ile Tyr Ile Arg  Asn Lys Phe
    1100                1105                1110


Ala His  Asn Gln Leu Pro Ile  Lys Glu Phe Phe Asp  Phe Cys Glu
    1115                1120                1125


Asn Asn  Tyr Arg Ser Ile Ser  Asp Asn Glu Tyr Tyr  Ala Glu Tyr
    1130                1135                1140


Tyr Met  Glu Ile Phe Arg Ser  Ile Lys Glu Lys Tyr  Thr Ser
    1145                1150                1155
```

<210> 94
<211> 1157
<212> PRT
<213> Myroides odoratimum

<400> 94

Met Lys Asp Ile Leu Thr Thr Asp Thr Thr Glu Lys Gln Asn Arg Phe
1 5 10 15

Tyr Ser His Lys Ile Ala Asp Lys Tyr Phe Phe Gly Gly Tyr Phe Asn
20 25 30

Leu Ala Ser Asn Asn Ile Tyr Glu Val Phe Glu Glu Val Asn Lys Arg
35 40 45

Asn Thr Phe Gly Lys Leu Ala Lys Arg Asp Asn Gly Asn Leu Lys Asn
50 55 60

Tyr Ile Ile His Val Phe Lys Asp Glu Leu Ser Ile Ser Asp Phe Glu
65 70 75 80

Lys Arg Val Ala Ile Phe Ala Ser Tyr Phe Pro Ile Leu Glu Thr Val
85 90 95

Asp Lys Lys Ser Ile Lys Glu Arg Asn Arg Thr Ile Asp Leu Thr Leu
100 105 110

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

Ala Val Asp Gln Leu Arg Asn Phe Tyr Thr His Tyr His His Ser Glu
130 135 140

Ile Val Ile Glu Asn Lys Val Leu Asp Phe Leu Asn Ser Ser Leu Val
145 150 155 160

Ser Thr Ala Leu His Val Lys Asp Lys Tyr Leu Lys Thr Asp Lys Thr
165 170 175

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

180 185 190

Glu Ala Tyr Lys Lys Lys Gln Ile Glu Lys Lys Asn Thr Arg Phe Lys
195 200 205

Ala Asn Lys Arg Glu Asp Ile Leu Asn Ala Ile Tyr Asn Glu Ala Phe
210 215 220

Trp Ser Phe Ile Asn Asp Lys Asp Lys Asp Lys Glu Thr Val Val Ala
225 230 235 240

Lys Gly Ala Asp Ala Tyr Phe Glu Lys Asn His His Lys Ser Asn Asp
245 250 255

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

Ser Phe Phe Leu Thr Asn Lys Glu Met Asp Ser Leu Lys Ala Asn Leu
275 280 285

Thr Gly Phe Lys Gly Lys Val Asp Arg Glu Ser Gly Asn Ser Ile Lys
290 295 300

Tyr Met Ala Thr Gln Arg Ile Tyr Ser Phe His Thr Tyr Arg Gly Leu
305 310 315 320

Lys Gln Lys Ile Arg Thr Ser Glu Glu Gly Val Lys Glu Thr Leu Leu
325 330 335

Met Gln Met Ile Asp Glu Leu Ser Lys Val Pro Asn Val Val Tyr Gln
340 345 350

His Leu Ser Thr Thr Gln Gln Asn Ser Phe Ile Glu Asp Trp Asn Glu
355 360 365

Tyr Tyr Lys Asp Tyr Glu Asp Asp Val Glu Thr Asp Asp Leu Ser Arg
370 375 380

Val Ile His Pro Val Ile Arg Lys Arg Tyr Glu Asp Arg Phe Asn Tyr

385 390 395 400

Phe Ala Ile Arg Phe Leu Asp Glu Phe Phe Asp Phe Pro Thr Leu Arg
405 410 415

Phe Gln Val His Leu Gly Asp Tyr Val His Asp Arg Arg Thr Lys Gln
420 425 430

Leu Gly Lys Val Glu Ser Asp Arg Ile Ile Lys Glu Lys Val Thr Val
435 440 445

Phe Ala Arg Leu Lys Asp Ile Asn Ser Ala Lys Ala Ser Tyr Phe His
450 455 460

Ser Leu Glu Glu Gln Asp Lys Glu Glu Leu Asp Asn Lys Trp Thr Leu
465 470 475 480

Phe Pro Asn Pro Ser Tyr Asp Phe Pro Lys Glu His Thr Leu Gln His
485 490 495

Gln Gly Glu Gln Lys Asn Ala Gly Lys Ile Gly Ile Tyr Val Lys Leu
500 505 510

Arg Asp Thr Gln Tyr Lys Glu Lys Ala Ala Leu Glu Glu Ala Arg Lys
515 520 525

Ser Leu Asn Pro Lys Glu Arg Ser Ala Thr Lys Ala Ser Lys Tyr Asp
530 535 540

Ile Ile Thr Gln Ile Ile Glu Ala Asn Asp Asn Val Lys Ser Glu Lys
545 550 555 560

Pro Leu Val Phe Thr Gly Gln Pro Ile Ala Tyr Leu Ser Met Asn Asp
565 570 575

Ile His Ser Met Leu Phe Ser Leu Leu Thr Asp Asn Ala Glu Leu Lys
580 585 590

Lys Thr Pro Glu Glu Val Glu Ala Lys Leu Ile Asp Gln Ile Gly Lys

595 600 605

Gln Ile Asn Glu Ile Leu Ser Lys Asp Thr Asp Thr Lys Ile Leu Lys
610 615 620

Lys Tyr Lys Asp Asn Asp Leu Lys Glu Thr Asp Thr Asp Lys Ile Thr
625 630 635 640

Arg Asp Leu Ala Arg Asp Lys Glu Glu Ile Glu Lys Leu Ile Leu Glu
645 650 655

Gln Lys Gln Arg Ala Asp Asp Tyr Asn Tyr Thr Ser Ser Thr Lys Phe
660 665 670

Asn Ile Asp Lys Ser Arg Lys Arg Lys His Leu Leu Phe Asn Ala Glu
675 680 685

Lys Gly Lys Ile Gly Val Trp Leu Ala Asn Asp Ile Lys Arg Phe Met
690 695 700

Phe Lys Glu Ser Lys Ser Lys Trp Lys Gly Tyr Gln His Thr Glu Leu
705 710 715 720

Gln Lys Leu Phe Ala Tyr Phe Asp Thr Ser Lys Ser Asp Leu Glu Leu
725 730 735

Ile Leu Ser Asp Met Val Met Val Lys Asp Tyr Pro Ile Glu Leu Ile
740 745 750

Asp Leu Val Arg Lys Ser Arg Thr Leu Val Asp Phe Leu Asn Lys Tyr
755 760 765

Leu Glu Ala Arg Leu Gly Tyr Ile Glu Asn Val Ile Thr Arg Val Lys
770 775 780

Asn Ser Ile Gly Thr Pro Gln Phe Lys Thr Val Arg Lys Glu Cys Phe
785 790 795 800

Ala Phe Leu Lys Glu Ser Asn Tyr Thr Val Ala Ser Leu Asp Lys Gln

805 810 815

Ile Glu Arg Ile Leu Ser Met Pro Leu Phe Ile Glu Arg Gly Phe Met
820 825 830

Asp Ser Lys Pro Thr Met Leu Glu Gly Lys Ser Tyr Gln Gln His Lys
835 840 845

Glu Asp Phe Ala Asp Trp Phe Val His Tyr Lys Glu Asn Ser Asn Tyr
850 855 860

Gln Asn Phe Tyr Asp Thr Glu Val Tyr Glu Ile Ile Thr Glu Asp Lys
865 870 875 880

Arg Glu Gln Ala Lys Val Thr Lys Lys Ile Lys Gln Gln Gln Lys Asn
885 890 895

Asp Val Phe Thr Leu Met Met Val Asn Tyr Met Leu Glu Glu Val Leu
900 905 910

Lys Leu Pro Ser Asn Asp Arg Leu Ser Leu Asn Glu Leu Tyr Gln Thr
915 920 925

Lys Glu Glu Arg Ile Val Asn Lys Gln Val Ala Lys Asp Thr Gln Glu
930 935 940

Arg Asn Lys Asn Tyr Ile Trp Asn Lys Val Val Asp Leu Gln Leu Cys
945 950 955 960

Glu Gly Leu Val Arg Ile Asp Lys Val Lys Leu Lys Asp Ile Gly Asn
965 970 975

Phe Arg Lys Tyr Glu Asn Asp Ser Arg Val Lys Glu Phe Leu Thr Tyr
980 985 990

Gln Ser Asp Ile Val Trp Ser Gly Tyr Leu Ser Asn Glu Val Asp Ser
995 1000 1005

Asn Lys Leu Tyr Val Ile Glu Arg Gln Leu Asp Asn Tyr Glu Ser

```
    1010                1015                1020


Ile Arg  Ser Lys Glu Leu Leu  Lys Glu Val Gln Glu  Ile Glu Cys
    1025                1030                1035


Ile Val  Tyr Asn Gln Val Ala  Asn Lys Glu Ser Leu  Lys Gln Ser
    1040                1045                1050


Gly Asn  Glu Asn Phe Lys Gln  Tyr Val Leu Gln Gly  Leu Leu Pro
    1055                1060                1065


Arg Gly  Thr Asp Val Arg Glu  Met Leu Ile Leu Ser  Thr Asp Val
    1070                1075                1080


Lys Phe  Lys Lys Glu Glu Ile  Met Gln Leu Gly Gln  Val Arg Glu
    1085                1090                1095


Val Glu  Gln Asp Leu Tyr Ser  Leu Ile Tyr Ile Arg  Asn Lys Phe
    1100                1105                1110


Ala His  Asn Gln Leu Pro Ile  Lys Glu Phe Phe Asp  Phe Cys Glu
    1115                1120                1125


Asn Asn  Tyr Arg Pro Ile Ser  Asp Asn Glu Tyr Tyr  Ala Glu Tyr
    1130                1135                1140


Tyr Met  Glu Ile Phe Arg Ser  Ile Lys Glu Lys Tyr  Ala Ser
    1145                1150                1155


<210>  95
<211>  1224
<212>  PRT
<213>  Bergeyella zoohelcum

<400>  95

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


Lys Pro Gln Asp Lys Ser Tyr Phe Ala Gly Tyr Phe Asn Ala Ala Met
            20                  25                  30
```

```
Glu Asn Thr Asp Ser Val Phe Arg Glu Leu Gly Lys Arg Leu Lys Gly
        35                  40                  45

Lys Glu Tyr Thr Ser Glu Asn Phe Phe Asp Ala Ile Phe Lys Glu Asn
    50                  55                  60

Ile Ser Leu Val Glu Tyr Glu Arg Tyr Val Lys Leu Leu Ser Asp Tyr
65                  70                  75                  80

Phe Pro Met Ala Arg Leu Leu Asp Lys Lys Glu Val Pro Ile Lys Glu
                85                  90                  95

Arg Lys Glu Asn Phe Lys Lys Asn Phe Lys Gly Ile Ile Lys Ala Val
            100                 105                 110

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

Ile Thr Asp Glu Ile Phe Gly Val Leu Asp Glu Met Leu Lys Ser Thr
    130                 135                 140

Val Leu Thr Val Lys Lys Lys Lys Val Lys Thr Asp Lys Thr Lys Glu
145                 150                 155                 160

Ile Leu Lys Lys Ser Ile Glu Lys Gln Leu Asp Ile Leu Cys Gln Lys
                165                 170                 175

Lys Leu Glu Tyr Leu Arg Asp Thr Ala Arg Lys Ile Glu Glu Lys Arg
            180                 185                 190

Arg Asn Gln Arg Glu Arg Gly Glu Lys Glu Leu Val Ala Pro Phe Lys
        195                 200                 205

Tyr Ser Asp Lys Arg Asp Asp Leu Ile Ala Ala Ile Tyr Asn Asp Ala
    210                 215                 220

Phe Asp Val Tyr Ile Asp Lys Lys Lys Asp Ser Leu Lys Glu Ser Ser
225                 230                 235                 240
```

```
Lys Ala Lys Tyr Asn Thr Lys Ser Asp Pro Gln Gln Glu Glu Gly Asp
                245                 250                 255

Leu Lys Ile Pro Ile Ser Lys Asn Gly Val Val Phe Leu Leu Ser Leu
            260                 265                 270

Phe Leu Thr Lys Gln Glu Ile His Ala Phe Lys Ser Lys Ile Ala Gly
        275                 280                 285

Phe Lys Ala Thr Val Ile Asp Glu Ala Thr Val Ser Glu Ala Thr Val
    290                 295                 300

Ser His Gly Lys Asn Ser Ile Cys Phe Met Ala Thr His Glu Ile Phe
305                 310                 315                 320

Ser His Leu Ala Tyr Lys Lys Leu Lys Arg Lys Val Arg Thr Ala Glu
                325                 330                 335

Ile Asn Tyr Gly Glu Ala Glu Asn Ala Glu Gln Leu Ser Val Tyr Ala
            340                 345                 350

Lys Glu Thr Leu Met Met Gln Met Leu Asp Glu Leu Ser Lys Val Pro
        355                 360                 365

Asp Val Val Tyr Gln Asn Leu Ser Glu Asp Val Gln Lys Thr Phe Ile
    370                 375                 380

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

Met Glu Glu Glu Gln Val Ile His Pro Val Ile Arg Lys Arg Tyr Glu
                405                 410                 415

Asp Lys Phe Asn Tyr Phe Ala Ile Arg Phe Leu Asp Glu Phe Ala Gln
            420                 425                 430

Phe Pro Thr Leu Arg Phe Gln Val His Leu Gly Asn Tyr Leu His Asp
        435                 440                 445
```

```
Ser Arg Pro Lys Glu Asn Leu Ile Ser Asp Arg Arg Ile Lys Glu Lys
    450                 455                 460

Ile Thr Val Phe Gly Arg Leu Ser Glu Leu Glu His Lys Lys Ala Leu
465                 470                 475                 480

Phe Ile Lys Asn Thr Glu Thr Asn Glu Asp Arg Glu His Tyr Trp Glu
                485                 490                 495

Ile Phe Pro Asn Pro Asn Tyr Asp Phe Pro Lys Glu Asn Ile Ser Val
            500                 505                 510

Asn Asp Lys Asp Phe Pro Ile Ala Gly Ser Ile Leu Asp Arg Glu Lys
        515                 520                 525

Gln Pro Val Ala Gly Lys Ile Gly Ile Lys Val Lys Leu Leu Asn Gln
    530                 535                 540

Gln Tyr Val Ser Glu Val Asp Lys Ala Val Lys Ala His Gln Leu Lys
545                 550                 555                 560

Gln Arg Lys Ala Ser Lys Pro Ser Ile Gln Asn Ile Ile Glu Glu Ile
                565                 570                 575

Val Pro Ile Asn Glu Ser Asn Pro Lys Glu Ala Ile Val Phe Gly Gly
            580                 585                 590

Gln Pro Thr Ala Tyr Leu Ser Met Asn Asp Ile His Ser Ile Leu Tyr
        595                 600                 605

Glu Phe Phe Asp Lys Trp Glu Lys Lys Lys Glu Lys Leu Glu Lys Lys
    610                 615                 620

Gly Glu Lys Glu Leu Arg Lys Glu Ile Gly Lys Glu Leu Glu Lys Lys
625                 630                 635                 640

Ile Val Gly Lys Ile Gln Ala Gln Ile Gln Gln Ile Ile Asp Lys Asp
                645                 650                 655
```

```
Thr Asn Ala Lys Ile Leu Lys Pro Tyr Gln Asp Gly Asn Ser Thr Ala
            660                 665                 670

Ile Asp Lys Glu Lys Leu Ile Lys Asp Leu Lys Gln Glu Gln Asn Ile
        675                 680                 685

Leu Gln Lys Leu Lys Asp Glu Gln Thr Val Arg Glu Lys Glu Tyr Asn
    690                 695                 700

Asp Phe Ile Ala Tyr Gln Asp Lys Asn Arg Glu Ile Asn Lys Val Arg
705                 710                 715                 720

Asp Arg Asn His Lys Gln Tyr Leu Lys Asp Asn Leu Lys Arg Lys Tyr
                725                 730                 735

Pro Glu Ala Pro Ala Arg Lys Glu Val Leu Tyr Tyr Arg Glu Lys Gly
            740                 745                 750

Lys Val Ala Val Trp Leu Ala Asn Asp Ile Lys Arg Phe Met Pro Thr
        755                 760                 765

Asp Phe Lys Asn Glu Trp Lys Gly Glu Gln His Ser Leu Leu Gln Lys
    770                 775                 780

Ser Leu Ala Tyr Tyr Glu Gln Cys Lys Glu Glu Leu Lys Asn Leu Leu
785                 790                 795                 800

Pro Glu Lys Val Phe Gln His Leu Pro Phe Lys Leu Gly Gly Tyr Phe
                805                 810                 815

Gln Gln Lys Tyr Leu Tyr Gln Phe Tyr Thr Cys Tyr Leu Asp Lys Arg
            820                 825                 830

Leu Glu Tyr Ile Ser Gly Leu Val Gln Gln Ala Glu Asn Phe Lys Ser
        835                 840                 845

Glu Asn Lys Val Phe Lys Lys Val Glu Asn Glu Cys Phe Lys Phe Leu
    850                 855                 860
```

```
Lys Lys Gln Asn Tyr Thr His Lys Glu Leu Asp Ala Arg Val Gln Ser
865                 870                 875                 880

Ile Leu Gly Tyr Pro Ile Phe Leu Glu Arg Gly Phe Met Asp Glu Lys
                885                 890                 895

Pro Thr Ile Ile Lys Gly Lys Thr Phe Lys Gly Asn Glu Ala Leu Phe
            900                 905                 910

Ala Asp Trp Phe Arg Tyr Tyr Lys Glu Tyr Gln Asn Phe Gln Thr Phe
        915                 920                 925

Tyr Asp Thr Glu Asn Tyr Pro Leu Val Glu Leu Glu Lys Lys Gln Ala
    930                 935                 940

Asp Arg Lys Arg Lys Thr Lys Ile Tyr Gln Gln Lys Lys Asn Asp Val
945                 950                 955                 960

Phe Thr Leu Leu Met Ala Lys His Ile Phe Lys Ser Val Phe Lys Gln
                965                 970                 975

Asp Ser Ile Asp Gln Phe Ser Leu Glu Asp Leu Tyr Gln Ser Arg Glu
            980                 985                 990

Glu Arg Leu Gly Asn Gln Glu Arg  Ala Arg Gln Thr Gly  Glu Arg Asn
        995                 1000                1005

Thr Asn  Tyr Ile Trp Asn Lys  Thr Val Asp Leu Lys  Leu Cys Asp
    1010                1015                1020

Gly Lys  Ile Thr Val Glu Asn  Val Lys Leu Lys Asn  Val Gly Asp
    1025                1030                1035

Phe Ile  Lys Tyr Glu Tyr Asp  Gln Arg Val Gln Ala  Phe Leu Lys
    1040                1045                1050

Tyr Glu  Glu Asn Ile Glu Trp  Gln Ala Phe Leu Ile  Lys Glu Ser
    1055                1060                1065
```

```
Lys Glu  Glu Glu Asn Tyr Pro  Tyr Val Val Glu Arg  Glu Ile Glu
    1070                1075                1080

Gln Tyr  Glu Lys Val Arg Arg  Glu Glu Leu Leu Lys  Glu Val His
    1085                1090                1095

Leu Ile  Glu Glu Tyr Ile Leu  Glu Lys Val Lys Asp  Lys Glu Ile
    1100                1105                1110

Leu Lys  Lys Gly Asp Asn Gln  Asn Phe Lys Tyr Tyr  Ile Leu Asn
    1115                1120                1125

Gly Leu  Leu Lys Gln Leu Lys  Asn Glu Asp Val Glu  Ser Tyr Lys
    1130                1135                1140

Val Phe  Asn Leu Asn Thr Glu  Pro Glu Asp Val Asn  Ile Asn Gln
    1145                1150                1155

Leu Lys  Gln Glu Ala Thr Asp  Leu Glu Gln Lys Ala  Phe Val Leu
    1160                1165                1170

Thr Tyr  Ile Arg Asn Lys Phe  Ala His Asn Gln Leu  Pro Lys Lys
    1175                1180                1185

Glu Phe  Trp Asp Tyr Cys Gln  Glu Lys Tyr Gly Lys  Ile Glu Lys
    1190                1195                1200

Glu Lys  Thr Tyr Ala Glu Tyr  Phe Ala Glu Val Phe  Lys Lys Glu
    1205                1210                1215

Lys Glu  Ala Leu Ile Lys
    1220
```

<210> 96
<211> 1150
<212> PRT
<213> Prevotella saccharolytica

<400> 96

```
Met Met Glu Lys Glu Asn Val Gln Gly Ser His Ile Tyr Tyr Glu Pro
1               5                  10                  15

Thr Asp Lys Cys Phe Trp Ala Ala Phe Tyr Asn Leu Ala Arg His Asn
            20                  25                  30

Ala Tyr Leu Thr Ile Ala His Ile Asn Ser Phe Val Asn Ser Lys Lys
        35                  40                  45

Gly Ile Asn Asn Asp Asp Lys Val Leu Asp Ile Ile Asp Asp Trp Ser
    50                  55                  60

Lys Phe Asp Asn Asp Leu Leu Met Gly Ala Arg Leu Asn Lys Leu Ile
65                  70                  75                  80

Leu Lys His Phe Pro Phe Leu Lys Ala Pro Leu Tyr Gln Leu Ala Lys
                85                  90                  95

Arg Lys Thr Arg Lys Gln Gln Gly Lys Glu Gln Gln Asp Tyr Glu Lys
            100                 105                 110

Lys Gly Asp Glu Asp Pro Glu Val Ile Gln Glu Ala Ile Ala Asn Ala
        115                 120                 125

Phe Lys Met Ala Asn Val Arg Lys Thr Leu His Ala Phe Leu Lys Gln
    130                 135                 140

Leu Glu Asp Leu Arg Asn His Phe Ser His Tyr Asn Tyr Asn Ser Pro
145                 150                 155                 160

Ala Lys Lys Met Glu Val Lys Phe Asp Asp Gly Phe Cys Asn Lys Leu
                165                 170                 175

Tyr Tyr Val Phe Asp Ala Ala Leu Gln Met Val Lys Asp Asp Asn Arg
            180                 185                 190

Met Asn Pro Glu Ile Asn Met Gln Thr Asp Phe Glu His Leu Val Arg
        195                 200                 205
```

```
Leu Gly Arg Asn Arg Lys Ile Pro Asn Thr Phe Lys Tyr Asn Phe Thr
    210                 215                 220

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

Leu Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Lys
                245                 250                 255

Gly Phe Lys Gly Gly Thr Glu Asn Tyr Met Arg Met Thr Asn Glu Val
            260                 265                 270

Phe Cys Arg Asn Arg Met Val Ile Pro Lys Leu Arg Leu Glu Thr Asp
        275                 280                 285

Tyr Asp Asn His Gln Leu Met Phe Asp Met Leu Asn Glu Leu Val Arg
    290                 295                 300

Cys Pro Leu Ser Leu Tyr Lys Arg Leu Lys Gln Glu Asp Gln Asp Lys
305                 310                 315                 320

Phe Arg Val Pro Ile Glu Phe Leu Asp Glu Asp Asn Glu Ala Asp Asn
                325                 330                 335

Pro Tyr Gln Glu Asn Ala Asn Ser Asp Glu Asn Pro Thr Glu Glu Thr
            340                 345                 350

Asp Pro Leu Lys Asn Thr Leu Val Arg His Gln His Arg Phe Pro Tyr
        355                 360                 365

Phe Val Leu Arg Tyr Phe Asp Leu Asn Glu Val Phe Lys Gln Leu Arg
    370                 375                 380

Phe Gln Ile Asn Leu Gly Cys Tyr His Phe Ser Ile Tyr Asp Lys Thr
385                 390                 395                 400

Ile Gly Glu Arg Thr Glu Lys Arg His Leu Thr Arg Thr Leu Phe Gly
                405                 410                 415
```

```
Phe Asp Arg Leu Gln Asn Phe Ser Val Lys Leu Gln Pro Glu His Trp
            420                 425                 430

Lys Asn Met Val Lys His Leu Asp Thr Glu Glu Ser Ser Asp Lys Pro
        435                 440                 445

Tyr Leu Ser Asp Ala Met Pro His Tyr Gln Ile Glu Asn Glu Lys Ile
    450                 455                 460

Gly Ile His Phe Leu Lys Thr Asp Thr Glu Lys Lys Glu Thr Val Trp
465                 470                 475                 480

Pro Ser Leu Glu Val Glu Glu Val Ser Ser Asn Arg Asn Lys Tyr Lys
                485                 490                 495

Ser Glu Lys Asn Leu Thr Ala Asp Ala Phe Leu Ser Thr His Glu Leu
            500                 505                 510

Leu Pro Met Met Phe Tyr Tyr Gln Leu Leu Ser Ser Glu Glu Lys Thr
        515                 520                 525

Arg Ala Ala Ala Gly Asp Lys Val Gln Gly Val Leu Gln Ser Tyr Arg
    530                 535                 540

Lys Lys Ile Phe Asp Ile Tyr Asp Asp Phe Ala Asn Gly Thr Ile Asn
545                 550                 555                 560

Ser Met Gln Lys Leu Asp Glu Arg Leu Ala Lys Asp Asn Leu Leu Arg
                565                 570                 575

Gly Asn Met Pro Gln Gln Met Leu Ala Ile Leu Glu His Gln Glu Pro
            580                 585                 590

Asp Met Glu Gln Lys Ala Lys Glu Lys Leu Asp Arg Leu Ile Thr Glu
        595                 600                 605

Thr Lys Lys Arg Ile Gly Lys Leu Glu Asp Gln Phe Lys Gln Lys Val
    610                 615                 620
```

```
Arg Ile Gly Lys Arg Arg Ala Asp Leu Pro Lys Val Gly Ser Ile Ala
625                 630                 635                 640

Asp Trp Leu Val Asn Asp Met Met Arg Phe Gln Pro Ala Lys Arg Asn
                645                 650                 655

Ala Asp Asn Thr Gly Val Pro Asp Ser Lys Ala Asn Ser Thr Glu Tyr
            660                 665                 670

Arg Leu Leu Gln Glu Ala Leu Ala Phe Tyr Ser Ala Tyr Lys Asp Arg
        675                 680                 685

Leu Glu Pro Tyr Phe Arg Gln Val Asn Leu Ile Gly Gly Thr Asn Pro
    690                 695                 700

His Pro Phe Leu His Arg Val Asp Trp Lys Lys Cys Asn His Leu Leu
705                 710                 715                 720

Ser Phe Tyr His Asp Tyr Leu Glu Ala Lys Glu Gln Tyr Leu Ser His
                725                 730                 735

Leu Ser Pro Ala Asp Trp Gln Lys His Gln His Phe Leu Leu Leu Lys
            740                 745                 750

Val Arg Lys Asp Ile Gln Asn Glu Lys Lys Asp Trp Lys Lys Ser Leu
        755                 760                 765

Val Ala Gly Trp Lys Asn Gly Phe Asn Leu Pro Arg Gly Leu Phe Thr
    770                 775                 780

Glu Ser Ile Lys Thr Trp Phe Ser Thr Asp Ala Asp Lys Val Gln Ile
785                 790                 795                 800

Thr Asp Thr Lys Leu Phe Glu Asn Arg Val Gly Leu Ile Ala Lys Leu
                805                 810                 815

Ile Pro Leu Tyr Tyr Asp Lys Val Tyr Asn Asp Lys Pro Gln Pro Phe
            820                 825                 830
```

```
Tyr Gln Tyr Pro Phe Asn Ile Asn Asp Arg Tyr Lys Pro Glu Asp Thr
        835                 840                 845

Arg Lys Arg Phe Thr Ala Ala Ser Ser Lys Leu Trp Asn Glu Lys Lys
    850                 855                 860

Met Leu Tyr Lys Asn Ala Gln Pro Asp Ser Ser Asp Lys Ile Glu Tyr
865                 870                 875                 880

Pro Gln Tyr Leu Asp Phe Leu Ser Trp Lys Lys Leu Glu Arg Glu Leu
                885                 890                 895

Arg Met Leu Arg Asn Gln Asp Met Met Val Trp Leu Met Cys Lys Asp
            900                 905                 910

Leu Phe Ala Gln Cys Thr Val Glu Gly Val Glu Phe Ala Asp Leu Lys
        915                 920                 925

Leu Ser Gln Leu Glu Val Asp Val Asn Val Gln Asp Asn Leu Asn Val
    930                 935                 940

Leu Asn Asn Val Ser Ser Met Ile Leu Pro Leu Ser Val Tyr Pro Ser
945                 950                 955                 960

Asp Ala Gln Gly Asn Val Leu Arg Asn Ser Lys Pro Leu His Thr Val
                965                 970                 975

Tyr Val Gln Glu Asn Asn Thr Lys Leu Leu Lys Gln Gly Asn Phe Lys
            980                 985                 990

Ser Leu Leu Lys Asp Arg Arg Leu  Asn Gly Leu Phe Ser  Phe Ile Ala
        995                 1000                1005

Ala Glu  Gly Glu Asp Leu Gln  Gln His Pro Leu Thr  Lys Asn Arg
    1010                1015                1020

Leu Glu  Tyr Glu Leu Ser Ile  Tyr Gln Thr Met Arg  Ile Ser Val
    1025                1030                1035
```

```
Phe Glu  Gln Thr Leu Gln Leu  Glu Lys Ala Ile Leu  Thr Arg Asn
    1040                1045                1050


Lys Thr  Leu Cys Gly Asn Asn  Phe Asn Asn Leu Leu  Asn Ser Trp
    1055                1060                1065


Ser Glu  His Arg Thr Asp Lys  Lys Thr Leu Gln Pro  Asp Ile Asp
    1070                1075                1080


Phe Leu  Ile Ala Val Arg Asn  Ala Phe Ser His Asn  Gln Tyr Pro
    1085                1090                1095


Met Ser  Thr Asn Thr Val Met  Gln Gly Ile Glu Lys  Phe Asn Ile
    1100                1105                1110


Gln Thr  Pro Lys Leu Glu Glu  Lys Asp Gly Leu Gly  Ile Ala Ser
    1115                1120                1125


Gln Leu  Ala Lys Lys Thr Lys  Asp Ala Ala Ser Arg  Leu Gln Asn
    1130                1135                1140


Ile Ile  Asn Gly Gly Thr Asn
    1145                1150


<210>  97
<211>  1175
<212>  PRT
<213>  Porphyromonas gingivalis

<400>  97

Met Thr Glu Gln Asn Glu Lys Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1               5                  10                  15


Glu Asp Lys His Phe Trp Ala Ala Phe Phe Asn Leu Ala Arg His Asn
            20                  25                  30


Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
        35                  40                  45
```

```
Ala Asp Ile Thr Asn Asp Glu Asp Ile Leu Phe Phe Lys Gly Gln Trp
    50                  55                  60

Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ala Arg Leu Arg Ser Leu
65                  70                  75                  80

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

Leu Phe Glu Ser Gln Ser Ser Gly Asn Lys Ser Ser Lys Lys Lys Glu
            100                 105                 110

Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
        115                 120                 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
    130                 135                 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Pro Glu Ser Ser Glu Leu
145                 150                 155                 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
                165                 170                 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val
            180                 185                 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
        195                 200                 205

Arg Cys Gly Asn Asn Asp Asn Pro Phe Phe Lys His His Phe Val Asp
    210                 215                 220

Arg Glu Glu Lys Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225                 230                 235                 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
                245                 250                 255
```

```
Phe Lys Gly Gly Thr Glu Thr Tyr Gln Gln Met Thr Asn Glu Val Phe
            260                 265                 270


Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
        275                 280                 285


Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys
    290                 295                 300


Pro Lys Ser Leu Tyr Asp Arg Leu Arg Glu Glu Asp Arg Ala Arg Phe
305                 310                 315                 320


Arg Val Pro Val Asp Ile Leu Ser Asp Glu Asp Asp Thr Asp Gly Thr
                325                 330                 335


Glu Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
            340                 345                 350


Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
        355                 360                 365


Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
    370                 375                 380


Lys Asn Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385                 390                 395                 400


Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
                405                 410                 415


Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
            420                 425                 430


Lys Pro Tyr Ile Thr Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly
        435                 440                 445


Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln Leu Leu Trp Pro Ser
    450                 455                 460
```

```
Pro Glu Val Gly Ala Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465                 470                 475                 480


Lys Arg Phe Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
                485                 490                 495


Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Ala
            500                 505                 510


Ser Ala Glu Arg Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
        515                 520                 525


Tyr Ala Val Tyr Asp Ala Phe Ala Arg Gly Glu Ile Asp Thr Leu Asp
    530                 535                 540


Arg Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu
545                 550                 555                 560


Pro Arg Gln Met Ile Ala Ile Leu Ser Gln Glu His Lys Asp Met Glu
                565                 570                 575


Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His
            580                 585                 590


Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly
        595                 600                 605


Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
    610                 615                 620


Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly
625                 630                 635                 640


Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
                645                 650                 655


Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro
            660                 665                 670
```

```
Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
        675                 680                 685

Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
    690                 695                 700

Arg Ser Tyr Leu Lys Ala Arg Lys Ala Phe Leu Gln Ser Ile Gly Arg
705                 710                 715                 720

Ser Asp Arg Val Glu Asn His Arg Phe Leu Leu Leu Lys Glu Pro Lys
                725                 730                 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Gly Glu Phe His Leu
            740                 745                 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
        755                 760                 765

Gly Leu Asp Glu Val Gly Ser Tyr Lys Glu Val Gly Phe Met Ala Lys
    770                 775                 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Lys Asp Arg Val Gln Pro
785                 790                 795                 800

Phe Tyr Asp Tyr Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
                805                 810                 815

Gly Arg Phe Leu Ser Lys Glu Lys Arg Ala Glu Glu Trp Glu Ser Gly
            820                 825                 830

Lys Glu Arg Phe Arg Asp Leu Glu Ala Trp Ser His Ser Ala Ala Arg
        835                 840                 845

Arg Ile Glu Asp Ala Phe Ala Gly Ile Glu Asn Ala Ser Arg Glu Asn
    850                 855                 860

Lys Lys Lys Ile Glu Gln Leu Leu Gln Asp Leu Ser Leu Trp Glu Thr
865                 870                 875                 880
```

```
Phe Glu Ser Lys Leu Lys Val Lys Ala Asp Lys Ile Asn Ile Ala Lys
                885                 890                 895


Leu Lys Lys Glu Ile Leu Glu Ala Lys Glu His Pro Tyr Leu Asp Phe
            900                 905                 910


Lys Ser Trp Gln Lys Phe Glu Arg Glu Leu Arg Leu Val Lys Asn Gln
        915                 920                 925


Asp Ile Ile Thr Trp Met Met Cys Arg Asp Leu Met Glu Glu Asn Lys
    930                 935                 940


Val Glu Gly Leu Asp Thr Gly Thr Leu Tyr Leu Lys Asp Ile Arg Thr
945                 950                 955                 960


Asp Val His Glu Gln Gly Ser Leu Asn Val Leu Asn Arg Val Lys Pro
                965                 970                 975


Met Arg Leu Pro Val Val Val Tyr Arg Ala Asp Ser Arg Gly His Val
            980                 985                 990


His Lys Glu Gln Ala Pro Leu Ala  Thr Val Tyr Ile Glu  Glu Arg Asp
        995                 1000                1005


Thr Lys  Leu Leu Lys Gln Gly  Asn Phe Lys Ser Phe  Val Lys Asp
    1010                1015                1020


Arg Arg  Leu Asn Gly Leu Phe  Ser Phe Val Asp Thr  Gly Ala Leu
    1025                1030                1035


Ala Met  Glu Gln Tyr Pro Ile  Ser Lys Leu Arg Val  Glu Tyr Glu
    1040                1045                1050


Leu Ala  Lys Tyr Gln Thr Ala  Arg Val Cys Ala Phe  Glu Gln Thr
    1055                1060                1065


Leu Glu  Leu Glu Glu Ser Leu  Leu Thr Arg Tyr Pro  His Leu Pro
    1070                1075                1080
```

```
Asp Lys  Asn Phe Arg Lys Met  Leu Glu Ser Trp Ser  Asp Pro Leu
    1085                1090                1095


Leu Asp  Lys Trp Pro Asp Leu  His Gly Asn Val Arg  Leu Leu Ile
    1100                1105                1110


Ala Val  Arg Asn Ala Phe Ser  His Asn Gln Tyr Pro  Met Tyr Asp
    1115                1120                1125


Glu Thr  Leu Phe Ser Ser Ile  Arg Lys Tyr Asp Pro  Ser Ser Pro
    1130                1135                1140


Asp Ala  Ile Glu Glu Arg Met  Gly Leu Asn Ile Ala  His Arg Leu
    1145                1150                1155


Ser Glu  Glu Val Lys Gln Ala  Lys Glu Met Val Glu  Arg Ile Ile
    1160                1165                1170


Gln Ala
    1175


<210>  98
<211>  1115
<212>  PRT
<213>  Bacterioides pyogenes

<400>  98

Met Glu Ser Ile Lys Asn Ser Gln Lys Ser Thr Gly Lys Thr Leu Gln
1               5                   10                  15


Lys Asp Pro Pro Tyr Phe Gly Leu Tyr Leu Asn Met Ala Leu Leu Asn
            20                  25                  30


Val Arg Lys Val Glu Asn His Ile Arg Lys Trp Leu Gly Asp Val Ala
        35                  40                  45


Leu Leu Pro Glu Lys Ser Gly Phe His Ser Leu Leu Thr Thr Asp Asn
    50                  55                  60


Leu Ser Ser Ala Lys Trp Thr Arg Phe Tyr Tyr Lys Ser Arg Lys Phe
```

65 70 75 80

Leu Pro Phe Leu Glu Met Phe Asp Ser Asp Lys Lys Ser Tyr Glu Asn
85 90 95

Arg Arg Glu Thr Thr Glu Cys Leu Asp Thr Ile Asp Arg Gln Lys Ile
100 105 110

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

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

Leu Ile Ile Ser Ser Glu Met His Arg Phe Ile Glu Asn Ala Tyr Ser
145 150 155 160

Phe Ala Leu Gln Lys Thr Arg Ala Arg Phe Thr Gly Val Phe Val Glu
165 170 175

Thr Asp Phe Leu Gln Ala Glu Glu Lys Gly Asp Asn Lys Lys Phe Phe
180 185 190

Ala Ile Gly Gly Asn Glu Gly Ile Lys Leu Lys Asp Asn Ala Leu Ile
195 200 205

Phe Leu Ile Cys Leu Phe Leu Asp Arg Glu Glu Ala Phe Lys Phe Leu
210 215 220

Ser Arg Ala Thr Gly Phe Lys Ser Thr Lys Glu Lys Gly Phe Leu Ala
225 230 235 240

Val Arg Glu Thr Phe Cys Ala Leu Cys Cys Arg Gln Pro His Glu Arg
245 250 255

Leu Leu Ser Val Asn Pro Arg Glu Ala Leu Leu Met Asp Met Leu Asn
260 265 270

Glu Leu Asn Arg Cys Pro Asp Ile Leu Phe Glu Met Leu Asp Glu Lys

```
        275                 280                 285

Asp Gln Lys Ser Phe Leu Pro Leu Leu Gly Glu Glu Glu Gln Ala His
    290                 295                 300

Ile Leu Glu Asn Ser Leu Asn Asp Glu Leu Cys Glu Ala Ile Asp Asp
305                 310                 315                 320

Pro Phe Glu Met Ile Ala Ser Leu Ser Lys Arg Val Arg Tyr Lys Asn
                325                 330                 335

Arg Phe Pro Tyr Leu Met Leu Arg Tyr Ile Glu Glu Lys Asn Leu Leu
            340                 345                 350

Pro Phe Ile Arg Phe Arg Ile Asp Leu Gly Cys Leu Glu Leu Ala Ser
        355                 360                 365

Tyr Pro Lys Lys Met Gly Glu Glu Asn Asn Tyr Glu Arg Ser Val Thr
    370                 375                 380

Asp His Ala Met Ala Phe Gly Arg Leu Thr Asp Phe His Asn Glu Asp
385                 390                 395                 400

Ala Val Leu Gln Gln Ile Thr Lys Gly Ile Thr Asp Glu Val Arg Phe
                405                 410                 415

Ser Leu Tyr Ala Pro Arg Tyr Ala Ile Tyr Asn Asn Lys Ile Gly Phe
            420                 425                 430

Val Arg Thr Gly Gly Ser Asp Lys Ile Ser Phe Pro Thr Leu Lys Lys
        435                 440                 445

Lys Gly Gly Glu Gly His Cys Val Ala Tyr Thr Leu Gln Asn Thr Lys
    450                 455                 460

Ser Phe Gly Phe Ile Ser Ile Tyr Asp Leu Arg Lys Ile Leu Leu Leu
465                 470                 475                 480

Ser Phe Leu Asp Lys Asp Lys Ala Lys Asn Ile Val Ser Gly Leu Leu
```

```
                485                 490                 495

Glu Gln Cys Glu Lys His Trp Lys Asp Leu Ser Glu Asn Leu Phe Asp
            500                 505                 510

Ala Ile Arg Thr Glu Leu Gln Lys Glu Phe Pro Val Pro Leu Ile Arg
        515                 520                 525

Tyr Thr Leu Pro Arg Ser Lys Gly Gly Lys Leu Val Ser Ser Lys Leu
    530                 535                 540

Ala Asp Lys Gln Glu Lys Tyr Glu Ser Glu Phe Glu Arg Arg Lys Glu
545                 550                 555                 560

Lys Leu Thr Glu Ile Leu Ser Glu Lys Asp Phe Asp Leu Ser Gln Ile
                565                 570                 575

Pro Arg Arg Met Ile Asp Glu Trp Leu Asn Val Leu Pro Thr Ser Arg
            580                 585                 590

Glu Lys Lys Leu Lys Gly Tyr Val Glu Thr Leu Lys Leu Asp Cys Arg
        595                 600                 605

Glu Arg Leu Arg Val Phe Glu Lys Arg Glu Lys Gly Glu His Pro Val
    610                 615                 620

Pro Pro Arg Ile Gly Glu Met Ala Thr Asp Leu Ala Lys Asp Ile Ile
625                 630                 635                 640

Arg Met Val Ile Asp Gln Gly Val Lys Gln Arg Ile Thr Ser Ala Tyr
                645                 650                 655

Tyr Ser Glu Ile Gln Arg Cys Leu Ala Gln Tyr Ala Gly Asp Asp Asn
            660                 665                 670

Arg Arg His Leu Asp Ser Ile Ile Arg Glu Leu Arg Leu Lys Asp Thr
        675                 680                 685

Lys Asn Gly His Pro Phe Leu Gly Lys Val Leu Arg Pro Gly Leu Gly
```

```
    690                 695                 700

His Thr Glu Lys Leu Tyr Gln Arg Tyr Phe Glu Glu Lys Lys Glu Trp
705                 710                 715                 720

Leu Glu Ala Thr Phe Tyr Pro Ala Ala Ser Pro Lys Arg Val Pro Arg
                725                 730                 735

Phe Val Asn Pro Pro Thr Gly Lys Gln Lys Glu Leu Pro Leu Ile Ile
            740                 745                 750

Arg Asn Leu Met Lys Glu Arg Pro Glu Trp Arg Asp Trp Lys Gln Arg
        755                 760                 765

Lys Asn Ser His Pro Ile Asp Leu Pro Ser Gln Leu Phe Glu Asn Glu
    770                 775                 780

Ile Cys Arg Leu Leu Lys Asp Lys Ile Gly Lys Glu Pro Ser Gly Lys
785                 790                 795                 800

Leu Lys Trp Asn Glu Met Phe Lys Leu Tyr Trp Asp Lys Glu Phe Pro
                805                 810                 815

Asn Gly Met Gln Arg Phe Tyr Arg Cys Lys Arg Arg Val Glu Val Phe
            820                 825                 830

Asp Lys Val Val Glu Tyr Glu Tyr Ser Glu Glu Gly Gly Asn Tyr Lys
        835                 840                 845

Lys Tyr Tyr Glu Ala Leu Ile Asp Glu Val Val Arg Gln Lys Ile Ser
    850                 855                 860

Ser Ser Lys Glu Lys Ser Lys Leu Gln Val Glu Asp Leu Thr Leu Ser
865                 870                 875                 880

Val Arg Arg Val Phe Lys Arg Ala Ile Asn Glu Lys Glu Tyr Gln Leu
                885                 890                 895

Arg Leu Leu Cys Glu Asp Asp Arg Leu Leu Phe Met Ala Val Arg Asp
```

```
            900                 905                 910

Leu Tyr Asp Trp Lys Glu Ala Gln Leu Asp Leu Asp Lys Ile Asp Asn
        915                 920                 925

Met Leu Gly Glu Pro Val Ser Val Ser Gln Val Ile Gln Leu Glu Gly
    930                 935                 940

Gly Gln Pro Asp Ala Val Ile Lys Ala Glu Cys Lys Leu Lys Asp Val
945                 950                 955                 960

Ser Lys Leu Met Arg Tyr Cys Tyr Asp Gly Arg Val Lys Gly Leu Met
                965                 970                 975

Pro Tyr Phe Ala Asn His Glu Ala Thr Gln Glu Gln Val Glu Met Glu
            980                 985                 990

Leu Arg His Tyr Glu Asp His Arg  Arg Arg Val Phe Asn  Trp Val Phe
        995                 1000                1005

Ala Leu  Glu Lys Ser Val Leu  Lys Asn Glu Lys Leu  Arg Arg Phe
    1010                1015                1020

Tyr Glu  Glu Ser Gln Gly Gly  Cys Glu His Arg Arg  Cys Ile Asp
    1025                1030                1035

Ala Leu  Arg Lys Ala Ser Leu  Val Ser Glu Glu Glu  Tyr Glu Phe
    1040                1045                1050

Leu Val  His Ile Arg Asn Lys  Ser Ala His Asn Gln  Phe Pro Asp
    1055                1060                1065

Leu Glu  Ile Gly Lys Leu Pro  Pro Asn Val Thr Ser  Gly Phe Cys
    1070                1075                1080

Glu Cys  Ile Trp Ser Lys Tyr  Lys Ala Ile Ile Cys  Arg Ile Ile
    1085                1090                1095

Pro Phe  Ile Asp Pro Glu Arg  Arg Phe Phe Gly Lys  Leu Leu Glu
```

1100 1105 1110

Gln Lys
1115

<210> 99
<211> 1135
<212> PRT
<213> Porphymonas gingivalis

<400> 99

Met Asn Thr Val Pro Ala Ser Glu Asn Lys Gly Gln Ser Arg Thr Val
1 5 10 15

Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu
20 25 30

Asn Leu Ile Glu Val Glu Ser His Val Arg Ile Lys Phe Gly Lys Lys
35 40 45

Lys Leu Asn Glu Glu Ser Leu Lys Gln Ser Leu Leu Cys Asp His Leu
50 55 60

Leu Ser Val Asp Arg Trp Thr Lys Val Tyr Gly His Ser Arg Arg Tyr
65 70 75 80

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

His Asp Ser Lys Thr Gly Val Asp Pro Asp Ser Ala Gln Arg Leu Ile
100 105 110

Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His
115 120 125

Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Glu Val Ser Pro Asp
130 135 140

Ile Ser Ser Phe Ile Thr Gly Thr Tyr Ser Leu Ala Cys Gly Arg Ala
145 150 155 160

```
Gln Ser Arg Phe Ala Asp Phe Phe Lys Pro Asp Asp Phe Val Leu Ala
                165                 170                 175

Lys Asn Arg Lys Glu Gln Leu Ile Ser Val Ala Asp Gly Lys Glu Cys
            180                 185                 190

Leu Thr Val Ser Gly Leu Ala Phe Phe Ile Cys Leu Phe Leu Asp Arg
        195                 200                 205

Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr
    210                 215                 220

Asp Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys
225                 230                 235                 240

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

Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu
            260                 265                 270

Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu
        275                 280                 285

Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asn Glu Glu
    290                 295                 300

Ser Arg Leu Leu Trp Asp Gly Ser Ser Asp Trp Ala Glu Ala Leu Thr
305                 310                 315                 320

Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe
                325                 330                 335

Ile Glu Glu Met Asp Leu Leu Lys Gly Ile Arg Phe Arg Val Asp Leu
            340                 345                 350

Gly Glu Ile Glu Leu Asp Ser Tyr Ser Lys Lys Val Gly Arg Asn Gly
        355                 360                 365
```

```
Glu Tyr Asp Arg Thr Ile Thr Asp His Ala Leu Ala Phe Gly Lys Leu
    370                 375                 380

Ser Asp Phe Gln Asn Glu Glu Glu Val Ser Arg Met Ile Ser Gly Glu
385                 390                 395                 400

Ala Ser Tyr Pro Val Arg Phe Ser Leu Phe Ala Pro Arg Tyr Ala Ile
                405                 410                 415

Tyr Asp Asn Lys Ile Gly Tyr Cys His Thr Ser Asp Pro Val Tyr Pro
            420                 425                 430

Lys Ser Lys Thr Gly Glu Lys Arg Ala Leu Ser Asn Pro Arg Ser Met
        435                 440                 445

Gly Phe Ile Ser Val His Asp Leu Arg Lys Leu Leu Leu Met Glu Leu
    450                 455                 460

Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Asp Phe Leu Arg Lys
465                 470                 475                 480

Ala Asn Arg Ile Leu Asp Glu Thr Ala Glu Gly Lys Leu Gln Phe Ser
                485                 490                 495

Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro
            500                 505                 510

Lys Ser Lys Asp Arg Arg Glu Lys Ala Glu Thr Thr Leu Glu Lys Tyr
        515                 520                 525

Lys Gln Glu Ile Lys Gly Arg Lys Asp Lys Leu Asn Ser Gln Leu Leu
    530                 535                 540

Ser Ala Phe Asp Met Asp Gln Arg Gln Leu Pro Ser Arg Leu Leu Asp
545                 550                 555                 560

Glu Trp Met Asn Ile Arg Pro Ala Ser His Ser Val Lys Leu Arg Thr
                565                 570                 575
```

```
Tyr Val Lys Gln Leu Asn Glu Asp Cys Arg Leu Arg Leu Gln Lys Phe
            580                 585                 590

Arg Lys Asp Gly Asp Gly Lys Ala Arg Ala Ile Pro Leu Val Gly Glu
        595                 600                 605

Met Ala Thr Phe Leu Ser Gln Asp Ile Val Arg Met Ile Ile Ser Glu
    610                 615                 620

Glu Thr Lys Lys Leu Ile Thr Ser Ala Tyr Tyr Asn Glu Met Gln Arg
625                 630                 635                 640

Ser Leu Ala Gln Tyr Ala Gly Glu Glu Asn Arg His Gln Phe Arg Ala
                645                 650                 655

Ile Val Ala Glu Leu Arg Leu Leu Asp Pro Ser Ser Gly His Pro Phe
            660                 665                 670

Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Asp Phe Tyr
        675                 680                 685

Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Thr Phe Tyr
    690                 695                 700

Arg Pro Glu Gln Asp Glu Asn Thr Lys Arg Arg Ile Ser Val Phe Phe
705                 710                 715                 720

Val Pro Asp Gly Glu Ala Arg Lys Leu Leu Pro Leu Leu Ile Arg Arg
                725                 730                 735

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

Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Ile Met
        755                 760                 765

Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe
    770                 775                 780
```

```
Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr
785                 790                 795                 800

Gly Leu Arg Arg Glu Leu Asn Ile His Gly Lys Ser Val Ser Tyr Ile
                805                 810                 815

Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu
            820                 825                 830

Lys Thr Val Gln Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro
        835                 840                 845

Val Pro Pro Asp Leu Ala Ala Asp Ile Lys Arg Ser Phe His Arg Ala
    850                 855                 860

Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg
865                 870                 875                 880

Leu Met Leu Met Ala Ile Asn Lys Met Met Thr Asp Arg Glu Glu Asp
                885                 890                 895

Ile Leu Pro Gly Leu Lys Asn Ile Asp Ser Ile Leu Asp Glu Glu Asn
            900                 905                 910

Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu
        915                 920                 925

Gly Gly Asp Asn Ser Leu Ser Leu Val Pro Ala Thr Ile Glu Ile Lys
    930                 935                 940

Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg
945                 950                 955                 960

Arg Val Pro Gly Leu Met Ser His Phe Pro Glu His Lys Ala Thr Leu
                965                 970                 975

Asp Glu Val Lys Thr Leu Leu Gly Glu Tyr Asp Arg Cys Arg Ile Lys
            980                 985                 990
```

```
Ile Phe Asp Trp Ala Phe Ala Leu  Glu Gly Ala Ile Met  Ser Asp Arg
        995                 1000                1005


Asp Leu  Lys Pro Tyr Leu His  Glu Ser Ser Ser Arg  Glu Gly Lys
    1010                1015                1020


Ser Gly  Glu His Ser Thr Leu  Val Lys Met Leu Val  Glu Lys Lys
    1025                1030                1035


Gly Cys  Leu Thr Pro Asp Glu  Ser Gln Tyr Leu Ile  Leu Ile Arg
    1040                1045                1050


Asn Lys  Ala Ala His Asn Gln  Phe Pro Cys Ala Ala  Glu Met Pro
    1055                1060                1065


Leu Ile  Tyr Arg Asp Val Ser  Ala Lys Val Gly Ser  Ile Glu Gly
    1070                1075                1080


Ser Ser  Ala Lys Asp Leu Pro  Glu Gly Ser Ser Leu  Val Asp Ser
    1085                1090                1095


Leu Trp  Lys Lys Tyr Glu Met  Ile Ile Arg Lys Ile  Leu Pro Ile
    1100                1105                1110


Leu Asp  Pro Glu Asn Arg Phe  Phe Gly Lys Leu Leu  Asn Asn Met
    1115                1120                1125


Ser Gln  Pro Ile Asn Asp Leu
    1130                1135


<210>  100
<211>  1135
<212>  PRT
<213>  Porphyromonas gingivalis

<400>  100

Met Asn Thr Val Pro Ala Ser Glu Asn Lys Gly Gln Ser Arg Thr Val
1               5                  10                  15
```

```
Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu
            20                  25                  30

Asn Leu Ile Glu Val Glu Ser His Val Arg Ile Lys Phe Gly Lys Lys
        35                  40                  45

Lys Leu Asn Glu Glu Ser Leu Lys Gln Ser Leu Leu Cys Asp His Leu
    50                  55                  60

Leu Ser Val Asp Arg Trp Thr Lys Val Tyr Gly His Ser Arg Arg Tyr
65                  70                  75                  80

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

His Asp Ser Lys Thr Gly Val Asp Pro Asp Ser Ala Gln Arg Leu Ile
            100                 105                 110

Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His
        115                 120                 125

Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Glu Val Ser Pro Asp
    130                 135                 140

Ile Ser Ser Phe Ile Thr Gly Thr Tyr Ser Leu Ala Cys Gly Arg Ala
145                 150                 155                 160

Gln Ser Arg Phe Ala Asp Phe Phe Lys Pro Asp Asp Phe Val Leu Ala
                165                 170                 175

Lys Asn Arg Lys Glu Gln Leu Ile Ser Val Ala Asp Gly Lys Glu Cys
            180                 185                 190

Leu Thr Val Ser Gly Leu Ala Phe Phe Ile Cys Leu Phe Leu Asp Arg
        195                 200                 205

Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr
    210                 215                 220
```

```
Asp Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys
225                 230                 235                 240

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

Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu
            260                 265                 270

Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu
        275                 280                 285

Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asn Glu Glu
    290                 295                 300

Ser Arg Leu Leu Trp Asp Gly Ser Ser Asp Trp Ala Glu Ala Leu Thr
305                 310                 315                 320

Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe
                325                 330                 335

Ile Glu Glu Met Asp Leu Leu Lys Gly Ile Arg Phe Arg Val Asp Leu
            340                 345                 350

Gly Glu Ile Glu Leu Asp Ser Tyr Ser Lys Lys Val Gly Arg Asn Gly
        355                 360                 365

Glu Tyr Asp Arg Thr Ile Thr Asp His Ala Leu Ala Phe Gly Lys Leu
    370                 375                 380

Ser Asp Phe Gln Asn Glu Glu Glu Val Ser Arg Met Ile Ser Gly Glu
385                 390                 395                 400

Ala Ser Tyr Pro Val Arg Phe Ser Leu Phe Ala Pro Arg Tyr Ala Ile
                405                 410                 415

Tyr Asp Asn Lys Ile Gly Tyr Cys His Thr Ser Asp Pro Val Tyr Pro
            420                 425                 430
```

```
Lys Ser Lys Thr Gly Glu Lys Arg Ala Leu Ser Asn Pro Gln Ser Met
        435                 440                 445


Gly Phe Ile Ser Val His Asp Leu Arg Lys Leu Leu Leu Met Glu Leu
    450                 455                 460


Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Gly Phe Leu Arg Lys
465                 470                 475                 480


Ala Asn Arg Ile Leu Asp Glu Thr Ala Glu Gly Lys Leu Gln Phe Ser
                485                 490                 495


Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro
            500                 505                 510


Lys Ser Lys Asp Arg Arg Glu Lys Ala Glu Thr Thr Leu Glu Lys Tyr
        515                 520                 525


Lys Gln Glu Ile Lys Gly Arg Lys Asp Lys Leu Asn Ser Gln Leu Leu
    530                 535                 540


Ser Ala Phe Asp Met Asn Gln Arg Gln Leu Pro Ser Arg Leu Leu Asp
545                 550                 555                 560


Glu Trp Met Asn Ile Arg Pro Ala Ser His Ser Val Lys Leu Arg Thr
                565                 570                 575


Tyr Val Lys Gln Leu Asn Glu Asp Cys Arg Leu Arg Leu Arg Lys Phe
            580                 585                 590


Arg Lys Asp Gly Asp Gly Lys Ala Arg Ala Ile Pro Leu Val Gly Glu
        595                 600                 605


Met Ala Thr Phe Leu Ser Gln Asp Ile Val Arg Met Ile Ile Ser Glu
    610                 615                 620


Glu Thr Lys Lys Leu Ile Thr Ser Ala Tyr Tyr Asn Glu Met Gln Arg
625                 630                 635                 640
```

```
Ser Leu Ala Gln Tyr Ala Gly Glu Glu Asn Arg Arg Gln Phe Arg Ala
                645                 650                 655

Ile Val Ala Glu Leu His Leu Leu Asp Pro Ser Ser Gly His Pro Phe
            660                 665                 670

Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Asp Phe Tyr
        675                 680                 685

Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Thr Phe Tyr
    690                 695                 700

Arg Pro Glu Gln Asp Glu Asn Thr Lys Arg Arg Ile Ser Val Phe Phe
705                 710                 715                 720

Val Pro Asp Gly Glu Ala Arg Lys Leu Leu Pro Leu Leu Ile Arg Arg
                725                 730                 735

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

Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Ile Met
        755                 760                 765

Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe
    770                 775                 780

Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr
785                 790                 795                 800

Gly Leu Arg Arg Glu Leu Asn Ile His Gly Lys Ser Val Ser Tyr Ile
                805                 810                 815

Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu
            820                 825                 830

Lys Thr Val Gln Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro
        835                 840                 845
```

```
Val Pro Pro Asp Leu Ala Ala Asp Ile Lys Arg Ser Phe His Arg Ala
    850                 855                 860

Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg
865                 870                 875                 880

Leu Met Leu Met Ala Ile Asn Lys Met Met Thr Asp Arg Glu Glu Asp
                885                 890                 895

Ile Leu Pro Gly Leu Lys Asn Ile Asp Ser Ile Leu Asp Glu Glu Asn
            900                 905                 910

Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu
        915                 920                 925

Gly Gly Asp Asn Ser Leu Ser Leu Val Pro Ala Thr Ile Glu Ile Lys
    930                 935                 940

Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg
945                 950                 955                 960

Arg Val Pro Gly Leu Met Ser His Phe Pro Glu His Lys Ala Thr Leu
                965                 970                 975

Asp Glu Val Lys Thr Leu Leu Gly Glu Tyr Asp Arg Cys Arg Ile Lys
            980                 985                 990

Ile Phe Asp Trp Ala Phe Ala Leu  Glu Gly Ala Ile Met  Ser Asp Arg
        995                 1000                1005

Asp Leu  Lys Pro Tyr Leu His  Glu Ser Ser Ser Arg  Glu Gly Lys
    1010                1015                1020

Ser Gly  Glu His Ser Thr Leu  Val Lys Met Leu Val  Glu Lys Lys
    1025                1030                1035

Gly Cys  Leu Thr Pro Asp Glu  Ser Gln Tyr Leu Ile  Leu Ile Arg
    1040                1045                1050
```

```
Asn Lys  Ala Ala His Asn Gln  Phe Pro Cys Ala Ala  Glu Met Pro
    1055                1060                1065


Leu Ile  Tyr Arg Asp Val Ser  Ala Lys Val Gly Ser  Ile Glu Gly
    1070                1075                1080


Ser Ser  Ala Lys Asp Leu Pro  Glu Gly Ser Ser Leu  Val Asp Ser
    1085                1090                1095


Leu Trp  Lys Lys Tyr Glu Met  Ile Ile Arg Lys Ile  Leu Pro Ile
    1100                1105                1110


Leu Asp  His Glu Asn Arg Phe  Phe Gly Lys Leu Leu  Asn Asn Met
    1115                1120                1125


Ser Gln  Pro Ile Asn Asp Leu
    1130                1135


<210>  101
<211>  1135
<212>  PRT
<213>  Porphyromonas gingivalis

<400>  101

Met Asn Thr Val Pro Ala Ser Glu Asn Lys Gly Gln Ser Arg Thr Val
1               5                   10                  15


Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu
            20                  25                  30


Asn Leu Ile Glu Val Glu Ser His Val Arg Ile Lys Phe Gly Lys Lys
        35                  40                  45


Lys Leu Asn Glu Glu Ser Leu Lys Gln Ser Leu Leu Cys Asp His Leu
    50                  55                  60


Leu Ser Val Asp Arg Trp Thr Lys Val Tyr Gly His Ser Arg Arg Tyr
65                  70                  75                  80
```

```
Leu Pro Phe Leu His Tyr Phe Asp Pro Asp Ser Gln Ile Glu Lys Asp
                85                  90                  95

His Asp Ser Lys Thr Gly Val Asp Pro Asp Ser Ala Gln Arg Leu Ile
            100                 105                 110

Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His
        115                 120                 125

Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Glu Val Ser Pro Asp
    130                 135                 140

Ile Ser Ser Phe Ile Thr Gly Thr Tyr Ser Leu Ala Cys Gly Arg Ala
145                 150                 155                 160

Gln Ser Arg Phe Ala Asp Phe Phe Lys Pro Asp Asp Phe Val Leu Ala
                165                 170                 175

Lys Asn Arg Lys Glu Gln Leu Ile Ser Val Ala Asp Gly Lys Glu Cys
            180                 185                 190

Leu Thr Val Ser Gly Leu Ala Phe Phe Ile Cys Leu Phe Leu Asp Arg
        195                 200                 205

Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr
    210                 215                 220

Asp Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys
225                 230                 235                 240

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

Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu
            260                 265                 270

Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu
        275                 280                 285
```

```
Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asn Glu Glu
    290                 295                 300

Ser Arg Leu Leu Trp Asp Gly Ser Ser Asp Trp Ala Glu Ala Leu Thr
305                 310                 315                 320

Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe
                325                 330                 335

Ile Glu Glu Met Asp Leu Leu Lys Gly Ile Arg Phe Arg Val Asp Leu
            340                 345                 350

Gly Glu Ile Glu Leu Asp Ser Tyr Ser Lys Lys Val Gly Arg Asn Gly
        355                 360                 365

Glu Tyr Asp Arg Thr Ile Thr Asp His Ala Leu Ala Phe Gly Lys Leu
    370                 375                 380

Ser Asp Phe Gln Asn Glu Glu Glu Val Ser Arg Met Ile Ser Gly Glu
385                 390                 395                 400

Ala Ser Tyr Pro Val Arg Phe Ser Leu Phe Ala Pro Arg Tyr Ala Ile
                405                 410                 415

Tyr Asp Asn Lys Ile Gly Tyr Cys His Thr Ser Asp Pro Val Tyr Pro
            420                 425                 430

Lys Ser Lys Thr Gly Glu Lys Arg Ala Leu Ser Asn Pro Arg Ser Met
        435                 440                 445

Gly Phe Ile Ser Val His Asp Leu Arg Lys Leu Leu Leu Met Glu Leu
    450                 455                 460

Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Asp Phe Leu Arg Lys
465                 470                 475                 480

Ala Asn Arg Ile Leu Asp Glu Thr Ala Glu Gly Lys Leu Gln Phe Ser
                485                 490                 495
```

```
Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro
            500                 505                 510

Lys Ser Lys Asp Arg Arg Glu Lys Ala Glu Thr Thr Leu Glu Lys Tyr
        515                 520                 525

Lys Gln Glu Ile Lys Gly Arg Lys Asp Lys Leu Asn Ser Gln Leu Leu
    530                 535                 540

Ser Ala Phe Asp Met Asp Gln Arg Gln Leu Pro Ser Arg Leu Leu Asp
545                 550                 555                 560

Glu Trp Met Asn Ile Arg Pro Ala Ser His Ser Val Lys Leu Arg Thr
                565                 570                 575

Tyr Val Lys Gln Leu Asn Glu Asp Cys Arg Leu Arg Leu Gln Lys Phe
            580                 585                 590

Arg Lys Asp Gly Asp Gly Lys Ala Arg Ala Ile Pro Leu Val Gly Glu
        595                 600                 605

Met Ala Thr Phe Leu Ser Gln Asp Ile Val Arg Met Ile Ile Ser Glu
    610                 615                 620

Glu Thr Lys Lys Leu Ile Thr Ser Ala Tyr Tyr Asn Glu Met Gln Arg
625                 630                 635                 640

Ser Leu Ala Gln Tyr Ala Gly Glu Glu Asn Arg His Gln Phe Arg Ala
                645                 650                 655

Ile Val Ala Glu Leu Arg Leu Leu Asp Pro Ser Ser Gly His Pro Phe
            660                 665                 670

Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Asp Phe Tyr
        675                 680                 685

Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Thr Phe Tyr
    690                 695                 700
```

```
Arg Pro Glu Gln Asp Glu Asn Thr Lys Arg Arg Ile Ser Val Phe Phe
705                 710                 715                 720


Val Pro Asp Gly Glu Ala Arg Lys Leu Leu Pro Leu Leu Ile Arg Arg
                725                 730                 735


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


Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Val Met
        755                 760                 765


Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe
    770                 775                 780


Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr
785                 790                 795                 800


Gly Leu Arg Arg Glu Leu Asn Ile His Gly Lys Ser Val Ser Tyr Ile
                805                 810                 815


Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu
            820                 825                 830


Lys Thr Val Arg Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro
        835                 840                 845


Val Pro Pro Asp Leu Ala Ala Tyr Ile Lys Arg Ser Phe His Arg Ala
    850                 855                 860


Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg
865                 870                 875                 880


Leu Met Leu Met Ala Ile Asn Lys Ile Met Thr Asp Arg Glu Glu Asp
                885                 890                 895


Ile Leu Pro Gly Leu Lys Asn Ile Asp Ser Ile Leu Asp Lys Glu Asn
            900                 905                 910
```

```
Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu
        915                 920                 925


Gly Gly Asp Asn Ser Leu Ser Leu Val Pro Ala Thr Ile Glu Ile Lys
    930                 935                 940


Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg
945                 950                 955                 960


Arg Val Pro Gly Leu Met Ser His Phe Pro Glu His Lys Ala Thr Leu
                965                 970                 975


Asp Glu Val Lys Thr Leu Leu Gly Glu Tyr Asp Arg Cys Arg Ile Lys
            980                 985                 990


Ile Phe Asp Trp Ala Phe Ala Leu  Glu Gly Ala Ile Met  Ser Asp Arg
        995                 1000                1005


Asp Leu  Lys Pro Tyr Leu His  Glu Ser Ser Ser Arg  Glu Gly Lys
    1010                1015                1020


Ser Gly  Glu His Ser Thr Leu  Val Lys Met Leu Val  Glu Lys Lys
    1025                1030                1035


Gly Cys  Leu Thr Pro Asp Glu  Ser Gln Tyr Leu Ile  Leu Ile Arg
    1040                1045                1050


Asn Lys  Ala Ala His Asn Gln  Phe Pro Cys Ala Ala  Glu Ile Pro
    1055                1060                1065


Leu Ile  Tyr Arg Asp Val Ser  Ala Lys Val Gly Ser  Ile Glu Gly
    1070                1075                1080


Ser Ser  Ala Lys Asp Leu Pro  Glu Gly Ser Ser Leu  Val Asp Ser
    1085                1090                1095


Leu Trp  Lys Lys Tyr Glu Met  Ile Ile Arg Lys Ile  Leu Pro Ile
    1100                1105                1110
```

Leu Asp Pro Glu Asn Arg Phe Phe Gly Lys Leu Leu Asn Asn Met
    1115   1120   1125

Ser Gln Pro Ile Asn Asp Leu
    1130   1135

<210> 102
<211> 1125
<212> PRT
<213> Prevotella intermedia

<400> 102

Met Lys Met Glu Asp Asp Lys Lys Thr Thr Glu Ser Thr Asn Met Leu
1   5   10   15

Asp Asn Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn
            20   25   30

Val Tyr Ile Thr Val Asn His Ile Asn Lys Val Leu Glu Leu Lys Asn
        35   40   45

Lys Lys Asp Gln Asp Ile Ile Ile Asp Asn Asp Gln Asp Ile Leu Ala
    50   55   60

Ile Lys Thr His Trp Glu Lys Val Asn Gly Asp Leu Asn Lys Thr Glu
65   70   75   80

Arg Leu Arg Glu Leu Met Thr Lys His Phe Pro Phe Leu Glu Thr Ala
                85   90   95

Ile Tyr Thr Lys Asn Lys Glu Asp Lys Glu Glu Val Lys Gln Glu Lys
            100   105   110

Gln Ala Glu Ala Gln Ser Leu Glu Ser Leu Lys Asp Cys Leu Phe Leu
        115   120   125

Phe Leu Glu Lys Leu Gln Glu Ala Arg Asn Tyr Tyr Ser His Tyr Lys
    130   135   140

Tyr Ser Glu Ser Thr Lys Glu Pro Met Leu Glu Glu Gly Leu Leu Glu

```
145                 150                 155                 160

Lys Met Tyr Asn Ile Phe Asp Asp Asn Ile Gln Leu Val Ile Lys Asp
                165                 170                 175

Tyr Gln His Asn Lys Asp Ile Asn Pro Asp Glu Asp Phe Lys His Leu
            180                 185                 190

Asp Arg Lys Gly Gln Phe Lys Tyr Ser Phe Ala Asp Asn Glu Gly Asn
        195                 200                 205

Ile Thr Glu Ser Gly Leu Leu Phe Phe Val Ser Leu Phe Leu Glu Lys
    210                 215                 220

Lys Asp Ala Ile Trp Met Gln Gln Lys Leu Thr Gly Phe Lys Asp Asn
225                 230                 235                 240

Arg Glu Ser Lys Lys Lys Met Thr His Glu Val Phe Cys Arg Arg Arg
                245                 250                 255

Met Leu Leu Pro Lys Leu Arg Leu Glu Ser Thr Gln Thr Gln Asp Trp
            260                 265                 270

Ile Leu Leu Asp Met Leu Asn Glu Leu Ile Arg Cys Pro Lys Ser Leu
        275                 280                 285

Tyr Glu Arg Leu Gln Gly Glu Tyr Arg Lys Lys Phe Asn Val Pro Phe
    290                 295                 300

Asp Ser Ala Asp Glu Asp Tyr Asp Ala Glu Gln Glu Pro Phe Lys Asn
305                 310                 315                 320

Thr Leu Val Arg His Gln Asp Arg Phe Pro Tyr Phe Ala Leu Arg Tyr
                325                 330                 335

Phe Asp Tyr Asn Glu Ile Phe Thr Asn Leu Arg Phe Gln Ile Asp Leu
            340                 345                 350

Gly Thr Tyr His Phe Ser Ile Tyr Lys Lys Leu Ile Gly Gly Gln Lys
```

```
        355                 360                 365

Glu Asp Arg His Leu Thr His Lys Leu Tyr Gly Phe Glu Arg Ile Gln
    370                 375                 380

Glu Phe Ala Lys Gln Asn Arg Pro Asp Glu Trp Lys Ala Leu Val Lys
385                 390                 395                 400

Asp Leu Asp Thr Tyr Glu Thr Ser Asn Glu Arg Tyr Ile Ser Glu Thr
                405                 410                 415

Thr Pro His Tyr His Leu Glu Asn Gln Lys Ile Gly Ile Arg Phe Arg
            420                 425                 430

Asn Gly Asn Lys Glu Ile Trp Pro Ser Leu Lys Thr Asn Gly Glu Asn
        435                 440                 445

Asn Glu Lys Ser Lys Tyr Lys Leu Asp Lys Pro Tyr Gln Ala Glu Ala
    450                 455                 460

Phe Leu Ser Val His Glu Leu Leu Pro Met Met Phe Tyr Tyr Leu Leu
465                 470                 475                 480

Leu Lys Lys Glu Glu Pro Asn Asn Asp Lys Lys Asn Ala Ser Ile Val
                485                 490                 495

Glu Gly Phe Ile Lys Arg Glu Ile Arg Asp Met Tyr Lys Leu Tyr Asp
            500                 505                 510

Ala Phe Ala Asn Gly Glu Ile Asn Asn Ile Gly Asp Leu Glu Lys Tyr
        515                 520                 525

Cys Glu Asp Lys Gly Ile Pro Lys Arg His Leu Pro Lys Gln Met Val
    530                 535                 540

Ala Ile Leu Tyr Asp Glu Pro Lys Asp Met Val Lys Glu Ala Lys Arg
545                 550                 555                 560

Lys Gln Lys Glu Met Val Lys Asp Thr Lys Lys Leu Leu Ala Thr Leu
```

```
                565                 570                 575

Glu Lys Gln Thr Gln Glu Glu Ile Glu Asp Gly Gly Arg Asn Ile Arg
            580                 585                 590

Leu Leu Lys Ser Gly Glu Ile Ala Arg Trp Leu Val Asn Asp Met Met
        595                 600                 605

Arg Phe Gln Pro Val Gln Lys Asp Asn Glu Gly Asn Pro Leu Asn Asn
    610                 615                 620

Ser Lys Ala Asn Ser Thr Glu Tyr Gln Met Leu Gln Arg Ser Leu Ala
625                 630                 635                 640

Leu Tyr Asn Lys Glu Glu Lys Pro Thr Arg Tyr Phe Arg Gln Val Asn
                645                 650                 655

Leu Ile Asn Ser Ser Asn Pro His Pro Phe Leu Lys Trp Thr Lys Trp
            660                 665                 670

Glu Glu Cys Asn Asn Ile Leu Ser Phe Tyr Arg Asn Tyr Leu Thr Lys
        675                 680                 685

Lys Ile Glu Phe Leu Asn Lys Leu Lys Pro Glu Asp Trp Glu Lys Asn
    690                 695                 700

Gln Tyr Phe Leu Lys Leu Lys Glu Pro Lys Thr Asn Arg Glu Thr Leu
705                 710                 715                 720

Val Gln Gly Trp Lys Asn Gly Phe Asn Leu Pro Arg Gly Ile Phe Thr
                725                 730                 735

Glu Pro Ile Arg Glu Trp Phe Lys Arg His Gln Asn Asp Ser Lys Glu
            740                 745                 750

Tyr Glu Lys Val Glu Ala Leu Lys Arg Val Gly Leu Val Thr Lys Val
        755                 760                 765

Ile Pro Leu Phe Phe Lys Glu Glu Tyr Phe Lys Glu Asp Ala Gln Lys
```

```
    770                 775                 780

Glu Ile Asn Asn Cys Val Gln Pro Phe Tyr Ser Phe Pro Tyr Asn Val
785                 790                 795                 800

Gly Asn Ile His Lys Pro Asp Glu Lys Asp Phe Leu Pro Ser Glu Glu
                805                 810                 815

Arg Lys Lys Leu Trp Gly Asp Lys Lys Asp Lys Phe Lys Gly Tyr Lys
            820                 825                 830

Ala Lys Val Lys Ser Lys Lys Leu Thr Asp Lys Glu Lys Glu Glu Tyr
        835                 840                 845

Arg Ser Tyr Leu Glu Phe Gln Ser Trp Asn Lys Phe Glu Arg Glu Leu
    850                 855                 860

Arg Leu Val Arg Asn Gln Asp Ile Val Thr Trp Leu Leu Cys Thr Glu
865                 870                 875                 880

Leu Ile Asp Lys Met Lys Val Glu Gly Leu Asn Val Glu Glu Leu Gln
                885                 890                 895

Lys Leu Arg Leu Lys Asp Ile Asp Thr Asp Thr Ala Lys Gln Glu Lys
            900                 905                 910

Asn Asn Ile Leu Asn Arg Ile Met Pro Met Gln Leu Pro Val Thr Val
        915                 920                 925

Tyr Glu Ile Asp Asp Ser His Asn Ile Val Lys Asp Arg Pro Leu His
    930                 935                 940

Thr Val Tyr Ile Glu Glu Thr Lys Thr Lys Leu Leu Lys Gln Gly Asn
945                 950                 955                 960

Phe Lys Ala Leu Val Lys Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe
                965                 970                 975

Val Asp Thr Ser Ser Lys Ala Glu Leu Lys Asp Lys Pro Ile Ser Lys
```

980 985 990

Ser Val Val Glu Tyr Glu Leu Gly Glu Tyr Gln Asn Ala Arg Ile Glu
995 1000 1005

Thr Ile Lys Asp Met Leu Leu Leu Glu Lys Thr Leu Ile Lys Lys
1010 1015 1020

Tyr Glu Lys Leu Pro Thr Asp Asn Phe Ser Asp Met Leu Asn Gly
1025 1030 1035

Trp Leu Glu Gly Lys Asp Glu Ser Asp Lys Ala Arg Phe Gln Asn
1040 1045 1050

Asp Val Lys Leu Leu Val Ala Val Arg Asn Ala Phe Ser His Asn
1055 1060 1065

Gln Tyr Pro Met Arg Asn Arg Ile Ala Phe Ala Asn Ile Asn Pro
1070 1075 1080

Phe Ser Leu Ser Ser Ala Asp Ile Ser Glu Glu Lys Lys Leu Asp
1085 1090 1095

Ile Ala Asn Gln Leu Lys Asp Lys Thr His Lys Ile Ile Lys Lys
1100 1105 1110

Ile Ile Glu Ile Glu Lys Pro Ile Glu Thr Lys Glu
1115 1120 1125

<210> 103
<211> 1161
<212> PRT
<213> Bacteroides bacterium

<400> 103

Met Glu Asn Gln Thr Gln Lys Gly Lys Gly Ile Tyr Tyr Tyr Tyr Thr
1 5 10 15

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

```
Asn Asn Ile Glu Gln Ile Ile Glu Glu Phe Arg Ile Arg Leu Ser Leu
        35                  40                  45

Lys Asp Glu Lys Asn Ile Lys Glu Ile Ile Asn Asn Tyr Phe Thr Asp
    50                  55                  60

Lys Lys Ser Tyr Thr Asp Trp Glu Arg Gly Ile Asn Ile Leu Lys Glu
65                  70                  75                  80

Tyr Leu Pro Val Ile Asp Tyr Leu Asp Leu Ala Ile Thr Asp Lys Glu
                85                  90                  95

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

Phe Arg Thr Asn Phe Ser Leu Leu Ile Asp Thr Ile Ile Asp Leu Arg
        115                 120                 125

Asn Phe Tyr Thr His Tyr Phe His Lys Pro Ile Ser Ile Asn Pro Asp
    130                 135                 140

Val Ala Lys Phe Leu Asp Lys Asn Leu Leu Asn Val Cys Leu Asp Ile
145                 150                 155                 160

Lys Lys Gln Lys Met Lys Thr Asp Lys Thr Lys Gln Ala Leu Lys Asp
                165                 170                 175

Gly Leu Asp Lys Glu Leu Lys Lys Leu Ile Glu Leu Lys Lys Ala Glu
            180                 185                 190

Leu Lys Glu Lys Lys Ile Lys Thr Trp Asn Ile Thr Glu Asn Val Glu
        195                 200                 205

Gly Ala Val Tyr Asn Asp Ala Phe Asn His Met Val Tyr Lys Asn Asn
    210                 215                 220

Ala Gly Val Thr Ile Leu Lys Asp Tyr His Lys Ser Ile Leu Pro Asp
225                 230                 235                 240
```

```
Asp Lys Ile Asp Ser Glu Leu Lys Leu Asn Phe Ser Ile Ser Gly Leu
                245                 250                 255

Val Phe Leu Leu Ser Met Phe Leu Ser Lys Lys Glu Ile Glu Gln Phe
            260                 265                 270

Lys Ser Asn Leu Glu Gly Phe Lys Gly Lys Val Ile Gly Glu Asn Gly
        275                 280                 285

Glu Tyr Glu Ile Ser Lys Phe Asn Asn Ser Leu Lys Tyr Met Ala Thr
    290                 295                 300

His Trp Ile Phe Ser Tyr Leu Thr Phe Lys Gly Leu Lys Gln Arg Val
305                 310                 315                 320

Lys Asn Thr Phe Asp Lys Glu Thr Leu Leu Met Gln Met Ile Asp Glu
                325                 330                 335

Leu Asn Lys Val Pro His Glu Val Tyr Gln Thr Leu Ser Lys Glu Gln
            340                 345                 350

Gln Asn Glu Phe Leu Glu Asp Ile Asn Glu Tyr Val Gln Asp Asn Glu
        355                 360                 365

Glu Asn Lys Lys Ser Met Glu Asn Ser Ile Val Val His Pro Val Ile
    370                 375                 380

Arg Lys Arg Tyr Asp Asp Lys Phe Asn Tyr Phe Ala Ile Arg Phe Leu
385                 390                 395                 400

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

Asn Phe Val His Asp Lys Arg Glu Lys Gln Ile Gln Gly Ser Met Leu
            420                 425                 430

Thr Ser Asp Arg Met Ile Lys Glu Lys Ile Asn Val Phe Gly Lys Leu
        435                 440                 445
```

```
Thr Glu Ile Ala Lys Tyr Lys Ser Asp Tyr Phe Ser Asn Glu Asn Thr
    450                 455                 460

Leu Glu Thr Ser Glu Trp Glu Leu Phe Pro Asn Pro Ser Tyr Leu Leu
465                 470                 475                 480

Ile Gln Asn Asn Ile Pro Val His Ile Asp Leu Ile His Asn Thr Glu
                485                 490                 495

Glu Ala Lys Gln Cys Gln Ile Ala Ile Asp Arg Ile Lys Cys Thr Thr
            500                 505                 510

Asn Pro Ala Lys Lys Arg Asn Thr Arg Lys Ser Lys Glu Glu Ile Ile
        515                 520                 525

Lys Ile Ile Tyr Gln Lys Asn Lys Asn Ile Lys Tyr Gly Asp Pro Thr
    530                 535                 540

Ala Leu Leu Ser Ser Asn Glu Leu Pro Ala Leu Ile Tyr Glu Leu Leu
545                 550                 555                 560

Val Asn Lys Lys Ser Gly Lys Glu Leu Glu Asn Ile Ile Val Glu Lys
                565                 570                 575

Ile Val Asn Gln Tyr Lys Thr Ile Ala Gly Phe Glu Lys Gly Gln Asn
            580                 585                 590

Leu Ser Asn Ser Leu Ile Thr Lys Lys Leu Lys Lys Ser Glu Pro Asn
        595                 600                 605

Glu Asp Lys Ile Asn Ala Glu Lys Ile Ile Leu Ala Ile Asn Arg Glu
    610                 615                 620

Leu Glu Ile Thr Glu Asn Lys Leu Asn Ile Ile Lys Asn Asn Arg Ala
625                 630                 635                 640

Glu Phe Arg Thr Gly Ala Lys Arg Lys His Ile Phe Tyr Ser Lys Glu
                645                 650                 655
```

```
Leu Gly Gln Glu Ala Thr Trp Ile Ala Tyr Asp Leu Lys Arg Phe Met
            660                 665                 670

Pro Glu Ala Ser Arg Lys Glu Trp Lys Gly Phe His His Ser Glu Leu
        675                 680                 685

Gln Lys Phe Leu Ala Phe Tyr Asp Arg Asn Lys Asn Asp Ala Lys Ala
    690                 695                 700

Leu Leu Asn Met Phe Trp Asn Phe Asp Asn Asp Gln Leu Ile Gly Asn
705                 710                 715                 720

Asp Leu Asn Ser Ala Phe Arg Glu Phe His Phe Asp Lys Phe Tyr Glu
                725                 730                 735

Lys Tyr Leu Ile Lys Arg Asp Glu Ile Leu Glu Gly Phe Lys Ser Phe
            740                 745                 750

Ile Ser Asn Phe Lys Asp Glu Pro Lys Leu Leu Lys Lys Gly Ile Lys
        755                 760                 765

Asp Ile Tyr Arg Val Phe Asp Lys Arg Tyr Tyr Ile Ile Lys Ser Thr
    770                 775                 780

Asn Ala Gln Lys Glu Gln Leu Leu Ser Lys Pro Ile Cys Leu Pro Arg
785                 790                 795                 800

Gly Ile Phe Asp Asn Lys Pro Thr Tyr Ile Glu Gly Val Lys Val Glu
                805                 810                 815

Ser Asn Ser Ala Leu Phe Ala Asp Trp Tyr Gln Tyr Thr Tyr Ser Asp
            820                 825                 830

Lys His Glu Phe Gln Ser Phe Tyr Asp Met Pro Arg Asp Tyr Lys Glu
        835                 840                 845

Gln Phe Glu Lys Phe Glu Leu Asn Asn Ile Lys Ser Ile Gln Asn Lys
    850                 855                 860
```

```
Lys Asn Leu Asn Lys Ser Asp Lys Phe Ile Tyr Phe Arg Tyr Lys Gln
865                 870                 875                 880

Asp Leu Lys Ile Lys Gln Ile Lys Ser Gln Asp Leu Phe Ile Lys Leu
                885                 890                 895

Met Val Asp Glu Leu Phe Asn Val Val Phe Lys Asn Asn Ile Glu Leu
            900                 905                 910

Asn Leu Lys Lys Leu Tyr Gln Thr Ser Asp Glu Arg Phe Lys Asn Gln
        915                 920                 925

Leu Ile Ala Asp Val Gln Lys Asn Arg Glu Lys Gly Asp Thr Ser Asp
    930                 935                 940

Asn Lys Met Asn Glu Asn Phe Ile Trp Asn Met Thr Ile Pro Leu Ser
945                 950                 955                 960

Leu Cys Asn Gly Gln Ile Glu Glu Pro Lys Val Lys Leu Lys Asp Ile
                965                 970                 975

Gly Lys Phe Arg Lys Leu Glu Thr Asp Asp Lys Val Ile Gln Leu Leu
            980                 985                 990

Glu Tyr Asp Lys Ser Lys Val Trp  Lys Lys Leu Glu Ile  Glu Asp Glu
        995                 1000                1005

Leu Glu  Asn Met Pro Asn Ser  Tyr Glu Arg Ile Arg  Arg Glu Lys
    1010                1015                1020

Leu Leu  Lys Gly Ile Gln Glu  Phe Glu His Phe Leu  Leu Glu Lys
    1025                1030                1035

Glu Lys  Phe Asp Gly Ile Asn  His Pro Lys His Phe  Glu Gln Asp
    1040                1045                1050

Leu Asn  Pro Asn Phe Lys Thr  Tyr Val Ile Asn Gly  Val Leu Arg
    1055                1060                1065
```

```
Lys Asn  Ser Lys Leu Asn Tyr  Thr Glu Ile Asp Lys  Leu Leu Asp
    1070                1075                1080


Leu Glu  His Ile Ser Ile Lys  Asp Ile Glu Thr Ser  Ala Lys Glu
    1085                1090                1095


Ile His  Leu Ala Tyr Phe Leu  Ile His Val Arg Asn  Lys Phe Gly
    1100                1105                1110


His Asn  Gln Leu Pro Lys Leu  Glu Ala Phe Glu Leu  Met Lys Lys
    1115                1120                1125


Tyr Tyr  Lys Lys Asn Asn Glu  Glu Thr Tyr Ala Glu  Tyr Phe His
    1130                1135                1140


Lys Val  Ser Ser Gln Ile Val  Asn Glu Phe Lys Asn  Ser Leu Glu
    1145                1150                1155


Lys His  Ser
    1160


<210>  104
<211>  1165
<212>  PRT
<213>  Chryseobacterium jejuense

<400>  104

Met Glu Lys Thr Gln Thr Gly Leu Gly Ile Tyr Tyr Asp His Thr Lys
1               5                   10                  15


Leu Gln Asp Lys Tyr Phe Phe Gly Gly Phe Phe Asn Leu Ala Gln Asn
            20                  25                  30


Asn Ile Asp Asn Val Ile Lys Ala Phe Ile Ile Lys Phe Phe Pro Glu
        35                  40                  45


Arg Lys Asp Lys Asp Ile Asn Ile Ala Gln Phe Leu Asp Ile Cys Phe
    50                  55                  60
```

```
Lys Asp Asn Asp Ala Asp Ser Asp Phe Gln Lys Lys Asn Lys Phe Leu
65                  70                  75                  80

Arg Ile His Phe Pro Val Ile Gly Phe Leu Thr Ser Asp Asn Asp Lys
                85                  90                  95

Ala Gly Phe Lys Lys Lys Phe Ala Leu Leu Leu Lys Thr Ile Ser Glu
            100                 105                 110

Leu Arg Asn Phe Tyr Thr His Tyr Tyr His Lys Ser Ile Glu Phe Pro
        115                 120                 125

Ser Glu Leu Phe Glu Leu Leu Asp Asp Ile Phe Val Lys Thr Thr Ser
    130                 135                 140

Glu Ile Lys Lys Leu Lys Lys Lys Asp Asp Lys Thr Gln Gln Leu Leu
145                 150                 155                 160

Asn Lys Asn Leu Ser Glu Glu Tyr Asp Ile Arg Tyr Gln Gln Gln Ile
                165                 170                 175

Glu Arg Leu Lys Glu Leu Lys Ala Gln Gly Lys Arg Val Ser Leu Thr
            180                 185                 190

Asp Glu Thr Ala Ile Arg Asn Gly Val Phe Asn Ala Ala Phe Asn His
        195                 200                 205

Leu Ile Tyr Arg Asp Gly Glu Asn Val Lys Pro Ser Arg Leu Tyr Gln
    210                 215                 220

Ser Ser Tyr Ser Glu Pro Asp Pro Ala Glu Asn Gly Ile Ser Leu Ser
225                 230                 235                 240

Gln Asn Ser Ile Leu Phe Leu Leu Ser Met Phe Leu Glu Arg Lys Glu
                245                 250                 255

Thr Glu Asp Leu Lys Ser Arg Val Lys Gly Phe Lys Ala Lys Ile Ile
            260                 265                 270
```

```
Lys Gln Gly Glu Glu Gln Ile Ser Gly Leu Lys Phe Met Ala Thr His
        275                 280                 285

Trp Val Phe Ser Tyr Leu Cys Phe Lys Gly Ile Lys Gln Lys Leu Ser
    290                 295                 300

Thr Glu Phe His Glu Glu Thr Leu Leu Ile Gln Ile Ile Asp Glu Leu
305                 310                 315                 320

Ser Lys Val Pro Asp Glu Val Tyr Ser Ala Phe Asp Ser Lys Thr Lys
                325                 330                 335

Glu Lys Phe Leu Glu Asp Ile Asn Glu Tyr Met Lys Glu Gly Asn Ala
            340                 345                 350

Asp Leu Ser Leu Glu Asp Ser Lys Val Ile His Pro Val Ile Arg Lys
        355                 360                 365

Arg Tyr Glu Asn Lys Phe Asn Tyr Phe Ala Ile Arg Phe Leu Asp Glu
    370                 375                 380

Tyr Leu Ser Ser Thr Ser Leu Lys Phe Gln Val His Val Gly Asn Tyr
385                 390                 395                 400

Val His Asp Arg Arg Val Lys His Ile Asn Gly Thr Gly Phe Gln Thr
                405                 410                 415

Glu Arg Ile Val Lys Asp Arg Ile Lys Val Phe Gly Arg Leu Ser Asn
            420                 425                 430

Ile Ser Asn Leu Lys Ala Asp Tyr Ile Lys Glu Gln Leu Glu Leu Pro
        435                 440                 445

Asn Asp Ser Asn Gly Trp Glu Ile Phe Pro Asn Pro Ser Tyr Ile Phe
    450                 455                 460

Ile Asp Asn Asn Val Pro Ile His Val Leu Ala Asp Glu Ala Thr Lys
465                 470                 475                 480
```

```
Lys Gly Ile Glu Leu Phe Lys Asp Lys Arg Arg Lys Glu Gln Pro Glu
                485                 490                 495

Glu Leu Gln Lys Arg Lys Gly Lys Ile Ser Lys Tyr Asn Ile Val Ser
            500                 505                 510

Met Ile Tyr Lys Glu Ala Lys Gly Lys Asp Lys Leu Arg Ile Asp Glu
        515                 520                 525

Pro Leu Ala Leu Leu Ser Leu Asn Glu Ile Pro Ala Leu Leu Tyr Gln
    530                 535                 540

Ile Leu Glu Lys Gly Ala Thr Pro Lys Asp Ile Glu Leu Ile Ile Lys
545                 550                 555                 560

Asn Lys Leu Thr Glu Arg Phe Glu Lys Ile Lys Asn Tyr Asp Pro Glu
                565                 570                 575

Thr Pro Ala Pro Ala Ser Gln Ile Ser Lys Arg Leu Arg Asn Asn Thr
            580                 585                 590

Thr Ala Lys Gly Gln Glu Ala Leu Asn Ala Glu Lys Leu Ser Leu Leu
        595                 600                 605

Ile Glu Arg Glu Ile Glu Asn Thr Glu Thr Lys Leu Ser Ser Ile Glu
    610                 615                 620

Glu Lys Arg Leu Lys Ala Lys Lys Glu Gln Arg Arg Asn Thr Pro Gln
625                 630                 635                 640

Arg Ser Ile Phe Ser Asn Ser Asp Leu Gly Arg Ile Ala Ala Trp Leu
                645                 650                 655

Ala Asp Asp Ile Lys Arg Phe Met Pro Ala Glu Gln Arg Lys Asn Trp
            660                 665                 670

Lys Gly Tyr Gln His Ser Gln Leu Gln Gln Ser Leu Ala Tyr Phe Glu
        675                 680                 685
```

```
Lys Arg Pro Gln Glu Ala Phe Leu Leu Leu Lys Glu Gly Trp Asp Thr
    690                 695                 700


Ser Asp Gly Ser Ser Tyr Trp Asn Asn Trp Val Met Asn Ser Phe Leu
705                 710                 715                 720


Glu Asn Asn His Phe Glu Lys Phe Tyr Lys Asn Tyr Leu Met Lys Arg
                725                 730                 735


Val Lys Tyr Phe Ser Glu Leu Ala Gly Asn Ile Lys Gln His Thr His
            740                 745                 750


Asn Thr Lys Phe Leu Arg Lys Phe Ile Lys Gln Gln Met Pro Ala Asp
        755                 760                 765


Leu Phe Pro Lys Arg His Tyr Ile Leu Lys Asp Leu Glu Thr Glu Lys
    770                 775                 780


Asn Lys Val Leu Ser Lys Pro Leu Val Phe Ser Arg Gly Leu Phe Asp
785                 790                 795                 800


Asn Asn Pro Thr Phe Ile Lys Gly Val Lys Val Thr Glu Asn Pro Glu
                805                 810                 815


Leu Phe Ala Glu Trp Tyr Ser Tyr Gly Tyr Lys Thr Glu His Val Phe
            820                 825                 830


Gln His Phe Tyr Gly Trp Glu Arg Asp Tyr Asn Glu Leu Leu Asp Ser
        835                 840                 845


Glu Leu Gln Lys Gly Asn Ser Phe Ala Lys Asn Ser Ile Tyr Tyr Asn
    850                 855                 860


Arg Glu Ser Gln Leu Asp Leu Ile Lys Leu Lys Gln Asp Leu Lys Ile
865                 870                 875                 880


Lys Lys Ile Lys Ile Gln Asp Leu Phe Leu Lys Arg Ile Ala Glu Lys
                885                 890                 895
```

```
Leu Phe Glu Asn Val Phe Asn Tyr Pro Thr Thr Leu Ser Leu Asp Glu
            900                 905                 910


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


Ala Gln Ser Leu Arg Glu Glu Gly Asp Asn Ser Pro Asn Ile Ile Lys
    930                 935                 940


Asp Asp Phe Ile Trp Ser Lys Thr Ile Ala Phe Arg Ser Lys Gln Ile
945                 950                 955                 960


Tyr Glu Pro Ala Ile Lys Leu Lys Asp Ile Gly Lys Phe Asn Arg Phe
                965                 970                 975


Val Leu Asp Asp Glu Glu Ser Lys Ala Ser Lys Leu Leu Ser Tyr Asp
            980                 985                 990


Lys Asn Lys Ile Trp Asn Lys Glu  Gln Leu Glu Arg Glu  Leu Ser Ile
        995                 1000                1005


Gly Glu  Asn Ser Tyr Glu Val  Ile Arg Arg Glu Lys  Leu Phe Lys
    1010                1015                1020


Glu Ile  Gln Asn Leu Glu Leu  Gln Ile Leu Ser Asn  Trp Ser Trp
    1025                1030                1035


Asp Gly  Ile Asn His Pro Arg  Glu Phe Glu Met Glu  Asp Gln Lys
    1040                1045                1050


Asn Thr  Arg His Pro Asn Phe  Lys Met Tyr Leu Val  Asn Gly Ile
    1055                1060                1065


Leu Arg  Lys Asn Ile Asn Leu  Tyr Lys Glu Asp Glu  Asp Phe Trp
    1070                1075                1080


Leu Glu  Ser Leu Lys Glu Asn  Asp Phe Lys Thr Leu  Pro Ser Glu
    1085                1090                1095
```

```
Val Leu  Glu Thr Lys Ser Glu  Met Val Gln Leu Leu  Phe Leu Val
    1100                1105                1110


Ile Leu  Ile Arg Asn Gln Phe  Ala His Asn Gln Leu  Pro Glu Ile
    1115                1120                1125


Gln Phe  Tyr Asn Phe Ile Arg  Lys Asn Tyr Pro Glu  Ile Gln Asn
    1130                1135                1140


Asn Thr  Val Ala Glu Leu Tyr  Leu Asn Leu Ile Lys  Leu Ala Val
    1145                1150                1155


Gln Lys  Leu Lys Asp Asn Ser
    1160                1165


<210>  105
<211>  1156
<212>  PRT
<213>  Chryseobacterium carnipullorum

<400>  105

Met Asn Thr Arg Val Thr Gly Met Gly Val Ser Tyr Asp His Thr Lys
1               5                   10                  15


Lys Glu Asp Lys His Phe Phe Gly Gly Phe Leu Asn Leu Ala Gln Asp
            20                  25                  30


Asn Ile Thr Ala Val Ile Lys Ala Phe Cys Ile Lys Phe Asp Lys Asn
        35                  40                  45


Pro Met Ser Ser Val Gln Phe Ala Glu Ser Cys Phe Thr Asp Lys Asp
    50                  55                  60


Ser Asp Thr Asp Phe Gln Asn Lys Val Arg Tyr Val Arg Thr His Leu
65                  70                  75                  80


Pro Val Ile Gly Tyr Leu Asn Tyr Gly Gly Asp Arg Asn Thr Phe Arg
                85                  90                  95
```

```
Gln Lys Leu Ser Thr Leu Leu Lys Ala Val Asp Ser Leu Arg Asn Phe
            100                 105                 110


Tyr Thr His Tyr Tyr His Ser Pro Leu Ala Leu Ser Thr Glu Leu Phe
        115                 120                 125


Glu Leu Leu Asp Thr Val Phe Ala Ser Val Ala Val Glu Val Lys Gln
    130                 135                 140


His Lys Met Lys Asp Asp Lys Thr Arg Gln Leu Leu Ser Lys Ser Leu
145                 150                 155                 160


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


Glu Leu Lys Glu Gln Gly Lys Asn Ile Asp Leu Arg Asp Glu Ala Gly
            180                 185                 190


Ile Arg Asn Gly Val Leu Asn Ala Ala Phe Asn His Leu Ile Tyr Lys
        195                 200                 205


Glu Gly Glu Ile Ala Lys Pro Thr Leu Ser Tyr Ser Ser Phe Tyr Tyr
    210                 215                 220


Gly Ala Asp Ser Ala Glu Asn Gly Ile Thr Ile Ser Gln Ser Gly Leu
225                 230                 235                 240


Leu Phe Leu Leu Ser Met Phe Leu Gly Lys Lys Glu Ile Glu Asp Leu
                245                 250                 255


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


Glu Asn Ile Ser Gly Leu Lys Phe Met Ala Thr His Trp Ile Phe Ser
        275                 280                 285


Tyr Leu Ser Phe Lys Gly Met Lys Gln Arg Leu Ser Thr Asp Phe His
    290                 295                 300
```

```
Glu Glu Thr Leu Leu Ile Gln Ile Ile Asp Glu Leu Ser Lys Val Pro
305                 310                 315                 320


Asp Glu Val Tyr His Asp Phe Asp Thr Ala Thr Arg Glu Lys Phe Val
                325                 330                 335


Glu Asp Ile Asn Glu Tyr Ile Arg Glu Gly Asn Glu Asp Phe Ser Leu
            340                 345                 350


Gly Asp Ser Thr Ile Ile His Pro Val Ile Arg Lys Arg Tyr Glu Asn
        355                 360                 365


Lys Phe Asn Tyr Phe Ala Val Arg Phe Leu Asp Glu Phe Ile Lys Phe
    370                 375                 380


Pro Ser Leu Arg Phe Gln Val His Leu Gly Asn Phe Val His Asp Arg
385                 390                 395                 400


Arg Ile Lys Asp Ile His Gly Thr Gly Phe Gln Thr Glu Arg Val Val
                405                 410                 415


Lys Asp Arg Ile Lys Val Phe Gly Lys Leu Ser Glu Ile Ser Ser Leu
            420                 425                 430


Lys Thr Glu Tyr Ile Glu Lys Glu Leu Asp Leu Asp Ser Asp Thr Gly
        435                 440                 445


Trp Glu Ile Phe Pro Asn Pro Ser Tyr Val Phe Ile Asp Asn Asn Ile
    450                 455                 460


Pro Ile Tyr Ile Ser Thr Asn Lys Thr Phe Lys Asn Gly Ser Ser Glu
465                 470                 475                 480


Phe Ile Lys Leu Arg Arg Lys Glu Lys Pro Glu Glu Met Lys Met Arg
                485                 490                 495


Gly Glu Asp Lys Lys Glu Lys Arg Asp Ile Ala Ser Met Ile Gly Asn
            500                 505                 510
```

```
Ala Gly Ser Leu Asn Ser Lys Thr Pro Leu Ala Met Leu Ser Leu Asn
        515                 520                 525


Glu Met Pro Ala Leu Leu Tyr Glu Ile Leu Val Lys Lys Thr Thr Pro
    530                 535                 540


Glu Glu Ile Glu Leu Ile Ile Lys Glu Lys Leu Asp Ser His Phe Glu
545                 550                 555                 560


Asn Ile Lys Asn Tyr Asp Pro Glu Lys Pro Leu Pro Ala Ser Gln Ile
                565                 570                 575


Ser Lys Arg Leu Arg Asn Asn Thr Thr Asp Lys Gly Lys Lys Val Ile
            580                 585                 590


Asn Pro Glu Lys Leu Ile His Leu Ile Asn Lys Glu Ile Asp Ala Thr
        595                 600                 605


Glu Ala Lys Phe Ala Leu Leu Ala Lys Asn Arg Lys Glu Leu Lys Glu
    610                 615                 620


Lys Phe Arg Gly Lys Pro Leu Arg Gln Thr Ile Phe Ser Asn Met Glu
625                 630                 635                 640


Leu Gly Arg Glu Ala Thr Trp Leu Ala Asp Asp Ile Lys Arg Phe Met
                645                 650                 655


Pro Asp Ile Leu Arg Lys Asn Trp Lys Gly Tyr Gln His Asn Gln Leu
            660                 665                 670


Gln Gln Ser Leu Ala Phe Phe Asn Ser Arg Pro Lys Glu Ala Phe Thr
        675                 680                 685


Ile Leu Gln Asp Gly Trp Asp Phe Ala Asp Gly Ser Ser Phe Trp Asn
    690                 695                 700


Gly Trp Ile Ile Asn Ser Phe Val Lys Asn Arg Ser Phe Glu Tyr Phe
705                 710                 715                 720
```

```
Tyr Glu Ala Tyr Phe Glu Gly Arg Lys Glu Tyr Phe Ser Ser Leu Ala
                725                 730                 735

Glu Asn Ile Lys Gln His Thr Ser Asn His Arg Asn Leu Arg Arg Phe
            740                 745                 750

Ile Asp Gln Gln Met Pro Lys Gly Leu Phe Glu Asn Arg His Tyr Leu
        755                 760                 765

Leu Glu Asn Leu Glu Thr Glu Lys Asn Lys Ile Leu Ser Lys Pro Leu
    770                 775                 780

Val Phe Pro Arg Gly Leu Phe Asp Thr Lys Pro Thr Phe Ile Lys Gly
785                 790                 795                 800

Ile Lys Val Asp Glu Gln Pro Glu Leu Phe Ala Glu Trp Tyr Gln Tyr
                805                 810                 815

Gly Tyr Ser Thr Glu His Val Phe Gln Asn Phe Tyr Gly Trp Glu Arg
            820                 825                 830

Asp Tyr Asn Asp Leu Leu Glu Ser Glu Leu Glu Lys Asp Asn Asp Phe
        835                 840                 845

Ser Lys Asn Ser Ile His Tyr Ser Arg Thr Ser Gln Leu Glu Leu Ile
    850                 855                 860

Lys Leu Lys Gln Asp Leu Lys Ile Lys Lys Ile Lys Ile Gln Asp Leu
865                 870                 875                 880

Phe Leu Lys Leu Ile Ala Gly His Ile Phe Glu Asn Ile Phe Lys Tyr
                885                 890                 895

Pro Ala Ser Phe Ser Leu Asp Glu Leu Tyr Leu Thr Gln Glu Glu Arg
            900                 905                 910

Leu Asn Lys Glu Gln Glu Ala Leu Ile Gln Ser Gln Arg Lys Glu Gly
        915                 920                 925
```

```
Asp His Ser Asp Asn Ile Ile Lys Asp Asn Phe Ile Gly Ser Lys Thr
    930                 935                 940

Val Thr Tyr Glu Ser Lys Gln Ile Ser Glu Pro Asn Val Lys Leu Lys
945                 950                 955                 960

Asp Ile Gly Lys Phe Asn Arg Phe Leu Leu Asp Asp Lys Val Lys Thr
                965                 970                 975

Leu Leu Ser Tyr Asn Glu Asp Lys Val Trp Asn Lys Asn Asp Leu Asp
            980                 985                 990

Leu Glu Leu Ser Ile Gly Glu Asn  Ser Tyr Glu Val Ile  Arg Arg Glu
        995                 1000                1005

Lys Leu  Phe Lys Lys Ile Gln  Asn Phe Glu Leu Gln  Thr Leu Thr
    1010                1015                1020

Asp Trp  Pro Trp Asn Gly Thr  Asp His Pro Glu Glu  Phe Gly Thr
    1025                1030                1035

Thr Asp  Asn Lys Gly Val Asn  His Pro Asn Phe Lys  Met Tyr Val
    1040                1045                1050

Val Asn  Gly Ile Leu Arg Lys  His Thr Asp Trp Phe  Lys Glu Gly
    1055                1060                1065

Glu Asp  Asn Trp Leu Glu Asn  Leu Asn Glu Thr His  Phe Lys Asn
    1070                1075                1080

Leu Ser  Phe Gln Glu Leu Glu  Thr Lys Ser Lys Ser  Ile Gln Thr
    1085                1090                1095

Ala Phe  Leu Ile Ile Met Ile  Arg Asn Gln Phe Ala  His Asn Gln
    1100                1105                1110

Leu Pro  Ala Val Gln Phe Phe  Glu Phe Ile Gln Lys  Lys Tyr Pro
    1115                1120                1125
```

Glu Ile Gln Gly Ser Thr Thr Ser Glu Leu Tyr Leu Asn Phe Ile
    1130   1135   1140

Asn Leu Ala Val Val Glu Leu Leu Glu Leu Leu Glu Lys
    1145   1150   1155

<210> 106
<211> 1036
<212> PRT
<213> Chryseobacterium ureilyticum

<400> 106

Met Glu Thr Gln Ile Leu Gly Asn Gly Ile Ser Tyr Asp His Thr Lys
1   5   10   15

Thr Glu Asp Lys His Phe Phe Gly Gly Phe Leu Asn Thr Ala Gln Asn
            20   25   30

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

Pro Arg Lys Leu Asn Ser Val Gln Phe Pro Asp Val Cys Phe Lys Lys
    50   55   60

Asn Asp Ser Asp Ala Asp Phe Gln His Lys Leu Gln Phe Ile Arg Lys
65   70   75   80

His Leu Pro Val Ile Gln Tyr Leu Lys Tyr Gly Gly Asn Arg Glu Val
                85   90   95

Leu Lys Glu Lys Phe Arg Leu Leu Leu Gln Ala Val Asp Ser Leu Arg
            100   105   110

Asn Phe Tyr Thr His Phe Tyr His Lys Pro Ile Gln Leu Pro Asn Glu
        115   120   125

Leu Leu Thr Leu Leu Asp Thr Ile Phe Gly Glu Ile Gly Asn Glu Val
    130   135   140

Arg Gln Asn Lys Met Lys Asp Asp Lys Thr Arg His Leu Leu Lys Lys

```
145                 150                 155                 160

Asn Leu Ser Glu Glu Leu Asp Phe Arg Tyr Gln Glu Gln Leu Glu Arg
                165                 170                 175

Leu Arg Lys Leu Lys Ser Glu Gly Lys Lys Val Asp Leu Arg Asp Thr
            180                 185                 190

Glu Ala Ile Arg Asn Gly Val Leu Asn Ala Ala Phe Asn His Leu Ile
        195                 200                 205

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

Tyr Tyr Asp Ser Asp Thr Ala Glu Asn Gly Ile Ser Ile Ser Gln Ser
225                 230                 235                 240

Gly Leu Leu Phe Leu Leu Ser Met Phe Leu Gly Arg Arg Glu Met Glu
                245                 250                 255

Asp Leu Lys Ser Arg Val Arg Gly Phe Lys Ala Arg Ile Ile Lys His
            260                 265                 270

Glu Glu Gln His Val Ser Gly Leu Lys Phe Met Ala Thr His Trp Val
        275                 280                 285

Phe Ser Glu Phe Cys Phe Lys Gly Ile Lys Thr Arg Leu Asn Ala Asp
    290                 295                 300

Tyr His Glu Glu Thr Leu Leu Ile Gln Leu Ile Asp Glu Leu Ser Lys
305                 310                 315                 320

Val Pro Asp Glu Leu Tyr Arg Ser Phe Asp Val Ala Thr Arg Glu Arg
                325                 330                 335

Phe Ile Glu Asp Ile Asn Glu Tyr Ile Arg Asp Gly Lys Glu Asp Lys
            340                 345                 350

Ser Leu Ile Glu Ser Lys Ile Val His Pro Val Ile Arg Lys Arg Tyr
```

```
        355                 360                 365

Glu Ser Lys Phe Asn Tyr Phe Ala Ile Arg Phe Leu Asp Glu Phe Val
    370                 375                 380

Asn Phe Pro Thr Leu Arg Phe Gln Val His Ala Gly Asn Tyr Val His
385                 390                 395                 400

Asp Arg Arg Ile Lys Ser Ile Glu Gly Thr Gly Phe Lys Thr Glu Arg
                405                 410                 415

Leu Val Lys Asp Arg Ile Lys Val Phe Gly Lys Leu Ser Thr Ile Ser
            420                 425                 430

Ser Leu Lys Ala Glu Tyr Leu Ala Lys Ala Val Asn Ile Thr Asp Asp
        435                 440                 445

Thr Gly Trp Glu Leu Leu Pro His Pro Ser Tyr Val Phe Ile Asp Asn
    450                 455                 460

Asn Ile Pro Ile His Leu Thr Val Asp Pro Ser Phe Lys Asn Gly Val
465                 470                 475                 480

Lys Glu Tyr Gln Glu Lys Arg Lys Leu Gln Lys Pro Glu Glu Met Lys
                485                 490                 495

Asn Arg Gln Gly Gly Asp Lys Met His Lys Pro Ala Ile Ser Ser Lys
            500                 505                 510

Ile Gly Lys Ser Lys Asp Ile Asn Pro Glu Ser Pro Val Ala Leu Leu
        515                 520                 525

Ser Met Asn Glu Ile Pro Ala Leu Leu Tyr Glu Ile Leu Val Lys Lys
    530                 535                 540

Ala Ser Pro Glu Glu Val Glu Ala Lys Ile Arg Gln Lys Leu Thr Ala
545                 550                 555                 560

Val Phe Glu Arg Ile Arg Asp Tyr Asp Pro Lys Val Pro Leu Pro Ala
```

```
                565                 570                 575

Ser Gln Val Ser Lys Arg Leu Arg Asn Asn Thr Asp Thr Leu Ser Tyr
            580                 585                 590

Asn Lys Glu Lys Leu Val Glu Leu Ala Asn Lys Glu Val Glu Gln Thr
        595                 600                 605

Glu Arg Lys Leu Ala Leu Ile Thr Lys Asn Arg Arg Glu Cys Arg Glu
    610                 615                 620

Lys Val Lys Gly Lys Phe Lys Arg Gln Lys Val Phe Lys Asn Ala Glu
625                 630                 635                 640

Leu Gly Thr Glu Ala Thr Trp Leu Ala Asn Asp Ile Lys Arg Phe Met
                645                 650                 655

Pro Glu Glu Gln Lys Lys Asn Trp Lys Gly Tyr Gln His Ser Gln Leu
            660                 665                 670

Gln Gln Ser Leu Ala Phe Phe Glu Ser Arg Pro Gly Glu Ala Arg Ser
        675                 680                 685

Leu Leu Gln Ala Gly Trp Asp Phe Ser Asp Gly Ser Ser Phe Trp Asn
    690                 695                 700

Gly Trp Val Met Asn Ser Phe Ala Arg Asp Asn Thr Phe Asp Gly Phe
705                 710                 715                 720

Tyr Glu Ser Tyr Leu Asn Gly Arg Met Lys Tyr Phe Leu Arg Leu Ala
                725                 730                 735

Asp Asn Ile Ala Gln Gln Ser Ser Thr Asn Lys Leu Ile Ser Asn Phe
            740                 745                 750

Ile Lys Gln Gln Met Pro Lys Gly Leu Phe Asp Arg Arg Leu Tyr Met
        755                 760                 765

Leu Glu Asp Leu Ala Thr Glu Lys Asn Lys Ile Leu Ser Lys Pro Leu
```

```
    770                 775                 780

Ile Phe Pro Arg Gly Ile Phe Asp Asp Lys Pro Thr Phe Lys Lys Gly
785                 790                 795                 800

Val Gln Val Ser Glu Glu Pro Glu Ala Phe Ala Asp Trp Tyr Ser Tyr
                805                 810                 815

Gly Tyr Asp Val Lys His Lys Phe Gln Glu Phe Tyr Ala Trp Asp Arg
            820                 825                 830

Asp Tyr Glu Glu Leu Leu Arg Glu Glu Leu Glu Lys Asp Thr Ala Phe
        835                 840                 845

Thr Lys Asn Ser Ile His Tyr Ser Arg Glu Ser Gln Ile Glu Leu Leu
    850                 855                 860

Ala Lys Lys Gln Asp Leu Lys Val Lys Lys Val Arg Ile Gln Asp Leu
865                 870                 875                 880

Tyr Leu Lys Leu Met Ala Glu Phe Leu Phe Glu Asn Val Phe Gly His
                885                 890                 895

Glu Leu Ala Leu Pro Leu Asp Gln Phe Tyr Leu Thr Gln Glu Glu Arg
            900                 905                 910

Leu Lys Gln Glu Gln Glu Ala Ile Val Gln Ser Gln Arg Pro Lys Gly
        915                 920                 925

Asp Asp Ser Pro Asn Ile Val Lys Glu Asn Phe Ile Trp Ser Lys Thr
    930                 935                 940

Ile Pro Phe Lys Ser Gly Arg Val Phe Glu Pro Asn Val Lys Leu Lys
945                 950                 955                 960

Asp Ile Gly Lys Phe Arg Asn Leu Leu Thr Asp Glu Lys Val Asp Ile
                965                 970                 975

Leu Leu Ser Tyr Asn Asn Thr Glu Ile Gly Lys Gln Val Ile Glu Asn
```

980 985 990

Glu Leu Ile Ile Gly Ala Gly Ser Tyr Glu Phe Ile Arg Arg Glu Gln
995 1000 1005

Leu Phe Lys Glu Ile Gln Gln Met Lys Arg Leu Ser Leu Arg Ser
1010 1015 1020

Val Arg Gly Met Gly Val Pro Ile Arg Leu Asn Leu Lys
1025 1030 1035

<210> 107
<211> 1127
<212> PRT
<213> Prevotella buccae

<400> 107

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

Phe Ala Phe Pro Lys Tyr Ile Thr Ile Met Glu Asn Gln Glu Lys Pro
20 25 30

Glu Pro Ile Tyr Tyr Glu Leu Thr Asp Lys His Phe Trp Ala Ala Phe
35 40 45

Leu Asn Leu Ala Arg His Asn Val Tyr Thr Thr Ile Asn His Ile Asn
50 55 60

Arg Arg Leu Glu Ile Ala Glu Leu Lys Asp Asp Gly Tyr Met Met Asp
65 70 75 80

Ile Lys Gly Ser Trp Asn Glu Gln Ala Lys Lys Leu Asp Lys Lys Val
85 90 95

Arg Leu Arg Asp Leu Ile Met Lys His Phe Pro Phe Leu Glu Ala Ala
100 105 110

Ala Tyr Glu Ile Thr Asn Ser Lys Ser Pro Asn Asn Lys Glu Gln Arg
115 120 125

```
Glu Lys Glu Gln Ser Glu Ala Leu Ser Leu Asn Asn Leu Lys Asn Val
    130                 135                 140

Leu Phe Ile Phe Leu Glu Lys Leu Gln Val Leu Arg Asn Tyr Tyr Ser
145                 150                 155                 160

His Tyr Lys Tyr Ser Glu Glu Ser Pro Lys Pro Ile Phe Glu Thr Ser
                165                 170                 175

Leu Leu Lys Asn Met Tyr Lys Val Phe Asp Ala Asn Val Arg Leu Val
            180                 185                 190

Lys Arg Asp Tyr Met His His Glu Asn Ile Asp Met Gln Arg Asp Phe
        195                 200                 205

Thr His Leu Asn Arg Lys Lys Gln Val Gly Arg Thr Lys Asn Ile Ile
    210                 215                 220

Asp Ser Pro Asn Phe His Tyr His Phe Ala Asp Lys Glu Gly Asn Met
225                 230                 235                 240

Thr Ile Ala Gly Leu Leu Phe Phe Val Ser Leu Phe Leu Asp Lys Lys
                245                 250                 255

Asp Ala Ile Trp Met Gln Lys Lys Leu Lys Gly Phe Lys Asp Gly Arg
            260                 265                 270

Asn Leu Arg Glu Gln Met Thr Asn Glu Val Phe Cys Arg Ser Arg Ile
        275                 280                 285

Ser Leu Pro Lys Leu Lys Leu Glu Asn Val Gln Thr Lys Asp Trp Met
    290                 295                 300

Gln Leu Asp Met Leu Asn Glu Leu Val Arg Cys Pro Lys Ser Leu Tyr
305                 310                 315                 320

Glu Arg Leu Arg Glu Lys Asp Arg Glu Ser Phe Lys Val Pro Phe Asp
                325                 330                 335
```

```
Ile Phe Ser Asp Asp Tyr Asp Ala Glu Glu Glu Pro Phe Lys Asn Thr
            340                 345                 350

Leu Val Arg His Gln Asp Arg Phe Pro Tyr Phe Val Leu Arg Tyr Phe
        355                 360                 365

Asp Leu Asn Glu Ile Phe Glu Gln Leu Arg Phe Gln Ile Asp Leu Gly
    370                 375                 380

Thr Tyr His Phe Ser Ile Tyr Asn Lys Arg Ile Gly Asp Glu Asp Glu
385                 390                 395                 400

Val Arg His Leu Thr His His Leu Tyr Gly Phe Ala Arg Ile Gln Asp
                405                 410                 415

Phe Ala Gln Gln Asn Gln Pro Glu Val Trp Arg Lys Leu Val Lys Asp
            420                 425                 430

Leu Asp Tyr Phe Glu Ala Ser Gln Glu Pro Tyr Ile Pro Lys Thr Ala
        435                 440                 445

Pro His Tyr His Leu Glu Asn Glu Lys Ile Gly Ile Lys Phe Cys Ser
    450                 455                 460

Thr His Asn Asn Leu Phe Pro Ser Leu Lys Thr Glu Lys Thr Cys Asn
465                 470                 475                 480

Gly Arg Ser Lys Phe Asn Leu Gly Thr Gln Phe Thr Ala Glu Ala Phe
                485                 490                 495

Leu Ser Val His Glu Leu Leu Pro Met Met Phe Tyr Tyr Leu Leu Leu
            500                 505                 510

Thr Lys Asp Tyr Ser Arg Lys Glu Ser Ala Asp Lys Val Glu Gly Ile
        515                 520                 525

Ile Arg Lys Glu Ile Ser Asn Ile Tyr Ala Ile Tyr Asp Ala Phe Ala
    530                 535                 540
```

```
Asn Gly Glu Ile Asn Ser Ile Ala Asp Leu Thr Cys Arg Leu Gln Lys
545                 550                 555                 560

Thr Asn Ile Leu Gln Gly His Leu Pro Lys Gln Met Ile Ser Ile Leu
                565                 570                 575

Glu Gly Arg Gln Lys Asp Met Glu Lys Glu Ala Glu Arg Lys Ile Gly
            580                 585                 590

Glu Met Ile Asp Asp Thr Gln Arg Arg Leu Asp Leu Leu Cys Lys Gln
        595                 600                 605

Thr Asn Gln Lys Ile Arg Ile Gly Lys Arg Asn Ala Gly Leu Leu Lys
    610                 615                 620

Ser Gly Lys Ile Ala Asp Trp Leu Val Asn Asp Met Met Arg Phe Gln
625                 630                 635                 640

Pro Val Gln Lys Asp Gln Asn Asn Ile Pro Ile Asn Asn Ser Lys Ala
                645                 650                 655

Asn Ser Thr Glu Tyr Arg Met Leu Gln Arg Ala Leu Ala Leu Phe Gly
            660                 665                 670

Ser Glu Asn Phe Arg Leu Lys Ala Tyr Phe Asn Gln Met Asn Leu Val
        675                 680                 685

Gly Asn Asp Asn Pro His Pro Phe Leu Ala Glu Thr Gln Trp Glu His
    690                 695                 700

Gln Thr Asn Ile Leu Ser Phe Tyr Arg Asn Tyr Leu Glu Ala Arg Lys
705                 710                 715                 720

Lys Tyr Leu Lys Gly Leu Lys Pro Gln Asn Trp Lys Gln Tyr Gln His
                725                 730                 735

Phe Leu Ile Leu Lys Val Gln Lys Thr Asn Arg Asn Thr Leu Val Thr
            740                 745                 750
```

```
Gly Trp Lys Asn Ser Phe Asn Leu Pro Arg Gly Ile Phe Thr Gln Pro
        755                 760                 765

Ile Arg Glu Trp Phe Glu Lys His Asn Asn Ser Lys Arg Ile Tyr Asp
    770                 775                 780

Gln Ile Leu Ser Phe Asp Arg Val Gly Phe Val Ala Lys Ala Ile Pro
785                 790                 795                 800

Leu Tyr Phe Ala Glu Glu Tyr Lys Asp Asn Val Gln Pro Phe Tyr Asp
                805                 810                 815

Tyr Pro Phe Asn Ile Gly Asn Lys Leu Lys Pro Gln Lys Gly Gln Phe
            820                 825                 830

Leu Asp Lys Lys Glu Arg Val Glu Leu Trp Gln Lys Asn Lys Glu Leu
        835                 840                 845

Phe Lys Asn Tyr Pro Ser Glu Lys Lys Lys Thr Asp Leu Ala Tyr Leu
    850                 855                 860

Asp Phe Leu Ser Trp Lys Lys Phe Glu Arg Glu Leu Arg Leu Ile Lys
865                 870                 875                 880

Asn Gln Asp Ile Val Thr Trp Leu Met Phe Lys Glu Leu Phe Asn Met
                885                 890                 895

Ala Thr Val Glu Gly Leu Lys Ile Gly Glu Ile His Leu Arg Asp Ile
            900                 905                 910

Asp Thr Asn Thr Ala Asn Glu Glu Ser Asn Asn Ile Leu Asn Arg Ile
        915                 920                 925

Met Pro Met Lys Leu Pro Val Lys Thr Tyr Glu Thr Asp Asn Lys Gly
    930                 935                 940

Asn Ile Leu Lys Glu Arg Pro Leu Ala Thr Phe Tyr Ile Glu Glu Thr
945                 950                 955                 960
```

```
Glu Thr Lys Val Leu Lys Gln Gly Asn Phe Lys Val Leu Ala Lys Asp
                965                 970                 975

Arg Arg Leu Asn Gly Leu Leu Ser Phe Ala Glu Thr Thr Asp Ile Asp
            980                 985                 990

Leu Glu Lys Asn Pro Ile Thr Lys  Leu Ser Val Asp His  Glu Leu Ile
        995                 1000                1005

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

Leu Glu  Lys Lys Leu Ile Asn  Lys Tyr Pro Thr Leu  Pro Thr Asp
    1025                1030                1035

Ser Phe  Arg Asn Met Leu Glu  Arg Trp Leu Gln Cys  Lys Ala Asn
    1040                1045                1050

Arg Pro  Glu Leu Lys Asn Tyr  Val Asn Ser Leu Ile  Ala Val Arg
    1055                1060                1065

Asn Ala  Phe Ser His Asn Gln  Tyr Pro Met Tyr Asp  Ala Thr Leu
    1070                1075                1080

Phe Ala  Glu Val Lys Lys Phe  Thr Leu Phe Pro Ser  Val Asp Thr
    1085                1090                1095

Lys Lys  Ile Glu Leu Asn Ile  Ala Pro Gln Leu Leu  Glu Ile Val
    1100                1105                1110

Gly Lys  Ala Ile Lys Glu Ile  Glu Lys Ser Glu Asn  Lys Asn
    1115                1120                1125
```

<210> 108
<211> 1135
<212> PRT
<213> Porphyromonas gingivalis

<400> 108

```
Met Asn Thr Val Pro Ala Ser Glu Asn Lys Gly Gln Ser Arg Thr Val
1               5                  10                  15

Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu
            20                  25                  30

Asn Leu Ile Glu Val Glu Ser His Val Arg Ile Lys Phe Gly Lys Lys
        35                  40                  45

Lys Leu Asn Glu Glu Ser Leu Lys Gln Ser Leu Leu Cys Asp His Leu
    50                  55                  60

Leu Ser Val Asp Arg Trp Thr Lys Val Tyr Gly His Ser Arg Arg Tyr
65                  70                  75                  80

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

His Asp Ser Lys Thr Gly Val Asp Pro Asp Ser Ala Gln Arg Leu Ile
            100                 105                 110

Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His
        115                 120                 125

Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Glu Val Ser Pro Asp
    130                 135                 140

Ile Ser Ser Phe Ile Thr Gly Thr Tyr Ser Leu Ala Cys Gly Arg Ala
145                 150                 155                 160

Gln Ser Arg Phe Ala Asp Phe Phe Lys Pro Asp Asp Phe Val Leu Ala
                165                 170                 175

Lys Asn Arg Lys Glu Gln Leu Ile Ser Val Ala Asp Gly Lys Glu Cys
            180                 185                 190

Leu Thr Val Ser Gly Leu Ala Phe Phe Ile Cys Leu Phe Leu Asp Arg
        195                 200                 205
```

```
Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr
    210                 215                 220

Asp Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys
225                 230                 235                 240

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

Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu
            260                 265                 270

Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu
        275                 280                 285

Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asn Glu Glu
    290                 295                 300

Ser Arg Leu Leu Trp Asp Gly Ser Ser Asp Trp Ala Glu Ala Leu Thr
305                 310                 315                 320

Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe
                325                 330                 335

Ile Glu Glu Met Asp Leu Leu Lys Gly Ile Arg Phe Arg Val Asp Leu
            340                 345                 350

Gly Glu Ile Glu Leu Asp Ser Tyr Ser Lys Lys Val Gly Arg Asn Gly
        355                 360                 365

Glu Tyr Asp Arg Thr Ile Thr Asp His Ala Leu Ala Phe Gly Lys Leu
    370                 375                 380

Ser Asp Phe Gln Asn Glu Glu Glu Val Ser Arg Met Ile Ser Gly Glu
385                 390                 395                 400

Ala Ser Tyr Pro Val Arg Phe Ser Leu Phe Ala Pro Arg Tyr Ala Ile
                405                 410                 415
```

```
Tyr Asp Asn Lys Ile Gly Tyr Cys His Thr Ser Asp Pro Val Tyr Pro
            420                 425                 430

Lys Ser Lys Thr Gly Glu Lys Arg Ala Leu Ser Asn Pro Gln Ser Met
        435                 440                 445

Gly Phe Ile Ser Val His Asn Leu Arg Lys Leu Leu Leu Met Glu Leu
    450                 455                 460

Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Asp Phe Leu Arg Lys
465                 470                 475                 480

Ala Asn Arg Ile Leu Asp Glu Thr Ala Glu Gly Lys Leu Gln Phe Ser
                485                 490                 495

Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro
            500                 505                 510

Lys Ser Lys Asp Arg Arg Glu Lys Ala Glu Thr Thr Leu Glu Lys Tyr
        515                 520                 525

Lys Gln Glu Ile Lys Gly Arg Lys Asp Lys Leu Asn Ser Gln Leu Leu
    530                 535                 540

Ser Ala Phe Asp Met Asn Gln Arg Gln Leu Pro Ser Arg Leu Leu Asp
545                 550                 555                 560

Glu Trp Met Asn Ile Arg Pro Ala Ser His Ser Val Lys Leu Arg Thr
                565                 570                 575

Tyr Val Lys Gln Leu Asn Glu Asp Cys Arg Leu Arg Leu Arg Lys Phe
            580                 585                 590

Arg Lys Asp Gly Asp Gly Lys Ala Arg Ala Ile Pro Leu Val Gly Glu
        595                 600                 605

Met Ala Thr Phe Leu Ser Gln Asp Ile Val Arg Met Ile Ile Ser Glu
    610                 615                 620
```

```
Glu Thr Lys Lys Leu Ile Thr Ser Ala Tyr Tyr Asn Glu Met Gln Arg
625                 630                 635                 640

Ser Leu Ala Gln Tyr Ala Gly Glu Glu Asn Arg Arg Gln Phe Arg Ala
                645                 650                 655

Ile Val Ala Glu Leu His Leu Leu Asp Pro Ser Ser Gly His Pro Phe
            660                 665                 670

Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Asp Phe Tyr
        675                 680                 685

Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Thr Phe Tyr
    690                 695                 700

Arg Pro Glu Gln Asp Glu Asn Thr Lys Arg Arg Ile Ser Val Phe Phe
705                 710                 715                 720

Val Pro Asp Gly Glu Ala Arg Lys Leu Leu Pro Leu Leu Ile Arg Arg
                725                 730                 735

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

Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Ile Met
        755                 760                 765

Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe
    770                 775                 780

Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr
785                 790                 795                 800

Gly Leu Arg Arg Glu Leu Asn Ile His Gly Lys Ser Val Ser Tyr Ile
                805                 810                 815

Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu
            820                 825                 830
```

```
Lys Thr Val Gln Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro
        835                 840                 845


Val Pro Pro Asp Leu Ala Ala Asp Ile Lys Arg Ser Phe His Arg Ala
    850                 855                 860


Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg
865                 870                 875                 880


Leu Met Leu Met Ala Ile Asn Lys Met Met Thr Asp Arg Glu Glu Asp
                885                 890                 895


Ile Leu Pro Gly Leu Lys Asn Ile Asp Ser Ile Leu Asp Glu Glu Asn
            900                 905                 910


Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu
        915                 920                 925


Gly Gly Asp Asn Ser Leu Ser Leu Val Pro Ala Thr Ile Glu Ile Lys
    930                 935                 940


Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg
945                 950                 955                 960


Arg Val Pro Gly Leu Met Ser His Phe Pro Glu His Lys Ala Thr Leu
                965                 970                 975


Asp Glu Val Lys Thr Leu Leu Gly Glu Tyr Asp Arg Cys Arg Ile Lys
            980                 985                 990


Ile Phe Asp Trp Ala Phe Ala Leu  Glu Gly Ala Ile Met  Ser Asp Arg
        995                 1000                1005


Asp Leu  Lys Pro Tyr Leu His  Glu Ser Ser Ser Arg  Glu Gly Lys
    1010                1015                1020


Ser Gly  Glu His Ser Thr Leu  Val Lys Met Leu Val  Glu Lys Lys
    1025                1030                1035
```

```
Gly Cys  Leu Thr Pro Asp Glu  Ser Gln Tyr Leu Ile  Leu Ile Arg
    1040                1045                1050


Asn Lys  Ala Ala His Asn Gln  Phe Pro Cys Ala Ala  Glu Met Pro
    1055                1060                1065


Leu Ile  Tyr Arg Asp Val Ser  Ala Lys Val Gly Ser  Ile Glu Gly
    1070                1075                1080


Ser Ser  Ala Lys Asp Leu Pro  Glu Gly Ser Ser Leu  Val Asp Ser
    1085                1090                1095


Leu Trp  Lys Lys Tyr Glu Met  Ile Ile Arg Lys Ile  Leu Pro Ile
    1100                1105                1110


Leu Asp  Pro Glu Asn Arg Phe  Phe Gly Lys Leu Leu  Asn Asn Met
    1115                1120                1125


Ser Gln  Pro Ile Asn Asp Leu
    1130                1135


<210>  109
<211>  1175
<212>  PRT
<213>  Porphyromonas gingivalis

<400>  109

Met Thr Glu Gln Asn Glu Lys Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1               5                  10                  15


Glu Asp Lys His Phe Trp Ala Ala Phe Phe Asn Leu Ala Arg His Asn
            20                  25                  30


Ala Tyr Ile Thr Leu Ala His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
        35                  40                  45


Ala Asp Ile Thr Asn Asp Glu Asp Ile Leu Phe Phe Lys Gly Gln Trp
    50                  55                  60
```

```
Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ala Arg Leu Arg Ser Leu
65                  70                  75                  80


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


Leu Phe Glu Ser Gln Ser Ser Gly Asn Lys Ser Ser Lys Lys Lys Glu
            100                 105                 110


Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
        115                 120                 125


Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
    130                 135                 140


Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Pro Glu Ser Ser Glu Leu
145                 150                 155                 160


Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
                165                 170                 175


Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val
            180                 185                 190


Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
        195                 200                 205


Lys Tyr Gly Asn Asn Asp Asn Pro Phe Phe Lys His His Phe Val Asp
    210                 215                 220


Arg Glu Glu Lys Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225                 230                 235                 240


Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
                245                 250                 255


Phe Lys Gly Gly Thr Glu Ala Tyr Gln Gln Met Thr Asn Glu Val Phe
            260                 265                 270
```

```
Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
        275                 280                 285

Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys
    290                 295                 300

Pro Lys Ser Leu Tyr Asp Arg Leu Arg Glu Glu Asp Arg Ala Arg Phe
305                 310                 315                 320

Arg Val Pro Val Asp Ile Leu Ser Asp Glu Asp Asp Thr Asp Gly Thr
                325                 330                 335

Glu Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
            340                 345                 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
        355                 360                 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
    370                 375                 380

Lys Asn Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385                 390                 395                 400

Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
                405                 410                 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
            420                 425                 430

Lys Pro Tyr Ile Thr Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly
        435                 440                 445

Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln Leu Leu Trp Pro Ser
    450                 455                 460

Pro Glu Val Gly Ala Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465                 470                 475                 480
```

```
Lys Arg Phe Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
                485                 490                 495

Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Ala
            500                 505                 510

Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
        515                 520                 525

Tyr Ala Val Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Arg Asp
    530                 535                 540

Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu
545                 550                 555                 560

Pro Arg Gln Met Ile Ala Ile Leu Ser Gln Glu His Lys Asp Met Glu
                565                 570                 575

Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His
            580                 585                 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly
        595                 600                 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
    610                 615                 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly
625                 630                 635                 640

Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
                645                 650                 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro
            660                 665                 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
        675                 680                 685
```

```
Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
    690                 695                 700

Arg Ser Tyr Leu Lys Ala Arg Lys Ala Phe Leu Gln Ser Ile Gly Arg
705                 710                 715                 720

Ser Asp Arg Glu Glu Asn His Arg Phe Leu Leu Leu Lys Glu Pro Lys
                725                 730                 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Ser Glu Phe His Leu
            740                 745                 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
        755                 760                 765

Gly Tyr Asp Glu Val Gly Ser Tyr Lys Glu Val Gly Phe Met Ala Lys
    770                 775                 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Lys Asp Arg Val Gln Pro
785                 790                 795                 800

Phe Tyr Asp Tyr Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
                805                 810                 815

Gly Arg Phe Leu Ser Lys Glu Lys Arg Ala Glu Glu Trp Glu Ser Gly
            820                 825                 830

Lys Glu Arg Phe Arg Asp Leu Glu Ala Trp Ser His Ser Ala Ala Arg
        835                 840                 845

Arg Ile Glu Asp Ala Phe Val Gly Ile Glu Tyr Ala Ser Trp Glu Asn
    850                 855                 860

Lys Lys Lys Ile Glu Gln Leu Leu Gln Asp Leu Ser Leu Trp Glu Thr
865                 870                 875                 880

Phe Glu Ser Lys Leu Lys Val Lys Ala Asp Lys Ile Asn Ile Ala Lys
                885                 890                 895
```

```
Leu Lys Lys Glu Ile Leu Glu Ala Lys Glu His Pro Tyr His Asp Phe
            900                 905                 910

Lys Ser Trp Gln Lys Phe Glu Arg Glu Leu Arg Leu Val Lys Asn Gln
        915                 920                 925

Asp Ile Ile Thr Trp Met Met Cys Arg Asp Leu Met Glu Glu Asn Lys
    930                 935                 940

Val Glu Gly Leu Asp Thr Gly Thr Leu Tyr Leu Lys Asp Ile Arg Thr
945                 950                 955                 960

Asp Val Gln Glu Gln Gly Ser Leu Asn Val Leu Asn His Val Lys Pro
                965                 970                 975

Met Arg Leu Pro Val Val Val Tyr Arg Ala Asp Ser Arg Gly His Val
            980                 985                 990

His Lys Glu Glu Ala Pro Leu Ala  Thr Val Tyr Ile Glu  Glu Arg Asp
        995                 1000                1005

Thr Lys  Leu Leu Lys Gln Gly  Asn Phe Lys Ser Phe  Val Lys Asp
    1010                1015                1020

Arg Arg  Leu Asn Gly Leu Phe  Ser Phe Val Asp Thr  Gly Ala Leu
    1025                1030                1035

Ala Met  Glu Gln Tyr Pro Ile  Ser Lys Leu Arg Val  Glu Tyr Glu
    1040                1045                1050

Leu Ala  Lys Tyr Gln Thr Ala  Arg Val Cys Ala Phe  Glu Gln Thr
    1055                1060                1065

Leu Glu  Leu Glu Glu Ser Leu  Leu Thr Arg Tyr Pro  His Leu Pro
    1070                1075                1080

Asp Glu  Ser Phe Arg Glu Met  Leu Glu Ser Trp Ser  Asp Pro Leu
    1085                1090                1095
```

Leu Asp Lys Trp Pro Asp Leu Gln Arg Glu Val Arg Leu Leu Ile
    1100   1105   1110

Ala Val Arg Asn Ala Phe Ser His Asn Gln Tyr Pro Met Tyr Asp
    1115   1120   1125

Glu Thr Ile Phe Ser Ser Ile Arg Lys Tyr Asp Pro Ser Ser Leu
    1130   1135   1140

Asp Ala Ile Glu Glu Arg Met Gly Leu Asn Ile Ala His Arg Leu
    1145   1150   1155

Ser Glu Glu Val Lys Leu Ala Lys Glu Met Val Glu Arg Ile Ile
    1160   1165   1170

Gln Ala
    1175

<210> 110
<211> 1125
<212> PRT
<213> Prevotella pallens

<400> 110

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

Lys Asp Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His
            20   25   30

Asn Val Tyr Ile Thr Val Asn His Ile Asn Lys Ile Leu Gly Glu Gly
        35   40   45

Glu Ile Asn Arg Asp Gly Tyr Glu Asn Thr Leu Glu Lys Ser Trp Asn
    50   55   60

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

Lys His Phe Pro Phe Leu Glu Val Thr Thr Tyr Gln Arg Asn Ser Ala

```
                85                  90                  95

Asp Thr Thr Lys Gln Lys Glu Glu Lys Gln Ala Glu Ala Gln Ser Leu
            100                 105                 110

Glu Ser Leu Lys Lys Ser Phe Phe Val Phe Ile Tyr Lys Leu Arg Asp
        115                 120                 125

Leu Arg Asn His Tyr Ser His Tyr Lys His Ser Lys Ser Leu Glu Arg
    130                 135                 140

Pro Lys Phe Glu Glu Asp Leu Gln Glu Lys Met Tyr Asn Ile Phe Asp
145                 150                 155                 160

Ala Ser Ile Gln Leu Val Lys Glu Asp Tyr Lys His Asn Thr Asp Ile
                165                 170                 175

Lys Thr Glu Glu Asp Phe Lys His Leu Asp Arg Lys Gly Gln Phe Lys
            180                 185                 190

Tyr Ser Phe Ala Asp Asn Glu Gly Asn Ile Thr Glu Ser Gly Leu Leu
        195                 200                 205

Phe Phe Val Ser Leu Phe Leu Glu Lys Lys Asp Ala Ile Trp Val Gln
    210                 215                 220

Lys Lys Leu Glu Gly Phe Lys Cys Ser Asn Glu Ser Tyr Gln Lys Met
225                 230                 235                 240

Thr Asn Glu Val Phe Cys Arg Ser Arg Met Leu Leu Pro Lys Leu Arg
                245                 250                 255

Leu Gln Ser Thr Gln Thr Gln Asp Trp Ile Leu Leu Asp Met Leu Asn
            260                 265                 270

Glu Leu Ile Arg Cys Pro Lys Ser Leu Tyr Glu Arg Leu Arg Glu Glu
        275                 280                 285

Asp Arg Lys Lys Phe Arg Val Pro Ile Glu Ile Ala Asp Glu Asp Tyr
```

```
    290                 295                 300

Asp Ala Glu Gln Glu Pro Phe Lys Asn Ala Leu Val Arg His Gln Asp
305                 310                 315                 320

Arg Phe Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Tyr Asn Glu Ile Phe
                325                 330                 335

Thr Asn Leu Arg Phe Gln Ile Asp Leu Gly Thr Tyr His Phe Ser Ile
            340                 345                 350

Tyr Lys Lys Gln Ile Gly Asp Tyr Lys Glu Ser His His Leu Thr His
        355                 360                 365

Lys Leu Tyr Gly Phe Glu Arg Ile Gln Glu Phe Thr Lys Gln Asn Arg
    370                 375                 380

Pro Asp Glu Trp Arg Lys Phe Val Lys Thr Phe Asn Ser Phe Glu Thr
385                 390                 395                 400

Ser Lys Glu Pro Tyr Ile Pro Glu Thr Thr Pro His Tyr His Leu Glu
                405                 410                 415

Asn Gln Lys Ile Gly Ile Arg Phe Arg Asn Asp Asn Asp Lys Ile Trp
            420                 425                 430

Pro Ser Leu Lys Thr Asn Ser Glu Lys Asn Glu Lys Ser Lys Tyr Lys
        435                 440                 445

Leu Asp Lys Ser Phe Gln Ala Glu Ala Phe Leu Ser Val His Glu Leu
    450                 455                 460

Leu Pro Met Met Phe Tyr Tyr Leu Leu Leu Lys Thr Glu Asn Thr Asp
465                 470                 475                 480

Asn Asp Asn Glu Ile Glu Thr Lys Lys Lys Glu Asn Lys Asn Asp Lys
                485                 490                 495

Gln Glu Lys His Lys Ile Glu Glu Ile Ile Glu Asn Lys Ile Thr Glu
```

500 505 510

Ile Tyr Ala Leu Tyr Asp Ala Phe Ala Asn Gly Lys Ile Asn Ser Ile
515 520 525

Asp Lys Leu Glu Glu Tyr Cys Lys Gly Lys Asp Ile Glu Ile Gly His
530 535 540

Leu Pro Lys Gln Met Ile Ala Ile Leu Lys Ser Glu His Lys Asp Met
545 550 555 560

Ala Thr Glu Ala Lys Arg Lys Gln Glu Glu Met Leu Ala Asp Val Gln
565 570 575

Lys Ser Leu Glu Ser Leu Asp Asn Gln Ile Asn Glu Glu Ile Glu Asn
580 585 590

Val Glu Arg Lys Asn Ser Ser Leu Lys Ser Gly Glu Ile Ala Ser Trp
595 600 605

Leu Val Asn Asp Met Met Arg Phe Gln Pro Val Gln Lys Asp Asn Glu
610 615 620

Gly Asn Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Gln Met
625 630 635 640

Leu Gln Arg Ser Leu Ala Leu Tyr Asn Lys Glu Glu Lys Pro Thr Arg
645 650 655

Tyr Phe Arg Gln Val Asn Leu Ile Glu Ser Ser Asn Pro His Pro Phe
660 665 670

Leu Asn Asn Thr Glu Trp Glu Lys Cys Asn Asn Ile Leu Ser Phe Tyr
675 680 685

Arg Ser Tyr Leu Glu Ala Lys Lys Asn Phe Leu Glu Ser Leu Lys Pro
690 695 700

Glu Asp Trp Glu Lys Asn Gln Tyr Phe Leu Met Leu Lys Glu Pro Lys

```
705                 710                 715                 720

Thr Asn Cys Glu Thr Leu Val Gln Gly Trp Lys Asn Gly Phe Asn Leu
                725                 730                 735

Pro Arg Gly Ile Phe Thr Glu Pro Ile Arg Lys Trp Phe Met Glu His
            740                 745                 750

Arg Lys Asn Ile Thr Val Ala Glu Leu Lys Arg Val Gly Leu Val Ala
        755                 760                 765

Lys Val Ile Pro Leu Phe Phe Ser Glu Glu Tyr Lys Asp Ser Val Gln
    770                 775                 780

Pro Phe Tyr Asn Tyr Leu Phe Asn Val Gly Asn Ile Asn Lys Pro Asp
785                 790                 795                 800

Glu Lys Asn Phe Leu Asn Cys Glu Glu Arg Arg Glu Leu Leu Arg Lys
                805                 810                 815

Lys Lys Asp Glu Phe Lys Lys Met Thr Asp Lys Glu Lys Glu Glu Asn
            820                 825                 830

Pro Ser Tyr Leu Glu Phe Gln Ser Trp Asn Lys Phe Glu Arg Glu Leu
        835                 840                 845

Arg Leu Val Arg Asn Gln Asp Ile Val Thr Trp Leu Leu Cys Met Glu
    850                 855                 860

Leu Phe Asn Lys Lys Lys Ile Lys Glu Leu Asn Val Glu Lys Ile Tyr
865                 870                 875                 880

Leu Lys Asn Ile Asn Thr Asn Thr Thr Lys Lys Glu Lys Asn Thr Glu
                885                 890                 895

Glu Lys Asn Gly Glu Glu Lys Ile Ile Lys Glu Lys Asn Asn Ile Leu
            900                 905                 910

Asn Arg Ile Met Pro Met Arg Leu Pro Ile Lys Val Tyr Gly Arg Glu
```

```
        915                 920                 925

Asn Phe Ser Lys Asn Lys Lys Lys Lys Ile Arg Arg Asn Thr Phe Phe
    930                 935                 940

Thr Val Tyr Ile Glu Glu Lys Gly Thr Lys Leu Leu Lys Gln Gly Asn
945                 950                 955                 960

Phe Lys Ala Leu Glu Arg Asp Arg Arg Leu Gly Gly Leu Phe Ser Phe
                965                 970                 975

Val Lys Thr His Ser Lys Ala Glu Ser Lys Ser Asn Thr Ile Ser Lys
            980                 985                 990

Ser Arg Val Glu Tyr Glu Leu Gly  Glu Tyr Gln Lys Ala  Arg Ile Glu
        995                 1000                1005

Ile Ile  Lys Asp Met Leu Ala  Leu Glu Glu Thr Leu  Ile Asp Lys
    1010                1015                1020

Tyr Asn  Ser Leu Asp Thr Asp  Asn Phe His Asn Met  Leu Thr Gly
    1025                1030                1035

Trp Leu  Lys Leu Lys Asp Glu  Pro Asp Lys Ala Ser  Phe Gln Asn
    1040                1045                1050

Asp Val  Asp Leu Leu Ile Ala  Val Arg Asn Ala Phe  Ser His Asn
    1055                1060                1065

Gln Tyr  Pro Met Arg Asn Arg  Ile Ala Phe Ala Asn  Ile Asn Pro
    1070                1075                1080

Phe Ser  Leu Ser Ser Ala Asn  Thr Ser Glu Glu Lys  Gly Leu Gly
    1085                1090                1095

Ile Ala  Asn Gln Leu Lys Asp  Lys Thr His Lys Thr  Ile Glu Lys
    1100                1105                1110

Ile Ile  Glu Ile Glu Lys Pro  Ile Glu Thr Lys Glu
```

1115 1120 1125

<210> 111
<211> 1157
<212> PRT
<213> Myoides odoratimum

<400> 111

Met Lys Asp Ile Leu Thr Thr Asp Thr Thr Glu Lys Gln Asn Arg Phe
1 5 10 15

Tyr Ser His Lys Ile Ala Asp Lys Tyr Phe Phe Gly Gly Tyr Phe Asn
20 25 30

Leu Ala Ser Asn Asn Ile Tyr Glu Val Phe Glu Glu Val Asn Lys Arg
35 40 45

Asn Thr Phe Gly Lys Leu Ala Lys Arg Asp Asn Gly Asn Leu Lys Asn
50 55 60

Tyr Ile Ile His Val Phe Lys Asp Glu Leu Ser Ile Ser Asp Phe Glu
65 70 75 80

Lys Arg Val Ala Ile Phe Ala Ser Tyr Phe Pro Ile Leu Glu Thr Val
85 90 95

Asp Lys Lys Ser Ile Lys Glu Arg Asn Arg Thr Ile Asp Leu Thr Leu
100 105 110

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

Ala Val Asp Gln Leu Arg Asn Phe Tyr Thr His Tyr His His Ser Glu
130 135 140

Ile Val Ile Glu Asn Lys Val Leu Asp Phe Leu Asn Ser Ser Leu Val
145 150 155 160

Ser Thr Ala Leu His Val Lys Asp Lys Tyr Leu Lys Thr Asp Lys Thr
165 170 175

```
Lys Glu Phe Leu Lys Glu Thr Ile Ala Ala Glu Leu Asp Ile Leu Ile
            180                 185                 190

Glu Ala Tyr Lys Lys Lys Gln Ile Glu Lys Lys Asn Thr Arg Phe Lys
        195                 200                 205

Ala Asn Lys Arg Glu Asp Ile Leu Asn Ala Ile Tyr Asn Glu Ala Phe
    210                 215                 220

Trp Ser Phe Ile Asn Asp Lys Asp Lys Asp Lys Glu Thr Val Val Ala
225                 230                 235                 240

Lys Gly Ala Asp Ala Tyr Phe Glu Lys Asn His His Lys Ser Asn Asp
                245                 250                 255

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

Ser Phe Phe Leu Thr Asn Lys Glu Met Asp Ser Leu Lys Ala Asn Leu
        275                 280                 285

Thr Gly Phe Lys Gly Lys Val Asp Arg Glu Ser Gly Asn Ser Ile Lys
    290                 295                 300

Tyr Met Ala Thr Gln Arg Ile Tyr Ser Phe His Thr Tyr Arg Gly Leu
305                 310                 315                 320

Lys Gln Lys Ile Arg Thr Ser Glu Glu Gly Val Lys Glu Thr Leu Leu
                325                 330                 335

Met Gln Met Ile Asp Glu Leu Ser Lys Val Pro Asn Val Val Tyr Gln
            340                 345                 350

His Leu Ser Thr Thr Gln Gln Asn Ser Phe Ile Glu Asp Trp Asn Glu
        355                 360                 365

Tyr Tyr Lys Asp Tyr Glu Asp Asp Val Glu Thr Asp Asp Leu Ser Arg
    370                 375                 380
```

```
Val Ile His Pro Val Ile Arg Lys Arg Tyr Glu Asp Arg Phe Asn Tyr
385                 390                 395                 400

Phe Ala Ile Arg Phe Leu Asp Glu Phe Phe Asp Phe Pro Thr Leu Arg
                405                 410                 415

Phe Gln Val His Leu Gly Asp Tyr Val His Asp Arg Arg Thr Lys Gln
            420                 425                 430

Leu Gly Lys Val Glu Ser Asp Arg Ile Ile Lys Glu Lys Val Thr Val
        435                 440                 445

Phe Ala Arg Leu Lys Asp Ile Asn Ser Ala Lys Ala Asn Tyr Phe His
    450                 455                 460

Ser Leu Glu Glu Gln Asp Lys Glu Glu Leu Asp Asn Lys Trp Thr Leu
465                 470                 475                 480

Phe Pro Asn Pro Ser Tyr Asp Phe Pro Lys Glu His Thr Leu Gln His
                485                 490                 495

Gln Gly Glu Gln Lys Asn Ala Gly Lys Ile Gly Ile Tyr Val Lys Leu
            500                 505                 510

Arg Asp Thr Gln Tyr Lys Glu Lys Ala Ala Leu Glu Glu Ala Arg Lys
        515                 520                 525

Ser Leu Asn Pro Lys Glu Arg Ser Ala Thr Lys Ala Ser Lys Tyr Asp
    530                 535                 540

Ile Ile Thr Gln Ile Ile Glu Ala Asn Asp Asn Val Lys Ser Glu Lys
545                 550                 555                 560

Pro Leu Val Phe Thr Gly Gln Pro Ile Ala Tyr Leu Ser Met Asn Asp
                565                 570                 575

Ile His Ser Met Leu Phe Ser Leu Leu Thr Asp Asn Ala Glu Leu Lys
            580                 585                 590
```

```
Lys Thr Pro Glu Glu Val Glu Ala Lys Leu Ile Asp Gln Ile Gly Lys
        595                 600                 605

Gln Ile Asn Glu Ile Leu Ser Lys Asp Thr Asp Thr Lys Ile Leu Lys
    610                 615                 620

Lys Tyr Lys Asp Asn Asp Leu Lys Glu Thr Asp Thr Asp Lys Ile Thr
625                 630                 635                 640

Arg Asp Leu Ala Arg Asp Lys Glu Glu Ile Glu Lys Leu Ile Leu Glu
                645                 650                 655

Gln Lys Gln Arg Ala Asp Asp Tyr Asn Tyr Thr Ser Ser Thr Lys Phe
            660                 665                 670

Asn Ile Asp Lys Ser Arg Lys Arg Lys His Leu Leu Phe Asn Ala Glu
        675                 680                 685

Lys Gly Lys Ile Gly Val Trp Leu Ala Asn Asp Ile Lys Arg Phe Met
    690                 695                 700

Thr Glu Glu Phe Lys Ser Lys Trp Lys Gly Tyr Gln His Thr Glu Leu
705                 710                 715                 720

Gln Lys Leu Phe Ala Tyr Tyr Asp Thr Ser Lys Ser Asp Leu Asp Leu
                725                 730                 735

Ile Leu Ser Asp Met Val Met Val Lys Asp Tyr Pro Ile Glu Leu Ile
            740                 745                 750

Ala Leu Val Lys Lys Ser Arg Thr Leu Val Asp Phe Leu Asn Lys Tyr
        755                 760                 765

Leu Glu Ala Arg Leu Gly Tyr Met Glu Asn Val Ile Thr Arg Val Lys
    770                 775                 780

Asn Ser Ile Gly Thr Pro Gln Phe Lys Thr Val Arg Lys Glu Cys Phe
785                 790                 795                 800
```

```
Thr Phe Leu Lys Lys Ser Asn Tyr Thr Val Val Ser Leu Asp Lys Gln
                805                 810                 815

Val Glu Arg Ile Leu Ser Met Pro Leu Phe Ile Glu Arg Gly Phe Met
            820                 825                 830

Asp Asp Lys Pro Thr Met Leu Glu Gly Lys Ser Tyr Gln Gln His Lys
        835                 840                 845

Glu Lys Phe Ala Asp Trp Phe Val His Tyr Lys Glu Asn Ser Asn Tyr
    850                 855                 860

Gln Asn Phe Tyr Asp Thr Glu Val Tyr Glu Ile Thr Thr Glu Asp Lys
865                 870                 875                 880

Arg Glu Lys Ala Lys Val Thr Lys Lys Ile Lys Gln Gln Gln Lys Asn
                885                 890                 895

Asp Val Phe Thr Leu Met Met Val Asn Tyr Met Leu Glu Glu Val Leu
            900                 905                 910

Lys Leu Ser Ser Asn Asp Arg Leu Ser Leu Asn Glu Leu Tyr Gln Thr
        915                 920                 925

Lys Glu Glu Arg Ile Val Asn Lys Gln Val Ala Lys Asp Thr Gln Glu
    930                 935                 940

Arg Asn Lys Asn Tyr Ile Trp Asn Lys Val Val Asp Leu Gln Leu Cys
945                 950                 955                 960

Glu Gly Leu Val Arg Ile Asp Lys Val Lys Leu Lys Asp Ile Gly Asn
                965                 970                 975

Phe Arg Lys Tyr Glu Asn Asp Ser Arg Val Lys Glu Phe Leu Thr Tyr
            980                 985                 990

Gln Ser Asp Ile Val Trp Ser Ala  Tyr Leu Ser Asn Glu  Val Asp Ser
        995                 1000                1005
```

```
Asn Lys  Leu Tyr Val Ile Glu  Arg Gln Leu Asp Asn  Tyr Glu Ser
    1010                1015                1020

Ile Arg  Ser Lys Glu Leu Leu  Lys Glu Val Gln Glu  Ile Glu Cys
    1025                1030                1035

Ser Val  Tyr Asn Gln Val Ala  Asn Lys Glu Ser Leu  Lys Gln Ser
    1040                1045                1050

Gly Asn  Glu Asn Phe Lys Gln  Tyr Val Leu Gln Gly  Leu Val Pro
    1055                1060                1065

Ile Gly  Met Asp Val Arg Glu  Met Leu Ile Leu Ser  Thr Asp Val
    1070                1075                1080

Lys Phe  Ile Lys Glu Glu Ile  Ile Gln Leu Gly Gln  Ala Gly Glu
    1085                1090                1095

Val Glu  Gln Asp Leu Tyr Ser  Leu Ile Tyr Ile Arg  Asn Lys Phe
    1100                1105                1110

Ala His  Asn Gln Leu Pro Ile  Lys Glu Phe Phe Asp  Phe Cys Glu
    1115                1120                1125

Asn Asn  Tyr Arg Ser Ile Ser  Asp Asn Glu Tyr Tyr  Ala Glu Tyr
    1130                1135                1140

Tyr Met  Glu Ile Phe Arg Ser  Ile Lys Glu Lys Tyr  Thr Ser
    1145                1150                1155
```

<210> 112
<211> 1157
<212> PRT
<213> Myroides odoratimum

<400> 112

```
Met Lys Asp Ile Leu Thr Thr Asp Thr Thr Glu Lys Gln Asn Arg Phe
1               5                  10                  15
```

```
Tyr Ser His Lys Ile Ala Asp Lys Tyr Phe Phe Gly Gly Tyr Phe Asn
            20                  25                  30

Leu Ala Ser Asn Asn Ile Tyr Glu Val Phe Glu Glu Val Asn Lys Arg
        35                  40                  45

Asn Thr Phe Gly Lys Leu Ala Lys Arg Asp Asn Gly Asn Leu Lys Asn
    50                  55                  60

Tyr Ile Ile His Val Phe Lys Asp Glu Leu Ser Ile Ser Asp Phe Glu
65                  70                  75                  80

Lys Arg Val Ala Ile Phe Ala Ser Tyr Phe Pro Ile Leu Glu Thr Val
                85                  90                  95

Asp Lys Lys Ser Ile Lys Glu Arg Asn Arg Thr Ile Asp Leu Thr Leu
            100                 105                 110

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

Ala Val Asp Gln Leu Arg Asn Phe Tyr Thr His Tyr His His Ser Glu
    130                 135                 140

Ile Val Ile Glu Asn Lys Val Leu Asp Phe Leu Asn Ser Ser Leu Val
145                 150                 155                 160

Ser Thr Ala Leu His Val Lys Asp Lys Tyr Leu Lys Thr Asp Lys Thr
                165                 170                 175

Lys Glu Phe Leu Lys Glu Thr Ile Ala Ala Glu Leu Asp Ile Leu Ile
            180                 185                 190

Glu Ala Tyr Lys Lys Lys Gln Ile Glu Lys Lys Asn Thr Arg Phe Lys
        195                 200                 205

Ala Asn Lys Arg Glu Asp Ile Leu Asn Ala Ile Tyr Asn Glu Ala Phe
    210                 215                 220
```

```
Trp Ser Phe Ile Asn Asp Lys Asp Lys Asp Lys Glu Thr Val Val Ala
225                 230                 235                 240


Lys Gly Ala Asp Ala Tyr Phe Glu Lys Asn His His Lys Ser Asn Asp
                245                 250                 255


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


Ser Phe Phe Leu Thr Asn Lys Glu Met Asp Ser Leu Lys Ala Asn Leu
        275                 280                 285


Thr Gly Phe Lys Gly Lys Val Asp Arg Glu Ser Gly Asn Ser Ile Lys
    290                 295                 300


Tyr Met Ala Thr Gln Arg Ile Tyr Ser Phe His Thr Tyr Arg Gly Leu
305                 310                 315                 320


Lys Gln Lys Ile Arg Thr Ser Glu Glu Gly Val Lys Glu Thr Leu Leu
                325                 330                 335


Met Gln Met Ile Asp Glu Leu Ser Lys Val Pro Asn Val Val Tyr Gln
            340                 345                 350


His Leu Ser Thr Thr Gln Gln Asn Ser Phe Ile Glu Asp Trp Asn Glu
        355                 360                 365


Tyr Tyr Lys Asp Tyr Glu Asp Asp Val Glu Thr Asp Asp Leu Ser Arg
    370                 375                 380


Val Ile His Pro Val Ile Arg Lys Arg Tyr Glu Asp Arg Phe Asn Tyr
385                 390                 395                 400


Phe Ala Ile Arg Phe Leu Asp Glu Phe Phe Asp Phe Pro Thr Leu Arg
                405                 410                 415


Phe Gln Val His Leu Gly Asp Tyr Val His Asp Arg Arg Thr Lys Gln
            420                 425                 430
```

```
Leu Gly Lys Val Glu Ser Asp Arg Ile Ile Lys Glu Lys Val Thr Val
        435                 440                 445

Phe Ala Arg Leu Lys Asp Ile Asn Ser Ala Lys Ala Ser Tyr Phe His
    450                 455                 460

Ser Leu Glu Glu Gln Asp Lys Glu Glu Leu Asp Asn Lys Trp Thr Leu
465                 470                 475                 480

Phe Pro Asn Pro Ser Tyr Asp Phe Pro Lys Glu His Thr Leu Gln His
                485                 490                 495

Gln Gly Glu Gln Lys Asn Ala Gly Lys Ile Gly Ile Tyr Val Lys Leu
            500                 505                 510

Arg Asp Thr Gln Tyr Lys Glu Lys Ala Ala Leu Glu Glu Ala Arg Lys
        515                 520                 525

Ser Leu Asn Pro Lys Glu Arg Ser Ala Thr Lys Ala Ser Lys Tyr Asp
    530                 535                 540

Ile Ile Thr Gln Ile Ile Glu Ala Asn Asp Asn Val Lys Ser Glu Lys
545                 550                 555                 560

Pro Leu Val Phe Thr Gly Gln Pro Ile Ala Tyr Leu Ser Met Asn Asp
                565                 570                 575

Ile His Ser Met Leu Phe Ser Leu Leu Thr Asp Asn Ala Glu Leu Lys
            580                 585                 590

Lys Thr Pro Glu Glu Val Glu Ala Lys Leu Ile Asp Gln Ile Gly Lys
        595                 600                 605

Gln Ile Asn Glu Ile Leu Ser Lys Asp Thr Asp Thr Lys Ile Leu Lys
    610                 615                 620

Lys Tyr Lys Asp Asn Asp Leu Lys Glu Thr Asp Thr Asp Lys Ile Thr
625                 630                 635                 640
```

```
Arg Asp Leu Ala Arg Asp Lys Glu Glu Ile Glu Lys Leu Ile Leu Glu
                645                 650                 655

Gln Lys Gln Arg Ala Asp Asp Tyr Asn Tyr Thr Ser Ser Thr Lys Phe
            660                 665                 670

Asn Ile Asp Lys Ser Arg Lys Arg Lys His Leu Leu Phe Asn Ala Glu
        675                 680                 685

Lys Gly Lys Ile Gly Val Trp Leu Ala Asn Asp Ile Lys Arg Phe Met
    690                 695                 700

Phe Lys Glu Ser Lys Ser Lys Trp Lys Gly Tyr Gln His Thr Glu Leu
705                 710                 715                 720

Gln Lys Leu Phe Ala Tyr Phe Asp Thr Ser Lys Ser Asp Leu Glu Leu
                725                 730                 735

Ile Leu Ser Asp Met Val Met Val Lys Asp Tyr Pro Ile Glu Leu Ile
            740                 745                 750

Asp Leu Val Arg Lys Ser Arg Thr Leu Val Asp Phe Leu Asn Lys Tyr
        755                 760                 765

Leu Glu Ala Arg Leu Gly Tyr Ile Glu Asn Val Ile Thr Arg Val Lys
    770                 775                 780

Asn Ser Ile Gly Thr Pro Gln Phe Lys Thr Val Arg Lys Glu Cys Phe
785                 790                 795                 800

Ala Phe Leu Lys Glu Ser Asn Tyr Thr Val Ala Ser Leu Asp Lys Gln
                805                 810                 815

Ile Glu Arg Ile Leu Ser Met Pro Leu Phe Ile Glu Arg Gly Phe Met
            820                 825                 830

Asp Ser Lys Pro Thr Met Leu Glu Gly Lys Ser Tyr Gln Gln His Lys
        835                 840                 845
```

```
Glu Asp Phe Ala Asp Trp Phe Val His Tyr Lys Glu Asn Ser Asn Tyr
    850                 855                 860

Gln Asn Phe Tyr Asp Thr Glu Val Tyr Glu Ile Ile Thr Glu Asp Lys
865                 870                 875                 880

Arg Glu Gln Ala Lys Val Thr Lys Lys Ile Lys Gln Gln Gln Lys Asn
                885                 890                 895

Asp Val Phe Thr Leu Met Met Val Asn Tyr Met Leu Glu Glu Val Leu
            900                 905                 910

Lys Leu Pro Ser Asn Asp Arg Leu Ser Leu Asn Glu Leu Tyr Gln Thr
        915                 920                 925

Lys Glu Glu Arg Ile Val Asn Lys Gln Val Ala Lys Asp Thr Gln Glu
    930                 935                 940

Arg Asn Lys Asn Tyr Ile Trp Asn Lys Val Val Asp Leu Gln Leu Cys
945                 950                 955                 960

Glu Gly Leu Val Arg Ile Asp Lys Val Lys Leu Lys Asp Ile Gly Asn
                965                 970                 975

Phe Arg Lys Tyr Glu Asn Asp Ser Arg Val Lys Glu Phe Leu Thr Tyr
            980                 985                 990

Gln Ser Asp Ile Val Trp Ser Gly  Tyr Leu Ser Asn Glu  Val Asp Ser
        995                 1000                1005

Asn Lys  Leu Tyr Val Ile Glu  Arg Gln Leu Asp Asn  Tyr Glu Ser
    1010                1015                1020

Ile Arg  Ser Lys Glu Leu Leu  Lys Glu Val Gln Glu  Ile Glu Cys
    1025                1030                1035

Ile Val  Tyr Asn Gln Val Ala  Asn Lys Glu Ser Leu  Lys Gln Ser
    1040                1045                1050
```

```
Gly Asn  Glu Asn Phe Lys Gln  Tyr Val Leu Gln Gly  Leu Leu Pro
    1055                 1060                 1065


Arg Gly  Thr Asp Val Arg Glu  Met Leu Ile Leu Ser  Thr Asp Val
    1070                 1075                 1080


Lys Phe  Lys Lys Glu Glu Ile  Met Gln Leu Gly Gln  Val Arg Glu
    1085                 1090                 1095


Val Glu  Gln Asp Leu Tyr Ser  Leu Ile Tyr Ile Arg  Asn Lys Phe
    1100                 1105                 1110


Ala His  Asn Gln Leu Pro Ile  Lys Glu Phe Phe Asp  Phe Cys Glu
    1115                 1120                 1125


Asn Asn  Tyr Arg Pro Ile Ser  Asp Asn Glu Tyr Tyr  Ala Glu Tyr
    1130                 1135                 1140


Tyr Met  Glu Ile Phe Arg Ser  Ile Lys Glu Lys Tyr  Ala Ser
    1145                 1150                 1155


<210>  113
<211>  1127
<212>  PRT
<213>  Prevotella sp.

<400>  113

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


Phe Ala Phe Pro Lys Tyr Ile Thr Ile Met Glu Asn Gln Glu Lys Pro
            20                  25                  30


Glu Pro Ile Tyr Tyr Glu Leu Thr Asp Lys His Phe Trp Ala Ala Phe
        35                  40                  45


Leu Asn Leu Ala Arg His Asn Val Tyr Thr Thr Ile Asn His Ile Asn
    50                  55                  60
```

```
Arg Arg Leu Glu Ile Ala Glu Leu Lys Asp Asp Gly Tyr Met Met Gly
65                  70                  75                  80


Ile Lys Gly Ser Trp Asn Glu Gln Ala Lys Lys Leu Asp Lys Lys Val
                85                  90                  95


Arg Leu Arg Asp Leu Ile Met Lys His Phe Pro Phe Leu Glu Ala Ala
            100                 105                 110


Ala Tyr Glu Ile Thr Asn Ser Lys Ser Pro Asn Asn Lys Glu Gln Arg
        115                 120                 125


Glu Lys Glu Gln Ser Glu Ala Leu Ser Leu Asn Asn Leu Lys Asn Val
    130                 135                 140


Leu Phe Ile Phe Leu Glu Lys Leu Gln Val Leu Arg Asn Tyr Tyr Ser
145                 150                 155                 160


His Tyr Lys Tyr Ser Glu Glu Ser Pro Lys Pro Ile Phe Glu Thr Ser
                165                 170                 175


Leu Leu Lys Asn Met Tyr Lys Val Phe Asp Ala Asn Val Arg Leu Val
            180                 185                 190


Lys Arg Asp Tyr Met His His Glu Asn Ile Asp Met Gln Arg Asp Phe
        195                 200                 205


Thr His Leu Asn Arg Lys Lys Gln Val Gly Arg Thr Lys Asn Ile Ile
    210                 215                 220


Asp Ser Pro Asn Phe His Tyr His Phe Ala Asp Lys Glu Gly Asn Met
225                 230                 235                 240


Thr Ile Ala Gly Leu Leu Phe Phe Val Ser Leu Phe Leu Asp Lys Lys
                245                 250                 255


Asp Ala Ile Trp Met Gln Lys Lys Leu Lys Gly Phe Lys Asp Gly Arg
            260                 265                 270
```

```
Asn Leu Arg Glu Gln Met Thr Asn Glu Val Phe Cys Arg Ser Arg Ile
        275                 280                 285

Ser Leu Pro Lys Leu Lys Leu Glu Asn Val Gln Thr Lys Asp Trp Met
    290                 295                 300

Gln Leu Asp Met Leu Asn Glu Leu Val Arg Cys Pro Lys Ser Leu Tyr
305                 310                 315                 320

Glu Arg Leu Arg Glu Lys Asp Arg Glu Ser Phe Lys Val Pro Phe Asp
                325                 330                 335

Ile Phe Ser Asp Asp Tyr Asp Ala Glu Glu Glu Pro Phe Lys Asn Thr
            340                 345                 350

Leu Val Arg His Gln Asp Arg Phe Pro Tyr Phe Val Leu Arg Tyr Phe
        355                 360                 365

Asp Leu Asn Glu Ile Phe Glu Gln Leu Arg Phe Gln Ile Asp Leu Gly
    370                 375                 380

Thr Tyr His Phe Ser Ile Tyr Asn Lys Arg Ile Gly Asp Glu Asp Glu
385                 390                 395                 400

Val Arg His Leu Thr His His Leu Tyr Gly Phe Ala Arg Ile Gln Asp
                405                 410                 415

Phe Ala Pro Gln Asn Gln Pro Glu Glu Trp Arg Lys Leu Val Lys Asp
            420                 425                 430

Leu Asp His Phe Glu Thr Ser Gln Glu Pro Tyr Ile Ser Lys Thr Ala
        435                 440                 445

Pro His Tyr His Leu Glu Asn Glu Lys Ile Gly Ile Lys Phe Cys Ser
    450                 455                 460

Thr His Asn Asn Leu Phe Pro Ser Leu Lys Arg Glu Lys Thr Cys Asn
465                 470                 475                 480
```

```
Gly Arg Ser Lys Phe Asn Leu Gly Thr Gln Phe Thr Ala Glu Ala Phe
                485                 490                 495

Leu Ser Val His Glu Leu Leu Pro Met Met Phe Tyr Tyr Leu Leu Leu
            500                 505                 510

Thr Lys Asp Tyr Ser Arg Lys Glu Ser Ala Asp Lys Val Glu Gly Ile
        515                 520                 525

Ile Arg Lys Glu Ile Ser Asn Ile Tyr Ala Ile Tyr Asp Ala Phe Ala
    530                 535                 540

Asn Asn Glu Ile Asn Ser Ile Ala Asp Leu Thr Cys Arg Leu Gln Lys
545                 550                 555                 560

Thr Asn Ile Leu Gln Gly His Leu Pro Lys Gln Met Ile Ser Ile Leu
                565                 570                 575

Glu Gly Arg Gln Lys Asp Met Glu Lys Glu Ala Glu Arg Lys Ile Gly
            580                 585                 590

Glu Met Ile Asp Asp Thr Gln Arg Arg Leu Asp Leu Leu Cys Lys Gln
        595                 600                 605

Thr Asn Gln Lys Ile Arg Ile Gly Lys Arg Asn Ala Gly Leu Leu Lys
    610                 615                 620

Ser Gly Lys Ile Ala Asp Trp Leu Val Ser Asp Met Met Arg Phe Gln
625                 630                 635                 640

Pro Val Gln Lys Asp Thr Asn Asn Ala Pro Ile Asn Asn Ser Lys Ala
                645                 650                 655

Asn Ser Thr Glu Tyr Arg Met Leu Gln His Ala Leu Ala Leu Phe Gly
            660                 665                 670

Ser Glu Ser Ser Arg Leu Lys Ala Tyr Phe Arg Gln Met Asn Leu Val
        675                 680                 685
```

```
Gly Asn Ala Asn Pro His Pro Phe Leu Ala Glu Thr Gln Trp Glu His
    690                 695                 700

Gln Thr Asn Ile Leu Ser Phe Tyr Arg Asn Tyr Leu Glu Ala Arg Lys
705                 710                 715                 720

Lys Tyr Leu Lys Gly Leu Lys Pro Gln Asn Trp Lys Gln Tyr Gln His
                725                 730                 735

Phe Leu Ile Leu Lys Val Gln Lys Thr Asn Arg Asn Thr Leu Val Thr
            740                 745                 750

Gly Trp Lys Asn Ser Phe Asn Leu Pro Arg Gly Ile Phe Thr Gln Pro
        755                 760                 765

Ile Arg Glu Trp Phe Glu Lys His Asn Asn Ser Lys Arg Ile Tyr Asp
    770                 775                 780

Gln Ile Leu Ser Phe Asp Arg Val Gly Phe Val Ala Lys Ala Ile Pro
785                 790                 795                 800

Leu Tyr Phe Ala Glu Glu Tyr Lys Asp Asn Val Gln Pro Phe Tyr Asp
                805                 810                 815

Tyr Pro Phe Asn Ile Gly Asn Lys Leu Lys Pro Gln Lys Gly Gln Phe
            820                 825                 830

Leu Asp Lys Lys Glu Arg Val Glu Leu Trp Gln Lys Asn Lys Glu Leu
        835                 840                 845

Phe Lys Asn Tyr Pro Ser Glu Lys Asn Lys Thr Asp Leu Ala Tyr Leu
    850                 855                 860

Asp Phe Leu Ser Trp Lys Lys Phe Glu Arg Glu Leu Arg Leu Ile Lys
865                 870                 875                 880

Asn Gln Asp Ile Val Thr Trp Leu Met Phe Lys Glu Leu Phe Lys Thr
                885                 890                 895
```

```
Thr Thr Val Glu Gly Leu Lys Ile Gly Glu Ile His Leu Arg Asp Ile
            900                 905                 910


Asp Thr Asn Thr Ala Asn Glu Glu Ser Asn Asn Ile Leu Asn Arg Ile
        915                 920                 925


Met Pro Met Lys Leu Pro Val Lys Thr Tyr Glu Thr Asp Asn Lys Gly
    930                 935                 940


Asn Ile Leu Lys Glu Arg Pro Leu Ala Thr Phe Tyr Ile Glu Glu Thr
945                 950                 955                 960


Glu Thr Lys Val Leu Lys Gln Gly Asn Phe Lys Val Leu Ala Lys Asp
                965                 970                 975


Arg Arg Leu Asn Gly Leu Leu Ser Phe Ala Glu Thr Thr Asp Ile Asp
            980                 985                 990


Leu Glu Lys Asn Pro Ile Thr Lys  Leu Ser Val Asp Tyr  Glu Leu Ile
        995                 1000                1005


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


Leu Glu  Lys Lys Leu Ile Asp  Lys Tyr Ser Thr Leu  Pro Thr Asp
    1025                1030                1035


Ser Phe  Arg Asn Met Leu Glu  Arg Trp Leu Gln Cys  Lys Ala Asn
    1040                1045                1050


Arg Pro  Glu Leu Lys Asn Tyr  Val Asn Ser Leu Ile  Ala Val Arg
    1055                1060                1065


Asn Ala  Phe Ser His Asn Gln  Tyr Pro Met Tyr Asp  Ala Thr Leu
    1070                1075                1080


Phe Ala  Glu Val Lys Lys Phe  Thr Leu Phe Pro Ser  Val Asp Thr
    1085                1090                1095
```

```
Lys Lys  Ile Glu Leu Asn Ile  Ala Pro Gln Leu Leu  Glu Ile Val
    1100                1105                1110


Gly Lys  Ala Ile Lys Glu Ile  Glu Lys Ser Glu Asn  Lys Asn
    1115                1120                1125


<210>  114
<211>  1119
<212>  PRT
<213>  Porphyromonas gingivalis

<400>  114

Met Thr Glu Gln Asn Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1               5                   10                  15


Glu Asp Lys His Phe Trp Ala Ala Phe Phe Asn Leu Ala Arg His Asn
            20                  25                  30


Ala Tyr Ile Thr Leu Ala His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
        35                  40                  45


Ala Asp Ile Thr Asn Asp Glu Asp Ile Leu Phe Phe Lys Gly Gln Trp
    50                  55                  60


Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ala Arg Leu Arg Ser Leu
65                  70                  75                  80


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


Leu Phe Glu Ser Gln Ser Ser Gly Asn Lys Ser Ser Lys Lys Lys Glu
            100                 105                 110


Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
        115                 120                 125


Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
    130                 135                 140


Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Pro Glu Ser Ser Glu Leu
```

```
145                 150                 155                 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
                165                 170                 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val
            180                 185                 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
        195                 200                 205

Arg Tyr Gly Asn Asn Asp Asn Pro Phe Phe Lys His His Phe Val Asp
    210                 215                 220

Arg Glu Glu Lys Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225                 230                 235                 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
                245                 250                 255

Phe Lys Gly Gly Thr Glu Thr Tyr Gln Gln Met Thr Asn Glu Val Phe
            260                 265                 270

Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
        275                 280                 285

Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys
    290                 295                 300

Pro Lys Ser Leu Tyr Asp Arg Leu Arg Glu Glu Asp Arg Ala Arg Phe
305                 310                 315                 320

Arg Val Pro Val Asp Ile Leu Ser Asp Glu Asp Asp Thr Asp Gly Thr
                325                 330                 335

Glu Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
            340                 345                 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
```

```
        355                 360                 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
    370                 375                 380

Lys Asn Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385                 390                 395                 400

Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
                405                 410                 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
            420                 425                 430

Lys Pro Tyr Ile Thr Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly
        435                 440                 445

Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln His Leu Trp Pro Ser
    450                 455                 460

Pro Glu Val Gly Ala Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465                 470                 475                 480

Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
                485                 490                 495

Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Asp Glu Ala
            500                 505                 510

Ser Ala Glu Arg Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
        515                 520                 525

Tyr Ala Val Tyr Asp Ala Phe Ala Arg Gly Glu Ile Asn Thr Arg Asp
    530                 535                 540

Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu
545                 550                 555                 560

Pro Arg Gln Met Ile Gly Ile Leu Ser Gln Glu His Lys Asp Met Glu
```

```
                565                 570                 575

Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Val Asp Thr Asp His
            580                 585                 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly
        595                 600                 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
    610                 615                 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly
625                 630                 635                 640

Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
                645                 650                 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro
            660                 665                 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
        675                 680                 685

Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
    690                 695                 700

Arg Ser Tyr Leu Lys Ala Arg Lys Ala Phe Leu Gln Ser Ile Gly Arg
705                 710                 715                 720

Ser Asp Arg Val Glu Asn His Arg Phe Leu Leu Leu Lys Glu Pro Lys
                725                 730                 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Gly Glu Phe His Leu
            740                 745                 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
        755                 760                 765

Gly Leu Asp Glu Val Gly Ser Tyr Lys Glu Val Gly Phe Met Ala Lys
```

```
    770                 775                 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Lys Asp Arg Val Gln Pro
785                 790                 795                 800

Phe Tyr Asp Tyr Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
                805                 810                 815

Gly Arg Phe Leu Ser Lys Glu Lys Arg Ala Glu Glu Trp Glu Ser Gly
            820                 825                 830

Lys Glu Arg Phe Arg Leu Ala Lys Leu Lys Lys Glu Ile Leu Glu Ala
        835                 840                 845

Lys Glu His Pro Tyr Leu Asp Phe Lys Ser Trp Gln Lys Phe Glu Arg
    850                 855                 860

Glu Leu Arg Leu Val Lys Asn Gln Asp Ile Ile Thr Trp Met Ile Cys
865                 870                 875                 880

Arg Asp Leu Met Glu Glu Asn Lys Val Glu Gly Leu Asp Thr Gly Thr
                885                 890                 895

Leu Tyr Leu Lys Asp Ile Arg Thr Asp Val Gln Glu Gln Gly Asn Leu
            900                 905                 910

Asn Val Leu Asn Arg Val Lys Pro Met Arg Leu Pro Val Val Val Tyr
        915                 920                 925

Arg Ala Asp Ser Arg Gly His Val His Lys Glu Gln Ala Pro Leu Ala
    930                 935                 940

Thr Val Tyr Ile Glu Glu Arg Asp Thr Lys Leu Leu Lys Gln Gly Asn
945                 950                 955                 960

Phe Lys Ser Phe Val Lys Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe
                965                 970                 975

Val Asp Thr Gly Ala Leu Ala Met Glu Gln Tyr Pro Ile Ser Lys Leu
```

980 985 990

Arg Val Glu Tyr Glu Leu Ala Lys Tyr Gln Thr Ala Arg Val Cys Ala
995 1000 1005

Phe Glu Gln Thr Leu Glu Leu Glu Glu Ser Leu Leu Thr Arg Tyr
1010 1015 1020

Pro His Leu Pro Asp Lys Asn Phe Arg Lys Met Leu Glu Ser Trp
1025 1030 1035

Ser Asp Pro Leu Leu Asp Lys Trp Pro Asp Leu His Gly Asn Val
1040 1045 1050

Arg Leu Leu Ile Ala Val Arg Asn Ala Phe Ser His Asn Gln Tyr
1055 1060 1065

Pro Met Tyr Asp Glu Ala Val Phe Ser Ser Ile Arg Lys Tyr Asp
1070 1075 1080

Pro Ser Ser Pro Asp Ala Ile Glu Glu Arg Met Gly Leu Asn Ile
1085 1090 1095

Ala His Arg Leu Ser Glu Glu Val Lys Gln Ala Lys Glu Met Ala
1100 1105 1110

Glu Arg Ile Ile Gln Ala
1115

<210> 115
<211> 1154
<212> PRT
<213> Paludibacter propionicigenes

<400> 115

Met Lys Thr Ser Ala Asn Asn Ile Tyr Phe Asn Gly Ile Asn Ser Phe
1 5 10 15

Lys Lys Ile Phe Asp Ser Lys Gly Ala Ile Ala Pro Ile Ala Glu Lys
20 25 30

```
Ser Cys Arg Asn Phe Asp Ile Lys Ala Gln Asn Asp Val Asn Lys Glu
        35                  40                  45

Gln Arg Ile His Tyr Phe Ala Val Gly His Thr Phe Lys Gln Leu Asp
    50                  55                  60

Thr Glu Asn Leu Phe Glu Tyr Val Leu Asp Glu Asn Leu Arg Ala Lys
65                  70                  75                  80

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

Glu Asn Ile Lys Arg Leu Ile Ser Asp Ile Arg Asn Ile Asn Ser His
            100                 105                 110

Tyr Ile His Arg Phe Asp Pro Leu Lys Ile Asp Ala Val Pro Thr Asn
        115                 120                 125

Ile Ile Asp Phe Leu Lys Glu Ser Phe Glu Leu Ala Val Ile Gln Ile
    130                 135                 140

Tyr Leu Lys Glu Lys Gly Ile Asn Tyr Leu Gln Phe Ser Glu Asn Pro
145                 150                 155                 160

His Ala Asp Gln Lys Leu Val Ala Phe Leu His Asp Lys Phe Leu Pro
                165                 170                 175

Leu Asp Glu Lys Lys Thr Ser Met Leu Gln Asn Glu Thr Pro Gln Leu
            180                 185                 190

Lys Glu Tyr Lys Glu Tyr Arg Lys Tyr Phe Lys Thr Leu Ser Lys Gln
        195                 200                 205

Ala Ala Ile Asp Gln Leu Leu Phe Ala Glu Lys Glu Thr Asp Tyr Ile
    210                 215                 220

Trp Asn Leu Phe Asp Ser His Pro Val Leu Thr Ile Ser Ala Gly Lys
225                 230                 235                 240
```

```
Tyr Leu Ser Phe Tyr Ser Cys Leu Phe Leu Leu Ser Met Phe Leu Tyr
                245                 250                 255

Lys Ser Glu Ala Asn Gln Leu Ile Ser Lys Ile Lys Gly Phe Lys Lys
            260                 265                 270

Asn Thr Thr Glu Glu Glu Lys Ser Lys Arg Glu Ile Phe Thr Phe Phe
        275                 280                 285

Ser Lys Arg Phe Asn Ser Met Asp Ile Asp Ser Glu Glu Asn Gln Leu
    290                 295                 300

Val Lys Phe Arg Asp Leu Ile Leu Tyr Leu Asn His Tyr Pro Val Ala
305                 310                 315                 320

Trp Asn Lys Asp Leu Glu Leu Asp Ser Ser Asn Pro Ala Met Thr Asp
                325                 330                 335

Lys Leu Lys Ser Lys Ile Ile Glu Leu Glu Ile Asn Arg Ser Phe Pro
            340                 345                 350

Leu Tyr Glu Gly Asn Glu Arg Phe Ala Thr Phe Ala Lys Tyr Gln Ile
        355                 360                 365

Trp Gly Lys Lys His Leu Gly Lys Ser Ile Glu Lys Glu Tyr Ile Asn
    370                 375                 380

Ala Ser Phe Thr Asp Glu Glu Ile Thr Ala Tyr Thr Tyr Glu Thr Asp
385                 390                 395                 400

Thr Cys Pro Glu Leu Lys Asp Ala His Lys Lys Leu Ala Asp Leu Lys
                405                 410                 415

Ala Ala Lys Gly Leu Phe Gly Lys Arg Lys Glu Lys Asn Glu Ser Asp
            420                 425                 430

Ile Lys Lys Thr Glu Thr Ser Ile Arg Glu Leu Gln His Glu Pro Asn
        435                 440                 445
```

```
Pro Ile Lys Asp Lys Leu Ile Gln Arg Ile Glu Lys Asn Leu Leu Thr
    450                 455                 460

Val Ser Tyr Gly Arg Asn Gln Asp Arg Phe Met Asp Phe Ser Ala Arg
465                 470                 475                 480

Phe Leu Ala Glu Ile Asn Tyr Phe Gly Gln Asp Ala Ser Phe Lys Met
                485                 490                 495

Tyr His Phe Tyr Ala Thr Asp Glu Gln Asn Ser Glu Leu Glu Lys Tyr
            500                 505                 510

Glu Leu Pro Lys Asp Lys Lys Lys Tyr Asp Ser Leu Lys Phe His Gln
        515                 520                 525

Gly Lys Leu Val His Phe Ile Ser Tyr Lys Glu His Leu Lys Arg Tyr
    530                 535                 540

Glu Ser Trp Asp Asp Ala Phe Val Ile Glu Asn Asn Ala Ile Gln Leu
545                 550                 555                 560

Lys Leu Ser Phe Asp Gly Val Glu Asn Thr Val Thr Ile Gln Arg Ala
                565                 570                 575

Leu Leu Ile Tyr Leu Leu Glu Asp Ala Leu Arg Asn Ile Gln Asn Asn
            580                 585                 590

Thr Ala Glu Asn Ala Gly Lys Gln Leu Leu Gln Glu Tyr Tyr Ser His
        595                 600                 605

Asn Lys Ala Asp Leu Ser Ala Phe Lys Gln Ile Leu Thr Gln Gln Asp
    610                 615                 620

Ser Ile Glu Pro Gln Gln Lys Thr Glu Phe Lys Lys Leu Leu Pro Arg
625                 630                 635                 640

Arg Leu Leu Asn Asn Tyr Ser Pro Ala Ile Asn His Leu Gln Thr Pro
                645                 650                 655
```

```
His Ser Ser Leu Pro Leu Ile Leu Glu Lys Ala Leu Leu Ala Glu Lys
            660                 665                 670

Arg Tyr Cys Ser Leu Val Val Lys Ala Lys Ala Glu Gly Asn Tyr Asp
        675                 680                 685

Asp Phe Ile Lys Arg Asn Lys Gly Lys Gln Phe Lys Leu Gln Phe Ile
    690                 695                 700

Arg Lys Ala Trp Asn Leu Met Tyr Phe Arg Asn Ser Tyr Leu Gln Asn
705                 710                 715                 720

Val Gln Ala Ala Gly His His Lys Ser Phe His Ile Glu Arg Asp Glu
                725                 730                 735

Phe Asn Asp Phe Ser Arg Tyr Met Phe Ala Phe Glu Glu Leu Ser Gln
            740                 745                 750

Tyr Lys Tyr Tyr Leu Asn Glu Met Phe Glu Lys Lys Gly Phe Phe Glu
        755                 760                 765

Asn Asn Glu Phe Lys Ile Leu Phe Gln Ser Gly Thr Ser Leu Glu Asn
    770                 775                 780

Leu Tyr Glu Lys Thr Lys Gln Lys Phe Glu Ile Trp Leu Ala Ser Asn
785                 790                 795                 800

Thr Ala Lys Thr Asn Lys Pro Asp Asn Tyr His Leu Asn Asn Tyr Glu
                805                 810                 815

Gln Gln Phe Ser Asn Gln Leu Phe Phe Ile Asn Leu Ser His Phe Ile
            820                 825                 830

Asn Tyr Leu Lys Ser Thr Gly Lys Leu Gln Thr Asp Ala Asn Gly Gln
        835                 840                 845

Ile Ile Tyr Glu Ala Leu Asn Asn Val Gln Tyr Leu Ile Pro Glu Tyr
    850                 855                 860
```

```
Tyr Tyr Thr Asp Lys Pro Glu Arg Ser Glu Ser Lys Ser Gly Asn Lys
865                 870                 875                 880


Leu Tyr Asn Lys Leu Lys Ala Thr Lys Leu Glu Asp Ala Leu Leu Tyr
                885                 890                 895


Glu Met Ala Met Cys Tyr Leu Lys Ala Asp Lys Gln Ile Ala Asp Lys
            900                 905                 910


Ala Lys His Pro Ile Thr Lys Leu Leu Thr Ser Asp Val Glu Phe Asn
        915                 920                 925


Ile Thr Asn Lys Glu Gly Ile Gln Leu Tyr His Leu Leu Val Pro Phe
    930                 935                 940


Lys Lys Ile Asp Ala Phe Ile Gly Leu Lys Met His Lys Glu Gln Gln
945                 950                 955                 960


Asp Lys Lys His Pro Thr Ser Phe Leu Ala Asn Ile Val Asn Tyr Leu
                965                 970                 975


Glu Leu Val Lys Asn Asp Lys Asp Ile Arg Lys Thr Tyr Glu Ala Phe
            980                 985                 990


Ser Thr Asn Pro Val Lys Arg Thr  Leu Thr Tyr Asp Asp  Leu Ala Lys
        995                 1000                1005


Ile Asp  Gly His Leu Ile Ser  Lys Ser Ile Lys Phe  Thr Asn Val
    1010                1015                1020


Thr Leu  Glu Leu Glu Arg Tyr  Phe Ile Phe Lys Glu  Ser Leu Ile
    1025                1030                1035


Val Lys  Lys Gly Asn Asn Ile  Asp Phe Lys Tyr Ile  Lys Gly Leu
    1040                1045                1050


Arg Asn  Tyr Tyr Asn Asn Glu  Lys Lys Lys Asn Glu  Gly Ile Arg
    1055                1060                1065
```

```
Asn Lys  Ala Phe His Phe Gly  Ile Pro Asp Ser Lys  Ser Tyr Asp
    1070                 1075                 1080


Gln Leu  Ile Arg Asp Ala Glu  Val Met Phe Ile Ala  Asn Glu Val
    1085                 1090                 1095


Lys Pro  Thr His Ala Thr Lys  Tyr Thr Asp Leu Asn  Lys Gln Leu
    1100                 1105                 1110


His Thr  Val Cys Asp Lys Leu  Met Glu Thr Val His  Asn Asp Tyr
    1115                 1120                 1125


Phe Ser  Lys Glu Gly Asp Gly  Lys Lys Lys Arg Glu  Ala Ala Gly
    1130                 1135                 1140


Gln Lys  Tyr Phe Glu Asn Ile  Ile Ser Ala Lys
    1145                 1150


<210>  116
<211>  1119
<212>  PRT
<213>  Porphyromonas gingivalis

<400>  116

Met Thr Glu Gln Asn Glu Lys Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1               5                   10                  15


Glu Asp Lys His Phe Trp Ala Ala Phe Phe Asn Leu Ala Arg His Asn
            20                  25                  30


Ala Tyr Ile Thr Leu Ala His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
        35                  40                  45


Ala Asp Ile Thr Asn Asp Glu Asp Ile Leu Phe Phe Lys Gly Gln Trp
    50                  55                  60


Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ala Arg Leu Arg Ser Leu
65                  70                  75                  80
```

```
Ile Leu Lys His Phe Ser Phe Leu Glu Gly Ala Ala Tyr Gly Lys Lys
                85                  90                  95

Leu Phe Glu Ser Gln Ser Ser Gly Asn Lys Ser Ser Lys Asn Lys Glu
            100                 105                 110

Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
        115                 120                 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
    130                 135                 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Pro Glu Ser Ser Glu Leu
145                 150                 155                 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
                165                 170                 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val
            180                 185                 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
        195                 200                 205

Arg Tyr Gly Asn Asn Asp Asn Pro Phe Phe Lys His His Phe Val Asp
    210                 215                 220

Arg Glu Gly Thr Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225                 230                 235                 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
                245                 250                 255

Phe Lys Gly Gly Thr Glu Thr Tyr Gln Gln Met Thr Asn Glu Val Phe
            260                 265                 270

Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
        275                 280                 285
```

```
Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys
    290                 295                 300

Pro Lys Ser Leu Tyr Asp Arg Leu Arg Glu Glu Asp Arg Ala Arg Phe
305                 310                 315                 320

Arg Val Pro Val Asp Ile Leu Ser Asp Glu Glu Asp Thr Asp Gly Ala
                325                 330                 335

Glu Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
            340                 345                 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
        355                 360                 365

Leu Arg Phe Gln Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
    370                 375                 380

Lys Asn Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385                 390                 395                 400

Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
                405                 410                 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
            420                 425                 430

Lys Pro Tyr Ile Thr Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly
        435                 440                 445

Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln His Leu Trp Pro Ser
    450                 455                 460

Pro Glu Val Gly Ala Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465                 470                 475                 480

Lys Arg Phe Thr Ala Glu Ala Phe Leu Ser Ala His Glu Leu Met Pro
                485                 490                 495
```

```
Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Ala
            500                 505                 510

Ser Ala Glu Arg Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
        515                 520                 525

Tyr Ala Val Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Arg Asp
    530                 535                 540

Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu
545                 550                 555                 560

Pro Arg Gln Met Ile Gly Ile Leu Ser Gln Glu His Lys Asp Met Glu
                565                 570                 575

Glu Lys Ile Arg Lys Lys Leu Gln Glu Met Met Ala Asp Thr Asp His
            580                 585                 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly
        595                 600                 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
    610                 615                 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly
625                 630                 635                 640

Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
                645                 650                 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro
            660                 665                 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
        675                 680                 685

Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
    690                 695                 700
```

```
Arg Ser Tyr Leu Lys Ala Arg Lys Ala Phe Leu Gln Ser Ile Gly Arg
705                 710                 715                 720


Ser Asp Arg Val Glu Asn His Arg Phe Leu Leu Leu Lys Glu Pro Lys
                725                 730                 735


Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Gly Glu Phe His Leu
            740                 745                 750


Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
        755                 760                 765


Gly Leu Asp Glu Val Gly Ser Tyr Lys Glu Val Gly Phe Met Ala Lys
    770                 775                 780


Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Lys Asp Trp Val Gln Pro
785                 790                 795                 800


Phe Tyr Asn Tyr Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
                805                 810                 815


Gly Arg Phe Leu Ser Lys Glu Lys Arg Ala Glu Glu Trp Glu Ser Gly
            820                 825                 830


Lys Glu Arg Phe Arg Leu Ala Lys Leu Lys Lys Glu Ile Leu Glu Ala
        835                 840                 845


Lys Glu His Pro Tyr Leu Asp Phe Lys Ser Trp Gln Lys Phe Glu Arg
    850                 855                 860


Glu Leu Arg Leu Val Lys Asn Gln Asp Ile Ile Thr Trp Met Ile Cys
865                 870                 875                 880


Gly Asp Leu Met Glu Glu Asn Lys Val Glu Gly Leu Asp Thr Gly Thr
                885                 890                 895


Leu Tyr Leu Lys Asp Ile Arg Thr Asp Val Gln Glu Gln Gly Ser Leu
            900                 905                 910
```

```
Asn Val Leu Asn Arg Val Lys Pro Met Arg Leu Pro Val Val Val Tyr
        915                 920                 925


Arg Ala Asp Ser Arg Gly His Val His Lys Glu Gln Ala Pro Leu Ala
    930                 935                 940


Thr Val Tyr Ile Glu Glu Arg Asp Thr Lys Leu Leu Lys Gln Gly Asn
945                 950                 955                 960


Phe Lys Ser Phe Val Lys Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe
                965                 970                 975


Val Asp Thr Gly Ala Leu Ala Met Glu Gln Tyr Pro Ile Ser Lys Leu
            980                 985                 990


Arg Val Glu Tyr Glu Leu Ala Lys  Tyr Gln Thr Ala Arg  Val Cys Ala
        995                 1000                1005


Phe Glu  Gln Thr Leu Glu Leu  Glu Glu Ser Leu Leu  Thr Arg Cys
    1010                1015                1020


Pro His  Leu Pro Asp Lys Asn  Phe Arg Lys Met Leu  Glu Ser Trp
    1025                1030                1035


Ser Asp  Pro Leu Leu Asp Lys  Trp Pro Asp Leu His  Arg Lys Val
    1040                1045                1050


Arg Leu  Leu Ile Ala Val Arg  Asn Ala Phe Ser His  Asn Gln Tyr
    1055                1060                1065


Pro Met  Tyr Asp Glu Ala Val  Phe Ser Ser Ile Arg  Lys Tyr Asp
    1070                1075                1080


Pro Ser  Phe Pro Asp Ala Ile  Glu Glu Arg Met Gly  Leu Asn Ile
    1085                1090                1095


Ala His  Arg Leu Ser Glu Glu  Val Lys Gln Ala Lys  Glu Thr Val
    1100                1105                1110
```

Glu Arg Ile Ile Gln Ala
    1115

<210> 117
<211> 1179
<212> PRT
<213> Flavobacterium columnare

<400> 117

Met Ser Ser Lys Asn Glu Ser Tyr Asn Lys Gln Lys Thr Phe Asn His
1   5   10   15

Tyr Lys Gln Glu Asp Lys Tyr Phe Phe Gly Gly Phe Leu Asn Asn Ala
            20   25   30

Asp Asp Asn Leu Arg Gln Val Gly Lys Glu Phe Lys Thr Arg Ile Asn
        35   40   45

Phe Asn His Asn Asn Asn Glu Leu Ala Ser Val Phe Lys Asp Tyr Phe
    50   55   60

Asn Lys Glu Lys Ser Val Ala Lys Arg Glu His Ala Leu Asn Leu Leu
65   70   75   80

Ser Asn Tyr Phe Pro Val Leu Glu Arg Ile Gln Lys His Thr Asn His
                85   90   95

Asn Phe Glu Gln Thr Arg Glu Ile Phe Glu Leu Leu Leu Asp Thr Ile
            100   105   110

Lys Lys Leu Arg Asp Tyr Tyr Thr His His Tyr His Lys Pro Ile Thr
        115   120   125

Ile Asn Pro Lys Ile Tyr Asp Phe Leu Asp Asp Thr Leu Leu Asp Val
    130   135   140

Leu Ile Thr Ile Lys Lys Lys Lys Val Lys Asn Asp Thr Ser Arg Glu
145   150   155   160

```
Leu Leu Lys Glu Lys Leu Arg Pro Glu Leu Thr Gln Leu Lys Asn Gln
                165                 170                 175

Lys Arg Glu Glu Leu Ile Lys Lys Gly Lys Lys Leu Leu Glu Glu Asn
            180                 185                 190

Leu Glu Asn Ala Val Phe Asn His Cys Leu Arg Pro Phe Leu Glu Glu
        195                 200                 205

Asn Lys Thr Asp Asp Lys Gln Asn Lys Thr Val Ser Leu Arg Lys Tyr
    210                 215                 220

Arg Lys Ser Lys Pro Asn Glu Glu Thr Ser Ile Thr Leu Thr Gln Ser
225                 230                 235                 240

Gly Leu Val Phe Leu Met Ser Phe Phe Leu His Arg Lys Glu Phe Gln
                245                 250                 255

Val Phe Thr Ser Gly Leu Glu Gly Phe Lys Ala Lys Val Asn Thr Ile
            260                 265                 270

Lys Glu Glu Glu Ile Ser Leu Asn Lys Asn Asn Ile Val Tyr Met Ile
        275                 280                 285

Thr His Trp Ser Tyr Ser Tyr Tyr Asn Phe Lys Gly Leu Lys His Arg
    290                 295                 300

Ile Lys Thr Asp Gln Gly Val Ser Thr Leu Glu Gln Asn Asn Thr Thr
305                 310                 315                 320

His Ser Leu Thr Asn Thr Asn Thr Lys Glu Ala Leu Leu Thr Gln Ile
                325                 330                 335

Val Asp Tyr Leu Ser Lys Val Pro Asn Glu Ile Tyr Glu Thr Leu Ser
            340                 345                 350

Glu Lys Gln Gln Lys Glu Phe Glu Glu Asp Ile Asn Glu Tyr Met Arg
        355                 360                 365
```

```
Glu Asn Pro Glu Asn Glu Asp Ser Thr Phe Ser Ser Ile Val Ser His
    370                 375                 380

Lys Val Ile Arg Lys Arg Tyr Glu Asn Lys Phe Asn Tyr Phe Ala Met
385                 390                 395                 400

Arg Phe Leu Asp Glu Tyr Ala Glu Leu Pro Thr Leu Arg Phe Met Val
                405                 410                 415

Asn Phe Gly Asp Tyr Ile Lys Asp Arg Gln Lys Lys Ile Leu Glu Ser
            420                 425                 430

Ile Gln Phe Asp Ser Glu Arg Ile Ile Lys Lys Glu Ile His Leu Phe
        435                 440                 445

Glu Lys Leu Ser Leu Val Thr Glu Tyr Lys Lys Asn Val Tyr Leu Lys
    450                 455                 460

Glu Thr Ser Asn Ile Asp Leu Ser Arg Phe Pro Leu Phe Pro Asn Pro
465                 470                 475                 480

Ser Tyr Val Met Ala Asn Asn Asn Ile Pro Phe Tyr Ile Asp Ser Arg
                485                 490                 495

Ser Asn Asn Leu Asp Glu Tyr Leu Asn Gln Lys Lys Lys Ala Gln Ser
            500                 505                 510

Gln Asn Lys Lys Arg Asn Leu Thr Phe Glu Lys Tyr Asn Lys Glu Gln
        515                 520                 525

Ser Lys Asp Ala Ile Ile Ala Met Leu Gln Lys Glu Ile Gly Val Lys
    530                 535                 540

Asp Leu Gln Gln Arg Ser Thr Ile Gly Leu Leu Ser Cys Asn Glu Leu
545                 550                 555                 560

Pro Ser Met Leu Tyr Glu Val Ile Val Lys Asp Ile Lys Gly Ala Glu
                565                 570                 575
```

```
Leu Glu Asn Lys Ile Ala Gln Lys Ile Arg Glu Gln Tyr Gln Ser Ile
            580                 585                 590


Arg Asp Phe Thr Leu Asp Ser Pro Gln Lys Asp Asn Ile Pro Thr Thr
        595                 600                 605


Leu Ile Lys Thr Ile Asn Thr Asp Ser Ser Val Thr Phe Glu Asn Gln
    610                 615                 620


Pro Ile Asp Ile Pro Arg Leu Lys Asn Ala Ile Gln Lys Glu Leu Thr
625                 630                 635                 640


Leu Thr Gln Glu Lys Leu Leu Asn Val Lys Glu His Glu Ile Glu Val
                645                 650                 655


Asp Asn Tyr Asn Arg Asn Lys Asn Thr Tyr Lys Phe Lys Asn Gln Pro
            660                 665                 670


Lys Asn Lys Val Asp Asp Lys Lys Leu Gln Arg Lys Tyr Val Phe Tyr
        675                 680                 685


Arg Asn Glu Ile Arg Gln Glu Ala Asn Trp Leu Ala Ser Asp Leu Ile
    690                 695                 700


His Phe Met Lys Asn Lys Ser Leu Trp Lys Gly Tyr Met His Asn Glu
705                 710                 715                 720


Leu Gln Ser Phe Leu Ala Phe Phe Glu Asp Lys Lys Asn Asp Cys Ile
                725                 730                 735


Ala Leu Leu Glu Thr Val Phe Asn Leu Lys Glu Asp Cys Ile Leu Thr
            740                 745                 750


Lys Gly Leu Lys Asn Leu Phe Leu Lys His Gly Asn Phe Ile Asp Phe
        755                 760                 765


Tyr Lys Glu Tyr Leu Lys Leu Lys Glu Asp Phe Leu Asn Thr Glu Ser
    770                 775                 780
```

```
Thr Phe Leu Glu Asn Gly Leu Ile Gly Leu Pro Pro Lys Ile Leu Lys
785                 790                 795                 800


Lys Glu Leu Ser Lys Arg Phe Lys Tyr Ile Phe Ile Val Phe Gln Lys
                805                 810                 815


Arg Gln Phe Ile Ile Lys Glu Leu Glu Glu Lys Lys Asn Asn Leu Tyr
            820                 825                 830


Ala Asp Ala Ile Asn Leu Ser Arg Gly Ile Phe Asp Glu Lys Pro Thr
        835                 840                 845


Met Ile Pro Phe Lys Lys Pro Asn Pro Asp Glu Phe Ala Ser Trp Phe
    850                 855                 860


Val Ala Ser Tyr Gln Tyr Asn Asn Tyr Gln Ser Phe Tyr Glu Leu Thr
865                 870                 875                 880


Pro Asp Ile Val Glu Arg Asp Lys Lys Lys Lys Tyr Lys Asn Leu Arg
                885                 890                 895


Ala Ile Asn Lys Val Lys Ile Gln Asp Tyr Tyr Leu Lys Leu Met Val
            900                 905                 910


Asp Thr Leu Tyr Gln Asp Leu Phe Asn Gln Pro Leu Asp Lys Ser Leu
        915                 920                 925


Ser Asp Phe Tyr Val Ser Lys Ala Glu Arg Glu Lys Ile Lys Ala Asp
    930                 935                 940


Ala Lys Ala Tyr Gln Lys Arg Asn Asp Ser Ser Leu Trp Asn Lys Val
945                 950                 955                 960


Ile His Leu Ser Leu Gln Asn Asn Arg Ile Thr Ala Asn Pro Lys Leu
                965                 970                 975


Lys Asp Ile Gly Lys Tyr Lys Arg Ala Leu Gln Asp Glu Lys Ile Ala
            980                 985                 990
```

```
Thr Leu Leu Thr Tyr Asp Asp Arg  Thr Trp Thr Tyr Ala  Leu Gln Lys
        995                 1000                1005


Pro Glu  Lys Glu Asn Glu Asn  Asp Tyr Lys Glu Leu  His Tyr Thr
    1010                1015                1020


Ala Leu  Asn Met Glu Leu Gln  Glu Tyr Glu Lys Val  Arg Ser Lys
    1025                1030                1035


Glu Leu  Leu Lys Gln Val Gln  Glu Leu Glu Lys Gln  Ile Leu Glu
    1040                1045                1050


Glu Tyr  Thr Asp Phe Leu Ser  Thr Gln Ile His Pro  Ala Asp Phe
    1055                1060                1065


Glu Arg  Glu Gly Asn Pro Asn  Phe Lys Lys Tyr Leu  Ala His Ser
    1070                1075                1080


Ile Leu  Glu Asn Glu Asp Asp  Leu Asp Lys Leu Pro  Glu Lys Val
    1085                1090                1095


Glu Ala  Met Arg Glu Leu Asp  Glu Thr Ile Thr Asn  Pro Ile Ile
    1100                1105                1110


Lys Lys  Ala Ile Val Leu Ile  Ile Ile Arg Asn Lys  Met Ala His
    1115                1120                1125


Asn Gln  Tyr Pro Pro Lys Phe  Ile Tyr Asp Leu Ala  Asn Arg Phe
    1130                1135                1140


Val Pro  Lys Lys Glu Glu Glu  Tyr Phe Ala Thr Tyr  Phe Asn Arg
    1145                1150                1155


Val Phe  Glu Thr Ile Thr Lys  Glu Leu Trp Glu Asn  Lys Glu Lys
    1160                1165                1170


Lys Asp  Lys Thr Gln Val
    1175
```

<210> 118
<211> 1145
<212> PRT
<213> Psychroflexus torquis

<400> 118

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

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

Leu Asn Ala Val Phe Ala Glu Phe Lys Glu Arg Ile Ser Tyr Lys Ala
35 40 45

Lys Asp Glu Asn Ile Ser Ser Leu Ile Glu Lys His Phe Ile Asp Asn
50 55 60

Met Ser Ile Val Asp Tyr Glu Lys Lys Ile Ser Ile Leu Asn Gly Tyr
65 70 75 80

Leu Pro Ile Ile Asp Phe Leu Asp Asp Glu Leu Glu Asn Asn Leu Asn
85 90 95

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

Ile Glu Lys Leu Arg Asp Tyr Tyr Thr His Phe Tyr His Asp Pro Ile
115 120 125

Thr Phe Glu Asp Asn Lys Glu Pro Leu Leu Glu Leu Leu Asp Glu Val
130 135 140

Leu Leu Lys Thr Ile Leu Asp Val Lys Lys Lys Tyr Leu Lys Thr Asp
145 150 155 160

Lys Thr Lys Glu Ile Leu Lys Asp Ser Leu Arg Glu Glu Met Asp Leu
165 170 175

Leu Val Ile Arg Lys Thr Asp Glu Leu Arg Glu Lys Lys Lys Thr Asn

```
            180                 185                 190

Pro Lys Ile Gln His Thr Asp Ser Ser Gln Ile Lys Asn Ser Ile Phe
        195                 200                 205

Asn Asp Ala Phe Gln Gly Leu Leu Tyr Glu Asp Lys Gly Asn Asn Lys
    210                 215                 220

Lys Thr Gln Val Ser His Arg Ala Lys Thr Arg Leu Asn Pro Lys Asp
225                 230                 235                 240

Ile His Lys Gln Glu Glu Arg Asp Phe Glu Ile Pro Leu Ser Thr Ser
                245                 250                 255

Gly Leu Val Phe Leu Met Ser Leu Phe Leu Ser Lys Lys Glu Ile Glu
            260                 265                 270

Asp Phe Lys Ser Asn Ile Lys Gly Phe Lys Gly Lys Val Val Lys Asp
        275                 280                 285

Glu Asn His Asn Ser Leu Lys Tyr Met Ala Thr His Arg Val Tyr Ser
    290                 295                 300

Ile Leu Ala Phe Lys Gly Leu Lys Tyr Arg Ile Lys Thr Asp Thr Phe
305                 310                 315                 320

Ser Lys Glu Thr Leu Met Met Gln Met Ile Asp Glu Leu Ser Lys Val
                325                 330                 335

Pro Asp Cys Val Tyr Gln Asn Leu Ser Glu Thr Lys Gln Lys Asp Phe
            340                 345                 350

Ile Glu Asp Trp Asn Glu Tyr Phe Lys Asp Asn Glu Glu Asn Thr Glu
        355                 360                 365

Asn Leu Glu Asn Ser Arg Val Val His Pro Val Ile Arg Lys Arg Tyr
    370                 375                 380

Glu Asp Lys Phe Asn Tyr Phe Ala Ile Arg Phe Leu Asp Glu Phe Ala
```

385 390 395 400

Asn Phe Lys Thr Leu Lys Phe Gln Val Phe Met Gly Tyr Tyr Ile His
405 410 415

Asp Gln Arg Thr Lys Thr Ile Gly Thr Thr Asn Ile Thr Thr Glu Arg
420 425 430

Thr Val Lys Glu Lys Ile Asn Val Phe Gly Lys Leu Ser Lys Met Asp
435 440 445

Asn Leu Lys Lys His Phe Phe Ser Gln Leu Ser Asp Asp Glu Asn Thr
450 455 460

Asp Trp Glu Phe Phe Pro Asn Pro Ser Tyr Asn Phe Leu Thr Gln Ala
465 470 475 480

Asp Asn Ser Pro Ala Asn Asn Ile Pro Ile Tyr Leu Glu Leu Lys Asn
485 490 495

Gln Gln Ile Ile Lys Glu Lys Asp Ala Ile Lys Ala Glu Val Asn Gln
500 505 510

Thr Gln Asn Arg Asn Pro Asn Lys Pro Ser Lys Arg Asp Leu Leu Asn
515 520 525

Lys Ile Leu Lys Thr Tyr Glu Asp Phe His Gln Gly Asp Pro Thr Ala
530 535 540

Ile Leu Ser Leu Asn Glu Ile Pro Ala Leu Leu His Leu Phe Leu Val
545 550 555 560

Lys Pro Asn Asn Lys Thr Gly Gln Gln Ile Glu Asn Ile Ile Arg Ile
565 570 575

Lys Ile Glu Lys Gln Phe Lys Ala Ile Asn His Pro Ser Lys Asn Asn
580 585 590

Lys Gly Ile Pro Lys Ser Leu Phe Ala Asp Thr Asn Val Arg Val Asn

595 600 605

Ala Ile Lys Leu Lys Lys Asp Leu Glu Ala Glu Leu Asp Met Leu Asn
610 615 620

Lys Lys His Ile Ala Phe Lys Glu Asn Gln Lys Ala Ser Ser Asn Tyr
625 630 635 640

Asp Lys Leu Leu Lys Glu His Gln Phe Thr Pro Lys Asn Lys Arg Pro
645 650 655

Glu Leu Arg Lys Tyr Val Phe Tyr Lys Ser Glu Lys Gly Glu Glu Ala
660 665 670

Thr Trp Leu Ala Asn Asp Ile Lys Arg Phe Met Pro Lys Asp Phe Lys
675 680 685

Thr Lys Trp Lys Gly Cys Gln His Ser Glu Leu Gln Arg Lys Leu Ala
690 695 700

Phe Tyr Asp Arg His Thr Lys Gln Asp Ile Lys Glu Leu Leu Ser Gly
705 710 715 720

Cys Glu Phe Asp His Ser Leu Leu Asp Ile Asn Ala Tyr Phe Gln Lys
725 730 735

Asp Asn Phe Glu Asp Phe Phe Ser Lys Tyr Leu Glu Asn Arg Ile Glu
740 745 750

Thr Leu Glu Gly Val Leu Lys Lys Leu His Asp Phe Lys Asn Glu Pro
755 760 765

Thr Pro Leu Lys Gly Val Phe Lys Asn Cys Phe Lys Phe Leu Lys Arg
770 775 780

Gln Asn Tyr Val Thr Glu Ser Pro Glu Ile Ile Lys Lys Arg Ile Leu
785 790 795 800

Ala Lys Pro Thr Phe Leu Pro Arg Gly Val Phe Asp Glu Arg Pro Thr

805 810 815

Met Lys Lys Gly Lys Asn Pro Leu Lys Asp Lys Asn Glu Phe Ala Glu
820 825 830

Trp Phe Val Glu Tyr Leu Glu Asn Lys Asp Tyr Gln Lys Phe Tyr Asn
835 840 845

Ala Glu Glu Tyr Arg Met Arg Asp Ala Asp Phe Lys Lys Asn Ala Val
850 855 860

Ile Lys Lys Gln Lys Leu Lys Asp Phe Tyr Thr Leu Gln Met Val Asn
865 870 875 880

Tyr Leu Leu Lys Glu Val Phe Gly Lys Asp Glu Met Asn Leu Gln Leu
885 890 895

Ser Glu Leu Phe Gln Thr Arg Gln Glu Arg Leu Lys Leu Gln Gly Ile
900 905 910

Ala Lys Lys Gln Met Asn Lys Glu Thr Gly Asp Ser Ser Glu Asn Thr
915 920 925

Arg Asn Gln Thr Tyr Ile Trp Asn Lys Asp Val Pro Val Ser Phe Phe
930 935 940

Asn Gly Lys Val Thr Ile Asp Lys Val Lys Leu Lys Asn Ile Gly Lys
945 950 955 960

Tyr Lys Arg Tyr Glu Arg Asp Glu Arg Val Lys Thr Phe Ile Gly Tyr
965 970 975

Glu Val Asp Glu Lys Trp Met Met Tyr Leu Pro His Asn Trp Lys Asp
980 985 990

Arg Tyr Ser Val Lys Pro Ile Asn Val Ile Asp Leu Gln Ile Gln Glu
995 1000 1005

Tyr Glu Glu Ile Arg Ser His Glu Leu Leu Lys Glu Ile Gln Asn

1010 1015 1020

Leu Glu Gln Tyr Ile Tyr Asp His Thr Thr Asp Lys Asn Ile Leu
1025 1030 1035

Leu Gln Asp Gly Asn Pro Asn Phe Lys Met Tyr Val Leu Asn Gly
1040 1045 1050

Leu Leu Ile Gly Ile Lys Gln Val Asn Ile Pro Asp Phe Ile Val
1055 1060 1065

Leu Lys Gln Asn Thr Asn Phe Asp Lys Ile Asp Phe Thr Gly Ile
1070 1075 1080

Ala Ser Cys Ser Glu Leu Glu Lys Lys Thr Ile Ile Leu Ile Ala
1085 1090 1095

Ile Arg Asn Lys Phe Ala His Asn Gln Leu Pro Asn Lys Met Ile
1100 1105 1110

Tyr Asp Leu Ala Asn Glu Phe Leu Lys Ile Glu Lys Asn Glu Thr
1115 1120 1125

Tyr Ala Asn Tyr Tyr Leu Lys Val Leu Lys Lys Met Ile Ser Asp
1130 1135 1140

Leu Ala
1145

<210> 119
<211> 948
<212> PRT
<213> Riemerella anatipestifer

<400> 119

Met Phe Phe Ser Phe His Asn Ala Gln Arg Val Ile Phe Lys His Leu
1 5 10 15

Tyr Lys Ala Phe Asp Ala Ser Leu Arg Met Val Lys Glu Asp Tyr Lys
20 25 30

```
Ala His Phe Thr Val Asn Leu Thr Arg Asp Phe Ala His Leu Asn Arg
        35                  40                  45

Lys Gly Lys Asn Lys Gln Asp Asn Pro Asp Phe Asn Arg Tyr Arg Phe
    50                  55                  60

Glu Lys Asp Gly Phe Phe Thr Glu Ser Gly Leu Leu Phe Phe Thr Asn
65                  70                  75                  80

Leu Phe Leu Asp Lys Arg Asp Ala Tyr Trp Met Leu Lys Lys Val Ser
                85                  90                  95

Gly Phe Lys Ala Ser His Lys Gln Arg Glu Lys Met Thr Thr Glu Val
            100                 105                 110

Phe Cys Arg Ser Arg Ile Leu Leu Pro Lys Leu Arg Leu Glu Ser Arg
        115                 120                 125

Tyr Asp His Asn Gln Met Leu Leu Asp Met Leu Ser Glu Leu Ser Arg
    130                 135                 140

Cys Pro Lys Leu Leu Tyr Glu Lys Leu Ser Glu Glu Asn Lys Lys His
145                 150                 155                 160

Phe Gln Val Glu Ala Asp Gly Phe Leu Asp Glu Ile Glu Glu Glu Gln
                165                 170                 175

Asn Pro Phe Lys Asp Thr Leu Ile Arg His Gln Asp Arg Phe Pro Tyr
            180                 185                 190

Phe Ala Leu Arg Tyr Leu Asp Leu Asn Glu Ser Phe Lys Ser Ile Arg
        195                 200                 205

Phe Gln Val Asp Leu Gly Thr Tyr His Tyr Cys Ile Tyr Asp Lys Lys
    210                 215                 220

Ile Gly Asp Glu Gln Glu Lys Arg His Leu Thr Arg Thr Leu Leu Ser
225                 230                 235                 240
```

```
Phe Gly Arg Leu Gln Asp Phe Thr Glu Ile Asn Arg Pro Gln Glu Trp
                245                 250                 255

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

Phe Ile Ser Lys Thr Thr Pro His Tyr His Ile Thr Asp Asn Lys Ile
        275                 280                 285

Gly Phe Arg Leu Gly Thr Ser Lys Glu Leu Tyr Pro Ser Leu Glu Ile
    290                 295                 300

Lys Asp Gly Ala Asn Arg Ile Ala Lys Tyr Pro Tyr Asn Ser Gly Phe
305                 310                 315                 320

Val Ala His Ala Phe Ile Ser Val His Glu Leu Leu Pro Leu Met Phe
                325                 330                 335

Tyr Gln His Leu Thr Gly Lys Ser Glu Asp Leu Leu Lys Glu Thr Val
            340                 345                 350

Arg His Ile Gln Arg Ile Tyr Lys Asp Phe Glu Glu Glu Arg Ile Asn
        355                 360                 365

Thr Ile Glu Asp Leu Glu Lys Ala Asn Gln Gly Arg Leu Pro Leu Gly
    370                 375                 380

Ala Phe Pro Lys Gln Met Leu Gly Leu Leu Gln Asn Lys Gln Pro Asp
385                 390                 395                 400

Leu Ser Glu Lys Ala Lys Ile Lys Ile Glu Lys Leu Ile Ala Glu Thr
                405                 410                 415

Lys Leu Leu Ser His Arg Leu Asn Thr Lys Leu Lys Ser Ser Pro Lys
            420                 425                 430

Leu Gly Lys Arg Arg Glu Lys Leu Ile Lys Thr Gly Val Leu Ala Asp
        435                 440                 445
```

```
Trp Leu Val Lys Asp Phe Met Arg Phe Gln Pro Val Ala Tyr Asp Ala
    450                 455                 460

Gln Asn Gln Pro Ile Lys Ser Ser Lys Ala Asn Ser Thr Glu Phe Trp
465                 470                 475                 480

Phe Ile Arg Arg Ala Leu Ala Leu Tyr Gly Gly Glu Lys Asn Arg Leu
                485                 490                 495

Glu Gly Tyr Phe Lys Gln Thr Asn Leu Ile Gly Asn Thr Asn Pro His
            500                 505                 510

Pro Phe Leu Asn Lys Phe Asn Trp Lys Ala Cys Arg Asn Leu Val Asp
        515                 520                 525

Phe Tyr Gln Gln Tyr Leu Glu Gln Arg Glu Lys Phe Leu Glu Ala Ile
    530                 535                 540

Lys His Gln Pro Trp Glu Pro Tyr Gln Tyr Cys Leu Leu Leu Lys Val
545                 550                 555                 560

Pro Lys Glu Asn Arg Lys Asn Leu Val Lys Gly Trp Glu Gln Gly Gly
                565                 570                 575

Ile Ser Leu Pro Arg Gly Leu Phe Thr Glu Ala Ile Arg Glu Thr Leu
            580                 585                 590

Ser Lys Asp Leu Thr Leu Ser Lys Pro Ile Arg Lys Glu Ile Lys Lys
        595                 600                 605

His Gly Arg Val Gly Phe Ile Ser Arg Ala Ile Thr Leu Tyr Phe Lys
    610                 615                 620

Glu Lys Tyr Gln Asp Lys His Gln Ser Phe Tyr Asn Leu Ser Tyr Lys
625                 630                 635                 640

Leu Glu Ala Lys Ala Pro Leu Leu Lys Lys Glu Glu His Tyr Glu Tyr
                645                 650                 655
```

```
Trp Gln Gln Asn Lys Pro Gln Ser Pro Thr Glu Ser Gln Arg Leu Glu
            660                 665                 670

Leu His Thr Ser Asp Arg Trp Lys Asp Tyr Leu Leu Tyr Lys Arg Trp
        675                 680                 685

Gln His Leu Glu Lys Lys Leu Arg Leu Tyr Arg Asn Gln Asp Ile Met
    690                 695                 700

Leu Trp Leu Met Thr Leu Glu Leu Thr Lys Asn His Phe Lys Glu Leu
705                 710                 715                 720

Asn Leu Asn Tyr His Gln Leu Lys Leu Glu Asn Leu Ala Val Asn Val
                725                 730                 735

Gln Glu Ala Asp Ala Lys Leu Asn Pro Leu Asn Gln Thr Leu Pro Met
            740                 745                 750

Val Leu Pro Val Lys Val Tyr Pro Thr Thr Ala Phe Gly Glu Val Gln
        755                 760                 765

Tyr His Glu Thr Pro Ile Arg Thr Val Tyr Ile Arg Glu Glu Gln Thr
    770                 775                 780

Lys Ala Leu Lys Met Gly Asn Phe Lys Ala Leu Val Lys Asp Arg Arg
785                 790                 795                 800

Leu Asn Gly Leu Phe Ser Phe Ile Lys Glu Glu Asn Asp Thr Gln Lys
                805                 810                 815

His Pro Ile Ser Gln Leu Arg Leu Arg Arg Glu Leu Glu Ile Tyr Gln
            820                 825                 830

Ser Leu Arg Val Asp Ala Phe Lys Glu Thr Leu Ser Leu Glu Glu Lys
        835                 840                 845

Leu Leu Asn Lys His Ala Ser Leu Ser Ser Leu Glu Asn Glu Phe Arg
    850                 855                 860
```

```
Thr Leu Leu Glu Glu Trp Lys Lys Lys Tyr Ala Ala Ser Ser Met Val
865                 870                 875                 880


Thr Asp Lys His Ile Ala Phe Ile Ala Ser Val Arg Asn Ala Phe Cys
                885                 890                 895


His Asn Gln Tyr Pro Phe Tyr Lys Glu Thr Leu His Ala Pro Ile Leu
            900                 905                 910


Leu Phe Thr Val Ala Gln Pro Thr Thr Glu Glu Lys Asp Gly Leu Gly
        915                 920                 925


Ile Ala Glu Ala Leu Leu Lys Val Leu Arg Glu Tyr Cys Glu Ile Val
    930                 935                 940


Lys Ser Gln Ile
945


<210>  120
<211>  1139
<212>  PRT
<213>  Prevotella pleuritidis

<400>  120

Met Glu Asn Asp Lys Arg Leu Glu Glu Ser Ala Cys Tyr Thr Leu Asn
1               5                   10                  15


Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn Val
            20                  25                  30


Tyr Ile Thr Val Asn His Ile Asn Lys Thr Leu Glu Leu Lys Asn Lys
        35                  40                  45


Lys Asn Gln Glu Ile Ile Ile Asp Asn Asp Gln Asp Ile Leu Ala Ile
    50                  55                  60


Lys Thr His Trp Ala Lys Val Asn Gly Asp Leu Asn Lys Thr Asp Arg
65                  70                  75                  80
```

```
Leu Arg Glu Leu Met Ile Lys His Phe Pro Phe Leu Glu Ala Ala Ile
                85                  90                  95

Tyr Ser Asn Asn Lys Glu Asp Lys Glu Glu Val Lys Glu Glu Lys Gln
            100                 105                 110

Ala Lys Ala Gln Ser Phe Lys Ser Leu Lys Asp Cys Leu Phe Leu Phe
        115                 120                 125

Leu Glu Lys Leu Gln Glu Ala Arg Asn Tyr Tyr Ser His Tyr Lys Tyr
    130                 135                 140

Ser Glu Ser Ser Lys Glu Pro Glu Phe Glu Glu Gly Leu Leu Glu Lys
145                 150                 155                 160

Met Tyr Asn Thr Phe Asp Ala Ser Ile Arg Leu Val Lys Glu Asp Tyr
                165                 170                 175

Gln Tyr Asn Lys Asp Ile Asp Pro Glu Lys Asp Phe Lys His Leu Glu
            180                 185                 190

Arg Lys Glu Asp Phe Asn Tyr Leu Phe Thr Asp Lys Asp Asn Lys Gly
        195                 200                 205

Lys Ile Thr Lys Asn Gly Leu Leu Phe Phe Val Ser Leu Phe Leu Glu
    210                 215                 220

Lys Lys Asp Ala Ile Trp Met Gln Gln Lys Phe Arg Gly Phe Lys Asp
225                 230                 235                 240

Asn Arg Gly Asn Lys Glu Lys Met Thr His Glu Val Phe Cys Arg Ser
                245                 250                 255

Arg Met Leu Leu Pro Lys Ile Arg Leu Glu Ser Thr Gln Thr Gln Asp
            260                 265                 270

Trp Ile Leu Leu Asp Met Leu Asn Glu Leu Ile Arg Cys Pro Lys Ser
        275                 280                 285
```

```
Leu Tyr Glu Arg Leu Gln Gly Ala Tyr Arg Glu Lys Phe Lys Val Pro
    290                 295                 300

Phe Asp Ser Ile Asp Glu Asp Tyr Asp Ala Glu Gln Glu Pro Phe Arg
305                 310                 315                 320

Asn Thr Leu Val Arg His Gln Asp Arg Phe Pro Tyr Phe Ala Leu Arg
                325                 330                 335

Tyr Phe Asp Tyr Asn Glu Ile Phe Lys Asn Leu Arg Phe Gln Ile Asp
            340                 345                 350

Leu Gly Thr Tyr His Phe Ser Ile Tyr Lys Lys Leu Ile Gly Gly Lys
        355                 360                 365

Lys Glu Asp Arg His Leu Thr His Lys Leu Tyr Gly Phe Glu Arg Ile
    370                 375                 380

Gln Glu Phe Thr Lys Gln Asn Arg Pro Asp Lys Trp Gln Ala Ile Ile
385                 390                 395                 400

Lys Asp Leu Asp Thr Tyr Glu Thr Ser Asn Glu Arg Tyr Ile Ser Glu
                405                 410                 415

Thr Thr Pro His Tyr His Leu Glu Asn Gln Lys Ile Gly Ile Arg Phe
            420                 425                 430

Arg Asn Asp Asn Asn Asp Ile Trp Pro Ser Leu Lys Thr Asn Gly Glu
        435                 440                 445

Lys Asn Glu Lys Ser Lys Tyr Asn Leu Asp Lys Pro Tyr Gln Ala Glu
    450                 455                 460

Ala Phe Leu Ser Val His Glu Leu Leu Pro Met Met Phe Tyr Tyr Leu
465                 470                 475                 480

Leu Leu Lys Met Glu Asn Thr Asp Asn Asp Lys Glu Asp Asn Glu Val
                485                 490                 495
```

```
Gly Thr Lys Lys Lys Gly Asn Lys Asn Asn Lys Gln Glu Lys His Lys
            500                 505                 510

Ile Glu Glu Ile Ile Glu Asn Lys Ile Lys Asp Ile Tyr Ala Leu Tyr
        515                 520                 525

Asp Ala Phe Thr Asn Gly Glu Ile Asn Ser Ile Asp Glu Leu Ala Glu
    530                 535                 540

Gln Arg Glu Gly Lys Asp Ile Glu Ile Gly His Leu Pro Lys Gln Leu
545                 550                 555                 560

Ile Val Ile Leu Lys Asn Lys Ser Lys Asp Met Ala Glu Lys Ala Asn
                565                 570                 575

Arg Lys Gln Lys Glu Met Ile Lys Asp Thr Lys Lys Arg Leu Ala Thr
            580                 585                 590

Leu Asp Lys Gln Val Lys Gly Glu Ile Glu Asp Gly Gly Arg Asn Ile
        595                 600                 605

Arg Leu Leu Lys Ser Gly Glu Ile Ala Arg Trp Leu Val Asn Asp Met
    610                 615                 620

Met Arg Phe Gln Pro Val Gln Lys Asp Asn Glu Gly Lys Pro Leu Asn
625                 630                 635                 640

Asn Ser Lys Ala Asn Ser Thr Glu Tyr Gln Met Leu Gln Arg Ser Leu
                645                 650                 655

Ala Leu Tyr Asn Lys Glu Glu Lys Pro Thr Arg Tyr Phe Arg Gln Val
            660                 665                 670

Asn Leu Ile Lys Ser Ser Asn Pro His Pro Phe Leu Glu Asp Thr Lys
        675                 680                 685

Trp Glu Glu Cys Tyr Asn Ile Leu Ser Phe Tyr Arg Asn Tyr Leu Lys
    690                 695                 700
```

```
Ala Lys Ile Lys Phe Leu Asn Lys Leu Lys Pro Glu Asp Trp Lys Lys
705                 710                 715                 720


Asn Gln Tyr Phe Leu Met Leu Lys Glu Pro Lys Thr Asn Arg Lys Thr
                725                 730                 735


Leu Val Gln Gly Trp Lys Asn Gly Phe Asn Leu Pro Arg Gly Ile Phe
            740                 745                 750


Thr Glu Pro Ile Lys Glu Trp Phe Lys Arg His Gln Asn Asp Ser Glu
        755                 760                 765


Glu Tyr Lys Lys Val Glu Ala Leu Asp Arg Val Gly Leu Val Ala Lys
    770                 775                 780


Val Ile Pro Leu Phe Phe Lys Glu Glu Tyr Phe Lys Glu Asp Ala Gln
785                 790                 795                 800


Lys Glu Ile Asn Asn Cys Val Gln Pro Phe Tyr Ser Phe Pro Tyr Asn
                805                 810                 815


Val Gly Asn Ile His Lys Pro Glu Glu Lys Asn Phe Leu His Cys Glu
            820                 825                 830


Glu Arg Arg Lys Leu Trp Asp Lys Lys Lys Asp Lys Phe Lys Gly Tyr
        835                 840                 845


Lys Ala Lys Glu Lys Ser Lys Lys Met Thr Asp Lys Glu Lys Glu Glu
    850                 855                 860


His Arg Ser Tyr Leu Glu Phe Gln Ser Trp Asn Lys Phe Glu Arg Glu
865                 870                 875                 880


Leu Arg Leu Val Arg Asn Gln Asp Ile Leu Thr Trp Leu Leu Cys Thr
                885                 890                 895


Lys Leu Ile Asp Lys Leu Lys Ile Asp Glu Leu Asn Ile Glu Glu Leu
            900                 905                 910
```

```
Gln Lys Leu Arg Leu Lys Asp Ile Asp Thr Asp Thr Ala Lys Lys Glu
        915                 920                 925


Lys Asn Asn Ile Leu Asn Arg Val Met Pro Met Arg Leu Pro Val Thr
    930                 935                 940


Val Tyr Glu Ile Asp Lys Ser Phe Asn Ile Val Lys Asp Lys Pro Leu
945                 950                 955                 960


His Thr Val Tyr Ile Glu Glu Thr Gly Thr Lys Leu Leu Lys Gln Gly
                965                 970                 975


Asn Phe Lys Ala Leu Val Lys Asp Arg Arg Leu Asn Gly Leu Phe Ser
            980                 985                 990


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


Lys Leu  Arg Val Glu Tyr Glu  Leu Gly Ala Tyr Gln  Lys Ala Arg
    1010                1015                1020


Ile Asp  Ile Ile Lys Asp Met  Leu Ala Leu Glu Lys  Thr Leu Ile
    1025                1030                1035


Asp Asn  Asp Glu Asn Leu Pro  Thr Asn Lys Phe Ser  Asp Met Leu
    1040                1045                1050


Lys Ser  Trp Leu Lys Gly Lys  Gly Glu Ala Asn Lys  Ala Arg Leu
    1055                1060                1065


Gln Asn  Asp Val Gly Leu Leu  Val Ala Val Arg Asn  Ala Phe Ser
    1070                1075                1080


His Asn  Gln Tyr Pro Met Tyr  Asn Ser Glu Val Phe  Lys Gly Met
    1085                1090                1095


Lys Leu  Leu Ser Leu Ser Ser  Asp Ile Pro Glu Lys  Glu Gly Leu
    1100                1105                1110
```

```
Gly Ile  Ala Lys Gln Leu Lys  Asp Lys Ile Lys Glu  Thr Ile Glu
    1115                1120                1125


Arg Ile  Ile Glu Ile Glu Lys  Glu Ile Arg Asn
    1130                1135


<210>  121
<211>  1135
<212>  PRT
<213>  Porphyromonas gingivalis

<400>  121

Met Asn Thr Val Pro Ala Ser Glu Asn Lys Gly Gln Ser Arg Thr Val
1               5                   10                  15


Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu
            20                  25                  30


Asn Leu Ile Glu Val Glu Ser His Val Arg Ile Lys Phe Gly Lys Lys
        35                  40                  45


Lys Leu Asn Glu Glu Ser Leu Lys Gln Ser Leu Leu Cys Asp His Leu
    50                  55                  60


Leu Ser Val Asp Arg Trp Thr Lys Val Tyr Gly His Ser Arg Arg Tyr
65                  70                  75                  80


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


His Asp Ser Lys Thr Gly Val Asp Pro Asp Ser Ala Gln Arg Leu Ile
            100                 105                 110


Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His
        115                 120                 125


Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Glu Val Ser Pro Asp
    130                 135                 140
```

```
Ile Ser Ser Phe Ile Thr Gly Thr Tyr Ser Leu Ala Cys Gly Arg Ala
145                 150                 155                 160


Gln Ser Arg Phe Ala Asp Phe Phe Lys Pro Asp Asp Phe Val Leu Ala
                165                 170                 175


Lys Asn Arg Lys Glu Gln Leu Ile Ser Val Ala Asp Gly Lys Glu Cys
            180                 185                 190


Leu Thr Val Ser Gly Leu Ala Phe Phe Ile Cys Leu Phe Leu Asp Arg
        195                 200                 205


Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr
    210                 215                 220


Asp Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys
225                 230                 235                 240


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


Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu
            260                 265                 270


Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu
        275                 280                 285


Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asn Glu Glu
    290                 295                 300


Ser Arg Leu Leu Trp Asp Gly Ser Ser Asp Trp Ala Glu Ala Leu Thr
305                 310                 315                 320


Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe
                325                 330                 335


Ile Glu Glu Met Asp Leu Leu Lys Gly Ile Arg Phe Arg Val Asp Leu
            340                 345                 350
```

```
Gly Glu Ile Glu Leu Asp Ser Tyr Ser Lys Lys Val Gly Arg Asn Gly
        355                 360                 365

Glu Tyr Asp Arg Thr Ile Thr Asp His Ala Leu Ala Phe Gly Lys Leu
    370                 375                 380

Ser Asp Phe Gln Asn Glu Glu Glu Val Ser Arg Met Ile Ser Gly Glu
385                 390                 395                 400

Ala Ser Tyr Pro Val Arg Phe Ser Leu Phe Ala Pro Arg Tyr Ala Ile
                405                 410                 415

Tyr Asp Asn Lys Ile Gly Tyr Cys His Thr Ser Asp Pro Val Tyr Pro
            420                 425                 430

Lys Ser Lys Thr Gly Glu Lys Arg Ala Leu Ser Asn Pro Arg Ser Met
        435                 440                 445

Gly Phe Ile Ser Val His Asp Leu Arg Lys Leu Leu Leu Met Glu Leu
    450                 455                 460

Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Asp Phe Leu Arg Lys
465                 470                 475                 480

Ala Asn Arg Ile Leu Asp Glu Thr Ala Glu Gly Lys Leu Gln Phe Ser
                485                 490                 495

Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro
            500                 505                 510

Lys Ser Lys Asp Arg Arg Glu Lys Ala Glu Thr Thr Leu Glu Lys Tyr
        515                 520                 525

Lys Gln Glu Ile Lys Gly Arg Lys Asp Lys Leu Asn Ser Gln Leu Leu
    530                 535                 540

Ser Ala Phe Asp Met Asp Gln Arg Gln Leu Pro Ser Arg Leu Leu Asp
545                 550                 555                 560
```

```
Glu Trp Met Asn Ile Arg Pro Ala Ser His Ser Val Lys Leu Arg Thr
                565                 570                 575


Tyr Val Lys Gln Leu Asn Glu Asp Cys Arg Leu Arg Leu Gln Lys Phe
            580                 585                 590


Arg Lys Asp Gly Asp Gly Lys Ala Arg Ala Ile Pro Leu Val Gly Glu
        595                 600                 605


Met Ala Thr Phe Leu Ser Gln Asp Ile Val Arg Met Ile Ile Ser Glu
    610                 615                 620


Glu Thr Lys Lys Leu Ile Thr Ser Ala Tyr Tyr Asn Glu Met Gln Arg
625                 630                 635                 640


Ser Leu Ala Gln Tyr Ala Gly Glu Glu Asn Arg His Gln Phe Arg Ala
                645                 650                 655


Ile Val Ala Glu Leu Arg Leu Leu Asp Pro Ser Ser Gly His Pro Phe
            660                 665                 670


Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Asp Phe Tyr
        675                 680                 685


Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Thr Phe Tyr
    690                 695                 700


Arg Pro Glu Gln Asp Glu Asn Thr Lys Arg Arg Ile Ser Val Phe Phe
705                 710                 715                 720


Val Pro Asp Gly Glu Ala Arg Lys Leu Leu Pro Leu Leu Ile Arg Arg
                725                 730                 735


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


Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Ile Met
        755                 760                 765
```

```
Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe
    770                 775                 780

Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr
785                 790                 795                 800

Gly Leu Arg Arg Glu Leu Asn Ile His Gly Lys Ser Val Ser Tyr Ile
                805                 810                 815

Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu
            820                 825                 830

Lys Thr Val Gln Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro
        835                 840                 845

Val Pro Pro Asp Leu Ala Ala Asp Ile Lys Arg Ser Phe His Arg Ala
    850                 855                 860

Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg
865                 870                 875                 880

Leu Met Leu Met Ala Ile Asn Lys Met Met Thr Asp Arg Glu Glu Asp
                885                 890                 895

Ile Leu Pro Gly Leu Lys Asn Ile Asp Ser Ile Leu Asp Glu Glu Asn
            900                 905                 910

Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu
        915                 920                 925

Gly Gly Asp Asn Ser Leu Ser Leu Val Pro Ala Thr Ile Glu Ile Lys
    930                 935                 940

Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg
945                 950                 955                 960

Arg Val Pro Gly Leu Met Ser His Phe Pro Glu His Lys Ala Thr Leu
                965                 970                 975
```

```
Asp Glu Val Lys Thr Leu Leu Gly Glu Tyr Asp Arg Cys Arg Ile Lys
            980                 985                 990


Ile Phe Asp Trp Ala Phe Ala Leu  Glu Gly Ala Ile Met  Ser Asp Arg
        995                 1000                1005


Asp Leu  Lys Pro Tyr Leu His  Glu Ser Ser Ser Arg  Glu Gly Lys
    1010                1015                1020


Ser Gly  Glu His Ser Thr Leu  Val Lys Met Leu Val  Glu Lys Lys
    1025                1030                1035


Gly Cys  Leu Thr Pro Asp Glu  Ser Gln Tyr Leu Ile  Leu Ile Arg
    1040                1045                1050


Asn Lys  Ala Ala His Asn Gln  Phe Pro Cys Ala Ala  Glu Met Pro
    1055                1060                1065


Leu Ile  Tyr Arg Asp Val Ser  Ala Lys Val Gly Ser  Ile Glu Gly
    1070                1075                1080


Ser Ser  Ala Lys Asp Leu Pro  Glu Gly Ser Ser Leu  Val Asp Ser
    1085                1090                1095


Leu Trp  Lys Lys Tyr Glu Met  Ile Ile Arg Lys Ile  Leu Pro Ile
    1100                1105                1110


Leu Asp  Pro Glu Asn Arg Phe  Phe Gly Lys Leu Leu  Asn Asn Met
    1115                1120                1125


Ser Gln  Pro Ile Asn Asp Leu
    1130                1135


<210>  122
<211>  1135
<212>  PRT
<213>  Porphyromonas gingivalis

<400>  122

Met Asn Thr Val Pro Ala Ser Glu Asn Lys Gly Gln Ser Arg Thr Val
```

```
1               5                   10                  15

Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu
            20                  25                  30

Asn Leu Ile Glu Val Glu Ser His Val Arg Ile Lys Phe Gly Lys Lys
        35                  40                  45

Lys Leu Asn Glu Glu Ser Leu Lys Gln Ser Leu Leu Cys Asp His Leu
    50                  55                  60

Leu Ser Val Asp Arg Trp Thr Lys Val Tyr Gly His Ser Arg Arg Tyr
65                  70                  75                  80

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

His Asp Ser Lys Thr Gly Val Asp Pro Asp Ser Ala Gln Arg Leu Ile
            100                 105                 110

Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His
        115                 120                 125

Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Glu Val Ser Pro Asp
    130                 135                 140

Ile Ser Ser Phe Ile Thr Gly Thr Tyr Ser Leu Ala Cys Gly Arg Ala
145                 150                 155                 160

Gln Ser Arg Phe Ala Asp Phe Phe Lys Pro Asp Asp Phe Val Leu Ala
                165                 170                 175

Lys Asn Arg Lys Glu Gln Leu Ile Ser Val Ala Asp Gly Lys Glu Cys
            180                 185                 190

Leu Thr Val Ser Gly Leu Ala Phe Phe Ile Cys Leu Phe Leu Asp Arg
        195                 200                 205

Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr
```

210 215 220

Asn Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys
225 230 235 240

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

Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu
260 265 270

Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu
275 280 285

Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asn Glu Glu
290 295 300

Ser Arg Leu Leu Trp Asp Gly Ser Ser Asp Trp Ala Glu Ala Leu Thr
305 310 315 320

Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe
325 330 335

Ile Glu Glu Met Asp Leu Leu Lys Gly Ile Arg Phe Arg Val Asp Leu
340 345 350

Gly Glu Ile Glu Leu Asp Ser Tyr Ser Lys Lys Val Gly Arg Asn Gly
355 360 365

Glu Tyr Asp Arg Thr Ile Thr Asp His Ala Leu Ala Phe Gly Lys Leu
370 375 380

Ser Asp Phe Gln Asn Glu Glu Glu Val Ser Arg Met Ile Ser Gly Glu
385 390 395 400

Ala Ser Tyr Pro Val Arg Phe Ser Leu Phe Ala Pro Arg Tyr Ala Ile
405 410 415

Tyr Asp Asn Lys Ile Gly Tyr Cys His Thr Ser Asp Pro Val Tyr Pro

```
            420                 425                 430

Lys Ser Lys Thr Gly Glu Lys Arg Ala Leu Ser Asn Pro Gln Ser Met
        435                 440                 445

Gly Phe Ile Ser Val His Asp Leu Arg Lys Leu Leu Leu Met Glu Leu
    450                 455                 460

Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Gly Phe Leu Arg Lys
465                 470                 475                 480

Ala Asn Arg Ile Leu Asp Glu Thr Ala Glu Gly Lys Leu Gln Phe Ser
                485                 490                 495

Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro
            500                 505                 510

Lys Ser Lys Asp Arg Arg Glu Lys Ala Glu Thr Thr Leu Glu Lys Tyr
        515                 520                 525

Lys Gln Glu Ile Lys Gly Arg Lys Asp Lys Leu Asn Ser Gln Leu Leu
    530                 535                 540

Ser Ala Phe Asp Met Asn Gln Arg Gln Leu Pro Ser Arg Leu Leu Asp
545                 550                 555                 560

Glu Trp Met Asn Ile Arg Pro Ala Ser His Ser Val Lys Leu Arg Thr
                565                 570                 575

Tyr Val Lys Gln Leu Asn Glu Asp Cys Arg Leu Arg Leu Arg Lys Phe
            580                 585                 590

Arg Lys Asp Gly Asp Gly Lys Ala Arg Ala Ile Pro Leu Val Gly Glu
        595                 600                 605

Met Ala Thr Phe Leu Ser Gln Asp Ile Val Arg Met Ile Ile Ser Glu
    610                 615                 620

Glu Thr Lys Lys Leu Ile Thr Ser Ala Tyr Tyr Asn Glu Met Gln Arg
```

625 630 635 640

Ser Leu Ala Gln Tyr Ala Gly Glu Glu Asn Arg Arg Gln Phe Arg Ala
645 650 655

Ile Val Ala Glu Leu His Leu Leu Asp Pro Ser Ser Gly His Pro Phe
660 665 670

Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Asp Phe Tyr
675 680 685

Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Thr Phe Tyr
690 695 700

Arg Pro Glu Gln Asp Glu Asn Thr Lys Arg Arg Ile Ser Val Phe Phe
705 710 715 720

Val Pro Asp Gly Glu Ala Arg Lys Leu Leu Pro Leu Leu Ile Arg Arg
725 730 735

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

Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Ile Met
755 760 765

Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe
770 775 780

Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr
785 790 795 800

Gly Leu Arg Arg Glu Leu Asn Ile His Gly Lys Ser Val Ser Tyr Ile
805 810 815

Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu
820 825 830

Lys Thr Val Gln Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro

```
        835                 840                 845

Val Pro Pro Asp Leu Ala Ala Asp Ile Lys Arg Ser Phe His Arg Ala
    850                 855                 860

Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg
865                 870                 875                 880

Leu Met Leu Met Ala Ile Asn Lys Met Met Thr Asp Arg Glu Glu Asp
                885                 890                 895

Ile Leu Pro Gly Leu Lys Asn Ile Asp Ser Ile Leu Asp Lys Glu Asn
            900                 905                 910

Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu
        915                 920                 925

Gly Gly Asp Asn Ser Leu Ser Leu Val Pro Ala Thr Ile Glu Ile Lys
    930                 935                 940

Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg
945                 950                 955                 960

Arg Val Pro Gly Leu Met Ser His Phe Pro Glu His Lys Ala Thr Leu
                965                 970                 975

Asp Glu Val Lys Thr Leu Leu Gly Glu Tyr Asp Arg Cys Arg Ile Lys
            980                 985                 990

Ile Phe Asp Trp Ala Phe Ala Leu  Glu Gly Ala Ile Met  Ser Asp Arg
        995                 1000                1005

Asp Leu  Lys Pro Tyr Leu His  Glu Ser Ser Ser Arg  Glu Gly Lys
    1010                1015                1020

Ser Gly  Glu His Ser Thr Leu  Val Lys Met Leu Val  Glu Lys Lys
    1025                1030                1035

Gly Cys  Leu Thr Pro Asp Glu  Ser Gln Tyr Leu Ile  Leu Ile Arg
```

1040 1045 1050

Asn Lys Ala Ala His Asn Gln Phe Pro Cys Ala Ala Glu Met Pro
1055 1060 1065

Leu Ile Tyr Arg Asp Val Ser Ala Lys Val Gly Ser Ile Glu Gly
1070 1075 1080

Ser Ser Ala Lys Asp Leu Pro Glu Gly Ser Ser Leu Val Asp Ser
1085 1090 1095

Leu Trp Lys Lys Tyr Glu Met Ile Ile Arg Lys Ile Leu Pro Ile
1100 1105 1110

Leu Asp His Glu Asn Arg Phe Phe Gly Lys Leu Leu Asn Asn Met
1115 1120 1125

Ser Gln Pro Ile Asn Asp Leu
1130 1135

<210> 123
<211> 1135
<212> PRT
<213> Porphyromonas gingivalis

<400> 123

Met Asn Thr Val Pro Ala Ser Glu Asn Lys Gly Gln Ser Arg Thr Val
1 5 10 15

Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu
20 25 30

Asn Leu Ile Glu Val Glu Ser His Val Arg Ile Lys Phe Gly Lys Lys
35 40 45

Lys Leu Asn Glu Glu Ser Leu Lys Gln Ser Leu Leu Cys Asp His Leu
50 55 60

Leu Ser Val Asp Arg Trp Thr Lys Val Tyr Gly His Ser Arg Arg Tyr
65 70 75 80

```
Leu Pro Phe Leu His Tyr Phe Asp Pro Asp Ser Gln Ile Glu Lys Asp
                85                  90                  95

His Asp Ser Lys Thr Gly Val Asp Pro Asp Ser Ala Gln Arg Leu Ile
            100                 105                 110

Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His
        115                 120                 125

Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Glu Val Ser Pro Asp
    130                 135                 140

Ile Ser Ser Phe Ile Thr Gly Thr Tyr Ser Leu Ala Cys Gly Arg Ala
145                 150                 155                 160

Gln Ser Arg Phe Ala Asp Phe Phe Lys Pro Asp Asp Phe Val Leu Ala
                165                 170                 175

Lys Asn Arg Lys Glu Gln Leu Ile Ser Val Ala Asp Gly Lys Glu Cys
            180                 185                 190

Leu Thr Val Ser Gly Leu Ala Phe Phe Ile Cys Leu Phe Leu Asp Arg
        195                 200                 205

Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr
    210                 215                 220

Asp Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys
225                 230                 235                 240

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

Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu
            260                 265                 270

Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu
        275                 280                 285
```

```
Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asn Glu Glu
    290                 295                 300

Ser Arg Leu Leu Trp Asp Gly Ser Ser Asp Trp Ala Glu Ala Leu Thr
305                 310                 315                 320

Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe
                325                 330                 335

Ile Glu Glu Met Asp Leu Leu Lys Gly Ile Arg Phe Arg Val Asp Leu
            340                 345                 350

Gly Glu Ile Glu Leu Asp Ser Tyr Ser Lys Lys Val Gly Arg Asn Gly
        355                 360                 365

Glu Tyr Asp Arg Thr Ile Thr Asp His Ala Leu Ala Phe Gly Lys Leu
    370                 375                 380

Ser Asp Phe Gln Asn Glu Glu Glu Val Ser Arg Met Ile Ser Gly Glu
385                 390                 395                 400

Ala Ser Tyr Pro Val Arg Phe Ser Leu Phe Ala Pro Arg Tyr Ala Ile
                405                 410                 415

Tyr Asp Asn Lys Ile Gly Tyr Cys His Thr Ser Asp Pro Val Tyr Pro
            420                 425                 430

Lys Ser Lys Thr Gly Glu Lys Arg Ala Leu Ser Asn Pro Gln Ser Met
        435                 440                 445

Gly Phe Ile Ser Val His Asp Leu Arg Lys Leu Leu Leu Met Glu Leu
    450                 455                 460

Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Gly Phe Leu Arg Lys
465                 470                 475                 480

Ala Asn Arg Ile Leu Asp Glu Thr Ala Glu Gly Lys Leu Gln Phe Ser
                485                 490                 495
```

```
Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro
            500                 505                 510

Lys Ser Lys Asp Arg Arg Glu Lys Ala Glu Thr Thr Leu Glu Lys Tyr
        515                 520                 525

Lys Gln Glu Ile Lys Gly Arg Lys Asp Lys Leu Asn Ser Gln Leu Leu
    530                 535                 540

Ser Ala Phe Asp Met Asn Gln Arg Gln Leu Pro Ser Arg Leu Leu Asp
545                 550                 555                 560

Glu Trp Met Asn Ile Arg Pro Ala Ser His Ser Val Lys Leu Arg Thr
                565                 570                 575

Tyr Val Lys Gln Leu Asn Glu Asp Cys Arg Leu Arg Leu Arg Lys Phe
            580                 585                 590

Arg Lys Asp Gly Asp Gly Lys Ala Arg Ala Ile Pro Leu Val Gly Glu
        595                 600                 605

Met Ala Thr Phe Leu Ser Gln Asp Ile Val Arg Met Ile Ile Ser Glu
    610                 615                 620

Glu Thr Lys Lys Leu Ile Thr Ser Ala Tyr Tyr Asn Glu Met Gln Arg
625                 630                 635                 640

Ser Leu Ala Gln Tyr Ala Gly Glu Glu Asn Arg Arg Gln Phe Arg Ala
                645                 650                 655

Ile Val Ala Glu Leu His Leu Leu Asp Pro Ser Ser Gly His Pro Phe
            660                 665                 670

Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Asp Phe Tyr
        675                 680                 685

Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Thr Phe Tyr
    690                 695                 700
```

```
Arg Pro Glu Gln Asp Glu Asn Thr Lys Arg Arg Ile Ser Val Phe Phe
705                 710                 715                 720

Val Pro Asp Gly Glu Ala Arg Lys Leu Leu Pro Leu Leu Ile Arg Arg
                725                 730                 735

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

Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Ile Met
        755                 760                 765

Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe
    770                 775                 780

Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr
785                 790                 795                 800

Gly Leu Arg Arg Glu Leu Asn Ile His Gly Lys Ser Val Ser Tyr Ile
                805                 810                 815

Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu
            820                 825                 830

Lys Thr Val Gln Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro
        835                 840                 845

Val Pro Pro Asp Leu Ala Ala Asp Ile Lys Arg Ser Phe His Arg Ala
    850                 855                 860

Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg
865                 870                 875                 880

Leu Met Leu Met Ala Ile Asn Lys Met Met Thr Asp Arg Glu Glu Asp
                885                 890                 895

Ile Leu Pro Gly Leu Lys Asn Ile Asp Ser Ile Leu Asp Glu Glu Asn
            900                 905                 910
```

```
Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu
        915                 920                 925

Gly Gly Asp Asn Ser Leu Ser Leu Val Pro Ala Thr Ile Glu Ile Lys
    930                 935                 940

Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg
945                 950                 955                 960

Arg Val Pro Gly Leu Met Ser His Phe Pro Glu His Lys Ala Thr Leu
                965                 970                 975

Asp Glu Val Lys Thr Leu Leu Gly Glu Tyr Asp Arg Cys Arg Ile Lys
            980                 985                 990

Ile Phe Asp Trp Ala Phe Ala Leu  Glu Gly Ala Ile Met  Ser Asp Arg
        995                 1000                1005

Asp Leu  Lys Pro Tyr Leu His  Glu Ser Ser Ser Arg  Glu Gly Lys
    1010                1015                1020

Ser Gly  Glu His Ser Thr Leu  Val Lys Met Leu Val  Glu Lys Lys
    1025                1030                1035

Gly Cys  Leu Thr Pro Asp Glu  Ser Gln Tyr Leu Ile  Leu Ile Arg
    1040                1045                1050

Asn Lys  Ala Ala His Asn Gln  Phe Pro Cys Ala Ala  Glu Met Pro
    1055                1060                1065

Leu Ile  Tyr Arg Asp Val Ser  Ala Lys Val Gly Ser  Ile Glu Gly
    1070                1075                1080

Ser Ser  Ala Lys Asp Leu Pro  Glu Gly Ser Ser Leu  Val Asp Ser
    1085                1090                1095

Leu Trp  Lys Lys Tyr Glu Met  Ile Ile Arg Lys Ile  Leu Pro Ile
    1100                1105                1110
```

```
Leu Asp  His Glu Asn Arg Phe  Phe Gly Lys Leu Leu  Asn Asn Met
    1115                1120                1125


Ser Gln  Pro Ile Asn Asp Leu
    1130                1135


<210>  124
<211>  1135
<212>  PRT
<213>  Porphyromonas gingivalis

<400>  124

Met Asn Thr Val Pro Ala Ser Glu Asn Lys Gly Gln Ser Arg Thr Val
1               5                   10                  15


Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu
            20                  25                  30


Asn Leu Ile Glu Val Glu Ser His Val Arg Ile Lys Phe Gly Lys Lys
        35                  40                  45


Lys Leu Asn Glu Glu Ser Leu Lys Gln Ser Leu Leu Cys Asp His Leu
    50                  55                  60


Leu Ser Val Asp Arg Trp Thr Lys Val Tyr Gly His Ser Arg Arg Tyr
65                  70                  75                  80


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


His Asp Ser Lys Thr Gly Val Asp Pro Asp Ser Ala Gln Arg Leu Ile
            100                 105                 110


Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His
        115                 120                 125


Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Glu Val Ser Pro Asp
    130                 135                 140
```

```
Ile Ser Ser Phe Ile Thr Gly Thr Tyr Ser Leu Ala Cys Gly Arg Ala
145                 150                 155                 160

Gln Ser Arg Phe Ala Asp Phe Phe Lys Pro Asp Asp Phe Val Leu Ala
                165                 170                 175

Lys Asn Arg Lys Glu Gln Leu Ile Ser Val Ala Asp Gly Lys Glu Cys
            180                 185                 190

Leu Thr Val Ser Gly Leu Ala Phe Phe Ile Cys Leu Phe Leu Asp Arg
        195                 200                 205

Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr
    210                 215                 220

Asp Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys
225                 230                 235                 240

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

Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu
            260                 265                 270

Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu
        275                 280                 285

Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asn Glu Glu
    290                 295                 300

Ser Arg Leu Leu Trp Asp Gly Ser Ser Asp Trp Ala Glu Ala Leu Thr
305                 310                 315                 320

Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe
                325                 330                 335

Ile Glu Glu Met Asp Leu Leu Lys Gly Ile Arg Phe Arg Val Asp Leu
            340                 345                 350
```

```
Gly Glu Ile Glu Leu Asp Ser Tyr Ser Lys Lys Val Gly Arg Asn Gly
        355                 360                 365

Glu Tyr Asp Arg Thr Ile Thr Asp His Ala Leu Ala Phe Gly Lys Leu
    370                 375                 380

Ser Asp Phe Gln Asn Glu Glu Glu Val Ser Arg Met Ile Ser Gly Glu
385                 390                 395                 400

Ala Ser Tyr Pro Val Arg Phe Ser Leu Phe Ala Pro Arg Tyr Ala Ile
                405                 410                 415

Tyr Asp Asn Lys Ile Gly Tyr Cys His Thr Ser Asp Pro Val Tyr Pro
            420                 425                 430

Lys Ser Lys Thr Gly Glu Lys Arg Ala Leu Ser Asn Pro Arg Ser Met
        435                 440                 445

Gly Phe Ile Ser Val His Asp Leu Arg Lys Leu Leu Leu Met Glu Leu
    450                 455                 460

Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Asp Phe Leu Arg Lys
465                 470                 475                 480

Ala Asn Arg Ile Leu Asp Glu Thr Ala Glu Gly Lys Leu Gln Phe Ser
                485                 490                 495

Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro
            500                 505                 510

Lys Ser Lys Asp Arg Arg Glu Lys Ala Glu Thr Thr Leu Glu Lys Tyr
        515                 520                 525

Lys Gln Glu Ile Lys Gly Arg Lys Asp Lys Leu Asn Ser Gln Leu Leu
    530                 535                 540

Ser Ala Phe Asp Met Asp Gln Arg Gln Leu Pro Ser Arg Leu Leu Asp
545                 550                 555                 560
```

```
Glu Trp Met Asn Ile Arg Pro Ala Ser His Ser Val Lys Leu Arg Thr
                565                 570                 575

Tyr Val Lys Gln Leu Asn Glu Asp Cys Arg Leu Arg Leu Gln Lys Phe
            580                 585                 590

Arg Lys Asp Gly Asp Gly Lys Ala Arg Ala Ile Pro Leu Val Gly Glu
        595                 600                 605

Met Ala Thr Phe Leu Ser Gln Asp Ile Val Arg Met Ile Ile Ser Glu
    610                 615                 620

Glu Thr Lys Lys Leu Ile Thr Ser Ala Tyr Tyr Asn Glu Met Gln Arg
625                 630                 635                 640

Ser Leu Ala Gln Tyr Ala Gly Glu Glu Asn Arg His Gln Phe Arg Ala
                645                 650                 655

Ile Val Ala Glu Leu Arg Leu Leu Asp Pro Ser Ser Gly His Pro Phe
            660                 665                 670

Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Asp Phe Tyr
        675                 680                 685

Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Thr Phe Tyr
    690                 695                 700

Arg Pro Glu Gln Asp Glu Asn Thr Lys Arg Arg Ile Ser Val Phe Phe
705                 710                 715                 720

Val Pro Asp Gly Glu Ala Arg Lys Leu Leu Pro Leu Leu Ile Arg Arg
                725                 730                 735

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

Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Val Met
        755                 760                 765
```

```
Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe
    770                 775                 780

Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr
785                 790                 795                 800

Gly Leu Arg Arg Glu Leu Asn Ile His Gly Lys Ser Val Ser Tyr Ile
                805                 810                 815

Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu
            820                 825                 830

Lys Thr Val Arg Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro
        835                 840                 845

Val Pro Pro Asp Leu Ala Ala Tyr Ile Lys Arg Ser Phe His Arg Ala
    850                 855                 860

Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg
865                 870                 875                 880

Leu Met Leu Met Ala Ile Asn Lys Ile Met Thr Asp Arg Glu Glu Asp
                885                 890                 895

Ile Leu Pro Gly Leu Lys Asn Ile Asp Ser Ile Leu Asp Lys Glu Asn
            900                 905                 910

Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu
        915                 920                 925

Gly Gly Asp Asn Ser Leu Ser Leu Val Pro Ala Thr Ile Glu Ile Lys
    930                 935                 940

Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg
945                 950                 955                 960

Arg Val Pro Gly Leu Met Ser His Phe Pro Glu His Lys Ala Thr Leu
                965                 970                 975
```

```
Asp Glu Val Lys Thr Leu Leu Gly Glu Tyr Asp Arg Cys Arg Ile Lys
            980                 985                 990


Ile Phe Asp Trp Ala Phe Ala Leu  Glu Gly Ala Ile Met  Ser Asp Arg
        995                 1000                1005


Asp Leu  Lys Pro Tyr Leu His  Glu Ser Ser Ser Arg  Glu Gly Lys
    1010                1015                1020


Ser Gly  Glu His Ser Thr Leu  Val Lys Met Leu Val  Glu Lys Lys
    1025                1030                1035


Gly Cys  Leu Thr Pro Asp Glu  Ser Gln Tyr Leu Ile  Leu Ile Arg
    1040                1045                1050


Asn Lys  Ala Ala His Asn Gln  Phe Pro Cys Ala Ala  Glu Ile Pro
    1055                1060                1065


Leu Ile  Tyr Arg Asp Val Ser  Ala Lys Val Gly Ser  Ile Glu Gly
    1070                1075                1080


Ser Ser  Ala Lys Asp Leu Pro  Glu Gly Ser Ser Leu  Val Asp Ser
    1085                1090                1095


Leu Trp  Lys Lys Tyr Glu Met  Ile Ile Arg Lys Ile  Leu Pro Ile
    1100                1105                1110


Leu Asp  Pro Glu Asn Arg Phe  Phe Gly Lys Leu Leu  Asn Asn Met
    1115                1120                1125


Ser Gln  Pro Ile Asn Asp Leu
    1130                1135


<210>  125
<211>  1135
<212>  PRT
<213>  Porphyromonas gingivalis

<400>  125
```

```
Met Asn Thr Val Pro Ala Ser Glu Asn Lys Gly Gln Ser Arg Thr Val
1               5                  10                  15

Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu
            20                  25                  30

Asn Leu Ile Glu Val Glu Ser His Val Arg Ile Lys Phe Gly Lys Lys
        35                  40                  45

Lys Leu Asn Glu Glu Ser Leu Lys Gln Ser Leu Leu Cys Asp His Leu
    50                  55                  60

Leu Ser Val Asp Arg Trp Thr Lys Val Tyr Gly His Ser Arg Arg Tyr
65                  70                  75                  80

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

His Asp Ser Lys Thr Gly Val Asp Pro Asp Ser Ala Gln Arg Leu Ile
            100                 105                 110

Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His
        115                 120                 125

Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Glu Val Ser Pro Asp
    130                 135                 140

Ile Ser Ser Phe Ile Thr Gly Thr Tyr Ser Leu Ala Cys Gly Arg Ala
145                 150                 155                 160

Gln Ser Arg Phe Ala Asp Phe Phe Lys Pro Asp Asp Phe Val Leu Ala
                165                 170                 175

Lys Asn Arg Lys Glu Gln Leu Ile Ser Val Ala Asp Gly Lys Glu Cys
            180                 185                 190

Leu Thr Val Ser Gly Leu Ala Phe Phe Ile Cys Leu Phe Leu Asp Arg
        195                 200                 205
```

```
Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr
    210                 215                 220

Asp Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys
225                 230                 235                 240

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

Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu
            260                 265                 270

Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu
        275                 280                 285

Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asn Glu Glu
    290                 295                 300

Ser Arg Leu Leu Trp Asp Gly Ser Ser Asp Trp Ala Glu Ala Leu Thr
305                 310                 315                 320

Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe
                325                 330                 335

Ile Glu Glu Met Asp Leu Leu Lys Gly Ile Arg Phe Arg Val Asp Leu
            340                 345                 350

Gly Glu Ile Glu Leu Asp Ser Tyr Ser Lys Lys Val Gly Arg Asn Gly
        355                 360                 365

Glu Tyr Asp Arg Thr Ile Thr Asp His Ala Leu Ala Phe Gly Lys Leu
    370                 375                 380

Ser Asp Phe Gln Asn Glu Glu Glu Val Ser Arg Met Ile Ser Gly Glu
385                 390                 395                 400

Ala Ser Tyr Pro Val Arg Phe Ser Leu Phe Ala Pro Arg Tyr Ala Ile
                405                 410                 415
```

```
Tyr Asp Asn Lys Ile Gly Tyr Cys His Thr Ser Asp Pro Val Tyr Pro
            420                 425                 430

Lys Ser Lys Thr Gly Glu Lys Arg Ala Leu Ser Asn Pro Arg Ser Met
        435                 440                 445

Gly Phe Ile Ser Val His Asp Leu Arg Lys Leu Leu Leu Met Glu Leu
    450                 455                 460

Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Asp Phe Leu Arg Lys
465                 470                 475                 480

Ala Asn Arg Ile Leu Asp Glu Thr Ala Glu Gly Lys Leu Gln Phe Ser
                485                 490                 495

Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro
            500                 505                 510

Lys Ser Lys Asp Arg Arg Glu Lys Ala Glu Thr Thr Leu Glu Lys Tyr
        515                 520                 525

Lys Gln Glu Ile Lys Gly Arg Lys Asp Lys Leu Asn Ser Gln Leu Leu
    530                 535                 540

Ser Ala Phe Asp Met Asn Gln Arg Gln Leu Pro Ser Arg Leu Leu Asp
545                 550                 555                 560

Glu Trp Met Asn Ile Arg Pro Ala Ser His Ser Val Lys Leu Arg Thr
                565                 570                 575

Tyr Val Lys Gln Leu Asn Glu Asp Cys Arg Leu Arg Leu Arg Lys Phe
            580                 585                 590

Arg Lys Asp Gly Asp Gly Lys Ala Arg Ala Ile Pro Leu Val Gly Glu
        595                 600                 605

Met Ala Thr Phe Leu Ser Gln Asp Ile Val Arg Met Ile Ile Ser Glu
    610                 615                 620
```

```
Glu Thr Lys Lys Leu Ile Thr Ser Ala Tyr Tyr Asn Glu Met Gln Arg
625                 630                 635                 640


Ser Leu Ala Gln Tyr Ala Gly Glu Glu Asn Arg Arg Gln Phe Arg Ala
                645                 650                 655


Ile Val Ala Glu Leu His Leu Leu Asp Pro Ser Ser Gly His Pro Phe
            660                 665                 670


Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Asp Phe Tyr
        675                 680                 685


Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Thr Phe Tyr
    690                 695                 700


Arg Pro Glu Gln Asp Glu Asn Thr Lys Arg Arg Ile Ser Val Phe Phe
705                 710                 715                 720


Val Pro Asp Gly Glu Ala Arg Lys Leu Leu Pro Leu Leu Ile Arg Arg
                725                 730                 735


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


Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Ile Met
        755                 760                 765


Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe
    770                 775                 780


Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr
785                 790                 795                 800


Gly Leu Arg Arg Glu Leu Asn Ile His Gly Lys Ser Val Ser Tyr Ile
                805                 810                 815


Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu
            820                 825                 830
```

```
Lys Thr Val Gln Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro
        835                 840                 845


Val Pro Pro Asp Leu Ala Ala Asp Ile Lys Arg Ser Phe His Arg Ala
    850                 855                 860


Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg
865                 870                 875                 880


Leu Met Leu Met Ala Ile Asn Lys Met Met Thr Asp Arg Glu Glu Asp
                885                 890                 895


Ile Leu Pro Gly Leu Lys Asn Ile Asp Ser Ile Leu Asp Glu Glu Asn
            900                 905                 910


Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu
        915                 920                 925


Gly Gly Asp Asn Ser Leu Ser Leu Val Pro Ala Thr Ile Glu Ile Lys
    930                 935                 940


Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg
945                 950                 955                 960


Arg Val Pro Gly Leu Met Ser His Phe Pro Glu His Lys Ala Thr Leu
                965                 970                 975


Asp Glu Val Lys Thr Leu Leu Gly Glu Tyr Asp Arg Cys Arg Ile Lys
            980                 985                 990


Ile Phe Asp Trp Ala Phe Ala Leu  Glu Gly Ala Ile Met  Ser Asp Arg
        995                 1000                1005


Asp Leu  Lys Pro Tyr Leu His  Glu Ser Ser Ser Arg  Glu Gly Lys
    1010                1015                1020


Ser Gly  Glu His Ser Thr Leu  Val Lys Met Leu Val  Glu Lys Lys
    1025                1030                1035
```

```
Gly Cys  Leu Thr Pro Asp Glu  Ser Gln Tyr Leu Ile  Leu Ile Arg
    1040                1045                1050


Asn Lys  Ala Ala His Asn Gln  Phe Pro Cys Ala Ala  Glu Met Pro
    1055                1060                1065


Leu Ile  Tyr Arg Asp Val Ser  Ala Lys Val Gly Ser  Ile Glu Gly
    1070                1075                1080


Ser Ser  Ala Lys Asp Leu Pro  Glu Gly Ser Ser Leu  Val Asp Ser
    1085                1090                1095


Leu Trp  Lys Lys Tyr Glu Met  Ile Ile Arg Lys Ile  Leu Pro Ile
    1100                1105                1110


Leu Asp  Pro Glu Asn Arg Phe  Phe Gly Lys Leu Leu  Asn Asn Met
    1115                1120                1125


Ser Gln  Pro Ile Asn Asp Leu
    1130                1135


<210>  126
<211>  1133
<212>  PRT
<213>  Prevotella falsenii

<400>  126

Met Lys Asn Asp Asn Asn Ser Thr Lys Ser Thr Asp Tyr Thr Leu Gly
1               5                   10                  15


Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn Val
            20                  25                  30


Tyr Ile Thr Val Asn His Ile Asn Lys Val Leu Glu Leu Lys Asn Lys
        35                  40                  45


Lys Asp Gln Glu Ile Ile Ile Asp Asn Asp Gln Asp Ile Leu Ala Ile
    50                  55                  60


Lys Thr Leu Trp Gly Lys Val Asp Thr Asp Ile Asn Lys Lys Asp Arg
```

65 70 75 80

Leu Arg Glu Leu Ile Met Lys His Phe Pro Phe Leu Glu Ala Ala Thr
85 90 95

Tyr Gln Gln Ser Ser Thr Asn Asn Thr Lys Gln Lys Glu Glu Glu Gln
100 105 110

Ala Lys Ala Gln Ser Phe Glu Ser Leu Lys Asp Cys Leu Phe Leu Phe
115 120 125

Leu Glu Lys Leu Arg Glu Ala Arg Asn Tyr Tyr Ser His Tyr Lys His
130 135 140

Ser Lys Ser Leu Glu Glu Pro Lys Leu Glu Glu Lys Leu Leu Glu Asn
145 150 155 160

Met Tyr Asn Ile Phe Asp Thr Asn Val Gln Leu Val Ile Lys Asp Tyr
165 170 175

Glu His Asn Lys Asp Ile Asn Pro Glu Glu Asp Phe Lys His Leu Gly
180 185 190

Arg Ala Glu Gly Glu Phe Asn Tyr Tyr Phe Thr Arg Asn Lys Lys Gly
195 200 205

Asn Ile Thr Glu Ser Gly Leu Leu Phe Phe Val Ser Leu Phe Leu Glu
210 215 220

Lys Lys Asp Ala Ile Trp Ala Gln Thr Lys Ile Lys Gly Phe Lys Asp
225 230 235 240

Asn Arg Glu Asn Lys Gln Lys Met Thr His Glu Val Phe Cys Arg Ser
245 250 255

Arg Met Leu Leu Pro Lys Leu Arg Leu Glu Ser Thr Gln Thr Gln Asp
260 265 270

Trp Ile Leu Leu Asp Met Leu Asn Glu Leu Ile Arg Cys Pro Lys Ser

```
        275                 280                 285

Leu Tyr Lys Arg Leu Gln Gly Glu Lys Arg Glu Lys Phe Arg Val Pro
    290                 295                 300

Phe Asp Pro Ala Asp Glu Asp Tyr Asp Ala Glu Gln Glu Pro Phe Lys
305                 310                 315                 320

Asn Thr Leu Val Arg His Gln Asp Arg Phe Pro Tyr Phe Ala Leu Arg
                325                 330                 335

Tyr Phe Asp Tyr Asn Glu Ile Phe Thr Asn Leu Arg Phe Gln Ile Asp
            340                 345                 350

Leu Gly Thr Tyr His Phe Ser Ile Tyr Lys Lys Gln Ile Gly Asp Lys
        355                 360                 365

Lys Glu Asp Arg His Leu Thr His Lys Leu Tyr Gly Phe Glu Arg Ile
    370                 375                 380

Gln Glu Phe Ala Lys Glu Asn Arg Pro Asp Glu Trp Lys Ala Leu Val
385                 390                 395                 400

Lys Asp Leu Asp Thr Phe Glu Glu Ser Asn Glu Pro Tyr Ile Ser Glu
                405                 410                 415

Thr Thr Pro His Tyr His Leu Glu Asn Gln Lys Ile Gly Ile Arg Asn
            420                 425                 430

Lys Asn Lys Lys Lys Lys Lys Thr Ile Trp Pro Ser Leu Glu Thr Lys
        435                 440                 445

Thr Thr Val Asn Glu Arg Ser Lys Tyr Asn Leu Gly Lys Ser Phe Lys
    450                 455                 460

Ala Glu Ala Phe Leu Ser Val His Glu Leu Leu Pro Met Met Phe Tyr
465                 470                 475                 480

Tyr Leu Leu Leu Asn Lys Glu Glu Pro Asn Asn Gly Lys Ile Asn Ala
```

485 490 495

Ser Lys Val Glu Gly Ile Ile Glu Lys Lys Ile Arg Asp Ile Tyr Lys
500 505 510

Leu Tyr Gly Ala Phe Ala Asn Glu Glu Ile Asn Asn Glu Glu Glu Leu
515 520 525

Lys Glu Tyr Cys Glu Gly Lys Asp Ile Ala Ile Arg His Leu Pro Lys
530 535 540

Gln Met Ile Ala Ile Leu Lys Asn Glu Tyr Lys Asp Met Ala Lys Lys
545 550 555 560

Ala Glu Asp Lys Gln Lys Lys Met Ile Lys Asp Thr Lys Lys Arg Leu
565 570 575

Ala Ala Leu Asp Lys Gln Val Lys Gly Glu Val Glu Asp Gly Gly Arg
580 585 590

Asn Ile Lys Pro Leu Lys Ser Gly Arg Ile Ala Ser Trp Leu Val Asn
595 600 605

Asp Met Met Arg Phe Gln Pro Val Gln Arg Asp Arg Asp Gly Tyr Pro
610 615 620

Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Gln Leu Leu Gln Arg
625 630 635 640

Thr Leu Ala Leu Phe Gly Ser Glu Arg Glu Arg Leu Ala Pro Tyr Phe
645 650 655

Arg Gln Met Asn Leu Ile Gly Lys Asp Asn Pro His Pro Phe Leu Lys
660 665 670

Asp Thr Lys Trp Lys Glu His Asn Asn Ile Leu Ser Phe Tyr Arg Ser
675 680 685

Tyr Leu Glu Ala Lys Lys Asn Phe Leu Gly Ser Leu Lys Pro Glu Asp

```
    690                 695                 700

Trp Lys Lys Asn Gln Tyr Phe Leu Lys Leu Lys Glu Pro Lys Thr Asn
705                 710                 715                 720

Arg Glu Thr Leu Val Gln Gly Trp Lys Asn Gly Phe Asn Leu Pro Arg
                725                 730                 735

Gly Ile Phe Thr Glu Pro Ile Arg Glu Trp Phe Ile Arg His Gln Asn
            740                 745                 750

Glu Ser Glu Glu Tyr Lys Lys Val Lys Asp Phe Asp Arg Ile Gly Leu
        755                 760                 765

Val Ala Lys Val Ile Pro Leu Phe Phe Lys Glu Asp Tyr Gln Lys Glu
    770                 775                 780

Ile Glu Asp Tyr Val Gln Pro Phe Tyr Gly Tyr Pro Phe Asn Val Gly
785                 790                 795                 800

Asn Ile His Asn Ser Gln Glu Gly Thr Phe Leu Asn Lys Lys Glu Arg
                805                 810                 815

Glu Glu Leu Trp Lys Gly Asn Lys Thr Lys Phe Lys Asp Tyr Lys Thr
            820                 825                 830

Lys Glu Lys Asn Lys Glu Lys Thr Asn Lys Asp Lys Phe Lys Lys Lys
        835                 840                 845

Thr Asp Glu Glu Lys Glu Glu Phe Arg Ser Tyr Leu Asp Phe Gln Ser
    850                 855                 860

Trp Lys Lys Phe Glu Arg Glu Leu Arg Leu Val Arg Asn Gln Asp Ile
865                 870                 875                 880

Val Thr Trp Leu Leu Cys Met Glu Leu Ile Asp Lys Leu Lys Ile Asp
                885                 890                 895

Glu Leu Asn Ile Glu Glu Leu Gln Lys Leu Arg Leu Lys Asp Ile Asp
```

900 905 910

Thr Asp Thr Ala Lys Lys Glu Lys Asn Asn Ile Leu Asn Arg Ile Met
915 920 925

Pro Met Glu Leu Pro Val Thr Val Tyr Glu Thr Asp Asp Ser Asn Asn
930 935 940

Ile Ile Lys Asp Lys Pro Leu His Thr Ile Tyr Ile Lys Glu Ala Glu
945 950 955 960

Thr Lys Leu Leu Lys Gln Gly Asn Phe Lys Ala Leu Val Lys Asp Arg
965 970 975

Arg Leu Asn Gly Leu Phe Ser Phe Val Glu Thr Ser Ser Glu Ala Glu
980 985 990

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

Glu Tyr Gln Arg Ala Arg Val Glu Ile Ile Lys Asp Met Leu Arg
1010 1015 1020

Leu Glu Glu Thr Leu Ile Gly Asn Asp Glu Lys Leu Pro Thr Asn
1025 1030 1035

Lys Phe Arg Gln Met Leu Asp Lys Trp Leu Glu His Lys Lys Glu
1040 1045 1050

Thr Asp Asp Thr Asp Leu Lys Asn Asp Val Lys Leu Leu Thr Glu
1055 1060 1065

Val Arg Asn Ala Phe Ser His Asn Gln Tyr Pro Met Arg Asp Arg
1070 1075 1080

Ile Ala Phe Ala Asn Ile Lys Pro Phe Ser Leu Ser Ser Ala Asn
1085 1090 1095

Thr Ser Asn Glu Glu Gly Leu Gly Ile Ala Lys Lys Leu Lys Asp

1100 1105 1110

Lys Thr Lys Glu Thr Ile Asp Arg Ile Ile Glu Ile Glu Glu Gln
1115 1120 1125

Thr Ala Thr Lys Arg
1130

<210> 127
<211> 1116
<212> PRT
<213> Prevotella pleuritidis

<400> 127

Met Glu Asn Asp Lys Arg Leu Glu Glu Ser Thr Cys Tyr Thr Leu Asn
1 5 10 15

Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn Val
20 25 30

Tyr Ile Thr Ile Asn His Ile Asn Lys Leu Leu Glu Ile Arg Gln Ile
35 40 45

Asp Asn Asp Glu Lys Val Leu Asp Ile Lys Ala Leu Trp Gln Lys Val
50 55 60

Asp Lys Asp Ile Asn Gln Lys Ala Arg Leu Arg Glu Leu Met Ile Lys
65 70 75 80

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

Lys Glu Glu Val Lys Glu Glu Lys Gln Ala Lys Ala Gln Ser Phe Lys
100 105 110

Ser Leu Lys Asp Cys Leu Phe Leu Phe Leu Glu Lys Leu Gln Glu Ala
115 120 125

Arg Asn Tyr Tyr Ser His Tyr Lys Ser Ser Glu Ser Ser Lys Glu Pro
130 135 140

```
Glu Phe Glu Glu Gly Leu Leu Glu Lys Met Tyr Asn Thr Phe Gly Val
145                 150                 155                 160

Ser Ile Arg Leu Val Lys Glu Asp Tyr Gln Tyr Asn Lys Asp Ile Asp
                165                 170                 175

Pro Glu Lys Asp Phe Lys His Leu Glu Arg Lys Glu Asp Phe Asn Tyr
            180                 185                 190

Leu Phe Thr Asp Lys Asp Asn Lys Gly Lys Ile Thr Lys Asn Gly Leu
        195                 200                 205

Leu Phe Phe Val Ser Leu Phe Leu Glu Lys Lys Asp Ala Ile Trp Met
    210                 215                 220

Gln Gln Lys Leu Arg Gly Phe Lys Asp Asn Arg Gly Asn Lys Glu Lys
225                 230                 235                 240

Met Thr His Glu Val Phe Cys Arg Ser Arg Met Leu Leu Pro Lys Ile
                245                 250                 255

Arg Leu Glu Ser Thr Gln Thr Gln Asp Trp Ile Leu Leu Asp Met Leu
            260                 265                 270

Asn Glu Leu Ile Arg Cys Pro Lys Ser Leu Tyr Glu Arg Leu Gln Gly
        275                 280                 285

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

Tyr Asp Ala Glu Gln Glu Pro Phe Arg Asn Thr Leu Val Arg His Gln
305                 310                 315                 320

Asp Arg Phe Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Tyr Asn Glu Ile
                325                 330                 335

Phe Lys Asn Leu Arg Phe Gln Ile Asp Leu Gly Thr Tyr His Phe Ser
            340                 345                 350
```

```
Ile Tyr Lys Lys Leu Ile Gly Asp Asn Lys Glu Asp Arg His Leu Thr
        355                 360                 365

His Lys Leu Tyr Gly Phe Glu Arg Ile Gln Glu Phe Ala Lys Gln Lys
    370                 375                 380

Arg Pro Asn Glu Trp Gln Ala Leu Val Lys Asp Leu Asp Ile Tyr Glu
385                 390                 395                 400

Thr Ser Asn Glu Gln Tyr Ile Ser Glu Thr Thr Pro His Tyr His Leu
                405                 410                 415

Glu Asn Gln Lys Ile Gly Ile Arg Phe Lys Asn Lys Lys Asp Lys Ile
            420                 425                 430

Trp Pro Ser Leu Glu Thr Asn Gly Lys Glu Asn Glu Lys Ser Lys Tyr
        435                 440                 445

Asn Leu Asp Lys Ser Phe Gln Ala Glu Ala Phe Leu Ser Ile His Glu
    450                 455                 460

Leu Leu Pro Met Met Phe Tyr Asp Leu Leu Leu Lys Lys Glu Glu Pro
465                 470                 475                 480

Asn Asn Asp Glu Lys Asn Ala Ser Ile Val Glu Gly Phe Ile Lys Lys
                485                 490                 495

Glu Ile Lys Arg Met Tyr Ala Ile Tyr Asp Ala Phe Ala Asn Glu Glu
            500                 505                 510

Ile Asn Ser Lys Glu Gly Leu Glu Glu Tyr Cys Lys Asn Lys Gly Phe
        515                 520                 525

Gln Glu Arg His Leu Pro Lys Gln Met Ile Ala Ile Leu Thr Asn Lys
    530                 535                 540

Ser Lys Asn Met Ala Glu Lys Ala Lys Arg Lys Gln Lys Glu Met Ile
545                 550                 555                 560
```

```
Lys Asp Thr Lys Lys Arg Leu Ala Thr Leu Asp Lys Gln Val Lys Gly
                565                 570                 575

Glu Ile Glu Asp Gly Gly Arg Asn Ile Arg Leu Leu Lys Ser Gly Glu
            580                 585                 590

Ile Ala Arg Trp Leu Val Asn Asp Met Met Arg Phe Gln Ser Val Gln
        595                 600                 605

Lys Asp Lys Glu Gly Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr
    610                 615                 620

Glu Tyr Gln Met Leu Gln Arg Ser Leu Ala Leu Tyr Asn Lys Glu Gln
625                 630                 635                 640

Lys Pro Thr Pro Tyr Phe Ile Gln Val Asn Leu Ile Lys Ser Ser Asn
                645                 650                 655

Pro His Pro Phe Leu Glu Glu Thr Lys Trp Glu Glu Cys Asn Asn Ile
            660                 665                 670

Leu Ser Phe Tyr Arg Ser Tyr Leu Glu Ala Lys Lys Asn Phe Leu Glu
        675                 680                 685

Ser Leu Lys Pro Glu Asp Trp Lys Lys Asn Gln Tyr Phe Leu Met Leu
    690                 695                 700

Lys Glu Pro Lys Thr Asn Arg Lys Thr Leu Val Gln Gly Trp Lys Asn
705                 710                 715                 720

Gly Phe Asn Leu Pro Arg Gly Ile Phe Thr Glu Pro Ile Lys Glu Trp
                725                 730                 735

Phe Lys Arg His Gln Asn Asp Ser Glu Glu Tyr Lys Lys Val Glu Ala
            740                 745                 750

Leu Asp Arg Val Gly Leu Val Ala Lys Val Ile Pro Leu Phe Phe Lys
        755                 760                 765
```

```
Glu Glu Tyr Phe Lys Glu Asp Ala Gln Lys Glu Ile Asn Asn Cys Val
    770                 775                 780

Gln Pro Phe Tyr Ser Phe Pro Tyr Asn Val Gly Asn Ile His Lys Pro
785                 790                 795                 800

Glu Glu Lys Asn Phe Leu His Cys Glu Glu Arg Arg Lys Leu Trp Asp
                805                 810                 815

Lys Lys Lys Asp Lys Phe Lys Gly Tyr Lys Ala Lys Glu Lys Ser Lys
            820                 825                 830

Lys Met Thr Asp Lys Glu Lys Glu Glu His Arg Ser Tyr Leu Glu Phe
        835                 840                 845

Gln Ser Trp Asn Lys Phe Glu Arg Glu Leu Arg Leu Val Arg Asn Gln
    850                 855                 860

Asp Ile Val Thr Trp Leu Leu Cys Thr Glu Leu Ile Asp Lys Leu Lys
865                 870                 875                 880

Ile Asp Glu Leu Asn Ile Glu Glu Leu Gln Lys Leu Arg Leu Lys Asp
                885                 890                 895

Ile Asp Thr Asp Thr Ala Lys Lys Glu Lys Asn Asn Ile Leu Asn Arg
            900                 905                 910

Ile Met Pro Met Gln Leu Pro Val Thr Val Tyr Glu Ile Asp Lys Ser
        915                 920                 925

Phe Asn Ile Val Lys Asp Lys Pro Leu His Thr Ile Tyr Ile Glu Glu
    930                 935                 940

Thr Gly Thr Lys Leu Leu Lys Gln Gly Asn Phe Lys Ala Leu Val Lys
945                 950                 955                 960

Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe Val Lys Thr Ser Ser Glu
                965                 970                 975
```

```
Ala Glu Ser Lys Ser Lys Pro Ile Ser Lys Leu Arg Val Glu Tyr Glu
            980                 985                 990


Leu Gly Ala Tyr Gln Lys Ala Arg  Ile Asp Ile Ile Lys  Asp Met Leu
        995                 1000                1005


Ala Leu  Glu Lys Thr Leu Ile  Asp Asn Asp Glu Asn  Leu Pro Thr
    1010                1015                1020


Asn Lys  Phe Ser Asp Met Leu  Lys Ser Trp Leu Lys  Gly Lys Gly
    1025                1030                1035


Glu Ala  Asn Lys Ala Arg Leu  Gln Asn Asp Val Asp  Leu Leu Val
    1040                1045                1050


Ala Ile  Arg Asn Ala Phe Ser  His Asn Gln Tyr Pro  Met Tyr Asn
    1055                1060                1065


Ser Glu  Val Phe Lys Gly Met  Lys Leu Leu Ser Leu  Ser Ser Asp
    1070                1075                1080


Ile Pro  Glu Lys Glu Gly Leu  Gly Ile Ala Lys Gln  Leu Lys Asp
    1085                1090                1095


Lys Ile  Lys Glu Thr Ile Glu  Arg Ile Ile Glu Ile  Glu Lys Glu
    1100                1105                1110


Ile Arg  Asn
    1115


<210>  128
<211>  1119
<212>  PRT
<213>  Porphyromonas gingivalis

<400>  128

Met Thr Glu Gln Asn Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1               5                  10                  15
```

```
Glu Asp Lys His Phe Trp Ala Ala Phe Phe Asn Leu Ala Arg His Asn
            20                  25                  30

Ala Tyr Ile Thr Leu Ala His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
        35                  40                  45

Ala Asp Ile Thr Asn Asp Glu Asp Ile Leu Phe Phe Lys Gly Gln Trp
    50                  55                  60

Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ala Arg Leu Arg Ser Leu
65                  70                  75                  80

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

Leu Phe Glu Ser Gln Ser Ser Gly Asn Lys Ser Ser Lys Lys Lys Glu
            100                 105                 110

Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
        115                 120                 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
    130                 135                 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Pro Glu Ser Ser Glu Leu
145                 150                 155                 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
                165                 170                 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val
            180                 185                 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
        195                 200                 205

Arg Tyr Gly Asn Asn Asp Asn Pro Phe Phe Lys His His Phe Val Asp
    210                 215                 220
```

```
Arg Glu Gly Thr Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225                 230                 235                 240


Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
                245                 250                 255


Phe Lys Gly Gly Thr Glu Ala Tyr Gln Gln Met Thr Asn Glu Val Phe
            260                 265                 270


Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
        275                 280                 285


Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys
    290                 295                 300


Pro Lys Ser Leu Tyr Asp Arg Leu Arg Glu Glu Asp Arg Ala Arg Phe
305                 310                 315                 320


Arg Val Pro Ile Asp Ile Leu Ser Asp Glu Asp Asp Thr Asp Gly Thr
                325                 330                 335


Glu Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
            340                 345                 350


Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
        355                 360                 365


Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
    370                 375                 380


Lys Asn Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385                 390                 395                 400


Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
                405                 410                 415


Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
            420                 425                 430
```

```
Lys Pro Tyr Ile Thr Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly
        435                 440                 445

Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln His Leu Trp Pro Ser
    450                 455                 460

Pro Glu Val Gly Ala Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465                 470                 475                 480

Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
                485                 490                 495

Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val
            500                 505                 510

Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
        515                 520                 525

Tyr Ala Val Tyr Asp Ala Phe Ala Arg Gly Glu Ile Asp Thr Leu Asp
    530                 535                 540

Arg Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu
545                 550                 555                 560

Pro Arg Gln Met Ile Ala Ile Leu Ser Gln Glu His Lys Asp Met Glu
                565                 570                 575

Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His
            580                 585                 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly
        595                 600                 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
    610                 615                 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly
625                 630                 635                 640
```

```
Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
                645                 650                 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro
            660                 665                 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
        675                 680                 685

Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
    690                 695                 700

Arg Ser Tyr Leu Lys Ala Arg Lys Ala Phe Leu Gln Ser Ile Gly Arg
705                 710                 715                 720

Ser Asp Arg Glu Glu Asn His Arg Phe Leu Leu Leu Lys Glu Pro Lys
                725                 730                 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Ser Glu Phe His Leu
            740                 745                 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
        755                 760                 765

Gly Tyr Asp Glu Val Gly Ser Tyr Lys Glu Val Gly Phe Met Ala Lys
    770                 775                 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Lys Asp Arg Val Gln Pro
785                 790                 795                 800

Phe Tyr Asp Tyr Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
                805                 810                 815

Gly Arg Phe Leu Ser Lys Glu Lys Arg Ala Glu Glu Trp Glu Ser Gly
            820                 825                 830

Lys Glu Arg Phe Arg Leu Ala Lys Leu Lys Lys Glu Ile Leu Glu Ala
        835                 840                 845
```

```
Lys Glu His Pro Tyr Leu Asp Phe Lys Ser Trp Gln Lys Phe Glu Arg
    850                 855                 860

Glu Leu Arg Leu Val Lys Asn Gln Asp Ile Ile Thr Trp Met Met Cys
865                 870                 875                 880

Arg Asp Leu Met Glu Glu Asn Lys Val Glu Gly Leu Asp Thr Gly Thr
                885                 890                 895

Leu Tyr Leu Lys Asp Ile Arg Thr Asp Val His Glu Gln Gly Ser Leu
            900                 905                 910

Asn Val Leu Asn Arg Val Lys Pro Met Arg Leu Pro Val Val Val Tyr
        915                 920                 925

Arg Ala Asp Ser Arg Gly His Val His Lys Glu Gln Ala Pro Leu Ala
    930                 935                 940

Thr Val Tyr Ile Glu Glu Arg Asp Thr Lys Leu Leu Lys Gln Gly Asn
945                 950                 955                 960

Phe Lys Ser Phe Val Lys Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe
                965                 970                 975

Val Asp Thr Gly Ala Leu Ala Met Glu Gln Tyr Pro Ile Ser Lys Leu
            980                 985                 990

Arg Val Glu Tyr Glu Leu Ala Lys  Tyr Gln Thr Ala Arg  Val Cys Ala
        995                 1000                1005

Phe Glu  Gln Thr Leu Glu Leu  Glu Glu Ser Leu Leu  Thr Arg Tyr
    1010                1015                1020

Pro His  Leu Pro Asp Lys Asn  Phe Arg Lys Met Leu  Glu Ser Trp
    1025                1030                1035

Ser Asp  Pro Leu Leu Asp Lys  Trp Pro Asp Leu His  Arg Lys Val
    1040                1045                1050
```

```
Arg Leu  Leu Ile Ala Val Arg  Asn Ala Phe Ser His  Asn Gln Tyr
    1055                1060                1065


Pro Met  Tyr Asp Glu Ala Val  Phe Ser Ser Ile Arg  Lys Tyr Asp
    1070                1075                1080


Pro Ser  Ser Pro Asp Ala Ile  Glu Glu Arg Met Gly  Leu Asn Ile
    1085                1090                1095


Ala His  Arg Leu Ser Glu Glu  Val Lys Gln Ala Lys  Glu Met Ala
    1100                1105                1110


Glu Arg  Ile Ile Gln Val
    1115


<210>  129
<211>  1175
<212>  PRT
<213>  Porphyromonas gulae

<400>  129

Met Thr Glu Gln Ser Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1               5                   10                  15


Glu Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn
            20                  25                  30


Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
        35                  40                  45


Ala Asp Ile Thr Asn Asp Gln Asp Val Leu Ser Phe Lys Ala Leu Trp
    50                  55                  60


Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ser Arg Leu Arg Ser Leu
65                  70                  75                  80


Ile Leu Lys His Phe Ser Phe Leu Glu Gly Ala Ala Tyr Gly Lys Lys
                85                  90                  95
```

```
Leu Phe Glu Ser Lys Ser Ser Gly Asn Lys Ser Ser Lys Asn Lys Glu
            100                 105                 110


Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
        115                 120                 125


Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
    130                 135                 140


Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Ser Gly Ser Ser Glu Leu
145                 150                 155                 160


Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
                165                 170                 175


Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val
            180                 185                 190


Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
        195                 200                 205


Arg Tyr Gly His Asn Asp Asn Pro Ser Phe Lys His His Phe Val Asp
    210                 215                 220


Ser Glu Gly Met Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225                 230                 235                 240


Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
                245                 250                 255


Phe Lys Gly Gly Thr Glu Thr Tyr Gln Gln Met Thr Asn Glu Val Phe
            260                 265                 270


Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
        275                 280                 285


Met Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys
    290                 295                 300
```

```
Pro Lys Pro Leu Tyr Asp Arg Leu Arg Glu Asp Asp Arg Ala Cys Phe
305                 310                 315                 320


Arg Val Pro Val Asp Ile Leu Pro Asp Glu Asp Asp Thr Asp Gly Gly
                325                 330                 335


Gly Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
            340                 345                 350


Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
        355                 360                 365


Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
    370                 375                 380


Lys Met Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385                 390                 395                 400


Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
                405                 410                 415


Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
            420                 425                 430


Lys Pro Tyr Ile Ser Gln Thr Ser Pro His Tyr His Ile Glu Lys Gly
        435                 440                 445


Lys Ile Gly Leu Arg Phe Met Pro Glu Gly Gln His Leu Trp Pro Ser
    450                 455                 460


Pro Glu Val Gly Thr Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465                 470                 475                 480


Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
                485                 490                 495


Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val
            500                 505                 510
```

```
Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
        515                 520                 525

Tyr Ala Ile Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Leu Lys
    530                 535                 540

Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu
545                 550                 555                 560

Pro Lys Gln Met Ile Ala Ile Leu Ser Gln Glu His Lys Asn Met Glu
                565                 570                 575

Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His
            580                 585                 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly
        595                 600                 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
    610                 615                 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Ala Ser Gly
625                 630                 635                 640

Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
                645                 650                 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro
            660                 665                 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
        675                 680                 685

Leu His Asp Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
    690                 695                 700

Arg Ser Tyr Leu Arg Ala Arg Lys Ala Phe Leu Glu Arg Ile Gly Arg
705                 710                 715                 720
```

```
Ser Asp Arg Met Glu Asn Arg Pro Phe Leu Leu Leu Lys Glu Pro Lys
                725                 730                 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Ser Glu Phe His Leu
            740                 745                 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
        755                 760                 765

Gly Tyr Asp Glu Val Gly Ser Tyr Arg Glu Val Gly Phe Met Ala Lys
    770                 775                 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Glu Asp Arg Val Gln Pro
785                 790                 795                 800

Phe Tyr Asp Ser Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
                805                 810                 815

Gly Arg Phe Leu Ser Lys Glu Glu Arg Ala Glu Glu Trp Glu Arg Gly
            820                 825                 830

Lys Glu Arg Phe Arg Asp Leu Glu Ala Trp Ser His Ser Ala Ala Arg
        835                 840                 845

Arg Ile Glu Asp Ala Phe Ala Gly Ile Glu Tyr Ala Ser Pro Gly Asn
    850                 855                 860

Lys Lys Lys Ile Glu Gln Leu Leu Arg Asp Leu Ser Leu Trp Glu Ala
865                 870                 875                 880

Phe Glu Ser Lys Leu Lys Val Arg Ala Asp Lys Ile Asn Leu Ala Lys
                885                 890                 895

Leu Lys Lys Glu Ile Leu Glu Ala Gln Glu His Pro Tyr His Asp Phe
            900                 905                 910

Lys Ser Trp Gln Lys Phe Glu Arg Glu Leu Arg Leu Val Lys Asn Gln
        915                 920                 925
```

```
Asp Ile Ile Thr Trp Met Met Cys Arg Asp Leu Met Glu Glu Asn Lys
    930                 935                 940

Val Glu Gly Leu Asp Thr Gly Thr Leu Tyr Leu Lys Asp Ile Arg Thr
945                 950                 955                 960

Asn Val Gln Glu Gln Gly Ser Leu Asn Val Leu Asn His Val Lys Pro
                965                 970                 975

Met Arg Leu Pro Val Val Val Tyr Arg Ala Asp Ser Arg Gly His Val
            980                 985                 990

His Lys Glu Glu Ala Pro Leu Ala  Thr Val Tyr Ile Glu  Glu Arg Asp
        995                 1000                1005

Thr Lys  Leu Leu Lys Gln Gly  Asn Phe Lys Ser Phe  Val Lys Asp
    1010                1015                1020

Arg Arg  Leu Asn Gly Leu Phe  Ser Phe Val Asp Thr  Gly Gly Leu
    1025                1030                1035

Ala Met  Glu Gln Tyr Pro Ile  Ser Lys Leu Arg Val  Glu Tyr Glu
    1040                1045                1050

Leu Ala  Lys Tyr Gln Thr Ala  Arg Val Cys Ala Phe  Glu Gln Thr
    1055                1060                1065

Leu Glu  Leu Glu Glu Ser Leu  Leu Thr Arg Tyr Pro  His Leu Pro
    1070                1075                1080

Asp Lys  Asn Phe Arg Lys Met  Leu Glu Ser Trp Ser  Asp Pro Leu
    1085                1090                1095

Leu Ala  Lys Trp Pro Glu Leu  His Gly Lys Val Arg  Leu Leu Ile
    1100                1105                1110

Ala Val  Arg Asn Ala Phe Ser  His Asn Gln Tyr Pro  Met Tyr Asp
    1115                1120                1125
```

```
Glu Ala  Val Phe Ser Ser Ile  Arg Lys Tyr Asp Pro  Ser Ser Pro
    1130                1135                1140


Asp Ala  Ile Glu Glu Arg Met  Gly Leu Asn Ile Ala  His Arg Leu
    1145                1150                1155


Ser Glu  Glu Val Lys Gln Ala  Lys Glu Thr Val Glu  Arg Ile Ile
    1160                1165                1170


Gln Ala
    1175


<210>  130
<211>  1119
<212>  PRT
<213>  Porphyromonas gulae

<400>  130

Met Thr Glu Gln Ser Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1               5                   10                  15


Glu Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn
            20                  25                  30


Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
        35                  40                  45


Ala Asp Ile Thr Asn Asp Gln Asp Val Leu Ser Phe Lys Ala Leu Trp
    50                  55                  60


Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ser Arg Leu Arg Ser Leu
65                  70                  75                  80


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


Leu Phe Glu Ser Lys Ser Ser Gly Asn Lys Ser Ser Lys Asn Lys Glu
            100                 105                 110


Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
```

```
        115                 120                 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
    130                 135                 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Ser Gly Ser Ser Glu Leu
145                 150                 155                 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
                165                 170                 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val
            180                 185                 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
        195                 200                 205

Arg Tyr Gly His Asn Asp Asn Pro Ser Phe Lys His His Phe Val Asp
    210                 215                 220

Gly Glu Gly Met Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225                 230                 235                 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
                245                 250                 255

Phe Lys Gly Gly Thr Glu Thr Tyr Gln Gln Met Thr Asn Glu Val Phe
            260                 265                 270

Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
        275                 280                 285

Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys
    290                 295                 300

Pro Lys Pro Leu Tyr Asp Arg Leu Arg Glu Lys Asp Arg Ala Arg Phe
305                 310                 315                 320

Arg Val Pro Val Asp Ile Leu Pro Asp Glu Asp Asp Thr Asp Gly Gly
```

```
                325                 330                 335

Gly Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
            340                 345                 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
        355                 360                 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
    370                 375                 380

Lys Val Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385                 390                 395                 400

Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
                405                 410                 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
            420                 425                 430

Lys Pro Tyr Ile Ser Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly
        435                 440                 445

Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln His Leu Trp Pro Ser
    450                 455                 460

Pro Glu Val Gly Thr Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465                 470                 475                 480

Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
                485                 490                 495

Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val
            500                 505                 510

Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
        515                 520                 525

Tyr Ala Ile Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Arg Asp
```

```
    530                 535                 540

Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu
545                 550                 555                 560

Pro Lys Gln Met Ile Gly Ile Leu Ser Gln Glu His Lys Asn Met Glu
                565                 570                 575

Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His
            580                 585                 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly
        595                 600                 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
    610                 615                 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly
625                 630                 635                 640

Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
                645                 650                 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro
            660                 665                 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
        675                 680                 685

Leu Asp Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
    690                 695                 700

Arg Ser Tyr Leu Arg Ala Arg Lys Ala Phe Leu Glu Arg Ile Gly Arg
705                 710                 715                 720

Ser Asp Arg Val Glu Asn Arg Pro Phe Leu Leu Leu Lys Glu Pro Lys
                725                 730                 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Ser Glu Phe His Leu
```

740 745 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
755 760 765

Gly Tyr Asp Glu Val Gly Ser Tyr Lys Glu Val Gly Phe Met Ala Lys
770 775 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Lys Asp Arg Val Gln Pro
785 790 795 800

Phe Tyr Asp Ser Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
805 810 815

Gly Arg Phe Leu Ser Lys Glu Lys Arg Ala Glu Glu Trp Glu Ser Gly
820 825 830

Lys Glu Arg Phe Arg Leu Ala Lys Leu Lys Lys Glu Ile Leu Glu Ala
835 840 845

Gln Glu His Pro Tyr His Asp Phe Lys Ser Trp Gln Lys Phe Glu Arg
850 855 860

Glu Leu Arg Leu Val Lys Asn Gln Asp Ile Ile Thr Trp Met Met Cys
865 870 875 880

Arg Asp Leu Met Glu Glu Asn Lys Val Glu Gly Leu Asp Thr Gly Thr
885 890 895

Leu Tyr Leu Lys Asp Ile Arg Pro Asn Val Gln Glu Gln Gly Ser Leu
900 905 910

Asn Val Leu Asn Arg Val Lys Pro Met Arg Leu Pro Val Val Val Tyr
915 920 925

Arg Ala Asp Ser Arg Gly His Val His Lys Glu Glu Ala Pro Leu Ala
930 935 940

Thr Val Tyr Ile Glu Glu Arg Asp Thr Lys Leu Leu Lys Gln Gly Asn

945 950 955 960

Phe Lys Ser Phe Val Lys Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe
965 970 975

Val Asp Thr Gly Gly Leu Ala Met Glu Gln Tyr Pro Ile Ser Lys Leu
980 985 990

Arg Val Glu Tyr Glu Leu Ala Lys Tyr Gln Thr Ala Arg Val Cys Val
995 1000 1005

Phe Glu Leu Thr Leu Arg Leu Glu Glu Ser Leu Leu Ser Arg Tyr
1010 1015 1020

Pro His Leu Pro Asp Glu Ser Phe Arg Glu Met Leu Glu Ser Trp
1025 1030 1035

Ser Asp Pro Leu Leu Ala Lys Trp Pro Glu Leu His Gly Lys Val
1040 1045 1050

Arg Leu Leu Ile Ala Val Arg Asn Ala Phe Ser His Asn Gln Tyr
1055 1060 1065

Pro Met Tyr Asp Glu Ala Val Phe Ser Ser Ile Arg Lys Tyr Asp
1070 1075 1080

Pro Ser Ser Pro Asp Ala Ile Glu Glu Arg Met Gly Leu Asn Ile
1085 1090 1095

Ala His Arg Leu Ser Glu Glu Val Lys Gln Ala Lys Glu Thr Val
1100 1105 1110

Glu Arg Ile Ile Gln Ala
1115

<210> 131
<211> 1119
<212> PRT
<213> Porphyromonas gulae

<400> 131

```
Met Thr Glu Gln Ser Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1               5                  10                  15

Glu Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn
            20                  25                  30

Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
        35                  40                  45

Ala Asp Ile Thr Asn Asp Gln Asp Val Leu Ser Phe Lys Ala Leu Trp
    50                  55                  60

Lys Asn Phe Asp Asn Asp Leu Glu Arg Lys Ser Arg Leu Arg Ser Leu
65                  70                  75                  80

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

Leu Phe Glu Ser Lys Ser Ser Gly Asn Lys Ser Ser Lys Asn Lys Glu
            100                 105                 110

Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
        115                 120                 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
    130                 135                 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Ser Gly Ser Ser Glu Leu
145                 150                 155                 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
                165                 170                 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val
            180                 185                 190

Asp Pro His Tyr His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
        195                 200                 205
```

```
Arg Tyr Gly His Asn Asp Asn Pro Ser Phe Lys His His Phe Val Asp
    210                 215                 220

Ser Glu Gly Met Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225                 230                 235                 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
                245                 250                 255

Phe Lys Gly Gly Thr Gly Pro Tyr Glu Gln Met Thr Asn Glu Val Phe
            260                 265                 270

Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
        275                 280                 285

Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys
    290                 295                 300

Pro Lys Pro Leu Tyr Asp Arg Leu Arg Glu Lys Asp Arg Ala Cys Phe
305                 310                 315                 320

Arg Val Pro Val Asp Ile Leu Pro Asp Glu Asp Asp Thr Asp Gly Gly
                325                 330                 335

Gly Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
            340                 345                 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
        355                 360                 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
    370                 375                 380

Lys Met Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385                 390                 395                 400

Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
                405                 410                 415
```

```
Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
            420                 425                 430

Lys Pro Tyr Ile Ser Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly
        435                 440                 445

Lys Ile Gly Leu Arg Phe Met Pro Glu Gly Gln His Leu Trp Pro Ser
    450                 455                 460

Pro Glu Val Gly Thr Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465                 470                 475                 480

Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
                485                 490                 495

Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val
            500                 505                 510

Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Lys Asp Val
        515                 520                 525

Tyr Ala Ile Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Leu Lys
    530                 535                 540

Glu Leu Asp Ala Cys Ser Ala Asp Lys Gly Ile Arg Arg Gly His Leu
545                 550                 555                 560

Pro Lys Gln Met Ile Gly Ile Leu Ser Gln Glu His Lys Asn Met Glu
                565                 570                 575

Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His
            580                 585                 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly
        595                 600                 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
    610                 615                 620
```

```
Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly
625                 630                 635                 640

Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
                645                 650                 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro
            660                 665                 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
        675                 680                 685

Leu Asp Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
    690                 695                 700

Arg Ser Tyr Leu Arg Ala Arg Lys Ala Phe Leu Glu Arg Ile Gly Arg
705                 710                 715                 720

Ser Asp Arg Val Glu Asn Arg Pro Phe Leu Leu Leu Lys Glu Pro Lys
                725                 730                 735

Asn Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Ser Glu Phe His Leu
            740                 745                 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
        755                 760                 765

Gly Tyr Asp Glu Val Gly Ser Tyr Lys Glu Val Gly Phe Met Ala Lys
    770                 775                 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Lys Asp Arg Val Gln Pro
785                 790                 795                 800

Phe Tyr Asp Ser Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
                805                 810                 815

Gly Arg Phe Leu Ser Lys Glu Lys Arg Ala Glu Glu Trp Glu Ser Gly
            820                 825                 830
```

```
Lys Glu Arg Phe Arg Leu Ala Lys Leu Lys Lys Glu Ile Leu Glu Ala
        835                 840                 845

Lys Glu His Pro Tyr His Asp Phe Lys Ser Trp Gln Lys Phe Glu Arg
    850                 855                 860

Glu Leu Arg Leu Val Lys Asn Gln Asp Ile Ile Thr Trp Met Met Cys
865                 870                 875                 880

Arg Asp Leu Met Glu Glu Asn Lys Val Glu Gly Leu Asp Thr Gly Thr
                885                 890                 895

Leu Tyr Leu Lys Asp Ile Arg Thr Asp Val His Glu Gln Gly Ser Leu
            900                 905                 910

Asn Val Leu Asn Arg Val Lys Pro Met Arg Leu Pro Val Val Val Tyr
        915                 920                 925

Arg Ala Asp Ser Arg Gly His Val His Lys Glu Gln Ala Pro Leu Ala
    930                 935                 940

Thr Val Tyr Ile Glu Glu Arg Asp Thr Lys Leu Leu Lys Gln Gly Asn
945                 950                 955                 960

Phe Lys Ser Phe Val Lys Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe
                965                 970                 975

Val Asp Thr Gly Gly Leu Ala Met Glu Gln Tyr Pro Ile Ser Lys Leu
            980                 985                 990

Arg Val Glu Tyr Glu Leu Ala Lys  Tyr Gln Thr Ala Arg  Val Cys Ala
        995                 1000                1005

Phe Glu  Gln Thr Leu Glu Leu  Glu Glu Ser Leu Leu  Thr Arg Tyr
    1010                1015                1020

Pro His  Leu Pro Asp Glu Asn  Phe Arg Glu Met Leu  Glu Ser Trp
    1025                1030                1035
```

```
Ser Asp  Pro Leu Leu Gly Lys  Trp Pro Asp Leu His  Gly Lys Val
    1040                 1045                 1050


Arg Leu  Leu Ile Ala Val Arg  Asn Ala Phe Ser His  Asn Gln Tyr
    1055                 1060                 1065


Pro Met  Tyr Asp Glu Ala Val  Phe Ser Ser Ile Arg  Lys Tyr Asp
    1070                 1075                 1080


Pro Ser  Ser Pro Asp Ala Ile  Glu Glu Arg Met Gly  Leu Asn Ile
    1085                 1090                 1095


Ala His  Arg Leu Ser Glu Glu  Val Lys Gln Ala Lys  Glu Thr Val
    1100                 1105                 1110


Glu Arg  Ile Ile Gln Ala
    1115
```

<210> 132
<211> 1175
<212> PRT
<213> Porphyromonas gulae

<400> 132

```
Met Thr Glu Gln Ser Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1               5                   10                  15


Glu Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn
            20                  25                  30


Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
        35                  40                  45


Ala Asp Ile Thr Asn Asp Gln Asp Val Leu Ser Phe Lys Ala Leu Trp
    50                  55                  60


Lys Asn Phe Asp Asn Asp Leu Glu Arg Lys Ser Arg Leu Arg Ser Leu
65                  70                  75                  80
```

```
Ile Leu Lys His Phe Ser Phe Leu Glu Gly Ala Ala Tyr Gly Lys Lys
                85                  90                  95

Leu Phe Glu Ser Lys Ser Ser Gly Asn Lys Ser Ser Lys Asn Lys Glu
            100                 105                 110

Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
        115                 120                 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
    130                 135                 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Ser Glu Ser Ser Glu Leu
145                 150                 155                 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
                165                 170                 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val
            180                 185                 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
        195                 200                 205

Arg Tyr Gly His Asn Asp Asn Pro Ser Phe Lys His His Phe Val Asp
    210                 215                 220

Gly Glu Gly Met Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225                 230                 235                 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
                245                 250                 255

Phe Lys Gly Gly Thr Glu Thr Tyr Gln Gln Met Thr Asn Glu Val Phe
            260                 265                 270

Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
        275                 280                 285
```

```
Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys
    290                 295                 300

Pro Lys Pro Leu Tyr Asp Arg Leu Arg Glu Asp Asp Arg Ala Cys Phe
305                 310                 315                 320

Arg Val Pro Val Asp Ile Leu Pro Asp Glu Asp Asp Thr Asp Gly Gly
                325                 330                 335

Gly Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
            340                 345                 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
        355                 360                 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
    370                 375                 380

Lys Met Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385                 390                 395                 400

Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
                405                 410                 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
            420                 425                 430

Lys Pro Tyr Ile Ser Gln Thr Ser Pro His Tyr His Ile Glu Lys Gly
        435                 440                 445

Lys Ile Gly Leu Arg Phe Met Pro Glu Gly Gln His Leu Trp Pro Ser
    450                 455                 460

Pro Glu Val Gly Thr Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465                 470                 475                 480

Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
                485                 490                 495
```

```
Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val
            500                 505                 510

Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
        515                 520                 525

Tyr Ala Ile Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Leu Lys
    530                 535                 540

Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu
545                 550                 555                 560

Pro Lys Gln Met Ile Ala Ile Leu Ser Gln Glu His Lys Asp Met Glu
                565                 570                 575

Glu Lys Ile Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His
            580                 585                 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly
        595                 600                 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
    610                 615                 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly
625                 630                 635                 640

Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
                645                 650                 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Lys Arg Leu Thr Pro
            660                 665                 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
        675                 680                 685

Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
    690                 695                 700
```

```
Arg Ser Tyr Leu Arg Ala Arg Lys Ala Phe Leu Glu Arg Ile Gly Arg
705                 710                 715                 720


Ser Asp Arg Met Glu Asn Arg Pro Phe Leu Leu Leu Lys Glu Pro Lys
                725                 730                 735


Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Ser Glu Phe His Leu
            740                 745                 750


Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
        755                 760                 765


Gly Tyr Asp Glu Val Gly Ser Tyr Arg Glu Val Gly Phe Met Ala Lys
    770                 775                 780


Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Glu Asp Arg Val Gln Pro
785                 790                 795                 800


Phe Tyr Asp Ser Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
                805                 810                 815


Gly Arg Phe Leu Ser Lys Glu Glu Arg Ala Glu Glu Trp Glu Arg Gly
            820                 825                 830


Lys Glu Arg Phe Arg Asp Leu Glu Ala Trp Ser His Ser Ala Ala Arg
        835                 840                 845


Arg Ile Glu Asp Ala Phe Ala Gly Ile Glu Tyr Ala Ser Pro Gly Asn
    850                 855                 860


Lys Lys Lys Ile Glu Gln Leu Leu Arg Asp Leu Ser Leu Trp Glu Ala
865                 870                 875                 880


Phe Glu Ser Lys Leu Lys Val Arg Ala Asp Lys Ile Asn Leu Ala Lys
                885                 890                 895


Leu Lys Lys Glu Ile Leu Glu Ala Gln Glu His Pro Tyr His Asp Phe
            900                 905                 910
```

```
Lys Ser Trp Gln Lys Phe Glu Arg Glu Leu Arg Leu Val Lys Asn Gln
        915                 920                 925


Asp Ile Ile Thr Trp Met Met Cys Arg Asp Leu Met Glu Glu Asn Lys
    930                 935                 940


Val Glu Gly Leu Asp Thr Gly Thr Leu Tyr Leu Lys Asp Ile Arg Pro
945                 950                 955                 960


Asn Val Gln Glu Gln Gly Ser Leu Asn Val Leu Asn Arg Val Lys Pro
                965                 970                 975


Met Arg Leu Pro Val Val Val Tyr Arg Ala Asp Ser Arg Gly His Val
            980                 985                 990


His Lys Glu Glu Ala Pro Leu Ala  Thr Val Tyr Ile Glu  Glu Arg Asp
        995                 1000                1005


Thr Lys  Leu Leu Lys Gln Gly  Asn Phe Lys Ser Phe  Val Lys Asp
    1010                1015                1020


Arg Arg  Leu Asn Gly Leu Phe  Ser Phe Val Asp Thr  Gly Gly Leu
    1025                1030                1035


Ala Met  Glu Gln Tyr Pro Ile  Ser Lys Leu Arg Val  Glu Tyr Glu
    1040                1045                1050


Leu Ala  Lys Tyr Gln Thr Ala  Arg Val Cys Val Phe  Glu Leu Thr
    1055                1060                1065


Leu Arg  Leu Glu Glu Ser Leu  Leu Thr Arg Tyr Pro  His Leu Pro
    1070                1075                1080


Asp Glu  Ser Phe Arg Lys Met  Leu Glu Ser Trp Ser  Asp Pro Leu
    1085                1090                1095


Leu Ala  Lys Trp Pro Glu Leu  His Gly Lys Val Arg  Leu Leu Ile
    1100                1105                1110
```

```
Ala Val  Arg Asn Ala Phe Ser  His Asn Gln Tyr Pro  Met Tyr Asp
    1115                1120                1125


Glu Ala  Val Phe Ser Ser Ile  Arg Lys Tyr Asp Pro  Ser Ser Pro
    1130                1135                1140


Asp Ala  Ile Glu Glu Arg Met  Gly Leu Asn Ile Ala  His Arg Leu
    1145                1150                1155


Ser Glu  Glu Val Lys Gln Ala  Lys Glu Thr Val Glu  Arg Ile Ile
    1160                1165                1170


Gln Val
    1175


<210>  133
<211>  1119
<212>  PRT
<213>  Porphyromonas gulae

<400>  133

Met Thr Glu Gln Ser Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1               5                   10                  15


Glu Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn
            20                  25                  30


Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
        35                  40                  45


Ala Asp Ile Thr Asn Asp Glu Asp Ile Leu Phe Phe Lys Gly Gln Trp
    50                  55                  60


Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ser Arg Leu Arg Ser Leu
65                  70                  75                  80


Ile Leu Lys His Phe Ser Phe Leu Glu Gly Ala Ala Tyr Gly Lys Lys
                85                  90                  95
```

```
Phe Phe Glu Ser Lys Ser Ser Gly Asn Lys Ser Ser Lys Asn Lys Glu
            100                 105                 110


Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
        115                 120                 125


Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
    130                 135                 140


Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Ser Gly Ser Ser Glu Leu
145                 150                 155                 160


Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
                165                 170                 175


Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Glu Val
            180                 185                 190


Asp Pro His Tyr His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
        195                 200                 205


Arg Tyr Gly His Asn Asp Asn Pro Ser Phe Lys His His Phe Val Asp
    210                 215                 220


Gly Glu Gly Met Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225                 230                 235                 240


Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
                245                 250                 255


Phe Lys Gly Gly Thr Glu Pro Tyr Glu Gln Met Thr Asn Glu Val Phe
            260                 265                 270


Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
        275                 280                 285


Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys
    290                 295                 300
```

```
Pro Lys Pro Leu Tyr Asp Arg Leu Arg Glu Lys Asp Arg Ala Cys Phe
305                 310                 315                 320


Arg Val Pro Val Asp Ile Leu Pro Asp Glu Asp Asp Thr Asp Gly Gly
                325                 330                 335


Gly Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
            340                 345                 350


Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
        355                 360                 365


Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
    370                 375                 380


Lys Met Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385                 390                 395                 400


Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
                405                 410                 415


Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
            420                 425                 430


Lys Pro Tyr Ile Ser Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly
        435                 440                 445


Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln His Leu Trp Pro Ser
    450                 455                 460


Pro Glu Val Gly Thr Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465                 470                 475                 480


Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
                485                 490                 495


Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val
            500                 505                 510
```

```
Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
        515                 520                 525

Tyr Ala Ile Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Leu Lys
    530                 535                 540

Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu
545                 550                 555                 560

Pro Lys Gln Met Ile Gly Ile Leu Ser Gln Glu Arg Lys Asp Met Glu
                565                 570                 575

Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His
            580                 585                 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly
        595                 600                 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
    610                 615                 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly
625                 630                 635                 640

Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
                645                 650                 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro
            660                 665                 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
        675                 680                 685

Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
    690                 695                 700

Arg Ser Tyr Leu Arg Ala Arg Lys Ala Phe Leu Glu Arg Ile Gly Arg
705                 710                 715                 720
```

```
Ser Asp Arg Val Glu Asn Cys Pro Phe Leu Leu Leu Lys Glu Pro Lys
                725                 730                 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Gly Glu Phe His Leu
            740                 745                 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
        755                 760                 765

Gly Tyr Asp Glu Val Gly Ser Tyr Arg Glu Val Gly Phe Met Ala Lys
    770                 775                 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Glu Asp Arg Val Gln Pro
785                 790                 795                 800

Phe Tyr Asp Ser Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
                805                 810                 815

Gly Arg Phe Leu Ser Lys Glu Lys Arg Ala Glu Glu Trp Glu Ser Gly
            820                 825                 830

Lys Glu Arg Phe Arg Leu Ala Lys Leu Lys Lys Glu Ile Leu Glu Ala
        835                 840                 845

Gln Glu His Pro Tyr His Asp Phe Lys Ser Trp Gln Lys Phe Glu Arg
    850                 855                 860

Glu Leu Arg Leu Val Lys Asn Gln Asp Ile Ile Thr Trp Met Met Cys
865                 870                 875                 880

Arg Asp Leu Met Glu Glu Asn Lys Val Glu Gly Leu Asp Thr Gly Thr
                885                 890                 895

Leu Tyr Leu Lys Asp Ile Arg Pro Asn Val Gln Glu Gln Gly Ser Leu
            900                 905                 910

Asn Val Leu Asn Arg Val Lys Pro Met Arg Leu Pro Val Val Val Tyr
        915                 920                 925
```

```
Arg Ala Asp Ser Arg Gly His Val His Lys Glu Glu Ala Pro Leu Ala
    930                 935                 940

Thr Val Tyr Ile Glu Glu Arg Asp Thr Lys Leu Leu Lys Gln Gly Asn
945                 950                 955                 960

Phe Lys Ser Phe Val Lys Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe
                965                 970                 975

Val Asp Thr Gly Ala Leu Ala Met Glu Gln Tyr Pro Ile Ser Lys Leu
            980                 985                 990

Arg Val Glu Tyr Glu Leu Ala Lys  Tyr Gln Thr Ala Arg  Val Cys Ala
        995                 1000                1005

Phe Glu  Gln Thr Leu Glu Leu  Glu Glu Ser Leu Leu  Thr Arg Tyr
    1010                1015                1020

Pro His  Leu Pro Asp Glu Ser  Phe Arg Glu Met Leu  Glu Ser Trp
    1025                1030                1035

Ser Asp  Pro Leu Leu Thr Lys  Trp Pro Glu Leu His  Gly Lys Val
    1040                1045                1050

Arg Leu  Leu Ile Ala Val Arg  Asn Ala Phe Ser His  Asn Gln Tyr
    1055                1060                1065

Pro Met  Tyr Asp Glu Ala Val  Phe Ser Ser Ile Trp  Lys Tyr Asp
    1070                1075                1080

Pro Ser  Ser Pro Asp Ala Ile  Glu Glu Arg Met Gly  Leu Asn Ile
    1085                1090                1095

Ala His  Arg Leu Ser Glu Glu  Val Lys Gln Ala Lys  Glu Thr Ile
    1100                1105                1110

Glu Arg  Ile Ile Gln Ala
    1115
```

<210> 134
<211> 1174
<212> PRT
<213> Porphyromonas gulae

<400> 134

Met Thr Glu Gln Ser Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1 5 10 15

Glu Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn
20 25 30

Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
35 40 45

Ala Asp Ile Thr Asn Asp Gln Asp Val Leu Ser Phe Lys Ala Leu Trp
50 55 60

Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ser Arg Leu Arg Ser Leu
65 70 75 80

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

Leu Phe Glu Ser Lys Ser Ser Gly Asn Lys Ser Ser Lys Asn Lys Glu
100 105 110

Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
115 120 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
130 135 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Ser Gly Ser Ser Glu Leu
145 150 155 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
165 170 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val

180 185 190

Asp Pro His Tyr His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
195 200 205

Arg Tyr Gly His Asn Asp Asn Pro Ser Phe Lys His His Phe Val Asp
210 215 220

Ser Glu Gly Met Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225 230 235 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
245 250 255

Phe Lys Gly Gly Thr Gly Pro Tyr Glu Gln Met Thr Asn Glu Val Phe
260 265 270

Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
275 280 285

Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys
290 295 300

Pro Lys Pro Leu Tyr Asp Arg Leu Arg Glu Lys Asp Arg Ala Cys Phe
305 310 315 320

Arg Val Pro Val Asp Ile Leu Pro Asp Glu Asp Asp Thr Asp Gly Gly
325 330 335

Gly Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
340 345 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
355 360 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
370 375 380

Lys Met Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu

385 390 395 400

Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
405 410 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Leu Glu Thr Gly Asp
420 425 430

Lys Pro Tyr Ile Ser Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly
435 440 445

Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln His Leu Trp Pro Ser
450 455 460

Pro Glu Val Gly Thr Thr Arg Thr Gly Arg Ser Lys Cys Ala Gln Asp
465 470 475 480

Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
485 490 495

Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val
500 505 510

Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
515 520 525

Tyr Ala Ile Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Leu Lys
530 535 540

Glu Leu Asp Thr Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu
545 550 555 560

Pro Lys Gln Met Ile Thr Ile Leu Ser Gln Glu Arg Lys Asp Met Lys
565 570 575

Glu Lys Ile Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His
580 585 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly

595 600 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
610 615 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Ala Ser Gly
625 630 635 640

Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
645 650 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro
660 665 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
675 680 685

Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
690 695 700

Arg Ser Tyr Leu Arg Ala Arg Lys Ala Phe Leu Glu Arg Ile Gly Arg
705 710 715 720

Ser Asp Arg Val Glu Asn Cys Pro Phe Leu Leu Leu Lys Glu Pro Lys
725 730 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Asp Glu Phe His Leu
740 745 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
755 760 765

Gly Tyr Asp Glu Val Gly Ser Tyr Arg Glu Val Gly Phe Met Ala Lys
770 775 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Glu Asp Arg Val Gln Pro
785 790 795 800

Phe Tyr Asp Ser Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys

805 810 815

Gly Arg Phe Leu Ser Lys Glu Asp Arg Ala Glu Glu Trp Glu Arg Gly
820 825 830

Met Glu Arg Phe Arg Asp Leu Glu Ala Trp Ser His Ser Ala Ala Arg
835 840 845

Arg Ile Lys Asp Ala Phe Ala Gly Ile Glu Tyr Ala Ser Pro Gly Asn
850 855 860

Lys Lys Lys Ile Glu Gln Leu Leu Arg Asp Leu Ser Leu Trp Glu Ala
865 870 875 880

Phe Glu Ser Lys Leu Lys Val Arg Ala Asp Lys Ile Asn Leu Ala Lys
885 890 895

Leu Lys Lys Glu Ile Leu Glu Ala Gln Glu His Pro Tyr His Asp Phe
900 905 910

Lys Ser Trp Gln Lys Phe Glu Arg Glu Leu Arg Leu Val Lys Asn Gln
915 920 925

Asp Ile Ile Thr Trp Met Met Cys Arg Asp Leu Met Glu Glu Asn Lys
930 935 940

Val Glu Gly Leu Asp Thr Gly Thr Leu Tyr Leu Lys Asp Ile Arg Pro
945 950 955 960

Asn Val Gln Glu Gln Gly Ser Leu Asn Val Leu Asn Arg Val Lys Pro
965 970 975

Met Arg Leu Pro Val Val Val Tyr Arg Ala Asp Ser Arg Gly His Val
980 985 990

His Lys Glu Ala Pro Leu Ala Thr Val Tyr Ile Glu Glu Arg Asn Thr
995 1000 1005

Lys Leu Leu Lys Gln Gly Asn Phe Lys Ser Phe Val Lys Asp Arg

```
    1010                1015                1020

Arg Leu  Asn Gly Leu Phe Ser  Phe Val Asp Thr Gly  Gly Leu Ala
    1025                1030                1035

Met Glu  Gln Tyr Pro Ile Ser  Lys Leu Arg Val Glu  Tyr Glu Leu
    1040                1045                1050

Ala Lys  Tyr Gln Thr Ala Arg  Val Cys Val Phe Glu  Leu Thr Leu
    1055                1060                1065

Arg Leu  Glu Glu Ser Leu Leu  Ser Arg Tyr Pro His  Leu Pro Asp
    1070                1075                1080

Glu Ser  Phe Arg Glu Met Leu  Glu Ser Trp Ser Asp  Pro Leu Leu
    1085                1090                1095

Ala Lys  Trp Pro Glu Leu His  Gly Lys Val Arg Leu  Leu Ile Ala
    1100                1105                1110

Val Arg  Asn Ala Phe Ser His  Asn Gln Tyr Pro Met  Tyr Asp Glu
    1115                1120                1125

Ala Val  Phe Ser Ser Ile Arg  Lys Tyr Asp Pro Ser  Ser Pro Asp
    1130                1135                1140

Ala Ile  Glu Glu Arg Met Gly  Leu Asn Ile Ala His  Arg Leu Ser
    1145                1150                1155

Glu Glu  Val Lys Gln Ala Lys  Glu Thr Val Glu Arg  Ile Ile Gln
    1160                1165                1170

Ala
```

<210> 135
<211> 1135
<212> PRT
<213> Porphyromonas gulae

<400> 135

```
Met Asn Thr Val Pro Ala Thr Glu Asn Lys Gly Gln Ser Arg Thr Val
1               5                  10                  15

Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu
            20                  25                  30

Asn Leu Ile Glu Val Glu Ser His Val Arg Ile Lys Phe Gly Lys Lys
        35                  40                  45

Lys Leu Asn Glu Glu Ser Leu Lys Gln Ser Leu Leu Cys Asp His Leu
    50                  55                  60

Leu Ser Ile Asp Arg Trp Thr Lys Val Tyr Gly His Ser Arg Arg Tyr
65                  70                  75                  80

Leu Pro Phe Leu His Cys Phe Asp Pro Asp Ser Gly Ile Glu Lys Asp
                85                  90                  95

His Asp Ser Lys Thr Gly Val Asp Pro Asp Ser Ala Gln Arg Leu Ile
            100                 105                 110

Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His
        115                 120                 125

Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Lys Val Ser Pro Asp
    130                 135                 140

Ile Ser Ser Phe Ile Thr Gly Ala Tyr Thr Phe Ala Cys Glu Arg Ala
145                 150                 155                 160

Gln Ser Arg Phe Ala Asp Phe Phe Lys Pro Asp Asp Phe Leu Leu Ala
                165                 170                 175

Lys Asn Arg Lys Glu Gln Leu Ile Ser Val Ala Asp Gly Lys Glu Cys
            180                 185                 190

Leu Thr Val Ser Gly Phe Ala Phe Phe Ile Cys Leu Phe Leu Asp Arg
        195                 200                 205
```

```
Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr
    210                 215                 220

Asp Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys
225                 230                 235                 240

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

Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu
            260                 265                 270

Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu
        275                 280                 285

Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asn Glu Glu
    290                 295                 300

Ser Arg Leu Leu Trp Asp Gly Ser Ser Asp Trp Ala Glu Ala Leu Thr
305                 310                 315                 320

Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe
                325                 330                 335

Ile Glu Glu Met Asp Leu Leu Lys Gly Ile Arg Phe Arg Val Asp Leu
            340                 345                 350

Gly Glu Ile Glu Leu Asp Ser Tyr Ser Lys Lys Val Gly Arg Asn Gly
        355                 360                 365

Glu Tyr Asp Arg Thr Ile Thr Asp His Ala Leu Ala Phe Gly Lys Leu
    370                 375                 380

Ser Asp Phe Gln Asn Glu Glu Glu Val Ser Arg Met Ile Ser Gly Glu
385                 390                 395                 400

Ala Ser Tyr Pro Val Arg Phe Ser Leu Phe Ala Pro Arg Tyr Ala Ile
                405                 410                 415
```

```
Tyr Asp Asn Lys Ile Gly Tyr Cys His Thr Ser Asp Pro Val Tyr Pro
            420                 425                 430

Lys Ser Lys Thr Gly Glu Lys Arg Ala Leu Ser Asn Pro Gln Ser Met
        435                 440                 445

Gly Phe Ile Ser Val His Asp Leu Arg Lys Leu Leu Leu Met Glu Leu
    450                 455                 460

Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Asp Phe Leu Arg Lys
465                 470                 475                 480

Ala Asn Arg Ile Leu Asp Glu Thr Ala Glu Gly Lys Leu Gln Phe Ser
                485                 490                 495

Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro
            500                 505                 510

Lys Ser Lys Asp Arg Arg Glu Lys Ala Glu Thr Thr Leu Glu Lys Tyr
        515                 520                 525

Lys Gln Glu Ile Lys Gly Arg Lys Asp Lys Leu Asn Ser Gln Leu Leu
    530                 535                 540

Ser Ala Phe Asp Met Asn Gln Arg Gln Leu Pro Ser Arg Leu Leu Asp
545                 550                 555                 560

Glu Trp Met Asn Ile Arg Pro Ala Ser His Ser Val Lys Leu Arg Thr
                565                 570                 575

Tyr Val Lys Gln Leu Asn Glu Asp Cys Arg Leu Arg Leu Arg Lys Phe
            580                 585                 590

Arg Lys Asp Gly Asp Gly Lys Ala Arg Ala Ile Pro Leu Val Gly Glu
        595                 600                 605

Met Ala Thr Phe Leu Ser Gln Asp Ile Val Arg Met Ile Ile Ser Glu
    610                 615                 620
```

```
Glu Thr Lys Lys Leu Ile Thr Ser Ala Tyr Tyr Asn Glu Met Gln Arg
625                 630                 635                 640

Ser Leu Ala Gln Tyr Ala Gly Glu Glu Asn Arg Arg Gln Phe Arg Ala
                645                 650                 655

Ile Val Ala Glu Leu His Leu Leu Asp Pro Ser Ser Gly His Pro Phe
            660                 665                 670

Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Asp Phe Tyr
        675                 680                 685

Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Thr Phe Tyr
    690                 695                 700

Arg Pro Glu Gln Asp Glu Asn Thr Lys Arg Arg Ile Ser Val Phe Phe
705                 710                 715                 720

Val Pro Asp Gly Glu Ala Arg Lys Leu Leu Pro Leu Leu Ile Arg Arg
                725                 730                 735

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

Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Ile Met
        755                 760                 765

Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe
    770                 775                 780

Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr
785                 790                 795                 800

Gly Leu Arg Arg Glu Leu Asn Ile His Gly Lys Ser Val Ser Tyr Ile
                805                 810                 815

Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu
            820                 825                 830
```

```
Lys Thr Val Arg Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro
        835                 840                 845

Val Pro Pro Asp Leu Ala Ala Tyr Ile Lys Arg Ser Phe His Arg Ala
    850                 855                 860

Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg
865                 870                 875                 880

Leu Met Leu Met Ala Ile Asn Lys Met Met Thr Asp Arg Glu Glu Asp
                885                 890                 895

Ile Leu Pro Gly Leu Lys Asn Ile Asp Ser Ile Leu Asp Glu Glu Asn
            900                 905                 910

Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu
        915                 920                 925

Gly Gly Asp Asn Ser Leu Ser Leu Val Pro Ala Thr Ile Glu Ile Lys
    930                 935                 940

Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg
945                 950                 955                 960

Arg Val Pro Gly Leu Met Ser His Phe Pro Glu His Lys Ala Thr Leu
                965                 970                 975

Asp Glu Val Lys Thr Leu Leu Gly Glu Tyr Asp Arg Cys Arg Ile Lys
            980                 985                 990

Ile Phe Asp Trp Ala Phe Ala Leu  Glu Gly Ala Ile Met  Ser Asp Arg
        995                 1000                1005

Asp Leu  Lys Pro Tyr Leu His  Glu Ser Ser Ser Arg  Glu Gly Lys
    1010                1015                1020

Ser Gly  Glu His Ser Thr Leu  Val Lys Met Leu Val  Glu Lys Lys
    1025                1030                1035
```

```
Gly Cys  Leu Thr Pro Asp Glu  Ser Gln Tyr Leu Ile  Leu Ile Arg
    1040                1045                1050


Asn Lys  Ala Ala His Asn Gln  Phe Pro Cys Ala Ala  Glu Met Pro
    1055                1060                1065


Leu Ile  Tyr Arg Asp Val Ser  Ala Lys Val Gly Ser  Ile Glu Gly
    1070                1075                1080


Ser Ser  Ala Lys Asp Leu Pro  Glu Gly Ser Ser Leu  Val Asp Ser
    1085                1090                1095


Leu Trp  Lys Lys Tyr Glu Met  Ile Ile Arg Lys Ile  Leu Pro Ile
    1100                1105                1110


Leu Asp  His Glu Asn Arg Phe  Phe Gly Lys Leu Leu  Asn Asn Met
    1115                1120                1125


Ser Gln  Pro Ile Asn Asp Leu
    1130                1135


<210>  136
<211>  1218
<212>  PRT
<213>  Capnocytophaga cynodegmi

<400>  136

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


Lys Pro Gln Asp Lys Ser Tyr Phe Ala Gly Tyr Leu Asn Ala Ala Met
            20                  25                  30


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


Lys Glu Tyr Thr Ser Glu Asn Phe Phe Asp Ala Ile Phe Lys Glu Asn
    50                  55                  60
```

```
Ile Ser Leu Val Glu Tyr Glu Arg Tyr Val Lys Leu Leu Ser Asp Tyr
65                  70                  75                  80

Phe Pro Met Ala Arg Leu Leu Asp Lys Lys Glu Val Pro Ile Lys Glu
                85                  90                  95

Arg Lys Glu Asn Phe Lys Lys Asn Phe Arg Gly Ile Ile Lys Ala Val
            100                 105                 110

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

Ile Thr Asp Glu Ile Phe Gly Val Leu Asp Glu Met Leu Lys Ser Thr
    130                 135                 140

Val Leu Thr Val Lys Lys Lys Lys Ile Lys Thr Asp Lys Thr Lys Glu
145                 150                 155                 160

Ile Leu Lys Lys Ser Ile Glu Lys Gln Leu Asp Ile Leu Cys Gln Lys
                165                 170                 175

Lys Leu Glu Tyr Leu Lys Asp Thr Ala Arg Lys Ile Glu Glu Lys Arg
            180                 185                 190

Arg Asn Gln Arg Glu Arg Gly Glu Lys Lys Leu Val Pro Arg Phe Glu
        195                 200                 205

Tyr Ser Asp Arg Arg Asp Asp Leu Ile Ala Ala Ile Tyr Asn Asp Ala
    210                 215                 220

Phe Asp Val Tyr Ile Asp Lys Lys Lys Asp Ser Leu Lys Glu Ser Ser
225                 230                 235                 240

Lys Thr Lys Tyr Asn Thr Glu Ser Tyr Pro Gln Gln Glu Glu Gly Asp
                245                 250                 255

Leu Lys Ile Pro Ile Ser Lys Asn Gly Val Val Phe Leu Leu Ser Leu
            260                 265                 270
```

```
Phe Leu Ser Lys Gln Glu Val His Ala Phe Lys Ser Lys Ile Ala Gly
        275                 280                 285

Phe Lys Ala Thr Val Ile Asp Glu Ala Thr Val Ser His Arg Lys Asn
    290                 295                 300

Ser Ile Cys Phe Met Ala Thr His Glu Ile Phe Ser His Leu Ala Tyr
305                 310                 315                 320

Lys Lys Leu Lys Arg Lys Val Arg Thr Ala Glu Ile Asn Tyr Ser Glu
                325                 330                 335

Ala Glu Asn Ala Glu Gln Leu Ser Ile Tyr Ala Lys Glu Thr Leu Met
            340                 345                 350

Met Gln Met Leu Asp Glu Leu Ser Lys Val Pro Asp Val Val Tyr Gln
        355                 360                 365

Asn Leu Ser Glu Asp Val Gln Lys Thr Phe Ile Glu Asp Trp Asn Glu
    370                 375                 380

Tyr Leu Lys Glu Asn Asn Gly Asp Val Gly Thr Met Glu Glu Glu Gln
385                 390                 395                 400

Val Ile His Pro Val Ile Arg Lys Arg Tyr Glu Asp Lys Phe Asn Tyr
                405                 410                 415

Phe Ala Ile Arg Phe Leu Asp Glu Phe Ala Gln Phe Pro Thr Leu Arg
            420                 425                 430

Phe Gln Val His Leu Gly Asn Tyr Leu His Asp Ser Arg Pro Lys Glu
        435                 440                 445

His Leu Ile Ser Asp Arg Arg Ile Lys Glu Lys Ile Thr Val Phe Gly
    450                 455                 460

Arg Leu Ser Glu Leu Glu His Lys Lys Ala Leu Phe Ile Lys Asn Thr
465                 470                 475                 480
```

```
Glu Thr Asn Glu Asp Arg Lys His Tyr Trp Glu Val Phe Pro Asn Pro
                485                 490                 495

Asn Tyr Asp Phe Pro Lys Glu Asn Ile Ser Val Asn Asp Lys Asp Phe
            500                 505                 510

Pro Ile Ala Gly Ser Ile Leu Asp Arg Glu Lys Gln Pro Thr Ala Gly
        515                 520                 525

Lys Ile Gly Ile Lys Val Asn Leu Leu Asn Gln Lys Tyr Ile Ser Glu
    530                 535                 540

Val Asp Lys Ala Val Lys Ala His Gln Leu Lys Gln Arg Asn Asn Lys
545                 550                 555                 560

Pro Ser Ile Gln Asn Ile Ile Glu Glu Ile Val Pro Ile Asn Gly Ser
                565                 570                 575

Asn Pro Lys Glu Ile Ile Val Phe Gly Gly Gln Pro Thr Ala Tyr Leu
            580                 585                 590

Ser Met Asn Asp Ile His Ser Ile Leu Tyr Glu Phe Phe Asp Lys Trp
        595                 600                 605

Glu Lys Lys Lys Glu Lys Leu Glu Lys Lys Gly Glu Lys Glu Leu Arg
    610                 615                 620

Lys Glu Ile Gly Lys Glu Leu Glu Glu Lys Ile Val Gly Lys Ile Gln
625                 630                 635                 640

Thr Gln Ile Gln Gln Ile Ile Asp Lys Asp Ile Asn Ala Lys Ile Leu
                645                 650                 655

Lys Pro Tyr Gln Asp Asp Asp Ser Thr Ala Ile Asp Lys Glu Lys Leu
            660                 665                 670

Ile Lys Asp Leu Lys Gln Glu Gln Lys Ile Leu Gln Lys Leu Lys Asn
        675                 680                 685
```

```
Glu Gln Thr Ala Arg Glu Lys Glu Tyr Gln Glu Cys Ile Ala Tyr Gln
    690                 695                 700

Glu Glu Ser Arg Lys Ile Lys Arg Ser Asp Lys Ser Arg Gln Lys Tyr
705                 710                 715                 720

Leu Arg Asn Gln Leu Lys Arg Lys Tyr Pro Glu Val Pro Thr Arg Lys
                725                 730                 735

Glu Ile Leu Tyr Tyr Gln Glu Lys Gly Lys Val Ala Val Trp Leu Ala
            740                 745                 750

Asn Asp Ile Lys Arg Phe Met Pro Thr Asp Phe Lys Asn Glu Trp Lys
        755                 760                 765

Gly Glu Gln His Ser Leu Leu Gln Lys Ser Leu Ala Tyr Tyr Glu Gln
    770                 775                 780

Cys Lys Glu Glu Leu Lys Asn Leu Leu Pro Gln Gln Lys Val Phe Lys
785                 790                 795                 800

His Leu Pro Phe Glu Leu Gly Gly His Phe Gln Gln Lys Tyr Leu Tyr
                805                 810                 815

Gln Phe Tyr Thr Arg Tyr Leu Asp Lys Arg Leu Glu His Ile Ser Gly
            820                 825                 830

Leu Val Gln Gln Ala Glu Asn Phe Lys Asn Glu Asn Lys Val Phe Lys
        835                 840                 845

Lys Val Glu Asn Glu Cys Phe Lys Phe Leu Lys Lys Gln Asn Tyr Thr
    850                 855                 860

His Lys Gly Leu Asp Ala Gln Ala Gln Ser Val Leu Gly Tyr Pro Ile
865                 870                 875                 880

Phe Leu Glu Arg Gly Phe Met Asp Glu Lys Pro Thr Ile Ile Lys Gly
                885                 890                 895
```

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


Tyr Lys Glu Tyr Gln Asn Phe Gln Thr Phe Tyr Asp Thr Glu Asn Tyr
        915                 920                 925


Pro Leu Val Glu Leu Glu Lys Lys Gln Ala Asp Arg Lys Arg Glu Thr
    930                 935                 940


Lys Ile Tyr Gln Gln Lys Lys Asn Asp Val Phe Thr Leu Leu Met Ala
945                 950                 955                 960


Lys His Ile Phe Lys Ser Val Phe Lys Gln Asp Ser Ile Asp Arg Phe
                965                 970                 975


Ser Leu Glu Asp Leu Tyr Gln Ser Arg Glu Glu Arg Leu Glu Asn Gln
            980                 985                 990


Glu Lys Ala Lys Gln Thr Gly Glu  Arg Asn Thr Asn Tyr  Ile Trp Asn
        995                 1000                1005


Lys Thr  Val Asp Leu Asn Leu  Cys Asp Gly Lys Val  Thr Val Glu
    1010                1015                1020


Asn Val  Lys Leu Lys Asn Val  Gly Asn Phe Ile Lys  Tyr Glu Tyr
    1025                1030                1035


Asp Gln  Arg Val Gln Thr Phe  Leu Lys Tyr Glu Glu  Asn Ile Lys
    1040                1045                1050


Trp Gln  Ala Phe Leu Ile Lys  Glu Ser Lys Glu Glu  Glu Asn Tyr
    1055                1060                1065


Pro Tyr  Ile Val Glu Arg Glu  Ile Glu Gln Tyr Glu  Lys Val Arg
    1070                1075                1080


Arg Glu  Glu Leu Leu Lys Glu  Val His Leu Ile Glu  Glu Tyr Ile
    1085                1090                1095
```

```
Leu Glu  Lys Val Lys Asp Lys  Glu Ile Leu Lys Lys  Gly Asp Asn
    1100                1105                1110


Gln Asn  Phe Lys Tyr Tyr Ile  Leu Asn Gly Leu Leu  Lys Gln Leu
    1115                1120                1125


Lys Asn  Glu Asp Val Glu Ser  Tyr Lys Val Phe Asn  Leu Asn Thr
    1130                1135                1140


Lys Pro  Glu Asp Val Asn Ile  Asn Gln Leu Lys Gln  Glu Ala Thr
    1145                1150                1155


Asp Leu  Glu Gln Lys Ala Phe  Val Leu Thr Tyr Ile  Arg Asn Lys
    1160                1165                1170


Phe Ala  His Asn Gln Leu Pro  Lys Lys Glu Phe Trp  Asp Tyr Cys
    1175                1180                1185


Gln Glu  Lys Tyr Gly Lys Ile  Glu Lys Glu Lys Thr  Tyr Ala Glu
    1190                1195                1200


Tyr Phe  Ala Glu Val Phe Lys  Arg Glu Lys Glu Ala  Leu Met Lys
    1205                1210                1215


<210>  137
<211>  1090
<212>  PRT
<213>  Prevotella sp.

<400>  137

Met Asn Ile Pro Ala Leu Val Glu Asn Gln Lys Lys Tyr Phe Gly Thr
1               5                  10                  15


Tyr Ser Val Met Ala Met Leu Asn Ala Gln Thr Val Leu Asp His Ile
            20                  25                  30


Gln Lys Val Ala Asp Ile Glu Gly Glu Gln Asn Glu Asn Asn Glu Asn
        35                  40                  45
```

```
Leu Trp Phe His Pro Val Met Ser His Leu Tyr Asn Ala Lys Asn Gly
    50                  55                  60

Tyr Asp Lys Gln Pro Glu Lys Thr Met Phe Ile Ile Glu Arg Leu Gln
65                  70                  75                  80

Ser Tyr Phe Pro Phe Leu Lys Ile Met Ala Glu Asn Gln Arg Glu Tyr
                85                  90                  95

Ser Asn Gly Lys Tyr Lys Gln Asn Arg Val Glu Val Asn Ser Asn Asp
            100                 105                 110

Ile Phe Glu Val Leu Lys Arg Ala Phe Gly Val Leu Lys Met Tyr Arg
        115                 120                 125

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

Cys Glu Phe Leu Thr Ser Thr Glu Gln Pro Leu Ser Gly Met Ile Ser
145                 150                 155                 160

Lys Tyr Tyr Thr Val Ala Leu Arg Asn Thr Lys Glu Arg Tyr Gly Tyr
                165                 170                 175

Lys Thr Glu Asp Leu Ala Phe Ile Gln Asp Asn Ile Lys Lys Ile Thr
            180                 185                 190

Lys Asp Ala Tyr Gly Lys Arg Lys Ser Gln Val Asn Thr Gly Phe Phe
        195                 200                 205

Leu Ser Leu Gln Asp Tyr Asn Gly Asp Thr Gln Lys Lys Leu His Leu
    210                 215                 220

Ser Gly Val Gly Ile Ala Leu Leu Ile Cys Leu Phe Leu Asp Lys Gln
225                 230                 235                 240

Tyr Ile Asn Ile Phe Leu Ser Arg Leu Pro Ile Phe Ser Ser Tyr Asn
                245                 250                 255
```

```
Ala Gln Ser Glu Glu Arg Arg Ile Ile Ile Arg Ser Phe Gly Ile Asn
            260                 265                 270

Ser Ile Lys Leu Pro Lys Asp Arg Ile His Ser Glu Lys Ser Asn Lys
        275                 280                 285

Ser Val Ala Met Asp Met Leu Asn Glu Val Lys Arg Cys Pro Asp Glu
    290                 295                 300

Leu Phe Thr Thr Leu Ser Ala Glu Lys Gln Ser Arg Phe Arg Ile Ile
305                 310                 315                 320

Ser Asp Asp His Asn Glu Val Leu Met Lys Arg Ser Thr Asp Arg Phe
                325                 330                 335

Val Pro Leu Leu Leu Gln Tyr Ile Asp Tyr Gly Lys Leu Phe Asp His
            340                 345                 350

Ile Arg Phe His Val Asn Met Gly Lys Leu Arg Tyr Leu Leu Lys Ala
        355                 360                 365

Asp Lys Thr Cys Ile Asp Gly Gln Thr Arg Val Arg Val Ile Glu Gln
    370                 375                 380

Pro Leu Asn Gly Phe Gly Arg Leu Glu Glu Ala Glu Thr Met Arg Lys
385                 390                 395                 400

Gln Glu Asn Gly Thr Phe Gly Asn Ser Gly Ile Arg Ile Arg Asp Phe
                405                 410                 415

Glu Asn Val Lys Arg Asp Asp Ala Asn Pro Ala Asn Tyr Pro Tyr Ile
            420                 425                 430

Val Asp Thr Tyr Thr His Tyr Ile Leu Glu Asn Asn Lys Val Glu Met
        435                 440                 445

Phe Ile Ser Asp Lys Gly Ser Ser Ala Pro Leu Leu Pro Leu Ile Glu
    450                 455                 460
```

```
Asp Asp Arg Tyr Val Val Lys Thr Ile Pro Ser Cys Arg Met Ser Thr
465                 470                 475                 480


Leu Glu Ile Pro Ala Met Ala Phe His Met Phe Leu Phe Gly Ser Lys
                485                 490                 495


Lys Thr Glu Lys Leu Ile Val Asp Val His Asn Arg Tyr Lys Arg Leu
            500                 505                 510


Phe Gln Ala Met Gln Lys Glu Glu Val Thr Ala Glu Asn Ile Ala Ser
        515                 520                 525


Phe Gly Ile Ala Glu Ser Asp Leu Pro Gln Lys Ile Leu Asp Leu Ile
    530                 535                 540


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


Val Asp Asp Met Leu Thr Asp Thr Glu Arg Arg Ile Lys Arg Phe Lys
                565                 570                 575


Asp Asp Arg Lys Ser Ile Arg Ser Ala Asp Asn Lys Met Gly Lys Arg
            580                 585                 590


Gly Phe Lys Gln Ile Ser Thr Gly Lys Leu Ala Asp Phe Leu Ala Lys
        595                 600                 605


Asp Ile Val Leu Phe Gln Pro Ser Val Asn Asp Gly Glu Asn Lys Ile
    610                 615                 620


Thr Gly Leu Asn Tyr Arg Ile Met Gln Ser Ala Ile Ala Val Tyr Asp
625                 630                 635                 640


Ser Gly Asp Asp Tyr Glu Ala Lys Gln Gln Phe Lys Leu Met Phe Glu
                645                 650                 655


Lys Ala Arg Leu Ile Gly Lys Gly Thr Thr Glu Pro His Pro Phe Leu
            660                 665                 670
```

```
Tyr Lys Val Phe Ala Arg Ser Ile Pro Ala Asn Ala Val Asp Phe Tyr
        675                 680                 685


Glu Arg Tyr Leu Ile Glu Arg Lys Phe Tyr Leu Thr Gly Leu Cys Asn
    690                 695                 700


Glu Ile Lys Arg Gly Asn Arg Val Asp Val Pro Phe Ile Arg Arg Asp
705                 710                 715                 720


Gln Asn Lys Trp Lys Thr Pro Ala Met Lys Thr Leu Gly Arg Ile Tyr
                725                 730                 735


Ser Glu Asp Leu Pro Val Glu Leu Pro Arg Gln Met Phe Asp Asn Glu
            740                 745                 750


Ile Lys Ser His Leu Lys Ser Leu Pro Gln Met Glu Gly Ile Asp Phe
        755                 760                 765


Asn Asn Ala Asn Val Thr Tyr Leu Ile Ala Glu Tyr Met Lys Arg Val
    770                 775                 780


Leu Asn Asp Asp Phe Gln Thr Phe Tyr Gln Trp Lys Arg Asn Tyr His
785                 790                 795                 800


Tyr Met Asp Met Leu Lys Gly Glu Tyr Asp Arg Lys Gly Ser Leu Gln
                805                 810                 815


His Cys Phe Thr Ser Val Glu Glu Arg Glu Gly Leu Trp Lys Glu Arg
            820                 825                 830


Ala Ser Arg Thr Glu Arg Tyr Arg Lys Leu Ala Ser Asn Lys Ile Arg
        835                 840                 845


Ser Asn Arg Gln Met Arg Asn Ala Ser Ser Glu Glu Ile Glu Thr Ile
    850                 855                 860


Leu Asp Lys Arg Leu Ser Asn Cys Arg Asn Glu Tyr Gln Lys Ser Glu
865                 870                 875                 880
```

```
Lys Val Ile Arg Arg Tyr Arg Val Gln Asp Ala Leu Leu Phe Leu Leu
                885                 890                 895


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


Lys Leu Lys Glu Ile Met Pro Asp Ala Glu Lys Gly Ile Leu Ser Glu
        915                 920                 925


Ile Met Pro Met Ser Phe Thr Phe Glu Lys Gly Gly Lys Lys Tyr Thr
    930                 935                 940


Ile Thr Ser Glu Gly Met Lys Leu Lys Asn Tyr Gly Asp Phe Phe Val
945                 950                 955                 960


Leu Ala Ser Asp Lys Arg Ile Gly Asn Leu Leu Glu Leu Val Gly Ser
                965                 970                 975


Asp Ile Val Ser Lys Glu Asp Ile Met Glu Glu Phe Asn Lys Tyr Asp
            980                 985                 990


Gln Cys Arg Pro Glu Ile Ser Ser  Ile Val Phe Asn Leu  Glu Lys Trp
        995                 1000                1005


Ala Phe  Asp Thr Tyr Pro Glu  Leu Ser Ala Arg Val  Asp Arg Glu
    1010                1015                1020


Glu Lys  Val Asp Phe Lys Ser  Ile Leu Lys Ile Leu  Leu Asn Asn
    1025                1030                1035


Lys Asn  Ile Asn Lys Glu Gln  Ser Asp Ile Leu Arg  Lys Ile Arg
    1040                1045                1050


Asn Ala  Phe Asp His Asn Asn  Tyr Pro Asp Lys Gly  Ile Val Glu
    1055                1060                1065


Ile Lys  Ala Leu Pro Glu Ile  Ala Met Ser Ile Lys  Lys Ala Phe
    1070                1075                1080
```

Gly Glu Tyr Ala Ile Met Lys
    1085   1090

<210> 138
<211> 1090
<212> PRT
<213> Prevotella sp.

<400> 138

Met Asn Ile Pro Ala Leu Val Glu Asn Gln Lys Lys Tyr Phe Gly Thr
1   5   10   15

Tyr Ser Val Met Ala Met Leu Asn Ala Gln Thr Val Leu Asp His Ile
            20   25   30

Gln Lys Val Ala Asp Ile Glu Gly Glu Gln Asn Glu Asn Asn Glu Asn
        35   40   45

Leu Trp Phe His Pro Val Met Ser His Leu Tyr Asn Ala Lys Asn Gly
    50   55   60

Tyr Asp Lys Gln Pro Glu Lys Thr Met Phe Ile Ile Glu Arg Leu Gln
65   70   75   80

Ser Tyr Phe Pro Phe Leu Lys Ile Met Ala Glu Asn Gln Arg Glu Tyr
                85   90   95

Ser Asn Gly Lys Tyr Lys Gln Asn Arg Val Glu Val Asn Ser Asn Asp
            100   105   110

Ile Phe Glu Val Leu Lys Arg Ala Phe Gly Val Leu Lys Met Tyr Arg
        115   120   125

Asp Gln Ala Ser His Tyr Lys Thr Tyr Asp Glu Lys Leu Ile Asp Gly
    130   135   140

Cys Glu Phe Leu Thr Ser Thr Glu Gln Pro Leu Ser Gly Met Ile Asn
145   150   155   160

Asn Tyr Tyr Thr Val Ala Leu Arg Asn Met Asn Glu Arg Tyr Gly Tyr

```
                165                 170                 175

Lys Thr Glu Asp Leu Ala Phe Ile Gln Asp Lys Arg Phe Lys Phe Val
            180                 185                 190

Lys Asp Ala Tyr Gly Lys Lys Lys Ser Gln Val Asn Thr Gly Phe Phe
        195                 200                 205

Leu Ser Leu Gln Asp Tyr Asn Gly Asp Thr Gln Lys Lys Leu His Leu
    210                 215                 220

Ser Gly Val Gly Ile Ala Leu Leu Ile Cys Leu Phe Leu Asp Lys Gln
225                 230                 235                 240

Tyr Ile Asn Ile Phe Leu Ser Arg Leu Pro Ile Phe Ser Ser Tyr Asn
                245                 250                 255

Ala Gln Ser Glu Glu Arg Arg Ile Ile Ile Arg Ser Phe Gly Ile Asn
            260                 265                 270

Ser Ile Lys Gln Pro Lys Asp Arg Ile His Ser Glu Lys Ser Asn Lys
        275                 280                 285

Ser Val Ala Met Asp Met Leu Asn Glu Ile Lys Arg Cys Pro Asn Glu
    290                 295                 300

Leu Phe Glu Thr Leu Ser Ala Glu Lys Gln Ser Arg Phe Arg Ile Ile
305                 310                 315                 320

Ser Asn Asp His Asn Glu Val Leu Met Lys Arg Ser Ser Asp Arg Phe
                325                 330                 335

Val Pro Leu Leu Leu Gln Tyr Ile Asp Tyr Gly Lys Leu Phe Asp His
            340                 345                 350

Ile Arg Phe His Val Asn Met Gly Lys Leu Arg Tyr Leu Leu Lys Ala
        355                 360                 365

Asp Lys Thr Cys Ile Asp Gly Gln Thr Arg Val Arg Val Ile Glu Gln
```

```
    370                 375                 380

Pro Leu Asn Gly Phe Gly Arg Leu Glu Glu Val Glu Thr Met Arg Lys
385                 390                 395                 400

Gln Glu Asn Gly Thr Phe Gly Asn Ser Gly Ile Arg Ile Arg Asp Phe
                405                 410                 415

Glu Asn Met Lys Arg Asp Asp Ala Asn Pro Ala Asn Tyr Pro Tyr Ile
            420                 425                 430

Val Asp Thr Tyr Thr His Tyr Ile Leu Glu Asn Asn Lys Val Glu Met
        435                 440                 445

Phe Ile Ser Asp Glu Glu Thr Pro Ala Pro Leu Leu Pro Val Ile Glu
    450                 455                 460

Asp Asp Arg Tyr Val Val Lys Thr Ile Pro Ser Cys Arg Met Ser Thr
465                 470                 475                 480

Leu Glu Ile Pro Ala Met Ala Phe His Met Phe Leu Phe Gly Ser Lys
                485                 490                 495

Lys Thr Glu Lys Leu Ile Val Asp Val His Asn Arg Tyr Lys Arg Leu
            500                 505                 510

Phe Lys Ala Met Gln Lys Glu Glu Val Thr Ala Glu Asn Ile Ala Ser
        515                 520                 525

Phe Gly Ile Ala Glu Ser Asp Leu Pro Gln Lys Ile Ile Asp Leu Ile
    530                 535                 540

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

Val Asp Asp Met Leu Ala Asp Thr Glu Arg Arg Ile Lys Arg Phe Lys
                565                 570                 575

Asp Asp Arg Lys Ser Ile Arg Ser Ala Asp Asn Lys Met Gly Lys Arg
```

```
            580                 585                 590

Gly Phe Lys Gln Ile Ser Thr Gly Lys Leu Ala Asp Phe Leu Ala Lys
        595                 600                 605

Asp Ile Val Leu Phe Gln Pro Ser Val Asn Asp Gly Glu Asn Lys Ile
    610                 615                 620

Thr Gly Leu Asn Tyr Arg Ile Met Gln Ser Ala Ile Ala Val Tyr Asn
625                 630                 635                 640

Ser Gly Asp Asp Tyr Glu Ala Lys Gln Gln Phe Lys Leu Met Phe Glu
                645                 650                 655

Lys Ala Arg Leu Ile Gly Lys Gly Thr Thr Glu Pro His Pro Phe Leu
            660                 665                 670

Tyr Lys Val Phe Val Arg Ser Ile Pro Ala Asn Ala Val Asp Phe Tyr
        675                 680                 685

Glu Arg Tyr Leu Ile Glu Arg Lys Phe Tyr Leu Ile Gly Leu Ser Asn
    690                 695                 700

Glu Ile Lys Lys Gly Asn Arg Val Asp Val Pro Phe Ile Arg Arg Asp
705                 710                 715                 720

Gln Asn Lys Trp Lys Thr Pro Ala Met Lys Thr Leu Gly Arg Ile Tyr
                725                 730                 735

Asp Glu Asp Leu Pro Val Glu Leu Pro Arg Gln Met Phe Asp Asn Glu
            740                 745                 750

Ile Lys Ser His Leu Lys Ser Leu Pro Gln Met Glu Gly Ile Asp Phe
        755                 760                 765

Asn Asn Ala Asn Val Thr Tyr Leu Ile Ala Glu Tyr Met Lys Arg Val
    770                 775                 780

Leu Asn Asp Asp Phe Gln Thr Phe Tyr Gln Trp Lys Arg Asn Tyr Arg
```

785 790 795 800

Tyr Met Asp Met Leu Arg Gly Glu Tyr Asp Arg Lys Gly Ser Leu Gln
805 810 815

Ser Cys Phe Thr Ser Val Glu Glu Arg Glu Gly Leu Trp Lys Glu Arg
820 825 830

Ala Ser Arg Thr Glu Arg Tyr Arg Lys Leu Ala Ser Asn Lys Ile Arg
835 840 845

Ser Asn Arg Gln Met Arg Asn Ala Ser Ser Glu Glu Ile Glu Thr Ile
850 855 860

Leu Asp Lys Arg Leu Ser Asn Ser Arg Asn Glu Tyr Gln Lys Ser Glu
865 870 875 880

Lys Val Ile Arg Arg Tyr Arg Val Gln Asp Ala Leu Leu Phe Leu Leu
885 890 895

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

Lys Leu Lys Glu Ile Met Pro Asp Ala Glu Lys Gly Ile Leu Ser Glu
915 920 925

Ile Met Pro Met Ser Phe Thr Phe Glu Lys Gly Gly Lys Lys Tyr Thr
930 935 940

Ile Thr Ser Glu Gly Met Lys Leu Lys Asn Tyr Gly Asp Phe Phe Val
945 950 955 960

Leu Ala Ser Asp Lys Arg Ile Gly Asn Leu Leu Glu Leu Val Gly Ser
965 970 975

Asp Thr Val Ser Lys Glu Asp Ile Met Glu Glu Phe Lys Lys Tyr Asp
980 985 990

Gln Cys Arg Pro Glu Ile Ser Ser Ile Val Phe Asn Leu Glu Lys Trp

995 1000 1005

Ala Phe Asp Thr Tyr Pro Glu Leu Ser Ala Arg Val Asp Arg Glu
1010 1015 1020

Glu Lys Val Asp Phe Lys Ser Ile Leu Lys Ile Leu Leu Asn Asn
1025 1030 1035

Lys Asn Ile Asn Lys Glu Gln Ser Asp Ile Leu Arg Lys Ile Arg
1040 1045 1050

Asn Ala Phe Asp His Asn Asn Tyr Pro Asp Lys Gly Val Val Glu
1055 1060 1065

Ile Arg Ala Leu Pro Glu Ile Ala Met Ser Ile Lys Lys Ala Phe
1070 1075 1080

Gly Glu Tyr Ala Ile Met Lys
1085 1090

<210> 139
<211> 1090
<212> PRT
<213> Prevotella sp.

<400> 139

Met Asn Ile Pro Ala Leu Val Glu Asn Gln Lys Lys Tyr Phe Gly Thr
1 5 10 15

Tyr Ser Val Met Ala Met Leu Asn Ala Gln Thr Val Leu Asp His Ile
20 25 30

Gln Lys Val Ala Asp Ile Glu Gly Glu Gln Asn Glu Asn Asn Glu Asn
35 40 45

Leu Trp Phe His Pro Val Met Ser His Leu Tyr Asn Ala Lys Asn Gly
50 55 60

Tyr Asp Lys Gln Pro Glu Lys Thr Met Phe Ile Ile Glu Arg Leu Gln
65 70 75 80

```
Ser Tyr Phe Pro Phe Leu Lys Ile Met Ala Glu Asn Gln Arg Glu Tyr
                85                  90                  95

Ser Asn Gly Lys Tyr Lys Gln Asn Arg Val Glu Val Asn Ser Asn Asp
            100                 105                 110

Ile Phe Glu Val Leu Lys Arg Ala Phe Gly Val Leu Lys Met Tyr Arg
        115                 120                 125

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

Cys Glu Phe Leu Thr Ser Thr Glu Gln Pro Phe Ser Gly Met Ile Ser
145                 150                 155                 160

Lys Tyr Tyr Thr Val Ala Leu Arg Asn Thr Lys Glu Arg Tyr Gly Tyr
                165                 170                 175

Lys Ala Glu Asp Leu Ala Phe Ile Gln Asp Asn Arg Tyr Lys Phe Thr
            180                 185                 190

Lys Asp Ala Tyr Gly Lys Arg Lys Ser Gln Val Asn Thr Gly Ser Phe
        195                 200                 205

Leu Ser Leu Gln Asp Tyr Asn Gly Asp Thr Thr Lys Lys Leu His Leu
    210                 215                 220

Ser Gly Val Gly Ile Ala Leu Leu Ile Cys Leu Phe Leu Asp Lys Gln
225                 230                 235                 240

Tyr Ile Asn Leu Phe Leu Ser Arg Leu Pro Ile Phe Ser Ser Tyr Asn
                245                 250                 255

Ala Gln Ser Glu Glu Arg Arg Ile Ile Ile Arg Ser Phe Gly Ile Asn
            260                 265                 270

Ser Ile Lys Gln Pro Lys Asp Arg Ile His Ser Glu Lys Ser Asn Lys
        275                 280                 285
```

```
Ser Val Ala Met Asp Met Leu Asn Glu Val Lys Arg Cys Pro Asp Glu
    290                 295                 300

Leu Phe Thr Thr Leu Ser Ala Glu Lys Gln Ser Arg Phe Arg Ile Ile
305                 310                 315                 320

Ser Asp Asp His Asn Glu Val Leu Met Lys Arg Ser Ser Asp Arg Phe
                325                 330                 335

Val Pro Leu Leu Leu Gln Tyr Ile Asp Tyr Gly Lys Leu Phe Asp His
            340                 345                 350

Ile Arg Phe His Val Asn Met Gly Lys Leu Arg Tyr Leu Leu Lys Ala
        355                 360                 365

Asp Lys Thr Cys Ile Asp Gly Gln Thr Arg Val Arg Val Ile Glu Gln
    370                 375                 380

Pro Leu Asn Gly Phe Gly Arg Leu Glu Glu Val Glu Thr Met Arg Lys
385                 390                 395                 400

Gln Glu Asn Gly Thr Phe Gly Asn Ser Gly Ile Arg Ile Arg Asp Phe
                405                 410                 415

Glu Asn Met Lys Arg Asp Asp Ala Asn Pro Ala Asn Tyr Pro Tyr Ile
            420                 425                 430

Val Glu Thr Tyr Thr His Tyr Ile Leu Glu Asn Asn Lys Val Glu Met
        435                 440                 445

Phe Ile Ser Asp Glu Glu Asn Pro Thr Pro Leu Leu Pro Val Ile Glu
    450                 455                 460

Asp Asp Arg Tyr Val Val Lys Thr Ile Pro Ser Cys Arg Met Ser Thr
465                 470                 475                 480

Leu Glu Ile Pro Ala Met Ala Phe His Met Phe Leu Phe Gly Ser Glu
                485                 490                 495
```

```
Lys Thr Glu Lys Leu Ile Ile Asp Val His Asp Arg Tyr Lys Arg Leu
            500                 505                 510

Phe Gln Ala Met Gln Lys Glu Glu Val Thr Ala Glu Asn Ile Ala Ser
        515                 520                 525

Phe Gly Ile Ala Glu Ser Asp Leu Pro Gln Lys Ile Met Asp Leu Ile
    530                 535                 540

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

Val Asp Asp Met Leu Thr Asp Thr Glu Arg Arg Ile Lys Arg Phe Lys
                565                 570                 575

Asp Asp Arg Lys Ser Ile Arg Ser Ala Asp Asn Lys Met Gly Lys Arg
            580                 585                 590

Gly Phe Lys Gln Ile Ser Thr Gly Lys Leu Ala Asp Phe Leu Ala Lys
        595                 600                 605

Asp Ile Val Leu Phe Gln Pro Ser Val Asn Asp Gly Glu Asn Lys Ile
    610                 615                 620

Thr Gly Leu Asn Tyr Arg Ile Met Gln Ser Ala Ile Ala Val Tyr Asp
625                 630                 635                 640

Ser Gly Asp Asp Tyr Glu Ala Lys Gln Gln Phe Lys Leu Met Phe Glu
                645                 650                 655

Lys Ala Arg Leu Ile Gly Lys Gly Thr Thr Glu Pro His Pro Phe Leu
            660                 665                 670

Tyr Lys Val Phe Val Arg Ser Ile Pro Ala Asn Ala Val Asp Phe Tyr
        675                 680                 685

Glu Arg Tyr Leu Ile Glu Arg Lys Phe Tyr Leu Ile Gly Leu Ser Asn
    690                 695                 700
```

```
Glu Ile Lys Lys Gly Asn Arg Val Asp Val Pro Phe Ile Arg Arg Asp
705                 710                 715                 720

Gln Asn Lys Trp Lys Thr Pro Ala Met Lys Thr Leu Gly Arg Ile Tyr
                725                 730                 735

Ser Glu Asp Leu Pro Val Glu Leu Pro Arg Gln Met Phe Asp Asn Glu
            740                 745                 750

Ile Lys Ser His Leu Lys Ser Leu Pro Gln Met Glu Gly Ile Asp Phe
        755                 760                 765

Asn Asn Ala Asn Val Thr Tyr Leu Ile Ala Glu Tyr Met Lys Arg Val
    770                 775                 780

Leu Asn Asp Asp Phe Gln Thr Phe Tyr Gln Trp Lys Arg Asn Tyr Arg
785                 790                 795                 800

Tyr Met Asp Met Leu Arg Gly Glu Tyr Asp Arg Lys Gly Ser Leu Gln
                805                 810                 815

His Cys Phe Thr Ser Ile Glu Glu Arg Glu Gly Leu Trp Lys Glu Arg
            820                 825                 830

Ala Ser Arg Thr Glu Arg Tyr Arg Lys Leu Ala Ser Asn Lys Ile Arg
        835                 840                 845

Ser Asn Arg Gln Met Arg Asn Ala Ser Ser Glu Glu Ile Glu Thr Ile
    850                 855                 860

Leu Asp Lys Arg Leu Ser Asn Cys Arg Asn Glu Tyr Gln Lys Ser Glu
865                 870                 875                 880

Lys Ile Ile Arg Arg Tyr Arg Val Gln Asp Ala Leu Leu Phe Leu Leu
                885                 890                 895

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

```
Lys Leu Lys Glu Ile Met Pro Asp Ala Glu Lys Gly Ile Leu Ser Glu
        915                 920                 925


Ile Met Pro Met Ser Phe Thr Phe Glu Lys Gly Gly Lys Ile Tyr Thr
    930                 935                 940


Ile Thr Ser Gly Gly Met Lys Leu Lys Asn Tyr Gly Asp Phe Phe Val
945                 950                 955                 960


Leu Ala Ser Asp Lys Arg Ile Gly Asn Leu Leu Glu Leu Val Gly Ser
                965                 970                 975


Asn Thr Val Ser Lys Glu Asp Ile Met Glu Glu Phe Lys Lys Tyr Asp
            980                 985                 990


Gln Cys Arg Pro Glu Ile Ser Ser  Ile Val Phe Asn Leu  Glu Lys Trp
        995                 1000                1005


Ala Phe  Asp Thr Tyr Pro Glu  Leu Pro Ala Arg Val  Asp Arg Lys
    1010                1015                1020


Glu Lys  Val Asp Phe Trp Ser  Ile Leu Asp Val Leu  Ser Asn Asn
    1025                1030                1035


Lys Asp  Ile Asn Asn Glu Gln  Ser Tyr Ile Leu Arg  Lys Ile Arg
    1040                1045                1050


Asn Ala  Phe Asp His Asn Asn  Tyr Pro Asp Lys Gly  Ile Val Glu
    1055                1060                1065


Ile Lys  Ala Leu Pro Glu Ile  Ala Met Ser Ile Lys  Lys Ala Phe
    1070                1075                1080


Gly Glu  Tyr Ala Ile Met Lys
    1085                1090


<210>  140
<211>  1131
```

<212> PRT
<213> Phaeodactylibacter xiamenensis

<400> 140

Met Thr Asn Thr Pro Lys Arg Arg Thr Leu His Arg His Pro Ser Tyr
1 5 10 15

Phe Gly Ala Phe Leu Asn Ile Ala Arg His Asn Ala Phe Met Ile Met
20 25 30

Glu His Leu Ser Thr Lys Tyr Asp Met Glu Asp Lys Asn Thr Leu Asp
35 40 45

Glu Ala Gln Leu Pro Asn Ala Lys Leu Phe Gly Cys Leu Lys Lys Arg
50 55 60

Tyr Gly Lys Pro Asp Val Thr Glu Gly Val Ser Arg Asp Leu Arg Arg
65 70 75 80

Tyr Phe Pro Phe Leu Asn Tyr Pro Leu Phe Leu His Leu Glu Lys Gln
85 90 95

Gln Asn Ala Glu Gln Ala Ala Thr Tyr Asp Ile Asn Pro Glu Asp Ile
100 105 110

Glu Phe Thr Leu Lys Gly Phe Phe Arg Leu Leu Asn Gln Met Arg Asn
115 120 125

Asn Tyr Ser His Tyr Ile Ser Asn Thr Asp Tyr Gly Lys Phe Asp Lys
130 135 140

Leu Pro Val Gln Asp Ile Tyr Glu Ala Ala Ile Phe Arg Leu Leu Asp
145 150 155 160

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

Arg His Leu Glu Ser Asn Asn Ser Glu Tyr Arg Pro Arg Ser Leu Ala
180 185 190

```
Asn Ser Pro Asp His Glu Asn Thr Val Ala Phe Val Thr Cys Leu Phe
        195                 200                 205

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

Arg Ser Thr Asn Asp Ala Ala Glu Gly Asp Pro Leu Ile Arg Lys Ala
225                 230                 235                 240

Ser His Glu Cys Tyr Thr Met Phe Cys Cys Arg Leu Pro Gln Pro Lys
                245                 250                 255

Leu Glu Ser Ser Asp Ile Leu Leu Asp Met Val Asn Glu Leu Gly Arg
            260                 265                 270

Cys Pro Ser Ala Leu Tyr Asn Leu Leu Ser Glu Glu Asp Gln Ala Arg
        275                 280                 285

Phe His Ile Lys Arg Glu Glu Ile Thr Gly Phe Glu Glu Asp Pro Asp
    290                 295                 300

Glu Glu Leu Glu Gln Glu Ile Val Leu Lys Arg His Ser Asp Arg Phe
305                 310                 315                 320

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Asp Thr Glu Ala Phe Gln Thr
                325                 330                 335

Leu Arg Phe Asp Val Tyr Leu Gly Arg Trp Arg Thr Lys Pro Val Tyr
            340                 345                 350

Lys Lys Arg Ile Tyr Gly Gln Glu Arg Asp Arg Val Leu Thr Gln Ser
        355                 360                 365

Ile Arg Thr Phe Thr Arg Leu Ser Arg Leu Leu Pro Ile Tyr Glu Asn
    370                 375                 380

Val Lys His Asp Ala Val Arg Gln Asn Glu Glu Asp Gly Lys Leu Val
385                 390                 395                 400
```

```
Asn Pro Asp Val Thr Ser Gln Phe His Lys Ser Trp Ile Gln Ile Glu
                405                 410                 415

Ser Asp Asp Arg Ala Phe Leu Ser Asp Arg Ile Glu His Phe Ser Pro
            420                 425                 430

His Tyr Asn Phe Gly Asp Gln Val Ile Gly Leu Lys Phe Ile Asn Pro
        435                 440                 445

Asp Arg Tyr Ala Ala Ile Gln Asn Val Phe Pro Lys Leu Pro Gly Glu
    450                 455                 460

Glu Lys Lys Asp Lys Asp Ala Lys Leu Val Asn Glu Thr Ala Asp Ala
465                 470                 475                 480

Ile Ile Ser Thr His Glu Ile Arg Ser Leu Phe Leu Tyr His Tyr Leu
                485                 490                 495

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

Asp Thr Glu Thr Phe Ile Lys Gln Tyr Ile Asp Thr Ile Lys Leu Phe
        515                 520                 525

Phe Glu Asp Ile Lys Ser Gly Glu Leu Gln Pro Ile Ala Asp Pro Pro
    530                 535                 540

Asn Tyr Gln Lys Asn Glu Pro Leu Pro Tyr Val Arg Gly Asp Lys Glu
545                 550                 555                 560

Lys Thr Gln Glu Glu Arg Ala Gln Tyr Arg Glu Arg Gln Lys Glu Ile
                565                 570                 575

Lys Glu Arg Arg Lys Glu Leu Asn Thr Leu Leu Gln Asn Arg Tyr Gly
            580                 585                 590

Leu Ser Ile Gln Tyr Ile Pro Ser Arg Leu Arg Glu Tyr Leu Leu Gly
        595                 600                 605
```

```
Tyr Lys Lys Val Pro Tyr Glu Lys Leu Ala Leu Gln Lys Leu Arg Ala
    610                 615                 620

Gln Arg Lys Glu Val Lys Lys Arg Ile Lys Asp Ile Glu Lys Met Arg
625                 630                 635                 640

Thr Pro Arg Val Gly Glu Gln Ala Thr Trp Leu Ala Glu Asp Ile Val
                645                 650                 655

Phe Leu Thr Pro Pro Lys Met His Thr Pro Glu Arg Lys Thr Thr Lys
            660                 665                 670

His Pro Gln Lys Leu Asn Asn Asp Gln Phe Arg Ile Met Gln Ser Ser
        675                 680                 685

Leu Ala Tyr Phe Ser Val Asn Lys Lys Ala Ile Lys Lys Phe Phe Gln
    690                 695                 700

Lys Glu Thr Gly Ile Gly Leu Ser Asn Arg Glu Thr Ser His Pro Phe
705                 710                 715                 720

Leu Tyr Arg Ile Asp Val Gly Arg Cys Arg Gly Ile Leu Asp Phe Tyr
                725                 730                 735

Thr Gly Tyr Leu Lys Tyr Lys Met Asp Trp Leu Asp Asp Ala Ile Lys
            740                 745                 750

Lys Val Asp Asn Arg Lys His Gly Lys Lys Glu Ala Lys Lys Tyr Glu
        755                 760                 765

Lys Tyr Leu Pro Ser Ser Ile Gln His Lys Thr Pro Leu Glu Leu Asp
    770                 775                 780

Tyr Thr Arg Leu Pro Val Tyr Leu Pro Arg Gly Leu Phe Lys Lys Ala
785                 790                 795                 800

Ile Val Lys Ala Leu Ala Ala His Ala Asp Phe Gln Val Glu Pro Glu
                805                 810                 815
```

```
Glu Asp Asn Val Ile Phe Cys Leu Asp Gln Leu Leu Asp Gly Asp Thr
            820                 825                 830

Gln Asp Phe Tyr Asn Trp Gln Arg Tyr Tyr Arg Ser Ala Leu Thr Glu
        835                 840                 845

Lys Glu Thr Asp Asn Gln Leu Val Leu Ala His Pro Tyr Ala Glu Gln
    850                 855                 860

Ile Leu Gly Thr Ile Lys Thr Leu Glu Gly Lys Gln Lys Asn Asn Lys
865                 870                 875                 880

Leu Gly Asn Lys Ala Lys Gln Lys Ile Lys Asp Glu Leu Ile Asp Leu
                885                 890                 895

Lys Arg Ala Lys Arg Arg Leu Leu Asp Arg Glu Gln Tyr Leu Arg Ala
            900                 905                 910

Val Gln Ala Glu Asp Arg Ala Leu Trp Leu Met Ile Gln Glu Arg Gln
        915                 920                 925

Lys Gln Lys Ala Glu His Glu Glu Ile Ala Phe Asp Gln Leu Asp Leu
    930                 935                 940

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

Ile Pro Asp Thr Lys Val Asp Ile Thr Asp Lys Leu Pro Leu Arg Arg
                965                 970                 975

Tyr Gly Asp Leu Arg Arg Val Ala Lys Asp Arg Arg Leu Val Asn Leu
            980                 985                 990

Ala Ser Tyr Tyr His Val Ala Gly  Leu Ser Glu Ile Pro  Tyr Asp Leu
        995                 1000                1005

Val Lys  Lys Glu Leu Glu Glu  Tyr Asp Arg Arg Arg  Val Ala Phe
    1010                1015                1020
```

```
Phe Glu  His Val Tyr Gln Phe  Glu Lys Glu Val Tyr  Asp Arg Tyr
    1025                1030                1035


Ala Ala  Glu Leu Arg Asn Glu  Asn Pro Lys Gly Glu  Ser Thr Tyr
    1040                1045                1050


Phe Ser  His Trp Glu Tyr Val  Ala Val Ala Val Lys  His Ser Ala
    1055                1060                1065


Asp Thr  His Phe Asn Glu Leu  Phe Lys Glu Lys Val  Met Gln Leu
    1070                1075                1080


Arg Asn  Lys Phe His His Asn  Glu Phe Pro Tyr Phe  Asp Trp Leu
    1085                1090                1095


Leu Pro  Glu Val Glu Lys Ala  Ser Ala Ala Leu Tyr  Ala Asp Arg
    1100                1105                1110


Val Phe  Asp Val Ala Glu Gly  Tyr Tyr Gln Lys Met  Arg Lys Leu
    1115                1120                1125


Met Arg  Gln
    1130


<210>  141
<211>  1155
<212>  PRT
<213>  Flavorbacterium sp.

<400>  141

Met Asp Asn Asn Ile Thr Val Glu Lys Thr Glu Leu Gly Leu Gly Ile
1               5                  10                  15


Thr Tyr Asn His Asp Lys Val Glu Asp Lys His Tyr Phe Gly Gly Phe
            20                  25                  30


Phe Asn Leu Ala Gln Asn Asn Ile Asp Leu Val Ala Gln Glu Phe Lys
        35                  40                  45
```

```
Lys Arg Leu Leu Ile Gln Gly Lys Asp Ser Ile Asn Ile Phe Ala Asn
    50                  55                  60

Tyr Phe Ser Asp Gln Cys Ser Ile Thr Asn Leu Glu Arg Gly Ile Lys
65                  70                  75                  80

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

Lys Asn Lys Ser Lys Ser Ile Arg Glu His Leu Ile Leu Leu Leu Glu
            100                 105                 110

Thr Ile Asn Asn Leu Arg Asn Tyr Tyr Thr His Tyr Tyr His Lys Lys
        115                 120                 125

Ile Ile Ile Asp Gly Ser Leu Phe Pro Leu Leu Asp Thr Ile Leu Leu
    130                 135                 140

Lys Val Val Leu Glu Ile Lys Lys Lys Lys Leu Lys Glu Asp Lys Thr
145                 150                 155                 160

Lys Gln Leu Leu Lys Lys Gly Leu Glu Lys Glu Met Thr Ile Leu Phe
                165                 170                 175

Asn Leu Met Lys Ala Glu Gln Lys Glu Lys Lys Ile Lys Gly Trp Asn
            180                 185                 190

Ile Asp Glu Asn Ile Lys Gly Ala Val Leu Asn Arg Ala Phe Ser His
        195                 200                 205

Leu Leu Tyr Asn Asp Glu Leu Ser Asp Tyr Arg Lys Ser Lys Tyr Asn
    210                 215                 220

Thr Glu Asp Glu Thr Leu Lys Asp Thr Leu Thr Glu Ser Gly Ile Leu
225                 230                 235                 240

Phe Leu Leu Ser Phe Phe Leu Asn Lys Lys Glu Gln Glu Gln Leu Lys
                245                 250                 255
```

```
Ala Asn Ile Lys Gly Tyr Lys Gly Lys Ile Ala Ser Ile Pro Asp Glu
            260                 265                 270

Glu Ile Thr Leu Lys Asn Asn Ser Leu Arg Asn Met Ala Thr His Trp
        275                 280                 285

Thr Tyr Ser His Leu Thr Tyr Lys Gly Leu Lys His Arg Ile Lys Thr
    290                 295                 300

Asp His Glu Lys Glu Thr Leu Leu Val Asn Met Val Asp Tyr Leu Ser
305                 310                 315                 320

Lys Val Pro His Glu Ile Tyr Gln Asn Leu Ser Glu Gln Asn Lys Ser
                325                 330                 335

Leu Phe Leu Glu Asp Ile Asn Glu Tyr Met Arg Asp Asn Glu Glu Asn
            340                 345                 350

His Asp Ser Ser Glu Ala Ser Arg Val Ile His Pro Val Ile Arg Lys
        355                 360                 365

Arg Tyr Glu Asn Lys Phe Ala Tyr Phe Ala Ile Arg Phe Leu Asp Glu
    370                 375                 380

Phe Ala Glu Phe Pro Thr Leu Arg Phe Met Val Asn Val Gly Asn Tyr
385                 390                 395                 400

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

Asn Arg Thr Ile Lys Gln Gln Ile Asn Val Phe Gly Asn Leu Thr Glu
            420                 425                 430

Ile His Lys Lys Lys Asn Asp Tyr Phe Glu Lys Glu Glu Asn Lys Glu
        435                 440                 445

Lys Thr Leu Glu Trp Glu Leu Phe Pro Asn Pro Ser Tyr His Phe Gln
    450                 455                 460
```

```
Lys Glu Asn Ile Pro Ile Phe Ile Asp Leu Glu Lys Ser Lys Glu Thr
465                 470                 475                 480


Asn Asp Leu Ala Lys Glu Tyr Ala Lys Glu Lys Lys Lys Ile Phe Gly
                485                 490                 495


Ser Ser Arg Lys Lys Gln Gln Asn Thr Ala Lys Lys Asn Arg Glu Thr
            500                 505                 510


Ile Ile Asn Leu Val Phe Asp Lys Tyr Lys Thr Ser Asp Arg Lys Thr
        515                 520                 525


Val Thr Phe Glu Gln Pro Thr Ala Leu Leu Ser Phe Asn Glu Leu Asn
    530                 535                 540


Ser Phe Leu Tyr Ala Phe Leu Val Glu Asn Lys Thr Gly Lys Glu Leu
545                 550                 555                 560


Glu Lys Ile Ile Ile Glu Lys Ile Ala Asn Gln Tyr Gln Ile Leu Lys
                565                 570                 575


Asn Cys Ser Ser Thr Val Asp Lys Thr Asn Asp Asn Ile Pro Lys Ser
            580                 585                 590


Ile Lys Lys Ile Val Asn Thr Thr Thr Asp Ser Phe Tyr Phe Glu Gly
        595                 600                 605


Lys Lys Ile Asp Ile Glu Lys Leu Glu Lys Asp Ile Thr Ile Glu Ile
    610                 615                 620


Glu Lys Thr Asn Glu Lys Leu Glu Thr Ile Lys Glu Asn Glu Glu Ser
625                 630                 635                 640


Ala Gln Asn Tyr Lys Arg Asn Glu Arg Asn Thr Gln Lys Arg Lys Leu
                645                 650                 655


Tyr Arg Lys Tyr Val Phe Phe Thr Asn Glu Ile Gly Ile Glu Ala Thr
            660                 665                 670
```

```
Trp Ile Thr Asn Asp Ile Leu Arg Phe Leu Asp Asn Lys Glu Asn Trp
        675                 680                 685

Lys Gly Tyr Gln His Ser Glu Leu Gln Lys Phe Ile Ser Gln Tyr Asp
    690                 695                 700

Asn Tyr Lys Lys Glu Ala Leu Gly Leu Leu Glu Ser Glu Trp Asn Leu
705                 710                 715                 720

Glu Ser Asp Ala Phe Phe Gly Gln Asn Leu Lys Arg Met Phe Gln Ser
                725                 730                 735

Asn Ser Thr Phe Glu Thr Phe Tyr Lys Lys Tyr Leu Asp Asn Arg Lys
            740                 745                 750

Asn Thr Leu Glu Thr Tyr Leu Ser Ala Ile Glu Asn Leu Lys Thr Met
        755                 760                 765

Thr Asp Val Arg Pro Lys Val Leu Lys Lys Lys Trp Thr Glu Leu Phe
    770                 775                 780

Arg Phe Phe Asp Lys Lys Ile Tyr Leu Leu Ser Thr Ile Glu Thr Lys
785                 790                 795                 800

Ile Asn Glu Leu Ile Thr Lys Pro Ile Asn Leu Ser Arg Gly Ile Phe
                805                 810                 815

Glu Glu Lys Pro Thr Phe Ile Asn Gly Lys Asn Pro Asn Lys Glu Asn
            820                 825                 830

Asn Gln His Leu Phe Ala Asn Trp Phe Ile Tyr Ala Lys Lys Gln Thr
        835                 840                 845

Ile Leu Gln Asp Phe Tyr Asn Leu Pro Leu Glu Gln Pro Lys Ala Ile
    850                 855                 860

Thr Asn Leu Lys Lys His Lys Tyr Lys Leu Glu Arg Ser Ile Asn Asn
865                 870                 875                 880
```

```
Leu Lys Ile Glu Asp Ile Tyr Ile Lys Gln Met Val Asp Phe Leu Tyr
                885                 890                 895


Gln Lys Leu Phe Glu Gln Ser Phe Ile Gly Ser Leu Gln Asp Leu Tyr
            900                 905                 910


Thr Ser Lys Glu Lys Arg Glu Ile Glu Lys Gly Lys Ala Lys Asn Glu
        915                 920                 925


Gln Thr Pro Asp Glu Ser Phe Ile Trp Lys Lys Gln Val Glu Ile Asn
    930                 935                 940


Thr His Asn Gly Arg Ile Ile Ala Lys Thr Lys Ile Lys Asp Ile Gly
945                 950                 955                 960


Lys Phe Lys Asn Leu Leu Thr Asp Asn Lys Ile Ala His Leu Ile Ser
                965                 970                 975


Tyr Asp Asp Arg Ile Trp Asp Phe Ser Leu Asn Asn Asp Gly Asp Ile
            980                 985                 990


Thr Lys Lys Leu Tyr Ser Ile Asn  Thr Glu Leu Glu Ser  Tyr Glu Thr
        995                 1000                1005


Ile Arg  Arg Glu Lys Leu Leu  Lys Gln Ile Gln Gln  Phe Glu Gln
    1010                1015                1020


Phe Leu  Leu Glu Gln Glu Thr  Glu Tyr Ser Ala Glu  Arg Lys His
    1025                1030                1035


Pro Glu  Lys Phe Glu Lys Asp  Cys Asn Pro Asn Phe  Lys Lys Tyr
    1040                1045                1050


Ile Ile  Glu Gly Val Leu Asn  Lys Ile Ile Pro Asn  His Glu Ile
    1055                1060                1065


Glu Glu  Ile Glu Ile Leu Lys  Ser Lys Glu Asp Val  Phe Lys Ile
    1070                1075                1080
```

```
Asn Phe  Ser Asp Ile Leu Ile  Leu Asn Asn Asp Asn  Ile Lys Lys
    1085                1090                1095


Gly Tyr  Leu Leu Ile Met Ile  Arg Asn Lys Phe Ala  His Asn Gln
    1100                1105                1110


Leu Ile  Asp Lys Asn Leu Phe  Asn Phe Ser Leu Gln  Leu Tyr Ser
    1115                1120                1125


Lys Asn  Glu Asn Glu Asn Phe  Ser Glu Tyr Leu Asn  Lys Val Cys
    1130                1135                1140


Gln Asn  Ile Ile Gln Glu Phe  Lys Glu Lys Leu Lys
    1145                1150                1155


<210>  142
<211>  1175
<212>  PRT
<213>  Porphyromonas gulae

<400>  142

Met Thr Glu Gln Ser Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1               5                   10                  15


Glu Asp Lys His Phe Trp Ala Ala Phe Leu Asn Leu Ala Arg His Asn
            20                  25                  30


Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
        35                  40                  45


Ala Asp Ile Thr Asn Asp Gln Asp Val Leu Ser Phe Lys Ala Leu Trp
    50                  55                  60


Lys Asn Phe Asp Asn Asp Leu Glu Arg Lys Ser Arg Leu Arg Ser Leu
65                  70                  75                  80


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


Leu Phe Glu Ser Lys Ser Ser Gly Asn Lys Ser Ser Lys Asn Lys Glu
```

100 105 110

Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
115 120 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
130 135 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Ser Glu Ser Ser Glu Leu
145 150 155 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
165 170 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val
180 185 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
195 200 205

Arg Tyr Gly His Asn Asp Asn Pro Ser Phe Lys His His Phe Val Asp
210 215 220

Ser Glu Gly Met Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225 230 235 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
245 250 255

Phe Lys Gly Gly Thr Glu Thr Tyr Gln Gln Met Thr Asn Glu Val Phe
260 265 270

Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
275 280 285

Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys
290 295 300

Pro Lys Pro Leu Tyr Asp Arg Leu Arg Glu Lys Asp Arg Ala Arg Phe

305 310 315 320

Arg Val Pro Val Asp Ile Leu Pro Asp Glu Asp Asp Thr Asp Gly Gly
325 330 335

Gly Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
340 345 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
355 360 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
370 375 380

Lys Met Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385 390 395 400

Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
405 410 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
420 425 430

Lys Pro Tyr Ile Ser Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly
435 440 445

Lys Ile Gly Leu Arg Phe Met Pro Glu Gly Gln His Leu Trp Pro Ser
450 455 460

Pro Glu Val Gly Thr Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465 470 475 480

Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
485 490 495

Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val
500 505 510

Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val

```
        515                 520                 525

Tyr Ala Ile Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Leu Lys
    530                 535                 540

Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu
545                 550                 555                 560

Pro Lys Gln Met Ile Ala Ile Leu Ser Gln Glu His Lys Asp Met Glu
                565                 570                 575

Glu Lys Ile Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His
            580                 585                 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly
        595                 600                 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
    610                 615                 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly
625                 630                 635                 640

Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
                645                 650                 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Lys Arg Leu Thr Pro
            660                 665                 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
        675                 680                 685

Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
    690                 695                 700

Arg Ser Tyr Leu Arg Ala Arg Lys Ala Phe Leu Glu Arg Ile Gly Arg
705                 710                 715                 720

Ser Asp Arg Met Glu Asn Arg Pro Phe Leu Leu Leu Lys Glu Pro Lys
```

```
                725                 730                 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Ser Glu Phe His Leu
            740                 745                 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
        755                 760                 765

Gly Tyr Asp Glu Val Gly Ser Tyr Arg Glu Val Gly Phe Met Ala Lys
    770                 775                 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Glu Asp Arg Val Gln Pro
785                 790                 795                 800

Phe Tyr Asp Ser Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
                805                 810                 815

Gly Arg Phe Leu Ser Lys Glu Glu Arg Ala Glu Glu Trp Glu Arg Gly
            820                 825                 830

Lys Glu Arg Phe Arg Asp Leu Glu Ala Trp Ser His Ser Ala Ala Arg
        835                 840                 845

Arg Ile Glu Asp Ala Phe Ala Gly Ile Glu Tyr Ala Ser Pro Gly Asn
    850                 855                 860

Lys Lys Lys Ile Glu Gln Leu Leu Arg Asp Leu Ser Leu Trp Glu Ala
865                 870                 875                 880

Phe Glu Ser Lys Leu Lys Val Arg Ala Asp Lys Ile Asn Leu Ala Lys
                885                 890                 895

Leu Lys Lys Glu Ile Leu Glu Ala Gln Glu His Pro Tyr His Asp Phe
            900                 905                 910

Lys Ser Trp Gln Lys Phe Glu Arg Glu Leu Arg Leu Val Lys Asn Gln
        915                 920                 925

Asp Ile Ile Thr Trp Met Met Cys Arg Asp Leu Met Glu Glu Asn Lys
```

```
    930                 935                 940

Val Glu Gly Leu Asp Thr Gly Thr Leu Tyr Leu Lys Asp Ile Arg Pro
945                 950                 955                 960

Asn Val Gln Glu Gln Gly Ser Leu Asn Val Leu Asn Arg Val Lys Pro
                965                 970                 975

Met Arg Leu Pro Val Val Val Tyr Arg Ala Asp Ser Arg Gly His Val
            980                 985                 990

His Lys Glu Glu Ala Pro Leu Ala  Thr Val Tyr Ile Glu  Glu Arg Asp
        995                 1000                1005

Thr Lys  Leu Leu Lys Gln Gly  Asn Phe Lys Ser Phe  Val Lys Asp
    1010                1015                1020

Arg Arg  Leu Asn Gly Leu Phe  Ser Phe Val Asp Thr  Gly Gly Leu
    1025                1030                1035

Ala Met  Glu Gln Tyr Pro Ile  Ser Lys Leu Arg Val  Glu Tyr Glu
    1040                1045                1050

Leu Ala  Lys Tyr Gln Thr Ala  Arg Val Cys Val Phe  Glu Leu Thr
    1055                1060                1065

Leu Arg  Leu Glu Glu Ser Leu  Leu Thr Arg Tyr Pro  His Leu Pro
    1070                1075                1080

Asp Glu  Ser Phe Arg Lys Met  Leu Glu Ser Trp Ser  Asp Pro Leu
    1085                1090                1095

Leu Ala  Lys Trp Pro Glu Leu  His Gly Lys Val Arg  Leu Leu Ile
    1100                1105                1110

Ala Val  Arg Asn Ala Phe Ser  His Asn Gln Tyr Pro  Met Tyr Asp
    1115                1120                1125

Glu Ala  Val Phe Ser Ser Ile  Arg Lys Tyr Asp Pro  Ser Ser Pro
```

1130 1135 1140

Asp Ala Ile Glu Glu Arg Met Gly Leu Asn Ile Ala His Arg Leu
1145 1150 1155

Ser Glu Glu Val Lys Gln Ala Lys Glu Thr Val Glu Arg Ile Ile
1160 1165 1170

Gln Val
1175

<210> 143
<211> 1145
<212> PRT
<213> Chryseobacterium sp.

<400> 143

Met Glu Thr Gln Thr Ile Gly His Gly Ile Ala Tyr Asp His Ser Lys
1 5 10 15

Ile Gln Asp Lys His Phe Phe Gly Gly Phe Leu Asn Leu Ala Glu Asn
20 25 30

Asn Ile Lys Ala Val Leu Lys Ala Phe Ser Glu Lys Phe Asn Val Gly
35 40 45

Asn Val Asp Val Lys Gln Phe Ala Asp Val Ser Leu Lys Asp Asn Leu
50 55 60

Pro Asp Asn Asp Phe Gln Lys Arg Val Ser Phe Leu Lys Met Tyr Phe
65 70 75 80

Pro Val Val Asp Phe Ile Asn Ile Pro Asn Asn Arg Ala Lys Phe Arg
85 90 95

Ser Asp Leu Thr Thr Leu Phe Lys Ser Val Asp Gln Leu Arg Asn Phe
100 105 110

Tyr Thr His Tyr Tyr His Lys Pro Leu Asp Phe Asp Ala Ser Leu Phe
115 120 125

```
Ile Leu Leu Asp Asp Ile Phe Ala Arg Thr Ala Lys Glu Val Arg Asp
    130                 135                 140

Gln Lys Met Lys Asp Asp Lys Thr Arg Gln Leu Leu Ser Lys Ser Leu
145                 150                 155                 160

Ser Glu Glu Leu Gln Lys Gly Tyr Glu Leu Gln Leu Glu Arg Leu Lys
                165                 170                 175

Glu Leu Asn Arg Leu Gly Lys Lys Val Asn Ile His Asp Gln Leu Gly
            180                 185                 190

Ile Lys Asn Gly Val Leu Asn Asn Ala Phe Asn His Leu Ile Tyr Lys
        195                 200                 205

Asp Gly Glu Ser Phe Lys Thr Lys Leu Thr Tyr Ser Ser Ala Leu Thr
    210                 215                 220

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

Leu Phe Leu Leu Ser Met Phe Leu Lys Arg Lys Glu Ile Glu Asp Leu
                245                 250                 255

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

Gly Lys Val Asn Gly Leu Lys Phe Met Ala Thr His Trp Val Phe Ser
        275                 280                 285

Tyr Leu Cys Phe Lys Gly Leu Lys Ser Lys Leu Ser Thr Glu Phe His
    290                 295                 300

Glu Glu Thr Leu Leu Ile Gln Ile Ile Asp Glu Leu Ser Lys Val Pro
305                 310                 315                 320

Asp Glu Leu Tyr Cys Ala Phe Asp Lys Glu Thr Arg Asp Lys Phe Ile
                325                 330                 335
```

```
Glu Asp Ile Asn Glu Tyr Val Lys Glu Gly His Gln Asp Phe Ser Leu
            340                 345                 350

Glu Asp Ala Lys Val Ile His Pro Val Ile Arg Lys Arg Tyr Glu Asn
        355                 360                 365

Lys Phe Asn Tyr Phe Ala Ile Arg Phe Leu Asp Glu Phe Val Lys Phe
    370                 375                 380

Pro Ser Leu Arg Phe Gln Val His Val Gly Asn Tyr Val His Asp Arg
385                 390                 395                 400

Arg Ile Lys Asn Ile Asp Gly Thr Thr Phe Glu Thr Glu Arg Val Val
                405                 410                 415

Lys Asp Arg Ile Lys Val Phe Gly Arg Leu Ser Glu Ile Ser Ser Tyr
            420                 425                 430

Lys Ala Gln Tyr Leu Ser Ser Val Ser Asp Lys His Asp Glu Thr Gly
        435                 440                 445

Trp Glu Ile Phe Pro Asn Pro Ser Tyr Val Phe Ile Asn Asn Asn Ile
    450                 455                 460

Pro Ile His Ile Ser Val Asp Thr Ser Phe Lys Lys Glu Ile Ala Asp
465                 470                 475                 480

Phe Lys Lys Leu Arg Arg Ala Gln Val Pro Asp Glu Leu Lys Ile Arg
                485                 490                 495

Gly Ala Glu Lys Lys Arg Lys Phe Glu Ile Thr Gln Met Ile Gly Ser
            500                 505                 510

Lys Ser Val Leu Asn Gln Glu Glu Pro Ile Ala Leu Leu Ser Leu Asn
        515                 520                 525

Glu Ile Pro Ala Leu Leu Tyr Glu Ile Leu Ile Asn Gly Lys Glu Pro
    530                 535                 540
```

```
Ala Glu Ile Glu Arg Ile Ile Lys Asp Lys Leu Asn Glu Arg Gln Asp
545                 550                 555                 560


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


Ser Arg Arg Leu Arg Ser Asn Lys Gly Glu Arg Ile Ile Asn Thr Asp
            580                 585                 590


Lys Leu Leu Gln Leu Val Thr Lys Glu Leu Leu Val Thr Glu Gln Lys
        595                 600                 605


Leu Lys Ile Ile Ser Asp Asn Arg Glu Ala Leu Lys Gln Lys Lys Glu
    610                 615                 620


Gly Lys Tyr Ile Arg Lys Phe Ile Phe Thr Asn Ser Glu Leu Gly Arg
625                 630                 635                 640


Glu Ala Ile Trp Leu Ala Asp Asp Ile Lys Arg Phe Met Pro Ala Asp
                645                 650                 655


Val Arg Lys Glu Trp Lys Gly Tyr Gln His Ser Gln Leu Gln Gln Ser
            660                 665                 670


Leu Ala Phe Tyr Asn Ser Arg Pro Lys Glu Ala Leu Ala Ile Leu Glu
        675                 680                 685


Ser Ser Trp Asn Leu Lys Asp Glu Lys Ile Ile Trp Asn Glu Trp Ile
    690                 695                 700


Leu Lys Ser Phe Thr Gln Asn Lys Phe Phe Asp Ala Phe Tyr Asn Glu
705                 710                 715                 720


Tyr Leu Lys Gly Arg Lys Lys Tyr Phe Ala Phe Leu Ser Glu His Ile
                725                 730                 735


Val Gln Tyr Thr Ser Asn Ala Lys Asn Leu Gln Lys Phe Ile Lys Gln
            740                 745                 750
```

```
Gln Met Pro Lys Asp Leu Phe Glu Lys Arg His Tyr Ile Ile Glu Asp
        755                 760                 765

Leu Gln Thr Glu Lys Asn Lys Ile Leu Ser Lys Pro Phe Ile Phe Pro
    770                 775                 780

Arg Gly Ile Phe Asp Lys Lys Pro Thr Phe Ile Lys Gly Val Lys Val
785                 790                 795                 800

Glu Asp Ser Pro Glu Ser Phe Ala Asn Trp Tyr Gln Tyr Gly Tyr Gln
                805                 810                 815

Lys Asp His Gln Phe Gln Lys Phe Tyr Asp Trp Lys Arg Asp Tyr Ser
            820                 825                 830

Asp Val Phe Leu Glu His Leu Gly Lys Pro Phe Ile Asn Asn Gly Asp
        835                 840                 845

Arg Arg Thr Leu Gly Met Glu Glu Leu Lys Glu Arg Ile Ile Ile Lys
    850                 855                 860

Gln Asp Leu Lys Ile Lys Lys Ile Lys Ile Gln Asp Leu Phe Leu Arg
865                 870                 875                 880

Leu Ile Ala Glu Asn Leu Phe Gln Lys Val Phe Lys Tyr Ser Ala Lys
                885                 890                 895

Leu Pro Leu Ser Asp Phe Tyr Leu Thr Gln Glu Glu Arg Met Glu Lys
            900                 905                 910

Glu Asn Met Ala Ala Leu Gln Asn Val Arg Glu Glu Gly Asp Lys Ser
        915                 920                 925

Pro Asn Ile Ile Lys Asp Asn Phe Ile Trp Ser Lys Met Ile Pro Tyr
    930                 935                 940

Lys Lys Gly Gln Ile Ile Glu Asn Ala Val Lys Leu Lys Asp Ile Gly
945                 950                 955                 960
```

```
Lys Leu Asn Val Leu Ser Leu Asp Asp Lys Val Gln Thr Leu Leu Ser
               965                 970                 975

Tyr Asp Asp Ala Lys Pro Trp Ser Lys Ile Ala Leu Glu Asn Glu Phe
           980                 985                 990

Ser Ile Gly Glu Asn Ser Tyr Glu  Val Ile Arg Arg Glu  Lys Leu Phe
        995                 1000                1005

Lys Glu  Ile Gln Gln Phe Glu  Ser Glu Ile Leu Phe  Arg Ser Gly
    1010                1015                1020

Trp Asp  Gly Ile Asn His Pro  Ala Gln Leu Glu Asp  Asn Arg Asn
    1025                1030                1035

Pro Lys  Phe Lys Met Tyr Ile  Val Asn Gly Ile Leu  Arg Lys Ser
    1040                1045                1050

Ala Gly  Leu Tyr Ser Gln Gly  Glu Asp Ile Trp Phe  Glu Tyr Asn
    1055                1060                1065

Ala Asp  Phe Asn Asn Leu Asp  Ala Asp Val Leu Glu  Thr Lys Ser
    1070                1075                1080

Glu Leu  Val Gln Leu Ala Phe  Leu Val Thr Ala Ile  Arg Asn Lys
    1085                1090                1095

Phe Ala  His Asn Gln Leu Pro  Ala Lys Glu Phe Tyr  Phe Tyr Ile
    1100                1105                1110

Arg Ala  Lys Tyr Gly Phe Ala  Asp Glu Pro Ser Val  Ala Leu Val
    1115                1120                1125

Tyr Leu  Asn Phe Thr Lys Tyr  Ala Ile Asn Glu Phe  Lys Lys Val
    1130                1135                1140

Met Ile
    1145
```

<210> 144
<211> 948
<212> PRT
<213> Riemerella anatipestifer

<400> 144

Met Phe Phe Ser Phe His Asn Ala Gln Arg Val Ile Phe Lys His Leu
1 5 10 15

Tyr Lys Ala Phe Asp Ala Ser Leu Arg Met Val Lys Glu Asp Tyr Lys
20 25 30

Ala His Phe Thr Val Asn Leu Thr Arg Asp Phe Ala His Leu Asn Arg
35 40 45

Lys Gly Lys Asn Lys Gln Asp Asn Pro Asp Phe Asn Arg Tyr Arg Phe
50 55 60

Glu Lys Asp Gly Phe Phe Thr Glu Ser Gly Leu Leu Phe Phe Thr Asn
65 70 75 80

Leu Phe Leu Asp Lys Arg Asp Ala Tyr Trp Met Leu Lys Lys Val Ser
85 90 95

Gly Phe Lys Ala Ser His Lys Gln Arg Glu Lys Met Thr Thr Glu Val
100 105 110

Phe Cys Arg Ser Arg Ile Leu Leu Pro Lys Leu Arg Leu Glu Ser Arg
115 120 125

Tyr Asp His Asn Gln Met Leu Leu Asp Met Leu Ser Glu Leu Ser Arg
130 135 140

Cys Pro Lys Leu Leu Tyr Glu Lys Leu Ser Glu Glu Asn Lys Lys His
145 150 155 160

Phe Gln Val Glu Ala Asp Gly Phe Leu Asp Glu Ile Glu Glu Glu Gln
165 170 175

```
Asn Pro Phe Lys Asp Thr Leu Ile Arg His Gln Asp Arg Phe Pro Tyr
            180                 185                 190

Phe Ala Leu Arg Tyr Leu Asp Leu Asn Glu Ser Phe Lys Ser Ile Arg
        195                 200                 205

Phe Gln Val Asp Leu Gly Thr Tyr His Tyr Cys Ile Tyr Asp Lys Lys
    210                 215                 220

Ile Gly Asp Glu Gln Glu Lys Arg His Leu Thr Arg Thr Leu Leu Ser
225                 230                 235                 240

Phe Gly Arg Leu Gln Asp Phe Thr Glu Ile Asn Arg Pro Gln Glu Trp
                245                 250                 255

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

Phe Ile Ser Lys Thr Thr Pro His Tyr His Ile Thr Asp Asn Lys Ile
        275                 280                 285

Gly Phe Arg Leu Gly Thr Ser Lys Glu Leu Tyr Pro Ser Leu Glu Ile
    290                 295                 300

Lys Asp Gly Ala Asn Arg Ile Ala Lys Tyr Pro Tyr Asn Ser Gly Phe
305                 310                 315                 320

Val Ala His Ala Phe Ile Ser Val His Glu Leu Leu Pro Leu Met Phe
                325                 330                 335

Tyr Gln His Leu Thr Gly Lys Ser Glu Asp Leu Leu Lys Glu Thr Val
            340                 345                 350

Arg His Ile Gln Arg Ile Tyr Lys Asp Phe Glu Glu Glu Arg Ile Asn
        355                 360                 365

Thr Ile Glu Asp Leu Glu Lys Ala Asn Gln Gly Arg Leu Pro Leu Gly
    370                 375                 380
```

```
Ala Phe Pro Lys Gln Met Leu Gly Leu Leu Gln Asn Lys Gln Pro Asp
385                 390                 395                 400


Leu Ser Glu Lys Ala Lys Ile Lys Ile Glu Lys Leu Ile Ala Glu Thr
                405                 410                 415


Lys Leu Leu Ser His Arg Leu Asn Thr Lys Leu Lys Ser Ser Pro Lys
            420                 425                 430


Leu Gly Lys Arg Arg Glu Lys Leu Ile Lys Thr Gly Val Leu Ala Asp
        435                 440                 445


Trp Leu Val Lys Asp Phe Met Arg Phe Gln Pro Val Ala Tyr Asp Ala
    450                 455                 460


Gln Asn Gln Pro Ile Lys Ser Ser Lys Ala Asn Ser Thr Glu Phe Trp
465                 470                 475                 480


Phe Ile Arg Arg Ala Leu Ala Leu Tyr Gly Gly Glu Lys Asn Arg Leu
                485                 490                 495


Glu Gly Tyr Phe Lys Gln Thr Asn Leu Ile Gly Asn Thr Asn Pro His
            500                 505                 510


Pro Phe Leu Asn Lys Phe Asn Trp Lys Ala Cys Arg Asn Leu Val Asp
        515                 520                 525


Phe Tyr Gln Gln Tyr Leu Glu Gln Arg Glu Lys Phe Leu Glu Ala Ile
    530                 535                 540


Lys Asn Gln Pro Trp Glu Pro Tyr Gln Tyr Cys Leu Leu Leu Lys Ile
545                 550                 555                 560


Pro Lys Glu Asn Arg Lys Asn Leu Val Lys Gly Trp Glu Gln Gly Gly
                565                 570                 575


Ile Ser Leu Pro Arg Gly Leu Phe Thr Glu Ala Ile Arg Glu Thr Leu
            580                 585                 590
```

```
Ser Glu Asp Leu Met Leu Ser Lys Pro Ile Arg Lys Glu Ile Lys Lys
        595                 600                 605

His Gly Arg Val Gly Phe Ile Ser Arg Ala Ile Thr Leu Tyr Phe Lys
    610                 615                 620

Glu Lys Tyr Gln Asp Lys His Gln Ser Phe Tyr Asn Leu Ser Tyr Lys
625                 630                 635                 640

Leu Glu Ala Lys Ala Pro Leu Leu Lys Arg Glu Glu His Tyr Glu Tyr
                645                 650                 655

Trp Gln Gln Asn Lys Pro Gln Ser Pro Thr Glu Ser Gln Arg Leu Glu
            660                 665                 670

Leu His Thr Ser Asp Arg Trp Lys Asp Tyr Leu Leu Tyr Lys Arg Trp
        675                 680                 685

Gln His Leu Glu Lys Lys Leu Arg Leu Tyr Arg Asn Gln Asp Val Met
    690                 695                 700

Leu Trp Leu Met Thr Leu Glu Leu Thr Lys Asn His Phe Lys Glu Leu
705                 710                 715                 720

Asn Leu Asn Tyr His Gln Leu Lys Leu Glu Asn Leu Ala Val Asn Val
                725                 730                 735

Gln Glu Ala Asp Ala Lys Leu Asn Pro Leu Asn Gln Thr Leu Pro Met
            740                 745                 750

Val Leu Pro Val Lys Val Tyr Pro Ala Thr Ala Phe Gly Glu Val Gln
        755                 760                 765

Tyr His Lys Thr Pro Ile Arg Thr Val Tyr Ile Arg Glu Glu His Thr
    770                 775                 780

Lys Ala Leu Lys Met Gly Asn Phe Lys Ala Leu Val Lys Asp Arg Arg
785                 790                 795                 800
```

```
Leu Asn Gly Leu Phe Ser Phe Ile Lys Glu Glu Asn Asp Thr Gln Lys
                805                 810                 815


His Pro Ile Ser Gln Leu Arg Leu Arg Arg Glu Leu Glu Ile Tyr Gln
            820                 825                 830


Ser Leu Arg Val Asp Ala Phe Lys Glu Thr Leu Ser Leu Glu Glu Lys
        835                 840                 845


Leu Leu Asn Lys His Thr Ser Leu Ser Ser Leu Glu Asn Glu Phe Arg
    850                 855                 860


Ala Leu Leu Glu Glu Trp Lys Lys Glu Tyr Ala Ala Ser Ser Met Val
865                 870                 875                 880


Thr Asp Glu His Ile Ala Phe Ile Ala Ser Val Arg Asn Ala Phe Cys
                885                 890                 895


His Asn Gln Tyr Pro Phe Tyr Lys Glu Ala Leu His Ala Pro Ile Pro
            900                 905                 910


Leu Phe Thr Val Ala Gln Pro Thr Thr Glu Glu Lys Asp Gly Leu Gly
        915                 920                 925


Ile Ala Glu Ala Leu Leu Lys Val Leu Arg Glu Tyr Cys Glu Ile Val
    930                 935                 940


Lys Ser Gln Ile
945


<210>  145
<211>  1175
<212>  PRT
<213>  Porhyromonas gingivalis

<400>  145

Met Thr Glu Gln Asn Glu Lys Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1               5                   10                  15
```

```
Glu Asp Lys His Phe Trp Ala Ala Phe Phe Asn Leu Ala Arg His Asn
            20                  25                  30

Ala Tyr Ile Thr Leu Ala His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
        35                  40                  45

Ala Asp Ile Thr Asn Asp Glu Asp Ile Leu Phe Phe Lys Gly Gln Trp
    50                  55                  60

Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ala Arg Leu Arg Ser Leu
65                  70                  75                  80

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

Leu Phe Glu Ser Gln Ser Ser Gly Asn Lys Ser Ser Lys Lys Lys Glu
            100                 105                 110

Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
        115                 120                 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
    130                 135                 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Pro Glu Ser Ser Glu Leu
145                 150                 155                 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
                165                 170                 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val
            180                 185                 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
        195                 200                 205

Lys Tyr Gly Asn Asn Asp Asn Pro Phe Phe Lys His His Phe Val Asp
    210                 215                 220
```

```
Arg Glu Glu Lys Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225                 230                 235                 240


Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
                245                 250                 255


Phe Lys Gly Gly Thr Glu Ala Tyr Gln Gln Met Thr Asn Glu Val Phe
            260                 265                 270


Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
        275                 280                 285


Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys
    290                 295                 300


Pro Lys Leu Leu Tyr Asp Arg Leu Arg Glu Glu Asp Arg Ala Arg Phe
305                 310                 315                 320


Arg Val Pro Val Asp Ile Leu Ser Asp Glu Asp Asp Thr Asp Gly Thr
                325                 330                 335


Glu Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
            340                 345                 350


Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
        355                 360                 365


Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
    370                 375                 380


Lys Asn Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385                 390                 395                 400


Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
                405                 410                 415


Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
            420                 425                 430
```

```
Lys Pro Tyr Ile Thr Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly
        435                 440                 445


Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln Leu Leu Trp Pro Ser
    450                 455                 460


Pro Glu Val Gly Ala Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465                 470                 475                 480


Lys Arg Phe Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
                485                 490                 495


Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Ala
            500                 505                 510


Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
        515                 520                 525


Tyr Ala Val Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Arg Asp
    530                 535                 540


Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu
545                 550                 555                 560


Pro Arg Gln Met Ile Ala Ile Leu Ser Gln Glu His Lys Asp Met Glu
                565                 570                 575


Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His
            580                 585                 590


Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly
        595                 600                 605


Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
    610                 615                 620


Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly
625                 630                 635                 640
```

```
Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
                645                 650                 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro
            660                 665                 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
        675                 680                 685

Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
    690                 695                 700

Arg Ser Tyr Leu Lys Ala Arg Lys Ala Phe Leu Gln Ser Ile Gly Arg
705                 710                 715                 720

Ser Asp Arg Glu Glu Asn His Arg Phe Leu Leu Leu Lys Glu Pro Lys
                725                 730                 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Ser Glu Phe His Leu
            740                 745                 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
        755                 760                 765

Gly Tyr Asp Glu Val Gly Ser Tyr Lys Glu Val Gly Phe Met Ala Lys
    770                 775                 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Lys Asp Arg Val Gln Pro
785                 790                 795                 800

Phe Tyr Asp Tyr Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
                805                 810                 815

Gly Arg Phe Leu Ser Lys Glu Lys Arg Ala Glu Glu Trp Glu Ser Gly
            820                 825                 830

Lys Glu Arg Phe Arg Asp Leu Glu Ala Trp Ser His Ser Ala Ala Arg
        835                 840                 845
```

```
Arg Ile Glu Asp Ala Phe Val Gly Ile Glu Tyr Ala Ser Trp Glu Asn
    850                 855                 860

Lys Lys Lys Ile Glu Gln Leu Leu Gln Asp Leu Ser Leu Trp Glu Thr
865                 870                 875                 880

Phe Glu Ser Lys Leu Lys Val Lys Ala Asp Lys Ile Asn Ile Ala Lys
                885                 890                 895

Leu Lys Lys Glu Ile Leu Glu Ala Lys Glu His Pro Tyr His Asp Phe
            900                 905                 910

Lys Ser Trp Gln Lys Phe Glu Arg Glu Leu Arg Leu Val Lys Asn Gln
        915                 920                 925

Asp Ile Ile Thr Trp Met Met Cys Arg Asp Leu Met Glu Glu Asn Lys
    930                 935                 940

Val Glu Gly Leu Asp Thr Gly Thr Leu Tyr Leu Lys Asp Ile Arg Thr
945                 950                 955                 960

Asp Val Gln Glu Gln Gly Ser Leu Asn Val Leu Asn His Val Lys Pro
                965                 970                 975

Met Arg Leu Pro Val Val Val Tyr Arg Ala Asp Ser Arg Gly His Val
            980                 985                 990

His Lys Glu Glu Ala Pro Leu Ala  Thr Val Tyr Ile Glu  Glu Arg Asp
        995                 1000                1005

Thr Lys  Leu Leu Lys Gln Gly  Asn Phe Lys Ser Phe  Val Lys Asp
    1010                1015                1020

Arg Arg  Leu Asn Gly Leu Phe  Ser Phe Val Asp Thr  Gly Ala Leu
    1025                1030                1035

Ala Met  Glu Gln Tyr Pro Ile  Ser Lys Leu Arg Val  Glu Tyr Glu
    1040                1045                1050
```

```
Leu Ala  Lys Tyr Gln Thr Ala  Arg Val Cys Ala Phe  Glu Gln Thr
    1055                 1060                 1065


Leu Glu  Leu Glu Glu Ser Leu  Leu Thr Arg Tyr Pro  His Leu Pro
    1070                 1075                 1080


Asp Glu  Ser Phe Arg Glu Met  Leu Glu Ser Trp Ser  Asp Pro Leu
    1085                 1090                 1095


Leu Asp  Lys Trp Pro Asp Leu  Gln Arg Glu Val Arg  Leu Leu Ile
    1100                 1105                 1110


Ala Val  Arg Asn Ala Phe Ser  His Asn Gln Tyr Pro  Met Tyr Asp
    1115                 1120                 1125


Glu Thr  Ile Phe Ser Ser Ile  Arg Lys Tyr Asp Pro  Ser Ser Leu
    1130                 1135                 1140


Asp Ala  Ile Glu Glu Arg Met  Gly Leu Asn Ile Ala  His Arg Leu
    1145                 1150                 1155


Ser Glu  Glu Val Lys Leu Ala  Lys Glu Met Val Glu  Arg Ile Ile
    1160                 1165                 1170


Gln Ala
    1175


<210>  146
<211>  1175
<212>  PRT
<213>  Porphyromonas gingivalis

<400>  146

Met Thr Glu Gln Asn Glu Lys Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1               5                  10                  15


Lys Asp Lys His Phe Trp Ala Ala Phe Phe Asn Leu Ala Arg His Asn
            20                  25                  30


Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
```

```
        35                  40                  45

Ala Asp Ile Thr Asn Asp Glu Asp Ile Leu Phe Phe Lys Gly Gln Trp
    50                  55                  60

Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ala Arg Leu Arg Ser Leu
65                  70                  75                  80

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

Leu Phe Glu Ser Gln Ser Ser Gly Asn Lys Ser Ser Lys Lys Lys Glu
            100                 105                 110

Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
        115                 120                 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
    130                 135                 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Pro Glu Ser Ser Glu Leu
145                 150                 155                 160

Pro Met Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
                165                 170                 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val
            180                 185                 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
        195                 200                 205

Arg Cys Gly Asn Asn Asp Asn Pro Phe Phe Lys His His Phe Val Asp
    210                 215                 220

Arg Glu Gly Lys Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225                 230                 235                 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
```

```
                245                 250                 255

Phe Lys Gly Gly Thr Glu Thr Tyr Gln Gln Met Thr Asn Glu Val Phe
            260                 265                 270

Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
        275                 280                 285

Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys
    290                 295                 300

Pro Lys Ser Leu Tyr Asp Arg Leu Arg Glu Glu Asp Arg Ala Cys Phe
305                 310                 315                 320

Arg Val Pro Val Asp Ile Leu Ser Asp Glu Asp Asp Thr Asp Gly Ala
                325                 330                 335

Glu Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
            340                 345                 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
        355                 360                 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
    370                 375                 380

Lys Asn Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385                 390                 395                 400

Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
                405                 410                 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Cys Phe Glu Thr Gly Asp
            420                 425                 430

Lys Pro Tyr Ile Thr Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly
        435                 440                 445

Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln His Leu Trp Pro Ser
```

```
    450                 455                 460

Pro Glu Val Gly Ala Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465                 470                 475                 480

Lys Arg Phe Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
                485                 490                 495

Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val
            500                 505                 510

Ser Ala Glu Arg Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
        515                 520                 525

Tyr Ala Val Tyr Asp Ala Phe Ala Arg Asp Glu Ile Asn Thr Arg Asp
    530                 535                 540

Glu Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu
545                 550                 555                 560

Pro Arg Gln Met Ile Ala Ile Leu Ser Gln Lys His Lys Asp Met Glu
                565                 570                 575

Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His
            580                 585                 590

Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly
        595                 600                 605

Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
    610                 615                 620

Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly
625                 630                 635                 640

Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
                645                 650                 655

Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro
```

660 665 670

Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
675 680 685

Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
690 695 700

Arg Ser Tyr Leu Lys Ala Arg Lys Ala Phe Leu Gln Ser Ile Gly Arg
705 710 715 720

Ser Asp Arg Val Glu Asn His Arg Phe Leu Leu Leu Lys Glu Pro Lys
725 730 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Gly Glu Phe His Leu
740 745 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
755 760 765

Gly Leu Asp Glu Val Gly Ser Tyr Lys Glu Val Gly Phe Met Ala Lys
770 775 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Lys Asp Arg Val Gln Pro
785 790 795 800

Phe Tyr Asp Tyr Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
805 810 815

Gly Arg Phe Leu Ser Lys Glu Lys Arg Ala Glu Glu Trp Glu Ser Gly
820 825 830

Lys Glu Arg Phe Arg Asp Leu Glu Ala Trp Ser His Ser Ala Ala Arg
835 840 845

Arg Ile Glu Asp Ala Phe Ala Gly Ile Glu Asn Ala Ser Arg Glu Asn
850 855 860

Lys Lys Lys Ile Glu Gln Leu Leu Gln Asp Leu Ser Leu Trp Glu Thr

865 870 875 880

Phe Glu Ser Lys Leu Lys Val Lys Ala Asp Lys Ile Asn Ile Ala Lys
885 890 895

Leu Lys Lys Glu Ile Leu Glu Ala Lys Glu His Pro Tyr Leu Asp Phe
900 905 910

Lys Ser Trp Gln Lys Phe Glu Arg Glu Leu Arg Leu Val Lys Asn Gln
915 920 925

Asp Ile Ile Thr Trp Met Met Cys Arg Asp Leu Met Glu Glu Asn Lys
930 935 940

Val Glu Gly Leu Asp Thr Gly Thr Leu Tyr Leu Lys Asp Ile Arg Thr
945 950 955 960

Asp Val Gln Glu Gln Gly Ser Leu Asn Val Leu Asn His Val Lys Pro
965 970 975

Met Arg Leu Pro Val Val Val Tyr Arg Ala Asp Ser Arg Gly His Val
980 985 990

His Lys Glu Gln Ala Pro Leu Ala Thr Val Tyr Ile Glu Glu Arg Asp
995 1000 1005

Thr Lys Leu Leu Lys Gln Gly Asn Phe Lys Ser Phe Val Lys Asp
1010 1015 1020

Arg Arg Leu Asn Gly Leu Phe Ser Phe Val Asp Thr Gly Ala Leu
1025 1030 1035

Ala Met Glu Gln Tyr Pro Ile Ser Lys Leu Arg Val Glu Tyr Glu
1040 1045 1050

Leu Ala Lys Tyr Gln Thr Ala Arg Val Cys Ala Phe Glu Gln Thr
1055 1060 1065

Leu Glu Leu Glu Glu Ser Leu Leu Thr Arg Tyr Pro His Leu Pro

```
    1070                1075                1080


Asp Glu  Asn Phe Arg Lys Met  Leu Glu Ser Trp Ser  Asp Pro Leu
    1085                1090                1095


Leu Asp  Lys Trp Pro Asp Leu  His Arg Lys Val Arg  Leu Leu Ile
    1100                1105                1110


Ala Val  Arg Asn Ala Phe Ser  His Asn Gln Tyr Pro  Met Tyr Asp
    1115                1120                1125


Glu Ala  Val Phe Ser Ser Ile  Arg Lys Tyr Asp Pro  Ser Ser Pro
    1130                1135                1140


Asp Ala  Ile Glu Glu Arg Met  Gly Leu Asn Ile Ala  His Arg Leu
    1145                1150                1155


Ser Glu  Glu Val Lys Gln Ala  Lys Glu Met Ala Glu  Arg Ile Ile
    1160                1165                1170


Gln Ala
    1175


<210>  147
<211>  1213
<212>  PRT
<213>  Flavobacterium colunare

<400>  147

Met Ser Ser Lys Asn Glu Ser Tyr Asn Lys Gln Lys Thr Phe Asn His
1               5                   10                  15


Tyr Lys Gln Glu Asp Lys Tyr Phe Phe Gly Gly Phe Leu Asn Asn Ala
            20                  25                  30


Asp Asp Asn Leu Arg Gln Val Gly Lys Glu Phe Lys Thr Arg Ile Asn
        35                  40                  45


Phe Asn His Asn Asn Asn Glu Leu Ala Ser Val Phe Lys Asp Tyr Phe
    50                  55                  60
```

```
Asn Lys Glu Lys Ser Val Ala Lys Arg Glu His Ala Leu Asn Leu Leu
65                  70                  75                  80

Ser Asn Tyr Phe Pro Val Leu Glu Arg Ile Gln Lys His Thr Asn His
                85                  90                  95

Asn Phe Glu Gln Thr Arg Glu Ile Phe Glu Leu Leu Leu Asp Thr Ile
            100                 105                 110

Lys Lys Leu Arg Asp Tyr Tyr Thr His His Tyr His Lys Pro Ile Thr
        115                 120                 125

Ile Asn Pro Lys Val Tyr Asp Phe Leu Asp Asp Thr Leu Leu Asp Val
    130                 135                 140

Leu Ile Thr Ile Lys Lys Lys Lys Val Lys Asn Asp Thr Ser Arg Glu
145                 150                 155                 160

Leu Leu Lys Glu Lys Phe Arg Pro Glu Leu Thr Gln Leu Lys Asn Gln
                165                 170                 175

Lys Arg Glu Glu Leu Ile Lys Lys Gly Lys Lys Leu Leu Glu Glu Asn
            180                 185                 190

Leu Glu Asn Ala Val Phe Asn His Cys Leu Arg Pro Phe Leu Glu Glu
        195                 200                 205

Asn Lys Thr Asp Asp Lys Gln Asn Lys Thr Val Ser Leu Arg Lys Tyr
    210                 215                 220

Arg Lys Ser Lys Pro Asn Glu Glu Thr Ser Ile Thr Leu Thr Gln Ser
225                 230                 235                 240

Gly Leu Val Phe Leu Ile Ser Phe Phe Leu His Arg Lys Glu Phe Gln
                245                 250                 255

Val Phe Thr Ser Gly Leu Glu Gly Phe Lys Ala Lys Val Asn Thr Ile
            260                 265                 270
```

```
Lys Glu Glu Glu Ile Ser Leu Asn Lys Asn Asn Ile Val Tyr Met Ile
        275                 280                 285

Thr His Trp Ser Tyr Ser Tyr Tyr Asn Phe Lys Gly Leu Lys His Arg
    290                 295                 300

Ile Lys Thr Asp Gln Gly Val Ser Thr Leu Glu Gln Asn Asn Thr Thr
305                 310                 315                 320

His Ser Leu Thr Asn Thr Asn Thr Lys Glu Ala Leu Leu Thr Gln Ile
                325                 330                 335

Val Asp Tyr Leu Ser Lys Val Pro Asn Glu Ile Tyr Glu Thr Leu Ser
            340                 345                 350

Glu Lys Gln Gln Lys Glu Phe Glu Glu Asp Ile Asn Glu Tyr Met Arg
        355                 360                 365

Glu Asn Pro Glu Asn Glu Asp Ser Thr Phe Ser Ser Ile Val Ser His
    370                 375                 380

Lys Val Ile Arg Lys Arg Tyr Glu Asn Lys Phe Asn Tyr Phe Ala Met
385                 390                 395                 400

Arg Phe Leu Asp Glu Tyr Ala Glu Leu Pro Thr Leu Arg Phe Met Val
                405                 410                 415

Asn Phe Gly Asp Tyr Ile Lys Asp Arg Gln Lys Lys Ile Leu Glu Ser
            420                 425                 430

Ile Gln Phe Asp Ser Glu Arg Ile Ile Lys Lys Glu Ile His Leu Phe
        435                 440                 445

Glu Lys Leu Gly Leu Val Thr Glu Tyr Lys Lys Asn Val Tyr Leu Lys
    450                 455                 460

Glu Thr Ser Asn Ile Asp Leu Ser Arg Phe Pro Leu Phe Pro Ser Pro
465                 470                 475                 480
```

```
Ser Tyr Val Met Ala Asn Asn Asn Ile Pro Phe Tyr Ile Asp Ser Arg
                485                 490                 495

Ser Asn Asn Leu Asp Glu Tyr Leu Asn Gln Lys Lys Lys Ala Gln Ser
            500                 505                 510

Gln Asn Arg Lys Arg Asn Leu Thr Phe Glu Lys Tyr Asn Lys Glu Gln
        515                 520                 525

Ser Lys Asp Ala Ile Ile Ala Met Leu Gln Lys Glu Ile Gly Val Lys
    530                 535                 540

Asp Leu Gln Gln Arg Ser Thr Ile Gly Leu Leu Ser Cys Asn Glu Leu
545                 550                 555                 560

Pro Ser Met Leu Tyr Glu Val Ile Val Lys Asp Ile Lys Gly Ala Glu
                565                 570                 575

Leu Glu Asn Lys Ile Ala Gln Lys Ile Arg Glu Gln Tyr Gln Ser Ile
            580                 585                 590

Arg Asp Phe Thr Leu Asp Ser Pro Gln Lys Asp Asn Ile Pro Thr Thr
        595                 600                 605

Leu Thr Lys Thr Ile Ser Thr Asp Thr Ser Val Thr Phe Glu Asn Gln
    610                 615                 620

Pro Ile Asp Ile Pro Arg Leu Lys Asn Ala Leu Gln Lys Glu Leu Thr
625                 630                 635                 640

Leu Thr Gln Glu Lys Leu Leu Asn Val Lys Gln His Glu Ile Glu Val
                645                 650                 655

Asp Asn Tyr Asn Arg Asn Lys Asn Thr Tyr Lys Phe Lys Asn Gln Pro
            660                 665                 670

Lys Asp Lys Val Asp Asp Asn Lys Leu Gln Arg Lys Tyr Val Phe Tyr
        675                 680                 685
```

```
Arg Asn Glu Ile Gly Gln Glu Ala Asn Trp Leu Ala Ser Asp Leu Ile
    690                 695                 700

His Phe Met Lys Asn Lys Ser Leu Trp Lys Gly Tyr Met His Asn Glu
705                 710                 715                 720

Leu Gln Ser Phe Leu Ala Phe Phe Glu Asp Lys Lys Asn Asp Cys Ile
                725                 730                 735

Ala Leu Leu Glu Thr Val Phe Asn Leu Lys Glu Asp Cys Ile Leu Thr
            740                 745                 750

Lys Asp Leu Lys Asn Leu Phe Leu Lys His Gly Asn Phe Ile Asp Phe
        755                 760                 765

Tyr Lys Glu Tyr Leu Lys Leu Lys Glu Asp Phe Leu Asn Thr Glu Ser
    770                 775                 780

Thr Phe Leu Glu Asn Gly Phe Ile Gly Leu Pro Pro Lys Ile Leu Lys
785                 790                 795                 800

Lys Glu Leu Ser Lys Arg Leu Asn Tyr Ile Phe Ile Val Phe Gln Lys
                805                 810                 815

Arg Gln Phe Ile Ile Lys Glu Leu Glu Glu Lys Lys Asn Asn Leu Tyr
            820                 825                 830

Ala Asp Ala Ile Asn Leu Ser Arg Gly Ile Phe Asp Glu Lys Pro Thr
        835                 840                 845

Met Ile Pro Phe Lys Lys Pro Asn Pro Asp Glu Phe Ala Ser Trp Phe
    850                 855                 860

Val Ala Ser Tyr Gln Tyr Asn Asn Tyr Gln Ser Phe Tyr Glu Leu Thr
865                 870                 875                 880

Pro Asp Lys Ile Glu Asn Asp Lys Lys Lys Lys Tyr Lys Asn Leu Arg
                885                 890                 895
```

```
Ala Ile Asn Lys Val Lys Ile Gln Asp Tyr Tyr Leu Lys Leu Met Val
            900                 905                 910

Asp Thr Leu Tyr Gln Asp Leu Phe Asn Gln Pro Leu Asp Lys Ser Leu
        915                 920                 925

Ser Asp Phe Tyr Val Ser Lys Thr Asp Arg Glu Lys Ile Lys Ala Asp
    930                 935                 940

Ala Lys Ala Tyr Gln Lys Arg Asn Asp Ser Phe Leu Trp Asn Lys Val
945                 950                 955                 960

Ile His Leu Ser Leu Gln Asn Asn Arg Ile Thr Ala Asn Pro Lys Leu
                965                 970                 975

Lys Asp Ile Gly Lys Tyr Lys Arg Ala Leu Gln Asp Glu Lys Ile Ala
            980                 985                 990

Thr Leu Leu Thr Tyr Asp Asp Arg  Thr Trp Thr Tyr Ala  Leu Gln Lys
        995                 1000                1005

Pro Glu  Lys Glu Asn Glu Asn  Asp Tyr Lys Glu Leu  His Tyr Thr
    1010                1015                1020

Ala Leu  Asn Met Glu Leu Gln  Glu Tyr Glu Lys Val  Arg Ser Lys
    1025                1030                1035

Lys Leu  Leu Lys Gln Val Gln  Glu Leu Glu Lys Gln  Ile Leu Asp
    1040                1045                1050

Lys Phe  Tyr Asp Phe Ser Asn  Asn Ala Thr His Pro  Glu Asp Leu
    1055                1060                1065

Glu Ile  Glu Asp Lys Lys Gly  Lys Arg His Pro Asn  Phe Lys Leu
    1070                1075                1080

Tyr Ile  Thr Lys Ala Leu Leu  Lys Asn Glu Ser Glu  Ile Ile Asn
    1085                1090                1095
```

```
Leu Glu  Asn Ile Asp Ile Glu  Ile Leu Ile Lys Tyr  Tyr Asp Tyr
    1100                1105                1110


Asn Thr  Glu Lys Leu Lys Glu  Lys Ile Lys Asn Met  Asp Glu Asp
    1115                1120                1125


Glu Lys  Ala Lys Ile Val Asn  Thr Lys Glu Asn Tyr  Asn Lys Ile
    1130                1135                1140


Thr Asn  Val Leu Ile Lys Lys  Ala Leu Val Leu Ile  Ile Ile Arg
    1145                1150                1155


Asn Lys  Met Ala His Asn Gln  Tyr Pro Pro Lys Phe  Ile Tyr Asp
    1160                1165                1170


Leu Ala  Thr Arg Phe Val Pro  Lys Lys Glu Glu Glu  Tyr Phe Ala
    1175                1180                1185


Cys Tyr  Phe Asn Arg Val Phe  Glu Thr Ile Thr Thr  Glu Leu Trp
    1190                1195                1200


Glu Asn  Lys Lys Lys Ala Lys  Glu Ile Val
    1205                1210


<210>  148
<211>  1119
<212>  PRT
<213>  Porphyromonas gingivalis

<400>  148

Met Thr Glu Gln Asn Glu Arg Pro Tyr Asn Gly Thr Tyr Tyr Thr Leu
1               5                   10                  15


Glu Asp Lys His Phe Trp Ala Ala Phe Phe Asn Leu Ala Arg His Asn
            20                  25                  30


Ala Tyr Ile Thr Leu Thr His Ile Asp Arg Gln Leu Ala Tyr Ser Lys
        35                  40                  45
```

```
Ala Asp Ile Thr Asn Asp Glu Asp Ile Leu Phe Phe Lys Gly Gln Trp
    50                  55                  60

Lys Asn Leu Asp Asn Asp Leu Glu Arg Lys Ala Arg Leu Arg Ser Leu
65                  70                  75                  80

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

Leu Phe Glu Asn Lys Ser Ser Gly Asn Lys Ser Ser Lys Lys Lys Glu
            100                 105                 110

Leu Thr Lys Lys Glu Lys Glu Glu Leu Gln Ala Asn Ala Leu Ser Leu
        115                 120                 125

Asp Asn Leu Lys Ser Ile Leu Phe Asp Phe Leu Gln Lys Leu Lys Asp
    130                 135                 140

Phe Arg Asn Tyr Tyr Ser His Tyr Arg His Pro Glu Ser Ser Glu Leu
145                 150                 155                 160

Pro Leu Phe Asp Gly Asn Met Leu Gln Arg Leu Tyr Asn Val Phe Asp
                165                 170                 175

Val Ser Val Gln Arg Val Lys Arg Asp His Glu His Asn Asp Lys Val
            180                 185                 190

Asp Pro His Arg His Phe Asn His Leu Val Arg Lys Gly Lys Lys Asp
        195                 200                 205

Arg Cys Gly Asn Asn Asp Asn Pro Phe Phe Lys His His Phe Val Asp
    210                 215                 220

Arg Glu Gly Lys Val Thr Glu Ala Gly Leu Leu Phe Phe Val Ser Leu
225                 230                 235                 240

Phe Leu Glu Lys Arg Asp Ala Ile Trp Met Gln Lys Lys Ile Arg Gly
                245                 250                 255
```

```
Phe Lys Gly Gly Thr Glu Ala Tyr Gln Gln Met Thr Asn Glu Val Phe
            260                 265                 270

Cys Arg Ser Arg Ile Ser Leu Pro Lys Leu Lys Leu Glu Ser Leu Arg
        275                 280                 285

Thr Asp Asp Trp Met Leu Leu Asp Met Leu Asn Glu Leu Val Arg Cys
    290                 295                 300

Pro Lys Ser Leu Tyr Asp Arg Leu Arg Glu Glu Asp Arg Ala Arg Phe
305                 310                 315                 320

Arg Val Pro Val Asp Ile Leu Ser Asp Glu Asp Asp Thr Asp Gly Thr
                325                 330                 335

Glu Glu Asp Pro Phe Lys Asn Thr Leu Val Arg His Gln Asp Arg Phe
            340                 345                 350

Pro Tyr Phe Ala Leu Arg Tyr Phe Asp Leu Lys Lys Val Phe Thr Ser
        355                 360                 365

Leu Arg Phe His Ile Asp Leu Gly Thr Tyr His Phe Ala Ile Tyr Lys
    370                 375                 380

Lys Asn Ile Gly Glu Gln Pro Glu Asp Arg His Leu Thr Arg Asn Leu
385                 390                 395                 400

Tyr Gly Phe Gly Arg Ile Gln Asp Phe Ala Glu Glu His Arg Pro Glu
                405                 410                 415

Glu Trp Lys Arg Leu Val Arg Asp Leu Asp Tyr Phe Glu Thr Gly Asp
            420                 425                 430

Lys Pro Tyr Ile Thr Gln Thr Thr Pro His Tyr His Ile Glu Lys Gly
        435                 440                 445

Lys Ile Gly Leu Arg Phe Val Pro Glu Gly Gln His Leu Trp Pro Ser
    450                 455                 460
```

```
Pro Glu Val Gly Ala Thr Arg Thr Gly Arg Ser Lys Tyr Ala Gln Asp
465                 470                 475                 480


Lys Arg Leu Thr Ala Glu Ala Phe Leu Ser Val His Glu Leu Met Pro
                485                 490                 495


Met Met Phe Tyr Tyr Phe Leu Leu Arg Glu Lys Tyr Ser Glu Glu Val
            500                 505                 510


Ser Ala Glu Lys Val Gln Gly Arg Ile Lys Arg Val Ile Glu Asp Val
        515                 520                 525


Tyr Ala Val Tyr Asp Ala Phe Ala Arg Gly Glu Ile Asp Thr Leu Asp
    530                 535                 540


Arg Leu Asp Ala Cys Leu Ala Asp Lys Gly Ile Arg Arg Gly His Leu
545                 550                 555                 560


Pro Arg Gln Met Ile Ala Ile Leu Ser Gln Glu His Lys Asp Met Glu
                565                 570                 575


Glu Lys Val Arg Lys Lys Leu Gln Glu Met Ile Ala Asp Thr Asp His
            580                 585                 590


Arg Leu Asp Met Leu Asp Arg Gln Thr Asp Arg Lys Ile Arg Ile Gly
        595                 600                 605


Arg Lys Asn Ala Gly Leu Pro Lys Ser Gly Val Ile Ala Asp Trp Leu
    610                 615                 620


Val Arg Asp Met Met Arg Phe Gln Pro Val Ala Lys Asp Thr Ser Gly
625                 630                 635                 640


Lys Pro Leu Asn Asn Ser Lys Ala Asn Ser Thr Glu Tyr Arg Met Leu
                645                 650                 655


Gln Arg Ala Leu Ala Leu Phe Gly Gly Glu Lys Glu Arg Leu Thr Pro
            660                 665                 670
```

```
Tyr Phe Arg Gln Met Asn Leu Thr Gly Gly Asn Asn Pro His Pro Phe
        675                 680                 685

Leu His Glu Thr Arg Trp Glu Ser His Thr Asn Ile Leu Ser Phe Tyr
    690                 695                 700

Arg Ser Tyr Leu Lys Ala Arg Lys Ala Phe Leu Gln Ser Ile Gly Arg
705                 710                 715                 720

Ser Asp Arg Glu Glu Asn His Arg Phe Leu Leu Leu Lys Glu Pro Lys
                725                 730                 735

Thr Asp Arg Gln Thr Leu Val Ala Gly Trp Lys Ser Glu Phe His Leu
            740                 745                 750

Pro Arg Gly Ile Phe Thr Glu Ala Val Arg Asp Cys Leu Ile Glu Met
        755                 760                 765

Gly Tyr Asp Glu Val Gly Ser Tyr Lys Glu Val Gly Phe Met Ala Lys
    770                 775                 780

Ala Val Pro Leu Tyr Phe Glu Arg Ala Cys Lys Asp Arg Val Gln Pro
785                 790                 795                 800

Phe Tyr Asp Tyr Pro Phe Asn Val Gly Asn Ser Leu Lys Pro Lys Lys
                805                 810                 815

Gly Arg Phe Leu Ser Lys Glu Lys Arg Ala Glu Glu Trp Glu Ser Gly
            820                 825                 830

Lys Glu Arg Phe Arg Leu Ala Lys Leu Lys Lys Glu Ile Leu Glu Ala
        835                 840                 845

Lys Glu His Pro Tyr Leu Asp Phe Lys Ser Trp Gln Lys Phe Glu Arg
    850                 855                 860

Glu Leu Arg Leu Val Lys Asn Gln Asp Ile Ile Thr Trp Met Met Cys
865                 870                 875                 880
```

```
Arg Asp Leu Met Glu Glu Asn Lys Val Glu Gly Leu Asp Thr Gly Thr
                885                 890                 895


Leu Tyr Leu Lys Asp Ile Arg Thr Glu Val Gln Glu Gln Gly Ser Leu
            900                 905                 910


Asn Val Leu Asn Arg Val Lys Pro Met Arg Leu Pro Val Val Val Tyr
        915                 920                 925


Arg Ala Asp Ser Arg Gly His Val His Lys Glu Gln Ala Pro Leu Ala
    930                 935                 940


Thr Val Tyr Ile Glu Glu Arg Asp Thr Lys Leu Leu Lys Gln Gly Asn
945                 950                 955                 960


Phe Lys Ser Phe Val Lys Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe
                965                 970                 975


Val Asp Thr Gly Gly Leu Ala Met Glu Gln Tyr Pro Ile Ser Lys Leu
            980                 985                 990


Arg Val Glu Tyr Glu Leu Ala Lys  Tyr Gln Thr Ala Arg  Val Cys Ala
        995                 1000                1005


Phe Glu  Gln Thr Leu Glu Leu  Glu Glu Ser Leu Leu  Thr Arg Cys
    1010                1015                1020


Pro His  Leu Pro Asp Lys Asn  Phe Arg Lys Met Leu  Glu Ser Trp
    1025                1030                1035


Ser Asp  Pro Leu Leu Asp Lys  Trp Pro Asp Leu Gln  Arg Glu Val
    1040                1045                1050


Trp Leu  Leu Ile Ala Val Arg  Asn Ala Phe Ser His  Asn Gln Tyr
    1055                1060                1065


Pro Met  Tyr Asp Glu Ala Val  Phe Ser Ser Ile Arg  Lys Tyr Asp
    1070                1075                1080
```

```
Pro Ser  Ser Pro Asp Ala Ile  Glu Glu Arg Met Gly  Leu Asn Ile
    1085                1090                1095


Ala His  Arg Leu Ser Glu Glu  Val Lys Gln Ala Lys  Glu Met Ala
    1100                1105                1110


Glu Arg  Ile Ile Gln Ala
    1115


<210>  149
<211>  1135
<212>  PRT
<213>  Porphyromonas gingivalis

<400>  149

Met Asn Thr Val Pro Ala Ser Glu Asn Lys Gly Gln Ser Arg Thr Val
1               5                   10                  15


Glu Asp Asp Pro Gln Tyr Phe Gly Leu Tyr Leu Asn Leu Ala Arg Glu
            20                  25                  30


Asn Leu Ile Glu Val Glu Ser His Val Arg Ile Lys Phe Gly Lys Lys
        35                  40                  45


Lys Leu Asn Glu Glu Ser Leu Lys Gln Ser Leu Leu Cys Asp His Leu
    50                  55                  60


Leu Ser Val Asp Arg Trp Thr Lys Val Tyr Gly His Ser Arg Arg Tyr
65                  70                  75                  80


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


His Asp Ser Lys Thr Gly Val Asp Pro Asp Ser Ala Gln Arg Leu Ile
            100                 105                 110


Arg Glu Leu Tyr Ser Leu Leu Asp Phe Leu Arg Asn Asp Phe Ser His
        115                 120                 125
```

```
Asn Arg Leu Asp Gly Thr Thr Phe Glu His Leu Glu Val Ser Pro Asp
    130                 135                 140

Ile Ser Ser Phe Ile Thr Gly Thr Tyr Ser Leu Ala Cys Gly Arg Ala
145                 150                 155                 160

Gln Ser Arg Phe Ala Asp Phe Phe Lys Pro Asp Asp Phe Val Leu Ala
                165                 170                 175

Lys Asn Arg Lys Glu Gln Leu Ile Ser Val Ala Asp Gly Lys Glu Cys
            180                 185                 190

Leu Thr Val Ser Gly Leu Ala Phe Phe Ile Cys Leu Phe Leu Asp Arg
        195                 200                 205

Glu Gln Ala Ser Gly Met Leu Ser Arg Ile Arg Gly Phe Lys Arg Thr
    210                 215                 220

Asp Glu Asn Trp Ala Arg Ala Val His Glu Thr Phe Cys Asp Leu Cys
225                 230                 235                 240

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

Leu Leu Leu Asp Met Leu Asn Glu Leu Asn Arg Cys Pro Arg Ile Leu
            260                 265                 270

Tyr Asp Met Leu Pro Glu Glu Glu Arg Ala Gln Phe Leu Pro Ala Leu
        275                 280                 285

Asp Glu Asn Ser Met Asn Asn Leu Ser Glu Asn Ser Leu Asn Glu Glu
    290                 295                 300

Ser Arg Leu Leu Trp Asp Gly Ser Ser Asp Trp Ala Glu Ala Leu Thr
305                 310                 315                 320

Lys Arg Ile Arg His Gln Asp Arg Phe Pro Tyr Leu Met Leu Arg Phe
                325                 330                 335
```

```
Ile Glu Glu Met Asp Leu Leu Lys Gly Ile Arg Phe Arg Val Asp Leu
            340                 345                 350

Gly Glu Ile Glu Leu Asp Ser Tyr Ser Lys Lys Val Gly Arg Asn Gly
        355                 360                 365

Glu Tyr Asp Arg Thr Ile Thr Asp His Ala Leu Ala Phe Gly Lys Leu
    370                 375                 380

Ser Asp Phe Gln Asn Glu Glu Glu Val Ser Arg Met Ile Ser Gly Glu
385                 390                 395                 400

Ala Ser Tyr Pro Val Arg Phe Ser Leu Phe Ala Pro Arg Tyr Ala Ile
                405                 410                 415

Tyr Asp Asn Lys Ile Gly Tyr Cys His Thr Ser Asp Pro Val Tyr Pro
            420                 425                 430

Lys Ser Lys Thr Gly Glu Lys Arg Ala Leu Ser Asn Pro Gln Ser Met
        435                 440                 445

Gly Phe Ile Ser Val His Asp Leu Arg Lys Leu Leu Leu Met Glu Leu
    450                 455                 460

Leu Cys Glu Gly Ser Phe Ser Arg Met Gln Ser Gly Phe Leu Arg Lys
465                 470                 475                 480

Ala Asn Arg Ile Leu Asp Glu Thr Ala Glu Gly Lys Leu Gln Phe Ser
                485                 490                 495

Ala Leu Phe Pro Glu Met Arg His Arg Phe Ile Pro Pro Gln Asn Pro
            500                 505                 510

Lys Ser Lys Asp Arg Arg Glu Lys Ala Glu Thr Thr Leu Glu Lys Tyr
        515                 520                 525

Lys Gln Glu Ile Lys Gly Arg Lys Asp Lys Leu Asn Ser Gln Leu Leu
    530                 535                 540
```

```
Ser Ala Phe Asp Met Asn Gln Arg Gln Leu Pro Ser Arg Leu Leu Asp
545                 550                 555                 560


Glu Trp Met Asn Ile Arg Pro Ala Ser His Ser Val Lys Leu Arg Thr
                565                 570                 575


Tyr Val Lys Gln Leu Asn Glu Asp Cys Arg Leu Arg Leu Arg Lys Phe
            580                 585                 590


Arg Lys Asp Gly Asp Gly Lys Ala Arg Ala Ile Pro Leu Val Gly Glu
        595                 600                 605


Met Ala Thr Phe Leu Ser Gln Asp Ile Val Arg Met Ile Ile Ser Glu
    610                 615                 620


Glu Thr Lys Lys Leu Ile Thr Ser Ala Tyr Tyr Asn Glu Met Gln Arg
625                 630                 635                 640


Ser Leu Ala Gln Tyr Ala Gly Glu Glu Asn Arg Arg Gln Phe Arg Ala
                645                 650                 655


Ile Val Ala Glu Leu His Leu Leu Asp Pro Ser Ser Gly His Pro Phe
            660                 665                 670


Leu Ser Ala Thr Met Glu Thr Ala His Arg Tyr Thr Glu Asp Phe Tyr
        675                 680                 685


Lys Cys Tyr Leu Glu Lys Lys Arg Glu Trp Leu Ala Lys Thr Phe Tyr
    690                 695                 700


Arg Pro Glu Gln Asp Glu Asn Thr Lys Arg Arg Ile Ser Val Phe Phe
705                 710                 715                 720


Val Pro Asp Gly Glu Ala Arg Lys Leu Leu Pro Leu Leu Ile Arg Arg
                725                 730                 735


Arg Met Lys Glu Gln Asn Asp Leu Gln Asp Trp Ile Arg Asn Lys Gln
            740                 745                 750
```

```
Ala His Pro Ile Asp Leu Pro Ser His Leu Phe Asp Ser Lys Ile Met
        755                 760                 765


Glu Leu Leu Lys Val Lys Asp Gly Lys Lys Lys Trp Asn Glu Ala Phe
    770                 775                 780


Lys Asp Trp Trp Ser Thr Lys Tyr Pro Asp Gly Met Gln Pro Phe Tyr
785                 790                 795                 800


Gly Leu Arg Arg Glu Leu Asn Ile His Gly Lys Ser Val Ser Tyr Ile
                805                 810                 815


Pro Ser Asp Gly Lys Lys Phe Ala Asp Cys Tyr Thr His Leu Met Glu
            820                 825                 830


Lys Thr Val Gln Asp Lys Lys Arg Glu Leu Arg Thr Ala Gly Lys Pro
        835                 840                 845


Val Pro Pro Asp Leu Ala Ala Asp Ile Lys Arg Ser Phe His Arg Ala
    850                 855                 860


Val Asn Glu Arg Glu Phe Met Leu Arg Leu Val Gln Glu Asp Asp Arg
865                 870                 875                 880


Leu Met Leu Met Ala Ile Asn Lys Met Met Thr Asp Arg Glu Glu Asp
                885                 890                 895


Ile Leu Pro Gly Leu Lys Asn Ile Asp Ser Ile Leu Asp Lys Glu Asn
            900                 905                 910


Gln Phe Ser Leu Ala Val His Ala Lys Val Leu Glu Lys Glu Gly Glu
        915                 920                 925


Gly Gly Asp Asn Ser Leu Ser Leu Val Pro Ala Thr Ile Glu Ile Lys
    930                 935                 940


Ser Lys Arg Lys Asp Trp Ser Lys Tyr Ile Arg Tyr Arg Tyr Asp Arg
945                 950                 955                 960
```

```
Arg Val Pro Gly Leu Met Ser His Phe Pro Glu His Lys Ala Thr Leu
                965                 970                 975


Asp Glu Val Lys Thr Leu Leu Gly Glu Tyr Asp Arg Cys Arg Ile Lys
            980                 985                 990


Ile Phe Asp Trp Ala Phe Ala Leu  Glu Gly Ala Ile Met  Ser Asp Arg
        995                 1000                1005


Asp Leu  Lys Pro Tyr Leu His  Glu Ser Ser Ser Arg  Glu Gly Lys
    1010                1015                1020


Ser Gly  Glu His Ser Thr Leu  Val Lys Met Leu Val  Glu Lys Lys
    1025                1030                1035


Gly Cys  Leu Thr Pro Asp Glu  Ser Gln Tyr Leu Ile  Leu Ile Arg
    1040                1045                1050


Asn Lys  Ala Ala His Asn Gln  Phe Pro Cys Ala Ala  Glu Met Pro
    1055                1060                1065


Leu Ile  Tyr Arg Asp Val Ser  Ala Lys Val Gly Ser  Ile Glu Gly
    1070                1075                1080


Ser Ser  Ala Lys Asp Leu Pro  Glu Gly Ser Ser Leu  Val Asp Ser
    1085                1090                1095


Leu Trp  Lys Lys Tyr Glu Met  Ile Ile Arg Lys Ile  Leu Pro Ile
    1100                1105                1110


Leu Asp  Pro Glu Asn Arg Phe  Phe Gly Lys Leu Leu  Asn Asn Met
    1115                1120                1125


Ser Gln  Pro Ile Asn Asp Leu
    1130                1135


<210>  150
<211>  950
<212>  PRT
<213>  Riemerella anatipestifer
```

<400> 150

```
Met Phe Phe Ser Phe His Asn Ala Gln Arg Val Ile Phe Lys His Leu
1               5                  10                  15


Tyr Lys Ala Phe Asp Ala Ser Leu Arg Met Val Lys Glu Asp Tyr Lys
            20                  25                  30


Ala His Phe Thr Val Asn Leu Thr Arg Asp Phe Ala His Leu Asn Arg
        35                  40                  45


Lys Gly Lys Asn Lys Gln Asp Asn Pro Asp Phe Asn Arg Tyr Arg Phe
    50                  55                  60


Glu Lys Asp Gly Phe Phe Thr Glu Ser Gly Leu Leu Phe Phe Thr Asn
65                  70                  75                  80


Leu Phe Leu Asp Lys Arg Asp Ala Tyr Trp Met Leu Lys Lys Val Ser
                85                  90                  95


Gly Phe Lys Ala Ser His Lys Gln Ser Glu Lys Met Thr Thr Glu Val
            100                 105                 110


Phe Cys Arg Ser Arg Ile Leu Leu Pro Lys Leu Arg Leu Glu Ser Arg
        115                 120                 125


Tyr Asp His Asn Gln Met Leu Leu Asp Met Leu Ser Glu Leu Ser Arg
    130                 135                 140


Cys Pro Lys Leu Leu Tyr Glu Lys Leu Ser Glu Lys Asp Lys Lys Cys
145                 150                 155                 160


Phe Gln Val Glu Ala Asp Gly Phe Leu Asp Glu Ile Glu Glu Glu Gln
                165                 170                 175


Asn Pro Phe Lys Asp Thr Leu Ile Arg His Gln Asp Arg Phe Pro Tyr
            180                 185                 190


Phe Ala Leu Arg Tyr Leu Asp Leu Asn Glu Ser Phe Lys Ser Ile Arg
```

```
        195                 200                 205

Phe Gln Val Asp Leu Gly Thr Tyr His Tyr Cys Ile Tyr Asp Lys Lys
    210                 215                 220

Ile Gly Tyr Glu Gln Glu Lys Arg His Leu Thr Arg Thr Leu Leu Asn
225                 230                 235                 240

Phe Gly Arg Leu Gln Asp Phe Thr Glu Ile Asn Arg Pro Gln Glu Trp
                245                 250                 255

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

Phe Ile Ser Lys Thr Thr Pro His Tyr His Ile Thr Asp Asn Lys Ile
        275                 280                 285

Gly Phe Arg Leu Arg Thr Ser Lys Glu Leu Tyr Pro Ser Leu Glu Val
    290                 295                 300

Lys Asp Gly Ala Asn Arg Ile Ala Lys Tyr Pro Tyr Asn Ser Asp Phe
305                 310                 315                 320

Val Ala His Ala Phe Ile Ser Ile Ser Val His Glu Leu Leu Pro Leu
                325                 330                 335

Met Phe Tyr Gln His Leu Thr Gly Lys Ser Glu Asp Leu Leu Lys Glu
            340                 345                 350

Thr Val Arg His Ile Gln Arg Ile Tyr Lys Asp Phe Glu Glu Glu Arg
        355                 360                 365

Ile Asn Thr Ile Glu Asp Leu Glu Lys Ala Asn Gln Gly Arg Leu Pro
    370                 375                 380

Leu Gly Ala Phe Pro Lys Gln Met Leu Gly Leu Leu Gln Asn Lys Gln
385                 390                 395                 400

Pro Asp Leu Ser Glu Lys Ala Lys Ile Lys Ile Glu Lys Leu Ile Ala
```

405 410 415

Glu Thr Lys Leu Leu Ser His Arg Leu Asn Thr Lys Leu Lys Ser Ser
420 425 430

Pro Lys Leu Gly Lys Arg Arg Glu Lys Leu Ile Lys Thr Gly Val Leu
435 440 445

Ala Asp Trp Leu Val Lys Asp Phe Met Arg Phe Gln Pro Val Val Tyr
450 455 460

Asp Ala Gln Asn Gln Pro Ile Lys Ser Ser Lys Ala Asn Ser Thr Glu
465 470 475 480

Ser Arg Leu Ile Arg Arg Ala Leu Ala Leu Tyr Gly Gly Glu Lys Asn
485 490 495

Arg Leu Glu Gly Tyr Phe Lys Gln Thr Asn Leu Ile Gly Asn Thr Asn
500 505 510

Pro His Pro Phe Leu Asn Lys Phe Asn Trp Lys Ala Cys Arg Asn Leu
515 520 525

Val Asp Phe Tyr Gln Gln Tyr Leu Glu Gln Arg Glu Lys Phe Leu Glu
530 535 540

Ala Ile Lys His Gln Pro Trp Glu Pro Tyr Gln Tyr Cys Leu Leu Leu
545 550 555 560

Lys Val Pro Lys Glu Asn Arg Lys Asn Leu Val Lys Gly Trp Glu Gln
565 570 575

Gly Gly Ile Ser Leu Pro Arg Gly Leu Phe Thr Glu Ala Ile Arg Glu
580 585 590

Thr Leu Ser Lys Asp Leu Thr Leu Ser Lys Pro Ile Arg Lys Glu Ile
595 600 605

Lys Lys His Gly Arg Val Gly Phe Ile Ser Arg Ala Ile Thr Leu Tyr

610 615 620

Phe Lys Glu Lys Tyr Gln Asp Lys His Gln Ser Phe Tyr Asn Leu Ser
625 630 635 640

Tyr Lys Leu Glu Ala Lys Ala Pro Leu Leu Lys Lys Glu Glu His Tyr
645 650 655

Glu Tyr Trp Gln Gln Asn Lys Pro Gln Ser Pro Thr Glu Ser Gln Arg
660 665 670

Leu Glu Leu His Thr Ser Asp Arg Trp Lys Asp Tyr Leu Leu Tyr Lys
675 680 685

Arg Trp Gln His Leu Glu Lys Lys Leu Arg Leu Tyr Arg Asn Gln Asp
690 695 700

Ile Met Leu Trp Leu Met Thr Leu Glu Leu Thr Lys Asn His Phe Lys
705 710 715 720

Glu Leu Asn Leu Asn Tyr His Gln Leu Lys Leu Glu Asn Leu Ala Val
725 730 735

Asn Val Gln Glu Ala Asp Ala Lys Leu Asn Pro Leu Asn Gln Thr Leu
740 745 750

Pro Met Val Leu Pro Val Lys Val Tyr Pro Thr Thr Ala Phe Gly Glu
755 760 765

Val Gln Tyr His Glu Thr Pro Ile Arg Thr Val Tyr Ile Arg Glu Glu
770 775 780

Gln Thr Lys Ala Leu Lys Met Gly Asn Phe Lys Ala Leu Val Lys Asp
785 790 795 800

Arg His Leu Asn Gly Leu Phe Ser Phe Ile Lys Glu Glu Asn Asp Thr
805 810 815

Gln Lys His Pro Ile Ser Gln Leu Arg Leu Arg Arg Glu Leu Glu Ile

820 825 830

Tyr Gln Ser Leu Arg Val Asp Ala Phe Lys Glu Thr Leu Ser Leu Glu
835 840 845

Glu Lys Leu Leu Asn Lys His Ala Ser Leu Ser Ser Leu Glu Asn Glu
850 855 860

Phe Arg Thr Leu Leu Glu Glu Trp Lys Lys Lys Tyr Ala Ala Ser Ser
865 870 875 880

Met Val Thr Asp Lys His Ile Ala Phe Ile Ala Ser Val Arg Asn Ala
885 890 895

Phe Cys His Asn Gln Tyr Pro Phe Tyr Lys Glu Thr Leu His Ala Pro
900 905 910

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

Leu Gly Ile Ala Glu Ala Leu Leu Arg Val Leu Arg Glu Tyr Cys Glu
930 935 940

Ile Val Lys Ser Gln Ile
945 950

<210> 151
<211> 1213
<212> PRT
<213> Flavobacterium columnare

<400> 151

Met Ser Ser Lys Asn Glu Ser Tyr Asn Lys Gln Lys Thr Phe Asn His
1 5 10 15

Tyr Lys Gln Glu Asp Lys Tyr Phe Phe Gly Gly Phe Leu Asn Asn Ala
20 25 30

Asp Asp Asn Leu Arg Gln Val Gly Lys Glu Phe Lys Thr Arg Ile Asn
35 40 45

```
Phe Asn His Asn Asn Asn Glu Leu Ala Ser Val Phe Lys Asp Tyr Phe
    50                  55                  60

Asn Lys Glu Lys Ser Val Ala Lys Arg Glu His Ala Leu Asn Leu Leu
65                  70                  75                  80

Ser Asn Tyr Phe Pro Val Leu Glu Arg Ile Gln Lys His Thr Asn His
                85                  90                  95

Asn Phe Glu Gln Thr Arg Glu Ile Phe Glu Leu Leu Leu Asp Thr Ile
            100                 105                 110

Lys Lys Leu Arg Asp Tyr Tyr Thr His His Tyr His Lys Pro Ile Thr
        115                 120                 125

Ile Asn Pro Lys Ile Tyr Asp Phe Leu Asp Asp Thr Leu Leu Asp Val
    130                 135                 140

Leu Ile Thr Ile Lys Lys Lys Lys Val Lys Asn Asp Thr Ser Arg Glu
145                 150                 155                 160

Leu Leu Lys Glu Lys Leu Arg Pro Glu Leu Thr Gln Leu Lys Asn Gln
                165                 170                 175

Lys Arg Glu Glu Leu Ile Lys Lys Gly Lys Lys Leu Leu Glu Glu Asn
            180                 185                 190

Leu Glu Asn Ala Val Phe Asn His Cys Leu Arg Pro Phe Leu Glu Glu
        195                 200                 205

Asn Lys Thr Asp Asp Lys Gln Asn Lys Thr Val Ser Leu Arg Lys Tyr
    210                 215                 220

Arg Lys Ser Lys Pro Asn Glu Glu Thr Ser Ile Thr Leu Thr Gln Ser
225                 230                 235                 240

Gly Leu Val Phe Leu Met Ser Phe Phe Leu His Arg Lys Glu Phe Gln
                245                 250                 255
```

```
Val Phe Thr Ser Gly Leu Glu Gly Phe Lys Ala Lys Val Asn Thr Ile
            260                 265                 270

Lys Glu Glu Lys Ile Ser Leu Asn Lys Asn Asn Ile Val Tyr Met Ile
        275                 280                 285

Thr His Trp Ser Tyr Ser Tyr Tyr Asn Phe Lys Gly Leu Lys His Arg
    290                 295                 300

Ile Lys Thr Asp Gln Gly Val Ser Thr Leu Glu Gln Asn Asn Thr Thr
305                 310                 315                 320

His Ser Leu Thr Asn Thr Asn Thr Lys Glu Ala Leu Leu Thr Gln Ile
                325                 330                 335

Val Asp Tyr Leu Ser Lys Val Pro Asn Glu Ile Tyr Glu Thr Leu Ser
            340                 345                 350

Glu Lys Gln Gln Lys Glu Phe Glu Glu Asp Ile Asn Glu Tyr Met Arg
        355                 360                 365

Glu Asn Pro Glu Asn Glu Asp Ser Thr Phe Ser Ser Ile Val Ser His
    370                 375                 380

Lys Val Ile Arg Lys Arg Tyr Glu Asn Lys Phe Asn Tyr Phe Ala Met
385                 390                 395                 400

Arg Phe Leu Asp Glu Tyr Ala Glu Leu Pro Thr Leu Arg Phe Met Val
                405                 410                 415

Asn Phe Gly Asp Tyr Ile Lys Asp Arg Gln Lys Lys Ile Leu Glu Ser
            420                 425                 430

Ile Gln Phe Asp Ser Glu Arg Ile Ile Lys Lys Glu Ile His Leu Phe
        435                 440                 445

Glu Lys Leu Gly Leu Val Thr Glu Tyr Lys Lys Asn Val Tyr Leu Lys
    450                 455                 460
```

```
Glu Thr Ser Asn Ile Asp Leu Ser Arg Phe Pro Leu Phe Pro Ser Pro
465                 470                 475                 480


Ser Tyr Val Met Ala Asn Asn Asn Ile Pro Phe Tyr Ile Asp Ser Arg
                485                 490                 495


Ser Asn Asn Leu Asp Glu Tyr Leu Asn Gln Lys Lys Lys Ala Gln Ser
            500                 505                 510


Gln Asn Arg Lys Arg Asn Leu Thr Phe Glu Lys Tyr Asn Lys Glu Gln
        515                 520                 525


Ser Lys Asp Ala Ile Ile Ala Met Leu Gln Lys Glu Ile Gly Val Lys
    530                 535                 540


Asp Leu Gln Gln Arg Ser Thr Ile Gly Leu Leu Ser Cys Asn Glu Leu
545                 550                 555                 560


Pro Ser Met Leu Tyr Glu Val Ile Val Lys Asp Ile Lys Gly Ala Glu
                565                 570                 575


Leu Glu Asn Lys Ile Ala Gln Lys Ile Arg Glu Gln Tyr Gln Ser Ile
            580                 585                 590


Arg Asp Phe Thr Leu Asn Ser Pro Gln Lys Asp Asn Ile Pro Thr Thr
        595                 600                 605


Leu Ile Lys Thr Ile Ser Thr Asp Thr Ser Val Thr Phe Glu Asn Gln
    610                 615                 620


Pro Ile Asp Ile Pro Arg Leu Lys Asn Ala Ile Gln Lys Glu Leu Ala
625                 630                 635                 640


Leu Thr Gln Glu Lys Leu Leu Asn Val Lys Gln His Glu Ile Glu Val
                645                 650                 655


Asn Asn Tyr Asn Arg Asn Lys Asn Thr Tyr Lys Phe Lys Asn Gln Pro
            660                 665                 670
```

```
Lys Asp Lys Val Asp Asp Asn Lys Leu Gln Arg Lys Tyr Val Phe Tyr
        675                 680                 685

Arg Asn Glu Ile Gly Gln Glu Ala Asn Trp Leu Ala Ser Asp Leu Ile
    690                 695                 700

His Phe Met Lys Asn Lys Ser Leu Trp Lys Gly Tyr Met His Asn Glu
705                 710                 715                 720

Leu Gln Ser Phe Leu Ala Phe Phe Glu Asp Lys Lys Asn Asp Cys Ile
                725                 730                 735

Ala Leu Leu Glu Thr Val Phe Asn Leu Lys Glu Asp Cys Ile Leu Thr
            740                 745                 750

Lys Asp Leu Lys Asn Leu Phe Leu Lys His Gly Asn Phe Ile Asp Phe
        755                 760                 765

Tyr Lys Glu Tyr Leu Lys Leu Lys Glu Asp Phe Leu Asn Thr Glu Ser
    770                 775                 780

Thr Phe Leu Glu Asn Gly Phe Ile Gly Leu Pro Pro Lys Ile Leu Lys
785                 790                 795                 800

Lys Glu Leu Ser Lys Arg Leu Asn Tyr Ile Phe Ile Val Phe Gln Lys
                805                 810                 815

Arg Gln Phe Ile Ile Lys Glu Leu Glu Glu Lys Lys Asn Asn Leu Tyr
            820                 825                 830

Ala Asp Ala Ile Asn Leu Ser Arg Gly Ile Phe Asp Glu Lys Pro Thr
        835                 840                 845

Met Ile Pro Phe Lys Lys Pro Asn Pro Asp Glu Phe Ala Ser Trp Phe
    850                 855                 860

Val Ala Ser Tyr Gln Tyr Asn Asn Tyr Gln Ser Phe Tyr Glu Leu Thr
865                 870                 875                 880
```

```
Pro Asp Lys Ile Glu Asn Asp Lys Lys Lys Lys Tyr Lys Asn Leu Arg
                885                 890                 895

Ala Ile Asn Lys Val Lys Ile Gln Asp Tyr Tyr Leu Lys Leu Met Val
            900                 905                 910

Asp Thr Leu Tyr Gln Asp Leu Phe Asn Gln Pro Leu Asp Lys Ser Leu
        915                 920                 925

Ser Asp Phe Tyr Val Ser Lys Thr Asp Arg Glu Lys Ile Lys Ala Asp
    930                 935                 940

Ala Lys Ala Tyr Gln Lys Arg Asn Asp Ser Phe Leu Trp Asn Lys Val
945                 950                 955                 960

Ile His Leu Ser Leu Gln Asn Asn Arg Ile Thr Ala Asn Pro Lys Leu
                965                 970                 975

Lys Asp Ile Gly Lys Tyr Lys Arg Ala Leu Gln Asp Glu Lys Ile Ala
            980                 985                 990

Thr Leu Leu Thr Tyr Asp Asp Arg  Thr Trp Thr Tyr Ala  Leu Gln Lys
        995                 1000                1005

Pro Glu  Lys Glu Asn Glu Asn  Asp Tyr Lys Glu Leu  His Tyr Thr
    1010                1015                1020

Ala Leu  Asn Met Glu Leu Gln  Glu Tyr Glu Lys Val  Arg Ser Lys
    1025                1030                1035

Lys Leu  Leu Lys Gln Val Gln  Glu Leu Glu Lys Gln  Ile Leu Asp
    1040                1045                1050

Lys Phe  Tyr Asp Phe Ser Asn  Asn Ala Thr His Pro  Glu Asp Leu
    1055                1060                1065

Glu Ile  Glu Asp Lys Lys Gly  Lys Arg His Pro Asn  Phe Lys Leu
    1070                1075                1080
```

```
Tyr Ile  Thr Lys Ala Leu Leu  Lys Asn Glu Ser Glu  Ile Ile Asn
    1085                1090                1095


Leu Glu  Asn Ile Asp Ile Glu  Ile Leu Ile Lys Tyr  Tyr Asp Tyr
    1100                1105                1110


Asn Thr  Glu Lys Leu Lys Glu  Lys Ile Lys Asn Met  Asp Glu Asp
    1115                1120                1125


Glu Lys  Ala Lys Ile Val Asn  Thr Lys Glu Asn Tyr  Asn Lys Ile
    1130                1135                1140


Thr Asn  Val Leu Ile Lys Lys  Ala Leu Val Leu Ile  Ile Ile Arg
    1145                1150                1155


Asn Lys  Met Ala His Asn Gln  Tyr Pro Pro Lys Phe  Ile Tyr Asp
    1160                1165                1170


Leu Ala  Thr Arg Phe Val Pro  Lys Lys Glu Glu Glu  Tyr Phe Ala
    1175                1180                1185


Cys Tyr  Phe Asn Arg Val Phe  Glu Thr Ile Thr Thr  Glu Leu Trp
    1190                1195                1200


Glu Asn  Lys Lys Lys Ala Lys  Glu Ile Val
    1205                1210


<210>  152
<211>  1095
<212>  PRT
<213>  Riemerella anatipestifer

<400>  152

Met Glu Lys Pro Leu Pro Pro Asn Val Tyr Thr Leu Lys His Lys Phe
1               5                  10                  15


Phe Trp Gly Ala Phe Leu Asn Ile Ala Arg His Asn Ala Phe Ile Thr
            20                  25                  30
```

```
Ile Cys His Ile Asn Glu Gln Leu Gly Leu Thr Thr Pro Pro Asn Asp
        35                  40                  45

Asp Lys Ile Ala Asp Val Val Cys Gly Thr Trp Asn Asn Ile Leu Asn
    50                  55                  60

Asn Asp His Asp Leu Leu Lys Lys Ser Gln Leu Thr Glu Leu Ile Leu
65                  70                  75                  80

Lys His Phe Pro Phe Leu Ala Ala Met Cys Tyr His Pro Pro Lys Lys
                85                  90                  95

Glu Gly Lys Lys Lys Gly Ser Gln Lys Glu Gln Gln Lys Glu Lys Glu
            100                 105                 110

Asn Glu Ala Gln Ser Gln Ala Glu Ala Leu Asn Pro Ser Glu Leu Ile
        115                 120                 125

Lys Val Leu Lys Thr Leu Val Lys Gln Leu Arg Thr Leu Arg Asn Tyr
    130                 135                 140

Tyr Ser His His Ser His Lys Lys Pro Asp Ala Glu Lys Asp Ile Phe
145                 150                 155                 160

Lys His Leu Tyr Lys Ala Phe Asp Ala Ser Leu Arg Met Val Lys Glu
                165                 170                 175

Asp Tyr Lys Ala His Phe Thr Val Asn Leu Thr Gln Asp Phe Ala His
            180                 185                 190

Leu Asn Arg Lys Gly Lys Asn Lys Gln Asp Asn Pro Asp Phe Asp Arg
        195                 200                 205

Tyr Arg Phe Glu Lys Asp Gly Phe Phe Thr Glu Ser Gly Leu Leu Phe
    210                 215                 220

Phe Thr Asn Leu Phe Leu Asp Lys Arg Asp Ala Tyr Trp Met Leu Lys
225                 230                 235                 240
```

```
Lys Val Ser Gly Phe Lys Ala Ser His Lys Gln Ser Glu Lys Met Thr
                245                 250                 255

Thr Glu Val Phe Cys Arg Ser Arg Ile Leu Leu Pro Lys Leu Arg Leu
            260                 265                 270

Glu Ser Arg Tyr Asp His Asn Gln Met Leu Leu Asp Met Leu Ser Glu
        275                 280                 285

Leu Ser Arg Tyr Pro Lys Leu Leu Tyr Glu Lys Leu Ser Glu Glu Asp
    290                 295                 300

Lys Lys Arg Phe Gln Val Glu Ala Asp Gly Phe Leu Asp Glu Ile Glu
305                 310                 315                 320

Glu Glu Gln Asn Pro Phe Lys Asp Thr Leu Ile Arg His Gln Asp Arg
                325                 330                 335

Phe Pro Tyr Phe Ala Leu Arg Tyr Leu Asp Leu Asn Glu Ser Phe Lys
            340                 345                 350

Ser Ile Arg Phe Gln Val Asp Leu Gly Thr Tyr His Tyr Cys Ile Tyr
        355                 360                 365

Asp Lys Lys Ile Gly Asp Glu Gln Glu Lys Arg His Leu Thr Arg Thr
    370                 375                 380

Leu Leu Ser Phe Gly Arg Leu Gln Asp Phe Thr Glu Ile Asn Arg Pro
385                 390                 395                 400

Gln Glu Trp Lys Ala Leu Thr Lys Asp Leu Asp Tyr Lys Glu Thr Ser
                405                 410                 415

Lys Gln Pro Phe Ile Ser Lys Thr Thr Pro His Tyr His Ile Thr Asp
            420                 425                 430

Asn Lys Ile Gly Phe Arg Leu Gly Thr Ser Lys Glu Leu Tyr Pro Ser
        435                 440                 445
```

```
Leu Glu Val Lys Asp Gly Ala Asn Arg Ile Ala Gln Tyr Pro Tyr Asn
    450                 455                 460

Ser Asp Phe Val Ala His Ala Phe Ile Ser Val His Glu Leu Leu Pro
465                 470                 475                 480

Leu Met Phe Tyr Gln His Leu Thr Gly Lys Ser Glu Asp Leu Leu Lys
                485                 490                 495

Glu Thr Val Arg His Ile Gln Arg Ile Tyr Lys Asp Phe Glu Glu Glu
            500                 505                 510

Arg Ile Asn Thr Ile Glu Asp Leu Glu Lys Ala Asn Gln Gly Arg Leu
        515                 520                 525

Pro Leu Gly Ala Phe Pro Lys Gln Met Leu Gly Leu Leu Gln Asn Lys
    530                 535                 540

Gln Pro Asp Leu Ser Glu Lys Ala Lys Ile Lys Ile Glu Lys Leu Ile
545                 550                 555                 560

Ala Glu Thr Lys Leu Leu Ser His Arg Leu Asn Thr Lys Leu Lys Ser
                565                 570                 575

Ser Pro Lys Leu Gly Lys Arg Arg Glu Lys Leu Ile Lys Thr Gly Val
            580                 585                 590

Leu Ala Asp Trp Leu Val Lys Asp Phe Met Arg Phe Gln Pro Val Ala
        595                 600                 605

Tyr Asp Ala Gln Asn Gln Pro Ile Glu Ser Ser Lys Ala Asn Ser Thr
    610                 615                 620

Glu Phe Gln Leu Ile Gln Arg Ala Leu Ala Leu Tyr Gly Gly Glu Lys
625                 630                 635                 640

Asn Arg Leu Glu Gly Tyr Phe Lys Gln Thr Asn Leu Ile Gly Asn Thr
                645                 650                 655
```

```
Asn Pro His Pro Phe Leu Asn Lys Phe Asn Trp Lys Ala Cys Arg Asn
            660                 665                 670

Leu Val Asp Phe Tyr Gln Gln Tyr Leu Glu Gln Arg Glu Lys Phe Leu
        675                 680                 685

Glu Ala Ile Lys Asn Gln Pro Trp Glu Pro Tyr Gln Tyr Cys Leu Leu
    690                 695                 700

Leu Lys Ile Pro Lys Glu Asn Arg Lys Asn Leu Val Lys Gly Trp Glu
705                 710                 715                 720

Gln Gly Gly Ile Ser Leu Pro Arg Gly Leu Phe Thr Glu Ala Ile Arg
                725                 730                 735

Glu Thr Leu Ser Lys Asp Leu Thr Leu Ser Lys Pro Ile Arg Lys Glu
            740                 745                 750

Ile Lys Lys His Gly Arg Val Gly Phe Ile Ser Arg Ala Ile Thr Leu
        755                 760                 765

Tyr Phe Arg Glu Lys Tyr Gln Asp Asp His Gln Ser Phe Tyr Asp Leu
    770                 775                 780

Pro Tyr Lys Leu Glu Ala Lys Ala Ser Pro Leu Pro Lys Lys Glu His
785                 790                 795                 800

Tyr Glu Tyr Trp Gln Gln Asn Lys Pro Gln Ser Pro Thr Glu Leu Gln
                805                 810                 815

Arg Leu Glu Leu His Thr Ser Asp Arg Trp Lys Asp Tyr Leu Leu Tyr
            820                 825                 830

Lys Arg Trp Gln His Leu Glu Lys Lys Leu Arg Leu Tyr Arg Asn Gln
        835                 840                 845

Asp Val Met Leu Trp Leu Met Thr Leu Glu Leu Thr Lys Asn His Phe
    850                 855                 860
```

```
Lys Glu Leu Asn Leu Asn Tyr His Gln Leu Lys Leu Glu Asn Leu Ala
865                 870                 875                 880


Val Asn Val Gln Glu Ala Asp Ala Lys Leu Asn Pro Leu Asn Gln Thr
                885                 890                 895


Leu Pro Met Val Leu Pro Val Lys Val Tyr Pro Ala Thr Ala Phe Gly
            900                 905                 910


Glu Val Gln Tyr Gln Glu Thr Pro Ile Arg Thr Val Tyr Ile Arg Glu
        915                 920                 925


Glu Gln Thr Lys Ala Leu Lys Met Gly Asn Phe Lys Ala Leu Val Lys
    930                 935                 940


Asp Arg Arg Leu Asn Gly Leu Phe Ser Phe Ile Lys Glu Glu Asn Asp
945                 950                 955                 960


Thr Gln Lys His Pro Ile Ser Gln Leu Arg Leu Arg Arg Glu Leu Glu
                965                 970                 975


Ile Tyr Gln Ser Leu Arg Val Asp Ala Phe Lys Glu Thr Leu Asn Leu
            980                 985                 990


Glu Glu Lys Leu Leu Lys Lys His  Thr Ser Leu Ser Ser  Val Glu Asn
        995                 1000                1005


Lys Phe  Arg Ile Leu Leu Glu  Glu Trp Lys Lys Glu  Tyr Ala Ala
    1010                1015                1020


Ser Ser  Met Val Thr Asp Glu  His Ile Ala Phe Ile  Ala Ser Val
    1025                1030                1035


Arg Asn  Ala Phe Cys His Asn  Gln Tyr Pro Phe Tyr  Glu Glu Ala
    1040                1045                1050


Leu His  Ala Pro Ile Pro Leu  Phe Thr Val Ala Gln  Gln Thr Thr
    1055                1060                1065
```

```
Glu Glu  Lys Asp Gly Leu Gly  Ile Ala Glu Ala Leu  Leu Arg Val
    1070                1075                1080


Leu Arg  Glu Tyr Cys Glu Ile  Val Lys Ser Gln Ile
    1085                1090                1095


<210>  153
<211>  1149
<212>  PRT
<213>  Sinomicrobium oceani

<400>  153

Met Glu Ser Thr Thr Thr Leu Gly Leu His Leu Lys Tyr Gln His Asp
1               5                   10                  15


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


Gln Asn Ile Glu Ser Ile Phe Gln Ala Phe Ala Glu Arg Tyr Gly Ile
        35                  40                  45


Gln Asn Pro Leu Arg Lys Asn Gly Val Pro Ala Ile Asn Asn Ile Phe
    50                  55                  60


His Asp Asn Ile Ser Ile Ser Asn Tyr Lys Glu Tyr Leu Lys Phe Leu
65                  70                  75                  80


Lys Gln Tyr Leu Pro Val Val Gly Phe Leu Glu Lys Ser Asn Glu Ile
                85                  90                  95


Asn Ile Phe Glu Phe Arg Glu Asp Phe Glu Ile Leu Ile Asn Ala Ile
            100                 105                 110


Tyr Lys Leu Arg His Phe Tyr Thr His Tyr Tyr His Ser Pro Ile Lys
        115                 120                 125


Leu Glu Asp Arg Phe Tyr Thr Cys Leu Asn Glu Leu Phe Val Ala Val
    130                 135                 140
```

```
Ala Ile Gln Val Lys Lys His Lys Met Lys Ser Asp Lys Thr Arg Gln
145                 150                 155                 160


Leu Leu Asn Lys Asn Leu His Gln Leu Leu Gln Gln Leu Ile Glu Gln
                165                 170                 175


Lys Arg Glu Lys Leu Lys Asp Lys Lys Ala Glu Gly Glu Lys Val Ser
            180                 185                 190


Leu Asp Thr Lys Ser Ile Glu Asn Ala Val Leu Asn Asp Ala Phe Val
        195                 200                 205


His Leu Leu Asp Lys Asp Glu Asn Ile Arg Leu Asn Tyr Ser Ser Arg
    210                 215                 220


Leu Ser Glu Asp Ile Ile Thr Lys Asn Gly Ile Thr Leu Ser Ile Ser
225                 230                 235                 240


Gly Leu Leu Phe Leu Leu Ser Leu Phe Leu Gln Arg Lys Glu Ala Glu
                245                 250                 255


Asp Leu Arg Ser Arg Ile Glu Gly Phe Lys Gly Lys Gly Asn Glu Leu
            260                 265                 270


Arg Phe Met Ala Thr His Trp Val Phe Ser Tyr Leu Asn Val Lys Arg
        275                 280                 285


Ile Lys His Arg Leu Asn Thr Asp Phe Gln Lys Glu Thr Leu Leu Ile
    290                 295                 300


Gln Ile Ala Asp Glu Leu Ser Lys Val Pro Asp Glu Val Tyr Lys Thr
305                 310                 315                 320


Leu Asp His Glu Asn Arg Ser Lys Phe Leu Glu Asp Ile Asn Glu Tyr
                325                 330                 335


Ile Arg Glu Gly Asn Glu Asp Ala Ser Leu Asn Glu Ser Thr Val Val
            340                 345                 350
```

```
His Gly Val Ile Arg Lys Arg Tyr Glu Asn Lys Phe His Tyr Leu Val
        355                 360                 365

Leu Arg Tyr Leu Asp Glu Phe Val Asp Phe Pro Ser Leu Arg Phe Gln
    370                 375                 380

Val His Leu Gly Asn Tyr Ile His Asp Arg Arg Asp Lys Val Ile Asp
385                 390                 395                 400

Gly Thr Asn Phe Ile Thr Asn Arg Val Ile Lys Glu Pro Ile Lys Val
                405                 410                 415

Phe Gly Lys Leu Ser His Val Ser Lys Leu Lys Ser Asp Tyr Met Glu
            420                 425                 430

Ser Leu Ser Arg Glu His Lys Asn Gly Trp Asp Val Phe Pro Asn Pro
        435                 440                 445

Ser Tyr Asn Phe Val Gly His Asn Ile Pro Ile Phe Ile Asn Leu Arg
    450                 455                 460

Ser Ala Ser Ser Lys Gly Lys Glu Leu Tyr Arg Asp Leu Met Lys Ile
465                 470                 475                 480

Lys Ser Glu Lys Lys Lys Lys Ser Arg Glu Glu Gly Ile Pro Met Glu
                485                 490                 495

Arg Arg Asp Gly Lys Pro Thr Lys Ile Glu Ile Ser Asn Gln Ile Asp
            500                 505                 510

Arg Asn Ile Lys Asp Asn Asn Phe Lys Asp Ile Tyr Pro Gly Glu Pro
        515                 520                 525

Leu Ala Met Leu Ser Leu Asn Glu Leu Pro Ala Leu Leu Phe Glu Leu
    530                 535                 540

Leu Arg Arg Pro Ser Ile Thr Pro Gln Asp Ile Glu Asp Arg Met Val
545                 550                 555                 560
```

```
Glu Lys Leu Tyr Glu Arg Phe Gln Ile Ile Arg Asp Tyr Lys Pro Gly
                565                 570                 575

Asp Gly Leu Ser Thr Ser Lys Ile Ser Lys Lys Leu Arg Lys Ala Asp
            580                 585                 590

Asn Ser Thr Arg Leu Asp Gly Lys Lys Leu Leu Arg Ala Ile Gln Thr
        595                 600                 605

Glu Thr Arg Asn Ala Arg Glu Lys Leu His Thr Leu Glu Glu Asn Lys
    610                 615                 620

Ala Leu Gln Lys Asn Arg Lys Arg Arg Thr Val Tyr Thr Thr Arg Glu
625                 630                 635                 640

Gln Gly Arg Glu Ala Ser Trp Leu Ala Gln Asp Leu Lys Arg Phe Met
                645                 650                 655

Pro Ile Ala Ser Arg Lys Glu Trp Arg Gly Tyr His His Ser Gln Leu
            660                 665                 670

Gln Gln Ile Leu Ala Phe Tyr Asp Gln Asn Pro Lys Gln Pro Leu Glu
        675                 680                 685

Leu Leu Glu Gln Phe Trp Asp Leu Lys Glu Asp Thr Tyr Val Trp Asn
    690                 695                 700

Ser Trp Ile His Lys Ser Leu Ser Gln His Asn Gly Phe Val Pro Met
705                 710                 715                 720

Tyr Glu Gly Tyr Leu Lys Gly Arg Leu Gly Tyr Tyr Lys Lys Leu Glu
                725                 730                 735

Ser Asp Ile Ile Gly Phe Leu Glu Glu His Lys Val Leu Lys Arg Tyr
            740                 745                 750

Tyr Thr Gln Gln His Leu Asn Val Ile Phe Arg Glu Arg Leu Tyr Phe
        755                 760                 765
```

```
Ile Lys Thr Glu Thr Lys Gln Lys Leu Glu Leu Leu Ala Arg Pro Leu
    770                 775                 780

Val Phe Pro Arg Gly Ile Phe Asp Asp Lys Pro Thr Phe Val Gln Asp
785                 790                 795                 800

Lys Lys Val Val Asp His Pro Glu Leu Phe Ala Asp Trp Tyr Val Tyr
                805                 810                 815

Ser Tyr Lys Asp Asp His Ser Phe Gln Glu Phe Tyr His Tyr Lys Arg
            820                 825                 830

Asp Tyr Asn Glu Ile Phe Glu Thr Glu Leu Ser Trp Asp Ile Asp Phe
        835                 840                 845

Lys Asp Asn Lys Arg Gln Leu Asn Pro Ser Glu Gln Met Asp Leu Phe
    850                 855                 860

Arg Met Lys Trp Asp Leu Lys Ile Lys Lys Ile Lys Ile Gln Asp Ile
865                 870                 875                 880

Phe Leu Lys Ile Val Ala Glu Asp Ile Tyr Leu Lys Ile Phe Gly His
                885                 890                 895

Lys Ile Pro Leu Ser Leu Ser Asp Phe Tyr Ile Ser Arg Gln Glu Arg
            900                 905                 910

Leu Thr Leu Asp Glu Gln Ala Val Ala Gln Ser Met Arg Leu Pro Gly
        915                 920                 925

Asp Thr Ser Glu Asn Gln Ile Lys Glu Ser Asn Leu Trp Gln Thr Thr
    930                 935                 940

Val Pro Tyr Glu Lys Glu Gln Ile Arg Glu Pro Lys Ile Lys Leu Lys
945                 950                 955                 960

Asp Ile Gly Lys Phe Lys Tyr Phe Leu Gln Gln Gln Lys Val Leu Asn
                965                 970                 975
```

```
Leu Leu Lys Tyr Asp Pro Gln His Val Trp Thr Lys Ala Glu Leu Glu
            980                 985                 990


Glu Glu Leu Tyr Ile Gly Lys His  Ser Tyr Glu Val Val  Arg Arg Glu
        995                 1000                1005


Met Leu  Leu Gln Lys Cys His  Gln Leu Glu Lys His  Ile Leu Glu
    1010                1015                1020


Gln Phe  Arg Phe Asp Gly Ser  Asn His Pro Arg Glu  Leu Glu Gln
    1025                1030                1035


Gly Asn  His Pro Asn Phe Lys  Met Tyr Ile Val Asn  Gly Ile Leu
    1040                1045                1050


Thr Lys  Arg Gly Glu Leu Glu  Ile Glu Ala Glu Asn  Trp Trp Leu
    1055                1060                1065


Glu Leu  Gly Asn Ser Lys Asn  Ser Leu Asp Lys Val  Glu Val Glu
    1070                1075                1080


Leu Leu  Thr Met Lys Thr Ile  Pro Glu Gln Lys Ala  Phe Leu Leu
    1085                1090                1095


Ile Leu  Ile Arg Asn Lys Phe  Ala His Asn Gln Leu  Pro Ala Asp
    1100                1105                1110


Asn Tyr  Phe His Tyr Ala Ser  Asn Leu Met Asn Leu  Lys Lys Ser
    1115                1120                1125


Asp Thr  Tyr Ser Leu Phe Trp  Phe Thr Val Ala Asp  Thr Ile Val
    1130                1135                1140


Gln Glu  Phe Met Ser Leu
    1145
```

<210> 154
<211> 1133
<212> PRT
<213> Reichenbachiella agariperforans

<400> 154

```
Met Lys Thr Asn Pro Leu Ile Ala Ser Ser Gly Glu Lys Pro Asn Tyr
1               5                  10                  15


Lys Lys Phe Asn Thr Glu Ser Asp Lys Ser Phe Lys Lys Ile Phe Gln
            20                  25                  30


Asn Lys Gly Ser Ile Ala Pro Ile Ala Glu Lys Ala Cys Lys Asn Phe
        35                  40                  45


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


Phe Ser Val Gly His Ala Phe Lys Asn Ile Asp Ser Lys Asn Val Phe
65                  70                  75                  80


Arg Tyr Glu Leu Asp Glu Ser Gln Met Asp Met Lys Pro Thr Gln Phe
                85                  90                  95


Leu Ala Leu Gln Lys Glu Phe Phe Asp Phe Gln Gly Ala Leu Asn Gly
            100                 105                 110


Leu Leu Lys His Ile Arg Asn Val Asn Ser His Tyr Val His Thr Phe
        115                 120                 125


Glu Lys Leu Glu Ile Gln Ser Ile Asn Gln Lys Leu Ile Thr Phe Leu
    130                 135                 140


Ile Glu Ala Phe Glu Leu Ala Val Ile His Ser Tyr Leu Asn Glu Glu
145                 150                 155                 160


Glu Leu Ser Tyr Glu Ala Tyr Lys Asp Asp Pro Gln Ser Gly Gln Lys
                165                 170                 175


Leu Val Gln Phe Leu Cys Asp Lys Phe Tyr Pro Asn Lys Glu His Glu
            180                 185                 190


Val Glu Glu Arg Lys Thr Ile Leu Ala Lys Asn Lys Arg Gln Ala Leu
```

```
        195                 200                 205

Glu His Leu Leu Phe Ile Glu Val Thr Ser Asp Ile Asp Trp Lys Leu
    210                 215                 220

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

Phe His Ala Cys Leu Phe Leu Leu Ser Leu Phe Leu Tyr Lys Ser Glu
                245                 250                 255

Ala Asn Gln Leu Ile Ser Lys Ile Lys Gly Phe Lys Arg Asn Asp Asp
            260                 265                 270

Asn Gln Tyr Arg Ser Lys Arg Gln Ile Phe Thr Phe Phe Ser Lys Lys
        275                 280                 285

Phe Thr Ser Gln Asp Val Asn Ser Glu Glu Gln His Leu Val Lys Phe
    290                 295                 300

Arg Asp Val Ile Gln Tyr Leu Asn His Tyr Pro Ser Ala Trp Asn Lys
305                 310                 315                 320

His Leu Glu Leu Lys Ser Gly Tyr Pro Gln Met Thr Asp Lys Leu Met
                325                 330                 335

Arg Tyr Ile Val Glu Ala Glu Ile Tyr Arg Ser Phe Pro Asp Gln Thr
            340                 345                 350

Asp Asn His Arg Phe Leu Leu Phe Ala Ile Arg Glu Phe Phe Gly Gln
        355                 360                 365

Ser Cys Leu Asp Thr Trp Thr Gly Asn Thr Pro Ile Asn Phe Ser Asn
    370                 375                 380

Gln Glu Gln Lys Gly Phe Ser Tyr Glu Ile Asn Thr Ser Ala Glu Ile
385                 390                 395                 400

Lys Asp Ile Glu Thr Lys Leu Lys Ala Leu Val Leu Lys Gly Pro Leu
```

405 410 415

Asn Phe Lys Glu Lys Lys Glu Gln Asn Arg Leu Glu Lys Asp Leu Arg
420 425 430

Arg Glu Lys Lys Glu Gln Pro Thr Asn Arg Val Lys Glu Lys Leu Leu
435 440 445

Thr Arg Ile Gln His Asn Met Leu Tyr Val Ser Tyr Gly Arg Asn Gln
450 455 460

Asp Arg Phe Met Asp Phe Ala Ala Arg Phe Leu Ala Glu Thr Asp Tyr
465 470 475 480

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

Glu Gln Arg Asp His Leu Lys Glu Gln Lys Lys Glu Leu Pro Lys Lys
500 505 510

Glu Phe Glu Lys Leu Lys Tyr His Gln Ser Lys Leu Val Asp Tyr Phe
515 520 525

Thr Tyr Ala Glu Gln Gln Ala Arg Tyr Pro Asp Trp Asp Thr Pro Phe
530 535 540

Val Val Glu Asn Asn Ala Ile Gln Ile Lys Val Thr Leu Phe Asn Gly
545 550 555 560

Ala Lys Lys Ile Val Ser Val Gln Arg Asn Leu Met Leu Tyr Leu Leu
565 570 575

Glu Asp Ala Leu Tyr Ser Glu Lys Arg Glu Asn Ala Gly Lys Gly Leu
580 585 590

Ile Ser Gly Tyr Phe Val His His Gln Lys Glu Leu Lys Asp Gln Leu
595 600 605

Asp Ile Leu Glu Lys Glu Thr Glu Ile Ser Arg Glu Gln Lys Arg Glu

610 615 620

Phe Lys Lys Leu Leu Pro Lys Arg Leu Leu His Arg Tyr Ser Pro Ala
625 630 635 640

Gln Ile Asn Asp Thr Thr Glu Trp Asn Pro Met Glu Val Ile Leu Glu
645 650 655

Glu Ala Lys Ala Gln Glu Gln Arg Tyr Gln Leu Leu Leu Glu Lys Ala
660 665 670

Ile Leu His Gln Thr Glu Glu Asp Phe Leu Lys Arg Asn Lys Gly Lys
675 680 685

Gln Phe Lys Leu Arg Phe Val Arg Lys Ala Trp His Leu Met Tyr Leu
690 695 700

Lys Glu Leu Tyr Met Asn Lys Val Ala Glu His Gly His His Lys Ser
705 710 715 720

Phe His Ile Thr Lys Glu Glu Phe Asn Asp Phe Cys Arg Trp Met Phe
725 730 735

Ala Phe Asp Glu Val Pro Lys Tyr Lys Glu Tyr Leu Cys Asp Tyr Phe
740 745 750

Ser Gln Lys Gly Phe Phe Asn Asn Ala Glu Phe Lys Asp Leu Ile Glu
755 760 765

Ser Ser Thr Ser Leu Asn Asp Leu Tyr Glu Lys Thr Lys Gln Arg Phe
770 775 780

Glu Gly Trp Ser Lys Asp Leu Thr Lys Gln Ser Asp Glu Asn Lys Tyr
785 790 795 800

Leu Leu Ala Asn Tyr Glu Ser Met Leu Lys Asp Asp Met Leu Tyr Val
805 810 815

Asn Ile Ser His Phe Ile Ser Tyr Leu Glu Ser Lys Gly Lys Ile Asn

820 825 830

Arg Asn Ala His Gly His Ile Ala Tyr Lys Ala Leu Asn Asn Val Pro
835 840 845

His Leu Ile Glu Glu Tyr Tyr Tyr Lys Asp Arg Leu Ala Pro Glu Glu
850 855 860

Tyr Lys Ser His Gly Lys Leu Tyr Asn Lys Leu Lys Thr Val Lys Leu
865 870 875 880

Glu Asp Ala Leu Leu Tyr Glu Met Ala Met His Tyr Leu Ser Leu Glu
885 890 895

Pro Ala Leu Val Pro Lys Val Lys Thr Lys Val Lys Asp Ile Leu Ser
900 905 910

Ser Asn Ile Ala Phe Asp Ile Lys Asp Ala Ala Gly His His Leu Tyr
915 920 925

His Leu Leu Ile Pro Phe His Lys Ile Asp Ser Phe Val Ala Leu Ile
930 935 940

Asn His Gln Ser Gln Gln Glu Lys Asp Pro Asp Lys Thr Ser Phe Leu
945 950 955 960

Ala Lys Ile Gln Pro Tyr Leu Glu Lys Val Lys Asn Ser Lys Asp Leu
965 970 975

Lys Ala Val Tyr His Tyr Tyr Lys Asp Thr Pro His Thr Leu Arg Tyr
980 985 990

Glu Asp Leu Asn Met Ile His Ser His Ile Val Ser Gln Ser Val Gln
995 1000 1005

Phe Thr Lys Val Ala Leu Lys Leu Glu Glu Tyr Phe Ile Ala Lys
1010 1015 1020

Lys Ser Ile Thr Leu Gln Ile Ala Arg Gln Ile Ser Tyr Ser Glu

1025 1030 1035

Ile Ala Asp Leu Ser Asn Tyr Phe Thr Asp Glu Val Arg Asn Thr
1040 1045 1050

Ala Phe His Phe Asp Val Pro Glu Thr Ala Tyr Ser Met Ile Leu
1055 1060 1065

Gln Gly Ile Glu Ser Glu Phe Leu Asp Arg Glu Ile Lys Pro Gln
1070 1075 1080

Lys Pro Lys Ser Leu Ser Glu Leu Ser Thr Gln Gln Val Ser Val
1085 1090 1095

Cys Thr Ala Phe Leu Glu Thr Leu His Asn Asn Leu Phe Asp Arg
1100 1105 1110

Lys Asp Asp Lys Lys Glu Arg Leu Ser Lys Ala Arg Glu Arg Tyr
1115 1120 1125

Phe Glu Gln Ile Asn
1130

<210> 155
<211> 1115
<212> PRT
<213> Fusobacterium necrophorum

<400> 155

Met Glu Lys Phe Arg Arg Gln Asn Arg Asn Ser Ile Ile Lys Ile Ile
1 5 10 15

Ile Ser Asn Tyr Asp Thr Lys Gly Ile Lys Glu Leu Lys Val Arg Tyr
20 25 30

Arg Lys Gln Ala Gln Leu Asp Thr Phe Ile Ile Lys Thr Glu Ile Val
35 40 45

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

```
Tyr Arg Tyr Ser Phe Leu Phe Asp Gly Glu Glu Lys Tyr His Phe Lys
65                  70                  75                  80

Asn Lys Ser Ser Val Glu Ile Val Lys Lys Asp Ile Phe Ser Gln Thr
                85                  90                  95

Pro Asp Asn Met Ile Arg Asn Tyr Lys Ile Thr Leu Lys Ile Ser Glu
            100                 105                 110

Lys Asn Pro Arg Val Val Glu Ala Glu Ile Glu Asp Leu Met Asn Ser
        115                 120                 125

Thr Ile Leu Lys Asp Gly Arg Arg Ser Ala Arg Arg Glu Lys Ser Met
    130                 135                 140

Thr Glu Arg Lys Leu Ile Glu Glu Lys Val Ala Lys Asn Tyr Ser Leu
145                 150                 155                 160

Leu Ala Asn Cys Pro Met Glu Glu Val Asp Ser Ile Lys Ile Tyr Lys
                165                 170                 175

Ile Lys Arg Phe Leu Thr Tyr Arg Ser Asn Met Leu Leu Tyr Phe Ala
            180                 185                 190

Ser Ile Asn Ser Phe Leu Cys Glu Gly Ile Lys Gly Lys Asp Asn Glu
        195                 200                 205

Thr Glu Glu Ile Trp His Leu Lys Asp Asn Asp Val Arg Lys Glu Lys
    210                 215                 220

Val Arg Glu Asn Phe Lys Asn Lys Leu Ile Gln Ser Thr Glu Asn Tyr
225                 230                 235                 240

Asn Ser Ser Leu Lys Asn Gln Ile Glu Glu Lys Glu Lys Leu Leu Arg
                245                 250                 255

Lys Glu Phe Lys Lys Gly Ala Phe Tyr Arg Thr Ile Ile Lys Lys Leu
            260                 265                 270
```

```
Gln Gln Glu Arg Ile Lys Glu Leu Ser Glu Lys Ser Leu Thr Glu Asp
        275                 280                 285

Cys Glu Lys Ile Ile Lys Leu Tyr Ser Lys Leu Arg His Ser Leu Met
    290                 295                 300

His Tyr Asp Tyr Gln Tyr Phe Glu Asn Leu Phe Glu Asn Lys Lys Asn
305                 310                 315                 320

Asp Asp Leu Met Lys Asp Leu Asn Leu Asp Leu Phe Lys Ser Leu Pro
                325                 330                 335

Leu Ile Arg Lys Met Lys Leu Asn Asn Lys Val Asn Tyr Leu Glu Asp
            340                 345                 350

Gly Asp Thr Leu Phe Val Leu Gln Lys Thr Lys Lys Ala Lys Thr Leu
        355                 360                 365

Tyr Gln Ile Tyr Asp Ala Leu Cys Glu Gln Lys Asn Gly Phe Asn Lys
    370                 375                 380

Phe Ile Asn Asp Phe Phe Val Ser Asp Gly Glu Glu Asn Thr Val Phe
385                 390                 395                 400

Lys Gln Ile Ile Asn Glu Lys Phe Gln Ser Glu Met Glu Phe Leu Glu
                405                 410                 415

Lys Arg Ile Ser Glu Ser Glu Lys Lys Asn Glu Lys Leu Lys Lys Lys
            420                 425                 430

Leu Asp Ser Met Lys Ala His Phe Arg Asn Ile Asn Ser Glu Asp Thr
        435                 440                 445

Lys Glu Ala Tyr Phe Trp Asp Ile His Ser Ser Arg Asn Tyr Lys Thr
    450                 455                 460

Lys Tyr Asn Glu Arg Lys Asn Leu Val Asn Glu Tyr Thr Glu Leu Leu
465                 470                 475                 480
```

```
Gly Ser Ser Lys Glu Lys Lys Leu Leu Arg Glu Glu Ile Thr Lys Ile
                485                 490                 495

Asn Arg Gln Leu Leu Lys Leu Lys Gln Glu Met Glu Glu Ile Thr Lys
            500                 505                 510

Lys Asn Ser Leu Phe Arg Leu Glu Tyr Lys Met Lys Ile Ala Phe Gly
        515                 520                 525

Phe Leu Phe Cys Glu Phe Asp Gly Asn Ile Ser Lys Phe Lys Asp Glu
    530                 535                 540

Phe Asp Ala Ser Asn Gln Glu Lys Ile Ile Gln Tyr His Lys Asn Gly
545                 550                 555                 560

Glu Lys Tyr Leu Thr Ser Phe Leu Lys Glu Glu Glu Lys Glu Lys Phe
                565                 570                 575

Asn Leu Glu Lys Met Gln Lys Ile Ile Gln Lys Thr Glu Glu Glu Asp
            580                 585                 590

Trp Leu Leu Pro Glu Thr Lys Asn Asn Leu Phe Lys Phe Tyr Leu Leu
        595                 600                 605

Thr Tyr Leu Leu Leu Pro Tyr Glu Leu Lys Gly Asp Phe Leu Gly Phe
    610                 615                 620

Val Lys Lys His Tyr Tyr Asp Ile Lys Asn Val Asp Phe Ile Asp Glu
625                 630                 635                 640

Asn Gln Asn Asn Ile Gln Val Ser Gln Thr Val Glu Lys Gln Glu Asp
                645                 650                 655

Tyr Phe Tyr His Lys Ile Arg Leu Phe Glu Lys Asn Thr Lys Lys Tyr
            660                 665                 670

Glu Ile Val Lys Tyr Ser Ile Val Pro Asn Glu Lys Leu Lys Gln Tyr
        675                 680                 685
```

```
Phe Glu Asp Leu Gly Ile Asp Ile Lys Tyr Leu Thr Val Glu Gln Lys
    690                 695                 700

Ser Glu Val Ser Glu Glu Lys Asn Lys Lys Val Ser Leu Lys Asn Asn
705                 710                 715                 720

Gly Met Phe Asn Lys Thr Ile Leu Leu Phe Val Phe Lys Tyr Tyr Gln
                725                 730                 735

Ile Ala Phe Lys Leu Phe Asn Asp Ile Glu Leu Tyr Ser Leu Phe Phe
            740                 745                 750

Leu Arg Glu Lys Ser Gly Lys Pro Leu Glu Ile Phe Arg Lys Glu Leu
        755                 760                 765

Glu Ser Lys Met Lys Asp Gly Tyr Leu Asn Phe Gly Gln Leu Leu Tyr
    770                 775                 780

Val Val Tyr Glu Val Leu Val Lys Asn Lys Asp Leu Asp Lys Ile Leu
785                 790                 795                 800

Ser Lys Lys Ile Asp Tyr Arg Lys Asp Lys Ser Phe Ser Pro Glu Ile
                805                 810                 815

Ala Tyr Leu Arg Asn Phe Leu Ser His Leu Asn Tyr Ser Lys Phe Leu
            820                 825                 830

Asp Asn Phe Met Lys Ile Asn Thr Asn Lys Ser Asp Glu Asn Lys Glu
        835                 840                 845

Val Leu Ile Pro Ser Ile Lys Ile Gln Lys Met Ile Gln Phe Ile Glu
    850                 855                 860

Lys Cys Asn Leu Gln Asn Gln Ile Asp Phe Asp Phe Asn Phe Val Asn
865                 870                 875                 880

Asp Phe Tyr Met Arg Lys Glu Lys Met Phe Phe Ile Gln Leu Lys Gln
                885                 890                 895
```

```
Ile Phe Pro Asp Ile Asn Ser Thr Glu Lys Gln Lys Met Asn Glu Lys
            900                 905                 910

Glu Glu Ile Leu Arg Asn Arg Tyr His Leu Thr Asp Lys Lys Asn Glu
        915                 920                 925

Gln Ile Lys Asp Glu His Glu Ala Gln Ser Gln Leu Tyr Glu Lys Ile
    930                 935                 940

Leu Ser Leu Gln Lys Ile Tyr Ser Ser Asp Lys Asn Asn Phe Tyr Gly
945                 950                 955                 960

Arg Leu Lys Glu Glu Lys Leu Leu Phe Leu Glu Lys Gln Gly Lys Lys
                965                 970                 975

Lys Leu Ser Met Glu Glu Ile Lys Asp Lys Ile Ala Gly Asp Ile Ser
            980                 985                 990

Asp Leu Leu Gly Ile Leu Lys Lys  Glu Ile Thr Arg Asp  Ile Lys Asp
        995                 1000                1005

Lys Leu  Thr Glu Lys Phe Arg  Tyr Cys Glu Glu Lys  Leu Leu Asn
    1010                1015                1020

Leu Ser  Phe Tyr Asn His Gln  Asp Lys Lys Lys Glu  Glu Ser Ile
    1025                1030                1035

Arg Val  Phe Leu Ile Arg Asp  Lys Asn Ser Asp Asn  Phe Lys Phe
    1040                1045                1050

Glu Ser  Ile Leu Asp Asp Gly  Ser Asn Lys Ile Phe  Ile Ser Lys
    1055                1060                1065

Asn Gly  Lys Glu Ile Thr Ile  Gln Cys Cys Asp Lys  Val Leu Glu
    1070                1075                1080

Thr Leu  Ile Ile Glu Lys Asn  Thr Leu Lys Ile Ser  Ser Asn Gly
    1085                1090                1095
```

```
Lys Ile  Ile Ser Leu Ile Pro  His Tyr Ser Tyr Ser  Ile Asp Val
    1100                1105                1110


Lys Tyr
    1115


<210>  156
<211>  1111
<212>  PRT
<213>  Fusobacterium necrophorum

<400>  156

Met Glu Lys Phe Arg Arg Gln Asn Arg Ser Ser Ile Ile Lys Ile Ile
1               5                  10                  15


Ile Ser Asn Tyr Asp Thr Lys Gly Ile Lys Glu Leu Lys Val Arg Tyr
            20                  25                  30


Arg Lys Gln Ala Gln Leu Asp Thr Phe Ile Ile Lys Thr Glu Ile Val
        35                  40                  45


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


Tyr Arg Tyr Ser Phe Leu Phe Asp Gly Glu Glu Lys Tyr His Phe Lys
65                  70                  75                  80


Asn Lys Ser Ser Val Glu Ile Val Lys Lys Asp Ile Phe Ser Gln Thr
                85                  90                  95


Pro Asp Asn Met Ile Arg Asn Tyr Lys Ile Thr Leu Lys Ile Ser Glu
            100                 105                 110


Lys Asn Pro Arg Val Val Glu Ala Glu Ile Glu Asp Leu Met Asn Ser
        115                 120                 125


Thr Ile Leu Lys Asp Gly Arg Arg Ser Ala Arg Arg Glu Lys Ser Met
    130                 135                 140
```

```
Thr Glu Arg Lys Leu Ile Glu Glu Lys Val Ala Glu Asn Tyr Ser Leu
145                 150                 155                 160


Leu Ala Asn Cys Pro Met Glu Glu Val Asp Ser Ile Lys Ile Tyr Lys
                165                 170                 175


Ile Lys Arg Phe Leu Thr Tyr Arg Ser Asn Met Leu Leu Tyr Phe Ala
            180                 185                 190


Ser Ile Asn Ser Phe Leu Cys Glu Gly Ile Lys Gly Lys Asp Asn Glu
        195                 200                 205


Thr Glu Glu Ile Trp His Leu Lys Asp Asn Asp Val Arg Lys Glu Lys
    210                 215                 220


Val Lys Glu Asn Phe Lys Asn Lys Leu Ile Gln Ser Thr Glu Asn Tyr
225                 230                 235                 240


Asn Ser Ser Leu Lys Asn Gln Ile Glu Glu Lys Glu Lys Leu Leu Arg
                245                 250                 255


Lys Glu Ser Lys Lys Gly Ala Phe Tyr Arg Thr Ile Ile Lys Lys Leu
            260                 265                 270


Gln Gln Glu Arg Ile Lys Glu Leu Ser Glu Lys Ser Leu Thr Glu Asp
        275                 280                 285


Cys Glu Lys Ile Ile Lys Leu Tyr Ser Glu Leu Arg His Pro Leu Met
    290                 295                 300


His Tyr Asp Tyr Gln Tyr Phe Glu Asn Leu Phe Glu Asn Lys Glu Asn
305                 310                 315                 320


Ser Glu Leu Thr Lys Asn Leu Asn Leu Asp Ile Phe Lys Ser Leu Pro
                325                 330                 335


Leu Val Arg Lys Met Lys Leu Asn Asn Lys Val Asn Tyr Leu Glu Asp
            340                 345                 350
```

```
Asn Asp Thr Leu Phe Val Leu Gln Lys Thr Lys Lys Ala Lys Thr Leu
        355                 360                 365

Tyr Gln Ile Tyr Asp Ala Leu Cys Glu Gln Lys Asn Gly Phe Asn Lys
    370                 375                 380

Phe Ile Asn Asp Phe Phe Val Ser Asp Gly Glu Glu Asn Thr Val Phe
385                 390                 395                 400

Lys Gln Ile Ile Asn Glu Lys Phe Gln Ser Glu Ile Glu Phe Leu Glu
                405                 410                 415

Lys Arg Ile Ser Glu Ser Glu Lys Lys Asn Glu Lys Leu Lys Lys Lys
            420                 425                 430

Leu Asp Ser Met Lys Ala His Phe Arg Asn Ile Asn Ser Glu Asp Thr
        435                 440                 445

Lys Glu Ala Tyr Phe Trp Asp Ile His Ser Ser Arg Asn Tyr Lys Thr
    450                 455                 460

Lys Tyr Asn Glu Arg Lys Asn Leu Val Asn Glu Tyr Thr Glu Leu Leu
465                 470                 475                 480

Gly Ser Ser Lys Glu Lys Lys Leu Leu Arg Glu Glu Ile Thr Lys Ile
                485                 490                 495

Asn Arg Gln Leu Leu Lys Leu Lys Gln Glu Met Glu Glu Ile Thr Lys
            500                 505                 510

Lys Asn Ser Leu Phe Arg Leu Glu Tyr Lys Met Lys Met Ala Phe Gly
        515                 520                 525

Phe Leu Phe Cys Glu Phe Asp Gly Asn Ile Ser Arg Phe Lys Asp Glu
    530                 535                 540

Phe Asp Ala Ser Asn Gln Glu Lys Ile Ile Gln Tyr His Lys Asn Gly
545                 550                 555                 560
```

```
Glu Lys Tyr Leu Thr Tyr Phe Leu Lys Glu Glu Glu Lys Glu Lys Phe
                565                 570                 575

Asn Leu Lys Lys Leu Gln Glu Thr Ile Gln Lys Thr Gly Glu Glu Asn
            580                 585                 590

Trp Leu Leu Pro Gln Asn Lys Asn Asn Leu Phe Lys Phe Tyr Leu Leu
        595                 600                 605

Thr Tyr Leu Leu Leu Pro Tyr Glu Leu Lys Gly Asp Phe Leu Gly Phe
    610                 615                 620

Val Lys Lys His Tyr Tyr Asp Ile Lys Asn Val Asp Phe Met Asp Glu
625                 630                 635                 640

Asn Gln Ser Ser Lys Ile Ile Glu Ser Lys Glu Asp Asp Phe Tyr His
                645                 650                 655

Lys Ile Arg Leu Phe Glu Lys Asn Thr Lys Lys Tyr Glu Ile Val Lys
            660                 665                 670

Tyr Ser Ile Val Pro Asp Lys Lys Leu Lys Gln Tyr Phe Lys Asp Leu
        675                 680                 685

Gly Ile Asp Thr Lys Tyr Leu Ile Leu Asp Gln Lys Ser Glu Val Ser
    690                 695                 700

Gly Glu Lys Asn Lys Lys Val Ser Leu Lys Asn Asn Gly Met Phe Asn
705                 710                 715                 720

Lys Thr Ile Leu Leu Phe Val Phe Lys Tyr Tyr Gln Ile Ala Phe Lys
                725                 730                 735

Leu Phe Asn Asp Ile Glu Leu Tyr Ser Leu Phe Phe Leu Arg Glu Lys
            740                 745                 750

Ser Gly Lys Pro Phe Glu Val Phe Leu Lys Glu Leu Lys Asp Lys Met
        755                 760                 765
```

```
Ile Gly Lys Gln Leu Asn Phe Gly Gln Leu Leu Tyr Val Val Tyr Glu
    770                 775                 780

Val Leu Val Lys Asn Lys Asp Leu Ser Glu Ile Leu Ser Glu Arg Ile
785                 790                 795                 800

Asp Tyr Arg Lys Asp Met Cys Phe Ser Ala Glu Ile Ala Asp Leu Arg
                805                 810                 815

Asn Phe Leu Ser His Leu Asn Tyr Ser Lys Phe Leu Asp Asn Phe Met
            820                 825                 830

Lys Ile Asn Thr Asn Lys Ser Asp Glu Asn Lys Glu Val Leu Ile Pro
        835                 840                 845

Ser Ile Lys Ile Gln Lys Met Ile Lys Phe Ile Glu Glu Cys Asn Leu
    850                 855                 860

Gln Ser Gln Ile Asp Phe Asp Phe Asn Phe Val Asn Asp Phe Tyr Met
865                 870                 875                 880

Arg Lys Glu Lys Met Phe Phe Ile Gln Leu Lys Gln Ile Phe Pro Asp
                885                 890                 895

Ile Asn Ser Thr Glu Lys Gln Lys Met Asn Glu Lys Glu Glu Ile Leu
            900                 905                 910

Arg Asn Arg Tyr His Leu Thr Asp Lys Lys Asn Glu Gln Ile Lys Asp
        915                 920                 925

Glu His Glu Ala Gln Ser Gln Leu Tyr Glu Lys Ile Leu Ser Leu Gln
    930                 935                 940

Lys Ile Tyr Ser Ser Asp Lys Asn Asn Phe Tyr Gly Arg Leu Lys Glu
945                 950                 955                 960

Glu Lys Leu Leu Phe Leu Glu Lys Gln Glu Lys Lys Lys Leu Ser Met
                965                 970                 975
```

```
Glu Glu Ile Lys Asp Lys Ile Ala Gly Asp Ile Ser Asp Leu Leu Gly
            980                 985                 990


Ile Leu Lys Lys Glu Ile Thr Arg  Asp Ile Lys Asp Lys  Leu Thr Glu
        995                 1000                1005


Lys Phe  Arg Tyr Cys Glu Glu  Lys Leu Leu Asn Leu  Ser Phe Tyr
    1010                1015                1020


Asn His  Gln Asp Lys Lys Lys  Glu Glu Ser Ile Arg  Val Phe Leu
    1025                1030                1035


Ile Arg  Asp Lys Asn Ser Asp  Asn Phe Lys Phe Glu  Ser Ile Leu
    1040                1045                1050


Asp Asp  Gly Ser Asn Lys Ile  Phe Ile Ser Lys Asn  Gly Lys Glu
    1055                1060                1065


Ile Thr  Ile Gln Cys Cys Asp  Lys Val Leu Glu Thr  Leu Ile Ile
    1070                1075                1080


Glu Lys  Asn Thr Leu Lys Ile  Ser Ser Asn Gly Lys  Ile Ile Ser
    1085                1090                1095


Leu Ile  Pro His Tyr Ser Tyr  Ser Ile Asp Val Lys  Tyr
    1100                1105                1110


<210>  157
<211>  1110
<212>  PRT
<213>  Fusobacterium necrophorum

<400>  157

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


Lys Thr Glu Ile Val Asn Asn Asp Ile Phe Ile Lys Ser Ile Ile Glu
            20                  25                  30
```

```
Lys Ala Arg Glu Lys Tyr Arg Tyr Ser Phe Leu Phe Asp Gly Glu Glu
        35                  40                  45

Lys Tyr His Phe Lys Asn Lys Ser Ser Val Glu Ile Val Lys Asn Asp
    50                  55                  60

Ile Phe Ser Gln Thr Pro Asp Asn Met Ile Arg Asn Tyr Lys Ile Thr
65                  70                  75                  80

Leu Lys Ile Ser Glu Lys Asn Pro Arg Val Val Glu Ala Glu Ile Glu
                85                  90                  95

Asp Leu Met Asn Ser Thr Ile Leu Lys Asp Gly Arg Arg Ser Ala Arg
            100                 105                 110

Arg Glu Lys Ser Met Thr Glu Arg Lys Leu Ile Glu Glu Lys Val Ala
        115                 120                 125

Glu Asn Tyr Ser Leu Leu Ala Asn Cys Pro Ile Glu Glu Val Asp Ser
    130                 135                 140

Ile Lys Ile Tyr Lys Ile Lys Arg Phe Leu Thr Tyr Arg Ser Asn Met
145                 150                 155                 160

Leu Leu Tyr Phe Ala Ser Ile Asn Ser Phe Leu Cys Glu Gly Ile Lys
                165                 170                 175

Gly Lys Asp Asn Glu Thr Glu Glu Ile Trp His Leu Lys Asp Asn Asp
            180                 185                 190

Val Arg Lys Glu Lys Val Lys Glu Asn Phe Lys Asn Lys Leu Ile Gln
        195                 200                 205

Ser Thr Glu Asn Tyr Asn Ser Ser Leu Lys Asn Gln Ile Glu Glu Lys
    210                 215                 220

Glu Lys Leu Ser Ser Lys Glu Phe Lys Lys Gly Ala Phe Tyr Arg Thr
225                 230                 235                 240
```

```
Ile Ile Lys Lys Leu Gln Gln Glu Arg Ile Lys Glu Leu Ser Glu Lys
                245                 250                 255

Ser Leu Thr Glu Asp Cys Glu Lys Ile Ile Lys Leu Tyr Ser Glu Leu
            260                 265                 270

Arg His Pro Leu Met His Tyr Asp Tyr Gln Tyr Phe Glu Asn Leu Phe
        275                 280                 285

Glu Asn Lys Glu Asn Ser Glu Leu Thr Lys Asn Leu Asn Leu Asp Ile
    290                 295                 300

Phe Lys Ser Leu Pro Leu Val Arg Lys Met Lys Leu Asn Asn Lys Val
305                 310                 315                 320

Asn Tyr Leu Glu Asp Asn Asp Thr Leu Phe Val Leu Gln Lys Thr Lys
                325                 330                 335

Lys Ala Lys Thr Leu Tyr Gln Ile Tyr Asp Ala Leu Cys Glu Gln Lys
            340                 345                 350

Asn Gly Phe Asn Lys Phe Ile Asn Asp Phe Phe Val Ser Asp Gly Glu
        355                 360                 365

Glu Asn Thr Val Phe Lys Gln Ile Ile Asn Glu Lys Phe Gln Ser Glu
    370                 375                 380

Met Glu Phe Leu Glu Lys Arg Ile Ser Glu Ser Glu Lys Lys Asn Glu
385                 390                 395                 400

Lys Leu Lys Lys Lys Leu Asp Ser Met Lys Ala His Phe Arg Asn Ile
                405                 410                 415

Asn Ser Glu Asp Thr Lys Glu Ala Tyr Phe Trp Asp Ile His Ser Ser
            420                 425                 430

Arg Asn Tyr Lys Thr Lys Tyr Asn Glu Arg Lys Asn Leu Val Asn Glu
        435                 440                 445
```

```
Tyr Thr Lys Leu Leu Gly Ser Ser Lys Glu Lys Lys Leu Leu Arg Glu
    450                 455                 460


Glu Ile Thr Lys Ile Asn Arg Gln Leu Leu Lys Leu Lys Gln Glu Met
465                 470                 475                 480


Glu Glu Ile Thr Lys Lys Asn Ser Leu Phe Arg Leu Glu Tyr Lys Met
                485                 490                 495


Lys Ile Ala Phe Gly Phe Leu Phe Cys Glu Phe Asp Gly Asn Ile Ser
            500                 505                 510


Lys Phe Lys Asp Glu Phe Asp Ala Ser Asn Gln Glu Lys Ile Ile Gln
        515                 520                 525


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


Glu Lys Glu Lys Phe Asn Leu Glu Lys Met Gln Lys Ile Ile Gln Lys
545                 550                 555                 560


Thr Glu Glu Glu Asp Trp Leu Leu Pro Glu Thr Lys Asn Asn Leu Phe
                565                 570                 575


Lys Phe Tyr Leu Leu Thr Tyr Leu Leu Leu Pro Tyr Glu Leu Lys Gly
            580                 585                 590


Asp Phe Leu Gly Phe Val Lys Lys His Tyr Tyr Asp Ile Lys Asn Val
        595                 600                 605


Asp Phe Met Asp Glu Asn Gln Asn Asn Ile Gln Val Ser Gln Thr Val
    610                 615                 620


Glu Lys Gln Glu Asp Tyr Phe Tyr His Lys Ile Arg Leu Phe Glu Lys
625                 630                 635                 640


Asn Thr Lys Lys Tyr Glu Ile Val Lys Tyr Ser Ile Val Pro Asn Glu
                645                 650                 655
```

```
Lys Leu Lys Gln Tyr Phe Glu Asp Leu Gly Ile Asp Ile Lys Tyr Leu
            660                 665                 670


Thr Gly Ser Val Glu Ser Gly Glu Lys Trp Leu Gly Glu Asn Leu Gly
        675                 680                 685


Ile Asp Ile Lys Tyr Leu Thr Val Glu Gln Lys Ser Glu Val Ser Glu
    690                 695                 700


Glu Lys Asn Lys Lys Val Ser Leu Lys Asn Asn Gly Met Phe Asn Lys
705                 710                 715                 720


Thr Ile Leu Leu Phe Val Phe Lys Tyr Tyr Gln Ile Ala Phe Lys Leu
                725                 730                 735


Phe Asn Asp Ile Glu Leu Tyr Ser Leu Phe Phe Leu Arg Glu Lys Ser
            740                 745                 750


Glu Lys Pro Phe Glu Val Phe Leu Glu Glu Leu Lys Asp Lys Met Ile
        755                 760                 765


Gly Lys Gln Leu Asn Phe Gly Gln Leu Leu Tyr Val Val Tyr Glu Val
    770                 775                 780


Leu Val Lys Asn Lys Asp Leu Asp Lys Ile Leu Ser Lys Lys Ile Asp
785                 790                 795                 800


Tyr Arg Lys Asp Lys Ser Phe Ser Pro Glu Ile Ala Tyr Leu Arg Asn
                805                 810                 815


Phe Leu Ser His Leu Asn Tyr Ser Lys Phe Leu Asp Asn Phe Met Lys
            820                 825                 830


Ile Asn Thr Asn Lys Ser Asp Glu Asn Lys Glu Val Leu Ile Pro Ser
        835                 840                 845


Ile Lys Ile Gln Lys Met Ile Gln Phe Ile Glu Lys Cys Asn Leu Gln
    850                 855                 860
```

```
Asn Gln Ile Asp Phe Asp Phe Asn Phe Val Asn Asp Phe Tyr Met Arg
865                 870                 875                 880


Lys Glu Lys Met Phe Phe Ile Gln Leu Lys Gln Ile Phe Pro Asp Ile
                885                 890                 895


Asn Ser Thr Glu Lys Gln Lys Lys Ser Glu Lys Glu Glu Ile Leu Arg
            900                 905                 910


Lys Arg Tyr His Leu Ile Asn Lys Lys Asn Glu Gln Ile Lys Asp Glu
        915                 920                 925


His Glu Ala Gln Ser Gln Leu Tyr Glu Lys Ile Leu Ser Leu Gln Lys
    930                 935                 940


Ile Phe Ser Cys Asp Lys Asn Asn Phe Tyr Arg Arg Leu Lys Glu Glu
945                 950                 955                 960


Lys Leu Leu Phe Leu Glu Lys Gln Gly Lys Lys Lys Ile Ser Met Lys
                965                 970                 975


Glu Ile Lys Asp Lys Ile Ala Ser Asp Ile Ser Asp Leu Leu Gly Ile
            980                 985                 990


Leu Lys Lys Glu Ile Thr Arg Asp  Ile Lys Asp Lys Leu  Thr Glu Lys
        995                 1000                1005


Phe Arg  Tyr Cys Glu Glu Lys  Leu Leu Asn Ile Ser  Phe Tyr Asn
    1010                1015                1020


His Gln  Asp Lys Lys Lys Glu  Glu Gly Ile Arg Val  Phe Leu Ile
    1025                1030                1035


Arg Asp  Lys Asn Ser Asp Asn  Phe Lys Phe Glu Ser  Ile Leu Asp
    1040                1045                1050


Asp Gly  Ser Asn Lys Ile Phe  Ile Ser Lys Asn Gly  Lys Glu Ile
    1055                1060                1065
```

```
Thr Ile  Gln Cys Cys Asp Lys  Val Leu Glu Thr Leu  Met Ile Glu
    1070                1075                1080


Lys Asn  Thr Leu Lys Ile Ser  Ser Asn Gly Lys Ile  Ile Ser Leu
    1085                1090                1095


Ile Pro  His Tyr Ser Tyr Ser  Ile Asp Val Lys Tyr
    1100                1105                1110


<210>  158
<211>  668
<212>  PRT
<213>  Fusobacterium necorphorum

<400>  158

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


Ile Val Lys Lys Asp Ile Phe Ser Gln Thr Pro Asp Asn Met Ile Arg
            20                  25                  30


Asn Tyr Lys Ile Thr Leu Lys Ile Ser Glu Lys Asn Pro Arg Val Val
        35                  40                  45


Glu Ala Glu Ile Glu Asp Leu Met Asn Ser Thr Ile Leu Lys Asp Gly
    50                  55                  60


Arg Arg Ser Ala Arg Arg Glu Lys Ser Met Thr Glu Arg Lys Leu Ile
65                  70                  75                  80


Glu Glu Lys Val Ala Glu Asn Tyr Ser Leu Leu Ala Asn Cys Pro Met
                85                  90                  95


Glu Glu Val Asp Ser Ile Lys Ile Tyr Lys Ile Lys Arg Phe Leu Thr
            100                 105                 110


Tyr Arg Ser Asn Met Leu Leu Tyr Phe Ala Ser Ile Asn Ser Phe Leu
        115                 120                 125


Cys Glu Gly Ile Lys Gly Lys Asp Asn Glu Thr Glu Glu Ile Trp His
```

```
    130                 135                 140

Leu Lys Asp Asn Asp Val Arg Lys Glu Lys Val Lys Glu Asn Phe Lys
145                 150                 155                 160

Asn Lys Leu Ile Gln Ser Thr Glu Asn Tyr Asn Ser Ser Leu Lys Asn
                165                 170                 175

Gln Ile Glu Glu Lys Glu Lys Leu Leu Arg Lys Glu Ser Lys Lys Gly
            180                 185                 190

Ala Phe Tyr Arg Thr Ile Ile Lys Lys Leu Gln Gln Glu Arg Ile Lys
        195                 200                 205

Glu Leu Ser Glu Lys Ser Leu Thr Glu Asp Cys Glu Lys Ile Ile Lys
    210                 215                 220

Leu Tyr Ser Glu Leu Arg His Pro Leu Met His Tyr Asp Tyr Gln Tyr
225                 230                 235                 240

Phe Glu Asn Leu Phe Glu Asn Lys Glu Asn Ser Glu Leu Thr Lys Asn
                245                 250                 255

Leu Asn Leu Asp Ile Phe Lys Ser Leu Pro Leu Val Arg Lys Met Lys
            260                 265                 270

Leu Asn Asn Lys Val Asn Tyr Leu Glu Asp Asn Asp Thr Leu Phe Val
        275                 280                 285

Leu Gln Lys Thr Lys Lys Ala Lys Thr Leu Tyr Gln Ile Tyr Asp Ala
    290                 295                 300

Leu Cys Glu Gln Lys Asn Gly Phe Asn Lys Phe Ile Asn Asp Phe Phe
305                 310                 315                 320

Val Ser Asp Gly Glu Glu Asn Thr Val Phe Lys Gln Ile Ile Asn Glu
                325                 330                 335

Lys Phe Gln Ser Glu Met Glu Phe Leu Glu Lys Arg Ile Ser Glu Ser
```

340 345 350

Glu Lys Lys Asn Glu Lys Leu Lys Lys Lys Phe Asp Ser Met Lys Ala
355 360 365

His Phe His Asn Ile Asn Ser Glu Asp Thr Lys Glu Ala Tyr Phe Trp
370 375 380

Asp Ile His Ser Ser Ser Asn Tyr Lys Thr Lys Tyr Asn Glu Arg Lys
385 390 395 400

Asn Leu Val Asn Glu Tyr Thr Glu Leu Leu Gly Ser Ser Lys Glu Lys
405 410 415

Lys Leu Leu Arg Glu Glu Ile Thr Gln Ile Asn Arg Lys Leu Leu Lys
420 425 430

Leu Lys Gln Glu Met Glu Glu Ile Thr Lys Lys Asn Ser Leu Phe Arg
435 440 445

Leu Glu Tyr Lys Met Lys Ile Ala Phe Gly Phe Leu Phe Cys Glu Phe
450 455 460

Asp Gly Asn Ile Ser Lys Phe Lys Asp Glu Phe Asp Ala Ser Asn Gln
465 470 475 480

Glu Lys Ile Ile Gln Tyr His Lys Asn Gly Glu Lys Tyr Leu Thr Tyr
485 490 495

Phe Leu Lys Glu Glu Glu Lys Glu Lys Phe Asn Leu Glu Lys Met Gln
500 505 510

Lys Ile Ile Gln Lys Thr Glu Glu Glu Asp Trp Leu Leu Pro Glu Thr
515 520 525

Lys Asn Asn Leu Phe Lys Phe Tyr Leu Leu Thr Tyr Leu Leu Leu Pro
530 535 540

Tyr Glu Leu Lys Gly Asp Phe Leu Gly Phe Val Lys Lys His Tyr Tyr

545 550 555 560

Asp Ile Lys Asn Val Asp Phe Met Asp Glu Asn Gln Asn Asn Ile Gln
565 570 575

Val Ser Gln Thr Val Glu Lys Gln Glu Asp Tyr Phe Tyr His Lys Ile
580 585 590

Arg Leu Phe Glu Lys Asn Thr Lys Lys Tyr Glu Ile Val Lys Tyr Ser
595 600 605

Ile Val Pro Asn Glu Lys Leu Lys Gln Tyr Phe Glu Asp Leu Gly Ile
610 615 620

Asp Ile Lys Tyr Leu Thr Gly Ser Val Glu Ser Gly Glu Lys Trp Leu
625 630 635 640

Gly Glu Asn Leu Gly Ile Asp Ile Lys Tyr Leu Thr Val Glu Gln Lys
645 650 655

Ser Glu Val Ser Glu Glu Lys Ile Lys Lys Phe Leu
660 665

<210> 159
<211> 1121
<212> PRT
<213> Fusobacterium perfoetens

<400> 159

Met Gly Lys Pro Asn Arg Ser Ser Ile Ile Lys Ile Ile Ile Ser Asn
1 5 10 15

Tyr Asp Asn Lys Gly Ile Lys Glu Val Lys Val Arg Tyr Asn Lys Gln
20 25 30

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

Phe Ile Leu Tyr Ser Ile Val Asp Lys Ala Arg Glu Lys Tyr Arg Tyr
50 55 60

```
Ser Phe Glu Ile Asp Lys Thr Asn Ile Asn Lys Asn Glu Ile Leu Ile
65                  70                  75                  80

Ile Lys Lys Asp Ile Tyr Ser Asn Lys Glu Asp Lys Val Ile Arg Lys
                85                  90                  95

Tyr Ile Leu Ser Phe Glu Val Ser Glu Lys Asn Asp Arg Thr Ile Val
            100                 105                 110

Thr Lys Ile Lys Asp Cys Leu Glu Thr Gln Lys Lys Glu Lys Phe Glu
        115                 120                 125

Arg Glu Asn Thr Arg Arg Leu Ile Ser Glu Thr Glu Arg Lys Leu Leu
    130                 135                 140

Ser Glu Glu Thr Gln Lys Thr Tyr Ser Lys Ile Ala Cys Cys Ser Pro
145                 150                 155                 160

Glu Asp Ile Asp Ser Val Lys Ile Tyr Lys Ile Lys Arg Tyr Leu Ala
                165                 170                 175

Tyr Arg Ser Asn Met Leu Leu Phe Phe Ser Leu Ile Asn Asp Ile Phe
            180                 185                 190

Val Lys Gly Val Val Lys Asp Asn Gly Glu Glu Val Gly Glu Ile Trp
        195                 200                 205

Arg Ile Ile Asp Ser Lys Glu Ile Asp Glu Lys Lys Thr Tyr Asp Leu
    210                 215                 220

Leu Val Glu Asn Phe Lys Lys Arg Met Ser Gln Glu Phe Ile Asn Tyr
225                 230                 235                 240

Lys Gln Ser Ile Glu Asn Lys Ile Glu Lys Asn Thr Asn Lys Ile Lys
                245                 250                 255

Glu Ile Glu Gln Lys Leu Lys Lys Glu Lys Tyr Lys Lys Glu Ile Asn
            260                 265                 270
```

```
Arg Leu Lys Lys Gln Leu Ile Glu Leu Asn Arg Glu Asn Asp Leu Leu
        275                 280                 285

Glu Lys Asp Lys Ile Glu Leu Ser Asp Glu Glu Ile Arg Glu Asp Ile
    290                 295                 300

Glu Lys Ile Leu Lys Ile Tyr Ser Asp Leu Arg His Lys Leu Met His
305                 310                 315                 320

Tyr Asn Tyr Gln Tyr Phe Glu Asn Leu Phe Glu Asn Lys Lys Ile Ser
                325                 330                 335

Lys Glu Lys Asn Glu Asp Val Asn Leu Thr Glu Leu Leu Asp Leu Asn
            340                 345                 350

Leu Phe Arg Tyr Leu Pro Leu Val Arg Gln Leu Lys Leu Glu Asn Lys
        355                 360                 365

Thr Asn Tyr Leu Glu Lys Glu Asp Lys Ile Thr Val Leu Gly Val Ser
    370                 375                 380

Asp Ser Ala Ile Lys Tyr Tyr Ser Tyr Tyr Asn Phe Leu Cys Glu Gln
385                 390                 395                 400

Lys Asn Gly Phe Asn Asn Phe Ile Asn Ser Phe Phe Ser Asn Asp Gly
                405                 410                 415

Glu Glu Asn Lys Ser Phe Lys Glu Lys Ile Asn Leu Ser Leu Glu Lys
            420                 425                 430

Glu Ile Glu Ile Met Glu Lys Glu Thr Asn Glu Lys Ile Lys Glu Ile
        435                 440                 445

Asn Lys Asn Glu Leu Gln Leu Met Lys Glu Gln Lys Glu Leu Gly Thr
    450                 455                 460

Ala Tyr Val Leu Asp Ile His Ser Leu Asn Asp Tyr Lys Ile Ser His
465                 470                 475                 480
```

```
Asn Glu Arg Asn Lys Asn Val Lys Leu Gln Asn Asp Ile Met Asn Gly
                485                 490                 495

Asn Arg Asp Lys Asn Ala Leu Asp Lys Ile Asn Lys Lys Leu Val Glu
            500                 505                 510

Leu Lys Ile Lys Met Asp Lys Ile Thr Lys Arg Asn Ser Ile Leu Arg
        515                 520                 525

Leu Lys Tyr Lys Leu Gln Val Ala Tyr Gly Phe Leu Met Glu Glu Tyr
    530                 535                 540

Lys Gly Asn Ile Lys Lys Phe Lys Asp Glu Phe Asp Ile Ser Lys Glu
545                 550                 555                 560

Lys Ile Lys Ser Tyr Lys Ser Lys Gly Glu Lys Tyr Leu Glu Val Lys
                565                 570                 575

Ser Glu Lys Lys Tyr Ile Thr Lys Ile Leu Asn Ser Ile Glu Asp Ile
            580                 585                 590

His Asn Ile Thr Trp Leu Lys Asn Gln Glu Glu Asn Asn Leu Phe Lys
        595                 600                 605

Phe Tyr Val Leu Thr Tyr Ile Leu Leu Pro Phe Glu Phe Arg Gly Asp
    610                 615                 620

Phe Leu Gly Phe Val Lys Lys His Tyr Tyr Asp Ile Lys Asn Val Glu
625                 630                 635                 640

Phe Leu Asp Glu Asn Asn Asp Arg Leu Thr Pro Glu Gln Leu Glu Lys
                645                 650                 655

Met Lys Asn Asp Ser Phe Phe Asn Lys Ile Arg Leu Phe Glu Lys Asn
            660                 665                 670

Ser Lys Lys Tyr Asp Ile Leu Lys Glu Ser Ile Leu Thr Ser Glu Arg
        675                 680                 685
```

```
Ile Gly Lys Tyr Phe Ser Leu Leu Asn Thr Gly Ala Lys Tyr Phe Glu
    690                 695                 700

Tyr Gly Gly Glu Glu Asn Arg Gly Ile Phe Asn Lys Asn Ile Ile Ile
705                 710                 715                 720

Pro Ile Phe Lys Tyr Tyr Gln Ile Val Leu Lys Leu Tyr Asn Asp Val
                725                 730                 735

Glu Leu Ala Met Leu Leu Thr Leu Ser Glu Ser Asp Glu Lys Asp Ile
            740                 745                 750

Asn Lys Ile Lys Glu Leu Val Thr Leu Lys Glu Lys Val Ser Pro Lys
        755                 760                 765

Lys Ile Asp Tyr Glu Lys Lys Tyr Lys Phe Ser Val Leu Leu Asp Cys
    770                 775                 780

Phe Asn Arg Ile Ile Asn Leu Gly Lys Lys Asp Phe Leu Ala Ser Glu
785                 790                 795                 800

Glu Val Lys Glu Val Ala Lys Thr Phe Thr Asn Leu Ala Tyr Leu Arg
                805                 810                 815

Asn Lys Ile Cys His Leu Asn Tyr Ser Lys Phe Ile Asp Asp Leu Leu
            820                 825                 830

Thr Ile Asp Thr Asn Lys Ser Thr Thr Asp Ser Glu Gly Lys Leu Leu
        835                 840                 845

Ile Asn Asp Arg Ile Arg Lys Leu Ile Lys Phe Ile Arg Glu Asn Asn
    850                 855                 860

Gln Lys Met Asn Ile Ser Ile Asp Tyr Asn Tyr Ile Asn Asp Tyr Tyr
865                 870                 875                 880

Met Lys Lys Glu Lys Phe Ile Phe Gly Gln Arg Lys Gln Ala Lys Thr
                885                 890                 895
```

```
Ile Ile Asp Ser Gly Lys Lys Ala Asn Lys Arg Asn Lys Ala Glu Glu
            900                 905                 910

Leu Leu Lys Met Tyr Arg Val Lys Lys Glu Asn Ile Asn Leu Ile Tyr
        915                 920                 925

Glu Leu Ser Lys Lys Leu Asn Glu Leu Thr Lys Ser Glu Leu Phe Leu
    930                 935                 940

Leu Asp Lys Lys Leu Leu Lys Asp Ile Asp Phe Thr Asp Val Lys Ile
945                 950                 955                 960

Lys Asn Lys Ser Phe Phe Glu Leu Lys Asn Asp Val Lys Glu Val Ala
                965                 970                 975

Asn Ile Lys Gln Ala Leu Gln Lys His Ser Ser Glu Leu Ile Gly Ile
            980                 985                 990

Tyr Lys Lys Glu Val Ile Met Ala  Ile Lys Arg Ser Ile  Val Ser Lys
        995                 1000                1005

Leu Ile  Tyr Asp Glu Glu Lys  Val Leu Ser Ile Ile  Ile Tyr Asp
    1010                1015                1020

Lys Thr  Asn Lys Lys Tyr Glu  Asp Phe Leu Leu Glu  Ile Arg Arg
    1025                1030                1035

Glu Arg  Asp Ile Asn Lys Phe  Gln Phe Leu Ile Asp  Glu Lys Lys
    1040                1045                1050

Glu Lys  Leu Gly Tyr Glu Lys  Ile Ile Glu Thr Lys  Glu Lys Lys
    1055                1060                1065

Lys Val  Val Val Lys Ile Gln  Asn Asn Ser Glu Leu  Val Ser Glu
    1070                1075                1080

Pro Arg  Ile Ile Lys Asn Lys  Asp Lys Lys Lys Ala  Lys Thr Pro
    1085                1090                1095
```

```
Glu Glu  Ile Ser Lys Leu Gly  Ile Leu Asp Leu Thr  Asn His Tyr
    1100                1105                1110


Cys Phe  Asn Leu Lys Ile Thr  Leu
    1115                1120


<210>  160
<211>  897
<212>  PRT
<213>  Fusobacterium ulcerans

<400>  160

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


Asn Ile Leu Val Ala Gln Ile Lys Glu Tyr Phe Thr Lys Glu Ile Glu
            20                  25                  30


Asn Tyr Asn Asn Arg Ile Asp Asn Ile Ile Asp Lys Lys Glu Leu Leu
        35                  40                  45


Lys Tyr Ser Glu Lys Lys Glu Glu Ser Glu Lys Asn Lys Lys Leu Glu
    50                  55                  60


Glu Leu Asn Lys Leu Lys Ser Gln Lys Leu Lys Ile Leu Thr Asp Glu
65                  70                  75                  80


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


Arg His Ser Leu Met His Tyr Glu Tyr Lys Tyr Phe Glu Asn Leu Phe
            100                 105                 110


Glu Asn Lys Lys Asn Glu Glu Leu Ala Glu Leu Leu Asn Leu Asn Leu
        115                 120                 125


Phe Lys Asn Leu Thr Leu Leu Arg Gln Met Lys Ile Glu Asn Lys Thr
    130                 135                 140
```

```
Asn Tyr Leu Glu Gly Arg Glu Glu Phe Asn Ile Ile Gly Lys Asn Ile
145                 150                 155                 160

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

Asn Gly Phe Asn Asn Phe Ile Asn Ser Phe Phe Val Gln Asp Gly Thr
            180                 185                 190

Glu Asn Leu Glu Phe Lys Lys Leu Ile Asp Glu His Phe Val Asn Ala
        195                 200                 205

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

Glu Leu Glu Lys Met Glu Gln His Tyr Gln Arg Leu Asn Cys Ala Tyr
225                 230                 235                 240

Val Trp Asp Ile His Thr Ser Thr Thr Tyr Lys Lys Leu Tyr Asn Lys
                245                 250                 255

Arg Lys Ser Leu Ile Glu Glu Tyr Asn Lys Gln Ile Asn Glu Ile Lys
            260                 265                 270

Asp Lys Glu Val Ile Thr Ala Ile Asn Val Glu Leu Leu Arg Ile Lys
        275                 280                 285

Lys Glu Met Glu Glu Ile Thr Lys Ser Asn Ser Leu Phe Arg Leu Lys
    290                 295                 300

Tyr Lys Met Gln Ile Ala Tyr Ala Phe Leu Glu Ile Glu Phe Gly Gly
305                 310                 315                 320

Asn Ile Ala Lys Phe Lys Asp Glu Phe Asp Cys Ser Lys Met Glu Glu
                325                 330                 335

Val Gln Lys Tyr Leu Lys Lys Gly Val Lys Tyr Leu Lys Tyr Tyr Lys
            340                 345                 350
```

```
Asp Lys Glu Ala Gln Lys Asn Tyr Glu Phe Pro Phe Glu Glu Ile Phe
        355                 360                 365

Glu Asn Lys Asp Thr His Asn Glu Glu Trp Leu Glu Asn Thr Ser Glu
    370                 375                 380

Asn Asn Leu Phe Lys Phe Tyr Ile Leu Thr Tyr Leu Leu Leu Pro Met
385                 390                 395                 400

Glu Phe Lys Gly Asp Phe Leu Gly Val Val Lys Lys His Tyr Tyr Asp
                405                 410                 415

Ile Lys Asn Val Asp Phe Thr Asp Glu Ser Glu Lys Glu Leu Ser Gln
            420                 425                 430

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

Leu Phe Glu Lys Asn Thr Lys Arg Tyr Glu Ile Ile Lys Tyr Ser Ile
    450                 455                 460

Leu Thr Ser Asp Glu Ile Lys Arg Tyr Phe Arg Leu Leu Glu Leu Asp
465                 470                 475                 480

Val Pro Tyr Phe Glu Tyr Glu Lys Gly Thr Asp Glu Ile Gly Ile Phe
                485                 490                 495

Asn Lys Asn Ile Ile Leu Thr Ile Phe Lys Tyr Tyr Gln Ile Ile Phe
            500                 505                 510

Arg Leu Tyr Asn Asp Leu Glu Ile His Gly Leu Phe Asn Ile Ser Ser
        515                 520                 525

Asp Leu Asp Lys Ile Leu Arg Asp Leu Lys Ser Tyr Gly Asn Lys Asn
    530                 535                 540

Ile Asn Phe Arg Glu Phe Leu Tyr Val Ile Lys Gln Asn Asn Asn Ser
545                 550                 555                 560
```

```
Ser Thr Glu Glu Glu Tyr Arg Lys Ile Trp Glu Asn Leu Glu Ala Lys
                565                 570                 575

Tyr Leu Arg Leu His Leu Leu Thr Pro Glu Lys Glu Glu Ile Lys Thr
            580                 585                 590

Lys Thr Lys Glu Glu Leu Glu Lys Leu Asn Glu Ile Ser Asn Leu Arg
        595                 600                 605

Asn Gly Ile Cys His Leu Asn Tyr Lys Glu Ile Ile Glu Glu Ile Leu
    610                 615                 620

Lys Thr Glu Ile Ser Glu Lys Asn Lys Glu Ala Thr Leu Asn Glu Lys
625                 630                 635                 640

Ile Arg Lys Val Ile Asn Phe Ile Lys Glu Asn Glu Leu Asp Lys Val
                645                 650                 655

Glu Leu Gly Phe Asn Phe Ile Asn Asp Phe Phe Met Lys Lys Glu Gln
            660                 665                 670

Phe Met Phe Gly Gln Ile Lys Gln Val Lys Glu Gly Asn Ser Asp Ser
        675                 680                 685

Ile Thr Thr Glu Arg Glu Arg Lys Glu Lys Asn Asn Lys Lys Leu Lys
    690                 695                 700

Glu Thr Tyr Glu Leu Asn Cys Asp Asn Leu Ser Glu Phe Tyr Glu Thr
705                 710                 715                 720

Ser Asn Asn Leu Arg Glu Arg Ala Asn Ser Ser Ser Leu Leu Glu Asp
                725                 730                 735

Ser Ala Phe Leu Lys Lys Ile Gly Leu Tyr Lys Val Lys Asn Asn Lys
            740                 745                 750

Val Asn Ser Lys Val Lys Asp Glu Glu Lys Arg Ile Glu Asn Ile Lys
        755                 760                 765
```

```
Arg Lys Leu Leu Lys Asp Ser Ser Asp Ile Met Gly Met Tyr Lys Ala
    770                 775                 780


Glu Val Val Lys Lys Leu Lys Glu Lys Leu Ile Leu Ile Phe Lys His
785                 790                 795                 800


Asp Glu Glu Lys Arg Ile Tyr Val Thr Val Tyr Asp Thr Ser Lys Ala
                805                 810                 815


Val Pro Glu Asn Ile Ser Lys Glu Ile Leu Val Lys Arg Asn Asn Ser
            820                 825                 830


Lys Glu Glu Tyr Phe Phe Glu Asp Asn Asn Lys Lys Tyr Val Thr Glu
        835                 840                 845


Tyr Tyr Thr Leu Glu Ile Thr Glu Thr Asn Glu Leu Lys Val Ile Pro
    850                 855                 860


Ala Lys Lys Leu Glu Gly Lys Glu Phe Lys Thr Glu Lys Asn Lys Glu
865                 870                 875                 880


Asn Lys Leu Met Leu Asn Asn His Tyr Cys Phe Asn Val Lys Ile Ile
                885                 890                 895


Tyr



<210>  161
<211>  1111
<212>  PRT
<213>  Anaerosalibacter sp.

<400>  161

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


Val Arg Val Ile Ile Ser Asn Phe Asp Asp Lys Gln Val Lys Glu Ile
            20                  25                  30
```

```
Lys Val Leu Tyr Thr Lys Gln Gly Gly Ile Asp Val Ile Lys Phe Lys
        35                  40                  45

Ser Thr Glu Lys Asp Glu Lys Gly Arg Met Lys Phe Asn Phe Asp Cys
    50                  55                  60

Ala Tyr Asn Arg Leu Glu Glu Glu Glu Phe Asn Ser Phe Gly Gly Lys
65                  70                  75                  80

Gly Lys Gln Ser Phe Phe Val Thr Thr Asn Glu Asp Leu Thr Glu Leu
                85                  90                  95

His Val Thr Lys Arg His Lys Thr Thr Gly Glu Ile Ile Lys Asp Tyr
            100                 105                 110

Thr Ile Gln Gly Lys Tyr Thr Pro Ile Lys Gln Asp Arg Thr Lys Val
        115                 120                 125

Thr Val Ser Ile Thr Asp Asn Lys Asp His Phe Asp Ser Asn Asp Leu
    130                 135                 140

Gly Asp Lys Ile Arg Leu Ser Arg Ser Leu Thr Gln Tyr Thr Asn Arg
145                 150                 155                 160

Ile Leu Leu Asp Ala Asp Val Met Lys Asn Tyr Arg Glu Ile Val Cys
                165                 170                 175

Ser Asp Ser Glu Lys Val Asp Glu Thr Ile Asn Ile Asp Ser Gln Glu
            180                 185                 190

Ile Tyr Lys Ile Asn Arg Phe Leu Ser Tyr Arg Ser Asn Met Ile Ile
        195                 200                 205

Tyr Tyr Gln Met Ile Asn Asn Phe Leu Leu His Tyr Asp Gly Glu Glu
    210                 215                 220

Asp Lys Gly Gly Asn Asp Ser Ile Asn Leu Ile Asn Glu Ile Trp Lys
225                 230                 235                 240
```

```
Tyr Glu Asn Lys Lys Asn Asp Glu Lys Glu Lys Ile Ile Glu Arg Ser
                245                 250                 255

Tyr Lys Ser Ile Glu Lys Ser Ile Asn Gln Tyr Ile Leu Asn His Asn
            260                 265                 270

Thr Glu Val Glu Ser Gly Asp Lys Glu Lys Lys Ile Asp Ile Ser Glu
        275                 280                 285

Glu Arg Ile Lys Glu Asp Leu Lys Lys Thr Phe Ile Leu Phe Ser Arg
    290                 295                 300

Leu Arg His Tyr Met Val His Tyr Asn Tyr Lys Phe Tyr Glu Asn Leu
305                 310                 315                 320

Tyr Ser Gly Lys Asn Phe Ile Ile Tyr Asn Lys Asp Lys Ser Lys Ser
                325                 330                 335

Arg Arg Phe Ser Glu Leu Leu Asp Leu Asn Ile Phe Lys Glu Leu Ser
            340                 345                 350

Lys Ile Lys Leu Val Lys Asn Arg Ala Val Ser Asn Tyr Leu Asp Lys
        355                 360                 365

Lys Thr Thr Ile His Val Leu Asn Lys Asn Ile Asn Ala Ile Lys Leu
    370                 375                 380

Leu Asp Ile Tyr Arg Asp Ile Cys Glu Thr Lys Asn Gly Phe Asn Asn
385                 390                 395                 400

Phe Ile Asn Asn Met Met Thr Ile Ser Gly Glu Glu Asp Lys Glu Tyr
                405                 410                 415

Lys Glu Met Val Thr Lys His Phe Asn Glu Asn Met Asn Lys Leu Ser
            420                 425                 430

Ile Tyr Leu Glu Asn Phe Lys Lys His Ser Asp Phe Lys Thr Asn Asn
        435                 440                 445
```

```
Lys Lys Lys Glu Thr Tyr Asn Leu Leu Lys Gln Glu Leu Asp Glu Gln
    450                 455                 460

Lys Lys Leu Arg Leu Trp Phe Asn Ala Pro Tyr Val Tyr Asp Ile His
465                 470                 475                 480

Ser Ser Lys Lys Tyr Lys Glu Leu Tyr Val Glu Arg Lys Lys Tyr Val
                485                 490                 495

Asp Ile His Ser Lys Leu Ile Glu Ala Gly Ile Asn Asn Asp Asn Lys
            500                 505                 510

Lys Lys Leu Asn Glu Ile Asn Val Lys Leu Cys Glu Leu Asn Thr Glu
        515                 520                 525

Met Lys Glu Met Thr Lys Leu Asn Ser Lys Tyr Arg Leu Gln Tyr Lys
    530                 535                 540

Leu Gln Leu Ala Phe Gly Phe Ile Leu Glu Glu Phe Asn Leu Asp Ile
545                 550                 555                 560

Asp Lys Phe Val Ser Ala Phe Asp Lys Asp Asn Asn Leu Thr Ile Ser
                565                 570                 575

Lys Phe Met Glu Lys Arg Glu Thr Tyr Leu Ser Lys Ser Leu Asp Arg
            580                 585                 590

Arg Asp Asn Arg Phe Lys Lys Leu Ile Lys Asp Tyr Lys Phe Arg Asp
        595                 600                 605

Thr Glu Asp Ile Phe Cys Ser Asp Arg Glu Asn Asn Leu Val Lys Leu
    610                 615                 620

Tyr Ile Leu Met Tyr Ile Leu Leu Pro Val Glu Ile Arg Gly Asp Phe
625                 630                 635                 640

Leu Gly Phe Val Lys Lys Asn Tyr Tyr Asp Leu Lys His Val Asp Phe
                645                 650                 655
```

```
Ile Asp Lys Arg Asn Asn Asp Asn Lys Asp Thr Phe Phe His Asp Leu
            660                 665                 670

Arg Leu Phe Glu Lys Asn Val Lys Arg Leu Glu Val Thr Ser Tyr Ser
        675                 680                 685

Leu Ser Asp Gly Phe Leu Gly Lys Lys Ser Arg Glu Lys Phe Gly Lys
    690                 695                 700

Glu Leu Glu Lys Phe Ile Tyr Lys Asn Val Ser Ile Ala Leu Pro Thr
705                 710                 715                 720

Asn Ile Asp Ile Lys Glu Phe Asn Lys Ser Leu Val Leu Pro Met Met
                725                 730                 735

Lys Asn Tyr Gln Ile Ile Phe Lys Leu Leu Asn Asp Ile Glu Ile Ser
            740                 745                 750

Ala Leu Phe Leu Ile Ala Lys Lys Glu Gly Asn Glu Gly Ser Ile Thr
        755                 760                 765

Phe Lys Lys Val Ile Asp Lys Val Arg Lys Glu Asp Met Asn Gly Asn
    770                 775                 780

Ile Asn Phe Ser Gln Val Met Lys Met Ala Leu Asn Glu Lys Val Asn
785                 790                 795                 800

Cys Gln Ile Arg Asn Ser Ile Ala His Ile Asn Met Lys Gln Leu Tyr
                805                 810                 815

Ile Glu Pro Leu Asn Ile Tyr Ile Asn Asn Asn Gln Asn Lys Lys Thr
            820                 825                 830

Ile Ser Glu Gln Met Glu Glu Ile Ile Asp Ile Cys Ile Thr Lys Gly
        835                 840                 845

Leu Thr Gly Lys Glu Leu Asn Lys Asn Ile Ile Asn Asp Tyr Tyr Met
    850                 855                 860
```

```
Lys Lys Glu Lys Leu Val Phe Asn Leu Lys Leu Arg Lys Arg Asn Asn
865                 870                 875                 880


Leu Val Ser Ile Asp Ala Gln Gln Lys Asn Met Lys Glu Lys Ser Ile
                885                 890                 895


Leu Asn Lys Tyr Asp Leu Asn Tyr Lys Asp Glu Asn Leu Asn Ile Lys
            900                 905                 910


Glu Ile Ile Leu Lys Val Asn Asp Leu Asn Asn Lys Gln Lys Leu Leu
        915                 920                 925


Lys Glu Thr Thr Glu Gly Glu Ser Asn Tyr Lys Asn Ala Leu Ser Lys
    930                 935                 940


Asp Ile Leu Leu Leu Asn Gly Ile Ile Arg Lys Asn Ile Asn Phe Lys
945                 950                 955                 960


Ile Lys Glu Met Ile Leu Gly Ile Ile Gln Gln Asn Glu Tyr Arg Tyr
                965                 970                 975


Val Asn Ile Asn Ile Tyr Asp Lys Ile Arg Lys Glu Asp His Asn Ile
            980                 985                 990


Asp Leu Lys Ile Asn Asn Lys Tyr  Ile Glu Ile Ser Cys  Tyr Glu Asn
        995                 1000                1005


Lys Ser  Asn Glu Ser Thr Asp  Glu Arg Ile Asn Phe  Lys Ile Lys
    1010                1015                1020


Tyr Met  Asp Leu Lys Val Lys  Asn Glu Leu Leu Val  Pro Ser Cys
    1025                1030                1035


Tyr Glu  Asp Ile Tyr Ile Lys  Lys Lys Ile Asp Leu  Glu Ile Arg
    1040                1045                1050


Tyr Ile  Glu Asn Cys Lys Val  Val Tyr Ile Asp Ile  Tyr Tyr Lys
    1055                1060                1065
```

Lys Tyr Asn Ile Asn Leu Glu Phe Asp Gly Lys Thr Leu Phe Val
1070 1075 1080

Lys Phe Asn Lys Asp Val Lys Lys Asn Asn Gln Lys Val Asn Leu
1085 1090 1095

Glu Ser Asn Tyr Ile Gln Asn Ile Lys Phe Ile Val Ser
1100 1105 1110

<210> 162
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 162
catagaatgt tctaaaccat cctgcggcct ctactctgca ttcaa 45

<210> 163
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 163
catagaatgt tcgaaaccat cctgcggcct ctactctgca ttcaa 45

<210> 164
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 164
catagaatgg tcgaaaccat cctgcggcct ctactctgca ttcaa 45

<210> 165
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 165
catagaatgt tcaaaaccat cctgcggcct ctactctgca ttcaa 45

<210> 166
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 166
catagaatga tcaaaaccat cctgcggcct ctactctgca ttcaa 45

<210> 167
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 167
ggtttctcca tcaattacca cttgcttcct gtaggaatcc tctatt 46

<210> 168
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 168
ggtttctcca tcaatgacca cttgcttcct gtaggaatcc tctatt 46

<210> 169
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 169
ggtttctcca tcaaggacca cttgcttcct gtaggaatcc tctatt 46

<210> 170
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 170
ggtttctcca tcaataacca cttgcttcct gtaggaatcc tctatt 46

<210> 171
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 171
ggtttctcca tcaaaaacca cttgcttcct gtaggaatcc tctatt 46

<210> 172
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 172
gcctttctct cctgtagcca aggccacaaa attatccact gttttt 46

<210> 173
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 173
gcctttctct cctggagcca aggccacaaa attatccact gttttt 46

<210> 174
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 174
gcctttctct cctgaagcca aggccacaaa attatccact gttttt 46

<210> 175
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 175
gcatagaatg ttctaaacca tcctgcggcc tctactctgc attcaa 46

<210> 176
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 176
gcatagaatg ttctaaacga tcctgcggcc tctactctgc attcaa 46

<210> 177
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 177
gcatagaatg ttctaaacta tcctgcggcc tctactctgc attcaa 46

<210> 178
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 178
gcatagaatg ttctaaacaa tcctgcggcc tctactctgc attcaa 46

<210> 179
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 179
gcatcctgcg gcctctactc tgcattcaat t 31

<210> 180
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 180
gaccatcctg cggcctctac tctgcattca a 31

<210> 181
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 181
gaaaccatcc tgcggcctct actctgcatt c 31

<210> 182
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 182
gctaaaccat cctgcggcct ctactctgca t 31

<210> 183
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 183
gttctaaacc atcctgcggc ctctactctg c 31

<210> 184
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 184
gtgttctaaa ccatcctgcg gcctctactc t 31

<210> 185
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 185
gaatgttcta aaccatcctg cggcctctac t 31

<210> 186
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 186
gagaatgttc taaaccatcc tgcggcctct a 31

<210> 187
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 187
gatagaatgt tctaaaccat cctgcggcct c 31

<210> 188
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 188
gccatagaat gttctaaacc atcctgcggc c 31

<210> 189
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 189
gttccataga atgttctaaa ccatcctgcg g 31

<210> 190
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 190
gctttccata gaatgttcta aaccatcctg c 31

<210> 191
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 191
gctctttcca tagaatgttc taaaccatcc t 31

<210> 192
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 192
gatctctttc catagaatgt tctaaaccat c 31

<210> 193
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 193
ggaatctctt tccatagaat gttctaaacc a 31

<210> 194
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 194
gcatcctgcg gcctctactc tgcattcaat tacatactga cacattcggc a 51

<210> 195
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 195
gaccatcctg cggcctctac tctgcattca attacatact gacacattcg g 51

<210> 196
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 196
gaaaccatcc tgcggcctct actctgcatt caattacata ctgacacatt c 51

<210> 197
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 197
gctaaaccat cctgcggcct ctactctgca ttcaattaca tactgacaca t 51

<210> 198
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 198
gttctaaacc atcctgcggc ctctactctg cattcaatta catactgaca c 51

<210> 199
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 199
gtgttctaaa ccatcctgcg gcctctactc tgcattcaat tacatactga c 51

<210> 200
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 200
gaatgttcta aaccatcctg cggcctctac tctgcattca attacatact g 51

<210> 201
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 201
gagaatgttc taaaccatcc tgcggcctct actctgcatt caattacata c 51

<210> 202
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 202
gatagaatgt tctaaaccat cctgcggcct ctactctgca ttcaattaca t 51

<210> 203
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 203
gccatagaat gttctaaacc atcctgcggc ctctactctg cattcaatta c 51

<210> 204
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 204
gttccataga atgttctaaa ccatcctgcg gcctctactc tgcattcaat t 51

<210> 205
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 205
gctttccata gaatgttcta aaccatcctg cggcctctac tctgcattca a 51

<210> 206
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 206
gctctttcca tagaatgttc taaaccatcc tgcggcctct actctgcatt c 51

<210> 207
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 207
gatctctttc catagaatgt tctaaaccat cctgcggcct ctactctgca t 51

<210> 208
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 208
ggaatctctt tccatagaat gttctaaacc atcctgcggc ctctactctg c 51

<210> 209
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 209
gtggaatctc tttccataga atgttctaaa ccatcctgcg gcctctactc t 51

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

<220>
<223> Synthetic or Unknown

<400> 210
gactggaatc tctttccata gaatgttcta aaccatcctg cggcctctac t 51

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

<220>
<223> Synthetic or Unknown

<400> 211
ggaactggaa tctctttcca tagaatgttc taaaccatcc tgcggcctct a 51

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

<220>
<223> Synthetic or Unknown

<400> 212
gtggaactgg aatctctttc catagaatgt tctaaaccat cctgcggcct c 51

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

<220>
<223> Synthetic or Unknown

<400> 213
gcctggaact ggaatctctt tccatagaat gttctaaacc atcctgcggc c 51

<210> 214
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 214
gttcctggaa ctggaatctc tttccataga atgttctaaa ccatcctgcg g 51

<210> 215
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 215
gggttcctgg aactggaatc tctttccata gaatgttcta aaccatcctg c 51

<210> 216
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 216
gcaggttcct ggaactggaa tctctttcca tagaatgttc taaaccatcc t 51

<210> 217
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 217
gaccaggttc ctggaactgg aatctctttc catagaatgt tctaaaccat c 51

<210> 218
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 218
ggtaccaggt tcctggaact ggaatctctt tccatagaat gttctaaacc a 51

<210> 219
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 219
gcatcctgcg gcctctactc tgcattcaat tacatactga cacattcggc aacatgtttt 60

tcctggttta t 71

<210> 220
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 220
gaccatcctg cggcctctac tctgcattca attacatact gacacattcg gcaacatgtt 60

tttcctggtt t 71

<210> 221
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 221
gaaaccatcc tgcggcctct actctgcatt caattacata ctgacacatt cggcaacatg 60

tttttcctgg t 71

<210> 222
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 222
gctaaaccat cctgcggcct ctactctgca ttcaattaca tactgacaca ttcggcaaca 60

tgtttttcct g 71

<210> 223
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 223
gttctaaacc atcctgcggc ctctactctg cattcaatta catactgaca cattcggcaa 60

catgtttttc c 71

<210> 224
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 224
gtgttctaaa ccatcctgcg gcctctactc tgcattcaat tacatactga cacattcggc 60

aacatgtttt t 71

<210> 225
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 225
gaatgttcta aaccatcctg cggcctctac tctgcattca attacatact gacacattcg 60

gcaacatgtt t 71

<210> 226
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 226
gagaatgttc taaaccatcc tgcggcctct actctgcatt caattacata ctgacacatt 60

cggcaacatg t 71

<210> 227
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 227
gatagaatgt tctaaaccat cctgcggcct ctactctgca ttcaattaca tactgacaca 60

ttcggcaaca t 71

<210> 228
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 228
gccatagaat gttctaaacc atcctgcggc ctctactctg cattcaatta catactgaca 60

cattcggcaa c 71

<210> 229
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 229
gttccataga atgttctaaa ccatcctgcg gcctctactc tgcattcaat tacatactga 60

cacattcggc a 71

<210> 230
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 230
gctttccata gaatgttcta aaccatcctg cggcctctac tctgcattca attacatact 60

gacacattcg g 71

<210> 231
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 231
gctctttcca tagaatgttc taaaccatcc tgcggcctct actctgcatt caattacata 60

ctgacacatt c 71

<210> 232
<211> 71
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 232
gatctctttc catagaatgt tctaaaccat cctgcggcct ctactctgca ttcaattaca 60

tactgacaca t 71

<210> 233
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 233
ggaatctctt tccatagaat gttctaaacc atcctgcggc ctctactctg cattcaatta 60

catactgaca c 71

<210> 234
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 234
gtggaatctc tttccataga atgttctaaa ccatcctgcg gcctctactc tgcattcaat 60

tacatactga c 71

<210> 235
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 235
gactggaatc tctttccata gaatgttcta aaccatcctg cggcctctac tctgcattca 60

attacatact g 71

<210> 236
<211> 71

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 236
ggaactggaa tctctttcca tagaatgttc taaaccatcc tgcggcctct actctgcatt 60

caattacata c 71

<210> 237
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 237
gtggaactgg aatctctttc catagaatgt tctaaaccat cctgcggcct ctactctgca 60

ttcaattaca t 71

<210> 238
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 238
gcctggaact ggaatctctt tccatagaat gttctaaacc atcctgcggc ctctactctg 60

cattcaatta c 71

<210> 239
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 239
gttcctggaa ctggaatctc tttccataga atgttctaaa ccatcctgcg gcctctactc 60

tgcattcaat t 71

<210> 240
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 240
gggttcctgg aactggaatc tctttccata gaatgttcta aaccatcctg cggcctctac 60

tctgcattca a 71

<210> 241
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 241
gcaggttcct ggaactggaa tctctttcca tagaatgttc taaaccatcc tgcggcctct 60

actctgcatt c 71

<210> 242
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 242
gaccaggttc ctggaactgg aatctctttc catagaatgt tctaaaccat cctgcggcct 60

ctactctgca t 71

<210> 243
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 243
ggtaccaggt tcctggaact ggaatctctt tccatagaat gttctaaacc atcctgcggc 60

ctctactctg c 71

<210> 244
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 244
gatgtaccag gttcctggaa ctggaatctc tttccataga atgttctaaa ccatcctgcg 60

gcctctactc t 71

<210> 245
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 245
ggtatgtacc aggttcctgg aactggaatc tctttccata gaatgttcta aaccatcctg 60

cggcctctac t 71

<210> 246
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 246
gacgtatgta ccaggttcct ggaactggaa tctctttcca tagaatgttc taaaccatcc 60

tgcggcctct a 71

<210> 247
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 247
gacacgtatg taccaggttc ctggaactgg aatctctttc catagaatgt tctaaaccat 60

cctgcggcct c 71

<210> 248
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 248
gcaacacgta tgtaccaggt tcctggaact ggaatctctt tccatagaat gttctaaacc 60

atcctgcggc c 71

<210> 249
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 249
gcccaacacg tatgtaccag gttcctggaa ctggaatctc tttccataga atgttctaaa 60

ccatcctgcg g 71

<210> 250
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 250
ggacccaaca cgtatgtacc aggttcctgg aactggaatc tctttccata gaatgttcta 60

aaccatcctg c 71

<210> 251
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 251
gttgacccaa cacgtatgta ccaggttcct ggaactggaa tctctttcca tagaatgttc 60

taaaccatcc t 71

<210> 252
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 252
gccttgaccc aacacgtatg taccaggttc ctggaactgg aatctctttc catagaatgt 60

tctaaaccat c 71

<210> 253
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 253
gttccttgac ccaacacgta tgtaccaggt tcctggaact ggaatctctt tccatagaat 60

gttctaaacc a 71

<210> 254
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 254
gcatcctgcg gcctctactc tgcattcaat tacatactga cacattcggc aacatgtttt 60

tcctggttta ttttcacaca gtcca 85

<210> 255
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 255
gaccatcctg cggcctctac tctgcattca attacatact gacacattcg gcaacatgtt 60

tttcctggtt tattttcaca cagtc 85

<210> 256
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 256
gaaaccatcc tgcggcctct actctgcatt caattacata ctgacacatt cggcaacatg 60

tttttcctgg tttattttca cacag 85

<210> 257
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 257
gctaaaccat cctgcggcct ctactctgca ttcaattaca tactgacaca ttcggcaaca 60

tgtttttcct ggtttatttt cacac 85

<210> 258
<211> 85
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 258
gttctaaacc atcctgcggc ctctactctg cattcaatta catactgaca cattcggcaa 60

catgtttttc ctggtttatt ttcac 85

<210> 259
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 259
gtgttctaaa ccatcctgcg gcctctactc tgcattcaat tacatactga cacattcggc 60

aacatgtttt tcctggttta ttttc 85

<210> 260
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 260
gaatgttcta aaccatcctg cggcctctac tctgcattca attacatact gacacattcg 60

gcaacatgtt tttcctggtt tattt 85

<210> 261
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 261
gagaatgttc taaaccatcc tgcggcctct actctgcatt caattacata ctgacacatt 60

cggcaacatg tttttcctgg tttat 85

<210> 262
<211> 85

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 262
gatagaatgt tctaaaccat cctgcggcct ctactctgca ttcaattaca tactgacaca 60

ttcggcaaca tgtttttcct ggttt 85

<210> 263
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 263
gccatagaat gttctaaacc atcctgcggc ctctactctg cattcaatta catactgaca 60

cattcggcaa catgtttttc ctggt 85

<210> 264
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 264
gttccataga atgttctaaa ccatcctgcg gcctctactc tgcattcaat tacatactga 60

cacattcggc aacatgtttt tcctg 85

<210> 265
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 265
gctttccata gaatgttcta aaccatcctg cggcctctac tctgcattca attacatact 60

gacacattcg gcaacatgtt tttcc 85

<210> 266
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 266
gctctttcca tagaatgttc taaaccatcc tgcggcctct actctgcatt caattacata 60

ctgacacatt cggcaacatg ttttt 85

<210> 267
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 267
gatctctttc catagaatgt tctaaaccat cctgcggcct ctactctgca ttcaattaca 60

tactgacaca ttcggcaaca tgttt 85

<210> 268
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 268
ggaatctctt tccatagaat gttctaaacc atcctgcggc ctctactctg cattcaatta 60

catactgaca cattcggcaa catgt 85

<210> 269
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 269
gtggaatctc tttccataga atgttctaaa ccatcctgcg gcctctactc tgcattcaat 60

tacatactga cacattcggc aacat 85

<210> 270
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 270
gactggaatc tctttccata gaatgttcta aaccatcctg cggcctctac tctgcattca 60

attacatact gacacattcg gcaac 85

<210> 271
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 271
ggaactggaa tctctttcca tagaatgttc taaaccatcc tgcggcctct actctgcatt 60

caattacata ctgacacatt cggca 85

<210> 272
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 272
gtggaactgg aatctctttc catagaatgt tctaaaccat cctgcggcct ctactctgca 60

ttcaattaca tactgacaca ttcgg 85

<210> 273
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 273
gcctggaact ggaatctctt tccatagaat gttctaaacc atcctgcggc ctctactctg 60

cattcaatta catactgaca cattc 85

<210> 274
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 274
gttcctggaa ctggaatctc tttccataga atgttctaaa ccatcctgcg gcctctactc 60

tgcattcaat tacatactga cacat 85

<210> 275
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 275
gggttcctgg aactggaatc tctttccata gaatgttcta aaccatcctg cggcctctac 60

tctgcattca attacatact gacac 85

<210> 276
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 276
gcaggttcct ggaactggaa tctctttcca tagaatgttc taaaccatcc tgcggcctct 60

actctgcatt caattacata ctgac 85

<210> 277
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 277
gaccaggttc ctggaactgg aatctctttc catagaatgt tctaaaccat cctgcggcct 60

ctactctgca ttcaattaca tactg 85

<210> 278
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 278
ggtaccaggt tcctggaact ggaatctctt tccatagaat gttctaaacc atcctgcggc 60

ctctactctg cattcaatta catac 85

<210> 279
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 279
gatgtaccag gttcctggaa ctggaatctc tttccataga atgttctaaa ccatcctgcg 60

gcctctactc tgcattcaat tacat 85

<210> 280
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 280
ggtatgtacc aggttcctgg aactggaatc tctttccata gaatgttcta aaccatcctg 60

cggcctctac tctgcattca attac 85

<210> 281
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 281
gacgtatgta ccaggttcct ggaactggaa tctctttcca tagaatgttc taaaccatcc 60

tgcggcctct actctgcatt caatt 85

<210> 282
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 282
gacacgtatg taccaggttc ctggaactgg aatctctttc catagaatgt tctaaaccat 60

cctgcggcct ctactctgca ttcaa 85

<210> 283
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 283
gcaacacgta tgtaccaggt tcctggaact ggaatctctt tccatagaat gttctaaacc 60

atcctgcggc ctctactctg cattc 85

<210> 284
<211> 85
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 284
gcccaacacg tatgtaccag gttcctggaa ctggaatctc tttccataga atgttctaaa 60

ccatcctgcg gcctctactc tgcat 85

<210> 285
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 285
ggacccaaca cgtatgtacc aggttcctgg aactggaatc tctttccata gaatgttcta 60

aaccatcctg cggcctctac tctgc 85

<210> 286
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 286
gttgacccaa cacgtatgta ccaggttcct ggaactggaa tctctttcca tagaatgttc 60

taaaccatcc tgcggcctct actct 85

<210> 287
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 287
gccttgaccc aacacgtatg taccaggttc ctggaactgg aatctctttc catagaatgt 60

tctaaaccat cctgcggcct ctact 85

<210> 288
<211> 85

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 288
gttccttgac ccaacacgta tgtaccaggt tcctggaact ggaatctctt tccatagaat 60

gttctaaacc atcctgcggc ctcta 85

<210> 289
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 289
gggttccttg acccaacacg tatgtaccag gttcctggaa ctggaatctc tttccataga 60

atgttctaaa ccatcctgcg gcctc 85

<210> 290
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 290
gttggttcct tgacccaaca cgtatgtacc aggttcctgg aactggaatc tctttccata 60

gaatgttcta aaccatcctg cggcc 85

<210> 291
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 291
gccttggttc cttgacccaa cacgtatgta ccaggttcct ggaactggaa tctctttcca 60

tagaatgttc taaaccatcc tgcgg 85

<210> 292
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 292
ggcccttggt tccttgaccc aacacgtatg taccaggttc ctggaactgg aatctctttc 60

catagaatgt tctaaaccat cctgc 85

<210> 293
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 293
gccgcccttg gttccttgac ccaacacgta tgtaccaggt tcctggaact ggaatctctt 60

tccatagaat gttctaaacc atcct 85

<210> 294
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 294
gcgccgccct tggttccttg acccaacacg tatgtaccag gttcctggaa ctggaatctc 60

tttccataga atgttctaaa ccatc 85

<210> 295
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 295
ggtcgccgcc cttggttcct tgacccaaca cgtatgtacc aggttcctgg aactggaatc 60

tctttccata gaatgttcta aacca 85

<210> 296
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 296

Gly Ser Gly Gly Gly Gly Ser
1 5

<210> 297
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 297

Glu Ala Ala Ala Lys
1 5

<210> 298
<211> 18
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 298

Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly
1 5 10 15

Gly Ser

<210> 299
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 299

Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser
1 5 10 15

<210> 300
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 300
gccagcuuuc cgggcauugg cuuccauc 28

<210> 301
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 301
gccagcuuuc cgggcauugg cuuccauc 28

<210> 302
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 302
acccaggaau cucaggaaug ucgacgau 28

<210> 303
<211> 28

<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 303
aggguuuucc cagucacgac guuguaaa 28

<210> 304
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 304
gggcauuggc uuccaucucu uugagcaccu 30

<210> 305
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 305
gugcagccag cuuuccgggc auuggcuucc 30

<210> 306
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 306
gcagccagcu uuccgggcau uggcuuccau 30

<210> 307
<211> 30
<212> RNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 307
ggcuuccauc ucuuugagca ccuccagcgg 30

<210> 308
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 308
ggaaugucga cgaucgccuc gccuaugccg 30

<210> 309
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 309
gaaugucgac gaucgccucg ccuaugccgc 30

<210> 310
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 310
gaccugugcg augaacugcu ccaugggcuc 30

<210> 311
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 311
guguggcagc guccugggau gaacuucuuc 30

<210> 312
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 312
guggcagcgu ccugggauga acuucuucau 30

<210> 313
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 313
gcuucuugcc gggcaacuuc ccgcggucag 30

<210> 314
<211> 31
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 314
gcaggguuuu cccagucacg acguuguaaa a 31

<210> 315
<211> 31
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 315
gcaggguuuu cccagucacg acguuguaaa a 31

<210> 316
<211> 28

<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 316
acccaggaau cucaggaaug ucgacgau 28

<210> 317
<211> 21
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 317
cagcuuuccg ggcauuggcu u 21

<210> 318
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 318
ccgcuggagg ugcucaaaga gauggaagcc 30

<210> 319
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 319
gccagcuuuc cgggcauugg cuuccauc 28

<210> 320
<211> 28
<212> RNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 320
acccaggaau cucaggaaug ucgacgau 28

<210> 321
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 321
gggcauuggc uuccaucucu uugagcaccu 30

<210> 322
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 322
gaaugucgac gaucgccucg ccuaugccgc 30

<210> 323
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 323
caaggcacuc uugccuacgc caccagcu 28

<210> 324
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 324
ucauauucgu ccacaaaaug auucugaa 28

<210> 325
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 325
auuauuuaug gcaaauacac aaagaaag 28

<210> 326
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 326
gaauaucuuc aaaugauuua guauuauu 28

<210> 327
<211> 28
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 327
accauaggua caucuucaga guccuuaa 28

<210> 328
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 328
gucaaggcac ucuugccuac gccaccagcu 30

<210> 329
<211> 30

<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 329
gaucauauuc guccacaaaa ugauucugaa 30

<210> 330
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 330
guauuauuua uggcaaauac acaaagaaag 30

<210> 331
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 331
gugaauaucu ucaaaugauu uaguauuauu 30

<210> 332
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 332
ggaccauagg uacaucuuca gaguccuuaa 30

<210> 333
<211> 50
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 333
aagagtgcct tgacgataca gcctcgaggc tgtatcgtca aggcactctt 50

<210> 334
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 334
aatcattttg tggacgaata tctcgagata ttcgtccaca aaatgatt 48

<210> 335
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 335
aaataatact aaatcatttg actcgagtca aatgatttag tattattt 48

<210> 336
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 336
aataatacta aatcatttga actcgagttc aaatgattta gtattatt 48

<210> 337
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 337
aaggactctg aagatgtacc tctcgagagg tacatcttca gagtcctt 48

<210> 338
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 338
gagaucaggg caaacagaac uuugacuccc 30

<210> 339
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 339
ggaugcagau cagggcaaac agaacuuuga 30

<210> 340
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 340
gcacagcgau gcagaucagg gcaaacagaa 30

<210> 341
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 341
gcucggccac agcgaugcag aucagggcaa 30

<210> 342
<211> 30

<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 342
ggggcuuggc cucggccaca gcgaugcaga 30

<210> 343
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 343
gugggcuugg ccucggccac agcgaugcag 30

<210> 344
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 344
gucucggugg gcuuggccuc ggccacagcg 30

<210> 345
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 345
guucguuguu cucggugggc uuggccucgg 30

<210> 346
<211> 30
<212> RNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 346
ggaagucuuc guuguucucg gugggcuugg 30

<210> 347
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 347
gauguugaag ucuucguugu ucucgguggg 30

<210> 348
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 348
gcggccacga uguugaaguc uucguuguuc 30

<210> 349
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 349
guggccacgg ccacgauguu gaagucuucg 30

<210> 350
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 350
gguugcuggc cacggccacg auguugaagu 30

<210> 351
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 351
gucgcgaagu ugcuggccac ggccacgaug 30

<210> 352
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 352
gccguggucg cgaaguugcu ggccacggcc 30

<210> 353
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 353
gcgagauccg uggucgcgaa guugcuggcc 30

<210> 354
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 354
gcagcaucga gauccguggu cgcgaaguug 30

<210> 355
<211> 30

<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 355
gggucagcau cgagauccgu ggucgcgaag 30

<210> 356
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 356
gcuucccgcg gucagcaucg agauccgugg 30

<210> 357
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 357
ggggcaacuu cccgcgguca gcaucgagau 30

<210> 358
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 358
gucuugccgg gcaacuuccc gcggucagca 30

<210> 359
<211> 30
<212> RNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 359
ggcagcuucu ugccgggcaa cuucccgcgg 30

<210> 360
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 360
gccagcggca gcuucuugcc gggcaacuuc 30

<210> 361
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 361
gcaccuccag cggcagcuuc uugccgggca 30

<210> 362
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 362
gcuuugagca ccuccagcgg cagcuucuug 30

<210> 363
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 363
gcaucucuuu gagcaccucc agcggcagcu 30

<210> 364
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 364
guccaucucu uugagcaccu ccagcggcag 30

<210> 365
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 365
gggcauuggc uuccaucucu uugagcaccu 30

<210> 366
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 366
guccgggcau uggcuuccau cucuuugagc 30

<210> 367
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 367
ggccagcuuu ccgggcauug gcuuccaucu 30

<210> 368
<211> 30

<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 368
gggugcagcc agcuuuccgg gcauuggcuu 30

<210> 369
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 369
gagccccugg ugcagccagc uuuccgggca 30

<210> 370
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 370
gaucagacag ccccuggugc agccagcuuu 30

<210> 371
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 371
ggcagaucag acagccccug gugcagccag 30

<210> 372
<211> 30
<212> RNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 372
gacaggcaga ucagacagcc ccuggugcag 30

<210> 373
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 373
gugauguggg acaggcagau cagacagccc 30

<210> 374
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 374
gacuugaugu gggacaggca gaucagacag 30

<210> 375
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 375
ggggcgugca cuugaugugg gacaggcaga 30

<210> 376
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 376
gcuucaucuu gggcgugcac uugauguggg 30

<210> 377
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 377
gugaacuucu ucaucuuggg cgugcacuug 30

<210> 378
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 378
gggaugaacu ucuucaucuu gggcgugcac 30

<210> 379
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 379
gugggaugaa cuucuucauc uugggcgugc 30

<210> 380
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 380
gggcagcguc cugggaugaa cuucuucauc 30

<210> 381
<211> 30

<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 381
ggguguggca gcguccuggg augaacuucu 30

<210> 382
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 382
guucguaggu guggcagcgu ccugggauga 30

<210> 383
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 383
gcgccuucgu agguguggca gcguccuggg 30

<210> 384
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 384
gucuuugucg ccuucguagg uguggcagcg 30

<210> 385
<211> 30
<212> RNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 385
gcuuugucgc cuucguaggu guggcagcgu 30

<210> 386
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 386
gugccgcccu gugcggacuc uuugucgccu 30

<210> 387
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 387
guaugccgcc cugugcggac ucuuugucgc 30

<210> 388
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 388
gccucgccua ugccgcccug ugcggacucu 30

<210> 389
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 389
ggaucgccuc gccuaugccg cccugugcgg 30

<210> 390
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 390
gaugucgacg aucgccucgc cuaugccgcc 30

<210> 391
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 391
gcaggaaugu cgacgaucgc cucgccuaug 30

<210> 392
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 392
gaaucucagg aaugucgacg aucgccucgc 30

<210> 393
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 393
gcccaggaau cucaggaaug ucgacgaucg 30

<210> 394
<211> 30

<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 394
gccuugaacc caggaaucuc aggaaugucg 30

<210> 395
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 395
gccaaguccu ugaacccagg aaucucagga 30

<210> 396
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 396
gugggcucca aguccuugaa cccaggaauc 30

<210> 397
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 397
gccaugggcu ccaaguccuu gaacccagga 30

<210> 398
<211> 30
<212> RNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 398
ggaacugcuc caugggcucc aaguccuuga 30

<210> 399
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 399
gugcgaugaa cugcuccaug ggcuccaagu 30

<210> 400
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 400
ggaccugugc gaugaacugc uccaugggcu 30

<210> 401
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 401
gacagaucga ccugugcgau gaacugcucc 30

<210> 402
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 402
gacacacaga ucgaccugug cgaugaacug 30

<210> 403
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 403
gugcagucca cacacagauc gaccugugcg 30

<210> 404
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 404
gccaguugug caguccacac acagaucgac 30

<210> 405
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 405
gggcagccag uugugcaguc cacacacaga 30

<210> 406
<211> 16
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 406
guugugcagu ccacac 16

<210> 407
<211> 30

<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 407
gaagcccuuu gaggcagcca guugugcagu 30

<210> 408
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 408
gcacguuggc aagcccuuug aggcagccag 30

<210> 409
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 409
gacugcacgu uggcaagccc uuugaggcag 30

<210> 410
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 410
gggucagaac acugcacguu ggcaagcccu 30

<210> 411
<211> 30
<212> RNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 411
gcaggucaga acacugcacg uuggcaagcc 30

<210> 412
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 412
gagcagguca gaacacugca cguuggcaag 30

<210> 413
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 413
ggccacuucu ugagcagguc agaacacugc 30

<210> 414
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 414
gcggcagcca cuucuugagc aggucagaac 30

<210> 415
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 415
gugcggcagc cacuucuuga gcaggucaga 30

<210> 416
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 416
gagcguugcg gcagccacuu cuugagcagg 30

<210> 417
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 417
gaaaggucgc acagcguugc ggcagccacu 30

<210> 418
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 418
gcuggcaaag gucgcacagc guugcggcag 30

<210> 419
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 419
gggcaaaggu cgcacagcgu ugcggcagcc 30

<210> 420
<211> 30

<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 420
guggaucuug cuggcaaagg ucgcacagcg 30

<210> 421
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 421
gcaccuggcc cuggaucuug cuggcaaagg 30

<210> 422
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 422
guggcccugg aucuugcugg caaaggucgc 30

<210> 423
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 423
gugaucuugu ccaccuggcc cuggaucuug 30

<210> 424
<211> 30
<212> RNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 424
gccccuugau cuuguccacc uggcccugga 30

<210> 425
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 425
gcccuugauc uuguccaccu ggcccuggau 30

<210> 426
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 426
gccuugaucu uguccaccug gcccuggauc 30

<210> 427
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 427
ggcaaagguc gcacagcguu gcggcagcca 30

<210> 428
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 428
gcaaaggucg cacagcguug cggcagccac 30

<210> 429
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 429
gaaggucgca cagcguugcg gcagccacuu 30

<210> 430
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 430
gaggucgcac agcguugcgg cagccacuuc 30

<210> 431
<211> 31
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 431
gguaaugccu ggcuugucga cgcauagucu g 31

<210> 432
<211> 31
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 432
gggaaccuug gccguuauaa agucugacca g 31

<210> 433
<211> 31

<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 433
ggagggugag aauuuagaac caagauuguu g 31

<210> 434
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 434
gaguccuggc aaugaacagu ggcgcaguag 30

<210> 435
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 435
gggugccaca gcugcuauca auacauucuc 30

<210> 436
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 436
guuacauacu gacacauucg gcaacauguu 30

<210> 437
<211> 30
<212> RNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 437
guauguacca gguuccugga acuggaaucu 30

<210> 438
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 438
gccuugguuc cauccagguu cuccagggug 30

<210> 439
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 439
gcagugaugg gauucucagu agcuugagcg 30

<210> 440
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 440
gagccuggca ucucaacaac agcgauggug 30

<210> 441
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 441
gugucugggg cgauucuuac agaucuuccu 30

<210> 442
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 442
gcuggaucug aagugaaguc uguaucuucc 30

<210> 443
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 443
ggcaacguca ucaggauuuc cauagagugg 30

<210> 444
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 444
gaggcgcagg agauggugua guaguagaag 30

<210> 445
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 445
gagggacccu ggaauuggua ucuugcuuug 30

<210> 446
<211> 30

<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 446
gguaagaguc aacauuccug ugugaaaccu 30

<210> 447
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 447
gaccagaauc uguuuuccau caacaaugag 30

<210> 448
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 448
gauggcugua gucaguaugu caccaucuug 30

<210> 449
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 449
guaccaucga auggaucucu aauauguacg 30

<210> 450
<211> 30
<212> RNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 450
gagaucacag gcuccuucag caucaaaaga 30

<210> 451
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 451
gcuuugaccg gcgaagagac uauugcagag 30

<210> 452
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 452
gccccucagg caauacucgu acaugcaucg 30

<210> 453
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 453
gcugguacuu cuaggguguc uccaugcuuu 30

<210> 454
<211> 31
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 454
gguaaugccu ggcuugucga cgcauagucu g 31

<210> 455
<211> 31
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 455
gggaaccuug gccguuauaa agucugacca g 31

<210> 456
<211> 31
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 456
ggagggugag aauuuagaac caagauuguu g 31

<210> 457
<211> 31
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 457
gcauccugcg gccucuacuc ugcauucaau u 31

<210> 458
<211> 31
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 458
gaccauccug cggccucuac ucugcauuca a 31

<210> 459
<211> 31

<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 459
gaaaccaucc ugcggccucu acucugcauu c 31

<210> 460
<211> 31
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 460
gcuaaaccau ccugcggccu cuacucugca u 31

<210> 461
<211> 31
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 461
guucuaaacc auccugcggc cucuacucug c 31

<210> 462
<211> 31
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 462
guguucuaaa ccauccugcg gccucuacuc u 31

<210> 463
<211> 31
<212> RNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 463
gaauguucua aaccauccug cggccucuac u 31

<210> 464
<211> 31
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 464
gagaauguuc uaaaccaucc ugcggccucu a 31

<210> 465
<211> 31
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 465
gauagaaugu ucuaaaccau ccugcggccu c 31

<210> 466
<211> 31
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 466
gccauagaau guucuaaacc auccugcggc c 31

<210> 467
<211> 31
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 467
guuccauaga auguucuaaa ccauccugcg g 31

<210> 468
<211> 31
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 468
gcuuuccaua gaauguucua aaccauccug c 31

<210> 469
<211> 31
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 469
gcucuuucca uagaauguuc uaaaccaucc u 31

<210> 470
<211> 31
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 470
gaucucuuuc cauagaaugu ucuaaaccau c 31

<210> 471
<211> 31
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 471
ggaaucucuu uccauagaau guucuaaacc a 31

<210> 472
<211> 51

<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 472
gcauccugcg gccucuacuc ugcauucaau uacauacuga cacauucggc a 51

<210> 473
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 473
gaccauccug cggccucuac ucugcauuca auuacauacu gacacauucg g 51

<210> 474
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 474
gaaaccaucc ugcggccucu acucugcauu caauuacaua cugacacauu c 51

<210> 475
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 475
gcuaaaccau ccugcggccu cuacucugca uucaauuaca uacugacaca u 51

<210> 476
<211> 51
<212> RNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 476
guucuaaacc auccugcggc cucuacucug cauucaauua cauacugaca c 51

<210> 477
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 477
guguucuaaa ccauccugcg gccucuacuc ugcauucaau uacauacuga c 51

<210> 478
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 478
gaauguucua aaccauccug cggccucuac ucugcauuca auuacauacu g 51

<210> 479
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 479
gagaauguuc uaaaccaucc ugcggccucu acucugcauu caauuacaua c 51

<210> 480
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 480
gauagaaugu ucuaaaccau ccugcggccu cuacucugca uucaauuaca u 51

<210> 481
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 481
gccauagaau guucuaaacc auccugcggc cucuacucug cauucaauua c 51

<210> 482
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 482
guuccauaga auguucuaaa ccauccugcg gccucuacuc ugcauucaau u 51

<210> 483
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 483
gcuuuccaua gaauguucua aaccauccug cggccucuac ucugcauuca a 51

<210> 484
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 484
gcucuuucca uagaauguuc uaaaccaucc ugcggccucu acucugcauu c 51

<210> 485
<211> 51

<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 485
gaucucuuuc cauagaaugu ucuaaaccau ccugcggccu cuacucugca u 51

<210> 486
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 486
ggaaucucuu uccauagaau guucuaaacc auccugcggc cucuacucug c 51

<210> 487
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 487
guggaaucuc uuuccauaga auguucuaaa ccauccugcg gccucuacuc u 51

<210> 488
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 488
gacuggaauc ucuuuccaua gaauguucua aaccauccug cggccucuac u 51

<210> 489
<211> 51
<212> RNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 489
ggaacuggaa ucucuuucca uagaauguuc uaaaccaucc ugcggccucu a 51

<210> 490
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 490
guggaacugg aaucucuuuc cauagaaugu ucuaaaccau ccugcggccu c 51

<210> 491
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 491
gccuggaacu ggaaucucuu uccauagaau guucuaaacc auccugcggc c 51

<210> 492
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 492
guuccuggaa cuggaaucuc uuuccauaga auguucuaaa ccauccugcg g 51

<210> 493
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 493
ggguuccugg aacuggaauc ucuuuccaua gaauguucua aaccauccug c 51

<210> 494
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 494
gcagguuccu ggaacuggaa ucucuuucca uagaauguuc uaaaccaucc u 51

<210> 495
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 495
gaccagguuc cuggaacugg aaucucuuuc cauagaaugu ucuaaaccau c 51

<210> 496
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 496
gguaccaggu uccuggaacu ggaaucucuu uccauagaau guucuaaacc a 51

<210> 497
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 497
gcauccugcg gccucuacuc ugcauucaau uacauacuga cacauucggc aacauguuuu 60

uccugguuua u 71

<210> 498
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 498
gaccauccug cggccucuac ucugcauuca auuacauacu gacacauucg gcaacauguu 60

uuuccugguu u 71

<210> 499
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 499
gaaaccaucc ugcggccucu acucugcauu caauuacaua cugacacauu cggcaacaug 60

uuuuuccugg u 71

<210> 500
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 500
gcuaaaccau ccugcggccu cuacucugca uucaauuaca uacugacaca uucggcaaca 60

uguuuuuccu g 71

<210> 501
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 501
guucuaaacc auccugcggc cucuacucug cauucaauua cauacugaca cauucggcaa 60

cauguuuuuc c 71

<210> 502
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 502
guguucuaaa ccauccugcg gccucuacuc ugcauucaau uacauacuga cacauucggc 60

aacauguuuu u 71

<210> 503
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 503
gaauguucua aaccauccug cggccucuac ucugcauuca auuacauacu gacacauucg 60

gcaacauguu u 71

<210> 504
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 504
gagaauguuc uaaaccaucc ugcggccucu acucugcauu caauuacaua cugacacauu 60

cggcaacaug u 71

<210> 505
<211> 71
<212> RNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 505
gauagaaugu ucuaaaccau ccugcggccu cuacucugca uucaauuaca uacugacaca 60

uucggcaaca u 71

<210> 506
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 506
gccauagaau guucuaaacc auccugcggc cucuacucug cauucaauua cauacugaca 60

cauucggcaa c 71

<210> 507
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 507
guuccauaga auguucuaaa ccauccugcg gccucuacuc ugcauucaau uacauacuga 60

cacauucggc a 71

<210> 508
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 508
gcuuuccaua gaauguucua aaccauccug cggccucuac ucugcauuca auuacauacu 60

gacacauucg g 71

<210> 509
<211> 71

<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 509
gcucuuucca uagaauguuc uaaaccaucc ugcggccucu acucugcauu caauuacaua 60

cugacacauu c 71

<210> 510
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 510
gaucucuuuc cauagaaugu ucuaaaccau ccugcggccu cuacucugca uucaauuaca 60

uacugacaca u 71

<210> 511
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 511
ggaaucucuu uccauagaau guucuaaacc auccugcggc cucuacucug cauucaauua 60

cauacugaca c 71

<210> 512
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 512
guggaaucuc uuuccauaga auguucuaaa ccauccugcg gccucuacuc ugcauucaau 60

uacauacuga c 71

<210> 513
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 513
gacuggaauc ucuuuccaua gaauguucua aaccauccug cggccucuac ucugcauuca 60

auuacauacu g 71

<210> 514
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 514
ggaacuggaa ucucuuucca uagaauguuc uaaaccaucc ugcggccucu acucugcauu 60

caauuacaua c 71

<210> 515
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 515
guggaacugg aaucucuuuc cauagaaugu ucuaaaccau ccugcggccu cuacucugca 60

uucaauuaca u 71

<210> 516
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 516
gccuggaacu ggaaucucuu uccauagaau guucuaaacc auccugcggc cucuacucug 60

cauucaauua c 71

<210> 517
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 517
guuccuggaa cuggaaucuc uuuccauaga auguucuaaa ccauccugcg gccucuacuc 60

ugcauucaau u 71

<210> 518
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 518
gggguuccugg aacuggaauc ucuuuccaua gaauguucua aaccauccug cggccucuac 60

ucugcauuca a 71

<210> 519
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 519
gcagguuccu ggaacuggaa ucucuuucca uagaauguuc uaaaccaucc ugcggccucu 60

acucugcauu c 71

<210> 520
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 520
gaccagguuc cuggaacugg aaucucuuuc cauagaaugu ucuaaaccau ccugcggccu 60

cuacucugca u 71

<210> 521
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 521
gguaccaggu uccuggaacu ggaaucucuu uccauagaau guucuaaacc auccugcggc 60

cucuacucug c 71

<210> 522
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 522
gauguaccag guuccuggaa cuggaaucuc uuuccauaga auguucuaaa ccauccugcg 60

gccucuacuc u 71

<210> 523
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 523
gguauguacc agguuccugg aacuggaauc ucuuuccaua gaauguucua aaccauccug 60

cggccucuac u 71

<210> 524
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 524
gacguaugua ccagguuccu ggaacuggaa ucucuuucca uagaauguuc uaaaccaucc 60

ugcggccucu a 71

<210> 525
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 525
gacacguaug uaccagguuc cuggaacugg aaucucuuuc cauagaaugu ucuaaaccau 60

ccugcggccu c 71

<210> 526
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 526
gcaacacgua uguaccaggu uccuggaacu ggaaucucuu uccauagaau guucuaaacc 60

auccugcggc c 71

<210> 527
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 527
gcccaacacg uauguaccag guuccuggaa cuggaaucuc uuuccauaga auguucuaaa 60

ccauccugcg g 71

<210> 528
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 528
ggacccaaca cguauguacc agguuccugg aacuggaauc ucuuuccaua gaauguucua 60

aaccauccug c 71

<210> 529
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 529
guugacccaa cacguaugua ccagguuccu ggaacuggaa ucucuuucca uagaauguuc 60

uaaaccaucc u 71

<210> 530
<211> 71
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 530
gccuugaccc aacacguaug uaccagguuc cuggaacugg aaucucuuuc cauagaaugu 60

ucuaaaccau c 71

<210> 531
<211> 71
<212> RNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 531
guuccuugac ccaacacgua uguaccaggu uccuggaacu ggaaucucuu uccauagaau 60

guucuaaacc a 71

<210> 532
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 532
gcauccugcg gccucuacuc ugcauucaau uacauacuga cacauucggc aacauguuuu 60

uccugguuua uuuucacaca gucca 85

<210> 533
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 533
gaccauccug cggccucuac ucugcauuca auuacauacu gacacauucg gcaacauguu 60

uuuccugguu uauuuucaca caguc 85

<210> 534
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 534
gaaaccaucc ugcggccucu acucugcauu caauuacaua cugacacauu cggcaacaug 60

uuuuuccugg uuuauuuuca cacag 85

<210> 535
<211> 85

<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 535
gcuaaaccau ccugcggccu cuacucugca uucaauuaca uacugacaca uucggcaaca 60

uguuuuuccu gguuuauuuu cacac 85

<210> 536
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 536
guucuaaacc auccugcggc cucuacucug cauucaauua cauacugaca cauucggcaa 60

cauguuuuuc cugguuuauu uucac 85

<210> 537
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 537
guguucuaaa ccauccugcg gccucuacuc ugcauucaau uacauacuga cacauucggc 60

aacauguuuu uccugguuua uuuuc 85

<210> 538
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 538
gaauguucua aaccauccug cggccucuac ucugcauuca auuacauacu gacacauucg 60

gcaacauguu uuuccugguu uauuu 85

<210> 539
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 539
gagaauguuc uaaaccaucc ugcggccucu acucugcauu caauuacaua cugacacauu 60

cggcaacaug uuuuuccugg uuuau 85

<210> 540
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 540
gauagaaugu ucuaaaccau ccugcggccu cuacucugca uucaauuaca uacugacaca 60

uucggcaaca uguuuuuccu gguuu 85

<210> 541
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 541
gccauagaau guucuaaacc auccugcggc cucuacucug cauucaauua cauacugaca 60

cauucggcaa cauguuuuuc cuggu 85

<210> 542
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 542
guuccauaga auguucuaaa ccauccugcg gccucuacuc ugcauucaau uacauacuga 60

cacauucggc aacauguuuu uccug 85

<210> 543
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 543
gcuuuccaua gaauguucua aaccauccug cggccucuac ucugcauuca auuacauacu 60

gacacauucg gcaacauguu uuucc 85

<210> 544
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 544
gcucuuucca uagaauguuc uaaaccaucc ugcggccucu acucugcauu caauuacaua 60

cugacacauu cggcaacaug uuuuu 85

<210> 545
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 545
gaucucuuuc cauagaaugu ucuaaaccau ccugcggccu cuacucugca uucaauuaca 60

uacugacaca uucggcaaca uguuu 85

<210> 546
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 546
ggaaucucuu uccauagaau guucuaaacc auccugcggc cucuacucug cauucaauua 60

cauacugaca cauucggcaa caugu 85

<210> 547
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 547
guggaaucuc uuuccauaga auguucuaaa ccauccugcg gccucuacuc ugcauucaau 60

uacauacuga cacauucggc aacau 85

<210> 548
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 548
gacuggaauc ucuuuccaua gaauguucua aaccauccug cggccucuac ucugcauuca 60

auuacauacu gacacauucg gcaac 85

<210> 549
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 549
ggaacuggaa ucucuuucca uagaauguuc uaaaccaucc ugcggccucu acucugcauu 60

caauuacaua cugacacauu cggca 85

<210> 550
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 550
guggaacugg aaucucuuuc cauagaaugu ucuaaaccau ccugcggccu cuacucugca 60

uucaauuaca uacugacaca uucgg 85

<210> 551
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 551
gccuggaacu ggaaucucuu uccauagaau guucuaaacc auccugcggc cucuacucug 60

cauucaauua cauacugaca cauuc 85

<210> 552
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 552
guuccuggaa cuggaaucuc uuuccauaga auguucuaaa ccauccugcg gccucuacuc 60

ugcauucaau uacauacuga cacau 85

<210> 553
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 553
gggguuccugg aacuggaauc ucuuuccaua gaauguucua aaccauccug cggccucuac 60

ucugcauuca auuacauacu gacac 85

<210> 554
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 554
gcagguuccu ggaacuggaa ucucuuucca uagaauguuc uaaaccaucc ugcggccucu 60

acucugcauu caauuacaua cugac 85

<210> 555
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 555
gaccagguuc cuggaacugg aaucucuuuc cauagaaugu ucuaaaccau ccugcggccu 60

cuacucugca uucaauuaca uacug 85

<210> 556
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 556
gguaccaggu uccuggaacu ggaaucucuu uccauagaau guucuaaacc auccugcggc 60

cucuacucug cauucaauua cauac 85

<210> 557
<211> 85
<212> RNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 557
gauguaccag guuccuggaa cuggaaucuc uuuccauaga auguucuaaa ccauccugcg 60

gccucuacuc ugcauucaau uacau 85

<210> 558
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 558
gguauguacc agguuccugg aacuggaauc ucuuuccaua gaauguucua aaccauccug 60

cggccucuac ucugcauuca auuac 85

<210> 559
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 559
gacguaugua ccagguuccu ggaacuggaa ucucuuucca uagaauguuc uaaaccaucc 60

ugcggccucu acucugcauu caauu 85

<210> 560
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 560
gacacguaug uaccagguuc cuggaacugg aaucucuuuc cauagaaugu ucuaaaccau 60

ccugcggccu cuacucugca uucaa 85

<210> 561
<211> 85

<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 561
gcaacacgua uguaccaggu uccuggaacu ggaaucucuu uccauagaau guucuaaacc 60

auccugcggc cucuacucug cauuc 85

<210> 562
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 562
gcccaacacg uauguaccag guuccuggaa cuggaaucuc uuuccauaga auguucuaaa 60

ccauccugcg gccucuacuc ugcau 85

<210> 563
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 563
ggacccaaca cguauguacc agguuccugg aacuggaauc ucuuuccaua gaauguucua 60

aaccauccug cggccucuac ucugc 85

<210> 564
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 564
guugacccaa cacguaugua ccagguuccu ggaacuggaa ucucuuucca uagaauguuc 60

uaaaccaucc ugcggccucu acucu 85

<210> 565
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 565
gccuugaccc aacacguaug uaccagguuc cuggaacugg aaucucuuuc cauagaaugu 60

ucuaaaccau ccugcggccu cuacu 85

<210> 566
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 566
guuccuugac ccaacacgua uguaccaggu uccuggaacu ggaaucucuu uccauagaau 60

guucuaaacc auccugcggc cucua 85

<210> 567
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 567
ggguuccuug acccaacacg uauguaccag guuccuggaa cuggaaucuc uuuccauaga 60

auguucuaaa ccauccugcg gccuc 85

<210> 568
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 568
guugguuccu ugacccaaca cguauguacc agguuccugg aacuggaauc ucuuuccaua 60

gaauguucua aaccauccug cggcc 85

<210> 569
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 569
gccuugguuc cuugacccaa cacguaugua ccagguuccu ggaacuggaa ucucuuucca 60

uagaauguuc uaaaccaucc ugcgg 85

<210> 570
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 570
ggcccuuggu uccuugaccc aacacguaug uaccagguuc cuggaacugg aaucucuuuc 60

cauagaaugu ucuaaaccau ccugc 85

<210> 571
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 571
gccgcccuug guuccuugac ccaacacgua uguaccaggu uccuggaacu ggaaucucuu 60

uccauagaau guucuaaacc auccu 85

<210> 572
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 572
gcgccgcccu ugguuccuug acccaacacg uauguaccag guuccuggaa cuggaaucuc 60

uuuccauaga auguucuaaa ccauc 85

<210> 573
<211> 85
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 573
ggucgccgcc cuugguuccu ugacccaaca cguauguacc agguuccugg aacuggaauc 60

ucuuuccaua gaauguucua aacca 85

<210> 574
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 574
guaaugccug gcuugucgac gcauagucug 30

<210> 575
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 575
gaaaacgcag guuccuccag uuucgggagc agcgcacguc ucccuguagu c 51

<210> 576
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 576
gacgcagguu ccucuagcuu cgggagcagc gcacgucucc cuguagucaa g 51

<210> 577
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 577
guaaugccug gcuugucgac gcauagucug 30

<210> 578
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 578
gauagaaugu ucuaaaccau ccugcggccu cuacucugca uucaauuaca u 51

<210> 579
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 579
guaaugccug gcuugucgac gcauagucug 30

<210> 580
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 580
gcaaggccac aaaauuaucc acuguuuuug gaacagucuu uccgaagaga c 51

<210> 581
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 581
gccuguagcc aaggccacaa aauuauccac uguuuuugga acagucuuuc c 51

<210> 582
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 582
gcuuucucuc cuguagccaa ggccacaaaa uuauccacug uuuuuggaac a 51

<210> 583
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 583
ggccaaaucc uuucucuccu guagccaagg ccacaaaauu auccacuguu u 51

<210> 584
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 584
guuuuuguag ccaaauccuu ucucuccugu agccaaggcc acaaaauuau c 51

<210> 585
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 585
gauuugcugu uuuuguagcc aaauccuuuc ucuccuguag ccaaggccac a 51

<210> 586
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 586
gacgauggaa uuugcuguuu uuguagccaa auccuuucuc uccuguagcc a 51

<210> 587
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 587
gauagaaugu ucuaaacgau ccugcggccu cuacucugca uucaauuaca u 51

<210> 588
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 588
gauagaaugu ucuaaacaau ccugcggccu cuacucugca uucaauuaca u 51

<210> 589
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 589
gauagaaugu ucuaaacuau ccugcggccu cuacucugca uucaauuaca u 51

<210> 590
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 590
ggucccacgc ggcccacagc ugcaccagga agaagggugc ccagcacagc a 51

<210> 591
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 591
ggggucccac gcggcccaca gcugcaccag gaagaagggu gcccagcaca g 51

<210> 592
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 592
gccggguccc acgcggccca cagcugcacc aggaagaagg gugcccagca c 51

<210> 593
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 593
gggugaugac aucccaggcg aucguguggc cuccaggagc ccagagcagg a 51

<210> 594
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 594
gagggugaug acaucccagg cgaucgugug gccuccagga gcccagagca g 51

<210> 595
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 595
gaucagggug augacauccc aggcgaucgu guggccucca ggagcccaga g 51

<210> 596
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 596
gguggcucca uucacuccaa ugcugagcac uuccacagag uggguuaaag c 51

<210> 597
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 597
guuucuaaua uauuuugcca gacugaugga cuauucucaa uuaauaauga u 51

<210> 598
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 598
gagauguugc uguggaucca guccacagcc agcccgucgg gggccuggau g 51

<210> 599
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 599
gcaggccggc ccagcugcca ggugcaccug cucggagcau cgggccggau c 51

<210> 600
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 600
gcaaagaacc ucugggucca aggguagacc accagcagcc ugcccagggc c 51

<210> 601
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 601
gaagagaaac uuaguuucca gggcuuuggu agagggcaaa gguugauagc a 51

<210> 602
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 602
gucagccuag ugcagagcca cugguaguug gugguuagag uuucaaguuc c 51

<210> 603
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 603
ggcucauugu gaacaggcca guaauguccg ggaugggcg gcauaggcgg g 51

<210> 604
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 604
guagcuaaag aacuugacca agacauauca ggauccaccu cagcuccuag a 51

<210> 605
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 605
ggggcauugu ucugugccca guccugcugg uagaccugcu ccccgguggc u 51

<210> 606
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 606
gagaagucgu ucaugugcca ccgugggagc guacagucau cauugaucuu g 51

<210> 607
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 607
gggauuaaug cugaacgcac caaaguucau cccaccaccc auauuacuac c 51

<210> 608
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 608
gcuccaaagg cuuuccucca cuguugcaaa guuauugaau cccaagacac a 51

<210> 609
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 609
gaugaaugaa cgauuuccca gaacucccua aucagaacag agucccuggu a 51

<210> 610
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 610
ggagccucug ccggagccca gagaacccga gagucagaca gagccagcgc c 51

<210> 611
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 611
ggcuuccgug gagacaccca aucaauuuga agagaucuug aagugaugcc a 51

<210> 612
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 612
gugggacugc ccuccuccca uuugcagaug ccgucguaga aucgcagcag g 51

<210> 613
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 613
gcuucuucaa uaguucucca gcuacacugg caggcauaug cccguguucc u 51

<210> 614
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 614
gauuccuuuu cuucguccca auucaccuca guggcuaguc gaagaaugaa g 51

<210> 615
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 615
gcagcuucag caccuuccag ucagacuccu gcuucaagca cugcagcagg a 51

<210> 616
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 616
gccauuugcu ugcagugcca cuccagagga uuccggauug ccauaaauac u 51

<210> 617
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 617
guucaauagu uuugguccag uaucguuuac agcccuucuu gguagauuuc a 51

<210> 618
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 618
ggcaaccguc uucugaccaa auggcagaac auuugucccc aacuuuccac u 51

<210> 619
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 619
gcgacuuucc aaugaaccac ugaagcccag guaugacaaa gccgaugauc u 51

<210> 620
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 620
guuuacacuc augcuuccac agcuuuaaca gaucauuugg uuccuugaug a 51

<210> 621
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 621
gcuuaagcuu ccguguccag ccuucaggca gggugggguc aucauacaug g 51

<210> 622
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 622
ggacagcugg gcugauccau gaugucaucc agaaacacug gggacccuca g 51

<210> 623
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 623
gucucaucuc aacuuuccau auccguauca uggaaucaua gcauccugua a 51

<210> 624
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 624
gcaugcagac gcgguuccac ucgcagccac aguuccagca ccacucgagc c 51

<210> 625
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 625
guugguuagg gucaacccag uauucuccac ucuugaguuc aggauggcag a 51

<210> 626
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 626
gcuacacugu ccaacaccca cucucggguc accacaggug ccucacacau c 51

<210> 627
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 627
gcugcacugu guacccccag agcuccgugu ugccgacauc cugggguggc u 51

<210> 628
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 628
gagcuuccug ccacucccaa cagguuucac aguaagcgcg uaucuguucc a 51

<210> 629
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 629
gacggcaaga gcuuacccag ucacuugugu ggagacuuaa auacuugcau a 51

<210> 630
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 630
gcaaggccac aaaauuaucc acuguuuuug gaacagucuu uccgaagaga c 51

<210> 631
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 631
gccuguagcc aaggccacaa aauuauccac uguuuuugga acagucuuuc c 51

<210> 632
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 632
gcuuucucuc cuguagccaa ggccacaaaa uuauccacug uuuuuggaac a 51

<210> 633
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 633
ggccaaaucc uuucucuccu guagccaagg ccacaaaauu auccacuguu u 51

<210> 634
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 634
guuuuuguag ccaaauccuu ucucuccugu agccaaggcc acaaaauuau c 51

<210> 635
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 635
gauuugcugu uuuuguagcc aaauccuuuc ucuccuguag ccaaggccac a 51

<210> 636
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 636
gacgauggaa uuugcuguuu uuguagccaa auccuuucuc uccuguagcc a 51

<210> 637
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 637
guaaugccug gcuugucgac gcauagucug 30

<210> 638
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 638
gauagaaugu ucuaaaccau ccugcggccu cuacucugca uucaauuaca u 51

<210> 639
<211> 34
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 639
gcaggguuuu cccagucacg acguuguaaa guug 34

<210> 640
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 640
gucaaggcac ucuugcccac gccaccagcu ccaacuacca caaguuuaua u 51

<210> 641
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 641
gcaaagauca cccggcccac aucuucaucu ccaauucgua ggucaaaaua c 51

<210> 642
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 642
gcgccaccag cuccaaccac cacaaguuua uauucaguca uuuucagcag g 51

<210> 643
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 643
guuucuccau caauuaccac uugcuuccug uaggaauccu cuauuguugg a 51

<210> 644
<211> 51
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 644
gcuuucucuc cuguagccaa ggccacaaaa uuauccacug uuuuuggaac a 51

<210> 645
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 645
guaaugccug gcuugucgac gcauagucug 30

<210> 646
<211> 83
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 646
ucuuuccaua ggcccugaaa aagggccugu ucuaaaccau ccugcggccu cuacucggcc 60

cugaaaaagg gccauucaau uac 83

<210> 647
<211> 83
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 647
cagcuggcga ggcccugaaa aagggccggg gaugugccgc aaggcgauua aguuggggcc 60

cugaaaaagg gccacgccag ggu 83

<210> 648
<211> 79
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 648
guggaauagu auaacaauau gcuaaauguu guuauaguau cccacucuaa accauccugc 60

ggggcccucu ucagggccc 79

<210> 649
<211> 79
<212> RNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 649
guggaauagu auaacaauau gcuaaauguu guuauaguau cccacacccu ggcguuaccc 60

agggcccucu ucagggccc 79

<210> 650
<211> 385
<212> PRT
<213> Homo sapiens

<400> 650

Gln Leu His Leu Pro Gln Val Leu Ala Asp Ala Val Ser Arg Leu Val
1 5 10 15

Leu Gly Lys Phe Gly Asp Leu Thr Asp Asn Phe Ser Ser Pro His Ala
20 25 30

Arg Arg Lys Val Leu Ala Gly Val Val Met Thr Thr Gly Thr Asp Val
35 40 45

Lys Asp Ala Lys Val Ile Ser Val Ser Thr Gly Thr Lys Cys Ile Asn
50 55 60

Gly Glu Tyr Met Ser Asp Arg Gly Leu Ala Leu Asn Asp Cys His Ala
65 70 75 80

Glu Ile Ile Ser Arg Arg Ser Leu Leu Arg Phe Leu Tyr Thr Gln Leu
85 90 95

Glu Leu Tyr Leu Asn Asn Lys Asp Asp Gln Lys Arg Ser Ile Phe Gln
100 105 110

Lys Ser Glu Arg Gly Gly Phe Arg Leu Lys Glu Asn Val Gln Phe His
115 120 125

Leu Tyr Ile Ser Thr Ser Pro Cys Gly Asp Ala Arg Ile Phe Ser Pro
130 135 140

His Glu Pro Ile Leu Glu Glu Pro Ala Asp Arg His Pro Asn Arg Lys

```
145                 150                 155                 160

Ala Arg Gly Gln Leu Arg Thr Lys Ile Glu Ser Gly Gln Gly Thr Ile
                165                 170                 175

Pro Val Arg Ser Asn Ala Ser Ile Gln Thr Trp Asp Gly Val Leu Gln
            180                 185                 190

Gly Glu Arg Leu Leu Thr Met Ser Cys Ser Asp Lys Ile Ala Arg Trp
        195                 200                 205

Asn Val Val Gly Ile Gln Gly Ser Leu Leu Ser Ile Phe Val Glu Pro
    210                 215                 220

Ile Tyr Phe Ser Ser Ile Ile Leu Gly Ser Leu Tyr His Gly Asp His
225                 230                 235                 240

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

Pro Leu Tyr Thr Leu Asn Lys Pro Leu Leu Ser Gly Ile Ser Asn Ala
            260                 265                 270

Glu Ala Arg Gln Pro Gly Lys Ala Pro Asn Phe Ser Val Asn Trp Thr
        275                 280                 285

Val Gly Asp Ser Ala Ile Glu Val Ile Asn Ala Thr Thr Gly Lys Asp
    290                 295                 300

Glu Leu Gly Arg Ala Ser Arg Leu Cys Lys His Ala Leu Tyr Cys Arg
305                 310                 315                 320

Trp Met Arg Val His Gly Lys Val Pro Ser His Leu Leu Arg Ser Lys
                325                 330                 335

Ile Thr Lys Pro Asn Val Tyr His Glu Ser Lys Leu Ala Ala Lys Glu
            340                 345                 350

Tyr Gln Ala Ala Lys Ala Arg Leu Phe Thr Ala Phe Ile Lys Ala Gly
```

355 360 365

Leu Gly Ala Trp Val Glu Lys Pro Thr Glu Gln Asp Gln Phe Ser Leu
370 375 380

Thr
385

<210> 651
<211> 336
<212> PRT
<213> Homo sapiens

<400> 651

Ser Leu Gly Thr Gly Asn Arg Cys Val Lys Gly Asp Ser Leu Ser Leu
1 5 10 15

Lys Gly Glu Thr Val Asn Asp Cys His Ala Glu Ile Ile Ser Arg Arg
20 25 30

Gly Phe Ile Arg Phe Leu Tyr Ser Glu Leu Met Lys Tyr Asn Ser Gln
35 40 45

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

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

Cys Gly Asp Gly Ala Leu Phe Asp Lys Ser Cys Ser Asp Arg Ala Met
85 90 95

Glu Ser Thr Glu Ser Arg His Tyr Pro Val Phe Glu Asn Pro Lys Gln
100 105 110

Gly Lys Leu Arg Thr Lys Val Glu Asn Gly Gln Gly Thr Ile Pro Val
115 120 125

Glu Ser Ser Asp Ile Val Pro Thr Trp Asp Gly Ile Arg Leu Gly Glu
130 135 140

```
Arg Leu Arg Thr Met Ser Cys Ser Asp Lys Ile Leu Arg Trp Asn Val
145                 150                 155                 160

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

Leu Lys Ser Val Thr Leu Gly Tyr Leu Phe Ser Gln Gly His Leu Thr
            180                 185                 190

Arg Ala Ile Cys Cys Arg Val Thr Arg Asp Gly Ser Ala Phe Glu Asp
        195                 200                 205

Gly Leu Arg His Pro Phe Ile Val Asn His Pro Lys Val Gly Arg Val
    210                 215                 220

Ser Ile Tyr Asp Ser Lys Arg Gln Ser Gly Lys Thr Lys Glu Thr Ser
225                 230                 235                 240

Val Asn Trp Cys Leu Ala Asp Gly Tyr Asp Leu Glu Ile Leu Asp Gly
                245                 250                 255

Thr Arg Gly Thr Val Asp Gly Pro Arg Asn Glu Leu Ser Arg Val Ser
            260                 265                 270

Lys Lys Asn Ile Phe Leu Leu Phe Lys Lys Leu Cys Ser Phe Arg Tyr
        275                 280                 285

Arg Arg Asp Leu Leu Arg Leu Ser Tyr Gly Glu Ala Lys Lys Ala Ala
    290                 295                 300

Arg Asp Tyr Glu Thr Ala Lys Asn Tyr Phe Lys Lys Gly Leu Lys Asp
305                 310                 315                 320

Met Gly Tyr Gly Asn Trp Ile Ser Lys Pro Gln Glu Glu Lys Asn Phe
                325                 330                 335
```

<210> 652
<211> 340

<212> PRT
<213> Octopus vulgaris

<400> 652

Ser Val Gly Thr Gly Asn Arg Cys Leu Thr Gly Asp His Leu Ser Leu
1 5 10 15

Glu Gly Asn Ser Val Asn Asp Ser His Ala Glu Met Ile Thr Arg Arg
20 25 30

Gly Phe Leu Arg Tyr Leu Tyr Arg His Leu Leu Glu Tyr Asp Ala Glu
35 40 45

Val Pro Asn Asp Leu Phe Glu Lys Gly Glu Arg Ser Ile Cys Arg Ile
50 55 60

Lys Thr Asn Ile Thr Phe His Leu Tyr Ile Ser Thr Ala Pro Cys Gly
65 70 75 80

Asp Gly Ala Leu Phe Ser Pro Arg Asp Thr Asp Ser Ser Asn Ala Lys
85 90 95

Met Glu Glu Glu Asn Lys His Ile His Asn Pro Thr Phe Ser Ser Ser
100 105 110

Val Gln Gly Leu Leu Arg Thr Lys Val Glu Gly Gly Gln Gly Thr Ile
115 120 125

Pro Ile Asp Ala Asp Phe Thr Glu Gln Thr Trp Asp Gly Ile Gln Arg
130 135 140

Gly Glu Arg Leu Arg Thr Met Ser Cys Ser Asp Lys Ile Cys Arg Trp
145 150 155 160

Asn Val Val Gly Leu Gln Gly Ala Leu Leu Ser His Phe Ile Glu Pro
165 170 175

Ile Tyr Leu Asp Ser Leu Thr Leu Gly Tyr Leu Tyr Asp His Gly His
180 185 190

```
Leu Ala Arg Ala Val Cys Cys Arg Ile Glu Arg Gly Glu Ala Ser Val
        195                 200                 205


Asn Gln Leu Leu Pro Glu Gly Tyr Arg Leu Asn His Pro Trp Leu Gly
    210                 215                 220


Arg Val Thr Ala Cys Asp Pro Pro Arg Glu Thr Gln Lys Thr Lys Ser
225                 230                 235                 240


Leu Ser Ile Asn Trp Cys Tyr Asp Asp Glu Lys Ser Glu Val Leu Asp
                245                 250                 255


Gly Thr Ala Gly Ile Cys Tyr Thr Ala Ile Glu Lys Asn Leu Phe Ser
            260                 265                 270


Arg Leu Thr Lys His Asn Leu Tyr Glu Glu Phe Lys Arg Val Cys Arg
        275                 280                 285


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


Lys Met Met Ala Thr Pro Phe Gln Thr Ala Lys Asn Val Met Leu Lys
305                 310                 315                 320


Lys Leu Lys Glu Asn Asn Cys Gly Thr Trp Val Ser Lys Pro Ile Glu
                325                 330                 335


Glu Glu Met Phe
            340
```

<210> 653
<211> 340
<212> PRT
<213> Sepia

<400> 653

```
Ser Val Gly Thr Gly Asn Arg Cys Leu Thr Gly Asp Arg Leu Ser Leu
1               5                   10                  15
```

```
Glu Gly Asn Ser Val Asn Asp Ser His Ala Glu Met Val Thr Arg Arg
            20                  25                  30

Gly Phe Leu Arg Tyr Leu Tyr Lys His Leu Leu Glu Tyr Asp Pro Glu
        35                  40                  45

Lys Pro His Asp Leu Phe Glu Lys Gly Glu Arg Ser Leu Cys Arg Ile
    50                  55                  60

Lys Thr Asn Ile Thr Phe His Leu Tyr Ile Ser Thr Ala Pro Cys Gly
65                  70                  75                  80

Asp Gly Ala Leu Phe Ser Pro Arg Asp Thr Asp Ser Ser Asn Val Lys
                85                  90                  95

Val Asp Glu Glu Asn Lys His Val His Asn Pro Thr Phe Ser Ser Ser
            100                 105                 110

Val Gln Gly Leu Leu Arg Thr Lys Val Glu Gly Gly Gln Gly Thr Ile
        115                 120                 125

Pro Ile Asp Ala Asp Phe Thr Glu Gln Thr Trp Asp Gly Ile Gln Arg
    130                 135                 140

Gly Glu Arg Leu Arg Thr Met Ser Cys Ser Asp Lys Ile Cys Arg Trp
145                 150                 155                 160

Asn Val Val Gly Leu Gln Gly Ala Leu Leu Ser His Phe Val Glu Pro
                165                 170                 175

Ile Tyr Leu Glu Ser Leu Thr Leu Gly Tyr Leu Tyr Asp His Gly His
            180                 185                 190

Leu Ala Arg Ala Val Cys Cys Arg Ile Glu Arg Gly Glu Ala Ser Val
        195                 200                 205

Asn Gln Leu Leu Pro Glu Gly Tyr Arg Leu Asn His Pro Trp Leu Gly
    210                 215                 220
```

```
Arg Val Thr Ala Cys Asp Pro Pro Arg Glu Thr Gln Lys Thr Lys Ser
225                 230                 235                 240


Leu Ser Ile Asn Trp Cys Tyr Asp Asp Glu Lys Ser Glu Val Leu Asp
                245                 250                 255


Gly Thr Ala Gly Ile Cys Tyr Thr Ala Ile Glu Lys Asn Leu Phe Ser
            260                 265                 270


Arg Leu Thr Lys His Ser Leu Tyr Glu Glu Phe Lys Lys Val Cys Gln
        275                 280                 285


Lys Phe Glu Arg Glu Asp Leu Leu Asn Val Thr Ser Tyr Asn Lys Ala
    290                 295                 300


Lys Met Met Ala Ile Pro Phe Gln Thr Ala Lys Asn Val Met Leu Lys
305                 310                 315                 320


Lys Leu Lys Glu Asn Asn Cys Gly Thr Trp Val Ser Lys Pro Ile Glu
                325                 330                 335


Glu Glu Met Phe
            340


<210>  654
<211>  328
<212>  PRT
<213>  Octopus vulgaris

<400>  654

Gly Ile Gly Thr Gly Thr Lys Cys Ile Asn Gly Glu His Met Ser Asp
1               5                   10                  15


Arg Gly Phe Gly Val Asn Asp Cys His Ala Glu Ile Ile Ala Arg Arg
            20                  25                  30


Cys Phe Leu Arg Tyr Ile Tyr Asp Gln Leu Glu Leu His Leu Ser Asp
        35                  40                  45


Asn Ser Asp Val Arg Asn Ser Ser Ile Phe Glu Leu Arg Asp Lys Gly
```

```
    50                  55                  60

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

Ala Pro Cys Gly Asp Ala Arg Ile Phe Ser Pro His Gly Gln Asp Val
                85                  90                  95

Glu Thr Gly Asp Arg His Pro Asn Arg Lys Ala Arg Gly Gln Leu Arg
            100                 105                 110

Thr Lys Ile Glu Ser Gly Gln Gly Thr Ile Pro Val Arg Thr Ser Gly
        115                 120                 125

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

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

Gly Ser Leu Leu Ser His Phe Met Asn Pro Ile Tyr Leu Glu Ser Ile
                165                 170                 175

Ile Leu Gly Ser Leu Tyr His Ser Asp His Leu Ser Arg Ala Met Tyr
            180                 185                 190

Ser Arg Ile Ser Ile Ile Glu Asn Leu Pro Glu Pro Phe His Leu Asn
        195                 200                 205

Arg Pro Phe Leu Ser Gly Ile Ser Ser Pro Glu Ser Arg Gln Pro Gly
    210                 215                 220

Lys Ala Pro Asn Phe Gly Ile Asn Trp Arg Lys Glu Asp Glu Thr Phe
225                 230                 235                 240

Glu Val Ile Asn Ala Met Thr Gly Arg Val Glu Gly Gly Ser Val Ser
                245                 250                 255

Arg Ile Cys Lys Gln Ala Leu Phe Gly Arg Phe Met Ser Leu Tyr Gly
```

260 265 270

Lys Leu Ser Ser Leu Thr Gly Gln Ser Val Thr Thr Arg Pro Thr His
275 280 285

Tyr Ser Asp Ala Lys Ala Ala Val Met Glu Tyr Gln Leu Ala Lys Gln
290 295 300

Cys Val Phe Gln Ala Phe Gln Lys Ala Gly Leu Gly Asn Trp Val Gln
305 310 315 320

Lys Pro Ile Glu Gln Asp Gln Phe
325

<210> 655
<211> 328
<212> PRT
<213> Sepia

<400> 655

Gly Ile Gly Thr Gly Thr Lys Cys Ile Asn Gly Glu Tyr Met Asn Asp
1 5 10 15

Arg Gly Phe Ala Val Asn Asp Cys His Ala Glu Ile Ile Ala Arg Arg
20 25 30

Cys Phe Leu Arg Phe Ile Tyr Asp Gln Leu Glu Met His Leu Ser Glu
35 40 45

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

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

Ala Pro Cys Gly Asp Ala Arg Ile Phe Ser Pro His Gly Gln Asp Val
85 90 95

Glu Thr Gly Asp Arg His Pro Asn Arg Lys Ala Arg Gly Gln Leu Arg
100 105 110

```
Thr Lys Ile Glu Ser Gly Gln Gly Thr Ile Pro Val Arg Ser Ser Gly
        115                 120                 125

Phe Ile Gln Thr Trp Asp Gly Val Leu Glu Gly Glu Arg Leu Leu Thr
    130                 135                 140

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

Gly Ala Leu Leu Cys His Phe Met His Pro Ile Tyr Leu Glu Ser Ile
                165                 170                 175

Ile Leu Gly Ser Leu Tyr His Ser Asp His Leu Ser Arg Ala Val Tyr
            180                 185                 190

Cys Arg Ile Ala Ser Ile Glu Asn Leu Pro Asp Leu Phe Gln Leu Asn
        195                 200                 205

Arg Pro Phe Leu Ser Gly Ile Ser Ser Pro Glu Ser Arg Gln Pro Gly
    210                 215                 220

Lys Ala Pro Asn Phe Gly Ile Asn Trp Arg Arg Asn Asp Asp Thr Phe
225                 230                 235                 240

Glu Val Ile Asn Ala Met Thr Gly Arg Val Glu Gly Gly Asn Met Ser
                245                 250                 255

Arg Ile Cys Lys Gln Ala Leu Phe Asp Arg Phe Met Asn Leu Tyr Gly
            260                 265                 270

Arg Leu Ser Ser Leu Thr Gly Gln Ser Val Thr Thr Arg Pro Thr Leu
        275                 280                 285

Tyr Ser Glu Ala Lys Ala Ala Val Met Glu Tyr Gln Leu Ala Lys Gln
    290                 295                 300

Cys Val Phe Gln Ala Phe Gln Lys Ala Gly Leu Gly Asn Trp Val Gln
305                 310                 315                 320
```

Lys Pro Ile Glu Gln Asp Gln Phe
325

<210> 656
<211> 328
<212> PRT
<213> Doryteusthis opalescens

<400> 656

Gly Ile Gly Thr Gly Thr Lys Cys Ile Asn Gly Glu Tyr Met Asn Asp
1 5 10 15

Arg Gly Phe Ala Val Asn Asp Cys His Ala Glu Ile Ile Ala Arg Arg
20 25 30

Cys Phe Leu Arg Phe Ile Tyr Asp Gln Leu Glu Leu His Leu Ser Asp
35 40 45

Asn Ala Glu Val Arg Gly Gln Ser Ile Phe Glu Leu Arg Asp Ala Gly
50 55 60

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

Ala Pro Cys Gly Asp Ala Arg Ile Phe Ser Pro His Gly Gln Asp Val
85 90 95

Glu Thr Gly Asp Arg His Pro Asn Arg Lys Ala Arg Gly Gln Leu Arg
100 105 110

Thr Lys Ile Glu Ser Gly Gln Gly Thr Ile Pro Val Arg Ser Ser Gly
115 120 125

Phe Ile Gln Thr Trp Asp Gly Val Leu Glu Gly Glu Arg Leu Leu Thr
130 135 140

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

```
Gly Ala Leu Leu Cys His Phe Met His Pro Ile Tyr Leu Glu Ser Ile
                165                 170                 175


Ile Leu Gly Ser Leu Tyr His Ser Asp His Leu Ser Arg Ala Val Tyr
            180                 185                 190


Cys Arg Ile Ala Ala Ile Glu Asn Leu Pro Asp Leu Phe Arg Leu Asn
        195                 200                 205


Arg Pro Phe Leu Ser Gly Ile Ser Ser Pro Glu Ser Arg Gln Pro Gly
    210                 215                 220


Lys Ala Pro Asn Phe Gly Ile Asn Trp Arg Arg Asn Asp Asp Ser Phe
225                 230                 235                 240


Glu Val Ile Asn Ala Met Thr Gly Arg Val Glu Gly Gly Ser Met Ser
                245                 250                 255


Arg Ile Cys Lys Gln Ala Leu Phe Asp Arg Phe Met Asn Leu Tyr Gly
            260                 265                 270


Lys Leu Ser Ser Leu Thr Gly Gln Ser Val Thr Thr Arg Pro Ala Leu
        275                 280                 285


Tyr Ser Glu Ala Lys Ala Thr Val Met Glu Tyr Gln Leu Ala Lys Gln
    290                 295                 300


Cys Val Phe Gln Ala Phe Gln Lys Ala Gly Leu Gly Asn Trp Val Gln
305                 310                 315                 320


Lys Pro Ile Glu Gln Asp Gln Phe
                325
```

<210> 657
<211> 57
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 657
gaatgcagag tagaggccgc aggatagttt agaacattct atggaaagag attccag 57

<210> 658
<211> 88
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 658
ttgccgaatg tgtcagtatg taattgaatg cagagtagag gccgcaggat agtttagaac 60

attctatgga aagagattcc agttccag 88

<210> 659
<211> 88
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 659
aacggcttac acagtcatac attaacttac gtctcatctc cggcgtccta tcaaatcttg 60

taagatacct ttctctaagg tcaaggtc 88

<210> 660
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 660
cttacgtctc atctccggcg tcctatcaaa tcttgtaaga tacctttctc t 51

<210> 661
<211> 19
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 661
augacaugga cugauuugc 19

<210> 662
<211> 19
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> modified_base
<222> (6)..(6)
<223> n at position 6 represents inosine

<220>
<221> misc_feature
<222> (6)..(6)
<223> n is a, c, g, or u

<400> 662
augacnugga cugauuugc 19

<210> 663
<211> 19
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 663
uacugcaccg gacgaaacg 19

<210> 664
<211> 19
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 664
uacugcaccg gacgaaacg 19

<210> 665
<211> 19
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 665
augacaugga cugauuugc 19

<210> 666
<211> 19
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> modified_base
<222> (6)..(6)
<223> n at position 6 represensts inosine

<220>
<221> misc_feature
<222> (6)..(6)
<223> n is a, c, g, or u

<400> 666
augacnugga cugauuugc 19

<210> 667
<211> 19
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 667
uacugcacca gacaaaacg 19

<210> 668
<211> 19
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 668
uacugcacca gacaaaacg 19

<210> 669
<211> 19
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 669
augacaugga cugauuugc 19

<210> 670
<211> 19
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> modified_base
<222> (6)..(6)
<223> n at position 6 represents inosine

<220>
<221> misc_feature
<222> (6)..(6)
<223> n is a, c, g, or u

<220>
<221> modified_base
<222> (14)..(14)
<223> n at position 6 represents inosine

<220>
<221> misc_feature
<222> (14)..(14)
<223> n is a, c, g, or u

<400> 670
augacnugga cugnuuugc 19

<210> 671
<211> 19
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 671
uacugcaccg gaccaaacg 19

<210> 672
<211> 19
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 672
uacugcaccg gaccaaacg 19

<210> 673
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 673
tggaggtgct caaagagatg gaagccaatg cccgga 36

<210> 674
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 674
agagatggaa gccaatgccc ggaaagctgg ctgc 34

<210> 675
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a, c, g, or t

<400> 675
nggaggtgct caaagagatg gaagccaatg cccgga 36

<210> 676
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (2)..(2)
<223> n is a, c, g, or t

<400> 676
tngaggtgct caaagagatg gaagccaatg cccgga 36

<210> 677
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (3)..(3)
<223> n is a, c, g, or t

<400> 677
tgnaggtgct caaagagatg gaagccaatg cccgga 36

<210> 678
<211> 36

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (4)..(4)
<223> n is a, c, g, or t

<400> 678
tggnggtgct caaagagatg gaagccaatg cccgga 36

<210> 679
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (4)..(4)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (36)..(36)
<223> n is a, c, g, or t

<400> 679
tggnggtgct caaagagatg gaagccaatg cccggn 36

<210> 680
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a, c, g, or t

<400> 680
ngagatggaa gccaatgccc ggaaagctgg ctgc 34

<210> 681
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (2)..(2)
<223> n is a, c, g, or t

<400> 681
anagatggaa gccaatgccc ggaaagctgg ctgc 34

<210> 682
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (3)..(3)
<223> n is a, c, g, or t

<400> 682
agngatggaa gccaatgccc ggaaagctgg ctgc 34

<210> 683
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature

<222> (4)..(4)
<223> n is a, c, g, or t

<400> 683
aganatggaa gccaatgccc ggaaagctgg ctgc 34

<210> 684
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (34)..(34)
<223> n is a, c, g, or t

<400> 684
agagatggaa gccaatgccc ggaaagctgg ctgn 34

<210> 685
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (1)..(2)
<223> n is a, c, g, or t

<400> 685
nngaggtgct caaagagatg gaagccaatg cccgga 36

<210> 686
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (3)..(3)
<223> n is a, c, g, or t

<400> 686
ngnaggtgct caaagagatg gaagccaatg cccgga 36

<210> 687
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (4)..(4)
<223> n is a, c, g, or t

<400> 687
nggnggtgct caaagagatg gaagccaatg cccgga 36

<210> 688
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a, c, g, or t

<220>

<221> misc_feature
<222> (5)..(5)
<223> n is a, c, g, or t

<400> 688
nggangtgct caaagagatg gaagccaatg cccgga 36

<210> 689
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (35)..(36)
<223> n is a, c, g, or t

<400> 689
tggaggtgct caaagagatg gaagccaatg cccgnn 36

<210> 690
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (1)..(2)
<223> n is a, c, g, or t

<400> 690
nnagatggaa gccaatgccc ggaaagctgg ctgc 34

<210> 691
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (3)..(3)
<223> n is a, c, g, or t

<400> 691
ngngatggaa gccaatgccc ggaaagctgg ctgc 34

<210> 692
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (4)..(4)
<223> n is a, c, g, or t

<400> 692
nganatggaa gccaatgccc ggaaagctgg ctgc 34

<210> 693
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (5)..(5)
<223> n is a, c, g, or t

<400> 693
ngagntggaa gccaatgccc ggaaagctgg ctgc 34

<210> 694
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (33)..(34)
<223> n is a, c, g, or t

<400> 694
agagatggaa gccaatgccc ggaaagctgg ctnn 34

<210> 695
<211> 55
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 695
aaugcagagu agaggccgca ggauaguuua gaacauucua uggaaagaga uucca 55

<210> 696
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 696
cgagcctcac acgtgctgct tgactacagg gagacgtgcg ctgctccc 48

<210> 697

<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 697

Arg Ala Ser His Val Leu Leu Asp Tyr Arg Glu Thr Cys Ala Ala Pro
1 5 10 15

<210> 698
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 698
cgagcctcac acgtggaaag ggactacagg gagacgtgcg ctgctccc 48

<210> 699
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 699
cgagcctcac acaagaccct tgactacagg gagacgtgcg ctgctccc 48

<210> 700
<211> 45
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 700
auugaaugca gaguagaggc cgcaggauag uuuagaacau ucuau 45

<210> 701
<211> 66
<212> RNA

<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 701
ucuaaaccau ccugcggccu cuacucugca guuguggaag guccaguuuu gggggcuauu 60

acaaca 66

<210> 702
<211> 105
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 702
tgccgaaugu gucaguaugu aauugaaugc agaguagagg ccgcaggaua guuuagaaca 60

uucuauggaa agagauucca guuccaggaa ccugguacau acgug 105

<210> 703
<211> 28
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 703

Cys Arg Met Cys Gln Tyr Val Ile Glu Cys Arg Val Glu Ala Ala Gly
1 5 10 15

Phe Arg Thr Phe Tyr Gly Lys Arg Phe Gln Phe Gln
20 25

<210> 704

<400> 704
000

<210> 705
<211> 65
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 705
ucgucuaugu gcugugcugg gcacccuucu uccuggugca gcuguaggcc gcgugggacc 60

cggag 65

<210> 706
<211> 50
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 706
gucccacgcg gcccacagcu gcaccaggaa gaagggugcc cagcacagca 50

<210> 707
<211> 50
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 707
gggucccacg cggcccacag cugcaccagg aagaagggug cccagcacag 50

<210> 708
<211> 50
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 708
ccggguccca cgcggcccac agcugcacca ggaagaaggg ugcccagcac 50

<210> 709
<211> 65
<212> RNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 709
ccucaacuuc cugcucuggg cuccuggagg ccacacgauc gccuaggaug ucaucacccu 60

gaugg 65

<210> 710
<211> 50
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 710
ggugaugaca ucccaggcga ucguguggcc uccaggagcc cagagcagga 50

<210> 711
<211> 50
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 711
agggugauga caucccaggc gaucgugugg ccuccaggag cccagagcag 50

<210> 712
<211> 50
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 712
aucaggguga ugacauccca ggcgaucgug uggccuccag gagcccagag 50

<210> 713
<211> 99
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 713
cccaggugcg ggagagaggc cugcugaaaa ugacugaaua uaaacuugug guaguuggag 60

cugguggcgu aggcaagagu gccuugacga uacagcuaa 99

<210> 714
<211> 50
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 714
caccacaagu uuauauucag ucauuuucag caggccucuc ucccgcaccu 50

<210> 715
<211> 50
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 715
gcuccaacca ccacaaguuu auauucaguc auuuucagca ggccucucuc 50

<210> 716
<211> 50
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 716
cgccaccagc uccaaccacc acaaguuuau auucagucau uuucagcagg 50

<210> 717
<211> 50
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 717
cuugccuacg ccaccagcuc caaccaccac aaguuuauau ucagucauuu 50

<210> 718
<211> 50
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 718
aaggcacucu ugccuacgcc accagcucca accaccacaa guuuauauuc 50

<210> 719
<211> 98
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 719
guacgucaag gcacucuugc cuacgccacc agcuccaacc accacaagug uacgucaagg 60

cacucuugcc uacgccacca gcuccaacca ccacaagu 98

<210> 720
<211> 50
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 720
aauuagcugu aucgucaagg cacucuugcc uacgccacca gcuccaacca 50

<210> 721
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 721
tatgtaattg aatgcagagt agaggccgca ggatagttta gaac 44

<210> 722
<211> 39
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (4)..(5)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (36)..(39)
<223> n is a, c, g, or u

<400> 722
augnnuugcc accgacgaca ccgugaggau aaauunnnn 39

<210> 723
<211> 36
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 723
uggaggugcu caaagagaug gaagccaaug cccgga 36

<210> 724
<211> 36
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a, c, g, or u

<400> 724
nggaggugcu caaagagaug gaagccaaug cccgga 36

<210> 725
<211> 36
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (2)..(2)
<223> n is a, c, g, or u

<400> 725
ungaggugcu caaagagaug gaagccaaug cccgga 36

<210> 726
<211> 36
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (3)..(3)
<223> n is a, c, g, or u

<400> 726
ugnaggugcu caaagagaug gaagccaaug cccgga 36

<210> 727
<211> 36
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (4)..(4)
<223> n is a, c, g, or u

<400> 727

uggnggugcu caaagagaug gaagccaaug cccgga 36

<210> 728
<211> 36
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (18)..(18)
<223> n is a, c, g, or u

<400> 728
uggaggugcu caaagagnug gaagccaaug cccgga 36

<210> 729
<211> 36
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (36)..(36)
<223> n is a, c, g, or u

<400> 729
uggaggugcu caaagagaug gaagccaaug cccggn 36

<210> 730
<211> 36
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (1)..(2)
<223> n is a, c, g, or u

<400> 730
nngaggugcu caaagagaug gaagccaaug cccgga 36

<210> 731
<211> 36
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (3)..(3)
<223> n is a, c, g, or u

<400> 731
ngnaggugcu caaagagaug gaagccaaug cccgga 36

<210> 732
<211> 36
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (4)..(4)
<223> n is a, c, g, or u

<400> 732
nggnggugcu caaagagaug gaagccaaug cccgga 36

<210> 733

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

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (5)..(5)
<223> n is a, c, g, or u

<400> 733
nggangugcu caaagagaug gaagccaaug cccgga 36

<210> 734
<211> 36
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (14)..(14)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (18)..(18)
<223> n is a, c, g, or u

<400> 734
uggaggugcu caangagnug gaagccaaug cccgga 36

<210> 735
<211> 36
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (35)..(36)
<223> n is a, c, g, or u

<400> 735
uggaggugcu caaagagaug gaagccaaug cccgnn 36

<210> 736
<211> 34
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 736
agagauggaa gccaaugccc ggaaagcugg cugc 34

<210> 737
<211> 107
<212> DNA
<213> Cypridinia luciferase

<400> 737
tgccgaaugu gucaguaugu aauugaaugc agaguagagg ccgcaggaua guuuagaaca 60

uucuauggaa agagauucca guuccaggaa ccugguacau acguguu 107

<210> 738
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 738
cauccugcgg ccucuacucu gcauucaauu 30

<210> 739
<211> 30
<212> RNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 739
aaaccauccu gcggccucua cucugcauuc 30

<210> 740
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 740
uucuaaacca uccugcggcc ucuacucugc 30

<210> 741
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 741
aauguucuaa accauccugc ggccucuacu 30

<210> 742
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 742
auagaauguu cuaaaccauc cugcggccuc 30

<210> 743
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 743
uuccauagaa uguucuaaac cauccugcgg 30

<210> 744
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 744
cucuuuccau agaauguucu aaaccauccu 30

<210> 745
<211> 30
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 745
gaaucucuuu ccauagaaug uucuaaacca 30

<210> 746
<211> 104
<212> RNA
<213> Homo sapiens

<400> 746
uuggucucuu cggaaagacu guuccaaaaa caguggauaa uuuuguggcc uuagcuacag 60

gagagaaagg auuuggcuac aaaaacagca aauuccaucg ugua 104

<210> 747
<211> 50
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 747
caaggccaca aaauuaucca cuguuuuugg aacagucuuu ccgaagagac 50

<210> 748
<211> 50
<212> RNA

<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 748
ccuguagcca aggccacaaa auuauccacu guuuuuggaa cagucuuucc 50

<210> 749
<211> 50
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 749
cuuucucucc uguagccaag gccacaaaau uauccacugu uuuuggaaca 50

<210> 750
<211> 50
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 750
gccaaauccu uucucuccug uagccaaggc cacaaaauua uccacuguuu 50

<210> 751
<211> 50
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 751
uuuuuguagc caaauccuuu cucuccugua gccaaggcca caaaauuauc 50

<210> 752
<211> 50
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 752
auuugcuguu uuuguagcca aauccuuucu cuccuguagc caaggccaca 50

<210> 753
<211> 50
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 753
acgauggaau uugcuguuuu uguagccaaa uccuuucucu ccuguagcca 50

<210> 754
<211> 50
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (17)..(17)
<223> n is a, c, g, or u

<400> 754
auagaauguu cuaaacnauc cugcggccuc uacucugcau ucaauuacau 50

<210> 755
<211> 50
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<220>
<221> misc_feature
<222> (17)..(17)
<223> n is a, c, g, or u

<400> 755
auagaauguu cuaaacnauc cugcggccuc uacucugcau ucaauuacau 50

<210> 756
<211> 59
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 756
auguaauuga augcagagua gaggccgcag gauaguuuag aacauucuau ggaaagaga 59

<210> 757
<211> 83
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 757
ucuuuccaua ggcccugaaa aagggccugu ucuaaaccau ccugcggccu cuacucggcc 60

cugaaaaagg gccauucaau uac 83

<210> 758
<211> 21
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 758
ggccgcagga uaguuuagaa c 21

<210> 759
<211> 52
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 759
aaauguuguu auaguauccc accucuaaac cauccugcgg ggccgggccc uu 52

<210> 760

<211> 26
<212> RNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 760
accgguggaa uaguauaaca auaugc 26

<210> 761
<211> 74
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 761

Gly Cys Cys Ala Gly Ala Gly Ala Thr Gly Ala Ala Gly Gly Ala Cys
1 5 10 15

Cys Ala Gly Gly Gly Thr Gly Gly Cys Thr Cys Ala Gly Gly Gly Gly
20 25 30

Ala Ala Gly Gly Gly Cys Thr Gly Ala Gly Cys Cys Gly Cys Ala Gly
35 40 45

Cys Gly Cys Cys Ala Ala Thr Gly Gly Ala Thr Thr Thr Cys Cys Ala
50 55 60

Thr Ala Cys Ala Gly Ala Gly Ala Gly Gly
65 70

<210> 762
<211> 668
<212> PRT
<213> Fusobacterium necrophorum

<400> 762

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

```
Ile Val Lys Lys Asp Ile Phe Ser Gln Thr Pro Asp Asn Met Ile Arg
            20                  25                  30

Asn Tyr Lys Ile Thr Leu Lys Ile Ser Glu Lys Asn Pro Arg Val Val
        35                  40                  45

Glu Ala Glu Ile Glu Asp Leu Met Asn Ser Thr Ile Leu Lys Asp Gly
    50                  55                  60

Arg Arg Ser Ala Arg Arg Glu Lys Ser Met Thr Glu Arg Lys Leu Ile
65                  70                  75                  80

Glu Glu Lys Val Ala Glu Asn Tyr Ser Leu Leu Ala Asn Cys Pro Met
                85                  90                  95

Glu Glu Val Asp Ser Ile Lys Ile Tyr Lys Ile Lys Arg Phe Leu Thr
            100                 105                 110

Tyr Arg Ser Asn Met Leu Leu Tyr Phe Ala Ser Ile Asn Ser Phe Leu
        115                 120                 125

Cys Glu Gly Ile Lys Gly Lys Asp Asn Glu Thr Glu Glu Ile Trp His
    130                 135                 140

Leu Lys Asp Asn Asp Val Arg Lys Glu Lys Val Lys Glu Asn Phe Lys
145                 150                 155                 160

Asn Lys Leu Ile Gln Ser Thr Glu Asn Tyr Asn Ser Ser Leu Lys Asn
                165                 170                 175

Gln Ile Glu Glu Lys Glu Lys Leu Leu Arg Lys Glu Ser Lys Lys Gly
            180                 185                 190

Ala Phe Tyr Arg Thr Ile Ile Lys Lys Leu Gln Gln Glu Arg Ile Lys
        195                 200                 205

Glu Leu Ser Glu Lys Ser Leu Thr Glu Asp Cys Glu Lys Ile Ile Lys
    210                 215                 220
```

```
Leu Tyr Ser Glu Leu Arg His Pro Leu Met His Tyr Asp Tyr Gln Tyr
225                 230                 235                 240


Phe Glu Asn Leu Phe Glu Asn Lys Glu Asn Ser Glu Leu Thr Lys Asn
                245                 250                 255


Leu Asn Leu Asp Ile Phe Lys Ser Leu Pro Leu Val Arg Lys Met Lys
            260                 265                 270


Leu Asn Asn Lys Val Asn Tyr Leu Glu Asp Asn Asp Thr Leu Phe Val
        275                 280                 285


Leu Gln Lys Thr Lys Lys Ala Lys Thr Leu Tyr Gln Ile Tyr Asp Ala
    290                 295                 300


Leu Cys Glu Gln Lys Asn Gly Phe Asn Lys Phe Ile Asn Asp Phe Phe
305                 310                 315                 320


Val Ser Asp Gly Glu Glu Asn Thr Val Phe Lys Gln Ile Ile Asn Glu
                325                 330                 335


Lys Phe Gln Ser Glu Met Glu Phe Leu Glu Lys Arg Ile Ser Glu Ser
            340                 345                 350


Glu Lys Lys Asn Glu Lys Leu Lys Lys Lys Phe Asp Ser Met Lys Ala
        355                 360                 365


His Phe His Asn Ile Asn Ser Glu Asp Thr Lys Glu Ala Tyr Phe Trp
    370                 375                 380


Asp Ile His Ser Ser Ser Asn Tyr Lys Thr Lys Tyr Asn Glu Arg Lys
385                 390                 395                 400


Asn Leu Val Asn Glu Tyr Thr Glu Leu Leu Gly Ser Ser Lys Glu Lys
                405                 410                 415


Lys Leu Leu Arg Glu Glu Ile Thr Gln Ile Asn Arg Lys Leu Leu Lys
            420                 425                 430
```

```
Leu Lys Gln Glu Met Glu Glu Ile Thr Lys Lys Asn Ser Leu Phe Arg
        435                 440                 445


Leu Glu Tyr Lys Met Lys Ile Ala Phe Gly Phe Leu Phe Cys Glu Phe
    450                 455                 460


Asp Gly Asn Ile Ser Lys Phe Lys Asp Glu Phe Asp Ala Ser Asn Gln
465                 470                 475                 480


Glu Lys Ile Ile Gln Tyr His Lys Asn Gly Glu Lys Tyr Leu Thr Tyr
                485                 490                 495


Phe Leu Lys Glu Glu Glu Lys Glu Lys Phe Asn Leu Glu Lys Met Gln
            500                 505                 510


Lys Ile Ile Gln Lys Thr Glu Glu Glu Asp Trp Leu Leu Pro Glu Thr
        515                 520                 525


Lys Asn Asn Leu Phe Lys Phe Tyr Leu Leu Thr Tyr Leu Leu Leu Pro
    530                 535                 540


Tyr Glu Leu Lys Gly Asp Phe Leu Gly Phe Val Lys Lys His Tyr Tyr
545                 550                 555                 560


Asp Ile Lys Asn Val Asp Phe Met Asp Glu Asn Gln Asn Asn Ile Gln
                565                 570                 575


Val Ser Gln Thr Val Glu Lys Gln Glu Asp Tyr Phe Tyr His Lys Ile
            580                 585                 590


Arg Leu Phe Glu Lys Asn Thr Lys Lys Tyr Glu Ile Val Lys Tyr Ser
        595                 600                 605


Ile Val Pro Asn Glu Lys Leu Lys Gln Tyr Phe Glu Asp Leu Gly Ile
    610                 615                 620


Asp Ile Lys Tyr Leu Thr Gly Ser Val Glu Ser Gly Glu Lys Trp Leu
625                 630                 635                 640
```

```
Gly Glu Asn Leu Gly Ile Asp Ile Lys Tyr Leu Thr Val Glu Gln Lys
                645                 650                 655


Ser Glu Val Ser Glu Glu Lys Ile Lys Lys Phe Leu
            660                 665


<210>  763
<211>  796
<212>  PRT
<213>  Tissierella sp.

<400>  763

Met Glu Lys Asp Lys Lys Gly Glu Lys Ile Asp Ile Ser Gln Glu Met
1               5                   10                  15


Ile Glu Glu Asp Leu Arg Lys Ile Leu Ile Leu Phe Ser Arg Leu Arg
            20                  25                  30


His Ser Met Val His Tyr Asp Tyr Glu Phe Tyr Gln Ala Leu Tyr Ser
        35                  40                  45


Gly Lys Asp Phe Val Ile Ser Asp Lys Asn Asn Leu Glu Asn Arg Met
    50                  55                  60


Ile Ser Gln Leu Leu Asp Leu Asn Ile Phe Lys Glu Leu Ser Lys Val
65                  70                  75                  80


Lys Leu Ile Lys Asp Lys Ala Ile Ser Asn Tyr Leu Asp Lys Asn Thr
                85                  90                  95


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


Ile Tyr Arg Asp Ile Cys Gly Ser Lys Asn Gly Phe Asn Lys Phe Ile
        115                 120                 125


Asn Thr Met Ile Thr Ile Ser Gly Glu Glu Asp Arg Glu Tyr Lys Glu
    130                 135                 140


Lys Val Ile Glu His Phe Asn Lys Lys Met Glu Asn Leu Ser Thr Tyr
```

```
145                 150                 155                 160

Leu Glu Lys Leu Glu Lys Gln Asp Asn Ala Lys Arg Asn Asn Lys Arg
                165                 170                 175

Val Tyr Asn Leu Leu Lys Gln Lys Leu Ile Glu Gln Gln Lys Leu Lys
            180                 185                 190

Glu Trp Phe Gly Gly Pro Tyr Val Tyr Asp Ile His Ser Ser Lys Arg
        195                 200                 205

Tyr Lys Glu Leu Tyr Ile Glu Arg Lys Lys Leu Val Asp Arg His Ser
    210                 215                 220

Lys Leu Phe Glu Glu Gly Leu Asp Glu Lys Asn Lys Lys Glu Leu Thr
225                 230                 235                 240

Lys Ile Asn Asp Glu Leu Ser Lys Leu Asn Ser Glu Met Lys Glu Met
                245                 250                 255

Thr Lys Leu Asn Ser Lys Tyr Arg Leu Gln Tyr Lys Leu Gln Leu Ala
            260                 265                 270

Phe Gly Phe Ile Leu Glu Glu Phe Asp Leu Asn Ile Asp Thr Phe Ile
        275                 280                 285

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

Lys Arg Asp Ile Tyr Leu Asn Arg Val Leu Asp Arg Gly Asp Asn Arg
305                 310                 315                 320

Leu Lys Asn Ile Ile Lys Glu Tyr Lys Phe Arg Asp Thr Glu Asp Ile
                325                 330                 335

Phe Cys Asn Asp Arg Asp Asn Asn Leu Val Lys Leu Tyr Ile Leu Met
            340                 345                 350

Tyr Ile Leu Leu Pro Val Glu Ile Arg Gly Asp Phe Leu Gly Phe Val
```

```
        355                 360                 365

Lys Lys Asn Tyr Tyr Asp Met Lys His Val Asp Phe Ile Asp Lys Lys
    370                 375                 380

Asp Lys Glu Asp Lys Asp Thr Phe Phe His Asp Leu Arg Leu Phe Glu
385                 390                 395                 400

Lys Asn Ile Arg Lys Leu Glu Ile Thr Asp Tyr Ser Leu Ser Ser Gly
                405                 410                 415

Phe Leu Ser Lys Glu His Lys Val Asp Ile Glu Lys Lys Ile Asn Asp
            420                 425                 430

Phe Ile Asn Arg Asn Gly Ala Met Lys Leu Pro Glu Asp Ile Thr Ile
        435                 440                 445

Glu Glu Phe Asn Lys Ser Leu Ile Leu Pro Ile Met Lys Asn Tyr Gln
    450                 455                 460

Ile Asn Phe Lys Leu Leu Asn Asp Ile Glu Ile Ser Ala Leu Phe Lys
465                 470                 475                 480

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

Lys Asn Glu Asp Ile Lys Lys Asn Ser Lys Lys Asn Asp Lys Asn Asn
            500                 505                 510

His Lys Asp Lys Asn Ile Asn Phe Thr Gln Leu Met Lys Arg Ala Leu
        515                 520                 525

His Glu Lys Ile Pro Tyr Lys Ala Gly Met Tyr Gln Ile Arg Asn Asn
    530                 535                 540

Ile Ser His Ile Asp Met Glu Gln Leu Tyr Ile Asp Pro Leu Asn Ser
545                 550                 555                 560

Tyr Met Asn Ser Asn Lys Asn Asn Ile Thr Ile Ser Glu Gln Ile Glu
```

565 570 575

Lys Ile Ile Asp Val Cys Val Thr Gly Gly Val Thr Gly Lys Glu Leu
580 585 590

Asn Asn Asn Ile Ile Asn Asp Tyr Tyr Met Lys Lys Glu Lys Leu Val
595 600 605

Phe Asn Leu Lys Leu Arg Lys Gln Asn Asp Ile Val Ser Ile Glu Ser
610 615 620

Gln Glu Lys Asn Lys Arg Glu Glu Phe Val Phe Lys Lys Tyr Gly Leu
625 630 635 640

Asp Tyr Lys Asp Gly Glu Ile Asn Ile Ile Glu Val Ile Gln Lys Val
645 650 655

Asn Ser Leu Gln Glu Glu Leu Arg Asn Ile Lys Glu Thr Ser Lys Glu
660 665 670

Lys Leu Lys Asn Lys Glu Thr Leu Phe Arg Asp Ile Ser Leu Ile Asn
675 680 685

Gly Thr Ile Arg Lys Asn Ile Asn Phe Lys Ile Lys Glu Met Val Leu
690 695 700

Asp Ile Val Arg Met Asp Glu Ile Arg His Ile Asn Ile His Ile Tyr
705 710 715 720

Tyr Lys Gly Glu Asn Tyr Thr Arg Ser Asn Ile Ile Lys Phe Lys Tyr
725 730 735

Ala Ile Asp Gly Glu Asn Lys Lys Tyr Tyr Leu Lys Gln His Glu Ile
740 745 750

Asn Asp Ile Asn Leu Glu Leu Lys Asp Lys Phe Val Thr Leu Ile Cys
755 760 765

Asn Met Asp Lys His Pro Asn Lys Asn Lys Gln Thr Ile Asn Leu Glu

770 775 780

Ser Asn Tyr Ile Gln Asn Val Lys Phe Ile Ile Pro
785 790 795

<210> 764
<211> 897
<212> PRT
<213> Fusobacterium ulcerans

<400> 764

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

Asn Ile Leu Val Ala Gln Ile Lys Glu Tyr Phe Thr Lys Glu Ile Glu
20 25 30

Asn Tyr Asn Asn Arg Ile Asp Asn Ile Ile Asp Lys Lys Glu Leu Leu
35 40 45

Lys Tyr Ser Glu Lys Lys Glu Glu Ser Glu Lys Asn Lys Lys Leu Glu
50 55 60

Glu Leu Asn Lys Leu Lys Ser Gln Lys Leu Lys Ile Leu Thr Asp Glu
65 70 75 80

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

Arg His Ser Leu Met His Tyr Glu Tyr Lys Tyr Phe Glu Asn Leu Phe
100 105 110

Glu Asn Lys Lys Asn Glu Glu Leu Ala Glu Leu Leu Asn Leu Asn Leu
115 120 125

Phe Lys Asn Leu Thr Leu Leu Arg Gln Met Lys Ile Glu Asn Lys Thr
130 135 140

Asn Tyr Leu Glu Gly Arg Glu Glu Phe Asn Ile Ile Gly Lys Asn Ile
145 150 155 160

```
Lys Ala Lys Glu Val Leu Gly His Tyr Asn Leu Leu Ala Glu Gln Lys
                165                 170                 175

Asn Gly Phe Asn Asn Phe Ile Asn Ser Phe Phe Val Gln Asp Gly Thr
            180                 185                 190

Glu Asn Leu Glu Phe Lys Lys Leu Ile Asp Glu His Phe Val Asn Ala
        195                 200                 205

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

Glu Leu Glu Lys Met Glu Gln His Tyr Gln Arg Leu Asn Cys Ala Tyr
225                 230                 235                 240

Val Trp Asp Ile His Thr Ser Thr Thr Tyr Lys Lys Leu Tyr Asn Lys
                245                 250                 255

Arg Lys Ser Leu Ile Glu Glu Tyr Asn Lys Gln Ile Asn Glu Ile Lys
            260                 265                 270

Asp Lys Glu Val Ile Thr Ala Ile Asn Val Glu Leu Leu Arg Ile Lys
        275                 280                 285

Lys Glu Met Glu Glu Ile Thr Lys Ser Asn Ser Leu Phe Arg Leu Lys
    290                 295                 300

Tyr Lys Met Gln Ile Ala Tyr Ala Phe Leu Glu Ile Glu Phe Gly Gly
305                 310                 315                 320

Asn Ile Ala Lys Phe Lys Asp Glu Phe Asp Cys Ser Lys Met Glu Glu
                325                 330                 335

Val Gln Lys Tyr Leu Lys Lys Gly Val Lys Tyr Leu Lys Tyr Tyr Lys
            340                 345                 350

Asp Lys Glu Ala Gln Lys Asn Tyr Glu Phe Pro Phe Glu Glu Ile Phe
        355                 360                 365
```

```
Glu Asn Lys Asp Thr His Asn Glu Glu Trp Leu Glu Asn Thr Ser Glu
    370                 375                 380

Asn Asn Leu Phe Lys Phe Tyr Ile Leu Thr Tyr Leu Leu Leu Pro Met
385                 390                 395                 400

Glu Phe Lys Gly Asp Phe Leu Gly Val Val Lys Lys His Tyr Tyr Asp
                405                 410                 415

Ile Lys Asn Val Asp Phe Thr Asp Glu Ser Glu Lys Glu Leu Ser Gln
            420                 425                 430

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

Leu Phe Glu Lys Asn Thr Lys Arg Tyr Glu Ile Ile Lys Tyr Ser Ile
    450                 455                 460

Leu Thr Ser Asp Glu Ile Lys Arg Tyr Phe Arg Leu Leu Glu Leu Asp
465                 470                 475                 480

Val Pro Tyr Phe Glu Tyr Glu Lys Gly Thr Asp Glu Ile Gly Ile Phe
                485                 490                 495

Asn Lys Asn Ile Ile Leu Thr Ile Phe Lys Tyr Tyr Gln Ile Ile Phe
            500                 505                 510

Arg Leu Tyr Asn Asp Leu Glu Ile His Gly Leu Phe Asn Ile Ser Ser
        515                 520                 525

Asp Leu Asp Lys Ile Leu Arg Asp Leu Lys Ser Tyr Gly Asn Lys Asn
    530                 535                 540

Ile Asn Phe Arg Glu Phe Leu Tyr Val Ile Lys Gln Asn Asn Asn Ser
545                 550                 555                 560

Ser Thr Glu Glu Glu Tyr Arg Lys Ile Trp Glu Asn Leu Glu Ala Lys
                565                 570                 575
```

```
Tyr Leu Arg Leu His Leu Leu Thr Pro Glu Lys Glu Glu Ile Lys Thr
            580                 585                 590

Lys Thr Lys Glu Glu Leu Glu Lys Leu Asn Glu Ile Ser Asn Leu Arg
        595                 600                 605

Asn Gly Ile Cys His Leu Asn Tyr Lys Glu Ile Ile Glu Glu Ile Leu
    610                 615                 620

Lys Thr Glu Ile Ser Glu Lys Asn Lys Glu Ala Thr Leu Asn Glu Lys
625                 630                 635                 640

Ile Arg Lys Val Ile Asn Phe Ile Lys Glu Asn Glu Leu Asp Lys Val
                645                 650                 655

Glu Leu Gly Phe Asn Phe Ile Asn Asp Phe Phe Met Lys Lys Glu Gln
            660                 665                 670

Phe Met Phe Gly Gln Ile Lys Gln Val Lys Glu Gly Asn Ser Asp Ser
        675                 680                 685

Ile Thr Thr Glu Arg Glu Arg Lys Glu Lys Asn Asn Lys Lys Leu Lys
    690                 695                 700

Glu Thr Tyr Glu Leu Asn Cys Asp Asn Leu Ser Glu Phe Tyr Glu Thr
705                 710                 715                 720

Ser Asn Asn Leu Arg Glu Arg Ala Asn Ser Ser Ser Leu Leu Glu Asp
                725                 730                 735

Ser Ala Phe Leu Lys Lys Ile Gly Leu Tyr Lys Val Lys Asn Asn Lys
            740                 745                 750

Val Asn Ser Lys Val Lys Asp Glu Glu Lys Arg Ile Glu Asn Ile Lys
        755                 760                 765

Arg Lys Leu Leu Lys Asp Ser Ser Asp Ile Met Gly Met Tyr Lys Ala
    770                 775                 780
```

```
Glu Val Val Lys Lys Leu Lys Glu Lys Leu Ile Leu Ile Phe Lys His
785                 790                 795                 800


Asp Glu Glu Lys Arg Ile Tyr Val Thr Val Tyr Asp Thr Ser Lys Ala
                805                 810                 815


Val Pro Glu Asn Ile Ser Lys Glu Ile Leu Val Lys Arg Asn Asn Ser
            820                 825                 830


Lys Glu Glu Tyr Phe Phe Glu Asp Asn Asn Lys Lys Tyr Val Thr Glu
        835                 840                 845


Tyr Tyr Thr Leu Glu Ile Thr Glu Thr Asn Glu Leu Lys Val Ile Pro
    850                 855                 860


Ala Lys Lys Leu Glu Gly Lys Glu Phe Lys Thr Glu Lys Asn Lys Glu
865                 870                 875                 880


Asn Lys Leu Met Leu Asn Asn His Tyr Cys Phe Asn Val Lys Ile Ile
                885                 890                 895


Tyr
```

<210> 765
<211> 919
<212> PRT
<213> Cetobacterium sp. ZOR0034

<400> 765

```
Met Glu Glu Ile Lys His Lys Lys Asn Lys Ser Ser Ile Ile Arg Val
1               5                   10                  15


Ile Val Ser Asn Tyr Asp Met Thr Gly Ile Lys Glu Ile Lys Val Leu
            20                  25                  30


Tyr Gln Lys Gln Gly Gly Val Asp Thr Phe Asn Leu Lys Thr Ile Ile
        35                  40                  45
```

```
Asn Leu Glu Ser Gly Asn Leu Glu Ile Ile Ser Cys Lys Pro Lys Glu
    50                  55                  60

Arg Glu Lys Tyr Arg Tyr Glu Phe Asn Cys Lys Thr Glu Ile Asn Thr
65                  70                  75                  80

Ile Ser Ile Thr Lys Lys Asp Lys Val Leu Lys Lys Glu Ile Arg Lys
                85                  90                  95

Tyr Ser Leu Glu Leu Tyr Phe Lys Asn Glu Lys Lys Asp Thr Val Val
            100                 105                 110

Ala Lys Val Thr Asp Leu Leu Lys Ala Pro Asp Lys Ile Glu Gly Glu
        115                 120                 125

Arg Asn His Leu Arg Lys Leu Ser Ser Ser Thr Glu Arg Lys Leu Leu
    130                 135                 140

Ser Lys Thr Leu Cys Lys Asn Tyr Ser Glu Ile Ser Lys Thr Pro Ile
145                 150                 155                 160

Glu Glu Ile Asp Ser Ile Lys Ile Tyr Lys Ile Lys Arg Phe Leu Asn
                165                 170                 175

Tyr Arg Ser Asn Phe Leu Ile Tyr Phe Ala Leu Ile Asn Asp Phe Leu
            180                 185                 190

Cys Ala Gly Val Lys Glu Asp Asp Ile Asn Glu Val Trp Leu Ile Gln
        195                 200                 205

Asp Lys Glu His Thr Ala Phe Leu Glu Asn Arg Ile Glu Lys Ile Thr
    210                 215                 220

Asp Tyr Ile Phe Asp Lys Leu Ser Lys Asp Ile Glu Asn Lys Lys Asn
225                 230                 235                 240

Gln Phe Glu Lys Arg Ile Lys Lys Tyr Lys Thr Ser Leu Glu Glu Leu
                245                 250                 255
```

```
Lys Thr Glu Thr Leu Glu Lys Asn Lys Thr Phe Tyr Ile Asp Ser Ile
            260                 265                 270

Lys Thr Lys Ile Thr Asn Leu Glu Asn Lys Ile Thr Glu Leu Ser Leu
        275                 280                 285

Tyr Asn Ser Lys Glu Ser Leu Lys Glu Asp Leu Ile Lys Ile Ile Ser
    290                 295                 300

Ile Phe Thr Asn Leu Arg His Ser Leu Met His Tyr Asp Tyr Lys Ser
305                 310                 315                 320

Phe Glu Asn Leu Phe Glu Asn Ile Glu Asn Glu Glu Leu Lys Asn Leu
                325                 330                 335

Leu Asp Leu Asn Leu Phe Lys Ser Ile Arg Met Ser Asp Glu Phe Lys
            340                 345                 350

Thr Lys Asn Arg Thr Asn Tyr Leu Asp Gly Thr Glu Ser Phe Thr Ile
        355                 360                 365

Val Lys Lys His Gln Asn Leu Lys Lys Leu Tyr Thr Tyr Tyr Asn Asn
    370                 375                 380

Leu Cys Asp Lys Lys Asn Gly Phe Asn Thr Phe Ile Asn Ser Phe Phe
385                 390                 395                 400

Val Thr Asp Gly Ile Glu Asn Thr Asp Phe Lys Asn Leu Ile Ile Leu
                405                 410                 415

His Phe Glu Lys Glu Met Glu Glu Tyr Lys Lys Ser Ile Glu Tyr Tyr
            420                 425                 430

Lys Ile Lys Ile Ser Asn Glu Lys Asn Lys Ser Lys Lys Glu Lys Leu
        435                 440                 445

Lys Glu Lys Ile Asp Leu Leu Gln Ser Glu Leu Ile Asn Met Arg Glu
    450                 455                 460
```

```
His Lys Asn Leu Leu Lys Gln Ile Tyr Phe Phe Asp Ile His Asn Ser
465                 470                 475                 480

Ile Lys Tyr Lys Glu Leu Tyr Ser Glu Arg Lys Asn Leu Ile Glu Gln
                485                 490                 495

Tyr Asn Leu Gln Ile Asn Gly Val Lys Asp Val Thr Ala Ile Asn His
            500                 505                 510

Ile Asn Thr Lys Leu Leu Ser Leu Lys Asn Lys Met Asp Lys Ile Thr
        515                 520                 525

Lys Gln Asn Ser Leu Tyr Arg Leu Lys Tyr Lys Leu Lys Ile Ala Tyr
    530                 535                 540

Ser Phe Leu Met Ile Glu Phe Asp Gly Asp Val Ser Lys Phe Lys Asn
545                 550                 555                 560

Asn Phe Asp Pro Thr Asn Leu Glu Lys Arg Val Glu Tyr Leu Asp Lys
                565                 570                 575

Lys Glu Glu Tyr Leu Asn Tyr Thr Ala Pro Lys Asn Lys Phe Asn Phe
            580                 585                 590

Ala Lys Leu Glu Glu Glu Leu Gln Lys Ile Gln Ser Thr Ser Glu Met
        595                 600                 605

Gly Ala Asp Tyr Leu Asn Val Ser Pro Glu Asn Asn Leu Phe Lys Phe
    610                 615                 620

Tyr Ile Leu Thr Tyr Ile Met Leu Pro Val Glu Phe Lys Gly Asp Phe
625                 630                 635                 640

Leu Gly Phe Val Lys Asn His Tyr Tyr Asn Ile Lys Asn Val Asp Phe
                645                 650                 655

Met Asp Glu Ser Leu Leu Asp Glu Asn Glu Val Asp Ser Asn Lys Leu
            660                 665                 670
```

```
Asn Glu Lys Ile Glu Asn Leu Lys Asp Ser Ser Phe Phe Asn Lys Ile
        675                 680                 685

Arg Leu Phe Glu Lys Asn Ile Lys Lys Tyr Glu Ile Val Lys Tyr Ser
    690                 695                 700

Val Ser Thr Gln Glu Asn Met Lys Glu Tyr Phe Lys Gln Leu Asn Leu
705                 710                 715                 720

Asp Ile Pro Tyr Leu Asp Tyr Lys Ser Thr Asp Glu Ile Gly Ile Phe
                725                 730                 735

Asn Lys Asn Met Ile Leu Pro Ile Phe Lys Tyr Tyr Gln Asn Val Phe
            740                 745                 750

Lys Leu Cys Asn Asp Ile Glu Ile His Ala Leu Leu Ala Leu Ala Asn
        755                 760                 765

Lys Lys Gln Gln Asn Leu Glu Tyr Ala Ile Tyr Cys Cys Ser Lys Lys
    770                 775                 780

Asn Ser Leu Asn Tyr Asn Glu Leu Leu Lys Thr Phe Asn Arg Lys Thr
785                 790                 795                 800

Tyr Gln Asn Leu Ser Phe Ile Arg Asn Lys Ile Ala His Leu Asn Tyr
                805                 810                 815

Lys Glu Leu Phe Ser Asp Leu Phe Asn Asn Glu Leu Asp Leu Asn Thr
            820                 825                 830

Lys Val Arg Cys Leu Ile Glu Phe Ser Gln Asn Asn Lys Phe Asp Gln
        835                 840                 845

Ile Asp Leu Gly Met Asn Phe Ile Asn Asp Tyr Tyr Met Lys Lys Thr
    850                 855                 860

Arg Phe Ile Phe Asn Gln Arg Arg Leu Arg Asp Leu Asn Val Pro Ser
865                 870                 875                 880
```

```
Lys Glu Lys Ile Ile Asp Gly Lys Arg Lys Gln Gln Asn Asp Ser Asn
                885                 890                 895


Asn Glu Leu Leu Lys Lys Tyr Gly Leu Ser Arg Thr Asn Ile Lys Asp
            900                 905                 910


Ile Phe Asn Lys Ala Trp Tyr
        915


<210>  766
<211>  1110
<212>  PRT
<213>  Fusobacterium necrophorum

<400>  766

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


Lys Thr Glu Ile Val Asn Asn Asp Ile Phe Ile Lys Ser Ile Ile Glu
            20                  25                  30


Lys Ala Arg Glu Lys Tyr Arg Tyr Ser Phe Leu Phe Asp Gly Glu Glu
        35                  40                  45


Lys Tyr His Phe Lys Asn Lys Ser Ser Val Glu Ile Val Lys Asn Asp
    50                  55                  60


Ile Phe Ser Gln Thr Pro Asp Asn Met Ile Arg Asn Tyr Lys Ile Thr
65                  70                  75                  80


Leu Lys Ile Ser Glu Lys Asn Pro Arg Val Val Glu Ala Glu Ile Glu
                85                  90                  95


Asp Leu Met Asn Ser Thr Ile Leu Lys Asp Gly Arg Arg Ser Ala Arg
            100                 105                 110


Arg Glu Lys Ser Met Thr Glu Arg Lys Leu Ile Glu Glu Lys Val Ala
        115                 120                 125
```

```
Glu Asn Tyr Ser Leu Leu Ala Asn Cys Pro Ile Glu Glu Val Asp Ser
    130                 135                 140

Ile Lys Ile Tyr Lys Ile Lys Arg Phe Leu Thr Tyr Arg Ser Asn Met
145                 150                 155                 160

Leu Leu Tyr Phe Ala Ser Ile Asn Ser Phe Leu Cys Glu Gly Ile Lys
                165                 170                 175

Gly Lys Asp Asn Glu Thr Glu Glu Ile Trp His Leu Lys Asp Asn Asp
            180                 185                 190

Val Arg Lys Glu Lys Val Lys Glu Asn Phe Lys Asn Lys Leu Ile Gln
        195                 200                 205

Ser Thr Glu Asn Tyr Asn Ser Ser Leu Lys Asn Gln Ile Glu Glu Lys
    210                 215                 220

Glu Lys Leu Ser Ser Lys Glu Phe Lys Lys Gly Ala Phe Tyr Arg Thr
225                 230                 235                 240

Ile Ile Lys Lys Leu Gln Gln Glu Arg Ile Lys Glu Leu Ser Glu Lys
                245                 250                 255

Ser Leu Thr Glu Asp Cys Glu Lys Ile Ile Lys Leu Tyr Ser Glu Leu
            260                 265                 270

Arg His Pro Leu Met His Tyr Asp Tyr Gln Tyr Phe Glu Asn Leu Phe
        275                 280                 285

Glu Asn Lys Glu Asn Ser Glu Leu Thr Lys Asn Leu Asn Leu Asp Ile
    290                 295                 300

Phe Lys Ser Leu Pro Leu Val Arg Lys Met Lys Leu Asn Asn Lys Val
305                 310                 315                 320

Asn Tyr Leu Glu Asp Asn Asp Thr Leu Phe Val Leu Gln Lys Thr Lys
                325                 330                 335
```

```
Lys Ala Lys Thr Leu Tyr Gln Ile Tyr Asp Ala Leu Cys Glu Gln Lys
            340                 345                 350

Asn Gly Phe Asn Lys Phe Ile Asn Asp Phe Phe Val Ser Asp Gly Glu
        355                 360                 365

Glu Asn Thr Val Phe Lys Gln Ile Ile Asn Glu Lys Phe Gln Ser Glu
    370                 375                 380

Met Glu Phe Leu Glu Lys Arg Ile Ser Glu Ser Glu Lys Lys Asn Glu
385                 390                 395                 400

Lys Leu Lys Lys Lys Leu Asp Ser Met Lys Ala His Phe Arg Asn Ile
                405                 410                 415

Asn Ser Glu Asp Thr Lys Glu Ala Tyr Phe Trp Asp Ile His Ser Ser
            420                 425                 430

Arg Asn Tyr Lys Thr Lys Tyr Asn Glu Arg Lys Asn Leu Val Asn Glu
        435                 440                 445

Tyr Thr Lys Leu Leu Gly Ser Ser Lys Glu Lys Lys Leu Leu Arg Glu
    450                 455                 460

Glu Ile Thr Lys Ile Asn Arg Gln Leu Leu Lys Leu Lys Gln Glu Met
465                 470                 475                 480

Glu Glu Ile Thr Lys Lys Asn Ser Leu Phe Arg Leu Glu Tyr Lys Met
                485                 490                 495

Lys Ile Ala Phe Gly Phe Leu Phe Cys Glu Phe Asp Gly Asn Ile Ser
            500                 505                 510

Lys Phe Lys Asp Glu Phe Asp Ala Ser Asn Gln Glu Lys Ile Ile Gln
        515                 520                 525

Tyr His Lys Asn Gly Glu Lys Tyr Leu Thr Ser Phe Leu Lys Glu Glu
    530                 535                 540
```

```
Glu Lys Glu Lys Phe Asn Leu Glu Lys Met Gln Lys Ile Ile Gln Lys
545                 550                 555                 560


Thr Glu Glu Glu Asp Trp Leu Leu Pro Glu Thr Lys Asn Asn Leu Phe
                565                 570                 575


Lys Phe Tyr Leu Leu Thr Tyr Leu Leu Leu Pro Tyr Glu Leu Lys Gly
            580                 585                 590


Asp Phe Leu Gly Phe Val Lys Lys His Tyr Tyr Asp Ile Lys Asn Val
        595                 600                 605


Asp Phe Met Asp Glu Asn Gln Asn Asn Ile Gln Val Ser Gln Thr Val
    610                 615                 620


Glu Lys Gln Glu Asp Tyr Phe Tyr His Lys Ile Arg Leu Phe Glu Lys
625                 630                 635                 640


Asn Thr Lys Lys Tyr Glu Ile Val Lys Tyr Ser Ile Val Pro Asn Glu
                645                 650                 655


Lys Leu Lys Gln Tyr Phe Glu Asp Leu Gly Ile Asp Ile Lys Tyr Leu
            660                 665                 670


Thr Gly Ser Val Glu Ser Gly Glu Lys Trp Leu Gly Glu Asn Leu Gly
        675                 680                 685


Ile Asp Ile Lys Tyr Leu Thr Val Glu Gln Lys Ser Glu Val Ser Glu
    690                 695                 700


Glu Lys Asn Lys Lys Val Ser Leu Lys Asn Asn Gly Met Phe Asn Lys
705                 710                 715                 720


Thr Ile Leu Leu Phe Val Phe Lys Tyr Tyr Gln Ile Ala Phe Lys Leu
                725                 730                 735


Phe Asn Asp Ile Glu Leu Tyr Ser Leu Phe Phe Leu Arg Glu Lys Ser
            740                 745                 750
```

```
Glu Lys Pro Phe Glu Val Phe Leu Glu Glu Leu Lys Asp Lys Met Ile
        755                 760                 765


Gly Lys Gln Leu Asn Phe Gly Gln Leu Leu Tyr Val Val Tyr Glu Val
    770                 775                 780


Leu Val Lys Asn Lys Asp Leu Asp Lys Ile Leu Ser Lys Lys Ile Asp
785                 790                 795                 800


Tyr Arg Lys Asp Lys Ser Phe Ser Pro Glu Ile Ala Tyr Leu Arg Asn
                805                 810                 815


Phe Leu Ser His Leu Asn Tyr Ser Lys Phe Leu Asp Asn Phe Met Lys
            820                 825                 830


Ile Asn Thr Asn Lys Ser Asp Glu Asn Lys Glu Val Leu Ile Pro Ser
        835                 840                 845


Ile Lys Ile Gln Lys Met Ile Gln Phe Ile Glu Lys Cys Asn Leu Gln
    850                 855                 860


Asn Gln Ile Asp Phe Asp Phe Asn Phe Val Asn Asp Phe Tyr Met Arg
865                 870                 875                 880


Lys Glu Lys Met Phe Phe Ile Gln Leu Lys Gln Ile Phe Pro Asp Ile
                885                 890                 895


Asn Ser Thr Glu Lys Gln Lys Lys Ser Glu Lys Glu Glu Ile Leu Arg
            900                 905                 910


Lys Arg Tyr His Leu Ile Asn Lys Lys Asn Glu Gln Ile Lys Asp Glu
        915                 920                 925


His Glu Ala Gln Ser Gln Leu Tyr Glu Lys Ile Leu Ser Leu Gln Lys
    930                 935                 940


Ile Phe Ser Cys Asp Lys Asn Asn Phe Tyr Arg Arg Leu Lys Glu Glu
945                 950                 955                 960
```

```
Lys Leu Leu Phe Leu Glu Lys Gln Gly Lys Lys Lys Ile Ser Met Lys
                965                 970                 975


Glu Ile Lys Asp Lys Ile Ala Ser Asp Ile Ser Asp Leu Leu Gly Ile
            980                 985                 990


Leu Lys Lys Glu Ile Thr Arg Asp  Ile Lys Asp Lys Leu  Thr Glu Lys
        995                 1000                1005


Phe Arg  Tyr Cys Glu Glu Lys  Leu Leu Asn Ile Ser  Phe Tyr Asn
    1010                1015                1020


His Gln  Asp Lys Lys Lys Glu  Glu Gly Ile Arg Val  Phe Leu Ile
    1025                1030                1035


Arg Asp  Lys Asn Ser Asp Asn  Phe Lys Phe Glu Ser  Ile Leu Asp
    1040                1045                1050


Asp Gly  Ser Asn Lys Ile Phe  Ile Ser Lys Asn Gly  Lys Glu Ile
    1055                1060                1065


Thr Ile  Gln Cys Cys Asp Lys  Val Leu Glu Thr Leu  Met Ile Glu
    1070                1075                1080


Lys Asn  Thr Leu Lys Ile Ser  Ser Asn Gly Lys Ile  Ile Ser Leu
    1085                1090                1095


Ile Pro  His Tyr Ser Tyr Ser  Ile Asp Val Lys Tyr
    1100                1105                1110


<210>  767
<211>  1111
<212>  PRT
<213>  Fusobacterium necrophorum

<400>  767

Met Glu Lys Phe Arg Arg Gln Asn Arg Ser Ser Ile Ile Lys Ile Ile
1               5                   10                  15


Ile Ser Asn Tyr Asp Thr Lys Gly Ile Lys Glu Leu Lys Val Arg Tyr
```

```
            20                  25                  30

Arg Lys Gln Ala Gln Leu Asp Thr Phe Ile Ile Lys Thr Glu Ile Val
        35                  40                  45

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

Tyr Arg Tyr Ser Phe Leu Phe Asp Gly Glu Glu Lys Tyr His Phe Lys
65                  70                  75                  80

Asn Lys Ser Ser Val Glu Ile Val Lys Lys Asp Ile Phe Ser Gln Thr
                85                  90                  95

Pro Asp Asn Met Ile Arg Asn Tyr Lys Ile Thr Leu Lys Ile Ser Glu
            100                 105                 110

Lys Asn Pro Arg Val Val Glu Ala Glu Ile Glu Asp Leu Met Asn Ser
        115                 120                 125

Thr Ile Leu Lys Asp Gly Arg Arg Ser Ala Arg Arg Glu Lys Ser Met
    130                 135                 140

Thr Glu Arg Lys Leu Ile Glu Glu Lys Val Ala Glu Asn Tyr Ser Leu
145                 150                 155                 160

Leu Ala Asn Cys Pro Met Glu Glu Val Asp Ser Ile Lys Ile Tyr Lys
                165                 170                 175

Ile Lys Arg Phe Leu Thr Tyr Arg Ser Asn Met Leu Leu Tyr Phe Ala
            180                 185                 190

Ser Ile Asn Ser Phe Leu Cys Glu Gly Ile Lys Gly Lys Asp Asn Glu
        195                 200                 205

Thr Glu Glu Ile Trp His Leu Lys Asp Asn Asp Val Arg Lys Glu Lys
    210                 215                 220

Val Lys Glu Asn Phe Lys Asn Lys Leu Ile Gln Ser Thr Glu Asn Tyr
```

225 230 235 240

Asn Ser Ser Leu Lys Asn Gln Ile Glu Glu Lys Glu Lys Leu Leu Arg
245 250 255

Lys Glu Ser Lys Lys Gly Ala Phe Tyr Arg Thr Ile Ile Lys Lys Leu
260 265 270

Gln Gln Glu Arg Ile Lys Glu Leu Ser Glu Lys Ser Leu Thr Glu Asp
275 280 285

Cys Glu Lys Ile Ile Lys Leu Tyr Ser Glu Leu Arg His Pro Leu Met
290 295 300

His Tyr Asp Tyr Gln Tyr Phe Glu Asn Leu Phe Glu Asn Lys Glu Asn
305 310 315 320

Ser Glu Leu Thr Lys Asn Leu Asn Leu Asp Ile Phe Lys Ser Leu Pro
325 330 335

Leu Val Arg Lys Met Lys Leu Asn Asn Lys Val Asn Tyr Leu Glu Asp
340 345 350

Asn Asp Thr Leu Phe Val Leu Gln Lys Thr Lys Lys Ala Lys Thr Leu
355 360 365

Tyr Gln Ile Tyr Asp Ala Leu Cys Glu Gln Lys Asn Gly Phe Asn Lys
370 375 380

Phe Ile Asn Asp Phe Phe Val Ser Asp Gly Glu Glu Asn Thr Val Phe
385 390 395 400

Lys Gln Ile Ile Asn Glu Lys Phe Gln Ser Glu Ile Glu Phe Leu Glu
405 410 415

Lys Arg Ile Ser Glu Ser Glu Lys Lys Asn Glu Lys Leu Lys Lys Lys
420 425 430

Leu Asp Ser Met Lys Ala His Phe Arg Asn Ile Asn Ser Glu Asp Thr

```
        435                 440                 445

Lys Glu Ala Tyr Phe Trp Asp Ile His Ser Ser Arg Asn Tyr Lys Thr
    450                 455                 460

Lys Tyr Asn Glu Arg Lys Asn Leu Val Asn Glu Tyr Thr Glu Leu Leu
465                 470                 475                 480

Gly Ser Ser Lys Glu Lys Lys Leu Leu Arg Glu Glu Ile Thr Lys Ile
                485                 490                 495

Asn Arg Gln Leu Leu Lys Leu Lys Gln Glu Met Glu Glu Ile Thr Lys
            500                 505                 510

Lys Asn Ser Leu Phe Arg Leu Glu Tyr Lys Met Lys Met Ala Phe Gly
        515                 520                 525

Phe Leu Phe Cys Glu Phe Asp Gly Asn Ile Ser Arg Phe Lys Asp Glu
    530                 535                 540

Phe Asp Ala Ser Asn Gln Glu Lys Ile Ile Gln Tyr His Lys Asn Gly
545                 550                 555                 560

Glu Lys Tyr Leu Thr Tyr Phe Leu Lys Glu Glu Glu Lys Glu Lys Phe
                565                 570                 575

Asn Leu Lys Lys Leu Gln Glu Thr Ile Gln Lys Thr Gly Glu Glu Asn
            580                 585                 590

Trp Leu Leu Pro Gln Asn Lys Asn Asn Leu Phe Lys Phe Tyr Leu Leu
        595                 600                 605

Thr Tyr Leu Leu Leu Pro Tyr Glu Leu Lys Gly Asp Phe Leu Gly Phe
    610                 615                 620

Val Lys Lys His Tyr Tyr Asp Ile Lys Asn Val Asp Phe Met Asp Glu
625                 630                 635                 640

Asn Gln Ser Ser Lys Ile Ile Glu Ser Lys Glu Asp Asp Phe Tyr His
```

```
                645                 650                 655

Lys Ile Arg Leu Phe Glu Lys Asn Thr Lys Lys Tyr Glu Ile Val Lys
            660                 665                 670

Tyr Ser Ile Val Pro Asp Lys Lys Leu Lys Gln Tyr Phe Lys Asp Leu
        675                 680                 685

Gly Ile Asp Thr Lys Tyr Leu Ile Leu Asp Gln Lys Ser Glu Val Ser
    690                 695                 700

Gly Glu Lys Asn Lys Lys Val Ser Leu Lys Asn Asn Gly Met Phe Asn
705                 710                 715                 720

Lys Thr Ile Leu Leu Phe Val Phe Lys Tyr Tyr Gln Ile Ala Phe Lys
                725                 730                 735

Leu Phe Asn Asp Ile Glu Leu Tyr Ser Leu Phe Phe Leu Arg Glu Lys
            740                 745                 750

Ser Gly Lys Pro Phe Glu Val Phe Leu Lys Glu Leu Lys Asp Lys Met
        755                 760                 765

Ile Gly Lys Gln Leu Asn Phe Gly Gln Leu Leu Tyr Val Val Tyr Glu
    770                 775                 780

Val Leu Val Lys Asn Lys Asp Leu Ser Glu Ile Leu Ser Glu Arg Ile
785                 790                 795                 800

Asp Tyr Arg Lys Asp Met Cys Phe Ser Ala Glu Ile Ala Asp Leu Arg
                805                 810                 815

Asn Phe Leu Ser His Leu Asn Tyr Ser Lys Phe Leu Asp Asn Phe Met
            820                 825                 830

Lys Ile Asn Thr Asn Lys Ser Asp Glu Asn Lys Glu Val Leu Ile Pro
        835                 840                 845

Ser Ile Lys Ile Gln Lys Met Ile Lys Phe Ile Glu Glu Cys Asn Leu
```

```
    850                 855                 860

Gln Ser Gln Ile Asp Phe Asp Phe Asn Phe Val Asn Asp Phe Tyr Met
865                 870                 875                 880

Arg Lys Glu Lys Met Phe Phe Ile Gln Leu Lys Gln Ile Phe Pro Asp
                885                 890                 895

Ile Asn Ser Thr Glu Lys Gln Lys Met Asn Glu Lys Glu Glu Ile Leu
            900                 905                 910

Arg Asn Arg Tyr His Leu Thr Asp Lys Lys Asn Glu Gln Ile Lys Asp
        915                 920                 925

Glu His Glu Ala Gln Ser Gln Leu Tyr Glu Lys Ile Leu Ser Leu Gln
    930                 935                 940

Lys Ile Tyr Ser Ser Asp Lys Asn Asn Phe Tyr Gly Arg Leu Lys Glu
945                 950                 955                 960

Glu Lys Leu Leu Phe Leu Glu Lys Gln Glu Lys Lys Lys Leu Ser Met
                965                 970                 975

Glu Glu Ile Lys Asp Lys Ile Ala Gly Asp Ile Ser Asp Leu Leu Gly
            980                 985                 990

Ile Leu Lys Lys Glu Ile Thr Arg  Asp Ile Lys Asp Lys  Leu Thr Glu
        995                 1000                1005

Lys Phe  Arg Tyr Cys Glu Glu  Lys Leu Leu Asn Leu  Ser Phe Tyr
    1010                1015                1020

Asn His  Gln Asp Lys Lys Lys  Glu Glu Ser Ile Arg  Val Phe Leu
    1025                1030                1035

Ile Arg  Asp Lys Asn Ser Asp  Asn Phe Lys Phe Glu  Ser Ile Leu
    1040                1045                1050

Asp Asp  Gly Ser Asn Lys Ile  Phe Ile Ser Lys Asn  Gly Lys Glu
```

1055 1060 1065

Ile Thr Ile Gln Cys Cys Asp Lys Val Leu Glu Thr Leu Ile Ile
1070 1075 1080

Glu Lys Asn Thr Leu Lys Ile Ser Ser Asn Gly Lys Ile Ile Ser
1085 1090 1095

Leu Ile Pro His Tyr Ser Tyr Ser Ile Asp Val Lys Tyr
1100 1105 1110

<210> 768
<211> 1111
<212> PRT
<213> Anaerosalibacter massiliensis

<400> 768

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

Val Arg Val Ile Ile Ser Asn Phe Asp Asp Lys Gln Val Lys Glu Ile
20 25 30

Lys Val Leu Tyr Thr Lys Gln Gly Gly Ile Asp Val Ile Lys Phe Lys
35 40 45

Ser Thr Glu Lys Asp Glu Lys Gly Arg Met Lys Phe Asn Phe Asp Cys
50 55 60

Ala Tyr Asn Arg Leu Glu Glu Glu Glu Phe Asn Ser Phe Gly Gly Lys
65 70 75 80

Gly Lys Gln Ser Phe Phe Val Thr Thr Asn Glu Asp Leu Thr Glu Leu
85 90 95

His Val Thr Lys Arg His Lys Thr Thr Gly Glu Ile Ile Lys Asp Tyr
100 105 110

Thr Ile Gln Gly Lys Tyr Thr Pro Ile Lys Gln Asp Arg Thr Lys Val
115 120 125

```
Thr Val Ser Ile Thr Asp Asn Lys Asp His Phe Asp Ser Asn Asp Leu
    130                 135                 140

Gly Asp Lys Ile Arg Leu Ser Arg Ser Leu Thr Gln Tyr Thr Asn Arg
145                 150                 155                 160

Ile Leu Leu Asp Ala Asp Val Met Lys Asn Tyr Arg Glu Ile Val Cys
                165                 170                 175

Ser Asp Ser Glu Lys Val Asp Glu Thr Ile Asn Ile Asp Ser Gln Glu
            180                 185                 190

Ile Tyr Lys Ile Asn Arg Phe Leu Ser Tyr Arg Ser Asn Met Ile Ile
        195                 200                 205

Tyr Tyr Gln Met Ile Asn Asn Phe Leu Leu His Tyr Asp Gly Glu Glu
    210                 215                 220

Asp Lys Gly Gly Asn Asp Ser Ile Asn Leu Ile Asn Glu Ile Trp Lys
225                 230                 235                 240

Tyr Glu Asn Lys Lys Asn Asp Glu Lys Glu Lys Ile Ile Glu Arg Ser
                245                 250                 255

Tyr Lys Ser Ile Glu Lys Ser Ile Asn Gln Tyr Ile Leu Asn His Asn
            260                 265                 270

Thr Glu Val Glu Ser Gly Asp Lys Glu Lys Lys Ile Asp Ile Ser Glu
        275                 280                 285

Glu Arg Ile Lys Glu Asp Leu Lys Lys Thr Phe Ile Leu Phe Ser Arg
    290                 295                 300

Leu Arg His Tyr Met Val His Tyr Asn Tyr Lys Phe Tyr Glu Asn Leu
305                 310                 315                 320

Tyr Ser Gly Lys Asn Phe Ile Ile Tyr Asn Lys Asp Lys Ser Lys Ser
                325                 330                 335
```

```
Arg Arg Phe Ser Glu Leu Leu Asp Leu Asn Ile Phe Lys Glu Leu Ser
            340                 345                 350

Lys Ile Lys Leu Val Lys Asn Arg Ala Val Ser Asn Tyr Leu Asp Lys
        355                 360                 365

Lys Thr Thr Ile His Val Leu Asn Lys Asn Ile Asn Ala Ile Lys Leu
    370                 375                 380

Leu Asp Ile Tyr Arg Asp Ile Cys Glu Thr Lys Asn Gly Phe Asn Asn
385                 390                 395                 400

Phe Ile Asn Asn Met Met Thr Ile Ser Gly Glu Glu Asp Lys Glu Tyr
                405                 410                 415

Lys Glu Met Val Thr Lys His Phe Asn Glu Asn Met Asn Lys Leu Ser
            420                 425                 430

Ile Tyr Leu Glu Asn Phe Lys Lys His Ser Asp Phe Lys Thr Asn Asn
        435                 440                 445

Lys Lys Lys Glu Thr Tyr Asn Leu Leu Lys Gln Glu Leu Asp Glu Gln
    450                 455                 460

Lys Lys Leu Arg Leu Trp Phe Asn Ala Pro Tyr Val Tyr Asp Ile His
465                 470                 475                 480

Ser Ser Lys Lys Tyr Lys Glu Leu Tyr Val Glu Arg Lys Lys Tyr Val
                485                 490                 495

Asp Ile His Ser Lys Leu Ile Glu Ala Gly Ile Asn Asn Asp Asn Lys
            500                 505                 510

Lys Lys Leu Asn Glu Ile Asn Val Lys Leu Cys Glu Leu Asn Thr Glu
        515                 520                 525

Met Lys Glu Met Thr Lys Leu Asn Ser Lys Tyr Arg Leu Gln Tyr Lys
    530                 535                 540
```

```
Leu Gln Leu Ala Phe Gly Phe Ile Leu Glu Glu Phe Asn Leu Asp Ile
545                 550                 555                 560


Asp Lys Phe Val Ser Ala Phe Asp Lys Asp Asn Asn Leu Thr Ile Ser
                565                 570                 575


Lys Phe Met Glu Lys Arg Glu Thr Tyr Leu Ser Lys Ser Leu Asp Arg
            580                 585                 590


Arg Asp Asn Arg Phe Lys Lys Leu Ile Lys Asp Tyr Lys Phe Arg Asp
        595                 600                 605


Thr Glu Asp Ile Phe Cys Ser Asp Arg Glu Asn Asn Leu Val Lys Leu
    610                 615                 620


Tyr Ile Leu Met Tyr Ile Leu Leu Pro Val Glu Ile Arg Gly Asp Phe
625                 630                 635                 640


Leu Gly Phe Val Lys Lys Asn Tyr Tyr Asp Leu Lys His Val Asp Phe
                645                 650                 655


Ile Asp Lys Arg Asn Asn Asp Asn Lys Asp Thr Phe Phe His Asp Leu
            660                 665                 670


Arg Leu Phe Glu Lys Asn Val Lys Arg Leu Glu Val Thr Ser Tyr Ser
        675                 680                 685


Leu Ser Asp Gly Phe Leu Gly Lys Lys Ser Arg Glu Lys Phe Gly Lys
    690                 695                 700


Glu Leu Glu Lys Phe Ile Tyr Lys Asn Val Ser Ile Ala Leu Pro Thr
705                 710                 715                 720


Asn Ile Asp Ile Lys Glu Phe Asn Lys Ser Leu Val Leu Pro Met Met
                725                 730                 735


Lys Asn Tyr Gln Ile Ile Phe Lys Leu Leu Asn Asp Ile Glu Ile Ser
            740                 745                 750
```

```
Ala Leu Phe Leu Ile Ala Lys Lys Glu Gly Asn Glu Gly Ser Ile Thr
        755                 760                 765

Phe Lys Lys Val Ile Asp Lys Val Arg Lys Glu Asp Met Asn Gly Asn
    770                 775                 780

Ile Asn Phe Ser Gln Val Met Lys Met Ala Leu Asn Glu Lys Val Asn
785                 790                 795                 800

Cys Gln Ile Arg Asn Ser Ile Ala His Ile Asn Met Lys Gln Leu Tyr
                805                 810                 815

Ile Glu Pro Leu Asn Ile Tyr Ile Asn Asn Asn Gln Asn Lys Lys Thr
            820                 825                 830

Ile Ser Glu Gln Met Glu Glu Ile Ile Asp Ile Cys Ile Thr Lys Gly
        835                 840                 845

Leu Thr Gly Lys Glu Leu Asn Lys Asn Ile Ile Asn Asp Tyr Tyr Met
    850                 855                 860

Lys Lys Glu Lys Leu Val Phe Asn Leu Lys Leu Arg Lys Arg Asn Asn
865                 870                 875                 880

Leu Val Ser Ile Asp Ala Gln Gln Lys Asn Met Lys Glu Lys Ser Ile
                885                 890                 895

Leu Asn Lys Tyr Asp Leu Asn Tyr Lys Asp Glu Asn Leu Asn Ile Lys
            900                 905                 910

Glu Ile Ile Leu Lys Val Asn Asp Leu Asn Asn Lys Gln Lys Leu Leu
        915                 920                 925

Lys Glu Thr Thr Glu Gly Glu Ser Asn Tyr Lys Asn Ala Leu Ser Lys
    930                 935                 940

Asp Ile Leu Leu Leu Asn Gly Ile Ile Arg Lys Asn Ile Asn Phe Lys
945                 950                 955                 960
```

```
Ile Lys Glu Met Ile Leu Gly Ile Ile Gln Gln Asn Glu Tyr Arg Tyr
                965                 970                 975


Val Asn Ile Asn Ile Tyr Asp Lys Ile Arg Lys Glu Asp His Asn Ile
            980                 985                 990


Asp Leu Lys Ile Asn Asn Lys Tyr  Ile Glu Ile Ser Cys  Tyr Glu Asn
        995                 1000                1005


Lys Ser  Asn Glu Ser Thr Asp  Glu Arg Ile Asn Phe  Lys Ile Lys
    1010                1015                1020


Tyr Met  Asp Leu Lys Val Lys  Asn Glu Leu Leu Val  Pro Ser Cys
    1025                1030                1035


Tyr Glu  Asp Ile Tyr Ile Lys  Lys Lys Ile Asp Leu  Glu Ile Arg
    1040                1045                1050


Tyr Ile  Glu Asn Cys Lys Val  Val Tyr Ile Asp Ile  Tyr Tyr Lys
    1055                1060                1065


Lys Tyr  Asn Ile Asn Leu Glu  Phe Asp Gly Lys Thr  Leu Phe Val
    1070                1075                1080


Lys Phe  Asn Lys Asp Val Lys  Lys Asn Asn Gln Lys  Val Asn Leu
    1085                1090                1095


Glu Ser  Asn Tyr Ile Gln Asn  Ile Lys Phe Ile Val  Ser
    1100                1105                1110


<210>  769
<211>  1111
<212>  PRT
<213>  Fusobacterium necrophorum

<400>  769

Met Glu Lys Phe Arg Arg Gln Asn Arg Asn Ser Ile Ile Lys Ile Ile
1               5                   10                  15
```

```
Ile Ser Asn Tyr Asp Thr Lys Gly Ile Lys Glu Leu Lys Val Arg Tyr
            20                  25                  30

Arg Lys Gln Ala Gln Leu Asp Thr Phe Ile Ile Lys Thr Glu Ile Val
        35                  40                  45

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

Tyr Arg Tyr Ser Phe Leu Phe Asp Gly Glu Glu Lys Tyr His Phe Lys
65                  70                  75                  80

Asn Lys Ser Ser Val Glu Ile Val Lys Lys Asp Ile Phe Ser Gln Thr
                85                  90                  95

Pro Asp Asn Met Ile Arg Asn Tyr Lys Ile Thr Leu Lys Ile Ser Glu
            100                 105                 110

Lys Asn Pro Arg Val Val Glu Ala Glu Ile Glu Asp Leu Met Asn Ser
        115                 120                 125

Thr Ile Leu Lys Asp Gly Arg Arg Ser Ala Arg Arg Glu Lys Ser Val
    130                 135                 140

Thr Glu Arg Lys Leu Ile Glu Glu Lys Val Ala Glu Asn Tyr Ser Leu
145                 150                 155                 160

Leu Ala Asn Cys Pro Met Glu Glu Val Asp Ser Ile Lys Ile Tyr Lys
                165                 170                 175

Ile Lys Arg Phe Leu Thr Tyr Arg Ser Asn Met Leu Leu Tyr Phe Ala
            180                 185                 190

Ser Ile Asn Ser Phe Leu Cys Glu Gly Ile Lys Gly Lys Glu Asn Glu
        195                 200                 205

Thr Glu Glu Ile Trp His Leu Lys Asp Asn Asp Val Arg Lys Glu Lys
    210                 215                 220
```

```
Val Lys Glu Asn Phe Lys Asn Lys Leu Ile Gln Ser Thr Glu Asn Tyr
225                 230                 235                 240

Asn Ser Ser Leu Lys Asn Gln Ile Glu Glu Lys Glu Lys Leu Leu Arg
                245                 250                 255

Lys Glu Ser Lys Lys Gly Ala Phe Tyr Arg Thr Ile Ile Lys Lys Leu
            260                 265                 270

Gln Gln Glu Arg Ile Lys Glu Leu Ser Glu Lys Ser Leu Thr Glu Asp
        275                 280                 285

Cys Glu Lys Ile Ile Lys Leu Tyr Ser Lys Leu Arg His Ser Leu Met
    290                 295                 300

His Tyr Asp Tyr Gln Tyr Phe Glu Asn Leu Phe Glu Asn Lys Glu Thr
305                 310                 315                 320

Pro Glu Leu Lys Asp Lys Leu Asp Leu His Leu Phe Lys Ser Leu Pro
                325                 330                 335

Leu Ile Arg Lys Met Lys Leu Asn Asn Lys Val Asn Tyr Leu Glu Asp
            340                 345                 350

Gly Asp Thr Leu Phe Val Leu Gln Lys Thr Lys Lys Ala Lys Thr Leu
        355                 360                 365

Tyr Gln Ile Tyr Asp Ala Leu Cys Glu Gln Lys Asn Gly Phe Asn Lys
    370                 375                 380

Phe Ile Asn Asp Phe Phe Val Ser Asp Gly Glu Glu Asn Thr Val Phe
385                 390                 395                 400

Lys Gln Ile Ile Asn Glu Lys Phe Gln Ser Glu Met Glu Phe Leu Gly
                405                 410                 415

Lys Arg Ile Ser Glu Ser Glu Glu Lys Asn Pro Lys Leu Lys Lys Lys
            420                 425                 430
```

```
Phe Asp Ser Met Lys Ala His Phe His Asn Ile Asn Ser Glu Asp Thr
        435                 440                 445

Lys Glu Ala Tyr Phe Trp Asp Ile His Ser Ser Ser Asn Tyr Lys Thr
    450                 455                 460

Lys Tyr Asn Glu Arg Lys Asn Leu Val Asn Glu Tyr Thr Glu Leu Leu
465                 470                 475                 480

Gly Ser Ser Lys Glu Lys Lys Leu Leu Arg Glu Glu Ile Thr Gln Ile
                485                 490                 495

Asn Arg Lys Leu Leu Lys Leu Lys Gln Glu Met Glu Glu Ile Thr Lys
            500                 505                 510

Lys Asn Ser Leu Phe Arg Leu Glu Tyr Lys Met Lys Met Ala Phe Gly
        515                 520                 525

Phe Leu Phe Cys Glu Phe Asp Gly Asn Ile Ser Arg Phe Lys Asp Glu
    530                 535                 540

Phe Asp Ala Ser Asn Gln Glu Lys Ile Ile Gln Tyr His Lys Asn Gly
545                 550                 555                 560

Glu Lys Tyr Leu Thr Tyr Phe Leu Lys Glu Glu Glu Lys Glu Lys Phe
                565                 570                 575

Asn Leu Lys Lys Leu Gln Glu Thr Ile Gln Lys Thr Gly Lys Glu Asn
            580                 585                 590

Trp Leu Leu Pro Gln Asn Lys Asn Asn Leu Phe Lys Phe Tyr Leu Leu
        595                 600                 605

Thr Tyr Leu Leu Leu Pro Tyr Glu Leu Lys Gly Asp Phe Leu Gly Phe
    610                 615                 620

Val Lys Lys His Tyr Tyr Asp Ile Lys Asn Val Asp Phe Met Asp Glu
625                 630                 635                 640
```

```
Asn Gln Ser Ser Lys Ile Ile Glu Ser Lys Glu Asp Asp Phe Tyr His
                645                 650                 655

Lys Ile Arg Leu Phe Glu Lys Asn Thr Lys Lys Tyr Glu Ile Val Lys
            660                 665                 670

Tyr Ser Ile Val Pro Asp Glu Lys Leu Lys Gln Tyr Phe Lys Asp Leu
        675                 680                 685

Gly Ile Asp Thr Lys Tyr Leu Ile Leu Glu Gln Lys Ser Glu Val Ser
    690                 695                 700

Gly Glu Lys Asn Lys Lys Val Ser Leu Lys Asn Asn Gly Met Phe Asn
705                 710                 715                 720

Lys Thr Ile Leu Leu Phe Val Phe Lys Tyr Tyr Gln Ile Ala Phe Lys
                725                 730                 735

Leu Phe Asn Asp Ile Glu Leu Tyr Ser Leu Phe Phe Leu Arg Glu Lys
            740                 745                 750

Ser Gly Lys Pro Phe Glu Val Phe Leu Lys Glu Leu Lys Asp Lys Met
        755                 760                 765

Ile Gly Lys Gln Leu Asn Phe Gly Gln Leu Leu Tyr Val Ile Tyr Glu
    770                 775                 780

Val Leu Val Lys Asn Lys Asp Leu Ser Glu Ile Leu Ser Glu Arg Ile
785                 790                 795                 800

Asp Tyr Arg Lys Asp Met Cys Phe Ser Ala Glu Ile Ala Asp Leu Arg
                805                 810                 815

Asn Phe Leu Ser His Leu Asn Tyr Ser Lys Phe Leu Asp Asn Phe Met
            820                 825                 830

Lys Ile Asn Thr Asn Lys Ser Asp Glu Asn Lys Glu Val Leu Ile Pro
        835                 840                 845
```

```
Ser Ile Lys Ile Gln Lys Met Ile Lys Phe Ile Glu Glu Cys Asn Leu
    850                 855                 860

Gln Ser Gln Ile Asp Phe Asp Phe Asn Phe Val Asn Asp Phe Tyr Met
865                 870                 875                 880

Arg Lys Glu Lys Met Phe Phe Ile Gln Leu Lys Gln Ile Phe Pro Asp
                885                 890                 895

Ile Asn Ser Thr Glu Lys Gln Lys Met Asn Glu Lys Glu Glu Ile Leu
            900                 905                 910

Arg Asn Arg Tyr His Leu Thr Asp Lys Lys Asn Glu Gln Ile Lys Asp
        915                 920                 925

Glu His Glu Ala Gln Ser Gln Leu Tyr Glu Lys Ile Leu Ser Leu Gln
    930                 935                 940

Lys Ile Tyr Ser Ser Asp Lys Asn Asn Phe Tyr Gly Arg Leu Lys Glu
945                 950                 955                 960

Glu Lys Leu Leu Phe Leu Gly Lys Gln Gly Lys Lys Lys Leu Ser Met
                965                 970                 975

Glu Glu Ile Lys Asp Lys Ile Ala Gly Asp Ile Ser Asp Leu Leu Gly
            980                 985                 990

Ile Leu Lys Lys Glu Ile Thr Arg  Asp Ile Lys Asp Lys  Leu Thr Glu
        995                 1000                1005

Lys Phe  Arg Tyr Cys Glu Glu  Lys Leu Leu Asn Leu  Ser Phe Tyr
    1010                1015                1020

Asn His  Gln Asp Lys Lys Lys  Glu Glu Ser Ile Arg  Val Phe Leu
    1025                1030                1035

Ile Arg  Asp Lys Asn Ser Asp  Asn Phe Lys Phe Glu  Ser Ile Leu
    1040                1045                1050
```

```
Asp Asp  Gly Ser Asn Lys Ile  Phe Ile Ser Lys Asn  Gly Lys Glu
    1055                1060                1065


Ile Thr  Ile Gln Cys Cys Asp  Lys Val Leu Glu Thr  Leu Met Ile
    1070                1075                1080


Glu Lys  Asn Thr Leu Lys Ile  Ser Ser Asn Gly Lys  Ile Ile Ser
    1085                1090                1095


Leu Val  Pro His Tyr Ser Tyr  Ser Ile Asp Val Lys  Tyr
    1100                1105                1110


<210>  770
<211>  1115
<212>  PRT
<213>  Fusobacterium necrophorum

<400>  770

Met Glu Lys Phe Arg Arg Gln Asn Arg Asn Ser Ile Ile Lys Ile Ile
1               5                   10                  15


Ile Ser Asn Tyr Asp Thr Lys Gly Ile Lys Glu Leu Lys Val Arg Tyr
            20                  25                  30


Arg Lys Gln Ala Gln Leu Asp Thr Phe Ile Ile Lys Thr Glu Ile Val
        35                  40                  45


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


Tyr Arg Tyr Ser Phe Leu Phe Asp Gly Glu Glu Lys Tyr His Phe Lys
65                  70                  75                  80


Asn Lys Ser Ser Val Glu Ile Val Lys Lys Asp Ile Phe Ser Gln Thr
                85                  90                  95


Pro Asp Asn Met Ile Arg Asn Tyr Lys Ile Thr Leu Lys Ile Ser Glu
            100                 105                 110
```

```
Lys Asn Pro Arg Val Val Glu Ala Glu Ile Glu Asp Leu Met Asn Ser
        115                 120                 125

Thr Ile Leu Lys Asp Gly Arg Arg Ser Ala Arg Arg Glu Lys Ser Met
    130                 135                 140

Thr Glu Arg Lys Leu Ile Glu Glu Lys Val Ala Lys Asn Tyr Ser Leu
145                 150                 155                 160

Leu Ala Asn Cys Pro Met Glu Glu Val Asp Ser Ile Lys Ile Tyr Lys
                165                 170                 175

Ile Lys Arg Phe Leu Thr Tyr Arg Ser Asn Met Leu Leu Tyr Phe Ala
            180                 185                 190

Ser Ile Asn Ser Phe Leu Cys Glu Gly Ile Lys Gly Lys Asp Asn Glu
        195                 200                 205

Thr Glu Glu Ile Trp His Leu Lys Asp Asn Asp Val Arg Lys Glu Lys
    210                 215                 220

Val Arg Glu Asn Phe Lys Asn Lys Leu Ile Gln Ser Thr Glu Asn Tyr
225                 230                 235                 240

Asn Ser Ser Leu Lys Asn Gln Ile Glu Glu Lys Glu Lys Leu Leu Arg
                245                 250                 255

Lys Glu Phe Lys Lys Gly Ala Phe Tyr Arg Thr Ile Ile Lys Lys Leu
            260                 265                 270

Gln Gln Glu Arg Ile Lys Glu Leu Ser Glu Lys Ser Leu Thr Glu Asp
        275                 280                 285

Cys Glu Lys Ile Ile Lys Leu Tyr Ser Lys Leu Arg His Ser Leu Met
    290                 295                 300

His Tyr Asp Tyr Gln Tyr Phe Glu Asn Leu Phe Glu Asn Lys Lys Asn
305                 310                 315                 320
```

```
Asp Asp Leu Met Lys Asp Leu Asn Leu Asp Leu Phe Lys Ser Leu Pro
                325                 330                 335

Leu Ile Arg Lys Met Lys Leu Asn Asn Lys Val Asn Tyr Leu Glu Asp
            340                 345                 350

Gly Asp Thr Leu Phe Val Leu Gln Lys Thr Lys Lys Ala Lys Thr Leu
        355                 360                 365

Tyr Gln Ile Tyr Asp Ala Leu Cys Glu Gln Lys Asn Gly Phe Asn Lys
    370                 375                 380

Phe Ile Asn Asp Phe Phe Val Ser Asp Gly Glu Glu Asn Thr Val Phe
385                 390                 395                 400

Lys Gln Ile Ile Asn Glu Lys Phe Gln Ser Glu Met Glu Phe Leu Glu
                405                 410                 415

Lys Arg Ile Ser Glu Ser Glu Lys Lys Asn Glu Lys Leu Lys Lys Lys
            420                 425                 430

Leu Asp Ser Met Lys Ala His Phe Arg Asn Ile Asn Ser Glu Asp Thr
        435                 440                 445

Lys Glu Ala Tyr Phe Trp Asp Ile His Ser Ser Arg Asn Tyr Lys Thr
    450                 455                 460

Lys Tyr Asn Glu Arg Lys Asn Leu Val Asn Glu Tyr Thr Glu Leu Leu
465                 470                 475                 480

Gly Ser Ser Lys Glu Lys Lys Leu Leu Arg Glu Glu Ile Thr Lys Ile
                485                 490                 495

Asn Arg Gln Leu Leu Lys Leu Lys Gln Glu Met Glu Glu Ile Thr Lys
            500                 505                 510

Lys Asn Ser Leu Phe Arg Leu Glu Tyr Lys Met Lys Ile Ala Phe Gly
        515                 520                 525
```

```
Phe Leu Phe Cys Glu Phe Asp Gly Asn Ile Ser Lys Phe Lys Asp Glu
    530                 535                 540

Phe Asp Ala Ser Asn Gln Glu Lys Ile Ile Gln Tyr His Lys Asn Gly
545                 550                 555                 560

Glu Lys Tyr Leu Thr Ser Phe Leu Lys Glu Glu Glu Lys Glu Lys Phe
                565                 570                 575

Asn Leu Glu Lys Met Gln Lys Ile Ile Gln Lys Thr Glu Glu Glu Asp
            580                 585                 590

Trp Leu Leu Pro Glu Thr Lys Asn Asn Leu Phe Lys Phe Tyr Leu Leu
        595                 600                 605

Thr Tyr Leu Leu Leu Pro Tyr Glu Leu Lys Gly Asp Phe Leu Gly Phe
    610                 615                 620

Val Lys Lys His Tyr Tyr Asp Ile Lys Asn Val Asp Phe Ile Asp Glu
625                 630                 635                 640

Asn Gln Asn Asn Ile Gln Val Ser Gln Thr Val Glu Lys Gln Glu Asp
                645                 650                 655

Tyr Phe Tyr His Lys Ile Arg Leu Phe Glu Lys Asn Thr Lys Lys Tyr
            660                 665                 670

Glu Ile Val Lys Tyr Ser Ile Val Pro Asn Glu Lys Leu Lys Gln Tyr
        675                 680                 685

Phe Glu Asp Leu Gly Ile Asp Ile Lys Tyr Leu Thr Val Glu Gln Lys
    690                 695                 700

Ser Glu Val Ser Glu Glu Lys Asn Lys Lys Val Ser Leu Lys Asn Asn
705                 710                 715                 720

Gly Met Phe Asn Lys Thr Ile Leu Leu Phe Val Phe Lys Tyr Tyr Gln
                725                 730                 735
```

```
Ile Ala Phe Lys Leu Phe Asn Asp Ile Glu Leu Tyr Ser Leu Phe Phe
            740                 745                 750


Leu Arg Glu Lys Ser Gly Lys Pro Leu Glu Ile Phe Arg Lys Glu Leu
        755                 760                 765


Glu Ser Lys Met Lys Asp Gly Tyr Leu Asn Phe Gly Gln Leu Leu Tyr
    770                 775                 780


Val Val Tyr Glu Val Leu Val Lys Asn Lys Asp Leu Asp Lys Ile Leu
785                 790                 795                 800


Ser Lys Lys Ile Asp Tyr Arg Lys Asp Lys Ser Phe Ser Pro Glu Ile
                805                 810                 815


Ala Tyr Leu Arg Asn Phe Leu Ser His Leu Asn Tyr Ser Lys Phe Leu
            820                 825                 830


Asp Asn Phe Met Lys Ile Asn Thr Asn Lys Ser Asp Glu Asn Lys Glu
        835                 840                 845


Val Leu Ile Pro Ser Ile Lys Ile Gln Lys Met Ile Gln Phe Ile Glu
    850                 855                 860


Lys Cys Asn Leu Gln Asn Gln Ile Asp Phe Asp Phe Asn Phe Val Asn
865                 870                 875                 880


Asp Phe Tyr Met Arg Lys Glu Lys Met Phe Phe Ile Gln Leu Lys Gln
                885                 890                 895


Ile Phe Pro Asp Ile Asn Ser Thr Glu Lys Gln Lys Met Asn Glu Lys
            900                 905                 910


Glu Glu Ile Leu Arg Asn Arg Tyr His Leu Thr Asp Lys Lys Asn Glu
        915                 920                 925


Gln Ile Lys Asp Glu His Glu Ala Gln Ser Gln Leu Tyr Glu Lys Ile
    930                 935                 940
```

```
Leu Ser Leu Gln Lys Ile Tyr Ser Ser Asp Lys Asn Asn Phe Tyr Gly
945                 950                 955                 960


Arg Leu Lys Glu Glu Lys Leu Leu Phe Leu Glu Lys Gln Gly Lys Lys
                965                 970                 975


Lys Leu Ser Met Glu Glu Ile Lys Asp Lys Ile Ala Gly Asp Ile Ser
            980                 985                 990


Asp Leu Leu Gly Ile Leu Lys Lys  Glu Ile Thr Arg Asp  Ile Lys Asp
        995                 1000                1005


Lys Leu  Thr Glu Lys Phe Arg  Tyr Cys Glu Glu Lys  Leu Leu Asn
    1010                1015                1020


Leu Ser  Phe Tyr Asn His Gln  Asp Lys Lys Lys Glu  Glu Ser Ile
    1025                1030                1035


Arg Val  Phe Leu Ile Arg Asp  Lys Asn Ser Asp Asn  Phe Lys Phe
    1040                1045                1050


Glu Ser  Ile Leu Asp Asp Gly  Ser Asn Lys Ile Phe  Ile Ser Lys
    1055                1060                1065


Asn Gly  Lys Glu Ile Thr Ile  Gln Cys Cys Asp Lys  Val Leu Glu
    1070                1075                1080


Thr Leu  Ile Ile Glu Lys Asn  Thr Leu Lys Ile Ser  Ser Asn Gly
    1085                1090                1095


Lys Ile  Ile Ser Leu Ile Pro  His Tyr Ser Tyr Ser  Ile Asp Val
    1100                1105                1110


Lys Tyr
    1115


<210>  771
<211>  1121
<212>  PRT
<213>  Fusobacterium perfoetens
```

<400> 771

```
Met Gly Lys Pro Asn Arg Ser Ser Ile Ile Lys Ile Ile Ile Ser Asn
1               5                  10                  15

Tyr Asp Asn Lys Gly Ile Lys Glu Val Lys Val Arg Tyr Asn Lys Gln
            20                  25                  30

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

Phe Ile Leu Tyr Ser Ile Val Asp Lys Ala Arg Glu Lys Tyr Arg Tyr
    50                  55                  60

Ser Phe Glu Ile Asp Lys Thr Asn Ile Asn Lys Asn Glu Ile Leu Ile
65                  70                  75                  80

Ile Lys Lys Asp Ile Tyr Ser Asn Lys Glu Asp Lys Val Ile Arg Lys
                85                  90                  95

Tyr Ile Leu Ser Phe Glu Val Ser Glu Lys Asn Asp Arg Thr Ile Val
            100                 105                 110

Thr Lys Ile Lys Asp Cys Leu Glu Thr Gln Lys Lys Glu Lys Phe Glu
        115                 120                 125

Arg Glu Asn Thr Arg Arg Leu Ile Ser Glu Thr Glu Arg Lys Leu Leu
    130                 135                 140

Ser Glu Glu Thr Gln Lys Thr Tyr Ser Lys Ile Ala Cys Cys Ser Pro
145                 150                 155                 160

Glu Asp Ile Asp Ser Val Lys Ile Tyr Lys Ile Lys Arg Tyr Leu Ala
                165                 170                 175

Tyr Arg Ser Asn Met Leu Leu Phe Phe Ser Leu Ile Asn Asp Ile Phe
            180                 185                 190

Val Lys Gly Val Val Lys Asp Asn Gly Glu Glu Val Gly Glu Ile Trp
```

```
        195                 200                 205

Arg Ile Ile Asp Ser Lys Glu Ile Asp Glu Lys Lys Thr Tyr Asp Leu
    210                 215                 220

Leu Val Glu Asn Phe Lys Lys Arg Met Ser Gln Glu Phe Ile Asn Tyr
225                 230                 235                 240

Lys Gln Ser Ile Glu Asn Lys Ile Glu Lys Asn Thr Asn Lys Ile Lys
                245                 250                 255

Glu Ile Glu Gln Lys Leu Lys Lys Glu Lys Tyr Lys Lys Glu Ile Asn
            260                 265                 270

Arg Leu Lys Lys Gln Leu Ile Glu Leu Asn Arg Glu Asn Asp Leu Leu
        275                 280                 285

Glu Lys Asp Lys Ile Glu Leu Ser Asp Glu Glu Ile Arg Glu Asp Ile
    290                 295                 300

Glu Lys Ile Leu Lys Ile Tyr Ser Asp Leu Arg His Lys Leu Met His
305                 310                 315                 320

Tyr Asn Tyr Gln Tyr Phe Glu Asn Leu Phe Glu Asn Lys Lys Ile Ser
                325                 330                 335

Lys Glu Lys Asn Glu Asp Val Asn Leu Thr Glu Leu Leu Asp Leu Asn
            340                 345                 350

Leu Phe Arg Tyr Leu Pro Leu Val Arg Gln Leu Lys Leu Glu Asn Lys
        355                 360                 365

Thr Asn Tyr Leu Glu Lys Glu Asp Lys Ile Thr Val Leu Gly Val Ser
    370                 375                 380

Asp Ser Ala Ile Lys Tyr Tyr Ser Tyr Tyr Asn Phe Leu Cys Glu Gln
385                 390                 395                 400

Lys Asn Gly Phe Asn Asn Phe Ile Asn Ser Phe Phe Ser Asn Asp Gly
```

405 410 415

Glu Glu Asn Lys Ser Phe Lys Glu Lys Ile Asn Leu Ser Leu Glu Lys
420 425 430

Glu Ile Glu Ile Met Glu Lys Glu Thr Asn Glu Lys Ile Lys Glu Ile
435 440 445

Asn Lys Asn Glu Leu Gln Leu Met Lys Glu Gln Lys Glu Leu Gly Thr
450 455 460

Ala Tyr Val Leu Asp Ile His Ser Leu Asn Asp Tyr Lys Ile Ser His
465 470 475 480

Asn Glu Arg Asn Lys Asn Val Lys Leu Gln Asn Asp Ile Met Asn Gly
485 490 495

Asn Arg Asp Lys Asn Ala Leu Asp Lys Ile Asn Lys Lys Leu Val Glu
500 505 510

Leu Lys Ile Lys Met Asp Lys Ile Thr Lys Arg Asn Ser Ile Leu Arg
515 520 525

Leu Lys Tyr Lys Leu Gln Val Ala Tyr Gly Phe Leu Met Glu Glu Tyr
530 535 540

Lys Gly Asn Ile Lys Lys Phe Lys Asp Glu Phe Asp Ile Ser Lys Glu
545 550 555 560

Lys Ile Lys Ser Tyr Lys Ser Lys Gly Glu Lys Tyr Leu Glu Val Lys
565 570 575

Ser Glu Lys Lys Tyr Ile Thr Lys Ile Leu Asn Ser Ile Glu Asp Ile
580 585 590

His Asn Ile Thr Trp Leu Lys Asn Gln Glu Glu Asn Asn Leu Phe Lys
595 600 605

Phe Tyr Val Leu Thr Tyr Ile Leu Leu Pro Phe Glu Phe Arg Gly Asp

```
    610                 615                 620

Phe Leu Gly Phe Val Lys Lys His Tyr Tyr Asp Ile Lys Asn Val Glu
625                 630                 635                 640

Phe Leu Asp Glu Asn Asn Asp Arg Leu Thr Pro Glu Gln Leu Glu Lys
                645                 650                 655

Met Lys Asn Asp Ser Phe Phe Asn Lys Ile Arg Leu Phe Glu Lys Asn
            660                 665                 670

Ser Lys Lys Tyr Asp Ile Leu Lys Glu Ser Ile Leu Thr Ser Glu Arg
        675                 680                 685

Ile Gly Lys Tyr Phe Ser Leu Leu Asn Thr Gly Ala Lys Tyr Phe Glu
    690                 695                 700

Tyr Gly Gly Glu Glu Asn Arg Gly Ile Phe Asn Lys Asn Ile Ile Ile
705                 710                 715                 720

Pro Ile Phe Lys Tyr Tyr Gln Ile Val Leu Lys Leu Tyr Asn Asp Val
                725                 730                 735

Glu Leu Ala Met Leu Leu Thr Leu Ser Glu Ser Asp Glu Lys Asp Ile
            740                 745                 750

Asn Lys Ile Lys Glu Leu Val Thr Leu Lys Glu Lys Val Ser Pro Lys
        755                 760                 765

Lys Ile Asp Tyr Glu Lys Lys Tyr Lys Phe Ser Val Leu Leu Asp Cys
    770                 775                 780

Phe Asn Arg Ile Ile Asn Leu Gly Lys Lys Asp Phe Leu Ala Ser Glu
785                 790                 795                 800

Glu Val Lys Glu Val Ala Lys Thr Phe Thr Asn Leu Ala Tyr Leu Arg
                805                 810                 815

Asn Lys Ile Cys His Leu Asn Tyr Ser Lys Phe Ile Asp Asp Leu Leu
```

820 825 830

Thr Ile Asp Thr Asn Lys Ser Thr Thr Asp Ser Glu Gly Lys Leu Leu
835 840 845

Ile Asn Asp Arg Ile Arg Lys Leu Ile Lys Phe Ile Arg Glu Asn Asn
850 855 860

Gln Lys Met Asn Ile Ser Ile Asp Tyr Asn Tyr Ile Asn Asp Tyr Tyr
865 870 875 880

Met Lys Lys Glu Lys Phe Ile Phe Gly Gln Arg Lys Gln Ala Lys Thr
885 890 895

Ile Ile Asp Ser Gly Lys Lys Ala Asn Lys Arg Asn Lys Ala Glu Glu
900 905 910

Leu Leu Lys Met Tyr Arg Val Lys Lys Glu Asn Ile Asn Leu Ile Tyr
915 920 925

Glu Leu Ser Lys Lys Leu Asn Glu Leu Thr Lys Ser Glu Leu Phe Leu
930 935 940

Leu Asp Lys Lys Leu Leu Lys Asp Ile Asp Phe Thr Asp Val Lys Ile
945 950 955 960

Lys Asn Lys Ser Phe Phe Glu Leu Lys Asn Asp Val Lys Glu Val Ala
965 970 975

Asn Ile Lys Gln Ala Leu Gln Lys His Ser Ser Glu Leu Ile Gly Ile
980 985 990

Tyr Lys Lys Glu Val Ile Met Ala Ile Lys Arg Ser Ile Val Ser Lys
995 1000 1005

Leu Ile Tyr Asp Glu Glu Lys Val Leu Ser Ile Ile Ile Tyr Asp
1010 1015 1020

Lys Thr Asn Lys Lys Tyr Glu Asp Phe Leu Leu Glu Ile Arg Arg

1025 1030 1035

Glu Arg Asp Ile Asn Lys Phe Gln Phe Leu Ile Asp Glu Lys Lys
1040 1045 1050

Glu Lys Leu Gly Tyr Glu Lys Ile Ile Glu Thr Lys Glu Lys Lys
1055 1060 1065

Lys Val Val Val Lys Ile Gln Asn Asn Ser Glu Leu Val Ser Glu
1070 1075 1080

Pro Arg Ile Ile Lys Asn Lys Asp Lys Lys Lys Ala Lys Thr Pro
1085 1090 1095

Glu Glu Ile Ser Lys Leu Gly Ile Leu Asp Leu Thr Asn His Tyr
1100 1105 1110

Cys Phe Asn Leu Lys Ile Thr Leu
1115 1120

<210> 772
<211> 1137
<212> PRT
<213> Fusobacterium necrophorum

<400> 772

Met Glu Lys Phe Arg Arg Gln Asn Arg Asn Ser Ile Ile Lys Ile Ile
1 5 10 15

Ile Ser Asn Tyr Asp Thr Lys Gly Ile Lys Glu Leu Lys Val Arg Tyr
20 25 30

Arg Lys Gln Ala Gln Leu Asp Thr Phe Ile Ile Lys Thr Glu Ile Val
35 40 45

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

Tyr Arg Tyr Ser Phe Leu Phe Asp Gly Glu Glu Lys Tyr His Phe Lys
65 70 75 80

```
Asn Lys Ser Ser Val Glu Ile Val Lys Lys Asp Ile Phe Ser Gln Thr
                85                  90                  95

Pro Asp Asn Met Ile Arg Asn Tyr Lys Ile Thr Leu Lys Ile Ser Glu
            100                 105                 110

Lys Asn Pro Arg Val Val Glu Ala Glu Ile Glu Asp Leu Met Asn Ser
        115                 120                 125

Thr Ile Leu Lys Asp Gly Arg Arg Ser Ala Arg Arg Glu Lys Ser Met
    130                 135                 140

Thr Glu Arg Lys Leu Ile Glu Glu Lys Val Ala Lys Asn Tyr Ser Leu
145                 150                 155                 160

Leu Ala Asn Cys Pro Met Glu Glu Val Asp Ser Ile Lys Ile Tyr Lys
                165                 170                 175

Ile Lys Arg Phe Leu Thr Tyr Arg Ser Asn Met Leu Leu Tyr Phe Ala
            180                 185                 190

Ser Ile Asn Ser Phe Leu Cys Glu Gly Ile Lys Gly Lys Asp Asn Glu
        195                 200                 205

Thr Glu Glu Ile Trp His Leu Lys Asp Asn Asp Val Arg Lys Glu Lys
    210                 215                 220

Val Arg Glu Asn Phe Lys Asn Lys Leu Ile Gln Ser Thr Glu Asn Tyr
225                 230                 235                 240

Asn Ser Ser Leu Lys Asn Gln Ile Glu Glu Lys Glu Lys Leu Leu Arg
                245                 250                 255

Lys Glu Phe Lys Lys Gly Ala Phe Tyr Arg Thr Ile Ile Lys Lys Leu
            260                 265                 270

Gln Gln Glu Arg Ile Lys Glu Leu Ser Glu Lys Ser Leu Thr Glu Asp
        275                 280                 285
```

```
Cys Glu Lys Ile Ile Lys Leu Tyr Ser Lys Leu Arg His Ser Leu Met
    290                 295                 300

His Tyr Asp Tyr Gln Tyr Phe Glu Asn Leu Phe Glu Asn Lys Lys Asn
305                 310                 315                 320

Asp Asp Leu Met Lys Asp Leu Asn Leu Asp Leu Phe Lys Ser Leu Pro
                325                 330                 335

Leu Ile Arg Lys Met Lys Leu Asn Asn Lys Val Asn Tyr Leu Glu Asp
            340                 345                 350

Gly Asp Thr Leu Phe Val Leu Gln Lys Thr Lys Lys Ala Lys Thr Leu
        355                 360                 365

Tyr Gln Ile Tyr Asp Ala Leu Cys Glu Gln Lys Asn Gly Phe Asn Lys
    370                 375                 380

Phe Ile Asn Asp Phe Phe Val Ser Asp Gly Glu Glu Asn Thr Val Phe
385                 390                 395                 400

Lys Gln Ile Ile Asn Glu Lys Phe Gln Ser Glu Met Glu Phe Leu Glu
                405                 410                 415

Lys Arg Ile Ser Glu Ser Glu Lys Lys Asn Glu Lys Leu Lys Lys Lys
            420                 425                 430

Leu Asp Ser Met Lys Ala His Phe Arg Asn Ile Asn Ser Glu Asp Thr
        435                 440                 445

Lys Glu Ala Tyr Phe Trp Asp Ile His Ser Ser Arg Asn Tyr Lys Thr
    450                 455                 460

Lys Tyr Asn Glu Arg Lys Asn Leu Val Asn Glu Tyr Thr Glu Leu Leu
465                 470                 475                 480

Gly Ser Ser Lys Glu Lys Lys Leu Leu Arg Glu Glu Ile Thr Lys Ile
                485                 490                 495
```

```
Asn Arg Gln Leu Leu Lys Leu Lys Gln Glu Met Glu Glu Ile Thr Lys
            500                 505                 510

Lys Asn Ser Leu Phe Arg Leu Glu Tyr Lys Met Lys Ile Ala Phe Gly
        515                 520                 525

Phe Leu Phe Cys Glu Phe Asp Gly Asn Ile Ser Lys Phe Lys Asp Glu
    530                 535                 540

Phe Asp Ala Ser Asn Gln Glu Lys Ile Ile Gln Tyr His Lys Asn Gly
545                 550                 555                 560

Glu Lys Tyr Leu Thr Ser Phe Leu Lys Glu Glu Glu Lys Glu Lys Phe
                565                 570                 575

Asn Leu Glu Lys Met Gln Lys Ile Ile Gln Lys Thr Glu Glu Glu Asp
            580                 585                 590

Trp Leu Leu Pro Glu Thr Lys Asn Asn Leu Phe Lys Phe Tyr Leu Leu
        595                 600                 605

Thr Tyr Leu Leu Leu Pro Tyr Glu Leu Lys Gly Asp Phe Leu Gly Phe
    610                 615                 620

Val Lys Lys His Tyr Tyr Asp Ile Lys Asn Val Asp Phe Ile Asp Glu
625                 630                 635                 640

Asn Gln Asn Asn Ile Gln Val Ser Gln Thr Val Glu Lys Gln Glu Asp
                645                 650                 655

Tyr Phe Tyr His Lys Ile Arg Leu Phe Glu Lys Asn Thr Lys Lys Tyr
            660                 665                 670

Glu Ile Val Lys Tyr Ser Ile Val Pro Asn Glu Lys Leu Lys Gln Tyr
        675                 680                 685

Phe Glu Asp Leu Gly Ile Asp Ile Lys Tyr Leu Thr Gly Ser Val Glu
    690                 695                 700
```

```
Ser Gly Glu Lys Trp Leu Gly Glu Asn Leu Gly Ile Asp Ile Lys Tyr
705                 710                 715                 720

Leu Thr Val Glu Gln Lys Ser Glu Val Ser Glu Glu Lys Asn Lys Lys
                725                 730                 735

Val Ser Leu Lys Asn Asn Gly Met Phe Asn Lys Thr Ile Leu Leu Phe
            740                 745                 750

Val Phe Lys Tyr Tyr Gln Ile Ala Phe Lys Leu Phe Asn Asp Ile Glu
        755                 760                 765

Leu Tyr Ser Leu Phe Phe Leu Arg Glu Lys Ser Gly Lys Pro Leu Glu
    770                 775                 780

Ile Phe Arg Lys Glu Leu Glu Ser Lys Met Lys Asp Gly Tyr Leu Asn
785                 790                 795                 800

Phe Gly Gln Leu Leu Tyr Val Val Tyr Glu Val Leu Val Lys Asn Lys
                805                 810                 815

Asp Leu Asp Lys Ile Leu Ser Lys Lys Ile Asp Tyr Arg Lys Asp Lys
            820                 825                 830

Ser Phe Ser Pro Glu Ile Ala Tyr Leu Arg Asn Phe Leu Ser His Leu
        835                 840                 845

Asn Tyr Ser Lys Phe Leu Asp Asn Phe Met Lys Ile Asn Thr Asn Lys
    850                 855                 860

Ser Asp Glu Asn Lys Glu Val Leu Ile Pro Ser Ile Lys Ile Gln Lys
865                 870                 875                 880

Met Ile Gln Phe Ile Glu Lys Cys Asn Leu Gln Asn Gln Ile Asp Phe
                885                 890                 895

Asp Phe Asn Phe Val Asn Asp Phe Tyr Met Arg Lys Glu Lys Met Phe
            900                 905                 910
```

```
Phe Ile Gln Leu Lys Gln Ile Phe Pro Asp Ile Asn Ser Thr Glu Lys
        915                 920                 925


Gln Lys Met Asn Glu Lys Glu Glu Ile Leu Arg Asn Arg Tyr His Leu
    930                 935                 940


Thr Asp Lys Lys Asn Glu Gln Ile Lys Asp Glu His Glu Ala Gln Ser
945                 950                 955                 960


Gln Leu Tyr Glu Lys Ile Leu Ser Leu Gln Lys Ile Tyr Ser Ser Asp
                965                 970                 975


Lys Asn Asn Phe Tyr Gly Arg Leu Lys Glu Glu Lys Leu Leu Phe Leu
            980                 985                 990


Glu Lys Gln Gly Lys Lys Lys Leu  Ser Met Glu Glu Ile  Lys Asp Lys
        995                 1000                1005


Ile Ala  Gly Asp Ile Ser Asp  Leu Leu Gly Ile Leu  Lys Lys Glu
    1010                1015                1020


Ile Thr  Arg Asp Ile Lys Asp  Lys Leu Thr Glu Lys  Phe Arg Tyr
    1025                1030                1035


Cys Glu  Glu Lys Leu Leu Asn  Leu Ser Phe Tyr Asn  His Gln Asp
    1040                1045                1050


Lys Lys  Lys Glu Glu Ser Ile  Arg Val Phe Leu Ile  Arg Asp Lys
    1055                1060                1065


Asn Ser  Asp Asn Phe Lys Phe  Glu Ser Ile Leu Asp  Asp Gly Ser
    1070                1075                1080


Asn Lys  Ile Phe Ile Ser Lys  Asn Gly Lys Glu Ile  Thr Ile Gln
    1085                1090                1095


Cys Cys  Asp Lys Val Leu Glu  Thr Leu Ile Ile Glu  Lys Asn Thr
    1100                1105                1110
```

```
Leu Lys  Ile Ser Ser Asn Gly  Lys Ile Ile Ser Leu  Ile Pro His
    1115                1120                1125


Tyr Ser  Tyr Ser Ile Asp Val  Lys Tyr
    1130                1135


<210>  773
<211>  1158
<212>  PRT
<213>  Fusobacterium varium

<400>  773

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


Ile Ser Asn Tyr Asp Met Lys Gly Ile Lys Glu Leu Lys Val Arg Tyr
            20                  25                  30


Arg Lys Gln Ala Gln Leu Asp Thr Phe Ile Leu Gln Thr Thr Leu Asp
        35                  40                  45


Lys Ser Asn Asn Ser Ile Leu Ile Asn Asp Phe Arg Val Lys Ala Arg
    50                  55                  60


Glu Lys Tyr Arg Tyr Ser Phe Thr Tyr Asp Gly Lys Glu Lys Phe Ser
65                  70                  75                  80


Val Pro Ser Asn Ser Ile Ile Val Thr Lys Ile Asp Asn Ala Ala Pro
                85                  90                  95


Glu Lys Ser Lys Glu Ile Arg Lys Tyr Lys Ile Thr Leu Gly Ile Asp
            100                 105                 110


Glu Lys Cys Lys Thr Gly Ser Met Ile Thr Ala Ala Ile Glu Asp Leu
        115                 120                 125


Leu Glu Asp Asp Arg Val Arg Glu Gly Ile Arg Asn Pro Arg Arg Lys
    130                 135                 140
```

```
Ala Ser Lys Thr Glu Arg Lys Leu Ile Thr Glu Ser Ile Cys His Asn
145                 150                 155                 160

Tyr Ala Gln Ile Thr Gln Cys Pro Val Glu Glu Ile Asp Ala Val Lys
                165                 170                 175

Ile Tyr Lys Val Lys Arg Phe Leu Ser Tyr Arg Ser Asn Met Leu Leu
            180                 185                 190

Phe Phe Ala Leu Ile Asn Asp Phe Leu Cys Lys Asn Leu Lys Asn Glu
        195                 200                 205

Lys Gly Glu Lys Ile Asn Glu Ile Trp Glu Met Glu Asn Lys Gly Asn
    210                 215                 220

Asn Lys Lys Ile Asp Phe Asp Glu Asn Tyr Asn Ile Leu Val Ala Gln
225                 230                 235                 240

Ile Lys Glu Tyr Phe Thr Lys Glu Ile Glu Asn Tyr Asn Asn Arg Ile
                245                 250                 255

Asp Asn Ile Ile Asp Lys Lys Glu Leu Leu Lys Tyr Ser Glu Glu Lys
            260                 265                 270

Glu Glu Ser Glu Lys Asn Lys Lys Leu Glu Glu Leu Asn Lys Leu Glu
        275                 280                 285

Ser Gln Lys Leu Lys Ile Leu Thr Asp Glu Glu Ile Lys Ala Asp Val
    290                 295                 300

Ile Lys Ile Ile Lys Ile Phe Ser Asp Leu Arg His Ser Leu Met His
305                 310                 315                 320

Tyr Glu Tyr Lys Tyr Phe Glu Asn Leu Phe Glu Asn Lys Lys Asn Glu
                325                 330                 335

Glu Leu Ala Glu Leu Leu Asn Leu Asn Leu Phe Lys Asn Leu Thr Leu
            340                 345                 350
```

```
Leu Arg Gln Met Lys Ile Glu Asn Lys Thr Asn Tyr Leu Glu Gly Asp
        355                 360                 365

Glu Lys Phe Asn Ile Leu Gly Lys Asp Val Arg Ala Lys Asn Ala Leu
    370                 375                 380

Gly His Tyr Asp Leu Leu Val Glu Gln Lys Asn Gly Phe Asn Asn Phe
385                 390                 395                 400

Ile Asn Ser Phe Phe Val Gln Asp Gly Thr Glu Asn Leu Glu Phe Lys
                405                 410                 415

Lys Phe Ile Asp Glu Asn Phe Ile Lys Ala Gln Lys Glu Leu Glu Glu
            420                 425                 430

Asp Ile Lys Asn Cys Lys Glu Ser Val Lys Lys Leu Glu Lys Lys Leu
        435                 440                 445

Lys Glu Asn Pro Lys Lys Ser Glu Asp Leu Glu Lys Lys Leu Glu Lys
    450                 455                 460

Lys Gln Lys Lys Leu Lys Glu Leu Lys Lys Glu Leu Glu Lys Met Lys
465                 470                 475                 480

Gln His Tyr Lys Arg Leu Asn Cys Ala Tyr Val Trp Asp Ile His Ser
                485                 490                 495

Ser Thr Val Tyr Lys Lys Leu Tyr Asn Glu Arg Lys Asn Leu Ile Glu
            500                 505                 510

Lys Tyr Asn Lys Gln Leu Asn Gly Leu Gln Asp Lys Asn Ala Ile Thr
        515                 520                 525

Gly Ile Asn Ala Gln Leu Leu Arg Ile Lys Lys Glu Met Glu Glu Ile
    530                 535                 540

Thr Lys Ser Asn Ser Leu Phe Arg Leu Lys Tyr Lys Met Gln Ile Ala
545                 550                 555                 560
```

```
Tyr Ala Phe Leu Glu Met Glu Tyr Glu Gly Asn Ile Ala Lys Phe Lys
                565                 570                 575

Asn Glu Phe Asp Cys Ser Lys Thr Glu Lys Ile Gln Glu Trp Leu Glu
            580                 585                 590

Lys Ser Glu Glu Tyr Leu Asn Tyr Cys Met Glu Lys Glu Glu Asp Gly
        595                 600                 605

Lys Asn Tyr Lys Phe His Phe Lys Glu Ile Ser Glu Ile Lys Asp Thr
    610                 615                 620

His Asn Glu Glu Trp Leu Glu Asn Thr Ser Glu Asn Asn Leu Phe Lys
625                 630                 635                 640

Phe Tyr Ile Leu Thr Tyr Leu Leu Leu Pro Met Glu Phe Lys Gly Asp
                645                 650                 655

Phe Leu Gly Val Val Lys Lys His Tyr Tyr Asp Ile Lys Asn Val Asp
            660                 665                 670

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

Met Ile Gly Asp Ser Phe Phe His Lys Ile Arg Leu Phe Glu Lys Asn
    690                 695                 700

Thr Lys Arg Tyr Glu Ile Ile Lys Tyr Ser Ile Leu Thr Ser Asp Glu
705                 710                 715                 720

Ile Lys Lys Tyr Phe Glu Leu Leu Glu Leu Lys Val Pro Tyr Leu Glu
                725                 730                 735

Tyr Lys Gly Ile Asp Glu Ile Gly Ile Phe Asn Lys Asn Ile Ile Leu
            740                 745                 750

Pro Ile Phe Lys Tyr Tyr Gln Ile Ile Phe Arg Leu Tyr Asn Asp Leu
        755                 760                 765
```

```
Glu Ile His Gly Leu Phe Asn Val Ser Phe Asp Ile Asn Lys Ile Leu
    770                 775                 780

Ser Asp Leu Lys Ser Tyr Gly Asn Glu Asn Ile Asn Phe Arg Glu Phe
785                 790                 795                 800

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

Gln Lys Ile Trp Glu Lys Leu Glu Ser Lys Tyr Leu Lys Glu Pro Leu
            820                 825                 830

Leu Thr Pro Glu Lys Lys Glu Ile Asn Lys Lys Thr Glu Lys Glu Leu
        835                 840                 845

Lys Lys Leu Asp Gly Ile Ser Phe Leu Arg Asn Lys Ile Ser His Leu
    850                 855                 860

Glu Tyr Glu Lys Ile Ile Glu Gly Val Leu Lys Thr Ala Val Asn Gly
865                 870                 875                 880

Glu Asn Lys Lys Thr Ser Glu Thr Asn Ala Asp Lys Val Phe Leu Asn
                885                 890                 895

Glu Lys Ile Lys Lys Ile Ile Asn Phe Ile Lys Glu Asn Glu Leu Asp
            900                 905                 910

Lys Ile Glu Leu Gly Phe Asn Phe Ile Asn Asp Phe Phe Met Lys Lys
        915                 920                 925

Glu Gln Phe Met Phe Gly Gln Ile Lys Gln Val Lys Glu Gly Asn Ser
    930                 935                 940

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

Leu Lys Ile Thr Tyr Gly Leu Asn Tyr Asn Asn Leu Ser Lys Ile Tyr
                965                 970                 975
```

```
Glu Phe Ser Asn Thr Leu Arg Glu Ile Val Asn Ser Pro Leu Phe Leu
            980                 985                 990


Lys Asp Ser Thr Leu Leu Lys Lys  Val Asp Leu Ser Lys  Val Met Leu
        995                 1000                1005


Lys Glu  Lys Pro Ile Cys Ser  Leu Gln Tyr Glu Asn  Asn Thr Lys
    1010                1015                1020


Leu Glu  Asp Asp Ile Lys Arg  Ile Leu Leu Lys Asp  Ser Ser Asp
    1025                1030                1035


Ile Met  Gly Ile Tyr Lys Ala  Glu Val Val Lys Lys  Leu Lys Glu
    1040                1045                1050


Lys Leu  Val Leu Ile Phe Lys  Tyr Asp Glu Glu Lys  Lys Ile Tyr
    1055                1060                1065


Val Thr  Val Tyr Asp Thr Ser  Lys Ala Val Pro Glu  Asn Ile Ser
    1070                1075                1080


Lys Glu  Ile Leu Val Lys Arg  Asn Asn Ser Lys Glu  Glu Tyr Phe
    1085                1090                1095


Phe Glu  Asp Asn Lys Lys Lys  Tyr Thr Thr Gln Tyr  Tyr Thr Leu
    1100                1105                1110


Glu Ile  Thr Lys Glu Asn Glu  Leu Lys Val Ile Pro  Ala Lys Lys
    1115                1120                1125


Leu Glu  Gly Lys Glu Phe Lys  Thr Glu Lys Lys Glu  Glu Asn Lys
    1130                1135                1140


Leu Met  Leu Asn Asn His Tyr  Cys Phe Asn Val Lys  Ile Ile Tyr
    1145                1150                1155


<210>  774
<211>  50
<212>  DNA
```

<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 774
ttcatcttgg gcgtgcactt gatgtgggac aggcagatca gacagcccct 50

<210> 775
<211> 50
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 775
ttcatcttgg gcgtgccctt gatgtgggac aggcagatca gacagcccct 50

<210> 776
<211> 50
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 776
ttcatcttgg gcgtgctctt gatgtgggac aggcagatca gacagcccct 50

<210> 777
<211> 50
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 777
ttcatcttgg gcgtgcgctt gatgtgggac aggcagatca gacagcccct 50

<210> 778
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 778
gcacttgatg tgggacaggc agatcagaca 30

<210> 779
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 779
gcccttgatg tgggacaggc agatcagaca 30

<210> 780
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 780
gctcttgatg tgggacaggc agatcagaca 30

<210> 781
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 781
gcgcttgatg tgggacaggc agatcagaca 30

<210> 782
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 782
gttcatcttg ggcgtgcact tgatgtggga caggcagatc agacagcccc t 51

<210> 783
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 783
gttcatcttg ggcgtgccct tgatgtggga caggcagatc agacagcccc t 51

<210> 784
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 784
gttcatcttg ggcgtgctct tgatgtggga caggcagatc agacagcccc t 51

<210> 785
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 785
gttcatcttg ggcgtgcgct tgatgtggga caggcagatc agacagcccc t 51

<210> 786

<400> 786
000

<210> 787
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 787
ggcccttgat gtgggacagg cagatcagac a 31

<210> 788
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 788
ggctcttgat gtgggacagg cagatcagac a 31

<210> 789
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 789
ggcgcttgat gtgggacagg cagatcagac a 31

<210> 790
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 790
aggggctgtc tgatctgcct gtcccacatc aagtgcacgc ccaagatgaa c 51

<210> 791
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 791
aggggctgtc tgatctgcct gtcccacatc aagggcacgc ccaagatgaa c 51

<210> 792
<211> 51

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 792
aggggctgtc tgatctgcct gtcccacatc aagagcacgc ccaagatgaa c 51

<210> 793
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 793
aggggctgtc tgatctgcct gtcccacatc aagcgcacgc ccaagatgaa c 51

<210> 794
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 794
tgtctgatct gcctgtccca catcaagtgc c 31

<210> 795
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 795
tgtctgatct gcctgtccca catcaagggc c 31

<210> 796
<211> 31
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 796
tgtctgatct gcctgtccca catcaagagc c 31

<210> 797
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 797
tgtctgatct gcctgtccca catcaagcgc c 31

<210> 798
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 798
caccgttcat cttgggcgtg cacttgatgt gggacaggca gatcagacag cccct 55

<210> 799
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 799
caccgttcat cttgggcgtg cccttgatgt gggacaggca gatcagacag cccct 55

<210> 800
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 800
caccgttcat cttgggcgtg ctcttgatgt gggacaggca gatcagacag cccct 55

<210> 801
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 801
caccgttcat cttgggcgtg cgcttgatgt gggacaggca gatcagacag cccct 55

<210> 802
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 802
caccggcact tgatgtggga caggcagatc agaca 35

<210> 803
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 803
caccggccct tgatgtggga caggcagatc agaca 35

<210> 804
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 804
caccggctct tgatgtggga caggcagatc agaca 35

<210> 805
<211> 35

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 805
caccggcgct tgatgtggga caggcagatc agaca 35

<210> 806
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 806
caacaggggc tgtctgatct gcctgtccca catcaagtgc acgcccaaga tgaac 55

<210> 807
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 807
caacaggggc tgtctgatct gcctgtccca catcaagggc acgcccaaga tgaac 55

<210> 808
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 808
caacaggggc tgtctgatct gcctgtccca catcaagagc acgcccaaga tgaac 55

<210> 809
<211> 55
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 809
caacaggggc tgtctgatct gcctgtccca catcaagcgc acgcccaaga tgaac 55

<210> 810
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 810
caactgtctg atctgcctgt cccacatcaa gtgcc 35

<210> 811
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 811
caactgtctg atctgcctgt cccacatcaa gggcc 35

<210> 812
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 812
caactgtctg atctgcctgt cccacatcaa gagcc 35

<210> 813
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 813
caactgtctg atctgcctgt cccacatcaa gcgcc 35

<210> 814
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 814
caccgttcat cttgggcgtg cacttgatgt gggacaggca gatcagacag cccct 55

<210> 815
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 815
caccgttcat cttgggcgtg cccttgatgt gggacaggca gatcagacag cccct 55

<210> 816
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 816
caccgttcat cttgggcgtg ctcttgatgt gggacaggca gatcagacag cccct 55

<210> 817
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 817
caccgttcat cttgggcgtg cgcttgatgt gggacaggca gatcagacag cccct 55

<210> 818
<211> 35

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 818
caccggcact tgatgtggga caggcagatc agaca 35

<210> 819
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 819
caccggccct tgatgtggga caggcagatc agaca 35

<210> 820
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 820
caccggctct tgatgtggga caggcagatc agaca 35

<210> 821
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 821
caccggcgct tgatgtggga caggcagatc agaca 35

<210> 822
<211> 55
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 822
caacaggggc tgtctgatct gcctgtccca catcaagtgc acgcccaaga tgaac 55

<210> 823
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 823
caacaggggc tgtctgatct gcctgtccca catcaagggc acgcccaaga tgaac 55

<210> 824
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 824
caacaggggc tgtctgatct gcctgtccca catcaagagc acgcccaaga tgaac 55

<210> 825
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 825
caacaggggc tgtctgatct gcctgtccca catcaagcgc acgcccaaga tgaac 55

<210> 826
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 826
caactgtctg atctgcctgt cccacatcaa gtgcc 35

<210> 827
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 827
caactgtctg atctgcctgt cccacatcaa gggcc 35

<210> 828
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 828
caactgtctg atctgcctgt cccacatcaa gagcc 35

<210> 829
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 829
caactgtctg atctgcctgt cccacatcaa gcgcc 35

<210> 830
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 830
ctctttgtcg ccttcgtagg tgtggcagcg 30

<210> 831
<211> 30

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 831
gtcgccttcg taggtgtggc agcgtcctgg 30

<210> 832
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 832
ttcgtaggtg tggcagcgtc ctgggatgaa 30

<210> 833
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 833
ggtgtggcag cgtcctggga tgaacttctt 30

<210> 834
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 834
gcagcgtcct gggatgaact tcttcatctt 30

<210> 835
<211> 30
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 835
tcctgggatg aacttcttca tcttgggcgt 30

<210> 836
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 836
gatgaacttc ttcatcttgg gcgtgcgctt 30

<210> 837
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 837
cttcttcatc ttgggcgtgc gcttgatgtg 30

<210> 838
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 838
catcttgggc gtgcgcttga tgtgggacag 30

<210> 839
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 839
gggcgtgcgc ttgatgtggg acaggcagat 30

<210> 840
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 840
gcgcttgatg tgggacaggc agatcagaca 30

<210> 841
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 841
gatgtgggac aggcagatca gacagcccct 30

<210> 842
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 842
ggacaggcag atcagacagc ccctggtgca 30

<210> 843
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 843
gcagatcaga cagcccctgg tgcagccagc 30

<210> 844
<211> 30

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 844
cagacagccc ctggtgcagc cagctttccg 30

<210> 845
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 845
gtaatgcctg gcttgtcgac gcatagtctg 30

<210> 846
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 846
gctctttgtc gccttcgtag gtgtggcagc g 31

<210> 847
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 847
ggtcgccttc gtaggtgtgg cagcgtcctg g 31

<210> 848
<211> 31
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 848
gttcgtaggt gtggcagcgt cctgggatga a 31

<210> 849
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 849
gggtgtggca gcgtcctggg atgaacttct t 31

<210> 850
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 850
ggcagcgtcc tgggatgaac ttcttcatct t 31

<210> 851
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 851
gtcctgggat gaacttcttc atcttgggcg t 31

<210> 852
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 852
ggatgaactt cttcatcttg ggcgtgcgct t 31

<210> 853
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 853
gcttcttcat cttgggcgtg cgcttgatgt g 31

<210> 854
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 854
gcatcttggg cgtgcgcttg atgtgggaca g 31

<210> 855
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 855
ggggcgtgcg cttgatgtgg gacaggcaga t 31

<210> 856
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 856
ggcgcttgat gtgggacagg cagatcagac a 31

<210> 857
<211> 31

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 857
ggatgtggga caggcagatc agacagcccc t 31

<210> 858
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 858
gggacaggca gatcagacag cccctggtgc a 31

<210> 859
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 859
ggcagatcag acagcccctg gtgcagccag c 31

<210> 860
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 860
gcagacagcc cctggtgcag ccagctttcc g 31

<210> 861
<211> 31
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 861
ggtaatgcct ggcttgtcga cgcatagtct g 31

<210> 862
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 862
cgctgccaca cctacgaagg cgacaaagag c 31

<210> 863
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 863
ccaggacgct gccacaccta cgaaggcgac c 31

<210> 864
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 864
ttcatcccag gacgctgcca cacctacgaa c 31

<210> 865
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 865
aagaagttca tcccaggacg ctgccacacc c 31

<210> 866
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 866
aagatgaaga agttcatccc aggacgctgc c 31

<210> 867
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 867
acgcccaaga tgaagaagtt catcccagga c 31

<210> 868
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 868
aagcgcacgc ccaagatgaa gaagttcatc c 31

<210> 869
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 869
cacatcaagc gcacgcccaa gatgaagaag c 31

<210> 870
<211> 31

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 870
ctgtcccaca tcaagcgcac gcccaagatg c 31

<210> 871
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 871
atctgcctgt cccacatcaa gcgcacgccc c 31

<210> 872
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 872
tgtctgatct gcctgtccca catcaagcgc c 31

<210> 873
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 873
aggggctgtc tgatctgcct gtcccacatc c 31

<210> 874
<211> 31
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 874
tgcaccaggg gctgtctgat ctgcctgtcc c 31

<210> 875
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 875
gctggctgca ccaggggctg tctgatctgc c 31

<210> 876
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 876
cggaaagctg gctgcaccag gggctgtctg c 31

<210> 877
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 877
cagactatgc gtcgacaagc caggcattac c 31

<210> 878
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 878
caccgctctt tgtcgccttc gtaggtgtgg cagcg 35

<210> 879
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 879
caccggtcgc cttcgtaggt gtggcagcgt cctgg 35

<210> 880
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 880
caccgttcgt aggtgtggca gcgtcctggg atgaa 35

<210> 881
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 881
caccgggtgt ggcagcgtcc tgggatgaac ttctt 35

<210> 882
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 882
caccggcagc gtcctgggat gaacttcttc atctt 35

<210> 883
<211> 35

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 883
caccgtcctg ggatgaactt cttcatcttg ggcgt 35

<210> 884
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 884
caccggatga acttcttcat cttgggcgtg cgctt 35

<210> 885
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 885
caccgcttct tcatcttggg cgtgcgcttg atgtg 35

<210> 886
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 886
caccgcatct tgggcgtgcg cttgatgtgg gacag 35

<210> 887
<211> 35
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 887
caccggggcg tgcgcttgat gtgggacagg cagat 35

<210> 888
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 888
caccggcgct tgatgtggga caggcagatc agaca 35

<210> 889
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 889
caccggatgt gggacaggca gatcagacag cccct 35

<210> 890
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 890
caccgggaca ggcagatcag acagcccctg gtgca 35

<210> 891
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 891
caccggcaga tcagacagcc cctggtgcag ccagc 35

<210> 892
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 892
caccgcagac agcccctggt gcagccagct ttccg 35

<210> 893
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 893
caccggtaat gcctggcttg tcgacgcata gtctg 35

<210> 894
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 894
caaccgctgc cacacctacg aaggcgacaa agagc 35

<210> 895
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 895
caacccagga cgctgccaca cctacgaagg cgacc 35

<210> 896
<211> 35

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 896
caacttcatc ccaggacgct gccacaccta cgaac 35

<210> 897
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 897
caacaagaag ttcatcccag gacgctgcca caccc 35

<210> 898
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 898
caacaagatg aagaagttca tcccaggacg ctgcc 35

<210> 899
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 899
caacacgccc aagatgaaga agttcatccc aggac 35

<210> 900
<211> 35
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic or Unknown

<400> 900
caacaagcgc acgcccaaga tgaagaagtt catcc 35

<210> 901
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 901
caaccacatc aagcgcacgc ccaagatgaa gaagc 35

<210> 902
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 902
caacctgtcc cacatcaagc gcacgcccaa gatgc 35

<210> 903
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 903
caacatctgc ctgtcccaca tcaagcgcac gcccc 35

<210> 904
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 904
caactgtctg atctgcctgt cccacatcaa gcgcc 35

<210> 905
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 905
caacaggggc tgtctgatct gcctgtccca catcc 35

<210> 906
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 906
caactgcacc aggggctgtc tgatctgcct gtccc 35

<210> 907
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 907
caacgctggc tgcaccaggg gctgtctgat ctgcc 35

<210> 908
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 908
caaccggaaa gctggctgca ccaggggctg tctgc 35

<210> 909
<211> 35

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic or Unknown

<400> 909
caaccagact atgcgtcgac aagccaggca ttacc 35

<210> 910
<211> 104
<212> RNA
<213> Cypridina luciferase

<400> 910
uguugccgaa ugugucagua uguaauugaa ugcagaguag aggccgcagg auaguuuaga 60

acauucuaug gaaagagauu ccaguuccag gaaccuggua caua 104

<210> 911
<211> 50
<212> RNA
<213> Cypridina luciferase

<400> 911
cauccugcgg ccucuacucu gcauucaauu acauacugac acauucggca 50

<210> 912
<211> 50
<212> RNA
<213> Cypridina luciferase

<400> 912
uucuaaacca uccugcggcc ucuacucugc auucaauuac auacugacac 50

<210> 913
<211> 50
<212> RNA
<213> Cypridina luciferase

<400> 913
auagaauguu cuaaaccauc cugcggccuc uacucugcau ucaauuacau 50

<210> 914
<211> 50
<212> RNA
<213> Cypridina luciferase

<400> 914
cucuuuccau agaauguucu aaaccauccu gcggccucua cucugcauuc 50

<210> 915
<211> 50
<212> RNA
<213> Cypridina luciferase

<400> 915
acuggaaucu cuuuccauag aauguucuaa accauccugc ggccucuacu 50

<210> 916
<211> 50
<212> RNA
<213> Cypridina luciferase

<400> 916
uuccuggaac uggaaucucu uuccauagaa uguucuaaac cauccugcgg 50

<210> 917
<211> 50
<212> RNA
<213> Cypridina luciferase

<400> 917
guaccagguu ccuggaacug gaaucucuuu ccauagaaug uucuaaacca 50

<210> 918
<211> 120
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 918
gacacagcag gtcaagagga gtacagtgca atgagggacc agtacatgag gactggggag 60

ggctttcttt gtgtatttgc cataaataat actaaatcat ttgaagatat tcaccattat 120