(12) STANDARD PATENT (11) Application No. **AU 2018320870 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
RNA targeting methods and compositions

(51) International Patent Classification(s)
C12N 9/22 (2006.01) ***C12N 15/55*** (2006.01)

(21) Application No: **2018320870** (22) Date of Filing: **2018.08.22**

(87) WIPO No: **WO19/040664**

(30) Priority Data

(31) Number	(32) Date	(33) Country
62/639,178	**2018.03.06**	**US**
62/548,846	**2017.08.22**	**US**
62/572,963	**2017.10.16**	**US**
15/937,699	**2018.03.27**	**US**

(43) Publication Date: **2019.02.28**
(44) Accepted Journal Date: **2025.04.17**

(71) Applicant(s)
Salk Institute for Biological Studies

(72) Inventor(s)
HSU, Patrick D.;KONERMANN, Silvana

(74) Agent / Attorney
Spruson & Ferguson, GPO Box 3898, Sydney, NSW, 2001, AU

(56) Related Art
WO 2018/183403 A1

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 February 2019 (28.02.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/040664 A1

(51) International Patent Classification:
C12N 9/22 (2006.01) *C12N 15/55* (2006.01)

(21) International Application Number:
PCT/US2018/047597

(22) International Filing Date:
22 August 2018 (22.08.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

62/548,846	22 August 2017 (22.08.2017)	US
62/572,963	16 October 2017 (16.10.2017)	US
62/639,178	06 March 2018 (06.03.2018)	US
15/937,699	27 March 2018 (27.03.2018)	US

(71) Applicant: **SALK INSTITUTE FOR BIOLOGICAL STUDIES** [US/US]; 10010 N. Torrey Pines Road, La Jolla, CA 92037-1002 (US).

(72) Inventors: **HSU, Patrick, D.**; 10010 N. Torrey Pines Road, La Jolla, CA 92037-1002 (US). **KONERMANN, Silvana**; 10010 N. Torrey Pines Road, La Jolla, CA 92037-1002 (US).

(74) Agent: **RYBAK, Sheree**; Klarquist Sparkman, LLP, One World Trade Center, Suite 1600, 121 SW Salmon Street, Portland, OR 97204 (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, 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).

Declarations under Rule 4.17:
— *of inventorship (Rule 4.17(iv))*

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: RNA TARGETING METHODS AND COMPOSITIONS

(57) Abstract: Provided herein are CRISPR/Cas methods and compositions for targeting RNA molecules, which be used to detect, edit, or modify a target RNA.

WO 2019/040664 A1

RNA TARGETING METHODS AND COMPOSITIONS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority to US Provisional Application No. 62/548,846 filed August 22, 2017, US Provisional Application No. 62/572,963 filed October 16, 2017, US Provisional Application No. 62/639,178, filed March 6, 2018, and US Application No. 15/937,699 filed March 27, 2018, all of which are herein incorporated by reference in their entireties.

FIELD

This disclosure relates to a CRISPR/Cas system for modifying (including detecting) RNA, which utilizes novel Cas13d proteins (also referred to CasR and nCas1) and guide RNAs.

ACKNOWLEDGMENT OF GOVERNMENT SUPPORT

This invention was made with government support under 5 DP5 OD021369-02 and 5 R21 AG056811-02 awarded by The National Institutes of Health. The Government has certain rights in the invention.

BACKGROUND

Mapping of transcriptome changes in cellular function and disease has been transformed by technological advances over the last two decades, from microarrays (Schena et al., 1995) to next-generation sequencing and single cell studies (Shendure et al., 2017). However, interrogating the function of individual transcript dynamics and establishing causal linkages between observed transcriptional changes and cellular phenotype requires the ability to actively control or modulate desired transcripts.

DNA engineering technologies such as CRISPR-Cas9 (Doudna and Charpentier, 2014; Hsu et al., 2014) enable researchers to dissect the function of specific genetic elements or correct disease-causing mutations. However, simple and scalable tools to study and manipulate RNA lag significantly behind their DNA counterparts. Existing RNA interference technologies, which enable cleavage or inhibition of desired transcripts, have significant off-target effects and remain challenging engineering targets due to their key role in endogenous processes (Birmingham et al., 2006; Jackson et al., 2003). As a result, methods for studying the functional role of RNAs directly have remained limited.

One of the key restrictions in RNA engineering has been the lack of RNA-binding domains that can be easily retargeted and introduced into target cells. The MS2 RNA-binding domain, for example, recognizes an invariant 21-nucleotide (nt) RNA sequence (Peabody, 1993), therefore requiring genomic modification to tag a desired transcript. Pumilio homology domains possess modular repeats with each protein module recognizing a separate RNA base, but they can only be targeted to short 8 nt RNA sequences (Cheong and Hall, 2006). While previously characterized type II (Batra et al., 2017; O'Connell et al., 2014) and VI (Abudayyeh et al., 2016; East-Seletsky et al., 2016) CRISPR-Cas systems can be reprogrammed to

recognize 20-30 nt RNAs, their large size (~1200 amino acids, aa) makes it difficult to package into AAV for primary cell and *in vivo* delivery.

SUMMARY

This application provides bioinformatic analysis of prokaryotic genomes to identify sequence signatures of CRISPR-Cas repeat arrays and mine previously uncharacterized, compact Cas ribonucleases that can be used for RNA targeting tools. Engineered Type VI-D CRISPR effectors can be used to efficiently knockdown endogenous RNAs in human cells and manipulate alternative splicing, paving the way for RNA targeting applications and further effector domain fusions as part of a transcriptome engineering toolbox.

Provided herein are methods of modifying one or more target RNA molecules, such as a clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR associated (Cas) system-mediated RNA editing method. Such methods can include contacting one or more target RNA molecules with a non-naturally occurring (*e.g.*, does not naturally occur in the cell or system into which it is introduced) or engineered CRISPR-Cas system. Such a CRISPR-Cas system can include (1) at least one Cas13d protein or at least one Cas13d nucleic acid coding sequence (such as a mRNA or a vector encoding the at least one Cas13d protein) provided herein and (2) at least one CRISPR-Cas system guide nucleic acid molecule (such as a guide RNA, gRNA) that hybridizes with the one or more target RNA molecules, or at least one nucleic acid molecule encoding the gRNA. The Cas13d protein forms a complex with the gRNA, and the gRNA directs the complex to the one or more target RNA molecules and modifies (*e.g.*, cuts, detects) the one or more target RNA molecules. In some examples, the one or more target RNA molecules (or a cell containing the one or more target RNA molecules) are contacted with a complex including the at least one Cas13d protein and the at least one gRNA. In some examples, the system includes Mg^{2+}. However, in some examples, the system does not include Mg^{2+}, such as if cleavage of the target RNA is not desired.

In some examples, contacting the one or more target RNA molecules with the non-naturally occurring or engineered CRISPR-Cas system includes introducing into a cell (such as a eukaryotic or prokaryotic cell) containing the one or more target RNA molecules the non-naturally occurring or engineered CRISPR-Cas system, for example using endocytosis, a liposome, a particle, an exosome, a microvesicle, a gene gun, electroporation, a virus, or combinations thereof. In some examples, contacting the one or more target RNA molecules with the non-naturally occurring or engineered CRISPR-Cas system includes contacting a cell-free system (such as a biological or environmental sample, or a cell lysate) containing the one or more target RNA molecules with the non-naturally occurring or engineered CRISPR-Cas system (for example in a diagnostic method to detect a target RNA).

In some examples, the least one Cas13d protein (1) includes one or more HEPN domains, (2) is no more than 150 kD (such as no more than 140 kD, no more than 130 kD, no more than 120 kD, such as about 90 to 120 kD, about 100 to 120 kD or about 110 kD), (3) includes one or more mutated HEPN domains, and can process the guide RNA, but cannot cleave or cut the one or more target RNA molecules, (4) includes a

Cas13d ortholog from a prokaryotic genome or metagenome, gut metagenome, an activated sludge metagenome, an anaerobic digester metagenome, a chicken gut metagenome, a human gut metagenome, a pig gut metagenome, a bovine gut metagenome, a sheep gut metagenome, a goat gut metagenome, a capybara gut metagenome, a primate gut metagenome, a termite gut metagenome, a fecal metagenome, a genome from the Order Clostridiales, or the Family Ruminococcaceae, (5) includes a Cas13d ortholog from Ruminococcus albus, Eubacterium siraeum, a flavefaciens strain XPD3002, Ruminococcus flavefaciens FD-1, uncultured Eubacterium sp TS28-c4095, uncultured Ruminococcus sp., Ruminococcus bicirculans, or Ruminococcus sp CAG57, (6) includes at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, 296 or (7) combinations thereof (*e.g.*, combinations of any of 1 to 6). In some examples, the least one Cas13d protein has at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296 and includes at least one motif shown in SEQ ID NO: 195, 196 or 197. In some examples, the least one Cas13d protein has at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296, and includes at least one consensus sequence shown in SEQ ID NO: 288, 289, 290 or 291. In some examples, the least one Cas13d protein further includes one or more other agents (*e.g.*, is a fusion protein),

such as one or more subcellular localization signals, one or more effector domains, or combinations thereof. In some examples, the least one Cas13d protein that includes one or more HEPN domains, is no more than 1500 aa, no more than 1200 aa, no more than 1100 aa, no more than 1000 aa, such as about 800 to 1500 aa, about 800 to 1250 aa or about 850 to 950 aa.

Also provided are isolated nucleic acid molecules encoding such Cas13d proteins, such as a cDNA, genomic DNA, RNA, or mRNA. Such isolated nucleic acid molecules can be part of a vector (such as a plasmid or viral vector), and can be operably linked to a promoter or enhancer element (or both). In some examples, an isolated nucleic acid molecule encoding a Cas13d protein has at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 124, 125, 126, 127, 128, 139, 140 or 141. In some examples, an isolated nucleic acid molecule encoding at least one Cas13d protein (which can be part of a vector) includes at least one Cas13d protein coding sequence codon optimized for expression in a eukaryotic cell, such as human cell, for example a Cas13d coding sequence having at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 142, 143, 144, or 145.

In some examples, the gRNA that hybridizes with the one or more target RNA molecules in a Cas13d-mediated manner includes one or more direct repeat (DR) sequences, one or more spacer sequences, such as one or more sequences comprising an array of DR-spacer-DR-spacer. In some examples, the one or more DR sequences have at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 129, 130, 131, 132, 133, 134, 135, 136, 137, 148, 150, 151, 152, 154, 156, 157, 159, 161, 163, 165, 167, 169, 176, 178, 180, 182, 184, 186, 188, 190, 191, 192, 193, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, or 254. In one example, the gRNA includes additional sequences, such as an aptamer sequence (*e.g.*, MS2 stem loops).

In some examples, a plurality of gRNAs are generated from a single array, wherein each gRNA can be different, for example target different RNAs or target multiple regions of a single RNA, or combinations thereof.

Methods of targeting one or more target RNA molecules are provided. In some examples, an entire RNA is targeted. In some examples, a portion of an RNA is targeted. Targeting an RNA molecule can include one or more of cutting or nicking one or more target RNA molecules, activating one or more target RNA molecules, deactivating the one or more target RNA molecules, visualizing or detecting the one or more target RNA molecules, labeling the one or more target RNA molecules, binding the one or more target RNA molecules, editing the one or more target RNA molecules, trafficking the one or more target RNA molecules, and masking the one or more target RNA molecules. In some example, modifying one or more target RNA molecules includes one or more of an RNA base substitution, an RNA base deletion, an RNA base insertion, a break in the target RNA, methylating RNA, and demethylating RNA.

In some examples, such methods are used to treat a disease, such as a disease in a human. In such examples, the one or more target RNA molecules is associated with the disease.

Also provided are isolated proteins, including non-naturally occurring proteins. In some examples, a protein has at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296. In some examples, an isolated Cas13d protein encompassed by this disclosure includes a sequence having at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 195, 196, 197, 288, 289, 290, or 291. In some examples, an isolated protein is a Cas13d ortholog from a prokaryotic genome or metagenome, gut metagenome, an activated sludge metagenome, an anaerobic digester metagenome, a chicken gut metagenome, a human gut metagenome, a pig gut metagenome, a bovine gut metagenome, a sheep gut metagenome, a goat gut metagenome, a capybara gut metagenome, a primate gut metagenome, a termite gut metagenome, a fecal metagenome, a genome from the Order Clostridiales, or the Family Ruminococcaceae. In some examples, a Cas13d ortholog includes a Cas13d ortholog from Ruminococcus albus, Eubacterium siraeum, a Ruminococcus flavefaciens strain XPD3002, Ruminococcus flavefaciens FD-1, uncultured Eubacterium sp TS28-c4095, uncultured Ruminococcus sp., Ruminococcus bicirculans, or Ruminococcus sp CAG57. The Cas13d protein can further include one or more other agents or domains (*e.g.*, is a fusion protein), such as one or more subcellular localization signals, one or more effector domains, or combinations thereof.

Also provided are isolated guide RNA (gRNA) molecules. In some examples, an isolated gRNA includes one or more direct repeat (DR) sequences, such as an unprocessed (*e.g.*, about 36 nt) or processed DR (*e.g.*, about 30 nt). In some examples a DR has at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 129, 130, 131, 132, 133, 134, 135, 136, 137, 148, 150, 151, 152, 154, 156, 157, 159, 161, 163, 165, 167, 169, 176, 178, 180, 182, 184, 186, 188, 190, 191, 192, 193, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, or 254. Such a gRNA can further include one or more spacer sequences specific for (*e.g.*, is complementary to) the target RNA.

Also provided are ribonucleoprotein (RNP) complexes, which include a Cas13d protein provided herein and a gRNA provided herein.

2018320870 28 Feb 2025

Also provided are recombinant cells that include any Cas13d protein (or nucleic acid molecule encoding Cas13d), any gRNA, any RNP complex, or any vector, provided herein. In one example, the cell is not a bacterial cell. In one example, the cell is a bacterial cell.

Also provided are compositions that include one or more of any Cas13d protein (or nucleic acid molecule encoding Cas13d), any gRNA, any RNP complex, any isolated nucleic acid molecule, any vector, or any cell, provided herein. Such compositions can include a pharmaceutically acceptable carrier.

Also provided are kits. Such kits can include one or more of any Cas13d protein (or nucleic acid molecule encoding Cas13d), any gRNA, any RNP complex, any isolated nucleic acid molecule, any vector, any cell, or any composition provided herein. Such reagents can be combined or in separate containers.

The foregoing and other objects and features of the disclosure will become more apparent from the following detailed description, which proceeds with reference to the accompanying figures.

Any reference to publications cited in this specification is not an admission that the disclosures constitute common general knowledge in Australia.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGS. 1A-1B: Bioinformatic pipeline for the identification of the RNA-targeting class 2 CRISPR system Cas13d. (A) Schematic describing a computational pipeline for CRISPR system identification. A minimal definition for a putative class 2 CRISPR locus was used, requiring only a CRISPR repeat array and a nearby protein >750 aa in length. The initial search was performed on prokaryotic genome assemblies derived from NCBI Genome, and later expanded via TBLASTN of predicted Cas13d proteins against public metagenome sequences without predicted open reading frames. DR, direct repeat. (B) Phylogenetic classification and alignment of full-length Cas13d effectors and metagenomic fragments. Cas13d effectors and metagenomic Cas13d protein fragments cluster into several distinct branches, which are colored for ease of interpretation. Shading indicates residue conservation using the Blosum62 matrix. Full-length Cas13d effectors used in this study were sampled from distinct branches of the Cas13d family. Alignment of Cas13d proteins and protein fragments was performed using ClustalOmega 1.2.4 and maximum-likelihood tree building was performed with PhyML 3.2.

FIGS. 2A-2C: Type VI CRISPR-Cas13d is a family of single effector CRISPR ribonucleases. (A) Maximum-likelihood phylogenetic tree of Cas13d effectors used herein, with the full Cas13d CRISPR locus depicted along with conserved HEPN RNase domains. Grey rectangles denote CRISPR direct repeats (DRs) and blue diamonds indicate spacer sequences. (B) RNA sequencing of a heterologously expressed Cas13d locus from an uncultured *Ruminococcus sp.* sample. Mature gRNAs mapping to the CRISPR array indicate a processed 30 nt DR and a variable spacer length from 14-26 nt. Co-fold analysis of direct repeat truncation indicates a strong hairpin structure. (C) Purified *E. siraeum* Cas13d and catalytically dead Cas13d (dCas13d) protein are each sufficient to process a guide array into its two component gRNAs. Addition of EDTA does not impair gRNA processing. 'd', dCas13d(R295A, H300A, R849A, H854A).

2018320870 28 Feb 2025

FIGS. 3A-3D. Phylogenetic classification of RNA-targeting class 2 CRISPR effectors and sequence conservation within the Cas13d family. (A) HEPN motif conservation in Cas13d effectors used herein with conserved residues shaded according to Blosum62. The *RxxxxH* HEPN motif is highlighted. (B)

Maximum-likelihood tree of type VI CRISPR-Cas families. Average amino acid lengths of Type VI Cas13 superfamily effectors are indicated in red. Alignment of previously described class 2 CRISPR RNA-targeting proteins (Abudayyeh et al., 2017; Cox et al., 2017; East-Seletsky et al., 2017; East-Seletsky et al., 2016; Smargon et al., 2017) and Cas13d effectors was performed using MAFFT 7.38 and maximum-likelihood tree building was performed with PhyML 3.2. Branch labels and scale bar indicate substitutions per site. (C) Predicted Cas13d direct repeat RNA secondary structure. (D) Sequence logo of full length 36 nt Cas13d direct repeats.

FIGS. 4A-4C: **Purification of recombinant Cas13d protein**. EsCas13d was expressed as N-terminal His-MBP fusion and purified by successive affinity, cation exchange, and size exclusion chromatography. The His-tag was removed by TEV protease cleavage. (A) Chromatogram from Superdex 200 column for EsCas13d. (B) SDS-PAGE gel of size exclusion chromatography fractions for *E. siraeum* Cas13d. (C) SDS-PAGE gel of purified *E. siraeum* Cas13d and dCas13d (R295A, H300A, R849A, H854A mutations of predicted catalytic residues in both HEPN motifs).

FIGS. 5A-5D: Programmable RNA targeting by Cas13d *in vitro*. (A) *E. siraeum* Cas13d requires a matching guide array or mature gRNA to efficiently cleave complementary ssRNA targets. Denaturing gel depicts cleavage reactions incubated at 37°C for 1 hour. NT, non-targeting. (B) Substitution with dCas13d or addition of EDTA abrogate Cas13d-mediated RNA targeting with both the guide and array. 'd', dead Cas13d. (C) Denaturing gel depicting guide-target match dependent activation of Cas13d cleavage activity. Scrambled target RNA ('A') is fluorescently labeled, while guide-complementary activator target RNA ('B') is unlabeled. RNA cleavage activity is abolished by the individual removal of guide RNA or complementary target RNA, as well as the addition of EDTA or the catalytic inactivation of Cas13d (indicated as 'd'). (D) A model for guide and target-dependent activation of Cas13d RNase activity. The ternary Cas13d:gRNA:target RNA complex is capable of cleaving the complementary target RNA or bystander RNAs.

FIGS. 6A-6H. *In vitro* characterization of Cas13d properties. (A) Schematic showing the length and sequence of gRNA spacer truncations and spacer position relative to the complementary ssRNA target. (B) Denaturing gel depicting EsCas13d cleavage activity of target RNA with different spacer lengths. (C) Denaturing gel depicting EsCas13d cleavage reactions paired with 12 guides from Figure 3A tiling a complementary ssDNA version of the ssRNA target. (D) Denaturing gel depicting cleavage reactions using EsCas13d paired with the same 12 guides tiling a dsDNA version of the complementary target. (E) Quantification of cleavage efficiency from Figure 3A. Each PFS base is the average of 3 different spacer sequences tiling a complementary target RNA. Cleavage percentage is determined by the ratio of cleaved band intensity divided by total lane intensity. Mean is depicted ± SD with each data point representing an independent replicate. (F) Cas13d-mediated cleavage of target RNA carrying different PFS bases given an invariant spacer sequence. Quantification of Cas13d cleavage efficiency and a representative denaturing gel depicting EsCas13d cleavage activity are shown. Differences are not significant (one-way ANOVA, P=0.768). Cleavage percentage is determined as above, and mean is depicted ± SD with n = 3. (G) and (H)

Optimal temperature range for Cas13d activity. Denaturing gels depicting EsCas13d cleavage activity at temperatures ranging from 16-62°C for two different target RNAs.

FIGS. 7A-7B: Characterization of Cas13d target substrate preference. (A) Cas13d can be generalizably reprogrammed with multiple guides and does not exhibit a protospacer flanking sequence (PFS) requirement. RNA cleavage by EsCas13d and 12 guides tiling the target RNA is shown. Control lanes are from a separate gel run in parallel. (B) Cas13d preferentially cleaves uracil bases in the loop of a hairpin or a linear homopolymer repeat, which is interrupted every 5 nt by a transition mutation (X) to enable synthesis.

FIGS. 8A-8D: RNA knockdown activity screen of engineered Cas13d orthologs in human cells. (A) Schematic for mammalian expression constructs encoding for engineered Cas13d effectors and guides. NLS, nuclear localization signal. pre-gRNA, artificial unprocessed guide RNA containing a single 30 nt spacer sequence flanked by 2 full length 36 nt DRs. gRNA, predicted mature guide RNA with a single 30 nt processed DR and 22 nt spacer sequence (B) Heatmap of mCherry protein knockdown in a Cas13d ortholog activity screen in human HEK 293FT cells using pools of 4 pre-gRNAs or gRNAs. Normalized MFI, median fluorescent intensity relative to non-targeting condition. Positions in gray were not tested, with n = 3. (C) Immunocytochemistry of Cas13d showing localization and expression of engineered constructs. Scale bar, 10 μm. Blue pseudocolor, DAPI staining of nuclei. (D) Comparison of AdmCas13d (SEQ ID NO: 143) and RfxCas13d (SEQ ID NO: 280) ortholog constructs for knockdown of endogenous *B4GALNT1* mRNA reveals RfxCas13d-NLS (CasRx; SEQ ID NO: 280) to be most effective for both guide RNA architectures. Pools of 4 guides were used for targeting. NT, non-targeting. Values are mean ± SEM with n = 3.

FIGS. 9A-9H: CasRx mediates efficient and specific knockdown of diverse human coding and noncoding transcripts. (A) Multiple guide RNAs tiling a target transcript can be expressed as a single array and processed by RfxCas13d-NLS (CasRx; SEQ ID NO: 280) into individual gRNAs within the same cell. (B) Arrays of 4 guides each mediate target knockdown by CasRx in 293FT cells via transient transfection. Knockdown relative to GFP vehicle control was determined by qPCR, with n = 3. (C) Schematic of CasRx target sequences and spacer position-matched shRNAs. (D) Relative target RNA knockdown by individual position-matched shRNAs and CasRx gRNAs. NT, non-targeting. CRISPRi, dCas9-mediated CRISPR interference for transcriptional repression (n = 3) (E) Volcano plot of differential transcript levels between *B4GALNT1* targeting and non-targeting (NT) shRNAs as determined by RNA sequencing (n = 3). 542 non-specific transcript changes were identified. (F) Volcano plot of differential transcript levels between *B4GALNT1*-targeting CasRx and non-targeting (NT) guide. Targeting guide position is matched to the shRNA shown in (E). *B4GALNT1* was the only transcript exhibiting a significant change, with n = 3. (G) Summary of significant off-target transcript perturbations by matched shRNAs and CasRx guides. (H) CasRx targeting of 11 endogenous transcripts, each with 3 guides and a non-targeting (NT) guide in 293FT cells. Transcript levels are relative to GFP vehicle control, mean ± SEM with n = 3.

FIGS. 10A-10D. Comparison of engineered Cas13 superfamily effectors for targeted knockdown and splicing. (A) Relative target RNA knockdown by individual position-matched gRNAs for

CasRx, NLS-LwaCas13a-msfGFP (Abudayyeh et al., 2017) and PspCas13b-NES (Cox et al., 2017) in HEK 293FT cells. NT, non-targeting. Values are mean ± SEM with n = 3. (B) Comparison of Cas13 median knockdown efficiencies. n = 3 per guide RNA. **** indicates $P<0.0001$ according to Friedman's test. (C) Exon exclusion by catalytically inactive NLS-dCas13a-msfGFP on the bichromatic splicing reporter. Guides are position-matched to those reported in Figure 6B for CasRx. Values are mean ± SEM with n = 3. (D) Comparison of splicing modulation by NLS-dCas13a-msfGFP and CasRx. Fold change in targeted exon exclusion relative to non-targeting guide is shown. **** indicates $P<0.0001$ according to two-way ANOVA.

FIGS. 11A-11B. RNA sequencing from CasRx and shRNA targeting of *ANXA4* in human cells. (A) Volcano plots of differential transcript levels between *ANXA4* targeting and non-targeting (NT) shRNAs as determined by RNA sequencing (n = 3). 915 non-specific transcript changes were identified. (B) Volcano plot of differential transcript levels for an *ANXA4* targeting CasRx array used in FIG. 9B containing a guide position matched to the shRNA shown in (A) and a non-targeting (NT) array. *ANXA4* was the only transcript exhibiting significant downregulation with n = 3. *HIST2HBE* was the only transcript identified to exhibit significant upregulation. H2B is a dimer partner of H2AX (Du et al., 2006) which has been shown to interact with ANXA4 (Yang et al., 2010).

FIGS. 12A-12F: AAV delivery of catalytically inactive dCasRx splice effectors to manipulate alternative splicing. (A) Schematic of bichromatic exon skipping reporter. +1 and +3, reading frame. BP, intronic branch point-targeting guide. SA, splice acceptor site-overlapping guide. EX, exonic guide. SD, splice donor site-overlapping guide. AUG, start codon. UGA, stop codon. Inclusion of the second exon leads to an out-of-frame (+3), non-fluorescent translation of dsRed followed by in-frame mTagBFP2. Exclusion of the targeted exon leads to an in frame translation of dsRed (+1) followed by a stop codon. (B) Induced exon exclusion by dCasRx and an N-terminal hnRNPa1-dCasRx fusion protein targeted to pre-mRNA. The Gly-rich C-terminal domain of hnRNPa1 is used as the effector domain. Exon skipping efficiency is depicted as a relative percentage of cells carrying primarily the dsRed or BFP isoform, determined through flow cytometry. NLS, nuclear localization signal. 'A', CRISPR array carrying all 4 guides. Values are mean ± SEM with n = 3. (C) AAV design carrying dCasRx and a three-guide array with total transgene size < 4.3 kb, including AAV inverted terminal repeats (ITRs). (D) Schematic of frontotemporal dementia (FTD) disease modeling. Neurons are generated via *Neurogenin-2* (*Ngn2*) directed differentiation of patient-derived and control iPSCs followed by transduction with dCasRx or vehicle control AAV (EFS-mTagBFP2). (E) FTD is associated with SNPs in a putative intronic splice enhancer following exon 10 of the *MAPT* transcript encoding for tau. Alternative splicing of *MAPT* exon 10 results in 4R tau (by inclusion) and 3R tau (by exclusion). SNPs in the intronic splice enhancer including the indicated IVS 10 + 16 mutation result in increased exon inclusion and higher levels of 4R tau. To facilitate reduction of 4R tau levels, gRNAs contained in a dCasRx array were targeted to the exon 10 splice acceptor (g1) as well as two putative exonic splice enhancers indicated in purple (g2, g3). (F) Relative 4R/3R tau transcript ratios in differentiated neurons were assayed via qPCR at 14 days following transduction with AAV. FTD, frontotemporal dementia cells carrying IVS 10+16. Values are mean ± S.D. with n = 3. **** indicates $P < 0.0001$.

FIG. 13 is a bar graph showing RNA targeting in human cells using the disclosed methods.

FIG. 14 is a schematic drawing showing how the disclosed Cas13d and DRs can be used to achieve alternative splicing.

FIGS. 15A-15D are a series of panels showing: (A) Cas13d is converted into an active RNase complex upon binding a target matching the spacer sequence of the guide RNA. It is capable of cleaving gRNA-complementary target RNA or non-complementary bystander RNAs. (B) Cas13d target-dependent RNase activity can be converted into a detectable signal, for example through cleavage of a labeled detector RNA that is cleaved only in the presence of a target matching the spacer of the Cas13d guide RNA. In this example, the detector RNA contains a fluorophore, 'F', and a quencher 'Q', that abolishes fluorescence. Only upon bystander RNA cleavage is the fluorophore liberated from the quencher and fluorescence is generated. (C) Cas13d from *E. siraeum* produces a visible signal only in the presence of a perfectly matched target and not in the presence of different mismatched targets. (D) Cas13d from *R. flavefaciens* strain XPD3002 produces a visible signal only in the presence of a perfectly matched target and not in the presence of different mismatched targets.

FIGS. 16A-16B: (A) Alignment of 7 orthologs shows regions with high (green bars) and low (red bars) conservation. Regions selected for deletion are marked 1-10. (B) knock-down (top) and splicing (bottom) activity of full-length CasRx and CasRx deletion variants. Deletion variant 5 is shown to retain full activity, demonstrating the feasibility of deleting areas of low conservation while retaining full activity.

FIGS. 17A and 17B show targeting of ccdB in bacterial cells. (A) construct introduced into bacterial cells expressing ccdB, and (B) relative expression of ccdB under various conditions.

FIGS. 18A-18MMM show an alignment of 53 different Cas13d proteins (SEQ ID NOs: 293, 183, 189, 220, 218, 222, 229, 216, 177, 200, 139, 179, 208, 166, 185, 202, 239, 294, 249, 210, 243, 212, 175, 164, 160, 295, 296, 187, 241, 140, 162, 155, 206, 181, 231, 128, 198, 237, 233, 253, 214, 224, 127, 235, 153, 125, 245, 247, 204, 226, 3, 126, and 149, from top to bottom).

FIG. 19A is a plot showing results of screening 50 Cas13d orthologs (see Table 1) with NLS (SEQ ID NOS: 25 and 256) and NES architectures (SEQ ID NO: 287) for knockdown in human cells via mCherry reporter assay, plotted by knockdown efficiency and genetic size of effector. Only a subset of orthologs (Ga0531-NLS, Ga7274-NLS, and k87-NES, SEQ ID NOS: 235, 198, and 189, respectively show comparable activity to the CasRx variant (SEQ ID NO: 280).

FIG. 19B is a bar graph showing the knock-down of endogenous CD81 in human HEK 293FT cells by multiple Cas13d orthologs with NLS and NES subcellular localization signals. The most active orthologs from the mCherry reporter assay (see Table 1) were selected for a second round of activity screening. CD81 knockdown was assessed via cell surface flow cytometry. The CasRx variant RfxCas13d-NLS (SEQ ID NO: 280) exhibited the highest level of knock-down, with some additional orthologs exhibiting >75% knock-down on this endogenous target.

FIG. 20 is a bar graph showing a dose-response comparison of CasRx (SEQ ID NO: 280; RfxCas13d-NLS design) to PspCas13b-NES (Cox *et al.*, *Science*. 2017 Nov 24;358(6366):1019-1027) for

knock-down of endogenous CD81 protein. CasRx displays significantly higher activity at equivalent protein levels (as measured by Cas13-GFP expression), indicating a higher activity per enzyme unit and therefore a lower required dose to reach a biologically/therapeutically relevant level of knock-down. In addition, CasRx is significantly smaller than PspCas13b. The small size and high activity at lower doses makes CasRx useful for therapeutic delivery. Readout is three days after transient transfection in human HEK 293FT cells.

FIG. 21 is a bar graph showing that multiple engineered deletions of Cas13d orthologs CasRx and Ga0531 maintain knock-down activity in human HEK 293FT cells.

FIG. 22 is a bar graph showing that dCas13d variants (with mutated HEPN domains to abolish target RNA cleavage) were targeted to the slice acceptor site of our bichromatic blue-to-red splice reporter to mediate exon exclusion (as measured by reduction in mTagBFP2 expression). Even for orthologs with high levels of endogenous protein knock down (>70% CD81 protein knock down), splicing modulation can be inefficient. This underscores the uniquely advantageous combination of high knock-down and high splicing modulation activity of certain Cas13d orthologs, such as CasRx and truncated CasRx variants such as CasRx-del13. Splicing readout by flow cytometry on day 3 following transient transfection of dCas13d variants into HEK293FT cells.

FIGS. 23A-23C: RNA editing with Cas13d. (A) Schematic of dRfxCas13d-ADAR2 effector fusion and inactivated GFP reporter. (B) Targeted A to I RNA editing of a premature stop codon in an GFP reporter reactivates GFP expression. Non-targeting guide does not lead to RNA editing. (C) Quantification of Cas13d-mediated RNA editing efficiency.

SEQUENCE LISTING

The nucleic and amino acid sequences listed in the accompanying sequence listing are shown using standard letter abbreviations for nucleotide bases, and three letter code for amino acids, as defined in 37 C.F.R. 1.822. Only one strand of each nucleic acid sequence is shown, but the complementary strand is understood as included by any reference to the displayed strand. The Sequence Listing is submitted as an ASCII text file, created on August 22, 2018, 0.99MB, which is incorporated by reference herein. In the accompanying sequence listing:

SEQ ID NO: 1 is an exemplary Cas13d sequence from Eubacterium siraeum containing a HEPN site.

SEQ ID NO: 2 is an exemplary Cas13d sequence from Eubacterium siraeum containing a mutated HEPN site.

SEQ ID NO: 3 is an exemplary Cas13d sequence from uncultured Ruminococcus sp. containing a HEPN site.

SEQ ID NO: 4 is an exemplary Cas13d sequence from uncultured Ruminococcus sp. containing a mutated HEPN site.

SEQ ID NO: 5 is an exemplary Cas13d sequence from Gut_metagenome_contig2791000549.

SEQ ID NO: 6 is an exemplary Cas13d sequence from Gut_metagenome_contig855000317.

SEQ ID NO: 7 is an exemplary Cas13d sequence from Gut_metagenome_contig3389000027.

SEQ ID NO: 8 is an exemplary Cas13d sequence from Gut_metagenome_contig8061000170.

SEQ ID NO: 9 is an exemplary Cas13d sequence from Gut_metagenome_contig1509000299.

SEQ ID NO: 10 is an exemplary Cas13d sequence from Gut_metagenome_contig9549000591.

SEQ ID NO: 11 is an exemplary Cas13d sequence from Gut_metagenome_contig71000500.

SEQ ID NO: 12 is an exemplary Cas13d sequence from human gut metagenome.

SEQ ID NO: 13 is an exemplary Cas13d sequence from Gut_metagenome_contig3915000357.

SEQ ID NO: 14 is an exemplary Cas13d sequence from Gut_metagenome_contig4719000173.

SEQ ID NO: 15 is an exemplary Cas13d sequence from Gut_metagenome_contig6929000468.

SEQ ID NO: 16 is an exemplary Cas13d sequence from Gut_metagenome_contig7367000486.

SEQ ID NO: 17 is an exemplary Cas13d sequence from Gut_metagenome_contig7930000403.

SEQ ID NO: 18 is an exemplary Cas13d sequence from Gut_metagenome_contig993000527.

SEQ ID NO: 19 is an exemplary Cas13d sequence from Gut_metagenome_contig6552000639.

SEQ ID NO: 20 is an exemplary Cas13d sequence from Gut_metagenome_contig11932000246.

SEQ ID NO: 21 is an exemplary Cas13d sequence from Gut_metagenome_contig12963000286.

SEQ ID NO: 22 is an exemplary Cas13d sequence from Gut_metagenome_contig2952000470.

SEQ ID NO: 23 is an exemplary Cas13d sequence from Gut_metagenome_contig451000394.

SEQ ID NO: 24 is an exemplary Cas13d sequence from Eubacterium_siraeum_DSM_15702.

SEQ ID NO: 25 is an exemplary Cas13d sequence from gut_metagenome_P19E0k2120140920,_c369000003.

SEQ ID NO: 26 is an exemplary Cas13d sequence from Gut_metagenome_contig7593000362.

SEQ ID NO: 27 is an exemplary Cas13d sequence from Gut_metagenome_contig12619000055.

SEQ ID NO: 28 is an exemplary Cas13d sequence from Gut_metagenome_contig1405000151.

SEQ ID NO: 29 is an exemplary Cas13d sequence from Chicken_gut_metagenome_c298474.

SEQ ID NO: 30 is an exemplary Cas13d sequence from Gut_metagenome_contig1516000227.

SEQ ID NO: 31 is an exemplary Cas13d sequence from Gut_metagenome_contig1838000319.

SEQ ID NO: 32 is an exemplary Cas13d sequence from Gut_metagenome_contig13123000268.

SEQ ID NO: 33 is an exemplary Cas13d sequence from Gut_metagenome_contig5294000434.

SEQ ID NO: 34 is an exemplary Cas13d sequence from Gut_metagenome_contig6415000192.

SEQ ID NO: 35 is an exemplary Cas13d sequence from Gut_metagenome_contig6144000300.

SEQ ID NO: 36 is an exemplary Cas13d sequence from Gut_metagenome_contig9118000041.

SEQ ID NO: 37 is an exemplary Cas13d sequence from Activated_sludge_metagenome_transcript_124486.

SEQ ID NO: 38 is an exemplary Cas13d sequence from Gut_metagenome_contig1322000437.

SEQ ID NO: 39 is an exemplary Cas13d sequence from Gut_metagenome_contig4582000531.

SEQ ID NO: 40 is an exemplary Cas13d sequence from Gut_metagenome_contig9190000283.

SEQ ID NO: 41 is an exemplary Cas13d sequence from Gut_metagenome_contig1709000510.

SEQ ID NO: 42 is an exemplary Cas13d sequence from M24_(LSQX01212483_Anaerobic_digester_metagenome) with a HEPN domain.

SEQ ID NO: 43 is an exemplary Cas13d sequence from Gut_metagenome_contig3833000494.

SEQ ID NO: 44 is an exemplary Cas13d sequence from Activated_sludge_metagenome_transcript_117355.

SEQ ID NO: 45 is an exemplary Cas13d sequence from Gut_metagenome_contig11061000330.

SEQ ID NO: 46 is an exemplary Cas13d sequence from Gut_metagenome_contig338000322 from sheep gut metagenome.

SEQ ID NO: 47 is an exemplary Cas13d sequence from human gut metagenome.

SEQ ID NO: 48 is an exemplary Cas13d sequence from Gut_metagenome_contig9530000097.

SEQ ID NO: 49 is an exemplary Cas13d sequence from Gut_metagenome_contig1750000258.

SEQ ID NO: 50 is an exemplary Cas13d sequence from Gut_metagenome_contig5377000274.

SEQ ID NO: 51 is an exemplary Cas13d sequence from gut_metagenome_P19E0k2120140920_c248000089.

SEQ ID NO: 52 is an exemplary Cas13d sequence from Gut_metagenome_contig11400000031.

SEQ ID NO: 53 is an exemplary Cas13d sequence from Gut_metagenome_contig7940000191.

SEQ ID NO: 54 is an exemplary Cas13d sequence from Gut_metagenome_contig6049000251.

SEQ ID NO: 55 is an exemplary Cas13d sequence from Gut_metagenome_contig1137000500.

SEQ ID NO: 56 is an exemplary Cas13d sequence from Gut_metagenome_contig9368000105.

SEQ ID NO: 57 is an exemplary Cas13d sequence from Gut_metagenome_contig546000275.

SEQ ID NO: 58 is an exemplary Cas13d sequence from Gut_metagenome_contig7216000573.

SEQ ID NO: 59 is an exemplary Cas13d sequence from Gut_metagenome_contig4806000409.

SEQ ID NO: 60 is an exemplary Cas13d sequence from Gut_metagenome_contig10762000480.

SEQ ID NO: 61 is an exemplary Cas13d sequence from Gut_metagenome_contig4114000374.

SEQ ID NO: 62 is an exemplary Cas13d sequence from Ruminococcus_flavefaciens_FD1.

SEQ ID NO: 63 is an exemplary Cas13d sequence from Gut_metagenome_contig7093000170.

SEQ ID NO: 64 is an exemplary Cas13d sequence from Gut_metagenome_contig11113000384.

SEQ ID NO: 65 is an exemplary Cas13d sequence from Gut_metagenome_contig6403000259.

SEQ ID NO: 66 is an exemplary Cas13d sequence from Gut_metagenome_contig6193000124.

SEQ ID NO: 67 is an exemplary Cas13d sequence from Gut_metagenome_contig721000619.

SEQ ID NO: 68 is an exemplary Cas13d sequence from Gut_metagenome_contig1666000270.

SEQ ID NO: 69 is an exemplary Cas13d sequence from Gut_metagenome_contig2002000411.

SEQ ID NO: 70 is an exemplary Cas13d sequence from Ruminococcus_albus.

SEQ ID NO: 71 is an exemplary Cas13d sequence from Gut_metagenome_contig13552000311.

SEQ ID NO: 72 is an exemplary Cas13d sequence from Gut_metagenome_contig10037000527.

SEQ ID NO: 73 is an exemplary Cas13d sequence from Gut_metagenome_contig238000329.

SEQ ID NO: 74 is an exemplary Cas13d sequence from Gut_metagenome_contig2643000492.

SEQ ID NO: 75 is an exemplary Cas13d sequence from Gut_metagenome_contig874000057.

SEQ ID NO: 76 is an exemplary Cas13d sequence from Gut_metagenome_contig4781000489.

SEQ ID NO: 77 is an exemplary Cas13d sequence from Gut_metagenome_contig12144000352.

SEQ ID NO: 78 is an exemplary Cas13d sequence from Gut_metagenome_contig5590000448.

SEQ ID NO: 79 is an exemplary Cas13d sequence from Gut_metagenome_contig9269000031.

SEQ ID NO: 80 is an exemplary Cas13d sequence from Gut_metagenome_contig8537000520.

SEQ ID NO: 81 is an exemplary Cas13d sequence from Gut_metagenome_contig1845000130.

SEQ ID NO: 82 is an exemplary Cas13d sequence from gut_metagenome_P13E0k2120140920_c3000072.

SEQ ID NO: 83 is an exemplary Cas13d sequence from gut_metagenome_P1E0k2120140920_c1000078.

SEQ ID NO: 84 is an exemplary Cas13d sequence from Gut_metagenome_contig12990000099.

SEQ ID NO: 85 is an exemplary Cas13d sequence from Gut_metagenome_contig525000349.

SEQ ID NO: 86 is an exemplary Cas13d sequence from Gut_metagenome_contig7229000302.

SEQ ID NO: 87 is an exemplary Cas13d sequence from Gut_metagenome_contig3227000343.

SEQ ID NO: 88 is an exemplary Cas13d sequence from Gut_metagenome_contig7030000469.

SEQ ID NO: 89 is an exemplary Cas13d sequence from Gut_metagenome_contig5149000068.

SEQ ID NO: 90 is an exemplary Cas13d sequence from Gut_metagenome_contig400200045.

SEQ ID NO: 91 is an exemplary Cas13d sequence from Gut_metagenome_contig10420000446.

SEQ ID NO: 92 is an exemplary Cas13d sequence from new_flavefaciens,_strain_XPD3002.

SEQ ID NO: 93 is an exemplary Cas13d sequence from M26_Gut_metagenome_contig698000307.

SEQ ID NO: 94 is an exemplary Cas13d sequence from M36_Uncultured_Eubacterium_sp_TS28_c40956.

SEQ ID NO: 95 is an exemplary Cas13d sequence from M12_gut_metagenome_P25C0k2120140920_c134000066.

SEQ ID NO: 96 is an exemplary Cas13d sequence from human gut metagenome.

SEQ ID NO: 97 is an exemplary Cas13d sequence from M10_gut_metagenome_P25C90k2120140920,_c28000041.

SEQ ID NO: 98 is an exemplary Cas13d sequence from M11_gut_metagenome_P25C7k2120140920_c4078000105.

SEQ ID NO: 99 is an exemplary Cas13d sequence from gut_metagenome_P25C0k2120140920_c32000045.

SEQ ID NO: 100 is an exemplary Cas13d sequence from M13_gut_metagenome_P23C7k2120140920_c3000067.

SEQ ID NO: 101 is an exemplary Cas13d sequence from M5_gut_metagenome_P18E90k2120140920.

SEQ ID NO: 102 is an exemplary Cas13d sequence from M21_gut_metagenome_P18E0k2120140920.

SEQ ID NO: 103 is an exemplary Cas13d sequence from M7_gut_metagenome_P38C7k2120140920_c4841000003.

SEQ ID NO: 104 is an exemplary Cas13d sequence from Ruminococcus_bicirculans.

SEQ ID NO: 105 is an exemplary Cas13d sequence.

SEQ ID NO: 106 is an exemplary Cas13d consensus sequence.

SEQ ID NO: 107 is an exemplary Cas13d sequence from M18_gut_metagenome_P22E0k2120140920_c3395000078.

SEQ ID NO: 108 is an exemplary Cas13d sequence from M17_gut_metagenome_P22E90k2120140920_c114.

SEQ ID NO: 109 is an exemplary Cas13d sequence from Ruminococcus_sp_CAG57.

SEQ ID NO: 110 is an exemplary Cas13d sequence from gut_metagenome_P11E90k2120140920_c43000123.

SEQ ID NO: 111 is an exemplary Cas13d sequence from M6_gut_metagenome_P13E90k2120140920_c7000009.

SEQ ID NO: 112 is an exemplary Cas13d sequence from M19_gut_metagenome_P17E90k2120140920.

SEQ ID NO: 113 is an exemplary Cas13d sequence from gut_metagenome_P17E0k2120140920,_c87000043.

SEQ ID NO: 114 is an exemplary human codon optimized Eubacterium siraeum Cas13d nucleic acid sequence.

SEQ ID NO: 115 is an exemplary human codon optimized Eubacterium siraeum Cas13d nucleic acid sequence with a mutant HEPN domain.

SEQ ID NO: 116 is an exemplary human codon-optimized Eubacterium siraeum Cas13d nucleic acid sequence with N-terminal NLS.

SEQ ID NO: 117 is an exemplary human codon-optimized Eubacterium siraeum Cas13d nucleic acid sequence with N- and C-terminal NLS tags.

SEQ ID NO: 118 is an exemplary human codon-optimized uncultured Ruminococcus sp. Cas13d nucleic acid sequence.

SEQ ID NO: 119 is an exemplary human codon-optimized uncultured Ruminococcus sp. Cas13d nucleic acid sequence with a mutant HEPN domain.

SEQ ID NO: 120 is an exemplary human codon-optimized uncultured Ruminococcus sp. Cas13d nucleic acid sequence with N-terminal NLS.

SEQ ID NO: 121 is an exemplary human codon-optimized uncultured Ruminococcus sp. Cas13d nucleic acid sequence with N- and C-terminal NLS tags.

SEQ ID NO: 122 is an exemplary human codon-optimized uncultured Ruminococcus flavefaciens FD1 Cas13d nucleic acid sequence.

SEQ ID NO: 123 is an exemplary human codon-optimized uncultured Ruminococcus flavefaciens FD1 Cas13d nucleic acid sequence with mutated HEPN domain.

SEQ ID NO: 124 is an exemplary Cas13d nucleic acid sequence from Ruminococcus bicirculans.

SEQ ID NO: 125 is an exemplary Cas13d nucleic acid sequence from Eubacterium siraeum.

SEQ ID NO: 126 is an exemplary Cas13d nucleic acid sequence from Ruminococcus flavefaciens FD1.

SEQ ID NO: 127 is an exemplary Cas13d nucleic acid sequence from Ruminococcus albus.

SEQ ID NO: 128 is an exemplary Cas13d nucleic acid sequence from Ruminococcus flavefaciens XPD.

SEQ ID NO: 129 is an exemplary consensus DR nucleic acid sequence for *E. siraeum* Cas13d.

SEQ ID NO: 130 is an exemplary consensus DR nucleic acid sequence for *Rum. Sp.* Cas13d.

SEQ ID NO: 131 is an exemplary consensus DR nucleic acid sequence for *Rum. Flavefaciens strain XPD3002* Cas13d and CasRx.

SEQ ID NOS: 132-137 are exemplary consensus DR nucleic acid sequences.

SEQ ID NO: 138 is an exemplary 50% consensus sequence for seven full-length Cas13d orthologues.

SEQ ID NO: 139 is an exemplary Cas13d nucleic acid sequence from Gut metagenome P1E0.

SEQ ID NO: 140 is an exemplary Cas13d nucleic acid sequence from Anaerobic digester.

SEQ ID NO: 141 is an exemplary Cas13d nucleic acid sequence from Ruminococcus sp. CAG:57.

SEQ ID NO: 142 is an exemplary human codon-optimized uncultured Gut metagenome P1E0 Cas13d nucleic acid sequence.

SEQ ID NO: 143 is an exemplary human codon-optimized Anaerobic Digester Cas13d nucleic acid sequence.

SEQ ID NO: 144 is an exemplary human codon-optimized Ruminococcus flavefaciens XPD Cas13d nucleic acid sequence.

SEQ ID NO: 145 is an exemplary human codon-optimized Ruminococcus albus Cas13d nucleic acid sequence.

SEQ ID NO: 146 is an exemplary processing of the *Ruminococcus* sp. CAG:57 CRISPR array.

SEQ ID NO: 147 is an exemplary Cas13d protein sequence from contig emb|OBVH01003037.1, human gut metagenome sequence (also found in WGS contigs emb|OBXZ01000094.1| and emb|OBJF01000033.1.

SEQ ID NO: 148 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 147).

SEQ ID NO: 149 is an exemplary Cas13d protein sequence from contig tpg|DBYI01000091.1| (Uncultivated Ruminococcus flavefaciens UBA1190 assembled from bovine gut metagenome).

SEQ ID NOS: 150-152 are exemplary consensus DR nucleic acid sequences (goes with SEQ ID NO: 149).

SEQ ID NO: 153 is an exemplary Cas13d protein sequence from contig tpg|DJXD01000002.1| (uncultivated Ruminococcus assembly, UBA7013, from sheep gut metagenome).

SEQ ID NO: 154 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 153).

SEQ ID NO: 155 is an exemplary Cas13d protein sequence from contig OGZC01000639.1 (human gut metagenome assembly).

SEQ ID NOS: 156-177 are exemplary consensus DR nucleic acid sequences (goes with SEQ ID NO: 155).

SEQ ID NO: 158 is an exemplary Cas13d protein sequence from contig emb|OHBM01000764.1 (human gut metagenome assembly).

SEQ ID NO: 159 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 158).

SEQ ID NO: 160 is an exemplary Cas13d protein sequence from contig emb|OHCP01000044.1 (human gut metagenome assembly).

SEQ ID NO: 161 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 160).

SEQ ID NO: 162 is an exemplary Cas13d protein sequence from contig emb|OGDF01008514.1| (human gut metagenome assembly).

SEQ ID NO: 163 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 162).

SEQ ID NO: 164 is an exemplary Cas13d protein sequence from contig emb|OGPN01002610.1 (human gut metagenome assembly).

SEQ ID NO: 165 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 164).

SEQ ID NO: 166 is an exemplary Cas13d protein sequence from contig NFIR01000008.1 (Eubacterium sp. An3, from chicken gut metagenome).

SEQ ID NO: 167 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 166).

SEQ ID NO: 168 is an exemplary Cas13d protein sequence from contig NFLV01000009.1 (Eubacterium sp. An11, from chicken gut metagenome).

SEQ ID NO: 169 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 168).

SEQ ID NOS: 171-174 are an exemplary Cas13d motif sequences.

SEQ ID NO: 175 is an exemplary Cas13d protein sequence from contig OJMM01002900 human gut metagenome sequence.

SEQ ID NO: 176 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 175).

SEQ ID NO: 177 is an exemplary Cas13d protein sequence from contig ODAI011611274.1 gut metagenome sequence.

SEQ ID NO: 178 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 177).

SEQ ID NO: 179 is an exemplary Cas13d protein sequence from contig OIZX01000427.1.

SEQ ID NO: 180 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 179).

SEQ ID NO: 181 is an exemplary Cas13d protein sequence from contig emb|OCVV012889144.1.

SEQ ID NO: 182 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 181).

SEQ ID NO: 183 is an exemplary Cas13d protein sequence from contig OCTW011587266.1

SEQ ID NO: 184 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 183).

SEQ ID NO: 185 is an exemplary Cas13d protein sequence from contig emb|OGNF01009141.1.

SEQ ID NO: 186 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 185).

SEQ ID NO: 187 is an exemplary Cas13d protein sequence from contig emb|OIEN01002196.1.

SEQ ID NO: 188 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 187).

SEQ ID NO: 189 is an exemplary Cas13d protein sequence from contig e-k87_11092736.

SEQ ID NOS: 190-193 are exemplary consensus DR nucleic acid sequences (goes with SEQ ID NO: 189).

SEQ ID NO: 194 is an exemplary Cas13d sequence from Gut_metagenome_contig6893000291.

SEQ ID NOS: 195-197 are exemplary Cas13d motif sequences.

SEQ ID NO: 198 is an exemplary Cas13d protein sequence from Ga0224415_10007274.

SEQ ID NO: 199 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 198).

SEQ ID NO: 200 is an exemplary Cas13d protein sequence from EMG_10003641.

SEQ ID NO: 201 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 200).

SEQ ID NO: 202 is an exemplary Cas13d protein sequence from Ga0129306_1000735.

SEQ ID NO: 203 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 202).

SEQ ID NO: 204 is an exemplary Cas13d protein sequence from Ga0129317_1008067.

SEQ ID NO: 205 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 204).

SEQ ID NO: 206 is an exemplary Cas13d protein sequence from Ga0224415_10048792.

SEQ ID NO: 207 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 206).

SEQ ID NO: 208 is an exemplary Cas13d protein sequence from 160582958 _gene49834.

SEQ ID NO: 209 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 208).

SEQ ID NO: 210 is an exemplary Cas13d protein sequence from 250twins_35838_GL0110300.

SEQ ID NO: 211 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 210).

SEQ ID NO: 212 is an exemplary Cas13d protein sequence from 250twins_36050_GL0158985.

SEQ ID NO: 213 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 212).

SEQ ID NO: 214 is an exemplary Cas13d protein sequence from 31009_GL0034153.

SEQ ID NO: 215 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 214).

SEQ ID NO: 216 is an exemplary Cas13d protein sequence from 530373_GL0023589.

SEQ ID NO: 217 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 216).

SEQ ID NO: 218 is an exemplary Cas13d protein sequence from BMZ-11B_GL0037771.

SEQ ID NO: 219 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 218).

SEQ ID NO: 220 is an exemplary Cas13d protein sequence from BMZ-11B_GL0037915.

SEQ ID NO: 221 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 220).

SEQ ID NO: 222 is an exemplary Cas13d protein sequence from BMZ-11B_GL0069617.

SEQ ID NO: 223 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 222).

SEQ ID NO: 224 is an exemplary Cas13d protein sequence from - DLF014_GL0011914.

SEQ ID NO: 225 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 224).

SEQ ID NO: 226 is an exemplary Cas13d protein sequence from EYZ-362B_GL0088915.

SEQ ID NO: 227-228 are exemplary consensus DR nucleic acid sequences (goes with SEQ ID NO: 226).

SEQ ID NO: 229 is an exemplary Cas13d protein sequence from Ga0099364_10024192.

SEQ ID NO: 230 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 229).

SEQ ID NO: 231 is an exemplary Cas13d protein sequence from Ga0187910_10006931.

SEQ ID NO: 232 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 231).

SEQ ID NO: 233 is an exemplary Cas13d protein sequence from Ga0187910_10015336.

SEQ ID NO: 234 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 233).

SEQ ID NO: 235 is an exemplary Cas13d protein sequence from Ga0187910_10040531.

SEQ ID NO: 236 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 235).

SEQ ID NO: 237 is an exemplary Cas13d protein sequence from Ga0187911_10069260.

SEQ ID NO: 238 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 237).

SEQ ID NO: 239 is an exemplary Cas13d protein sequence from MH0288_GL0082219.

SEQ ID NO: 240 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 239).

SEQ ID NO: 241 is an exemplary Cas13d protein sequence from O2.UC29-0_GL0096317.

SEQ ID NO: 242 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 241).

SEQ ID NO: 243 is an exemplary Cas13d protein sequence from PIG-014_GL0226364.

SEQ ID NO: 244 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 243).

SEQ ID NO: 245 is an exemplary Cas13d protein sequence from PIG-018_GL0023397.

SEQ ID NO: 246 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 245).

SEQ ID NO: 247 is an exemplary Cas13d protein sequence from PIG-025_GL0099734.

SEQ ID NO: 248 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 247).

SEQ ID NO: 249 is an exemplary Cas13d protein sequence from PIG-028_GL0185479.

SEQ ID NO: 250 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 249).

SEQ ID NO: 251 is an exemplary Cas13d protein sequence from - Ga0224422_10645759.

SEQ ID NO: 252 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 251).

SEQ ID NO: 253 is an exemplary Cas13d protein sequence from ODAI chimera.

SEQ ID NO: 254 is an exemplary consensus DR nucleic acid sequence (goes with SEQ ID NO: 253).

SEQ ID NO: 255 is an HEPN motif.

SEQ ID NOs: 256 and 257 are exemplary Cas13d nuclear localization signal amino acid and nucleic acid sequences, respectively.

SEQ ID NOs: 258 and 260 are exemplary SV40 large T antigen nuclear localization signal amino acid and nucleic acid sequences, respectively.

SEQ ID NO: 259 is a dCas9 target sequence.

SEQ ID NO: 261 is an artificial Eubacterium siraeum nCas1 array targeting ccdB.

SEQ ID NO: 262 is a full 36 nt direct repeat.

SEQ ID NOs: 263-266 are spacer sequences.

SEQ ID NO: 267 is an artificial uncultured Ruminoccus sp. nCas1 array targeting ccdB.

SEQ ID NO: 268 is a full 36 nt direct repeat.

SEQ ID NOs: 269-272 are spacer sequences.

SEQ ID NO: 273 is a ccdB target RNA sequence.

SEQ ID NOs: 274-277 are spacer sequences.

SEQ ID NO: 278 is a mutated Cas13d sequence, NLS-Ga_0531(trunc)-NLS-HA. This mutant has a deletion of the non-conserved N-terminus.

SEQ ID NO: 279 is a mutated Cas13d sequence, NES-Ga_0531(trunc)-NES-HA. This mutant has a deletion of the non-conserved N-terminus.

SEQ ID NO: 280 is a full-length Cas13d sequence, NLS-RfxCas13d-NLS-HA.

SEQ ID NO: 281 is a mutated Cas13d sequence, NLS-RfxCas13d(del5)-NLS-HA. This mutant has a deletion of amino acids 558-587.

SEQ ID NO: 282 is a mutated Cas13d sequence, NLS-RfxCas13d(del5.12)-NLS-HA. This mutant has a deletion of amino acids 558-587 and 953-966.

SEQ ID NO: 283 is a mutated Cas13d sequence, NLS-RfxCas13d(del5.13)-NLS-HA. This mutant has a deletion of amino acids 376-392 and 558-587.

SEQ ID NO: 284 is a mutated Cas13d sequence, NLS-RfxCas13d(del5.12+5.13)-NLS-HA. This mutant has a deletion of amino acids 376-392, 558-587, and 953-966.

SEQ ID NO: 285 is a mutated Cas13d sequence, NLS-RfxCas13d(del13)-NLS-HA. This mutant has a deletion of amino acids 376-392.

SEQ ID NO: 286 is an effector sequence used to edit expression of ADAR2. Amino acids 1 to 969 are dRfxCas13, aa 970 to 991 are an NLS sequence, and amino acids 992 to 1378 are $ADAR2_{DD}$.

SEQ ID NO: 287 is an exemplary HIV NES protein sequence.

SEQ ID NOS: 288-291 are exemplary Cas13d motif sequences.

SEQ ID NO: 292 is Cas13d ortholog sequence MH_4866.

SEQ ID NO: 293 is an exemplary Cas13d protein sequence from 037_-_emb|OIZA01000315.1|

SEQ ID NO: 294 is an exemplary Cas13d protein sequence from PIG-022_GL0026351.

SEQ ID NO: 295 is an exemplary Cas13d protein sequence from PIG-046_GL0077813.

SEQ ID NO: 296 is an exemplary Cas13d protein sequence from pig_chimera.

DETAILED DESCRIPTION

Unless otherwise noted, technical terms are used according to conventional usage. Definitions of common terms in molecular biology can be found in Benjamin Lewin, *Genes VII*, published by Oxford University Press, 1999; Kendrew *et al.* (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Science Ltd., 1994; and Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by VCH Publishers, Inc., 1995; and other similar references.

As used herein, the singular forms "a," "an," and "the," refer to both the singular as well as plural, unless the context clearly indicates otherwise. As used herein, the term "comprises" means "includes." Thus, "comprising a nucleic acid molecule" means "including a nucleic acid molecule" without excluding other elements. It is further to be understood that any and all base sizes given for nucleic acids are approximate, and are provided for descriptive purposes, unless otherwise indicated. Although many methods and materials similar or equivalent to those described herein can be used, particular suitable methods and materials are described below. In case of conflict, the present specification, including explanations of terms, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting. All references, including patent applications and patents, are herein incorporated by reference in their entireties.

In order to facilitate review of the various embodiments of the disclosure, the following explanations of specific terms are provided:

I. Terms

Administration: To provide or give a subject an agent, such as a Cas13d protein (or Cas13d coding sequence) or guide molecule (or coding sequence) disclosed herein, by any effective route. Exemplary routes of administration include, but are not limited to, injection (such as subcutaneous, intramuscular, intradermal, intraperitoneal, intratumoral, and intravenous), transdermal, intranasal, and inhalation routes.

Cas13d (also referred to as CasR, for CRISPR-associated RNase): An RNA-guided RNA endonuclease enzyme that can cut or bind RNA. Cas13d proteins include one or two HEPN domains (*e.g.*, see SEQ ID NOS: 1-3, 42, 62, 70, 82, 83, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, and 296). Native HEPN domains include the sequence RXXXXH (SEQ ID NO: 255). Cas13d proteins that include mutated HEPN domain(s), and thus cannot cut RNA, but can process

guide RNA, are also encompassed by this disclosure (*e.g.*, see SEQ ID NOS: 2 and 4). An alignment of native Cas13d proteins is shown in FIGS. 18A-18MMM. In addition, Cas13d proteins specifically recognize direct repeat sequences of gRNA having a particular secondary structure (*e.g.*, see FIGS.S2C). In one example, Cas13d proteins recognize and/or bind a DR having (1) a loop of about 4 to 8 nt, (2) a stem of 4 to 12 nt, stem formed of complementary nucleotides, which can include a small (*e.g.*, 1 or 2 bp) bulge due to a nt mismatch in the stem, and (3) a bulge or overhang formed of unpaired nt, which can be about 10 to 14 nt (*e.g.*, 5 to 7 to on either side).

In one example, the full length (non-truncated) Cas13d protein is between 870-1080 amino acids long. In one example, the Cas13d protein is derived from a genome sequence of a bacterium from the Order Clostridiales or a metagenomic sequence. In one example, the corresponding DR sequence of a Cas13d protein is located at the 5' end of the spacer sequence in the molecule that includes the Cas13d gRNA. In one example, the DR sequence in the Cas13d gRNA is truncated at the 5' end relative to the DR sequence in the unprocessed Cas13d guide array transcript (such as truncated by at least 1 nt, at least 2 nt, at least 3 nt, at least 4 nt, at least 5 nt, such as 1-3 nt, 3-6 nt, 5-7 nt, or 5-10 nt). In one example, the DR sequence in the Cas13d gRNA is truncated by 5-7 nt at the 5' end by the Cas13d protein. In one example, the Cas13d protein can cut a target RNA flanked at the 3' end of the spacer-target duplex by any of a A, U, G or C ribonucleotide and flanked at the 5' end by any of a A, U, G or C ribonucleotide.

In one example, a Cas13d protein has at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296.

In one example, a Cas13d protein contains two HEPN Rnase domains which contain a RXXXXH amino acid motif (SEQ ID NO: 255; where X indicates any amino acid). In addition, a Cas13d protein can include one or more of the following amino acid motifs written in the common Prosite format:

Motif 1: L-x(5)-[FWY]-x(3)-K-[NQS]-[ILM]-[ILMV]-x(2)-N-x(2)-[FWY]-x(2)-[AG]-x(4)-[DE]-x-D (SEQ ID NO: 195)

Motif 2: [FWY]-[ILV]-x(2)-[NQS]-[ILV]-x(2)-[DNST]-x(2)-F-x-Y-x(2)-[HKR]-[FHY] (SEQ ID NO: 196)

Motif 3: Y-[CDNSV]-x(2)-R-[FWY]-x-[ADNT]-[LM]-[ST]-x(4)-[FWY] (SEQ ID NO: 197)

In one example, a Cas13d protein include one or more of the following amino acid consensus sequences, written in the common Prosite format:

L-[DE]-G-K-E-I-N-x-[LM]-x-[CST]-[ANST]-L-I-x-K-F-[DE]-[GN]-I-x-[ADS]-[FIL]-x-[DEHK]-[AILV] (SEQ ID NO: 288)

R-N-[FY]-I-[AIL]-x-N-V-[ILV]-x-[NS]-x(2)-F-x-Y-[ILV]-[AIV]-[KR]-[HY]-x(2)-[PT]-x(3)-[HKR]-x-[IL]-x(2)-N-x(2)-[ILV]-x(2)-[FY]-V-L-x(2)-[ILM]-[AGP]-[DE]-x(4)-[NQRS] (SEQ ID NO: 289)

L-x-[LM]-[ST]-[IV]-x-Y-x(3)-K-[NS]-[ILM]-V-x-[IV]-N-[AST]-R-[FY]-x(2)-[AG]-[FY]-x-[ACV]-[LV]-E-R-D-x(2)-[FL] (SEQ ID NO: 290)

P-F-[AGS]-Y-[CNV]-x-[AP]-R-[FY]-K-N-L-[ST]-x-[EN]-x-L-F-x-[DKR]-[DEN] (SEQ ID NO: 291)

Thus, in some examples, a Cas13d protein having at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296 includes the motif of SEQ ID NO: 195, 196 or 197.

In addition, in some examples, a Cas13d protein having at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296 and includes the consensus sequence of SEQ ID NO: 288, 289, 290, or 291.

Complementarity: The ability of a nucleic acid to form hydrogen bond(s) with another nucleic acid sequence by either traditional Watson-Crick base pairing or other non-traditional types. A percent complementarity indicates the percentage of residues in a nucleic acid molecule which can form hydrogen bonds (*e.g.*, Watson-Crick base pairing) with a second nucleic acid sequence (*e.g.*, 5, 6, 7, 8, 9, 10 out of 10 being 50%, 60%, 70%, 80%, 90%, and 100% complementary). "Perfectly complementary" means that all the contiguous residues of a nucleic acid sequence will hydrogen bond with the same number of contiguous residues in a second nucleic acid sequence. "Substantially complementary" as used herein refers to a degree of complementarity that is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% over a region of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, or more nucleotides, or refers to two nucleic acids that hybridize under stringent conditions.

CRISPRs (**C**lustered **R**egularly **I**nter**S**paced **R**epeats): The CRISPR RNA array is a defining feature of CRISPR systems. The term "CRISPR" refers to the architecture of the array which includes constant direct repeats (DRs) interspaced with the variable spacers. In some examples, a CRISPR array includes at least a DR-spacer-DR-spacer (see **FIG. 1A**). This feature was used in the disclosed computational pipeline to identify the novel Cas13d protein family (**FIG. 1A**). In bacteria, the array is transcribed as one single transcript (containing multiple crRNA units), which is then processed by the Cas13d protein and other RNases into individual crRNAs. CRISPRs are found in approximately 40% of sequenced bacteria genomes and 90% of sequenced archaea. CRISPRs are often associated with cas genes that code for proteins related to CRISPRs (such as the Cas13d proteins provided herein). The disclosed CRISPR/Cas system can be used for RNA targeting, for example to detect a target RNA, modify a target RNA at any desired location, or cut the target RNA at any desired location.

Downregulated or knocked down: When used in reference to the expression of a molecule, such as a target RNA, refers to any process which results in a decrease in production of the target RNA, but in some examples not complete elimination of the target RNA product or target RNA function. In one example, downregulation or knock down does not result in complete elimination of detectable target RNA expression or target RNA activity. In some examples, the target RNA is a coding RNA. In some examples, the target RNA is non-coding RNA. Specific examples of RNA molecules that can be targeted for downregulation include mRNA, miRNA, rRNA, tRNA, nuclear RNA, lincRNA, circular RNA, and structural RNA. In some examples, downregulation or knock down of a target RNA includes processes that decrease translation of the target RNA and thus can decrease the presence of corresponding proteins. The disclosed CRISPR/Cas system can be used to downregulate any target RNA of interest.

Downregulation or knock down includes any detectable decrease in the target RNA. In certain examples, detectable target RNA in a cell or cell free system decreases by at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% (such as a decrease of 40% to 90%, 40% to 80% or 50% to 95%) as compared to a control (such an amount of target RNA detected in a corresponding normal cell or sample). In one example, a control is a relative amount of expression in a normal cell (*e.g.*, a non-recombinant cell that does not include Cas13d or guide RNA).

Effective amount: The amount of an agent (such as the CRISPR/Cas agents provided herein) that is sufficient to effect beneficial or desired results.

A therapeutically effective amount may vary depending upon one or more of: the subject and disease condition being treated, the weight and age of the subject, the severity of the disease condition, the manner of administration and the like, which can readily be determined by one of ordinary skill in the art. The beneficial therapeutic effect can include enablement of diagnostic determinations; amelioration of a disease, symptom, disorder, or pathological condition; reducing or preventing the onset of a disease, symptom, disorder or condition; and generally counteracting a disease, symptom, disorder or pathological condition. In one embodiment, an "effective amount" is an amount sufficient to reduce symptoms of a

disease, for example by at least 10%, at least 20%, at least 50%, at least 70%, or at least 90% (as compared to no administration of the therapeutic agent).

The term also applies to a dose that will allow for expression of a Cas13d and/or gRNA herein, and that allows for targeting (*e.g.*, detection or modification) of a target RNA.

Increase or Decrease: A statistically significant positive or negative change, respectively, in quantity from a control value. An increase is a positive change, such as an increase at least 50%, at least 100%, at least 200%, at least 300%, at least 400% or at least 500% as compared to the control value. A decrease is a negative change, such as a decrease of at least 20%, at least 25%, at least 50%, at least 75%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99%, or at least 100% decrease as compared to a control value. In some examples the decrease is less than 100%, such as a decrease of no more than 90%, no more than 95% or no more than 99%.

Isolated: An "isolated" biological component (such as a Cas13d protein or nucleic acid, gRNA, or cell containing such) has been substantially separated, produced apart from, or purified away from other biological components in the cell or tissue of an organism in which the component occurs, such as other cells, chromosomal and extrachromosomal DNA and RNA, and proteins. Nucleic acids and proteins that have been "isolated" include nucleic acids and proteins purified by standard purification methods. The term also embraces nucleic acids and proteins prepared by recombinant expression in a host cell as well as chemically synthesized nucleic acids and proteins. Isolated Cas13d proteins or nucleic acids, or cells containing such, in some examples are at least 50% pure, such as at least 75%, at least 80%, at least 90%, at least 95%, at least 98%, or at least 100% pure.

Label: A compound or composition that is conjugated directly or indirectly to another molecule (such as a nucleic acid molecule) to facilitate detection of that molecule. Specific, non-limiting examples of labels include fluorescent and fluorogenic moieties, chromogenic moieties, haptens, affinity tags, and radioactive isotopes. The label can be directly detectable (*e.g.*, optically detectable) or indirectly detectable (for example, via interaction with one or more additional molecules that are in turn detectable).

Modulate: A change in the content of RNA. Modulation can include, but is not limited to, RNA activation (*e.g.*, upregulation), RNA repression (*e.g.*, downregulation), ribonucleotide deletion, ribonucleotide insertion, ribonucleotide chemical modification, ribonucleotide covalent or non-covalent linkage, and/or ribonucleotide substitution.

Non-naturally occurring or engineered: Terms used herein as interchangeably and indicate the involvement of the hand of man. The terms, when referring to nucleic acid molecules or polypeptides indicate that the nucleic acid molecule or the polypeptide is at least substantially free from at least one other component with which they are naturally associated in nature and as found in nature. In addition, the terms can indicate that the nucleic acid molecules or polypeptides have a sequence not found in nature.

Operably linked: A first nucleic acid sequence is operably linked with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence (such as a coding sequence

of a Cas13d protein) if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences are contiguous and, where necessary to join two protein-coding regions, in the same reading frame.

Pharmaceutically acceptable carriers: The pharmaceutically acceptable carriers useful in this invention are conventional. *Remington's Pharmaceutical Sciences*, by E. W. Martin, Mack Publishing Co., Easton, PA, 15th Edition (1975), describes compositions and formulations suitable for pharmaceutical delivery of a Cas13d protein or nucleic acid molecule (or other molecules needed for modifying RNA using the disclosed CRISPR/Cas system with the disclosed Cas13d proteins).

In general, the nature of the carrier will depend on the particular mode of administration being employed. For instance, parenteral formulations usually comprise injectable fluids that include pharmaceutically and physiologically acceptable fluids such as water, physiological saline, balanced salt solutions, aqueous dextrose, glycerol or the like as a vehicle. In addition to biologically-neutral carriers, pharmaceutical compositions to be administered can contain minor amounts of non-toxic auxiliary substances, such as wetting or emulsifying agents, preservatives, and pH buffering agents and the like, for example sodium acetate or sorbitan monolaurate.

Polypeptide, peptide and protein: Refer to polymers of amino acids of any length. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non-amino acids. The terms also encompass an amino acid polymer that has been modified; for example, disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation, such as conjugation with a labeling component. As used herein the term "amino acid" includes natural and/or unnatural or synthetic amino acids, including glycine and both the D or L optical isomers, and amino acid analogs and peptidomimetics.

Promoter: An array of nucleic acid control sequences which direct transcription of a nucleic acid. A promoter includes necessary nucleic acid sequences near the start site of transcription. A promoter also optionally includes distal enhancer or repressor elements. A "constitutive promoter" is a promoter that is continuously active and is not subject to regulation by external signals or molecules. In contrast, the activity of an "inducible promoter" is regulated by an external signal or molecule (for example, a transcription factor).

Recombinant or host cell: A cell that has been genetically altered, or is capable of being genetically altered by introduction of an exogenous polynucleotide, such as a recombinant plasmid or vector. Typically, a host cell is a cell in which a vector can be propagated and its nucleic acid expressed. Such cells can be eukaryotic or prokaryotic. The term also includes any progeny of the subject host cell. It is understood that all progeny may not be identical to the parental cell since there may be mutations that occur during replication. However, such progeny are included when the term "host cell" is used.

Regulatory element: A phrase that includes 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) which is hereby incorporated by reference in its entirety. Regulatory elements include those that direct constitutive expression of a nucleotide sequence in many types of host cells 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 provided herein includes a pol III promoter (*e.g.*, U6 and H1 promoters), a pol II promoter (*e.g.*, the retroviral Rous sarcoma virus (RSV) LTR promoter (optionally with the RSV enhancer), the cytomegalovirus (CMV) promoter (optionally with the CMV enhancer), the SV40 promoter, the dihydrofolate reductase promoter, the β-actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1α promoter), or both.

Also encompassed by the term "regulatory element" are enhancer elements, such as WPRE; CMV enhancers; the R-U5' segment in LTR of HTLV-I; SV40 enhancer; and the intron sequence between exons 2 and 3 of rabbit β-globin.

In some examples, a Cas13d coding sequence is operably linked to a promoter, enhancer element, or both.

RNA Editing: A type of genetic engineering in which a RNA molecule (or ribonucleotides of the RNA) is inserted, deleted or replaced in the genome of an organism using engineered nucleases (such as the Cas13d proteins provided herein), which create site-specific strand breaks at desired locations in the RNA. The induced breaks are repaired resulting in targeted mutations or repairs. The CRISPR/Cas methods disclosed herein, such as those that use a Cas13d, can be used to edit the sequence of one or more target RNAs, such as one associated with cancer (*e.g.*, breast cancer, colon cancer, lung cancer, prostate cancer, melanoma), infectious disease (such as HIV, hepatitis, HPV, and West Nile virus), or neurodegenerative disorder (*e.g.*, Huntington's disease or ALS). For example, RNA editing can be used to treat a disease or viral infection.

RNA insertion site: A site of the RNA that is targeted for, or has undergone, insertion of an exogenous polynucleotide. The disclosed methods include use of a disclosed Cas13d protein, which can be used to target a RNA for manipulation at a RNA insertion site.

Sequence identity/similarity: The similarity between amino acid (or nucleotide) sequences is expressed in terms of the similarity between the sequences, otherwise referred to as sequence identity. Sequence identity is frequently measured in terms of percentage identity (or similarity or homology); the higher the percentage, the more similar the two sequences are.

Methods of alignment of sequences for comparison are well known in the art. Various programs and alignment algorithms are described in: Smith and Waterman, *Adv. Appl. Math.* 2:482, 1981; Needleman and

Wunsch, *J. Mol. Biol.* 48:443, 1970; Pearson and Lipman, *Proc. Natl. Acad. Sci. U.S.A.* 85:2444, 1988; Higgins and Sharp, *Gene* 73:237, 1988; Higgins and Sharp, *CABIOS* 5:151, 1989; Corpet *et al.*, *Nucleic Acids Research* 16:10881, 1988; and Pearson and Lipman, *Proc. Natl. Acad. Sci. U.S.A.* 85:2444, 1988. Altschul *et al.*, *Nature Genet.* 6:119, 1994, presents a detailed consideration of sequence alignment methods and homology calculations.

The NCBI Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, *J. Mol. Biol.* 215:403, 1990) is available from several sources, including the National Center for Biotechnology Information (NCBI, Bethesda, MD) and on the internet, for use in connection with the sequence analysis programs blastp, blastn, blastx, tblastn and tblastx. A description of how to determine sequence identity using this program is available on the NCBI website on the internet.

Variants of protein and nucleic acid sequences known in the art and disclosed herein are typically characterized by possession of at least about 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% sequence identity counted over the full length alignment with the amino acid sequence using the NCBI Blast 2.0, gapped blastp set to default parameters. For comparisons of amino acid sequences of greater than about 30 amino acids, the Blast 2 sequences function is employed using the default BLOSUM62 matrix set to default parameters, (gap existence cost of 11, and a per residue gap cost of 1). When aligning short peptides (fewer than around 30 amino acids), the alignment should be performed using the Blast 2 sequences function, employing the PAM30 matrix set to default parameters (open gap 9, extension gap 1 penalties). Proteins with even greater similarity to the reference sequences will show increasing percentage identities when assessed by this method, such as at least 95%, at least 98%, or at least 99% sequence identity. When less than the entire sequence is being compared for sequence identity, homologs and variants will typically possess at least 80% sequence identity over short windows of 10-20 amino acids, and may possess sequence identities of at least 85% or at least 90% or at least 95% depending on their similarity to the reference sequence. Methods for determining sequence identity over such short windows are available at the NCBI website on the internet. One of skill in the art will appreciate that these sequence identity ranges are provided for guidance only; it is entirely possible that strongly significant homologs could be obtained that fall outside of the ranges provided.

Thus, in one example, a Cas13d protein has at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 19, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296.

Subject: A vertebrate, such as a mammal, for example a human. Mammals include, but are not limited to, murines, simians, humans, farm animals, sport animals, and pets. In one embodiment, the subject is a non-human mammalian subject, such as a monkey or other non-human primate, mouse, rat, rabbit, pig, goat, sheep, dog, cat, horse, or cow. In some examples, the subject has a disorder (e.g., viral infection) or genetic disease that can be treated using methods provided herein. In some examples, the subject has a disorder (*e.g.*, viral infection) or genetic disease that can be diagnosed using methods provided herein. In some examples, the subject is a laboratory animal/organism, such as a zebrafish, *Xenopus*, *C. elegans*, *Drosophila*, mouse, rabbit, or rat. Tissues, cells and their progeny of a biological entity obtained *in vivo* or cultured *in vitro* are also encompassed.

Therapeutic agent: Refers to one or more molecules or compounds that confer some beneficial effect upon administration to a subject. The beneficial therapeutic effect can include enablement of diagnostic determinations; amelioration of a disease, symptom, disorder, or pathological condition; reducing or preventing the onset of a disease, symptom, disorder or condition; and generally counteracting a disease, symptom, disorder or pathological condition.

Transduced, **Transformed** and **Transfected:** A virus or vector "transduces" a cell when it transfers nucleic acid molecules into a cell. A cell is "transformed" or "transfected" by a nucleic acid transduced into the cell when the nucleic acid becomes stably replicated by the cell, either by incorporation of the nucleic acid into the cellular genome, or by episomal replication.

These terms encompasses all techniques by which a nucleic acid molecule can be introduced into such a cell, including transfection with viral vectors, transformation with plasmid vectors, and introduction of naked DNA by electroporation, lipofection, particle gun acceleration and other methods in the art. In some example the method is a chemical method (*e.g.*, calcium-phosphate transfection), physical method (*e.g.*, electroporation, microinjection, particle bombardment), fusion (*e.g.*, liposomes), receptor-mediated endocytosis (*e.g.*, DNA-protein complexes, viral envelope/capsid-DNA complexes) and biological infection by viruses such as recombinant viruses (Wolff, J. A., ed, *Gene Therapeutics*, Birkhauser, Boston, USA, 1994). Methods for the introduction of nucleic acid molecules into cells are known (*e.g.*, see U.S. Patent No. 6,110,743). These methods can be used to transduce a cell with the disclosed agents to manipulate its genome.

Transgene: An exogenous gene.

Treating, Treatment, and Therapy: Any success or indicia of success in the attenuation or amelioration of an injury, pathology or condition, including any objective or subjective parameter such as abatement, remission, diminishing of symptoms or making the condition more tolerable to the patient, slowing in the rate of degeneration or decline, making the final point of degeneration less debilitating, improving a subject's physical or mental well-being, or prolonging the length of survival. The treatment may be assessed by objective or subjective parameters; including the results of a physical examination, blood and other clinical tests, and the like. For prophylactic benefit, the disclosed compositions may be administered to a subject at risk of developing a particular disease, condition, or symptom, or to a subject

reporting one or more of the physiological symptoms of a disease, even though the disease, condition, or symptom may not have yet been manifested.

Upregulated: When used in reference to the expression of a molecule, such as a target RNA, refers to any process which results in an increase in production of the target RNA. In one example, includes direct upregulation, for example if the target RNA participates in a feedback loop with its own transcription. In one example, includes indirect upregulation, such as by knockdown of an inhibitory miRNA that leads to the activation of a target of that miRNA.

In some examples, the target RNA is a coding RNA. In some examples, the target RNA is non-coding RNA. Specific examples of RNA molecules that can be targeted for upregulation include mRNA, miRNA, rRNA, tRNA, nuclear RNA, and structural RNA. In some examples, upregulation or activation of a target RNA includes processes that increase translation of the target RNA and thus can increase the presence of corresponding proteins. The disclosed CRISPR/Cas system can be used to upregulate any target RNA of interest.

Upregulation includes any detectable increase in target RNA. In certain examples, detectable target RNA expression in a cell or cell free system (such as a cell expressing a Cas13d protein and gRNA disclosed herein) increases by at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, at least 100%, at least 200%, at least 400%, or at least 500% as compared to a control(such an amount of target RNA detected in a corresponding normal cell or sample). In one example, a control is a relative amount of expression in a normal cell (*e.g.*, a non-recombinant cell that does not include Cas13d or guide RNA).

Under conditions sufficient for: A phrase that is used to describe any environment that permits a desired activity. In one example the desired activity is expression of a Cas13d protein disclosed herein, in combination with other necessary elements, for example to modulate a target RNA.

Vector: A nucleic acid molecule into which a foreign nucleic acid molecule can be introduced without disrupting the ability of the vector to replicate and/or integrate in a host cell. 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.

A vector can include nucleic acid sequences that permit it to replicate in a host cell, such as an origin of replication. A vector can also include one or more selectable marker genes and other genetic elements known in the art. An integrating vector is capable of integrating itself into a host nucleic acid. An expression vector is a vector that contains the necessary regulatory sequences to allow transcription and translation of inserted gene or genes.

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. In some embodiments, the vector is a lentivirus (such as an integration-deficient lentiviral vector) or adeno-associated viral (AAV) vector.

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.

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 are often in the form of plasmids. Recombinant expression vectors can comprise a nucleic acid provided herein (such as a guide RNA [which can be expressed from an RNA sequence or a RNA sequence], nucleic acid encoding a Cas13d protein) 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). 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.).

II. Overview of Several Embodiments

Class 2 CRISPR-Cas systems endow microbes with diverse mechanisms for adaptive immunity. Provided herein is an analysis of prokaryotic genome and metagenome sequences to identify an uncharacterized family of RNA-guided, RNA-targeting CRISPR systems which is classified as Type VI-D. Biochemical characterization and protein engineering of seven distinct orthologs generated a ribonuclease effector derived from *Ruminococcus flavefaciens* XPD3002 (CasRx) with robust activity in human cells. CasRx-mediated knockdown exhibits high efficiency and specificity relative to RNA interference across diverse endogenous transcripts. As one of the most compact single effector Cas enzymes, CasRx can also be flexibly packaged into adeno-associated virus. Virally encoded, catalytically inactive CasRx can be targeted to cis-elements of pre-mRNA to manipulate alternative splicing, alleviating dysregulated tau isoform ratios in a neuronal model of frontotemporal dementia. The results herein present CasRx as a programmable RNA-binding module for efficient targeting of cellular RNA, enabling a general platform for transcriptome engineering and therapeutic methods.

Class 2 CRISPR systems are found throughout diverse bacterial and archaeal life. Using a minimal definition of the CRISPR locus for bioinformatic mining of prokaryotic genome and metagenome sequences, which requires only a CRISPR repeat array and a nearby protein, provided herein is the identification of an uncharacterized, remarkably compact family of RNA-targeting class 2 CRISPR systems designated as Type VI CRISPR-Cas13d.

Because CRISPR systems generally exist as a functional operon within 20 kilobases of genome sequence, even fragmented metagenome reads may be sufficient to recover useful Cas enzymes for bioengineering purposes. CRISPR genome mining strategies described herein and by others (Shmakov et al., 2015), combined with ongoing efforts to profile microbial populations via next-generation sequencing, should contribute mechanistically diverse additions to the genome engineering toolbox.

Two distinct ribonuclease properties of the Cas13d effector, which processes a CRISPR repeat array into mature guides via a HEPN domain-independent mechanism followed by guide sequence-dependent recognition of a complementary activator RNA, were biochemically characterized. This triggers HEPN-mediated RNase activity, enabling Cas13d to cleave both activator and bystander RNAs, a property shared by other RNA-targeting CRISPR systems. Cas13d additionally exhibits no apparent flanking sequence requirements and was found to be active across crRNAs tiling a target RNA, suggesting the ability to target arbitrary single-stranded RNA sequences.

A comprehensive activity reporter screen in human cells of Cas13d orthologs sampled from distinct branches of the Cas13d family revealed that NLS fusions to Cas13d from *Ruminococcus flavefaciens* strain XPD3002 (CasRx) can be engineered for programmable RNA targeting in a eukaryotic context (**FIG. 8D**). CasRx fusions knocked down a diverse set of 14 endogenous mRNAs and lncRNAs, consistently achieving >90% knockdown with favorable efficiency relative to RNA interference, dCas9-mediated CRISPR interference, and other members of the Cas13 superfamily (**FIGS. 10A-10C**). Additionally, CasRx interference is markedly more specific than spacer-matching shRNAs, with no detectable off-target changes compared with hundreds for RNA interference.

CasRx is a minimal two-component platform, including an engineered CRISPR-Cas13d effector and an associated guide RNA, and can be fully genetically encoded. Because CasRx is an orthogonally delivered protein, HEPN-inactive dCasRx can be engineered as a flexible RNA-binding module to target specific RNA elements. Importantly, because CasRx uses a distinct ribonuclease activity to process guide RNAs, dCasRx can still be paired with a repeat array for multiplexing applications. The utility of this concept is shown herein by creating a dCasRx splice effector fusion for tuning alternative splicing and resulting protein isoform ratios, applying it in a neuronal model of frontotemporal dementia.

At an average size of 930 aa, Cas13d is the smallest class 2 CRISPR effector characterized in mammalian cells. This allows CasRx effector domain fusions to be paired with a CRISPR array encoding multiple guide RNAs while remaining under the packaging size limit of the versatile adeno-associated virus (AAV) delivery vehicle (Naldini, 2015) for primary cell and *in vivo* delivery. Further, targeted AAV delivery of CasRx to specific postmitotic cell types such as neurons can mediate long-term expression of a

corrective payload that avoids permanent genetic modifications or frequent re-administration (Chiriboga et al., 2016), complementing other nucleic acid targeting technologies such as DNA nuclease editing or antisense oligonucleotides. RNA mis-splicing diseases have been estimated to account for up to 15% of genetic diseases (Hammond and Wood, 2011), highlighting the potential for engineered splice effectors capable of multiplexed targeting. The materials provided herein can be used for RNA targeting for knockdown and splicing, such as live cell labeling and genetic screens to transcript imaging, trafficking, or regulation. CRISPR-Cas13d and engineered variants such as CasRx collectively enable flexible nucleic acid engineering, transcriptome-related study, and therapeutics, expanding the genome editing toolbox beyond DNA to RNA.

In a first aspect of the invention, there is provided an isolated CRISPR-Cas system guide RNA (gRNA), or isolated nucleic acid

encoding the gRNA, comprising one or more direct repeat (DR) sequences and one or more

spacer sequences,

wherein the one or more DR sequences comprise:

one or more DR sequences comprising or consisting of the nucleic acid sequence of SEQ ID NO: 131,129, 130,132, 133, 134, 135, 136, 137, 148, 150, 151, 152, 154, 156, 157, 159, 161, 163, 165, 167, 169, 176, 178, 180, 182, 184, 186, 188, 190, 191, 192, 193, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252 or 254;

one or more DR sequences comprising a 5'-end truncation of 1-10 nucleotides of one or more nucleic acid sequences of SEQ ID NOs: 131, 129, 130, 132, 133, 134, 135, 136, 137, 148, 150, 151, 152, 154, 156, 157, 159, 161, 163, 165, 167, 169, 176, 178, 180, 182, 184, 186, 188, 190, 191, 192, 193, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, or 254; or

one or more DR sequences comprising at least 80% sequence identity to the nucleic acid sequence of SEQ ID NO: 131, 129, 130, 132, 133, 134, 135, 136, 137, 148, 150, 151, 152, 154, 156, 157, 159, 161, 165, 167, 169, 176, 178, 180, 182, 184, 186, 188, 190, 191, 192, 193, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, or 254,

wherein the one or more spacer sequences hybridize with one or more target RNA molecules,

wherein the gRNA is capable of forming a complex with a Cas13d protein, and

wherein the gRNA directs the complex to the one or more target RNA molecules.

In a second aspect of the invention, there is provided an isolated protein comprising at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% sequence identity to SEQ ID NO: 208.

In a third aspect of the invention, there is provided an isolated protein, comprising at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% sequence identity to SEQ ID NO: 212.

In a fourth aspect of the invention, there is provided an isolated nucleic acid molecule selected from the group:

i) a nucleic acid molecule encoding the protein as described herein;

ii) a nucleic acid molecule that is codon optimised for expression in a host cell, optionally a eukaryotic cell, optionally a human cell, and comprises the amino acid sequence in SEQ ID NO: 208; or

iii) a nucleic acid molecule that is codon optimised for expression in a host cell, optionally a eukaryotic cell, optionally a human cell, and comprises the amino acid sequence in SEQ ID NO: 212.

In a fifth aspect of the invention, there is provided a recombinant vector comprising the isolated nucleic acid molecule as described herein.

In a sixth aspect of the invention, there is provided a pharmaceutical composition comprising an isolated protein, nucleic acid molecule or recombinant vector as described herein.

In a seventh aspect of the invention, there is provided a method of treating a neurodegenerative disorder, the method comprising administering an effective amount of the isolated protein, th nucleic acid molecule or recombinant vector as described herein to a subject in need thereof.

In an eight aspect of the invention, there is provided a method of treating a cancer, the method comprising administering an effective amount of the isolated protein, nucleic acid molecule or recombinant vector as described herein to a subject in need thereof.

In a ninth aspect of the invention, there is provided a non-naturally occurring or engineered clustered regularly interspaced short palindromic repeats (CRISPR)-associated (Cas) system for targeting one or more target RNA molecules, comprising:

(a) at least one catalytically inactive Cas13d protein comprising one or more mutated HEPN domains, or a nucleic acid molecule encoding the at least one catalytically inactive Cas13d protein; and

(b) at least one CRISPR-Cas system guide RNA (gRNA) comprising one or more direct repeat (DR) sequences and one or more spacer sequences, which hybridizes with the one or more target RNA molecules, or at least one nucleic acid molecule encoding the gRNA,

whereby the catalytically inactive Cas13d protein forms a complex with the gRNA, wherein the gRNA directs the complex to the one or more target RNA molecules and targets the one or more target RNA molecules,

wherein the at least one catalytically inactive Cas13d protein comprises a motif sequence of one or more of SEQ ID NOS: 195, 196, 197, 288, 289, 290, and 291, and,

wherein the one or more mutated HEPN domains comprise the amino acid sequence RXXXXH (SEQ ID NO: 255), and wherein the R, the H, or both the R and the H, of the amino acid sequence RXXXXH (SEQ ID NO: 255) is mutated.

In a tenth aspect of the invention, there is provided a method of targeting one or more target RNA molecules, comprising:

contacting one or more target RNA molecules with the non-naturally occurring or engineered CRISPR-Cas system as described herein.

In a eleventh aspect of the invention, there is provided a ribonucleoprotein (RNP) complex comprising:

the system as described herein, wherein the ribonucleoprotein complex comprises the (a) at least one catalytically inactive Cas13d protein comprising one or more mutated HEPN domains and the (b) at least one CRISPR-Cas system gRNA comprising one or more DR sequences and one or more spacer sequences, which hybridizes with the one or more target RNA molecules,

wherein the gRNA directs the ribonucleoprotein complex to the one or more target RNA molecules and targets the one or more target RNA molecules.

In an twelfth aspect of the invention, there is provided a non-bacterial cell comprising the non-naturally occurring or engineered CRISPR-Cas system as described herein.

In a thirteenth aspect of the invention, there is provided a bacterial cell comprising the non-naturally occurring or engineered CRISPR-Cas system as described herein, wherein the at least one catalytically inactive Cas13d protein and the gRNA are not native to the bacterial cell.

In a fourteenth aspect of the invention, there is provided a method of splicing or perturbing the splicing of one or more target RNA molecules, comprising:

contacting one or more target RNA molecules with the non-naturally occurring or engineered CRISPR-Cas system as described herein.

In a fifteenth aspect of the invention, there is provided a use of the isolated protein, nucleic acid molecule or recombinant vector as described herein in the manufacture of a medicament for the treatment of a neurodegenerative disorder.

In an sixteenth aspect of the invention, there is provided a use of the isolated protein, nucleic acid molecule or recombinant vector as described herein in the manufacture of a medicament for the treatment of a neurodegenerative disorder.

Provided herein are methods of targeting (*e.g.*, modifying, detecting) one or more target RNA molecules, such as a clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR associated (Cas) system-mediated RNA editing method. Such methods can include contacting one or more target RNA molecules with a non-naturally occurring or engineered (*e.g.*, does not naturally occur in the cell or system into which it is introduced) CRISPR-Cas system. Thus in some examples, the disclosed CRISPR-Cas system includes a naturally occurring Cas13d protein (or coding sequence) and a naturally occurring gRNA, but is used in a system or cell in which the Cas1 protein (or coding sequence) and the gRNA are not naturally found. Furthermore, the spacer sequence within the gRNA molecule is not naturally occurring, and has been modified to be complementary to the target RNA molecule.

In some examples, a target RNA is a coding RNA. In some examples, the RNA is non-coding RNA.

2018320870 28 Feb 2025

The disclosed CRISPR-Cas system can include (1) at least one Cas13d protein or at least one Cas13d nucleic acid coding sequence (such as a mRNA or a vector encoding the at least one Cas13d protein) and (2) at least one CRISPR-Cas system guide nucleic acid molecule (*e.g.*, gRNA) (or at least one nucleic acid molecule encoding the gRNA) having sufficient complementary to a target RNA such that it can hybridize to a target RNA molecule. The Cas13d protein forms a complex with the gRNA, and the gRNA directs the complex to the one or more target RNA molecules. This targeting can allow the Cas13d-gRNA complex to modify or detect the one or more target RNA molecules. In some examples, the one or more target RNA molecules (or a cell containing the one or more target RNA molecules) are contacted with a complex comprising the at least one Cas13d protein and the at least one gRNA. In some examples, the system includes Mg^{2+}. However, in some examples, the system does not require Mg^{2+}, such as if cleavage of the target RNA is not desired.

In some examples, contacting the one or more target RNA molecules with the non-naturally occurring or engineered CRISPR-Cas system includes introducing into a cell (such as a eukaryotic or prokaryotic cell) containing the one or more target RNA molecules the non-naturally occurring or engineered CRISPR-Cas system, for example using endocytosis (*e.g.*, receptor-mediated endocytosis, micropinocytosis), a liposome, a particle, an exosome, a microvesicle, a gene gun, electroporation, a virus, RNP-antibody fusion (*e.g.*, by tethering a Cas13d RNP to an antibody, antibody fragment, or other targeting

moiety [such as ScFv, aptamers, DARPins, nanobodies, affibodies, etc.], the RNP can be endocytosed into the cell, The RNP could conceivably be tethered to many things other than), or combinations thereof. Thus, cells can be transformed, transduced, transfected, or otherwise contacted with appropriate nucleic acid molecules of the disclosed CRISPR-Cas system. The resulting cells are recombinant cells. In some examples, contacting the one or more target RNA molecules with the non-naturally occurring or engineered CRISPR-Cas system includes contacting a cell-free system (such as a biological or environmental sample, or a cell lysate) containing the one or more target RNA molecules the non-naturally occurring or engineered CRISPR-Cas system (for example in a diagnostic method to detect a target RNA).

In some examples, at least 2, at least 3, at least 4, at least 5, at least 10, or at least 20 different gRNAs are used. For example, such a method could include targeting at least 2, at least 3, at least 4, at least 5, at least 10, or at least 20 different target RNA molecules, targeting at least 2, at least 3, at least 4, at least 5, at least 10, or at least 20 different regions of one or more RNA molecules, or combinations thereof.

Also provided are isolated nucleic acid molecules encoding such Cas13d proteins, such as a cDNA, genomic DNA, RNA, or mDNA. Such isolated nucleic acid molecules can be part of a vector (such as a plasmid or viral vector), and can be operably linked to a promoter. In some examples, an isolated nucleic acid molecule encoding a Cas13d protein has at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 124, 125, 126, 127, 128, 139, 140 or 141; or at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 142, 143, 144, or 145. In an additional example, an isolated nucleic acid encodes a Cas13d protein having at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296.

In some examples, an isolated nucleic acid molecule encoding at least one Cas13d protein (which can be part of a vector) includes at least one Cas13d protein coding sequence that is codon optimized for expression in a eukaryotic cell, at least one Cas13d protein coding sequence codon optimized for expression in a human cell. In one example, such a Cas13d coding sequence has at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 142, 143, 144, or 145, or has at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 124, 125, 126, 127, 128, 139, 140 or 141. In an

additional example, a eukaryotic cell codon optimized nucleic acid sequence encodes a Cas13d protein having at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296.

In some examples, the gRNA that hybridizes with the one or more target RNA molecules includes one or more direct repeat (DR) sequences, one or more spacer sequences, or one or more sequences comprising DR-spacer-DR-spacer. In some examples, the one or more DR sequences have at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 129, 130, 131, 132, 133, 134, 135, 136, 137, 148, 150, 151, 152, 154, 156, 157, 159, 161, 163, 165, 167, 169, 176, 178, 180, 182, 184, 186, 188, 190, 191, 192, 193, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, or 254. In one example, the gRNA includes additional sequences, such as an aptamer sequence.

In some examples, a plurality of gRNAs are processed from a single array transcript, wherein each gRNA can be different, for example to target different RNAs or target multiple regions of a single RNA.

In some examples, the DRs are truncated by 1-10 nucleotides (such as 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides) at the 5'end, for example to be expressed as mature pre-processed guide RNAs.

Methods of targeting one or more target RNA molecules are provided. Targeting an RNA molecule can include one or more of cutting or nicking one or more target RNA molecules, activating or upregulating one or more target RNA molecules, activating or suppressing translation the one or more target RNA molecules, deactivating the one or more target RNA molecules, visualizing, labeling, or detecting the one or more target RNA molecules, binding the one or more target RNA molecules, editing the one or more target RNA molecules, trafficking the one or more target RNA molecules, and masking the one or more target RNA molecules. In some example, modifying one or more target RNA molecules includes one or more of an RNA base substitution, an RNA base deletion, an RNA base insertion, a break in the target RNA, methylating RNA, and demethylating RNA.

In some examples, such methods are used to treat a disease, such as a disease in a human. In such examples, the one or more target RNA molecules is associated with the disease.

Also provided are isolated proteins, including non-naturally occurring proteins. In some examples, a protein has at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,

14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296. In some examples, an isolated protein is a Cas13d ortholog from a prokaryotic genome or metagenome, gut metagenome, an activated sludge metagenome, an anaerobic digester metagenome, a chicken gut metagenome, a human gut metagenome, a pig gut metagenome, a bovine gut metagenome, a sheep gut metagenome, a goat gut metagenome, a capybara gut metagenome, a primate gut metagenome, a termite gut metagenome, a fecal metagenome, a genome from the Order Clostridiales, or the Family Ruminococcaceae. In some examples, a Cas13d ortholog includes a Cas13d ortholog from Ruminococcus albus, Eubacterium siraeum, a Ruminococcus flavefaciens strain XPD3002, Ruminococcus flavefaciens FD-1, uncultured Eubacterium sp TS28-c4095, uncultured Ruminococcus sp., Ruminococcus bicirculans, or Ruminococcus sp CAG57. Such proteins can include a subcellular localization signal. In some examples, such proteins include a mutation in at least one native HEPN domain.

Also provided are isolated guide RNA (gRNA) molecules. In some examples, an isolated gRNA includes one or more direct repeat (DR) sequences, such as one having at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 129, 130, 131, 132, 133, 134, 135, 136, 137, 148, 150, 151, 152, 154, 156, 157, 159, 161, 163, 165, 167, 169, 176, 178, 180, 182, 184, 186, 188, 190, 191, 192, 193, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, or 254. Such a gRNA can further include one or more spacer sequences specific for (*e.g.*, is complementary to) the target RNA. Such guide RNAs can further be optionally truncated by 1-10 nucleotides (such as 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides) at the 5'end of the DR, for example to generate pre-processed guide RNAs.

Also provided are ribonucleoprotein (RNP) complexes, which include one or more Cas13d proteins provided herein and one or more gRNAs provided herein.

Also provided are recombinant cells that include any Cas13d protein (or nucleic acid molecule encoding Cas13d), any gRNA, any RNP complex, or any vector, provided herein. In one example, the cell is not a bacterial cell. In one example, the cell is a bacterial cell.

Also provided are compositions that include one or more of any Cas13d protein (or nucleic acid molecule encoding Cas13d), any gRNA or array, any RNP complex, any isolated nucleic acid molecule, any vector, or any cell, provided herein. Such compositions can include a pharmaceutically acceptable carrier.

Also provided are kits. Such kits can include one or more of any Cas13d protein (or nucleic acid molecule encoding Cas13d), any gRNA or array, any RNP complex, any isolated nucleic acid molecule, any

vector, any cell, or any composition provided herein. Such reagents can be combined or in separate containers.

In some examples, a Cas13d protein is programmed toward its RNA target by combining the protein (or nucleic acid encoding the protein) with an engineered RNA guide (or nucleic acid encoding RNA guide) consisting of a full or partial direct repeat sequence followed by a "spacer" sequence complementary to the RNA target(s) (or variations thereof, i.e. arrays (DR-spacer-DR-spacer-DR-spacer..etc.) or pre-guide RNAs (DR-spacer-DR). Cas13d Proteins can be catalytically inactivated and transformed into RNA binding modules by mutating the conserved RNAse HEPN motif (RXXXXH). Exemplary Cas13d proteins and corresponding guides are provided herein (e.g., SEQ ID NOS: 147-170, 175-193 and SEQ ID NOS: 198-254).

A. Cas13d Proteins

Provided herein are novel Cas13d proteins, such as those as shown in in the sequence listing. SEQ ID NOS: 1, 3, 42, 62, 70, 82, 83, 92, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 292, 293, 294, 295, and 296 provide distinct full length proteins, and SEQ ID NOS: 2, 4-41, 43-61, 63-69, 71-81, 84-91, 93-113, 194, 278, 279, 280, 281, 282, 283, 284, and 285 provide Cas13d variants and fragments. Such proteins can be used in the disclosed methods, compositions, and kits.

In some examples, a Cas13d protein includes one or more (such as 1 or 2) native HEPN domains. In some examples, a Cas13d protein includes one or more mutated HEPN domains (such as mutant Cas13d protein can process the gRNA, but cannot modify the target RNA). In some examples, a Cas13d protein is no more than 150 kD, no more than 140 kD, no more than 130 kD, no more than 120 kD, such as about 90 to 120 kD, about 100 to 120 kD or about 110 kD.

In addition to the Cas13d proteins provided herein, the disclosure encompasses Cas13d orthologs from a prokaryotic genome or metagenome, gut metagenome, an activated sludge metagenome, an anaerobic digester metagenome, a chicken gut metagenome, a human gut metagenome, a pig gut metagenome, a bovine gut metagenome, a sheep gut metagenome, a goat gut metagenome, a capybara gut metagenome, a primate gut metagenome, a termite gut metagenome, a fecal metagenome, a genome from the Order Clostridiales, or the Family Ruminococcaceae, such as a Cas13d ortholog from Ruminococcus albus, Eubacterium siraeum, a Ruminococcus flavefaciens strain XPD3002, Ruminococcus flavefaciens FD-1, uncultured Eubacterium sp TS28-c4095, uncultured Ruminococcus sp., Ruminococcus bicirculans, or Ruminococcus sp CAG57.

In some examples, a Cas13d protein is at least 800 aa, at least 900 aa, or at least 1000 aa, such as 800 to 1200 aa, 850 to 1050 aa, or 860-1040 aa.

1. Variant Cas13d Sequences

Cas13d proteins, including variants of the sequences provided herein (such as variants of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296) are encompassed within this disclosure. In some examples, Cas13d proteins provided herein can contain one or more mutations, such as a single insertion, a single deletion, a single substitution, or combinations thereof. In some examples, the Cas13d protein includes at least 1, at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 40, at least 50, at least 75, at least 100, at least 200 or at least 300 aa insertions, such as 1-20 insertions (for example at the N- or C-terminus or within the protein, such as insertion of a whole small domain), at least 1, at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 40, at least 50, at least 75, at least 100, at least 200 or at least 300 aa deletions (such as deletion of a whole small domain), such as 1-20 deletions (for example at the N- or C-terminus or within the protein), at least 1, at least 5, at least 10, at least 15, at least 20, at least 25, at least 30 aa substitutions, such as 1-20 substitutions, or any combination thereof (*e.g.*, single insertion together with 1-19 substitutions), but retain the ability to bind target RNA molecules complementary to the spacer sequence within the gRNA molecule and/or process an guide array RNA transcript into gRNA molecules and/or retain the ability to cleave target RNA. Exemplary Cas13 proteins with deletions in non-conserved regions are shown in SEQ ID NOS: 278, 279, 280, 281, 282, 283, 284, and 285. In some examples, the disclosure provides a variant of any disclosed Cas13d protein (such as SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296) having 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 or 30 amino acid changes, but retain the ability to bind target RNA molecules complementary to the spacer sequence within the gRNA molecule and/or process an guide array RNA transcript into gRNA molecules. In some examples, any disclosed Cas13d protein (such as SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138,

147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, or 25, or variant thereof, such as SEQ ID NO: 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296 further includes 1-8 amino acid insertions, 1-15 amino acid deletions, 1-10 amino acid substitutions, or any combination thereof (*e.g.*, 1-15, 1-4, or 1-5 amino acid deletions together with 1-10, 1-5 or 1-7 amino acid substitutions), with the retained ability to bind target RNA molecules complementary to the spacer sequence within the gRNA molecule and/or process an guide array RNA transcript into gRNA molecules. In one example, such variant peptides are produced by manipulating the nucleotide sequence encoding a peptide using standard procedures such as site-directed mutagenesis or PCR. Such variants can also be chemically synthesized.

In some examples, a Cas13d protein includes a motif shown in SEQ ID NO: 195, 196 or 197. Thus, a Cas13d protein having at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296 in some examples includes at least one motif shown in SEQ ID NO: 195, 196 or 197.

In some examples, a Cas13d protein includes a conserved sequence shown in SEQ ID NO: 288, 289, 290, or 291. Thus, a Cas13d protein having at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296 in some examples includes at least one motif shown in SEQ ID NO: 288, 289, 290, or 291.

One type of modification or mutation includes the substitution of amino acids for amino acid residues having a similar biochemical property, that is, a conservative substitution (such as 1-4, 1-8, 1-10, or 1-20 conservative substitutions). Typically, conservative substitutions have little to no impact on the activity of a resulting peptide. For example, a conservative substitution is an amino acid substitution in a Cas13d protein (such as SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22,

23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296) that does not substantially affect the ability of the Cas13d protein to bind target RNA molecules complementary to the spacer sequence within the gRNA molecule and/or process an guide array RNA transcript into gRNA molecules. An alanine scan can be used to identify which amino acid residues in a Cas13d protein (such as SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296), can tolerate an amino acid substitution. In one example, the ability of a variant Cas13d protein (such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296) to modify gene expression in a CRISPR/Cas system, is not altered by more than 25%, for example not more than 20%, for example not more than 10%, when an alanine, or other conservative amino acid, is substituted for 1-4, 1-8, 1-10, or 1-20 native amino acids. Examples of amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative substitutions include: Ser for Ala; Lys for Arg; Gln or His for Asn; Glu for Asp; Ser for Cys; Asn for Gln; Asp for Glu; Pro for Gly; Asn or Gln for His; Leu or Val for Ile; Ile or Val for Leu; Arg or Gln for Lys; Leu or Ile for Met; Met, Leu or Tyr for Phe; Thr for Ser; Ser for Thr; Tyr for Trp; Trp or Phe for Tyr; and Ile or Leu for Val.

One method for identifying regions particularly amenable to insertions, substitutions or deletion is to target stretches of amino acids exhibiting low levels of conservation between orthologs. Such regions are indicated in the conservation graph of the alignment of Cas13d proteins provided in FIG. 1B. Conserved residues of Cas13d are further marked in the Cas13d protein alignment provided in Fig 18A-18MMM (indicated by symbols "." ":" or "*" below aligned conserved residues). Examples of deletions and their

functional testing are further provided in FIGS. 16A-16B. Such methods were used to generate the functional deletion variants of Cas13d are shown in SEQ ID NOS: 278, 279, 280, 281, 282, 283, 284, and 285. Such as variants can further include or changes, such as 1-30, 1 to 20, or 5 to 30, amino acid substitutions, such as conservative amino acid substitutions.

Another type of substitution can be achieved by swapping out parts of one ortholog with the homologous region of another ortholog to obtain a combined "chimeric" protein. Such a chimeric protein may combine favorable properties of multiple Cas13d orthologs.

More substantial changes can be made by using substitutions that are less conservative, *e.g.*, selecting residues that differ more significantly in their effect on maintaining: (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation; (b) the charge or hydrophobicity of the polypeptide at the target site; or (c) the bulk of the side chain. The substitutions that in general are expected to produce the greatest changes in polypeptide function are those in which: (a) a hydrophilic residue, *e.g.*, serine or threonine, is substituted for (or by) a hydrophobic residue, *e.g.*, leucine, isoleucine, phenylalanine, valine or alanine; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, *e.g.*, lysine, arginine, or histidine, is substituted for (or by) an electronegative residue, *e.g.*, glutamic acid or aspartic acid; or (d) a residue having a bulky side chain, *e.g.*, phenylalanine, is substituted for (or by) one not having a side chain, *e.g.*, glycine.

Thus, the disclosure provides Cas13d proteins having at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296 or combinations (*e.g.*, chimeras) thereof.

In one example, a Cas13d protein includes non-naturally occurring amino acids.

2. Cas13d proteins with Other Elements

A Cas13d protein (such as any of SEQ ID NOS: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280,

281, 282, 283, 284, 285, 292, 293, 294, 295, or 296) can include other elements or domains, for example at the N- or C-terminus (or both). The resulting protein can be referred to as a Cas13d fusion protein.

In one example, a Cas13d protein provided herein (such as a native Cas13d, a truncated Cas13d, or a Cas13d with mutated HEPN domain(s)) includes a subcellular localization signal. Exemplary subcellular localization signals include an organelle localization signal, such as a nuclear localization signal (NLS), nuclear export signal (NES), or mitochondrial localization signal. In one example, a Cas13d protein includes an NLS, such as SPKKKRKVEAS (SEQ ID NO: 256; *e.g.,* encoded by AGCCCCAAGAAgAAGAGaAAGGTGGAGGCCAGC, SEQ ID NO: 257) or GPKKKRKVAAA (SV40 large T antigen NLS, SEQ ID NO: 258; *e.g.,* encoded by ggacctaagaaaaagaggaaggtggcggccgct, SEQ ID NO: 260). Exemplary NES that can be part of a Cas13d protein include an adenovirus type 5 E1B nuclear export sequence, an HIV nuclear export sequence (*e.g.*, see SEQ ID NO: 287), a MAPK nuclear export sequence, or a PTK2 nuclear export sequence.

In some examples, the at least one Cas13d protein (such as a native Cas13d, a truncated Cas13d, or a Cas13d with mutated HEPN domain(s)) further includes one or more effector domains. Exemplary effector domains include proteins and/or enzymes, such that those can cleave RNA (*e.g.*, a PIN endonuclease domain, an NYN domain, an SMR domain from SOT1, or an RNase domain from Staphylococcal nuclease), those that can affect RNA stability (e.g., tristetraprolin (TTP) or domains from UPF1, EXOSC5, and STAU1), those that can edit a nucleotide or ribonucleotide (*e.g.*, a cytidine deaminase, PPR protein, adenosine deaminase, ADAR family protein, or APOBEC family protein), those that can activate translation (*e.g.*, eIF4E and other translation initiation factors, a domain of the yeast poly(A)-binding protein or GLD2), those that can repress translation (*e.g.*, Pumilio or FBF PUF proteins, deadenylases, CAF1, Argonaute proteins), those that can methylate RNA (*e.g.*, domains from m6A methyltransferase factors such as METTL14, METTL3, or WTAP), those that can demethylate RNA (*e.g.*, human alkylation repair homolog 5 or Alkbh5), those that can affect splicing (*e.g.*, the RS-rich domain of SRSF1, the Gly-rich domain of hnRNP A1, the alanine-rich motif of RBM4, or the proline-rich motif of DAZAP1), those that can enable affinity purification or immunoprecipitation (*e.g.*, FLAG, HA, biotin, or HALO tags), and those that can enable proximity-based protein labeling and identification (e.g., a biotin ligase (such as BirA) or a peroxidase (such as APEX2) in order to biotinylate proteins that interact with the target RNA).

In some examples, the Cas13d protein and effector module combination can constitute a transcriptional sensor. For example, the transcriptional sensor can be comprised of at least one Cas13d protein with a mutated HEPN domain (*e.g.*, SEQ ID NO: 2 or 4), at least one gRNA containing at least one spacer sequence specific for the target RNA, and an effector module such as an optionally split fluorescent protein or probe (*e.g.*, a split Venus fluorescent protein, a split GFP, a split enhanced GFP, a split mCherry, a split super-folder mCherry, and other fluorescent protein variants such as ECFP, YFP, RFP, and derivatives or fragments thereof); an optionally split luminescent protein or probe (e.g. Gaussia, Firefly, NanoLuc, or Renilla variants); an optionally split enzyme (*e.g.*, ubiquitin or TEV protease); a FRET-compatible protein pair; one or more transcription factor(s) fused to Cas13d via cleavable linkers (*e.g.*, an

artificial GAL4, zinc finger, transcriptional activator like effector (TALE), CRISPR-Cas9, CRISPR-Cpf1, or TetR-based transcription factor or an endogenous transcription factor); a split intein that trans-splices a protein to restore its function such as a transcription factor (*e.g.*, an intein from Rhodothermus marinus or DnaE); a kinase-substrate pair that activates upon phosphorylation (*e.g.*, TYK2-STAT3); one, two, or more monomers that activate upon dimerization or multimerization (*e.g.*, caspase 9); or one or more proteins that induce conformational and functional change upon interaction. In one example, the spatial proximity of two or more Cas13d proteins and gRNAs due to binding a particular transcript would activate the effector module, resulting in a detectable signal or detectable activity in the cell.

In one example, the effector domain is fused to a protein that specifically recognizes and binds an RNA aptamer, such as one that can be appended to or inserted within a gRNA molecule (*e.g.*, an MS2, PP7, Qβ, and other aptamers). This aptamer-effector domain fusion can be used to target the target RNA because the Cas13d and gRNA complex will guide the aptamer protein-effector domain in proximity to the target RNA.

In another example, the aptamer can be directly inserted into the gRNA molecule to permit detection of a target RNA, such as a fluorophore aptamer (*e.g.*, Spinach, Mango, etc.).

In some examples, the Cas13d protein (such as a native Cas13d, a truncated Cas13d, or a Cas13d with mutated HEPN domain(s)) includes a purification tag, such as an HA-tag, His-tag (such as 6-His), Myc-tag, E-tag, S-tag, calmodulin tag, FLAG-tag, GST-tag, MBP-tag, and the like. Such tags are in some examples at the N- or C-terminal end of the Cas13d protein.

In some examples, a Cas13d protein (such as a native Cas13d, a truncated Cas13d, or a Cas13d with mutated HEPN domain(s)) includes one or more subcellular localization signals, effector domains, and purification tags.

In some examples, a Cas13d protein may be split into multiple fragments, which are then expressed individually. Such fragments of Cas13d may be optionally fused to other protein domains. In one example, a Cas13d can be split into two halves, which are then fused to two parts of an inducible heterodimer pair. Upon induction of heterodimer binding, the Cas13d halves are recruited to each other to form an active protein. Such a system would allow for the inducible control of Cas13d activity. Useful heterodimer pairs include two proteins that dimerize upon light illumination or through administration of a small molecule compound, amongst others. Specific examples of heterodimer pair include but are not limited to: light inducible Magnets proteins, the light inducible iLID-SspB pair, the light inducible Cryptochrome2-CIB1 dimer and the small molecule inducible FKBP protein. In another example of a split Cas13d design, two halved of the Cas13d protein may be fused to protein trans-splicing domains. Such a design would enable the separate expression of two halves which are reconstituted into a full length protein once expressed inside a cell. An example of such transsplicing domains includes the Intein system.

One method for identifying regions particularly amenable to splitting of the protein, is to identify stretches of amino acids exhibiting low levels of conservation between orthologs. Such regions are indicated in the conservation graph of the alignment of Cas13d proteins provided in FIG. 1B. Regions of conserved

residues of Cas13d are further marked in the Cas13d protein alignment provided in Fig 18A-18MMM (indicated by symbols . : or * below aligned conserved residues).

3. Generation of Cas13d Proteins

In one example, the Cas13d protein is expressed *in vitro*, for example, in a prokaryotic cell (*e.g.*, bacteria such as *Lactobacillus*, *Lactococcus*, *Bacillus* (such as *B. subtilis*), *Escherichia* (such as *E. coli*), *Salmonella typhimurium,* and *Clostridium*), archea cell, plant or plant cell, fungal cell (*e.g.*, *Neurospora*), yeast cell (*e.g.*, *Saccharomyces or Pichia* (such as *S. cerevisiae* or *P. pastoris*), *Kluyveromyces lactis*), insect cell (*e.g.*, SF9 cells), or mammalian cells (*e.g.*, 293 cells, or immortalized mammalian myeloid and lymphoid cell lines). Once expressed, the Cas13d protein can be isolated and/or purified (e.g., using chromatography or immunological separation). In some examples, as tag on the Cas13d protein permits isolation of the protein from a culture media. Exemplary procedures include ammonium sulfate precipitation, affinity columns, column chromatography, and the like (see, generally, R. Scopes, *Protein Purification*, Springer-Verlag, N.Y., 1982). Substantially pure compositions of at least about 90 to 95% homogeneity, such as 98% to 99% homogeneity, can be used in the methods provided herein. For example, a purified preparation of a Cas13d protein can be used as an alternative to expressing the Cas13d protein from a nucleic acid molecule in the CRISPR/Cas system.

In addition to recombinant methods, Cas13d proteins disclosed herein can also be constructed in whole or in part using native chemical ligation and/or expressed protein ligation.

B. Nucleic Acid Molecules Encoding Cas13d Proteins

Nucleic acid molecules encoding a Cas13d protein are encompassed by this disclosure. Nucleic acid molecules include DNA, genomic DNA, cDNA, mRNA, and RNA sequences which encode a Cas13d peptide. Such nucleic acid molecules can include naturally occurring or non-naturally occurring nucleotides or ribonucleotides. Exemplary nucleic acid molecules that encode the novel Cas13d proteins of SEQ ID NO: 1, 3, 42, 62, 70, 82, 83, 92 and 104, are shown in SEQ ID NOS: 124-128, 139, 140, and 141. Also provided are codon optimized nucleic acid molecules that encode the novel Cas13d proteins, for example those optimized for expression in a mammalian cells, such as a human cell (SEQ ID NOS: 114-123 and 142-145). For example, SEQ ID NOS: 114, 118, and 122 provide nucleic acid molecules optimized for expression in human cells. SEQ ID NOS: 115, 119 and 123 provide nucleic acid molecules optimized for expression in human cells, and which encode for mutant HEPN sites. SEQ ID NOS: 116 and 120 provide nucleic acid molecules optimized for expression in human cells, and which includes an N-terminal nuclear localization (NLS) coding sequence (namely, SPKKKRKVEAS). SEQ ID NO: 117 and 121 provide nucleic acid molecules optimized for expression in human cells, and which include N-terminal and C-terminal NLS coding sequences (namely, SPKKKRKVEAS, SEQ ID NO: 256, and GPKKKRKVAAA SEQ ID NO: 258, respectively).

In one example, a nucleic acid sequence encodes a Cas13d protein having at least 60%, at least 70%, at least 75%, at least 80%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 99% or at least 99% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296. Such nucleic acid sequences can be generated based on the amino acid sequences provided herein, and the genetic code. In one example, a Cas13d nucleic acid sequence has at least 70%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 98%, or at least 99% sequence identity to SEQ ID NO: 124, 125, 126, 127, 128, 139, 140 or 141. In one example, a Cas13d nucleic acid sequence is optimized for expression in mammalian cells, such as human cells, such as one having at least 70%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 98%, or at least 99% sequence identity to SEQ ID NO: 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 142, 143, 144, or 145.

One of skill can readily construct a variety of clones containing functionally equivalent nucleic acids, such as nucleic acids which differ in sequence but which encode the same Cas13d protein sequence. Silent mutations in the coding sequence result from the degeneracy (*i.e.*, redundancy) of the genetic code, whereby more than one codon can encode the same amino acid residue. Thus, for example, leucine can be encoded by CTT, CTC, CTA, CTG, TTA, or TTG; serine can be encoded by TCT, TCC, TCA, TCG, AGT, or AGC; asparagine can be encoded by AAT or AAC; aspartic acid can be encoded by GAT or GAC; cysteine can be encoded by TGT or TGC; alanine can be encoded by GCT, GCC, GCA, or GCG; glutamine can be encoded by CAA or CAG; tyrosine can be encoded by TAT or TAC; and isoleucine can be encoded by ATT, ATC, or ATA. Tables showing the standard genetic code can be found in various sources (see, for example, Stryer, 1988, Biochemistry, 3rd Edition, W.H. 5 Freeman and Co., NY).

Based on the genetic code, nucleic acid sequences coding for any Cas13d sequence can be generated. In some examples, such a sequence is optimized for expression in a host or target cell, such as a host cell used to express the Cas13d protein or a cell in which the disclosed methods are practice (such as in a mammalian cell, such as a human cell). Codon preferences and codon usage tables for a particular species can be used to engineer isolated nucleic acid molecules encoding a Cas13d (such as one encoding a protein having at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101,

102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, or 253 that takes advantage of the codon usage preferences of that particular species). For example, the Cas13d proteins disclosed herein can be designed to have codons that are preferentially used by a particular organism of interest. In one example, a Cas13d nucleic acid sequence is optimized for expression in human cells, such as one having at least 70%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 98%, or at least 99% sequence identity to SEQ ID NO: 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 142, 143, 144, or 145.

A nucleic acid encoding a Cas13d protein (such as one encoding a protein having at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296) can be cloned or amplified by *in vitro* methods, such as the polymerase chain reaction (PCR), the ligase chain reaction (LCR), the transcription-based amplification system (TAS), the self-sustained sequence replication system (3SR) and the Qβ replicase amplification system (QB). In addition, nucleic acids encoding a Cas13d protein (such as one encoding a protein having at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296) can be prepared by cloning techniques. Examples of appropriate cloning and sequencing techniques, and instructions sufficient to direct persons of skill through cloning are found in Sambrook *et al.* (ed.), Molecular Cloning: A Laboratory Manual 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring, Harbor, N.Y., 1989, and Ausubel et al., (1987) in "Current Protocols in Molecular Biology," John Wiley and Sons, New York, N.Y..

Nucleic acid sequences encoding a Cas13d protein (such as one encoding a protein having at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least

99%, or 100% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296) can be prepared by any suitable method including, for example, cloning of appropriate sequences or by direct chemical synthesis by methods such as the phosphotriester method of Narang *et al., Meth. Enzymol.* 68:90-99, 1979; the phosphodiester method of Brown *et al., Meth. Enzymol.* 68:109-151, 1979; the diethylphosphoramidite method of Beaucage *et al., Tetra. Lett.* 22:1859-1862, 1981; the solid phase phosphoramidite triester method described by Beaucage & Caruthers, *Tetra. Letts.* 22(20):1859-1862, 1981, for example, using an automated synthesizer as described in, for example, Needham-VanDevanter *et al., Nucl. Acids Res.* 12:6159-6168, 1984; and, the solid support method of U.S. Patent No. 4,458,066. Chemical synthesis produces a single stranded oligonucleotide. This can be converted into double stranded DNA by hybridization with a complementary sequence, or by polymerization with a DNA polymerase using the single strand as a template. One of skill would recognize that while chemical synthesis of DNA is generally limited to sequences of about 100 bases, longer sequences may be obtained by the ligation of shorter sequences.

In one example, a Cas13d protein (such as a protein having at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296) is prepared by inserting the cDNA which encodes the Cas13d protein into a plasmid or vector. The insertion can be made so that the Cas13d protein is read in frame so that the Cas13d protein is produced.

The Cas13d nucleic acid coding sequence (such as one having at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181,

183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296) can be inserted into an expression vector including, but not limited to a plasmid, virus or other vehicle that can be manipulated to allow insertion or incorporation of sequences and can be expressed in either prokaryotes or eukaryotes. Hosts can include microbial, yeast, insect, plant and mammalian organisms. The vector can encode a selectable marker, such as a thymidine kinase gene or antibiotic resistance gene.

Nucleic acid sequences encoding a Cas13d protein (such as one encoding a protein having at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296) can be operatively linked to expression control sequences. An expression control sequence operatively linked to a Cas13d coding sequence is ligated such that expression of the Cas13d protein coding sequence is achieved under conditions compatible with the expression control sequences. The expression control sequences include, but are not limited to appropriate promoters, enhancers, transcription terminators, a start codon (*i.e.*, ATG) in front of a Cas13d protein-encoding gene, splicing signal for introns, maintenance of the correct reading frame of that gene to permit proper translation of mRNA, and stop codons.

In one embodiment, vectors are used for expression in yeast such as *S. cerevisiae*, *P. pastoris*, or *Kluyveromyces lactis*. Exemplary promoters for use in yeast expression systems include but are not limited to: the constitutive promoters plasma membrane H^+-ATPase (*PMA1*), glyceraldehyde-3-phosphate dehydrogenase (*GPD*), phosphoglycerate kinase-1 (*PGK1*), alcohol dehydrogenase-1 (*ADH1*), and pleiotropic drug-resistant pump (*PDR5*). In addition, many inducible promoters are of use, such as *GAL1–10* (induced by galactose), *PHO5* (induced by low extracellular inorganic phosphate), and tandem heat shock *HSE* elements (induced by temperature elevation to 37°C). Promoters that direct variable expression in response to a titratable inducer include the methionine-responsive *MET3* and *MET25* promoters and copper-dependent *CUP1* promoters. Any of these promoters may be cloned into multicopy (2μ) or single copy (*CEN*) plasmids to give an additional level of control in expression level. The plasmids can include nutritional markers (such as *URA3*, *ADE3*, *HIS1*, and others) for selection in yeast and antibiotic resistance (*AMP*) for propagation in bacteria. Plasmids for expression on *K. lactis* are known, such as pKLAC1.

Viral vectors can also be prepared that encode a Cas13d (such as one encoding a protein having at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at

least 99%, or 100% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296). Exemplary viral vectors include polyoma, SV40, adenovirus, vaccinia virus, adeno-associated virus, herpes viruses including HSV and EBV, lentivirus, Sindbis viruses, alphaviruses and retroviruses of avian, murine, and human origin. Baculovirus (Autographa californica multinuclear polyhedrosis virus; AcMNPV) vectors can be used and obtained from commercial sources. Other suitable vectors include retrovirus vectors, orthopox vectors, avipox vectors, fowlpox vectors, capripox vectors, suipox vectors, adenoviral vectors, herpes virus vectors, alpha virus vectors, baculovirus vectors, Sindbis virus vectors, vaccinia virus vectors and poliovirus vectors. Specific exemplary vectors are poxvirus vectors such as vaccinia virus, fowlpox virus and a highly attenuated vaccinia virus (MVA), adenovirus, baculovirus and the like. Pox viruses of use include orthopox, suipox, avipox, and capripox virus. Orthopox include vaccinia, ectromelia, and raccoon pox. One example of an orthopox of use is vaccinia. Avipox includes fowlpox, canary pox and pigeon pox. Capripox include goatpox and sheeppox. In one example, the suipox is swinepox. Other viral vectors that can be used include other DNA viruses such as herpes simplex virus and adenoviruses, and RNA viruses such as retroviruses and polio.

Viral vectors that encode a Cas13d protein (such as one encoding a protein having at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283, 284, 285, 292, 293, 294, 295, or 296) can include at least one expression control element operationally linked to the nucleic acid sequence encoding the Cas13d protein. The expression control elements control and regulate the expression of the Cas13d nucleic acid sequence. Exemplary expression control elements that can be used include, but are not limited to, lac system, operator and promoter regions of phage lambda, yeast promoters and promoters derived from polyoma, adenovirus, retrovirus or SV40. In one example the promoter is CMV, U6, CBh, CMW, Cbh, EF1a. In one example, the promoter is a cell type specific promoter, such as synapsin or GFAP, or an inducible promoter, such as a tetracycline inducible promoter. Additional operational elements include, but are not limited to, leader sequence, termination codons,

polyadenylation signals and any other sequences necessary for the appropriate transcription and subsequent translation of the nucleic acid sequence encoding the Cas13d protein in the host system. The expression vector can contain additional elements necessary for the transfer and subsequent replication of the expression vector containing the nucleic acid sequence in the host system. Examples of such elements include, but are not limited to, origins of replication and selectable markers.

In one example, the vector includes a polyA signal after the Cas13d protein coding sequence, a WPRE signal for expression in viral vectors, or combinations thereof.

In one example, the method uses direct delivery of an mRNA that encodes for a Cas13d protein.

C. Guide Nucleic Acid Molecules

The disclosure provides guide nucleic acid molecules, such as guide RNA (gRNA or crRNA, CRISPR (guide) RNA), which can be used in the methods, compositions, and kits provided herein. Such molecules can include naturally occurring or non-naturally occurring nucleotides or ribonucleotides (such as LNAs or other chemically modified nucleotides or ribonucleotides, for example to protect a guide RNA from degradation). In some examples, the guide sequence is RNA. The guide nucleic acid can include modified bases or chemical modifications (*e.g.*, see Latorre *et al.*, *Angewandte Chemie* 55:3548-50, 2016). A guide sequence directs a Cas13d protein to a target RNA, thereby targeting the RNA (*e.g.*, modifying or detecting the RNA).

Guide molecules include one or more regions referred to as spacers. A spacer has sufficient complementarity with a target RNA sequence to hybridize with the target RNA and direct sequence-specific binding of a Cas13d protein to the target RNA. Thus, the spacer is the variable portion of the guide sequence. In some examples, a spacer has 100% complementarity to a target RNA (or region of the RNA to be target), but a spacer can have less than 100% complementarity to a target RNA, such as at least 80%, at least 85%, at least 90%, at least 95%, at least 98% or at least 99% complementarity to a target RNA.

A guide sequence can also include one or more direct repeats (DRs). The DR is the constant portion of the guide, which contains strong secondary structure (**FIG. 3C**), which facilitate interaction between a Cas13d protein and the guide molecule. Each ortholog has a slightly different DR sequence (*e.g.*, SEQ ID NOS: 129, 130, 131, 132, 133, 134, 135, 136, 137, 148, 150, 151, 152, 154, 156, 157, 159, 161, 163, 165, 167, and 169). In one example, the gRNA includes at least one DR sequence having at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 129, 130, 131, 132, 133, 134, 135, 136, 137, 148, 150, 151, 152, 154, 156, 157, 159, 161, 163, 165, 167, 169, 176, 178, 180, 182, 184, 186, 188, 190, 191, 192, 193, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, or 254 (such as 1, 2, 3, or 4 of such DR sequences).

In one example, a guide sequence includes a constant DR on its 5'-end and a variable spacer on its 3' end. In one example includes the sequence DR-spacer-DR-spacer. In some examples, the sequence DR-

spacer is repeated two or more times, such as at least 3 times or at least 4 times. This type of sequence is called a guide array.

Guide molecules generally exist in various states of processing. In one example, an unprocessed guide RNA is 36nt of DR followed by 30-32 nt of spacer. The guide RNA is processed (truncated/modified) by Cas13d itself or other RNases into the shorter "mature" form. In some embodiments, an unprocessed guide sequence is about, or at least about 30, 35, 40, 45, 50, 55, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, or more nucleotides (nt) in length. In some embodiments, a processed guide sequence is about 44 to 60 nt (such as 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, or 70 nt). In some embodiments, an unprocessed spacer is about 28-32 nt long (such as 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, or 35 nt) while the mature (processed) spacer can be about 10 to 30 nt, 10 to 25 nt, 14 to 25 nt, 20 to 22 nt, or 14-30 nt (such as 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, or 35 nt). In some embodiments, an unprocessed DR is about 36 nt (such as 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40 or 41 nt), while the processed DR is about 30 nt (such as 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, or 35 nt).

The ability of a guide sequence to direct sequence-specific binding of a CRISPR complex to a target RNA 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 RNA molecule, such as by transfection with vectors encoding the components of the CRISPR sequence, followed by an assessment of preferential cleavage within the target sequence. Similarly, cleavage of a target RNA sequence may be evaluated in a test tube by providing the target RNA, 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 RNA between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art.

Also provided are vectors, such as a viral vector or plasmid (*e.g.*, retrovirus, lentivirus, adenovirus, adeno-associated virus, or herpes simplex virus), that includes a guide nucleic acid molecule. Exemplary vectors are described herein. In some examples, the guide nucleic acid molecule is operably linked to a promoter or expression control element (examples of which are provided elsewhere in this application). As described elsewhere herein, such vectors can include other elements, such as a gene encoding a selectable marker, such as an antibiotic, such as puromycin, hygromycin, or a detectable marker such as GFP or other fluorophore.

In one example, a plurality of gRNAs are part of an array (which can be part of a vector, such as a viral vector or plasmid). For example, a guide array including the sequence DR-spacer-DR-spacer-DR-spacer, can include three unique unprocessed gRNAs (one for each DR-spacer sequence). Once introduced into a cell or cell-free system, the array is processed by the Cas13d protein into the three individual mature gRNAs. This allows for multiplexing, e.g. the delivery of multiple gRNAs to a cell or system to target multiple target RNAs or multiple positions within a single target RNA (or combinations thereof).

D. Vectors that encode Cas13d and Guide Nucleic Acid Molecules

The disclosure provides vectors, such as plasmids and viral vectors as described elsewhere herein, which include one or more guide molecule coding sequences (*e.g.*, to permit targeting of one or more RNA molecules), and one or more Cas13d protein coding sequences. Such vectors can be used in the methods, compositions, and kits provided herein. Such vectors can include naturally occurring or non-naturally occurring nucleotides or ribonucleotides. Such vectors can include a single promoter operably linked to the guide molecule (which can be part of an array that includes at least two different guide molecules) and the Cas13d protein coding sequence. Alternatively, the guide molecule (which can be part of an array that includes at least two different guide molecules) and the Cas13d protein coding sequence can be operably linked to different promoters. In some examples, , the guide molecule (which can be part of an array that includes at least two different guide molecules) and the Cas13d protein coding sequence are operably linked to a promoter, an enhancer, or both.

E. Recombinant Cells and Cell-Free Systems

Cells that include a non-native Cas13d protein, a non-native Cas13d protein coding sequence, a guide molecule (or coding sequence), or combinations thereof, are provided. Such recombinant cells can be used in the methods, compositions, and kits provided herein. Nucleic acid molecules encoding a Cas13d protein disclosed herein and/or nucleic acid molecules encoding a guide molecule can be introduced into cells to generate transformed (*e.g.*, recombinant) cells. In some examples, such cells are generated by introducing one or more non-native Cas13d proteins and one or more guide molecules (*e.g.*, gRNAs) into the cell, for example as a ribonucleoprotein (RNP) complex.

Similarly, cell free systems, such as those generated from lysed cells (or those that include a Cas13d RNP in a test tube or other vessel, into which *in vitro* transcribed or chemically synthesized target RNAs are added), which include a, Cas13d protein, a Cas13d protein coding sequence, a guide molecule (or coding sequence), or combinations thereof, are provided. Such cell free systems can be used in the methods, compositions, and kits provided herein. In some examples, one or more non-native Cas13d proteins and one or more guide molecules (*e.g.*, gRNAs) are added to a cell free system, for example as a RNP complex.

Thus, cells and cell-free systems containing a Cas13d protein (such as a protein having at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 138, 147, 149, 153, 155, 158, 160, 162, 164, 166, 168, 170, 175, 177, 179, 181, 183, 185, 187, 189, 194, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 278, 279, 280, 281, 282, 283,

284, 285, 292, 293, 294, 295, or 296) are disclosed. Similarly, cells and cell-free systems containing a guide molecule, such as one having at least one DR sequence having at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 129, 130, 131, 132, 133, 134, 135, 136, or 137, and in some examples also at least one spacer sequence complementary to a target RNA, are provided.

Such recombinant cells (e.g., which can be used to generate a cell-free system) can be eukaryotic or prokaryotic. Examples of such cells include, but are not limited to bacteria, archaea, plant, fungal, yeast, insect, and mammalian cells, such as *Lactobacillus*, *Lactococcus*, *Bacillus* (such as *B. subtilis*), *Escherichia* (such as *E. coli*), *Clostridium*, *Saccharomyces or Pichia* (such as *S. cerevisiae* or *P. pastoris*), *Kluyveromyces lactis*, *Salmonella typhimurium*, *Drosophila* cells, *C. elegans* cells, *Xenopus* cells, SF9 cells, C129 cells, 293 cells, *Neurospora*, and immortalized mammalian cell lines (*e.g.*, Hela cells, myeloid cell lines, and lymphoid cell lines).

In one example, the cell is a prokaryotic cell, such as a bacterial cell, such as *E. coli*.

In one example, the cell is a eukaryotic cell, such as a mammalian cell, such as a human cell. In one example, the cell is primary eukaryotic cell, a stem cell, a tumor/cancer cell, a circulating tumor cell (CTC), a blood cell (*e.g.*, T cell, B cell, NK cell, Tregs, etc.), hematopoietic stem cell, specialized immune cell (*e.g.*, tumor-infiltrating lymphocyte or tumor-suppressed lymphocytes), a stromal cell in the tumor microenvironment (*e.g.*, cancer-associated fibroblasts, etc.) In one example, the cell is a brain cell (*e.g.*, neurons, astrocytes, microglia, retinal ganglion cells, rods/cones, etc.) of the central or peripheral nervous system).

In one example, a cell is part of (or obtained from) a biological sample, such as a biological specimen containing genomic DNA, RNA (*e.g.*, mRNA), protein, or combinations thereof, obtained from a subject. Examples include, but are not limited to, peripheral blood, serum, plasma, urine, saliva, sputum, tissue biopsy, fine needle aspirate, surgical specimen, and autopsy material. Such cells can also be used to generate a cell free system.

In one example the cell (or cell free system) is from a tumor, such as a hematological tumor (*e.g.*, leukemias, including acute leukemias (such as acute lymphocytic leukemia, acute myelocytic leukemia, acute myelogenous leukemia and myeloblastic, promyelocytic, myelomonocytic, monocytic and erythroleukemia), chronic leukemias (such as chronic myelocytic (granulocytic) leukemia, chronic myelogenous leukemia, and chronic lymphocytic leukemia), polycythemia vera, lymphoma, Hodgkin's disease, non-Hodgkin's lymphoma (including low-, intermediate-, and high-grade), multiple myeloma, Waldenström's macroglobulinemia, heavy chain disease, myelodysplastic syndrome, mantle cell lymphoma and myelodysplasia) or solid tumor (*e.g.*, sarcomas and carcinomas: fibrosarcoma, myxosarcoma, liposarcoma, chondrosarcoma, osteogenic sarcoma, and other sarcomas, synovioma, mesothelioma, Ewing's tumor, leiomyosarcoma, rhabdomyosarcoma, colon carcinoma, lymphoid malignancy, pancreatic cancer, breast cancer, lung cancers, ovarian cancer, prostate cancer, hepatocellular carcinoma, squamous cell carcinoma, basal cell carcinoma, adenocarcinoma, sweat gland carcinoma, sebaceous gland carcinoma,

papillary carcinoma, papillary adenocarcinomas, medullary carcinoma, bronchogenic carcinoma, renal cell carcinoma, hepatoma, bile duct carcinoma, choriocarcinoma, Wilms' tumor, cervical cancer, testicular tumor, bladder carcinoma, and CNS tumors (such as a glioma, astrocytoma, medulloblastoma, craniopharyogioma, ependymoma, pinealoma, hemangioblastoma, acoustic neuroma, oligodendroglioma, menangioma, melanoma, neuroblastoma and retinoblastoma).

In one example the cell (or cell free system) is obtained from an environmental sample, such as a water, soil, or air sample.

F. Compositions & Kits

Compositions and kits that include a Cas13d protein, a Cas13d protein coding sequence, a guide molecule (or coding sequence), or combinations thereof, are provided. In one example, the composition or kit includes an RNP complex composed of one or more Cas13d proteins and one or more guide molecules (*e.g.*, gRNAs). In one example, the composition or kit includes a vector encoding a Cas13d protein, a guide molecule, or both. In one example, the composition or kit includes a cell, such as a bacterial cell or eukaryotic cell, that includes a non-native Cas13d protein, a non-native Cas13d protein coding sequence, a guide molecule (or coding sequence), or combinations thereof. In one example, the composition or kit includes a cell-free system that includes a non-native Cas13d protein, a non-native Cas13d protein coding sequence, a guide molecule (or coding sequence), or combinations thereof.

Such compositions can include a pharmaceutically acceptable carrier (*e.g.*, saline, water, glycerol, DMSO, or PBS). In some examples, the composition is a liquid, lyophilized powder, or cryopreserved.

In some examples, the kit includes a delivery system (*e.g.*, liposome, a particle, an exosome, a microvesicle, a viral vector, or a plasmid), and/or a label (*e.g.*, a peptide or antibody that can be conjugated either directly to a Cas13d RNP or to a particle containing the Cas13d RNP to direct cell type specific uptake/enhance endosomal escape/enable blood-brain barrier crossing etc.). In some examples, the kits further include cell culture or growth media, such as media appropriate for growing bacterial, plant, insect, or mammalian cells.

In some examples, such parts of a kit are in separate containers.

G. Targeting RNA

The disclosed Cas13d proteins (and coding sequences), and guide molecules (e.g., gRNA and coding sequences) can be used in a CRISPR/Cas system to target one or more RNA molecules, such as those present in a sample (such as a biological sample, environmental sample (*e.g.*, soil, air or water sample), and the like. In one example, the target RNA is a coding RNA. In one example, the target RNA is a nuclear RNA. In other examples, the target RNA is non-coding RNA (such as functional RNA, siRNA, microRNA, snRNA, snoRNA, piRNA, scaRNA, tRNA, rRNA, lncRNA, or lincRNA). Such RNA targeting methods can be performed *in vitro* (such as in cell culture or in a cell-free system), or *in vivo* (such as in an organism, embryo, or mammal).

The CRISPR/Cas system provided herein includes two general components: (1) a Cas13d protein or its coding sequence (whose expression can be driven by a promoter) and (2) a guide nucleic acid molecule, such as RNA (gRNA), which is specific for the target RNA (whose expression can also be driven by promoter. When introduced into cells (or to a cell free system) for example (1) as Cas13d mRNA and Cas13d gRNA, (2) as part of a single vector or plasmid or divided into multiple vectors or plasmids, (3) as separate Cas13d protein and guide molecules, or (4) as an RNP complex of the Cas13d protein and guide molecule, the guide molecule guides the Cas13d to the target RNA. If the Cas13d protein has a native HEPN domain(s) or is fused to an appropriate effector domain bearing RNase activity, the RNA can be cut. If the Cas13d protein has a mutated HEPN domain(s), a guide array can be processed into mature gRNAs, but the target RNA is not cut. Using this system, RNA sequences are easily targeted, for example edited or detected, optionally with an effector domain.

1. Introduction of Cas13d Protein Directly into a Cell

In one example, the Cas13d protein is expressed in a recombinant cell, such as *E. coli*, and purified. The resulting purified Cas13d protein, along with an appropriate guide molecule specific for the target RNA, is then introduced into a cell or organism where one or more RNAs can be targeted. In some examples, the Cas13d protein and guide nucleic acid molecule are introduced as separate components into the target cell/organism. In other examples, the purified Cas13d protein is complexed with the guide nucleic acid (*e.g.*, gRNA), and this ribonucleoprotein (RNP) complex is introduced into target cells (*e.g.*, using transfection or injection). In some examples, the Cas13d protein and guide molecule are injected into an embryo (such as a human, mouse, zebrafish, or *Xenopus* embryo).

Once the Cas13d protein and guide nucleic acid molecule are in the cell, one or more RNAs can be targeted.

2. Expression of Cas13d from Nucleic Acids

In one example, the Cas13d protein is expressed from a nucleic acid molecule in a cell containing a target RNA, for example an RNA to be detected or modified. In some such examples, the Cas13d protein is expressed from a vector, such as a viral vector or plasmid introduced into a cell or into a cell-free system. This results in the production of the Cas13d protein in the cell, organism, or system. In addition, these nucleic acid molecules can be co-expressed in the cell/organism/system with the guide nucleic acid molecule (*e.g.*, gRNA) specific for the target RNA.

In one example, multiple plasmids or vectors are used for RNA targeting. The nucleic acid molecule encoding the Cas13d can be provided for example on one vector or plasmid, and the guide nucleic acid molecule (*e.g.*, gRNA) on another plasmid or vector. Multiple plasmids or viral vectors can be mixed and introduced into cells (or a cell free system) at the same time, or separately.

In some examples, multiple nucleic acid molecules are expressed from a single vector or plasmid. For example, a single vector can include the nucleic acid molecule encoding the Cas13d, and a separate vector can include the guide molecule.

In some examples a plurality of different guide molecules (*e.g.*, gRNAs), one for each target (such as 1, 2, 3, 4, 5, or 10 different targets), are present on a single array and/or vector. In one example, the method includes delivering a plurality of gRNAs (such as at least 2, at least 3, at least 4, at least 5, at least 10, at least 20, or at least 50 different gRNAs), which are part of an array (which can be part of a vector, such as a viral vector or plasmid). Once introduced into a cell or cell-free system, the array is processed by the Cas13d protein into the individual mature gRNAs.

The nucleic acid molecules expressed from the vector can be under the control of a promoter and optionally contain selection markers (such as antibiotic resistance).

In some examples, the protein and guide molecules are expressed by an embryo (such as a zebrafish or *Xenopus* embryo). The Cas13d protein can be expressed from injected plasmid DNA, injected mRNA, or stably integrated copies into the animal genome. The gRNA can be directly injected or expressed from a vector or stably integrated copies into the animal genome.

3. Targets

One or more RNAs can be targeted by the disclosed methods, such as at least 1, at least 2, at least 3, at least 4 or at least 5 different RNAs in a cell, cell-free system, or organism, such as 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 different RNAs. In one example, the RNA is associated with a disease such as cystic fibrosis, Huntington's disease, Tay-Sachs, Fragile X syndrome, Fragile X-associated tremor/ataxia syndrome, Duchenne muscular dystrophy, myotonic dystrophy, spinal muscular atrophy, spinocerebellar ataxia, or familial ALS. In one example, the RNA is associated with cancer (*e.g.*, a cancer of the lung, breast, colon, liver, pancreas, prostate, bone, brain, skin (*e.g.*, melanoma), or kidney). Examples of target RNAs include, but are not limited to those associated with cancer (*e.g.*, BCR-ABL, Ras, Raf, p53, BRCA1, BRCA2, CXCR4, beta-catenin, HER2, and CDK4).

In one example, the RNA is associated with viral infection, such as infection by a positive-strand RNA viruses, such as Picornaviruses (such as Aphthoviridae [for example foot-and-mouth-disease virus (FMDV)]), Cardioviridae; Enteroviridae (such as Coxsackie viruses, Echoviruses, Enteroviruses, and Polioviruses); Rhinoviridae (Rhinoviruses)); Hepataviridae (Hepatitis A viruses); Togaviruses (examples of which include rubella; alphaviruses (such as Western equine encephalitis virus, Eastern equine encephalitis virus, and Venezuelan equine encephalitis virus)); Flaviviruses (examples of which include Dengue virus, West Nile virus, and Japanese encephalitis virus); Calciviridae (which includes Norovirus and Sapovirus); or Coronaviruses (examples of which include SARS coronaviruses, such as the Urbani strain), or a negative-strand RNA virus, such as Orthomyxyoviruses (such as the influenza virus), Rhabdoviruses (such as Rabies virus), and Paramyxoviruses (examples of which include measles virus, respiratory syncytial virus, or parainfluenza viruses), or a DNA viral infection (such as infection by Herpesviruses (such as Varicella-

zoster virus, for example the Oka strain; cytomegalovirus; and Herpes simplex virus (HSV) types 1 and 2), Adenoviruses (such as Adenovirus type 1 and Adenovirus type 41), Poxviruses (such as Vaccinia virus), or Parvoviruses (such as Parvovirus B19).

In one example, the RNA is associated with a bacterial infection or property of a bacterial infection, such as bacterial resistance, persistence, or antibiotic resistance. Detection of these RNAs can be used for diagnostic methods, while editing these RNAs in cell-based or cell-free systems can be used for therapeutic methods.

4. Methods of Detecting RNA

In one example, the method of targeting an RNA results in detecting, visualizing, or labeling a target RNA. For example, by using at least one Cas13d protein with a mutated HEPN domain (*e.g.*, SEQ ID NO: 2 or 4), at least one gRNA containing at least one spacer sequence specific for the target RNA, and an effector module, the target RNA will be recognized by Cas13d but will not be cut or nicked while the effector module becomes activated. In some examples, such a method is used to detect a target RNA. Such a method can be used in a cell or cell free system to determine if a target RNA is present, such as in a tumor cell. In some examples, the cell or cell free system is obtained from a tissue sample, blood sample or saliva sample.

In one example, the method of detecting RNA comprises of a Cas13d protein fused to a fluorescent protein or other detectable label along with a gRNA containing a spacer sequence specific for the target RNA. Binding of Cas13d to the target RNA can be visualized by microscopy or other methods of imaging. In another example, RNA aptamer sequences can be appended to or inserted within the gRNA molecule, such as MS2, PP7, Qβ, and other aptamers. The introduction of proteins that specifically bind to these aptamers, e.g. the MS2 phage coat protein, fused to a fluorescent protein or other detectable label can be used to detect the target RNA because the Cas13d-gRNA-target RNA complex will be labeled via the aptamer interaction.

In another example, the method of detecting RNA is a transcriptional sensor (*e.g.*, as part of a synthetic circuit) for diagnostics or therapeutics. For example, the transcriptional sensor can be comprised of at least one Cas13d protein with a mutated HEPN domain (*e.g.*, SEQ ID NO: 2 or 4), at least one gRNA containing at least one spacer sequence specific for the target RNA, and an effector module such as an optionally split fluorescent protein or probe; an optionally split luminescent protein or probe; an optionally split enzyme that catalyzes a detectable reaction such as ubiquitin or TEV protease; a FRET-compatible protein pair; one or more transcription factor(s) fused to Cas13d via cleavable linkers; a split intein that trans-splices a protein to restore its function such as a transcription factor; a kinase-substrate pair that activates upon phosphorylation; one, two, or more monomers that activate upon dimerization or multimerization; or one or more proteins that induce conformational and functional change upon interaction. In one example, the spatial proximity of two or more Cas13d proteins and gRNAs due to binding a particular transcript would activate the effector module, resulting in a detectable signal or detectable activity in the cell.

For example, the transcriptional sensor could allow a cancer-specific transcript, inflammation-specific transcript, disease-specific transcript, or cell state-specific transcript to be detected. A synthetic circuit containing a Cas13d-based system that is able to sense particular transcripts could encode conditional logic, e.g. requiring target detection to up- or down-regulate a gene for therapeutic application.

In one example, the method results in a detectable agent being bound to the target RNA, which can be detected. For example, two separate Cas13d fusion proteins that each include part of a fluorophore (*e.g.*, GFP), and two different gRNAs with different spacer sequences that target regions of an RNA in close proximity, can be used. When the two parts bind to the target RNA in proximity, the two parts of the fluorophore form a complete fluorophore, thereby generating a detectable signal.

In one example, the method results in RNA detection, for example by triggering a response such as expression of a second gene, modification of a protein, translocation of a protein or RNA to a different location, induction of cell death via suicide gene, induction of cell proliferation, induction of a transgene that enables a secondary function, induction of a permanent change in DNA sequence to enable storing a memory of past transcriptional events, or altering the RNA to enable pulldown.

In one embodiment, two halves of a transcription factor could be linked to two separate Cas13d via a split intein system. The Cas13d proteins are provided with two different gRNAs with different spacer sequences that target regions of an RNA in close proximity. Upon binding to the target RNA in proximity, the split inteins trans-splice a reconstituted transcription factor (TF) so that it can translocate to the nucleus and turn on a target gene or cluster of target genes. In one example, the target gene could be an endogenous gene in the cell. In another example, the target gene could be a transgene expressed on a vector or introduced through genetic engineering, such as a fluorescent protein or toxin.

5. Methods of Detecting RNA in Cell-Free Systems

In one example, the method of detecting a target RNA in a cell-free system results in a detectable label or enzyme activity. For example, by using at least one Cas13d protein (*e.g.*, SEQ ID NO: 3, 42, 62, 70, 82, 83, and 92), at least one gRNA containing at least one spacer sequence specific for the target RNA, and a detectable label, the target RNA will be recognized by Cas13d. The binding of the target RNA by Cas13d triggers its RNase activity, which can lead to the cleavage of the target RNA as well as the detectable label.

In one example, the detectable label is an RNA linked to a fluorescent probe and quencher. The intact detectable RNA links the fluorescent probe and quencher, suppressing fluorescence. Upon cleavage by Cas13d of the detectable RNA, the fluorescent probe is released from the quencher and displays fluorescent activity. Such a method can be used to determine if a target RNA is present in a lysed cell sample, lysed tissue sample, blood sample, saliva sample, environmental sample (such as a water, soil, or air sample), or other lysed cell or cell-free sample. Such a method can also be used to detect a pathogen, such as a virus or bacteria, or diagnose a disease state, such as a cancer.

In one example, the detection of the target RNA aids in the diagnosis of disease and/or pathological state, or the existence of a viral or bacterial infection. For example, Cas13d-mediated detection of non-

coding RNAs such as PCA3 can be used to diagnose prostate cancer if detected in patient urine. In another example, Cas13d-mediated detection of the lncRNA-AA174084, which is a biomarker of gastric cancer, can be used to diagnose gastric cancer.

6. Methods of Editing Target RNA

In one example, the method of targeting an RNA results in editing the sequence of a target RNA. For example, by using a Cas13d protein with a non-mutated HEPN domain (*e.g.*, SEQ ID NO: 1, 3, 42, 62, 70, 82, 83, or 92), and a gRNA containing at least one a spacer sequence specific for the target RNA, the target RNA can be cut or nicked at a precise location. In some examples, such a method is used to decrease expression of a target RNA, which will decrease translation of the corresponding protein. Such a method can be used in a cell where increased expression of an RNA is not desired. In one example, the RNA is associated with a disease such as cystic fibrosis, Huntington's disease, Tay-Sachs, Fragile X syndrome, Fragile X-associated tremor/ataxia syndrome, muscular dystrophy, myotonic dystrophy, spinal muscular atrophy, spinocerebellar ataxia, or familial ALS. In another example, the RNA is associated with cancer (*e.g.*, a cancer of the lung, breast, colon, liver, pancreas, prostate, bone, brain, skin (*e.g.*, melanoma), or kidney). Examples of target RNAs include, but are not limited to those associated with cancer (*e.g.*, PD-L1, BCR-ABL, Ras, Raf, p53, BRCA1, BRCA2, CXCR4, beta-catenin, HER2, and CDK4). Editing such target RNAs can have a therapeutic effect.

In another example, the RNA is expressed in an immune cell. The target RNA could, for example, code for a protein leading to the repression of a desirable immune response, such as infiltration of a tumor. Knock-down of such an RNA could enable progression of such a desirable immune response (*e.g.*, PD1, CTLA4, LAG3, TIM3). In another example, the target RNA encodes a protein resulting in the undesirable activation of an immune response, for example in the context of an autoimmune disease such as multiple sclerosis, Crohn's disease, lupus, or rheumatoid arthritis.

In one example, targeting the target RNA allows for decreasing expression of the target protein encoded by the RNA. For example, by using a Cas13d fusion protein with a mutated HEPN domain (*e.g.*, SEQ ID NO: 2 or 4) and a translational repression domain (such as Pumilio or FBF PUF proteins, deadenylases, CAF1, Argonaute proteins, and others), and a guide RNA containing at least one spacer sequence specific for the target RNA, expression of a target RNA can be decreased.

In some examples, Cas13d can be fused to a ribonuclease (such as a PIN endonuclease domain, an NYN domain, an SMR domain from SOT1, or an RNase domain from Staphylococcal nuclease) or a domain that affects RNA stability (such as tristetraprolin or domains from UPF1, EXOSC5, and STAU1).

In another example, RNA aptamer sequences can be appended to or inserted within the gRNA molecule, such as MS2, PP7, Qβ, and other aptamers. Proteins that specifically bind to these aptamers, *e.g.* the MS2 phage coat protein, can be fused to a translational repression domain, a ribonuclease, or a domain that affects RNA stability. This aptamer-effector domain fusion can be used to target the target RNA

because the Cas13d and gRNA complex will guide the aptamer protein-effector domain in proximity to the target RNA.

Such a method can be used in a cell where increased expression of an RNA is not desired, such as when an expressed RNA is associated with a disease such as cystic fibrosis, Huntington's disease, Tay-Sachs, Fragile X syndrome, Fragile X-associated tremor/ataxia syndrome, muscular dystrophy, myotonic dystrophy, spinal muscular atrophy, spinocerebellar ataxia, or familial ALS. In another example, the target RNA is associated with cancer (*e.g.*, a cancer of the lung, breast, colon, liver, pancreas, prostate, bone, brain, skin (*e.g.*, melanoma), or kidney). Examples of target RNAs include, but are not limited to those associated with cancer (*e.g.*, PD-L1, BCR-ABL, Ras, Raf, p53, BRCA1, BRCA2, CXCR4, beta-catenin, HER2, and CDK4). Editing such target RNAs would have a therapeutic effect.

In another example, the RNA is expressed in an immune cell. The target RNA could, for example, code for a protein leading to the repression of a desirable immune response, such as infiltration of a tumor. Knock-down of such an RNA could enable progression of such a desirable immune response (*e.g.*, PD1, CTLA4, LAG3, TIM3). In another example, the target RNA could encode a protein resulting in the undesirable activation of an immune response, for example in the context of an autoimmune disease such as multiple sclerosis, Crohn's disease, lupus, or rheumatoid arthritis.

In one example, targeting the target RNA allows for activating or increasing expression of the target RNA. For example, by using a Cas13d fusion protein with a mutated HEPN domain (e.g., SEQ ID NO: 2 or 4) and a translational activation domain (such as eIF4E and other translation initiation factors, a domain of the yeast poly(A)-binding protein or GLD2), and a guide RNA containing at least one a spacer sequence specific for the target RNA, expression of a target RNA can be increased. Aptamer introduction into the gRNA with a cognate aptamer-binding protein fused to a translational activation domain can also be used. In one example, RNA aptamer sequences are appended to or inserted within the gRNA molecule, such as MS2, PP7, Qβ, and other aptamers. The introduction of proteins that specifically bind to these aptamers, e.g. the MS2 phage coat protein, fused to a translational activation domain can be used to target the target RNA because the Cas13d and gRNA complex will bring the aptamer protein-translational activation domain in proximity to the target RNA.

In some examples, such a method is used to increase the activity or expression of a target RNA, which will increase translation of the corresponding protein (if the RNA is a coding RNA). Such a method can be used in a cell where increased expression of an RNA is desired, such as a heterozygous genetic disease or disorders caused by copy number variation. Increasing translation of a desired protein product could be therapeutic in nature.

In another example, increasing the expression of a target RNA (such as Cyclin B1) can render the target cell (such as cancers) more sensitive to drugs (such as chemotherapeutic agents).

In one example, targeting the target RNA allows for one or more RNA base substitutions, RNA base edits, RNA base deletions, RNA base insertions, or combinations thereof, in the target RNA. In some examples, the Cas13d protein with a mutated HEPN domain is associated, either via direct fusion or a

gRNA-aptamer modification, an effector domain that allows base edits (such as a cytidine deaminase, PPR protein, adenosine deaminase, ADAR family protein, or APOBEC family protein). In some examples, such a method is used to modify an RNA sequence, edit an RNA mutation, or modify an RNA transcript (*e.g.*, gene therapy), for example to treat diseases such as ALS and melanoma or genetic disorders caused by undesired splice sites, such as Leber congenital amaurosis.

In one example, targeting the target RNA allows for methylating the target RNA. Some examples may use a chimeric Cas13d protein with a mutated HEPN domain (*e.g.*, SEQ ID NO: 2 or 4) associated either via direct fusion or a gRNA-aptamer modification with a methylation domain (*e.g.*, m6A), and a guide RNA containing at least one a spacer sequence specific for the target RNA. In some examples, such a method is used to combat aberrant RNA demethylation. In one example, such a method is used modify the methylation levels of pluripotency transcripts such as NANOG or KLF4 for example to decrease their stability in breast cancer cells, which can suppress the acquisition of breast cancer stem cell phenotypes that are associated with increased proliferation and cancer stem cell formation.

In one example, targeting the target RNA allows for demethylating the target RNA. Some examples can use Cas13d protein with a mutated HEPN domain (*e.g.*, SEQ ID NO: 2 or 4), a guide RNA containing at least one a spacer sequence specific for the target RNA, and a demethylation domain (e.g., human alkylation repair homolog 5 or Alkbh5). The demethylation domain can be associated either via direct fusion to the Cas13d protein or via a gRNA-aptamer modification. In some examples, such a method is used to reverse aberrant RNA methylation, for example to treat myeloid leukemia by decreasing m^6A levels.

In one example, targeting the target RNA allows for binding to the target RNA. For example, by using a Cas13d protein with a mutated HEPN domain (*e.g.*, SEQ ID NO: 2 or 4) and a guide RNA containing at least one a spacer sequence specific for the target RNA, molecules can be bound or tethered to a target RNA. In some examples, such a method is used to capture the target RNA (*e.g.*, immuno-precipitation). This can be used as part of a kit to identify the proteins interacting with a specific RNA transcript. In one example, an epitope tagged Cas13d (*e.g.*, FLAG, HA, biotin, HALO tag) can be targeted to specific target RNAs and cross-linked via fixation (e.g. with paraformaldehyde or glutaraldehyde). Immunoprecipitation of Cas13d with an epitope-recognizing antibody allows for the identification of co-immunoprecipitated proteins via Western blot or mass spectrometry.

In another example, Cas13d can be fused to a biotin ligase (such as BirA) or a peroxidase (such as APEX2) in order to biotinylate proteins that interact with the target RNA. Labeled proteins can then be pulled down with streptavidin beads followed by mass spectrometry or Western blot.

In some examples, biotinylated Cas13d could be targeted to ribosomal RNA sequences with a gRNA. Streptavidin bead-mediated pulldown can be used to deplete rRNA for RNA sequencing library preparation.

In one example, targeting the target RNA allows for masking the target RNA. For example, by using a Cas13d protein with a mutated or intact HEPN domain and a guide RNA containing at least one a spacer

sequence specific for the target RNA, a target RNA can be masked from RNA-binding proteins or RNA-binding elements such as miRNAs.

In some examples, the Cas13d can be used to mask RNA binding sites from RNA-binding proteins (RBPs). In another example, Cas13d can mask miRNA binding sites. For example, the liver-specific miR-122 forms a complex with Hepatitis C viral RNA which protects it from degradation. A HEPN-active Cas13d protein could be targeted to the miRNA-122 binding site on the viral RNA to synergistically combat HCV infections by simultaneously reversing miRNA-122-mediated protection and directly degrading HCV RNA. In some examples, such a method is used to preserve or protect the target RNA molecule, for example to protect the target RNA from degradation. For example, by targeting AU-rich elements in the 3' UTR of a target gene, a HEPN-mutated Cas13d can block binding of RNA-binding proteins such as tristetraprolin (TTR) or AUF1, which lead to degradation of the target transcript.

In one example, targeting the target RNA allows for changing splicing of the target RNA. Both the direct binding of splice acceptor and/or donor sites as well as splice effector domains can be used to manipulate splicing. For example, by using a Cas13d protein with a mutated HEPN domain (*e.g.*, SEQ ID NO: 2 or 4), a guide RNA containing at least one a spacer sequence specific for the target RNA, and optionally an effector domain that affects splicing (such as the RS-rich domain of SRSF1, the Gly-rich domain of hnRNP A1, the alanine-rich motif of RBM4, or the proline-rich motif of DAZAP1), alternative splicing of the RNA can be achieved.

In some examples, such a method is used for exon inclusion, for example to include exon 2 of acid alpha-glucosidase (GAA) to treat Pompe disease or to include exon 7 of SMN2 to treat spinal muscular atrophy (SMA). In some examples, such a method is used for exon exclusion, for example to restore the reading frame of dystrophin to treat Duchenne muscular dystrophy or to shift the splicing of the Bcl-x pre-mRNA from the antiapoptotic long isoform to the proapoptotic short isoform to treat cancer.

In some examples, the method uses the Cas13d protein with a mutated HEPN domain to mask splice acceptor or donor sites, for example to create neoantigens to make cold tumors hot. By affecting the splicing of certain target pre-mRNAs, this method can generate novel exon-exon junctions that can lead to the creation of neo-epitopes in cancer cells. This can make a cancer cell vulnerable to the immune system due to the display of unnatural antigens. In other examples, this method can be used to dynamically manipulate isoform ratios or to restore reading frame of a protein (e.g., dystrophin for Duchenne's muscular dystrophy).

In one example, targeting the target RNA allows for controlling transcript trafficking of the target RNA. For example, by using a Cas13d fusion protein with a mutated HEPN domain (*e.g.*, SEQ ID NO: 2 or 4) and a subcellular localization signal or export sequence, a guide RNA containing at least one a spacer sequence specific for the target RNA. In some examples, such a method is used to traffic the target RNA molecule to a particular organelle or cytosolic compartment, or even export the target RNA transcript, for example to endosomes for extracellular release

In another example, the method can affect RNA trafficking. For example, the zipcode binding protein ZBP1 specifically recognizes an RNA sequence 5'-CGGAC(C/A-CA-C/U) that leads to localization

of certain transcripts to the leading edge of fibroblasts. By masking or manipulating particular RNA zipcodes or regulatory sequences from recognition by regulatory protein complexes, this method can affect RNA localization or trafficking within a cell.

In one example, the target RNA is a nuclearly localized RNA. For example, by using a Cas13d protein with a non-mutated HEPN domain (*e.g.*, SEQ ID NOS: 1, 3, 42, 62, 70, 82, 83, and 92) fused to a nuclear localization signal and a guide RNA containing at least one a spacer sequence specific for the target nuclear RNA, the nuclearly localized RNA can be targeted and degraded. In some examples, such a method is used to degrade the target nuclear RNA molecule, for example to knock-down a non-coding nuclear RNA such as HOTAIR, which is associated with metastatic progression in breast cancer.

In one example, the target RNA is viral RNA or transcript of a DNA virus. For example, a Cas13d protein with a non-mutated HEPN domain (*e.g.*, SEQ ID NOS: 1, 3, 42, 62, 70, 82, 83, and 92) and a guide RNA containing at least one spacer sequence specific for the target RNA can be used. In some examples, such a method is used to treat an RNA viral infection (such as infection by a positive-strand RNA viruses, such as Picornaviruses (such as Aphthoviridae [for example foot-and-mouth-disease virus (FMDV)]), Cardioviridae; Enteroviridae (such as Coxsackie viruses, Echoviruses, Enteroviruses, and Polioviruses); Rhinoviridae (Rhinoviruses)); Hepataviridae (Hepatitis A viruses); Togaviruses (examples of which include rubella; alphaviruses (such as Western equine encephalitis virus, Eastern equine encephalitis virus, and Venezuelan equine encephalitis virus)); Flaviviruses (examples of which include Dengue virus, West Nile virus, and Japanese encephalitis virus); Calciviridae (which includes Norovirus and Sapovirus); or Coronaviruses (examples of which include SARS coronaviruses, such as the Urbani strain), or a negative-strand RNA virus, such as Orthomyxyoviruses (such as the influenza virus), Rhabdoviruses (such as Rabies virus), and Paramyxoviruses (examples of which include measles virus, respiratory syncytial virus, or parainfluenza viruses), or a DNA viral infection (such as infection by Herpesviruses (such as Varicella-zoster virus, for example the Oka strain; cytomegalovirus; and Herpes simplex virus (HSV) types 1 and 2), Adenoviruses (such as Adenovirus type 1 and Adenovirus type 41), Poxviruses (such as Vaccinia virus), or Parvoviruses (such as Parvovirus B19), for example by cutting the viral RNA or transcript of a DNA virus. Thus, such methods can be used as an RNA-based antiviral or antimicrobial.

Example 1

Materials and Methods

This example describes the materials and methods used to obtain the results shown in Examples 2 - 7.

Cell culture of Human Embryonic Kidney (HEK) cell line 293FT

Human embryonic kidney (HEK) cell line 293FT (Thermo Fisher) was maintained in DMEM (4.5 g/L glucose), supplemented with 10% FBS (GE Life Sciences) and 10 mM HEPES at 37°C with 5% CO_2. Upon reaching 80-90% confluency, cells were dissociated using TrypLE Express (Life Technologies) and passaged at a ratio of 1:2.

Cell culture of human bone osteosarcoma epithelial cell line U2OS

Human bone osteosarcoma epithelial U2OS were maintained in DMEM (4.5 g/L glucose) supplemented with 10% FBS and 10 mM HEPES at 37°C with 5% CO_2. Cells were passaged at a 1:3 ratio upon reaching 70% confluence. This cell line was not authenticated.

Maintenance of induced pluripotent stem cells and neuronal differentiation

Stable human iPSC lines containing the FTDP-17 IVS10+16 mutation or age- and sex-matched control lines were obtained from the laboratory of Fen-Biao Gao (Biswas et al., 2016). Briefly, cells obtained from one male patient with the *MAPT* IVS10 + 16 mutation and two separate lines from one male control patient were reprogrammed into hiPSCs (Almeida et al., 2012). iPSCs were transduced with lentivirus containing a doxycycline-inducible *Ngn2* cassette. Lentiviral plasmids were a gift from S. Schafer and F. Gage. iPSCs were then passaged with Accutase and plated into a Matrigel-coated 6-well plate with mTESR media containing ROCK inhibitor Y-27632 (10 μM, Cayman) at 500,000 cells per well. On day 1, media was changed with mTESR. On day 2, media was changed to mTESR containing doxycycline (2 μg/ml, Sigma) to induce *Ngn2* expression. On day 3, culture media was replaced with Neural Induction media (NIM, DMEM/F12 (Life Technologies) containing BSA (0.1 mg/ml, Sigma), apo-transferrin (0.1 mg/ml, Sigma), putrescine (16 μg/ml, Sigma), progesterone (0.0625 μg/ml, Sigma), sodium selenite (0.0104 μg/ml, Sigma), insulin (5 μg/ml, Roche), BDNF (10 ng/ml, Peprotech), SB431542 (10 μM, Cayman), LDN-193189 (0.1 μM, Sigma), laminin (2 μg/ml, Life Technologies), doxycycline (2 μg/ml, Sigma) and puromycin (Life Technologies)). NIM media was changed daily. Following 3 days of puromycin selection, immature neuronal cells were passaged with Accumax (Innovative Cell Technologies) and plated onto 96-well plates coated with poly-D-lysine and Matrigel in Neural Maturation media (NMM; 1:1 Neurobasal/DMEM (Life Technologies) containing B27 (Life Technologies), BDNF (10 ng/ml, Peprotech), N-Acetyl cysteine (Sigma), laminin (2 μg/ml, Life Technologies), dbcAMP (49 μg/ml, Sigma) and doxycycline (2 μg/ml, Sigma). Media was replaced the next day (day 7) with NMM containing AraC (2 μg/ml, Sigma) to eliminate any remaining non-differentiated cells. On day 8, AraC was removed and astrocytes were plated on top of neurons to support neuron cultures in NMM containing hbEGF (5 ng/ml, Peprotech). Cells were transduced with AAV on day 10 and assayed on day 24.

Computational pipeline for Cas13d identification

We obtained whole genome, chromosome, and scaffold-level prokaryotic genome assemblies from NCBI Genome in June 2016 and compared CRISPRfinder, PILER-CR, and CRT for identifying CRISPR repeats. The 20 kilobase flanking regions around each putative CRISPR repeat was extracted to identify nearby proteins and predicted proteins using Python. Candidate Cas proteins were required to be >750 aa in length and within 5 proteins of the repeat array, and extracted CRISPR loci were filtered out if they contained Cas genes associated with known CRISPR systems such as types I-III CRISPR. Putative effectors

were clustered into families via all-by-all BLASTp analysis followed by single-linkage hierarchical clustering where a bit score of at least 60 was required for cluster assignment. Each cluster of at least 2 proteins was subjected to BLAST search against the NCBI non-redundant (nr) protein database, requiring a bit score >200 to assign similarity. The co-occurrence of homologous proteins in each expanded cluster to a CRISPR array was analyzed and required to be >70%. Protein families were sorted by average amino acid length and multiple sequence alignment for each cluster was performed using Clustal Omega and the Geneious aligner with a Blosum62 cost matrix. The *RxxxxH* HEPN motif was identified in the Cas13d family on the basis of this alignment. TBLASTN was performed on all predicted Cas13d effectors against public metagenome whole genome shotgun sequences without predicted open reading frames (ORFs). The Cas13d family was regularly updated via monthly BLAST search on genome and metagenome databases to identify any newly deposited sequences. New full-length homologs and homologous fragments were aligned using Clustal Omega and clustered using PhyML 3.2. CRISPRDetect was used to predict the direction of direct repeats in the Cas13d array and DR fold predictions were performed using the Andronescu 2007 RNA energy model at 37°C. Sequence logos for Cas13d direct repeats were generated using Geneious 10.

Protein expression and purification

Recombinant Cas13d proteins were PCR amplified from genomic DNA extractions of cultured isolates or metagenomic samples and cloned into a pET-based vector with an N-terminal His-MBP fusion and TEV protease cleavage site. The resulting plasmids were transformed into Rosetta2(DE3) cells (Novagen), induced with 200 μM IPTG at OD_{600} 0.5, and grown for 20 hours at 18°C. Cells were then pelleted, freeze-thawed, and resuspended in Lysis Buffer (50 mM HEPES, 500 mM NaCl, 2 mM MgCl2, 20 mM Imidazole, 1% v/v Triton X-100, 1 mM DTT) supplemented with 1X protease inhibitor tablets, 1 mg/mL lysozyme, 2.5U/mL Turbo DNase (Life Technologies), and 2.5U/mL salt active nuclease (Sigma Aldrich). Lysed samples were then sonicated and clarified via centrifugation (18,000 x g for 1 hour at 4°C), filtered with 0.45 μM PVDF filter and incubated with 50 mL of Ni-NTA Superflow resin (Qiagen) per 10 L of original bacterial culture for 1 hour. The bead-lysate mixture was applied to a chromatography column, washed with 5 column volumes of Lysis Buffer, and 3 column volumes of Elution Buffer (50 mM HEPES, 500 mM NaCl, 300 mM Imidazole, 0.01% v/v Triton X-100, 10% glycerol, 1 mM DTT). The samples were then dialyzed overnight into TEV Cleavage Buffer (50 mM Tris-HCl, 250 mM KCl, 7.5% v/v glycerol, 0.2 mM TCEP, 0.8 mM DTT, TEV protease) before cation exchange (HiTrap SP, GE Life Sciences) and gel filtration (Superdex 200 16/600, GE Life Sciences). Purified, eluted protein fractions were pooled and frozen at 4 mg/mL in Protein Storage Buffer (50 mM Tris-HCl, 1M NaCl, 10% glycerol, 2 mM DTT).

Preparation of guide and target RNAs

Oligonucleotides carrying the T7 promoter and appropriate downstream sequence were synthesized (IDT) and annealed with an antisense T7 oligo for crRNAs and PCR-amplified for target and array templates. Homopolymer target RNAs were synthesized by Synthego. The oligo anneal and PCR templates

were *in vitro* transcribed with the Hiscribe T7 High Yield RNA Synthesis kit (New England Biolabs) at 31°C for 12 hours. For labeled targets, fluorescently labelled aminoallyl-UTP atto 680 (Jena Biosciences) was additionally added at 2 mM. Guide RNAs were purified with RNA-grade Agencourt AMPure XP beads (Beckman Coulter) and arrays and targets were purified with MEGAclear Transcription Clean-Up Kit (Thermo Fisher) and frozen at -80°C. For ssDNA and dsDNA targets, corresponding oligonucleotide sequences were synthesized (IDT) and either gel purified, or PCR amplified and then subsequently gel purified respectively.

Biochemical cleavage reactions

Purified EsCas13d protein and guide RNA were mixed (unless otherwise indicated) at 2:1 molar ratio in RNA Cleavage Buffer (25mM Tris pH 7.5, 15mM Tris pH 7.0, 1mM DTT, 6mM MgCl2). The reaction was prepared on ice and incubated at 37°C for 15 minutes prior to the addition of target at 1:2 molar ratio relative to EsCas13d. The reaction was subsequently incubated at 37°C for 45 minutes and quenched with 1 µL of enzyme stop solution (10 mg/mL Proteinase K, 4M Urea, 80mM EDTA, 20mM Tris pH 8.0) at 37°C for 15 minutes. The reaction was then denatured with 2X RNA loading buffer (2X: 13mM Ficoll, 8M Urea, 25 mM EDTA), at 85°C for 10 minutes, and separated on a 10% TBE-Urea gel (Life Technologies). Gels containing labeled targets were visualized on the Odyssey Imager (Li-Cor); unlabeled array or target cleavage gels were stained with SYBR Gold prior to imaging via Gel Doc EZ system (Bio-Rad).

Transient transfection of human cell lines

Engineered Cas13 coding sequences were cloned into a standardized plasmid expression backbone containing an EF1a promoter and prepared using the Nucleobond Xtra Midi EF Kit (Machery Nagel) according to the manufacturer's protocol. NLS-LwaCas13a-msfGFP and PspCas13b-NES-HIV were PCR amplified from Addgene #103854, and #103862, respectively, a gift from Feng Zhang. Cas13d pre-gRNAs and gRNAs were cloned into a minimal backbone containing a U6 promoter. shRNAs and guides for LwaCas13a were cloned into the same backbone and position matched to their corresponding guide RNA at the 3' of the target sequence. Matched gRNAs for PspCas13b were moved to the closest 5'-G nucleotide.

For transient transfection, HEK 293FT cells were plated at a density of 20,000 cells per well in a 96-well plate and transfected at >90% confluence with 200 ng of Cas13 expression plasmid and 200 ng of gRNA expression plasmid using Lipofectamine 2000 (Life Technologies) according to the manufacturer's protocol. Transfected cells were harvested 48-72 hours post-transfection for flow cytometry, gene expression analysis, or other downstream processing.

For reporter assays, HEK 293FT cells were transfected in 96-well format with 192ng of Cas13d expression plasmid, 192ng of guide expression plasmid, and 12ng of mCherry expression plasmid with Lipofectamine 2000 (Life Technologies). Cells were harvested after 48 hours and analyzed by flow cytometry.

U2OS cells were plated at a density of 20,000 cells per well in a 96-well plate and transfected at >90% confluence with 100 ng of Cas13d expression plasmid using Lipofectamine 3000 (Life Technologies) according to the manufacturer's protocol and processed for immunocytochemistry after 48h.

Flow cytometry

Cells were dissociated 48 hours post-transfection with TrypLE Express and resuspended in FACS Buffer (1X DPBS$^{-/-}$, 0.2% BSA, 2 mM EDTA). Flow cytometry was performed in 96-well plate format using a MACSQuant VYB (Miltenyi Biotec) and analyzed using FlowJo 10. RG6 was a gift from Thomas Cooper (Addgene plasmid # 80167) and modified to replace EGFP with mTagBFP2. All represented samples were assayed with three biological replicates. In the mCherry reporter assay, data is representative of at least 20,000 gated events per condition. In the splicing reporter assay, data is representative of at least 2,500 gated events per condition.

Gene expression analysis

Cells were lysed 48 hours post-transfection with DTT-supplemented RLT buffer and total RNA was extracted using RNeasy Mini Plus columns (Qiagen). 200 ng of total RNA was then reverse transcribed using random hexamer primers and Revertaid Reverse Transcriptase (Thermo Fisher) at 25°C for 10 min, 37°C for 60 min, and 95°C for 5 min followed by qPCR using 2X Taqman Fast Advanced Master Mix (Life Technologies) and Taqman probes for GAPDH and the target gene as appropriate (Life Technologies and IDT). Taqman probe and primer sets were generally selected to amplify cDNA across the Cas13 or shRNA target site position to prevent detection of cleaved transcript fragments (see Table S4 of Konermann *et al.*, Cell 173:1-12, 2018, herein incorporated by reference in its entirety). qPCR was carried out in 5 μL multiplexed reactions and 384-well format using the LightCycler 480 Instrument II (Roche). Fold-change was calculated relative to GFP-transfected vehicle controls using the ddCt method. One-way or two-way ANOVA with multiple comparison correction was used to assess statistical significance of transcript changes using Prism 7.

Immunohistochemistry

For immunohistochemical analysis, U2OS cells were cultured on 96-well optically clear plates (Greiner Bio-One), transfected as previously described, then fixed in 4% PFA (Electron Microscopy Sciences) diluted in PBS (Gibco) and washed with 0.3M glycine (Sigma) in PBS to quench PFA. Samples were blocked and permeabilized in a PBS solution containing 8% donkey serum (Jackson ImmunoResearch), 8% goat serum (Cell Signaling Technologies), and 0.3% Triton-X 100 (Sigma) for one hour, followed by primary antibody incubation in 1% BSA (Fisher Bioreagents), 1% goat serum, and 0.25% Triton-X overnight at 4°C. Samples were washed 3 times with PBS containing 0.1% BSA and 0.1% Triton-X 100 before incubating with fluorophore-conjugated secondary antibodies in PBS with 0.05% Triton-X 100 and 1% BSA at room temperature for one hour. Cells were washed with PBS with 0.1% Triton-X, stained

with DAPI, and then covered with Mounting Media (Ibidi) before imaging. Primary antibody, HA-Tag 6E2 (Cell Signaling, 2367), was used at a 1:100 dilution as per manufacturer's instructions. Secondary antibodies used were goat anti-mouse IgG1-Alexa-Fluor 647 (Thermo Fisher, A21240) and Anti-Mouse IgG1 CF 633 (Sigma, SAB4600335). Confocal images were taken using a Zeiss Airyscan LSM 880 followed by image processing in Zen 2.3 (Zeiss).

Bacterial small RNA sequencing and analysis

E. coli DH5α cells were transformed with pACYC184 carrying the CRISPR-Cas13d locus derived from an uncultured *Ruminococcus sp.* strain. Cells were harvested in stationary phase, rinsed in PBS, resuspended in TRIzol (Life Technologies), transferred to Lysing Matrix B tubes containing 0.1 mm silica beads (MP Biomedicals), and homogenized on a Bead Mill 24 (Fisher Scientific) for three 30-second cycles. Total RNA was isolated by phenol-chloroform extraction, then purified using the DirectZol Miniprep Kit (Zymo Research). RNA quality was assessed on an Agilent 2200 Tapestation followed by Turbo DNase treatment (Ambion). Total RNA was treated with T4 Polynucleotide Kinase (NEB) and rRNA-depleted using the Ribo-Zero rRNA Removal Kit for bacteria (Illumina). RNA was treated with RNA 5' polyphosphatase, poly(A)-tailed with E. coli poly(A) polymerase, and ligated with 5' RNA sequencing adapters using T4 RNA ligase 1 (NEB). cDNA was generated via reverse transcription using an oligo-dT primer and M-MLV RT/RNase Block (AffinityScript, Agilent) followed by PCR amplification and barcoding. Resulting libraries were sequenced on Illumina MiSeq, demultiplexed using custom Python scripts, and aligned to the Cas13d CRISPR locus using Bowtie 2. Alignments were visualized with Geneious.

***Ngn2* lentivirus preparation**

Low passage HEK 293FT cells were transfected with Polyethylenimine Max (PEI, Polysciences) and *Ngn2* target plasmid plus pMDG.2 and psPAX2 packaging plasmids (a gift from Didier Trono, Addgene #12259 and #12260) in DMEM + 10% FBS media during plating. The following day, media was changed to serum-free chemically defined minimal medium (Ultraculture supplemented with Glutamax, Lonza). Viral supernatant was harvested 48h later, clarified through a 0.45 micron PVDF filter (Millipore) and concentrated using ultracentrifugation.

AAV preparation

Low passage HEK 293FT cells were transfected with Polyethylenimine Max (PEI, Polysciences) and AAV target plasmid plus AAV1 serotype and pAdDeltaF6 helper packaging plasmids (UPenn Vector Core) in DMEM + 10% FBS media during plating. The following day, 60% of the media was changed to chemically defined minimal medium (Ultraculture supplemented with Glutamax, Lonza). 48h later, AAV-containing supernatant was harvested and clarified through a 0.45μm PVDF filter (Millipore) and

concentrated using precipitation by polyethylene glycol (PEG virus precipitation kit #K904, Biovision) following the manufacturer's protocol.

RNA-seq library preparation and sequencing

48h after transfection, total RNA was extracted from 293FT cells using the RNeasy Plus Mini kit from Qiagen. Stranded mRNA libraries were prepared using the NEBNext II Ultra Directional RNA Library Prep Kit from New England Biolabs (Cat# E7760S) and sequenced on an Illumina NextSeq500 with 42 nt paired end reads. ~15M total reads were demultiplexed per condition.

RNA-seq analysis

Sequenced reads were quality-tested using FASTQC and aligned to the hg19 human genome using the 2.5.1b STAR aligner (Dobin et al., 2013). Mapping was carried out using default parameters (up to 10 mismatches per read, and up to 9 multi-mapping locations per read). The genome index was constructed using the gene annotation supplied with the hg19 Illumina iGenomes collection (Illumina) and sjdbOverhang value of 100. Uniquely mapped reads were quantified across all gene exons using the top-expressed isoform as proxy for gene expression with the HOMER analysis suite (Heinz et al., 2010), and differential gene expression was carried out with DESeq2 v 1.14.1 (Love et al., 2014) using triplicates to compute within-group dispersion and contrasts to compare between targeting and non-targeting conditions. Significant differentially expressed genes were defined as having a false discovery rate (FDR) <0.01 and a log2 fold change >0.75. Volcano plots were generated in R 3.3.2 using included plotting libraries and the alpha() color function from the scales 0.5.0 package.

Statistics

All values are reported as mean ± SD or mean ± SEM as indicated in the appropriate figure legends. For comparing two groups, a one-tailed student's t-test was used and statistical significance was determined using the Holm-Sidak method with alpha = 0.05. A one-way ANOVA with Tukey multiple hypothesis correction was used to assess significance between more than two groups. Two-way ANOVA was used when comparing across two factors (i.e., RNA targeting modality and guide position) and adjusted for multiple hypothesis correction by Sidak's multiple comparisons test. For comparing groups that were found to not meet the assumption of a normal distribution by a D'Agostino and Pearson normality test, the non-parametric Friedman test with Dunn's multiple comparison adjustment was performed. PRISM 7.0 was used for all statistical analysis. Sample sizes were not determined *a priori.* At least three biological replicates were used for each experiment, as indicated specifically in each figure.

Sequencing data reported herein can be found in the NCBI Gene Expression Omnibus under GEO Series accession number GSE108519.

Additional details on the materials and methods used, such as sequences can be found in Konermann *et al.*, Cell 173:1-12, 2018, herein incorporated by reference in its entirety.

Example 2

Computational identification of a Type VI-like Cas ribonuclease family

This example describes methods used to identify previously undetected or uncharacterized RNA-targeting CRISPR-Cas systems by developing a computational pipeline for class 2 CRISPR-Cas loci, which require only a single nuclease for CRISPR interference such as Cas9, Cas12a (formerly Cpf1), or Cas13a (formerly C2c2) (Makarova et al., 2015; Shmakov et al., 2015). To improve upon previous strategies for bioinformatic mining of CRISPR systems, which focus on discovering sets of conserved Cas genes involved in spacer acquisition (Shmakov et al., 2015), the minimal requirements for a CRISPR locus to be the presence of a CRISPR repeat array and a nearby effector nuclease were defined. Using the CRISPR array as a search anchor, all prokaryotic genome assemblies and scaffolds were obtained from the NCBI WGS database and adapted algorithms for *de novo* CRISPR array detection (Bland et al., 2007; Edgar, 2007; Grissa et al., 2007) to identify 21,175 putative CRISPR repeat arrays **(FIG. 1A)**.

Up to 20 kilobases (kb) of genomic DNA sequence flanking each CRISPR array was extracted to identify predicted protein-coding genes in the immediate vicinity. Candidate loci containing signature genes of known class 1 and class 2 CRISPR-Cas systems such as Cas3 or Cas9 were excluded from further analysis, except for Cas12a and Cas13a to judge the ability of the pipeline to detect and cluster these known class 2 effector families. To identify new class 2 Cas effectors, it was required that candidate proteins be >750 residues in length and within 5 protein-coding genes of the repeat array, as large proteins closely associated with CRISPR repeats are key characteristics of known single effectors. The resulting proteins were classified into 408 putative protein families using single-linkage hierarchical clustering based on homology.

To discard protein clusters that reside in close proximity to CRISPR arrays due to chance or overall abundance in the genome, additional homologous proteins to each cluster were identified from the NCBI non-redundant protein database and their proximity to a CRISPR array determined. Reasoning that true Cas genes would have a high co-occurrence rate with CRISPR repeats, >70% of the proteins for each expanded cluster were required to exist within 20 kb of a CRISPR repeat. These remaining protein families were analyzed for nuclease domains and motifs.

Among the candidates, which include the recently described Cas13b system (Smargon et al., 2017), a family of uncharacterized putative class 2 CRISPR-Cas systems encoding a candidate CRISPR-associated ribonuclease containing 2 predicted HEPN ribonuclease motifs (Anantharaman et al., 2013) were identified **(FIG. 2A)**. Importantly, they are among the smallest class 2 CRISPR effectors described to date (~930 aa). The Type VI CRISPR-Cas13 superfamily is exemplified by sequence-divergent, single-effector signature nucleases and the presence of two HEPN domains. Other than these two *RxxxxH* HEPN motifs **(FIG. 3A),** the candidate effectors have no significant sequence similarity to previously described Cas13 enzymes, so his family of putative CRISPR ribonucleases is designated as Type VI Cas13d, or Type VI-D **(FIG. 3B)**.

CRISPR-Cas13d systems are derived from gut-resident microbes, so we sought to expand the Cas13d family via alignment to metagenomic contigs from recent large-scale microbiome sequencing efforts. Comparison of Cas13d proteins against public metagenome sequences without predicted open reading frames (ORFs) identified additional full-length systems as well as multiple effector and array fragments that cluster in several distinct branches **(FIG. 1B)**. To generate full-length Cas13d ortholog proteins and loci from the different branches of the Cas13d protein family, genomic DNA samples we43 obtained from associated assemblies and performed targeted Sanger sequencing to fill in gaps due to incomplete sequencing coverage, such as for the metagenomic ortholog 'Anaerobic digester metagenome' (Adm) (Treu et al., 2016).

Cas13d CRISPR loci are largely clustered within benign, Gram-positive gut bacteria of the genus *Ruminococcus*, and exhibit a surprising diversity of CRISPR locus architectures **(FIG. 2A)**. With the exception of the metagenomic AdmCas13d system, Cas13d systems lack the key spacer acquisition protein Cas1 (Yosef et al., 2012) within their CRISPR locus, highlighting the utility of a class 2 CRISPR discovery pipeline without Cas1 or Cas2 gene requirements. Cas13d direct repeats (DRs) are highly conserved in length and predicted secondary structure **(FIG. 3C)**, with a 36 nt length, an 8-10 nt stem with A/U-rich loop, and a 5′-AAAAC motif at the 3′ end of the direct repeat **(FIG. 3D)**. This conserved 5′-AAAAC motif has been shown to be specifically recognized by a type II Cas1/2 spacer acquisition complex (Wright and Doudna, 2016). In fact, Cas1 can be found in relative proximity to some Cas13d systems (within 10-30 kb for P1E0 and Rfx) while the remaining Cas13d-containing bacteria contain Cas1 elsewhere in their genomes, likely as part of another CRISPR locus.

Example 3

CRISPR-Cas13d possesses dual RNase activities

To demonstrate that the Cas13d repeat array is transcribed and processed into CRISPR guide RNAs (gRNA), the Cas13d CRISPR locus was cloned from an uncultured *Ruminococcus sp.* sample (Ur) into a bacterial expression plasmid. CRISPR systems tend to form self-contained operons with the necessary regulatory sequences for independent expression, facilitating heterologous expression in *E. coli* (Gasiunas et al., 2012). RNA sequencing (Heidrich et al., 2015) revealed processing of the array into ~52nt mature gRNAs, with a 30 nt 5′ direct repeat followed by a variable 3′ spacer that ranged from 14-26 nt in length **(FIG. 2B).**

To characterize Cas13d properties *in vitro*, *Eubacterium siraeum* Cas13d protein (EsCas13d) was purified based on its robust recombinant expression in *E. coli* **(FIGS. 4A-4C)** and found that EsCas13d was solely sufficient to process its matching CRISPR array into constituent guides without additional helper ribonucleases **(FIG. 2C, Table S1** of Konermann *et al.*, "Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors," Cell 173:1-12, 2018, herein incorporated by reference in its entirety), a property shared by some class 2 CRISPR-Cas systems (East-Seletsky et al., 2016; Fonfara et al., 2016; Smargon et al., 2017). Furthermore, inactivating the positively charged catalytic residues of the HEPN

motifs (Anantharaman et al., 2013) (dCas13d: R295A, H300A, R849A, H854A) did not affect array processing, indicating a distinct RNase activity dictating gRNA biogenesis analogous to Cas13a (East-Seletsky et al., 2016; Liu et al., 2017).

Cas effector proteins typically form a binary complex with mature gRNA to generate an RNA-guided surveillance ribonucleoprotein capable of cleaving foreign nucleic acids for immune defense (van der Oost et al., 2014). To assess if Cas13d has programmable RNA targeting activity as indicated by the presence of two HEPN motifs, EsCas13d protein was paired with an array or a mature gRNA along with a cognate *in vitro*-transcribed target. Based on the RNA sequencing results, a mature gRNA containing a 30 nt direct repeat and an intermediate spacer length of 22 nt was selected (nucleotides 6-36 of SEQ ID NO: 129, followed by 22 bases complementary to the RNA target).

Cas13d was able to efficiently cleave the complementary target ssRNA with both the unprocessed array and mature gRNA in a guide-sequence dependent manner, while non-matching spacer sequences abolished Cas13d activity **(FIG. 5A).** Substitution with dCas13d or the addition of EDTA to the cleavage reaction also abolished guide-dependent RNA targeting, indicating that Cas13d targeting is HEPN- and Mg^{2+}-dependent **(FIG. 5B)**. To determine the minimal spacer length for efficient Cas13d targeting, a series of spacer truncations ranging from the unprocessed 30 nt length down to 10 nt were generated **(FIG. 6A)**. Cleavage activity dropped significantly below a 21 nt spacer length, confirming the choice of a 22 nt spacer **(FIG. 6B)**.

RNA-targeting class 2 CRISPR systems have been proposed to act as sensors of foreign RNAs (Abudayyeh et al., 2016; East-Seletsky et al., 2016), where general RNase activity of the effector nuclease is triggered by a guide-matching target. To assay for a similar property in Cas13d, RNase activity of the binary EsCas13d:gRNA complex was monitored in the presence of a matching RNA target. It was observed that EsCas13d can be activated by target RNA to cleave bystander RNA targets **(FIG. 3C)**, albeit inefficiently relative to its activity on the complementary ssRNA target. Bystander cleavage is guide sequence- and HEPN-dependent, as the presence of non-matching bystander target alone was insufficient to induce cleavage while substitution of dCas13d or addition of EDTA abolished activity. These results indicate that bystander RNase activity may be a general property of RNA-targeting class 2 systems in CRISPR adaptive bacterial immunity **(FIG. 3D).**

To assess the generalizability of Cas13d reprogramming, twelve guides tiling a complementary RNA target were generated and efficient cleavage in all cases was observed **(FIG. 7A)**. Cas13d was unable to cleave a ssDNA **(FIG. 6C)** or dsDNA **(FIG. 6D)** version of the ssRNA target, indicating that Cas13d is an RNA-specific nuclease. Further, RNA target cleavage did not depend on the protospacer flanking sequence (PFS) **(FIG. 7A)** in contrast to other RNA-targeting class 2 systems, which require a 3′-H (Abudayyeh et al., 2016) or a double-sided, DR-proximal 5′-D and 3′-NAN or NNA (Smargon et al., 2017). Although a slight bias against an adenine PFS was initially observed **(FIG. 6E)**, varying the target PFS base with a constant guide sequence resulted in no significant differences (P=0.768) in targeting efficiency **(FIG. 6F)**.

While DNA-targeting class 2 CRISPR systems (Gasiunas et al., 2012; Jinek et al., 2012; Zetsche et al., 2015) and some RNA-targeting class 1 systems tend to cleave at defined positions relative to the target-guide duplex (Samai et al., 2015; Zhang et al., 2016), the Cas13d cleavage pattern varies for different targets **(FIGS. 5A, 5C, 6H)** and remains remarkably similar despite the guide sequence position **(FIG. 7A)**. This indicates that Cas13d may preferentially cleave specific sequences or structurally accessible regions in the target RNA. Cas13d was tested activity on targets containing variable homopolymer repeats in the loop region of a hairpin or as a linear single-stranded repeat. EsCas13d exhibited significant preference for uracil bases in both target structures, with lower but detectable activity at all other bases **(FIG. 7B)**.

Cas enzymes are found in nearly all archaea and about half of bacteria (Hsu et al., 2014; van der Oost et al., 2014), spanning a wide range of environmental temperatures. To determine the optimal temperature range for Cas13d activity, a spectrum of cleavage temperature conditions from 16-62°C was tested and observed maximal activity in the 24-41°C range **(FIG. 6G, 6H)**. This temperature range is compatible with a wide range of prokaryotic and eukaryotic hosts, indicating Cas13d can be adapted for RNA targeting in different cells and organisms.

Example 4

Cell-based activity screen of engineered orthologs

The Cas13d nuclease was used as a flexible tool for programmable RNA targeting in mammalian cells. CRISPR orthologs from distinct bacterial species commonly exhibit variable activity (Abudayyeh et al., 2017; East-Seletsky et al., 2017), especially upon heterologous expression in human cells (Ran et al., 2015; Zetsche et al., 2017). Highly active Cas13d orthologs were identified in a eukaryotic cell-based mCherry reporter screen.

By synthesizing human codon-optimized versions of 7 orthologs from distinct branches within the Cas13d family **(FIG. 1B)**, mammalian expression plasmids carrying the catalytically active and HEPN-inactive proteins were generated . Each protein was then optionally fused to N- and C-terminal nuclear localization signals (NLS). These Cas13d effector designs were HA-tagged and paired with two distinct guide RNA architectures, either with a 30 nt spacer flanked by two direct repeat sequences to mimic an unprocessed guide RNA (pre-gRNA) or a 30 nt direct repeat with 22 nt spacer (gRNA) predicted to mimic mature guide RNAs **(FIG. 8A)**. For each guide design, four distinct spacer sequences complementary to the mCherry transcript were then pooled to minimize potential spacer-dependent variability in targeting efficiency. The ability of Cas13d to knockdown mCherry protein levels was determined in a human embryonic kidney (HEK) 293FT cell-based reporter assay.

48 hours post-transfection, flow cytometry indicated that RfxCas13d and AdmCas13d efficiently knocked down mCherry protein levels by up to 92% and 87% ($P<0.0003$), respectively, relative to a non-targeting control guide **(FIG. 8B).** In contrast, EsCas13d along with RaCas13d and RffCas13d exhibited limited activity in human cells. Furthermore, none of the HEPN-inactive Rfx-dCas13d constructs significantly affected mCherry fluorescence, indicating HEPN-dependent knockdown ($P>0.43$ for all cases).

Robust nuclear translocation of the Rfx and AdmCas13d NLS fusion constructs was observed via immunocytochemistry, while the wild-type effectors remain primarily extra-nuclear **(FIG. 8C)**.

Proceeding with RfxCas13d and AdmCas13d as lead candidates, we next compared their ability to knockdown endogenous transcripts. To determine the optimal ortholog and guide architecture, the capability of Rfx and AdmCas13d construct variants to target β-1,4-N-acetyl-galactosaminyl transferase 1 (*B4GALNT1*) transcripts was systematically assayed. In each condition, four guides containing distinct spacer sequences tiling the *B4GALNT1* transcript were pooled. The RfxCas13d-NLS fusion targeted *B4GALNT1* more efficiently than wild-type RfxCas13d and both variants of AdmCas13d, with both the gRNA and pre-gRNA mediating potent knockdown (~82%, $P<0.0001$) **(FIG. 8D)**. Cas13d-NLS from *Ruminococcus flavefaciens* strain XPD3002 was therefore selected for the remaining experiments (CasRx).

Example 5

Programmable RNA knockdown in human cells with CasRx

Because Cas13d is capable of processing its own CRISPR array, this property was leveraged for the simultaneous delivery of multiple targeting guides in a simple single-vector system **(FIG. 9A)**. Arrays encoding four spacers that each tile the transcripts of mRNAs (*B4GALNT1 and ANXA4)* or nuclear localized lncRNAs *(HOTTIP and MALAT1)* consistently facilitated robust (>90%) RNA knockdown by CasRx ($P<0.0001$) **(FIG. 9B)**.

CasRx was compared to more established technologies for transcript knockdown or repression, by comparing CasRx-mediated RNA interference to dCas9-mediated CRISPR interference (Gilbert et al., 2014; Gilbert et al., 2013) and spacer sequence-matched shRNAs via transient transfection **(FIG. 9C)**. For CRISPRi-based repression, the most potent dCas9 guide for *B4GALNT1* from previous reports was analyzed (Gilbert et al., 2014; Zalatan et al., 2015). Across 3 endogenous transcripts, CasRx outperformed shRNAs (11/11) and CRISPRi (4/4) in each case **(FIG. 9D)**, exhibiting a median knockdown of 96% compared to 65% for shRNA and 53% for CRISPRi after 48 hours. In addition, knockdown by CasRx was compared to two recently described Cas13a and Cas13b effectors (Abudayyeh et al., 2017; Cox et al., 2017) **(FIG. 10A)**. Across three genes and eight guide RNAs, CasRx mediated significantly greater transcript knockdown than both LwaCas13a-msfGFP-NLS and PspCas13b-NES (median: 97% compared to 80% and 66% respectively, $P< 0.0001$) **(FIG. 10B)**.

RNAi has been widely used to disrupt any gene of interest due to a combination of simple re-targeting principles, scalable synthesis, knockdown potency, and ease of reagent delivery. However, widespread off-target transcript silencing has been a consistent concern (Jackson et al., 2003; Sigoillot et al., 2012), possibly due to the entry of RNAi reagents into the endogenous miRNA pathway (Doench et al., 2003; Smith et al., 2017). Consistent with these reports, upon RNA sequencing of human cells transfected with a *B4GALNT1*-targeting shRNA, widespread off-target transcriptional changes relative to a non-targeting shRNA were observed (>500 significant off-target changes, $P<0.01$, **FIGS. 9E, 9G)**. In contrast, transcriptome profiling of spacer-matched CasRx guide RNAs revealed no significant off-target changes

other than the targeted transcript (**FIG. 9F**). This indicates that the moderate bystander cleavage observed *in vitro* (**FIG. 5C**) may not result in observable off-target transcriptome perturbation in mammalian cells. A similar pattern was observed when targeting *ANXA4* (**FIGS. 11A-11B**), with over 900 significant off-target changes resulting from shRNA targeting compared to zero with CasRx (**FIG. 9G**).

To confirm that CasRx interference is broadly applicable, a panel of 11 additional genes with diverse roles in cancer, cell signaling, and epigenetic regulation were selected and 3 guides per gene were screened. CasRx consistently mediated high levels of transcript knockdown across genes with a median reduction of 96% (**FIG. 9H**). Each tested guide mediated at least 80% knockdown, underscoring the consistency of the CasRx system for RNA interference.

Example 6

Splice isoform engineering with dCasRx

The experiments on RNA targeting with CasRx revealed that target RNA and protein knockdown is dependent on the catalytic activity of the HEPN domains **(FIGS. 8B, 5B).** The same guide sequences mediating efficient knockdown with CasRx failed to significantly reduce mCherry levels when paired with catalytically inactive dCasRx **(FIG. 8B)**, indicating that targeting of dCasRx to the coding portion of mRNA does not necessarily perturb protein translation. This observation indicated the possibility of utilizing dCasRx for targeting of specific coding and non-coding elements within a transcript to study and manipulate RNA. To validate this concept, the utility of the dCasRx system was expanded by creating a splice effector.

Alternative splicing is generally regulated by the interaction of cis-acting elements in the pre-mRNA with positive or negative trans-acting splicing factors, which can mediate exon inclusion or exclusion (Matera and Wang, 2014; Wang et al., 2015). It was reasoned that dCasRx binding to such motifs may be sufficient for targeted isoform perturbation. For proof-of-concept, distinct splice elements were identified in a bichromatic splicing reporter containing DsRed upstream of mTagBFP2 in two different reading frames following an alternatively spliced exon (Orengo et al., 2006) (**FIG. 12A**). Inclusion or exclusion of this second exon toggles the reading frame and resulting fluorescence, facilitating quantitative readout of splicing patterns by flow cytometry . To mediate exon skipping, four guide RNAs were designed to target the intronic branchpoint nucleotide, splice acceptor site, putative exonic splice enhancer, and splice donor of exon 2.

One widespread family of negative splice factors are the highly conserved heterogeneous nuclear ribonucleoproteins (hnRNPs), which typically inhibit exon inclusion via a C-terminal, glycine-rich domain (Wang et al., 2015). The splicing reporter was targeted with dCasRx and engineered fusions to the Gly-rich C-terminal domain of hnRNPa1, one of the most abundant hnRNP family members (**FIG. 12B**).

Guide position appears to be a major determinant of the efficiency of engineered exon skipping. While each guide position mediated a significant increase in exon exclusion ($P<0.0001$ in all cases) relative to the non-targeting guide, targeting the splice acceptor resulted in the most potent exon exclusion (increase from 8% basal skipping to 65% for dCasRx alone and 75% with hnRNPa1 fusion). By comparison,

dLwaCas13a-msfGFP-NLS mediated significantly lower levels of exon skipping across all four positions (19% skipping for splice acceptor guide) (**FIGS.** 10**C** and 10**D,** $P<0.0001$).

Targeting all 4 positions simultaneously with a CRISPR array achieved higher levels of exon skipping than individual guides alone (81% for dCasRx and 85% for hnRNPa1 fusion, $P<0.006$ compared to SA guide) (**FIG. 12B**) . These results indicate that dCasRx allows for tuning of isoform ratios through varying guide placement and suggest that it can be leveraged as an efficient RNA binding module in human cells for targeting and manipulation of specific RNA elements.

Example 7

Viral delivery of dCasRx to a neuronal model of frontotemporal dementia

The Cas13d family averages 930 amino acids in length, in contrast to Cas9 (~1100 aa to ~1400 aa depending on subtype, with compact outliers such as CjCas9 or SaCas9), Cas13a (1250 aa), Cas13b (1150 aa), and Cas13c (1120 aa) (**FIG. 3B**) (Chylinski et al., 2013; Cox et al., 2017; Hsu et al., 2014; Kim et al., 2017; Shmakov et al., 2015; Smargon et al., 2017). Although adeno-associated virus (AAV) is a versatile vehicle for transgene delivery and gene therapy due to its broad range of capsid serotypes, low levels of insertional mutagenesis, and lack of apparent pathogenicity, its limited packaging capacity (~4.7kb) makes it challenging to effectively deliver many single effector CRISPR enzymes (Abudayyeh et al., 2017; Ran et al., 2015; Swiech et al., 2015). The remarkably small size of Cas13d effectors render them uniquely suited for all-in-one AAV delivery with a CRISPR array, an optional effector domain, and requisite expression or regulatory elements (**FIG. 12C**).

Frontotemporal Dementia with Parkinsonism linked to Chromosome 17 (FTDP-17) is an autosomal dominant major neurodegenerative disease caused by diverse point mutations in *MAPT,* the gene encoding for tau. Tau exists as two major isoforms in human neurons, 4R and 3R, which are distinguished by the presence or absence of tau exon 10 and thus contain 4 or 3 microtubule binding domains. The balance of these two isoforms is generally perturbed in FTDP-17 as well as other tauopathies, driving the progression of neurodegeneration (Boeve and Hutton, 2008). Some forms of FTD are caused by mutations in the intron following MAPT exon 10 which disrupt an intronic splice silencer and elevate the expression of 4R tau (Kar et al., 2005), thereby inducing pathological changes (Schoch et al., 2016).

It was reasoned that dCasRx targeted to *MAPT* exon 10 could induce exon exclusion to alleviate dysregulated 4R/3R tau ratios. Patient-derived human induced pluripotent stem cells (hiPSCs) were differentiated into cortical neurons via *Neurogenin-2* directed differentiation for 2 weeks (Zhang et al., 2013). Postmitotic neurons were then transduced with AAV1 carrying dCasRx (**FIG. 12D**) paired with a repeat array containing 3 spacers that target the exon 10 splice acceptor and two putative exonic splice enhancers (**FIG. 12E**). dCasRx-mediated exon exclusion was able to reduce the relative 4R/3R tau ratio by nearly 50% relative to a BFP vehicle control (**FIG. 12F**) and to a level similar to unaffected control neurons, demonstrating that CasRx can be exploited for transcriptional modulation in primary cell types via AAV delivery.

Example 8

RNA Targeting in Human Cells Using Cas13d

RNA can be targeted in human cells using the active Cas13d nuclease. As a proof of concept, human U-2 OS bone osteosarcoma cells were stably integrated with an mCherry reporter and transfected with plasmids encoding human codon optimized Cas13d and guide RNAs targeting the mCherry transcript **(FIG. 13).**

Cas13d proteins were also fused with N- and C-terminal NLS sequences (SPKKKRKVEAS, SEQ ID NO: 256, for the N-terminal NLS and GPKKKRKVAAA, SEQ ID NO: 258, C-terminal NLS) to understand if nuclear localization can affect mCherry knockdown (these are denoted the 2x NLS constructs). Guide RNAs were either provided in a vector with a U6 promoter operably linked to a 36 nt DR-30 nt spacer-36 nt DR sequence, which mimics the unprocessed CRISPR guide array (denoted DR36), or to a 30nt DR-22nt spacer to mimic the processed, mature gRNA (denoted gRNA). The DR36 construct is presumed to be processed by Cas13d into mature gRNAs within the cell. The spacer sequences within the DR36 or gRNA molecules were either complementary to the mCherry target RNA (on-target mCherry) or computationally optimized to avoid complementarity to mCherry or any endogenous human transcript (non-targeting mCherry).

mCherry knockdown was quantified via flow cytometry and normalized to a transfection control. The non-targeting mCherry guides do not affect mCherry protein levels via flow cytometry, presumably because the mCherry transcripts are not targeted. However, the on-target mCherry guides paired with 4 different Cas13d orthologues exhibited significant mCherry knockdown **(FIG. 13).** "XPD" refers to Ruminococcus flavefaciens XPD3002 Cas13d; "P1E0" refers to Gut metagenome P1E0 Cas13d; "AnDig" refers to Anaerobic digester gut metagenome Cas13d; "Uncultured" refers to uncultured Ruminococcus sp. Cas13d.

Example 9

in vivo RNA Targeting Using Cas13d

RNA can be targeted in mouse models of cancer. To observe which cells in the mouse are expressing EGFR, a guide RNA is designed that includes one or more spacer regions complementary to mouse EGFR and is combined with Cas13d having a mutated HEPN domain (such as SEQ ID NO: 2 or 4), and a biotin label. The gRNA and Cas13d coding sequence are cloned into a viral vector (such as a lentivirus) which is used to infect the mice by tail vein injection at a titer to insure 100% infection rates. A fluorescent streptavidin label is administered to the mice. Cells expressing EGFR are visualized and detected with the appropriate excitation frequency for the fluorescent label. Alternatively, Cas13d is delivered in its active form *in vivo* to mediate target knock-down.

Example 10

Treatment of Cancer

Human subjects with histologically confirmed stage 1, EGFR+ breast cancer can be treated with the disclosed methods. Each subject is administered a complex comprising an active Cas13d or a Cas13d protein mutated in the HEPN domain (such as SEQ ID NO: 2 or 4), a guide RNA targeting EGFR, and a toxin, after receiving lumpectomy surgery. Treated individuals are monitored for breast cancer recurrence.

Example 11

Treatment of HIV Infection

Human subjects with HIV infection can be treated with the disclosed methods. Each subject is administered a construct comprising an active Cas13d or a Cas13d protein mutated in the HEPN domain (such as SEQ ID NO: 2 or 4), a guide RNA targeting HIV Nef protein, and a toxin. Treated individuals are monitored for HIV progression.

Example 12

Treatment of Huntington's Disease

Human subjects with Huntington's Disease can be treated with the disclosed methods. Each individual is administered a construct comprising Cas13d, a guide RNA targeting the Huntington mutation. Treated individuals are monitored for disease progression.

Example 13

Alternative Splicing Using Cas13d

Cas13d splice effectors can be used for therapeutic protein restoration (for example that results from a mutation or deletion), gene knockdown via frameshift induction, tuning or restoring a desired isoform ratio, or inducing a desired dominant splice isoform (FIG. 14). Alternative splicing is generally regulated by the interaction of cis-acting elements in the pre-mRNA with positive or negative trans-acting splicing factors, which can mediate exon inclusion or exclusion. dCas13d and Cas13d, with optional fusion to positive or negative splicing factors, can be used as splicing effectors that target to said cis-acting elements in the pre-mRNA to manipulate splicing. Such elements can include exonic splicing enhancers or suppressors, intronic splicing enhancers or suppressors, splice acceptor and splice donor sites, and more generally protein- or RNA-interacting motifs or elements on that particular pre-mRNA, mRNA, or other RNA species, such as a non-coding RNA, tRNA, miRNA, and the like.

Additionally, that the effects of Cas13d-based splice effectors can be guide position-dependent. This can be exploited to perturb or discover particular motifs or sites in an RNA transcript such as protein-binding sites, via steric hindrance, blocking, recruitment, or effector-mediated interaction. For example, the interaction between non-coding RNAs and particular chromatin remodeling complexes can be perturbed.

Access of the ribosomal binding site and other elements can be blocked in a 5' or 3' UTR (or to recruit appropriate effector domains) to decrease, increase, or otherwise manipulate translation.

Targeting or tiling Cas13d guides along a pre-mRNA can be used to discover or map new cis-acting elements such as intronic or exonic splice enhancers. This has been exploited in a therapeutic context in the case of the dystrophin gene for optimal antisense oligonucleotide positioning and can also be used for optimal Cas13dpositioning. This can be also used to map, mask, or otherwise perturb RNA zipcodes or other cis-acting elements to affect trafficking and localization, chromatin remodeling, polyadenylation, RNA stability and half-life, or levels of nonsense-mediated decay.

In one example, targeting an RNA allows for changing splicing of the target RNA. Both the direct binding of splice acceptor and/or donor sites as well as splice effector domains can be used to manipulate splicing. For example, by using a dCas13d protein with a mutated HEPN domain (*e.g.*, SEQ ID NO: 2 or 4), a guide RNA containing at least one a spacer sequence specific for the target RNA, and optionally an effector domain that affects splicing (such as the RS-rich domain of SRSF1, the Gly-rich domain of hnRNPA1, the alanine-rich motif of RBM4, or the proline-rich motif of DAZAP1), alternative splicing of the RNA can be achieved.

In some examples, such a method is used for exon inclusion, for example to include exon 2 of acid alpha-glucosidase (GAA) to treat Pompe disease or to include exon 7 of SMN2 to treat spinal muscular atrophy (SMA). In some examples, such a method is used for exon exclusion, for example to restore the reading frame of dystrophin to treat Duchenne muscular dystrophy, to shift the splicing of the Bcl-x pre-mRNA from the antiapoptotic long isoform to the proapoptotic short isoform to treat cancer, to shift the splicing of the MAPT transcript to affect ratios of 3R and 4R tau, or to manipulate the splicing of the lamin A transcript in the case of Hutchinson-Gilford progeria syndrome or other genetic diseases of aging.

In some examples, the method uses a Cas13dprotein, optionally with a mutated HEPN domain, to mask splice acceptor or donor sites, for example to create neoantigens to make cold tumors hot. By affecting the splicing of certain target pre-mRNAs, this method can generate novel exon-exon junctions that can lead to the creation of neo-epitopes in cancer cells. This can make a cancer cell vulnerable to the immune system due to the display of unnatural antigens. In other examples, this method can be used to dynamically manipulate isoform ratios or to restore reading frame of a protein (*e.g.*, dystrophin for Duchenne's muscular dystrophy).

Example 14

AAV delivery of Cas13d

As described in the examples above, Cas13d can be effectively packaged into AAV to mediate expression in cell types that are not amenable to plasmid delivery or for *in vivo* delivery of Cas13d. AAV delivery of nuclease active Cas13d can be used to mediate RNA target knock-down in the cell type of interest. Due to its small size compared to other single-effector CRISPR nucleases, Cas13d can be packaged together with a guide RNA or an array containing multiple guide RNAs in a single AAV vector.

Example 15

Nucleic acid-based diagnostics with Cas13d

Cas13d enzymes can be exploited for nucleic acid-based diagnostics within the context of a cell, using cell-free lysate derived from a cell, or a cell-free system containing an engineered Cas13d enzyme and guide RNA to facilitate formation of a ribonucleoprotein complex. Said guide RNA can be provided in the form of a pre-guide RNA, a mature guide RNA, or an array containing one or more spacer sequences. The components can also be provided in the form of a DNA or RNA precursor encoding for the Cas13d enzyme and appropriate guide RNA design via an *in vitro* transcription/translation system to facilitate the generation of the necessary components. These components of the diagnostic kit comprise the "sensor" module.

Such a method can be used to determine if a target RNA is present in a test sample. Such a method can also be used to detect a pathogen, such as a virus or bacteria, or diagnose a disease state, such as a cancer (e.g., wherein the target RNA is specific for a particular microbe or disease). Such a method can also be used to test the purity or identity of an environmental sample or agricultural sample, such as seed or soil.

The "sensor" module will then be challenged with a test sample in the form of RNA. Said test sample can be, but is not limited to, a genomic DNA sample that is converted into RNA – for example via in vitro transcription - or a direct RNA sample. These samples can be extracted from biological material such as patient samples (*e.g.*, cells, tissue, blood, plasma, serum, saliva, urine, tumor biopsy, cell free DNA or RNA, exosomes, carrier vesicles or particles) and environmental samples (*e.g.*, soil, water, air, seed, or plant samples). In one embodiment to improve diagnostic sensitivity, nucleic acid molecules in the sample are amplified using amplification techniques, such as polymerase chain reaction, recombinase polymerase amplification, loop mediated isothermal amplification, nucleic acid sequence based amplification, strand displacement amplification, rolling circle amplification, ligase chain reaction, and others (e.g., those that use isothermal amplification). Said amplification techniques can optionally employ nucleic acid conversion techniques such as transcription or reverse transcription with randomized primers or targeted primers.

If the sensor module recognizes a cognate target in the test sample, it will activate an RNase activity. This RNase activity can be detected by using a detectable label. In one example, the detectable label includes an RNA linked to a fluorophore and quencher. The intact detectable RNA links the fluorophore and quencher, suppressing fluorescence. Upon cleavage by Cas13d of the detectable RNA, the fluorophore is released from the quencher and displays detectable fluorescent activity.

In another example, cleavage of the reporter RNA releases a non-fluorescent molecule, which can be converted into a visible signal (e.g., visible by eye). In one example, cleavage of the reporter RNA releases a molecule that can be detected via lateral flow. A molecule that can be detected by lateral flow is any molecule that can be bound specifically by antibodies. In one example, the Cas13d protein along with the guide RNA detecting the target and the reporter RNA conjugated to the reporter molecule can be delivered as a single system in the form of a dry test strip. Upon incubation with the test sample, Cas1Cas13d protein a3d, guide RNA and the reporter RNA are rehydrated and in the presence of the RNA target. The Cas13d

protein will cleave the reporter RNA, resulting in the migration of the reporter molecule in the test strip via lateral flow and a resulting positive test line signal by binding to antibodies localized there. Such a shelf stable dry detection system not requiring special (frozen) storage could for example prove advantageous in situations where detection of a target RNA or DNA is performed outside a centralized laboratory facility such as a doctor's office, a hospital, a pharmacy, during field work, in an agricultural setting and so forth.

Cas13d is active over a broad range of temperatures making such an application outside of a controlled laboratory environment feasible.

Example 16

Cas13d as a diagnostic for RNA or DNA transcribed into RNA *in vitro*

Cas13d is capable of converting the presence of a matched target RNA into a visible signal in a minimal diagnostic *in vitro* system. **FIGS. 15A-15D** (A) Cas13d is converted into an active RNase complex upon binding a target matching the spacer sequence of the guide RNA. It is capable of cleaving gRNA-complementary target RNA or non-complementary bystander RNAs. (B) Cas13d target-dependent RNase activity can be converted into a detectable signal, for example through cleavage of a labeled detector RNA that is cleaved only in the presence of a target matching the spacer of the Cas13d guide RNA. In this example, the detector RNA contains a fluorophore, 'F', and a quencher 'Q', that abolishes fluorescence. Only upon bystander RNA cleavage is the fluorophore liberated from the quencher and fluorescence is generated. (C) Cas13d from *E. siraeum* produces a visible signal only in the presence of a perfectly matched target and not in the presence of different mismatched targets. (D) Cas13d from *R. flavefaciens* strain XPD3002 produces a visible signal only in the presence of a perfectly matched target and not in the presence of different mismatched targets.

Thus, the system disclosed herein can be part of a lateral flow device (or other solid support), that can be used in diagnostics. The presence of an RNA or DNA sequence can be converted into a signal than can then be detected by conventional lateral flow.

Example 17

Cas13d Modifications

FIGS. 16A-16B, Cas13d is amenable to modifications including truncations of regions with low conservation among orthologs. The alignment of Cas13d orthologs in FIG. 16A shows regions with high (green bars) and low (red bars) conservation.

Example 18

Targeting Transcripts *in vivo*

The ccdB gene in bacterial cells was targeted *in vivo* using different nCas1 orthologues, *Eubacterium siraeum* nCas1 (Es_nCas1; SEQ ID NO: 1); *Eubacterium siraeum* nCas1 with mutated HEPN domains (Es_nCas1 HEPN -/-; SEQ ID NO: 2); uncultured Ruminococcus sp. nCas1 (uncul_nCas1; SEQ ID

NO: 3) and uncultured Ruminococcus sp. nCas1 with mutated HEPN domains (uncul_nCas1 HEPN -/-; SEQ ID NO: 4) (**FIG. 17A**). Chemically competent *E. coli* (strain BW25141-DE3) cells were transformed with (1) an arabinose-inducible ccdB plasmid, and (2) a second plasmid (targeting vector) carrying a compatible origin of replication, nCas1 protein coding sequence, and nCas1 guide array containing 4 spacer sequences targeting the ccdB transcript (FIG. 17A).

1. artificial Eubacterium siraeum nCas1 array targeting ccdB
    a. **GAACUACACCCGUGCAAAAAUGCAGGGGUCUAAAAC**UAACGGCUCUCUCUU UUAUAGGUGUAAACC**GAACUACACCCGUGCAAAAAUGCAGGGGUCUAAAA CCUUUAUCUGACAGCAGACGUGCACUGGCCA**GAACUACACCCGUGCAAAAA **UGCAGGGGUCUAAAAC**CAUCAUGCGCCAGCUUUCAUCCCCGAUAUG**GAACU ACACCCGUGCAAAAAUGCAGGGGUCUAAAAC**UAAUGGCGUUUUUGAUGUCA UUUUCGCGGUCCGCUGA (SEQ ID NO: 261)
        i. Full 36 nt direct repeat: **GAACUACACCCGUGCAAAAAUGCAGGGGUCUAAAAC** (SEQ ID NO: 262)
        ii. Spacer 1: UAACGGCUCUCUCUUUUAUAGGUGUAAACC (SEQ ID NO: 263)
        iii. Spacer 2: CUUUAUCUGACAGCAGACGUGCACUGGCCA (SEQ ID NO: 264)
        iv. Spacer 3: CAUCAUGCGCCAGCUUUCAUCCCCGAUAUG (SEQ ID NO: 265)
        v. Spacer 4: UAAUGGCGUUUUUGAUGUCAUUUUCGCGGUCCGCUGA (SEQ ID NO: 266)
2. artificial uncultured Ruminoccus sp. nCas1 array targeting ccdB
    a. **CUACUACACUGGUGCAAAUUUGCACUAGUCUAAAAC**UAACGGCUCUCUCUU UUAUAGGUGUAAACC**CUACUACACUGGUGCAAAUUUGCACUAGUCUAAAAC** CUUUAUCUGACAGCAGACGUGCACUGGCCA**CUACUACACUGGUGCAAAUUU GCACUAGUCUAAAAC**CAUCAUGCGCCAGCUUUCAUCCCCGAUAUG**CUACUAC ACUGGUGCAAAUUUGCACUAGUCUAAAAC**UAAUGGCGUUUUUGAUGUCAUU UUCGCGGUCCGC (SEQ ID NO: 267)
        i. Full 36 nt direct repeat: **CUACUACACUGGUGCAAAUUUGCACUAGUCUAAAAC** (SEQ ID NO: 268)
        ii. Spacer 1: UAACGGCUCUCUCUUUUAUAGGUGUAAACC (SEQ ID NO: 269)
        iii. Spacer 2: CUUUAUCUGACAGCAGACGUGCACUGGCCA (SEQ ID NO: 270)
        iv. Spacer 3: CAUCAUGCGCCAGCUUUCAUCCCCGAUAUG (SEQ ID NO: 271)
        v. Spacer 4: UAAUGGCGUUUUUGAUGUCAUUUUCGCGGUCCGCUGA (SEQ ID NO: 272)
3. Target RNA (ccdB sequence)

a. AUGCAGUUUAA**GGUUUACACCUAUAAAAGAGAGAGCCG**UUAUCGUCUGUUU GUGGAUGUACAGAGUGAUAUUAUUGACACGCCCGGGCGACGGAUGGUGAUCC CCC**UGGCCAGUGCACGUCUGCUGUCAGAUAAAG**UCUCCCGUGAACUUUACC CGGUGGUG**CAUAUCGGGGAUGAAAGCUGGCGCAUGAUG**ACCACCGAUAUG GCCAGUGUGCCGGUCUCCGUUAUCGGGGAAGAAGUGGCUGAUCUCAGCC**ACC GCGAAAAUGACAUCAAAAACGCCAUUA**ACCUGAUGUUUUGGGGAAUA (SEQ ID NO: 273)

i. spacer 1 target: **GGUUUACACCUAUAAAAGAGAGAGCCGUUA** (SEQ ID NO: 274)

ii. spacer 2 target: **UGGCCAGUGCACGUCUGCUGUCAGAUAAAG** (SEQ ID NO: 275)

iii. spacer 3 target: **CAUAUCGGGGAUGAAAGCUGGCGCAUGAUG** (SEQ ID NO: 276)

iv. spacer 4 target: **ACCGCGAAAAUGACAUCAAAAACGCCAUUA** (SEQ ID NO: 277)

v. dCas9 target (underlined): UUAUCGUCUGUUUGUGGAUG (SEQ ID NO: 259)

The transformed bacteria were plated on 2 mM arabinose plates to induce ccdB expression and harvested after 24 hours. Total RNA was extracted with Trizol followed by random hexamer-mediated reverse transcription and Taqman probe-based qPCR.

Mutation of the two HEPN domains in each nCas1 protein demonstrate no targeting of ccdB (**FIG. 17B**, HEPN -/-), while active wild-type nCas1 proteins knock downed the expression of ccdB (**FIG. 17B**, Es_nCas1; uncul_nCas1). Because dCas9-(catalytically inactive, or "dead" Cas9)-mediated transcriptional repression through binding of the DNA sequence downstream of the transcriptional start site is a current standard in the field, we validate the assay by targeting dCas9 to the ccdB promoter to repress transcription of the ccdB gene. Taken together, this data demonstrates guide RNA-specific degradation of target transcripts *in vivo* within prokaryotic cells in a HEPN domain-dependent manner.

Example 19

Cas13d Orthologs Decrease Expression in Human Cells

Fifty different Cas13d orthologs (some of which included NLS tags SPKKKRKVEAS SEQ ID NO: 256 and GPKKKRKVAAA SEQ ID NO: 25, or two HIV NES sequences (LQLPPLERLTL, SEQ ID NO: 287, see Table 1) were tested for their ability to decrease mCherry or endogenous CD81 expression in 293 cells.

Cell culture

Human embryonic kidney (HEK) cell line 293FT (Thermo Fisher) was maintained in DMEM (4.5 g/L glucose), supplemented with 10% FBS (GE Life Sciences) and 10 mM HEPES at 37°C with 5% CO_2. Upon reaching 80-90% confluency, cells were dissociated using TrypLE Express (Life Technologies) and passaged at a ratio of 1:2.

Transient transfection of human cell lines

Cas13d orthologs were cloned into common expression backbones flanked either by two HIV NES sequences (SEQ ID NO: 287) or two SV40 NLS sequences (SPKKKRKVEAS, SEQ ID NO: 256 and GPKKKRKVAAA, SEQ ID NO: 25), and a T2A-linked EGFP. Spacers targeting CD81 were cloned into guide expression constructs in which an array of two spacers was driven by a U6 promoter.

Prior to transient transfection, HEK 293FT cells were plated at a density of 20,000 cells per well in a 96-well plate and transfected at >90% confluence with 200 ng of Cas13 expression plasmid and 200 ng of gRNA expression plasmid using Lipofectamine 2000 (Life Technologies) according to the manufacturer's protocol. Transfected cells were harvested 72 hours post-transfection for flow cytometry.

Flow cytometry

Prior to analysis by flow cytometry, 293FT cells were blocked with Human Trustain FcX Blocking Solution (Biolegend) for 10 minutes at room temperature in FACS Buffer (1X $DPBS^{-/-}$, 0.2% BSA, 2 mM EDTA). Cells were then stained for 15 minutes on ice with α-CD81-APC (1:400 for 293FTs Miltenyi Biotec) or REA293 (S) Isotype Control (Miltenyi Biotec) antibodies. Flow cytometry was performed in 96-well plate format using a MACSQuant VYB (Miltenyi Biotec) and analyzed using FlowJo 10.

As shown in Table 1, several of orthologs had significant knowndown activity (*i.e.*, a lower value for the knockdown efficiency). In Table 1, (trunc) indicates a version of the ortholog with a truncation of the non-conserved N-terminus. NES indicates orthologs with a dual HIV NES fusion (SEQ ID NO: 287) and NLS indicates orthologs with a dual SV40 NLS fusion (SEQ ID NOS: 256 and 25).

Table 1: mCherry activity screen of Cas13d orthologs in human cells.

Ortholog	Knockdown Efficiency	Standard Deviation	Seq ID No.
Ga7274-NLS	12.7	1.0	198
Rfx-NLS	18.6	2.6	280
Ur(trunc)-NLS	19.0	1.2	3
twins_985-NES	20.7	1.3	212
Ga_0531-NES	22.3	2.8	235
Ga_0531-NLS	22.6	0.5	235
k87-NES	24.7	3.1	189
twins_985-NLS	26.4	1.8	212

Ur(trunc)-NES	27.7	1.0	3
stool_9834-NLS	29.0	2.4	208
OGPN-NES	32.6	1.9	164
Ur-NLS	33.9	0.7	3
OGZC-NES	34.7	1.7	155
Ga_9260-NLS	36.5	6.6	237
NFIR-NLS	36.7	1.5	166
Ga_9260-NES	38.3	2.9	237
DJXD-NLS	39.2	3.0	153
OGZC-trunc	40.3	1.1	155
k87-NLS	41.1	4.6	189
PIG_6364-NLS	41.2	6.3	243
twins_300-NLS	42.8	5.9	210
DLF_1914-NES	44.5	10.5	224
BMZ_7771-NLS	46.4	5.9	218
PIG_3397-NES	46.9	3.2	245
Ga7274-NES	47.7	3.6	198
PIG_5479-NES	49.1	3.6	249
OCTW-NLS	50.1	10.8	183
PIG_6364-NES	50.3	2.1	243
Piggy-NES	51.6	5.4	170
OGNF-NLS	52.3	3.2	185
Ga_8792-NES	52.9	1.9	206
OJMM-NES	53.1	1.5	175
OGDF-NLS	53.6	3.4	162
GL_3589-NLS	53.7	3.5	216
Ga_5336-NLS	53.7	0.6	233
OGDF-NES	54.6	1.4	162
BMZ_7915-NES	55.7	8.0	220
OCTW-NES	55.9	10.8	183
OBVH-NLS	56.5	5.2	147
Ga8792-NLS	59.0	8.8	206
OJMM-NLS	59.1	1.6	175
BMZ_7915-NLS	59.9	9.7	220
BMZ_7771-NES	59.9	4.6	218
Rfx-NES	60.5	7.7	92

Ga_4192	62.2	5.8	229
twins_300-NES	62.4	8.4	210
DJXD-NES	63.5	8.2	153
Ga_8067-NES	63.7	10.9	204
PIG_5479-NLS	64.4	20.8	249
Ga0735-NLS	65.2	12.4	202
OGPN-NLS	66.2	6.4	164
Piggy-NLS	66.3	8.0	147
MH_4866-NLS	67.1	4.1	292
OIEN-NES	67.8	16.6	187
Ga_4192-NLS	68.7	2.2	229
Ga_0735-NES	69.2	12.0	202
OHCP-NLS	69.2	4.1	160
OIEN-NLS	69.6	7.2	187
DBYI-NLS	70.5	13.7	149
Ga5759-trunc	71.2	15.2	251
DBYI-NES	71.5	7.1	149
stool_9834-NES	72.6	3.9	208
Ga_6931-NLS	73.4	5.2	231
MH_4866-NES	76.6	3.1	292
O2UC_6317-NES	79.2	2.7	241
OGNF-NES	80.2	10.2	185
Ga_6931-NES	83.3	15.9	231
OCVV-NES	84.8	14.7	181
O2UC_6317-NLS	86.2	6.6	241
ODAI_557-NES	86.5	14.4	
Ga5759-NLS	91.1	10.3	251
ODAI1274-NLS	92.0	9.0	177
Ga_5336-NES	93.5	2.8	233
OCVV-NLS	93.5	5.6	181
GL_3589-NES	95.4	3.5	216
PIG_9734-NLS	97.1	10.6	247
EMG3641-NES	105.8	7.6	200
GL_4153-NLS	106.7	5.3	214
BMZ_9617-NES	110.0	7.8	222
EMG3641-NLS	112.6	11.3	200

Ga8067-NLS	111.4	16.1	204

The data in Table 1 were plotted as relative mCherry expression (knockdown efficiency) vs. size of the ortholog, and compared to CasRx (SEQ ID NO: 280). As shown in FIG. 19A, three orthologs (Ga0531-NLS, k87-NES, and Ga7274-NLS, SEQ ID NOS: 235, 189, and 198, respectively) show comparable activity to the CasRx variant (SEQ ID NO: 280).

The 25 most active orthologs shown in Table 1 and FIG. 19A were tested for their ability to decrease expression of an endogenous gene, CD81, in HEK293 cells. As shown in FIG. 19B, the CasRx variant RfxCas13d-NLS (SEQ ID NO: 280) exhibited the highest level of knock-down, with several additional orthologs exhibiting >75% knock-down on this endogenous target.

Analysis of the five orthologs exhibiting >70% knock-down indicated the following consensus motifs are present at a higher frequency in Cas13d orthologs with high activity in human cells compared to Cas13d orthologs with low activity (<50% knock-down). Thus, provided herein are Cas13d proteins having such a consensus motif (*e.g.*, protein sequences comprising SEQ ID NO: 288, 289, 290, or 291, as well as nucleic acids encoding such, vectors encoding such nucleic acids, and methods of using such proteins/ nucleic acids):

1. L-[DE]-G-K-E-I-N-x-[LM]-x-[CST]-[ANST]-L-I-x-K-F-[DE]-[GN]-I-x-[ADS]-[FIL]-x-[DEHK]-[AILV] (SEQ ID NO: 288)
2. R-N-[FY]-I-[AIL]-x-N-V-[ILV]-x-[NS]-x(2)-F-x-Y-[ILV]-[AIV]-[KR]-[HY]-x(2)-[PT]-x(3)-[HKR]-x-[IL]-x(2)-N-x(2)-[ILV]-x(2)-[FY]-V-L-x(2)-[ILM]-[AGP]-[DE]-x(4)-[NQRS] (SEQ ID NO: 289)
3. L-x-[LM]-[ST]-[IV]-x-Y-x(3)-K-[NS]-[ILM]-V-x-[IV]-N-[AST]-R-[FY]-x(2)-[AG]-[FY]-x-[ACV]-[LV]-E-R-D-x(2)-[FL] (SEQ ID NO: 290)
4. P-F-[AGS]-Y-[CNV]-x-[AP]-R-[FY]-K-N-L-[ST]-x-[EN]-x-L-F-x-[DKR]-[DEN] (SEQ ID NO: 291)

Example 20

Comparison of Knock-Down Activity of CasRx to PspCas13b-NES

The ability of CasRx (RfxCas13d-NLS design; SEQ ID NO: 280) knock-down endogenous CD81 protein expression in 293FT cells was compared to PspCas13b-NES (see Cox *et al.*, *Science*. 2017 Nov 24;358(6366):1019-1027). 293FT cells were cultured, transiently transfected, and three days later, CD81 expression was detected using flow cytometry, as described in Example 19.

As shown in FIG. 20, CasRx (SEQ ID NO: 280) displays significantly higher activity at equivalent protein levels (as measured by Cas13-GFP expression), indicating a higher activity per enzyme unit and therefore a lower required dose to reach a biologically/therapeutically relevant level of knock-down. In addition, CasRx is significantly smaller than PspCas13b. The small size and high activity at lower doses makes CasRx useful for therapeutic delivery.

Example 21

Effect of Deletions on Cas13d Orthologs

The effect of deletions on Cas13d orthologs was tested as follows. Briefly, amino acid (aa) deletions were engineered into two different Cas13d orthologs, Ga-531 (SEQ ID NO: 235) and RfxCas13d (SEQ ID NO: 280). The amino acid deletions were made at the N-terminus, C-terminus, or internally. Regions for truncation were chosen based on their low conservation between Cas13d orthologs (*e.g.*, see FIGS. 18A-18MMM). Up to 50 aa deletions were made in a single protein. The resulting protein sequences are shown in Table 2.

Table 2: Cas13d mutations (**Bold**= NLS/NES tag, underlined = inserted linker sequence; *italics* = flanking sequence)

Effector	Sequence (SEQ ID NO:)	Description
NLS-Ga_0531(trunc)-NLS-HA	MS**PKKKRR***KVEASKQSKK*TFAKASGLKST LAVDNSAVMTVFGRGNEAKLDHRINAD LQSESLHPQAALKNVHAPNKQKIHFIGR MQDMNLTADHPLHSHDGERAVGADLL CAKDKLEQLYFGRTFNDNIHIQLIYQILD IQKILALHANNIIFALDNLLHKKNDELSD DFVGMGRMRATIGYDAFRNSTNQKVQE TYREFQEFVRRKELLYFGSAFYNGDTRR DEKVIYHILSLAASVRQFCFHNDYTSDD GKGFIKADWMYRLEEALPAEYKDTLDA LYLEGVEGLDQSFLKNNTVNIQILCSIFN HDDPNKIAEEYYGFLMTKEYKNMGFSIK KLRECMLELPELSGYKEDQYNSVRSKLY KLFDFIIAHYFRKHPEKGEEMVDCLRLC MTEDEKDSHYEGTAKKLVRELAYDMQ EAAEQANGSNITQMQKNEQQGKTKGMF AIRDEIRVSRKPVSYFSKVIYVMTLLLDG KEINDLLTTLINKFENIVSFEDVLRQLNV DCTFKPEFAFFGYDRCRNISGELRLINSF ARMQKPSAKAKHVMYRDALRILGLDN GMSEEALDQEVRRILQIGADGKPIKNAN KGFRNFIASNVIESSRFRYLVRYNNPHKT RMIAQNEAIVRFVLSEIPDEQIRRYYDVC RDPKLPRSSSREAQVDILTGIITDVNYRIF EDVPQSKKINKDRPDANDRMTLKKQRY QAIVSLYLTVMYLVTKNLVYVNSRYVM AFHALERDAYLYGITNIKGDYRKLTDNL LADENYKKFGHFKNKKWRGIAEQNLRN SDVPVIKSFRNMAAHISVIRNIDLYIGDIQ KVDSYFALYHFLMQKLIQRVVPENTKG LSDQTKKYYDALEQYNTYCKDFVKAYC TPFAYVTPRYKNLTIDGLFDRNRPGEDK GSG**PKKKRKV**AAAYPYDVPDYA (SEQ ID NO: 278)	N-terminal deletion
NES-Ga_0531(tr	M**LQLPPLERL***TLGSGKQSKK*TFAKASGL KSTLAVDNSAVMTVFGRGNEAKLDHRI	N-terminal deletion

unc)-NES-HA	NADLQSESLHPQAALKNVHAPNKQKIHF IGRMQDMNLTADHPLHSHDGERAVGAD LLCAKDKLEQLYFGRTFNDNIHIQLIYQI LDIQKILALHANNIIFALDNLLHKKNDEL SDDFVGMGRMRATIGYDAFRNSTNQKV QETYREFQEFVRRKELLYFGSAFYNGDT RRDEKVIYHILSLAASVRQFCFHNDYTS DDGKGFIKADWMYRLEEALPAEYKDTL DALYLEGVEGLDQSFLKNNTVNIQILCSI FNHDDPNKIAEEYYGFLMTKEYKNMGF SIKKLRECMLELPELSGYKEDQYNSVRS KLYKLFDFIIAHYFRKHPEKGEEMVDCL RLCMTEDEKDSHYEGTAKKLVRELAYD MQEAAEQANGSNITQMQKNEQQGKTK GMFAIRDEIRVSRKPVSYFSKVIYVMTLL LDGKEINDLLTTLINKFENIVSFEDVLRQ LNVDCTFKPEFAFFGYDRCRNISGELRLI NSFARMQKPSAKAKHVMYRDALRILGL DNGMSEEALDQEVRRILQIGADGKPIKN ANKGFRNFIASNVIESSRFRYLVRYNNPH KTRMIAQNEAIVRFVLSEIPDEQIRRYYD VCRDPKLPRSSSREAQVDILTGIITDVNY RIFEDVPQSKKINKDRPDANDRMTLKKQ RYQAIVSLYLTVMYLVTKNLVYVNSRY VMAFHALERDAYLYGITNIKGDYRKLT DNLLADENYKKFGHFKNKKWRGIAEQN LRNSDVPVIKSFRNMAAHISVIRNIDLYI GDIQKVDSYFALYHFLMQKLIQRVVPEN TKGLSDQTKKYYDALEQYNTYCKDFVK AYCTPFAYVTPRYKNLTIDGLFDRNRPG EDKGSG**LQLPPLERLTL**YPYDVPDYA (SEQ ID NO: 279)	
NLS-RfxCas13d-NLS-HA	MS**PKKKRKV**EASIEKKKSFAKGMGVKS TLVSGSKVYMTTFAEGSDARLEKIVEGD SIRSVNEGEAFSAEMADKNAGYKIGNAK FSHPKGYAVVANNPLYTGPVQQDMLGL KETLEKRYFGESADGNDNICIQVIHNILD IEKILAEYITNAAYAVNNISGLDKDIIGFG KFSTVYTYDEFKDPEHHRAAFNNNDKLI NAIKAQYDEFDNFLDNPRLGYFGQAFFS KEGRNYIINYGNECYDILALLSGLRHWV VHNNEEESRISRTWLYNLDKNLDNEYIS TLNYLYDRITNELTNSFSKNSAANVNYI AETLGINPAEFAEQYFRFSIMKEQKNLGF NITKLREVMLDRKDMSEIRKNHKVFDSI RTKVYTMMDFVIYRYYIEEDAKVAAAN KSLPDNEKSLSEKDIFVINLRGSFNDDQK DALYYDEANRIWRKLENIMHNIKEFRGN KTREYKKKDAPRLPRILPAGRDVSAFSK LMYALTMFLDGKEINDLLTTLINKFDNI QSFLKVMPLIGVNAKFVEEYAFFKDSAK IADELRLIKSFARMGEPIADARRAMYIDA IRILGTNLSYDELKALADTFSLDENGNKL KKGKHGMRNFIINNVISNKRFHYLIRYG DPAHLHEIAKNEAVVKFVLGRIADIQKK QGQNGKNQIDRYYETCIGKDKGKSVSE	No Deletion

	KVDALTKIITGMNYDQFDKKRSVIEDTG RENAEREKFKKIISLYLTVIYHILKNIVNI NARYVIGFHCVERDAQLYKEKGYDINL KKLEEKGFSSVTKLCAGIDETAPDKRKD VEKEMAERAKESIDSLESANPKLYANYI KYSDEKKAEEFTRQINREKAKTALNAYL RNTKWNVIIREDLLRIDNKTCTLFRNKA VHLEVARYVHAYINDIAEVNSYFQLYH YIMQRIIMNERYEKSSGKVSEYFDAVND EKKYNDRLLKLLCVPFGYCIPRFKNLSIE ALFDRNEAAKFDKEKKKVSGNSGSG**PK KKRKV**AAAYPYDVPDYA (SEQ ID NO: 280)	
NLS- RfxCas13d (del5)- NLS-HA	MS**PKKKRKV**EASIEKKKSFAKGMGVKS TLVSGSKVYMTTFAEGSDARLEKIVEGD SIRSVNEGEAFSAEMADKNAGYKIGNAK FSHPKGYAVVANNPLYTGPVQQDMLGL KETLEKRYFGESADGNDNICIQVIHNILD IEKILAEYITNAAYAVNNISGLDKDIIGFG KFSTVYTYDEFKDPEHHRAAFNNNDKLI NAIKAQYDEFDNFLDNPRLGYFGQAFFS KEGRNYIINYGNECYDILALLSGLRHWV VHNNEEESRISRTWLYNLDKNLDNEYIS TLNYLYDRITNELTNSFSKNSAANVNYI AETLGINPAEFAEQYFRFSIMKEQKNLGF NITKLREVMLDRKDMSEIRKNHKVFDSI RTKVYTMMDFVIYRYYIEEDAKVAAAN KSLPDNEKSLSEKDIFVINLRGSFNDDQK DALYYDEANRIWRKLENIMHNIKEFRGN KTREYKKKDAPRLPRILPAGRDVSAFSK LMYALTMFLDGKEINDLLTTLINKFDNI QSFLKVMPLIGVNAKFVEEYAFFKDSAK IADELRLIKSFARMGEPIADARRAMYIDA *IRILG*GGGGSGGGGS*MRNFI*INNVISNKRF HYLIRYGDPAHLHEIAKNEAVVKFVLGR IADIQKKQGQNGKNQIDRYYETCIGKDK GKSVSEKVDALTKIITGMNYDQFDKKRS VIEDTGRENAEREKFKKIISLYLTVIYHIL KNIVNINARYVIGFHCVERDAQLYKEKG YDINLKKLEEKGFSSVTKLCAGIDETAPD KRKDVEKEMAERAKESIDSLESANPKLY ANYIKYSDEKKAEEFTRQINREKAKTAL NAYLRNTKWNVIIREDLLRIDNKTCTLF RNKAVHLEVARYVHAYINDIAEVNSYF QLYHYIMQRIIMNERYEKSSGKVSEYFD AVNDEKKYNDRLLKLLCVPFGYCIPRFK NLSIEALFDRNEAAKFDKEKKKVSGNSG SG**PKKKRKV**AAAYPYDVPDYA (SEQ ID NO: 281)	Amino acids 558-587 deleted
NLS- RfxCas13d (del5.12)- NLS-HA	MS**PKKKRKV**EASIEKKKSFAKGMGVKS TLVSGSKVYMTTFAEGSDARLEKIVEGD SIRSVNEGEAFSAEMADKNAGYKIGNAK FSHPKGYAVVANNPLYTGPVQQDMLGL KETLEKRYFGESADGNDNICIQVIHNILD IEKILAEYITNAAYAVNNISGLDKDIIGFG KFSTVYTYDEFKDPEHHRAAFNNNDKLI	Amino acids 558-587 and 953-966 deleted

	NAIKAQYDEFDNFLDNPRLGYFGQAFFS KEGRNYIINYGNECYDILALLSGLRHWV VHNNEEESRISRTWLYNLDKNLDNEYIS TLNYLYDRITNELTNSFSKNSAANVNYI AETLGINPAEFAEQYFRFSIMKEQKNLGF NITKLREVMLDRKDMSEIRKNHKVFDSI RTKVYTMMDFVIYRYYIEEDAKVAAAN KSLPDNEKSLSEKDIFVINLRGSFNDDQK DALYYDEANRIWRKLENIMHNIKEFRGN KTREYKKKDAPRLPRILPAGRDVSAFSK LMYALTMFLDGKEINDLLTTLINKFDNI QSFLKVMPLIGVNAKFVEEYAFFKDSAK IADELRLIKSFARMGEPIADARRAMYIDA *IRILG*GGGGSGGGGS*MRNFI*INNVISNKRF HYLIRYGDPAHLHEIAKNEAVVKFVLGR IADIQKKQGQNGKNQIDRYYETCIGKDK GKSVSEKVDALTKIITGMNYDQFDKKRS VIEDTGRENAEREKFKKIISLYLTVIYHIL KNIVNINARYVIGFHCVERDAQLYKEKG YDINLKKLEEKGFSSVTKLCAGIDETAPD KRKDVEKEMAERAKESIDSLESANPKLY ANYIKYSDEKKAEEFTRQINREKAKTAL NAYLRNTKWNVIIREDLLRIDNKTCTLF RNKAVHLEVARYVHAYINDIAEVNSYF QLYHYIMQRIIMNERYEKSSGKVSEYFD AVNDEKKYNDRLLKLLCVPFGYCIPRFK NLSIEAL*FDRNEGSG**PK***KKRKV*AAAYPY DVPDYA (SEQ ID NO: 282)	
NLS-RfxCas13d (del5.13)-NLS-HA	MS***PKKKRKV***EASIEKKKSFAKGMGVKS TLVSGSKVYMTTFAEGSDARLEKIVEGD SIRSVNEGEAFSAEMADKNAGYKIGNAK FSHPKGYAVVANNPLYTGPVQQDMLGL KETLEKRYFGESADGNDNICIQVIHNILD IEKILAEYITNAAYAVNNISGLDKDIIGFG KFSTVYTYDEFKDPEHHRAAFNNNDKLI NAIKAQYDEFDNFLDNPRLGYFGQAFFS KEGRNYIINYGNECYDILALLSGLRHWV VHNNEEESRISRTWLYNLDKNLDNEYIS TLNYLYDRITNELTNSFSKNSAANVNYI AETLGINPAEFAEQYFRFSIMKEQKNLGF NITKLREVMLDRKDMSEIRKNHKVFDSI RTKVYTMMDFVIYRY*YIEEDSEKDI*FVIN LRGSFNDDQKDALYYDEANRIWRKLEN IMHNIKEFRGNKTREYKKKDAPRLPRILP AGRDVSAFSKLMYALTMFLDGKEINDL LTTLINKFDNIQSFLKVMPLIGVNAKFVE EYAFFKDSAKIADELRLIKSFARMGEPIA DARRAMYIDA*IRILG*GGGGSGGGGS*MRN FI*INNVISNKRFHYLIRYGDPAHLHEIAK NEAVVKFVLGRIADIQKKQGQNGKNQI DRYYETCIGKDKGKSVSEKVDALTKIIT GMNYDQFDKKRSVIEDTGRENAEREKF KKIISLYLTVIYHILKNIVNINARYVIGFH CVERDAQLYKEKGYDINLKKLEEKGFSS VTKLCAGIDETAPDKRKDVEKEMAERA KESIDSLESANPKLYANYIKYSDEKKAEE	Amino acids 376-392 and 558-587 deleted

	FTRQINREKAKTALNAYLRNTKWNVIIR EDLLRIDNKTCTLFRNKAVHLEVARYVH AYINDIAEVNSYFQLYHYIMQRIIMNERY EKSSGKVSEYFDAVNDEKKYNDRLLKL LCVPFGYCIPRFKNLSIEALFDRNEAAKF DKEKKKVSGNSGSG**PKKKRKV**AAAYP YDVPDYA (SEQ ID NO: 283)	
NLS-RfxCas13d (del5.12+5.13)-NLS-HA	MS**PKKKRKV**EASIEKKKSFAKGMGVKS TLVSGSKVYMTTFAEGSDARLEKIVEGD SIRSVNEGEAFSAEMADKNAGYKIGNAK FSHPKGYAVVANNPLYTGPVQQDMLGL KETLEKRYFGESADGNDNICIQVIHNILD IEKILAEYITNAAYAVNNISGLDKDIIGFG KFSTVYTYDEFKDPEHHRAAFNNNDKLI NAIKAQYDEFDNFLDNPRLGYFGQAFFS KEGRNYIINYGNECYDILALLSGLRHWV VHNNEEESRISRTWLYNLDKNLDNEYIS TLNYLYDRITNELTNSFSKNSAANVNYI AETLGINPAEFAEQYFRFSIMKEQKNLGF NITKLREVMLDRKDMSEIRKNHKVFDSI RTKVYTMMDFVIYRY*YIEEDSEKDI*FVIN LRGSFNDDQKDALYYDEANRIWRKLEN IMHNIKEFRGNKTREYKKKDAPRLPRILP AGRDVSAFSKLMYALTMFLDGKEINDL LTTLINKFDNIQSFLKVMPLIGVNAKFVE EYAFFKDSAKIADELRLIKSFARMGEPIA DARRAMYIDA*IRILG*GGGGSGGGGS*MRN FI*INNVISNKRFHYLIRYGDPAHLHEIAK NEAVVKFVLGRIADIQKKQGQNGKNQI DRYYETCIGKDKGKSVSEKVDALTKIIT GMNYDQFDKKRSVIEDTGRENAEREKF KKIISLYLTVIYHILKNIVNINARYVIGFH CVERDAQLYKEKGYDINLKKLEEKGFSS VTKLCAGIDETAPDKRKDVEKEMAERA KESIDSLESANPKLYANYIKYSDEKKAEE FTRQINREKAKTALNAYLRNTKWNVIIR EDLLRIDNKTCTLFRNKAVHLEVARYVH AYINDIAEVNSYFQLYHYIMQRIIMNERY EKSSGKVSEYFDAVNDEKKYNDRLLKL LCVPFGYCIPRFKNLSIEAL*FDRNEGSG**P*** ***K*KKRKV**AAAYPYDVPDYA (SEQ ID NO: 284)	Amino acids 376-392, 558-587, and 953-966 deleted
NLS-RfxCas13d (del13)-NLS-HA	MS**PKKKRKV**EASIEKKKSFAKGMGVKS TLVSGSKVYMTTFAEGSDARLEKIVEGD SIRSVNEGEAFSAEMADKNAGYKIGNAK FSHPKGYAVVANNPLYTGPVQQDMLGL KETLEKRYFGESADGNDNICIQVIHNILD IEKILAEYITNAAYAVNNISGLDKDIIGFG KFSTVYTYDEFKDPEHHRAAFNNNDKLI NAIKAQYDEFDNFLDNPRLGYFGQAFFS KEGRNYIINYGNECYDILALLSGLRHWV VHNNEEESRISRTWLYNLDKNLDNEYIS TLNYLYDRITNELTNSFSKNSAANVNYI AETLGINPAEFAEQYFRFSIMKEQKNLGF NITKLREVMLDRKDMSEIRKNHKVFDSI RTKVYTMMDFVIYRY*YIEEDSEKDI*FVIN	Amino acids 376-392 deleted

	LRGSFNDDQKDALYYDEANRIWRKLEN IMHNIKEFRGNKTREYKKKDAPRLPRILP AGRDVSAFSKLMYALTMFLDGKEINDL LTTLINKFDNIQSFLKVMPLIGVNAKFVE EYAFFKDSAKIADELRLIKSFARMGEPIA DARRAMYIDAIRILGTNLSYDELKALAD TFSLDENGNKLKKGKHGMRNFIINNVIS NKRFHYLIRYGDPAHLHEIAKNEAVVKF VLGRIADIQKKQGQNGKNQIDRYYETCI GKDKGKSVSEKVDALTKIITGMNYDQF DKKRSVIEDTGRENAEREKFKKIISLYLT VIYHILKNIVNINARYVIGFHCVERDAQL YKEKGYDINLKKLEEKGFSSVTKLCAGI DETAPDKRKDVEKEMAERAKESIDSLES ANPKLYANYIKYSDEKKAEEFTRQINRE KAKTALNAYLRNTKWNVIIREDLLRIDN KTCTLFRNKAVHLEVARYVHAYINDIAE VNSYFQLYHYIMQRIIMNERYEKSSGKV SEYFDAVNDEKKYNDRLLKLLCVPFGY CIPRFKNLSIEALFDRNEAAKFDKEKKK VSGNSGSG**PKKKRKV**AAAYPYDVPDY A (SEQ ID NO: 285)	

293FT cells were cultured, transiently transfected, and three days later, CD81 expression was detected using flow cytometry, as described in Example 19.

As shown in FIG. 21, Cas13d orthologs CasRx and Ga0531 containing deletions in low-conserved regions maintain knock-down activity in human HEK 293FT cells. Based on these results, similar deletions can be made to the other Cas13d orthologs provided herein using the sequence alignments provided herein.

Example 22

dCas13d Mediated Splicing

The ability of several Cas13d orthologs (native and truncated variants, see Example 21) that exhibited high levels of protein knock-down (see Examples 19-20) to direct splicing was tested as follows. Cas13d orthologs (with mutated HEPN domains to abolish target RNA cleavage) were targeted to the slice acceptor site of a bichromatic blue-to-red splice reporter to mediate exon exclusion (as measured by reduction in mTagBFP2 expression). 293FT cells were cultured, transiently transfected, and three days later, CD81 expression was detected using flow cytometry, as described in Example 19.

As shown in FIG. 22, even for orthologs with high levels of endogenous protein knock down (>70% CD81 protein knock down), splicing modulation was in some cases inefficient. However, dCasRx (SEQ ID NO: 280), dCasRx-del13 (SEQ ID NO: 285), and dGa7274 (SEQ ID NO: 198), all exhibited good splicing activity. These three sequences are comparatively closely related and share >70% sequence identity. Thus, Cas13d orthologs, such as CasRx and truncated CasRx variants such as CasRx-del13, can be used to modulate splicing.

Example 23

RNA Editing using Cas13d

This example describes the use of Cas13d to edit RNA expression in human cells.

Cell culture

Human embryonic kidney (HEK) cell line 293FT (Thermo Fisher) was maintained in DMEM (4.5 g/L glucose), supplemented with 10% FBS (GE Life Sciences) and 10 mM HEPES at 37°C with 5% CO_2. Upon reaching 80-90% confluency, cells were dissociated using TrypLE Express (Life Technologies) and passaged at a ratio of 1:2.

Cloning of dCas13d-ADAR mammalian expression constructs

As shown in FIG. 23A, dCas13b and dCas13d coding sequences were cloned into mammalian expression vectors with N- or C-terminal fusions to $ADAR2_{DD}$(T375G) as well as a C-terminal NLS or NES sequence (*e.g.,* see SEQ ID NO: 286). The EGFP reporter used to assess A-to-I RNA editing by dCas13-ADAR fusions was engineered to contain a premature stop codon (UAG) at tryptophan 57 (W57X) and was linked to mCherry by a P2A peptide linker. Guides targeting this EGFP reporter were designed with variable spacer lengths and with A-C mismatches at defined distances away from the Direct Repeat sequence (5' for RfxCas13d and 3' for PspCas13b).

Transient transfection of human cell lines

Prior to transient transfection, HEK 293FT cells were plated at a density of 20,000 cells per well in a 96-well plate and transfected at >90% confluence with 500 ng of total plasmid DNA with 40 ng of EGFP(W57X) reporter plasmid and varying molar ratios of dCas13d-ADAR vector to guide expression vector using Lipofectamine 2000 (Life Technologies) according to the manufacturer's protocol. Transfected cells were harvested 48 hours post-transfection for flow cytometry.

Flow cytometry

Cells were dissociated 48 hours post-transfection with TrypLE Express and resuspended in FACS Buffer (1X $DPBS^{-/-}$, 0.2% BSA, 2 mM EDTA). Flow cytometry was performed in 96-well plate format using a MACSQuant VYB (Miltenyi Biotec) and analyzed using FlowJo 10. Analysis is representative of BFP^+ (expressing dCas13d-ADAR-P2A-mTagBFP2) and $mCherry^+$ (expressing mCherry-P2A-EGFP(W47X)) populations, which were assessed for RNA editing by observing the % of cells expressing EGFP above baseline.

As shown in FIG. 23B, targeted A to I RNA editing of a premature stop codon using Cas13d in an GFP reporter reactivates GFP expression, while the non-targeting guide did not result in RNA editing. FIG. 23C provides quantification of Cas13d-mediated RNA editing efficiency based on the amount of dCas13d-ADAR vector transfected into the cells.

Example 24

Cas13-mediated Identification of Antisense Oligonucleotides

Discovery of effective antisense oligonucleotide (ASO)-binding sites can be a low-throughput and expensive process, due to the lack of genetic encoding, the laborious nature of ASO generation, and the frequent need for chemical modifications. The Cas13 enzymes and variants thereof, such as the Cas13d proteins and variants disclosed herein, can be used to simplify and accelerate the identification of ASO binding sites as the binding of a catalytically active or inactive Cas13/dCas13 can mimic the effect of ASO-mediated RNA cutting and/or binding.

In one example, libraries of Cas13 guide RNAs are generated, for example in well-by-well or pooled oligonucleotide format, and designed to target specific cis-elements of desired RNAs (*e.g.*, tiling the intron of pre-mRNA of a therapeutically important protein, such as dystrophin or SMN1/SMN2, or desired locations within a UTR or gene body to identify effective and site-accessible ASO binding sites). This method can be exploited for ASO lead finding, interrogation of *in vivo* RNA secondary structure, and the identification of known or novel splicing regulatory elements.

Example 25

Increasing Translation with Cas13

The Cas13d enzymes and variants disclosed herein can be used to increase translation of a target protein, by targeted manipulation of post-transcriptional regulatory processes. Cas13d-mediated inhibition (*e.g.*, via steric blocking, or other mechanisms) of upstream open reading frames (uORFs), can be used to achieve this manipulation. For example, Cas13d-mediated inhibition using the Cas13d enzymes and variants disclosed herein can serve as an alternative start site for translation of a human or mammalian gene, or translation-inhibiting motifs and elements within the UTR of a target gene. By inhibiting uORFs, decreased ribosomal competition and other mechanisms can increase in translation of a target gene from a canonical or desired start codon. Nearly half of human transcripts present uORFs (Barbosa *et al.*, *Plos Genetics* 9:e1003529, 2013), which can cause widespread reduction of protein expression. Similarly, by inactivating inhibitory elements within a 5' UTR, such as secondary structure or G-quadruplexes, translation can be increased.

This can be achieved by using a guide RNA to target Cas13d effectors, optionally fused to effector proteins that can increase translation inhibition (*e.g.*, the C-terminal domain MA3 domain of Pdcd4), to uORFs or translation-inhibitory elements. Said guide RNAs can also be delivered in a multiplexed fashion, either via multiple guide RNAs or a pre-gRNA array. Such a method can be used to increase the translation of therapeutically or biotechnologically useful proteins, such as LDLR in familial hypercholesterolemia, CFTR in cystic fibrosis, KCNJ11 in congenital hyperinsulinism, HBB in sickle cell disease, PRKAR1A, IRF6, and others (Barbosa *et al.*, *Plos Genetics* 9:e1003529, 2013; Calvo *et al.*, *PNAS* 106:7507-12, 2009).

Example 26

Synthetic Lethal Genetic Screens with Cas13

The Cas13d enzymes and variants disclosed herein can be used in screening methods to aid in the identification of therapeutic compounds, for example by miming a lead small molecule or inhibitory RNA (RNAi). Exploiting synthetic lethal interactions, such as for a tumor suppressor or oncogene, permits identification of compounds (*e.g.*, small molecule or inhibitory RNA) that selectively inhibit cancer cells or other cells of interest. Conventional high-throughput screening for the discovery of synthetic lethal genetic interactions with chemical compounds, CRISPR-Cas9, and RNAi all have limitations including unknown or multiple target genes (for chemical compounds), pleiotropic effect on off-target genes (chemical compounds and RNAi), and inappropriate dual loss-of-function (for CRISPR-Cas9 or other nuclease-based approaches).

Methods are provided for using the Cas13d enzymes and variants disclosed herein in oncology lead-finding to address these challenges in modeling the "second hit" for pharmacologically-oriented synthetic lethality, due to 1) its specificity for targeting a known target gene or set of genes and 2) its knockdown mechanism, which more accurately reflects the inhibitory mechanism of a small molecule lead. For example, Cas13d or other Cas13 enzymes can be introduced along with cognate guide RNAs into cells that carry a particular mutation (*e.g.,* KRAS mutant lung cancer cells, BRCA-null ovarian cancer cells, and many others). Phenotypic readouts of growth inhibition or cell loss can then be performed either in an arrayed format or a guide RNA-pooled genetic screen, followed by genetic analysis to identify the genes of interest.

References

Abudayyeh (2017). RNA targeting with CRISPR-Cas13 . Nature. 550:280-4.

Abudayye *et al.,* (2016). Science *353*, aaf5573.

Almeida *et al.,* (2012). Cell Rep *2*, 789-798.

Anantharaman *et al.,* (2013) Biol Direct *8*, 15.

Batra *et al.,* (2017). Cell *170*, 899-912 e810.

Birmingham *et al.,* (2006) Nat Methods *3*, 199-204.

Biswas *et al.,* (2016) Stem Cell Reports *7*, 316-324.

Bland *et al.,* (2007). BMC Bioinformatics *8*, 209.

Boeve, B.F., and Hutton, M. (2008). Arch Neurol *65*, 460-464.

Cheong, C.G., and Hall, T.M. (2006). Proc Natl Acad Sci U S A *103*, 13635-13639.

Chiriboga *et al.,* (2016). Neurology *86*, 890-897.

Chylinski *et al.*, (2013). RNA Biol *10*, 726-737.

Cox *et al.*, (2017). Science *358*, 1019-1027.

Deltcheva *et al.*, (2011). Nature *471*, 602-607.

Dobin *et al.*, (2013). Bioinformatics *29*, 15-21.

Doench *et al.*, . (2003). Genes Dev *17*, 438-442.

Doudna, J.A., and Charpentier, E. (2014). Science *346*, 1258096.

Du *et al.*, (2006). Mol Cell Proteomics *5*, 1033-1044.

East-Seletsky *et al.*, (2017). Mol Cell *66*, 373-383 e373.

East-Seletsky *et al.*, (2016). Nature *538*, 270-273.

Edgar, R.C. (2007). BMC Bioinformatics *8*, 18.

Fonfara *et al.*, (2016). Nature *532*, 517-521.

Gasiunas *et al.*, (2012). Proc Natl Acad Sci U S A *109*, E2579-2586.

Gilbert *et al.*, (2014). Cell *159*, 647-661.

Gilbert *et al.* (2013). Cell *154*, 442-451.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). Nucleic Acids Res *35*, W52-57.

Hammond, S.M., and Wood, M.J. (2011). Trends Genet *27*, 196-205.

Heidrich *et al.*,. (2015). Methods Mol Biol *1311*, 1-21.

Heinz *et al.*, (2010). Mol Cell *38*, 576-589.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Cell *157*, 1262-1278.

Jackson *et al.*, (2003). Nat Biotechnol *21*, 635-637.

Jinek *et al.*, (2012). Science *337*, 816-821.

Ka *et al.*, (2005) Alzheimer Dis Assoc Disord *19 Suppl 1*, S29-36.

Kim *et al.* (2017). Nat Commun *8*, 14500.

Liu *et al.,* (2017). Cell *168*, 121-134 e112.

Love, M.I., Huber, W., and Anders, S. (2014). Genome Biol *15*, 550.

Makarova *et al.* (2015). Nat Rev Microbiol *13*, 722-736.

Matera, A.G., and Wang, Z. (2014). Nat Rev Mol Cell Biol *15*, 108-121.

Naldini, L. (2015). Gene therapy returns to centre stage. Nature *526*, 351-360.

O'Connell *et al.,* (2014). Nature *516*, 263-266.

Orengo, J.P., Bundman, D., and Cooper, T.A. (2006). Nucleic Acids Res *34*, e148.

Peabody, D.S. (1993). EMBO J *12*, 595-600.

Ran *et al.,* (2015). Nature *520*, 186-191.

Samai *et al.,* (2015). Cell *161*, 1164-1174.

Schena *et al.,* (1995). Science *270*, 467-470.

Schoch *et al.,* (2016). I Neuron *90*, 941-947.

Shendure *et al.,* (2017). Nature 550:345-53.

Shmakov *et al.,* (2015) Mol Cell *60*, 385-397.

Sigoillot *et al.,* (2012). Nat Methods *9*, 363-366.

Smargon *et al.,* (2017). Mol Cell *65*, 618-630 e617.

Smith *et al.,* (2017). PLoS Biol *15*, e2003213.

Swiech *et al.,* (2015). Nat Biotechnol *33*, 102-106.

Treu *et al.,* (2016). Bioresour Technol *216*, 260-266.

van der Oos *et al.,* (2014).. Nat Rev Microbiol *12*, 479-492.

Wang *et al.,* (2015). Biomed Rep *3*, 152-158.

Wright, A.V., and Doudna, J.A. (2016). Nat Struct Mol Biol *23*, 876-883.

Yang *et al.,* (2010). J Proteome Res *9*, 1402-1415.

Yosef *et al.,* (2012). Nucleic Acids Res *40*, 5569-5576.

Zalatan *et al.,* (2015). Cell *160*, 339-350.

Zetsche *et al.,*. (2015). Cell *163*, 759-771.

Zetsche, *et al.,* (2017). A Survey of Genome Editing Activity for 16 Cpf1 orthologs. bioRxiv. doi: https://doi.org/10.1101/134015

Zhang *et al.,* (2016). M Nucleic Acids Res *44*, 1789-1799.

Zhang *et al., (*2013). RNeuron *78*, 785-798.

In view of the many possible embodiments to which the principles of the disclosure may be applied, it should be recognized that the illustrated embodiments are only examples of the invention and should not be taken as limiting the scope of the invention. Rather, the scope of the invention is defined by the following claims. We therefore claim as our invention all that comes within the scope and spirit of these claims.

2018320870 27 Mar 2025

We claim:

1. An isolated CRISPR-Cas system guide RNA (gRNA), or isolated nucleic acid encoding the gRNA, comprising one or more direct repeat (DR) sequences and one or more spacer sequences,

wherein the one or more DR sequences comprise:

one or more DR sequences comprising or consisting of the nucleic acid sequence of SEQ ID NO: 131,129, 130,132, 133, 134, 135, 136, 137, 148, 150, 151, 152, 154, 156, 157, 159, 161, 163, 165, 167, 169, 176, 178, 180, 182, 184, 186, 188, 190, 191, 192, 193, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252 or 254;

one or more DR sequences comprising a 5'-end truncation of 1-10 nucleotides of one or more nucleic acid sequences of SEQ ID NOs: 131, 129, 130, 132, 133, 134, 135, 136, 137, 148, 150, 151, 152, 154, 156, 157, 159, 161, 163, 165, 167, 169, 176, 178, 180, 182, 184, 186, 188, 190, 191, 192, 193, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, or 254; or

one or more DR sequences comprising at least 80% sequence identity to the nucleic acid sequence of SEQ ID NO: 131, 129, 130, 132, 133, 134, 135, 136, 137, 148, 150, 151, 152, 154, 156, 157, 159, 161, 165, 167, 169, 176, 178, 180, 182, 184, 186, 188, 190, 191, 192, 193, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, or 254,

wherein the one or more spacer sequences hybridize with one or more target RNA molecules,

wherein the gRNA is capable of forming a complex with a Cas13d protein, and

wherein the gRNA directs the complex to the one or more target RNA molecules.

2. The isolated gRNA according to claim 1, comprising:

one or more direct repeat (DR) sequences comprising at least 80% sequence identity to SEQ ID NO: 131,129, 130, 132, 133, 134, 135, 136, 137, 148, 150, 151, 152, 154, 156, 157, 159, 161, 165, 167, 169, 176, 178, 180, 182, 184, 186, 188, 190, 191, 192, 193, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252 or 254.

3. The isolated gRNA according to claim 1, comprising:

27 Mar 2025

2018320870

one or more DR sequences comprising a 5'-end truncation of 1-10 nucleotides of one or more nucleic acid sequences of SEQ ID NO: 131, 129, 130, 132, 133, 134, 135, 136, 137, 148, 150, 151, 152, 154, 156, 157, 159, 161, 163, 165, 167, 169, 176, 178, 180, 182, 184, 186, 188, 190, 191, 192, 193, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252 or 254.

4. The isolated gRNA according to claim 1, comprising:

one or more DR sequences comprising or consisting of the nucleic acid sequence of SEQ ID NO: 131,129, 130, 132, 133, 134, 135, 136, 137, 148, 150, 151, 152, 154, 156, 157, 159, 161, 163, 165, 167, 169, 176, 178, 180, 182, 184, 186, 188, 190, 191, 192, 193, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, or 254.

5. The isolated gRNA according to claim 1, or isolated nucleic acid encoding the gRNA, wherein the gRNA comprises:

one or more DR sequences comprising or consisting of the nucleic acid sequence of SEQ ID NO: 131;

one or more DR sequences comprising a 5'-end truncation of 1-10 nucleotides of the nucleic acid sequence of SEQ ID NO: 131; or

one or more DR sequences comprising at least 80% sequence identity to the nucleic acid sequence of SEQ ID NO: 131.

6. The isolated gRNA according to claim 1, or isolated nucleic acid encoding the gRNA, wherein the gRNA comprises:

one or more DR sequences comprising or consisting of the nucleic acid sequence of SEQ ID NO: 135;

one or more DR sequences comprising a 5'-end truncation of 1-10 nucleotides of the nucleic acid sequence of SEQ ID NO: 135; or

one or more DR sequences comprising at least 80% sequence identity to the nucleic acid sequence of SEQ ID NO: 135.

7. The isolated gRNA according to claim 1, or isolated nucleic acid encoding the gRNA, wherein the gRNA comprises:

one or more DR sequences comprising or consisting of the nucleic acid sequence of SEQ ID NO: 180;

2018320870 27 Mar 2025

one or more DR sequences comprising a 5'-end truncation of 1-10 nucleotides of the nucleic acid sequence of SEQ ID NO: 180; or

one or more DR sequences comprising at least 80% sequence identity to the nucleic acid sequence of SEQ ID NO: 180.

8. The isolated gRNA according to claim 1, or isolated nucleic acid encoding the gRNA, wherein the gRNA comprises:

one or more DR sequences comprising or consisting of the nucleic acid sequence of SEQ ID NO: 190;

one or more DR sequences comprising a 5'-end truncation of 1-10 nucleotides of the nucleic acid sequence of SEQ ID NO: 190; or

one or more DR sequences comprising at least 80% sequence identity to the nucleic acid sequence of SEQ ID NO: 190.

9. The isolated gRNA according to claim 1, or isolated nucleic acid encoding the gRNA, wherein the gRNA comprises:

one or more DR sequences comprising or consisting of the nucleic acid sequence of SEQ ID NO: 199;

one or more DR sequences comprising a 5'-end truncation of 1-10 nucleotides of the nucleic acid sequence of SEQ ID NO: 199; or

one or more DR sequences comprising at least 80% sequence identity to the nucleic acid sequence of SEQ ID NO: 199.

10. The isolated gRNA according to claim 1, or isolated nucleic acid encoding the gRNA, wherein the gRNA comprises:

one or more DR sequences comprising or consisting of the nucleic acid sequence of SEQ ID NO: 209;

one or more DR sequences comprising a 5'-end truncation of 1-10 nucleotides of the nucleic acid sequence of SEQ ID NO: 209; or

one or more DR sequences comprising at least 80% sequence identity to the nucleic acid sequence of SEQ ID NO: 209.

11. The isolated gRNA according to claim 1, or isolated nucleic acid encoding the gRNA, wherein the gRNA comprises:

one or more DR sequences comprising or consisting of the nucleic acid sequence of SEQ ID NO: 213;

one or more DR sequences comprising a 5'-end truncation of 1-10 nucleotides of the nucleic acid sequence of SEQ ID NO: 213; or

one or more DR sequences comprising at least 80% sequence identity to the nucleic acid sequence of SEQ ID NO: 213.

12. The isolated gRNA according to claim 1, comprising:
one or more DR sequences comprising or consisting of the nucleic acid sequence of SEQ: 131.

13. The isolated gRNA according to claim 1, comprising:
one or more DR sequences comprising or consisting of the nucleic acid sequence of SEQ: 135.

14. The isolated gRNA according to claim 1, comprising:
one or more DR sequences comprising or consisting of the nucleic acid sequence of SEQ: 180.

15. The isolated gRNA according to claim 1, comprising:
one or more DR sequences comprising or consisting of the nucleic acid sequence of SEQ: 190.

16. The isolated gRNA according to claim 1, comprising:
one or more DR sequences comprising or consisting of the nucleic acid sequence of SEQ: 199.

17. The isolated gRNA according to claim 1, comprising:
one or more DR sequences comprising or consisting of the nucleic acid sequence of SEQ: 209.

18. The isolated gRNA according to claim 1, comprising:
one or more DR sequences comprising or consisting of the nucleic acid sequence of SEQ: 213.

FIG. 1A

CRISPR system identification
Obtain bacterial and archaeal genome assemblies and scaffolds from NCBI WGS database
Identify interspaced repeat sequence locations
spacer
DR'
DR
0.6X DR-1X DR 23-55 bp
0.6-2.5X DR
21175 'CRISPRs'
Identify putative proteins in 20kb window around CRISPRs
Eliminate candidate arrays with proteins associated with known CRISPR subtypes
11441 'CRISPRs'
cluster of Cas proteins
CRISPR array
single effector?
Extract proteins >750 aa length and ± 5 proteins from CRISPR array
CRISPR effector identification
8977 proteins proximal to CRISPR array
all v. all BLASTp
single-linkage hierarchical clustering
408 clusters
cluster expansion: BLASTp each cluster against all proteins (-nr database)
sort clusters by member protein proximity to CRISPR array (>70% co-occurrence)
candidate novel single effector CRISPR proteins
expand protein family via tblastn of metagenome contigs, custom sequencing of genomic DNA fragments
Type VI Cas13d family

FIG. 1B

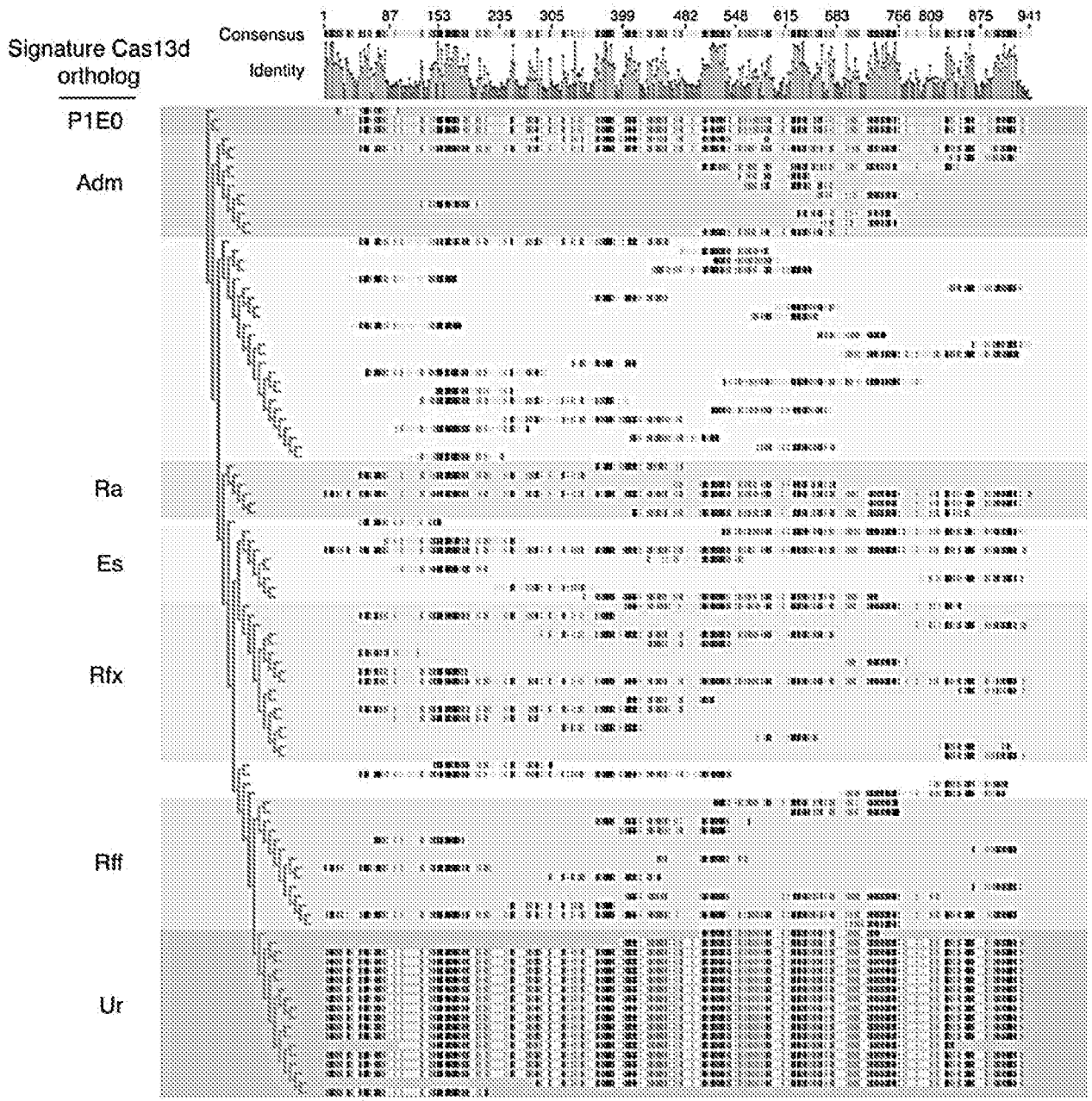

Signature Cas13d ortholog
Consensus
Identity
1
87
153
235
305
399
482
548
615
683
766
809
875
941
P1E0
Adm
Ra
Es
Rfx
Rff
Ur

FIG. 2A

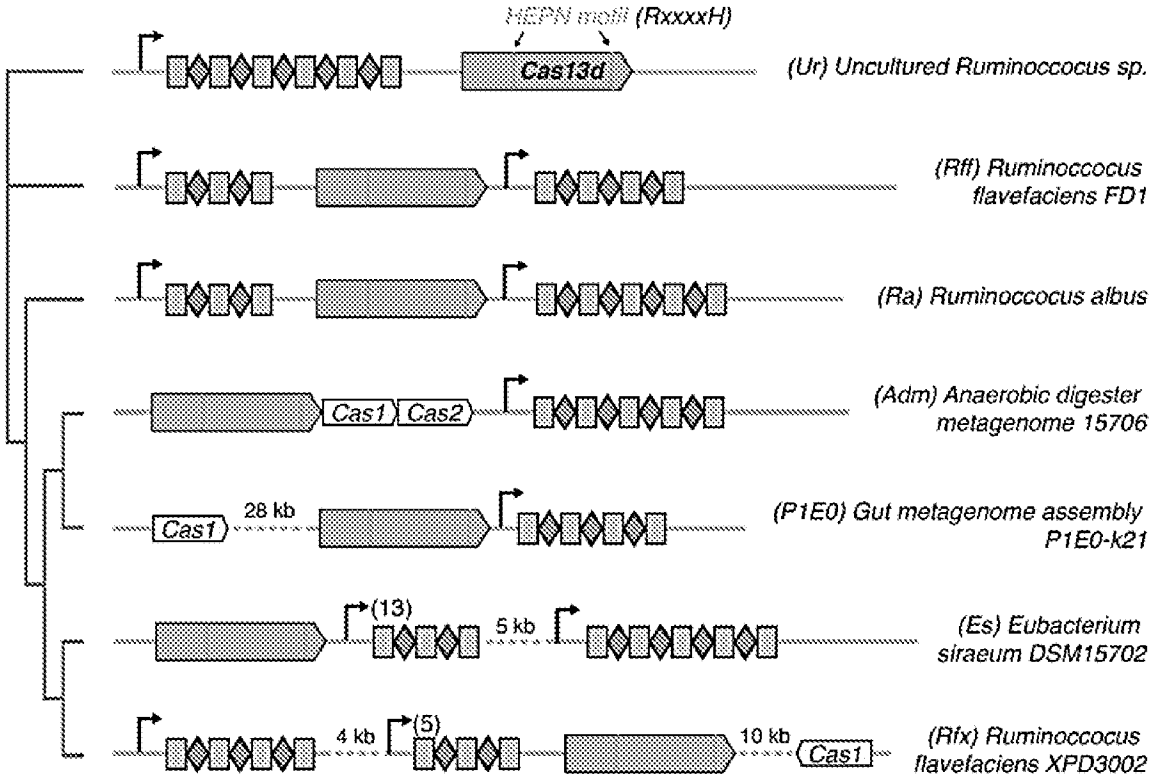


FIG. 2B

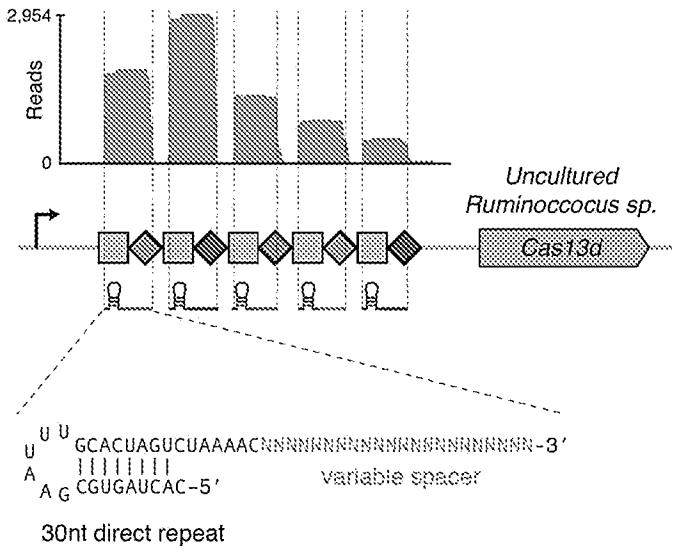


FIG. 2C

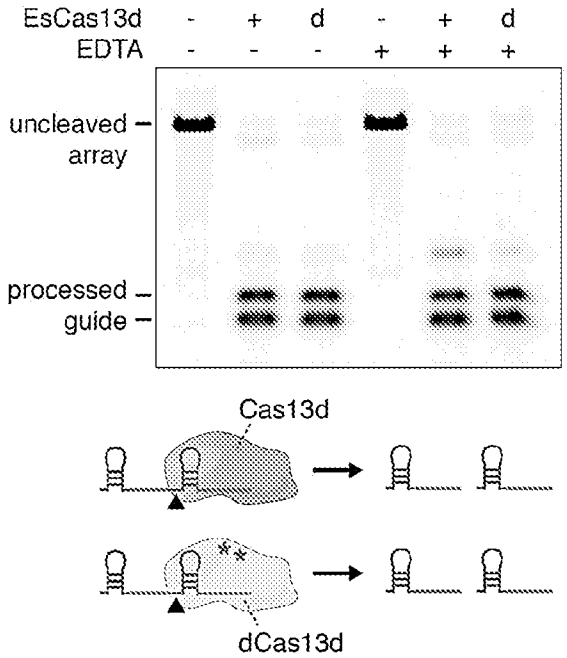


FIG. 3A

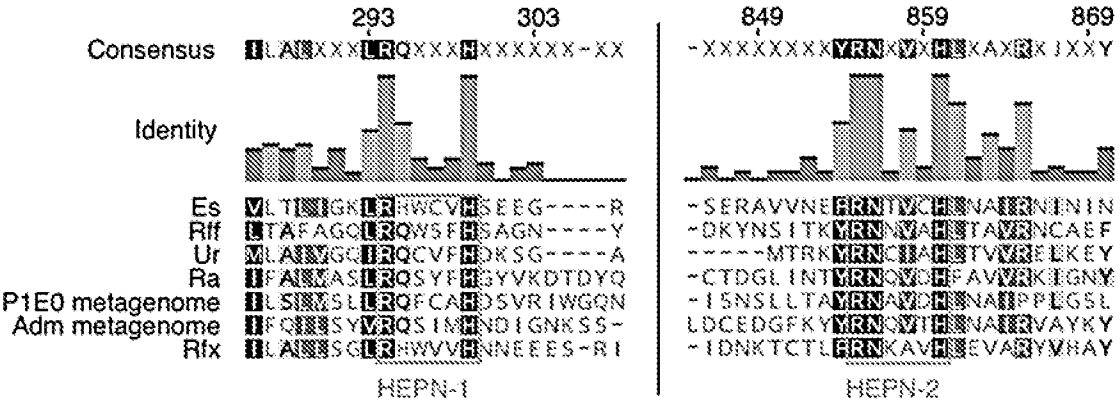
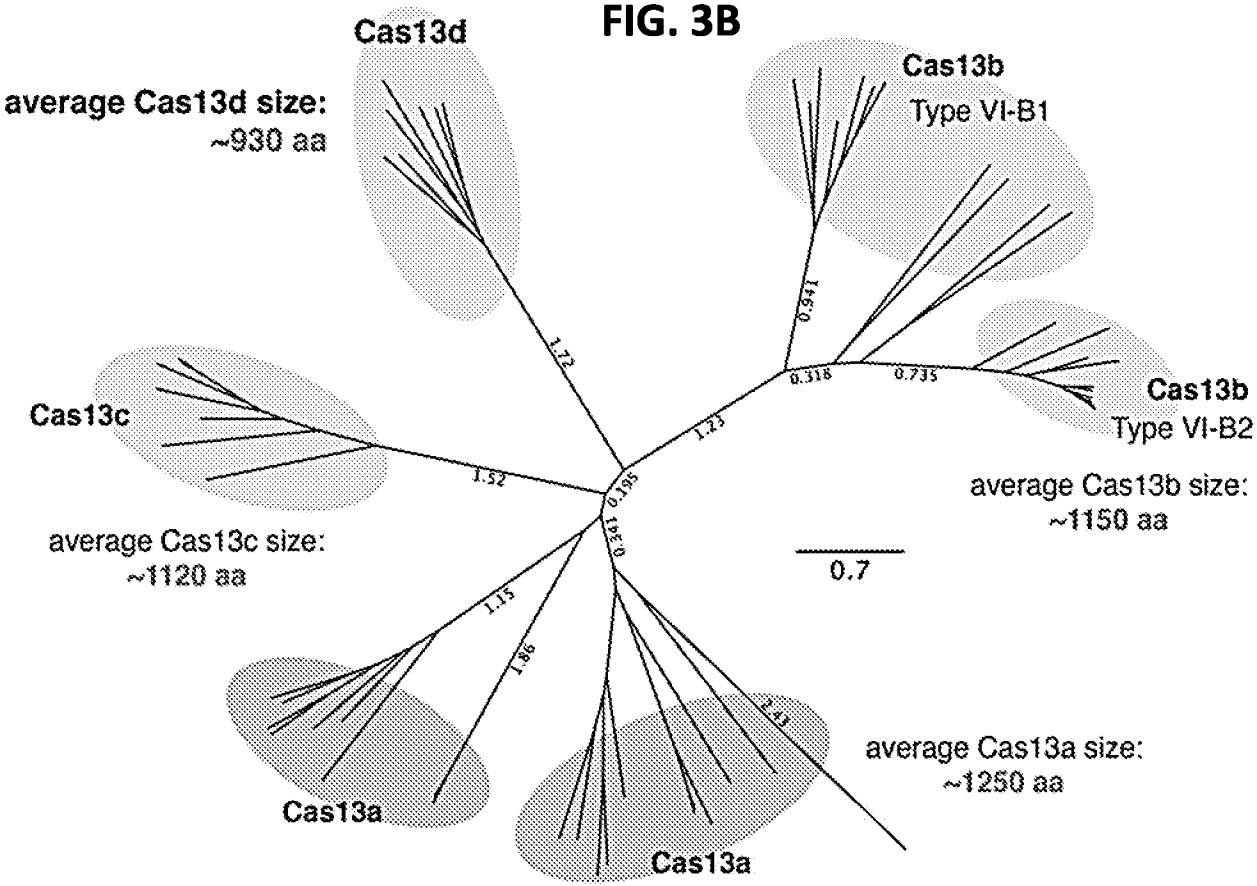

293
303
849
859
869
Consensus
Identity
Es
Rff
Ur
Ra
P1E0 metagenome
Adm metagenome
Rfx
HEPN-1
HEPN-2

FIG. 3B


Cas13d
average Cas13d size:
~930 aa
Cas13b
Type VI-B1
Cas13c
Cas13b
Type VI-B2
average Cas13b size:
~1150 aa
average Cas13c size:
~1120 aa
0.7
Cas13a
Cas13a
average Cas13a size:
~1250 aa

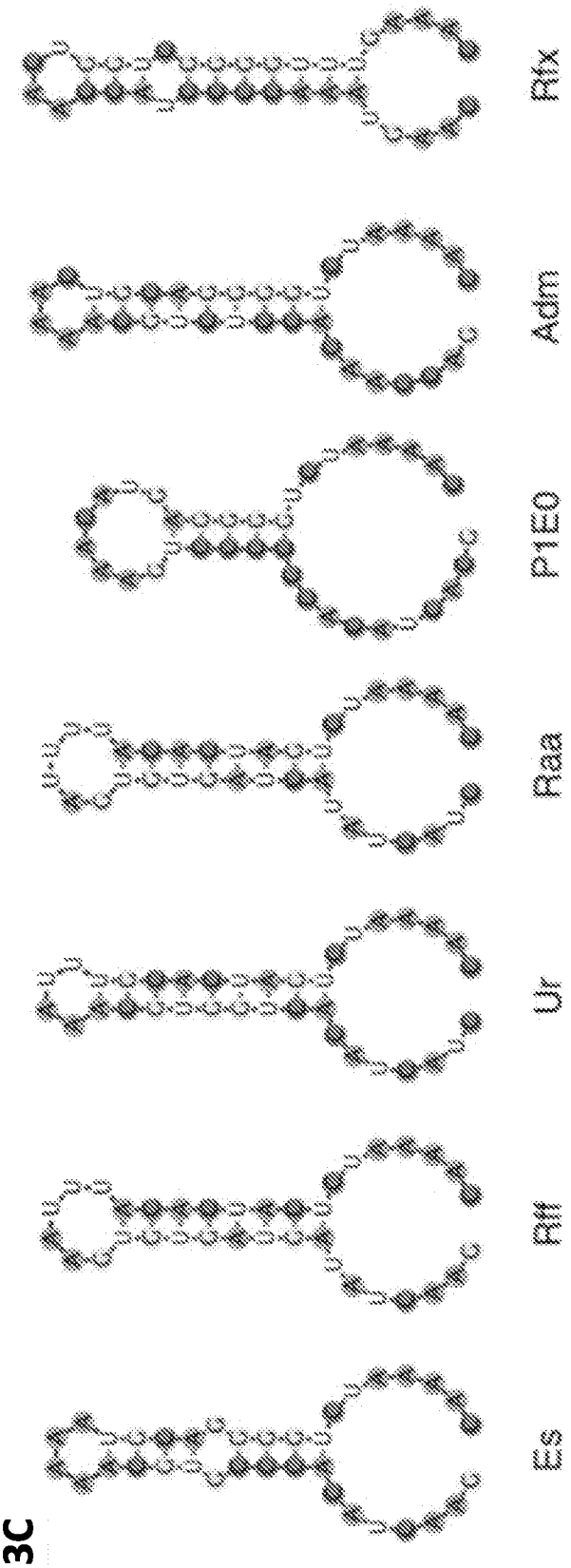

Es
Rff
Ur
Raa
P1E0
Adm
Rfx

FIG. 3C

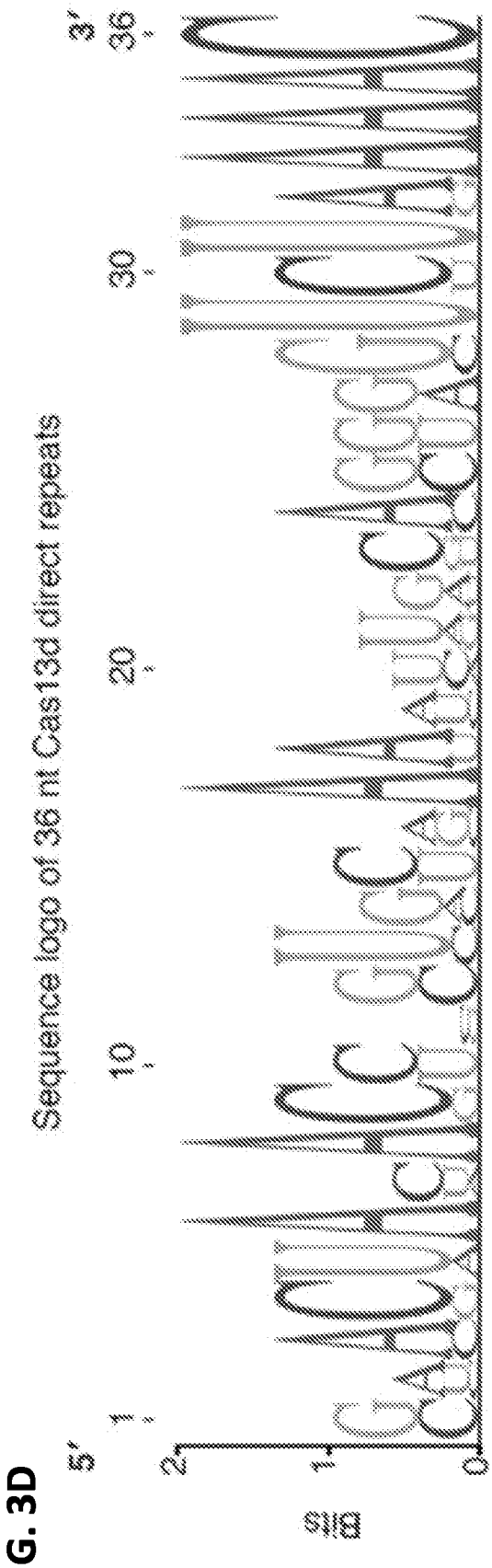

Sequence logo of 36 nt Cas13d direct repeats
5'
3'
1
10
20
30
36
Bits
2
1
0

FIG. 3D

FIG. 4A

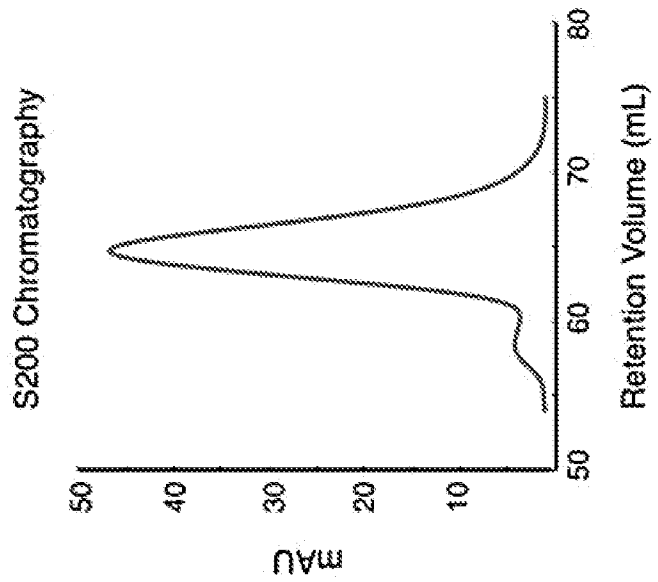


FIG. 4B

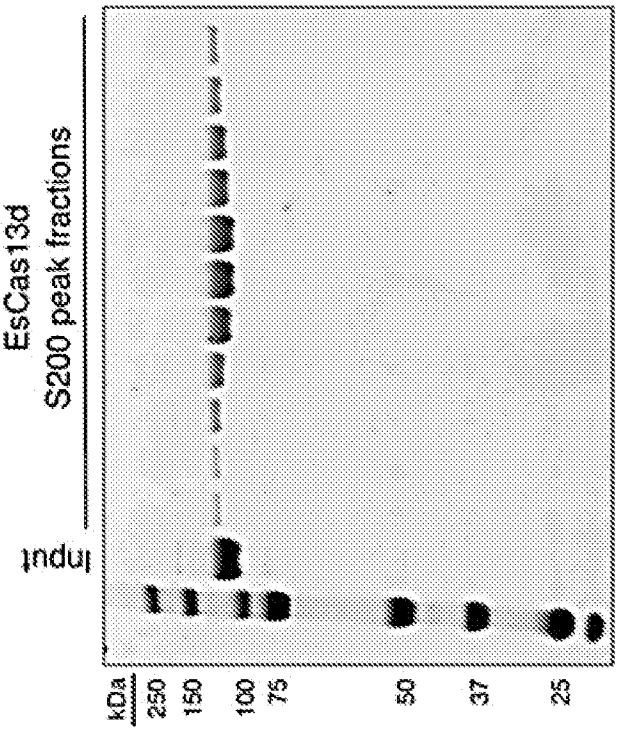


FIG. 4C

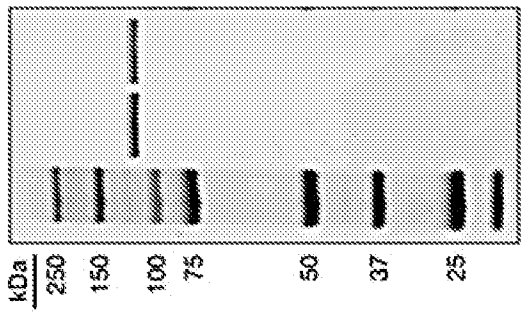


FIG. 5A

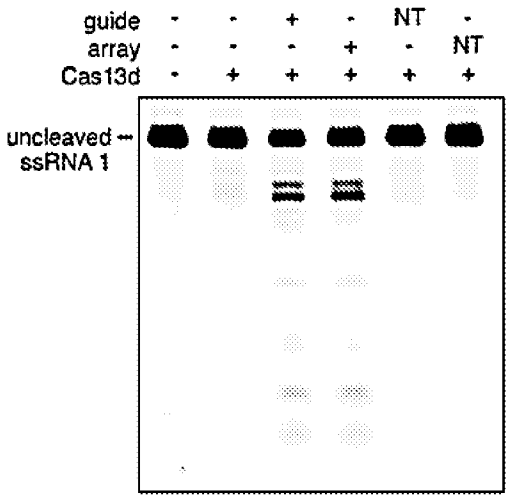


FIG. 5B

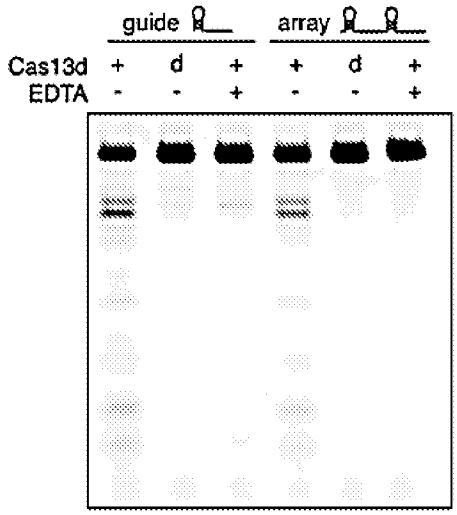


FIG. 5C

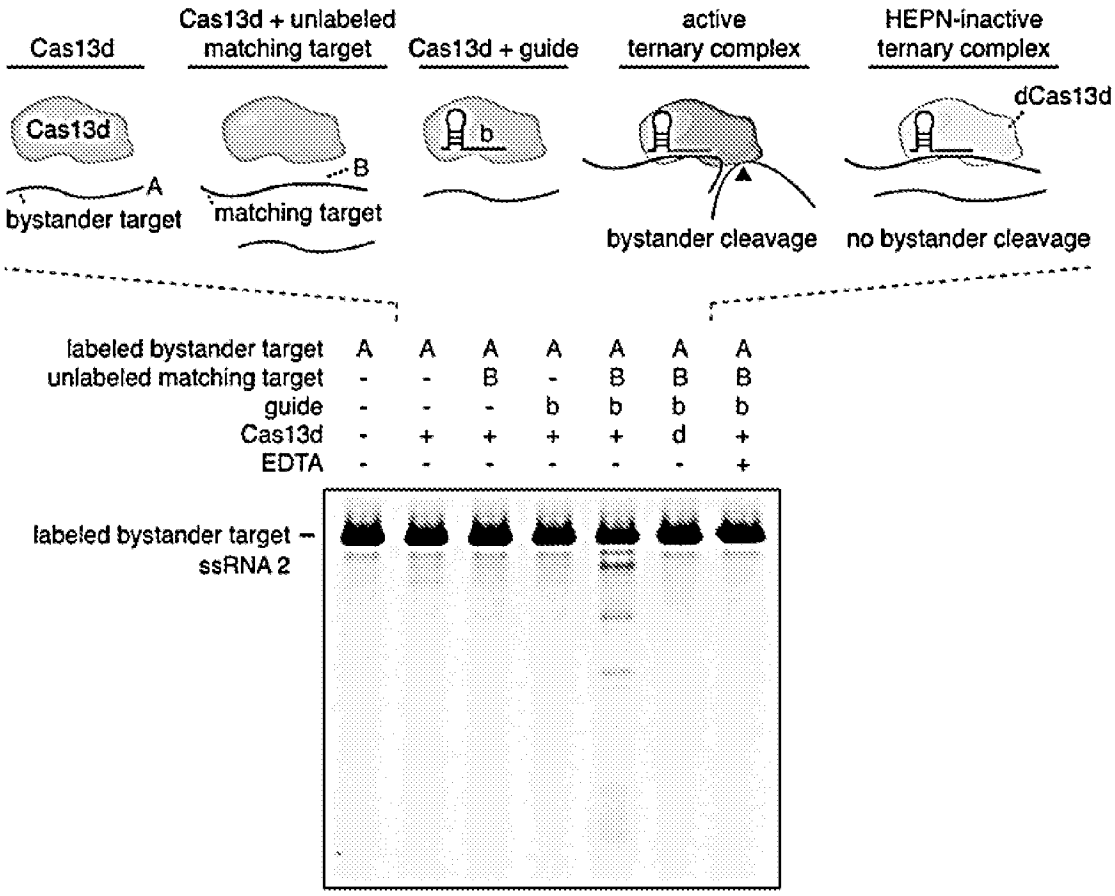


FIG. 5D

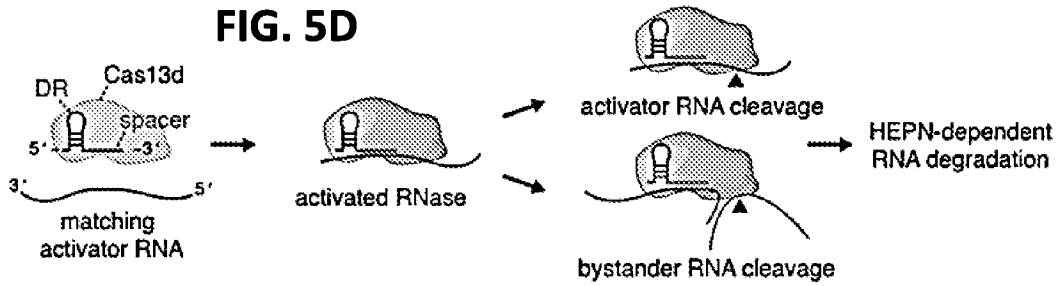


FIG. 6B

FIG. 6A

FIG. 6D

FIG. 6C

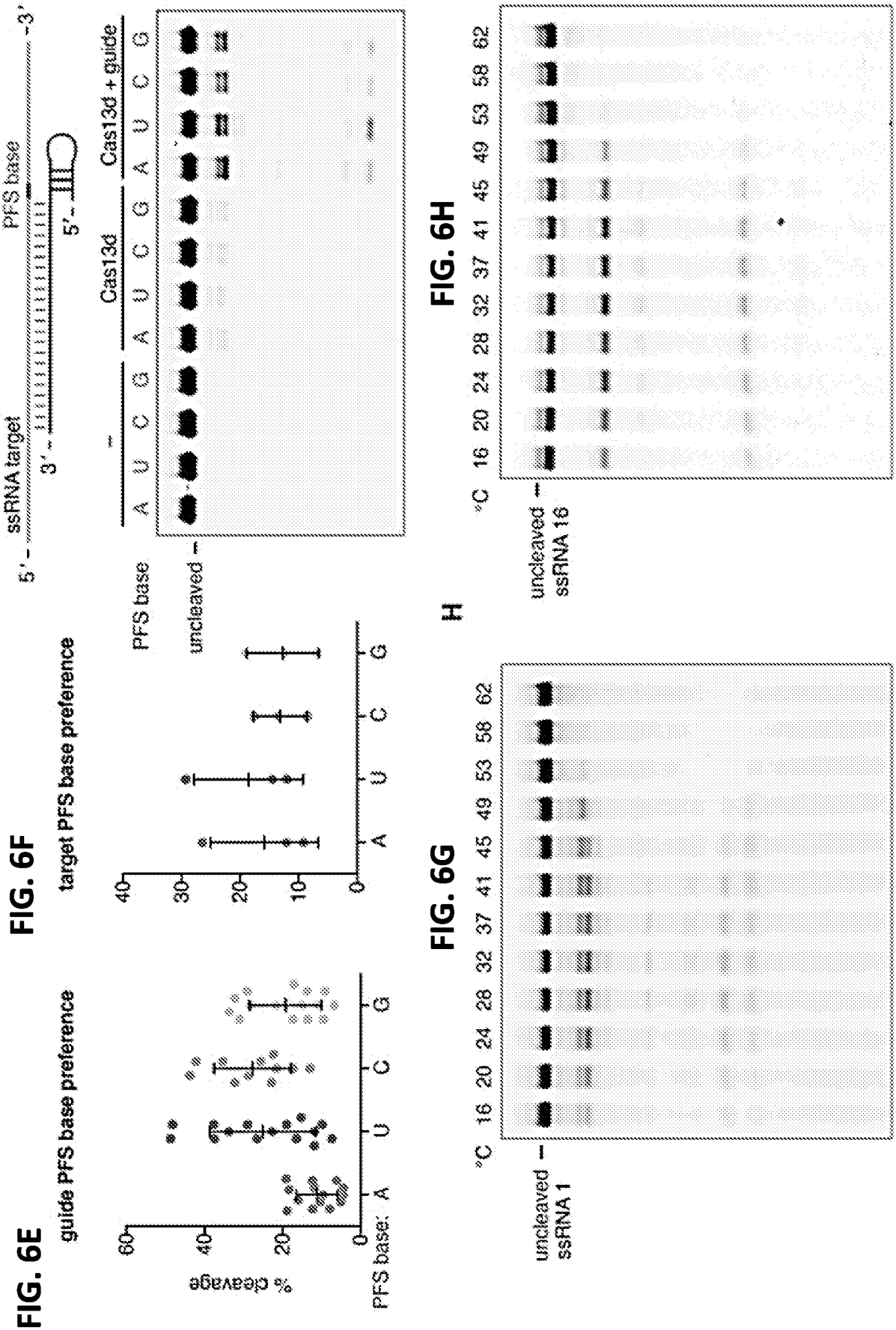


FIG. 6E
guide PFS base preference
% cleavage
PFS base: A U C G
FIG. 6F
target PFS base preference
PFS base
A U C G
FIG. 6G
°C 16 20 24 28 32 37 41 45 49 53 58 62
uncleaved ssRNA 1
FIG. 6H
°C 16 20 24 28 32 37 41 45 49 53 58 62
uncleaved ssRNA 16

FIG. 7A

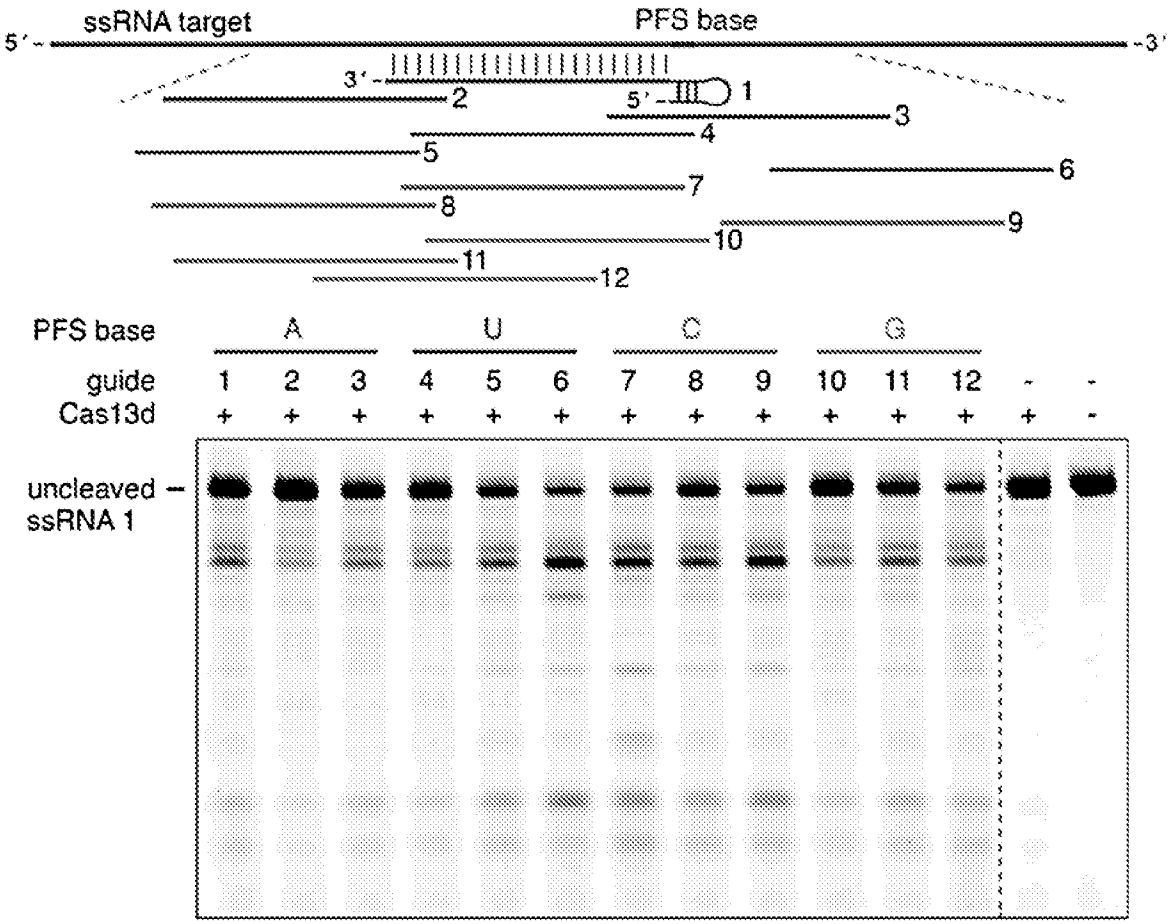
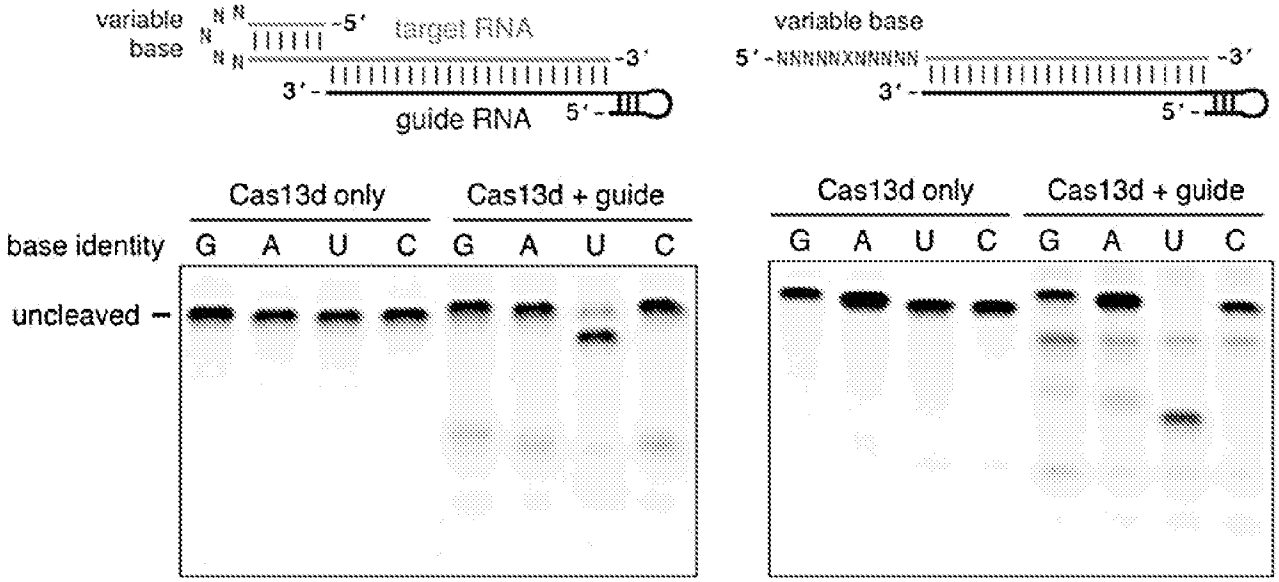

ssRNA target
PFS base
5'
3'
1
2
3
4
5
6
7
8
9
10
11
12
PFS base
A
U
C
G
guide
1 2 3 4 5 6 7 8 9 10 11 12 - -
Cas13d
+ + + + + + + + + + + + + -
uncleaved –
ssRNA 1

FIG. 7B


variable base
N N
5'
target RNA
3'
guide RNA
variable base
5' - NNNNNXNNNNN
3'
Cas13d only
Cas13d + guide
base identity
G A U C G A U C
uncleaved –
Cas13d only
Cas13d + guide
G A U C G A U C

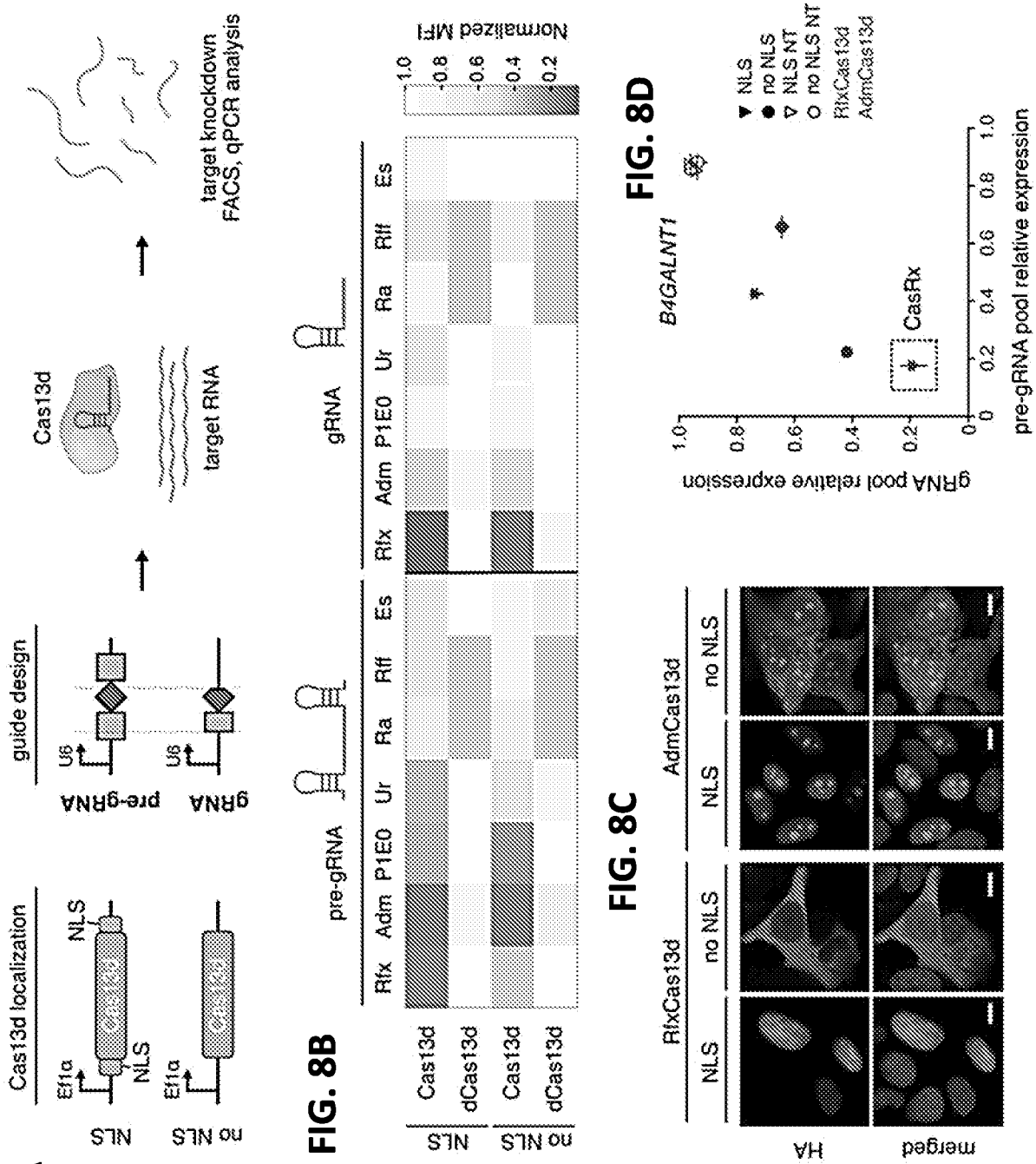


FIG. 8A

FIG. 8B

FIG. 8C

FIG. 8D

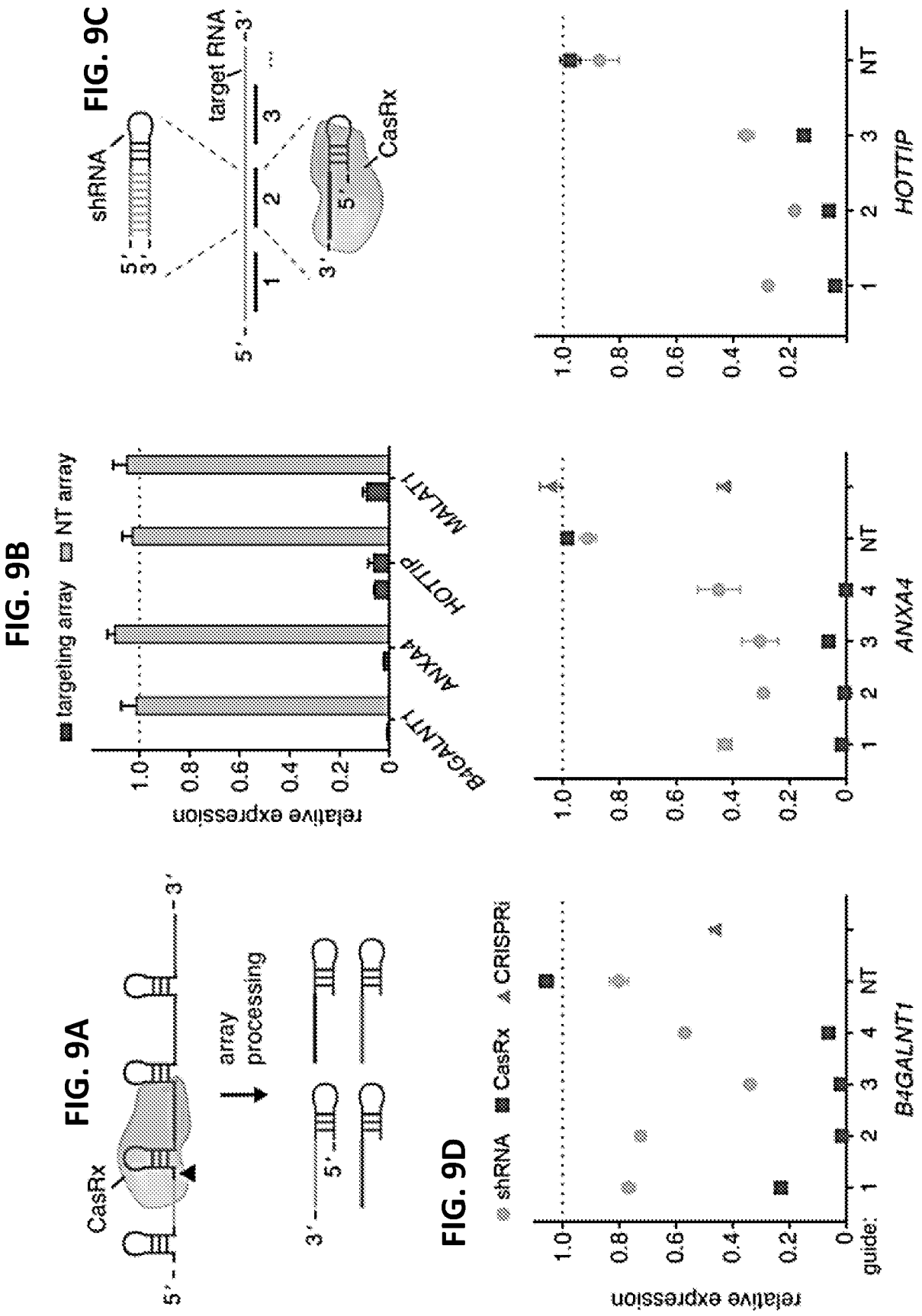


FIG. 9A
CasRx
5'
3'
array processing
3'
5'
FIG. 9B
targeting array
NT array
relative expression
1.0
0.8
0.6
0.4
0.2
0
B4GALNT1
ANXA4
HOTTIP
MALAT1
FIG. 9C
shRNA
5'
3'
target RNA
5'
3'
1
2
3
CasRx
FIG. 9D
shRNA
CasRx
CRISPRi
relative expression
guide:
1
2
3
4
NT
B4GALNT1
ANXA4
HOTTIP

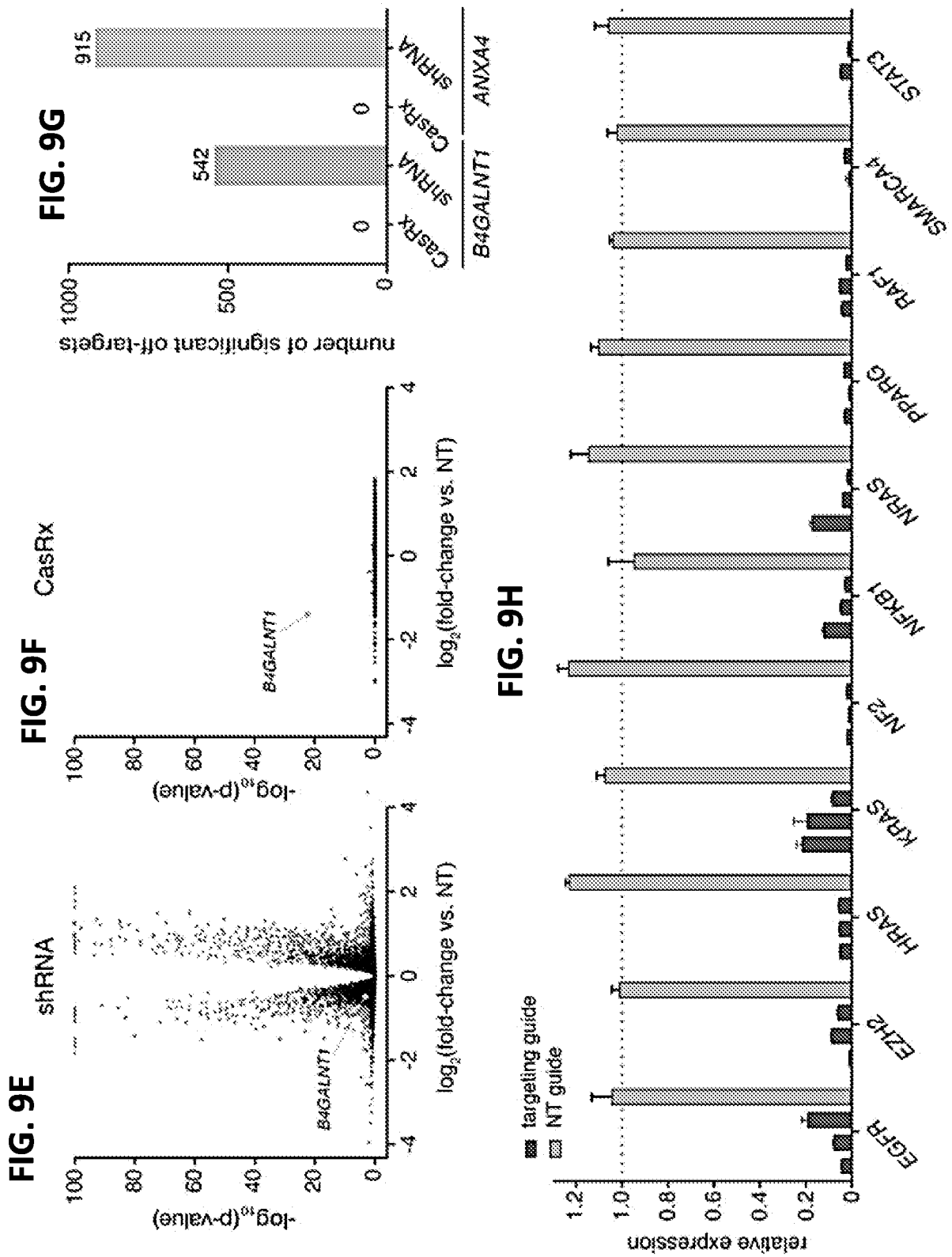


FIG. 9E
shRNA
-log10(p-value)
log2(fold-change vs. NT)
B4GALNT1
FIG. 9F
CasRx
-log10(p-value)
log2(fold-change vs. NT)
B4GALNT1
FIG. 9G
number of significant off-targets
915
542
0
0
0
CasRx shRNA CasRx shRNA
B4GALNT1
ANXA4
FIG. 9H
relative expression
targeting guide
NT guide
EGFR
EZH2
HRAS
KRAS
NF2
NFKB1
NRAS
PPARG
RAF1
SMARCA4
STAT3

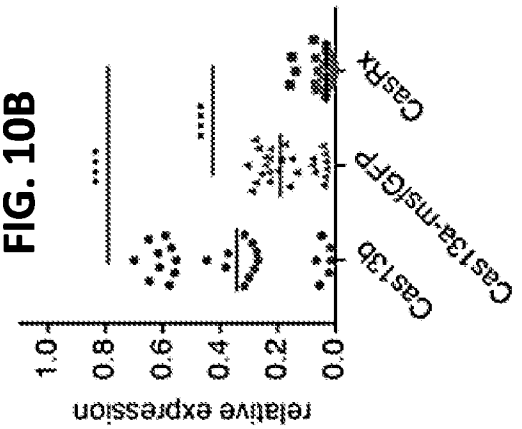


FIG. 10B
relative expression
Cas13b
Cas13a-msfGFP
CasRx

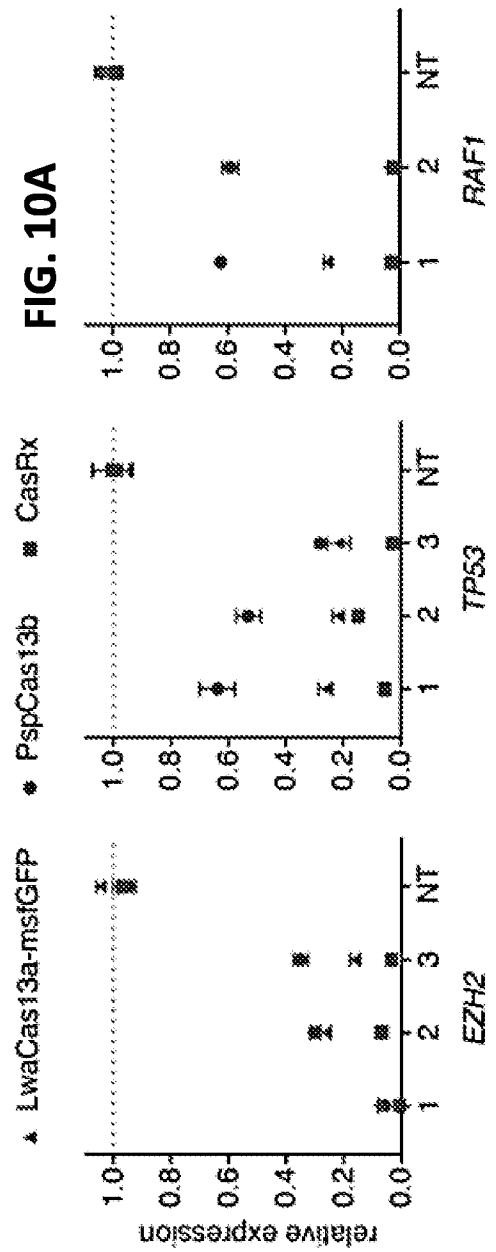


FIG. 10A
LwaCas13a-msfGFP
PspCas13b
CasRx
relative expression
EZH2
TP53
RAF1
NT

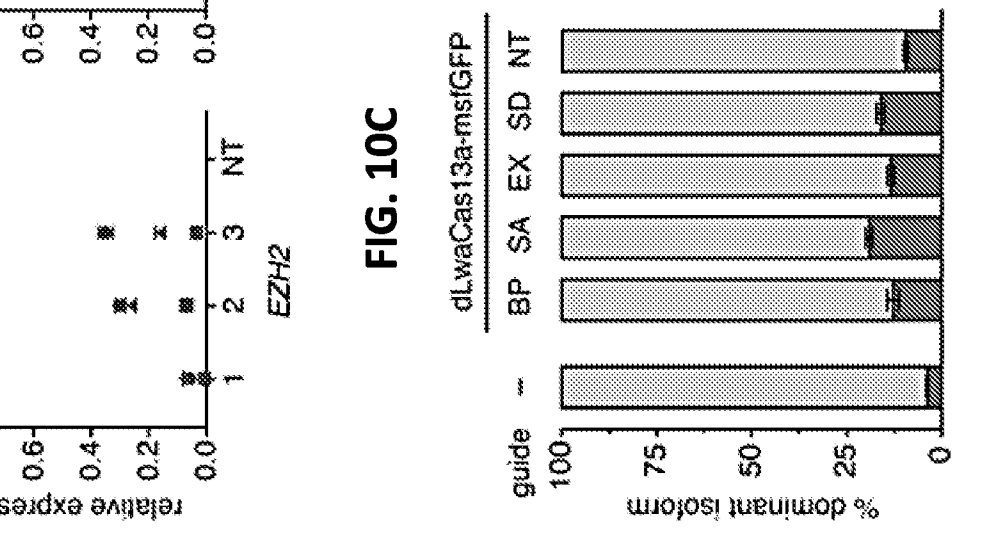


FIG. 10D
dCasRx
dCas13a-msfGFP
fold-change in exon exclusion
guide
BP
SA
EX
SD

FIG. 10C
dLwaCas13a-msfGFP
guide
BP
SA
EX
SD
NT
% dominant isoform

FIG. 11A

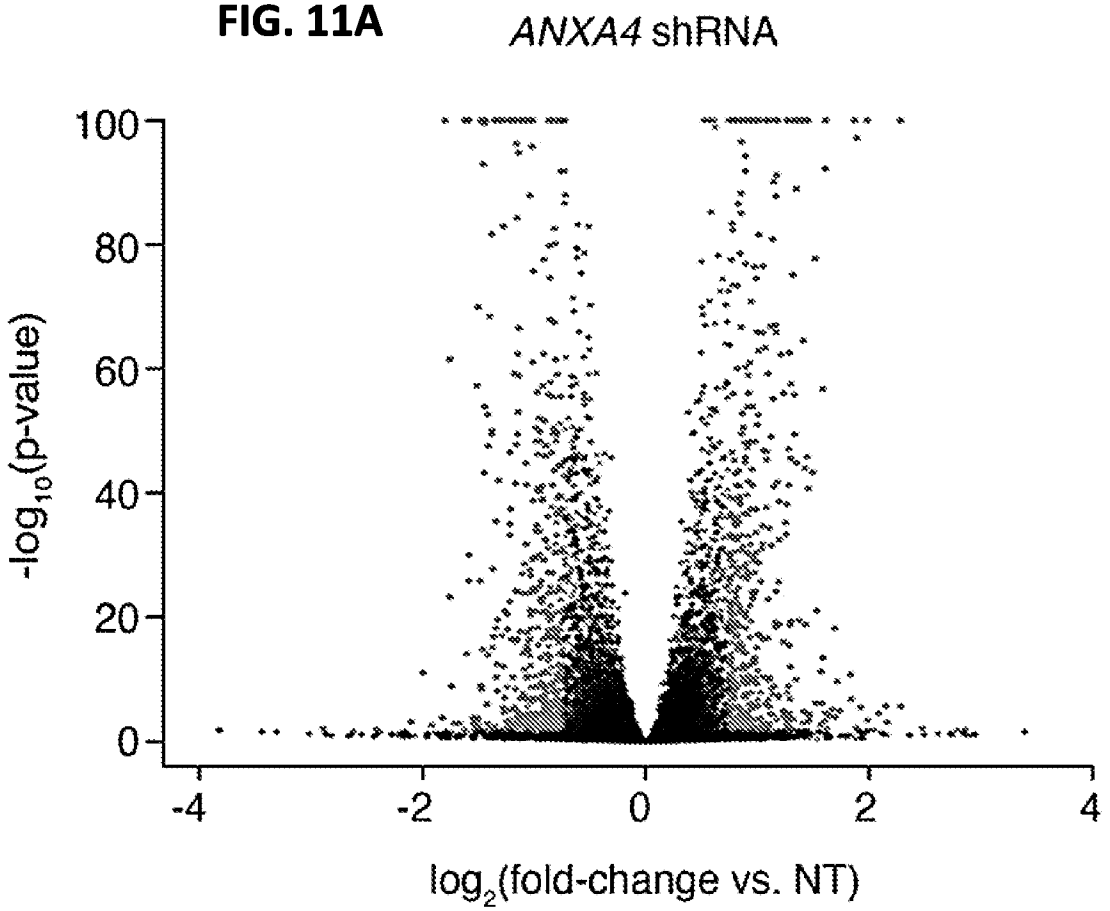


FIG. 11B

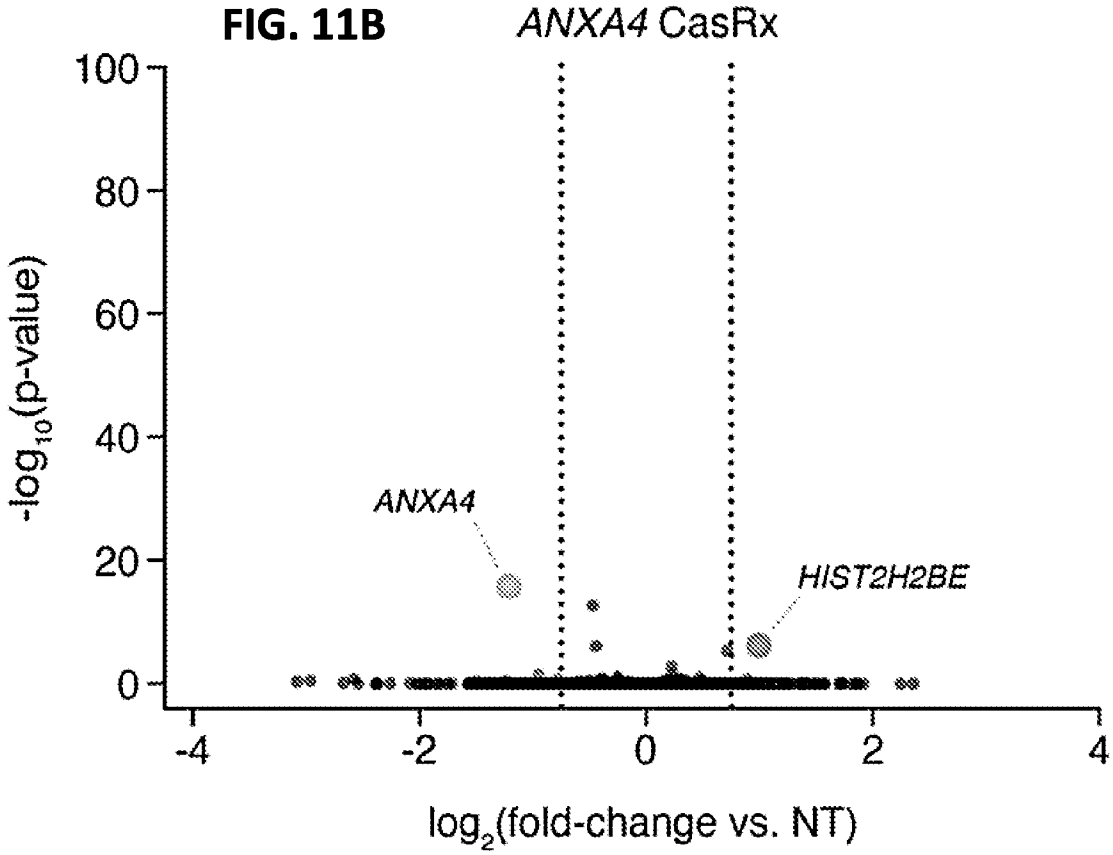


FIG. 12A

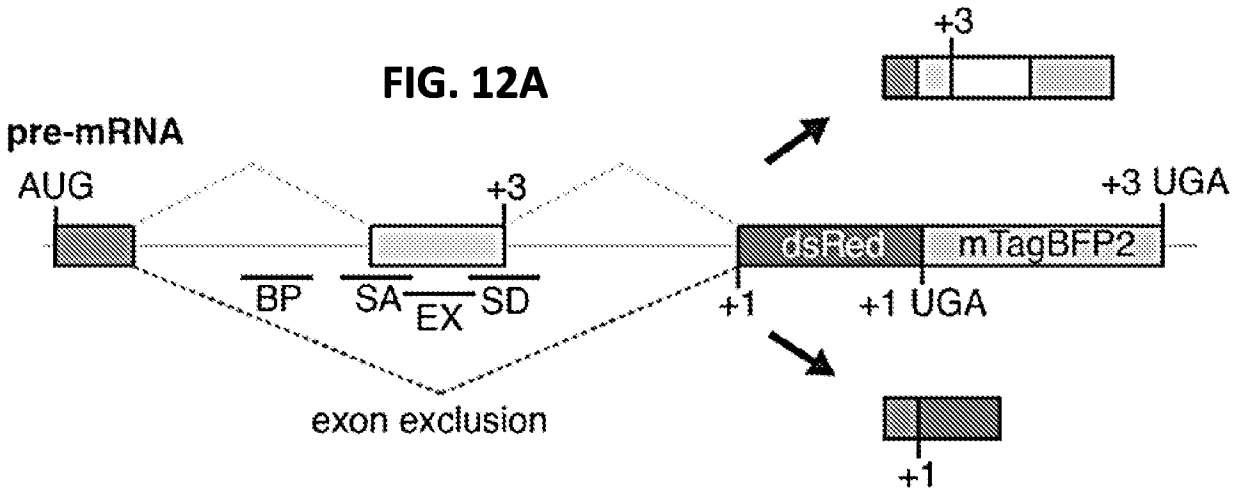


FIG. 12B

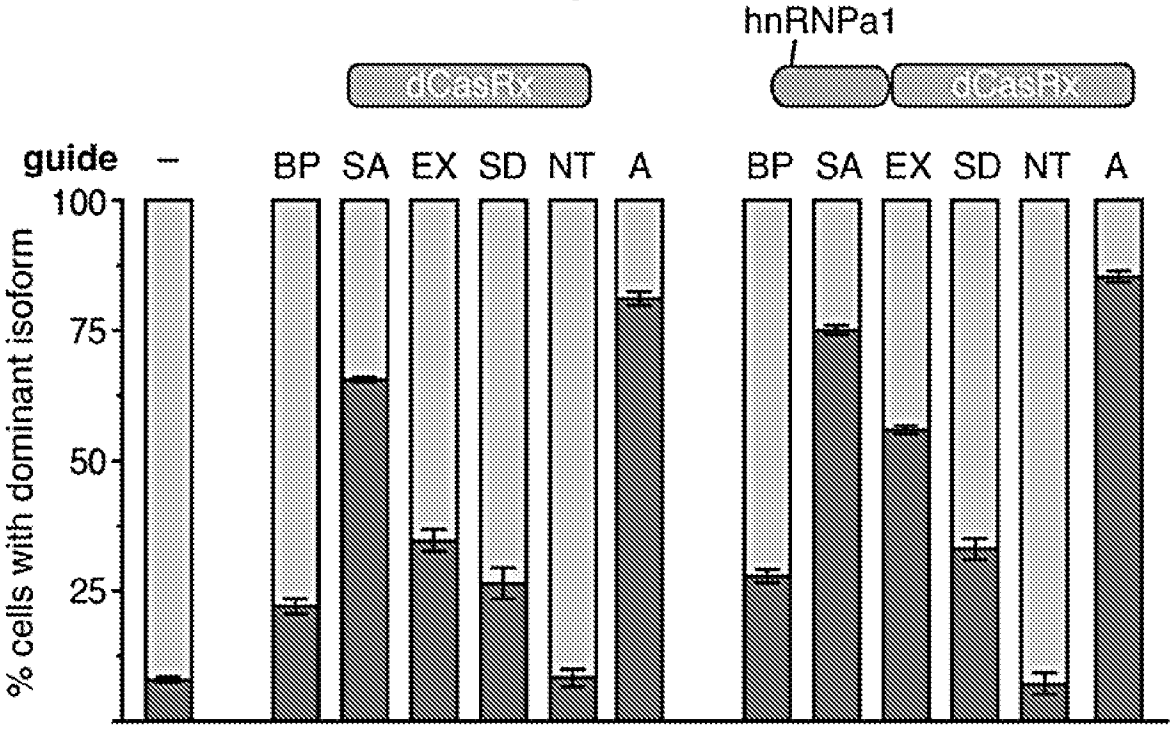
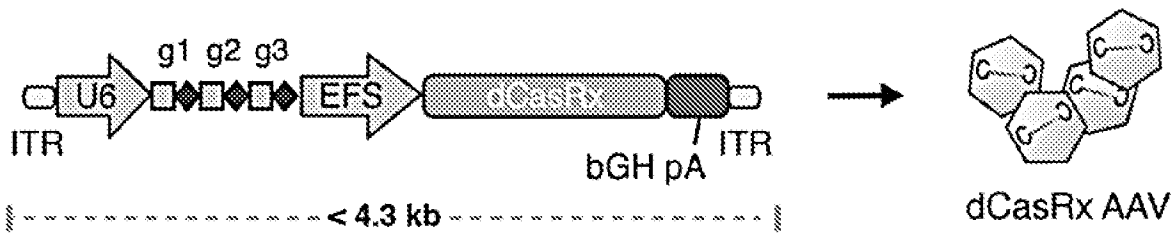


FIG. 12C



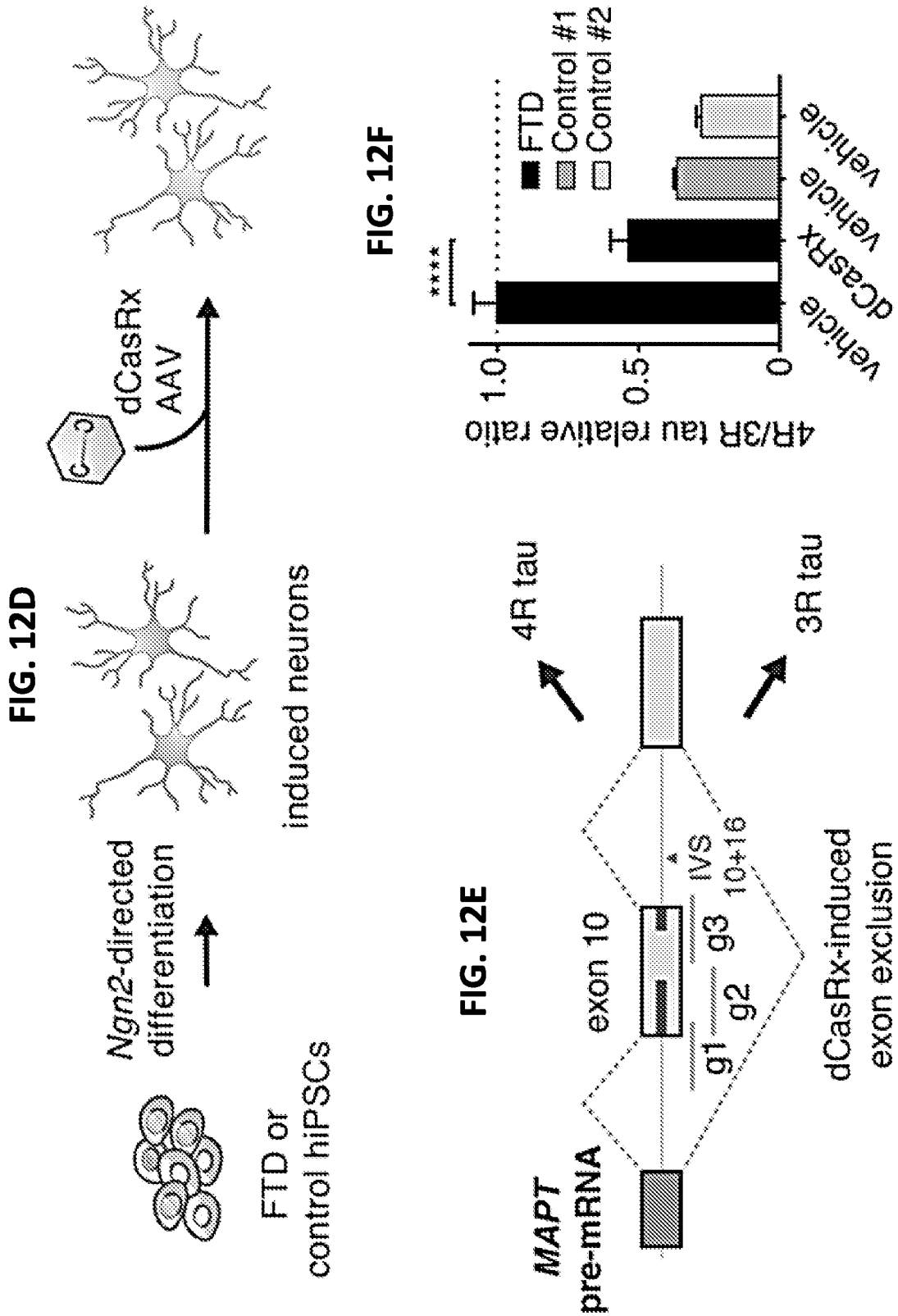


FIG. 12D
FTD or control hiPSCs
Ngn2-directed differentiation
induced neurons
dCasRx AAV
FIG. 12E
MAPT pre-mRNA
exon 10
g1
g2
g3
IVS 10+16
4R tau
3R tau
dCasRx-induced exon exclusion
FIG. 12F
4R/3R tau relative ratio
1.0
0.5
0
****
FTD
Control #1
Control #2
vehicle
dCasRx
vehicle
vehicle

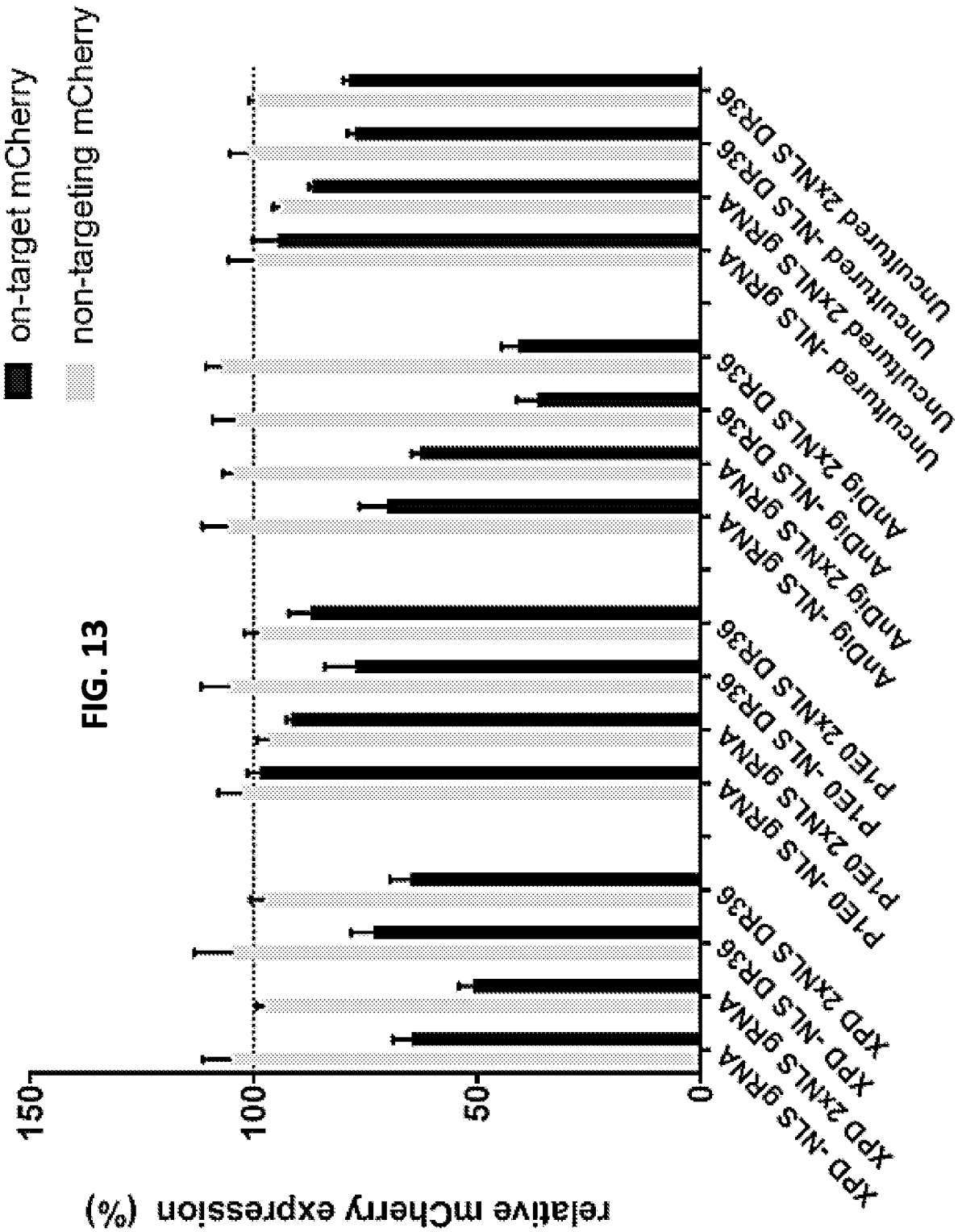
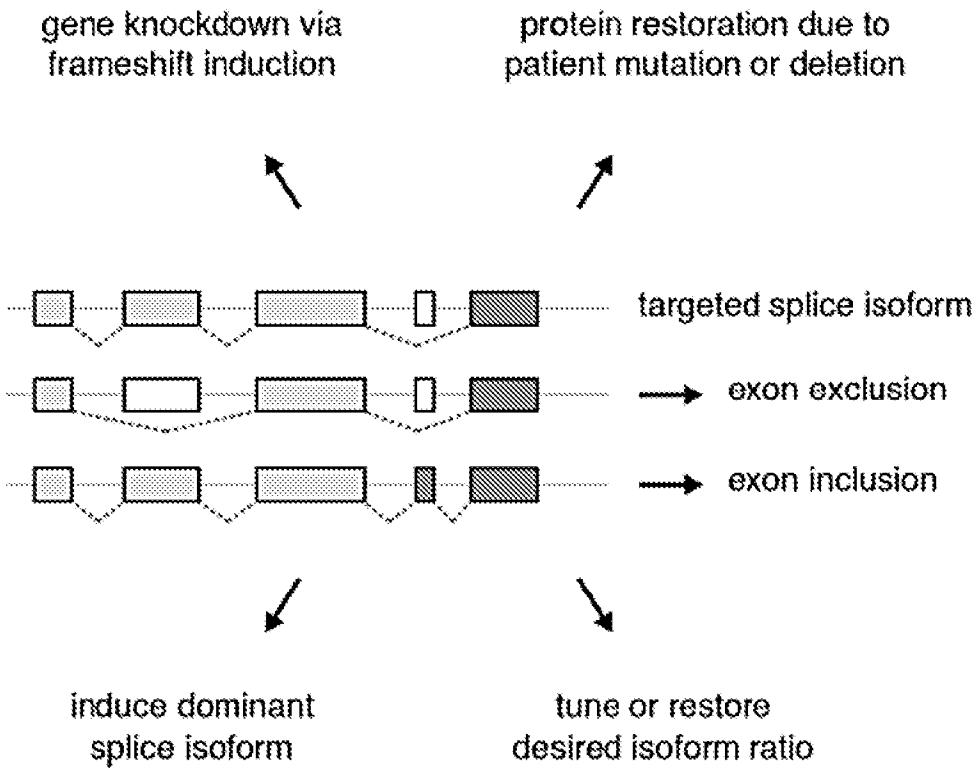

on-target mCherry
non-targeting mCherry
relative mCherry expression (%)
150
100
50
0
XPD -NLS gRNA
XPD 2xNLS gRNA
XPD -NLS DR36
XPD 2xNLS DR36
P1E0 -NLS gRNA
P1E0 2xNLS gRNA
P1E0 -NLS DR36
P1E0 2xNLS DR36
AnDig -NLS gRNA
AnDig 2xNLS gRNA
AnDig -NLS DR36
AnDig 2xNLS DR36
Uncultured -NLS gRNA
Uncultured 2xNLS gRNA
Uncultured -NLS DR36
Uncultured 2xNLS DR36

FIG. 13

FIG. 14


gene knockdown via frameshift induction
protein restoration due to patient mutation or deletion
targeted splice isoform
exon exclusion
exon inclusion
induce dominant splice isoform
tune or restore desired isoform ratio

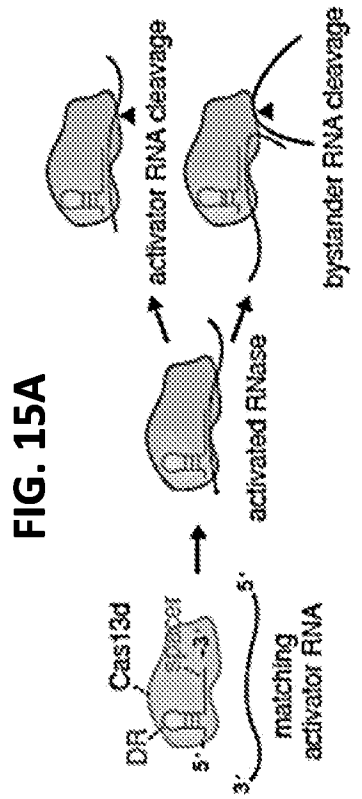


FIG. 15A

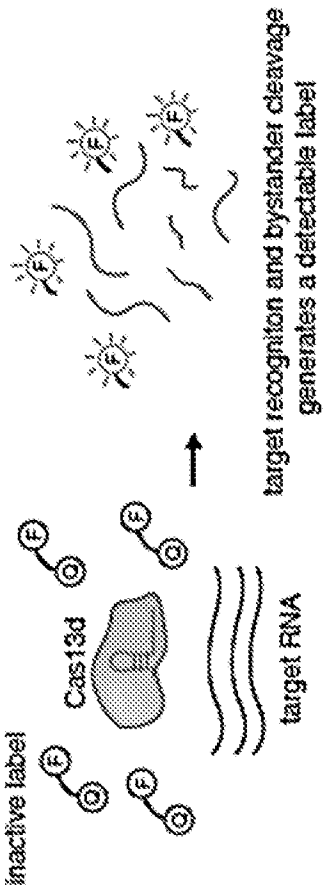


FIG. 15B

FIG. 15C

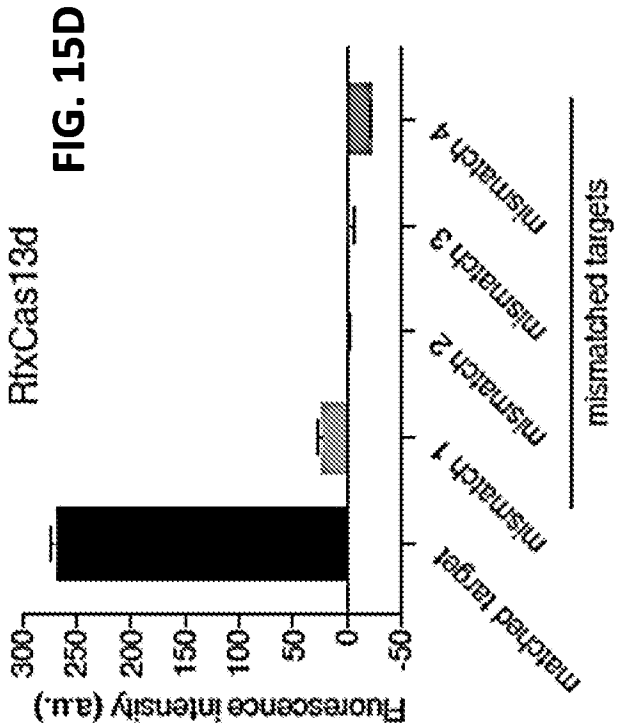
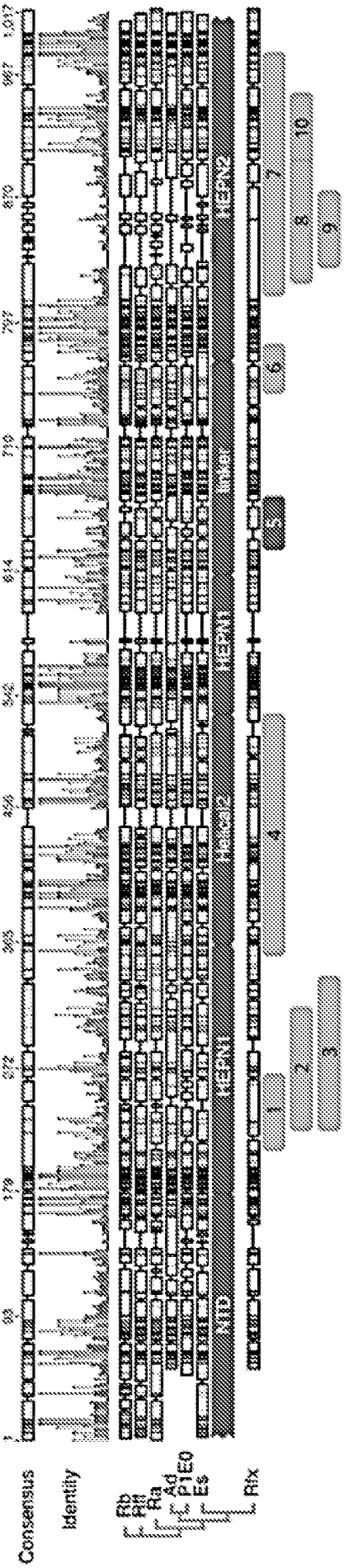


FIG. 15D

FIG. 16A


Consensus
Identity
Rb
Rh
Ra
Ad
P1E0
Es
Rfx
NTD
MEAN1
MEAN2
1
2
3
4
5
6
7
8
9
10

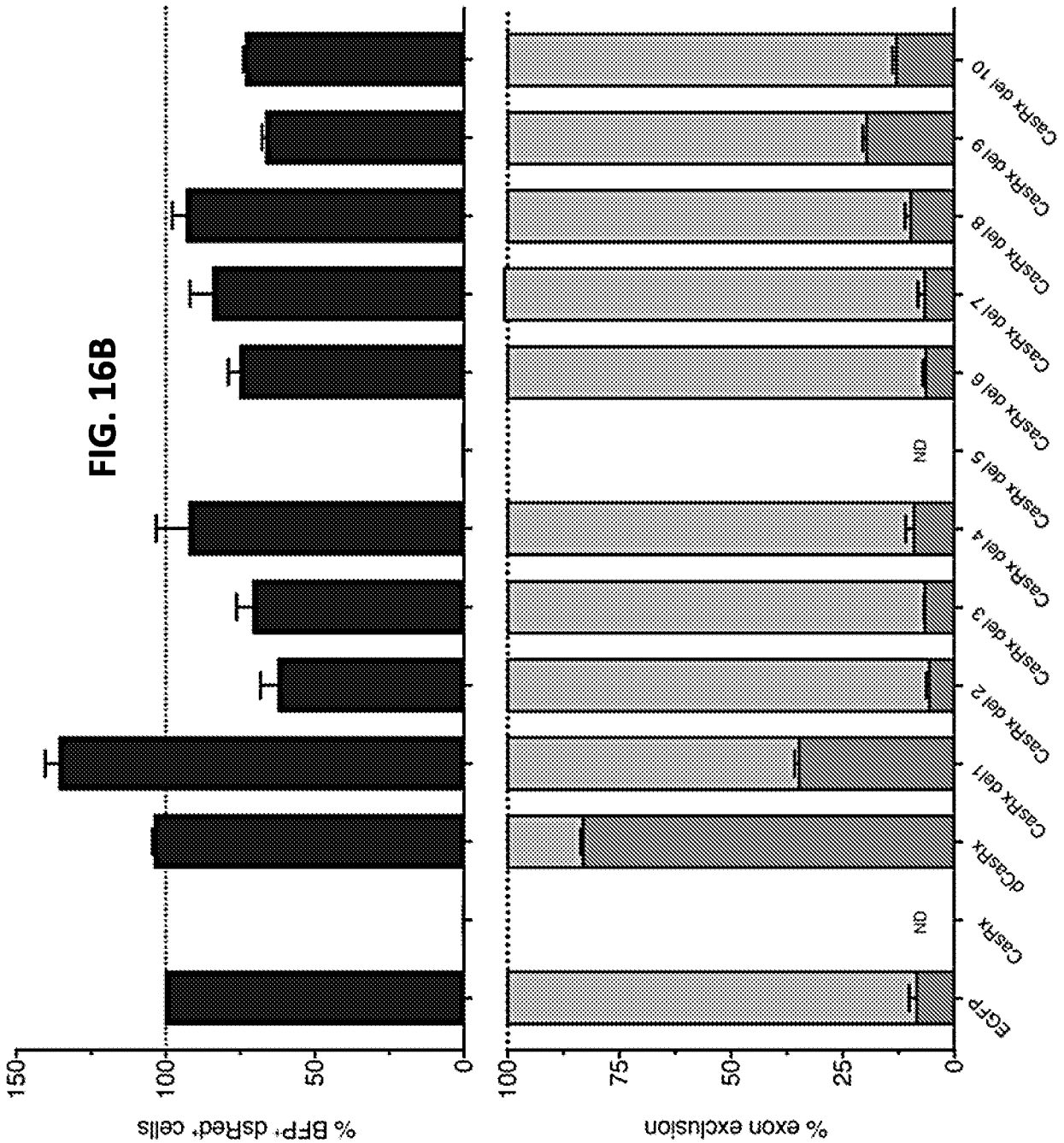

% BFP+ dsRed+ cells
150
100
50
0
% exon exclusion
100
75
50
25
0
ND
ND
EGFP
CasRx
dCasRx
CasRx del 1
CasRx del 2
CasRx del 3
CasRx del 4
CasRx del 5
CasRx del 6
CasRx del 7
CasRx del 8
CasRx del 9
CasRx del 10

FIG. 16B

FIG. 17A

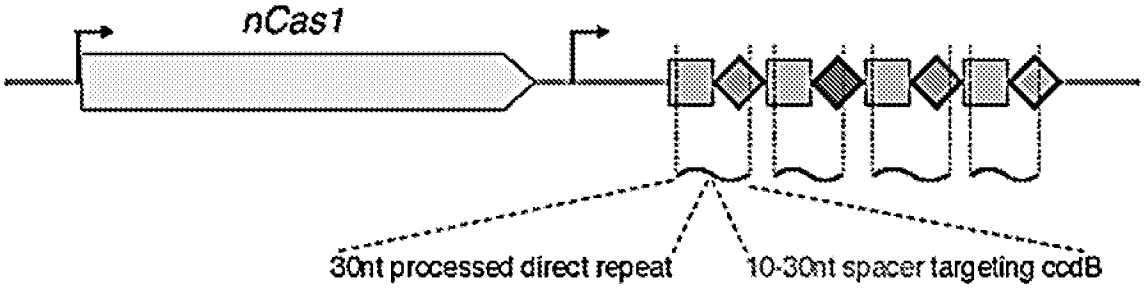
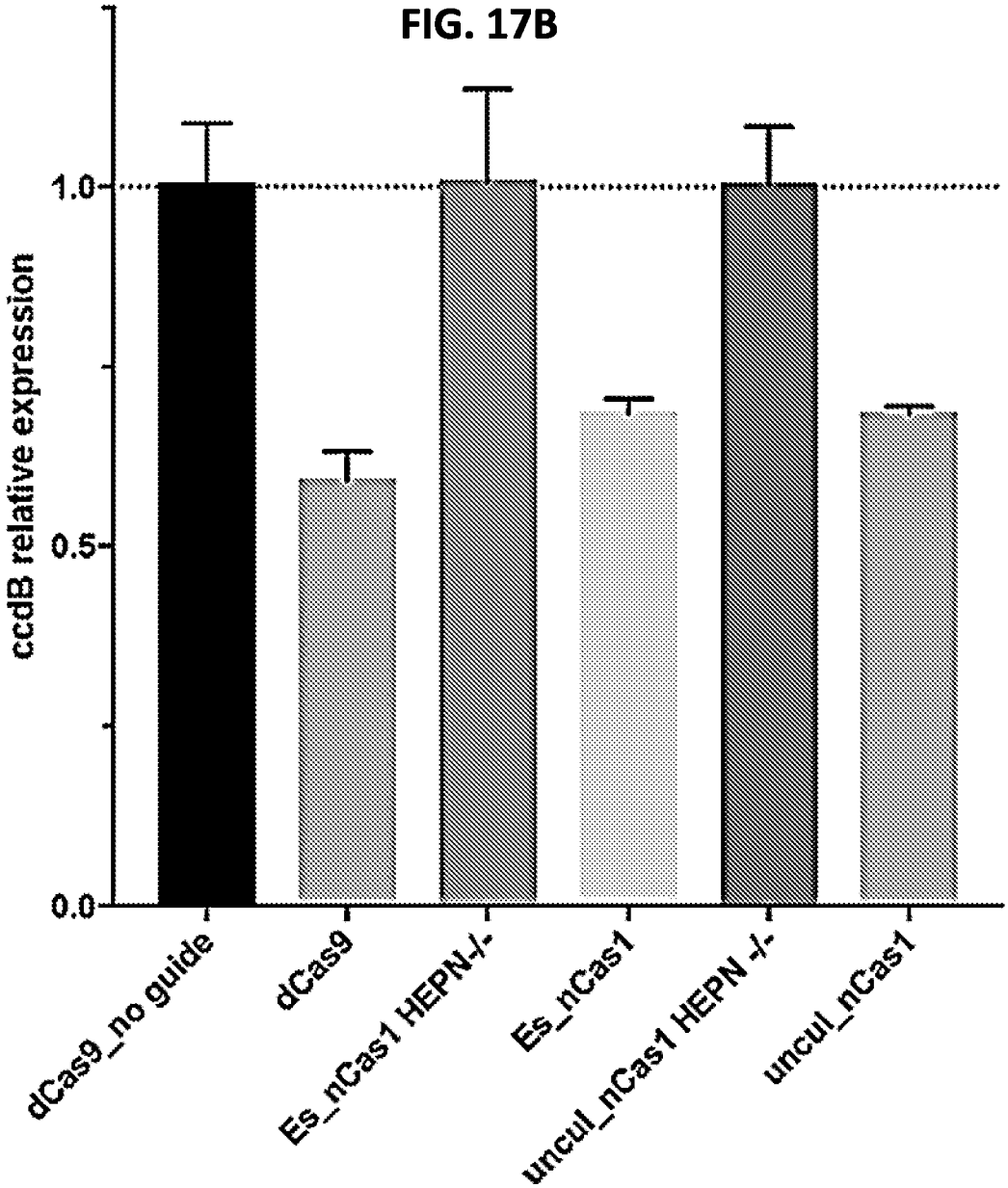


FIG. 17B



```
CLUSTAL O(1.2.4) multiple sequence alignment

037_-_emb|OIZA01000315.1|          ---------------------------MAKKKRMTAKERKQNHRESL-----------MK    22
emb|OCTW011587266.1|               ------------------------------------------------------------    0
k87_11092736                       ------------------------------------------------------------    0
BMZ-11B_GL0037915                  ------------------------------------------------------------    0
BMZ-11B_GL0037771                  ------------------------------------------------------------    0
BMZ-11B_GL0069617                  ------------------------------------------------------------    0
Ga0099364_10024192                 ------------------------------------------------------------    0
530373_GL0023589                   ------------------------------------------------------------    0
emb|ODAI011611274.1|               ----------------------------MKKKISLKEQRNTKKAEN-----------KL    20
EMG_10003641                       ------------------------------------------------------------    0
CasR_P1E0_metageno                 ------------------------------------------------------------    0
emb|OIZX01000427.1|                --------------------------MAKKKKTARQLREEMQQQR-----------KQ    21
160582958_gene49834                ----------------------------MKNSVTFKLIQAQENKEA-----------AR    20
gi|1198542314|gb|NFIR01000008.1|   ----------------------------MAKKLRPKELREKRRMAE-----------KE    20
emb|OGNF01009141.1                 ------------------------------------------------------------    0
Ga0129306_1000735                  ------------------------------------------------------------    0
MH0288_GL0082219                   ----------------------------MKKMTAKERREQEKQKR-----------AE    19
PIG-022_GL0026351                  ------------------------------------------------------------    0
PIG-028_GL0185479                  ------------------------------------------------------------    0
250twins_35838_GL0110300           ---------------------------MGNKQRVSAQKRRENAKLCN-----------QQ    22
PIG-014_GL0226364                  ------------------------------------------------------------    0
250twins_36050_GL0158985           ---------------------------MKKKHQSAAEKRQVKKLKN-----------QE    21
```

FIG. 18A

```
emb|OJMM01002900.1|                ----------------------------MGKHLNAKQRKLEKKLKN----------QQ    20
emb|OGPN01002610.1|                ----------------------------MAKKITAKQKREEKERLN----------KQ    20
emb|OHCP01000044.1|                ----------------------------MAKKITAKQKREEKERLN----------KQ    20
PIG-046_GL0077813                  ----------------------------MAKKITAKQRREERERQN----------KQ    20
pig_chimera                        ----------------------------MAKKITAKQRREERERQN----------KQ    20
emb|OIEN01002196.1|                ------------------------------------------------------------    0
O2.UC29-0_GL0096317                ------------------------------------------------------------    0
CasR_Anaerobic_dig                 ------------------------------------------------------------    0
emb|OGDF01008514.1|                ------------------------------------------------------------    0
emb|OGZC01000639.1                 ----------------------------MKKKNIRATREALKAQK----------IK    19
Ga0224415_10048792_chimera         ------------------------------------------------------------    0
emb|OCVV012889144.1|               ------------------------------------------------------------    0
Ga0187910_10006931                 ------------------------------------------------------------    0
R._flav_XPD_                       ------------------------------------------------------------    0
Ga0224415_10007274                 ------------------------------------------------------------    0
Ga0187911_10069260                 ------------------------------------------------------------    0
Ga0187910_10015336                 ----------------------------MENVKNPKEIVKGAKVRFEIYENPKPKKQNQ    31
ODAI_chimera                       MLKKCSVCGKKFETNYQGGKPVCRNCRSNKSKPVNPKSSRKPESIDK----------AI    49
31009_GL0034153                    ------------------------------------------------------------    0
DLF014_GL0011914_                  ----------------------------MAKKKNAKQRREEEKNRL----------KA    20
CasR_R._albus                      --------------------------MAKKSKGMSLREKRELEKQKR----------IQ    23
Ga0187910_10040531                 ----------------------------MAKKKKAKQRREEQEAAR----------MN    20
tpg|DJXD01000002.1|                ------------------------------------------------------------    0
CasR_E._siraeum                    ----------------------------MGKKIHARDLREQRKTDR----------TE    20
```

FIG. 18B

```
PIG-018_GL0023397                    ----------------------------MAKKLSPRELREQRERER-----------QA    20
PIG-025_GL0099734                    ----------------------------MAKKMNAKEQRAAIQAEK-----------NK    20
Ga0129317_1008067_chimera            ------------------------------------------------------------    0
EYZ-362B_GL0088915                   ----------------------------MKKKLNIREQREAKKAAA-----------KQ    20
uncultured_Ru_sp                     --------------------------MAKKNKMKPRELREAQKKAR-----------QL    22
R._flav._FD-1                        ----------------------------MKKKMSLREKREAEKQAK-----------KA    20
tpg|DBYI01000091.1|                  ----------------------------MKKKIKARDLREAKKQEK-----------LA    20

037_-_emb|OIZA01000315.1|            KADSNAEKEKAKKPVVENK------------------PDTAISKDNKPKPNKEIKKSKAK    64
emb|OCTW011587266.1|                 ----------------------------------------MKQNDRENNNKIKKSAAK    18
k87_11092736                         ------------------------------------------------MKRQKTFAK     9
BMZ-11B_GL0037915                    ---------------------------------------------MAEHSKKTAAK    11
BMZ-11B_GL0037771                    ------------------------------------------------MSKDKKTKAK    10
BMZ-11B_GL0069617                    ------------------------------------------------MPHNKKSTIK    10
Ga0099364_10024192                   ------------------------------------------------MSQSTKTKAK    10
530373_GL0023589                     ----------------------------------MKTEMKNGEVKTSAKKENCSIAK    23
emb|ODAI011611274.1|                 KYQKAQAERAAAAQQTAAGAESEENPCFDVVKDT---KRKALNPLHVEIEAPSAKKSSVK    77
EMG_10003641                         ----------------------MEET-----KVT---KETTIEKQSTKRHKQKSKKTATK    30
CasR_P1E0_metageno                   ------------------------------------------MEREVKKPPKKSLAK    15
emb|OIZX01000427.1|                  AIQKQQEQRQ------------EKAA-----AAR---ETAAPEQPAAAPVPKRQRKSLAK    61
160582958_gene49834                  KKAKDIAEQARIAKRNGVVKKEENRI-----NRI---QIEIQTQKKSNTQNAYHLKSLAK    72
gi|1198542314|gb|NFIR01000008.1|     EHKKQEK----LRKEQEELRKKQEKQ-----RED---QKELEKIKKEEGGEGEKKKSGAK    68
emb|OGNF01009141.1                   ------------------------------------------------MADIDKKKSSAK    12
```

FIG. 18C

```
Ga0129306_1000735                 ------------------------------MQK---QREQQTVTDESERKKKPLKSGAK    26
MH0288_GL0082219                  V-----------------IRGKQEAN-----RKK---AQAAATQDATAPVHHKNAKSLAK    54
PIG-022_GL0026351                 ---------------------------------------------MAESYKIKSKKTRTK    15
PIG-028_GL0185479                 ----------------------------------------MSESDKIKIKSKKTRTK      17
250twins_35838_GL0110300          KARQAESQRDKIKNMNVEK------------M-----------KNINTNDIKHTKTTAK    58
PIG-014_GL0226364                 ----------------------------------------MTKPEKKQIKQATLTK      16
250twins_36050_GL0158985          KAQKYASEPSPLQSDTAGV------------EC--------SQKKTVVSHIASSKTLAK    60
emb|OJMM01002900.1|               KDMMYTKSTDAVSVPTLKA------------APTKAEMSQDTAEASTLITPGTLKTKAK    67
emb|OGPN01002610.1|               KWAKQDTPVVPKSKTEEKP--------------V---AASDDKLLKTTQVKKVQTKSKAK    63
emb|OHCP01000044.1|               KWAKNDSVIIVPETKEEIK----------------------TGEIQDNNRKRSRQKSQAK    58
PIG-046_GL0077813                 KWAKKQDDATAVFECEANI------------KP---ADSKDEDCANT--SIKRKKTQAK    62
pig_chimera                       KWAKKQADATAVFECEADI------------KP---ADSKDEDCTNIYIKREKKKTQAK    64
emb|OIEN01002196.1|               ---------------------------------------------MERQKRKMKSKSK     13
O2.UC29-0_GL0096317               ----------------------------------------MERQKRRKKSCAK          13
CasR_Anaerobic_dig                ------------------------------------------------MNNKRKTKAK     10
emb|OGDF01008514.1|               -----------------------------------MTETKPKREDIAKTPAAKSRSK      22
emb|OGZC01000639.1                KSQENEALKKQKLAEEAAQ-------------KRREELEKKNLAQWEETSAEGRRSRVK    65
Ga0224415_10048792_chimera        ----------------------------------------MSKKENRKSYVK          12
emb|OCVV012889144.1|              ------------------------------------------------MSKNKESYAK     10
Ga0187910_10006931                ------------------------------------------------MIEKKKSYAK     10
R._flav_XPD_                      ------------------------------------------------MIEKKKSFAK     10
Ga0224415_10007274                ------------------------------------------------MIEKKKSYAK     10
Ga0187911_10069260                ----------------------------------------MSTKKRFRYSVAAK          14
Ga0187910_10015336                RFSAINVEVITDETHKEAKIQKNEFKTFDQFTKE-LQETQKVNGETKKEHITKNKHTNVK    90
```

FIG. 18D

```
ODAI_chimera                    EILTEKLISDCKK----------------------------NDCASECGKGRRKSKAK    79
31009_GL0034153                 ----------------------------------------MGSRGVIIMTEKRKSRAK    18
DLF014_GL0011914_               KIEKFKIKYAEQETVDKI---------------TEEIAQSDKIMRTESILVQHTKKSLAK    65
CasR_R._albus                   KAAVNSVNDTPEKTEEAN--------------------VVSVNVRTSAENKHSKKSAAK    62
Ga0187910_10040531              KIQSAVKAKAETAPA-------------------------VSSAFVEKRKDKQSKKTFAK    55
tpg|DJXD01000002.1|             ------------------------------------------------MKKQKSKKTVSK    12
CasR_E._siraeum                 KFADQNKKREAERAVPKK-----------------DAAVSVKSVSSVSSKKDNVTKSMAK    63
PIG-018_GL0023397               KFAKQVQEREAEKAKREK-----------------QVSVSDTVGAAAAVQNENGKKSMAK    63
PIG-025_GL0099734               KIQMANAAREKQAEEARLKAEAERIEKEENKKREAELKRFKLDQETRGKLPKNERKSLAK    80
Ga0129317_1008067_chimera       ------------------------------------------------MKKQKKSLVK    10
EYZ-362B_GL0088915              IYAAKASAAQKKADETV----------------------VAAQPAGDTANSKPIKSRNK    57
uncultured_Ru_sp                KAAEINNNAAPAI----A----------------------AMPAAEVIAPAAEKKKSSVK    56
R._flav._FD-1                   AYSAASKNTDSKPAEKKA---------------------ETPKPAEIISDNSRNKTAVK    58
tpg|DBYI01000091.1|             AFSAKANTVYENEDK-----------------------NVEAFPEALNLRSIKKSMNK    55
                                                                                          *

037_-_emb|OIZA01000315.1|       LAGVKWIIKAN--DD-----VTYISSFGKGN--------NSVLEKRIIGDVSGDVN----   105
emb|OCTW011587266.1|            AVGVKSLARLS--DG-----STVVSSFGKGA--------AAELESLITGGEIRKLS----    59
k87_11092736                    RIGIKSTVAYG--QG-----KYAITTFGKGS--------KAEIAVRSADPPEETLPTESD    54
BMZ-11B_GL0037915               RVGIKSVIAHD--GG-----KLAVTTFGKKS--------AAKLAYRDGDQTPFDEEKKSE    56
BMZ-11B_GL0037771               RMGVKALLAHG--ED-----KLTMTTFGKGN--------RSKIEFTEGYHGRALETPKHF    55
BMZ-11B_GL0069617               RIGVKSAFSHG--NS-----RLTITQFGRGN--------TADIAVETDSRGRSLSLPHRV    55
Ga0099364_10024192              RMGVKSVLAHG--KDEKGHIKLAITAFGKGN--------KAELAIQTDEKGSNLAKTYKE    60
530373_GL0023589                ANGLKSTFKLA--EN-----TYYMTSFGKNN--------TADPEKNIVRMNVENIR----    64
```

FIG. 18E

```
emb|ODAI011611274.1|                ANGLKSLLLT---DG-----KTVMTSFGRGS--------EANVEKRFDETGTKTFD----   117
EMG_10003641                        MSGLKSALVIN--NH-----EMLLTSFGKGN--------NAIAEKRYILDGDIETI---N    72
CasR_P1E0_metageno                  AAGLKSTFVISPQEK-----ELAMTAFGRGN--------DALLQKRIVDGVVRDVA----    58
emb|OIZX01000427.1|                 AAGLKSNFILDPQRR-----TTVMTAFGQGS--------TAILEKQIVDRAISDLQ----   104
160582958_gene49834                 AAGVKSVFAI---GN-----DLLMTGFGPGN--------DATIEKRVFQNRAIETL----   112
gi|1198542314|gb|NFIR01000008.1|    ALGLKSTFILDRDEQ-----KMLMTSFGRGN--------KAVRDKYIIGDKVSDIDDSWE   115
emb|OGNF01009141.1                  AAGLKSTFVLE--NN-----KLLMTSFGNGN--------KAVIEKIIDEKVDSIN-----    52
Ga0129306_1000735                   AAGLKSVFVLS--EG-----KELLTSFGRGN--------EAVPEKRVTGGTIANAR---T    68
MH0288_GL0082219                    AAGLKSTFTLP--TG-----ELLMTSFGRGS--------EALIQKRIMKDVISDVQ----    95
PIG-022_GL0026351                   RAGLKSSFVN---EN-----KGLMTAFGKGN--------NAVIEKRISGNLIENVS----    55
PIG-028_GL0185479                   RAGLKSSFVS---EG-----KGLMTAFGKGN--------NAVIEKRISGDLIENVS----    57
250twins_35838_GL0110300            KLGLKSTIIAD--K------KIILTSFINEQ-----SSKTANIEKVAGFKGDTIDT--IS   103
PIG-014_GL0226364                   RMGLKSTLVFD--N------KITVTSFANSS---NSEKASAHIEKRTDFWGKTI-----M    60
250twins_36050_GL0158985            AMGLKSTLVMG--D------KLVITSFAASKAVGGAGYKSANIEKITDLQGRVI-----E   107
emb|OJMM01002900.1|                 AMGLKSTLVFD--D------KIVVTSFLNSKT--EENEKCAHIEKIADCNGQTI-----V   112
emb|OGPN01002610.1|                 AMGLKTVLSFD--D------KIAIASFVNDK-----KTKLPHIERITDKSGTTI-----H   105
emb|OHCP01000044.1|                 AMGLKAVLSFD--N------KIAIASFVSSK-----NAKSSHIERITDKEGTTI-----S   100
PIG-046_GL0077813                   AMGLKTVLGFD--N------KIAIASFTSSK-----DSKSSHIERITDPNGKTI-----R   104
pig_chimera                         AMGLKTVLGFD--N------KIAIASFMSSK-----DSKSSHIERITDPNGKTI-----R   106
emb|OIEN01002196.1|                 MAGVKSVFVIG--D------ELLMTSFGDGD--------DAVLEKDIDENGVVNDC---R    54
O2.UC29-0_GL0096317                 MAGVKSTFVID--S------DLLMTSFGERD--------KAVREKYIAQGNVKSLR---K    54
CasR_Anaerobic_dig                  AAGLKSVFFDQ--K------QAVLTTFAKGN--------NSQIEKKVVNSEVKDLR---Q    51
emb|OGDF01008514.1|                 AAGLKSTFAVN--G------SVLLTSFGRGN--------DAVPEKLITEKAVSEIN---T    63
emb|OGZC01000639.1                  AVGVKSVFVVG--D------DLYLATFGNGN--------ETVLEKKITPDGKITTF---P   106
```

FIG. 18F

```
Ga0224415_10048792_chimera          GLGLKSTLVSD--S------KVYLTTFADGS--------NAKLEKCVENNKIICIS---N    53
emb|OCVV012889144.1|                GMGLKSALVSG--S------KVYMTSFEGGN--------DAKLEKVVENSEIVSLA----    50
Ga0187910_10006931                  GMGLKSTLVSD--S------KVYMTSFGNGN--------DARLEKVVENNAISCLV---D    51
R._flav_XPD_                        GMGVKSTLVSG--S------KVYMTTFAEGS--------DARLEKIVEGDSIRSVN----    50
Ga0224415_10007274                  GMGLKSTIVSG--S------KVYMTTFGDGS--------EARLEKVVENDSIKTLH----    50
Ga0187911_10069260                  AAGLKSSLAVD--TD-----RTVMTSFGHGN--------AAILEKEIVDGEISVLN---I    56
Ga0187910_10015336                  AAGVKSVFAVD--DG-----NVLITSFGRGN--------AADIETLKSDDDKTINL---T   132
ODAI_chimera                        AAGLKSALDLG--NG-----QALITSFGKGN--------EAVPEKRVEGDKIISLR---E   121
31009_GL0034153                     ALGVKSAFLKD--D------DVIMTSFGRGN--------DAVVEKIIKNNEVQNIN---I    59
DLF014_GL0011914_                   ASGVKSVFIKD--D------EVIMTSFGRGN--------DAVVEKIIKDNNVDEK----N   105
CasR_R._albus                       ALGLKSGLVIG--D------ELYLTSFGRGN--------EAKLEKKISGDTVEKLG---I   103
Ga0187910_10040531                  ASGLKSTLAVD--N------SAVMTVFGRGN--------EAKLDHRINADLQSESL---H    96
tpg|DJXD01000002.1|                 TSGLKEALSVQ--G------TVIMTSFGKGN--------MANLSYKIPSSQKPQNL---N    53
CasR_E._siraeum                     AAGVKSVFAVG--N------TVYMTSFGRGN--------DAVLEQKIVDTSHEPLN---I   104
PIG-018_GL0023397                   AAGVKSVFAVG--D------TVYMTSFGRGN--------DAVLEKKIVDESHESLN---K   104
PIG-025_GL0099734                   AAGLKSTFAVG--T------DLYMTSFGKGN--------DAIVEKKISNKQVTQIN---T   121
Ga0129317_1008067_chimera           AAGLKSAFVVG--D------SVYLTSFGKGN--------AARLDTKINPDNSTERY---V    51
EYZ-362B_GL0088915                  AAGLKSTFVVG--D------SLYMTSFGKGN--------LSVPEYRVDTGDAYRVD---P    98
uncultured_Ru_sp                    AAGMKSILVSE--N------KMYITSFGKGN--------SAVLEYEVDNNDYNQTQ---L    97
R._flav._FD-1                       AAGLKSTIISG--D------KLYMTSFGKGN--------AAVIEQKIDINDYSFSA----    98
tpg|DBYI01000091.1|                 AAGLKSTLIDG--K------SLYLTAFGKGN--------NAVVEHMIATDDSYSLK---T    96
                                      *:*                 :: *

037_-_emb|OIZA01000315.1|           ---KDSHMYVNPKYTKKNYEIKNGF-------------------------SSGSSLVTY   136
```

FIG. 18G

```
emb|OCTW011587266.1|              ---D-KAILEITDDTQNKNAYNVKS-------------------------SRIPNLTAR    89
k87_11092736                      ATLSIHAKFAKAGRDGREFKCGDVD----------------------ETRI---HTSRS    88
BMZ-11B_GL0037915                 -----LGGFNVTGVDENTVDAERGI----------------------PGEGLIEASVAN    88
BMZ-11B_GL0037771                 G----KRGFEVRKIDENVDLYGDL-------------------------------EEGK    79
BMZ-11B_GL0069617                 -----EKSFTVNKIDDEIDVQRS-------------------------------------    73
Ga0099364_10024192                -----RNITANKIVSEGIQTSGTI--------------------------------AGE    82
530373_GL0023589                  ------NTFSATPISERIVDITGET--------------------------------------    83
emb|ODAI011611274.1|              ---RDPELFSAKPLETGYRIQRFNA-------------------S------PKDAG----    145
EMG_10003641                      ---NKNKKFDANNDSKVVVIKGISN---------------------------PNGQLTN    101
CasR_P1E0_metageno                ---GEKQQFQVQRQDESRFRLQNSR-------------------------LADRTVTADD    90
emb|OIZX01000427.1|               ----PVQQFQVEPASAAKYRLKNSR-------------------------VRFPNVTADD    135
160582958_gene49834               ---SSPEQYSAEFQNKQFKIKGNIK-------------------VLNHS-TQKMEEIQTE    149
gi|1198542314|gb|NFIR01000008.1|  ---NKKAALSVEVCGKSFNISKKEN-------------------------DDCEPVKVNN    147
emb|OGNF01009141.1                ----EPEVFSVTPCDKKFELQPAKR-------------------------GLAADSLVDN    83
Ga0129306_1000735                 ---DNKEAFSAALQNKRFEVFGR-------------------------------TAGSSDD    95
MH0288_GL0082219                  ---EGDPAFSVIPMPKQFEIQGRAI-------------------DVQKEGSSVPEVLTDD    133
PIG-022_GL0026351                 ----NKPRFSVEFGNKEYIINDK-------------------------------KLGIVSDD    82
PIG-028_GL0185479                 ----NKPRFSVEFGNKEYIINDK-------------------------------KLGIVSDD    84
250twins_35838_GL0110300          ---YTPRMFRSEINPGEIVISKGD-------------------------------DLSEFAN    131
PIG-014_GL0226364                 ---QDEKMFRTDVEKQQIYIEKEFPDENNKII------------T--------QTAMFEN    97
250twins_36050_GL0158985          ---EHERMFSADVGEKNIELSKND-------------------------------CHTNVNN    135
emb|OJMM01002900.1|               ---ERPRMFNTSINAKKVDLSKDN-------------------------------DETNYPN    140
emb|OGPN01002610.1|               ---ENARMFDSSVDEQNVNIEKRMTIEEKQND------------GTFKKDEKDVKATICN    150
emb|OHCP01000044.1|               ---VNSKMFESSVNKRDINIEKRITIEEPQQD------------GTIKKEEKGVKSTTCN    145
```

FIG. 18H

```
PIG-046_GL0077813                ---EDVRMFDSNVDECSINLEKRMTVEERQKD-----------GTIKKDEKDVKSTICN   149
pig_chimera                      ---EDVRMFDSNVDECSINLEKRMTVEERQKD-----------GTIKKDEKDVKSTICN   151
emb|OIEN01002196.1|              ---N-PAAYDAVYGTDSIRVKKTN-------------------------NNIRAKVNN   83
O2.UC29-0_GL0096317              ---Q-EQAFDAQYGKEKIFVTNAD-------------------------NKLHVAINN   83
CasR_Anaerobic_dig               ---P--PAFDLELKEKTFYISGKNNINTSRENPLASASLPLSKRQRIRAERIKRAREENR   106
emb|OGDF01008514.1|              ---V-KPRFSVEKPATSYSSSF-------------------------GIKSHISATADN   93
emb|OGZC01000639.1               ---E-EETFTAKLKFAQTEPTVATSI-----------------GISNGRIVLPEISVDN   144
Ga0224415_10048792_chimera       ---D-KEAFAASIANKNVGYKIKND-------------------EKFRHPKGYDIISNN   89
emb|OCVV012889144.1|             ---E-KESFSAEIFKKNIGCKIE-N-------------------KKFKHPKRYDVIADN   85
Ga0187910_10006931               ---K-KEAFVAEITDKNAGYKII-N-------------------KKFGHPKGYDVVANN   86
R._flav_XPD_                     ---E-GEAFSAEMADKNAGYKIG-N-------------------AKFSHPKGYAVVANN   85
Ga0224415_10007274               ---E-GEAFSAEMTDNNAGYKIG-N-------------------VKFSHPNGYDVVANN   85
Ga0187911_10069260               ---E-NPAFDAVIND---KKYALTG-------------------HH-----AGVHALVDQ   85
Ga0187910_10015336               ---ETENQ----------KKYVVTN-------------------K-----RSNVKGLADN   155
ODAI_chimera                     ---G--EAFALAKEPVNHKLVLKSG----------------------RIKAEGGPVEFDG   154
31009_GL0034153                  ---D-KPVYDVVGVEKETGIINVVG-------------------TKRGINTINKKKQYAN   96
DLF014_GL0011914_                ---T-KPAYEIVDIEKNGDLIKVQ-------------------------SRRCKNIESAN   136
CasR_R._albus                    ---G---AFEVAERD--ESTLTLES---------------------GRI--KDKTARPKD   132
Ga0187910_10040531               ---PQAALKN--VHAPNKQKIHFIG-------------------RM-----QDMNLTADH   127
tpg|DJXD01000002.1|              ---SSAGLKN--VEVS-GKKIKFQG-------------------R-----HPKIATTDN   82
CasR_E._siraeum                  ---D---DPAYQLNVVTMNGYSVTG-------------------HR--G--ETVSAVTDN   135
PIG-018_GL0023397                ---A---NPAYEITDISSVNYSVEG-------------------HR--G--QSVCATADN   135
PIG-025_GL0099734                ---E-KENFTVDTDGITDTIVPVQS-------------------KR--I--ACLNTQADN   154
Ga0129317_1008067_chimera        ---SDSEKHTLKINSITDTELRLSG-------------------P------FPKQAEAKN   83
```

FIG. 18I

```
EYZ-362B_GL0088915                  ---IS-EHSELHVLGADDLAVSFSG-------------------TRRMK--DVSELKADN   133
uncultured_Ru_sp                    ---SSKDNSNIQLGGVNEVNITFSS-------------------KHGFE--SGVEINTSN   133
R._flav._FD-1                       ---M-KDTPSLEVDKAESKEISFSS-------------------HHPFV--KNDKLTTYN   133
tpg|DBYI01000091.1|                 ---L-ENEPSLKVKAADELKVTFMS-------------------RRPFV--QESELSAVN   131


037_-_emb|OIZA01000315.1|           PNK-----------PDKNSGMDALCLKPYFEIDFFG-----HIF--TDNMHIQAIYNIFD   178
emb|OCTW011587266.1|                TDK-----------LSDKSGMDDLGFKRELELEVFG-----QCF--DDSIHIQIAHAVFD   131
k87_11092736                        EYES------LISNPAESPREDYLGLKGTLERKFFG-----DEY-PKDNLRIQIIYSILD   136
BMZ-11B_GL0037915                   PVND------LIDESRKEYGQDYLMLKDQLEKVFFG-----QTF-PGDNLRIQIIYNILD   136
BMZ-11B_GL0037771                   TIEA------LLINPSEKVGEDYLKLKGTLEKRFFG-----REF-PHDNIRIQLIYNILD   127
BMZ-11B_GL0069617                   SLLG------TLLNPAENDPVDYLGLKSTLEKEFFG-----QEF-PNDSIRIQIIHNILD   121
Ga0099364_10024192                  GHAT------FLNNPAEHVGTDYLKLKETLEMEFFG-----KSF-PGDSVRIQIIHQILD   130
530373_GL0023589                    --------KKTAQVVLPACSDNQLHAKDTIEQMYFG-----KTY--TDNIHIQIAYNIMD   128
emb|ODAI011611274.1|                -----------LAYRPAGVRPDQIGAKAALEKRYFG-----KET-PGDNIHVQIAYQIQD   188
EMG_10003641                        PLFD---QSPTAIQPNRTSGNDMIGIRRMLERKYFVHNEENKEF--QDNIRIQIAYCILD   156
CasR_P1E0_metageno                  PLHR---A-ETPRRQPLGAGMDQLRRKAILEQKYFG-----RTF--DDNIHIQLIYNILD   139
emb|OIZX01000427.1|                 PLYR---R-KD---GGFVPGMDALRRKNVLEQRFFG-----KSF--ADNIHIQMIYSILD   181
160582958_gene49834                 LQD-----------NYNRPHFDLLGCKNVLEQKYFG-----RTF--SDNIHVQIAYNIMD   191
gi|1198542314|gb|NFIR01000008.1|    PVL-----------SGNKKDDDLIHCRKNLEEMYFG-----QQF--KDNIHIQLIYNILD   189
emb|OGNF01009141.1                  PLK-----------SKKTAGDDAIHSRKFLERQFFDG----NTF--NDNIHIQLIYNILD   126
Ga0129306_1000735                   PLA-----------VSRAPGQDLIGAKTALEERYFG-----RAF--ADNIHMQVIYAIQD   137
MH0288_GL0082219                    PRA-----------SQKPAGADLIGAKPALEKRYFG-----KTF--TDNIHIQLIYRILD   175
PIG-022_GL0026351                   PRE-----------SQKPVGEDGIHCKRALEERYFG-----KTF--NDNIHIQLIYSILD   124
```

FIG. 18J

PIG-028_GL0185479	PRE----------SQKPVGEDGIHCKRALEERYFG-----KTF--NDNIHIQLIYSILD	126
250twins_35838_GL0110300	PAN----------FPIGRDYVKIRSALEKQYFG-----KEF-PEDNLHVQIAYNVAD	172
PIG-014_GL0226364	PAL----------RNCGKDYIGMKNSVEKVMFG-----KEF-PNDNLHIQIAYNILD	138
250twins_36050_GL0158985	PVV----------TNIGKDYIGLKSRLEQEFFG-----KTF-ENDNLHVQLAYNILD	176
emb|OJMM01002900.1|	PAF----------EDCGRDYINVKSALEKRVFG-----KTY-NKDNLHVQIAYNIFD	181
emb|OGPN01002610.1|	PYF----------KTCGKDYIGIKDVAEKYFFG-----KTF-PNENLRVQIAYNVFD	191
emb|OHCP01000044.1|	PYF----------KVGGKDYIGIKEIAEEHFFG-----RAF-PNENLRVQIAYNIFD	186
PIG-046_GL0077813	PYS----------NECGKDYIGIKSVAEELFFG-----RTF-PNDNLRVQIAYNIFD	190
pig_chimera	PYS----------NECGKDYIGIKSVAEELFFG-----RTF-PNDNLRVQIAYNIFD	192
emb|OIEN01002196.1|	PLAKSNIRSEESALFRTRVNEYKREQKDKYETLFFG-----KTF--DDNIHIQLISKILD	136
O2.UC29-0_GL0096317	PLCGNKKQSPEAQEYSRRVLEHKKAINQKYETIVFG-----KPS--NDNIHIQVVSKIMD	136
CasR_Anaerobic_dig	PYHNV----------KRVGEDDLRAKADLEKHYFG-----KEY--SDNLKIQIIYNILD	148
emb|OGDF01008514.1|	PLAGR----------APVGEDAIHAKEVLEQRVFG-----KTF-SDDNIHIQLIYNILD	136
emb|OGZC01000639.1	PLHTT--MQK--NTIKRSAGEDILQLKDVLENRYFD-----RSF--NDDLHIRLIYNILD	193
Ga0224415_10048792_chimera	PLLHN----------NSVQQDMLGLKNVLEKRYFG-----KSSGGDNNLCIQIIHNIID	133
emb|OCVV012889144.1|	PLYK----------GSVRQDMLGLKETLEKRYFN-----SAD-GTDNVCIQVIHNILD	127
Ga0187910_10006931	PLYT----------GPVQQDMLGLKETLEKRYFG-----SSVSGNDNICIQVIHNILD	129
R._flav_XPD_	PLYT----------GPVQQDMLGLKETLEKRYFG-----ESADGNDNICIQVIHNILD	128
Ga0224415_10007274	PFYT----------GPVQQDMLGLKEILERRYFG-----SSTDGNNTICIQIIHNILD	128
Ga0187911_10069260	PQN----------RSDAVHIRGALEKKYFG-----DTF--ADNIHVQIAYNILD	122
Ga0187910_10015336	PTKV----------ESIIPGETQIGFKSILEKHFFG-----RTF--NDNIHIQIIHNILD	198
ODAI_chimera	PQKNS----------SMSAVRLDEYNNQIERKFFRGN---KVL--QDNIRIQIAHNIMD	198
31009_GL0034153	TQV----------SQKRMGMDLIQRKDKLEKVFFN-----ENF--NDNIHIQLIYNILD	138
DLF014_GL0011914_	TEL----------PPEKIGMDLIRCKDKLEEIYFG-----RTF--NDNIHIQLIYNILD	178

FIG. 18K

```
CasR_R._albus                    PRHI-----T--VDTQGKFKEDMLGIRSVLEKKIFG-----KTF--DDNIHVQLAYNILD   178
Ga0187910_10040531               PLHSH--------DGERAVGADLLCAKDKLEQLYFG-----RTF--NDNIHIQLIYQILD   172
tpg|DJXD01000002.1|              PLFK------------PQPGMDLLCLKDKLEMHYFG-----KTF--DDNIHIQLIYQILD   123
CasR_E._siraeum                  PLRRFNGRKK--DEPEQSVPTDMLCLKPTLEKKFFG-----KEF--DDNIHIQLIYNILD   186
PIG-018_GL0023397                PLHHD---------GGSSVPTDMLCLKPTLERVYFG-----KTF--DDNLHIQLIYNILD   179
PIG-025_GL0099734                PLYRK--------DSENQVQPDKLLLKDTLEKVYFG-----KTF--NDTLHIQLIYNILD   199
Ga0129317_1008067_chimera        PTHK-----K--DNEQKNTRQDMLGLKSTLEKFYFG-----STF--DDNIHIQIIHNIQD   129
EYZ-362B_GL0088915               PLHSG----K--NNLVPKNGTDILGLKGILEQRYYG-----RNF--DDNVHIQIIYNILD   180
uncultured_Ru_sp                 PTHRS--------GESSPVRGDMLGLKSELEKRFFG-----KTF--DDNIHIQLIYNILD   178
R._flav._FD-1                    PLYGG----K--DNPEKPVGRDMLGLKDKLEERYFG-----CTF--NDNLHIQIIYNILD   180
tpg|DBYI01000091.1|              PLHSG----K--DKPNKSAGQDMLGLKSELEKRYFG-----KIF--DDNLHIQIIHNILD   178
                                                .   *   :            : : ::    : *

037_-_emb|OIZA01000315.1|        IEKILAKHITNIIYTVNSFDRNYNQ------------SGNDTIGF---------DLNYRV   217
emb|OCTW011587266.1|             IQKSLAAVIPNVLYTLNNLDRSYST--------DNTSDKKDIIGN---------TLNYQH   174
k87_11092736                     IQKILGLYVEDILHFVDGLQDEP----------------EDLVGL---------GLG--D   169
BMZ-11B_GL0037915                MQKAFGLYVNDVLYAVNNLQNTGRL-------LTDDDWTSDVIGL---------ALG--K   178
BMZ-11B_GL0037771                IYKILGMNVADILYALGNMQDTE--------------LDIDMFGQ---------SLN--N   162
BMZ-11B_GL0069617                IQKILGIYINDVLYSIANLQNVE--------------TVEDCLGQ---------SLS--K   156
Ga0099364_10024192               IQKLLGIYITDIIYCINNLRDETHL-----------DHESDIVGL---------SMS--N   168
530373_GL0023589                 MKKIFGVYSNNIIYSINNLLEE--------------REGSDFLGV----------FYTEN   164
emb|ODAI011611274.1|             IEKLLAVYISNIIYAVNNVTGVSAM-------KDSKGRPVDLLGD--------YGILGEE   233
EMG_10003641                     IEKILMPHINNICFEINNMLRL-E-----------GYQEDSFMGS----------FNLYK   194
CasR_P1E0_metageno               IHKMLAVPANHIVHTLNLLGGY---------------GETDFVGM----------LPAGL   174
```

FIG. 18L

```
emb|OIZX01000427.1|                IHKILAAASGHIVHLLNIVNGS--------------KDRDFIGM----------LAAHV   216
160582958_gene49834                IEKLLTPYINNIIYTLNELMRD--------------NSKDDFFGCDSHFSVAYLYDELKA   237
gi|1198542314|gb|NFIR01000008.1|   IEKILAVQINNIVFILNNLLRW-S-----------GEEEFDLIGS----------LGVNH   227
emb|OGNF01009141.1                 IEKILSVHVNDIVYSVNNILSRGE----------GMEYNDYIGT--------L---NLK   164
Ga0129306_1000735                  INKILAVHANNIVYTLNNLDRE-A----------DPETDDFIGS--------GYLTLKN   177
MH0288_GL0082219                   IKKVLAIHANNIVYTLDNLSRR-E----------GTEGTDFIGI--------GGLTLKN   215
PIG-022_GL0026351                  LEKILAVHSNNIVYTFDNLRRK-E----------TDEYDDFIGY----------MSTVN   162
PIG-028_GL0185479                  LEKILAVHSNNIVYTFDNLRRK-E----------TDEYDDFIGY----------MSTVN   164
250twins_35838_GL0110300           IKKILSVYINNIIYMFYNLARSEEYDIFYNSQSENSGRDCDVIG----------SLYYQA   222
PIG-014_GL0226364                  IKKIICPYINNVVNIFYNLGRGKLK----------DDKNDLLG----------KQIGP-   176
250twins_36050_GL0158985           IKKILGTYVNNIIYIFYNLNRAGTG--------RDERMYDDLIG----------TLYAYK   218
emb|OJMM01002900.1|                IKKIIGAYINNIIYIFYNLGREEYD------------AKKDIIG----------TQ----   215
emb|OGPN01002610.1|                IQKILGTYVNNIIYSFYNLRRDG-K------------SDVDIIG----------SLYAFA   228
emb|OHCP01000044.1|                VQKILGTFVNNIIYSFYNLSRDEVQ------------SDNDVIG----------MLYSIS   224
PIG-046_GL0077813                  IQKIFGTYINNIIYSFYNLSRDEFQ------------SDNDVIG----------TLYMLK   228
pig_chimera                        IQKILGTYINNIIYSFYNLSRDESQ------------SDNDVIG----------TLYMLK   230
emb|OIEN01002196.1|                IEKTFSVVIGNIVYAINNLSLEQS----------IDRPIDIFGD--------KNTQ---   174
O2.UC29-0_GL0096317                IEKTFSFVMGNIVYAINNLDPEQN----------IDNPIDYMSN--------KART---   174
CasR_Anaerobic_dig                 INKIISPYINDIVYSMNNLARNDEY----------IDGKIDVIG----------SLSSTT   188
emb|OGDF01008514.1|                IRKILSTYANNVVFTINSMRRLDE----------YDREQDYLG----------YLYTGN   175
emb|OGZC01000639.1                 IEKILAEYTTNAVFAIDNVSGCS----------------DDFLS----------NFSTRN   227
Ga0224415_10048792_chimera         IEKILSEYIPNVVYAFNNIAGFKDE----------HNNIIDIIGT--------QTYNSSY   175
emb|OCVV012889144.1|               IEKILAEYITNAVYSFDNIAGFG----------------EDIIGM--------GGFKPIY   163
Ga0187910_10006931                 IEKILAEYITNAAYAVNNIAGLD----------------KDIIGF--------GKFSTVY   165
```

FIG. 18M

```
R._flav_XPD_                    IEKILAEYITNAAYAVNNISGLD----------------KDIIGF--------GKFSTVY    164
Ga0224415_10007274              IEKILAEYITNAVYATNNIIDPD----------------NDVIGG--------KKFTSIK    164
Ga0187911_10069260              ITKILTVYANNVVYALNNLVHADDD--------T-QADELDSLG----------NFSAGT    163
Ga0187910_10015336              IKKILAVHTNNIVYALDNIHERGRE--------NSAEKPIDMIGA--------GGISTSK    242
ODAI_chimera                    IEKILAVHINNIVYSVNNLFRVQG------------NEGDDFLG----------YRGLDK    236
31009_GL0034153                 IEKILSVYVNNIVYALGNLERDNS------------DDEKDLIG----------YASTLN    176
DLF014_GL0011914_               IEKILSVYINNAVYALGNLERKEN------------DKEKDLVG----------YLSARA    216
CasR_R._albus                   VEKIMAQYVSDIVYMLHNTDKTER--------------NDNLMG----------YMSIRN    214
Ga0187910_10040531              IQKILALHANNIIFALDNLLHKKN-----------DELSDDFVGM--------GRMRATI    213
tpg|DJXD01000002.1|             IEKILAVHVNNIVFTLDNVLHPQK-----------EELTEDFIGA--------GGWRINL    164
CasR_E._siraeum                 IEKILAVYSTNAIYALNNMSADEN------------IENSDFFM----------KRTTDE    224
PIG-018_GL0023397               IEKILAIYSTNAVYALNNMIADED------------DESSDLFQ----------NMTTDY    217
PIG-025_GL0099734               IQKILTVYSVNTVYTLNNLFGKEI------------SERGDFIG----------ALSFQQ    237
Ga0129317_1008067_chimera       IAKILAAHSNNAGYALDNMLAYQG------------VEFSDMIG----------YMGTSR    167
EYZ-362B_GL0088915              IEKNLAIYVTNVVAALDHMCDETL-----------DTEHEDFIG----------YMNTMN    219
uncultured_Ru_sp                IEKILAVYVTNIVYALNNMLGVKG-----------SESHDDFIG----------YLSTNN    217
R._flav._FD-1                   IEKILAVHSANITTALDHMVDEDDE----------KYLNSDYIG----------YMNTIN    220
tpg|DBYI01000091.1|             IEKIIAVYATNITAAIDHMVDDDNE----------QYLQGDFIG----------YMNTLN    218
                                : * :     .                              . .

037_-_emb|OIZA01000315.1|       PYLEYGGGKDSNGK--------------PNKISAWKKRENFINFYN-----EAKPHLGY-    257
emb|OCTW011587266.1|            SYESFNVE----------------------------KRGEFTEYYN-----AAKDRFSY-    200
k87_11092736                    ------------------------------------EKMQKLLS-----KALPYMGF-    185
BMZ-11B_GL0037915               ------------------------------------GRGKEIIE-----SMSDALGF-    194
```

FIG. 18N

```
BMZ-11B_GL0037771                    ----------------------------------P--EKAKDVVQ-----RMKPYMGF-  179
BMZ-11B_GL0069617                    ----------------------------------GQDEKLKKILN-----SITPYLGF-  175
Ga0099364_10024192                   -------------------------------------TNVNLALN-----QMRPYFGF-  184
530373_GL0023589                     SYQAMNFADAI----------IKSKLLTKNSKYLNLDMNNVWF-NK-EVSELSSFRKYV-  211
emb|ODAI011611274.1|                 GLTKRLQRIPE----------Q-----------ADEEAKALQAFL-------CSERLSY-  264
EMG_10003641                         PYDAFIATTDD----------KE-----------SSRRDNFAKLM-------TSKQVRY-  225
CasR_P1E0_metageno                   PYDKLRVVKKK----------NGDT---------VDIKADIAAYA-------KRPQLAY-  207
emb|OIZX01000427.1|                  LYNELNEEA----------------------------KRSIADFC-------KSPRLIY-  240
160582958_gene49834                  GYSDRLKTKPN----------LS--------KNIDRIWNNFCNYMN-SDSGNTEARLAY-  277
gi|1198542314|gb|NFIR01000008.1|     TYEEFRGRNK-------------------------NYGKFSELI-------KQSQMRY-  253
emb|OGNF01009141.1                   SFETYKNNLV-----------------------NKKKFDLDRVK-------KIPQLAY-  192
Ga0129306_1000735                    TFETYCDPAAL----------NER-----EREKVTVSKQHFDAFM-------QNPRLAY-  214
MH0288_GL0082219                     SYAAYCEPRTF----------FDEGKEKNIIENIEKSHGEFLEYM-------KNPRLAY-  257
PIG-022_GL0026351                    SYEEFMNPEKL----------L-PGKDQSVYDNINRSRETFLEYR-------KNPRLGY-  203
PIG-028_GL0185479                    SYEEFISPEKL----------L-PGKEQSVYDNINRSQETFLEYR-------KNPRLGY-  205
250twins_35838_GL0110300             SYRNQDANRFE-------------------KD---GKKKAIDSLLD-----DTRAYYTY-  254
PIG-014_GL0226364                    ----------------------------------KSEEMVEDLLK-----NSSAYFVY-  195
250twins_36050_GL0158985             PMEAQQTYLLK-------------------GDKDMRRFEEVKQLLQ-----NTSAYYVY-  253
emb|OJMM01002900.1|                  -------------------------------------DSAYISKILN-----NTSAYFTY-  233
emb|OGPN01002610.1|                  DFDNQLKDK-------------------------PAFREAKDLLK-----NTEAYFSY-  256
emb|OHCP01000044.1|                  DYDRQKET-------------------------ETFLQAKSLLK-----QTEAYYAY-  251
PIG-046_GL0077813                    DFDGQKET-------------------------DTFRQARALLE-----RTEAYYSY-  255
pig_chimera                          DFDGQKET-------------------------DTFRQARALLE-----RTEAYYSY-  257
emb|OIEN01002196.1|                  ----------------------------------GISLREDNDYLK-----TMLPRCEYL  195
```

FIG. 18O

```
O2.UC29-0_GL0096317                ----------------------------------------DENLK-----RVLRRCMYL   188
CasR_Anaerobic_dig                 DYSSFMSPNKD----------------LEKEKKFSFHRENYKKFVE-----ASKPYMRY-   226
emb|OGDF01008514.1|                SYERLLDIADKYAVDGEDWRNTAAGISNDFEKKQFQTINGFWDLLD-----MIEPYMCY-   229
emb|OGZC01000639.1                 QWDEFQNPEQH-----REHF----GNKDNVICSVKKQQDLFFNFFK-------NNRIGY-   270
Ga0224415_10048792_chimera         TYADFSKDKSD-----K-------------------KYIEFQKLLK-------NKRLGY-   203
emb|OCVV012889144.1|               TYKQFKEPDK-------------------------YNKKFDDILN-------NSRLGY-   189
Ga0187910_10006931                 TFDEFAEPDRH-----KERFIKDGKLDTKLINQLKNQYDEFDAFLD-------DTRFGY-   212
R._flav_XPD_                       TYDEFKDPEHH-----RAAFN----NNDKLINAIKAQYDEFDNFLD-------NPRLGY-   207
Ga0224415_10007274                 TFAQFSASD--------------------------SSNDFEQFLK-------NPRLGY-   189
Ga0187911_10069260                 SYAKSKSKSKS------------------------KQQDFVELFI-----KKKEIHGY-   192
Ga0187910_10015336                 EYEQYCSEKSD------------------------YEDNFLKQLI------NNERIAY-   270
ODAI_chimera                       NYSDFCKNKPE------------------------DKERFFNFIGKGDGKVTNPRLSY-   270
31009_GL0034153                    SYEDFINKNL-----------------------VKQNENIDKFIE------NGNRLGY-   205
DLF014_GL0011914_                  GYNKVINKAIN-------------------DEKIKDRIEDLKEFIE------NGNRLGY-   250
CasR_R._albus                      TYKTFCDTSN---------------LPDDTKQKVENQKREFDKIIK-------SGRLGY-   251
Ga0187910_10040531                 GYDAFRNSTNQ---------------------KVQETYREFQEFVR-------RKELLY-   244
tpg|DJXD01000002.1|                DYQTLRGQTN-------------------------KYDRFKNYIK-------RKELLY-   190
CasR_E._siraeum                    TFDDFEKKKES-----T------------NSREKADFDAFEKFIG-------NYRLAY-   258
PIG-018_GL0023397                  TYDKFRGRSE------------------------RIVKSFNDFVS-------NPRLAY-   244
PIG-025_GL0099734                  TYDDYTSDSKN-------------------------KEIFEDFYS-------KRSLGY-   263
Ga0129317_1008067_chimera          TFDNYDPNHKN-------------------------NKDFFRFLK-------LPRLGY-   193
EYZ-362B_GL0088915                 SFDVFMHPSQS--------------SRADIYKKINDSRRRFENLLR-------TKRLGY-   257
uncultured_Ru_sp                   IYDVFIDPDN------S-------SLSDDKKANVRKSLSKFNALLK-------TKRLGY-   256
R._flav._FD-1                      TYDVFMDPSKN-----S-------SLSPKDRKNIDNSRAKFEKLLS-------TKRLGY-   260
```

FIG. 18P

```
tpg|DBYI01000091.1|                  TYEVFMEPSKN-----P-------RLDSNARKNIENSREKFEYLLD-------TQRLGY-    258

037_-_emb|OIZA01000315.1|            YENIFYDHG-------------------------------------------------    266
emb|OCTW011587266.1|                 FPDILCVLEKVNGKDR-----------------------------------------    216
k87_11092736                         FGSTDVFKVTKKRE-------------------------------------------    199
BMZ-11B_GL0037915                    FGSAFKPIPKKPGDKTSGKYAGGKLKER-----------------------------    222
BMZ-11B_GL0037771                    FGGIFRTQKKEDRKKDSNLSEEEKEEKKKK---------------------------    209
BMZ-11B_GL0069617                    LGAGFKQSERPKHQKNSRPDPRQKIE-------------------------------    201
Ga0099364_10024192                   FGEAFRPVGDDKVKEITLSDEVRKNIEKIIALEEQKRNPSTPRFKQENINLEIENAMGKF    244
530373_GL0023589                     YSRSYANKTSTAKKICICLQEDFHFFDYA------------TVKKAADQFTLLKRIQRS    258
emb|ODAI011611274.1|                 FGKEFCLVRNSPKQP-------------------------------------------    279
EMG_10003641                         LGNALYSDSLSNLTKDE-----------------------------------------    242
CasR_P1E0_metageno                   LGAAFYDVTPGKSKR------------------------------------------    222
emb|OIZX01000427.1|                  YSAAFYETLDNGK--------------------------------------------    253
160582958_gene49834                  FGELFYKPKETGDAK-----------------------------------------S    293
gi|1198542314|gb|NFIR01000008.1|     FGSTFCLFNENEERITS-------------------------------------    270
emb|OGNF01009141.1                   FGSAFYNTPEDTSAK-----------------------------------------    207
Ga0129306_1000735                    YGNAFFRKLSKAERLARG----REIFDK---------------ESPERRQEILGSRGKN    254
MH0288_GL0082219                     YGEAFYRRLSNNERSKLGMERYERMSDA---------------QRDAFDRKYPIDFCDG    301
PIG-022_GL0026351                    FGKVFYEKVTDPKNK-----------------------------------------    218
PIG-028_GL0185479                    FGKVFYEKSTDAKDK-----------------------------------------    220
250twins_35838_GL0110300             FDGLFSVPKREDDGKI----------------------------------------    270
PIG-014_GL0226364                    FGDVFKQVKLSKEQQE----------------------------------------    211
```

FIG. 18Q

250twins_36050_GL0158985	YGTLFEKVKAKSKKEQ----------	269
emb|OJMM01002900.1|	FDGVFKQITDRDSNK----------	248
emb|OGPN01002610.1|	FGDVFKKSKKGKKDE----------	271
emb|OHCP01000044.1|	FDDVFKKNKKPDKNKE----------	267
PIG-046_GL0077813	FDNVFKKIDKKKKRN----------	270
pig_chimera	FDNVFKKIDKNKKKS----------	272
emb|OIEN01002196.1|	FHNILNSDSDNNSKMN----------	211
O2.UC29-0_GL0096317	FDLGFEYAEPSPYKKS----------	204
CasR_Anaerobic_dig	YGKVFIRDVKKSKLST----------	242
emb|OGDF01008514.1|	FSEAFFCETTVKDPDS----------	245
emb|OGZC01000639.1	FGKAFFHAESE----------	281
Ga0224415_10048792_chimera	WGKAFFTGQGN----------	214
emb|OCVV012889144.1|	YGKAFFEKNDLKHNPN----------	205
Ga0187910_10006931	FGKAFFCKEGD----------	223
R._flav_XPD_	FGQAFFSKEGR----------	218
Ga0224415_10007274	LGKAFFYKDGK----------	200
Ga0187911_10069260	FGDTFAFLDKRI----------	204
Ga0187910_10015336	FGNAFFKDEGNK----------	282
ODAI_chimera	YGAAFAEKVQPT----------	282
31009_GL0034153	FGNVFYKNKNE----------	216
DLF014_GL0011914_	FGDIFFKKEEVSKNGK----------	266
CasR_R._albus	FGEAFMVNSGNS----------	263
Ga0187910_10040531	FGSAFYNGDT----------	254
tpg|DJXD01000002.1|	FGEAFYHENE----------	200

FIG. 18R

```
CasR_E._siraeum                  FADAFYVNKKNPKGK-----------------------------------------------  273
PIG-018_GL0023397                YSDAFYEKKGKG--------------------------------------------------  256
PIG-025_GL0099734                YGHIFFKDKK----------------------------------------------------  273
Ga0129317_1008067_chimera        FGSAFYSQKG-KD-------------------------------------------------  205
EYZ-362B_GL0088915               FGFAYNPKDR----------------------------------------------------  267
uncultured_Ru_sp                 FGLEEPKTKD-NR-------------------------------------------------  268
R._flav._FD-1                    FGFDYDANGKDKK-------------------------------------------------  273
tpg|DBYI01000091.1|              LSLEYDKRSKDKR----------------------------------------------------  271


037_-_emb|OIZA01000315.1|        --------------EPISEEKFYNYLNILNFIRNNTFHYKDN------------------  294
emb|OCTW011587266.1|             -------------YQPKSEKDAFNVLSSVNMLRNSLFHFAPKSND------------GKA  251
k87_11092736                     -------------ERAAADEHNAKVFRALGAIRQKLAHFKWK-----------------E  229
BMZ-11B_GL0037915                ---------------DAAIAEDVAVLRVLCGLRQISAHFRSEKKG--------------E  253
BMZ-11B_GL0037771                ----------QEELLQKDLQHNMDVARCISALRHATAHNKPATAHDK----------QDE  249
BMZ-11B_GL0069617                ---------------DPNWAYNKSALRILSTLRNKLAHFEDS------------------  228
Ga0099364_10024192               KSKDAFETAKKKYNRIVADETNAKTLRILGAMRQITAHFKDQ------------------  286
530373_GL0023589                 NYFFDMLCDSAGNYQ---CKNAQILLSILGNLRQESFHENRS-----------------T  298
emb|ODAI011611274.1|             ------------DKEEKRQYKLMRVLCLLGELRQFLVHGKKK---------------EKE  312
EMG_10003641                     ---------ILDGKRSKELKKYYQELCLLGMVRQSMIHSNQF------------------  275
CasR_P1E0_metageno               ---------DAARGRVKREQDVYAILSLMSLLRQFCAHDSVRIWG------------QNT  261
emb|OIZX01000427.1|              -------------SERRSNEDIFNILALMTCLRNFSSHHSIAIKVK-----------DYS  289
160582958_gene49834              DYKTHLSNNQKEEWELKSDKEVYNIFAILCDLRHFCTHGESIT--------------PSG  339
gi|1198542314|gb|NFIR01000008.1|  ---------ENKKEWKRFEKKCYHLLAVLGMMRQATAHGDSK-----------------R  304
```

FIG. 18S

```
emb|OGNF01009141.1                ----------ITKTKIKSNEEIYYTFMLLSTARNFSAHYLDRNRAKS----SDAEDFDGT   253
Ga0129306_1000735                 KSVDDEIRALAPEWVKREERDVYSELVLMSELRQSCFHGQQK----------------N   297
MH0288_GL0082219                  ELHFREIEKLAPEYVRKSEKEIYYILAMLSELRQCVFHDSAA----------------S   344
PIG-022_GL0026351                 ---------SKKKMRLISEKEMYYIFALLGGLRQFCTHDKNR----------------A   252
PIG-028_GL0185479                 ---------NKKKMRLISEKEMYYIFALLGGLRQFCTHDKNR----------------A   254
250twins_35838_GL0110300          ----------KESEKEKAKDQNFDVLRLLSVGRQLTFHSDKSNNE---------------   305
PIG-014_GL0226364                 -KSEEEKRRIWQELWNKNFDHNKNIVNSLIKFRQMCFHQNYENSD----------KEVS   259
250twins_36050_GL0158985          --------RAKEAEIDACTAHNYDVLRLLSLMRQLCMHSVAGTAF----------KLA   309
emb|OJMM01002900.1|               -------------DREIKNSYNALVLKVLYYLRQFCMHGNTYTKR--------NEESFLS   287
emb|OGPN01002610.1|               ----------NNEDYEKNLRHNFNVLRVLSFLRQICTHAYVKCTGGAKNNGDSTKVEAES   321
emb|OHCP01000044.1|               --------GDNSKQYQENLRHNFNILRVLSFLRQICMHAEVHVSD--DEGCTRTQNYTDS   317
PIG-046_GL0077813                 --------DDYEGKRNEILRYNFNVLRVLSFLRQICAHAKVKVSS--DNDREKGGVFVDS   320
pig_chimera                       --------DDCKRERNEILRYNFNVLRVLSFLRQICAHAQVKISN--EHDREKGGGLVDS   322
emb|OIEN01002196.1|               YNKVNKGKEEKDNRNNENIEKLKKALEVIKIIRVDSFHGVDGIKGD----------QKF   260
O2.UC29-0_GL0096317               EQ---QIEQEREEREKENLETLKKALDVISVMRTDSFHGYDMCNSTS--------AIVKF   253
CasR_Anaerobic_dig                ------GKGEKIEVMYRSDEEIFTIFQILSYVRQSIMHNDI----------------GN   279
emb|OGDF01008514.1|               -------GRIVPCLEQRSDGDIYNILRILSIVRQTCMHDNA----------------SM   281
emb|OGZC01000639.1                ----------RKIVKKTEKEVYHILTLIGSLRQWITHSTEG---------------GIS   315
Ga0224415_10048792_chimera        ------------NAKVRQENQCFHIIALLISLRNWATHSNELDK-------------HTK   249
emb|OCVV012889144.1|              -----KKKRDKNPYILKYDNECYYIIALLSGLRHWNIHSHAKDD-------------LVS   247
Ga0187910_10006931                ------------KYLNKQDNERYHILALLSGLRNWVVHNNEVES-------------KID   258
R._flav_XPD_                      ------------NYIINYGNECYDILALLSGLRHWVVHNNEEES-------------RIS   253
Ga0224415_10007274                ----------RNNRQKDPIECYHLLALLCGLRNWVVHNNEEKD-------------LIK   236
Ga0187911_10069260                -------------ADADKEKQVYAMLACLGSLRQACSHYRIRYSVNG------KNVDADA   245
```

FIG. 18T

```
Ga0187910_10015336          ------------KVMRTEKEIYYILGMLNEVRNVSTHFTEEDN-----------RDWA   317
ODAI_chimera                ------------LKYREAEDIYYIICMLGEARQMFAHGND----------------DRT   313
31009_GL0034153             ---------------LKDKKDIFDMLCLLGSLRQFCFHFKGAKFEGV---------EQIS   252
DLF014_GL0011914_           --------KPKKVEVLKSEKEIFDILALLGSLRQFCFHYDEAVFENEDGV---IDKEYND   315
CasR_R._albus               ------------TKLRPEKEIYHIFALMASLRQSYFHGYVKDT-----------DYQG   298
Ga0187910_10040531          ---------------RRDEKVIYHILSLAASVRQFCFHNDYTSDDGK---------GFIK   290
tpg|DJXD01000002.1|         ---------------RRYEEDIFAILTLLSALRQFCFHSDLSSDES----------DHVN   235
CasR_E._siraeum             -----------AKNVLREDKELYSVLTLIGKLRHWCVHSEE----------------GRA   306
PIG-018_GL0023397           ------------SKRRSEKEIYSMLAMLGKLRHWCVHSES----------------GEA   287
PIG-025_GL0099734           ---------------KRSKREIYDILALIATIRQWCIHSEE----------------DK   301
Ga0129317_1008067_chimera   ------------FEKRSDEEVYNICALMGQIRQCCFHGKQ---------------EKYQ   237
EYZ-362B_GL0088915          ----------------NQKKRLYYLMSMIGQLRQFAFHKDS----------------PFN   295
uncultured_Ru_sp            ------------VSQAYKKRVYHMLAIVGQIRQCVFHDKS----------------GAK   299
R._flav._FD-1               ------------KNEEIKKRLYHLTAFAGQLRQWSFHSAG----------------NYP   304
tpg|DBYI01000091.1|         ------------KSEEIKKRLYHLVAFAGQLRQWSFHSVE----------------GLP   302
                                                                    *    *

037_-_emb|OIZA01000315.1|   DIELYSENYS-E-----------------------EYVFINCLNEFVKNKFKNVNKN-F   328
emb|OCTW011587266.1|        RIAVFKNQFDSD-----------------------FSHITSTVNKIYSAKIAGVNEN-F   286
k87_11092736                SLAIFGANANMPI-------------RFFQGATGGRQLWNDVIAPLWKKRIERVRKS-F   274
BMZ-11B_GL0037915           SLWASDLGVEK-----------------------YFKNEM-SVVDANYKARAKKIDAN-F   288
BMZ-11B_GL0037771           YPWFKSSDIYET--------------------KIFKAGMW-SVIEADYKKKIQDVNKQ-F   287
BMZ-11B_GL0069617           -SILFCDNLKG------------------Q--FLGSNGNW-ITIETHYDEMITRINES-F   265
Ga0099364_10024192          -ATLFMSDVELPK-------------ILK--KEFSKADW-QTVEDYYAKLVDRINEG-F   327
```

FIG. 18U

```
530373_GL0023589                    SAWMFNLEKVLT------------------------PERKKLLTALIDSRTNHINN-HF   332
emb|ODAI011611274.1|                FAWLYRLDRQLS------------------------QEYRKLLGEFYDAQVDKVNK-SF   346
EMG_10003641                        NSSIYTLDSSYDSTMNTAELLGKGDDSSLVALATDARVEARAILDEIYKKGVDSINN-SF   334
CasR_P1E0_metageno                  TAALYHLQALP------------------------QDMKDLLDDGWRRALGGVND-HF   294
emb|OIZX01000427.1|                 AAGLYNLRRLG------------------------PDMKKMLDTFYTEAFIQLNQ-SF   322
160582958_gene49834                 KPFPYNLEKNLF------------------------PEAKQVLNSLFEEKAESLGAEAF   374
gi|1198542314|gb|NFIR01000008.1|    RAEIYKLGKEFDRL-----------------EARGCRPEARKELDELYKKKIHEMNQ-GF   346
emb|OGNF01009141.1                  SVIMYNLDNEELYK------------KLYNKKVHMALTGMKKVLDANFNKKVEHLNN-SF   300
Ga0129306_1000735                   SARIFRLDNDLG----------------------PGVDGARELLDRLYAEKINDLR--SF   333
MH0288_GL0082219                    RSWLYNLEKVLP------------------------AEAMQVLDGVYSQKVAALN--NF   377
PIG-022_GL0026351                   RNWLYSLDTMLN------------------------DDAKAVLDKYYVSTVSSLNK-HF   286
PIG-028_GL0185479                   RNWLYSLDTMLN------------------------DDAKAVLDKYYVSTVSSLNK-HF   288
250twins_35838_GL0110300            -AYLFDLSKLTRA--------------AQDENRRQDIQSLLNILNSTCRSNLEGVNGD-F   349
PIG-014_GL0226364                   STALFNIEQFL------------------------NKEEKASLDEIYTGAIDKVNCD-F   293
250twins_36050_GL0158985            ESALFNIEDVL------------------------SADLKEILDEAFSGAVNKLNDG-F   343
emb|OJMM01002900.1|                 DTALYNAKEFFAK----------------------ADPQINELIDAVYADGIKTINSD-F   324
emb|OGPN01002610.1|                 LDALFNITEYFAK----------------------TAPELSKTINEIYKEGIDRINND-F   358
emb|OHCP01000044.1|                 LEALFNISKAFGK----------------------KMPELKTLIDNIYSKGINAINDE-F   354
PIG-046_GL0077813                   LDALFNISSFFDA----------------------VAPELNEVINSVYSKGIYDINDN-F   357
pig_chimera                         LDALFNISRFFDA----------------------VAPELNEVINSVYSKGIDDINDN-F   359
emb|OIEN01002196.1|                 PRSKYNLAVN------------------------YNEEIQKTISEPFNRKVEEVQQD-F   294
O2.UC29-0_GL0096317                 PSKKFNLDQ--------------------------KGKNLIYRNAFKYKTDTIKKQ-Y   284
CasR_Anaerobic_dig                  KSSILAIEK------------------------YPARFVGFLSDLLKTKTNDVNRM-F   312
emb|OGDF01008514.1|                 RTVMFTLGQNSVR----------------DRKNGFDELAELLDYLYDEKIDIVNRD-F   322
```

FIG. 18V

```
emb|OGZC01000639.1                RLWLYQLEDA-----------------------LSREYQETMNNCYNSTIYGLQKD-F   349
Ga0224415_10048792_chimera        RTWLYKLDDTNI---------------------LNAEYVKTLNYLYDTIADELTKS-F   285
emb|OCVV012889144.1|              YRWLYNLDSI-----------------------LNREYISTLNYLYDDIADELTES-F   281
Ga0187910_10006931                RKWLYNLDKN-----------------------LDKEYITTLDYMYSDIADELTKS-F   292
R._flav_XPD_                      RTWLYNLDKN-----------------------LDNEYISTLNYLYDRITNELTNS-F   287
Ga0224415_10007274                YTWLYNLDKY-----------------------LDAEYITTLNYMYNDIGDELTDS-F   270
Ga0187911_10069260                DTWLFSSAQLDQ---------------------TDPLFSEMLNRIYSHKIKTVNQN-F   281
Ga0187910_10015336                KANLYNLSNR-----------------------LKVGSKEVLNQLYKEKIDKIDANGF   352
ODAI_chimera                      KTFLYNIVGYRN-------------------GTEINSKFTAVLDGIYSKKVEELNGN-F   352
31009_GL0034153                   NSWLYNLSQ------------------------LHNEFKSTLDSFYNEKIEDINGN-F   285
DLF014_GL0011914_                 NSWLYNLEK------------------------LEPEFKDTLDYFYNRKINDINKG-F   348
CasR_R._albus                     TTWAYTLEDKLK--------------------GPSHEFRETIDKIFDEGFSKISKD-F   335
Ga0187910_10040531                ADWMYRLEEA-----------------------LPAEYKDTLDALYLEGVEGLDQS-F   324
tpg|DJXD01000002.1|               SFWLYQLEDQ-----------------------LSDEFKETLSILWEEVTERIDSE-F   269
CasR_E._siraeum                   EFWLYKLDE------------------------LKDDFKNVLDVVYNRPVEEINNR-F   339
PIG-018_GL0023397                 EFWLYKLDS------------------------LNGEFIDVLNTLYNRSVNEINRG-F   320
PIG-025_GL0099734                 KTWIYNAENV-----------------------LDSEYLDILDDVYGGLVEKINKN-F   335
Ga0129317_1008067_chimera         LKWLYNFHNFK----------------------SNKPFLDTLDKHFDEMIDRINKN-F   272
EYZ-362B_GL0088915                RVWLYQLDEK-----------------------LSKEHKDTLNYFFDQRYDEIHEN-F   329
uncultured_Ru_sp                  RFDLYSFINN-----------------------IDPEYRDTLDYLVEERLKSINKD-F   333
R._flav._FD-1                     RTWLYKLDS------------------------LDKEYLDTLDHYFDKRFNDINDD-F   337
tpg|DBYI01000091.1|               RTWIYQLDNPK----------------------LAQEYRDTLDYFFNERFDAINKD-F   337
                                                                                  :    :
```

FIG. 18W

```
037_-_emb|OIZA01000315.1|          I-SNE-KNNLYIILKAYGEDTENVEVVKKYSKELYKLSVLKTNKNLGVNVKKLRESAIEY   386
emb|OCTW011587266.1|               L-NNE-GNNLYIILKATNWD------IKKIVPQLYRFSVLKSDKNMGFNMRKLREFAVES   338
k87_11092736                       L-SNS-AKNLWVLYQVFKDDTD--EKKKARARQYYHFSVLKEGKNLGFNLTKTREYFLDK   330
BMZ-11B_GL0037915                  M-ATS-RANLVLLFRILNAKTK--EEKEKIAVDYYKFAIRKEGKNLGVNLKKAREILLDK   344
BMZ-11B_GL0037771                  F-SKN-KVNLAILFDLLDVRDV--KQKKRISDEFYRFTIRKDGKNLGMNLVKIREIIIDR   343
BMZ-11B_GL0069617                  F-NNS-RMNLKILYLVLNAATL--EDKKKITEEYYRFSILKEGKNLGVNMRKLREMMVDM   321
Ga0099364_10024192                 C-KNA-ATNVHFLTELLPE-----ESKKQLTEDYFRFAILKEGKNLGVNMKRLREVMFAL   380
530373_GL0023589                   I-ENS-NKNLNAIFRLFPEKS-----QADLIVEYYDFSVRKAYKNLGFSVKLLREMLLKT   385
emb|ODAI011611274.1|               L-TNS-TVNLEVLFRALKTGTD--PERKTVTQEYYQFTVRKEDGNLGFSLKTLREILLSA   402
EMG_10003641                       L-SNS-INDLENLFKIYKCDSS--EKKTELIKQYYDFCIRKPQMNMGFSITTIREGMFTR   390
CasR_P1E0_metageno                 L-DTN-KVNLLTLFEYYGAETK--QARVALTQDFYRFVVLKEQKNMGFSLRRLREELLKL   350
emb|OIZX01000427.1|                Q-DHN-TTNLTCLFDILNISDS--ARQKQLAEEFYRYVVFKEQKNLGFSVRKLREEMLLL   378
160582958_gene49834                G-KTAGKTDVSILLKVFEKEQASQKEQQALLKEYYDFKVQKTYKNMGFSIKKLREAIMEI   433
gi|1198542314|gb|NFIR01000008.1|   L-KNS-KSDILMLFRIYNAESK--EAKRKLAQEYYEFIMLKSYKNTGFSIKHLRETMIDK   402
emb|OGNF01009141.1                 I-KNS-AKDFVILCEVLGIKSR--DEKTKFVKDYYDFVVRKNYKHLGFSVKELRELLFAN   356
Ga0129306_1000735                  D-KTSASSNFRLLFNAYHADN---EKKKELAQEFYRFSVLKVSKNTGFSIRTLREKIIED   389
MH0288_GL0082219                   S-KTSSKSNFRLLFNALHADTD--AQKKQIAEDFYKFSVRKTYKNIGFSIKLLRESMLDT   434
PIG-022_GL0026351                  V-DNSAKKDFSIIFNALNITDI--ELKKNIAKEYYEFVCRKNFKNIGFSIKKLREYLLEK   343
PIG-028_GL0185479                  V-ENSAKKDFSIIFNALNITDI--ELKKNIAKEYYEFVCRKNFKNIGFSIKKLREYLLEK   345
250twins_35838_GL0110300           V-KHA-KNNLYVLNQLYPS-----LKANDLIGEYYNFIVKKENRNIGIRLITVRELIIEH   402
PIG-014_GL0226364                  N-KNS-SNNLYILEQIYCN-----D--KNLGQRYYDFVVKKSQCNLGLDTKLIRELLIRK   344
250twins_36050_GL0158985           V-QHS-GNNLYVLQQLYPN-----ETIERIAEKYYRLTVRKEDLNMGVNIKKLRELIVGQ   396
emb|OJMM01002900.1|                M-AHA-KNNMYIICEVYKN-----EAEDSLMKEYYDFVVRKEGNNLGFNTRQLREILIDK   377
emb|OGPN01002610.1|                V-TNG-KNNLYILSKVYPD-----MQRNELVKKYYQFVVCKEGNNVGINTRKLKESIISQ   411
```

FIG. 18X

emb|OHCP01000044.1|	V-KNG-KNNLYILSKVYPN-----EKREVLLREYYNFVVCKEGSNIGISTRKLKETMIAQ	407
PIG-046_GL0077813	V-KND-KNNFYILSKIYPE-----VARENLLKEYYYFVVSKEGNNIGISTKKLKEAIIAQ	410
pig_chimera	V-KNG-KNNFYILSKIYPE-----VAREDLLREYYYFVVSKEGNNIGISTKKLKEAIIVQ	412
emb|OIEN01002196.1|	Y-RNS-CVNIDFLKEIMYGSNYTDRGSDSLECSYFNFAILKQNKNMGFSITSIRECLLDL	352
O2.UC29-0_GL0096317	A-QNS-KVNINFLSYVLYGNGCTEEQKKELSREYFNFAILKQYKNLGFSITSLREQLLSL	342
CasR_Anaerobic_dig	IDNNS-QTNFWVLFSIFGL-QDHTSGADKICRNFYDFVIKADSKNLGFSLKKIRELMLDL	370
emb|OGDF01008514.1|	L-RNQ-KNNIELLSRIYGS-SADSPERDRLVQNFYDFRVLSQDKNLGFSIKKLREKLLDS	379
emb|OGZC01000639.1	E-KTN-APNLNFLAEILGK-N-----ASELAEPYFRFIITKEYKNLGFSIKTLREMLLDQ	401
Ga0224415_10048792_chimera	S-KNG-AVNVNYLAKKYNI-K---DDLPGFSEQYFRFSIMKEQKNLGFNISKLRENMLDF	339
emb|OCVV012889144.1|	S-KNS-SANVNYIAETLNI-D-----PSEFAQQYFRFSIMKEQKNMGFNVSKLREIMLDR	333
Ga0187910_10006931	S-KNS-AANVNYIAEILNI-D-----SKTFAEQYFRFSIMKEQKNLGFTLTKLRECMLDR	344
R._flav_XPD_	S-KNS-AANVNYIAETLGI-N-----PAEFAEQYFRFSIMKEQKNLGFNITKLREVMLDR	339
Ga0224415_10007274	S-KNS-AANINYIAETLGI-D-----PKTFAEQYFRFSIMKEQKNLGFNLTKLREVMLDR	322
Ga0187911_10069260	FENNR-KANFPILKKMYPE-----TTLKVLMNEYYDFSIRKGYKNFGFSIKSLREALLSP	335
Ga0187910_10015336	VNKGC-KRDFSILFKSLNL-T-TDKDKGELVVGFYDFSIRKNYKNIGFSIKTLREYMLKI	409
ODAI_chimera	I-KNS-KKDLTILFSAFGC-Q-DEASKKKLIEKYYDFIVKKTNKNMGFSVTKLREKIIEG	408
31009_GL0034153	I-EKN-KVNIQILCNDLKL------DKNKLVQDYYNFLISKNHKNMGFSIKKIREYILDI	337
DLF014_GL0011914_	I-KTN-QINLHIICDELGL-N---WDKEQVVGDYYDFLIAKKYKNIGFSIKKIREHMFDI	402
CasR_R._albus	G-KMN-KVNLQILEQMIGE-LYGSIERQNLTCDYYDFIQLKKHKYLGFSIKRLRETMLET	392
Ga0187910_10040531	L-KNN-TVNIQILCSIFNH-----DDPNKIAEEYYGFLMTKEYKNMGFSIKKLRECMLEL	377
tpg|DJXD01000002.1|	L-KTN-TVNLHILCHVFPK-----ESKETIVRAYYEFLIKKSFKNMGFSIKKLREIMLEQ	322
CasR_E._siraeum	I-ENN-KVNIQILGSVYKN-----TDIAELVRSYYEFLITKKYKNMGFSIKKLRESMLEG	392
PIG-018_GL0023397	V-DTN-KVNIQILQAVCKN-----MDLISLIKAYYEFLITKKYKNMGFSIKKLREMMLAE	373
PIG-025_GL0099734	I-KDN-KVNLILLSDILDM-K-SKDDFEELIRQYYRFIITKEQKLLGFSITKIREAMLED	391

FIG. 18Y

```
Ga0129317_1008067_chimera          I-KNN-TPDLIILSGLYPD-----MAKKELVRLFYDFTTVKEYKNMGFSVKKLREKMLES   325
EYZ-362B_GL0088915                 N-KTN-KVNLELLAEAFPS-----EDYATLANQYYDFTVFKAHKNLGFSIKKLREKLIEN   382
uncultured_Ru_sp                   I-EDN-KVNISLLIDMMKG-----YEADDIIRLYYDFIVLKSQKNLGFSIKKLREKMLDE   386
R._flav._FD-1                      V-TKN-ATNLYILKEVFPE-----ANFKDIADLYYDFIVIKSHKNMGFSIKKLREKMLEC   390
tpg|DBYI01000091.1|                I-ETN-NINLHILKEVFPA-----EDFQKLAALYYDFIVKKTFKNIGFSIKNLREQMLEC   390
                                          :.  :                    :          *.     :*  .

037_-_emb|OIZA01000315.1|          GYCPLP-----YDKEKEVAKLSSVKHKLYKTYDFVITHYLNSN----------------   424
emb|OCTW011587266.1|               KNIDL--------SRLNDKFLTNNRKKLYKVIDFIIYYHLNKVL---------------K   375
k87_11092736                       FFPIFH-----SSAPDVKRKVDTFRSKFYAILDFIIYEASVSV-----ANSGQMGKVAPW   380
BMZ-11B_GL0037915                  HYPD-----------AKNKQFDSYRGKLYTITDFIIYRMAAED----------------   376
BMZ-11B_GL0037771                  YASG-----------LRDKKHDPHRQKINVIADFLIFRALSQN----------------   375
BMZ-11B_GL0069617                  FYSE-----------IKDKQYDSYRQKIYTILDFLLFREIKDS----------------   353
Ga0099364_10024192                 FVPELT-------APETKKRYDSYRAKIYGLTDFLLFKHIHNT----------------   416
530373_GL0023589                   QQELS---------PLNDKAYDSIRSKLYTMFDFVLYKLYVD----------------N   419
emb|ODAI011611274.1|               YKH-----------EVRDKEYDSIRHKLYQLFSFALYHYYKTG---------------VG   436
EMG_10003641                       CSEANTLLLCDEGSTVKLNVHDTMKSKFYKNLDFMIYKYYKYE---------------N   434
CasR_P1E0_metageno                 PDAA----------YLTGQEYDSVRQKLYMLLDFLLCRLYAQE---------------R   384
emb|OIZX01000427.1|                PDAA----------VIADKRYDTCRSKLYNLMDFLILRVYRTG---------------R   412
160582958_gene49834                PDAA----------KFKDDLYSSLRHKLYGLFDFILVKHFLDT------------SDSE   470
gi|1198542314|gb|NFIR01000008.1|   MDEDK-------KEKLKDDKYNPIRRKIYRIMDFMIYQYYQEP---------------EH   440
emb|OGNF01009141.1                 HDSNKYIK---EFDKISNKKFDSVRSRLNRLADYIIYDYYNK----------------N   396
Ga0129306_1000735                  HAA-----------QYRDKIYDSMRKKLFSTFDFFLWRFYEE----------------R   421
MH0288_GL0082219                   FAP-----------EYKEKSYDSVRGKLYQLVDFLIWNIYDK----------------N   466
```

FIG. 18Z

PIG-022_GL0026351	PEFC----------NLKEDKYNSVRSKLYRLIDFIIYLNYKDAKF----------ENDN	382
PIG-028_GL0185479	PEFR----------NLKDDKYNSVRSKLYRLIDFIIYLNYKDAKF----------KNDN	384
250twins_35838_GL0110300	NYT----------NLKDSKYDTYRNKIYTVLNFILFREIQEN----------	434
PIG-014_GL0226364	LSD--------EGLDINDSRYDTYRNKIYTVFGFILYYELTKN----------	379
250twins_36050_GL0158985	YFPE--VLDKEYDLSKNGDSVVTYRSKIYTVMNYILLYYLEDH----------	437
emb|OJMM01002900.1|	YVG----------NLRGKKYNTFRNKLYTVLGFILVKEIKRN----------	409
emb|OGPN01002610.1|	HPWI--TTPQD---NNKANDYESCRHKLYTIMCFILVAELDAH----------	449
emb|OHCP01000044.1|	NMP----------SLKEENTYRNKLYTVMNFILVRELKNC----------	437
PIG-046_GL0077813	DMS----------YIKSKDYDTYRNKLYTVLCFILVKELNER----------	442
pig_chimera	DMS----------YIKSEDYDTYRNKLYTVLCFILVKELNER----------	444
emb|OIEN01002196.1|	YEL----------NFESMQNLRPRANSFCDFLIYDYYCKNE----------	383
O2.UC29-0_GL0096317	YDI----------DFECKRNLRSRANSFCDYVLYYRYAVQF----------	373
CasR_Anaerobic_dig	PNA----------NMLRDHQFDTVRSKFYTLLDFIIYQHYLEE----------	403
emb|OGDF01008514.1|	PAL----------SVVRSKKYDTMRSKIYSLIDFMIYRKFSEN----------	412
emb|OGZC01000639.1	PDLQ--E------IRENHNVYDSIRSKLYKMIDFVLVYAYSNERK----------	438
Ga0224415_10048792_chimera	KDMS--V------IRDDHNRYDKDRSKIYTMMDFVIYRYYIDNNN----------	376
emb|OCVV012889144.1|	KELS--D------IRDNHRVFDSIRSKLYTMMDFVIYRYYIEEAAKTEAENRNLPENEKK	385
Ga0187910_10006931	EELS--D------IRDNHKVFDSIRSKLYTMMDFVIYRYYIEEAKKIENENKTLSDDKKK	396
R._flav_XPD_	KDMS--E------IRKNHKVFDSIRTKVYTMMDFVIYRYYIEEDAKVAAANKSLPDNEKS	391
Ga0224415_10007274	KDMS--E------IRENHNDFDSIRAKVYTMMDFVIYRYYIEEAAKVNAANKSLPDNEKS	374
Ga0187911_10069260	QYES--LIGV---QIKDNKEYDTVRSKLYQLFDFALTRYFNQH----------	373
Ga0187910_10015336	SNS------T---LCADTISNNAIRPKAYKLYDFIIWHYYMNK----------	443
ODAI_chimera	YRFR--G------NPVTGTDYDSVRHKLYRTFDFIICSYYDDH----------	443
31009_GL0034153	DYA----------SYIKEQGYDSVRSKLFTLIDFIIYDLYVNKE----------	371

FIG. 18AA

```
DLF014_GL0011914_                YEA----------FDIKDKRFDSVRSIMYKIIDFVIYYSFIHYK----------------  436
CasR_R._albus                    TPA----------ECYKAECYNSERQKLYKLIDFLIYDLYYNRK----------------  426
Ga0187910_10040531               PEL----------SGYKEDQYNSVRSKLYKLFDFIIAHYFRKH----------------  410
tpg|DJXD01000002.1|              SDL----------KSFKEDKYNSVRAKLYKLFDFIITYYYDHH----------------  355
CasR_E._siraeum                  KGY-------------ADKEYDSVRNKLYQMTDFILYTGYINED----------------  423
PIG-018_GL0023397                TAYT--SNPD---DRLYDKEINSVRNKLYQMTDFIIYLEYRDN----------------  411
PIG-025_GL0099734                TV------------FKTDKKYDSIRSKLYKLIDFMMFYKYTTVDSE--------------  425
Ga0129317_1008067_chimera        EEA----------SDFRDKDYDSVRRKLYKLMDFCIYYLYYS-D----------------  358
EYZ-362B_GL0088915               --------------ALRDEKIDSFRSKLYRLIDFCLFRIYLK-D----------------  411
uncultured_Ru_sp                 YG-----------FRFKDKQYDSVRSKMYKLMDFLLFCNYYRND----------------  419
R._flav._FD-1                    DGA----------DRIKEQDMDSVRSKLYKLIDFCIFKYYHE-F----------------  423
tpg|DBYI01000091.1|              DEA----------EKIRSKDMNSVRSKLYKLFDFCIFYQYFI-D----------------  423
                                                        :       : :

037_-_emb|OIZA01000315.1|        DKLLLEIVETLRLSK-----ND-DEKENVYKKYA---E-KLFKADDVINPIKAISKLFAE  474
emb|OCTW011587266.1|             DSFVDDFVAALRASQ-----SE-EEKEKLYAQYS---E-RLFADEGLKSAIKKAVDMISD  425
k87_11092736                     KGAIDNALVKLREAP-----DE-EAKEKIYNVLA---A-SIRNDSLFL--RLKSACDKFG  428
BMZ-11B_GL0037915                GAFLEESVQKLRSTP-----SE-EEKEKVYDVLA---E-TIWNATKDD--VKKVFDIRQQ  424
BMZ-11B_GL0037771                QGIIDKTVSSLRLTK-----DE-EEKDHVYQNAA---E-LVWGMVSNC--LTPYFNDPKN  423
BMZ-11B_GL0069617                -EELETMVADLRESE-----SD-EKKEELYHDYA---E-LFWDGVKDL--IVPFFAAFDG  400
Ga0099364_10024192               -KQLEEWVAVLRETS-----NE-DAKENLYDEFA---R-TAWNTVGDS--AKQLIENMQS  463
530373_GL0023589                 EEQIDQFVRKLRAST-----SN-VEKEALYQQES---E-RTWNEIRGQ--IIDNVIPFID  467
emb|ODAI011611274.1|             AERREAFVAKLRAVM-----TA-EAKQRAYADEA---A-EIWNDEGSG--IRAAFLEILE  484
EMG_10003641                     PEKGEKLIEDLRSKIKGKKKED-EDKKQRYAEES---A-CILKAKRDI--IKKDLTEAAN  487
```

FIG. 18BB

CasR_P1E0_metageno	ADRCEELVSALRCAL-----SD-EEKDTVYQAEA---A-ALWQALGDT--LRRKLLPLLK	432
emb|OIZX01000427.1|	ADRCDKLPEALRAAL-----TD-EEKAVVYHKEA---L-SLWNEMRTL--ILDGLLPQMT	460
160582958_gene49834	NLQNNDIFRQLRACR-----CE-EEKDQVYRSIA---V-KVWEKVKKK--ELNMFKQVVV	518
gi|1198542314|gb|NFIR01000008.1|	QEEAEELVRKLRNAE-----IE-AKKELAYRKEA---E-KLKKELEKI--IFNSVLPSCD	488
emb|OGNF01009141.1	NAKVSDLVKYLRAAA-----DD-EQKKKIYLNES---INLVKSGILER--IKKILPKLNG	445
Ga0129306_1000735	EDEAEELRACLRAAR-----SD-EEKEQIYAEAA---A-SCWPSVKPF--VESVAATLCD	469
MH0288_GL0082219	EKMADALVAQLRAEVK---RDR-DEKSPVYDREA---A-RLWTMIREA--VAQTAKTLRA	516
PIG-022_GL0026351	DEKIERFVSKLRKCP-----DE-ASKETVYQTES---D-IIKKSIYGS--INSIKRGMM-	429
PIG-028_GL0185479	DEKIERFVSKLRKCP-----DE-ASKETVYQTEA---D-IIKKSISGS--INSIKRDMM-	431
250twins_35838_GL0110300	SIAIKNFREKLRSTE-----K--AEQPALYQAFA---N-KIYPMVQAK--FAKAIDLFEE	481
PIG-014_GL0226364	ETFCQEIVAKLRRFG-----GDEIEKNKIYNEYFLRKE-AVYNQIAEK--FKQSVCWFKK	431
250twins_36050_GL0158985	DSSRESMVEALRQNR-----EGDEGKEEIYRQFA---K-KVWNGVSGL--FGVCLNLFKT	486
emb|OJMM01002900.1|	PKIQDSFIAKLRANQ-----NGDEGKLNIYNEFA---P-KIWSVVSSK--LNSAITCFDE	458
emb|OGPN01002610.1|	ESIRDNMVAELRANM-----DGDDGRDAIYEKYA---K-DIYHIVKDK--LLAMQKVFDE	498
emb|OHCP01000044.1|	ATIREQMIKELRANM-----DEEEGRDRIYSKYA---K-EIYLYVKDK--LKLMLNVFKE	486
PIG-046_GL0077813	TTIREQMVADLRANM-----NGDIGREDIYSKYA---K-IIYAQVKPR--FDGMKNVFEE	491
pig_chimera	TTIREQMVADLRANM-----NGDIGREDIYSKYA---K-IIYAQVKPR--FDTMKSAFEE	493
emb|OIEN01002196.1|	-SERANLVDCLRSAA-----SE-EEKKNIYFQTA---E-RVKEKFRNA--FNRISRFDAS	430
O2.UC29-0_GL0096317	PERIVDLVDKLRCSA-----NE-KEKTNVYIEEA---N-RIKEDLRIA--FNRINNFDAK	421
CasR_Anaerobic_dig	KSRIDNMVEKLRMTL-----KE-EEKEVLYAAEA---K-IVWNAIGAK--VINKLVPMMN	451
emb|OGDF01008514.1|	HVAVDDFVEELRSLL-----TE-DEKESAYSRWA---ETLINDGFAQE--ILVKLLPQTD	461
emb|OGZC01000639.1	-SKADALASNLRSAI-----TE-DAKKRIYQNEA---D-QLWTSYQEL--FKRIRGFKGA	485
Ga0224415_10048792_chimera	--DSIDFINKLRSSI-----DE-KSKEKLYNEEA---N-RLWNKLKEY--MLYIKEFNGK	422
emb|OCVV012889144.1|	ISEKDFFVINLRGSF-----DE-NQKEKLYIEEA---K-RLWEKLKDI--MLKIKEFRGE	433

FIG. 18CC

```
Ga0187910_10006931          LSEKDIFIISLRGSF-----SE-EQKDKLYSDEA---E-RLWAKLGKL--MLEIKKFRGQ  444
R._flav_XPD_                LSEKDIFVINLRGSF-----ND-DQKDALYYDEA---N-RIWRKLENI--MHNIKEFRGN  439
Ga0224415_10007274          LSEKDIFVISLRGSF-----NE-DQKDRLYYDEA---Q-RLWSKVGKL--MLKIKKFRGK  422
Ga0187911_10069260          PDMVDAFVVELRSLA-----KDEDAKNAVYEKYA---K-AVWNDVKQP--IAVMLSYMNG  422
Ga0187910_10015336          PDKINDFVEKLRTQN-----KN-DEKIKLYYDEA---V-CLLSELGRE--IHTMTSCVHN  491
ODAI_chimera                AKEKDTLVSSLRASL-----ND-TKKENLYAGEA---K-TIAESIKDR--MDTVCRMYNK  491
31009_GL0034153             KDRIDDIVSKLRGAI-----YD-YKKDDIYKAEA---D-WVWKKLNRQ--IKALTKLLSN  419
DLF014_GL0011914_           QDEVDEIVNYLRSTL-----SE-EQKKEKYIKLA---G-NIWAEYKTQ--FNKLKDILLA  484
CasR_R._albus               PARIEEIVDKLRESV-----ND-EEKESIYSVEA---K-YVYESLSKV--LDKSLKNSVS  474
Ga0187910_10040531          PEKGEEMVDCLRLCM-----TE-DEKDSHYEGTA---K-KLVRELAYD--MQEAAEQANG  458
tpg|DJXD01000002.1|         AFEKEALVSSLRSSL-----TE-ENKEEIYIKTA---R-TLASALGAD--FKKAAADVNA  403
CasR_E._siraeum             SDRADDLVNTLRSSL-----KE-DDKTTVYCKEA---D-YLWKKYRES--IREVADALDG  471
PIG-018_GL0023397           KSIIDTLVSKLRSSV-----TD-EAKEAVYREEA---A-NLRGRLLDA--VALIAKSLNG  459
PIG-025_GL0099734           KEHADSIVEKLRATL-----SG-ESKEEIYKAEA---N-RIWNLYEDI--IMNKIEPALD  473
Ga0129317_1008067_chimera   SERNENLVSRLRESL-----TD-ENKDIIYSKEA---K-IVWNELRKK--FSTILDNVKG  406
EYZ-362B_GL0088915          PDRAEKNVRFLRSCI-----NE-QEKEQFYTQEA---N-ALKSENPGL--VNDFCMQLKD  459
uncultured_Ru_sp            IAAGESLVRKLRFSM-----TD-DEKEGIYADEA---A-KLWGKFRND--FENIADHMNG  467
R._flav._FD-1               PELSEKNVDILRAAV-----SD-TKKDNLYSDEA---A-RLWSIFKEK--FLGFCDKIVV  471
tpg|DBYI01000091.1|         EERSRENVNYLRSTL-----ND-EQKDAFYEEEG---K-RLWSENRKK--FIYFCDNINK  471
                                      **            :   *

037_-_emb|OIZA01000315.1|   KGNKLFKEK---------------------IIIKKEYIEDVSIDKNIYDFTKVIFFMTCF  513
emb|OCTW011587266.1|        TKSNIFKMK---------------------TPLDKALIENIKVNSDASDFCKLIYVFTRF  464
k87_11092736                AEQNRPVFPNELR-------------NNRDIRNVRSEWLEATQDVDAAAFVQLIAFLCNF  475
```

FIG. 18DD

```
BMZ-11B_GL0037915                  DIKNAAAQS----------------------LDDVNLDSVTVEKEKEFPVVKLLAFLCNF   462
BMZ-11B_GL0037771                  KYILKYKDA----------------------KTPGDFEDWITSKISEDDGEPFVKVLSFLCNF   464
BMZ-11B_GL0069617                  DFTVFTTEK----------------------VPADTSKEPFLSSANTEPLVKLLAFLCNF   438
Ga0099364_10024192                 YFTKKEKEIT--K------------TAQPVLSTSSIAHTSKKITQFSSFAKLLAFLCNF   508
530373_GL0023589                   SIQKQENEDKLTKR-------------IDQSIIDKAVPHFVNSRDLSYFAKVVYVVASF   513
emb|ODAI011611274.1|               AVDFGSAVKGIKARSS----------VAGDKRFAEWLEEVRIRPEGVSCFTKLMYLLTRF   534
EMG_10003641                       KDLFADLVKSNKNEKQK------------FKNEYEELLKPFMIPVKVDYFSELIYLVTRF   535
CasR_P1E0_metageno                 GKKLQDKDK-KKSDELGLSRDVLDGVLFR--------PAQQGSRANADYFCRLMHLSTWF   483
emb|OIZX01000427.1|                PENLSRLSGQKRKGELSLDDAMLKECLYE---PGPVPEDAAPEEANAEYFCRMIYLATLF   517
160582958_gene49834                IPSLSKDE----------------LKQMEMTKNTELLSSIETISTQASLFSEMIFMMTYL   562
gi|1198542314|gb|NFIR01000008.1|   RILSEMDERRNKKVNQESSDTDKEEPL-----DSEIAEGITFIKETAHSFSEMIYLLTVF   543
emb|OGNF01009141.1                 KIIGNMQPDSTI----------------TASMLHNTGKDWHPISENAHYFTKWIYTLTLF   489
Ga0129306_1000735                  VVKGR-------------TKLNKLKLSADESTLVRNAIDGVRISPRASYFTKLIYLMTLF   516
MH0288_GL0082219                   VTLSNNTPGSIKIKELALSKYDRDLVENAIDSAVDERGRRLKITDEASYFSKLIYLLTLF   576
PIG-022_GL0026351                  ---------------SVKSKYGKLELDKQTINQLKRAVDDISINTKGDYFTKLMYLISLL   474
PIG-028_GL0185479                  ---------------LVKSKYGKLELDTQTVNQLKRAVDDISINTKGDYFTKLMYLISLL   476
250twins_35838_GL0110300           QYKTKFKSEFKGGI----------------SIENMQQQNILLQTENIDYFSKYVLFLTKF   525
PIG-014_GL0226364                  EKEKKKEFSS--------------------KTKIDLRSVQKIEDQNASVFAKMVYLISRF   471
250twins_36050_GL0158985           EKRNKFRSKV--------------------ALPDVSGAAYMLSSENIDYFVKMLFFVCKF   526
emb|OJMM01002900.1|                ESLSKFKGYK--------------------DIDESLISRYGITVANTDTLVKILYFLCKF   498
emb|OGPN01002610.1|                ELVPVKVEGKNDPQ--Q---------FTHGKLGKKEIESFCLSDKNTSDIAKVVYFLCNF   547
emb|OHCP01000044.1|                EAEGIIIPGKEDPV--K---------FSHGKLDKKEIESFCLTTKNTEDITKVIYFLCKF   535
PIG-046_GL0077813                  EAKGVIVPDKEEAV--K---------FSNGALDKSEIERFCITSANTDSVAKIIYFLCKF   540
pig_chimera                        EAKDVIVPDKKKPV--K---------FSHGKLDKNEIERFCITSANTDSVAKIIYFLCKF   542
```

FIG. 18EE

```
emb|OIEN01002196.1|              YIKNSREKNLSGGS--SLPK-----YSFIEGFTKRSKKINDNDEKNADLFCNMLYYLAQF  483
O2.UC29-0_GL0096317              SIKLYSTKNNVDDQ--PMLP-----FEQVSK----DIKGGDAAREKADAFCDMTYYLTQF  470
CasR_Anaerobic_dig               GDALKEIKRKN-----------------RDRKLPQSVIATVQVNSDANVFSGLIYFLTLF  494
emb|OGDF01008514.1|              PAVIGKIKG---------------------KKLLNDSIAGIKLKKDASFFTKIINVLCMF  500
emb|OGZC01000639.1               QVKEYSSKN-M--------------P----IPIQKQIQNILKPAEQVTYFTKLMYLLTMF  526
Ga0224415_10048792_chimera       LASRTPDRD-G--------------NISEFVESLPKIHRLLPRGQKISNFSKLMYLLTMF  467
emb|OCVV012889144.1|             KVKEYK------------------------KEVPRIERILPQGKDISAFSKLMYMLSMF  468
Ga0187910_10006931               MTRDYK-----------------------KSDTPTLNRILPESEDVSTFSKLMYALTMF  480
R._flav_XPD_                     KTREYK-----------------------KKDAPRLPRILPAGRDVSAFSKLMYALTMF  475
Ga0224415_10007274               DTRKYK-----------------------NMGTPRIRRLIPEGRDISTFSKLMYALTMF  458
Ga0187911_10069260               SAIKNIKAFEL------------------KPDQKELNGIMNSNALDVPHFCKLVYFLTRF  464
Ga0187910_10015336               IENTSYEITDK-----------------KQKEYYKMQINSLNSADKVSDFSKVIYLVTLF  534
ODAI_chimera                     GEGIKHITVDR----------------FIAGAGAGKLIDKCSISTECDYFCKFIYLMTLF  535
31009_GL0034153                  RRIFSQDND-----------------ITNWNGVEFEKIGEQKLEDNADYFCKLMYFVTLF  462
DLF014_GL0011914_                NIKEFSDAKEDDV------------YYEEFKNFKFNEVGKEKLGENAEYFCKLMYLLTLF  532
CasR_R._albus                    GETIKDLQKRY----------------DDETANRIWDISQHSISGNVNCFCKLIYIMTLM  518
Ga0187910_10040531               SNITQMQKNEQQG--------------KT-KGMFAIRDEIRVSRKPVSYFSKVIYVMTLL  503
tpg|DJXD01000002.1|              KNIRDYQKKAN--------------------DYRISFEDIKIGNTGIGYFSELIYMLTLL  443
CasR_E._siraeum                  DNIKKLSKS------------------NIEIQEDKLRKCFISYADSVSEFTKLIYLLTRF  513
PIG-018_GL0023397                VNIKKLQQN------------------TLEIPDSKLENCFIKSAEKVSAFTKLNYLLTRF  501
PIG-025_GL0099734                EGVISGLKK------------------NKSYDNMSIEGIVAEETANVSYFSKIIYLLAQF  515
Ga0129317_1008067_chimera        SNIKKLEN-------------------VKEKFISEDEFDDIKLDIDISYFSKLMYVMCYF  447
EYZ-362B_GL0088915               RYSSGSNSTENAE--------------KKEDFPVFPHCHDKQKVSNVSLFAKLMYAFCLF  505
uncultured_Ru_sp                 DVIKELGK-------------------ADMDFDEKILDSEKKNASDLLYFSKMIYMLTYF  508
```

FIG. 18FF

```
R._flav._FD-1                          WVTGEHEK----------------------DITSVIDKDAYRNRSNVSYFSKLMYAMCFF    509
tpg|DBYI01000091.1|                    WVKNDYSD----------------------EVAKCIDLNEFRVNSNVSYFSKLLYAMSFF    509
                                                                                .         :


037_-_emb|OIZA01000315.1|              LDGKEINDLLTNIISKLQVIEDHNNVIKFIFH-------------------------NKD    548
emb|OCTW011587266.1|                   LDGKEINILLNSLIKKFQDIHSFNTTVKKLSE-------------------------NNL    499
k87_11092736                           LEGKEINELVTALIKKFEGIQALIDLLRNLEGV-------------------------     508
BMZ-11B_GL0037915                      LDGKEINELLSTFVNRLENVQALIDLAKNLGET-------------------------     495
BMZ-11B_GL0037771                      LEGKEINELLTAYIHKFECIQDFLNVISSLGEN-------------------------     497
BMZ-11B_GL0069617                      WDGKEVNEVLSAYIHKFESIQAFIDVLESKELN-----------------------EKV    474
Ga0099364_10024192                     WEGKEINELLSAYIHKFENIQEFINLLEKLEGK-------------------------     541
530373_GL0023589                       LDSKEINMFFDALINELENIASFNDVLEQLDM---------------------------     545
emb|ODAI011611274.1|                  LDGKEINELLTGLINKLENIQSFLDVMQQEHA---------------------------     566
EMG_10003641                           LSGKEINDLLTQLINKFENIAAFIRMYQNDQG---------------------------     567
CasR_P1E0_metageno                     MDGKEINTLLTTLISKLENIDSLRSVLESMGL---------------------------     515
emb|OIZX01000427.1|                    MDGKEINTLLTTLISKFENIAAFLQTMEQLNI---------------------------     549
160582958_gene49834                    LDGKEINLLCTSLIEKFENIASFNEVLKSPQIG---------------------------     595
gi|1198542314|gb|NFIR01000008.1|       LDGKEINILLTQLIHCFDNISSFMDTMEEENL---------------------------     575
emb|OGNF01009141.1                     MDGKEINDLVTTLINKFDNIASFIEVLKSQSV---------------------------     521
Ga0129306_1000735                      LDGKEINDLLTTLIHAFENIDSFLSVLGSERL---------------------------     548
MH0288_GL0082219                       LDGKEINDLITTLVNSFDEIDSFCCVLQKAKL---------------------------     608
PIG-022_GL0026351                      IDGKEINDLLTTLIHQFENIASFIDVMKKENI---------------------------     506
PIG-028_GL0185479                      IDGKEINDLLTTLIHQFENIASFIDVMKKENI---------------------------     508
250twins_35838_GL0110300               LDGKEINELLCALINKFDNIADLLDISKQIGT---------------------------     557
```

FIG. 18GG

```
PIG-014_GL0226364                LDGKEINELCTKLLNKFKNINELVETAKRCGE---------------------------  503
250twins_36050_GL0158985         LDGKEINELLCALINKFDNIADILDAAAQCGS---------------------------  558
emb|OJMM01002900.1|              LDGKEINELCCAMINKFDNINDLIKTAAQCGE---------------------------  530
emb|OGPN01002610.1|              LDGKEINELCCAMMNKFDGIGDLIDTAKQCGE---------------------------  579
emb|OHCP01000044.1|              LDGKEINELCCAMMNKLDGISDLIETAKQCGE---------------------------  567
PIG-046_GL0077813                LDGKEINEFCCAMMNKLDGINDLMETAEQCGA---------------------------  572
pig_chimera                      LDGKEINELCCAMMNKLDGINDLIETAEQCGA---------------------------  574
emb|OIEN01002196.1|              LDGKEINIFLTSIHNIFQNIDSFLKVMKEKGM---------------------------  515
O2.UC29-0_GL0096317              LDGKEINVFLTTLHNLFENIASFLDVMHKQNI---------------------------  502
CasR_Anaerobic_dig               LDGKEINEMVSNLITKFENIDSLLHVDREIYKSDEKDLDLEIEKLALFFKGVVRPNAKTD  554
emb|OGDF01008514.1|              QDGKEINELVSSLVNKFANIQSFVDVMRSQGI---------------------------  532
emb|OGZC01000639.1               LDGKEINDLLTTLINKFDNISSLLKTMEQLEL---------------------------  558
Ga0224415_10048792_chimera       LDGKEINDLLTTLINKFENIQGFLDIMPEINV---------------------------  499
emb|OCVV012889144.1|             LDGKEINDLFTTLINKFDNIQCFLKIMPLIGV---------------------------  500
Ga0187910_10006931               LDGKEINELLTTLINKFDNIQSMLKIMPLIGV---------------------------  512
R._flav_XPD_                     LDGKEINDLLTTLINKFDNIQSFLKVMPLIGV---------------------------  507
Ga0224415_10007274               LDGKEINDLLTTLINKFDNIQSFLKVMPLIGV---------------------------  490
Ga0187911_10069260               LDGKEINDLLTTLVNKFDNIHSFNQVLTALGL---------------------------  496
Ga0187910_10015336               LDGKEINDLLTTLINKFDNIASLLSVLEKQSG---------------------------  566
ODAI_chimera                     IDGKEINDLLTTLINKFDNIDSPSVDKAVNNQ---------------------------  567
31009_GL0034153                  LDGKEINDLLTTLINKFDNIRSFIEVMEEMSL---------------------------  494
DLF014_GL0011914_                LDGKEINDLLTTLINKFDNIRSFVELMKEENI---------------------------  564
CasR_R._albus                    LDGKEINDLLTTLVNKFDNIASFIDVMDELGL---------------------------  550
Ga0187910_10040531               LDGKEINDLLTTLINKFENIVSFEDVLRQLNV---------------------------  535
```

FIG. 18HH

```
tpg|DJXD01000002.1|                LDGKEINDLLTTLINKFDNIISFIDILKKLNL---------------------------  475
CasR_E._siraeum                    LSGKEINDLVTTLINKFDNIRSFLEIMDELGL---------------------------  545
PIG-018_GL0023397                  INGKEINDLLTTLINKFDNIRSFLETMDELGL---------------------------  533
PIG-025_GL0099734                  IDGKEVNDLTTTLINKFDNIRSFIDTANQIGL---------------------------  547
Ga0129317_1008067_chimera          LDGKEINDLLTTLVSKFDNIGSIIEAATQIGI---------------------------  479
EYZ-362B_GL0088915                 LDGKEINELLSALLNKFDTIACMTKCTKDLGI---------------------------  537
uncultured_Ru_sp                   LDGKEINDLLTTLISKFDNIKEFLKIMKSSAVD---------------------------  541
R._flav._FD-1                      LDGKEINDLLTTLINKFDNIANQIKTAKELGI---------------------------  541
tpg|DBYI01000091.1|                LDGKEINDLLTTLINKFDNIRSFIDTANFLNI---------------------------  541
                                    ..**:* .       :  :

037_-_emb|OIZA01000315.1|          AVYKDYSDKYAIFR----------------NA---------------GKIATELEAIKSI  577
emb|OCTW011587266.1|               IINADYVDDYSLFE----------------QS---------------GTVARELMLIKSI  528
k87_11092736                       -DSIRFENEFALFN----------------DD----------KGNMAGRIARQLRLLASV  541
BMZ-11B_GL0037915                  ---VNFSKNYAVFN----------------DC----------GSGAAGRIAERLRLLASI  526
BMZ-11B_GL0037771                  ---VQFQPRFALFN----------------NA------------SFAQNVAVQLRILASI  526
BMZ-11B_GL0069617                  VFSAKKGRNFSAFN----------------EH------------GFAGRVADQLRVLASI  506
Ga0099364_10024192                 --KPQFTENYALFN----------------EA----------AGQRAGEIAQNLRILASI  573
530373_GL0023589                   --PLDYQEKYRFFE----------------RS---------------REEAAGLRFIKSI  572
emb|ODAI011611274.1|               --ETGLSDAFSFFE----------------YS---------------GEIAAELRMTRSF  593
EMG_10003641                       --KLEFTANYKMFEIDPQKDIPKDGKRVLSGS---------------AKIAYYLRTINYI  610
CasR_P1E0_metageno                 --AYSFVPAYAMFD----------------HS---------------RYIAGQLRVVNNI  542
emb|OIZX01000427.1|                --EAELGPEYAMFT----------------RS---------------RAVAEQLRVINSF  576
160582958_gene49834                -YETKYTEGYAFFK----------------NA---------------DKTAKELRQVNNM  623
```

FIG. 18II

```
gi|1198542314|gb|NFIR01000008.1|    --LTKLKEDYEIFE----------ES----------KEISRELRIINSF   602
emb|OGNF01009141.1                  --CTHFSEERKMFI----------DS----------AEICSELSAMNSF   548
Ga0129306_1000735                   --ERTFDANYRIFA----------DS----------GVIAQELRAVNSF   575
MH0288_GL0082219                    --DRGFTEGYRFFA----------HS----------AEISQELRAINSF   635
PIG-022_GL0026351                   --NYEFAEKYKFFN----------TS----------KNIACELRTVNSF   533
PIG-028_GL0185479                   --ECEFTEKFAFFK----------NS----------EIIAHELRTVNSF   535
250twins_35838_GL0110300            --PVVFCADYESLN----------DA----------AKIAENIRLIKNI   584
PIG-014_GL0226364                   --TIEFCPEYKMFK----------NS----------AKLAEEMRFVQSV   530
250twins_36050_GL0158985            --SVWFVDSYRFFE----------RS----------RRISAQIRIVKNI   585
emb|OJMM01002900.1|                 --DIEFVKEYKLFI----------NS----------NDLSDQIRIVKSI   557
emb|OGPN01002610.1|                 --EVKFIEEFACLS----------NC----------RKITNDIRVAKSI   606
emb|OHCP01000044.1|                 --DVEFVDQFKCLS----------KC----------ATMSNQIRIVKNI   594
PIG-046_GL0077813                   --KIEFVDNFNVLS----------KC----------GTISGQIRVVKSI   599
pig_chimera                         --KVEFVDKFSVLS----------NC----------ETISDQIRIVKSI   601
emb|OIEN01002196.1|                 --ECKFQKDFKMFS----------HA----------GHVAKKIEIVISL   542
O2.UC29-0_GL0096317                 --SYDFTEQYVMFY----------DS----------AYIVKKLETVISV   529
CasR_Anaerobic_dig                  TGAGEISKSFSIFQ----------SA----------ERIIEELKFIKNV   583
emb|OGDF01008514.1|                 --DSGFTADYAMFA----------ES----------GRISRELHILKGI   559
emb|OGZC01000639.1                  --QTTFKEDYTFFQ----------QS----------SRLCKEITQLKSF   585
Ga0224415_10048792_chimera          --NAKFEPEYVFFN----------KS----------HEIAGELKLIKGF   526
emb|OCVV012889144.1|                --NVKFTEDYAFFN----------KS----------EKIADELRLIKSF   527
Ga0187910_10006931                  --NAKFSSDYAFFN----------NS----------EKIADELKLIKSF   539
R._flav_XPD_                        --NAKFVEEYAFFK----------DS----------AKIADELRLIKSF   534
Ga0224415_10007274                  --NAKFAEEYSFFN----------NS----------EKIADELRLIKSF   517
```

FIG. 18JJ

```
Ga0187911_10069260                --SASYEADYKIFE----------------DS--------------GRVVEYLREINSF    523
Ga0187910_10015336                -KKVEFVENYSFFN----------------SSNLLKEKTLNKSENYTCKIVEELREINSF    609
ODAI_chimera                      -NRVEFGDKYDMFN----------------KA--------------GEICESLRSINSF    595
31009_GL0034153                   --KCEFMEEYKFFN----------------ES--------------KDICLRLKEIHSF    521
DLF014_GL0011914_                 --ECIFDKKFAFFD----------------ES--------------KNVCSTLREINSF    591
CasR_R._albus                     --EHSFTDNYKMFA----------------DS--------------KAICLDLQFINSF    577
Ga0187910_10040531                --DCTFKPEFAFFG----------------Y-D-----------RCRNISGELRLINSF    564
tpg|DJXD01000002.1|               --EFKFKPEYADFF----------------NMT-----------NCRYTLEELRVINSI    505
CasR_E._siraeum                   --DRTFTAEYSFFE----------------GS--------------TKYLAELVELNSF    572
PIG-018_GL0023397                 --ERTFTENYRFFE----------------DS--------------QKYRDELIELNSF    560
PIG-025_GL0099734                 --ECNLLDEYQFFN----------------TS--------------DTVKKELHVIKNL    574
Ga0129317_1008067_chimera         --NIEFIDDFKFFD----------------RS--------------KDISVELNIIRNF    506
EYZ-362B_GL0088915                --YAPFSKDFNFFE----------------DS--------------EQRAEEMNLVKSI    564
uncultured_Ru_sp                  -VECELTAGYKLFN----------------DS--------------QRITNELFIVKNI    569
R._flav._FD-1                     --NTAFVKNYDFFN----------------HS--------------EKYVDELNIVKNI    568
tpg|DBYI01000091.1|               --DVKFTKDYDFFN----------------II--------------CDYAGELNIIKNI    568
                                              :                                 :     .

037_-_emb|OIZA01000315.1|         ARMENKIE--------------NAPKEPLLKDALLALGVSSNDFDEKYEKYFKTDVD---    620
emb|OCTW011587266.1|              SKMDFGLD--------------NINLSFMYDDALRTLGVSDENLPEVKREYF--------    566
k87_11092736                      GKMKPDMT---------------DAKRVLYKSALEILGAPPD--------EVS-------    571
BMZ-11B_GL0037915                 GKMDAGLD---------------SLKPLFIRFAIQILGGARIS------DEDV-------    558
BMZ-11B_GL0037771                 GKMKPDLT---------------EAKRPLYKAAIRMLCPPEK-----WEKYTS-------    559
BMZ-11B_GL0069617                 GKMKPDLE---------------GAKRALFRDAVETLGIVENSNWLNEDGTIS-------    544
```

FIG. 18KK

```
Ga0099364_10024192                  GKMKPDLG---------------DAKRQLYKAAIEMLGIDTE-------EYIS-------   604
530373_GL0023589                    AKMNKTRVA--------VKNDTKTVKLKQYCDAGAVLGEINAEKIKAAFGLGD-------   617
emb|ODAI011611274.1|                ARMAAADP---------------EAKRFMVVDGAKLLGFNPKDTESEDEGIIRAIYGDAC   638
EMG_10003641                        ARMESFEI---------------KSDKTAINDAISLLGYNSNEHRDEFITYTMAKHVVDK   655
CasR_P1E0_metageno                  ARMRKPAI---------------GAKREMYRAAVVLLGVDSPEAAAAITDDLLQI-----   582
emb|OIZX01000427.1|                 ALMKKPQV---------------NAKQQLYRAAVTLLGTEDPDGVTDEMLCIDP------   615
160582958_gene49834                 ARMTKPLG-------------GVNTKCVMYNEAAKILGAKPMSKAELESVFNLDNHD---   667
gi|1198542314|gb|NFIR01000008.1|    ARMTEPVP---------------KTERIMFIEAAQILGYSNGEKELEGYVDALLD-----   642
emb|OGNF01009141.1                  ARMEAPGA---------------SSKRAMFVEAARILGDNRSKEELEEYFDTL-------   586
Ga0129306_1000735                   ARMTTEPF---------------NSKLVMFEDAAQLFGMSGGLVEHAEELREYLDNK---   617
MH0288_GL0082219                    ARMTTETV---------------QGKPCMFEDAADLLGMDTDLEEVREALEAFYNKE---   677
PIG-022_GL0026351                   ARMTGFMEN------------DESTKFIMFREAADILGTSYTDDELTDHLNL--------   573
PIG-028_GL0185479                   ARMTGFMEN------------DESTKFIMFREAADILGTSYTDDELTDHLNL--------   575
250twins_35838_GL0110300            AHLRPAIQ---EAQSSKDNADAAGTPATLLIDAYNMLNTDIQL-----------------   624
PIG-014_GL0226364                   SKMKKLTK---KE--------DKSIKPQVIDDALELLGLKIEKYKKDDQGNFVLDDQ---   576
250twins_36050_GL0158985            ASKDFKKS---KKD------SDESYPEQLYLDALALLGDVISKYKQNRDGSVVIDDQ---   633
emb|OJMM01002900.1|                 SKMKPELS---KI------------GEALILDAIDILGYKINKYKYDAAGNRLVDSN---   599
emb|OGPN01002610.1|                 SKMKNKVN---ID------------NDIIYLDAIELLGRKIEKYQKDENGKILLGTD---   648
emb|OHCP01000044.1|                 SRMKKEMT---ID------------NDTIFLDALELLGRKIEKYQKDKNGDYVKDEK---   636
PIG-046_GL0077813                   SKMKKEID---VD------------NETIFLDAFELLGKRIEKYKKDDSGNVLRNDD---   641
pig_chimera                         SKMKKEIA---ID------------NDTIFLDALELLGRKIDKYKKDATGKYLKDEN---   643
emb|OIEN01002196.1|                 AKMKKTL---------------DFYNAQALKDAVTILGVSKKHQYLDMNSYLDFYMF---   584
O2.UC29-0_GL0096317                 SRMKKAL---------------DFYNGQALRDAIRVLGISERHVDMDMKEYVDKYLF---   571
CasR_Anaerobic_dig                  TRMDNEI---------------FPSEGVFLDAANVLGVRGDDFDFSNEFVGDDLHSD--   625
```

FIG. 18LL

```
emb|OGDF01008514.1|              ARMQHSIAGLGDVKIYGSDDKFHGVSRRVYTDAAYILGFGERSEDNDGY--VDD------   611
emb|OGZC01000639.1               ARMGNPI---------------SNLKEVMMVDAIQILGTEKSEQELQSM--ACF------   622
Ga0224415_10048792_chimera       AQMGEPA---------------ATLKLEMTADAIKILGTEKEDAELIKL--AES------   563
emb|OCVV012889144.1|             ARMGEPV---------------ANTKRVMMIDAVKILGTDLSDDELVNM--ADS------   564
Ga0187910_10006931               ARMGEPV---------------ANAKRDMMIDAIKILGTDLDDNEIKKL--ADS------   576
R._flav_XPD_                     ARMGEPI---------------ADARRAMYIDAIRILGTNLSYDELKAL--ADT------   571
Ga0224415_10007274               ARMGEPV---------------ADARRAMYIDAIRILGTDLSDDELKAL--ADS------   554
Ga0187911_10069260               ARMTVDM---------------EKIKRSAYKKALLILGSSKYSDEDLDARVDEMLGV---   565
Ga0187910_10015336               ARMTGDC----------------KIRKSAFEDASQLLGYHDKTVNNLFEVLRLKELESK-   652
ODAI_chimera                     ARMESPA---------------PAAKRTMFVDAIRMFGFADVSDNDAEKFYKENIEP-F-   638
31009_GL0034153                  ARMKRPLD-------------NKNVEKEMYKDAISILMKDSSIENSNIDVIFNTYIY---   565
DLF014_GL0011914_                ARMQKPLN-------------NVSVKREMYKDAIKILLKDTRAEEQNIDRILNEYIP---   635
CasR_R._albus                    ARMSKID--------------DEKSKRQLFRDALVILDIGNKDETWINNYLDSDIFK---   620
Ga0187910_10040531               ARMQKPS---------------AKAKHVMYRDALRILGLDNGMSEEALDQEVRRILQ---   606
tpg|DJXD01000002.1|              ARMQKPS---------------ADARKIMYRDALRILGMDNRPDEEIDRELERTM-----   545
CasR_E._siraeum                  VKSCSFD---------------INAKRTMYRDALDILGIESDKTEEDIEKMIDNILQ---   614
PIG-018_GL0023397                AKSCSFD---------------ISAKRIMYRDALDILGIQAELSEDEIGRMIDNILQ---   602
PIG-025_GL0099734                TGMGYYD---------------VSIKHQMYKDAIDILGLHDNLSDEEINQLIDTNIL-F-   617
Ga0129317_1008067_chimera        ARMQAPV---------------PNAKRAMQEDAIRILGGSEEDIFSILDDMTG-------   544
EYZ-362B_GL0088915               ARMTKPD---------------IQAKEQLFRDALDILGVEVGVSEEALQKEIH-------   602
uncultured_Ru_sp                 ASMRKPA---------------ASAKLTMFRDALTILGIDDKITDDRISGILK-------   607
R._flav._FD-1                    ARMKKPS---------------SNAKKAMYHDALTILGIPEDMDEKALDEELDLILE---   610
tpg|DBYI01000091.1|              ARMKKPS---------------PSAKKNMYRDALTILGIPTEMSDEQLDAEIDKILE---   610
```

FIG. 18MM

```
037_-_emb|OIZA01000315.1|          ------------------------------------------------------------  620
emb|OCTW011587266.1|               ------------------------------------------------------------  566
k87_11092736                       ----------DEWLAENILLDKSNNDYQKAK----------------------------  592
BMZ-11B_GL0037915                  ----------DAYIKENILVRKDDPEKK------------------------------  576
BMZ-11B_GL0037771                  ----------DEWLEKNMLLNSEDRKNDKKK----------------------------  580
BMZ-11B_GL0069617                  ----------NDWLETYVLPGDKVTEEEK------------------------------  563
Ga0099364_10024192                 ----------DEWLEPNMLLAQPPKEPKKDNEKYRKEPHKYSYEKDMETYRKKLREYEET  654
530373_GL0023589                   ------------------------------------------------------------  617
emb|ODAI011611274.1|               AEYLQFSEEEK-------------------------------------------------  649
EMG_10003641                       YQNTDYQKIVKDFLSANKTLDCKS----KNMQAFVSELKNAHLSENYEQREKEIYELADT  711
CasR_P1E0_metageno                 ------------------------------------------------------------  582
emb|OIZX01000427.1|                ------------------------------------------------------------  615
160582958_gene49834                ------------------------------------------------------------  667
gi|1198542314|gb|NFIR01000008.1|   ------------------------------------------------------------  642
emb|OGNF01009141.1                 ------------------------------------------------------------  586
Ga0129306_1000735                  ------------------------------------------------------------  617
MH0288_GL0082219                   ------------------------------------------------------------  677
PIG-022_GL0026351                  ------------------------------------------------------------  573
PIG-028_GL0185479                  ------------------------------------------------------------  575
250twins_35838_GL0110300           ----------------------------------------VYGEAAYEELRKDLFERKNG  644
PIG-014_GL0226364                  ---------------------------------------NHNVWTDEYKSYRKECFTPKNS  598
250twins_36050_GL0158985           ---------------------------------------GNAVLTEQYKRFRYEFFEEIKR  655
emb|OJMM01002900.1|                ---------------------------------------NKPVYSEEYCAFKKDFFETCEL  621
```

FIG. 18NN

```
emb|OGPN01002610.1|                 ------------------------------------GKRLYTQEYKYFNDMFFNA---  667
emb|OHCP01000044.1|                 ------------------------------------GKKVYTKDYNNFQDMFFEG---  655
PIG-046_GL0077813                   ------------------------------------GKVLRTKEYNNIRKMLCKK---  660
pig_chimera                         ------------------------------------GKYLYSKEYDDFQYMFFKD---  662
emb|OIEN01002196.1|                 ----------------------------------------------------D-----  585
O2.UC29-0_GL0096317                 ----------------------------------------------------D-----  572
CasR_Anaerobic_dig                  ---------------------------------------------------AN-----  627
emb|OGDF01008514.1|                 ----------------------------------------------------------  611
emb|OGZC01000639.1                  ----------------------------------------------------F-----  623
Ga0224415_10048792_chimera          ----------------------------------------------------L-----  564
emb|OCVV012889144.1|                ----------------------------------------------------F-----  565
Ga0187910_10006931                  ----------------------------------------------------F-----  577
R._flav_XPD_                        ----------------------------------------------------F-----  572
Ga0224415_10007274                  ----------------------------------------------------F-----  555
Ga0187911_10069260                  ----------------------------------------------------D-----  566
Ga0187910_10015336                  ------------------------------------DWKKRTDDEQQEYDRLLN-----  670
ODAI_chimera                        ------------------------------------IYDRNGDIED------EW-----  650
31009_GL0034153                     ----------------------------------------------------------  565
DLF014_GL0011914_                   ----------------------------------------------------------  635
CasR_R._albus                       ----------------------------------------------------------  620
Ga0187910_10040531                  ----------------------------------------------------------  606
tpg|DJXD01000002.1|                 ----------------------------------------------------------  545
CasR_E._siraeum                     ----------------------------------------------------I-----  615
PIG-018_GL0023397                   ----------------------------------------------------L-----  603
```

FIG. 18OO

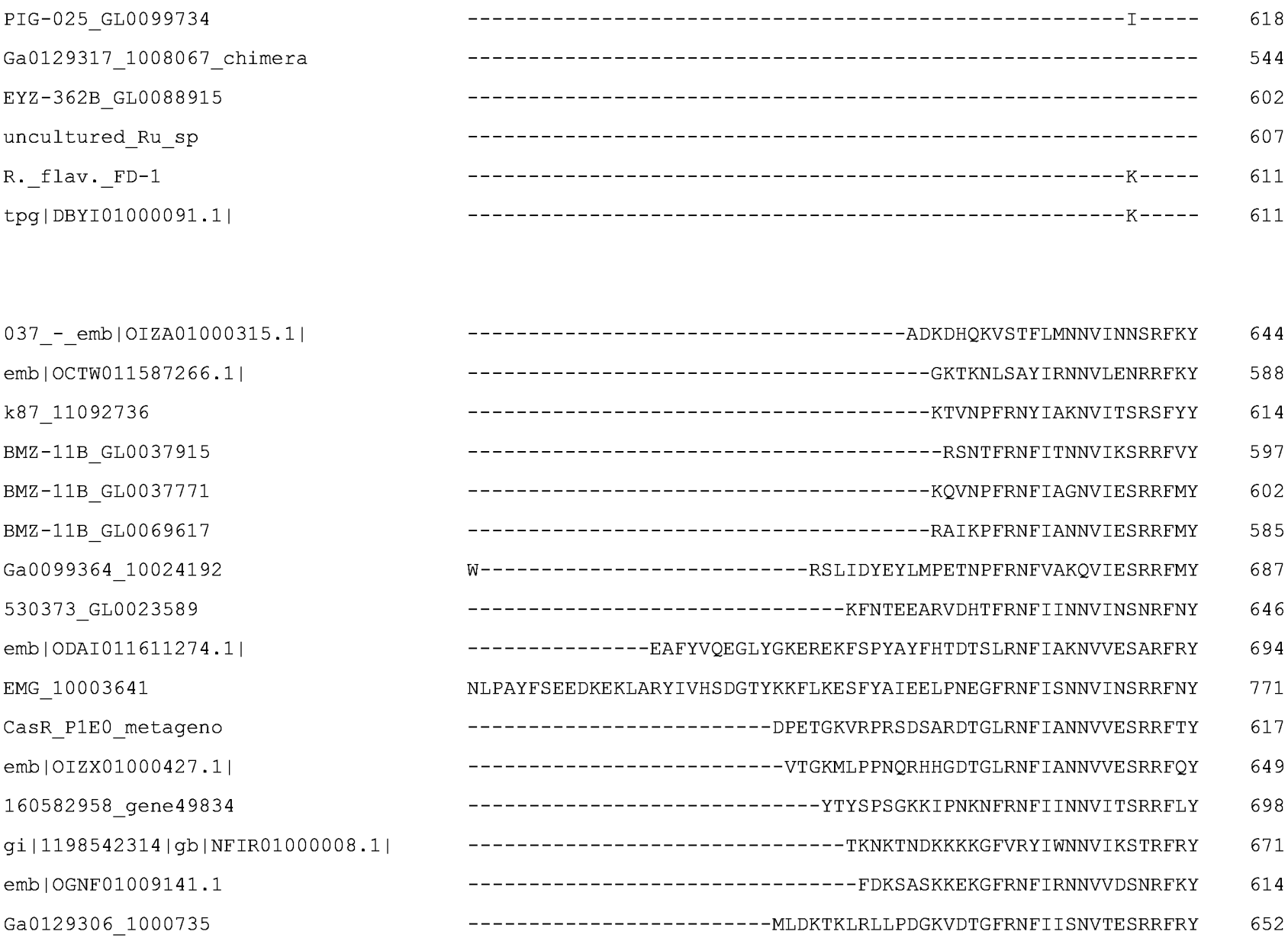

PIG-025_GL0099734 618
Ga0129317_1008067_chimera 544
EYZ-362B_GL0088915 602
uncultured_Ru_sp 607
R._flav._FD-1 611
tpg|DBYI01000091.1| 611
037_-_emb|OIZA01000315.1| 644
emb|OCTW011587266.1| 588
k87_11092736 614
BMZ-11B_GL0037915 597
BMZ-11B_GL0037771 602
BMZ-11B_GL0069617 585
Ga0099364_10024192 687
530373_GL0023589 646
emb|ODAI011611274.1| 694
EMG_10003641 771
CasR_P1E0_metageno 617
emb|OIZX01000427.1| 649
160582958_gene49834 698
gi|1198542314|gb|NFIR01000008.1| 671
emb|OGNF01009141.1 614
Ga0129306_1000735 652

FIG. 18PP

```
MH0288_GL0082219                          ------------------------K--MSLKWTKDGQEADTGFRNFIISNVLKSSRFRY    710
PIG-022_GL0026351                         ----------------------------KKTRLNARSKADTGLRNFIISNVIKSSRFKY    604
PIG-028_GL0185479                         ----------------------------KKTRLIARSKADMGLRNFIISNVIKSSRFKY    606
250twins_35838_GL0110300                  T-----------------------KY---NKKGKKVDVYDHKFRNFLINNVIKSKWFFY    677
PIG-014_GL0226364                         D-----------------------SM---NDAEKSFVAKGHQKRNFIISNVVLNKRFLY    631
250twins_36050_GL0158985                  D-----------------------ESGGIKYKKSGKPEYNHQRRNFILNNVLKSKWFFY    691
emb|OJMM01002900.1|                       D-----------------------EFGRVKYNKKGKPVINHRRRNFIINNVLSSKWFFY    657
emb|OGPN01002610.1|                       ---------------------------------------GNHKVRNFIANNVMQSKWFFY    688
emb|OHCP01000044.1|                       ----------------------------------------KNHRVRNFVSNNVIKSKWFSY    676
PIG-046_GL0077813                         -----------------------------------------DRRLNFIANSIIKNKWFSY    679
pig_chimera                               ----------------------------------------SHRVRNFISNSVIKSKWFSY    682
emb|OIEN01002196.1|                       ----------------------------NRSGATGKNAGKDHNLRNFLVSNVIRSRKFNY    617
O2.UC29-0_GL0096317                       ----------------------------DC----KDVKHADHNLRNFIVTNVIKSRKFNY    600
CasR_Anaerobic_dig                        ----------------------------KKIINKINGTKEDRNLRNFIINNVVKSRRFQY    659
emb|OGDF01008514.1|                       -----------------------------YVSSKLLGGADKNLRNFITNNVIKNRRFLY    641
emb|OGZC01000639.1                        ----------------------------FRDKNGKKLNTGEHGMRNFIGNNVISNTRFQY    655
Ga0224415_10048792_chimera                ----------------------------FKDENGKLLGNKQHGMRNFIGNNVIKSKRFHY    596
emb|OCVV012889144.1|                      ----------------------------FNDSEGNLLKKGKHGMRNFIINNVIENKRFHY    597
Ga0187910_10006931                        ----------------------------FKDSNGKLLHKGKHGMRNFIINNVVNNKRFHY    609
R._flav_XPD_                              ----------------------------SLDENGNKLKKGKHGMRNFIINNVISNKRFHY    604
Ga0224415_10007274                        ----------------------------SLDENGNKLGKGKHGMRNFIINNVITNKRFHY    587
Ga0187911_10069260                        ----------------------------YNQNGEKIKVRVDTGFRNFIANNVVESSRFHY    598
Ga0187910_10015336                        ----------------------------KHHYFKSGKKLPDTGLRNFIINNVIESRRFNY    702
ODAI_chimera                              ----------------------------KKGKLVHKKDEKKKGVYNFIATNVIESNRFQY    682
```

FIG. 18QQ

```
31009_GL0034153                 ---------------------------------QKDNKNAKKDFRNFIIKNVIKSNRFIY   592
DLF014_GL0011914_               ---------------------------------SKEDKGIKKDFRNFIIKNIIKSNRFMY   662
CasR_R._albus                   -------------------------------LDKEGNKLKGARHDFRNFIANNVIKSSRFKY   651
Ga0187910_10040531              -------------------------------IGADGKPIKNANKGFRNFIASNVIESSRFRY   637
tpg|DJXD01000002.1|             -------------------------------PVGADGKFIKGKQGFRNFIASNVIESSRFHY   576
CasR_E._siraeum                 -------------------------------DANGD--KKLKKNNGLRNFIASNVIDSNRFKY   645
PIG-018_GL0023397               -------------------------------DANGERISKNQRENGLRNFIASNVIDSNRFKY   635
PIG-025_GL0099734               -------------------------------GADGKPVKMAKGKRGFRNFIISNVLKSSRFRY   650
Ga0129317_1008067_chimera       -------------------------------YDKSGKKLAQSKKGFRNFIINNVVESSRFKY   575
EYZ-362B_GL0088915              ---------------------------------QYFQKGAKHDFRNFIINNVINSTRFIY   629
uncultured_Ru_sp                ---------------------------------LKEKGKGIHGLRNFITNNVIESSRFVY   634
R._flav._FD-1                   -------------------------------KTDPVTGKPLKGKNPLRNFIANNVIENSRFIY   643
tpg|DBYI01000091.1|             -------------------------------KINPVTGKTEKGKNPFRNFIANNVIENKRFIY   643
                                                                                  ::  .:  .  * *

037_-_emb|OIZA01000315.1|       VVKYINPADINGLAKNRYLVKFVLSKIPE--E----------QIDSYYKLFSNEEE----   688
emb|OCTW011587266.1|            VIKYIHPSDVQKIACNKAIAGFVLNRMPD--T----------QIKRYYDSLINKGA----   632
k87_11092736                    LVRYAKPTAVRKLMSNPKIVRYVLKRLPE-----------KQVASYYSAIWTQSES---   659
BMZ-11B_GL0037915               LVRYTKADAVRAVMKNGKIVRYVLARIPE-----------TQVDSYYKSVVDSSD----   641
BMZ-11B_GL0037771               LVRYSKPKAVRAIMQNRSIVNYVLHRLPS-----------EQIKTYSRVFPEDF-----   645
BMZ-11B_GL0069617               LARYAKPKSVRELMKHEEIIRYVLTRLPE-----------KQIDTYFSNISEGDE----   629
Ga0099364_10024192              LVRYTKPKTVRALMSNRAIVHYVLSRIAD--I--QDHHMTESQIDRYYQNLPQYNEQ---   740
530373_GL0023589                VMRFMNPKSARRIMQCESLVAFSLKDIPD--T----------QIERYCKSVNIPFDS---   691
emb|ODAI011611274.1|            VIRYVSPEIARKYARQEALVRFALHRVPL--L----------QLRRYYQSCCGPKKD---   739
```

FIG. 18RR

EMG_10003641	IMRFCNPEKIANIGKNKVLISFALSSLAEKTD----------MIAKYYRVFCDRID----	817
CasR_P1E0_metageno	LLRYMTPEQARVLAQNEKLIAFVLSTVPD--T----------QLERYCRTCGREDI----	661
emb|OIZX01000427.1|	LIRYSDPAQLHQLASNKKLVRFVLSSIPD--T----------QINRYYETCGQTRL----	693
160582958_gene49834	LIRYGNPEKIRKIAINPSIISFVLKQIPD--E----------QIKRYYPPCIGKRT----	742
gi|1198542314|gb|NFIR01000008.1|	LVRYADPKKVRAFAANKKVVAFVLKDIPD--D----------QIRAYYNSCFRQNSDSSS	719
emb|OGNF01009141.1	LTRYTDTSSVKAFSNNKALVKFAIKDIPQ--E----------QILRYYNSCFGASER---	659
Ga0129306_1000735	LVRYCEPRAVRDYMSCRPLIRLTLRDMPD--T----------ILRRYYEQSVGAAT----	696
MH0288_GL0082219	VVRYGNPKKLREVMQNEAVVKFVLNDLPE--E----------QINRYSENILGLDA----	754
PIG-022_GL0026351	LIRYSNPKKVKAIAENKKLVRFVLDKIPD--P----------QIIRYCNTCGF-KV----	647
PIG-028_GL0185479	LIRYSNPKKVKAIAENKKLVSFVLDKIPD--T----------QIIRYCNTCGF-KV----	649
250twins_35838_GL0110300	IAKYVKPADCAKMMSNKKMIEFALRDLPE--T----------QIKRYYYTITGNEA----	721
PIG-014_GL0226364	VAKYGSPTACNQMMKNQKIVRMALESISE--D----------QLRKYARRVYDNKEH---	676
250twins_36050_GL0158985	VVKYNRPSSCRELMKNKEILRFVLRDIPD--S----------QVRRYFKAVQGEEA----	735
emb|OJMM01002900.1|	VAKYNRPSECQKFMKSKKLIALVLKDVPE--T----------QIARYYQSVTGGRT----	701
emb|OGPN01002610.1|	VVRYNKPAECQIIMRNKTLVKFTLDDLPD--M----------QIQRYYSSVFGDNN----	732
emb|OHCP01000044.1|	VVRYNKPAECQALMRNSKLVKFALDELPD--S----------QIEKYYISVFGEKS----	720
PIG-046_GL0077813	IVRYNQPSECQAIMKNKTLVKFALDELPD--L----------QIQRYFVALYGDED----	723
pig_chimera	IVRYNQPSECRAIMKNKTLVKFALDELPD--L----------QIQRYFVALYGDED----	726
emb|OIEN01002196.1|	LSRYSNLAEVKKLAQNPSLVQFVLSRIEP--S----------LICRYYESSQGIS-----	660
O2.UC29-0_GL0096317	LARYSDMATVSKLSKNKEVVKFVLKRVDP--K----------QIERYYRSVCDVR-----	643
CasR_Anaerobic_dig	IARHMNTHYVKQLANNETLNRFVLNKMGD--A---------KIINRYYESISGNT-----	703
emb|OGDF01008514.1|	TVRYMNPKRAKKLVQNDALVVLALSGIPE--T----------QIDRYYKSCIEKR-----	684
emb|OGZC01000639.1	LIRYGNPQKLHTLSQNETVVRFVLSRIAK--NQRVQGMNGKNQIDRYYETCGGTN-----	708
Ga0224415_10048792_chimera	LIRYGDPAHLHKIATNKNVVRFVLGRIAD--MQKKQGQKGKNQIDRYYEVCVGNK-----	649

FIG. 18SS

```
emb|OCVV012889144.1|                LIRYGDPAHLHEIAKNEAVVRFVLGRIAD--IQKKQGKDGKNQIDRYYEICVGKD-----   650
Ga0187910_10006931                  IIRYGDPAHLHEIAKNEVVVRFVLGRIAD--IQKKQGKGGKNQIDRYYEICIGNG-----   662
R._flav_XPD_                        LIRYGDPAHLHEIAKNEAVVKFVLGRIAD--IQKKQGQNGKNQIDRYYETCIGKD-----   657
Ga0224415_10007274                  LIRYGNPVHLHEIAKNEAVVKFVLGRIAD--IQKKQGQNGKNQIDRYYETCIGKD-----   640
Ga0187911_10069260                  LIRYCHPRKIRNLAGNAALIEYQLRRLPE--L----------QILRYYEACTEPIK----   642
Ga0187910_10015336                  IVRYADPKKIRKCTENNELLKFAFKDVPD--S----------QVDRYYNICVTNK-----   745
ODAI_chimera                        LVRYANTEKIAALARNETIVRFVIKKLPA--A----------QIVRYYIACNRIREN---   727
31009_GL0034153                     LIKYSNPTNVRKFASNTNVIKFVLGTLPD--K----------QLERYYKSCGLSD-----   635
DLF014_GL0011914_                   LVKYSNPTDVRNLASNRNVVKFVLNAIPD--T----------QIDRYYKSCIQHY-----   705
CasR_R._albus                       LVKYSSADGMIKLKTNEKLIGFVLDKLPE--T----------QIDRYYESCGLD------   693
Ga0187910_10040531                  LVRYNNPHKTRMIAQNEAIVRFVLSEIPD--E----------QIRRYYDVCRDPK-----   680
tpg|DJXD01000002.1|                 LVRYNNPHKTRTLVKNPNVVKFVLEGIPE--T----------QIKRYFDVCKGQE-----   619
CasR_E._siraeum                     LVRYGNPKKIRETAKCKPAVRFVLNEIPD--A----------QIERYYEACCPKN-----   688
PIG-018_GL0023397                   LVRYGNPKKIRETAKCPPAVKFVLDGIPD--S----------QIERYYNSCFPED-----   678
PIG-025_GL0099734                   LVKYCNPKKIRKIANNEKIVKFVLGRITE--T----------QIERYYYSCNPELK----   694
Ga0129317_1008067_chimera           IVRYSNPQKIRKLANNSVVVGFVLGKLPD--A----------QIESYFNSCLPN------   617
EYZ-362B_GL0088915                  VIKFCNPKTAGQIIRNENILRFVLQRMPQ--E----------QIERYYHACTKN------   671
uncultured_Ru_sp                    LIKYANAQKIREVAKNEKVVMFVLGGIPD--T----------QIERYYKSCVEFP-----   677
R._flav._FD-1                       LIKFCNPENVRKIVNNTKVTEFVLKRIPD--A----------QIERYYKSCTDS------   685
tpg|DBYI01000091.1|                 VIKFCNPKNVRKLVNNTKVTEFVLKRMPE--T----------QIDRYFESCIEG------   685
                                      :.                  :  :               :  *

037_-_emb|OIZA01000315.1|           ---------------PSCEEKIKLLTKKISKLNFQTLFENNK-----------------   715
emb|OCTW011587266.1|                ---------------TDIQAQAKALLDCITGISFDAIKDDKHLHKS-----------KE   665
```

FIG. 18TT

```
k87_11092736                          ----------NSNEMVKLIEMIDRLTTEIAGFSFAVLKDKKDSIVSA----------SR    698
BMZ-11B_GL0037915                     ----------AGNGKPIKEKIKALAEKLGEFSFDLIAENRDGIIDNA----------KL    680
BMZ-11B_GL0037771                     ---------------SDTEAEIDFLVNKLSEFSFETFISNQQTILNNS------KRGFSP    684
BMZ-11B_GL0069617                     ---------------SEISRKIDTLTKALTGFSFNSLFINRDKIVENT----------KL    664
Ga0099364_10024192                    ----------QHKNVSLETKIDALADYLCKYTFEKNVLKQKNGIV-L----------NT    778
530373_GL0023589                      -------------ANPNYAHMKEQLTYKLLQVQFCNFSKVSNSSHQ--------------    724
emb|ODAI011611274.1|                  ---------------PDAAECVDFLAGVVNRVDFANFTDVRTGDSSK-------------    771
EMG_10003641                          ----------------DQKTMEDYLVNKLTKISYTEFLNVNQKANAE-------------    848
CasR_P1E0_metageno                    ---------------TGRPAQIRYLTAQIMGVRYESFTDVEQRGRG--------------    692
emb|OIZX01000427.1|                   ---------------AGRAAKVEFLTDMIAAIRFDQFRDVNQKERG--------------    724
160582958_gene49834                   ---------------DDVTLMRDELGKMLQSVNFEQFSRVNNKQNAK-------------    774
gi|1198542314|gb|NFIR01000008.1|      NNSNASWDADSNKRDISVSDMRKALTEKITGLNFGDFEEESKKGIR--------------    765
emb|OGNF01009141.1                    ---------------YYNDGMSDKLVEAIGKINLMQFNGVIQQADRN---------MLPE    695
Ga0129306_1000735                     ---------------VDRERILDTLADKLLSLRFTDFENVNQRANAE-------------    728
MH0288_GL0082219                      ---------------AAKEKKVSELARFLRALSLEDFRDVEQSANAA-------------    786
PIG-022_GL0026351                     ---------------GERQENINTLAKAITNFSFSAFEKVKQSANVE---------N---    680
PIG-028_GL0185479                     ---------------GKRQENINTLAEAITYFSFSNFEKVKQSANVE---------N---    682
250twins_35838_GL0110300              --------------LGDAESLKGVIIEQLHAFSIKNTLLSIKNM--G---------EGEY    756
PIG-014_GL0226364                     ------------FQNQTADEIRENLAKKLKNFNIGAELKTVGNL------------SEKD    712
250twins_36050_GL0158985              --------------YASAEAMRTRLVDALSQFSVTACLDEVGGM--T---------DKEF    770
emb|OJMM01002900.1|                   --------------QANSEAMRMTLIKLLHEFSIKNVLSDVGTM--T---------ASEN    736
emb|OGPN01002610.1|                   --------------MPAVDEMRKRLLDKINQFSVRGFLDELDEIVLM---------SDEE    769
emb|OHCP01000044.1|                   --------------SSSNEEMRRELLKKLCDFSVRGFLDEIVLL--S---------EDEM    755
PIG-046_GL0077813                     --------------LPSYGEMRKILLKKLHDFSIRDFDEEIGSL--F---------D-EM    757
```

FIG. 18UU

```
pig_chimera                          --------------LPSYGEMRKILLKKLHDFSIKGFLDEIVLL--S---------DLDM    761
emb|OIEN01002196.1|                  ------------SEGITIDEQIKKLTGIIVDMNIDSFENINNGEIGMR------Y---SK    699
O2.UC29-0_GL0096317                  --------------NISAEDEIEKLVDIIYNFNIDELEQVNNAEIGTG------HSGRKV    683
CasR_Anaerobic_dig                   -------------PNIEVRSQIDYLVKRLRSFSFEDLNDVKQKVRPG-------------    737
emb|OGDF01008514.1|                  ------------SFNPDLNEKIAALSEMITTLKIDDFEDVKQNPEKNA---------NYE    723
emb|OGZC01000639.1                   ------------SWSVSEEEKINFLCKILTNMSYDQFQDVKQSGAEI-------------    743
Ga0224415_10048792_chimera           ------------DIKKTIEEKIDALTDIIVNMNYDQFEKKKAVIENQ---NRGKTFEEKN    694
emb|OCVV012889144.1|                 -------------RGKTVSQKLDELTKIIVNMNYDQFEQKKEIIENTRTKTKYMRYDEEQ    697
Ga0187910_10006931                   -------------YGKSVSEKIDALTKVIINMNYDQFEAKRKVIENT-------------    696
R._flav_XPD_                         -------------KGKSVSEKVDALTKIITGMNYDQFDKKRSVIEDT-------------    691
Ga0224415_10007274                   -------------SSKSVAEKVNALTKIITGMNYDQFDSRRNVIENT-------------    674
Ga0187911_10069260                   ------------RTARTMDEKIGTLIDLIVKMDFSQFEDVQQNDRVRV-----FSDAEKK    685
Ga0187910_10015336                   ------------ITNATREEKIERLVDIIKSMNLSKVATVKQRDKQ--------------    779
ODAI_chimera                         ----------INSNVLTSDDMAESLVKLVTGMDFGKIKGVRNDDR---------------    762
31009_GL0034153                      --------------GIGKTEWVRELSEIITGIDYSDFLNVKQRAKS--------------    667
DLF014_GL0011914_                    ------------SNEVDKETQIQELTDIITGIDYLDFLDVRQSYR---------------    738
CasR_R._albus                        ------------NAVVDKKVRIEKLSGLIRDMKFDDFSGVKTSNKA--------------    727
Ga0187910_10040531                   -----------LPRSSSREAQVDILTGIITDVNYRIFEDVPQSKKINK-----DRPD---    721
tpg|DJXD01000002.1|                  -----------IPPTSDKSAQIDVLARIISSVDYKIFEDVPQSAKINK-----DDPS---    660
CasR_E._siraeum                      ------------TALCSANKRREKLADMIAEIKFENFSDAGNYQKANV-----TS-----    726
PIG-018_GL0023397                    ------------TAICSADKKRRELAEMISNISFDWFKDAGNVQRVNA-----ED-----    716
PIG-025_GL0099734                    -----------RAAYPGREAAIDDLTKIIINMKFDDFKNVKQSANV-------ED-----    731
Ga0129317_1008067_chimera            ------------RVYSTPDKARESLRDMLHNISFNDFADVKQDDRR-------A------    652
EYZ-362B_GL0088915                   ------------PSLCTHDEKIQCLSNIIMEVRYKKFCNVAQKSRD-------T------    706
```

FIG. 18VV

```
uncultured_Ru_sp                    ----------DMNSSLGVKRSELARMIKNISFDDFKNVKQQAKG--------------   711
R._flav._FD-1                       ----------EMNPPTEKKITELAGKLKDMNFGNFRNVRQSAK---------------   718
tpg|DBYI01000091.1|                 ----------NLNPTTEKKIEKLAEMIKNIKFEEFRNVKQKVRD-------N------   720
                                                                 :   :

037_-_emb|OIZA01000315.1|           ---IPNVEKERKKAIITLYFTIVYILVKNLVNINGLYTLALYFVERDGFFYKQICEKKLI   772
emb|OCTW011587266.1|                KSPQRSADRERKKAMLTLYYTIVYIFVKQMLHINSLYTIGFFYLERDQRFIYSRAKKEN-   724
k87_11092736                        ESRAVNLEVERLKKLTTLYMSIAYIAVKSLVKVNARYFIAYSALERDLYFFNEKYGEEFR   758
BMZ-11B_GL0037915                   DEEKKNVEIERLKAVTGLYLTVAFIAIKNLVKANARYYIAYSAFERDSELLKEKLGEEEF   740
BMZ-11B_GL0037771                   NRRETAEEIERLKAITGLYLSVAYIAIKNIVKANARYYIAFAVFERDKELVKAKDAR---   741
BMZ-11B_GL0069617                   DTSNKNVKIERLKALTGLYLTVAFVAVKNLVKANARYFIAFSAFERDYGFMKSHDPD--V   722
Ga0099364_10024192                  KSATKNVEIEHLKALTGLYLTVAYIAVKNLVKANARYYIAFSIFERDYALFEKKLGKDTL   838
530373_GL0023589                    -----AVEKERLKALVGLYLSVLYRIVKSIVRINTSYTIAFAAYERDCVILNRKMGKNLY   779
emb|ODAI011611274.1|                ---SEQEKKQKYQAIVGLYLTVVYWIVKNLVNVNSRYVMAFHILERDTVLLEGKRLFVGG   828
EMG_10003641                        ----KNKEKDRSQKLIGLYITLLYEIVKNLVNINSRYNIAFQRCDNDSIMILQGQYDER-   903
CasR_P1E0_metageno                  ----DNPKKERYKALIGLYLTVLYLAVKNMVNCNARYVIAFYCRDRDTALYQKEVCWYDL   748
emb|OIZX01000427.1|                 ----ANTQKERYKAMLGLYQTVLYLAVKNLVNINARYVMAFHCVERDMFLYDGELTDPKG   780
160582958_gene49834                 --QNPNGEKARLQACVRLYLTVPYLFIKNMVNINARYVLAFHCLERDHALCFNSRKLN--   830
gi|1198542314|gb|NFIR01000008.1|    -------KEESDKNIIRLYLTVLYLVQKNLIYVNSRYFLAFHCAERDEMLYNGETIDNNK   818
emb|OGNF01009141.1                  EKKKANAQKEKYKSIIRLYLTVCYLFFKNLVYVNSRYYSAFYNLEKDRSLFEINGELKPT   755
Ga0129306_1000735                   ----RNREKQKMMGIISLYLNVAYQIVKNLVYVNARYTMAYHCAERDTELLLNAAGEGNL   784
MH0288_GL0082219                    ----RNARKARYQGAMSLYLNVIYQIIKNMVYINSRYTLAFAKYEHDASLLLDNWVYDPY   842
PIG-022_GL0026351                   -DTPDAIDKAQKQGLISLYLTVLYLFIKNMVYVNSRYVLAFHCLERDSQLLKITSGYN--   737
PIG-028_GL0185479                   -DTPDAIDKAQKQGLISLYLTVLYLFIKNMVYVNSRYVLAFHCLERDSQLLKITSGYN--   739
```

FIG. 18WW

250twins_35838_GL0110300	KIQQIGSSKEKLKAIVNLYLTVAYLLTKSLVKVNIRFSIAFGCLERDLVLQKKS------	810
PIG-014_GL0226364	FRSSKNALKDKSKSLIGLYLTVAYLITKNMIKINTRFSIAFGAYERDMEFLLQQ------	766
250twins_36050_GL0158985	ASQRAVDSKEKLRAIIRLYLTVAYLITKSMVKVNTRFSIAFSVLERDYYLLIDG------	824
emb|OJMM01002900.1|	KRQIENSRKERMKAIVKLYLTVVYLIAKSLVKVNTRFSIAFSAYERDVSLLADENELIA-	795
emb|OGPN01002610.1|	SKRNKSSEKEQKKSLIRLYLTIAYLITKSMVKINTRFSIACAMYERDYALLCQS------	823
emb|OHCP01000044.1|	KQKDKFSEKEKKKSLIRLYLTIVYLITKSMVKINTRFSIACATYERDYILLCQS------	809
PIG-046_GL0077813	KRKDKFCESDRKKCLFRLYLTIAYLITKSMVKINTRFSIACATYERDYALLCAS------	811
pig_chimera	ESQDKYCEKEQKKSLFRLYLTIAYLITKSMVKINTRFSIACATYERDYALLCAS------	815
emb|OIEN01002196.1|	ATPQSIERRNQMRVCVGLYLNVLYQIEKNLMNVNARYVLAFAFAERDALMLNFTLEECK-	758
O2.UC29-0_GL0096317	ASKRSVQERNQARACVSLYLNVLYQIVKNLMNVNSRYSIAFAMTERDMIVEGKGIEAIET	743
CasR_Anaerobic_dig	--TNESIEKEKKKALVGLCLTIQYLVYKNLVNINARYTTAFYCLERDSKLKGFGVDVWR-	794
emb|OGDF01008514.1|	AKKNQRISKERYKACIGLYLTVLYLICKNLVKINARYSIAIGCLERDTQLHGVDFKGAA-	782
emb|OGZC01000639.1	-TAEEKRKKERYKAIISLYLTVLYQLIKNLVNINARYIIAFHCLERDAILYSSKFNTSIN	802
Ga0224415_10048792_chimera	KYKRDNAEREKFKKIISLYLTVIYHILKNIVNVNSRYILGFHCLERDKQLYIEKYNKD--	752
emb|OCVV012889144.1|	KLRKDNAEREKFKKIISLYLTVLYYVVKNIVNINSRYVIGFHCLERDAQLYSEKYNKE--	755
Ga0187910_10006931	--GRDNAEREKYKKIISLYLTVIYQILKNLVNVNSRYVIGFHCVERDAQLYKEKGYDINT	754
R._flav_XPD_	--GRENAEREKFKKIISLYLTVIYHILKNIVNINARYVIGFHCVERDAQLYKEKGYDINL	749
Ga0224415_10007274	--GAGNAEREKYKKIISLYLTVIYHILKNIVNINSRYVIGFHCVERDAQLYKEKGYDINL	732
Ga0187911_10069260	EKIRKMREKQRYQSIISLYLTMLYLIVKNLVNINARYVMAFQAWERDNYLLLQLSGKE--	743
Ga0187910_10015336	----DNVEKQKQLAIMSLYLNILYQIAKNLVYVNSRYVMAFHSLERDSQMLFDAYYDV--	833
ODAI_chimera	-----SGDKEMKKAIVTLYLTILYHLTKNLVYVNSRYFMAFHCLERDMNLFGEGQKQALR	817
31009_GL0034153	--KEEKEDKAKKQAIVTLYLTILYILTKNLVNVNSRYVIAVHCIERDSELYKTELYEKLK	725
DLF014_GL0011914_	-----NEDKSKKQAVVTLYLTILYILTKNLVNVNSRYVIALHCLERDSKLYGYTKLSKKN	793
CasR_R._albus	-GDNDKQDKAKYQAIISLYLMVLYQIVKNMIYVNSRYVIAFHCLERDFGMYGKDFGKY--	784

FIG. 18XX

```
Ga0187910_10040531                ANDRMTLKKQRYQAIVSLYLTVMYLVTKNLVYVNSRYVMAFHALERDAYLYGITNI----   777
tpg|DJXD01000002.1|               RNFSDALKKQRYQAIVSLYLTVMYLITKNLVYVNSRYVIAFHCLERDAFLHGVTLPKM--   718
CasR_E._siraeum                   -RTSEAEIKRKNQAIIRLYLTVMYIMLKNLVNVNARYVIAFHCVERDTKLYAESGLEVG-   784
PIG-018_GL0023397                 -KSRNAEIKRKNQAIIRLYLTVMYLMLKNLVNVNSRYVIGFHCVERDAKLYKLSGLTIE-   774
PIG-025_GL0099734                 -SSKEAQEKMKYQTIISLYLTVCYHLVKNLVNINARYAMAFHSLERDARLYGIYGSEKDY   790
Ga0129317_1008067_chimera         -TPEEKVEKERYKAIIGLYLTVMYHLVKNLVYVNSRYVMAFHCLERDAMHYDVS------   705
EYZ-362B_GL0088915                ---TKQMEKERFKAIIRLYLTAIYLVVKNLVNVNARYVLAFHCLERDADLFGLQIKQTLP   763
uncultured_Ru_sp                  ---RENVAKERAKAVIGLYLTVMYLLVKNLVNVNARYVIAIHCLERDFGLYKEIIPEL-A   767
R._flav._FD-1                     ----ENMEKERFKAVIGLYLTVVYRVVKNLVDVNSRYIMAFHSLERDSQLYNVS------   768
tpg|DBYI01000091.1|               --SQEAVEKERFKAIIGLYLTVIYLLVKNLVNVNSRYVMAFHCLERDAKLYGVQ------   772
                                                *     :   *.::  *  :  .    :.*

037_-_emb|OIZA01000315.1|         ---------------ETLKKKDKKQLYNDDDYLLLPEIFSGSKYRE--------------   803
emb|OCTW011587266.1|              --------------------KNPSKNSYLNDFRSVTAYFIPSEIMKRIEKNE--------   756
k87_11092736                      ----------LHFIP-----YELNGKTCQFEYLAILKYYLARDEETLKRKCEICEEIKVG   803
BMZ-11B_GL0037915                 ----------YKLCIEFKN---KKGETGYNNAFALLKYYIDRDEDIWRKEEATHRGRDEN   787
BMZ-11B_GL0037771                 -------------------IQTTIPNTKFTNYFCLTQYYLDRDEEKKFPGDPRDKEAFFE   782
BMZ-11B_GL0069617                 ----------AQSCLAYEEIMDGKSKTKYNEIYALTEYYLKKEEENDYHPEPGQPFDKEA   772
Ga0099364_10024192                ----------EKYVKPFKYIDKGEEKEGKNNFFALTEYLLDKDNSLRYQWNNDLSDEENK   888
530373_GL0023589                  GL-------------------------KTDPMGITNDFMHMPTVAPSNPH---------   804
emb|ODAI011611274.1|              MK-------------------------AEDPFLLTDGYVSRQDAYVRK-----------   851
EMG_10003641                      ----------------------------AVQESKLTKKFISNQK---------------   919
CasR_P1E0_metageno                EE-----------------DKKSGKQRQVEDYTALTRYFVSQGY---------------   775
emb|OIZX01000427.1|               ES-------------VSAFLAVNGKKGVQPQYLLLTQLFIRRDY---------------   811
```

FIG. 18YY

```
160582958_gene49834                ----------------------------DDSYNEMANKFQMVRKAKKEQYEKEYKCK---   859
gi|1198542314|gb|NFIR01000008.1|   EK-----------------------GSEKDWRKFAKQFIMEHS----------------   838
emb|OGNF01009141.1                 GK-----------------------FDEGHYTGLVKLFIDNGW----------------   775
Ga0129306_1000735                  LRRDRSWPARLHLPRRALARRRDRVEVMERDVARGPEAYNRDEWLG-------------   830
MH0288_GL0082219                   ----------------------------KTVFSDLTEDSLQKGK----------------   858
PIG-022_GL0026351                  ----------------------------NSSYNNLTNYFIDNKY----------------   753
PIG-028_GL0185479                  ----------------------------NLSYNNLTNHFIEKEY----------------   755
250twins_35838_GL0110300           ----------------------------EKKFDAIINEILLEDDKIRKECDKERAQAKTL   842
PIG-014_GL0226364                  ----------------------KKEKGKRFDWLALTQKFLSEDKNYYFEHDLRRKEINKI   804
250twins_36050_GL0158985           --------------------KKKSSDYTGEDMLALTRKFVGEDAGLYREWKEKNAEAKDK   864
emb|OJMM01002900.1|                -----------------LANNEDDKWKKGNYVFALTKHFWDNDEPYFDKYNNALQQIRSI   838
emb|OGPN01002610.1|                -------------------EMKGGPWDGGAQALAVTRKFLNHDREVFDRYCAREAEIARL   864
emb|OHCP01000044.1|                -------------------EKAERAWEKGATAFALTRKFLNHDKPTFEQYYTREREISAM   850
PIG-046_GL0077813                  -------------------NKQERAWSSGATALALTRKFLNQDKLIFEKHYAREGEISKL   852
pig_chimera                        -------------------NKQERAWSSGATALALTRRFLNQDKLIFEKHYAREGEISKL   856
emb|OIEN01002196.1|                --------------------KNKKRSSGGFSFIEMTQFFIDKKLFKVAT-----------   787
O2.UC29-0_GL0096317                FKDDP-----------DAGKKSKKQIAYRTKYIALTEEYLDKGYIKCAT-----------   781
CasR_Anaerobic_dig                 ----------------------------DFESYTALTNHFIKEGY----------------   811
emb|OGDF01008514.1|                --------------------------------YMTRDVFIAKGWI---------------   795
emb|OGZC01000639.1                 ----------------------------LKKRYTALTEMILGYETDEKAR-----------   824
Ga0224415_10048792_chimera         ----------------------------KLDGFVALTKFCLGDEERFEDL--------KAK   777
emb|OCVV012889144.1|               ----------------------------IDKKISSLTKFCLGESDEKKLK---------AL   779
Ga0187910_10006931                 ------------------------NNLESKGFTSVTKLCVGIADDDPVKYKNVEIELKER   790
R._flav_XPD_                       ------------------------KKLEEKGFSSVTKLCAGIDETAPDKRKDVEKEMAER   785
```

FIG. 18ZZ


Ga0224415_10007274 -----------------------KKLKDKGFTSVTKLCAGIDEECK----DVEKEMTER 764
Ga0187911_10069260 --------------------------AEAEYLNLTRHFIEPLDGAKP------------ 764
Ga0187910_10015336 ----------------------------KRGYCDLSTVLLFGVDDLQNRNRGSYKYLRDN 865
ODAI_chimera NTGDY-----------PADN--ADLRRRDIKYCVLTKAHIKQEEVKSRRND--------- 855
31009_GL0034153 PEQPE-----------KLYKIKPDKPLQPAQYFLFTKYLLQNDFFGE------------- 761
DLF014_GL0011914_ --------------------NKKLEKIKPPKYHELTKYFIDNRYFDRK------------ 821
CasR_R._albus ----------------------------YQGCRKLTDHFIEEKYMKE------------- 803
Ga0187910_10040531 ----------------------------KGDYRKLTDNLLADENYKKF------------ 797
tpg|DJXD01000002.1| --------------------------NKKIVYSQLTTHLLTDKNYTTY------------ 740
CasR_E._siraeum -------------------------NIEKNKTNLTMAVMGVKLENG----IIKT----- 809
PIG-018_GL0023397 ------------------------GNTDKNRTLLTATILGITSADRAPSGEIVA----- 804
PIG-025_GL0099734 V----------------------EKLKKEQAVLVEKLLSDGPETAG------------ 814
Ga0129317_1008067_chimera ----------------------------LDNYRDLIRHLISEGDSSCN------------ 725
EYZ-362B_GL0088915 D----------------------RKKKSDYLKLAVDQCGDNPENKP------------ 787
uncultured_Ru_sp S----------------------KNLKNDYRILSQTLCELCDKSPN------------ 791
R._flav._FD-1 --------------------------VDNDYLALTDTLVKEGDNSRS------------ 789
tpg|DBYI01000091.1| -------------------------NIGGDYLALTAKLCAEGDDYGKKLSEAKQ----- 801
037_-_emb|OIZA01000315.1| ---------- 803
emb|OCTW011587266.1| ---------- 756
k87_11092736 CEKHKKNANPPYEYDQEWIDKKKALNSE--RKACERRLHF----------STH 844
BMZ-11B_GL0037915 GKLDYDPDP---------- 796
BMZ-11B_GL0037771 ---------HL---------- 784

FIG. 18AAA

```
BMZ-11B_GL0069617                CRKHLD----------------------------------------------------------  778
Ga0099364_10024192               QALRKHLDKKE------------------------------------------------------  899
530373_GL0023589                 ----------------------------------------------------------------  804
emb|ODAI011611274.1|             ----------------------------------------------------------------  851
EMG_10003641                     ----------------------------------------------------------------  919
CasR_P1E0_metageno               ----------------------------------------------------------------  775
emb|OIZX01000427.1|              ----------------------------------------------------------------  811
160582958_gene49834              ----------------------------------------------------------------  859
gi|1198542314|gb|NFIR01000008.1| ----------------------------------------------------------------  838
emb|OGNF01009141.1               ----------------------------------------------------------------  775
Ga0129306_1000735                ----------------------------------------------------------------  830
MH0288_GL0082219                 ----------------------------------------------------------------  858
PIG-022_GL0026351                ----------------------------------------------------------------  753
PIG-028_GL0185479                ----------------------------------------------------------------  755
250twins_35838_GL0110300         P---------------------------------------------------RELAQE------  849
PIG-014_GL0226364                SD--------------------------------------------------RDERKK------  812
250twins_36050_GL0158985         YF--------------------------------------------------DKAERK------  872
emb|OJMM01002900.1|              V---------------------------------------------------DPNERR------  845
emb|OGPN01002610.1|              P---------------------------------------------------S-EERK------  870
emb|OHCP01000044.1|              P---------------------------------------------------Q-EKRK------  856
PIG-046_GL0077813                P---------------------------------------------------K-EERK------  858
pig_chimera                      P---------------------------------------------------K-EERK------  862
emb|OIEN01002196.1|              ----------------------------------------------------------EA------  789
O2.UC29-0_GL0096317              ----------------------------------------------------------ER------  783
```

FIG. 18BBB

CasR_Anaerobic_dig	----------	811
emb|OGDF01008514.1|	----------	795
emb|OGZC01000639.1	----------RKD--------TR	829
Ga0224415_10048792_chimera	AQASI----------QALETANPKLYAKYMNYSDEEKKEEFK--------KQ	811
emb|OCVV012889144.1|	AKKSL----------EELKTTNSKLYENYIKYSDERKAEEAK--------RQ	813
Ga0187910_10006931	ALASF----------DALEKENPELYEKYNMYSEKQKEAELE--------KQ	824
R._flav_XPD_	AKESI----------DSLESANPKLYANYIKYSDEKKAEEFT--------RQ	819
Ga0224415_10007274	AKASF----------AALETANPKLYATYINYSDEEKNAELR--------KQ	798
Ga0187911_10069260	----------	764
Ga0187910_10015336	RRSNKDVIETFGDFKGKVSKVVEKKNQGLTNEIYDSLCNVAGTTKTEVQNEIKSILKSNG	925
ODAI_chimera	----------	855
31009_GL0034153	----------	761
DLF014_GL0011914_	----------	821
CasR_R._albus	----------	803
Ga0187910_10040531	----------	797
tpg|DJXD01000002.1|	----------	740
CasR_E._siraeum	----------EF----	811
PIG-018_GL0023397	----------DC----	806
PIG-025_GL0099734	----------	814
Ga0129317_1008067_chimera	----------	725
EYZ-362B_GL0088915	----------	787
uncultured_Ru_sp	----------	791
R._flav._FD-1	----------	789
tpg|DBYI01000091.1|	----------NI----	803

FIG. 18CCC

```
037_-_emb|OIZA01000315.1|          --------------------ET--KNLKLPKEKDRDIMKKYLPNDKDREEYNKFFKQYRN    841
emb|OCTW011587266.1|               -------NKGFLEDFEALWNSC--GKTSRLRKEDVLLYARYISPDHALKNYKMILNSYRN    807
k87_11092736                       WAQYATKRDENMAKHPQKWYDILASHYDELLAL---------------QATGWLATQARN    889
BMZ-11B_GL0037915                  ---------RRKHYPKRRWLNILKDNLAEMERI---------------DPTGMLAAQMRN    832
BMZ-11B_GL0037771                  --------RKTKRHFSKQWREWLNEKIADAKSS---------------QATGLLLREARN    821
BMZ-11B_GL0069617                  ---------SIRRHFTKKWRDIFRKEIDEAVKI---------------HFTGYLGVWARN    814
Ga0099364_10024192                 --------IRSQRHFSQYWLDIFARQIENAKKT---------------SESGYLLTAARN    936
530373_GL0023589                   ------NDFINETKLSKRVCALVNKNRTL--------------------CTNRTYQVYRN    838
emb|ODAI011611274.1|               -------RIGENKRANRHGLNCVLENRNALGS-D----------PASTDAAASLIWSYRN    893
EMG_10003641                       --------------LNSYSCRYLTHNISQLD------------------RCNDFIRQYRN    947
CasR_P1E0_metageno                 --------------LNRHACGYLRSNMNG--------------------ISNSLLTAYRN    801
emb|OIZX01000427.1|                --------------LKRSACEQIQHNMEN--------------------ISDRLLREYRN    837
160582958_gene49834                -KQETGTAHTKKIEKLNQQIAYIDKDIKN--------------------MHSYTCRNYRN    898
gi|1198542314|gb|NFIR01000008.1|   --------------PKKKVKDYLAKNFEY--------------------SNKWSLKEFRN    864
emb|OGNF01009141.1                 --------------INPRASAYLTVNLAN--------------------SDETAIRTFRN    801
Ga0129306_1000735                  --------LVRTLRREKRVCDNLHNNYAYLCGAD----------AEPGDASLSLLFVYRN    872
MH0288_GL0082219                   --------------LNRRAAKYLQTNLEH--------------------SDNTLIRQYRN    884
PIG-022_GL0026351                  --------------LNPHASRYLKVNISN--------------------ADEWSIVKFRN    779
PIG-028_GL0185479                  --------------LNSHARNYLRVNISN--------------------SDEWSIVKFRN    781
250twins_35838_GL0110300           RFAQIKRRESGCYFKSYHVYDYLSKNSNEFKQ----------------NHIDFAVTSYRN    893
PIG-014_GL0226364                  QHRENDKYYDSCMHFSRHSFEYISKNFAEVQSH---------------PDPLKLFIKYRN    857
250twins_36050_GL0158985           KVLRQNDKMIRKMHFTPHSLNYVQKNLESVQS----------------NGLAAVIKEYRN    916
```

FIG. 18DDD

emb|OJMM01002900.1|	LAYRANDKVVKHTHFNLHSYKYVKHNYEEISK----------ASKIITAYRN	887
emb|OGPN01002610.1|	PLRKANDKLLKQTHYTNHSYTYIVNNLNSFTDIDYCAKDVGLPAPNDKNDNASILGEMRN	930
emb|OHCP01000044.1|	ELRKENDQLLKKTHYSKHAYCYIVDNVNNLTGAVANDNGRGLPCLSEKNDNANLFLEMRN	916
PIG-046_GL0077813	AMRKVNDQLLKQTHFSKHSYCYIVDNVNRLTGGECRTDKRVLPVLNEKNDNAGILSDFRK	918
pig_chimera	AMRKVNDQLLKRTHFSKHSYCYIVDNVNRLTGGECRTDKRVLPVLNEKNDNAGILLDFRK	922
emb|OIEN01002196.1|	IKKNVLKYNGNPESLNHIPGEYICKNMEGYHEN----------TVRNFRN	829
O2.UC29-0_GL0096317	IKENIKKYGGDPKSIDHIPGEYIKKNLEGYCAG----------STNTFRN	823
CasR_Anaerobic_dig	----------LPVRKAEILRANLKHLDCED----------GFKYYRN	838
emb|OGDF01008514.1|	----------NPKKPTVKSIKEQYAFLTPY----------IFTTYRN	822
emb|OGZC01000639.1	TVYEKAEAAKNRHLKNVKWNCKTRENLENA----------DKNAIVAFRN	869
Ga0224415_10048792_chimera	LNRERVKNARNAYLKNIKNYIMIRLQLRDQTDS----------SGYLCGEFRD	854
emb|OCVV012889144.1|	INRERAKTAMNAHLRNTKWNDIMYGQLKDLADS----------KSRICSEFRN	856
Ga0187910_10006931	INREKAKTALNAHLRNTKWNVIIRENIRNT----------EKDACKQFRN	864
R._flav_XPD_	INREKAKTALNAYLRNTKWNVIIREDLLRI----------DNKTCTLFRN	859
Ga0224415_10007274	INREKAKTALNAHLRNTKWNVIIREDLLRR----------DNKACKIFRN	838
Ga0187911_10069260	----------YLKKRPVEYLKKDMAMVGNS----------SIRHFRN	791
Ga0187910_10015336	LDESASSYLSHKLVNKVHSYKYLKQNLDCADNT----------MINQFRN	965
ODAI_chimera	---TDSKSQQKAKSRRRRWCLLLKEDIDKLNGC----------ESLRILYRN	894
31009_GL0034153	----------YTSKKNQYIQKNMKIYSKI----------NKNFNVFVRYRN	792
DLF014_GL0011914_	-----KKNKKYGEYTSKKISTYIETNMKNFIGCE-----QFRTPERYRDTQIDMFVKYRN	871
CasR_R._albus	----------GKLGCNKKVGRYLKNNISCCTDG----------LINTYRN	833
Ga0187910_10040531	-------GH----FKNKKWRGIAEQNLRNSDVP----------VIKSFRN	826
tpg|DJXD01000002.1|	-------GHLKNQKGHRKWYVLVKNNLQNSDIT----------AVSSFRN	773
CasR_E._siraeum	DKSF-AENAANRYLRNARWYKLILDNLKKSERA----------VVNEFRN	850

FIG. 18EEE

```
PIG-018_GL0023397                  DLGR-AETAANRHLRKKRWYVQIFNNLKQADRN--------------------VINEFRN   845
PIG-025_GL0099734                  ----------NLHLRSKKWNRLMRVNLENYSAA--------------------ASQNFRN   844
Ga0129317_1008067_chimera          ----------HFISHNRRMRDCIEENVKNSEQLI--------------FGKEDAVIRFRN   761
EYZ-362B_GL0088915                 ---------KGYLARNKRLCQCVREDIDNAKHL------------------NGDITDYRN   820
uncultured_Ru_sp                   ----------LFLKKNERLRKCVEVDINNADSS--------------------MTRKYRN   821
R._flav._FD-1                      ----------RYLAGNKRLRDCVKQDIDNAKKWF-------------VSDKYNSITKYRN   826
tpg|DBYI01000091.1|                NQDKVQMPKNYFLARNKRWREAIEQDIDNAKKWF-------------IGEKFNNVKNYRN   850
                                                                    .                          *.

037_-_emb|OIZA01000315.1|          NIVHLKIIA-KLSEL---TKNIDK-----------DINSYFDIYHYCTQRVMFDYCKKNN   886
emb|OCTW011587266.1|               KIAHINVIM-SAGKY---TGGIK------------RMDSYFSVFQHLVQCDILSNPNNKG   851
k87_11092736                       DAEHLNPVN-EFDVY---IEDLRRYPEGTPKNKDYHIGSYFEIYHYIRQRAYLEEVLAKR   945
BMZ-11B_GL0037915                  DAMHLNAMF-KIAKF---VGDFRK---G----SSHEMRSYFELFHFLEQKTLLDSKEDFV   881
BMZ-11B_GL0037771                  DVEHLNVLR-AIPDY---IQDFRH---GEKG--ETAMNSYFELYHYLMQRLMLKNTELDL   872
BMZ-11B_GL0069617                  QSAHLNVLTQALPQF---IGQFAR---NSKTGKTARMTSYFQLYHYLLQRLLCEDLNFDV   868
Ga0099364_10024192                 CALHLNVLT-ALPEF---VGEFRK---T-----GDKMTSYFELYHFLLQKLMLAEAGLNL   984
530373_GL0023589                   IIMHLNAVS-AFPQY---LEGLK------------VVHSYFDLYHYTVMMLLHRQNEKE-   881
emb|ODAI011611274.1|               AAAHLTAVA-AAQEY---VSELR------------EIHSYFEVYHYAMQRYLKSGAEFA-   936
EMG_10003641                       KVAHLEVVS-NIDEY---LSGIK------------HIESYYALYHYLMQKCLLKNYRIE-   990
CasR_P1E0_metageno                 AVDHLNAIP-PLGSL---CRDIG------------RVDSYFALYHYAVQQYLNGRYYRK-   844
emb|OIZX01000427.1|                AVAHLNVIA-HLADY---SADMR------------EITSYYGLYHYLMQRHLFKRHAWQI   881
160582958_gene49834                LVAHLNVVS-KLQNYVSELPNDY------------QITSYFSFYHYCMQLGLMEKVSSK-   944
gi|1198542314|gb|NFIR01000008.1|   SVQHLNVIR-DAHKYIKYINDNK------------DVQSYFALYHYLVQRYISERAANRT   911
emb|OGNF01009141.1                 TAEHLEALR-NADKY---LNDLK------------QFDSYFEIYHYITQRNIKEKCEML-   844
```

FIG. 18FFF

```
Ga0129306_1000735                KAAHLSVLN-KGGRL---SGDLK-----------EAKSWFYVYHFLMQRVLEEEFRNT-    915
MH0288_GL0082219                 SVAHLNAVR-NMNLY---IGDLR-----------EITSYFGIYHYIMQRDLITRCQKV-    927
PIG-022_GL0026351                GTAHLNSVR-NIDTC---IEGLK-----------DFKSYFEIYAYTVQKSIKKYFDAEF    823
PIG-028_GL0185479                GVAHLNSVR-NIDTS---IKGLR-----------EFKSYFEIYAYTVQKSIKNYYDAES    825
250twins_35838_GL0110300         NVEHLNVVH-CMTKY---FSEVK-----------DVKSYYGVYCYIMQRMLCDELIIKN    937
PIG-014_GL0226364                NVMHLNIIR-DLNDY---LPDFKR---G---KAFEKTDTFYDLYCYCLQKMLLSEVD---    904
250twins_36050_GL0158985         AVAHLNIIN-RLDEY---IGSAR-------------ADSYYSLYCYCLQMYLSKNFSVG-    958
emb|OJMM01002900.1|              NVQHLNVMN-SITKY---LGDIS-----------EVTSYYSLYCYTLQRLLLDDNNND-    930
emb|OGPN01002610.1|              DIAHLNIVH-DMVKY---IEELK-----------DISSYYAFYCYVLQRRLVGKDPNC-    973
emb|OHCP01000044.1|              KIVHLNVVH-DMVKY---INEIK-----------NITSYYAFFCYVLQRMIIGNNSNE-    959
PIG-046_GL0077813                TIAHLNVVH-KMVDY---VAEIK-----------GITSYYAFFCYVLQRMPVGNKN-E-    960
pig_chimera                      TIAHLNVVH-KMVDY---VDEIK-----------GITSYYAFFCYVLQRMLVGNNLNE-    965
emb|OIEN01002196.1|              MVAHLTAVA-RVPLY---ISEVT-----------QIDSYYALYHYCMQMNILQGIEQSG    873
O2.UC29-0_GL0096317              AVAHLALMA-RVSEY---ASDIK-----------NVNSYFSLYHYCMQKHILYCARRFN    867
CasR_Anaerobic_dig               QVTHLNAIR-VAYKY---INEIK-----------SVHSYFALYHYIMQRHLYDSLQAKA    882
emb|OGDF01008514.1|              MIAHLAAVT-NAYKY---IPQMD-----------RFKSWFHLYHTVIQHSLIQQYEYDR    866
emb|OGZC01000639.1               IVAHLWIIR-DADRF---ITGMG-----------AMKRYFDCYHYLLQRELGYILEKSN    913
Ga0224415_10048792_chimera       KVAHLEVAR-HAHEY---IGNIK-----------EVNSYFQLYHYIMQCRLYDVLKNNT    898
emb|OCVV012889144.1|             KAAHLEVAR-YAHMY---INDIS-----------EVKSYFRLYHYIMQRRIIDVIENNP    900
Ga0187910_10006931               KADHLEVAR-YAYKY---INDIS-----------EVNSYFQLYHYIMQRIIIDSSGN--    906
R._flav_XPD_                     KAVHLEVAR-YVHAY---INDIA-----------EVNSYFQLYHYIMQRIIMNERYE--    901
Ga0224415_10007274               KVAHLEAIR-YAHLY---INDIA-----------EVNSYFQFYHYIMQRRIMAERYD--    880
Ga0187911_10069260               ATVHLNVIM-EAHRY---TKDIK-----------YIGSYYALYHYILQRHLLDKIEEDS    835
Ga0187910_10015336               NVAHLNTIR-NMDG----IENVT-----------GITSYFQIYHYLMQKALYKEFKKCR   1008
```

FIG. 18GGG

```
ODAI_chimera                    AVAHLNIPR-NAIDY---IGGIK-----------KADSYFAIYHYMMQQYLSKKHGLP-    937
31009_GL0034153                 NIAHLNSIK-NATKY---VNDIG-----------TFTSYFQLYHYIMQCDLKDWVKKEV    836
DLF014_GL0011914_               SIAHLNAVR-KASKY---INDIR-----------GFDTYFELYHYIMQRDLKDILEKIA    915
CasR_R._albus                   QVDHFAVVR-KIGNY---AAYIK-----------SIGSWFELYHYVIQRIVFDEYRFAL    877
Ga0187910_10040531              MAAHISVIR-NIDLY---IGDIQ-----------KVDSYFALYHFLMQKLIQRVVPENT    870
tpg|DJXD01000002.1|             IVAHISVVR-NSNEY---ISGIG-----------ELHSYFELYHYLVQSMIAKNNWYDT    817
CasR_E._siraeum                 TVCHLNAIR-NININ---IKEIK-----------EVENYFALYHYLIQKHLENRFADKK    894
PIG-018_GL0023397               SVCHLNAIR-NINTN---IEGIG-----------FVKNYFALYHYLIQRHISSKRIS--    887
PIG-025_GL0099734               AVAHLNPIR-NADVL---IEGIE-----------NITSYFDIYHYIVQRSVLNRTNKT-    887
Ga0129317_1008067_chimera       NVAHLSAIR-NANEY---IGDIR-----------EITSYFALYHYLMQRKLIDDCKVN-    804
EYZ-362B_GL0088915              CVAHLNVVR-NCDQY---LENID-----------TVHSYFSLYQYLVQHYLLKRFEKDK    864
uncultured_Ru_sp                CIAHLTVVR-ELKEY---IGDIC-----------TVDSYFSIYHYVMQRCITKRENDTK    865
R._flav._FD-1                   NVAHLTAVR-NCAEF---IGDIT-----------KIDSYFALYHYLIQRQLAKGLDHER    870
tpg|DBYI01000091.1|             NVAHLTAIR-NCAEF---IGEIT-----------KIDSYFALYHYLIQRQLAGRLDPNH    894
                                   *:                                  ::  :

037_-_emb|OIZA01000315.1|       NVVLA-----------------------KMKDLAHIKSDCDEFSSKHTYPYSSAVLRFM    922
emb|OCTW011587266.1|            KCFES-----------------------ESL--KPLLL-DMKFDGTDEKLYSKRLTRAL    884
k87_11092736                    KEYRD---------------------SGSFTDEQLDK-LQKILDDIRARGSYDKNLLKLE    983
BMZ-11B_GL0037915               QGMRE---------------------S----------------YGDRIRVEPCVDLIKYL    904
BMZ-11B_GL0037771               S-HWS---------------------G-----------------WIMRSGRPDRDLIQIA    893
BMZ-11B_GL0069617               PEEWK---------------------N-----------------KIQETGLPCNDLIHIC    890
Ga0099364_10024192              D-EYR---------------------E-----------------RIDTYQTACKDLINIT    1005
530373_GL0023589                ---------------KKRNEKEKKQNEKENNMILSTA-ALEKYDMVRKYQSANKDFIYAI    925
```

FIG. 18HHH

```
emb|ODAI011611274.1|                -------------------------ELVSKNGPASGK-IAAWANAVDRCHSFCKDWLWLL    970
EMG_10003641                        ----------------------------DHSQNEYKN-LNDFSSKLDKHGTYVKDFVKAL    1021
CasR_P1E0_metageno                  ----------------------------------TPR-EQELFAAMAQHRTWCSDLVKAL    869
emb|OIZX01000427.1|                 RQPERP--------------TEEEQKLIEQEQKQLAW-EKALFDKTLQYHSYNKDLVKAL    926
160582958_gene49834                 ---------------NIPLVESLKNEANDAQSYSAKK-TLEYFDLIEKNRTYCKDFLKAL    988
gi|1198542314|gb|NFIR01000008.1|    ----------------------------DKESLSEGR-LQYYLSQVKEYRTYCKDFVKAL    942
emb|OGNF01009141.1                  ----------------------------------KEQ-TVKYNNDLLKYHGYSKDFVKAL    869
Ga0129306_1000735                   -------------------------------QALPER-LRELLMMAERYRGCSKDLIKVL    943
MH0288_GL0082219                    -------------------------------KMKNPV-VVGYFDKVRTYRTCCKDLVKAL    955
PIG-022_GL0026351                   SEDRNIEDGKISLCNGVYIHERQKNGKPKITVKINPK-TIEYMESVDKYGSYTKDFVKAL    882
PIG-028_GL0185479                   P-----------------------------VQINPK-TIEYMESVDKYGSYTKDFVKAL    854
250twins_35838_GL0110300            QDKPD----------------------------VRQT-FEEYNRLLKDHGTYSKNLMWLL    968
PIG-014_GL0226364                   ------------------------------------G-FSAINENIDKYCTASKDFMNLL    927
250twins_36050_GL0158985            -----------------------------------Y-LINVQKQLEEHHTYMKDLMWLL    981
emb|OJMM01002900.1|                 ------------------------------------K-FASIKGNLRKFGIYNKDFMWLL    953
emb|OGPN01002610.1|                 ---------------------------------QNKF-KAKYAKELNDYGTYNKNLMWML    999
emb|OHCP01000044.1|                 ---------------------------------QNKF-KAKYSKTLQEFGTYSKDLMWVL    985
PIG-046_GL0077813                   ---------------------------------KNAL-KEKYSETAKSYGTYSKDFMWLI    986
pig_chimera                         ---------------------------------KNAI-KEKYSATVKSFGTYSKDFMWLI    991
emb|OIEN01002196.1|                 KIL----------------------------DNIK-LKNALENARVHRTYSKDAVKYL    902
O2.UC29-0_GL0096317                 KQ-----------------------------GNEK-MAAALNDAEIHRSYSKDAVKFL    895
CasR_Anaerobic_dig                  KDS----------------------------SGF-VIDALKKSFEHKIYSKDLLHVL    910
emb|OGDF01008514.1|                 DYGRK----------------------GAPVVSER-VLQLLEQCREHSNYSRDLLHIL    901
emb|OGZC01000639.1                  ----Q----------------------------GSEY-TKKSLEKVQQYHSYCKDFLHML    940
```

FIG. 18III

```
Ga0224415_10048792_chimera      KAEAM----------------------------VKGK-AKEYFEALEKEGTYNDKLLKIA    929
emb|OCVV012889144.1|            K--AK----------------------------YEGK-VKVYFEDVKKNKKYNKNLLKLM    929
Ga0187910_10006931              ----N----------------------------ANGM-IKKYYESVISDKKYNDRLLKLL    933
R._flav_XPD_                    ----K----------------------------SSGK-VSEYFDAVNDEKKYNDRLLKLL    928
Ga0224415_10007274              ----K----------------------------SSGK-VREYFDAVNNEKKYNDRLLKLL    907
Ga0187911_10069260              YAEKT----------------------------VS-EKLWESQISQYGTYSKDFVKAL    864
Ga0187910_10015336              ENAVRKWIPYITENAEPKYVYWNKKEQQEVEVSFNPKIFGYMENIKNHSNTYCKDFVKAL   1068
ODAI_chimera                    ----------------------------------------EGLGCTHGVYCKDLVKAL    955
31009_GL0034153                 EEKKI----------------------------SDKY-YNYLEKYYKNNNIYLKDFVKAI    867
DLF014_GL0011914_               EGAAKET--------------------EEGQKQDEL-LKTYFDNLDKYGTYVKDFVKAL    953
CasR_R._albus                   N---------------------------------NT-ESNYKNSIIKHHTYCKDMVKAL    902
Ga0187910_10040531              KGL-------------------------------SDQ-TKKYYDALEQYNTYCKDFVKAY    898
tpg|DJXD01000002.1|             S-H-------------------------------QPK-TAEYLNNLKKHHTYCKDFVKAY    844
CasR_E._siraeum                 VE--------------------------------RD-TGDFISKLEEHKTYCKDFVKAY    920
PIG-018_GL0023397               FN--------------------------------SF-SAEYISNLEKYHTPSEYFIKAY    913
PIG-025_GL0099734               ---------------------------------PK-MKEYEELINKYHTYNKDFVKAL    911
Ga0129317_1008067_chimera       ---------------------------------DT-AHKYFEQLTKYKTYVMDMVKAL    828
EYZ-362B_GL0088915              YRCK----------------------------DDL-AIHYYSQVKQYRSYVKDLVKAL    893
uncultured_Ru_sp                QEE---------------------------------KIKYEDDLLKNHGYTKDFVKAL    890
R._flav._FD-1                   SGF--------------------------------DR-NYPQYAPLFKWHTYVKDVVKAL    897
tpg|DBYI01000091.1|             PGF--------------------------------EK-NYPQYAPLFKWNTYVKDMVKAL    921


037_-_emb|OIZA01000315.1|       NLPFAYNVPRFKNLSYKKFFDKQ-----------------------------------    945
```

FIG. 18JJJ

emb|OCTW011587266.1|	NIPFGYNVPRYKNLTFEKIYLKSSINE----------	911
k87_11092736	YLPFAYNLPRYKNLTTEALFDDDSVSGKKRVAEWREREKTREAEREQRRQR--------	1034
BMZ-11B_GL0037915	YITLGYSLPRYKTMTVSGLFDPDSEDAKERAKGKRA----------	940
BMZ-11B_GL0037771	FVSLAYNLPRYRNLTKEHHFDDTVLQKIREKESLD----------	928
BMZ-11B_GL0069617	FVPLGYNLPRYKNLTIECLFDEDSVSAKEGKKA----------	923
Ga0099364_10024192	YVSLGYNLPRYKNLTCEPLFDEESATGKERQTRLDEKSKEKKQRKGGQK----------	1054
530373_GL0023589	NSPLAYNAARYINLSCRTKFEEGFGK----------	951
emb|ODAI011611274.1|	NVPFAYNPARYKNLSIANLFDKNEAAPVTEDASEQKEDE----------	1009
EMG_10003641	NVPFGYNLPRYKNLSIDELFDRNKLKTGGTIEMKGE----------	1057
CasR_P1E0_metageno	NTPFGYNLARYKNLSIDGLFDREGDHVVREDGEKPAE----------	906
emb|OIZX01000427.1|	NAPFGYNLARYKNLSIEPLFSKEAAPAAEIKATHA----------	961
160582958_gene49834	NAPFSYNLPRFKNLSIEALFDKNIVYEQADLKKE----------	1022
gi|1198542314|gb|NFIR01000008.1|	NVPFAYNLPRYKNLSIDELFDRNNYLPNKAKKWIPEKKENGEYVMEDCGNKDAGQVENA	1001
emb|OGNF01009141.1	CVPFGYNLPRFKNLSIDALFDKNDKREKLKKGFED----------	904
Ga0129306_1000735	NLTFAYNLPRYKNLSIDGRFDKNHPDPSDE----------	973
MH0288_GL0082219	NITFGYNLPRYKNLSIEGLFDKNRPGDKKESALIAE----------	991
PIG-022_GL0026351	NTPFGYCLARYKNLSIEPLFDRNTVYNKEDVTEEFSVNSKEQ----------	924
PIG-028_GL0185479	NTPFGYCLARYKNLSIEPLFDRNTVYNKKDVKEEFAVNSK----------	894
250twins_35838_GL0110300	NFPFAYNLARYKNLSNEDLFNAKNNDQKSK----------	998
PIG-014_GL0226364	NIIFAYNVPRYKNLSNKELFYRLDSEQ----------	954
250twins_36050_GL0158985	NIPFAYNLARYKNLSNEKLFYDEEAAAEKADKAENERGE----------	1020
emb|OJMM01002900.1|	NIPFAYNLPRYKNLSNEEIFYDELQK----------	979
emb|OGPN01002610.1|	NLPFAYNLPRYKNLSSEFLFYDMEYNKKDDE----------	1030
emb|OHCP01000044.1|	NLPFAYNLPRYKNLSNEQLFYDEEERMEKIVGRKNDSR----------	1023

FIG. 18KKK

PIG-046_GL0077813	NLPFAYNLPRYKNLSNEQLFYDEEERNETEGQIDRL----------	1022
pig_chimera	NLPFAYNLPRYKNLSNEQLFYDEEERNETEEQIDRL----------	1027
emb|OIEN01002196.1|	CLPFAYNISRYKALTIKDLFDWTEYSCKKDE----------	933
O2.UC29-0_GL0096317	CLPFAYNLPRYKALTVEGLFDWNEYGQDDD----------	925
CasR_Anaerobic_dig	HSPFGYNTARYKNLSIEALFDKNESRPEVNPLSTND----------	946
emb|OGDF01008514.1|	NLPFGYNLPRYLNLSSEKYFDANAI----------	926
emb|OGZC01000639.1	CLPFAYCIPRYKNLSIAELFDRHEPEAEPKEEASSVNNSQFITT----------	984
Ga0224415_10048792_chimera	CVPFGYCIPRYKNLSMEELFDMNEEKKFKKKAPENT----------	965
emb|OCVV012889144.1|	CVPFGYCIPRFKNLSIEQMFDMNETDNSDKKKEK----------	963
Ga0187910_10006931	CVPFGYCIPRFKNLSIEALFDKNEAAKYDKIKKKVAVR----------	971
R._flav_XPD_	CVPFGYCIPRFKNLSIEALFDRNEAAKFDKEKKKVSGNS----------	967
Ga0224415_10007274	CVPFGYCIPRFKNLSIEALFDMNEAVKFDKEKK----------	940
Ga0187911_10069260	CCPFGYNLPRFKNLSIEQLFDRNESKEITDATAPRQ----------	900
Ga0187910_10015336	CAPFAYNLPRFKNLSIEELFDMHELSEEPKESMKLTD----------	1105
ODAI_chimera	NAPFGYNLPRFKNLSVEALFDMNEEPAPEGKERSKNADGCN----------	996
31009_GL0034153	NVPFAYNYSRFKNLSIDKLFDKNDPRKKKWIVIFNF----------	903
DLF014_GL0011914_	NIPFAYNYPRYKNLSIDELFDKNNTRKTIKIQDKKLEERT----------	993
CasR_R._albus	NTPFGYDLPRYKNLSIGDLFDRNNYLNKTKESIDANSSIDSQ----------	944
Ga0187910_10040531	CTPFAYVTPRYKNLTIDGLFDRNRPGEDK----------	927
tpg|DJXD01000002.1|	CIPFGYVVPRYKNLTINELFDRNNPNPEPKEEV----------	877
CasR_E._siraeum	CTPFGYNLVRYKNLTIDGLFDKNYPGKDDSDEQK----------	954
PIG-018_GL0023397	CTPFAYNLVRYKNLTIDGLFDRNAPSEK----------	941
PIG-025_GL0099734	CAPFGYNVVRFKALTIYEMFDRNYPEEKRAEEARKSC----------	948
Ga0129317_1008067_chimera	CSPFGYNLPRFKNLSIEGKFDMHESK----------	854

FIG. 18LLL

```
EYZ-362B_GL0088915      NAPFGYNIPRFKNLTVEQLFDRNADPETITA---------------------------- 924
uncultured_Ru_sp        NSPFGYNIPRFKNLSIEQLFDRNEYLTEK----------------------------- 919
R._flav._FD-1           NAPFGYNIPRFKNLSIDALFDRNEIKKNDGEKKSDD---------------------- 933
tpg|DBYI01000091.1|     NSPFGYNIPRFKDLSIDALFDRNEMKEETDDEKKIQT------------------- 958
                           :.*  *:  ::   :
```

FIG. 18MMM

FIG. 19A

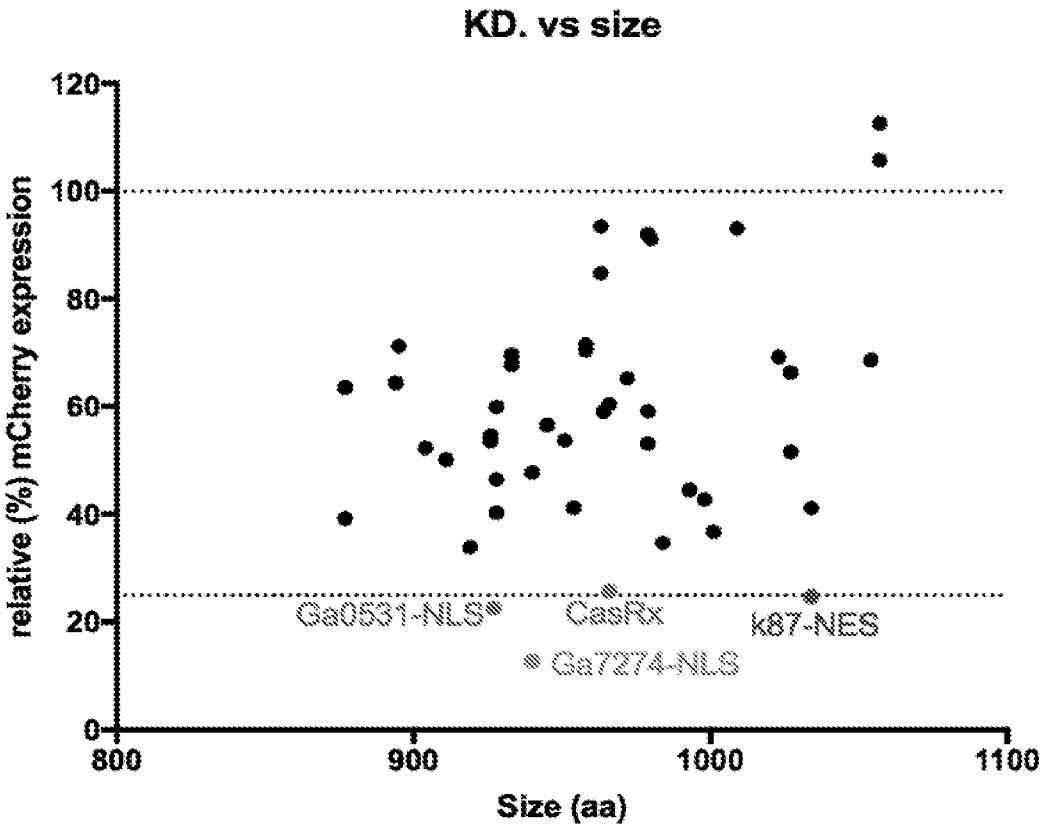

KD. vs size
relative (%) mCherry expression
120
100
80
60
40
20
0
800
900
1000
1100
Size (aa)
Ga0531-NLS
CasRx
k87-NES
Ga7274-NLS

FIG. 19B

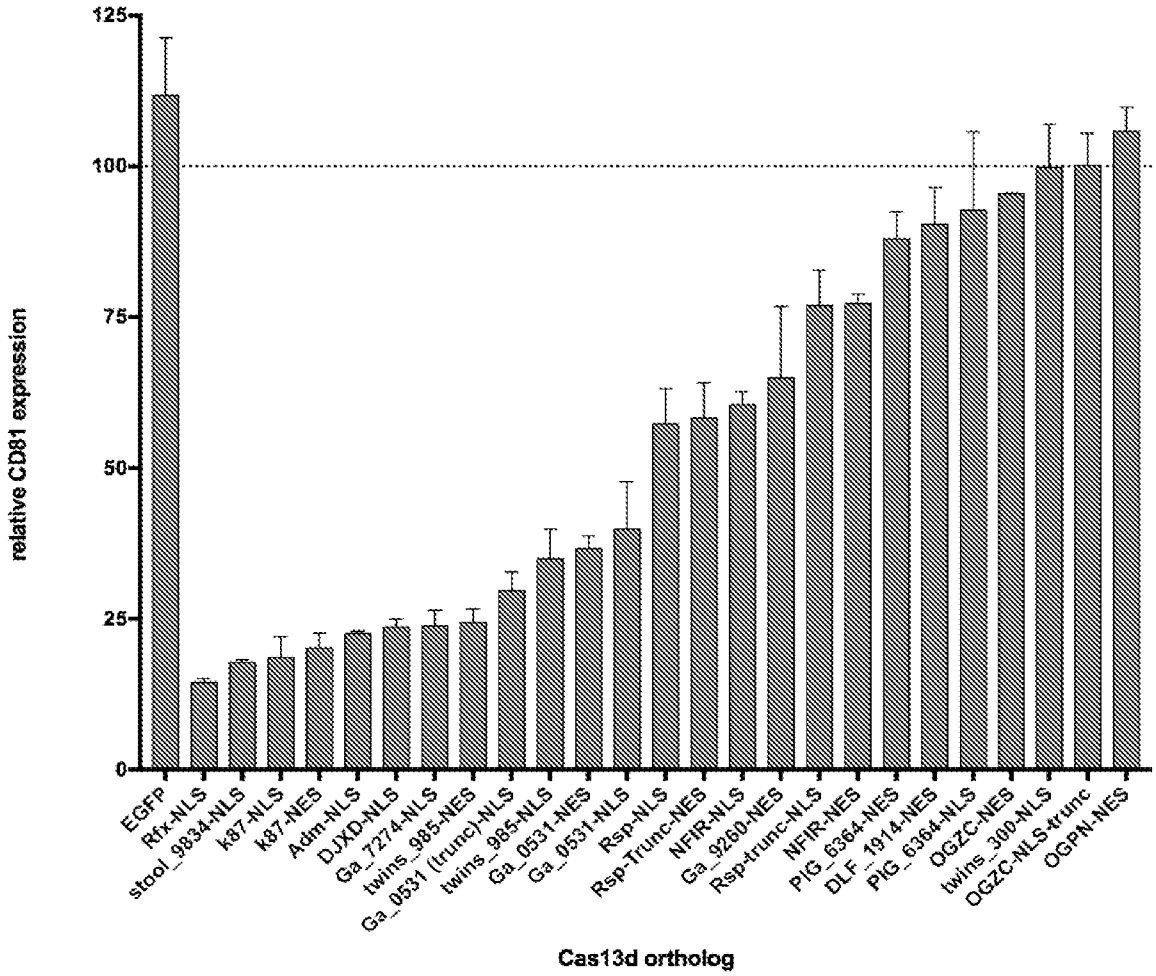

125
100
75
50
25
0
relative CD81 expression
EGFP
Rfx-NLS
stool_9834-NLS
k87-NLS
k87-NES
Adm-NLS
DJXD-NLS
Ga_7274-NLS
twins_985-NES
Ga_0531 (trunc)-NLS
twins_985-NLS
Ga_0531-NES
Ga_0531-NLS
Rsp-NLS
Rsp-Trunc-NES
NFIR-NLS
Ga_9260-NES
Rsp-trunc-NLS
NFIR-NES
PIG_6364-NES
DLF_1914-NES
PIG_6364-NLS
OGZC-NES
twins_300-NLS
OGZC-NLS-trunc
OGPN-NES
Cas13d ortholog

FIG. 20

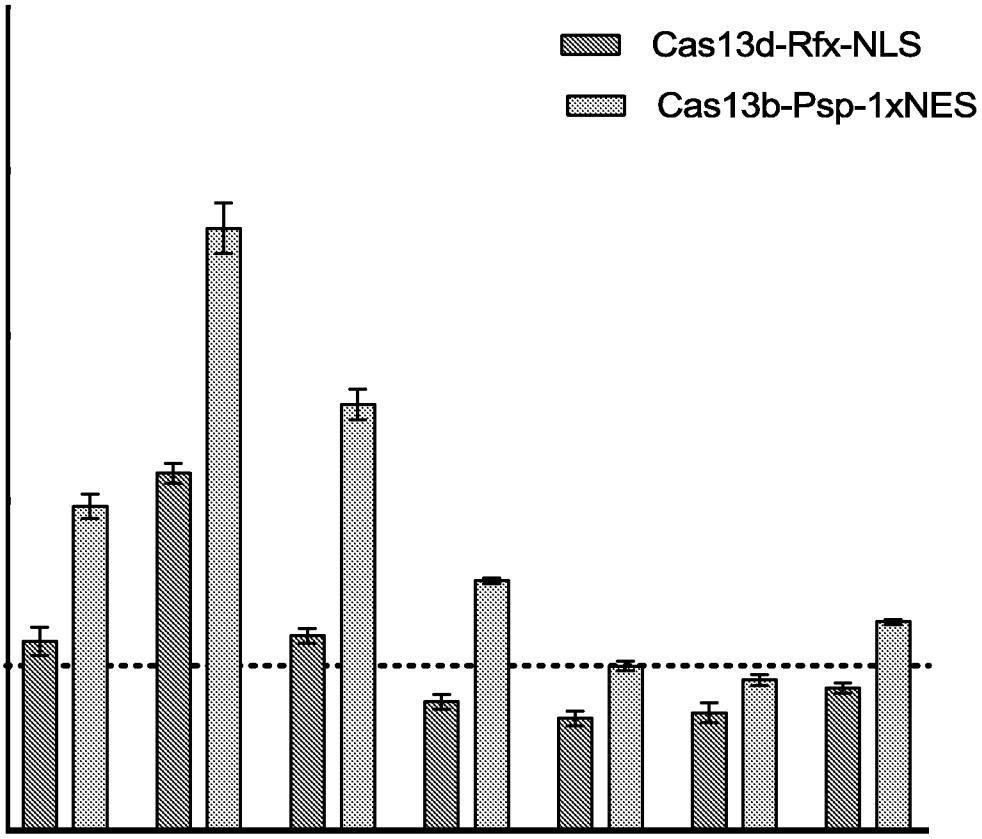


FIG. 21

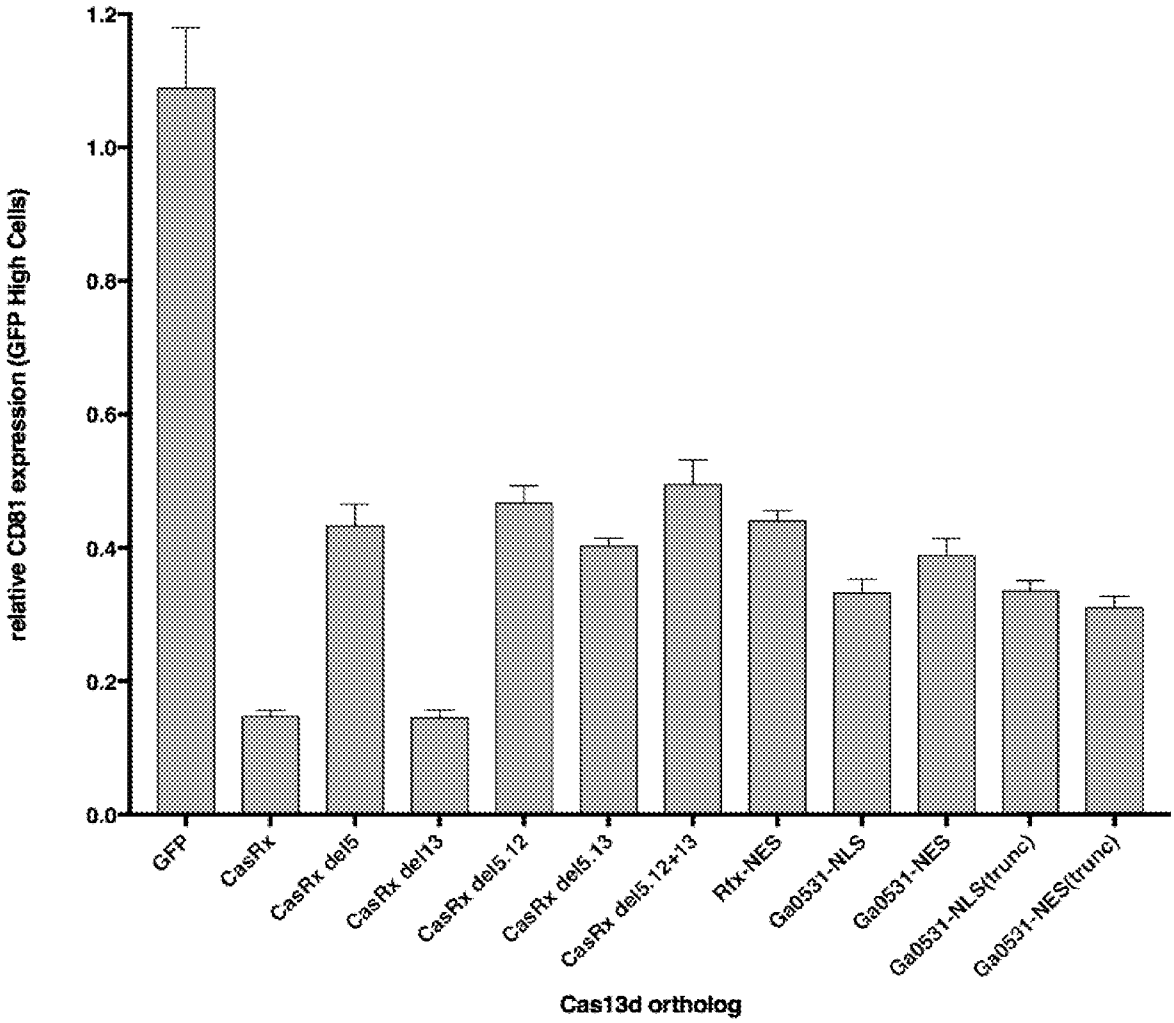

relative CD81 expression (GFP High Cells)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
GFP
CasRx
CasRx del5
CasRx del13
CasRx del5.12
CasRx del5.13
CasRx del5.12+13
RfxNES
Ga0531-NLS
Ga0531-NES
Ga0531-NLS(trunc)
Ga0531-NES(trunc)
Cas13d ortholog

FIG. 22

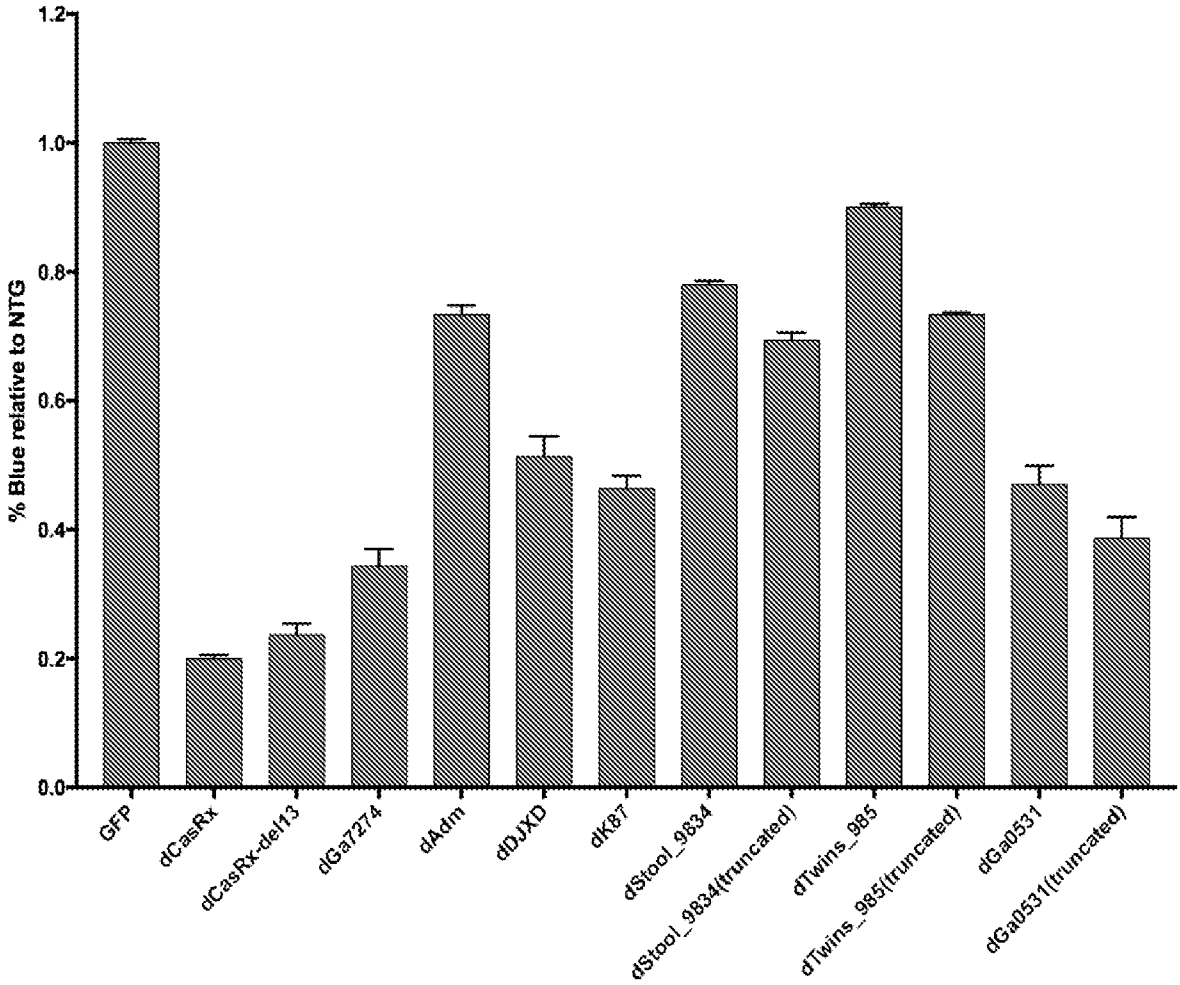

% Blue relative to NTG
1.2
1.0
0.8
0.6
0.4
0.2
0.0
GFP
dCasRx
dCasRx-del13
dGa7274
dAdm
dDJXD
dK87
dStool_9834
dStool_9834(truncated)
dTwins_985
dTwins_985(truncated)
dGa0531
dGa0531(truncated)

FIG. 23A

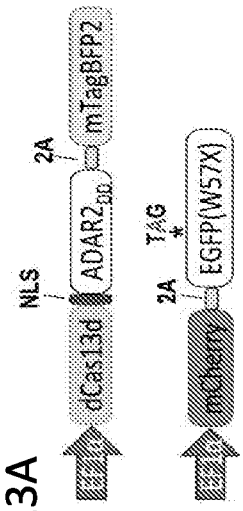


FIG. 23B

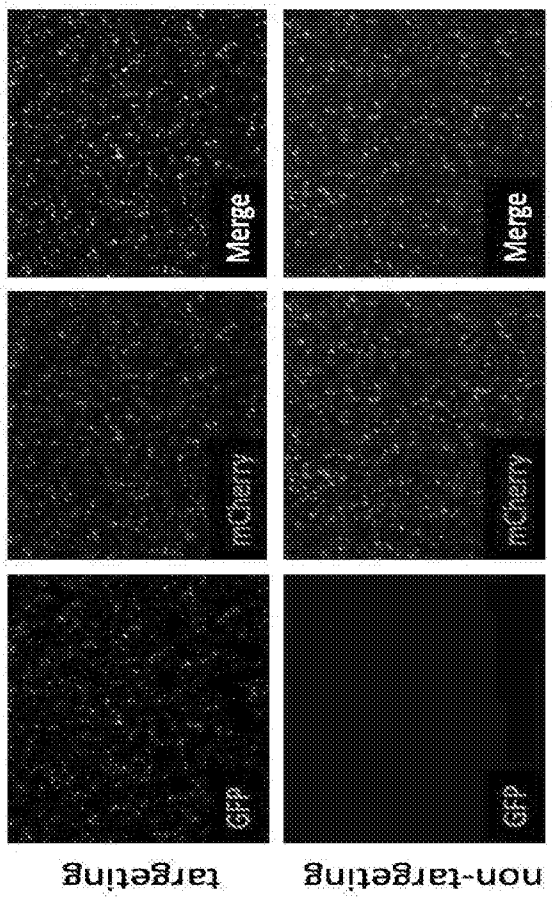
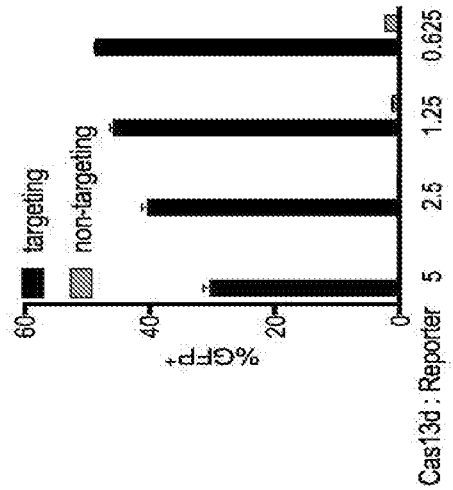


FIG. 23C



SEQUENCE LISTING

<110> Salk Institute for Biological Studies

<120> RNA TARGETING METHODS AND COMPOSITIONS

<130> 7158-99284-05

<150> 62/548846
<151> 2017-08-22

<150> 62/572963
<151> 2017-10-16

<150> 62/639178
<151> 2018-03-06

<150> 15/937,699
<151> 2018-03-27

<160> 296

<170> PatentIn version 3.5

<210> 1
<211> 954
<212> PRT
<213> Artificial Sequence

<220>
<223> Eubacterium siraeum nCas1

<400> 1

Met Gly Lys Lys Ile His Ala Arg Asp Leu Arg Glu Gln Arg Lys Thr
1 5 10 15

Asp Arg Thr Glu Lys Phe Ala Asp Gln Asn Lys Lys Arg Glu Ala Glu
20 25 30

Arg Ala Val Pro Lys Lys Asp Ala Ala Val Ser Val Lys Ser Val Ser
35 40 45

Ser Val Ser Ser Lys Lys Asp Asn Val Thr Lys Ser Met Ala Lys Ala
50 55 60

Ala Gly Val Lys Ser Val Phe Ala Val Gly Asn Thr Val Tyr Met Thr
65 70 75 80

```
Ser Phe Gly Arg Gly Asn Asp Ala Val Leu Glu Gln Lys Ile Val Asp
                85                  90                  95

Thr Ser His Glu Pro Leu Asn Ile Asp Asp Pro Ala Tyr Gln Leu Asn
            100                 105                 110

Val Val Thr Met Asn Gly Tyr Ser Val Thr Gly His Arg Gly Glu Thr
        115                 120                 125

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

Lys Asp Glu Pro Glu Gln Ser Val Pro Thr Asp Met Leu Cys Leu Lys
145                 150                 155                 160

Pro Thr Leu Glu Lys Lys Phe Phe Gly Lys Glu Phe Asp Asp Asn Ile
                165                 170                 175

His Ile Gln Leu Ile Tyr Asn Ile Leu Asp Ile Glu Lys Ile Leu Ala
            180                 185                 190

Val Tyr Ser Thr Asn Ala Ile Tyr Ala Leu Asn Asn Met Ser Ala Asp
        195                 200                 205

Glu Asn Ile Glu Asn Ser Asp Phe Phe Met Lys Arg Thr Thr Asp Glu
    210                 215                 220

Thr Phe Asp Asp Phe Glu Lys Lys Lys Glu Ser Thr Asn Ser Arg Glu
225                 230                 235                 240

Lys Ala Asp Phe Asp Ala Phe Glu Lys Phe Ile Gly Asn Tyr Arg Leu
                245                 250                 255

Ala Tyr Phe Ala Asp Ala Phe Tyr Val Asn Lys Lys Asn Pro Lys Gly
            260                 265                 270

Lys Ala Lys Asn Val Leu Arg Glu Asp Lys Glu Leu Tyr Ser Val Leu
        275                 280                 285
```

```
Thr Leu Ile Gly Lys Leu Arg His Trp Cys Val His Ser Glu Glu Gly
    290                 295                 300

Arg Ala Glu Phe Trp Leu Tyr Lys Leu Asp Glu Leu Lys Asp Asp Phe
305                 310                 315                 320

Lys Asn Val Leu Asp Val Val Tyr Asn Arg Pro Val Glu Glu Ile Asn
                325                 330                 335

Asn Arg Phe Ile Glu Asn Asn Lys Val Asn Ile Gln Ile Leu Gly Ser
            340                 345                 350

Val Tyr Lys Asn Thr Asp Ile Ala Glu Leu Val Arg Ser Tyr Tyr Glu
        355                 360                 365

Phe Leu Ile Thr Lys Lys Tyr Lys Asn Met Gly Phe Ser Ile Lys Lys
    370                 375                 380

Leu Arg Glu Ser Met Leu Glu Gly Lys Gly Tyr Ala Asp Lys Glu Tyr
385                 390                 395                 400

Asp Ser Val Arg Asn Lys Leu Tyr Gln Met Thr Asp Phe Ile Leu Tyr
                405                 410                 415

Thr Gly Tyr Ile Asn Glu Asp Ser Asp Arg Ala Asp Asp Leu Val Asn
            420                 425                 430

Thr Leu Arg Ser Ser Leu Lys Glu Asp Asp Lys Thr Thr Val Tyr Cys
        435                 440                 445

Lys Glu Ala Asp Tyr Leu Trp Lys Lys Tyr Arg Glu Ser Ile Arg Glu
    450                 455                 460

Val Ala Asp Ala Leu Asp Gly Asp Asn Ile Lys Lys Leu Ser Lys Ser
465                 470                 475                 480

Asn Ile Glu Ile Gln Glu Asp Lys Leu Arg Lys Cys Phe Ile Ser Tyr
                485                 490                 495
```

```
Ala Asp Ser Val Ser Glu Phe Thr Lys Leu Ile Tyr Leu Leu Thr Arg
            500                 505                 510

Phe Leu Ser Gly Lys Glu Ile Asn Asp Leu Val Thr Thr Leu Ile Asn
        515                 520                 525

Lys Phe Asp Asn Ile Arg Ser Phe Leu Glu Ile Met Asp Glu Leu Gly
    530                 535                 540

Leu Asp Arg Thr Phe Thr Ala Glu Tyr Ser Phe Phe Glu Gly Ser Thr
545                 550                 555                 560

Lys Tyr Leu Ala Glu Leu Val Glu Leu Asn Ser Phe Val Lys Ser Cys
                565                 570                 575

Ser Phe Asp Ile Asn Ala Lys Arg Thr Met Tyr Arg Asp Ala Leu Asp
            580                 585                 590

Ile Leu Gly Ile Glu Ser Asp Lys Thr Glu Glu Asp Ile Glu Lys Met
        595                 600                 605

Ile Asp Asn Ile Leu Gln Ile Asp Ala Asn Gly Asp Lys Lys Leu Lys
    610                 615                 620

Lys Asn Asn Gly Leu Arg Asn Phe Ile Ala Ser Asn Val Ile Asp Ser
625                 630                 635                 640

Asn Arg Phe Lys Tyr Leu Val Arg Tyr Gly Asn Pro Lys Lys Ile Arg
                645                 650                 655

Glu Thr Ala Lys Cys Lys Pro Ala Val Arg Phe Val Leu Asn Glu Ile
            660                 665                 670

Pro Asp Ala Gln Ile Glu Arg Tyr Tyr Glu Ala Cys Cys Pro Lys Asn
        675                 680                 685

Thr Ala Leu Cys Ser Ala Asn Lys Arg Arg Glu Lys Leu Ala Asp Met
    690                 695                 700
```

```
Ile Ala Glu Ile Lys Phe Glu Asn Phe Ser Asp Ala Gly Asn Tyr Gln
705                 710                 715                 720

Lys Ala Asn Val Thr Ser Arg Thr Ser Glu Ala Glu Ile Lys Arg Lys
                725                 730                 735

Asn Gln Ala Ile Ile Arg Leu Tyr Leu Thr Val Met Tyr Ile Met Leu
            740                 745                 750

Lys Asn Leu Val Asn Val Asn Ala Arg Tyr Val Ile Ala Phe His Cys
        755                 760                 765

Val Glu Arg Asp Thr Lys Leu Tyr Ala Glu Ser Gly Leu Glu Val Gly
    770                 775                 780

Asn Ile Glu Lys Asn Lys Thr Asn Leu Thr Met Ala Val Met Gly Val
785                 790                 795                 800

Lys Leu Glu Asn Gly Ile Ile Lys Thr Glu Phe Asp Lys Ser Phe Ala
                805                 810                 815

Glu Asn Ala Ala Asn Arg Tyr Leu Arg Asn Ala Arg Trp Tyr Lys Leu
            820                 825                 830

Ile Leu Asp Asn Leu Lys Lys Ser Glu Arg Ala Val Val Asn Glu Phe
        835                 840                 845

Arg Asn Thr Val Cys His Leu Asn Ala Ile Arg Asn Ile Asn Ile Asn
    850                 855                 860

Ile Lys Glu Ile Lys Glu Val Glu Asn Tyr Phe Ala Leu Tyr His Tyr
865                 870                 875                 880

Leu Ile Gln Lys His Leu Glu Asn Arg Phe Ala Asp Lys Lys Val Glu
                885                 890                 895

Arg Asp Thr Gly Asp Phe Ile Ser Lys Leu Glu Glu His Lys Thr Tyr
            900                 905                 910
```

```
Cys Lys Asp Phe Val Lys Ala Tyr Cys Thr Pro Phe Gly Tyr Asn Leu
        915                 920                 925


Val Arg Tyr Lys Asn Leu Thr Ile Asp Gly Leu Phe Asp Lys Asn Tyr
    930                 935                 940


Pro Gly Lys Asp Asp Ser Asp Glu Gln Lys
945                 950


<210>  2
<211>  954
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Eubacterium siraeum nCas1 HEPN-/-

<400>  2

Met Gly Lys Lys Ile His Ala Arg Asp Leu Arg Glu Gln Arg Lys Thr
1               5                   10                  15


Asp Arg Thr Glu Lys Phe Ala Asp Gln Asn Lys Lys Arg Glu Ala Glu
            20                  25                  30


Arg Ala Val Pro Lys Lys Asp Ala Ala Val Ser Val Lys Ser Val Ser
        35                  40                  45


Ser Val Ser Ser Lys Lys Asp Asn Val Thr Lys Ser Met Ala Lys Ala
    50                  55                  60


Ala Gly Val Lys Ser Val Phe Ala Val Gly Asn Thr Val Tyr Met Thr
65                  70                  75                  80


Ser Phe Gly Arg Gly Asn Asp Ala Val Leu Glu Gln Lys Ile Val Asp
                85                  90                  95


Thr Ser His Glu Pro Leu Asn Ile Asp Asp Pro Ala Tyr Gln Leu Asn
            100                 105                 110
```

```
Val Val Thr Met Asn Gly Tyr Ser Val Thr Gly His Arg Gly Glu Thr
        115                 120                 125

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

Lys Asp Glu Pro Glu Gln Ser Val Pro Thr Asp Met Leu Cys Leu Lys
145                 150                 155                 160

Pro Thr Leu Glu Lys Lys Phe Phe Gly Lys Glu Phe Asp Asp Asn Ile
                165                 170                 175

His Ile Gln Leu Ile Tyr Asn Ile Leu Asp Ile Glu Lys Ile Leu Ala
            180                 185                 190

Val Tyr Ser Thr Asn Ala Ile Tyr Ala Leu Asn Asn Met Ser Ala Asp
        195                 200                 205

Glu Asn Ile Glu Asn Ser Asp Phe Phe Met Lys Arg Thr Thr Asp Glu
    210                 215                 220

Thr Phe Asp Asp Phe Glu Lys Lys Lys Glu Ser Thr Asn Ser Arg Glu
225                 230                 235                 240

Lys Ala Asp Phe Asp Ala Phe Glu Lys Phe Ile Gly Asn Tyr Arg Leu
                245                 250                 255

Ala Tyr Phe Ala Asp Ala Phe Tyr Val Asn Lys Lys Asn Pro Lys Gly
            260                 265                 270

Lys Ala Lys Asn Val Leu Arg Glu Asp Lys Glu Leu Tyr Ser Val Leu
        275                 280                 285

Thr Leu Ile Gly Lys Leu Ala His Trp Cys Val Ala Ser Glu Glu Gly
    290                 295                 300

Arg Ala Glu Phe Trp Leu Tyr Lys Leu Asp Glu Leu Lys Asp Asp Phe
305                 310                 315                 320
```

```
Lys Asn Val Leu Asp Val Val Tyr Asn Arg Pro Val Glu Glu Ile Asn
                325                 330                 335

Asn Arg Phe Ile Glu Asn Asn Lys Val Asn Ile Gln Ile Leu Gly Ser
            340                 345                 350

Val Tyr Lys Asn Thr Asp Ile Ala Glu Leu Val Arg Ser Tyr Tyr Glu
        355                 360                 365

Phe Leu Ile Thr Lys Lys Tyr Lys Asn Met Gly Phe Ser Ile Lys Lys
    370                 375                 380

Leu Arg Glu Ser Met Leu Glu Gly Lys Gly Tyr Ala Asp Lys Glu Tyr
385                 390                 395                 400

Asp Ser Val Arg Asn Lys Leu Tyr Gln Met Thr Asp Phe Ile Leu Tyr
                405                 410                 415

Thr Gly Tyr Ile Asn Glu Asp Ser Asp Arg Ala Asp Asp Leu Val Asn
            420                 425                 430

Thr Leu Arg Ser Ser Leu Lys Glu Asp Asp Lys Thr Thr Val Tyr Cys
        435                 440                 445

Lys Glu Ala Asp Tyr Leu Trp Lys Lys Tyr Arg Glu Ser Ile Arg Glu
    450                 455                 460

Val Ala Asp Ala Leu Asp Gly Asp Asn Ile Lys Lys Leu Ser Lys Ser
465                 470                 475                 480

Asn Ile Glu Ile Gln Glu Asp Lys Leu Arg Lys Cys Phe Ile Ser Tyr
                485                 490                 495

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

Phe Leu Ser Gly Lys Glu Ile Asn Asp Leu Val Thr Thr Leu Ile Asn
        515                 520                 525
```

```
Lys Phe Asp Asn Ile Arg Ser Phe Leu Glu Ile Met Asp Glu Leu Gly
    530                 535                 540

Leu Asp Arg Thr Phe Thr Ala Glu Tyr Ser Phe Phe Glu Gly Ser Thr
545                 550                 555                 560

Lys Tyr Leu Ala Glu Leu Val Glu Leu Asn Ser Phe Val Lys Ser Cys
                565                 570                 575

Ser Phe Asp Ile Asn Ala Lys Arg Thr Met Tyr Arg Asp Ala Leu Asp
            580                 585                 590

Ile Leu Gly Ile Glu Ser Asp Lys Thr Glu Glu Asp Ile Glu Lys Met
        595                 600                 605

Ile Asp Asn Ile Leu Gln Ile Asp Ala Asn Gly Asp Lys Lys Leu Lys
    610                 615                 620

Lys Asn Asn Gly Leu Arg Asn Phe Ile Ala Ser Asn Val Ile Asp Ser
625                 630                 635                 640

Asn Arg Phe Lys Tyr Leu Val Arg Tyr Gly Asn Pro Lys Lys Ile Arg
                645                 650                 655

Glu Thr Ala Lys Cys Lys Pro Ala Val Arg Phe Val Leu Asn Glu Ile
            660                 665                 670

Pro Asp Ala Gln Ile Glu Arg Tyr Tyr Glu Ala Cys Cys Pro Lys Asn
        675                 680                 685

Thr Ala Leu Cys Ser Ala Asn Lys Arg Arg Glu Lys Leu Ala Asp Met
    690                 695                 700

Ile Ala Glu Ile Lys Phe Glu Asn Phe Ser Asp Ala Gly Asn Tyr Gln
705                 710                 715                 720

Lys Ala Asn Val Thr Ser Arg Thr Ser Glu Ala Glu Ile Lys Arg Lys
                725                 730                 735
```

```
Asn Gln Ala Ile Ile Arg Leu Tyr Leu Thr Val Met Tyr Ile Met Leu
            740                 745                 750

Lys Asn Leu Val Asn Val Asn Ala Arg Tyr Val Ile Ala Phe His Cys
        755                 760                 765

Val Glu Arg Asp Thr Lys Leu Tyr Ala Glu Ser Gly Leu Glu Val Gly
    770                 775                 780

Asn Ile Glu Lys Asn Lys Thr Asn Leu Thr Met Ala Val Met Gly Val
785                 790                 795                 800

Lys Leu Glu Asn Gly Ile Ile Lys Thr Glu Phe Asp Lys Ser Phe Ala
                805                 810                 815

Glu Asn Ala Ala Asn Arg Tyr Leu Arg Asn Ala Arg Trp Tyr Lys Leu
            820                 825                 830

Ile Leu Asp Asn Leu Lys Lys Ser Glu Arg Ala Val Val Asn Glu Phe
        835                 840                 845

Ala Asn Thr Val Cys Ala Leu Asn Ala Ile Arg Asn Ile Asn Ile Asn
    850                 855                 860

Ile Lys Glu Ile Lys Glu Val Glu Asn Tyr Phe Ala Leu Tyr His Tyr
865                 870                 875                 880

Leu Ile Gln Lys His Leu Glu Asn Arg Phe Ala Asp Lys Lys Val Glu
                885                 890                 895

Arg Asp Thr Gly Asp Phe Ile Ser Lys Leu Glu Glu His Lys Thr Tyr
            900                 905                 910

Cys Lys Asp Phe Val Lys Ala Tyr Cys Thr Pro Phe Gly Tyr Asn Leu
        915                 920                 925

Val Arg Tyr Lys Asn Leu Thr Ile Asp Gly Leu Phe Asp Lys Asn Tyr
    930                 935                 940
```

Pro Gly Lys Asp Asp Ser Asp Glu Gln Lys
945 950

<210> 3
<211> 919
<212> PRT
<213> Artificial Sequence

<220>
<223> Uncultured Ruminococcus sp. nCas1

<400> 3

Met Ala Lys Lys Asn Lys Met Lys Pro Arg Glu Leu Arg Glu Ala Gln
1 5 10 15

Lys Lys Ala Arg Gln Leu Lys Ala Ala Glu Ile Asn Asn Asn Ala Ala
20 25 30

Pro Ala Ile Ala Ala Met Pro Ala Ala Glu Val Ile Ala Pro Ala Ala
35 40 45

Glu Lys Lys Lys Ser Ser Val Lys Ala Ala Gly Met Lys Ser Ile Leu
50 55 60

Val Ser Glu Asn Lys Met Tyr Ile Thr Ser Phe Gly Lys Gly Asn Ser
65 70 75 80

Ala Val Leu Glu Tyr Glu Val Asp Asn Asn Asp Tyr Asn Gln Thr Gln
85 90 95

Leu Ser Ser Lys Asp Asn Ser Asn Ile Gln Leu Gly Gly Val Asn Glu
100 105 110

Val Asn Ile Thr Phe Ser Ser Lys His Gly Phe Glu Ser Gly Val Glu
115 120 125

Ile Asn Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser Ser Pro Val
130 135 140

Arg Gly Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys Arg Phe Phe
145 150 155 160

```
Gly Lys Thr Phe Asp Asp Asn Ile His Ile Gln Leu Ile Tyr Asn Ile
                165                 170                 175

Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn Ile Val Tyr
            180                 185                 190

Ala Leu Asn Asn Met Leu Gly Val Lys Gly Ser Glu Ser His Asp Asp
        195                 200                 205

Phe Ile Gly Tyr Leu Ser Thr Asn Asn Ile Tyr Asp Val Phe Ile Asp
    210                 215                 220

Pro Asp Asn Ser Ser Leu Ser Asp Asp Lys Lys Ala Asn Val Arg Lys
225                 230                 235                 240

Ser Leu Ser Lys Phe Asn Ala Leu Leu Lys Thr Lys Arg Leu Gly Tyr
                245                 250                 255

Phe Gly Leu Glu Glu Pro Lys Thr Lys Asp Asn Arg Val Ser Gln Ala
            260                 265                 270

Tyr Lys Lys Arg Val Tyr His Met Leu Ala Ile Val Gly Gln Ile Arg
        275                 280                 285

Gln Cys Val Phe His Asp Lys Ser Gly Ala Lys Arg Phe Asp Leu Tyr
    290                 295                 300

Ser Phe Ile Asn Asn Ile Asp Pro Glu Tyr Arg Asp Thr Leu Asp Tyr
305                 310                 315                 320

Leu Val Glu Glu Arg Leu Lys Ser Ile Asn Lys Asp Phe Ile Glu Asp
                325                 330                 335

Asn Lys Val Asn Ile Ser Leu Leu Ile Asp Met Met Lys Gly Tyr Glu
            340                 345                 350

Ala Asp Asp Ile Ile Arg Leu Tyr Tyr Asp Phe Ile Val Leu Lys Ser
        355                 360                 365
```

```
Gln Lys Asn Leu Gly Phe Ser Ile Lys Lys Leu Arg Glu Lys Met Leu
    370                 375                 380

Asp Glu Tyr Gly Phe Arg Phe Lys Asp Lys Gln Tyr Asp Ser Val Arg
385                 390                 395                 400

Ser Lys Met Tyr Lys Leu Met Asp Phe Leu Leu Phe Cys Asn Tyr Tyr
                405                 410                 415

Arg Asn Asp Ile Ala Ala Gly Glu Ser Leu Val Arg Lys Leu Arg Phe
            420                 425                 430

Ser Met Thr Asp Asp Glu Lys Glu Gly Ile Tyr Ala Asp Glu Ala Ala
        435                 440                 445

Lys Leu Trp Gly Lys Phe Arg Asn Asp Phe Glu Asn Ile Ala Asp His
    450                 455                 460

Met Asn Gly Asp Val Ile Lys Glu Leu Gly Lys Ala Asp Met Asp Phe
465                 470                 475                 480

Asp Glu Lys Ile Leu Asp Ser Glu Lys Lys Asn Ala Ser Asp Leu Leu
                485                 490                 495

Tyr Phe Ser Lys Met Ile Tyr Met Leu Thr Tyr Phe Leu Asp Gly Lys
            500                 505                 510

Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Ser Lys Phe Asp Asn Ile
        515                 520                 525

Lys Glu Phe Leu Lys Ile Met Lys Ser Ser Ala Val Asp Val Glu Cys
    530                 535                 540

Glu Leu Thr Ala Gly Tyr Lys Leu Phe Asn Asp Ser Gln Arg Ile Thr
545                 550                 555                 560

Asn Glu Leu Phe Ile Val Lys Asn Ile Ala Ser Met Arg Lys Pro Ala
                565                 570                 575
```

```
Ala Ser Ala Lys Leu Thr Met Phe Arg Asp Ala Leu Thr Ile Leu Gly
            580                 585                 590

Ile Asp Asp Lys Ile Thr Asp Asp Arg Ile Ser Gly Ile Leu Lys Leu
        595                 600                 605

Lys Glu Lys Gly Lys Gly Ile His Gly Leu Arg Asn Phe Ile Thr Asn
    610                 615                 620

Asn Val Ile Glu Ser Ser Arg Phe Val Tyr Leu Ile Lys Tyr Ala Asn
625                 630                 635                 640

Ala Gln Lys Ile Arg Glu Val Ala Lys Asn Glu Lys Val Val Met Phe
                645                 650                 655

Val Leu Gly Gly Ile Pro Asp Thr Gln Ile Glu Arg Tyr Tyr Lys Ser
            660                 665                 670

Cys Val Glu Phe Pro Asp Met Asn Ser Ser Leu Gly Val Lys Arg Ser
        675                 680                 685

Glu Leu Ala Arg Met Ile Lys Asn Ile Ser Phe Asp Asp Phe Lys Asn
    690                 695                 700

Val Lys Gln Gln Ala Lys Gly Arg Glu Asn Val Ala Lys Glu Arg Ala
705                 710                 715                 720

Lys Ala Val Ile Gly Leu Tyr Leu Thr Val Met Tyr Leu Leu Val Lys
                725                 730                 735

Asn Leu Val Asn Val Asn Ala Arg Tyr Val Ile Ala Ile His Cys Leu
            740                 745                 750

Glu Arg Asp Phe Gly Leu Tyr Lys Glu Ile Ile Pro Glu Leu Ala Ser
        755                 760                 765

Lys Asn Leu Lys Asn Asp Tyr Arg Ile Leu Ser Gln Thr Leu Cys Glu
    770                 775                 780
```

```
Leu Cys Asp Lys Ser Pro Asn Leu Phe Leu Lys Lys Asn Glu Arg Leu
785                 790                 795                 800


Arg Lys Cys Val Glu Val Asp Ile Asn Asn Ala Asp Ser Ser Met Thr
                805                 810                 815


Arg Lys Tyr Arg Asn Cys Ile Ala His Leu Thr Val Val Arg Glu Leu
            820                 825                 830


Lys Glu Tyr Ile Gly Asp Ile Cys Thr Val Asp Ser Tyr Phe Ser Ile
        835                 840                 845


Tyr His Tyr Val Met Gln Arg Cys Ile Thr Lys Arg Glu Asn Asp Thr
    850                 855                 860


Lys Gln Glu Glu Lys Ile Lys Tyr Glu Asp Asp Leu Leu Lys Asn His
865                 870                 875                 880


Gly Tyr Thr Lys Asp Phe Val Lys Ala Leu Asn Ser Pro Phe Gly Tyr
                885                 890                 895


Asn Ile Pro Arg Phe Lys Asn Leu Ser Ile Glu Gln Leu Phe Asp Arg
            900                 905                 910


Asn Glu Tyr Leu Thr Glu Lys
        915


<210>  4
<211>  919
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Uncultured Ruminococcus sp. nCas1 HEPN-/-

<400>  4

Met Ala Lys Lys Asn Lys Met Lys Pro Arg Glu Leu Arg Glu Ala Gln
1               5                   10                  15
```

```
Lys Lys Ala Arg Gln Leu Lys Ala Ala Glu Ile Asn Asn Asn Ala Ala
            20                  25                  30

Pro Ala Ile Ala Ala Met Pro Ala Ala Glu Val Ile Ala Pro Ala Ala
        35                  40                  45

Glu Lys Lys Lys Ser Ser Val Lys Ala Ala Gly Met Lys Ser Ile Leu
    50                  55                  60

Val Ser Glu Asn Lys Met Tyr Ile Thr Ser Phe Gly Lys Gly Asn Ser
65                  70                  75                  80

Ala Val Leu Glu Tyr Glu Val Asp Asn Asn Asp Tyr Asn Gln Thr Gln
                85                  90                  95

Leu Ser Ser Lys Asp Asn Ser Asn Ile Gln Leu Gly Gly Val Asn Glu
            100                 105                 110

Val Asn Ile Thr Phe Ser Ser Lys His Gly Phe Glu Ser Gly Val Glu
        115                 120                 125

Ile Asn Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser Ser Pro Val
    130                 135                 140

Arg Gly Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys Arg Phe Phe
145                 150                 155                 160

Gly Lys Thr Phe Asp Asp Asn Ile His Ile Gln Leu Ile Tyr Asn Ile
                165                 170                 175

Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn Ile Val Tyr
            180                 185                 190

Ala Leu Asn Asn Met Leu Gly Val Lys Gly Ser Glu Ser His Asp Asp
        195                 200                 205

Phe Ile Gly Tyr Leu Ser Thr Asn Asn Ile Tyr Asp Val Phe Ile Asp
    210                 215                 220
```

```
Pro Asp Asn Ser Ser Leu Ser Asp Asp Lys Lys Ala Asn Val Arg Lys
225                 230                 235                 240


Ser Leu Ser Lys Phe Asn Ala Leu Leu Lys Thr Lys Arg Leu Gly Tyr
                245                 250                 255


Phe Gly Leu Glu Glu Pro Lys Thr Lys Asp Asn Arg Val Ser Gln Ala
            260                 265                 270


Tyr Lys Lys Arg Val Tyr His Met Leu Ala Ile Val Gly Gln Ile Ala
        275                 280                 285


Gln Cys Val Phe Ala Asp Lys Ser Gly Ala Lys Arg Phe Asp Leu Tyr
    290                 295                 300


Ser Phe Ile Asn Asn Ile Asp Pro Glu Tyr Arg Asp Thr Leu Asp Tyr
305                 310                 315                 320


Leu Val Glu Glu Arg Leu Lys Ser Ile Asn Lys Asp Phe Ile Glu Asp
                325                 330                 335


Asn Lys Val Asn Ile Ser Leu Leu Ile Asp Met Met Lys Gly Tyr Glu
            340                 345                 350


Ala Asp Asp Ile Ile Arg Leu Tyr Tyr Asp Phe Ile Val Leu Lys Ser
        355                 360                 365


Gln Lys Asn Leu Gly Phe Ser Ile Lys Lys Leu Arg Glu Lys Met Leu
    370                 375                 380


Asp Glu Tyr Gly Phe Arg Phe Lys Asp Lys Gln Tyr Asp Ser Val Arg
385                 390                 395                 400


Ser Lys Met Tyr Lys Leu Met Asp Phe Leu Leu Phe Cys Asn Tyr Tyr
                405                 410                 415


Arg Asn Asp Ile Ala Ala Gly Glu Ser Leu Val Arg Lys Leu Arg Phe
            420                 425                 430
```

```
Ser Met Thr Asp Asp Glu Lys Glu Gly Ile Tyr Ala Asp Glu Ala Ala
        435                 440                 445

Lys Leu Trp Gly Lys Phe Arg Asn Asp Phe Glu Asn Ile Ala Asp His
    450                 455                 460

Met Asn Gly Asp Val Ile Lys Glu Leu Gly Lys Ala Asp Met Asp Phe
465                 470                 475                 480

Asp Glu Lys Ile Leu Asp Ser Glu Lys Lys Asn Ala Ser Asp Leu Leu
                485                 490                 495

Tyr Phe Ser Lys Met Ile Tyr Met Leu Thr Tyr Phe Leu Asp Gly Lys
            500                 505                 510

Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Ser Lys Phe Asp Asn Ile
        515                 520                 525

Lys Glu Phe Leu Lys Ile Met Lys Ser Ser Ala Val Asp Val Glu Cys
    530                 535                 540

Glu Leu Thr Ala Gly Tyr Lys Leu Phe Asn Asp Ser Gln Arg Ile Thr
545                 550                 555                 560

Asn Glu Leu Phe Ile Val Lys Asn Ile Ala Ser Met Arg Lys Pro Ala
                565                 570                 575

Ala Ser Ala Lys Leu Thr Met Phe Arg Asp Ala Leu Thr Ile Leu Gly
            580                 585                 590

Ile Asp Asp Lys Ile Thr Asp Asp Arg Ile Ser Gly Ile Leu Lys Leu
        595                 600                 605

Lys Glu Lys Gly Lys Gly Ile His Gly Leu Arg Asn Phe Ile Thr Asn
    610                 615                 620

Asn Val Ile Glu Ser Ser Arg Phe Val Tyr Leu Ile Lys Tyr Ala Asn
625                 630                 635                 640
```

```
Ala Gln Lys Ile Arg Glu Val Ala Lys Asn Glu Lys Val Val Met Phe
                645                 650                 655

Val Leu Gly Gly Ile Pro Asp Thr Gln Ile Glu Arg Tyr Tyr Lys Ser
            660                 665                 670

Cys Val Glu Phe Pro Asp Met Asn Ser Ser Leu Gly Val Lys Arg Ser
        675                 680                 685

Glu Leu Ala Arg Met Ile Lys Asn Ile Ser Phe Asp Asp Phe Lys Asn
    690                 695                 700

Val Lys Gln Gln Ala Lys Gly Arg Glu Asn Val Ala Lys Glu Arg Ala
705                 710                 715                 720

Lys Ala Val Ile Gly Leu Tyr Leu Thr Val Met Tyr Leu Leu Val Lys
                725                 730                 735

Asn Leu Val Asn Val Asn Ala Arg Tyr Val Ile Ala Ile His Cys Leu
            740                 745                 750

Glu Arg Asp Phe Gly Leu Tyr Lys Glu Ile Ile Pro Glu Leu Ala Ser
        755                 760                 765

Lys Asn Leu Lys Asn Asp Tyr Arg Ile Leu Ser Gln Thr Leu Cys Glu
    770                 775                 780

Leu Cys Asp Lys Ser Pro Asn Leu Phe Leu Lys Lys Asn Glu Arg Leu
785                 790                 795                 800

Arg Lys Cys Val Glu Val Asp Ile Asn Asn Ala Asp Ser Ser Met Thr
                805                 810                 815

Arg Lys Tyr Ala Asn Cys Ile Ala Ala Leu Thr Val Val Arg Glu Leu
            820                 825                 830

Lys Glu Tyr Ile Gly Asp Ile Cys Thr Val Asp Ser Tyr Phe Ser Ile
        835                 840                 845
```

Tyr His Tyr Val Met Gln Arg Cys Ile Thr Lys Arg Glu Asn Asp Thr
850 855 860

Lys Gln Glu Glu Lys Ile Lys Tyr Glu Asp Asp Leu Leu Lys Asn His
865 870 875 880

Gly Tyr Thr Lys Asp Phe Val Lys Ala Leu Asn Ser Pro Phe Gly Tyr
885 890 895

Asn Ile Pro Arg Phe Lys Asn Leu Ser Ile Glu Gln Leu Phe Asp Arg
900 905 910

Asn Glu Tyr Leu Thr Glu Lys
915

<210> 5
<211> 408
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig2791000549

<400> 5

Met Lys Gln Asn Asp Arg Glu Asn Asn Asn Lys Ile Lys Lys Ser Ala
1 5 10 15

Ala Lys Ala Val Gly Val Lys Ser Leu Ala Arg Leu Ser Asp Gly Ser
20 25 30

Thr Val Val Ser Ser Phe Gly Lys Gly Ala Ala Ala Glu Leu Glu Ser
35 40 45

Leu Ile Thr Gly Gly Glu Ile Arg Lys Leu Ser Asp Lys Ala Ile Leu
50 55 60

Glu Ile Thr Asp Asp Thr Gln Asn Lys Asn Ala Tyr Asn Val Lys Ser
65 70 75 80

Ser Arg Ile Pro Asn Leu Thr Ala Arg Thr Asp Lys Leu Ser Asp Lys
85 90 95

```
Ser Gly Met Asp Asp Leu Gly Phe Lys Arg Glu Leu Glu Leu Glu Val
            100                 105                 110

Phe Gly Gln Cys Phe Asp Asp Ser Ile His Ile Gln Ile Ala His Ala
        115                 120                 125

Val Phe Asp Ile Gln Lys Ser Leu Ala Ala Val Ile Pro Asn Val Leu
    130                 135                 140

Tyr Thr Leu Asn Asn Leu Asp Arg Ser Tyr Ser Thr Asp Asn Thr Ser
145                 150                 155                 160

Asp Lys Lys Asp Ile Ile Gly Asn Thr Leu Asn Tyr Gln His Ser Tyr
                165                 170                 175

Glu Ser Phe Asn Val Glu Lys Arg Gly Glu Phe Thr Glu Tyr Tyr Asn
            180                 185                 190

Ala Ala Lys Asp Arg Phe Ser Tyr Phe Pro Asp Ile Leu Cys Val Leu
        195                 200                 205

Glu Lys Val Asn Gly Lys Asp Arg Tyr Gln Pro Lys Ser Glu Lys Asp
    210                 215                 220

Ala Phe Asn Val Leu Ser Ser Val Asn Met Leu Arg Asn Ser Leu Phe
225                 230                 235                 240

His Phe Ala Pro Lys Ser Asn Asp Gly Lys Ala Arg Ile Ala Val Phe
                245                 250                 255

Lys Asn Gln Phe Asp Ser Asp Phe Ser His Ile Thr Ser Thr Val Asn
            260                 265                 270

Lys Ile Tyr Ser Ala Lys Ile Ala Gly Val Asn Glu Asn Phe Leu Asn
        275                 280                 285

Asn Glu Gly Asn Asn Leu Tyr Ile Ile Leu Lys Ala Thr Asn Trp Asp
    290                 295                 300
```

```
Ile Lys Lys Ile Val Pro Gln Leu Tyr Arg Phe Ser Val Leu Lys Ser
305                 310                 315                 320


Asp Lys Asn Met Gly Phe Asn Met Arg Lys Leu Arg Glu Phe Ala Val
                325                 330                 335


Glu Ser Lys Asn Ile Asp Leu Ser Arg Leu Asn Asp Lys Phe Leu Thr
            340                 345                 350


Asn Asn Arg Lys Lys Leu Tyr Lys Val Ile Asp Phe Ile Ile Tyr Tyr
        355                 360                 365


His Leu Asn Lys Val Leu Lys Asp Ser Phe Val Asp Asp Phe Val Ala
    370                 375                 380


Ala Leu Arg Ala Ser Gln Ser Glu Glu Glu Lys Glu Lys Leu Tyr Ala
385                 390                 395                 400


Gln Tyr Ser Glu Arg Leu Phe Ala
                405


<210>  6
<211>  295
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Gut_metagenome_contig855000317

<400>  6

Thr Asp Asn Pro Leu His Gly Ala Gly Gly Ile Ala Arg Asp His Ser
1               5                   10                  15


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


Thr Phe Glu Asp Asn Ile His Ile Gln Leu Ile Tyr Asn Ile Leu Asp
        35                  40                  45
```

```
Ile Glu Lys Ile Leu Ser Val His Ile Asn Asn Ile Val Tyr Thr Leu
    50                  55                  60

Asn Asn Leu Leu Arg Pro Leu Thr Asp Asn Asn Phe Asp Phe Ile Gly
65                  70                  75                  80

Glu Leu Ser Thr Tyr Arg Asp Tyr Glu Lys Phe Leu Glu Val Ser Glu
                85                  90                  95

Ile Glu Tyr Ile Ile Asp Glu Asn Gly Glu Glu Lys Gln Asn Lys Glu
            100                 105                 110

Tyr Thr Lys Arg Lys Leu Phe Phe Arg Leu Cys Tyr Met Pro Gln Leu
        115                 120                 125

Arg Tyr Phe Asn Ile Ile Val Glu Gly Pro His Glu Pro Val Val Gln
    130                 135                 140

Val Glu Asn Pro Glu Leu Leu Lys Lys Ile Glu Ala Ser Arg Ile Ser
145                 150                 155                 160

Asp Arg Ala Lys Thr Val Thr Phe Asn Glu Lys Glu Phe Tyr Asn Val
                165                 170                 175

Leu Ile Ala Leu Gly Thr Leu Arg Gln Met Glu Ala His Gly Asn Pro
            180                 185                 190

Leu Ile Asn Ile Tyr Gln Phe Asp Lys Lys Gln Asn Val Tyr Ser Thr
        195                 200                 205

Leu Thr Lys Lys Tyr Lys Glu Arg Val Ile Thr Leu Asp Asn Gly Phe
    210                 215                 220

Leu Ser Arg Ala Lys Lys Asn Leu Ile Ile Leu Phe Arg Ala Tyr Gly
225                 230                 235                 240

Ile Glu Asp Ala Glu Glu Lys Lys Gln Tyr Ile Arg Asn Tyr Tyr Asp
                245                 250                 255
```

Phe Thr Val Arg Lys Asn Tyr Lys Asn Gln Gly Phe Ser Ile Lys Thr
260 265 270

Leu Arg Glu His Ile Thr Thr Asp Ile Glu Glu Ala Ile Gln Leu Arg
275 280 285

Ser Thr Ala Phe Asp Ser Val
290 295

<210> 7
<211> 159
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig3389000027

<400> 7

Pro Glu Lys Ala Ala Phe Thr Val His Asn Asp Cys Glu Ala Phe Phe
1 5 10 15

Glu Ile Asp Gly Arg Val Gln Arg Ala Val Ala Asp Asn Pro Leu His
20 25 30

Lys Tyr Glu Lys Asp Asp His Gly Gln Asp Arg Ile Ser Arg Glu Lys
35 40 45

Leu Glu Met Arg Tyr Tyr Gly Arg Thr Phe Glu Asp Asn Ile His Ile
50 55 60

Gln Val Ile Tyr Asn Ile Leu Asp Ile Glu Lys Ile Met Ser Ile His
65 70 75 80

Ile Asn Asn Ile Ile Phe Thr Ile Asn Asn Leu Leu Arg Arg Glu Glu
85 90 95

Glu Asp Phe Ser Asp Leu Ile Gly Tyr Ile Thr Leu Cys Glu Lys Tyr
100 105 110

Glu Asp Leu Ser Lys Lys Gly Arg Glu Glu Ile Arg Asp Ser Phe Asp
115 120 125

Arg Leu Cys Lys Ala Pro Gln Leu Ala Tyr Tyr Asn Phe Gln Ile Glu
130 135 140

Lys Lys Gly Lys Lys Asp Pro Asn Ala Ile Ile Leu Thr Glu Lys
145 150 155

<210> 8
<211> 108
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig8061000170

<400> 8

Glu Val Thr Gly Arg Gly Gln Gly Ala Thr Ala Asp Asn Pro Leu Tyr
1 5 10 15

Lys Tyr Ser Gln Lys Lys Ala Trp Asp Arg Gln Asp Arg Ile Ser Lys
20 25 30

Asp Lys Leu Glu Leu Arg Tyr Tyr Gly Gln Thr Phe Glu Asp Asn Ile
35 40 45

His Ile Gln Thr Ile Tyr Asn Ile Leu Asp Ile Glu Lys Ile Leu Ser
50 55 60

Ile His Ile Asn Asn Ile Val Phe Thr Ile Asn Asn Leu Leu Arg Asn
65 70 75 80

Asp Asn Asp Glu Gln Asp Asp Ile Val Gly Tyr Ile Ser Leu Cys Lys
85 90 95

Lys Tyr Ala Asp Leu Ser Lys Glu Lys Asn Ala Glu
100 105

<210> 9
<211> 237
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig1509000299

<400> 9

```
Met Lys Ser Thr Phe Val Leu Gly Glu Asn Asn Leu Leu Met Thr Ser
1               5                   10                  15


Phe Gly Lys Gly Asn Lys Ala Val Val Glu Lys Gln Ile Arg Asp Gln
            20                  25                  30


Gln Ile Thr Asp Leu Ser Asp Pro Ser Ala Phe Thr Val Lys Gln Asp
        35                  40                  45


Lys Asp Leu His Phe Asn Ile Asp Gly Arg Val Arg Gly Ala Ser Thr
    50                  55                  60


Asp Asn Pro Phe His His Asn Pro Glu Lys Thr Ala Gln Pro Pro Val
65                  70                  75                  80


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


Gly Glu Asn Phe Thr Asp Asn Ile His Ile Gln Cys Ile Tyr Ser Ile
            100                 105                 110


Leu Asp Val Glu Lys Ile Leu Val Ser His Ile Asn Asn Ile Val Tyr
        115                 120                 125


Met Leu Asn Asn Phe Leu Arg Glu Glu Gly Glu Thr Leu Thr Asp Leu
    130                 135                 140


Ile Gly Tyr Leu Ser Ile Phe Lys Ser Tyr Glu Gly Leu Met Asn Ser
145                 150                 155                 160


Thr Arg Asn Asp Ala Ala Glu Met Arg Arg His Phe Phe Ser Leu Cys
                165                 170                 175


Gln Ala Lys Gln Leu Gly Tyr Leu Asn Leu Glu Val Leu Lys Pro Gln
            180                 185                 190
```

```
Pro Asp Lys Pro Lys Gln Gly Lys Lys Lys Pro Ala Asp Pro Asn Thr
        195                 200                 205


Ile Lys Leu Thr Glu Glu Glu Phe Phe Asn Val Leu Leu Ala Leu Ser
    210                 215                 220


Asn Met Arg Gln Met Leu Val His Gly Asn Ala Lys Gln
225                 230                 235


<210>  10
<211>  214
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Gut_metagenome_contig9549000591

<400>  10

Lys Ala Lys Leu Glu Lys Arg Tyr Phe Gly Gln Thr Phe Glu Asp Asn
1               5                   10                  15


Ile His Ile Gln Val Ile Tyr Asn Met Leu Asp Ile Glu Lys Ile Leu
            20                  25                  30


Thr Ile His Val Asn Asn Ile Val Phe Thr Ile Asn Asn Phe Leu Arg
        35                  40                  45


Glu Glu Gly Glu Asp Leu Thr Asp Ile Val Gly Tyr Ile Thr Lys Tyr
    50                  55                  60


Lys Ser Leu Ala Lys Phe Lys Glu Asn Lys Asn Ala Arg Glu Tyr Tyr
65                  70                  75                  80


Glu Leu Phe Cys Lys Leu Cys Asp Ala Lys Gln Leu Gly Tyr Leu Asn
                85                  90                  95


Leu Glu Tyr Asn Ser Val Arg Ser Lys Leu Tyr Lys Leu Phe Asp Phe
            100                 105                 110
```

```
Ile Leu Val His His Tyr Lys Asn Asp Ser Ser Pro Glu Ser Thr Ala
        115                 120                 125


Leu Lys Asn Leu Leu Arg Thr Cys Leu Thr Asp Glu Glu Lys Glu Lys
    130                 135                 140


Val Tyr Val Gln Thr Ala Lys Glu Leu Thr Ala Leu Tyr Ser Tyr His
145                 150                 155                 160


Phe Lys Glu Ala Ser Val Ser Val Asn Ala Arg Asn Ile Lys Glu Tyr
                165                 170                 175


Gln Glu Lys Ala Gln Val Tyr His Ile Ser Ile Ala Asp Lys Ile Ser
            180                 185                 190


Ser Thr Gly Ile Gly Leu Phe Ser Gln Leu Ile Tyr Met Leu Thr Leu
        195                 200                 205


Leu Leu Asp Gly Lys Glu
    210


<210>  11
<211>  422
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Gut_metagenome_contig71000500

<400>  11

Met Lys Lys Gln Lys Ser Lys Lys Thr Val Ser Lys Thr Ser Gly Leu
1               5                   10                  15


Lys Glu Ala Leu Ser Val Gln Gly Thr Val Ile Met Thr Ser Phe Gly
            20                  25                  30


Lys Gly Asn Met Ala Asn Leu Ser Tyr Lys Ile Pro Ser Ser Gln Lys
        35                  40                  45


Pro Gln Asn Leu Asn Ser Ser Ala Gly Leu Lys Asn Val Glu Val Ser
    50                  55                  60
```

```
Gly Lys Lys Ile Lys Phe Gln Gly Arg His Pro Lys Ile Ala Thr Thr
65                  70                  75                  80

Asp Asn Pro Leu Phe Lys Pro Gln Pro Gly Met Asp Leu Leu Cys Leu
                85                  90                  95

Lys Asp Lys Leu Glu Met His Tyr Phe Gly Lys Thr Phe Asp Asp Asn
            100                 105                 110

Ile His Ile Gln Leu Ile Tyr Gln Ile Leu Asp Ile Glu Lys Ile Leu
        115                 120                 125

Ala Val His Val Asn Asn Ile Val Phe Thr Leu Asp Asn Val Leu His
    130                 135                 140

Pro Gln Lys Glu Glu Leu Thr Glu Asp Phe Ile Gly Ala Gly Gly Trp
145                 150                 155                 160

Arg Ile Asn Leu Asp Tyr Gln Thr Leu Arg Gly Gln Thr Asn Lys Tyr
                165                 170                 175

Asp Arg Phe Lys Asn Tyr Ile Lys Arg Lys Glu Leu Leu Tyr Phe Gly
            180                 185                 190

Glu Ala Phe Tyr His Glu Asn Glu Arg Arg Tyr Glu Glu Asp Ile Phe
        195                 200                 205

Ala Ile Leu Thr Leu Leu Ser Ala Leu Arg Gln Phe Cys Phe His Ser
    210                 215                 220

Asp Leu Ser Ser Asp Glu Ser Asp His Val Asn Ser Phe Trp Leu Tyr
225                 230                 235                 240

Gln Leu Glu Asp Gln Leu Ser Asp Glu Phe Lys Glu Thr Leu Ser Ile
                245                 250                 255

Leu Trp Glu Glu Val Thr Glu Arg Ile Asp Ser Glu Phe Leu Lys Thr
            260                 265                 270
```

```
Asn Thr Val Asn Leu His Ile Leu Cys His Val Phe Pro Lys Glu Ser
        275                 280                 285

Lys Glu Thr Ile Val Arg Ala Tyr Tyr Glu Phe Leu Ile Lys Lys Ser
    290                 295                 300

Phe Lys Asn Met Gly Phe Ser Ile Lys Lys Leu Arg Glu Ile Met Leu
305                 310                 315                 320

Glu Gln Ser Asp Leu Lys Ser Phe Lys Glu Asp Lys Tyr Asn Ser Val
                325                 330                 335

Arg Ala Lys Leu Tyr Lys Leu Phe Asp Phe Ile Ile Thr Tyr Tyr Tyr
            340                 345                 350

Asp His His Ala Phe Glu Lys Glu Ala Leu Val Ser Ser Leu Arg Ser
        355                 360                 365

Ser Leu Thr Glu Glu Asn Lys Glu Glu Ile Tyr Ile Lys Thr Ala Arg
    370                 375                 380

Thr Leu Ala Ser Ala Leu Gly Ala Asp Phe Lys Lys Ala Ala Ala Asp
385                 390                 395                 400

Val Asn Ala Lys Asn Ile Arg Asp Tyr Gln Lys Lys Ala Asn Asp Tyr
                405                 410                 415

Arg Ile Ser Phe Glu Asp
            420


<210>  12
<211>  49
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Artificial sequence

<400>  12
```

Ile Lys Ile Gly Asn Thr Gly Ile Gly Tyr Phe Ser Glu Leu Ile Tyr
1 5 10 15

Met Leu Thr Leu Leu Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr
20 25 30

Thr Leu Ile Asn Lys Phe Asp Asn Ile Ile Ser Phe Ile Asp Ile Leu
35 40 45

Lys

<210> 13
<211> 148
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig3915000357

<400> 13

Met Lys Asp Lys Leu Glu Gln Trp Tyr Phe Asp Lys Thr Phe Asp Asp
1 5 10 15

Asn Ile His Ile Gln Leu Ile Tyr Gln Ile Gln Asp Ile Glu Lys Ile
20 25 30

Leu Ala Ser His Val Asn Asn Ile Val Phe Ala Leu Asp Asn Ile Leu
35 40 45

His Ala Thr Glu Ser Glu Thr Gly Asp Phe Leu Gly Met Gly Tyr Met
50 55 60

Arg Ala Thr Asn Thr Phe Glu Glu Tyr Arg Asp Ser Ala Ile Ser Asn
65 70 75 80

Val Lys Lys Ser Tyr Glu Thr Phe Glu Thr Tyr Ile Lys Lys Lys Glu
85 90 95

Leu Leu Tyr Phe Gly Glu Ala Phe Tyr His Gly Asn Thr Cys Arg Lys
100 105 110

```
Pro Glu Asp Ile Tyr His Ile Leu Ala Met Val Ser Ser Leu Arg Gln
        115                 120                 125


Phe Cys Phe His Ser Thr Phe Asp Asp Lys Ser Lys Ser Ile Lys Gly
    130                 135                 140


Ser Trp Leu Tyr
145


<210>  14
<211>  137
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Gut_metagenome_contig4719000173

<400>  14

Lys Asn Val Ile Thr Ser Arg Ser Phe Tyr Tyr Leu Val Arg Tyr Ala
1               5                   10                  15


Lys Pro Thr Ala Val Arg Lys Leu Met Ser Asn Pro Lys Ile Val Arg
            20                  25                  30


Tyr Val Leu Lys Arg Leu Pro Glu Asn Gln Val Ala Ser Tyr Tyr Ser
        35                  40                  45


Ala Ile Trp Thr Gln Ser Glu Ser Asn Ser Asn Glu Thr Val Lys Leu
    50                  55                  60


Ile Glu Met Ile Asp Arg Leu Thr Thr Glu Ile Ala Gly Phe Ser Phe
65                  70                  75                  80


Ala Val Leu Lys Asp Lys Lys Asp Ser Ile Val Ser Ala Ser Arg Glu
                85                  90                  95


Ser Arg Ala Val Asn Leu Glu Val Glu Arg Leu Lys Lys Leu Thr Thr
            100                 105                 110
```

Leu Tyr Met Ser Ile Ala Tyr Ile Ala Val Lys Ser Leu Val Lys Val
115 120 125

Asn Ala Arg Tyr Phe Ile Ala Tyr Ser
130 135

<210> 15
<211> 95
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig6929000468

<400> 15

Met Lys Thr Ile Lys Lys Ser Ser Lys Ala Tyr Gly Leu Lys Ser Thr
1 5 10 15

Phe Ala Tyr Asn Asp Asp Val Tyr Met Thr Ser Phe Gly Lys Gly Asn
20 25 30

Asp Phe Val Met Glu Lys Leu Ile Tyr Pro Asn Asp Asn Val Glu Asp
35 40 45

Leu Ser Asp Asn Pro Lys Phe Asp Val Glu Val Glu Asn Lys Asp Ile
50 55 60

Thr Leu Lys Tyr Lys Lys Ser Glu Ile Asn Val Asn Asn Pro Ala Lys
65 70 75 80

Ser Gly Ile Lys Gln Asp Tyr Leu Met Leu Lys Asp Lys Leu Glu
85 90 95

<210> 16
<211> 114
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig7367000486

<400> 16

```
Asn Arg Leu Lys Thr Lys Ala Asp Leu Glu Arg Leu Phe Ser Asn Arg
1               5                   10                  15


Arg Ile Gly Tyr Phe Gly Gln Ala Phe Tyr Glu Ser Asp Lys Gln Gln
            20                  25                  30


Arg Asn Glu Met Glu Ile Tyr Asp Asn Leu Ala Leu Ile Gly Ser Leu
        35                  40                  45


Arg Gln Trp Ala Val His Gly Asn Lys Thr Asn Asp Pro Ala Trp Leu
    50                  55                  60


Tyr Lys Leu Glu Ser Leu Pro Asp Glu Phe Lys Ser Val Leu Asp Lys
65                  70                  75                  80


Val Tyr Ala Glu Thr Val Arg Lys Val Asn Tyr Lys Phe Ile Lys Thr
                85                  90                  95


Asn Ala Val Asn Leu Gln Ile Leu Gln Glu Leu Tyr Pro Lys Glu Ser
            100                 105                 110


Phe Ala



<210>  17
<211>  107
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Gut_metagenome_contig7930000403

<400>  17

Tyr Ala Lys Ile Ile Ile Gly Glu Ile Thr Ser Lys Tyr Asn Lys Phe
1               5                   10                  15


Asp Asp Ile Ile Asn Leu Val Tyr Arg Glu Ile Arg Asn Arg Asn Asn
            20                  25                  30


Thr Ile Asn Phe Lys Asn Glu Ile Lys Asn Ile Ser Thr Ile Thr Asn
        35                  40                  45
```

Tyr Ser Tyr Phe Thr Lys Phe Ile Tyr Phe Ile Thr Lys Phe Leu Asp
    50   55   60

Gly Lys Glu Ile Asn Glu Leu Leu Thr Ser Leu Ile Asn Lys Phe Asp
65   70   75   80

Asn Ile Gln Ala Phe Asn Asp Ile Leu Leu Glu Lys Leu Gln Asp Lys
                85   90   95

Asn Glu Ser Lys Phe Ala Ile Phe Asn Gly Lys
            100   105

<210> 18
<211> 282
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig993000527

<400> 18

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

Arg Val Glu Phe Gly Asp Lys Tyr Asp Met Phe Asn Lys Ala Gly Glu
            20   25   30

Ile Cys Glu Ser Leu Arg Ser Ile Asn Ser Phe Ala Arg Met Glu Ser
        35   40   45

Pro Ala Pro Ala Ala Lys Arg Thr Met Phe Val Asp Ala Ile Arg Met
    50   55   60

Phe Gly Phe Ala Asp Val Ser Asp Asn Asp Ala Glu Lys Phe Tyr Lys
65   70   75   80

Glu Asn Ile Glu Pro Phe Ile Tyr Asp Arg Asn Gly Asp Ile Glu Asp
                85   90   95

```
Glu Trp Lys Lys Gly Lys Leu Val His Lys Lys Asp Glu Lys Lys Lys
            100                 105                 110

Gly Val Tyr Asn Phe Ile Ala Thr Asn Val Ile Glu Ser Asn Arg Phe
        115                 120                 125

Gln Tyr Leu Val Arg Tyr Ala Asn Thr Glu Lys Ile Ala Ala Leu Ala
    130                 135                 140

Arg Asn Glu Thr Ile Val Arg Phe Val Ile Lys Lys Leu Pro Ala Ala
145                 150                 155                 160

Gln Ile Val Arg Tyr Tyr Ile Ala Cys Asn Arg Ile Arg Glu Asn Ile
                165                 170                 175

Asn Ser Asn Val Leu Thr Ser Asp Asp Met Ala Glu Ser Leu Val Lys
            180                 185                 190

Leu Val Thr Gly Met Asp Phe Gly Lys Ile Lys Gly Val Arg Asn Asp
        195                 200                 205

Asp Arg Ser Gly Asp Lys Glu Met Lys Lys Ala Ile Val Thr Leu Tyr
    210                 215                 220

Leu Thr Ile Leu Tyr His Leu Thr Lys Asn Leu Val Tyr Val Asn Ser
225                 230                 235                 240

Arg Tyr Phe Met Ala Phe His Cys Leu Glu Arg Asp Met Asn Leu Phe
                245                 250                 255

Gly Glu Gly Gln Lys Gln Ala Leu Arg Asn Thr Gly Asp Tyr Pro Ala
            260                 265                 270

Asp Asn Ala Asp Leu Arg Arg Ser Asp Ile
        275                 280
```

<210> 19
<211> 118
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig6552000639

<400> 19

Ala Val Asp Leu Asn Pro Asp Asp Pro Ala Phe Asn Val Asn Glu Val
1   5   10   15

Thr Glu Glu Arg Phe Ile Leu Asp His Gly Arg Ile Lys Gly Cys Ile
            20   25   30

Ala Thr Ala Asp Asn Pro Leu His Ala Ser Ser Ala Pro Gln Asp Met
        35   40   45

Leu Leu Leu Lys Ser Val Leu Glu Glu Arg Phe Phe Gly Met Ser Phe
    50   55   60

Pro Glu Asp Asn Val His Ile Gln Leu Ile Tyr Asn Ile Leu Asp Ile
65   70   75   80

Glu Lys Ile Leu Thr Val Tyr Ser Thr Asn Ala Val Tyr Ala Ile Asn
                85   90   95

Asn Leu Phe Arg Ala Asp Asp Gln Glu Asp Asp Glu Phe Ile Gly Asp
            100   105   110

Leu Ser Val Asp Thr Ser
        115

<210> 20
<211> 87
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig11932000246

<400> 20

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

```
Lys Lys Ile Met Phe Asp Glu Ala Leu Ile Val Leu Gly Ile Asp Asp
            20                  25                  30


Ala Glu Lys Glu Arg Glu Leu Asp Gln Met Leu Asp Lys Asp Ala Arg
        35                  40                  45


Lys Gln Asn Pro Gln Arg Ala Gly Leu Arg Asn Phe Ile Ala Asn Asn
    50                  55                  60


Val Ile Glu Ser Asp Arg Phe Lys Tyr Leu Val Arg Tyr Gly Asn Val
65                  70                  75                  80


Lys Arg Ile Lys Gly Ile Ala
                85


<210>  21
<211>  85
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Gut_metagenome_contig12963000286

<400>  21

Glu Lys Glu Thr Glu Leu Asn Leu Met Leu Asp Lys Asp Ala Arg Lys
1               5                   10                  15


Gln Asn Arg Gln Lys Ala Gly Leu Arg Asn Phe Ile Ala Asn Asn Val
            20                  25                  30


Ile Glu Ser Glu Arg Phe Lys Tyr Leu Val Arg Tyr Gly Asn Val Lys
        35                  40                  45


Arg Ile Lys Gly Leu Ser Thr Asn Ser Gln Val Met Gly Phe Val Leu
    50                  55                  60


Lys Asp Ile Ala Asp Ala Gln Ile Val Arg Tyr Tyr Arg Ser Ile Tyr
65                  70                  75                  80


Gly Asp Asp Asn Cys
                85
```

<210> 22
<211> 155
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig2952000470

<400> 22

```
Ile Arg Leu Pro Asn Gly Thr Ile Met Asp Glu Cys Asp Val Ser Lys
1               5                  10                  15


Ala Thr Thr Val Lys Asn Arg Tyr Leu His Asn Lys Lys Trp Tyr Ser
            20                  25                  30


Leu Ala Phe Thr Asn Tyr Lys Asn Ser Asp Ser Asp Val Ile Val Asn
        35                  40                  45


Phe Arg Asn Ser Val Ala His Leu Asn Ala Ile Arg Asn Ile Asp Glu
    50                  55                  60


Tyr Ile Asp Gly Ile Gly Glu Ile Asn Ser His Thr Gly Tyr Phe Glu
65                  70                  75                  80


Ile Tyr His Tyr Ile Ile Gln Lys Arg Met Gln Ala Val Thr Thr Asp
                85                  90                  95


Lys Ser Lys Leu Gln Pro Lys Thr Glu Glu Tyr Phe Glu Asn Leu Ser
            100                 105                 110


Lys Tyr His Thr Tyr Cys Lys Asp Phe Val Lys Ala Tyr Cys Val Pro
        115                 120                 125


Met Ala Tyr Asn Leu Pro Arg Tyr Lys Asn Leu Thr Ile Asp Gly Gln
    130                 135                 140


Phe Asp Arg Asn Ile His Thr Glu Thr Lys Glu
145                 150                 155
```

<210> 23
<211> 406
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig451000394

<400> 23

```
Leu Gln Glu Leu Tyr Pro Lys Glu Ser Phe Ala Ser Ile Val Lys Glu
1               5                  10                  15


Tyr Tyr Gly Phe Leu Val Thr Lys Lys Tyr Lys Asn Ile Gly Phe Ser
            20                  25                  30


Ile Lys His Leu Arg Glu Ile Leu Leu Asp Ser Thr Asp Tyr Arg Asp
        35                  40                  45


Lys Glu Tyr Asp Ser Val Arg His Lys Leu Tyr Gln Leu Leu Asp Phe
    50                  55                  60


Cys Ile Tyr His Gly Tyr Leu Tyr Glu Asp Ala Glu Phe Ala Ala Asp
65                  70                  75                  80


Leu Val Asn Gln Leu Arg Ala Ser Leu Ser Glu Leu Asp Lys Glu Lys
                85                  90                  95


Ile Tyr Asn Ile Glu Ala Ala Arg Leu Ala Asp Lys Leu Gly Ser Ser
            100                 105                 110


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


Leu Lys Lys Asn Ala Glu Asn Leu Glu Lys Val Lys Leu Glu Asp Ile
    130                 135                 140


Gln Ile Ser Ser Gly Lys Thr Val Ser Tyr Phe Thr Lys Leu Met Tyr
145                 150                 155                 160


Leu Leu Thr Leu Phe Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr
                165                 170                 175
```

```
Thr Leu Ile Asn Lys Phe Asp Asn Ile Arg Val Phe Asn Lys Thr Leu
            180                 185                 190

Thr Gln Leu Asn Gln Ser Thr Glu Phe Val Glu Asp Tyr Ser Phe Phe
        195                 200                 205

Gly Ser Ser Gly Thr Val Phe Glu Glu Leu Thr Val Leu Asn Ser Phe
    210                 215                 220

Ala Arg Met Ser Arg Ile Asp Pro Ser Ala Arg Lys Asp Met Tyr Arg
225                 230                 235                 240

Asp Ala Ile Glu Ile Leu Gly Ile Ser Met Glu Met Ser His Glu Lys
                245                 250                 255

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

Pro Ile Ile Asn Pro Asp Gly Thr Lys Lys Ser Gly Met Arg Asn Phe
        275                 280                 285

Ile Ala Thr Asn Ile Ile Ser Ser Asn Arg Phe Lys Tyr Leu Val Arg
    290                 295                 300

Tyr Gly Asn Ala Thr Lys Ile Arg Ala Leu Ala Asn Cys Glu Asn Ala
305                 310                 315                 320

Val Lys Phe Val Leu Ser Thr Ile Pro Asp Glu Gln Ile Val Arg Tyr
                325                 330                 335

Phe Asp Ser Cys Arg Glu Pro Thr Asp His Ile Glu Asn Asn Ile Glu
            340                 345                 350

Arg Lys Arg Glu Tyr Leu Ala Ser Ile Ile Lys Glu Met Ser Phe Glu
        355                 360                 365

Arg Phe Arg Asp Ala Glu Lys Leu Gln Arg Tyr His Val Asp Arg Pro
    370                 375                 380
```

```
Thr Asn Gly Ser Ser Val Glu Gly Glu Gln Lys Gln Arg Tyr Gln Ala
385                 390                 395                 400


Val Ile Arg Leu Tyr Leu
                405


<210>  24
<211>  954
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Eubacterium_siraeum_DSM_15702

<400>  24

Met Gly Lys Lys Ile His Ala Arg Asp Leu Arg Glu Gln Arg Lys Thr
1               5                   10                  15


Asp Arg Thr Glu Lys Phe Ala Asp Gln Asn Lys Lys Arg Glu Ala Glu
            20                  25                  30


Arg Ala Val Pro Lys Lys Asp Ala Ala Val Ser Val Lys Ser Val Ser
        35                  40                  45


Ser Val Ser Ser Lys Lys Asp Asn Val Thr Lys Ser Met Ala Lys Ala
    50                  55                  60


Ala Gly Val Lys Ser Val Phe Ala Val Gly Asn Thr Val Tyr Met Thr
65                  70                  75                  80


Ser Phe Gly Arg Gly Asn Asp Ala Val Leu Glu Gln Lys Ile Val Asp
                85                  90                  95


Thr Ser His Glu Pro Leu Asn Ile Asp Asp Pro Ala Tyr Gln Leu Asn
            100                 105                 110


Val Val Thr Met Asn Gly Tyr Ser Val Thr Gly His Arg Gly Glu Thr
        115                 120                 125
```

```
Val Ser Ala Val Thr Asp Asn Pro Leu Arg Arg Phe Asn Gly Arg Lys
    130                 135                 140

Lys Asp Glu Pro Glu Gln Ser Val Pro Thr Asp Met Leu Cys Leu Lys
145                 150                 155                 160

Pro Thr Leu Glu Lys Lys Phe Phe Gly Lys Glu Phe Asp Asp Asn Ile
                165                 170                 175

His Ile Gln Leu Ile Tyr Asn Ile Leu Asp Ile Glu Lys Ile Leu Ala
            180                 185                 190

Val Tyr Ser Thr Asn Ala Ile Tyr Ala Leu Asn Asn Met Ser Ala Asp
        195                 200                 205

Glu Asn Ile Glu Asn Ser Asp Phe Phe Met Lys Arg Thr Thr Asp Glu
    210                 215                 220

Thr Phe Asp Asp Phe Glu Lys Lys Lys Glu Ser Thr Asn Ser Arg Glu
225                 230                 235                 240

Lys Ala Asp Phe Asp Ala Phe Glu Lys Phe Ile Gly Asn Tyr Arg Leu
                245                 250                 255

Ala Tyr Phe Ala Asp Ala Phe Tyr Val Asn Lys Lys Asn Pro Lys Gly
            260                 265                 270

Lys Ala Lys Asn Val Leu Arg Glu Asp Lys Glu Leu Tyr Ser Val Leu
        275                 280                 285

Thr Leu Ile Gly Lys Leu Arg His Trp Cys Val His Ser Glu Glu Gly
    290                 295                 300

Arg Ala Glu Phe Trp Leu Tyr Lys Leu Asp Glu Leu Lys Asp Asp Phe
305                 310                 315                 320

Lys Asn Val Leu Asp Val Val Tyr Asn Arg Pro Val Glu Glu Ile Asn
                325                 330                 335
```

```
Asn Arg Phe Ile Glu Asn Asn Lys Val Asn Ile Gln Ile Leu Gly Ser
            340                 345                 350

Val Tyr Lys Asn Thr Asp Ile Ala Glu Leu Val Arg Ser Tyr Tyr Glu
        355                 360                 365

Phe Leu Ile Thr Lys Lys Tyr Lys Asn Met Gly Phe Ser Ile Lys Lys
    370                 375                 380

Leu Arg Glu Ser Met Leu Glu Gly Lys Gly Tyr Ala Asp Lys Glu Tyr
385                 390                 395                 400

Asp Ser Val Arg Asn Lys Leu Tyr Gln Met Thr Asp Phe Ile Leu Tyr
                405                 410                 415

Thr Gly Tyr Ile Asn Glu Asp Ser Asp Arg Ala Asp Asp Leu Val Asn
            420                 425                 430

Thr Leu Arg Ser Ser Leu Lys Glu Asp Asp Lys Thr Thr Val Tyr Cys
        435                 440                 445

Lys Glu Ala Asp Tyr Leu Trp Lys Lys Tyr Arg Glu Ser Ile Arg Glu
    450                 455                 460

Val Ala Asp Ala Leu Asp Gly Asp Asn Ile Lys Lys Leu Ser Lys Ser
465                 470                 475                 480

Asn Ile Glu Ile Gln Glu Asp Lys Leu Arg Lys Cys Phe Ile Ser Tyr
                485                 490                 495

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

Phe Leu Ser Gly Lys Glu Ile Asn Asp Leu Val Thr Thr Leu Ile Asn
        515                 520                 525

Lys Phe Asp Asn Ile Arg Ser Phe Leu Glu Ile Met Asp Glu Leu Gly
    530                 535                 540
```

```
Leu Asp Arg Thr Phe Thr Ala Glu Tyr Ser Phe Phe Glu Gly Ser Thr
545                 550                 555                 560

Lys Tyr Leu Ala Glu Leu Val Glu Leu Asn Ser Phe Val Lys Ser Cys
                565                 570                 575

Ser Phe Asp Ile Asn Ala Lys Arg Thr Met Tyr Arg Asp Ala Leu Asp
            580                 585                 590

Ile Leu Gly Ile Glu Ser Asp Lys Thr Glu Glu Asp Ile Glu Lys Met
        595                 600                 605

Ile Asp Asn Ile Leu Gln Ile Asp Ala Asn Gly Asp Lys Lys Leu Lys
    610                 615                 620

Lys Asn Asn Gly Leu Arg Asn Phe Ile Ala Ser Asn Val Ile Asp Ser
625                 630                 635                 640

Asn Arg Phe Lys Tyr Leu Val Arg Tyr Gly Asn Pro Lys Lys Ile Arg
                645                 650                 655

Glu Thr Ala Lys Cys Lys Pro Ala Val Arg Phe Val Leu Asn Glu Ile
            660                 665                 670

Pro Asp Ala Gln Ile Glu Arg Tyr Tyr Glu Ala Cys Cys Pro Lys Asn
        675                 680                 685

Thr Ala Leu Cys Ser Ala Asn Lys Arg Arg Glu Lys Leu Ala Asp Met
    690                 695                 700

Ile Ala Glu Ile Lys Phe Glu Asn Phe Ser Asp Ala Gly Asn Tyr Gln
705                 710                 715                 720

Lys Ala Asn Val Thr Ser Arg Thr Ser Glu Ala Glu Ile Lys Arg Lys
                725                 730                 735

Asn Gln Ala Ile Ile Arg Leu Tyr Leu Thr Val Met Tyr Ile Met Leu
            740                 745                 750
```

```
Lys Asn Leu Val Asn Val Asn Ala Arg Tyr Val Ile Ala Phe His Cys
        755                 760                 765

Val Glu Arg Asp Thr Lys Leu Tyr Ala Glu Ser Gly Leu Glu Val Gly
    770                 775                 780

Asn Ile Glu Lys Asn Lys Thr Asn Leu Thr Met Ala Val Met Gly Val
785                 790                 795                 800

Lys Leu Glu Asn Gly Ile Ile Lys Thr Glu Phe Asp Lys Ser Phe Ala
                805                 810                 815

Glu Asn Ala Ala Asn Arg Tyr Leu Arg Asn Ala Arg Trp Tyr Lys Leu
            820                 825                 830

Ile Leu Asp Asn Leu Lys Lys Ser Glu Arg Ala Val Val Asn Glu Phe
        835                 840                 845

Arg Asn Thr Val Cys His Leu Asn Ala Ile Arg Asn Ile Asn Ile Asn
    850                 855                 860

Ile Lys Glu Ile Lys Glu Val Glu Asn Tyr Phe Ala Leu Tyr His Tyr
865                 870                 875                 880

Leu Ile Gln Lys His Leu Glu Asn Arg Phe Ala Asp Lys Lys Val Glu
                885                 890                 895

Arg Asp Thr Gly Asp Phe Ile Ser Lys Leu Glu Glu His Lys Thr Tyr
            900                 905                 910

Cys Lys Asp Phe Val Lys Ala Tyr Cys Thr Pro Phe Gly Tyr Asn Leu
        915                 920                 925

Val Arg Tyr Lys Asn Leu Thr Ile Asp Gly Leu Phe Asp Lys Asn Tyr
    930                 935                 940

Pro Gly Lys Asp Asp Ser Asp Glu Gln Lys
945                 950
```

<210> 25
<211> 173
<212> PRT
<213> Artificial Sequence

<220>
<223> gut_metagenome_P19E0k2120140920,_c369000003

<400> 25

```
Asp Gly Lys Glu Ile Asn Asp Leu Val Thr Thr Leu Ile Asn Lys Thr
1               5                  10                  15


Glu Asn Ile Ala Ser Phe Asn Glu Val Leu Lys Gln Leu Gly Tyr Glu
            20                  25                  30


Leu Asp Tyr Ser Glu Glu Tyr Ser Val Phe Ala Ala Ser Asp Glu Met
        35                  40                  45


Ala Ser Gln Leu Arg Ala Leu Asn Ser Phe Ala Arg Met Glu Lys Pro
    50                  55                  60


Asp Ala Ser Ala Lys Arg Ile Met Phe Val Glu Ala Ala Arg Met Leu
65                  70                  75                  80


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


Leu Ala Lys Lys Ser Lys Lys Glu Phe Asp Ala Ala Thr Asp Asn Glu
            100                 105                 110


Ala Ile Arg Ala Thr Asp Lys Lys Lys Lys Gly Phe Arg Asn Phe Ile
        115                 120                 125


Ile Asn Asn Val Ile Glu Ser Asp Arg Phe Lys Tyr Leu Ile Arg Tyr
    130                 135                 140


Cys Asn Val Glu Ser Val Gln Ala Phe Ser Lys Asn Arg Lys Leu Val
145                 150                 155                 160


Lys Phe Val Leu Lys Asp Ile Pro Glu Ser Gln Ile Leu
                165                 170
```

<210> 26
<211> 111
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig7593000362

<400> 26

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

Ala Ala Lys Val Leu Gly Tyr Ser Glu Pro Glu Glu Asp Leu Glu Thr
20 25 30

Tyr Ile Glu Gly Ile Leu Asp Lys Lys Lys Asn Val Asp Ala Ser Gly
35 40 45

Lys Lys Asn Met Asn Phe Arg Asn Phe Ile Ala Asn Asn Val Ile Glu
50 55 60

Ser Arg Arg Phe Lys Tyr Leu Val Arg Tyr Cys Asn Pro Glu Lys Ile
65 70 75 80

Arg Ala Met Ala Glu Asn Arg Lys Val Val Ser Phe Val Leu Lys Gly
85 90 95

Ile Pro Asp Ala Gln Ile Arg Ala Tyr Tyr Asn Ser Cys Asn Gly
100 105 110

<210> 27
<211> 86
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig12619000055

<400> 27

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

```
Leu Leu Asp Val Leu Lys Ile Lys Gly Gln Glu Lys Glu Ile Glu Ile
            20                  25                  30


Ala Gln Asn Tyr Tyr Glu Phe Ile Val Lys Lys Gly Tyr Lys Asn Glu
        35                  40                  45


Gly Phe Ser Ile Lys Lys Leu Arg Gln Lys Leu Thr Ser Leu Ser Glu
    50                  55                  60


Ser Glu Lys Glu Pro Leu Asn Ser Met Arg Tyr Asp Ser Val Arg Pro
65                  70                  75                  80


Lys Leu Asn His Leu Leu
                85


<210>  28
<211>  236
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Gut_metagenome_contig1405000151

<400>  28

Lys Glu Leu Phe Gln Lys Leu Cys Lys Ala Pro Gln Leu Ala Tyr Phe
1               5                   10                  15


Asn Leu Gln Ile Glu Thr Glu Gly Lys Arg Lys Ala Asp Pro Asn Ala
            20                  25                  30


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


Tyr Met Arg Gln Met Leu Met His Gly Ala Pro Asn Gln Asn Ile Tyr
    50                  55                  60


Lys Met Asp Pro Leu Lys Asp Lys Thr Phe Lys Asn Cys Leu Asp Cys
65                  70                  75                  80
```

```
Leu Asn Arg Leu Tyr Asn Glu Arg Val Met Ala Leu Asn Asn Asn Phe
                85                  90                  95

Leu Asp Leu Ala Lys Lys Asn Leu Ser Ile Leu Phe Asp Val Phe Ala
            100                 105                 110

Val Lys Thr Gln Glu Glu Lys Ala Glu Tyr Val Arg Asp Tyr Tyr Asp
        115                 120                 125

Phe Val Val Arg Lys Gln Phe Lys Asn Gln Gly Phe Ser Ile Lys Cys
    130                 135                 140

Leu Arg Glu His Met Thr Asp Asp Ile Glu Glu Ala Gly Thr Leu Arg
145                 150                 155                 160

Asp Gln Lys Tyr Asp Ser Val Arg Gly Lys Leu Tyr Pro Phe Ile Asn
                165                 170                 175

Phe Ala Ile Phe Arg Tyr Tyr Asn Ser His Ser Ser Glu Ala Glu Ala
            180                 185                 190

Leu Val Glu Arg Leu Arg Ala Ser Met Asn Glu Glu Glu Lys Asp Lys
        195                 200                 205

Val Tyr Arg His Glu Ala Val Ile Leu Trp Ser Lys Ile Arg Asn Leu
    210                 215                 220

Val Leu Lys His Ile Leu Pro Glu Met Glu Gly Ser
225                 230                 235


<210>  29
<211>  103
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Chicken_gut_metagenome_c298474

<400>  29

Ser Asp His Asp Lys Lys Ala Leu Val Gln Glu Tyr Tyr Glu Phe Val
1               5                   10                  15
```

```
Val Leu Lys Asn Tyr Lys Asn Leu Gly Phe Ser Ile Arg Arg Leu Arg
            20                  25                  30

Asp Thr Ile Ile Ala Arg Glu Thr Gly Ala Asp Ile Leu Ser Asp Thr
        35                  40                  45

Glu Tyr Asp Thr Val Arg Ser Lys Leu Tyr Arg Ala Met Asp Phe Lys
    50                  55                  60

Ile Tyr Leu Tyr Tyr Gln Glu His Pro Glu Lys Ala Lys Thr Leu Ile
65                  70                  75                  80

Glu Cys Leu Arg Ala Ser Thr Glu Glu Ala Glu Lys Glu Trp Ile Tyr
                85                  90                  95

Arg Arg Glu Ala Ser Ala Val
            100


<210>  30
<211>  225
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Gut_metagenome_contig1516000227

<400>  30

Asn Lys Phe Asp Gly Glu Ser Cys Lys Ser Phe Pro Leu Leu Met Tyr
1               5                   10                  15

Leu Leu Cys Lys Phe Leu Asp Ala Lys Glu Val Asn Glu Leu Leu Thr
            20                  25                  30

Ser Ile Ile Asn Lys Leu Glu Asn Ile Glu Ser Leu Ile Asp Ala Leu
        35                  40                  45

Val Thr Leu Asn Lys Trp Asp Gly Phe Ser Ser Asn Tyr Ser Leu Phe
    50                  55                  60
```

```
Asp Lys Gln Asn Ile Arg Ser Leu Ile Asn Asp Phe Arg Leu Val Lys
65                  70                  75                  80

Asn Leu Ser Ser Ser Lys Arg Lys Leu Lys Lys Ala Lys Gln Tyr Glu
                85                  90                  95

Tyr Gln Lys Asn Lys Gln Leu Tyr Ala Asp Ala Ile Asn Leu Phe Lys
            100                 105                 110

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

Leu Tyr Leu Ser Thr Leu Leu Thr Lys Lys Ser Gly Asn Arg Gly Asp
    130                 135                 140

Lys Val Lys Asn Phe Ile Glu Ser Ala Ile Ile Lys Asn Arg Arg Phe
145                 150                 155                 160

Val Tyr Leu Ile Arg Tyr Ala Asp Pro Glu Lys Cys Tyr Lys Leu Val
                165                 170                 175

His Asn Glu Lys Met Leu Arg Phe Val Leu Gly Glu Tyr Asn Glu Glu
            180                 185                 190

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

Asn Asn Asp Ser Asn Cys Glu Asn Arg Asn Lys Met Ile Asp Thr Leu
    210                 215                 220

Val
225
```

<210> 31
<211> 210
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig1838000319

<400> 31

```
Leu Glu Tyr Gln Ile Thr Pro Glu Asn Asp Val Val Lys Leu Pro Val
1               5                  10                  15

Ser Asn Ser Ser Ser Ile Thr Val Asn Ser Ala Asn Asp Asn Ile Phe
            20                  25                  30

Glu Val Thr Thr Ser Arg Ser Asp Leu Lys Asp Leu Glu Ala Lys Pro
        35                  40                  45

Ser Asn Pro Thr Lys Lys Ser Lys Ser Ser Glu Lys Ser Pro Arg Gln
    50                  55                  60

Asp Met Leu Gly Leu Lys Ser Ala Leu Glu Lys Glu Tyr Phe Gly Lys
65                  70                  75                  80

Glu Phe Asn Asp Asn Ile His Ile Gln Ile Ile His Cys Ile Gln Asp
                85                  90                  95

Ile Glu Lys Ile Leu Ala Val Tyr Ser Thr Asn Ala Asn Tyr Ala Leu
            100                 105                 110

Val Asn Leu Leu Gly His Asn Phe Lys Asp Glu Ser Gly Lys Arg Glu
        115                 120                 125

Asp Phe Leu Gly Ser Asp Ala Phe Lys Ile Glu Met Ser Tyr Glu Ser
    130                 135                 140

Phe Ser Ala Pro Ala Asn Glu Thr Lys Leu Ser Ser Glu Lys Glu Arg
145                 150                 155                 160

Thr Arg Glu Asn Glu Glu Lys Arg Leu Lys Tyr Leu Glu Asp Phe Asn
                165                 170                 175

Asn Met Thr Asp Pro Asp Asn Tyr Asp Lys Arg Leu Gly Tyr Phe Gly
            180                 185                 190

Glu Ala Phe Phe Lys Pro Ser Lys Asn Arg Gly Lys Asp Thr Val Ala
        195                 200                 205
```

Asp Thr
    210

<210> 32
<211> 84
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig13123000268

<400> 32

Ser Pro Lys Asp Ala Gly Leu Ala Tyr Arg Pro Ala Gly Val Arg Pro
1   5   10   15

Asp Gln Ile Gly Ala Lys Ala Ala Leu Glu Lys Arg Tyr Phe Gly Lys
            20   25   30

Glu Thr Pro Gly Asp Asn Ile His Val Gln Ile Ala Tyr Gln Ile Gln
        35   40   45

Asp Ile Glu Lys Leu Leu Ala Val Tyr Ile Ser Asn Ile Ile Tyr Ala
    50   55   60

Val Asn Asn Val Thr Gly Val Ser Ala Met Lys Asp Ser Lys Gly Arg
65   70   75   80

Pro Val Asp Leu

<210> 33
<211> 115
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig5294000434

<400> 33

Asp Lys Val Val His Val Asn Val Ala Gln Gly Ile Arg Asp Phe Ile
1   5   10   15

```
Glu Asp Phe Lys Ser Asn Ser Asp Phe Ser Met Asn Ser Tyr Phe Glu
            20                  25                  30

Leu Tyr Asn Tyr Leu Met Gln Arg Met Leu Val Asn Glu Tyr Thr Lys
        35                  40                  45

Lys Ala Asn Asp Phe Lys Val Glu Val Pro Glu Leu Leu Ile Thr Tyr
    50                  55                  60

Lys Asn Asn Leu Asp Lys Tyr Lys Thr Tyr Asp Arg Asn Phe Leu His
65                  70                  75                  80

Ile Ile Asn Met Pro Phe Ala Tyr Asn Leu Ala Arg Tyr Lys Asn Leu
                85                  90                  95

Thr Ile Glu Asp Ile Phe Tyr Asp Asn Tyr Lys Asp Lys Leu Lys Thr
            100                 105                 110

Lys Arg Gly
        115
```

<210> 34
<211> 377
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig6415000192

<400> 34

```
Val Asn Ser Asp Ala Ser Asp Phe Cys Lys Leu Ile Tyr Val Phe Thr
1               5                   10                  15

Arg Phe Leu Asp Gly Lys Glu Ile Asn Ile Leu Leu Asn Ser Leu Ile
            20                  25                  30

Lys Lys Phe Gln Asp Ile His Ser Phe Asn Thr Thr Val Lys Lys Leu
        35                  40                  45
```

```
Ser Glu Asn Asn Leu Ile Ile Asn Ala Asp Tyr Val Asp Asp Tyr Ser
    50                  55                  60

Leu Phe Glu Gln Ser Gly Thr Val Ala Arg Glu Leu Met Leu Ile Lys
65                  70                  75                  80

Ser Ile Ser Lys Met Asp Phe Gly Leu Asp Asn Ile Asn Leu Ser Phe
                85                  90                  95

Met Tyr Asp Asp Ala Leu Arg Thr Leu Gly Val Ser Asp Glu Asn Leu
            100                 105                 110

Pro Glu Val Lys Arg Glu Tyr Phe Gly Lys Thr Lys Asn Leu Ser Ala
        115                 120                 125

Tyr Ile Arg Asn Asn Val Leu Glu Asn Arg Arg Phe Lys Tyr Val Ile
    130                 135                 140

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

Ile Ala Gly Phe Val Leu Asn Arg Met Pro Asp Thr Gln Ile Lys Arg
                165                 170                 175

Tyr Tyr Asp Ser Leu Ile Asn Lys Gly Ala Thr Asp Ile Gln Ala Gln
            180                 185                 190

Ala Lys Ala Leu Leu Asp Cys Ile Thr Gly Ile Ser Phe Asp Ala Ile
        195                 200                 205

Lys Asp Asp Lys His Leu His Lys Ser Lys Glu Lys Ser Pro Gln Arg
    210                 215                 220

Ser Ala Asp Arg Glu Arg Lys Lys Ala Met Leu Thr Leu Tyr Tyr Thr
225                 230                 235                 240

Ile Val Tyr Ile Phe Val Lys Gln Met Leu His Ile Asn Ser Leu Tyr
                245                 250                 255
```

```
Thr Ile Gly Phe Phe Tyr Leu Glu Arg Asp Gln Arg Phe Ile Tyr Ser
            260                 265                 270


Arg Ala Lys Lys Glu Asn Lys Asn Pro Ser Lys Asn Ser Tyr Leu Asn
        275                 280                 285


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


Arg Ile Glu Lys Asn Glu Asn Lys Gly Phe Leu Glu Asp Phe Glu Ala
305                 310                 315                 320


Leu Trp Asn Ser Cys Gly Lys Thr Ser Arg Leu Arg Lys Glu Asp Val
                325                 330                 335


Leu Leu Tyr Ala Arg Tyr Ile Ser Pro Asp His Ala Leu Lys Asn Tyr
            340                 345                 350


Lys Met Ile Leu Asn Ser Tyr Arg Asn Lys Ile Ala His Ile Asn Val
        355                 360                 365


Ile Met Ser Ala Gly Lys Tyr Thr Gly
    370                 375


<210>  35
<211>  123
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Gut_metagenome_contig6144000300

<400>  35

Asn Lys Asp Ile Leu Lys Leu Val Ile Gln Glu Leu Pro Glu Thr Gln
1               5                   10                  15


Ile Leu Arg Tyr Tyr Ser Thr Ile Gly Lys Asn Lys Thr Thr Ser Ile
            20                  25                  30


Asp Glu Met Lys Gln Cys Leu Ile Asp Ala Ile Tyr Asn Phe Asn Ile
        35                  40                  45
```

```
Asp Asp Thr Leu Asn Ala Ile Gly Asp Ser Arg Val Ile Asn Glu Lys
    50                  55                  60

Lys Glu Gln Ser Lys Ala Ile Ile Asn Leu Tyr Leu Thr Val Ile Tyr
65                  70                  75                  80

Leu Val Thr Lys Asn Met Val Lys Ile Asn Thr Arg Phe Ser Ile Ala
                85                  90                  95

Phe Ala Cys Leu Glu Arg Asp Tyr Tyr Leu His Phe Gly Gln Pro Asn
            100                 105                 110

Gly Lys Leu Leu Tyr Leu Gln Leu Thr Glu Glu
        115                 120
```

<210> 36
<211> 101
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig9118000041

<400> 36

```
Asp Thr Asp Cys Asp Thr Gln Tyr Glu Leu Phe Asn Gly Asp Asn Pro
1               5                   10                  15

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

Lys Pro Glu Met Asp Glu Ile Lys Lys Pro Leu Tyr Lys Asp Ala Leu
        35                  40                  45

Asn Ile Leu Asn Ile Gly Asn Glu Glu Phe Asp Glu Asn Thr Phe Glu
    50                  55                  60

Phe Leu Asn Arg Glu Lys Gly Asp His Lys Leu Arg Asn Phe Ile Val
65                  70                  75                  80
```

Asn Asn Val Ile Glu Ser Thr Arg Phe Met Tyr Val Val Lys Tyr Met
                85   90   95

Asn Pro Ser Asn Cys
            100

<210> 37
<211> 331
<212> PRT
<213> Artificial Sequence

<220>
<223> Activated_sludge_metagenome_transcript_124486

<400> 37

Met Val Lys Thr Lys Ala Glu Ile His Gln Ile Leu Arg Phe Leu Gly
1   5   10   15

Thr Ile Arg Gln Cys Leu Ala His Lys Lys Asp Ala Lys Trp Asp Ile
            20   25   30

Ile Phe Asn Pro Asp Ser Met Ser Asn Glu Phe Lys Thr Leu Ile Asp
        35   40   45

Ser Tyr Tyr Asn Gly Arg Val Lys Ala Leu Asn Asp Asp Phe Ile Lys
    50   55   60

His Asn Arg Ser Ser Ser Leu Pro Ile Leu Phe His Tyr Tyr Gly Ala
65   70   75   80

Thr Glu Glu Asn Arg Arg Ala Glu Leu Thr Glu Gly Phe Tyr Asp Phe
                85   90   95

Ile Ile Lys Arg Glu Asn Lys Asn Met Gly Phe Ser Leu Lys Lys Ile
            100   105   110

Arg Glu Lys Met Leu Glu Lys Gln Glu Ala Ser Phe Ile Lys Glu Asp
        115   120   125

Gly Tyr Asn Ser Val Arg His Lys Leu Tyr Ile Leu Met Asp Phe Ile
    130   135   140

```
Ile Tyr Lys Thr Tyr Ala Ala Asp Ile Glu Leu Gln Glu Ser Ile Val
145                 150                 155                 160

Ala Thr Leu Arg Ser Asn Leu Thr Asp Glu Asp Lys Glu Ala Thr Tyr
                165                 170                 175

Glu Ile Gln Ala Lys Ala Leu Trp Asn Lys Ile Gly Asp Asp Ile Val
            180                 185                 190

Lys Lys Leu Met Pro Leu Ile Asp Gly Asn Lys Ile Lys Gly Tyr Lys
        195                 200                 205

Asp Glu Lys Ile Glu Ile Lys Lys Glu Trp Leu Lys Asp Val Met Ile
    210                 215                 220

Ser Thr Glu Asn Ala Ser Tyr Met Thr Lys Met Val Tyr Phe Met Thr
225                 230                 235                 240

Leu Phe Leu Asp Gly Lys Glu Ile Asn Glu Phe Leu Thr Val Leu Ile
                245                 250                 255

Asn Lys Phe Glu Asn Ile Gln Gly Phe Phe Asp Val Leu Gly His Asn
            260                 265                 270

Ala Leu Ala Glu Lys Ile Glu Glu Pro Lys Glu Ile Ser Lys Arg Pro
        275                 280                 285

Leu Lys Gly Phe Ala Glu Leu Ala Ala Phe Thr Gln Lys Ser Thr Asp
    290                 295                 300

Val Gly Leu Thr Glu His Leu Arg Leu Phe Glu Asn Ser Val Asn Ile
305                 310                 315                 320

Ser Asp Glu Ile Arg Val Thr Lys Asn Phe Ala
                325                 330
```

<210> 38
<211> 245

<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig1322000437

<400> 38

```
Ile Asn Glu Leu Leu Thr Thr Leu Ile Asn Lys Phe Asp Asp Ile Gln
1               5                  10                  15


Ser Leu Met Asp Ile Ala Thr Ser Lys Gln Val Asn Val Pro Cys Lys
            20                  25                  30


Phe Thr Ser Asp Tyr Ser Ile Phe Asn Asn Ser Ser Arg Ile Val Lys
        35                  40                  45


Glu Leu Asn Val Val Lys Asn Val Ala Arg Met Gln Lys Pro Thr Ala
    50                  55                  60


Ser Val Lys Thr Asp Met Leu Met Asp Ala Leu Thr Leu Leu Gly Thr
65                  70                  75                  80


Asp Asp Thr Met Asp Ser Glu Lys Leu Arg Asn Tyr Val Arg Val Gln
                85                  90                  95


Ile Met Lys Leu Asp Ala Glu Gly Lys Thr Ser Lys Ser Asp Asp His
            100                 105                 110


Thr Leu Arg Asn Phe Leu Ile Asn Asn Val Ile Asn Ser Ser Arg Phe
        115                 120                 125


Ile Tyr Val Ile Lys Tyr Val Lys Ala Asp Thr Ala Tyr Ala Leu Gly
    130                 135                 140


His Asn Arg Lys Ala Val Ala Tyr Val Leu His Arg Ile Pro Glu Gln
145                 150                 155                 160


Gln Ile Asp Arg Tyr Tyr Ser Ser Cys Ile Asp Arg Asp Ala Gly Asn
                165                 170                 175
```

```
Arg Met Glu Lys Glu Asn Ala Leu Val Asp Met Ile Val Lys Leu Ser
            180                 185                 190


Phe Arg Asp Ile Glu Asn Val Ser Asn Asp Asn Glu Lys Ala Ser Glu
        195                 200                 205


Lys Gln Arg Tyr Arg Ala Val Val Gly Leu Tyr Leu Thr Val Ile Tyr
    210                 215                 220


Glu Leu Val Lys Asn Ile Val Asn Ile Asn Ser Arg Tyr Val Ile Ala
225                 230                 235                 240


Phe His Ser Leu Glu
                245


<210>  39
<211>  139
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Gut_metagenome_contig4582000531

<400>  39

Ser Asn Asp His Ala Leu Arg Asn Phe Ile Val Asn Asn Val Ile Ser
1               5                   10                  15


Ser Ser Arg Phe Thr Tyr Ile Ile Lys Phe Ile Asn Ala Asn Ile Ala
            20                  25                  30


Tyr Lys Ile Gly His Asn Arg Lys Ala Val Ser Tyr Val Ile Lys Lys
        35                  40                  45


Leu Pro Pro Asp Gln Ile Asp Arg Tyr Phe Thr Ser Cys Ile Gly Gln
    50                  55                  60


Glu Ser Lys Gln Ala Glu Ile Ser Asp Lys Arg Gln Ala Leu Ile Asp
65                  70                  75                  80


Met Ile Thr Thr Leu Asp Tyr Tyr Arg Phe Thr Asp Ile Asn Asn Glu
                85                  90                  95
```

```
Asp Ile Gln Lys Lys Glu Gln Tyr Lys Ala Ile Ile Gly Leu Tyr Leu
            100                 105                 110


Thr Val Val Tyr Arg Leu Val Lys Asn Ile Val Asn Ile Asn Ser Arg
        115                 120                 125


Tyr Val Ile Ala Phe His Cys Leu Glu Arg Asp
    130                 135


<210>  40
<211>  100
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Gut_metagenome_contig9190000283

<400>  40

Phe Val Leu Asn Ser Ile Pro Asp Leu Gln Leu Asp Arg Tyr Tyr Lys
1               5                   10                  15


Ala Ile Thr Gly Lys Ser Asp Gly Asn Arg Val Glu Lys Thr Asp Ile
            20                  25                  30


Ile Met Asp Lys Leu Met Lys Leu Asp Phe Asn Gln Leu Lys Val Asn
        35                  40                  45


Glu Asp Gln Met Gly Lys Lys Asn Ile Tyr Phe Val Glu Lys Leu Lys
    50                  55                  60


Ala Leu Val Gly Leu Tyr Leu Thr Ile Ile Tyr Leu Leu Thr Lys Ser
65                  70                  75                  80


Leu Val Lys Ile Asn Ser Ile Tyr Ile Leu Ala Phe Ser Cys Phe Glu
                85                  90                  95


Arg Asp Tyr Arg
            100
```

<210> 41
<211> 213
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig1709000510

<400> 41

```
Pro Lys Phe Lys Ser Gly Ser His Ala Gln Lys Tyr Tyr Gly Phe Ile
1               5                   10                  15


Ile Phe Lys Ser His Lys Asn Leu Gly Phe Ser Ile Lys Glu Leu Arg
            20                  25                  30


Lys Lys Leu Phe Asp Ile Glu Asp Ala Lys Ala Phe Ser Lys Lys Gln
        35                  40                  45


Tyr Asp Ser Val Arg Phe Lys Met Asn Lys Leu Ile Asp Phe Phe Ile
    50                  55                  60


Thr Tyr His Tyr Thr Thr Ile Asp Lys Lys Gln Ala Asp Lys Leu Val
65                  70                  75                  80


Ser Lys Leu Arg Tyr Ser Ile Asp Glu Lys Glu Lys Asp Asp Ile Tyr
                85                  90                  95


Thr Thr Glu Ala Glu Asp Leu Trp Asn Lys Tyr Lys Ser Ser Phe Glu
            100                 105                 110


Gln Ile Arg Ala Tyr Val Ser Pro Asp Ile Ile Thr Glu Tyr Glu His
        115                 120                 125


Gln Ala Thr Asn Phe Glu Tyr Glu Ile Lys Lys Phe Asp Pro Ser Thr
    130                 135                 140


Glu Leu Lys Thr Phe Phe Ser Lys Met Ile Tyr Val Leu Cys Met Phe
145                 150                 155                 160


Leu Asp Gly Lys Glu Ile Asn Glu Leu Leu Ser Thr Leu Ile Asn Lys
                165                 170                 175
```

```
Phe Asp Asn Ile Asp Ser Phe Ile Thr Thr Ser Glu Gly Ile Thr Asp
            180                 185                 190


Glu Cys Arg Phe Thr Gly Thr Gly Glu Lys Lys Asn Ile Tyr Asp Ile
        195                 200                 205


Phe Asn Ala Ser Gly
    210


<210>  42
<211>  946
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  M24_(LSQX01212483_Anaerobic_digester_metagenome)

<400>  42

Met Asn Asn Lys Arg Lys Thr Lys Ala Lys Ala Ala Gly Leu Lys Ser
1               5                   10                  15


Val Phe Phe Asp Gln Lys Gln Ala Val Leu Thr Thr Phe Ala Lys Gly
            20                  25                  30


Asn Asn Ser Gln Ile Glu Lys Lys Val Val Asn Ser Glu Val Lys Asp
        35                  40                  45


Leu Arg Gln Pro Pro Ala Phe Asp Leu Glu Leu Lys Glu Lys Thr Phe
    50                  55                  60


Tyr Ile Ser Gly Lys Asn Asn Ile Asn Thr Ser Arg Glu Asn Pro Leu
65                  70                  75                  80


Ala Ser Ala Ser Leu Pro Leu Ser Lys Arg Gln Arg Ile Arg Ala Glu
                85                  90                  95


Arg Ile Lys Arg Ala Arg Glu Glu Asn Arg Pro Tyr His Asn Val Lys
            100                 105                 110
```

```
Arg Val Gly Glu Asp Asp Leu Arg Ala Lys Ala Asp Leu Glu Lys His
        115                 120                 125

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

Asn Ile Leu Asp Ile Asn Lys Ile Ile Ser Pro Tyr Ile Asn Asp Ile
145                 150                 155                 160

Val Tyr Ser Met Asn Asn Leu Ala Arg Asn Asp Glu Tyr Ile Asp Gly
                165                 170                 175

Lys Ile Asp Val Ile Gly Ser Leu Ser Ser Thr Thr Asp Tyr Ser Ser
            180                 185                 190

Phe Met Ser Pro Asn Lys Asp Leu Glu Lys Glu Lys Lys Phe Ser Phe
        195                 200                 205

His Arg Glu Asn Tyr Lys Lys Phe Val Glu Ala Ser Lys Pro Tyr Met
    210                 215                 220

Arg Tyr Tyr Gly Lys Val Phe Ile Arg Asp Val Lys Lys Ser Lys Leu
225                 230                 235                 240

Ser Thr Gly Lys Gly Glu Lys Ile Glu Val Met Tyr Arg Ser Asp Glu
                245                 250                 255

Glu Ile Phe Thr Ile Phe Gln Ile Leu Ser Tyr Val Arg Gln Ser Ile
            260                 265                 270

Met His Asn Asp Ile Gly Asn Lys Ser Ser Ile Leu Ala Ile Glu Lys
        275                 280                 285

Tyr Pro Ala Arg Phe Val Gly Phe Leu Ser Asp Leu Leu Lys Thr Lys
    290                 295                 300

Thr Asn Asp Val Asn Arg Met Phe Ile Asp Asn Asn Ser Gln Thr Asn
305                 310                 315                 320
```

```
Phe Trp Val Leu Phe Ser Ile Phe Gly Leu Gln Asp His Thr Ser Gly
                325                 330                 335

Ala Asp Lys Ile Cys Arg Asn Phe Tyr Asp Phe Val Ile Lys Ala Asp
            340                 345                 350

Ser Lys Asn Leu Gly Phe Ser Leu Lys Lys Ile Arg Glu Leu Met Leu
        355                 360                 365

Asp Leu Pro Asn Ala Asn Met Leu Arg Asp His Gln Phe Asp Thr Val
    370                 375                 380

Arg Ser Lys Phe Tyr Thr Leu Leu Asp Phe Ile Ile Tyr Gln His Tyr
385                 390                 395                 400

Leu Glu Glu Lys Ser Arg Ile Asp Asn Met Val Glu Lys Leu Arg Met
                405                 410                 415

Thr Leu Lys Glu Glu Glu Lys Glu Val Leu Tyr Ala Ala Glu Ala Lys
            420                 425                 430

Ile Val Trp Asn Ala Ile Gly Ala Lys Val Ile Asn Lys Leu Val Pro
        435                 440                 445

Met Met Asn Gly Asp Ala Leu Lys Glu Ile Lys Arg Lys Asn Arg Asp
    450                 455                 460

Arg Lys Leu Pro Gln Ser Val Ile Ala Thr Val Gln Val Asn Ser Asp
465                 470                 475                 480

Ala Asn Val Phe Ser Gly Leu Ile Tyr Phe Leu Thr Leu Phe Leu Asp
                485                 490                 495

Gly Lys Glu Ile Asn Glu Met Val Ser Asn Leu Ile Thr Lys Phe Glu
            500                 505                 510

Asn Ile Asp Ser Leu Leu His Val Asp Arg Glu Ile Tyr Lys Ser Asp
        515                 520                 525
```

```
Glu Lys Asp Leu Asp Leu Glu Ile Glu Lys Leu Ala Leu Phe Phe Lys
    530                 535                 540

Gly Val Val Arg Pro Asn Ala Lys Thr Asp Thr Gly Ala Gly Glu Ile
545                 550                 555                 560

Ser Lys Ser Phe Ser Ile Phe Gln Ser Ala Glu Arg Ile Ile Glu Glu
                565                 570                 575

Leu Lys Phe Ile Lys Asn Val Thr Arg Met Asp Asn Glu Ile Phe Pro
            580                 585                 590

Ser Glu Gly Val Phe Leu Asp Ala Ala Asn Val Leu Gly Val Arg Gly
        595                 600                 605

Asp Asp Phe Asp Phe Ser Asn Glu Phe Val Gly Asp Asp Leu His Ser
    610                 615                 620

Asp Ala Asn Lys Lys Ile Ile Asn Lys Ile Asn Gly Thr Lys Glu Asp
625                 630                 635                 640

Arg Asn Leu Arg Asn Phe Ile Ile Asn Asn Val Val Lys Ser Arg Arg
                645                 650                 655

Phe Gln Tyr Ile Ala Arg His Met Asn Thr His Tyr Val Lys Gln Leu
            660                 665                 670

Ala Asn Asn Glu Thr Leu Asn Arg Phe Val Leu Asn Lys Met Gly Asp
        675                 680                 685

Ala Lys Ile Ile Asn Arg Tyr Tyr Glu Ser Ile Ser Gly Asn Thr Pro
    690                 695                 700

Asn Ile Glu Val Arg Ser Gln Ile Asp Tyr Leu Val Lys Arg Leu Arg
705                 710                 715                 720

Ser Phe Ser Phe Glu Asp Leu Asn Asp Val Lys Gln Lys Val Arg Pro
                725                 730                 735
```

```
Gly Thr Asn Glu Ser Ile Glu Lys Glu Lys Lys Lys Ala Leu Val Gly
            740                 745                 750

Leu Cys Leu Thr Ile Gln Tyr Leu Val Tyr Lys Asn Leu Val Asn Ile
        755                 760                 765

Asn Ala Arg Tyr Thr Thr Ala Phe Tyr Cys Leu Glu Arg Asp Ser Lys
    770                 775                 780

Leu Lys Gly Phe Gly Val Asp Val Trp Arg Asp Phe Glu Ser Tyr Thr
785                 790                 795                 800

Ala Leu Thr Asn His Phe Ile Lys Glu Gly Tyr Leu Pro Val Arg Lys
                805                 810                 815

Ala Glu Ile Leu Arg Ala Asn Leu Lys His Leu Asp Cys Glu Asp Gly
            820                 825                 830

Phe Lys Tyr Tyr Arg Asn Gln Val Thr His Leu Asn Ala Ile Arg Val
        835                 840                 845

Ala Tyr Lys Tyr Ile Asn Glu Ile Lys Ser Val His Ser Tyr Phe Ala
    850                 855                 860

Leu Tyr His Tyr Ile Met Gln Arg His Leu Tyr Asp Ser Leu Gln Ala
865                 870                 875                 880

Lys Ala Lys Asp Ser Ser Gly Phe Val Ile Asp Ala Leu Lys Lys Ser
                885                 890                 895

Phe Glu His Lys Ile Tyr Ser Lys Asp Leu Leu His Val Leu His Ser
            900                 905                 910

Pro Phe Gly Tyr Asn Thr Ala Arg Tyr Lys Asn Leu Ser Ile Glu Ala
        915                 920                 925

Leu Phe Asp Lys Asn Glu Ser Arg Pro Glu Val Asn Pro Leu Ser Thr
    930                 935                 940
```

Asn Asp
945

<210> 43
<211> 150
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig3833000494

<400> 43

Glu Glu Ala Ser Asp Phe Lys Asp Lys Lys Asn Asp Ser Asp Arg Arg
1 5 10 15

Lys Leu Tyr Lys Leu Met Asp Phe Cys Ile Phe Tyr Leu Tyr Tyr Ser
20 25 30

Asp Ser Lys Arg Asn Glu Asp Leu Val Ser Arg Leu Arg Glu Ser Leu
35 40 45

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

Trp Asn Glu Leu Arg Lys Lys Phe Ser Thr Ile Leu Asp Asn Val Lys
65 70 75 80

Gly Ala Asn Ile Lys Lys Leu Glu Asp Val Lys Glu Lys Phe Ile Ser
85 90 95

Glu Asp Glu Phe Asp Asp Ile Lys Leu Asp Ile Asp Ile Ser Tyr Phe
100 105 110

Ser Lys Leu Met Tyr Val Met Cys Tyr Phe Leu Asp Gly Lys Lys Ile
115 120 125

Asn Asp Leu Leu Thr Thr Leu Val Ser Lys Phe Asp Asn Ile Gly Ser
130 135 140

Ile Ile Glu Ala Ala Lys
145 150

<210> 44
<211> 123
<212> PRT
<213> Artificial Sequence

<220>
<223> Activated_sludge_metagenome_transcript_117355

<400> 44

Met Glu Lys Met Glu Lys Ile Arg Lys Ser Glu Asn Arg Lys Ser Arg
1 5 10 15

Ala Lys Ala Ala Gly Leu Lys Ser Thr Phe Val Val Ala Asp Lys Leu
20 25 30

Ala Leu Thr Ser Phe Gly Lys Gly Asn Lys Ala Asn Leu Glu Lys Ile
35 40 45

Ile Lys Asn Glu Glu Val Glu Glu Ile Asn Lys Pro Pro Lys Phe Asp
50 55 60

Ala Lys Val Glu Glu Lys Gln Ile His Ile Ile Gly Lys Thr Ile Lys
65 70 75 80

Asp Gly Ile Thr Glu Lys Pro Ala Gly Met Arg Asn Asp Leu Ile Gly
85 90 95

Ala Lys Gln Glu Leu Glu Lys Met Phe Phe Gly Lys Thr Phe Asp Asp
100 105 110

Asn Ile His Ile Gln Leu Ile Tyr Asn Ile Leu
115 120

<210> 45
<211> 84
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig11061000330

<400> 45

Ser Gln Arg Asp Ser Thr Ile Asn Pro Gly Glu Lys Gln Ile Ser Val
1 5 10 15

Glu Lys Leu Asn Pro Lys Met Val Gln Tyr Ala Lys Ala Val Ile Asn
20 25 30

His Gly Thr Tyr Cys Lys Asp Ala Val Lys Ala Leu Cys Val Pro Phe
35 40 45

Ala Tyr Gln Leu Val Arg Tyr Lys Asn Leu Thr Ile Asp Gly Leu Phe
50 55 60

Asp Lys Asn His Pro Gly Asp Lys Lys Glu Ala Lys Thr Leu Ala Leu
65 70 75 80

Asp Asp Asp Val

<210> 46
<211> 418
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig338000322

<400> 46

Phe Asp Asn Ile Ile Ser Phe Ile Asp Ile Leu Lys Lys Leu Asn Leu
1 5 10 15

Glu Phe Lys Phe Lys Pro Glu Tyr Ala Asp Phe Phe Asn Met Thr Asn
20 25 30

Cys Arg Tyr Thr Leu Glu Glu Leu Arg Val Ile Asn Ser Ile Ala Arg
35 40 45

Met Gln Lys Pro Ser Ala Asp Ala Arg Lys Ile Met Tyr Arg Asp Ala
50 55 60

```
Leu Arg Ile Leu Gly Met Asp Asn Arg Pro Asp Glu Glu Ile Asp Arg
65                  70                  75                  80

Glu Leu Glu Arg Thr Met Pro Val Gly Ala Asp Gly Lys Phe Ile Lys
                85                  90                  95

Gly Lys Gln Gly Phe Arg Asn Phe Ile Ala Ser Asn Val Ile Glu Ser
            100                 105                 110

Ser Arg Phe His Tyr Leu Val Arg Tyr Asn Asn Pro His Lys Thr Arg
        115                 120                 125

Thr Leu Val Lys Asn Pro Asn Val Val Lys Phe Val Leu Glu Gly Ile
    130                 135                 140

Pro Glu Thr Gln Ile Lys Arg Tyr Phe Asp Val Cys Lys Gly Gln Glu
145                 150                 155                 160

Ile Pro Pro Thr Ser Asp Lys Ser Ala Gln Ile Asp Val Leu Ala Arg
                165                 170                 175

Ile Ile Ser Ser Val Asp Tyr Lys Ile Phe Glu Asp Val Pro Gln Ser
            180                 185                 190

Ala Lys Ile Asn Lys Asp Asp Pro Ser Arg Asn Phe Ser Asp Ala Leu
        195                 200                 205

Lys Lys Gln Arg Tyr Gln Ala Ile Val Ser Leu Tyr Leu Thr Val Met
    210                 215                 220

Tyr Leu Ile Thr Lys Asn Leu Val Tyr Val Asn Ser Arg Tyr Val Ile
225                 230                 235                 240

Ala Phe His Cys Leu Glu Arg Asp Ala Phe Leu His Gly Val Thr Leu
                245                 250                 255

Pro Lys Met Asn Lys Lys Ile Val Tyr Ser Gln Leu Thr Thr His Leu
            260                 265                 270
```

```
Leu Thr Asp Lys Asn Tyr Thr Thr Tyr Gly His Leu Lys Asn Gln Lys
        275                 280                 285


Gly His Arg Lys Trp Tyr Val Leu Val Lys Asn Asn Leu Gln Asn Ser
    290                 295                 300


Asp Ile Thr Ala Val Ser Ser Phe Arg Asn Ile Val Ala His Ile Ser
305                 310                 315                 320


Val Val Arg Asn Ser Asn Glu Tyr Ile Ser Gly Ile Gly Glu Leu His
                325                 330                 335


Ser Tyr Phe Glu Leu Tyr His Tyr Leu Val Gln Ser Met Ile Ala Lys
            340                 345                 350


Asn Asn Trp Tyr Asp Thr Ser His Gln Pro Lys Thr Ala Glu Tyr Leu
        355                 360                 365


Asn Asn Leu Lys Lys His His Thr Tyr Cys Lys Asp Phe Val Lys Ala
    370                 375                 380


Tyr Cys Ile Pro Phe Gly Tyr Val Val Pro Arg Tyr Lys Asn Leu Thr
385                 390                 395                 400


Ile Asn Glu Leu Phe Asp Arg Asn Asn Pro Asn Pro Glu Pro Lys Glu
                405                 410                 415


Glu Val



<210>  47
<211>  418
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Artificial sequence

<400>  47

Phe Asp Asn Ile Ile Ser Phe Ile Asp Ile Leu Lys Lys Leu Asn Leu
1               5                   10                  15
```

```
Glu Phe Lys Phe Lys Pro Glu Tyr Ala Asp Phe Phe Asn Met Thr Asn
            20                  25                  30

Cys Arg Tyr Thr Leu Glu Glu Leu Arg Val Ile Asn Ser Ile Ala Arg
        35                  40                  45

Met Gln Lys Pro Ser Ala Asp Ala Arg Lys Ile Met Tyr Arg Asp Ala
    50                  55                  60

Leu Arg Ile Leu Gly Met Asp Asn Arg Pro Asp Glu Glu Ile Asp Arg
65                  70                  75                  80

Glu Leu Glu Arg Thr Met Pro Val Gly Ala Asp Gly Lys Phe Ile Lys
                85                  90                  95

Gly Lys Gln Gly Phe Arg Asn Phe Ile Ala Ser Asn Val Ile Glu Ser
            100                 105                 110

Ser Arg Phe His Tyr Leu Val Arg Tyr Asn Asn Pro His Lys Thr Arg
        115                 120                 125

Thr Leu Val Lys Asn Pro Asn Val Val Lys Phe Val Leu Glu Gly Ile
    130                 135                 140

Pro Glu Thr Gln Ile Lys Arg Tyr Phe Asp Val Cys Lys Gly Gln Glu
145                 150                 155                 160

Ile Pro Pro Thr Ser Asp Lys Ser Ala Gln Ile Asp Val Leu Ala Arg
                165                 170                 175

Ile Ile Ser Ser Val Asp Tyr Lys Ile Phe Glu Asp Val Pro Gln Ser
            180                 185                 190

Ala Lys Ile Asn Lys Asp Asp Pro Ser Arg Asn Phe Ser Asp Ala Leu
        195                 200                 205

Lys Lys Gln Arg Tyr Gln Ala Ile Val Ser Leu Tyr Leu Thr Val Met
    210                 215                 220
```

```
Tyr Leu Ile Thr Lys Asn Leu Val Tyr Val Asn Ser Arg Tyr Val Ile
225                 230                 235                 240

Ala Phe His Cys Leu Glu Arg Asp Ala Phe Leu His Gly Val Thr Leu
                245                 250                 255

Pro Lys Met Asn Lys Lys Ile Val Tyr Ser Gln Leu Thr Thr His Leu
            260                 265                 270

Leu Thr Asp Lys Asn Tyr Thr Thr Tyr Gly His Leu Lys Asn Gln Lys
        275                 280                 285

Gly His Arg Lys Trp Tyr Val Leu Val Lys Asn Asn Leu Gln Asn Ser
    290                 295                 300

Asp Ile Thr Ala Val Ser Ser Phe Arg Asn Ile Val Ala His Ile Ser
305                 310                 315                 320

Val Val Arg Asn Ser Asn Glu Tyr Ile Ser Gly Ile Gly Glu Leu His
                325                 330                 335

Ser Tyr Phe Glu Leu Tyr His Tyr Leu Val Gln Ser Met Ile Ala Lys
            340                 345                 350

Asn Asn Trp Tyr Asp Thr Ser His Gln Pro Lys Thr Ala Glu Tyr Leu
        355                 360                 365

Asn Asn Leu Lys Lys His His Thr Tyr Cys Lys Asp Phe Val Lys Ala
    370                 375                 380

Tyr Cys Ile Pro Phe Gly Tyr Val Val Pro Arg Tyr Lys Asn Leu Thr
385                 390                 395                 400

Ile Asn Glu Leu Phe Asp Arg Asn Asn Pro Asn Pro Glu Pro Lys Glu
                405                 410                 415

Glu Val
```

<210> 48
<211> 109
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig9530000097

<400> 48

Lys Ser Lys His Tyr Arg Gln Cys Val Lys Glu Asp Leu Leu Asn Ala
1 5 10 15

Arg Lys Leu His Thr Asp Asp Tyr Asn Ile Asp Ala Ile Arg Ala Phe
20 25 30

Arg Asn Ala Val Ala His Leu Asp Val Val Arg Asn Met Ser Ala Tyr
35 40 45

Ile Glu Gly Ile Arg Glu Ile Asn Ser Tyr Tyr Gly Leu Tyr His Tyr
50 55 60

Leu Ala Gln Arg Cys Leu Gln Lys Arg Cys Asn Gly Ala Leu Asp Asp
65 70 75 80

Asn Thr Asp Tyr Phe Arg Thr Leu Asp Lys Tyr Asn Thr Tyr Val Lys
85 90 95

Asp Phe Val Lys Ala Leu Cys Val Pro Phe Ala Tyr Asn
100 105

<210> 49
<211> 213
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig1750000258

<400> 49

Ser Thr Pro Asp Lys Ala Arg Glu Ser Leu Arg Asp Met Leu His Asn
1 5 10 15

```
Ile Ser Phe Asn Asp Phe Ala Asp Val Lys Gln Asp Asp Arg Arg Ala
            20                  25                  30

Thr Pro Glu Glu Lys Val Glu Lys Glu Arg Tyr Lys Ala Ile Ile Gly
        35                  40                  45

Leu Tyr Leu Thr Val Met Tyr His Leu Val Lys Asn Leu Val Tyr Val
    50                  55                  60

Asn Ser Arg Tyr Val Met Ala Phe His Cys Leu Glu Arg Asp Ala Met
65                  70                  75                  80

His Tyr Asp Val Cys Leu Asp Asn Tyr Arg Asp Leu Ile Arg His Leu
                85                  90                  95

Ile Ser Asp Gly Asp Ser Ser Cys Asn His Phe Ile Ser His Asn Lys
            100                 105                 110

Arg Met Arg Gly Cys Ile Glu Glu Asp Val Asn Asn Ser Glu Gln Leu
        115                 120                 125

Ile Phe Gly Lys Glu Asp Ala Val Ile Arg Phe Arg Asn Asn Val Ala
    130                 135                 140

His Leu Ser Ala Ile Arg Asn Ala Asn Glu Tyr Ile Asp Asp Ile Arg
145                 150                 155                 160

Glu Ile Thr Ser Tyr Phe Ala Leu Tyr His Tyr Leu Met Gln Arg Lys
                165                 170                 175

Leu Ile Asp Asp Cys Lys Val Asn Asp Thr Ala His Lys Tyr Phe Glu
            180                 185                 190

Lys Leu Thr Lys Tyr Lys Thr Tyr Val Met Asp Met Val Lys Ala Leu
        195                 200                 205

Cys Ser Pro Phe Gly
    210
```

<210> 50
<211> 131
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig5377000274

<400> 50

Lys Glu Asn Asp Ile Leu Lys Lys Ser Arg Lys Lys Ser Ser Val Lys
1 5 10 15

Ala Asn Gly Leu Lys Ser Val Phe Val Met Glu Asn Lys Ala Leu Met
20 25 30

Thr Ser Phe Gly Lys Gly Asn Lys Ala Asn Pro Glu Lys Glu Leu Val
35 40 45

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

Asn Leu Gly Asn Ile Ile Glu Ile Ser Arg Arg Gly Val Arg Ile Glu
65 70 75 80

Lys Lys Gly Tyr Ile Ala Ser Asp Tyr Glu Ile His Lys Asp Val Ile
85 90 95

Asn Ala Lys Asp Lys Leu Glu Glu Arg Tyr Phe Gly Ser Ala Met Asp
100 105 110

Asp Asn Ile His Ile Gln Met Val Tyr Gln Ile Leu Asp Ile Glu Lys
115 120 125

Ile Leu Ser
130

<210> 51
<211> 255
<212> PRT
<213> Artificial Sequence

<220>
<223> gut_metagenome_P19E0k2120140920_c248000089

<400> 51

```
Tyr Ser Ala Asp Met Ile Asp Glu Leu Ala Gly Ile Ile His Asp Met
1               5                  10                  15


Lys Phe Lys Gln Phe Glu Asn Val Asp Gln Ser Ala Lys Arg Gly Thr
            20                  25                  30


Pro Glu Ala Ala Asp Lys Gln Gln Lys Gln Asn Val Ile Arg Leu Tyr
        35                  40                  45


Leu Thr Val Cys Tyr Leu Phe Phe Lys Asn Leu Ile Tyr Val Asn Ser
    50                  55                  60


Arg Tyr Phe Leu Ala Phe Tyr Cys Leu Glu Arg Asp Val Lys Leu Trp
65                  70                  75                  80


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


Val Leu Thr Arg Glu Phe Met Asp Ser Gly Lys Ile Ser Thr Lys Ser
            100                 105                 110


Lys Ser Lys Asn Gly Ala Glu Ser Pro Asp Glu Ala Ala Lys His Ile
        115                 120                 125


Pro Arg Asn Tyr Ile Gln Lys Asn Leu Asp Asn Ser Asp Asp Val Ala
    130                 135                 140


Ile Arg Tyr Phe Arg Asn Ala Ala Glu His Ile Asn Ala Leu Gln Cys
145                 150                 155                 160


Ala Ala Asn Tyr Leu Glu Asn Val Glu Asp Thr Glu Ser Tyr Tyr Arg
                165                 170                 175


Ile Tyr His Tyr Ile Met Gln Arg Tyr Ile Met Gln Ser Tyr Ser Gln
            180                 185                 190
```

```
Asp Lys Ser Gly Gly Lys Tyr Asp Lys Ala Asp Ala Arg Asp Ile Asn
        195                 200                 205


Asp Lys Thr Lys Glu Tyr Phe Glu Leu Val Glu Asn His Asn Thr Tyr
    210                 215                 220


Cys Lys Asp Phe Val Lys Ala Leu Cys Val Pro Phe Ala Tyr Asn Leu
225                 230                 235                 240


Ala Arg Phe Lys Asn Leu Ser Ile Asp Gly Leu Phe Asp Lys Asn
                245                 250                 255


<210>  52
<211>  89
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Gut_metagenome_contig11400000031

<400>  52

Asn Ile Val Lys His Val Leu Lys Asp Ile Pro Asp Thr Gln Ile Glu
1               5                   10                  15


Arg Tyr Tyr Lys Ser Tyr Phe Asn Thr Glu Asn Ala Ala Asn Ala Glu
            20                  25                  30


Ile Met Arg Glu Lys Leu Ala Ser Ala Leu Met Gln Val Lys Tyr Asp
        35                  40                  45


Glu Phe Lys Asn Val Lys Gln Lys Val Lys Asn Gly Thr Glu Glu Ala
    50                  55                  60


Phe Glu Lys Glu Arg Lys Lys Ala Leu Val Arg Leu Tyr Leu Thr Val
65                  70                  75                  80


Val Tyr Ile Phe Val Lys Asn Leu Val
                85
```

<210> 53
<211> 109
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig7940000191

<400> 53

Asn His Ile Asp His Leu Asn Ala Val Arg Asn Ala Asp Lys Tyr Leu
1 5 10 15

Asn Asp Val Arg Gln Val Thr Ser Trp Tyr Glu Leu Phe His Tyr Ile
20 25 30

Val Gln Arg Ala Leu Phe Asp Asp Tyr Leu Lys Ala Leu Ser Tyr Asn
35 40 45

Pro Asp Ala Gly Lys Glu Leu Asn Pro Arg Thr Leu Glu Tyr Phe Ala
50 55 60

Lys Val Gln Lys Tyr Gly Thr Cys Cys Lys Asp Phe Ile Lys Ala Leu
65 70 75 80

Asn Val Pro Phe Ala Tyr Asn Leu Pro Arg Tyr Lys Asn Leu Ser Ile
85 90 95

Asp Gly Leu Phe Asp Met Asn Arg Pro Gly Glu Lys Gly
100 105

<210> 54
<211> 124
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig6049000251

<400> 54

Leu Tyr Leu Thr Ser Phe Gly Lys Gly Asn Ala Ala Val Ile Glu Gln
1 5 10 15

```
Lys Ile Glu Pro Glu Asn Gly Tyr Arg Val Thr Gly Met Gln Ile Thr
            20                  25                  30

Pro Ser Ile Thr Val Asn Lys Ala Thr Asp Glu Ser Val Arg Phe Arg
        35                  40                  45

Val Lys Arg Lys Ile Ala Gln Lys Asp Glu Phe Ile Ala Asp Asn Pro
    50                  55                  60

Met His Glu Gly Arg His Arg Ile Glu Pro Ser Ala Gly Ser Asp Met
65                  70                  75                  80

Leu Gly Leu Lys Thr Lys Leu Glu Lys Tyr Tyr Phe Gly Lys Glu Phe
                85                  90                  95

Asp Asp Asn Leu His Ile Gln Ile Ile Tyr Asn Ile Leu Asp Ile Glu
            100                 105                 110

Lys Ile Leu Ala Val Tyr Ser Thr Asn Ile Thr Ala
        115                 120
```

<210> 55
<211> 233
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig1137000500

<400> 55

```
Met Lys Lys Lys Lys Asn Leu Arg Glu Gln Arg Glu Ala Lys Lys Gln
1               5                   10                  15

Glu Arg Leu Ala Ala Cys Ser Ala Lys Thr Lys Ala Ala Glu Ile Pro
            20                  25                  30

Ala Ala Glu Lys Pro Val Ala Thr Thr Pro Ile Ala Glu Pro Val Lys
        35                  40                  45

Thr Asp Ser Lys Lys Lys Ser Ser Val Lys Ala Ala Gly Leu Lys Ser
    50                  55                  60
```

```
Thr Leu Val Leu Asp Asn Lys Leu Tyr Thr Thr Ser Phe Gly Lys Gly
65                  70                  75                  80

Asn Glu Ala Val Ile Glu Gln Glu Val Asn Thr Ala Asp Tyr Thr Val
                85                  90                  95

Ala Asn Val Asn Ser Thr Pro Ser Leu Glu Ile Arg Ser Ala Gly Asp
            100                 105                 110

Lys Glu Leu Ser Phe Ser Ser His Arg Gln Phe Val Asn Lys Asp Glu
        115                 120                 125

Leu Ser Ala Val Asn Pro Leu Tyr Gly Gly Lys Asp Lys Pro Arg Lys
    130                 135                 140

Pro Ala Gly Lys Asp Met Leu Gly Leu Lys Asp Lys Leu Glu Glu Arg
145                 150                 155                 160

Tyr Phe Gly Lys Ser Phe Asn Asp Asn Ile His Ile Gln Ile Ile Tyr
                165                 170                 175

Asn Ile Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Ala Thr Asn Ile
            180                 185                 190

Thr Val Ala Leu Asp His Met Ile Asp Lys Asp Val Asn Ile Glu Lys
        195                 200                 205

Glu Asp Phe Ile Gly Tyr Met Asn Thr Leu Asn Thr Tyr Glu Val Phe
    210                 215                 220

Met Asp Pro Ser Lys Asn Pro Lys Leu
225                 230


<210>  56
<211>  69
<212>  PRT
<213>  Artificial Sequence

<220>
```

<223> Gut_metagenome_contig9368000105

<400> 56

Gln Arg His Val Ala Glu Gly Val Glu Pro Asn Arg His Asp Phe Ala
1   5   10   15

Glu Arg Phe Pro Tyr Tyr Ala Ala Leu Tyr Gln Tyr His Ser Tyr Val
            20   25   30

Lys Asp Phe Val Lys Ala Leu Asn Ala Pro Phe Gly Tyr Asn Ile Pro
        35   40   45

Arg Phe Lys Asn Leu Thr Ile Glu Glu Leu Phe Asp Arg Asn Ile His
    50   55   60

Ile Glu Thr Tyr Val
65

<210> 57
<211> 358
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig546000275

<400> 57

Met Asp Ser Tyr Arg Pro Lys Leu Tyr Lys Leu Ile Asp Phe Cys Ile
1   5   10   15

Phe Lys His Tyr His Glu Tyr Thr Glu Ile Ser Glu Lys Asn Val Asp
            20   25   30

Thr Leu Arg Ala Ala Val Ser Glu Glu Gln Lys Glu Ser Phe Tyr Ala
        35   40   45

Asp Glu Ala Lys Arg Leu Trp Gly Ile Phe Asp Lys Gln Phe Leu Gly
    50   55   60

Phe Cys Lys Lys Ile Asn Val Trp Val Asn Gly Ser His Glu Lys Glu
65   70   75   80

```
Ile Leu Gly Tyr Ile Asp Lys Asp Ala Tyr Arg Lys Lys Ser Asp Val
                85                  90                  95

Ser Tyr Phe Ser Lys Phe Leu Tyr Ala Met Ser Phe Phe Leu Asp Gly
            100                 105                 110

Lys Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Asn Lys Phe Asp Asn
        115                 120                 125

Ile Ala Ser Phe Ile Ser Thr Ala Lys Glu Leu Asp Ala Glu Ile Asp
    130                 135                 140

Arg Ile Leu Glu Lys Lys Leu Asp Pro Val Thr Gly Lys Pro Leu Lys
145                 150                 155                 160

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

Lys Arg Phe Ile Tyr Val Ile Lys Phe Cys Asn Pro Lys Asn Val Leu
            180                 185                 190

Lys Leu Val Lys Asn Thr Lys Val Thr Glu Phe Val Leu Lys Arg Met
        195                 200                 205

Pro Glu Ser Gln Ile Asp Arg Tyr Tyr Ser Ser Cys Ile Asp Thr Glu
    210                 215                 220

Lys Asn Pro Ser Val Asp Lys Lys Ile Ser Asp Leu Ala Glu Met Ile
225                 230                 235                 240

Lys Lys Ile Ala Phe Asp Asp Phe Arg Asn Val Arg Gln Lys Thr Arg
                245                 250                 255

Thr Arg Glu Glu Ser Leu Glu Lys Glu Arg Phe Lys Ala Val Ile Gly
            260                 265                 270

Leu Tyr Leu Thr Val Val Tyr Leu Leu Ile Lys Asn Leu Val Asn Val
        275                 280                 285
```

```
Asn Ser Arg Tyr Val Met Ala Phe His Cys Leu Glu Arg Asp Ala Lys
    290                 295                 300

Leu Tyr Gly Ile Asn Ile Gly Lys Asn Tyr Ile Glu Leu Thr Glu Asp
305                 310                 315                 320

Leu Cys Arg Glu Asn Glu Asn Ser Arg Ser Ala Tyr Leu Ala Arg Asn
                325                 330                 335

Lys Arg Leu Arg Asp Cys Val Lys Gln Asn Ile Asp Asn Ala Lys Asn
            340                 345                 350

Met Lys Ser Lys Glu Lys
        355
```

<210> 58
<211> 114
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig7216000573

<400> 58

```
Lys Glu Met Ile Tyr Ala Asn Glu Ala Thr Phe Leu Trp Asp Lys Leu
1               5                   10                  15

Gly Arg Thr Phe Leu Asn Phe Cys Asp Asn Ile Gly Ala Trp Val Lys
            20                  25                  30

Asn Asp His Thr Glu Glu Ile Glu Thr Tyr Ile Asp Lys Asp Ala Tyr
        35                  40                  45

Lys Thr Ser Ser Lys Val Ser Tyr Phe Ser Lys Leu Leu Tyr Ala Met
    50                  55                  60

Cys Phe Phe Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr Thr Leu
65                  70                  75                  80
```

Ile Asn Lys Phe Asp Asn Ile Gly Ser Phe Ile Glu Thr Ala Lys Glu
85 90 95

Leu Gly Leu Ser Val Thr Phe Ile Lys Asp Tyr Ala Phe Phe Asn Ala
100 105 110

Ser Lys

<210> 59
<211> 136
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig4806000409

<400> 59

Leu Ile Asp Ser Lys Arg Leu Arg Tyr Phe Gly Phe Glu Tyr Asp Pro
1 5 10 15

Asp Gly Lys Asn Lys Asn Glu Glu Met Lys Lys Arg Leu Tyr His Leu
20 25 30

Val Ala Phe Ala Gly Gln Leu Arg Gln Trp Ser Phe His Ser Glu Gly
35 40 45

Asn Phe Gln Leu Glu Trp Leu Tyr Lys Leu Asp Asp Ser Arg Ile Ala
50 55 60

Gln Glu Tyr Arg Asp Thr Leu Asp Tyr Phe Phe Asp Arg Arg Phe Asp
65 70 75 80

Glu Leu Asn Asn Asn Phe Val Glu Gln Asn Ala Thr Asn Leu Phe Ile
85 90 95

Leu Lys Glu Thr Phe Pro Gly Glu Asp Leu Lys Ala Val Thr Asp Leu
100 105 110

Tyr Tyr Asp Phe Ile Ile Val Lys Ser Gln Lys Asn Ile Gly Phe Ser
115 120 125

Ile Lys Lys Leu Arg Glu Lys Met
130 135

<210> 60
<211> 94
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig10762000480

<400> 60

Cys Phe Asp Thr Glu Leu Asn Pro Met Thr Glu Ile Lys Ile Gln Asn
1 5 10 15

Leu Ala Gln Lys Met Lys Asp Met Asp Phe Thr Asn Phe Arg Asp Val
20 25 30

Gln Gln Lys Ser Arg Asp Pro Glu Lys Asn Arg Glu Lys Glu Arg Phe
35 40 45

Lys Ala Ile Ile Gly Leu Tyr Leu Thr Val Val Tyr Arg Leu Val Lys
50 55 60

Asn Leu Ala Asp Val Asn Ala Arg Tyr Val Met Ala Phe His Ser Leu
65 70 75 80

Glu Arg Asp Ser Ser Leu Tyr Asn Glu Pro Val Val Trp Lys
85 90

<210> 61
<211> 149
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig4114000374

<400> 61

Asp Thr Lys Ile Asn Pro Gln Thr Trp Leu Tyr Gln Leu Glu Asn Thr
1 5 10 15

```
Pro Asp Leu Asp Asn Glu Tyr Arg Asp Thr Leu Asp His Phe Phe Asp
            20                  25                  30

Glu Arg Phe Asn Glu Ile Asn Glu His Phe Val Thr Gln Asn Ala Thr
        35                  40                  45

Asn Leu Cys Ile Met Lys Glu Val Phe Pro Asp Glu Asp Phe Lys Ser
    50                  55                  60

Ile Ala Asp Leu Tyr Tyr Asp Phe Ile Val Val Lys Ser Tyr Lys Asn
65                  70                  75                  80

Ile Gly Phe Ser Ile Lys Lys Leu Arg Glu Lys Met Leu Glu Leu Pro
                85                  90                  95

Glu Ala Lys Arg Val Thr Ser Thr Glu Met Asp Ser Val Arg Ser Lys
            100                 105                 110

Leu Tyr Lys Leu Ile Asp Phe Cys Ile Phe Lys His Tyr His Glu Lys
        115                 120                 125

Pro Glu Thr Val Glu Met Ile Val Ser Met Leu Arg Ala Tyr Thr Ser
    130                 135                 140

Glu Asp Met Lys Glu
145


<210>  62
<211>  933
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Ruminococcus_flavefaciens_FD1

<400>  62

Met Lys Lys Lys Met Ser Leu Arg Glu Lys Arg Glu Ala Glu Lys Gln
1               5                   10                  15
```

```
Ala Lys Lys Ala Ala Tyr Ser Ala Ala Ser Lys Asn Thr Asp Ser Lys
            20                  25                  30

Pro Ala Glu Lys Lys Ala Glu Thr Pro Lys Pro Ala Glu Ile Ile Ser
        35                  40                  45

Asp Asn Ser Arg Asn Lys Thr Ala Val Lys Ala Ala Gly Leu Lys Ser
    50                  55                  60

Thr Ile Ile Ser Gly Asp Lys Leu Tyr Met Thr Ser Phe Gly Lys Gly
65                  70                  75                  80

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

Ser Ala Met Lys Asp Thr Pro Ser Leu Glu Val Asp Lys Ala Glu Ser
            100                 105                 110

Lys Glu Ile Ser Phe Ser Ser His His Pro Phe Val Lys Asn Asp Lys
        115                 120                 125

Leu Thr Thr Tyr Asn Pro Leu Tyr Gly Gly Lys Asp Asn Pro Glu Lys
    130                 135                 140

Pro Val Gly Arg Asp Met Leu Gly Leu Lys Asp Lys Leu Glu Glu Arg
145                 150                 155                 160

Tyr Phe Gly Cys Thr Phe Asn Asp Asn Leu His Ile Gln Ile Ile Tyr
                165                 170                 175

Asn Ile Leu Asp Ile Glu Lys Ile Leu Ala Val His Ser Ala Asn Ile
            180                 185                 190

Thr Thr Ala Leu Asp His Met Val Asp Glu Asp Asp Glu Lys Tyr Leu
        195                 200                 205

Asn Ser Asp Tyr Ile Gly Tyr Met Asn Thr Ile Asn Thr Tyr Asp Val
    210                 215                 220
```

```
Phe Met Asp Pro Ser Lys Asn Ser Ser Leu Ser Pro Lys Asp Arg Lys
225                 230                 235                 240


Asn Ile Asp Asn Ser Arg Ala Lys Phe Glu Lys Leu Leu Ser Thr Lys
                245                 250                 255


Arg Leu Gly Tyr Phe Gly Phe Asp Tyr Asp Ala Asn Gly Lys Asp Lys
            260                 265                 270


Lys Lys Asn Glu Glu Ile Lys Lys Arg Leu Tyr His Leu Thr Ala Phe
        275                 280                 285


Ala Gly Gln Leu Arg Gln Trp Ser Phe His Ser Ala Gly Asn Tyr Pro
    290                 295                 300


Arg Thr Trp Leu Tyr Lys Leu Asp Ser Leu Asp Lys Glu Tyr Leu Asp
305                 310                 315                 320


Thr Leu Asp His Tyr Phe Asp Lys Arg Phe Asn Asp Ile Asn Asp Asp
                325                 330                 335


Phe Val Thr Lys Asn Ala Thr Asn Leu Tyr Ile Leu Lys Glu Val Phe
            340                 345                 350


Pro Glu Ala Asn Phe Lys Asp Ile Ala Asp Leu Tyr Tyr Asp Phe Ile
        355                 360                 365


Val Ile Lys Ser His Lys Asn Met Gly Phe Ser Ile Lys Lys Leu Arg
    370                 375                 380


Glu Lys Met Leu Glu Cys Asp Gly Ala Asp Arg Ile Lys Glu Gln Asp
385                 390                 395                 400


Met Asp Ser Val Arg Ser Lys Leu Tyr Lys Leu Ile Asp Phe Cys Ile
                405                 410                 415


Phe Lys Tyr Tyr His Glu Phe Pro Glu Leu Ser Glu Lys Asn Val Asp
            420                 425                 430
```

```
Ile Leu Arg Ala Ala Val Ser Asp Thr Lys Lys Asp Asn Leu Tyr Ser
        435                 440                 445

Asp Glu Ala Ala Arg Leu Trp Ser Ile Phe Lys Glu Lys Phe Leu Gly
    450                 455                 460

Phe Cys Asp Lys Ile Val Val Trp Val Thr Gly Glu His Glu Lys Asp
465                 470                 475                 480

Ile Thr Ser Val Ile Asp Lys Asp Ala Tyr Arg Asn Arg Ser Asn Val
                485                 490                 495

Ser Tyr Phe Ser Lys Leu Met Tyr Ala Met Cys Phe Phe Leu Asp Gly
            500                 505                 510

Lys Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Asn Lys Phe Asp Asn
        515                 520                 525

Ile Ala Asn Gln Ile Lys Thr Ala Lys Glu Leu Gly Ile Asn Thr Ala
    530                 535                 540

Phe Val Lys Asn Tyr Asp Phe Phe Asn His Ser Glu Lys Tyr Val Asp
545                 550                 555                 560

Glu Leu Asn Ile Val Lys Asn Ile Ala Arg Met Lys Lys Pro Ser Ser
                565                 570                 575

Asn Ala Lys Lys Ala Met Tyr His Asp Ala Leu Thr Ile Leu Gly Ile
            580                 585                 590

Pro Glu Asp Met Asp Glu Lys Ala Leu Asp Glu Glu Leu Asp Leu Ile
        595                 600                 605

Leu Glu Lys Lys Thr Asp Pro Val Thr Gly Lys Pro Leu Lys Gly Lys
    610                 615                 620

Asn Pro Leu Arg Asn Phe Ile Ala Asn Asn Val Ile Glu Asn Ser Arg
625                 630                 635                 640
```

```
Phe Ile Tyr Leu Ile Lys Phe Cys Asn Pro Glu Asn Val Arg Lys Ile
                645                 650                 655

Val Asn Asn Thr Lys Val Thr Glu Phe Val Leu Lys Arg Ile Pro Asp
            660                 665                 670

Ala Gln Ile Glu Arg Tyr Tyr Lys Ser Cys Thr Asp Ser Glu Met Asn
        675                 680                 685

Pro Pro Thr Glu Lys Lys Ile Thr Glu Leu Ala Gly Lys Leu Lys Asp
    690                 695                 700

Met Asn Phe Gly Asn Phe Arg Asn Val Arg Gln Ser Ala Lys Glu Asn
705                 710                 715                 720

Met Glu Lys Glu Arg Phe Lys Ala Val Ile Gly Leu Tyr Leu Thr Val
                725                 730                 735

Val Tyr Arg Val Val Lys Asn Leu Val Asp Val Asn Ser Arg Tyr Ile
            740                 745                 750

Met Ala Phe His Ser Leu Glu Arg Asp Ser Gln Leu Tyr Asn Val Ser
        755                 760                 765

Val Asp Asn Asp Tyr Leu Ala Leu Thr Asp Thr Leu Val Lys Glu Gly
    770                 775                 780

Asp Asn Ser Arg Ser Arg Tyr Leu Ala Gly Asn Lys Arg Leu Arg Asp
785                 790                 795                 800

Cys Val Lys Gln Asp Ile Asp Asn Ala Lys Lys Trp Phe Val Ser Asp
                805                 810                 815

Lys Tyr Asn Ser Ile Thr Lys Tyr Arg Asn Asn Val Ala His Leu Thr
            820                 825                 830

Ala Val Arg Asn Cys Ala Glu Phe Ile Gly Asp Ile Thr Lys Ile Asp
        835                 840                 845
```

```
Ser Tyr Phe Ala Leu Tyr His Tyr Leu Ile Gln Arg Gln Leu Ala Lys
    850                 855                 860


Gly Leu Asp His Glu Arg Ser Gly Phe Asp Arg Asn Tyr Pro Gln Tyr
865                 870                 875                 880


Ala Pro Leu Phe Lys Trp His Thr Tyr Val Lys Asp Val Val Lys Ala
                885                 890                 895


Leu Asn Ala Pro Phe Gly Tyr Asn Ile Pro Arg Phe Lys Asn Leu Ser
            900                 905                 910


Ile Asp Ala Leu Phe Asp Arg Asn Glu Ile Lys Lys Asn Asp Gly Glu
        915                 920                 925


Lys Lys Ser Asp Asp
    930


<210>  63
<211>  70
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Gut_metagenome_contig7093000170

<400>  63

Arg Gln Leu Ala Val Gly Leu Asp Pro Lys Arg Ser Gly Phe Glu Arg
1               5                   10                  15


Asn Tyr Pro Tyr Phe Ala Lys Leu Phe Glu Trp His Thr Tyr Val Lys
            20                  25                  30


Asp Met Val Lys Ala Leu Asn Ala Pro Phe Gly Tyr Asn Ile Pro Arg
        35                  40                  45


Phe Lys Asn Leu Ser Ile Asp Asp Leu Phe Asp Met Asn Glu Glu Arg
    50                  55                  60


Val Lys Asp Ala Lys Ala
65                  70
```

<210> 64
<211> 91
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig11113000384

<400> 64

Ala Ser Glu Asn Glu Arg Lys Ala Lys Gln Gln Ala Gly Gln Pro Ala
1 5 10 15

Ala Lys Gly Asn Gly Glu Glu Lys Gln Lys Pro Ile Leu Lys Leu Asn
20 25 30

Pro Val Arg Glu Arg Lys Leu Ser Thr Thr Leu Lys Ser Arg Ala Lys
35 40 45

Ala Ala Gly Leu Lys Ser Thr Phe Val Ile Asn Asp Lys Glu Leu Leu
50 55 60

Met Thr Ala Phe Gly Arg Gly Asn Asp Ala Ile Pro Glu Lys Arg Ile
65 70 75 80

Leu Gly Ala Gln Val Glu Thr Leu Ser Glu Lys
85 90

<210> 65
<211> 119
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig6403000259

<400> 65

Asn Lys Val Lys Glu Phe Glu Leu Asp Ala Ser Val Lys Tyr Arg Leu
1 5 10 15

Ser Lys Val Val Glu Ser Val Arg Leu Lys Asp Lys Ala Thr Tyr Phe
20 25 30

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


Asn Asp Leu Leu Thr Thr Leu Ile His Gln Phe Glu Asn Ile Ala Ser
    50                  55                  60


Phe Ile Asp Val Met Lys Asp Arg Gly Ile Ala Asp Gly Phe Ser Gly
65                  70                  75                  80


Glu Tyr Lys Phe Phe Glu Ser Ser Lys Gln Val Ala Phe Glu Leu Arg
                85                  90                  95


Asn Val Asn Ser Phe Ala Arg Met Thr Lys Ala Ser Lys Asp Asp Asp
            100                 105                 110


Asn Ala Lys Gln Ile Met Phe
        115


<210>  66
<211>  123
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Gut_metagenome_contig6193000124

<400>  66

Ala Glu Asp Tyr Tyr Asp Phe Leu Phe Lys Lys Thr Tyr Lys Tyr Leu
1               5                   10                  15


Gly Phe Ser Ile Lys Thr Leu Arg Glu Glu Phe Leu Thr Ile Lys Asp
            20                  25                  30


Ala Glu Lys Leu Lys Asp Lys Lys Tyr Asp Ser Val Arg Ser Lys Leu
        35                  40                  45


Tyr Lys Val Ile Asp Phe Leu Ile Phe Asp Leu Tyr Gln Asn Arg Glu
    50                  55                  60
```

Cys Leu Arg Ile Asn Glu Leu Val Gln Arg Leu Arg Leu Ser Thr Asp
65 70 75 80

Glu Leu Thr Lys Ala Lys Ile Tyr Ala Asp Glu Ala His Phe Val Tyr
85 90 95

Lys Lys Leu Trp Asp Ile Ile Asp Thr Glu Ile Val Pro Gln Leu Asp
100 105 110

Gln Lys Gly Ile Lys Ala Leu Glu Glu Ala Tyr
115 120

<210> 67
<211> 320
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig721000619

<400> 67

Lys Glu Gly Ser Thr Met Ala Lys Asn Glu Lys Lys Lys Ser Thr Ala
1 5 10 15

Lys Ala Leu Gly Leu Lys Ser Ser Phe Val Val Asn Asn Asp Ile Tyr
20 25 30

Met Thr Ser Phe Gly Lys Gly Asn Lys Ala Val Leu Glu Lys Lys Ile
35 40 45

Thr Glu Asn Thr Ile Glu Asn Lys Ser Asp Thr Thr Tyr Phe Asp Val
50 55 60

Ile Asn Arg Asp Pro Lys Gly Phe Thr Leu Glu Gly Arg Arg Ile Ala
65 70 75 80

Asp Met Thr Ala Phe Ser Asn Asp Pro Lys Tyr His Val Asn Val Val
85 90 95

Asn Gly Lys Phe Leu Glu Asp Gln Leu Gly Ala Arg Ser Glu Leu Glu
100 105 110

```
Lys Lys Val Phe Gly Arg Thr Phe Asp Asp Asn Val His Ile Gln Leu
        115                 120                 125

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

Asp Ile Val Tyr Leu Leu His Asn Thr Ile Lys Arg Asp Met Asn Asp
145                 150                 155                 160

Asp Ile Met Gly Tyr Ile Ser Ile Arg Asn Ser Phe Asp Asp Phe Cys
                165                 170                 175

His Pro Glu Arg Ile Pro Asp Arg Lys Ala Lys Asp Asn Leu Gln Lys
            180                 185                 190

Gln His Asp Ile Phe Phe Asp Glu Ile Leu Lys Cys Gly Arg Leu Ala
        195                 200                 205

Tyr Phe Gly Asn Ala Phe Phe Glu Asp Gly Ser Asp Asn Lys Glu Ile
    210                 215                 220

Ala Lys Leu Lys Arg Tyr Lys Glu Ile Tyr His Ile Ile Ala Leu Met
225                 230                 235                 240

Gly Ser Leu Arg Gln Ser Tyr Phe His Gly Glu Asn Ser Asp Lys Asn
                245                 250                 255

Phe Gln Gly Pro Thr Trp Ala Tyr Thr Leu Glu Ser Asn Leu Thr Gly
            260                 265                 270

Lys Tyr Lys Glu Phe Lys Asp Thr Leu Asp Lys Thr Phe Asp Glu Arg
        275                 280                 285

Tyr Glu Met Ile Ser Lys Asp Phe Gly Ser Thr Asn Met Val Asn Leu
    290                 295                 300

Gln Ile Leu Glu Glu Leu Leu Lys Met Leu Tyr Gly Asn Val Ser Pro
305                 310                 315                 320
```

<210> 68
<211> 220
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig1666000270

<400> 68

Lys Glu Leu Lys Phe Arg Ile Asn Gly Lys Phe Ile Lys Asn Tyr His
1 5 10 15

Asp Arg Tyr Asp Glu Asp Ser Ala Asn Arg Arg Trp Lys Val Val Thr
20 25 30

Lys Gln Asn Asp Val Asn Leu Phe Cys Lys Leu Ile Tyr Met Met Thr
35 40 45

Leu Leu Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr Thr Leu Val
50 55 60

Asn Lys Phe Asp Asn Ile Ala Ser Phe Ile Asp Val Met Glu Asp Met
65 70 75 80

Gly Ile Glu His Arg Phe Ile Gly Gly Tyr Asp Met Phe Ala Asp Ser
85 90 95

Lys Lys Ile Cys Val Asp Leu Gln Met Ile Asn Ser Phe Ala Arg Met
100 105 110

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

Val Ile Leu Gly Met Asp Gln Ser Glu Glu Glu Ile Glu Arg Phe Met
130 135 140

Leu Glu Asn Asn Ile Ile Lys Asp Asn Asn Glu Asp Lys Ser Glu Lys
145 150 155 160

Lys Gln His Asp Phe Arg Asn Phe Leu Ile Asn Ser Val Ile Asn Ser
165 170 175

Ser Arg Phe Arg Tyr Met Val Arg Tyr Ser Ser Ala Asp Gly Leu Arg
180 185 190

Lys Leu Lys Lys Ser Arg Lys Leu Ile Gly Phe Val Leu Asn Ser Leu
195 200 205

Pro Glu Thr Leu Ile Asp Arg Tyr Tyr His Ser Cys
210 215 220

<210> 69
<211> 204
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig2002000411

<400> 69

Glu Lys Gln Asn Lys Ala Lys Tyr Gln Ala Ile Ile Ser Leu Tyr Leu
1 5 10 15

Met Val Met Tyr Gln Ile Val Lys Asn Met Ile Tyr Val Asn Ser Arg
20 25 30

Tyr Val Ile Ala Phe His Cys Leu Glu Arg Asp Ser Asn Gln Leu Leu
35 40 45

Gly Arg Phe Asn Ser Arg Asp Ala Ser Met Tyr Asn Lys Leu Thr Gln
50 55 60

Lys Phe Ile Thr Asp Lys Tyr Leu Asn Asp Gly Ala Gln Gly Cys Ser
65 70 75 80

Lys Lys Val Gly Asn Tyr Leu Ser His Asn Ile Thr Cys Cys Ser Asp
85 90 95

Glu Leu Arg Lys Glu Tyr Arg Asn Gln Val Asp His Phe Ala Val Val
100 105 110

```
Arg Met Ile Gly Lys Tyr Ala Ala Asp Ile Gly Lys Phe Ser Thr Trp
        115                 120                 125


Phe Glu Leu Tyr His Tyr Val Met Gln Arg Ile Ile Phe Asp Lys Arg
    130                 135                 140


Asn Pro Leu Ser Glu Thr Glu Arg Thr Tyr Lys Gln Leu Ile Ala Lys
145                 150                 155                 160


His His Thr Tyr Cys Lys Asp Leu Val Lys Ala Leu Asn Thr Pro Phe
                165                 170                 175


Gly Tyr Asn Leu Ala Arg Tyr Lys Asn Leu Ser Ile Gly Glu Leu Phe
            180                 185                 190


Asp Arg Asn Asn Tyr Asn Ala Lys Thr Lys Glu Thr
        195                 200


<210>  70
<211>  944
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Ruminococcus_albus

<400>  70

Met Ala Lys Lys Ser Lys Gly Met Ser Leu Arg Glu Lys Arg Glu Leu
1               5                   10                  15


Glu Lys Gln Lys Arg Ile Gln Lys Ala Ala Val Asn Ser Val Asn Asp
            20                  25                  30


Thr Pro Glu Lys Thr Glu Glu Ala Asn Val Val Ser Val Asn Val Arg
        35                  40                  45


Thr Ser Ala Glu Asn Lys His Ser Lys Lys Ser Ala Ala Lys Ala Leu
    50                  55                  60
```

```
Gly Leu Lys Ser Gly Leu Val Ile Gly Asp Glu Leu Tyr Leu Thr Ser
65                  70                  75                  80


Phe Gly Arg Gly Asn Glu Ala Lys Leu Glu Lys Lys Ile Ser Gly Asp
                85                  90                  95


Thr Val Glu Lys Leu Gly Ile Gly Ala Phe Glu Val Ala Glu Arg Asp
            100                 105                 110


Glu Ser Thr Leu Thr Leu Glu Ser Gly Arg Ile Lys Asp Lys Thr Ala
        115                 120                 125


Arg Pro Lys Asp Pro Arg His Ile Thr Val Asp Thr Gln Gly Lys Phe
    130                 135                 140


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


Gly Lys Thr Phe Asp Asp Asn Ile His Val Gln Leu Ala Tyr Asn Ile
                165                 170                 175


Leu Asp Val Glu Lys Ile Met Ala Gln Tyr Val Ser Asp Ile Val Tyr
            180                 185                 190


Met Leu His Asn Thr Asp Lys Thr Glu Arg Asn Asp Asn Leu Met Gly
        195                 200                 205


Tyr Met Ser Ile Arg Asn Thr Tyr Lys Thr Phe Cys Asp Thr Ser Asn
    210                 215                 220


Leu Pro Asp Asp Thr Lys Gln Lys Val Glu Asn Gln Lys Arg Glu Phe
225                 230                 235                 240


Asp Lys Ile Ile Lys Ser Gly Arg Leu Gly Tyr Phe Gly Glu Ala Phe
                245                 250                 255


Met Val Asn Ser Gly Asn Ser Thr Lys Leu Arg Pro Glu Lys Glu Ile
            260                 265                 270
```

```
Tyr His Ile Phe Ala Leu Met Ala Ser Leu Arg Gln Ser Tyr Phe His
        275                 280                 285

Gly Tyr Val Lys Asp Thr Asp Tyr Gln Gly Thr Thr Trp Ala Tyr Thr
    290                 295                 300

Leu Glu Asp Lys Leu Lys Gly Pro Ser His Glu Phe Arg Glu Thr Ile
305                 310                 315                 320

Asp Lys Ile Phe Asp Glu Gly Phe Ser Lys Ile Ser Lys Asp Phe Gly
                325                 330                 335

Lys Met Asn Lys Val Asn Leu Gln Ile Leu Glu Gln Met Ile Gly Glu
            340                 345                 350

Leu Tyr Gly Ser Ile Glu Arg Gln Asn Leu Thr Cys Asp Tyr Tyr Asp
        355                 360                 365

Phe Ile Gln Leu Lys Lys His Lys Tyr Leu Gly Phe Ser Ile Lys Arg
    370                 375                 380

Leu Arg Glu Thr Met Leu Glu Thr Thr Pro Ala Glu Cys Tyr Lys Ala
385                 390                 395                 400

Glu Cys Tyr Asn Ser Glu Arg Gln Lys Leu Tyr Lys Leu Ile Asp Phe
                405                 410                 415

Leu Ile Tyr Asp Leu Tyr Tyr Asn Arg Lys Pro Ala Arg Ile Glu Glu
            420                 425                 430

Ile Val Asp Lys Leu Arg Glu Ser Val Asn Asp Glu Glu Lys Glu Ser
        435                 440                 445

Ile Tyr Ser Val Glu Ala Lys Tyr Val Tyr Glu Ser Leu Ser Lys Val
    450                 455                 460

Leu Asp Lys Ser Leu Lys Asn Ser Val Ser Gly Glu Thr Ile Lys Asp
465                 470                 475                 480
```

```
Leu Gln Lys Arg Tyr Asp Asp Glu Thr Ala Asn Arg Ile Trp Asp Ile
                485                 490                 495

Ser Gln His Ser Ile Ser Gly Asn Val Asn Cys Phe Cys Lys Leu Ile
            500                 505                 510

Tyr Ile Met Thr Leu Met Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu
        515                 520                 525

Thr Thr Leu Val Asn Lys Phe Asp Asn Ile Ala Ser Phe Ile Asp Val
    530                 535                 540

Met Asp Glu Leu Gly Leu Glu His Ser Phe Thr Asp Asn Tyr Lys Met
545                 550                 555                 560

Phe Ala Asp Ser Lys Ala Ile Cys Leu Asp Leu Gln Phe Ile Asn Ser
                565                 570                 575

Phe Ala Arg Met Ser Lys Ile Asp Asp Glu Lys Ser Lys Arg Gln Leu
            580                 585                 590

Phe Arg Asp Ala Leu Val Ile Leu Asp Ile Gly Asn Lys Asp Glu Thr
        595                 600                 605

Trp Ile Asn Asn Tyr Leu Asp Ser Asp Ile Phe Lys Leu Asp Lys Glu
    610                 615                 620

Gly Asn Lys Leu Lys Gly Ala Arg His Asp Phe Arg Asn Phe Ile Ala
625                 630                 635                 640

Asn Asn Val Ile Lys Ser Ser Arg Phe Lys Tyr Leu Val Lys Tyr Ser
                645                 650                 655

Ser Ala Asp Gly Met Ile Lys Leu Lys Thr Asn Glu Lys Leu Ile Gly
            660                 665                 670

Phe Val Leu Asp Lys Leu Pro Glu Thr Gln Ile Asp Arg Tyr Tyr Glu
        675                 680                 685
```

```
Ser Cys Gly Leu Asp Asn Ala Val Val Asp Lys Lys Val Arg Ile Glu
    690                 695                 700

Lys Leu Ser Gly Leu Ile Arg Asp Met Lys Phe Asp Asp Phe Ser Gly
705                 710                 715                 720

Val Lys Thr Ser Asn Lys Ala Gly Asp Asn Asp Lys Gln Asp Lys Ala
                725                 730                 735

Lys Tyr Gln Ala Ile Ile Ser Leu Tyr Leu Met Val Leu Tyr Gln Ile
            740                 745                 750

Val Lys Asn Met Ile Tyr Val Asn Ser Arg Tyr Val Ile Ala Phe His
        755                 760                 765

Cys Leu Glu Arg Asp Phe Gly Met Tyr Gly Lys Asp Phe Gly Lys Tyr
    770                 775                 780

Tyr Gln Gly Cys Arg Lys Leu Thr Asp His Phe Ile Glu Glu Lys Tyr
785                 790                 795                 800

Met Lys Glu Gly Lys Leu Gly Cys Asn Lys Lys Val Gly Arg Tyr Leu
                805                 810                 815

Lys Asn Asn Ile Ser Cys Cys Thr Asp Gly Leu Ile Asn Thr Tyr Arg
            820                 825                 830

Asn Gln Val Asp His Phe Ala Val Val Arg Lys Ile Gly Asn Tyr Ala
        835                 840                 845

Ala Tyr Ile Lys Ser Ile Gly Ser Trp Phe Glu Leu Tyr His Tyr Val
    850                 855                 860

Ile Gln Arg Ile Val Phe Asp Glu Tyr Arg Phe Ala Leu Asn Asn Thr
865                 870                 875                 880

Glu Ser Asn Tyr Lys Asn Ser Ile Ile Lys His His Thr Tyr Cys Lys
                885                 890                 895
```

Asp Met Val Lys Ala Leu Asn Thr Pro Phe Gly Tyr Asp Leu Pro Arg
900 905 910

Tyr Lys Asn Leu Ser Ile Gly Asp Leu Phe Asp Arg Asn Asn Tyr Leu
915 920 925

Asn Lys Thr Lys Glu Ser Ile Asp Ala Asn Ser Ser Ile Asp Ser Gln
930 935 940

<210> 71
<211> 449
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig13552000311

<400> 71

Leu Ile Asp Phe Leu Ile Tyr Asp Leu Tyr Tyr Asn Arg Lys Pro Ala
1 5 10 15

Arg Ile Glu Glu Ile Val Asp Lys Leu Arg Glu Ser Val Asn Asp Glu
20 25 30

Glu Lys Glu Ser Ile Tyr Ser Ala Glu Thr Lys Tyr Val Tyr Glu Ala
35 40 45

Leu Gly Lys Val Leu Val Arg Ser Leu Lys Lys Tyr Leu Asn Gly Ala
50 55 60

Thr Ile Arg Asp Leu Lys Asn Arg Tyr Asp Ala Lys Thr Ala Asn Arg
65 70 75 80

Ile Trp Asp Ile Ser Glu His Ser Lys Ser Gly His Val Asn Cys Phe
85 90 95

Cys Lys Leu Ile Tyr Met Met Thr Leu Met Leu Asp Gly Lys Glu Ile
100 105 110

Asn Asp Leu Leu Thr Thr Leu Val Asn Lys Phe Asp Asn Ile Ala Ser
115 120 125

```
Phe Ile Asp Val Met Asp Glu Leu Gly Leu Glu His Ser Phe Thr Asp
    130                 135                 140

Asn Tyr Lys Met Phe Ala Asp Ser Lys Ala Ile Cys Leu Asp Leu Gln
145                 150                 155                 160

Phe Ile Asn Ser Phe Ala Arg Met Ser Lys Ile Asp Asp Glu Lys Ser
                165                 170                 175

Lys Arg Gln Leu Phe Arg Asp Ala Leu Val Val Leu Asp Ile Gly Asp
            180                 185                 190

Lys Asn Glu Asp Trp Ile Glu Lys Tyr Leu Thr Ser Asp Ile Phe Lys
        195                 200                 205

Arg Asp Glu Asn Gly Asn Lys Ile Asp Gly Glu Lys Arg Asp Phe Arg
    210                 215                 220

Asn Phe Ile Ala Asn Asn Val Ile Lys Ser Ala Arg Phe Lys Tyr Leu
225                 230                 235                 240

Val Lys Tyr Ser Ser Ala Asp Gly Met Ile Lys Leu Lys Lys Asn Glu
                245                 250                 255

Lys Leu Ile Ser Phe Val Leu Glu Gln Leu Pro Glu Thr Gln Ile Asp
            260                 265                 270

Arg Tyr Tyr Glu Ser Cys Gly Leu Asp Cys Ala Val Ala Asp Arg Lys
        275                 280                 285

Val Arg Ile Glu Lys Leu Thr Gly Leu Ile Arg Asp Met Arg Phe Asp
    290                 295                 300

Asn Phe Arg Gly Val Asn Tyr Ser Asn Asp Ala Cys Lys Lys Asp Lys
305                 310                 315                 320

Gln Ala Lys Ala Lys Tyr Gln Ala Ile Ile Ser Leu Tyr Leu Met Val
                325                 330                 335
```

```
Leu Tyr Gln Ile Val Lys Asn Met Ile Tyr Val Asn Ser Arg Tyr Val
            340                 345                 350


Ile Ala Phe His Cys Leu Glu Arg Asp Leu Leu Phe Phe Asn Ile Glu
        355                 360                 365


Leu Asp Asn Ser Tyr Gln Tyr Ser Asn Cys Asn Glu Leu Thr Glu Lys
    370                 375                 380


Phe Ile Lys Asp Lys Tyr Met Lys Glu Gly Ala Leu Gly Phe Asn Met
385                 390                 395                 400


Lys Ala Gly Arg Tyr Leu Thr Lys Asn Ile Gly Asn Cys Ser Asn Glu
                405                 410                 415


Leu Arg Lys Ile Tyr Arg Asn Gln Val Asp His Phe Ala Val Val Arg
            420                 425                 430


Lys Ile Gly Asn Tyr Ala Ala Asp Ile Ala Ser Val Gly Ser Trp Phe
        435                 440                 445


Glu



<210>  72
<211>  96
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Gut_metagenome_contig10037000527

<400>  72

Tyr Met Asp Gln Asn Phe Ala Asn Ser Asp Ala Trp Ala Ile His Val
1               5                   10                  15


Tyr Arg Asn Lys Ile Gln His Leu Asp Ala Val Arg His Ala Asp Met
            20                  25                  30
```

Tyr Ile Gly Asp Ile Arg Glu Phe His Ser Trp Phe Glu Leu Tyr His
        35   40   45

Tyr Ile Ile Gln Arg Arg Ile Ile Asp Gln Tyr Ala Tyr Glu Ser Thr
    50   55   60

Pro Gly Ser Ser Arg Asp Gly Ser Ala Ile Ile Asp Glu Glu Arg Leu
65   70   75   80

Asn Pro Ala Thr Arg Arg Tyr Phe Arg Leu Ile Thr Thr Tyr Lys Thr
                85   90   95

<210> 73
<211> 519
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig238000329

<400> 73

Arg Tyr Asp Lys Asp Arg Ser Lys Ile Tyr Thr Met Met Asp Phe Val
1   5   10   15

Ile Tyr Arg Tyr Tyr Ile Asp Asn Asn Asn Asp Ser Ile Asp Phe Ile
            20   25   30

Asn Lys Leu Arg Ser Ser Ile Asp Glu Lys Ser Lys Glu Lys Leu Tyr
        35   40   45

Asn Glu Glu Ala Asn Arg Leu Trp Asn Lys Leu Lys Glu Tyr Met Leu
    50   55   60

Tyr Ile Lys Glu Phe Asn Gly Lys Leu Ala Ser Arg Thr Pro Asp Arg
65   70   75   80

Asp Gly Asn Ile Ser Glu Phe Val Glu Ser Leu Pro Lys Ile His Arg
                85   90   95

Leu Leu Pro Arg Gly Gln Lys Ile Ser Asn Phe Ser Lys Leu Met Tyr
            100   105   110

```
Leu Leu Thr Met Phe Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr
        115                 120                 125

Thr Leu Ile Asn Lys Phe Glu Asn Ile Gln Gly Phe Leu Asp Ile Met
    130                 135                 140

Pro Glu Ile Asn Val Asn Ala Lys Phe Glu Pro Glu Tyr Val Phe Phe
145                 150                 155                 160

Asn Lys Ser His Glu Ile Ala Gly Glu Leu Lys Leu Ile Lys Gly Phe
                165                 170                 175

Ala Gln Met Gly Glu Pro Ala Ala Thr Leu Lys Leu Glu Met Thr Ala
            180                 185                 190

Asp Ala Ile Lys Ile Leu Gly Thr Glu Lys Glu Asp Ala Glu Leu Ile
        195                 200                 205

Lys Leu Ala Glu Ser Leu Phe Lys Asp Glu Asn Gly Lys Leu Leu Gly
    210                 215                 220

Asn Lys Gln His Gly Met Arg Asn Phe Ile Gly Asn Asn Val Ile Lys
225                 230                 235                 240

Ser Lys Arg Phe His Tyr Leu Ile Arg Tyr Gly Asp Pro Ala His Leu
                245                 250                 255

His Lys Ile Ala Thr Asn Lys Asn Val Val Arg Phe Val Leu Gly Arg
            260                 265                 270

Ile Ala Asp Met Gln Lys Lys Gln Gly Gln Lys Gly Lys Asn Gln Ile
        275                 280                 285

Asp Arg Tyr Tyr Glu Val Cys Val Gly Asn Lys Asp Ile Lys Lys Thr
    290                 295                 300

Ile Glu Glu Lys Ile Asp Ala Leu Thr Asp Ile Ile Val Asn Met Asn
305                 310                 315                 320
```

```
Tyr Asp Gln Phe Glu Lys Lys Lys Ala Val Ile Glu Asn Gln Asn Arg
                325                 330                 335

Gly Lys Thr Phe Glu Glu Lys Asn Lys Tyr Lys Arg Asp Asn Ala Glu
            340                 345                 350

Arg Glu Lys Phe Lys Lys Ile Ile Ser Leu Tyr Leu Thr Val Ile Tyr
        355                 360                 365

His Ile Leu Lys Asn Ile Val Asn Val Asn Ser Arg Tyr Ile Leu Gly
    370                 375                 380

Phe His Cys Leu Glu Arg Asp Lys Gln Leu Tyr Ile Glu Lys Tyr Asn
385                 390                 395                 400

Lys Asp Lys Leu Asp Gly Phe Val Ala Leu Thr Lys Phe Cys Leu Gly
                405                 410                 415

Asp Glu Glu Arg Tyr Glu Asp Leu Lys Ala Lys Ala Gln Ala Ser Ile
            420                 425                 430

Gln Ala Leu Glu Thr Ala Asn Pro Lys Leu Tyr Ala Lys Tyr Met Asn
        435                 440                 445

Tyr Ser Asp Glu Glu Lys Lys Glu Glu Phe Lys Lys Gln Leu Asn Arg
    450                 455                 460

Glu Arg Val Lys Asn Ala Arg Asn Ala Tyr Leu Lys Asn Ile Lys Asn
465                 470                 475                 480

Tyr Ile Met Ile Arg Leu Gln Leu Arg Asp Gln Thr Asp Ser Ser Gly
                485                 490                 495

Tyr Leu Cys Gly Glu Phe Arg Asp Lys Val Ala His Leu Glu Val Ala
            500                 505                 510

Arg His Ala His Glu Tyr Ile
        515
```

<210> 74
<211> 181
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig2643000492

<400> 74

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

Ile Ala Asp Lys Asn Ile Gly Cys Lys Ile Glu Asn Lys Gln Phe Arg
20 25 30

His Pro Lys Gly Tyr Asp Val Ile Ala Asp Asn Pro Ile Tyr Lys Gly
35 40 45

Ser Pro Arg Gln Asp Met Leu Gly Leu Lys Glu Thr Leu Glu Lys Arg
50 55 60

Tyr Phe Ser Pro Ser Asp Ser Ile Asp Asn Val Arg Val Gln Val Ala
65 70 75 80

His Asn Ile Leu Asp Ile Glu Lys Ile Leu Ala Glu Tyr Ile Thr Asn
85 90 95

Ala Val Tyr Ser Phe Asp Asn Ile Ala Gly Phe Gly Lys Asp Ile Ile
100 105 110

Gly Asp Asp Phe Ser Pro Val Tyr Thr Tyr Asp Lys Phe Glu Lys Ser
115 120 125

Asp Arg Tyr Glu Tyr Phe Lys Asn Leu Leu Asn Asn Ser Arg Leu Gly
130 135 140

Tyr Tyr Gly Gln Ala Phe Phe Glu Cys Asp Asp Ser Lys Glu Asn Lys
145 150 155 160

Lys Lys Lys Asp Ala Ile Lys Cys Tyr Asn Ile Ile Ala Leu Leu Ser
165 170 175

Gly Leu Arg His Trp
180

<210> 75
<211> 440
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig874000057

<400> 75

Met Ser Lys Asn Lys Glu Ser Tyr Ala Lys Gly Met Gly Leu Lys Ser
1 5 10 15

Ala Leu Val Ser Gly Ser Lys Val Tyr Met Thr Ser Phe Glu Gly Gly
20 25 30

Asn Asp Ala Lys Leu Glu Lys Val Val Glu Asn Ser Glu Ile Val Ser
35 40 45

Leu Ala Glu Lys Glu Ser Phe Ser Ala Glu Ile Phe Lys Lys Asn Ile
50 55 60

Gly Cys Lys Ile Glu Asn Lys Lys Phe Lys His Pro Lys Arg Tyr Asp
65 70 75 80

Val Ile Ala Asp Asn Pro Leu Tyr Lys Gly Ser Val Arg Gln Asp Met
85 90 95

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

Gly Thr Asp Asn Val Cys Ile Gln Val Ile His Asn Ile Leu Asp Ile
115 120 125

Glu Lys Ile Leu Ala Glu Tyr Ile Thr Asn Ala Val Tyr Ser Phe Asp
130 135 140

```
Asn Ile Ala Gly Phe Gly Glu Asp Ile Ile Gly Met Gly Gly Phe Lys
145                 150                 155                 160

Pro Ile Tyr Thr Tyr Lys Gln Phe Lys Glu Pro Asp Lys Tyr Asn Lys
                165                 170                 175

Lys Phe Asp Asp Ile Leu Asn Asn Ser Arg Leu Gly Tyr Tyr Gly Lys
            180                 185                 190

Ala Phe Phe Glu Lys Asn Asp Leu Lys His Asn Pro Asn Lys Lys Lys
        195                 200                 205

Arg Asp Lys Asn Pro Tyr Ile Leu Lys Tyr Asp Asn Glu Cys Tyr Tyr
    210                 215                 220

Ile Ile Ala Leu Leu Ser Gly Leu Arg His Trp Asn Ile His Ser His
225                 230                 235                 240

Ala Lys Asp Asp Leu Val Ser Tyr Arg Trp Leu Tyr Asn Leu Asp Ser
                245                 250                 255

Ile Leu Asn Arg Glu Tyr Ile Ser Thr Leu Asn Tyr Leu Tyr Asp Asp
            260                 265                 270

Ile Ala Asp Glu Leu Thr Glu Ser Phe Ser Lys Asn Ser Ser Ala Asn
        275                 280                 285

Val Asn Tyr Ile Ala Glu Thr Leu Asn Ile Asp Pro Ser Glu Phe Ala
    290                 295                 300

Gln Gln Tyr Phe Arg Phe Ser Ile Met Lys Glu Gln Lys Asn Met Gly
305                 310                 315                 320

Phe Asn Val Ser Lys Leu Arg Glu Ile Met Leu Asp Arg Lys Glu Leu
                325                 330                 335

Ser Asp Ile Arg Asp Asn His Arg Val Phe Asp Ser Ile Arg Ser Lys
            340                 345                 350
```

```
Leu Tyr Thr Met Met Asp Phe Val Ile Tyr Arg Tyr Tyr Ile Glu Glu
        355                 360                 365


Ala Ala Lys Thr Glu Ala Glu Asn Arg Asn Leu Pro Glu Asn Glu Lys
    370                 375                 380


Lys Ile Ser Glu Lys Asp Phe Phe Val Ile Asn Leu Arg Gly Ser Phe
385                 390                 395                 400


Asp Glu Asn Gln Lys Glu Lys Leu Tyr Ile Glu Glu Ala Lys Arg Leu
                405                 410                 415


Trp Glu Lys Leu Lys Asp Ile Met Leu Lys Ile Lys Glu Phe Arg Gly
            420                 425                 430


Glu Lys Val Lys Glu Tyr Lys Lys
        435                 440


<210>  76
<211>  137
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Gut_metagenome_contig4781000489

<400>  76

Leu Asp Lys Gln Leu Asp Tyr Glu Tyr Ile Arg Thr Leu Asn Tyr Met
1               5                   10                  15


Phe Asn Asp Ile Ala Asp Glu Leu Thr Arg Thr Phe Ser Lys Asn Ser
            20                  25                  30


Ala Ala Asn Val Asn Tyr Ile Ala Glu Thr Leu Asn Ile Asp Pro Asn
        35                  40                  45


Lys Phe Ala Glu Gln Tyr Phe Arg Phe Ser Ile Met Lys Glu Gln Lys
    50                  55                  60
```

```
Asn Leu Gly Phe Asn Leu Thr Lys Leu Arg Glu Ser Met Leu Asp Arg
65                  70                  75                  80


Arg Glu Leu Ser Asp Ile Arg Asp Asn His Asn Val Phe Asp Ser Ile
                85                  90                  95


Arg Pro Lys Leu Tyr Thr Met Met Asp Phe Val Ile Tyr Lys His Tyr
            100                 105                 110


Ile Asp Glu Ala Lys Lys Thr Glu Ala Glu Asn Lys Ser Leu Pro Asp
        115                 120                 125


Asp Arg Lys Asn Leu Ser Glu Lys Asp
    130                 135


<210>  77
<211>  87
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Gut_metagenome_contig12144000352

<400>  77

Arg Met Gly Glu Pro Val Ala Asn Thr Lys Arg Val Met Met Ile Asp
1               5                   10                  15


Ala Val Lys Ile Leu Gly Thr Asp Leu Ser Asp Asp Glu Leu Lys Glu
            20                  25                  30


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


Gly Lys His Gly Met Arg Asn Phe Ile Thr Asn Asn Val Ile Lys Asn
    50                  55                  60


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


Glu Ile Ala Lys Asn Glu Ala
                85
```

<210> 78
<211> 414
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig5590000448

<400> 78

```
Val His Asn Asn Glu Glu Lys Asp Leu Ile Lys Tyr Thr Trp Leu Tyr
1               5                   10                  15


Asn Leu Asp Lys Tyr Leu Asp Ala Glu Tyr Ile Thr Thr Leu Asn Tyr
            20                  25                  30


Met Tyr Asn Asp Ile Gly Asp Glu Leu Thr Asp Ser Phe Ser Lys Asn
        35                  40                  45


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


Lys Thr Phe Ala Glu Gln Tyr Phe Arg Phe Ser Ile Met Lys Glu Gln
65                  70                  75                  80


Lys Asn Leu Gly Phe Asn Leu Thr Lys Leu Arg Glu Val Met Leu Asp
                85                  90                  95


Arg Lys Asp Met Ser Glu Ile Arg Glu Asn His Asn Asp Phe Asp Ser
            100                 105                 110


Ile Arg Ala Lys Val Tyr Thr Met Met Asp Phe Val Ile Tyr Arg Tyr
        115                 120                 125


Tyr Ile Glu Glu Ala Ala Lys Val Asn Ala Ala Asn Lys Ser Leu Pro
    130                 135                 140


Asp Asn Glu Lys Ser Leu Ser Glu Lys Asp Ile Phe Val Ile Ser Leu
145                 150                 155                 160
```

```
Arg Gly Ser Phe Asn Glu Asp Gln Lys Asp Arg Leu Tyr Tyr Asp Glu
                165                 170                 175

Ala Gln Arg Leu Trp Ser Lys Val Gly Lys Leu Met Leu Lys Ile Lys
            180                 185                 190

Lys Phe Arg Gly Lys Asp Thr Arg Lys Tyr Lys Asn Met Gly Thr Pro
        195                 200                 205

Arg Ile Arg Arg Leu Ile Pro Glu Gly Arg Asp Ile Ser Thr Phe Ser
    210                 215                 220

Lys Leu Met Tyr Ala Leu Thr Met Phe Leu Asp Gly Lys Glu Ile Asn
225                 230                 235                 240

Asp Leu Leu Thr Thr Leu Ile Asn Lys Phe Asp Asn Ile Gln Ser Phe
                245                 250                 255

Leu Lys Val Met Pro Leu Ile Gly Val Asn Ala Lys Phe Ala Glu Glu
            260                 265                 270

Tyr Ser Phe Phe Asn Asn Ser Glu Lys Ile Ala Asp Glu Leu Arg Leu
        275                 280                 285

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

Ala Met Tyr Ile Asp Ala Ile Arg Ile Leu Gly Thr Asp Leu Ser Asp
305                 310                 315                 320

Asp Glu Leu Lys Ala Leu Ala Asp Ser Phe Ser Leu Asp Glu Asn Gly
                325                 330                 335

Asn Lys Leu Gly Lys Gly Lys His Gly Met Arg Asn Phe Ile Ile Asn
            340                 345                 350

Asn Val Ile Thr Asn Lys Arg Phe His Tyr Leu Ile Arg Tyr Gly Asn
        355                 360                 365
```

Pro Val His Leu His Glu Ile Ala Lys Asn Glu Ala Val Val Lys Phe
370 375 380

Val Leu Gly Arg Ile Ala Asp Ile Gln Lys Lys Gln Gly Gln Asn Gly
385 390 395 400

Lys Asn Gln Ile Asp Arg Tyr Tyr Glu Thr Cys Ile Gly Lys
405 410

<210> 79
<211> 109
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig9269000031

<400> 79

Leu Ser Glu Lys Asp Ile Phe Val Ile Ser Leu Arg Gly Ser Phe Asn
1 5 10 15

Glu Asp Gln Lys Asp Arg Leu Tyr Tyr Asp Glu Ala Gln Arg Leu Trp
20 25 30

Ser Lys Leu Gly Lys Leu Met Leu Lys Ile Lys Glu Phe Lys Gly Lys
35 40 45

Asp Thr Lys Lys Tyr Arg Asp Met Gly Lys Pro Arg Leu Arg Arg Ile
50 55 60

Leu Pro Glu Gly Arg Asp Ile Ser Thr Phe Ser Lys Leu Met Tyr Ala
65 70 75 80

Leu Thr Met Phe Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr Thr
85 90 95

Leu Ile Asn Lys Phe Asp Asn Ile Gln Ser Phe Leu Lys
100 105

<210> 80
<211> 105

<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig8537000520

<400> 80

Ser Ala Arg Leu Ala Asp Glu Leu Arg Ile Val Lys Ser Val Ser Arg
1 5 10 15

Met Glu Met Pro Met Pro Ala Ala Lys Thr Leu Met Phe Glu Asp Ala
20 25 30

Leu Lys Leu Phe Trp Ile Asp Ser Pro Glu Glu Ala Lys Lys Ile Ile
35 40 45

Ser Gly Asp Asn Lys Lys Ile Arg Asn Phe Ile Ala Asn Asn Val Val
50 55 60

Glu Ser Ser Trp Phe Asn Tyr Ile Val Arg Tyr Thr Asp Leu Lys Asn
65 70 75 80

Ala Arg Asp Phe Ile Arg Cys Arg Pro Ile Val Glu Phe Val Leu Asn
85 90 95

Lys Val Pro Glu Thr Gln Ile Asp Arg
100 105

<210> 81
<211> 211
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig1845000130

<400> 81

Ser Ser Leu Asn Glu Val Glu Lys Asp Ala Ile Tyr Ala Glu Glu Ala
1 5 10 15

Lys Gln Ala Trp Val Asn Ile Arg Asp Leu Val Ile Arg His Ile Leu
20 25 30

```
Pro Met Met Lys Gly Asp Tyr Ile Ala Gln Ile Ser Val Glu Asp Glu
        35                  40                  45

Lys Ile Asp Gln Ser Ile Leu Glu Asp Val Phe Ile Ser Lys Ile Thr
    50                  55                  60

Pro Ser Ala Leu Ala Glu Phe Ile Tyr Ile Val Thr Leu Phe Ile Asp
65                  70                  75                  80

Gly Lys Glu Ile Asn Asp Leu Leu Thr Gln Met Ile His Lys Phe Asp
                85                  90                  95

Ser Ile Ala Ala Phe Gln Glu Val Met Arg Gln Arg Cys Met Val Thr
            100                 105                 110

Gln Phe Asn Gln Asp Tyr Ala Ile Phe Glu Gln Ser Gly Lys Ile Ala
        115                 120                 125

Glu Asp Leu Arg Val Val Asn Ser Phe Ala Arg Met Ser Ile Ser Lys
    130                 135                 140

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

Leu Gly Leu Ala Ser Pro Asp Glu Thr Met Glu Ser Val Leu Gln Lys
                165                 170                 175

Tyr Met Asn Ser Asn Leu Asp Arg Glu Arg Lys Ser Val Arg Asn Phe
            180                 185                 190

Val Ile Asn Asn Val Ile Gln Ser Arg Arg Phe His Tyr Leu Val Arg
        195                 200                 205

Tyr Gly Asn
    210
```

<210> 82
<211> 906

<212> PRT
<213> Artificial Sequence

<220>
<223> gut_metagenome_P13E0k2120140920_c3000072

<400> 82

Met Glu Arg Glu Val Lys Lys Pro Pro Lys Lys Ser Leu Ala Lys Ala
1 5 10 15

Ala Gly Leu Lys Ser Thr Phe Val Ile Ser Pro Gln Glu Lys Glu Leu
20 25 30

Ala Met Thr Ala Phe Gly Arg Gly Asn Asp Ala Leu Leu Gln Lys Arg
35 40 45

Ile Val Asp Gly Val Val Arg Asp Val Ala Gly Glu Lys Gln Gln Phe
50 55 60

Gln Val Gln Arg Gln Asp Glu Ser Arg Phe Arg Leu Gln Asn Ser Arg
65 70 75 80

Leu Ala Asp Arg Thr Val Thr Ala Asp Asp Pro Leu His Arg Ala Glu
85 90 95

Thr Pro Arg Arg Gln Pro Leu Gly Ala Gly Met Asp Gln Leu Arg Arg
100 105 110

Lys Ala Ile Leu Glu Gln Lys Tyr Phe Gly Arg Thr Phe Asp Asp Asn
115 120 125

Ile His Ile Gln Leu Ile Tyr Asn Ile Leu Asp Ile His Lys Met Leu
130 135 140

Ala Val Pro Ala Asn His Ile Val His Thr Leu Asn Leu Leu Gly Gly
145 150 155 160

Tyr Gly Glu Thr Asp Phe Val Gly Met Leu Pro Ala Gly Leu Pro Tyr
165 170 175

```
Asp Lys Leu Arg Val Val Lys Lys Lys Asn Gly Asp Thr Val Asp Ile
            180                 185                 190

Lys Ala Asp Ile Ala Ala Tyr Ala Lys Arg Pro Gln Leu Ala Tyr Leu
        195                 200                 205

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

Ala Arg Gly Arg Val Lys Arg Glu Gln Asp Val Tyr Thr Ile Leu Ser
225                 230                 235                 240

Leu Met Ser Leu Leu Arg Gln Phe Cys Ala His Asp Ser Val Arg Ile
                245                 250                 255

Trp Gly Gln Asn Thr Pro Ala Ala Leu Tyr Gly Leu Gln Ala Leu Pro
            260                 265                 270

Gln Asp Met Lys Asp Leu Leu Asp Asp Gly Trp Arg Arg Ala Leu Gly
        275                 280                 285

Gly Val Asn Asp His Phe Leu Asp Thr Asn Lys Val Asn Leu Leu Thr
    290                 295                 300

Leu Phe Glu Tyr Tyr Gly Ala Glu Thr Lys Gln Glu Arg Val Ala Leu
305                 310                 315                 320

Thr Gln Asp Phe Tyr Arg Phe Val Val Leu Lys Glu Gln Lys Asn Met
                325                 330                 335

Gly Phe Ser Leu Arg Arg Leu Arg Glu Glu Leu Leu Lys Leu Pro Asp
            340                 345                 350

Ala Ala Tyr Leu Thr Gly Gln Glu Tyr Asp Ser Val Arg Gln Lys Leu
        355                 360                 365

Tyr Met Leu Leu Asp Phe Leu Leu Cys Arg Leu Tyr Ala Gln Glu Arg
    370                 375                 380
```

```
Ala Asp Arg Cys Glu Glu Leu Val Ser Ala Leu Arg Cys Ala Leu Ser
385                 390                 395                 400


Asp Glu Glu Lys Asp Ala Val Tyr Gln Ala Glu Ala Ala Ala Leu Trp
                405                 410                 415


Gln Ala Leu Gly Asp Thr Leu Arg Arg Glu Leu Leu Pro Leu Leu Lys
            420                 425                 430


Gly Lys Lys Leu Gln Asp Lys Asp Lys Lys Lys Leu Asp Glu Leu Gly
        435                 440                 445


Leu Ser Arg Asp Val Leu Asp Gly Val Leu Phe Arg Pro Ala Gln Gln
    450                 455                 460


Gly Ser Arg Ala Asn Ala Asp Tyr Phe Cys Arg Leu Met His Leu Ser
465                 470                 475                 480


Thr Trp Phe Met Asp Gly Lys Glu Ile Asn Thr Leu Leu Thr Thr Leu
                485                 490                 495


Ile Ser Lys Leu Glu Asn Ile Asp Ser Leu Arg Ser Val Leu Glu Ser
            500                 505                 510


Met Gly Leu Ala Tyr Ser Phe Val Pro Ala Tyr Ala Met Phe Asp His
        515                 520                 525


Ser Arg Tyr Ile Ala Gly Gln Leu Arg Val Val Asn Asn Ile Ala Arg
    530                 535                 540


Met Arg Lys Pro Ala Ile Gly Ala Lys Arg Glu Met Tyr Arg Ala Ala
545                 550                 555                 560


Val Val Leu Leu Gly Val Asp Ser Pro Glu Ala Ala Ala Ala Ile Thr
                565                 570                 575


Asp Asp Leu Leu Gln Ile Asp Pro Glu Thr Gly Lys Val Arg Pro Arg
            580                 585                 590
```

```
Gly Asp Ser Ala Arg Asp Thr Gly Leu Arg Asn Phe Val Ala Asn Asn
        595                 600                 605

Val Val Glu Ser Arg Arg Phe Thr Tyr Leu Leu Arg Tyr Met Thr Pro
    610                 615                 620

Glu Gln Ala Arg Val Leu Ala Gln Asn Glu Lys Leu Ile Ala Phe Val
625                 630                 635                 640

Leu Ser Thr Val Pro Ser Ala Gln Leu Glu Arg Tyr Cys Arg Thr Cys
                645                 650                 655

Gly Arg Glu Asp Ile Thr Gly Arg Pro Ala Gln Ile Arg Tyr Leu Thr
            660                 665                 670

Ala Gln Ile Met Gly Val Arg Tyr Glu Ser Phe Thr Asp Val Glu Gln
        675                 680                 685

Arg Gly Arg Gly Asp Asn Pro Lys Lys Glu Arg Tyr Lys Ala Leu Ile
    690                 695                 700

Gly Leu Tyr Leu Thr Val Leu Tyr Leu Ala Val Lys Asn Met Val Asn
705                 710                 715                 720

Cys Asn Ala Arg Tyr Val Ile Ala Phe Tyr Cys Arg Asp Arg Asp Thr
                725                 730                 735

Ala Leu Tyr Gln Lys Glu Val Cys Trp Tyr Asp Leu Glu Glu Asp Lys
            740                 745                 750

Lys Ser Gly Lys Gln Arg Gln Val Glu Asp Tyr Thr Ala Leu Thr Arg
        755                 760                 765

Tyr Phe Val Ser Gln Gly Tyr Leu Asn Arg His Ala Cys Gly Tyr Leu
    770                 775                 780

Arg Ser Asn Met Asn Gly Ile Ser Asn Gly Leu Leu Ala Ala Tyr Arg
785                 790                 795                 800
```

Asn Ala Val Asp His Leu Asn Val Ile Pro Pro Leu Gly Ser Leu Cys
805 810 815

Arg Asp Ile Gly Arg Val Asp Ser Tyr Phe Ala Leu Tyr His Tyr Ala
820 825 830

Val Gln Gln Tyr Leu Asn Gly Arg Tyr Tyr Arg Lys Thr Pro Arg Glu
835 840 845

Gln Glu Leu Phe Ala Ala Met Ala Gln His Arg Thr Trp Cys Ser Asp
850 855 860

Leu Val Lys Ala Leu Asn Thr Pro Phe Gly Tyr Asn Leu Ala Arg Tyr
865 870 875 880

Lys Asn Leu Ser Ile Asp Gly Leu Phe Asp Arg Glu Gly Asp His Val
885 890 895

Val Arg Glu Asp Gly Glu Lys Pro Ala Glu
900 905

<210> 83
<211> 906
<212> PRT
<213> Artificial Sequence

<220>
<223> gut_metagenome_P1E0k2120140920_c1000078

<400> 83

Met Glu Arg Glu Val Lys Lys Pro Pro Lys Lys Ser Leu Ala Lys Ala
1 5 10 15

Ala Gly Leu Lys Ser Thr Phe Val Ile Ser Pro Gln Glu Lys Glu Leu
20 25 30

Ala Met Thr Ala Phe Gly Arg Gly Asn Asp Ala Leu Leu Gln Lys Arg
35 40 45

Ile Val Asp Gly Val Val Arg Asp Val Ala Gly Glu Lys Gln Gln Phe
50 55 60

```
Gln Val Gln Arg Gln Asp Glu Ser Arg Phe Arg Leu Gln Asn Ser Arg
65                  70                  75                  80

Leu Ala Asp Arg Thr Val Thr Ala Asp Asp Pro Leu His Arg Ala Glu
                85                  90                  95

Thr Pro Arg Arg Gln Pro Leu Gly Ala Gly Met Asp Gln Leu Arg Arg
            100                 105                 110

Lys Ala Ile Leu Glu Gln Lys Tyr Phe Gly Arg Thr Phe Asp Asp Asn
        115                 120                 125

Ile His Ile Gln Leu Ile Tyr Asn Ile Leu Asp Ile His Lys Met Leu
    130                 135                 140

Ala Val Pro Ala Asn His Ile Val His Thr Leu Asn Leu Leu Gly Gly
145                 150                 155                 160

Tyr Gly Glu Thr Asp Phe Val Gly Met Leu Pro Ala Gly Leu Pro Tyr
                165                 170                 175

Asp Lys Leu Arg Val Val Lys Lys Lys Asn Gly Asp Thr Val Asp Ile
            180                 185                 190

Lys Ala Asp Ile Ala Ala Tyr Ala Lys Arg Pro Gln Leu Ala Tyr Leu
        195                 200                 205

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

Ala Arg Gly Arg Val Lys Arg Glu Gln Asp Val Tyr Ala Ile Leu Ser
225                 230                 235                 240

Leu Met Ser Leu Leu Arg Gln Phe Cys Ala His Asp Ser Val Arg Ile
                245                 250                 255

Trp Gly Gln Asn Thr Thr Ala Ala Leu Tyr His Leu Gln Ala Leu Pro
            260                 265                 270
```

```
Gln Asp Met Lys Asp Leu Leu Asp Asp Gly Trp Arg Arg Ala Leu Gly
        275                 280                 285

Gly Val Asn Asp His Phe Leu Asp Thr Asn Lys Val Asn Leu Leu Thr
    290                 295                 300

Leu Phe Glu Tyr Tyr Gly Ala Glu Thr Lys Gln Ala Arg Val Ala Leu
305                 310                 315                 320

Thr Gln Asp Phe Tyr Arg Phe Val Val Leu Lys Glu Gln Lys Asn Met
                325                 330                 335

Gly Phe Ser Leu Arg Arg Leu Arg Glu Glu Leu Leu Lys Leu Pro Asp
            340                 345                 350

Ala Ala Tyr Leu Thr Gly Gln Glu Tyr Asp Ser Val Arg Gln Lys Leu
        355                 360                 365

Tyr Met Leu Leu Asp Phe Leu Leu Cys Arg Leu Tyr Ala Gln Glu Arg
    370                 375                 380

Ala Asp Arg Cys Glu Glu Leu Val Ser Ala Leu Arg Cys Ala Leu Ser
385                 390                 395                 400

Asp Glu Glu Lys Asp Thr Val Tyr Gln Ala Glu Ala Ala Ala Leu Trp
                405                 410                 415

Gln Ala Leu Gly Asp Thr Leu Arg Arg Lys Leu Leu Pro Leu Leu Lys
            420                 425                 430

Gly Lys Lys Leu Gln Asp Lys Asp Lys Lys Lys Ser Asp Glu Leu Gly
        435                 440                 445

Leu Ser Arg Asp Val Leu Asp Gly Val Leu Phe Arg Pro Ala Gln Gln
    450                 455                 460

Gly Ser Arg Ala Asn Ala Asp Tyr Phe Cys Arg Leu Met His Leu Ser
465                 470                 475                 480
```

```
Thr Trp Phe Met Asp Gly Lys Glu Ile Asn Thr Leu Leu Thr Thr Leu
                485                 490                 495

Ile Ser Lys Leu Glu Asn Ile Asp Ser Leu Arg Ser Val Leu Glu Ser
            500                 505                 510

Met Gly Leu Ala Tyr Ser Phe Val Pro Ala Tyr Ala Met Phe Asp His
        515                 520                 525

Ser Arg Tyr Ile Ala Gly Gln Leu Arg Val Val Asn Asn Ile Ala Arg
    530                 535                 540

Met Arg Lys Pro Ala Ile Gly Ala Lys Arg Glu Met Tyr Arg Ala Ala
545                 550                 555                 560

Val Val Leu Leu Gly Val Asp Ser Pro Glu Ala Ala Ala Ala Ile Thr
                565                 570                 575

Asp Asp Leu Leu Gln Ile Asp Pro Glu Thr Gly Lys Val Arg Pro Arg
            580                 585                 590

Ser Asp Ser Ala Arg Asp Thr Gly Leu Arg Asn Phe Ile Ala Asn Asn
        595                 600                 605

Val Val Glu Ser Arg Arg Phe Thr Tyr Leu Leu Arg Tyr Met Thr Pro
    610                 615                 620

Glu Gln Ala Arg Val Leu Ala Gln Asn Glu Lys Leu Ile Ala Phe Val
625                 630                 635                 640

Leu Ser Thr Val Pro Asp Thr Gln Leu Glu Arg Tyr Cys Arg Thr Cys
                645                 650                 655

Gly Arg Glu Asp Ile Thr Gly Arg Pro Ala Gln Ile Arg Tyr Leu Thr
            660                 665                 670

Ala Gln Ile Met Gly Val Arg Tyr Glu Ser Phe Thr Asp Val Glu Gln
        675                 680                 685
```

```
Arg Gly Arg Gly Asp Asn Pro Lys Lys Glu Arg Tyr Lys Ala Leu Ile
    690                 695                 700

Gly Leu Tyr Leu Thr Val Leu Tyr Leu Ala Val Lys Asn Met Val Asn
705                 710                 715                 720

Cys Asn Ala Arg Tyr Val Ile Ala Phe Tyr Cys Arg Asp Arg Asp Thr
                725                 730                 735

Ala Leu Tyr Gln Lys Glu Val Cys Trp Tyr Asp Leu Glu Glu Asp Lys
            740                 745                 750

Lys Ser Gly Lys Gln Arg Gln Val Glu Asp Tyr Thr Ala Leu Thr Arg
        755                 760                 765

Tyr Phe Val Ser Gln Gly Tyr Leu Asn Arg His Ala Cys Gly Tyr Leu
    770                 775                 780

Arg Ser Asn Met Asn Gly Ile Ser Asn Ser Leu Leu Thr Ala Tyr Arg
785                 790                 795                 800

Asn Ala Val Asp His Leu Asn Ala Ile Pro Pro Leu Gly Ser Leu Cys
                805                 810                 815

Arg Asp Ile Gly Arg Val Asp Ser Tyr Phe Ala Leu Tyr His Tyr Ala
            820                 825                 830

Val Gln Gln Tyr Leu Asn Gly Arg Tyr Tyr Arg Lys Thr Pro Arg Glu
        835                 840                 845

Gln Glu Leu Phe Ala Ala Met Ala Gln His Arg Thr Trp Cys Ser Asp
    850                 855                 860

Leu Val Lys Ala Leu Asn Thr Pro Phe Gly Tyr Asn Leu Ala Arg Tyr
865                 870                 875                 880

Lys Asn Leu Ser Ile Asp Gly Leu Phe Asp Arg Glu Gly Asp His Val
                885                 890                 895
```

Val Arg Glu Asp Gly Glu Lys Pro Ala Glu
900 905

<210> 84
<211> 84
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig12990000099

<400> 84

Ile Asn Asp Leu Leu Thr Thr Leu Ile His Lys Phe Glu Asn Ile Ala
1 5 10 15

Ala Phe Leu His Val Ile Gln Ala Glu Lys Leu Pro Val Glu Phe Val
20 25 30

Pro Glu Phe Gln Val Phe Ser Asp Ser Gln Lys Ile Ala Glu Glu Leu
35 40 45

Arg Ile Ile Asn Ser Phe Ala Arg Met Asn Lys Pro Ser Ala Asn Ala
50 55 60

Arg Glu Ile Met Tyr Ile Glu Ala Phe Lys Val Leu Gly Met Ser Gly
65 70 75 80

Asp Glu Lys Gln

<210> 85
<211> 345
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig525000349

<400> 85

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

```
Lys Ser Thr Leu Val Ser Asp Ser Lys Val Tyr Leu Thr Thr Phe Ala
            20                  25                  30

Asp Gly Ser Asn Ala Lys Leu Glu Lys Cys Val Glu Asn Asn Lys Ile
        35                  40                  45

Ile Cys Ile Ser Asn Asp Lys Glu Ala Phe Ala Ala Ser Ile Ala Asn
    50                  55                  60

Lys Asn Val Gly Tyr Lys Ile Lys Asn Asp Glu Lys Phe Arg His Pro
65                  70                  75                  80

Lys Gly Tyr Asp Ile Ile Ser Asn Asn Pro Leu Leu His Asn Asn Ser
                85                  90                  95

Val Gln Gln Asp Met Leu Gly Leu Lys Asn Val Leu Glu Lys Arg Tyr
            100                 105                 110

Phe Gly Lys Ser Ser Gly Gly Asp Asn Asn Leu Cys Ile Gln Ile Ile
        115                 120                 125

His Asn Ile Ile Asp Ile Glu Lys Ile Leu Ser Glu Tyr Ile Pro Asn
    130                 135                 140

Val Val Tyr Ala Phe Asn Asn Ile Ala Gly Phe Lys Asp Glu His Asn
145                 150                 155                 160

Asn Ile Ile Asp Ile Ile Gly Thr Gln Thr Tyr Asn Ser Ser Tyr Thr
                165                 170                 175

Tyr Ala Asp Phe Ser Lys Asp Lys Ser Asp Lys Lys Tyr Ile Glu Phe
            180                 185                 190

Gln Lys Leu Leu Lys Asn Lys Arg Leu Gly Tyr Trp Gly Lys Ala Phe
        195                 200                 205

Phe Thr Gly Gln Gly Asn Asn Ala Lys Val Arg Gln Glu Asn Gln Cys
    210                 215                 220
```

```
Phe His Ile Ile Ala Leu Leu Ile Ser Leu Arg Asn Trp Ala Thr His
225                 230                 235                 240


Ser Asn Glu Leu Asp Lys His Thr Lys Arg Thr Trp Leu Tyr Lys Leu
                245                 250                 255


Asp Asp Thr Asn Ile Leu Asn Ala Glu Tyr Val Lys Thr Leu Asn Tyr
            260                 265                 270


Leu Tyr Asp Thr Ile Ala Asp Glu Leu Thr Lys Ser Phe Ser Lys Asn
        275                 280                 285


Gly Ala Val Asn Val Asn Tyr Leu Ala Lys Lys Tyr Asn Ile Lys Asp
    290                 295                 300


Asp Leu Pro Gly Phe Ser Glu Gln Tyr Phe Arg Phe Ser Ile Met Lys
305                 310                 315                 320


Glu Gln Lys Asn Leu Gly Phe Asn Ile Ser Lys Leu Arg Glu Asn Met
                325                 330                 335


Leu Asp Phe Lys Asp Met Ser Val Ile
            340                 345


<210>  86
<211>  206
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Gut_metagenome_contig7229000302

<400>  86

Lys Lys Ile Ser Ser Leu Thr Lys Phe Cys Leu Gly Glu Ser Asp Glu
1               5                   10                  15


Lys Lys Leu Lys Ala Leu Ala Lys Lys Ser Leu Glu Glu Leu Lys Thr
            20                  25                  30
```

```
Thr Asn Ser Lys Leu Tyr Glu Asn Tyr Ile Lys Tyr Ser Asp Glu Arg
        35                  40                  45

Lys Ala Glu Glu Ala Lys Arg Gln Ile Asn Arg Glu Arg Ala Lys Thr
    50                  55                  60

Ala Met Asn Ala His Leu Arg Asn Thr Lys Trp Asn Asp Ile Met Tyr
65                  70                  75                  80

Gly Gln Leu Lys Asp Leu Ala Asp Ser Lys Ser Arg Ile Cys Ser Glu
                85                  90                  95

Phe Arg Asn Lys Ala Ala His Leu Glu Val Ala Arg Tyr Ala His Met
            100                 105                 110

Tyr Ile Asn Asp Ile Ser Glu Val Lys Ser Tyr Phe Arg Leu Tyr His
        115                 120                 125

Tyr Ile Met Gln Arg Arg Ile Ile Asp Val Ile Glu Asn Asn Pro Lys
    130                 135                 140

Ala Lys Tyr Glu Gly Lys Val Lys Val Tyr Phe Glu Asp Val Lys Lys
145                 150                 155                 160

Asn Lys Lys Tyr Asn Lys Asn Leu Leu Lys Leu Met Cys Val Pro Phe
                165                 170                 175

Gly Tyr Cys Ile Pro Arg Phe Lys Asn Leu Ser Ile Glu Gln Met Phe
            180                 185                 190

Asp Met Asn Glu Thr Asp Asn Ser Asp Lys Lys Lys Glu Lys
        195                 200                 205
```

<210> 87
<211> 95
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig3227000343

<400> 87

Ile Gly Asp Ile Ser Glu Val Asn Ser Tyr Phe Gln Leu Tyr His Tyr
1 5 10 15

Ile Met Gln Arg Ile Leu Ile Asp Lys Ile Gly Ser Lys Thr Thr Gly
20 25 30

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

Asp Arg Leu Leu Lys Leu Leu Cys Ser Pro Leu Gly Tyr Cys Leu Thr
50 55 60

Arg Tyr Lys Asp Leu Ser Ile Glu Ala Leu Phe Asp Met Asn Glu Ala
65 70 75 80

Ala Lys Tyr Asp Lys Leu Asn Lys Glu Arg Lys Asn Lys Lys Lys
85 90 95

<210> 88
<211> 115
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig7030000469

<400> 88

Ser Ile Arg Ser Lys Leu Tyr Thr Met Met Asp Phe Val Ile Tyr Arg
1 5 10 15

Tyr Tyr Ile Glu Glu Ser Ala Lys Ala Ala Ala Glu Asn Lys Pro Ser
20 25 30

Glu Ser Asp Ser Phe Val Ile Arg Leu Arg Gly Ser Phe Asn Glu Asn
35 40 45

Gln Lys Glu Glu Leu Tyr Ile Glu Glu Ala Glu Arg Leu Trp Lys Lys
50 55 60

Phe Gly Glu Ile Met Leu Lys Ile Lys Glu Phe Arg Gly Glu Lys Val
65 70 75 80

Lys Glu Tyr Lys Lys Glu Val Pro Arg Ile Glu Arg Ile Leu Pro His
85 90 95

Gly Lys Asp Ile Ser Ala Phe Ser Lys Leu Met Tyr Met Leu Ser Met
100 105 110

Phe Leu Asp
115

<210> 89
<211> 110
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig5149000068

<400> 89

Phe Arg Asn Lys Val Ala His Leu Glu Ala Val Arg Tyr Ala His Leu
1 5 10 15

Tyr Ile Asn Asp Ile Ala Glu Val Asn Ser Tyr Phe Gln Leu Tyr His
20 25 30

Tyr Ile Ile Gln Arg Arg Ile Met Ala Glu Cys Tyr Asp Lys Ser Ser
35 40 45

Gly Lys Val Arg Glu Tyr Phe Asn Ala Val Asn Lys Glu Lys Lys Tyr
50 55 60

Asn Asp Arg Leu Leu Lys Leu Leu Cys Val Pro Phe Gly Tyr Cys Val
65 70 75 80

Thr Arg Phe Lys Asp Leu Ser Ile Glu Ala Leu Phe Asp Lys Asn Glu
85 90 95

Ala Asp Arg Phe Asp Lys Glu Lys Lys Lys Ala Glu Lys Lys
100 105 110

<210> 90
<211> 84
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig400200045

<400> 90

Met Ile Glu Lys Lys Lys Ser Tyr Ala Lys Gly Met Gly Leu Lys Ser
1 5 10 15

Thr Ile Val Ser Gly Ser Lys Val Tyr Met Thr Thr Phe Gly Asp Gly
20 25 30

Ser Glu Ala Arg Leu Glu Lys Val Val Glu Asn Asp Ser Ile Lys Thr
35 40 45

Leu His Glu Gly Glu Ala Phe Ser Ala Glu Met Thr Asp Asn Asn Ala
50 55 60

Gly Tyr Lys Ile Gly Asn Val Lys Phe Ser His Pro Asn Gly Tyr Asp
65 70 75 80

Val Val Ala Asn

<210> 91
<211> 94
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig10420000446

<400> 91

Ser Ser Lys Ser Val Ala Glu Lys Val Asn Ala Leu Thr Lys Ile Ile
1 5 10 15

Thr Gly Met Asn Tyr Asp Gln Phe Asp Ser Arg Arg Asn Val Ile Glu
20 25 30

```
Asn Thr Gly Ala Gly Asn Ala Glu Arg Glu Lys Tyr Lys Lys Ile Ile
        35                  40                  45


Ser Leu Tyr Leu Thr Val Ile Tyr His Ile Leu Lys Asn Ile Val Asn
    50                  55                  60


Ile Asn Ser Arg Tyr Val Ile Gly Phe His Cys Val Glu Arg Asp Ala
65                  70                  75                  80


Gln Leu Tyr Lys Glu Lys Gly Tyr Asp Ile Asn Leu Lys Lys
                85                  90


<210>  92
<211>  967
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  new_flavefaciens,_strain_XPD3002

<400>  92

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


Thr Leu Val Ser Gly Ser Lys Val Tyr Met Thr Thr Phe Ala Glu Gly
            20                  25                  30


Ser Asp Ala Arg Leu Glu Lys Ile Val Glu Gly Asp Ser Ile Arg Ser
        35                  40                  45


Val Asn Glu Gly Glu Ala Phe Ser Ala Glu Met Ala Asp Lys Asn Ala
    50                  55                  60


Gly Tyr Lys Ile Gly Asn Ala Lys Phe Ser His Pro Lys Gly Tyr Ala
65                  70                  75                  80


Val Val Ala Asn Asn Pro Leu Tyr Thr Gly Pro Val Gln Gln Asp Met
                85                  90                  95
```

```
Leu Gly Leu Lys Glu Thr Leu Glu Lys Arg Tyr Phe Gly Glu Ser Ala
            100                 105                 110

Asp Gly Asn Asp Asn Ile Cys Ile Gln Val Ile His Asn Ile Leu Asp
        115                 120                 125

Ile Glu Lys Ile Leu Ala Glu Tyr Ile Thr Asn Ala Ala Tyr Ala Val
    130                 135                 140

Asn Asn Ile Ser Gly Leu Asp Lys Asp Ile Ile Gly Phe Gly Lys Phe
145                 150                 155                 160

Ser Thr Val Tyr Thr Tyr Asp Glu Phe Lys Asp Pro Glu His His Arg
                165                 170                 175

Ala Ala Phe Asn Asn Asn Asp Lys Leu Ile Asn Ala Ile Lys Ala Gln
            180                 185                 190

Tyr Asp Glu Phe Asp Asn Phe Leu Asp Asn Pro Arg Leu Gly Tyr Phe
        195                 200                 205

Gly Gln Ala Phe Phe Ser Lys Glu Gly Arg Asn Tyr Ile Ile Asn Tyr
    210                 215                 220

Gly Asn Glu Cys Tyr Asp Ile Leu Ala Leu Leu Ser Gly Leu Arg His
225                 230                 235                 240

Trp Val Val His Asn Asn Glu Glu Glu Ser Arg Ile Ser Arg Thr Trp
                245                 250                 255

Leu Tyr Asn Leu Asp Lys Asn Leu Asp Asn Glu Tyr Ile Ser Thr Leu
            260                 265                 270

Asn Tyr Leu Tyr Asp Arg Ile Thr Asn Glu Leu Thr Asn Ser Phe Ser
        275                 280                 285

Lys Asn Ser Ala Ala Asn Val Asn Tyr Ile Ala Glu Thr Leu Gly Ile
    290                 295                 300
```

```
Asn Pro Ala Glu Phe Ala Glu Gln Tyr Phe Arg Phe Ser Ile Met Lys
305                 310                 315                 320


Glu Gln Lys Asn Leu Gly Phe Asn Ile Thr Lys Leu Arg Glu Val Met
                325                 330                 335


Leu Asp Arg Lys Asp Met Ser Glu Ile Arg Lys Asn His Lys Val Phe
            340                 345                 350


Asp Ser Ile Arg Thr Lys Val Tyr Thr Met Met Asp Phe Val Ile Tyr
        355                 360                 365


Arg Tyr Tyr Ile Glu Glu Asp Ala Lys Val Ala Ala Ala Asn Lys Ser
    370                 375                 380


Leu Pro Asp Asn Glu Lys Ser Leu Ser Glu Lys Asp Ile Phe Val Ile
385                 390                 395                 400


Asn Leu Arg Gly Ser Phe Asn Asp Asp Gln Lys Asp Ala Leu Tyr Tyr
                405                 410                 415


Asp Glu Ala Asn Arg Ile Trp Arg Lys Leu Glu Asn Ile Met His Asn
            420                 425                 430


Ile Lys Glu Phe Arg Gly Asn Lys Thr Arg Glu Tyr Lys Lys Lys Asp
        435                 440                 445


Ala Pro Arg Leu Pro Arg Ile Leu Pro Ala Gly Arg Asp Val Ser Ala
    450                 455                 460


Phe Ser Lys Leu Met Tyr Ala Leu Thr Met Phe Leu Asp Gly Lys Glu
465                 470                 475                 480


Ile Asn Asp Leu Leu Thr Thr Leu Ile Asn Lys Phe Asp Asn Ile Gln
                485                 490                 495


Ser Phe Leu Lys Val Met Pro Leu Ile Gly Val Asn Ala Lys Phe Val
            500                 505                 510
```

```
Glu Glu Tyr Ala Phe Phe Lys Asp Ser Ala Lys Ile Ala Asp Glu Leu
        515                 520                 525


Arg Leu Ile Lys Ser Phe Ala Arg Met Gly Glu Pro Ile Ala Asp Ala
    530                 535                 540


Arg Arg Ala Met Tyr Ile Asp Ala Ile Arg Ile Leu Gly Thr Asn Leu
545                 550                 555                 560


Ser Tyr Asp Glu Leu Lys Ala Leu Ala Asp Thr Phe Ser Leu Asp Glu
                565                 570                 575


Asn Gly Asn Lys Leu Lys Lys Gly Lys His Gly Met Arg Asn Phe Ile
            580                 585                 590


Ile Asn Asn Val Ile Ser Asn Lys Arg Phe His Tyr Leu Ile Arg Tyr
        595                 600                 605


Gly Asp Pro Ala His Leu His Glu Ile Ala Lys Asn Glu Ala Val Val
    610                 615                 620


Lys Phe Val Leu Gly Arg Ile Ala Asp Ile Gln Lys Lys Gln Gly Gln
625                 630                 635                 640


Asn Gly Lys Asn Gln Ile Asp Arg Tyr Tyr Glu Thr Cys Ile Gly Lys
                645                 650                 655


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


Ile Thr Gly Met Asn Tyr Asp Gln Phe Asp Lys Lys Arg Ser Val Ile
        675                 680                 685


Glu Asp Thr Gly Arg Glu Asn Ala Glu Arg Glu Lys Phe Lys Lys Ile
    690                 695                 700


Ile Ser Leu Tyr Leu Thr Val Ile Tyr His Ile Leu Lys Asn Ile Val
705                 710                 715                 720
```

```
Asn Ile Asn Ala Arg Tyr Val Ile Gly Phe His Cys Val Glu Arg Asp
                725                 730                 735

Ala Gln Leu Tyr Lys Glu Lys Gly Tyr Asp Ile Asn Leu Lys Lys Leu
            740                 745                 750

Glu Glu Lys Gly Phe Ser Ser Val Thr Lys Leu Cys Ala Gly Ile Asp
        755                 760                 765

Glu Thr Ala Pro Asp Lys Arg Lys Asp Val Glu Lys Glu Met Ala Glu
    770                 775                 780

Arg Ala Lys Glu Ser Ile Asp Ser Leu Glu Ser Ala Asn Pro Lys Leu
785                 790                 795                 800

Tyr Ala Asn Tyr Ile Lys Tyr Ser Asp Glu Lys Lys Ala Glu Glu Phe
                805                 810                 815

Thr Arg Gln Ile Asn Arg Glu Lys Ala Lys Thr Ala Leu Asn Ala Tyr
            820                 825                 830

Leu Arg Asn Thr Lys Trp Asn Val Ile Ile Arg Glu Asp Leu Leu Arg
        835                 840                 845

Ile Asp Asn Lys Thr Cys Thr Leu Phe Arg Asn Lys Ala Val His Leu
    850                 855                 860

Glu Val Ala Arg Tyr Val His Ala Tyr Ile Asn Asp Ile Ala Glu Val
865                 870                 875                 880

Asn Ser Tyr Phe Gln Leu Tyr His Tyr Ile Met Gln Arg Ile Ile Met
                885                 890                 895

Asn Glu Arg Tyr Glu Lys Ser Ser Gly Lys Val Ser Glu Tyr Phe Asp
            900                 905                 910

Ala Val Asn Asp Glu Lys Lys Tyr Asn Asp Arg Leu Leu Lys Leu Leu
        915                 920                 925
```

Cys Val Pro Phe Gly Tyr Cys Ile Pro Arg Phe Lys Asn Leu Ser Ile
    930   935   940

Glu Ala Leu Phe Asp Arg Asn Glu Ala Ala Lys Phe Asp Lys Glu Lys
945   950   955   960

Lys Lys Val Ser Gly Asn Ser
                965

<210> 93
<211> 146
<212> PRT
<213> Artificial Sequence

<220>
<223> M26_Gut_metagenome_contig698000307

<400> 93

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

Thr Leu Val Ser Gly Ser Lys Val Tyr Met Thr Thr Phe Ala Glu Gly
            20   25   30

Ser Asp Ala Lys Leu Glu Lys Ile Val Glu Gly Asp Ser Ile Arg Ser
        35   40   45

Val Asn Glu Gly Glu Ala Phe Ser Ala Glu Met Ala Asp Lys Asn Ala
    50   55   60

Gly Tyr Lys Ile Gly Asn Ala Lys Phe Ser His Pro Lys Gly Tyr Ala
65   70   75   80

Val Val Ala Asn Asn Pro Leu Tyr Thr Gly Pro Val Gln Gln Asp Met
                85   90   95

Leu Gly Leu Lys Glu Thr Leu Glu Lys Arg Tyr Phe Gly Glu Ser Ala
            100   105   110

Asp Gly Asn Asp Asn Ile Cys Ile Gln Val Ile His Asn Ile Leu Asp
        115   120   125

```
Ile Glu Lys Ile Leu Ala Glu Tyr Ile Thr Asn Ala Ala Tyr Ser Val
    130                 135                 140


Thr Asp
145
```

<210> 94
<211> 724
<212> PRT
<213> Artificial Sequence

<220>
<223> M36_Uncultured_Eubacterium_sp_TS28_c40956

<400> 94

```
Arg Ile Arg Phe Val Leu Ser Gly Ile Phe Val Asn Ala Phe Ser Ser
1               5                   10                  15


Lys His Gly Phe Glu Ser Gly Val Glu Ile Asn Thr Ser Asn Pro Thr
            20                  25                  30


His Arg Ser Gly Glu Ser Ser Pro Val Arg Gly Asp Met Leu Gly Leu
        35                  40                  45


Lys Ser Glu Leu Glu Lys Arg Phe Phe Gly Lys Thr Phe Asp Asp Asn
    50                  55                  60


Ile His Ile Gln Leu Ile Tyr Asn Ile Leu Asp Ile Glu Lys Ile Leu
65                  70                  75                  80


Ala Val Tyr Val Thr Asn Ile Val Tyr Ala Leu Asn Asn Met Leu Gly
                85                  90                  95


Val Lys Gly Ser Glu Ser His Asp Asp Phe Ile Gly Tyr Leu Ser Thr
            100                 105                 110


Asn Asn Thr Tyr Asp Val Phe Ile Asp Pro Asp Asn Ser Ser Leu Ser
        115                 120                 125
```

```
Asp Asp Lys Lys Ala Asn Val Arg Lys Ser Leu Ser Lys Phe Asn Val
    130                 135                 140

Leu Leu Lys Thr Lys Arg Leu Gly Tyr Phe Gly Leu Glu Glu Pro Lys
145                 150                 155                 160

Thr Lys Asp Asn Arg Val Ser Glu Ala Tyr Lys Lys Arg Val Tyr His
                165                 170                 175

Met Leu Ala Ile Val Gly Gln Ile Arg Gln Cys Val Phe His Asp Lys
            180                 185                 190

Ser Gly Ala Lys Arg Phe Asp Leu Tyr Ser Phe Ile Asn Asn Ile Asp
        195                 200                 205

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

Ser Ile Asn Lys Asp Phe Ile Gln Gly Asn Lys Val Asn Ile Ser Leu
225                 230                 235                 240

Leu Ile Asp Met Met Lys Gly Tyr Glu Ala Asp Asp Ile Ile Arg Leu
                245                 250                 255

Tyr Tyr Asp Phe Ile Val Leu Lys Ser Gln Lys Asn Leu Gly Phe Ser
            260                 265                 270

Ile Lys Lys Leu Arg Glu Lys Met Leu Glu Glu Tyr Gly Phe Arg Phe
        275                 280                 285

Lys Asp Lys Gln Tyr Asp Ser Val Arg Ser Lys Met Tyr Lys Leu Met
    290                 295                 300

Asp Phe Leu Leu Phe Cys Asn Tyr Tyr Arg Asn Asp Ile Ala Ala Gly
305                 310                 315                 320

Glu Ala Leu Val Arg Lys Leu Arg Phe Ser Met Thr Asp Asp Glu Lys
                325                 330                 335
```

```
Glu Gly Leu Tyr Ala Asp Glu Ala Ala Lys Leu Trp Gly Lys Phe Arg
            340                 345                 350

Asn Asp Phe Glu Asn Ile Ala Asp His Met Asn Gly Asp Val Ile Lys
        355                 360                 365

Glu Leu Gly Lys Ala Asp Met Asp Phe Asp Glu Lys Ile Leu Asp Ser
    370                 375                 380

Glu Lys Lys Asn Ala Ser Asp Leu Leu Tyr Phe Ser Lys Met Ile Tyr
385                 390                 395                 400

Met Leu Thr Tyr Phe Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr
                405                 410                 415

Thr Leu Ile Ser Lys Phe Asp Asn Ile Lys Glu Phe Leu Lys Ile Met
            420                 425                 430

Lys Ser Ser Ala Val Asn Val Glu Cys Glu Leu Thr Ala Gly Tyr Lys
        435                 440                 445

Leu Phe Asn Asp Ser Gln Arg Ile Thr Asn Glu Leu Phe Ile Val Lys
    450                 455                 460

Asn Ile Ala Ser Met Arg Lys Pro Ala Ala Ser Ala Lys Leu Thr Met
465                 470                 475                 480

Phe Arg Asp Ala Leu Thr Ile Leu Gly Ile Asp Asp Lys Ile Thr Asp
                485                 490                 495

Asp Arg Ile Ser Glu Ile Leu Lys Leu Lys Glu Lys Gly Lys Gly Ile
            500                 505                 510

His Gly Leu Arg Asn Phe Ile Thr Asn Asn Val Ile Glu Ser Ser Arg
        515                 520                 525

Phe Val Tyr Leu Ile Lys Tyr Ala Asn Ala Gln Lys Ile Arg Lys Val
    530                 535                 540
```

```
Ala Glu Asn Glu Lys Val Val Met Phe Val Leu Gly Gly Ile Pro Asp
545                 550                 555                 560


Thr Gln Ile Glu Arg Tyr Tyr Lys Ser Cys Val Glu Phe Pro Asp Met
                565                 570                 575


Asn Ser Ser Leu Glu Val Lys Arg Ser Glu Leu Ala Arg Met Ile Lys
            580                 585                 590


Asn Ile Ser Phe Asp Asp Phe Lys Asn Val Lys Gln Gln Ala Lys Gly
        595                 600                 605


Gln Arg Lys Arg Arg Leu Arg Arg Gly Gln Lys Ala Val Ile Gly Leu
    610                 615                 620


Tyr Leu Thr Val Met Tyr Leu Leu Val Lys Asn Leu Val Asn Val Asn
625                 630                 635                 640


Ala Arg Tyr Val Ile Ala Ile His Cys Leu Glu Arg Asp Phe Gly Leu
                645                 650                 655


Tyr Lys Glu Ile Ile Pro Glu Leu Ala Ser Lys Asn Leu Lys Asn Asp
            660                 665                 670


Tyr Arg Ile Leu Ser Gln Thr Leu Cys Glu Leu Cys Asp Lys Ser Pro
        675                 680                 685


Asn Leu Phe Leu Lys Lys Asn Glu Arg Leu Arg Lys Cys Val Glu Val
    690                 695                 700


Asp Ile Asn Asn Ala Asp Ser Ser Met Thr Arg Lys Tyr Arg Asn Cys
705                 710                 715                 720


Ile Ala His Leu



<210>  95
<211>  219
<212>  PRT
<213>  Artificial Sequence
```

<220>
<223> M12_gut_metagenome_P25C0k2120140920_c134000066

<400> 95

```
Met Ala Lys Lys Asn Lys Met Lys Pro Arg Glu Leu Arg Glu Ala Gln
1               5                  10                  15


Lys Lys Ala Arg Gln Phe Lys Ala Ala Glu Ile Asn Asn Asn Ala Ala
            20                  25                  30


Pro Ala Ile Ala Ala Met Pro Ala Ala Glu Val Ile Ala Pro Val Ala
        35                  40                  45


Glu Lys Lys Lys Ser Ser Val Lys Ala Ala Gly Met Lys Ser Ile Leu
    50                  55                  60


Val Ser Glu Asn Lys Met Tyr Ile Thr Ser Phe Gly Lys Gly Asn Ser
65                  70                  75                  80


Ala Val Leu Glu Tyr Glu Val Asp Asn Asn Asp Tyr Asn Lys Thr Gln
                85                  90                  95


Leu Ser Ser Lys Asp Asn Ser Asn Ile Glu Leu Cys Gly Val Asn Glu
            100                 105                 110


Val Asn Ile Thr Phe Ser Ser Lys His Gly Phe Glu Ser Gly Val Glu
        115                 120                 125


Ile Asn Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser Ser Pro Val
    130                 135                 140


Arg Trp Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys Arg Phe Phe
145                 150                 155                 160


Gly Lys Thr Phe Asp Asp Asn Ile His Ile Gln Leu Ile Tyr Asn Ile
                165                 170                 175


Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn Ile Val Tyr
            180                 185                 190
```

```
Ala Leu Asn Asn Met Leu Gly Val Lys Gly Ser Glu Ser His Asp Asp
        195                 200                 205


Phe Ile Gly Tyr Leu Ser Thr Asn Asn Ile Tyr
    210                 215


<210>  96
<211>  915
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Artificial sequence

<400>  96

Met Ala Lys Lys Asn Lys Met Lys Pro Arg Glu Leu Arg Glu Ala Gln
1               5                   10                  15


Lys Lys Ala Arg Gln Phe Lys Ala Ala Glu Ile Asn Asn Asn Ala Ala
            20                  25                  30


Pro Ala Ile Ala Ala Met Pro Ala Ala Glu Val Ile Ala Pro Val Ala
        35                  40                  45


Glu Lys Lys Lys Ser Ser Val Lys Ala Ala Gly Met Lys Ser Ile Leu
    50                  55                  60


Val Ser Glu Asn Lys Met Tyr Ile Thr Ser Phe Gly Lys Gly Asn Ser
65                  70                  75                  80


Ala Val Leu Glu Tyr Glu Val Asp Asn Asn Asp Tyr Asn Lys Thr Gln
                85                  90                  95


Leu Ser Ser Lys Asp Asn Ser Asn Ile Glu Leu Cys Gly Val Asn Glu
            100                 105                 110


Val Asn Ile Thr Phe Ser Ser Lys His Gly Phe Glu Ser Gly Val Glu
        115                 120                 125
```

```
Ile Asn Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser Ser Pro Val
    130                 135                 140

Arg Trp Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys Arg Phe Phe
145                 150                 155                 160

Gly Lys Thr Phe Asp Asp Asn Ile His Ile Gln Leu Ile Tyr Asn Ile
                165                 170                 175

Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn Ile Val Tyr
            180                 185                 190

Ala Leu Asn Asn Met Leu Gly Val Lys Gly Ser Glu Ser His Asp Asp
        195                 200                 205

Phe Ile Gly Tyr Leu Ser Thr Asn Asn Ile Tyr Asp Val Phe Ile Asp
    210                 215                 220

Pro Asp Asn Ser Ser Leu Ser Asp Asp Lys Lys Ala Asn Val Arg Lys
225                 230                 235                 240

Ser Leu Ser Lys Phe Asn Val Leu Leu Lys Thr Lys Arg Leu Gly Tyr
                245                 250                 255

Phe Gly Leu Glu Glu Pro Lys Thr Lys Asp Thr Arg Val Ser Gln Ala
            260                 265                 270

Tyr Lys Lys Arg Val Tyr His Met Leu Ala Ile Val Gly Gln Ile Arg
        275                 280                 285

Gln Cys Val Phe His Asp Lys Ser Gly Ala Lys Arg Phe Asp Leu Tyr
    290                 295                 300

Ser Phe Ile Asn Asn Ile Asp Pro Glu Tyr Arg Glu Thr Leu Asp Tyr
305                 310                 315                 320

Leu Val Asp Glu Arg Phe Asp Ser Ile Asn Lys Gly Phe Ile Gln Gly
                325                 330                 335
```

```
Asn Lys Val Asn Ile Ser Leu Leu Ile Asp Met Met Lys Gly Tyr Glu
            340                 345                 350

Ala Asp Asp Ile Ile Arg Leu Tyr Tyr Asp Phe Ile Val Leu Lys Ser
        355                 360                 365

Gln Lys Asn Leu Gly Phe Ser Ile Lys Lys Leu Arg Glu Lys Met Leu
    370                 375                 380

Glu Glu Tyr Gly Tyr Arg Phe Lys Asp Lys Gln Met Tyr Lys Leu Met
385                 390                 395                 400

Asp Phe Leu Leu Phe Cys Asn Tyr Tyr Arg Asn Asp Val Val Ala Gly
                405                 410                 415

Glu Ala Leu Val Arg Lys Leu Arg Phe Ser Met Thr Asp Asp Glu Lys
            420                 425                 430

Glu Gly Ile Tyr Ala Asp Glu Ala Ser Lys Leu Trp Gly Lys Phe Arg
        435                 440                 445

Asn Asp Phe Glu Asn Ile Ala Asp His Met Asn Gly Asp Val Ile Lys
    450                 455                 460

Glu Leu Gly Lys Ala Asp Met Asp Phe Asp Glu Lys Ile Leu Asp Ser
465                 470                 475                 480

Glu Lys Lys Asn Ala Ser Asp Leu Leu Tyr Phe Ser Lys Met Ile Tyr
                485                 490                 495

Met Leu Thr Tyr Phe Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr
            500                 505                 510

Thr Leu Ile Ser Lys Phe Asp Asn Ile Lys Glu Phe Leu Lys Ile Met
        515                 520                 525

Lys Ser Ser Ala Val Asp Val Glu Cys Glu Leu Thr Ala Gly Tyr Lys
    530                 535                 540
```

```
Leu Phe Asn Asp Ser Gln Arg Ile Thr Asn Glu Leu Phe Ile Val Lys
545                 550                 555                 560


Asn Ile Ala Ser Met Arg Lys Pro Ala Ser Ser Ala Lys Leu Thr Met
                565                 570                 575


Phe Arg Asp Ala Leu Thr Ile Leu Gly Ile Asp Asp Asn Ile Thr Asp
            580                 585                 590


Asp Arg Ile Ser Glu Ile Leu Lys Leu Lys Glu Lys Gly Lys Gly Ile
        595                 600                 605


His Gly Leu Arg Asn Phe Ile Thr Asn Asn Val Ile Glu Ser Ser Arg
    610                 615                 620


Phe Val Tyr Leu Ile Lys Tyr Ala Asn Ala Gln Lys Ile Arg Glu Val
625                 630                 635                 640


Ala Lys Asn Glu Lys Val Val Met Phe Val Leu Gly Gly Ile Pro Asp
                645                 650                 655


Thr Gln Ile Glu Arg Tyr Tyr Lys Ser Cys Val Glu Phe Pro Asp Met
            660                 665                 670


Asn Ser Ser Leu Glu Val Lys Arg Ser Glu Leu Ala Arg Met Ile Lys
        675                 680                 685


Asn Ile Ser Phe Asp Asp Phe Lys Asn Val Lys Gln Gln Ala Lys Gly
    690                 695                 700


Arg Glu Asn Val Ala Lys Glu Arg Ala Lys Ala Val Ile Gly Leu Tyr
705                 710                 715                 720


Leu Thr Val Met Tyr Leu Leu Val Lys Asn Leu Val Asn Val Asn Ala
                725                 730                 735


Arg Tyr Val Ile Ala Ile His Cys Leu Glu Arg Asp Phe Gly Leu Tyr
            740                 745                 750
```

```
Lys Glu Ile Ile Pro Glu Leu Ala Ser Lys Asn Leu Lys Asn Asp Tyr
        755                 760                 765


Arg Ile Leu Ser Gln Thr Leu Cys Glu Leu Cys Asp Asn Gly Asp Glu
    770                 775                 780


Ser Pro Asn Leu Phe Leu Lys Lys Asn Lys Arg Leu Arg Lys Cys Val
785                 790                 795                 800


Glu Val Asp Ile Asn Asn Ala Asp Ser Asn Met Thr Arg Lys Tyr Arg
                805                 810                 815


Asn Cys Ile Ala His Leu Thr Val Val Arg Glu Leu Lys Glu Tyr Ile
            820                 825                 830


Gly Asp Ile Arg Thr Val Asp Ser Tyr Phe Ser Ile Tyr His Tyr Val
        835                 840                 845


Met Gln Arg Cys Ile Thr Lys Arg Glu Asp Asp Thr Lys Gln Glu Glu
    850                 855                 860


Lys Ile Lys Tyr Glu Asp Asp Leu Leu Lys Asn His Gly Tyr Thr Lys
865                 870                 875                 880


Asp Phe Val Lys Ala Leu Asn Ser Pro Phe Gly Tyr Asn Ile Pro Arg
                885                 890                 895


Phe Lys Asn Leu Ser Ile Glu Gln Leu Phe Asp Arg Asn Glu Tyr Leu
            900                 905                 910


Thr Glu Lys
        915


<210>  97
<211>  922
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  M10_gut_metagenome_P25C90k2120140920,_c28000041
```

<400> 97

```
Met Ala Lys Lys Asn Lys Met Lys Pro Arg Glu Leu Arg Glu Ala Gln
1               5                  10                  15

Lys Lys Ala Arg Gln Phe Lys Ala Ala Glu Ile Asn Asn Asn Ala Ala
            20                  25                  30

Pro Ala Ile Ala Ala Met Pro Ala Ala Glu Val Ile Ala Pro Val Ala
        35                  40                  45

Glu Lys Lys Lys Ser Ser Val Lys Ala Ala Gly Met Lys Ser Ile Leu
    50                  55                  60

Val Ser Glu Asn Lys Met Tyr Ile Thr Ser Phe Gly Lys Gly Asn Ser
65                  70                  75                  80

Ala Val Leu Glu Tyr Glu Val Asp Asn Asn Asp Tyr Asn Lys Thr Gln
                85                  90                  95

Leu Ser Ser Lys Asp Asn Ser Asn Ile Glu Leu Cys Gly Val Asn Glu
            100                 105                 110

Val Asn Ile Thr Phe Ser Ser Lys His Gly Phe Glu Ser Gly Val Glu
        115                 120                 125

Ile Asn Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser Ser Pro Val
    130                 135                 140

Arg Trp Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys Arg Phe Phe
145                 150                 155                 160

Gly Lys Thr Phe Asp Asp Asn Ile His Ile Gln Leu Ile Tyr Asn Ile
                165                 170                 175

Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn Ile Val Tyr
            180                 185                 190

Ala Leu Asn Asn Met Leu Gly Val Lys Gly Ser Glu Ser His Asp Asp
        195                 200                 205
```

```
Phe Ile Gly Tyr Leu Ser Thr Asn Asn Ile Tyr Asp Val Phe Ile Asp
    210                 215                 220

Pro Asp Asn Ser Ser Leu Ser Asp Asp Lys Lys Ala Asn Val Arg Lys
225                 230                 235                 240

Ser Leu Ser Lys Phe Asn Val Leu Leu Lys Thr Lys Arg Leu Gly Tyr
                245                 250                 255

Phe Gly Leu Glu Glu Pro Lys Thr Lys Asp Thr Arg Val Ser Gln Ala
            260                 265                 270

Tyr Lys Lys Arg Val Tyr His Met Leu Ala Ile Val Gly Gln Ile Arg
        275                 280                 285

Gln Cys Val Phe His Asp Lys Ser Gly Ala Lys Arg Phe Asp Leu Tyr
    290                 295                 300

Ser Phe Ile Asn Asn Ile Asp Pro Glu Tyr Arg Glu Thr Leu Asp Tyr
305                 310                 315                 320

Leu Val Asp Glu Arg Phe Asp Ser Ile Asn Lys Gly Phe Ile Gln Gly
                325                 330                 335

Asn Lys Val Asn Ile Ser Leu Leu Ile Asp Met Met Lys Gly Tyr Glu
            340                 345                 350

Ala Asp Asp Ile Ile Arg Leu Tyr Tyr Asp Phe Ile Val Leu Lys Ser
        355                 360                 365

Gln Lys Asn Leu Gly Phe Ser Ile Lys Lys Leu Arg Glu Lys Met Leu
    370                 375                 380

Glu Glu Tyr Gly Tyr Arg Phe Lys Asp Lys Gln Tyr Asp Ser Val Arg
385                 390                 395                 400

Ser Lys Met Tyr Lys Leu Met Asp Phe Leu Leu Phe Cys Asn Tyr Tyr
                405                 410                 415
```

```
Arg Asn Asp Val Val Ala Gly Glu Ala Leu Val Arg Lys Leu Arg Phe
            420                 425                 430

Ser Met Thr Asp Asp Glu Lys Glu Gly Ile Tyr Ala Asp Glu Ala Ser
        435                 440                 445

Lys Leu Trp Gly Lys Phe Arg Asn Asp Phe Glu Asn Ile Ala Asp His
    450                 455                 460

Met Asn Gly Asp Val Ile Lys Glu Leu Gly Lys Ala Asp Met Asp Phe
465                 470                 475                 480

Asp Glu Lys Ile Leu Asp Ser Glu Lys Lys Asn Ala Ser Asp Leu Leu
                485                 490                 495

Tyr Phe Ser Lys Met Ile Tyr Met Leu Thr Tyr Phe Leu Asp Gly Lys
            500                 505                 510

Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Ser Lys Phe Asp Asn Ile
        515                 520                 525

Lys Glu Phe Leu Lys Ile Met Lys Ser Ser Ala Val Asp Val Glu Cys
    530                 535                 540

Glu Leu Thr Ala Gly Tyr Lys Leu Phe Asn Asp Ser Gln Arg Ile Thr
545                 550                 555                 560

Asn Glu Leu Phe Ile Val Lys Asn Ile Ala Ser Met Arg Lys Pro Ala
                565                 570                 575

Ser Ser Ala Lys Leu Thr Met Phe Arg Asp Ala Leu Thr Ile Leu Gly
            580                 585                 590

Ile Asp Asp Asn Ile Thr Asp Asp Arg Ile Ser Glu Ile Leu Lys Leu
        595                 600                 605

Lys Glu Lys Gly Lys Gly Ile His Gly Leu Arg Asn Phe Ile Thr Asn
    610                 615                 620
```

```
Asn Val Ile Glu Ser Ser Arg Phe Val Tyr Leu Ile Lys Tyr Ala Asn
625                 630                 635                 640

Ala Gln Lys Ile Arg Glu Val Ala Lys Asn Glu Lys Val Val Met Phe
                645                 650                 655

Val Leu Gly Gly Ile Pro Asp Thr Gln Ile Glu Arg Tyr Tyr Lys Ser
            660                 665                 670

Cys Val Glu Phe Pro Asp Met Asn Ser Ser Leu Glu Val Lys Arg Ser
        675                 680                 685

Glu Leu Ala Arg Met Ile Lys Asn Ile Ser Phe Asp Asp Phe Lys Asn
    690                 695                 700

Val Lys Gln Gln Ala Lys Gly Arg Glu Asn Val Ala Lys Glu Arg Ala
705                 710                 715                 720

Lys Ala Val Ile Gly Leu Tyr Leu Thr Val Met Tyr Leu Leu Val Lys
                725                 730                 735

Asn Leu Val Asn Val Asn Ala Arg Tyr Val Ile Ala Ile His Cys Leu
            740                 745                 750

Glu Arg Asp Phe Gly Leu Tyr Lys Glu Ile Ile Pro Glu Leu Ala Ser
        755                 760                 765

Lys Asn Leu Lys Asn Asp Tyr Arg Ile Leu Ser Gln Thr Leu Cys Glu
    770                 775                 780

Leu Cys Asp Asn Gly Asp Glu Ser Pro Asn Leu Phe Leu Lys Lys Asn
785                 790                 795                 800

Lys Arg Leu Arg Lys Cys Val Glu Val Asp Ile Asn Asn Ala Asp Ser
                805                 810                 815

Asn Met Thr Arg Lys Tyr Arg Asn Cys Ile Ala His Leu Thr Val Val
            820                 825                 830
```

```
Arg Glu Leu Lys Glu Tyr Ile Gly Asp Ile Arg Thr Val Asp Ser Tyr
        835                 840                 845

Phe Ser Ile Tyr His Tyr Val Met Gln Arg Cys Ile Thr Lys Arg Glu
    850                 855                 860

Asp Asp Thr Lys Gln Glu Glu Lys Ile Lys Tyr Glu Asp Asp Leu Leu
865                 870                 875                 880

Lys Asn His Gly Tyr Thr Lys Asp Phe Val Lys Ala Leu Asn Ser Pro
                885                 890                 895

Phe Gly Tyr Asn Ile Pro Arg Phe Lys Asn Leu Ser Ile Glu Gln Leu
            900                 905                 910

Phe Asp Arg Asn Glu Tyr Leu Thr Glu Lys
        915                 920


<210>  98
<211>  922
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  M11_gut_metagenome_P25C7k2120140920_c4078000105

<400>  98

Met Ala Lys Lys Asn Lys Met Lys Pro Arg Glu Leu Arg Glu Ala Gln
1               5                   10                  15

Lys Lys Ala Arg Gln Phe Lys Ala Ala Glu Ile Asn Asn Asn Ala Ala
            20                  25                  30

Pro Ala Ile Ala Ala Met Pro Ala Ala Glu Val Ile Ala Pro Val Ala
        35                  40                  45

Glu Lys Lys Lys Ser Ser Val Lys Ala Ala Gly Met Lys Ser Ile Leu
    50                  55                  60
```

```
Val Ser Glu Asn Lys Met Tyr Ile Thr Ser Phe Gly Lys Gly Asn Ser
65                  70                  75                  80


Ala Val Leu Glu Tyr Glu Val Asp Asn Asn Asp Tyr Asn Lys Thr Gln
                85                  90                  95


Leu Ser Ser Lys Asp Asn Ser Asn Ile Glu Leu Cys Gly Val Asn Glu
            100                 105                 110


Val Asn Ile Thr Phe Ser Ser Lys His Gly Phe Glu Ser Gly Val Glu
        115                 120                 125


Ile Asn Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser Ser Pro Val
    130                 135                 140


Arg Trp Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys Arg Phe Phe
145                 150                 155                 160


Gly Lys Thr Phe Asp Asp Asn Ile His Ile Gln Leu Ile Tyr Asn Ile
                165                 170                 175


Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn Ile Val Tyr
            180                 185                 190


Ala Leu Asn Asn Met Leu Gly Val Lys Gly Ser Glu Ser His Asp Asp
        195                 200                 205


Phe Ile Gly Tyr Leu Ser Thr Asn Asn Ile Tyr Asp Val Phe Ile Asp
    210                 215                 220


Pro Asp Asn Ser Ser Leu Ser Asp Asp Lys Lys Ala Asn Val Arg Lys
225                 230                 235                 240


Ser Leu Ser Lys Phe Asn Val Leu Leu Lys Thr Lys Arg Leu Gly Tyr
                245                 250                 255


Phe Gly Leu Glu Glu Pro Lys Thr Lys Asp Thr Arg Val Ser Gln Ala
            260                 265                 270
```

```
Tyr Lys Lys Arg Val Tyr His Met Leu Ala Ile Val Gly Gln Ile Arg
        275                 280                 285

Gln Cys Val Phe His Asp Lys Ser Gly Ala Lys Arg Phe Asp Leu Tyr
    290                 295                 300

Ser Phe Ile Asn Asn Ile Asp Pro Glu Tyr Arg Glu Thr Leu Asp Tyr
305                 310                 315                 320

Leu Val Asp Glu Arg Phe Asp Ser Ile Asn Lys Gly Phe Ile Gln Gly
                325                 330                 335

Asn Lys Val Asn Ile Ser Leu Leu Ile Asp Met Met Lys Gly Tyr Glu
            340                 345                 350

Ala Asp Asp Ile Ile Arg Leu Tyr Tyr Asp Phe Ile Val Leu Lys Ser
        355                 360                 365

Gln Lys Asn Leu Gly Phe Ser Ile Lys Lys Leu Arg Glu Lys Met Leu
    370                 375                 380

Glu Glu Tyr Gly Tyr Arg Phe Lys Asp Lys Gln Tyr Asp Ser Val Arg
385                 390                 395                 400

Ser Lys Met Tyr Lys Leu Met Asp Phe Leu Leu Phe Cys Asn Tyr Tyr
                405                 410                 415

Arg Asn Asp Val Val Ala Gly Glu Ala Leu Val Arg Lys Leu Arg Phe
            420                 425                 430

Ser Met Thr Asp Asp Glu Lys Glu Gly Ile Tyr Ala Asp Glu Ala Ser
        435                 440                 445

Lys Leu Trp Gly Lys Phe Arg Asn Asp Phe Glu Asn Ile Ala Asp His
    450                 455                 460

Met Asn Gly Asp Val Ile Lys Glu Leu Gly Lys Ala Asp Met Asp Phe
465                 470                 475                 480
```

```
Asp Glu Lys Ile Leu Asp Ser Glu Lys Lys Asn Ala Ser Asp Leu Leu
                485                 490                 495

Tyr Phe Ser Lys Met Ile Tyr Met Leu Thr Tyr Phe Leu Asp Gly Lys
            500                 505                 510

Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Ser Lys Phe Asp Asn Ile
        515                 520                 525

Lys Glu Phe Leu Lys Ile Met Lys Ser Ser Ala Val Asp Val Glu Cys
    530                 535                 540

Glu Leu Thr Ala Gly Tyr Lys Leu Phe Asn Asp Ser Gln Arg Ile Thr
545                 550                 555                 560

Asn Glu Leu Phe Ile Val Lys Asn Ile Ala Ser Met Arg Lys Pro Ala
                565                 570                 575

Ser Ser Ala Lys Leu Thr Met Phe Arg Asp Ala Leu Thr Ile Leu Gly
            580                 585                 590

Ile Asp Asp Asn Ile Thr Asp Asp Arg Ile Ser Glu Ile Leu Lys Leu
        595                 600                 605

Lys Glu Lys Gly Lys Gly Ile His Gly Leu Arg Asn Phe Ile Thr Asn
    610                 615                 620

Asn Val Ile Glu Ser Ser Arg Phe Val Tyr Leu Ile Lys Tyr Ala Asn
625                 630                 635                 640

Ala Gln Lys Ile Arg Glu Val Ala Lys Asn Glu Lys Val Val Met Phe
                645                 650                 655

Val Leu Gly Gly Ile Pro Asp Thr Gln Ile Glu Arg Tyr Tyr Lys Ser
            660                 665                 670

Cys Val Glu Phe Pro Asp Met Asn Ser Ser Leu Glu Val Lys Arg Ser
        675                 680                 685
```

```
Glu Leu Ala Arg Met Ile Lys Asn Ile Ser Phe Asp Asp Phe Lys Asn
    690                 695                 700

Val Lys Gln Gln Ala Lys Gly Arg Glu Asn Val Ala Lys Glu Arg Ala
705                 710                 715                 720

Lys Ala Val Ile Gly Leu Tyr Leu Thr Val Met Tyr Leu Leu Val Lys
                725                 730                 735

Asn Leu Val Asn Val Asn Ala Arg Tyr Val Ile Ala Ile His Cys Leu
            740                 745                 750

Glu Arg Asp Phe Gly Leu Tyr Lys Glu Ile Ile Pro Glu Leu Ala Ser
        755                 760                 765

Lys Asn Leu Lys Asn Asp Tyr Arg Ile Leu Ser Gln Thr Leu Cys Glu
    770                 775                 780

Leu Cys Asp Asn Gly Asp Glu Ser Pro Asn Leu Phe Leu Lys Lys Asn
785                 790                 795                 800

Lys Arg Leu Arg Lys Cys Val Glu Val Asp Ile Asn Asn Ala Asp Ser
                805                 810                 815

Asn Met Thr Arg Lys Tyr Arg Asn Cys Ile Ala His Leu Thr Val Val
            820                 825                 830

Arg Glu Leu Lys Glu Tyr Ile Gly Asp Ile Arg Thr Val Asp Ser Tyr
        835                 840                 845

Phe Ser Ile Tyr His Tyr Val Met Gln Arg Cys Ile Thr Lys Arg Glu
    850                 855                 860

Asp Asp Thr Lys Gln Glu Glu Lys Ile Lys Tyr Glu Asp Asp Leu Leu
865                 870                 875                 880

Lys Asn His Gly Tyr Thr Lys Asp Phe Val Lys Ala Leu Asn Ser Pro
                885                 890                 895
```

Phe Gly Tyr Asn Ile Pro Arg Phe Lys Asn Leu Ser Ile Glu Gln Leu
            900   905   910

Phe Asp Arg Asn Glu Tyr Leu Thr Glu Lys
        915   920

<210> 99
<211> 636
<212> PRT
<213> Artificial Sequence

<220>
<223> gut_metagenome_P25C0k2120140920_c32000045

<400> 99

Ile Arg Gln Cys Val Phe His Asp Lys Ser Gly Ala Lys Arg Phe Asp
1   5   10   15

Leu Tyr Ser Phe Ile Asn Asn Ile Asp Pro Glu Tyr Arg Glu Thr Leu
            20   25   30

Asp Tyr Leu Val Asp Glu Arg Phe Asp Ser Ile Asn Lys Gly Phe Ile
        35   40   45

Gln Gly Asn Lys Val Asn Ile Ser Leu Leu Ile Asp Met Met Lys Gly
    50   55   60

Tyr Glu Ala Asp Asp Ile Ile Arg Leu Tyr Tyr Asp Phe Ile Val Leu
65   70   75   80

Lys Ser Gln Lys Asn Leu Gly Phe Ser Ile Lys Lys Leu Arg Glu Lys
                85   90   95

Met Leu Glu Glu Tyr Gly Tyr Arg Phe Lys Asp Lys Gln Tyr Asp Ser
            100   105   110

Val Arg Ser Lys Met Tyr Lys Leu Met Asp Phe Leu Leu Phe Cys Asn
        115   120   125

Tyr Tyr Arg Asn Asp Val Val Ala Gly Glu Ala Leu Val Arg Lys Leu
    130   135   140

```
Arg Phe Ser Met Thr Asp Asp Glu Lys Glu Gly Ile Tyr Ala Asp Glu
145                 150                 155                 160


Ala Ser Lys Leu Trp Gly Lys Phe Arg Asn Asp Phe Glu Asn Ile Ala
                165                 170                 175


Asp His Met Asn Gly Asp Val Ile Lys Glu Leu Gly Lys Ala Asp Met
            180                 185                 190


Asp Phe Asp Glu Lys Ile Leu Asp Ser Glu Lys Lys Asn Ala Ser Asp
        195                 200                 205


Leu Leu Tyr Phe Ser Lys Met Ile Tyr Met Leu Thr Tyr Phe Leu Asp
    210                 215                 220


Gly Lys Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Ser Lys Phe Asp
225                 230                 235                 240


Asn Ile Lys Glu Phe Leu Lys Ile Met Lys Ser Ser Ala Val Asp Val
                245                 250                 255


Glu Cys Glu Leu Thr Ala Gly Tyr Lys Leu Phe Asn Asp Ser Gln Arg
            260                 265                 270


Ile Thr Asn Glu Leu Phe Ile Val Lys Asn Ile Ala Ser Met Arg Lys
        275                 280                 285


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


Leu Gly Ile Asp Asp Asn Ile Thr Asp Asp Arg Ile Ser Glu Ile Leu
305                 310                 315                 320


Lys Leu Lys Glu Lys Gly Lys Gly Ile His Gly Leu Arg Asn Phe Ile
                325                 330                 335


Thr Asn Asn Val Ile Glu Ser Ser Arg Phe Val Tyr Leu Ile Lys Tyr
            340                 345                 350
```

```
Ala Asn Ala Gln Lys Ile Arg Glu Val Ala Lys Asn Glu Lys Val Val
        355                 360                 365

Met Phe Val Leu Gly Gly Ile Pro Asp Thr Gln Ile Glu Arg Tyr Tyr
    370                 375                 380

Lys Ser Cys Val Glu Phe Pro Asp Met Asn Ser Ser Leu Glu Val Lys
385                 390                 395                 400

Arg Ser Glu Leu Ala Arg Met Ile Lys Asn Ile Ser Phe Asp Asp Phe
                405                 410                 415

Lys Asn Val Lys Gln Gln Ala Lys Gly Arg Glu Asn Val Ala Lys Glu
            420                 425                 430

Arg Ala Lys Ala Val Ile Gly Leu Tyr Leu Thr Val Met Tyr Leu Leu
        435                 440                 445

Val Lys Asn Leu Val Asn Val Asn Ala Arg Tyr Val Ile Ala Ile His
    450                 455                 460

Cys Leu Glu Arg Asp Phe Gly Leu Tyr Lys Glu Ile Ile Pro Glu Leu
465                 470                 475                 480

Ala Ser Lys Asn Leu Lys Asn Asp Tyr Arg Ile Leu Ser Gln Thr Leu
                485                 490                 495

Cys Glu Leu Cys Asp Asn Gly Asp Glu Ser Pro Asn Leu Phe Leu Lys
            500                 505                 510

Lys Asn Lys Arg Leu Arg Lys Cys Val Glu Val Asp Ile Asn Asn Ala
        515                 520                 525

Asp Ser Asn Met Thr Arg Lys Tyr Arg Asn Cys Ile Ala His Leu Thr
    530                 535                 540

Val Val Arg Glu Leu Lys Glu Tyr Ile Gly Asp Ile Arg Thr Val Asp
545                 550                 555                 560
```

```
Ser Tyr Phe Ser Ile Tyr His Tyr Val Met Gln Arg Cys Ile Thr Lys
                565                 570                 575


Arg Glu Asp Asp Thr Lys Gln Glu Glu Lys Ile Lys Tyr Glu Asp Asp
            580                 585                 590


Leu Leu Lys Asn His Gly Tyr Thr Lys Asp Phe Val Lys Ala Leu Asn
        595                 600                 605


Ser Pro Phe Gly Tyr Asn Ile Pro Arg Phe Lys Asn Leu Ser Ile Glu
    610                 615                 620


Gln Leu Phe Asp Arg Asn Glu Tyr Leu Thr Glu Lys
625                 630                 635


<210>  100
<211>  929
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  M13_gut_metagenome_P23C7k2120140920_c3000067

<400>  100

Met Ala Lys Lys Asn Lys Met Lys Pro Arg Glu Leu Arg Glu Ala Gln
1               5                   10                  15


Lys Lys Ala Arg Gln Leu Lys Ala Ala Glu Ile Asn Asn Asn Ala Ala
            20                  25                  30


Pro Ala Ile Ala Ala Met Pro Ala Ala Gln Val Ile Ala Pro Ala Ala
        35                  40                  45


Glu Lys Lys Lys Ser Ser Val Lys Ala Ala Gly Met Lys Ser Ile Leu
    50                  55                  60


Val Ser Glu Asn Lys Met Tyr Ile Thr Ser Phe Gly Lys Gly Asn Ser
65                  70                  75                  80
```

```
Ala Val Leu Glu Tyr Glu Val Asp Lys Val Asp Asn Asn Asn Tyr Asn
                85                  90                  95

Lys Thr Gln Leu Ser Ser Lys Asp Asn Ser Asn Ile Glu Leu Gly Asp
            100                 105                 110

Val Asn Glu Val Asn Ile Thr Phe Ser Ser Lys His Gly Phe Glu Ser
        115                 120                 125

Gly Val Glu Ile Asn Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser
    130                 135                 140

Ser Pro Val Arg Trp Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys
145                 150                 155                 160

Arg Phe Phe Gly Lys Thr Phe Asp Asp Asn Ile His Ile Gln Leu Ile
                165                 170                 175

Tyr Asn Ile Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn
            180                 185                 190

Ile Val Tyr Ala Leu Asn Asn Met Leu Gly Ile Lys Gly Ser Glu Ser
        195                 200                 205

Tyr Asp Asp Phe Met Gly Tyr Leu Ser Ala Arg Asn Thr Tyr Glu Val
    210                 215                 220

Phe Thr His Pro Asp Lys Ser Asn Leu Ser Asp Lys Val Lys Gly Asn
225                 230                 235                 240

Ile Lys Lys Ser Phe Ser Thr Phe Asn Asp Leu Leu Lys Thr Lys Arg
                245                 250                 255

Leu Gly Tyr Phe Gly Leu Glu Glu Pro Lys Thr Lys Asp Thr Arg Val
            260                 265                 270

Ser Glu Ala Tyr Lys Lys Arg Val Tyr His Met Leu Ala Ile Val Gly
        275                 280                 285
```

```
Gln Ile Arg Gln Cys Val Phe His Asp Leu Ser Glu His Ser Glu Tyr
    290                 295                 300

Asp Leu Tyr Ser Phe Ile Asp Asn Ser Lys Lys Val Tyr Arg Glu Cys
305                 310                 315                 320

Arg Glu Thr Leu Asn Tyr Leu Val Asp Glu Arg Phe Asp Ser Ile Asn
                325                 330                 335

Lys Gly Phe Ile Gln Gly Asn Lys Val Asn Ile Ser Leu Leu Ile Asp
            340                 345                 350

Met Met Lys Asp Asp Tyr Glu Ala Asp Asp Ile Ile His Leu Tyr Tyr
        355                 360                 365

Asp Phe Ile Val Leu Lys Ser Gln Lys Asn Leu Gly Phe Ser Ile Lys
    370                 375                 380

Lys Leu Arg Glu Lys Met Leu Asp Glu Tyr Gly Phe Arg Phe Lys Asp
385                 390                 395                 400

Lys Gln Tyr Asp Ser Val Arg Ser Lys Met Tyr Lys Leu Met Asp Phe
                405                 410                 415

Leu Leu Phe Cys Asn Tyr Tyr Arg Asn Asp Val Val Ala Gly Glu Ala
            420                 425                 430

Leu Val Arg Lys Leu Arg Phe Ser Met Thr Asp Asp Glu Lys Glu Gly
        435                 440                 445

Ile Tyr Ala Asp Glu Ala Glu Lys Leu Trp Gly Lys Phe Arg Asn Asp
    450                 455                 460

Phe Glu Asn Ile Ala Asp His Met Asn Gly Asp Val Ile Lys Glu Leu
465                 470                 475                 480

Gly Lys Ala Asp Met Asp Phe Asp Glu Lys Ile Leu Asp Ser Glu Lys
                485                 490                 495
```

```
Lys Asn Ala Ser Asp Leu Leu Tyr Phe Ser Lys Met Ile Tyr Met Leu
            500                 505                 510

Thr Tyr Phe Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr Thr Leu
        515                 520                 525

Ile Ser Lys Phe Asp Asn Ile Lys Glu Phe Leu Lys Ile Met Lys Ser
    530                 535                 540

Ser Ala Val Asp Val Glu Cys Glu Leu Thr Ala Gly Tyr Lys Leu Phe
545                 550                 555                 560

Asn Asp Ser Gln Arg Ile Thr Asn Glu Leu Phe Ile Val Lys Asn Ile
                565                 570                 575

Ala Ser Met Arg Lys Pro Ala Ala Ser Ala Lys Leu Thr Met Phe Arg
            580                 585                 590

Asp Ala Leu Thr Ile Leu Gly Ile Asp Asp Lys Ile Thr Asp Asp Arg
        595                 600                 605

Ile Ser Glu Ile Leu Lys Leu Lys Glu Lys Gly Lys Gly Ile His Gly
    610                 615                 620

Leu Arg Asn Phe Ile Thr Asn Asn Val Ile Glu Ser Ser Arg Phe Val
625                 630                 635                 640

Tyr Leu Ile Lys Tyr Ala Asn Ala Gln Lys Ile Arg Glu Val Ala Lys
                645                 650                 655

Asn Glu Lys Val Val Met Phe Val Leu Gly Gly Ile Pro Asp Thr Gln
            660                 665                 670

Ile Glu Arg Tyr Tyr Lys Ser Cys Val Glu Phe Pro Asp Met Asn Ser
        675                 680                 685

Ser Leu Lys Val Lys Arg Ser Glu Leu Ala Arg Met Ile Lys Asn Ile
    690                 695                 700
```

```
Arg Phe Asp Asp Phe Lys Asn Val Lys Gln Gln Ala Lys Gly Arg Glu
705                 710                 715                 720


Asn Val Ala Lys Glu Arg Ala Lys Ala Val Ile Gly Leu Tyr Leu Thr
                725                 730                 735


Val Met Tyr Leu Leu Val Lys Asn Leu Val Asn Val Asn Ala Arg Tyr
            740                 745                 750


Val Ile Ala Ile His Cys Leu Glu Arg Asp Phe Gly Leu Tyr Lys Glu
        755                 760                 765


Ile Ile Pro Glu Leu Ala Ser Lys Asn Leu Lys Asn Asp Tyr Arg Ile
    770                 775                 780


Leu Ser Gln Thr Leu Cys Glu Leu Cys Asp Asp Arg Asp Glu Ser Pro
785                 790                 795                 800


Asn Leu Phe Leu Lys Lys Asn Arg Arg Leu Arg Lys Cys Val Glu Val
                805                 810                 815


Asp Ile Asn Asn Ala Asp Ser Ser Met Thr Arg Lys Tyr Arg Asn Cys
            820                 825                 830


Ile Ala His Leu Thr Val Val Arg Glu Leu Lys Glu Tyr Ile Gly Asp
        835                 840                 845


Ile Arg Thr Val Asp Ser Tyr Phe Ser Ile Tyr His Tyr Val Met Gln
    850                 855                 860


Arg Cys Ile Thr Lys Arg Glu Asp Asp Thr Lys Gln Glu Glu Lys Ile
865                 870                 875                 880


Lys Tyr Glu Asp Asp Leu Leu Lys Asn His Gly Tyr Thr Lys Asp Phe
                885                 890                 895


Val Lys Ala Leu Asn Ser Pro Phe Gly Tyr Asn Ile Pro Arg Phe Lys
            900                 905                 910
```

Asn Leu Ser Ile Glu Gln Leu Phe Asp Arg Asn Glu Tyr Leu Thr Glu
        915   920   925

Lys

<210> 101
<211> 923
<212> PRT
<213> Artificial Sequence

<220>
<223> M5_gut_metagenome_P18E90k2120140920

<400> 101

Met Ala Lys Lys Asn Lys Met Lys Pro Arg Glu Leu Arg Glu Ala Gln
1   5   10   15

Lys Lys Ala Arg Gln Leu Lys Ala Ala Glu Ile Asn Asn Asn Ala Val
            20   25   30

Pro Ala Ile Ala Ala Met Pro Ala Ala Glu Val Ile Ala Pro Ala Ala
        35   40   45

Glu Lys Lys Lys Ser Ser Val Lys Ala Ala Gly Met Lys Ser Ile Leu
    50   55   60

Val Ser Glu Asn Lys Met Tyr Ile Thr Ser Phe Gly Lys Gly Asn Ser
65   70   75   80

Ala Val Leu Glu Tyr Glu Val Asp Lys Val Asp Asp Asn Asp Tyr Asn
                85   90   95

Lys Thr Gln Leu Ser Ser Lys Asp Asn Ser Asn Ile Glu Leu Gly Asn
            100   105   110

Val Asn Glu Val Asn Ile Thr Phe Ser Ser Arg Arg Gly Phe Glu Ser
        115   120   125

Gly Val Glu Ile Asn Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser
    130   135   140

```
Ser Ser Val Arg Gly Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys
145                 150                 155                 160

Arg Phe Phe Gly Lys Thr Phe Asp Asp Asn Ile His Ile Gln Leu Ile
                165                 170                 175

Tyr Asn Ile Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn
            180                 185                 190

Ile Val Tyr Ala Leu Asn Asn Met Leu Gly Val Lys Gly Ser Glu Ser
        195                 200                 205

Tyr Asp Asp Phe Met Gly Tyr Leu Ser Ala Gln Asn Thr Tyr Tyr Ile
    210                 215                 220

Phe Thr His Pro Asp Lys Ser Asn Leu Ser Asp Lys Val Lys Gly Asn
225                 230                 235                 240

Ile Lys Lys Ser Leu Ser Lys Phe Asn Asp Leu Leu Lys Thr Lys Arg
                245                 250                 255

Leu Gly Tyr Phe Gly Leu Glu Glu Pro Lys Thr Lys Asp Lys Arg Val
            260                 265                 270

Ser Glu Ala Tyr Lys Lys Arg Val Tyr His Met Leu Ala Ile Val Gly
        275                 280                 285

Gln Ile Arg Gln Ser Val Phe His Asp Lys Ser Asn Glu Leu Asp Glu
    290                 295                 300

Tyr Leu Tyr Ser Phe Ile Asp Ile Ile Asp Ser Glu Tyr Arg Asp Thr
305                 310                 315                 320

Leu Asp Tyr Leu Val Asp Glu Arg Phe Asp Ser Ile Asn Lys Gly Phe
                325                 330                 335

Ile Gln Gly Asn Lys Val Asn Ile Ser Leu Leu Ile Asp Met Met Lys
            340                 345                 350
```

```
Asp Asp Tyr Glu Ala Asp Asp Ile Ile Arg Leu Tyr Tyr Asp Phe Ile
        355                 360                 365

Val Leu Lys Ser Gln Lys Asn Leu Gly Phe Ser Ile Lys Lys Leu Arg
    370                 375                 380

Glu Lys Met Leu Glu Glu Tyr Gly Phe Arg Phe Lys Asp Lys Gln Tyr
385                 390                 395                 400

Asp Ser Val Arg Ser Lys Met Tyr Lys Ile Met Asp Phe Leu Leu Phe
                405                 410                 415

Cys Asn Tyr Tyr Arg Asn Asp Val Val Ala Gly Glu Ala Leu Val Arg
            420                 425                 430

Lys Leu Arg Phe Ser Met Thr Asp Asp Glu Lys Glu Gly Ile Tyr Ala
        435                 440                 445

Asp Glu Ala Ala Lys Leu Trp Gly Lys Phe Arg Asn Asp Phe Glu Asn
    450                 455                 460

Ile Ala Asp His Met Asn Gly Asp Val Ile Lys Glu Leu Gly Lys Ala
465                 470                 475                 480

Asp Met Asp Phe Asp Glu Lys Ile Leu Asp Ser Glu Lys Lys Asn Ala
                485                 490                 495

Ser Asp Leu Leu Tyr Phe Ser Lys Met Ile Tyr Met Leu Thr Tyr Phe
            500                 505                 510

Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Ser Lys
        515                 520                 525

Phe Asp Asn Ile Lys Glu Phe Leu Lys Ile Met Lys Ser Ser Ala Val
    530                 535                 540

Asp Val Glu Cys Glu Leu Thr Ala Gly Tyr Lys Leu Phe Asn Asp Ser
545                 550                 555                 560
```

```
Gln Arg Ile Thr Asn Glu Leu Phe Ile Val Lys Asn Ile Ala Ser Met
                565                 570                 575

Arg Lys Pro Ala Ala Ser Ala Lys Leu Thr Met Phe Arg Asp Ala Leu
            580                 585                 590

Thr Ile Leu Gly Ile Asp Asp Asn Ile Thr Asp Asp Arg Ile Ser Glu
        595                 600                 605

Ile Leu Lys Leu Lys Glu Lys Gly Lys Gly Ile His Gly Leu Arg Asn
    610                 615                 620

Phe Ile Thr Asn Asn Val Ile Glu Ser Ser Arg Phe Val Tyr Leu Ile
625                 630                 635                 640

Lys Tyr Ala Asn Ala Gln Lys Ile Arg Glu Val Ala Lys Asn Glu Lys
                645                 650                 655

Val Val Met Phe Val Leu Gly Gly Ile Pro Asp Thr Gln Ile Glu Arg
            660                 665                 670

Tyr Tyr Lys Ser Cys Val Glu Phe Pro Asp Met Asn Ser Ser Leu Glu
        675                 680                 685

Val Lys Arg Ser Glu Leu Ala Arg Met Ile Lys Asn Ile Cys Phe Asp
    690                 695                 700

Asp Phe Lys Asn Val Lys Gln Gln Ala Lys Gly Arg Glu Asn Val Ala
705                 710                 715                 720

Lys Glu Arg Ala Lys Ala Val Ile Gly Leu Tyr Leu Thr Val Met Tyr
                725                 730                 735

Leu Leu Val Lys Asn Leu Val Asn Val Asn Ala Arg Tyr Val Ile Ala
            740                 745                 750

Ile His Cys Leu Glu Arg Asp Phe Gly Leu Tyr Lys Glu Ile Val Ser
        755                 760                 765
```

```
Glu Leu Ala Ser Lys Asn Leu Lys Asn Asp Tyr Arg Ile Leu Ser Gln
    770                 775                 780

Thr Leu Cys Glu Leu Cys Asp Lys Ser Pro Asn Leu Phe Leu Lys Lys
785                 790                 795                 800

Asn Glu Arg Leu Arg Lys Cys Val Glu Val Asp Ile Asn Asn Ala Asp
                805                 810                 815

Ser Ser Met Thr Arg Lys Tyr Arg Asn Cys Ile Ala His Leu Thr Val
            820                 825                 830

Val Arg Glu Leu Lys Glu Tyr Ile Gly Asp Ile Arg Ala Val Asp Ser
        835                 840                 845

Tyr Phe Ser Ile Tyr His Tyr Val Met Gln Arg Cys Ile Thr Lys Arg
    850                 855                 860

Gly Asn Asp Thr Lys Gln Glu Asp Lys Ile Lys Tyr Glu Asp Asp Leu
865                 870                 875                 880

Leu Lys Asn His Gly Tyr Thr Lys Asp Phe Val Lys Ala Leu Asn Ser
                885                 890                 895

Pro Phe Gly Tyr Asn Ile Pro Arg Phe Lys Asn Leu Ser Ile Glu Gln
            900                 905                 910

Leu Phe Asp Arg Asn Glu Tyr Leu Thr Glu Lys
        915                 920
```

<210> 102
<211> 923
<212> PRT
<213> Artificial Sequence

<220>
<223> M21_gut_metagenome_P18E0k2120140920

<400> 102

```
Met Ala Lys Lys Asn Lys Met Lys Pro Arg Glu Leu Arg Glu Ala Gln
1               5                  10                  15

Lys Lys Ala Arg Gln Leu Lys Ala Ala Glu Ile Asn Asn Asn Ala Val
            20                  25                  30

Pro Ala Ile Ala Ala Met Pro Ala Ala Glu Val Ile Ala Pro Ala Ala
        35                  40                  45

Glu Lys Lys Lys Ser Ser Val Lys Ala Ala Gly Met Lys Ser Ile Leu
    50                  55                  60

Val Ser Glu Asn Lys Met Tyr Ile Thr Ser Phe Gly Lys Gly Asn Ser
65                  70                  75                  80

Ala Val Leu Glu Tyr Glu Val Asp Lys Val Asp Asp Asn Asp Tyr Asn
                85                  90                  95

Lys Thr Gln Leu Ser Ser Lys Asp Asn Ser Asn Ile Glu Leu Gly Asn
            100                 105                 110

Val Asn Glu Val Asn Ile Thr Phe Ser Ser Arg Arg Gly Phe Glu Ser
        115                 120                 125

Gly Val Glu Ile Asn Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser
    130                 135                 140

Ser Ser Val Arg Gly Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys
145                 150                 155                 160

Arg Phe Phe Gly Lys Thr Phe Asp Asp Asn Ile His Ile Gln Leu Ile
                165                 170                 175

Tyr Asn Ile Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn
            180                 185                 190

Ile Val Tyr Ala Leu Asn Asn Met Leu Gly Val Lys Gly Ser Glu Ser
        195                 200                 205
```

```
Tyr Asp Asp Phe Met Gly Tyr Leu Ser Ala Gln Asn Thr Tyr Tyr Ile
    210                 215                 220

Phe Thr His Pro Asp Lys Ser Asn Leu Ser Asp Lys Val Lys Gly Asn
225                 230                 235                 240

Ile Lys Lys Ser Leu Ser Lys Phe Asn Asp Leu Leu Lys Thr Lys Arg
                245                 250                 255

Leu Gly Tyr Phe Gly Leu Glu Glu Pro Lys Thr Lys Asp Lys Arg Val
            260                 265                 270

Ser Glu Ala Tyr Lys Lys Arg Val Tyr His Met Leu Ala Ile Val Gly
        275                 280                 285

Gln Ile Arg Gln Ser Val Phe His Asp Lys Ser Asn Glu Leu Asp Glu
    290                 295                 300

Tyr Leu Tyr Ser Phe Ile Asp Ile Ile Asp Ser Glu Tyr Arg Asp Thr
305                 310                 315                 320

Leu Asp Tyr Leu Val Asp Glu Arg Phe Asp Ser Ile Asn Lys Gly Phe
                325                 330                 335

Ile Gln Gly Asn Lys Val Asn Ile Ser Leu Leu Ile Asp Met Met Lys
            340                 345                 350

Asp Asp Tyr Glu Ala Asp Asp Ile Ile Arg Leu Tyr Tyr Asp Phe Ile
        355                 360                 365

Val Leu Lys Ser Gln Lys Asn Leu Gly Phe Ser Ile Lys Lys Leu Arg
    370                 375                 380

Glu Lys Met Leu Glu Glu Tyr Gly Phe Arg Phe Lys Asp Lys Gln Tyr
385                 390                 395                 400

Asp Ser Val Arg Ser Lys Met Tyr Lys Ile Met Asp Phe Leu Leu Phe
                405                 410                 415
```

```
Cys Asn Tyr Tyr Arg Asn Asp Val Val Ala Gly Glu Ala Leu Val Arg
            420                 425                 430

Lys Leu Arg Phe Ser Met Thr Asp Asp Glu Lys Glu Gly Ile Tyr Ala
        435                 440                 445

Asp Glu Ala Ala Lys Leu Trp Gly Lys Phe Arg Asn Asp Phe Glu Asn
    450                 455                 460

Ile Ala Asp His Met Asn Gly Asp Val Ile Lys Glu Leu Gly Lys Ala
465                 470                 475                 480

Asp Met Asp Phe Asp Glu Lys Ile Leu Asp Ser Glu Lys Lys Asn Ala
                485                 490                 495

Ser Asp Leu Leu Tyr Phe Ser Lys Met Ile Tyr Met Leu Thr Tyr Phe
            500                 505                 510

Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Ser Lys
        515                 520                 525

Phe Asp Asn Ile Lys Glu Phe Leu Lys Ile Met Lys Ser Ser Ala Val
    530                 535                 540

Asp Val Glu Cys Glu Leu Thr Ala Gly Tyr Lys Leu Phe Asn Asp Ser
545                 550                 555                 560

Gln Arg Ile Thr Asn Glu Leu Phe Ile Val Lys Asn Ile Ala Ser Met
                565                 570                 575

Arg Lys Pro Ala Ala Ser Ala Lys Leu Thr Met Phe Arg Asp Ala Leu
            580                 585                 590

Thr Ile Leu Gly Ile Asp Asp Asn Ile Thr Asp Asp Arg Ile Ser Glu
        595                 600                 605

Ile Leu Lys Leu Lys Glu Lys Gly Lys Gly Ile His Gly Leu Arg Asn
    610                 615                 620
```

```
Phe Ile Thr Asn Asn Val Ile Glu Ser Ser Arg Phe Val Tyr Leu Ile
625                 630                 635                 640

Lys Tyr Ala Asn Ala Gln Lys Ile Arg Glu Val Ala Lys Asn Glu Lys
                645                 650                 655

Val Val Met Phe Val Leu Gly Gly Ile Pro Asp Thr Gln Ile Glu Arg
            660                 665                 670

Tyr Tyr Lys Ser Cys Val Glu Phe Pro Asp Met Asn Ser Ser Leu Glu
        675                 680                 685

Val Lys Arg Ser Glu Leu Ala Arg Met Ile Lys Asn Ile Cys Phe Asp
    690                 695                 700

Asp Phe Lys Asn Val Lys Gln Gln Ala Lys Gly Arg Glu Asn Val Ala
705                 710                 715                 720

Lys Glu Arg Ala Lys Ala Val Ile Gly Leu Tyr Leu Thr Val Met Tyr
                725                 730                 735

Leu Leu Val Lys Asn Leu Val Asn Val Asn Ala Arg Tyr Val Ile Ala
            740                 745                 750

Ile His Cys Leu Glu Arg Asp Phe Gly Leu Tyr Lys Glu Ile Val Ser
        755                 760                 765

Glu Leu Ala Ser Lys Asn Leu Lys Asn Asp Tyr Arg Ile Leu Ser Gln
    770                 775                 780

Thr Leu Cys Glu Leu Cys Asp Lys Ser Pro Asn Leu Phe Leu Lys Lys
785                 790                 795                 800

Asn Glu Arg Leu Arg Lys Cys Val Glu Val Asp Ile Asn Asn Ala Asp
                805                 810                 815

Ser Ser Met Thr Arg Lys Tyr Arg Asn Cys Ile Ala His Leu Thr Val
            820                 825                 830
```

```
Val Arg Glu Leu Lys Glu Tyr Ile Gly Asp Ile Arg Ala Val Asp Ser
        835                 840                 845


Tyr Phe Ser Ile Tyr His Tyr Val Met Gln Arg Cys Ile Thr Lys Arg
    850                 855                 860


Gly Asn Asp Thr Lys Gln Glu Asp Lys Ile Lys Tyr Glu Asp Asp Leu
865                 870                 875                 880


Leu Lys Asn His Gly Tyr Thr Lys Asp Phe Val Lys Ala Leu Asn Ser
                885                 890                 895


Pro Phe Gly Tyr Asn Ile Pro Arg Phe Lys Asn Leu Ser Ile Glu Gln
            900                 905                 910


Leu Phe Asp Arg Asn Glu Tyr Leu Thr Glu Lys
        915                 920


<210>  103
<211>  926
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  M7_gut_metagenome_P38C7k2120140920_c4841000003


<220>
<221>  misc_feature
<222>  (551)..(551)
<223>  Xaa can be any naturally occurring amino acid

<400>  103

Met Ala Lys Lys Asn Lys Met Lys Pro Arg Glu Leu Arg Glu Ala Gln
1               5                   10                  15


Lys Lys Ala Arg Gln Leu Lys Ala Ala Glu Ile Asn Asn Asn Ala Ala
            20                  25                  30


Pro Ala Ile Ala Ala Met Pro Ala Ala Glu Ala Ile Ala Pro Ala Ala
        35                  40                  45
```

```
Glu Lys Lys Lys Ser Ser Val Lys Ala Ala Gly Met Lys Ser Ile Phe
    50                  55                  60

Val Ser Glu Asn Lys Met Tyr Ile Thr Ser Phe Gly Lys Gly Asn Ser
65                  70                  75                  80

Ala Val Leu Glu Tyr Glu Val Asp Lys Val Asp Asn Asn Val Tyr Asn
                85                  90                  95

Gln Thr Gln Leu Ser Ser Glu Asp Ser Ser Asn Ile Glu Leu Cys Gly
            100                 105                 110

Val Thr Lys Val Asn Ile Thr Phe Ser Ser Lys His Gly Leu Glu Ser
        115                 120                 125

Gly Val Glu Ile Ser Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser
    130                 135                 140

Ser Pro Val Arg Trp Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys
145                 150                 155                 160

Arg Phe Phe Gly Lys Thr Phe Asp Asp Asn Ile His Ile Gln Leu Ile
                165                 170                 175

Tyr Asn Ile Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn
            180                 185                 190

Ile Val Tyr Ala Leu Asn Asn Met Leu Gly Ile Lys Asp Ser Glu Ser
        195                 200                 205

Tyr Asp Asp Phe Met Gly Tyr Leu Ser Ala Arg Asn Thr Tyr Lys Val
    210                 215                 220

Phe Thr His Pro Asp Lys Ser Asn Leu Ser Asp Lys Val Lys Gly Asn
225                 230                 235                 240

Ile Lys Lys Ser Phe Ser Thr Phe Asn Asp Leu Leu Lys Thr Lys Arg
                245                 250                 255
```

```
Leu Gly Tyr Phe Gly Leu Glu Glu Pro Lys Thr Lys Asp Thr Arg Val
            260                 265                 270

Ser Glu Ala Tyr Lys Lys Arg Val Tyr His Met Leu Ala Ile Val Gly
        275                 280                 285

Gln Ile Arg Gln Cys Val Phe His Asp Leu Ser Glu His Ser Glu Tyr
    290                 295                 300

Asp Leu Tyr Ser Phe Ile Asp Asn Ser Lys Lys Val Tyr Arg Glu Cys
305                 310                 315                 320

Arg Glu Thr Leu Asn Tyr Leu Val Asp Glu Arg Phe Asp Ser Ile Asn
                325                 330                 335

Lys Gly Phe Ile Gln Gly Asn Lys Val Asn Ile Ser Leu Leu Ile Asp
            340                 345                 350

Met Met Lys Asp Asp Tyr Glu Ala Asp Asp Ile Ile His Leu Tyr Tyr
        355                 360                 365

Asp Phe Ile Val Leu Lys Ser Gln Lys Asn Leu Gly Phe Ser Ile Lys
    370                 375                 380

Lys Leu Arg Glu Lys Met Leu Asp Glu Tyr Gly Phe Arg Phe Lys Asp
385                 390                 395                 400

Lys Gln Tyr Asp Ser Val Arg Ser Lys Met Tyr Lys Leu Met Asp Phe
                405                 410                 415

Leu Leu Phe Cys Asn Tyr Tyr Arg Asn Asp Val Val Ala Gly Glu Ala
            420                 425                 430

Leu Val Arg Lys Leu Arg Phe Ser Met Thr Asp Asp Glu Lys Glu Gly
        435                 440                 445

Ile Tyr Ala Asp Glu Ala Glu Lys Leu Trp Gly Lys Phe Arg Asn Asp
    450                 455                 460
```

```
Phe Glu Asn Ile Ala Asp His Met Asn Gly Asp Val Ile Lys Glu Leu
465                 470                 475                 480


Gly Lys Ala Asp Met Asp Phe Asp Glu Lys Ile Ile Asp Ser Glu Lys
                485                 490                 495


Lys Asn Ala Ser Asp Leu Leu Tyr Phe Ser Lys Met Ile Tyr Met Leu
            500                 505                 510


Thr Tyr Phe Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr Thr Leu
        515                 520                 525


Ile Ser Lys Phe Asp Asn Ile Lys Glu Phe Leu Lys Ile Met Lys Ser
    530                 535                 540


Ser Ala Val Asp Val Glu Xaa Glu Leu Thr Ala Gly Tyr Lys Leu Phe
545                 550                 555                 560


Asn Asp Ser Gln Arg Ile Thr Asn Glu Leu Phe Ile Val Lys Asn Ile
                565                 570                 575


Ala Ser Met Arg Lys Pro Ala Ala Ser Ala Lys Leu Thr Met Phe Arg
            580                 585                 590


Asp Ala Leu Thr Ile Leu Gly Ile Asp Asp Asn Ile Thr Asp Asp Arg
        595                 600                 605


Ile Ser Glu Ile Leu Lys Leu Lys Glu Lys Gly Lys Gly Ile His Gly
    610                 615                 620


Leu Arg Asn Phe Ile Thr Asn Asn Val Ile Glu Ser Ser Arg Phe Val
625                 630                 635                 640


Tyr Leu Ile Lys Tyr Ala Asn Ala Gln Lys Ile Arg Glu Val Ala Lys
                645                 650                 655


Asn Glu Lys Val Val Met Phe Val Leu Gly Gly Ile Pro Asp Thr Gln
            660                 665                 670
```

```
Ile Glu Arg Tyr Tyr Lys Ser Cys Val Glu Phe Pro Asp Met Asn Ser
        675                 680                 685

Ser Leu Gly Val Lys Arg Ser Glu Leu Ala Arg Met Ile Lys Asn Ile
    690                 695                 700

Ser Phe Asp Asp Phe Lys Asn Val Lys Gln Gln Ala Lys Gly Arg Glu
705                 710                 715                 720

Asn Val Ala Lys Glu Arg Ala Lys Ala Val Ile Gly Leu Tyr Leu Thr
                725                 730                 735

Val Met Tyr Leu Leu Val Lys Asn Leu Val Asn Val Asn Ala Arg Tyr
            740                 745                 750

Val Ile Ala Ile His Cys Leu Glu Arg Asp Phe Gly Leu Tyr Lys Glu
        755                 760                 765

Ile Ile Pro Glu Leu Ala Ser Lys Asn Leu Lys Asn Asp Tyr Arg Ile
    770                 775                 780

Leu Ser Gln Thr Leu Cys Glu Leu Cys Asp Lys Ser Pro Asn Leu Phe
785                 790                 795                 800

Leu Lys Lys Asn Glu Arg Leu Arg Lys Cys Val Glu Val Asp Ile Asn
                805                 810                 815

Asn Ala Asp Ser Ser Met Thr Arg Lys Tyr Arg Asn Cys Ile Ala His
            820                 825                 830

Leu Thr Val Val Arg Glu Leu Lys Glu Tyr Ile Gly Asp Ile Cys Thr
        835                 840                 845

Val Asp Ser Tyr Phe Ser Ile Tyr His Tyr Val Met Gln Arg Cys Ile
    850                 855                 860

Thr Lys Arg Glu Asn Asp Thr Lys Gln Glu Glu Lys Ile Lys Tyr Glu
865                 870                 875                 880
```

Asp Asp Leu Leu Lys Asn His Gly Tyr Thr Lys Asp Phe Val Lys Ala
885 890 895

Leu Asn Ser Pro Phe Gly Tyr Asn Ile Pro Arg Phe Lys Asn Leu Ser
900 905 910

Ile Glu Gln Leu Phe Asp Arg Asn Glu Tyr Leu Thr Glu Lys
915 920 925

<210> 104
<211> 918
<212> PRT
<213> Artificial Sequence

<220>
<223> Ruminococcus_bicirculans

<400> 104

Met Ala Lys Lys Asn Lys Met Lys Pro Arg Glu Leu Arg Glu Ala Gln
1 5 10 15

Lys Lys Ala Arg Gln Leu Lys Ala Ala Glu Ile Asn Asn Asn Ala Val
20 25 30

Pro Ala Ile Ala Ala Met Pro Ala Ala Glu Ala Ala Ala Pro Ala Ala
35 40 45

Glu Lys Lys Lys Ser Ser Val Lys Ala Ala Gly Met Lys Ser Ile Leu
50 55 60

Val Ser Glu Asn Lys Met Tyr Ile Thr Ser Phe Gly Lys Gly Asn Ser
65 70 75 80

Ala Val Leu Glu Tyr Glu Val Asp Asn Asn Asp Tyr Asn Lys Thr Gln
85 90 95

Leu Ser Ser Lys Asp Asn Ser Asn Ile Glu Leu Cys Asp Val Gly Lys
100 105 110

Val Asn Ile Thr Phe Ser Ser Arg Arg Gly Phe Glu Ser Gly Val Glu
115 120 125

```
Ile Asn Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser Ser Ser Val
    130                 135                 140

Arg Gly Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys Arg Phe Phe
145                 150                 155                 160

Gly Lys Asn Phe Asp Asp Asn Ile His Ile Gln Leu Ile Tyr Asn Ile
                165                 170                 175

Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn Ile Val Tyr
            180                 185                 190

Ala Leu Asn Asn Met Leu Gly Glu Gly Asp Glu Ser Asn Tyr Asp Phe
        195                 200                 205

Met Gly Tyr Leu Ser Thr Phe Asn Thr Tyr Lys Val Phe Thr Asn Pro
    210                 215                 220

Asn Gly Ser Thr Leu Ser Asp Asp Lys Lys Glu Asn Ile Arg Lys Ser
225                 230                 235                 240

Leu Ser Lys Phe Asn Ala Leu Leu Lys Thr Lys Arg Leu Gly Tyr Phe
                245                 250                 255

Gly Leu Glu Glu Pro Lys Thr Lys Asp Thr Arg Ala Ser Glu Ala Tyr
            260                 265                 270

Lys Lys Arg Val Tyr His Met Leu Ala Ile Val Gly Gln Ile Arg Gln
        275                 280                 285

Cys Val Phe His Asp Lys Ser Gly Ala Lys Arg Phe Asp Leu Tyr Ser
    290                 295                 300

Phe Ile Asn Asn Ile Asp Pro Glu Tyr Arg Glu Thr Leu Asp Tyr Leu
305                 310                 315                 320

Val Asp Glu Arg Phe Asp Ser Ile Asn Lys Gly Phe Ile Gln Gly Asn
                325                 330                 335
```

```
Lys Val Asn Ile Ser Leu Leu Ile Asp Met Met Lys Gly Tyr Glu Ala
            340                 345                 350

Asp Asp Ile Ile Arg Leu Tyr Tyr Asp Phe Ile Val Leu Lys Ser Gln
        355                 360                 365

Lys Asn Leu Gly Phe Ser Ile Lys Lys Leu Arg Glu Lys Met Leu Asp
    370                 375                 380

Glu Tyr Gly Phe Arg Phe Lys Asp Lys Gln Tyr Asp Ser Val Arg Ser
385                 390                 395                 400

Lys Met Tyr Lys Leu Met Asp Phe Leu Leu Phe Cys Asn Tyr Tyr Arg
                405                 410                 415

Asn Asp Ile Ala Ala Gly Glu Ser Leu Val Arg Lys Leu Arg Phe Ser
            420                 425                 430

Met Thr Asp Asp Glu Lys Glu Gly Ile Tyr Ala Asp Glu Ala Ala Lys
        435                 440                 445

Leu Trp Gly Lys Phe Arg Asn Asp Phe Glu Asn Ile Ala Asp His Met
    450                 455                 460

Asn Gly Asp Val Ile Lys Glu Leu Gly Lys Ala Asp Met Asp Phe Asp
465                 470                 475                 480

Glu Lys Ile Leu Asp Ser Glu Lys Lys Asn Ala Ser Asp Leu Leu Tyr
                485                 490                 495

Phe Ser Lys Met Ile Tyr Met Leu Thr Tyr Phe Leu Asp Gly Lys Glu
            500                 505                 510

Ile Asn Asp Leu Leu Thr Thr Leu Ile Ser Lys Phe Asp Asn Ile Lys
        515                 520                 525

Glu Phe Leu Lys Ile Met Lys Ser Ser Ala Val Asp Val Glu Cys Glu
    530                 535                 540
```

```
Leu Thr Ala Gly Tyr Lys Leu Phe Asn Asp Ser Gln Arg Ile Thr Asn
545                 550                 555                 560

Glu Leu Phe Ile Val Lys Asn Ile Ala Ser Met Arg Lys Pro Ala Ala
                565                 570                 575

Ser Ala Lys Leu Thr Met Phe Arg Asp Ala Leu Thr Ile Leu Gly Ile
            580                 585                 590

Asp Asp Lys Ile Thr Asp Asp Arg Ile Ser Glu Ile Leu Lys Leu Lys
        595                 600                 605

Glu Lys Gly Lys Gly Ile His Gly Leu Arg Asn Phe Ile Thr Asn Asn
    610                 615                 620

Val Ile Glu Ser Ser Arg Phe Val Tyr Leu Ile Lys Tyr Ala Asn Ala
625                 630                 635                 640

Gln Lys Ile Arg Glu Val Ala Lys Asn Glu Lys Val Val Met Phe Val
                645                 650                 655

Leu Gly Gly Ile Pro Asp Thr Gln Ile Glu Arg Tyr Tyr Lys Ser Cys
            660                 665                 670

Val Glu Phe Pro Asp Met Asn Ser Ser Leu Gly Val Lys Arg Ser Glu
        675                 680                 685

Leu Ala Arg Met Ile Lys Asn Ile Ser Phe Asp Asp Phe Lys Asn Val
    690                 695                 700

Lys Gln Gln Ala Lys Gly Arg Glu Asn Val Ala Lys Glu Arg Ala Lys
705                 710                 715                 720

Ala Val Ile Gly Leu Tyr Leu Thr Val Met Tyr Leu Leu Val Lys Asn
                725                 730                 735

Leu Val Asn Val Asn Ala Arg Tyr Val Ile Ala Ile His Cys Leu Glu
            740                 745                 750
```

```
Arg Asp Phe Gly Leu Tyr Lys Glu Ile Ile Pro Glu Leu Ala Ser Lys
        755                 760                 765

Asn Leu Lys Asn Asp Tyr Arg Ile Leu Ser Gln Thr Leu Cys Glu Leu
    770                 775                 780

Cys Asp Lys Ser Pro Asn Leu Phe Leu Lys Lys Asn Glu Arg Leu Arg
785                 790                 795                 800

Lys Cys Val Glu Val Asp Ile Asn Asn Ala Asp Ser Ser Met Thr Arg
                805                 810                 815

Lys Tyr Arg Asn Cys Ile Ala His Leu Thr Val Val Arg Glu Leu Lys
            820                 825                 830

Glu Tyr Ile Gly Asp Ile Cys Thr Val Asp Ser Tyr Phe Ser Ile Tyr
        835                 840                 845

His Tyr Val Met Gln Arg Cys Ile Thr Lys Arg Glu Asn Asp Thr Lys
    850                 855                 860

Gln Glu Glu Lys Ile Lys Tyr Glu Asp Asp Leu Leu Lys Asn His Gly
865                 870                 875                 880

Tyr Thr Lys Asp Phe Val Lys Ala Leu Asn Ser Pro Phe Gly Tyr Asn
                885                 890                 895

Ile Pro Arg Phe Lys Asn Leu Ser Ile Glu Gln Leu Phe Asp Arg Asn
            900                 905                 910

Glu Tyr Leu Thr Glu Lys
        915


<210>  105
<211>  920
<212>  PRT
<213>  Artificial Sequence

<220>
```

<223> our_cloned_CAG_Ncas1

<400> 105

```
Met Ala Lys Lys Asn Lys Met Lys Pro Arg Glu Leu Arg Glu Ala Gln
1               5                  10                  15


Lys Lys Ala Arg Gln Leu Lys Ala Ala Glu Ile Asn Asn Asn Ala Ala
            20                  25                  30


Pro Ala Ile Ala Ala Met Pro Ala Ala Glu Ala Ala Ala Pro Ala Ala
        35                  40                  45


Glu Lys Lys Lys Ser Ser Val Lys Ala Ala Gly Met Lys Ser Ile Leu
    50                  55                  60


Val Ser Glu Asn Lys Met Tyr Ile Thr Ser Phe Gly Lys Gly Asn Ser
65                  70                  75                  80


Ala Val Leu Glu Tyr Glu Val Asp Asn Asn Asp Tyr Asn Gln Thr Gln
                85                  90                  95


Leu Ser Ser Lys Gly Ser Ser Asn Ile Glu Leu His Gly Val Asn Glu
            100                 105                 110


Val Asn Ile Thr Phe Ser Ser Lys His Gly Phe Glu Ser Gly Val Glu
        115                 120                 125


Ile Asn Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser Ser Pro Val
    130                 135                 140


Arg Trp Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys Arg Phe Phe
145                 150                 155                 160


Gly Lys Thr Phe Asp Asp Asn Ile His Ile Gln Leu Ile Tyr Asn Ile
                165                 170                 175


Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn Ile Val Tyr
            180                 185                 190
```

```
Ala Leu Asn Asn Met Leu Gly Val Lys Gly Ser Glu Ser His Asp Asp
        195                 200                 205

Phe Ile Gly Tyr Leu Ser Thr Asn Asn Thr Tyr Asp Val Phe Ile Asp
    210                 215                 220

Pro Asp Asn Ser Ser Leu Ser Asp Asp Lys Lys Ala Asn Val Arg Lys
225                 230                 235                 240

Ser Leu Ser Lys Phe Asn Val Leu Leu Lys Thr Lys Arg Leu Gly Tyr
                245                 250                 255

Phe Gly Leu Glu Glu Pro Lys Thr Lys Asp Thr Arg Val Ser Gln Ala
            260                 265                 270

Tyr Lys Lys Arg Val Tyr Leu Ala Ile Val Gly Gln Ile Arg Gln Cys
        275                 280                 285

Val Phe His Asp Lys Ser Gly Ala Lys Arg Phe Asp Leu Tyr Ser Phe
    290                 295                 300

Ile Asn Asn Ile Asp Pro Glu Tyr Arg Glu Thr Leu Asp Tyr Leu Val
305                 310                 315                 320

Asp Glu Arg Phe Asp Ser Ile Asn Lys Gly Phe Ile Glu Gly Asn Lys
                325                 330                 335

Ile Asn Ile Ser Leu Leu Ile Asp Met Met Lys Gly Tyr Glu Ala Asp
            340                 345                 350

Asp Ile Ile Arg Leu Tyr Tyr Asp Phe Ile Val Leu Lys Ser Gln Lys
        355                 360                 365

Asn Leu Gly Phe Ser Ile Lys Lys Leu Arg Glu Lys Met Leu Asp Glu
    370                 375                 380

Tyr Gly Phe Arg Phe Lys Asp Lys Gln Tyr Asp Pro Val Arg Ser Lys
385                 390                 395                 400
```

```
Met Tyr Lys Leu Met Asp Phe Leu Leu Phe Cys Asn His Tyr Arg Asn
                405                 410                 415

Asp Val Ala Ala Gly Glu Ala Leu Val Arg Lys Leu Arg Phe Ser Met
            420                 425                 430

Thr Asp Asp Glu Lys Glu Gly Ile Tyr Ala Asp Glu Ala Ala Lys Leu
        435                 440                 445

Trp Gly Lys Phe Arg Asn Asp Phe Glu Asn Ile Ala Asp His Met Asn
    450                 455                 460

Gly Asp Val Ile Lys Glu Leu Gly Lys Ala Asp Met Asp Phe Asp Glu
465                 470                 475                 480

Lys Ile Leu Asp Ser Glu Lys Lys Asn Ala Ser Asp Leu Leu Tyr Phe
                485                 490                 495

Ser Lys Met Ile Tyr Met Leu Thr Tyr Phe Leu Asp Gly Lys Glu Ile
            500                 505                 510

Asn Asp Leu Leu Thr Thr Leu Ile Ser Lys Phe Asp Asn Ile Lys Glu
        515                 520                 525

Phe Leu Lys Ile Met Lys Ser Ser Ala Val Asp Val Glu Cys Glu Leu
    530                 535                 540

Thr Ala Gly Tyr Lys Leu Phe Asn Asp Ser Gln Arg Ile Thr Asn Glu
545                 550                 555                 560

Leu Phe Ile Val Lys Asn Ile Ala Ser Met Arg Lys Pro Ala Ala Ser
                565                 570                 575

Ala Lys Leu Thr Met Phe Arg Asp Ala Leu Thr Ile Leu Gly Ile Asp
            580                 585                 590

Asp Asn Ile Thr Asp Asp Arg Ile Ser Glu Ile Leu Lys Leu Lys Glu
        595                 600                 605
```

```
Lys Gly Lys Gly Ile His Gly Leu Arg Asn Phe Ile Thr Asn Asn Val
    610                 615                 620

Ile Glu Ser Ser Arg Phe Val Tyr Leu Ile Lys Tyr Ala Asn Ala Gln
625                 630                 635                 640

Lys Ile Arg Glu Val Ala Lys Asn Glu Lys Val Val Met Phe Val Leu
                645                 650                 655

Gly Gly Ile Pro Asp Thr His Ile Glu Arg Tyr Tyr Lys Ser Cys Val
            660                 665                 670

Glu Phe Pro Asp Met Asn Ser Ser Leu Glu Ala Lys Arg Ser Glu Leu
        675                 680                 685

Ala Arg Met Ile Lys Asn Ile Ser Phe Asp Asp Phe Lys Asn Val Lys
    690                 695                 700

Gln Gln Ala Lys Gly Arg Glu Asn Val Ala Lys Glu Arg Ala Lys Ala
705                 710                 715                 720

Val Ile Gly Leu Tyr Leu Thr Val Met Tyr Leu Leu Val Lys Asn Leu
                725                 730                 735

Val Asn Val Asn Ala Arg Tyr Val Ile Ala Ile His Cys Leu Glu Arg
            740                 745                 750

Asp Phe Gly Leu Tyr Lys Glu Ile Ile Pro Glu Leu Ala Ser Lys Asn
        755                 760                 765

Leu Lys Asn Asp Tyr Arg Ile Leu Ser Gln Thr Leu Cys Glu Leu Cys
    770                 775                 780

Asp Asp Arg Asp Glu Ser Pro Asn Leu Phe Leu Lys Lys Asn Lys Arg
785                 790                 795                 800

Leu Arg Lys Cys Val Glu Val Asp Ile Asn Asn Ala Asp Ser Ser Met
                805                 810                 815
```

```
Thr Arg Lys Tyr Arg Asn Cys Ile Ala His Leu Thr Val Val Arg Glu
            820                 825                 830


Leu Lys Glu Tyr Ile Gly Asp Ile Arg Thr Val Asp Ser Tyr Phe Ser
        835                 840                 845


Ile Tyr His Tyr Val Met Gln Arg Cys Ile Thr Lys Arg Glu Asn Asp
    850                 855                 860


Thr Lys Gln Glu Glu Lys Ile Lys Tyr Glu Asp Asp Leu Leu Lys Asn
865                 870                 875                 880


His Gly Tyr Thr Lys Asp Phe Val Lys Ala Leu Asn Ser Pro Phe Gly
                885                 890                 895


Tyr Asn Ile Pro Arg Phe Lys Asn Leu Ser Ile Glu Gln Leu Phe Asp
            900                 905                 910


Arg Asn Glu Tyr Leu Thr Glu Lys
        915                 920


<210>  106
<211>  922
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  consensus_1


<220>
<221>  misc_feature
<222>  (89)..(89)
<223>  Xaa can be any naturally occurring amino acid

<220>
<221>  misc_feature
<222>  (386)..(386)
<223>  Xaa can be any naturally occurring amino acid

<220>
<221>  misc_feature
<222>  (789)..(790)
<223>  Xaa can be any naturally occurring amino acid
```

<400> 106

```
Met Ala Lys Lys Asn Lys Met Lys Pro Arg Glu Leu Arg Glu Ala Gln
1               5                  10                  15

Lys Lys Ala Arg Gln Leu Lys Ala Ala Glu Ile Asn Asn Asn Ala Ala
            20                  25                  30

Pro Ala Ile Ala Ala Met Pro Ala Ala Glu Val Ile Ala Pro Ala Ala
        35                  40                  45

Glu Lys Lys Lys Ser Ser Val Lys Ala Ala Gly Met Lys Ser Ile Leu
    50                  55                  60

Val Ser Glu Asn Lys Met Tyr Ile Thr Ser Phe Gly Lys Gly Asn Ser
65                  70                  75                  80

Ala Val Leu Glu Tyr Glu Val Asp Xaa Asn Asn Asp Tyr Asn Lys Thr
                85                  90                  95

Gln Leu Ser Ser Lys Asp Asn Ser Asn Ile Glu Leu Gly Asp Val Asn
            100                 105                 110

Glu Val Asn Ile Thr Phe Ser Ser Lys His Gly Phe Glu Ser Gly Val
        115                 120                 125

Glu Ile Asn Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser Ser Pro
    130                 135                 140

Val Arg Trp Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys Arg Phe
145                 150                 155                 160

Phe Gly Lys Thr Phe Asp Asp Asn Ile His Ile Gln Leu Ile Tyr Asn
                165                 170                 175

Ile Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn Ile Val
            180                 185                 190

Tyr Ala Leu Asn Asn Met Leu Gly Ile Lys Gly Ser Glu Ser Tyr Asp
        195                 200                 205
```

```
Asp Phe Met Gly Tyr Leu Ser Ala Arg Asn Thr Tyr Asp Val Phe Thr
    210                 215                 220

His Pro Asp Lys Ser Asn Leu Ser Asp Lys Val Lys Gly Asn Ile Lys
225                 230                 235                 240

Lys Ser Leu Ser Lys Phe Asn Asp Leu Leu Lys Thr Lys Arg Leu Gly
                245                 250                 255

Tyr Phe Gly Leu Glu Glu Pro Lys Thr Lys Asp Thr Arg Val Ser Glu
            260                 265                 270

Ala Tyr Lys Lys Arg Val Tyr His Met Leu Ala Ile Val Gly Gln Ile
        275                 280                 285

Arg Gln Cys Val Phe His Asp Lys Ser Gly Ala Lys Arg Phe Asp Leu
    290                 295                 300

Tyr Ser Phe Ile Asp Asn Ile Asp Pro Glu Tyr Arg Asp Thr Leu Asp
305                 310                 315                 320

Tyr Leu Val Asp Glu Arg Phe Asp Ser Ile Asn Lys Gly Phe Ile Gln
                325                 330                 335

Gly Asn Lys Val Asn Ile Ser Leu Leu Ile Asp Met Met Lys Gly Tyr
            340                 345                 350

Glu Ala Asp Asp Ile Ile Arg Leu Tyr Tyr Asp Phe Ile Val Leu Lys
        355                 360                 365

Ser Gln Lys Asn Leu Gly Phe Ser Ile Lys Lys Leu Arg Glu Lys Met
    370                 375                 380

Leu Xaa Glu Tyr Gly Phe Arg Phe Lys Asp Lys Gln Tyr Asp Ser Val
385                 390                 395                 400

Arg Ser Lys Met Tyr Lys Leu Met Asp Phe Leu Leu Phe Cys Asn Tyr
                405                 410                 415
```

```
Tyr Arg Asn Asp Val Val Ala Gly Glu Ala Leu Val Arg Lys Leu Arg
            420                 425                 430

Phe Ser Met Thr Asp Asp Glu Lys Glu Gly Ile Tyr Ala Asp Glu Ala
        435                 440                 445

Ala Lys Leu Trp Gly Lys Phe Arg Asn Asp Phe Glu Asn Ile Ala Asp
    450                 455                 460

His Met Asn Gly Asp Val Ile Lys Glu Leu Gly Lys Ala Asp Met Asp
465                 470                 475                 480

Phe Asp Glu Lys Ile Leu Asp Ser Glu Lys Lys Asn Ala Ser Asp Leu
                485                 490                 495

Leu Tyr Phe Ser Lys Met Ile Tyr Met Leu Thr Tyr Phe Leu Asp Gly
            500                 505                 510

Lys Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Ser Lys Phe Asp Asn
        515                 520                 525

Ile Lys Glu Phe Leu Lys Ile Met Lys Ser Ser Ala Val Asp Val Glu
    530                 535                 540

Cys Glu Leu Thr Ala Gly Tyr Lys Leu Phe Asn Asp Ser Gln Arg Ile
545                 550                 555                 560

Thr Asn Glu Leu Phe Ile Val Lys Asn Ile Ala Ser Met Arg Lys Pro
                565                 570                 575

Ala Ala Ser Ala Lys Leu Thr Met Phe Arg Asp Ala Leu Thr Ile Leu
            580                 585                 590

Gly Ile Asp Asp Asn Ile Thr Asp Asp Arg Ile Ser Glu Ile Leu Lys
        595                 600                 605

Leu Lys Glu Lys Gly Lys Gly Ile His Gly Leu Arg Asn Phe Ile Thr
    610                 615                 620
```

```
Asn Asn Val Ile Glu Ser Ser Arg Phe Val Tyr Leu Ile Lys Tyr Ala
625                 630                 635                 640

Asn Ala Gln Lys Ile Arg Glu Val Ala Lys Asn Glu Lys Val Val Met
                645                 650                 655

Phe Val Leu Gly Gly Ile Pro Asp Thr Gln Ile Glu Arg Tyr Tyr Lys
            660                 665                 670

Ser Cys Val Glu Phe Pro Asp Met Asn Ser Ser Leu Glu Val Lys Arg
        675                 680                 685

Ser Glu Leu Ala Arg Met Ile Lys Asn Ile Ser Phe Asp Asp Phe Lys
    690                 695                 700

Asn Val Lys Gln Gln Ala Lys Gly Arg Glu Asn Val Ala Lys Glu Arg
705                 710                 715                 720

Ala Lys Ala Val Ile Gly Leu Tyr Leu Thr Val Met Tyr Leu Leu Val
                725                 730                 735

Lys Asn Leu Val Asn Val Asn Ala Arg Tyr Val Ile Ala Ile His Cys
            740                 745                 750

Leu Glu Arg Asp Phe Gly Leu Tyr Lys Glu Ile Ile Pro Glu Leu Ala
        755                 760                 765

Ser Lys Asn Leu Lys Asn Asp Tyr Arg Ile Leu Ser Gln Thr Leu Cys
    770                 775                 780

Glu Leu Cys Asp Xaa Xaa Glu Ser Pro Asn Leu Phe Leu Lys Lys Asn
785                 790                 795                 800

Lys Arg Leu Arg Lys Cys Val Glu Val Asp Ile Asn Asn Ala Asp Ser
                805                 810                 815

Ser Met Thr Arg Lys Tyr Arg Asn Cys Ile Ala His Leu Thr Val Val
            820                 825                 830
```

```
Arg Glu Leu Lys Glu Tyr Ile Gly Asp Ile Arg Thr Val Asp Ser Tyr
        835                 840                 845

Phe Ser Ile Tyr His Tyr Val Met Gln Arg Cys Ile Thr Lys Arg Glu
    850                 855                 860

Asp Asp Thr Lys Gln Glu Glu Lys Ile Lys Tyr Glu Asp Asp Leu Leu
865                 870                 875                 880

Lys Asn His Gly Tyr Thr Lys Asp Phe Val Lys Ala Leu Asn Ser Pro
                885                 890                 895

Phe Gly Tyr Asn Ile Pro Arg Phe Lys Asn Leu Ser Ile Glu Gln Leu
            900                 905                 910

Phe Asp Arg Asn Glu Tyr Leu Thr Glu Lys
        915                 920


<210>  107
<211>  922
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  M18_gut_metagenome_P22E0k2120140920_c3395000078

<400>  107

Met Ala Lys Lys Asn Lys Met Lys Pro Arg Glu Leu Arg Glu Ala Gln
1               5                   10                  15

Lys Lys Ala Arg Gln Leu Lys Ala Ala Glu Ile Asn Asn Asn Ala Ala
            20                  25                  30

Pro Ala Ile Ala Ala Met Pro Ala Ala Glu Val Ile Ala Pro Val Ala
        35                  40                  45

Glu Lys Lys Lys Ser Ser Val Lys Ala Ala Gly Met Lys Ser Ile Leu
    50                  55                  60
```

```
Val Ser Glu Asn Lys Met Tyr Ile Thr Ser Phe Gly Lys Gly Asn Ser
65                  70                  75                  80

Ala Val Leu Glu Tyr Glu Val Asp Asn Asn Asp Tyr Asn Lys Thr Gln
                85                  90                  95

Leu Ser Ser Lys Asp Asn Ser Asn Ile Glu Leu Gly Asp Val Asn Glu
            100                 105                 110

Val Asn Ile Thr Phe Ser Ser Lys His Gly Phe Gly Ser Gly Val Glu
        115                 120                 125

Ile Asn Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser Ser Pro Val
    130                 135                 140

Arg Gly Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys Arg Phe Phe
145                 150                 155                 160

Gly Lys Thr Phe Asp Asp Asn Ile His Ile Gln Leu Ile Tyr Asn Ile
                165                 170                 175

Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn Ile Val Tyr
            180                 185                 190

Ala Leu Asn Asn Met Leu Gly Ile Lys Asp Ser Glu Ser Tyr Asp Asp
        195                 200                 205

Phe Met Gly Tyr Leu Ser Ala Arg Asn Thr Tyr Glu Val Phe Thr His
    210                 215                 220

Pro Asp Lys Ser Asn Leu Ser Asp Lys Val Lys Gly Asn Ile Lys Lys
225                 230                 235                 240

Ser Leu Ser Lys Phe Asn Asp Leu Leu Lys Thr Lys Arg Leu Gly Tyr
                245                 250                 255

Phe Gly Leu Glu Glu Pro Lys Thr Lys Asp Thr Arg Ala Ser Glu Ala
            260                 265                 270
```

```
Tyr Lys Lys Arg Val Tyr His Met Leu Ala Ile Val Gly Gln Ile Arg
        275                 280                 285

Gln Cys Val Phe His Asp Lys Ser Gly Ala Lys Arg Phe Asp Leu Tyr
    290                 295                 300

Ser Phe Ile Asn Asn Ile Asp Pro Glu Tyr Arg Asp Thr Leu Asp Tyr
305                 310                 315                 320

Leu Val Glu Glu Arg Leu Lys Ser Ile Asn Lys Asp Phe Ile Glu Gly
                325                 330                 335

Asn Lys Val Asn Ile Ser Leu Leu Ile Asp Met Met Lys Gly Tyr Glu
            340                 345                 350

Ala Asp Asp Ile Ile Arg Leu Tyr Tyr Asp Phe Ile Val Leu Lys Ser
        355                 360                 365

Gln Lys Asn Leu Gly Phe Ser Ile Lys Lys Leu Arg Glu Lys Met Leu
    370                 375                 380

Glu Glu Tyr Gly Phe Arg Phe Lys Asp Lys Gln Tyr Asp Ser Val Arg
385                 390                 395                 400

Ser Lys Met Tyr Lys Leu Met Asp Phe Leu Leu Phe Cys Asn Tyr Tyr
                405                 410                 415

Arg Asn Asp Val Ala Ala Gly Glu Ala Leu Val Arg Lys Leu Arg Phe
            420                 425                 430

Ser Met Thr Asp Asp Glu Lys Glu Gly Ile Tyr Ala Asp Glu Ala Ala
        435                 440                 445

Lys Leu Trp Gly Lys Phe Arg Asn Asp Phe Glu Asn Ile Ala Asp His
    450                 455                 460

Met Asn Gly Asp Val Ile Lys Glu Leu Gly Lys Ala Asp Met Asp Phe
465                 470                 475                 480
```

```
Asp Glu Lys Ile Leu Asp Ser Glu Lys Lys Asn Ala Ser Asp Leu Leu
                485                 490                 495

Tyr Phe Ser Lys Met Ile Tyr Met Leu Thr Tyr Phe Leu Asp Gly Lys
            500                 505                 510

Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Ser Lys Phe Asp Asn Ile
        515                 520                 525

Lys Glu Phe Leu Lys Ile Met Lys Ser Ser Ala Val Asp Val Glu Cys
    530                 535                 540

Glu Leu Thr Ala Gly Tyr Lys Leu Phe Asn Asp Ser Gln Arg Ile Thr
545                 550                 555                 560

Asn Glu Leu Phe Ile Val Lys Asn Ile Ala Ser Met Arg Lys Pro Ala
                565                 570                 575

Ala Ser Ala Lys Leu Thr Met Phe Arg Asp Ala Leu Thr Ile Leu Gly
            580                 585                 590

Ile Asp Asp Asn Ile Thr Asp Asp Arg Ile Ser Glu Ile Leu Lys Leu
        595                 600                 605

Lys Glu Lys Gly Lys Gly Ile His Gly Leu Arg Asn Phe Ile Thr Asn
    610                 615                 620

Asn Val Ile Glu Ser Ser Arg Phe Val Tyr Leu Ile Lys Tyr Ala Asn
625                 630                 635                 640

Ala Gln Lys Ile Arg Glu Val Ala Lys Asn Glu Lys Val Val Met Phe
                645                 650                 655

Val Leu Gly Gly Ile Pro Asp Thr Gln Ile Glu Arg Tyr Tyr Lys Ser
            660                 665                 670

Cys Val Glu Phe Pro Asp Met Asn Ser Ser Leu Glu Ala Lys Arg Ser
        675                 680                 685
```

```
Glu Leu Ala Arg Met Ile Lys Asn Ile Ser Phe Asp Asp Phe Lys Asn
    690                 695                 700


Val Lys Gln Gln Ala Lys Gly Arg Glu Asn Val Ala Lys Glu Arg Ala
705                 710                 715                 720


Lys Ala Val Ile Gly Leu Tyr Leu Thr Val Met Tyr Leu Leu Val Lys
                725                 730                 735


Asn Leu Val Asn Val Asn Ala Arg Tyr Val Ile Ala Ile His Cys Leu
            740                 745                 750


Glu Arg Asp Phe Gly Leu Tyr Lys Glu Ile Ile Pro Glu Leu Ala Ser
        755                 760                 765


Lys Asn Leu Lys Asn Asp Tyr Arg Ile Leu Ser Gln Thr Leu Cys Glu
    770                 775                 780


Leu Cys Asp Asp Arg Asn Glu Ser Ser Asn Leu Phe Leu Lys Lys Asn
785                 790                 795                 800


Lys Arg Leu Arg Lys Cys Val Glu Val Asp Ile Asn Asn Ala Asp Ser
                805                 810                 815


Ser Met Thr Arg Lys Tyr Arg Asn Cys Ile Ala His Leu Thr Val Val
            820                 825                 830


Arg Glu Leu Lys Glu Tyr Ile Gly Asp Ile Arg Thr Val Asp Ser Tyr
        835                 840                 845


Phe Ser Ile Tyr His Tyr Val Met Gln Arg Cys Ile Thr Lys Arg Gly
    850                 855                 860


Asp Asp Thr Lys Gln Glu Glu Lys Ile Lys Tyr Glu Asp Asp Leu Leu
865                 870                 875                 880


Lys Asn His Gly Tyr Thr Lys Asp Phe Val Lys Ala Leu Asn Ser Pro
                885                 890                 895
```

Phe Gly Tyr Asn Ile Pro Arg Phe Lys Asn Leu Ser Ile Glu Gln Leu
            900   905   910

Phe Asp Arg Asn Glu Tyr Leu Thr Glu Lys
        915   920

<210> 108
<211> 922
<212> PRT
<213> Artificial Sequence

<220>
<223> M17_gut_metagenome_P22E90k2120140920_c114

<400> 108

Met Ala Lys Lys Asn Lys Met Lys Pro Arg Glu Leu Arg Glu Ala Gln
1   5   10   15

Lys Lys Ala Arg Gln Leu Lys Ala Ala Glu Ile Asn Asn Asn Ala Ala
            20   25   30

Pro Ala Ile Ala Ala Met Pro Ala Ala Glu Val Ile Ala Pro Val Ala
        35   40   45

Glu Lys Lys Lys Ser Ser Val Lys Ala Ala Gly Met Lys Ser Ile Leu
    50   55   60

Val Ser Glu Asn Lys Met Tyr Ile Thr Ser Phe Gly Lys Gly Asn Ser
65   70   75   80

Ala Val Leu Glu Tyr Glu Val Asp Asn Asn Asp Tyr Asn Lys Thr Gln
                85   90   95

Leu Ser Ser Lys Asp Asn Ser Asn Ile Glu Leu Gly Asp Val Asn Glu
            100   105   110

Val Asn Ile Thr Phe Ser Ser Lys His Gly Phe Gly Ser Gly Val Glu
        115   120   125

Ile Asn Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser Ser Pro Val
    130   135   140

```
Arg Gly Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys Arg Phe Phe
145                 150                 155                 160

Gly Lys Thr Phe Asp Asp Asn Ile His Ile Gln Leu Ile Tyr Asn Ile
                165                 170                 175

Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn Ile Val Tyr
            180                 185                 190

Ala Leu Asn Asn Met Leu Gly Ile Lys Asp Ser Glu Ser Tyr Asp Asp
        195                 200                 205

Phe Met Gly Tyr Leu Ser Ala Arg Asn Thr Tyr Glu Val Phe Thr His
    210                 215                 220

Pro Asp Lys Ser Asn Leu Ser Asp Lys Val Lys Gly Asn Ile Lys Lys
225                 230                 235                 240

Ser Leu Ser Lys Phe Asn Asp Leu Leu Lys Thr Lys Arg Leu Gly Tyr
                245                 250                 255

Phe Gly Leu Glu Glu Pro Lys Thr Lys Asp Thr Arg Ala Ser Glu Ala
            260                 265                 270

Tyr Lys Lys Arg Val Tyr His Met Leu Ala Ile Val Gly Gln Ile Arg
        275                 280                 285

Gln Cys Val Phe His Asp Lys Ser Gly Ala Lys Arg Phe Asp Leu Tyr
    290                 295                 300

Ser Phe Ile Asn Asn Ile Asp Pro Glu Tyr Arg Asp Thr Leu Asp Tyr
305                 310                 315                 320

Leu Val Glu Glu Arg Leu Lys Ser Ile Asn Lys Asp Phe Ile Glu Gly
                325                 330                 335

Asn Lys Val Asn Ile Ser Leu Leu Ile Asp Met Met Lys Gly Tyr Glu
            340                 345                 350
```

```
Ala Asp Asp Ile Ile Arg Leu Tyr Tyr Asp Phe Ile Val Leu Lys Ser
        355                 360                 365

Gln Lys Asn Leu Gly Phe Ser Ile Lys Lys Leu Arg Glu Lys Met Leu
    370                 375                 380

Glu Glu Tyr Gly Phe Arg Phe Lys Asp Lys Gln Tyr Asp Ser Val Arg
385                 390                 395                 400

Ser Lys Met Tyr Lys Leu Met Asp Phe Leu Leu Phe Cys Asn Tyr Tyr
                405                 410                 415

Arg Asn Asp Val Ala Ala Gly Glu Ala Leu Val Arg Lys Leu Arg Phe
            420                 425                 430

Ser Met Thr Asp Asp Glu Lys Glu Gly Ile Tyr Ala Asp Glu Ala Ala
        435                 440                 445

Lys Leu Trp Gly Lys Phe Arg Asn Asp Phe Glu Asn Ile Ala Asp His
    450                 455                 460

Met Asn Gly Asp Val Ile Lys Glu Leu Gly Lys Ala Asp Met Asp Phe
465                 470                 475                 480

Asp Glu Lys Ile Leu Asp Ser Glu Lys Lys Asn Ala Ser Asp Leu Leu
                485                 490                 495

Tyr Phe Ser Lys Met Ile Tyr Met Leu Thr Tyr Phe Leu Asp Gly Lys
            500                 505                 510

Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Ser Lys Phe Asp Asn Ile
        515                 520                 525

Lys Glu Phe Leu Lys Ile Met Lys Ser Ser Ala Val Asp Val Glu Cys
    530                 535                 540

Glu Leu Thr Ala Gly Tyr Lys Leu Phe Asn Asp Ser Gln Arg Ile Thr
545                 550                 555                 560
```

```
Asn Glu Leu Phe Ile Val Lys Asn Ile Ala Ser Met Arg Lys Pro Ala
                565                 570                 575

Ala Ser Ala Lys Leu Thr Met Phe Arg Asp Ala Leu Thr Ile Leu Gly
            580                 585                 590

Ile Asp Asp Asn Ile Thr Asp Asp Arg Ile Ser Glu Ile Leu Lys Leu
        595                 600                 605

Lys Glu Lys Gly Lys Gly Ile His Gly Leu Arg Asn Phe Ile Thr Asn
    610                 615                 620

Asn Val Ile Glu Ser Ser Arg Phe Val Tyr Leu Ile Lys Tyr Ala Asn
625                 630                 635                 640

Ala Gln Lys Ile Arg Glu Val Ala Lys Asn Glu Lys Val Val Met Phe
                645                 650                 655

Val Leu Gly Gly Ile Pro Asp Thr Gln Ile Glu Arg Tyr Tyr Lys Ser
            660                 665                 670

Cys Val Glu Phe Pro Asp Met Asn Ser Ser Leu Glu Ala Lys Arg Ser
        675                 680                 685

Glu Leu Ala Arg Met Ile Lys Asn Ile Ser Phe Asp Asp Phe Lys Asn
    690                 695                 700

Val Lys Gln Gln Ala Lys Gly Arg Glu Asn Val Ala Lys Glu Arg Ala
705                 710                 715                 720

Lys Ala Val Ile Gly Leu Tyr Leu Thr Val Met Tyr Leu Leu Val Lys
                725                 730                 735

Asn Leu Val Asn Val Asn Ala Arg Tyr Val Ile Ala Ile His Cys Leu
            740                 745                 750

Glu Arg Asp Phe Gly Leu Tyr Lys Glu Ile Ile Pro Glu Leu Ala Ser
        755                 760                 765
```

```
Lys Asn Leu Lys Asn Asp Tyr Arg Ile Leu Ser Gln Thr Leu Cys Glu
    770                 775                 780

Leu Cys Asp Asp Arg Asn Glu Ser Ser Asn Leu Phe Leu Lys Lys Asn
785                 790                 795                 800

Lys Arg Leu Arg Lys Cys Val Glu Val Asp Ile Asn Asn Ala Asp Ser
                805                 810                 815

Ser Met Thr Arg Lys Tyr Arg Asn Cys Ile Ala His Leu Thr Val Val
            820                 825                 830

Arg Glu Leu Lys Glu Tyr Ile Gly Asp Ile Arg Thr Val Asp Ser Tyr
        835                 840                 845

Phe Ser Ile Tyr His Tyr Val Met Gln Arg Cys Ile Thr Lys Arg Gly
    850                 855                 860

Asp Asp Thr Lys Gln Glu Glu Lys Ile Lys Tyr Glu Asp Asp Leu Leu
865                 870                 875                 880

Lys Asn His Gly Tyr Thr Lys Asp Phe Val Lys Ala Leu Asn Ser Pro
                885                 890                 895

Phe Gly Tyr Asn Ile Pro Arg Phe Lys Asn Leu Ser Ile Glu Gln Leu
            900                 905                 910

Phe Asp Arg Asn Glu Tyr Leu Thr Glu Lys
        915                 920


<210>  109
<211>  922
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Ruminococcus_sp_CAG57

<400>  109
```

```
Met Ala Lys Lys Asn Lys Met Lys Pro Arg Glu Leu Arg Glu Ala Gln
1               5                  10                  15

Lys Lys Ala Arg Gln Leu Lys Ala Ala Glu Ile Asn Asn Asn Ala Ala
            20                  25                  30

Pro Ala Ile Ala Ala Met Pro Ala Ala Glu Val Ile Ala Pro Val Ala
        35                  40                  45

Glu Lys Lys Lys Ser Ser Val Lys Ala Ala Gly Met Lys Ser Ile Leu
    50                  55                  60

Val Ser Glu Asn Lys Met Tyr Ile Thr Ser Phe Gly Lys Gly Asn Ser
65                  70                  75                  80

Ala Val Leu Glu Tyr Glu Val Asp Asn Asn Asp Tyr Asn Lys Thr Gln
                85                  90                  95

Leu Ser Ser Lys Asp Asn Ser Asn Ile Glu Leu Gly Asp Val Asn Glu
            100                 105                 110

Val Asn Ile Thr Phe Ser Ser Lys His Gly Phe Gly Ser Gly Val Glu
        115                 120                 125

Ile Asn Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser Ser Pro Val
    130                 135                 140

Arg Gly Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys Arg Phe Phe
145                 150                 155                 160

Gly Lys Thr Phe Asp Asp Asn Ile His Ile Gln Leu Ile Tyr Asn Ile
                165                 170                 175

Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn Ile Val Tyr
            180                 185                 190

Ala Leu Asn Asn Met Leu Gly Ile Lys Asp Ser Glu Ser Tyr Asp Asp
        195                 200                 205
```

```
Phe Met Gly Tyr Leu Ser Ala Arg Asn Thr Tyr Glu Val Phe Thr His
    210                 215                 220

Pro Asp Lys Ser Asn Leu Ser Asp Lys Val Lys Gly Asn Ile Lys Lys
225                 230                 235                 240

Ser Leu Ser Lys Phe Asn Asp Leu Leu Lys Thr Lys Arg Leu Gly Tyr
                245                 250                 255

Phe Gly Leu Glu Glu Pro Lys Thr Lys Asp Thr Arg Ala Ser Glu Ala
            260                 265                 270

Tyr Lys Lys Arg Val Tyr His Met Leu Ala Ile Val Gly Gln Ile Arg
        275                 280                 285

Gln Cys Val Phe His Asp Lys Ser Gly Ala Lys Arg Phe Asp Leu Tyr
    290                 295                 300

Ser Phe Ile Asn Asn Ile Asp Pro Glu Tyr Arg Asp Thr Leu Asp Tyr
305                 310                 315                 320

Leu Val Glu Glu Arg Leu Lys Ser Ile Asn Lys Asp Phe Ile Glu Gly
                325                 330                 335

Asn Lys Val Asn Ile Ser Leu Leu Ile Asp Met Met Lys Gly Tyr Glu
            340                 345                 350

Ala Asp Asp Ile Ile Arg Leu Tyr Tyr Asp Phe Ile Val Leu Lys Ser
        355                 360                 365

Gln Lys Asn Leu Gly Phe Ser Ile Lys Lys Leu Arg Glu Lys Met Leu
    370                 375                 380

Glu Glu Tyr Gly Phe Arg Phe Lys Asp Lys Gln Tyr Asp Ser Val Arg
385                 390                 395                 400

Ser Lys Met Tyr Lys Leu Met Asp Phe Leu Leu Phe Cys Asn Tyr Tyr
                405                 410                 415
```

```
Arg Asn Asp Val Ala Ala Gly Glu Ala Leu Val Arg Lys Leu Arg Phe
            420                 425                 430

Ser Met Thr Asp Asp Glu Lys Glu Gly Ile Tyr Ala Asp Glu Ala Ala
        435                 440                 445

Lys Leu Trp Gly Lys Phe Arg Asn Asp Phe Glu Asn Ile Ala Asp His
    450                 455                 460

Met Asn Gly Asp Val Ile Lys Glu Leu Gly Lys Ala Asp Met Asp Phe
465                 470                 475                 480

Asp Glu Lys Ile Leu Asp Ser Glu Lys Lys Asn Ala Ser Asp Leu Leu
                485                 490                 495

Tyr Phe Ser Lys Met Ile Tyr Met Leu Thr Tyr Phe Leu Asp Gly Lys
            500                 505                 510

Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Ser Lys Phe Asp Asn Ile
        515                 520                 525

Lys Glu Phe Leu Lys Ile Met Lys Ser Ser Ala Val Asp Val Glu Cys
    530                 535                 540

Glu Leu Thr Ala Gly Tyr Lys Leu Phe Asn Asp Ser Gln Arg Ile Thr
545                 550                 555                 560

Asn Glu Leu Phe Ile Val Lys Asn Ile Ala Ser Met Arg Lys Pro Ala
                565                 570                 575

Ala Ser Ala Lys Leu Thr Met Phe Arg Asp Ala Leu Thr Ile Leu Gly
            580                 585                 590

Ile Asp Asp Asn Ile Thr Asp Asp Arg Ile Ser Glu Ile Leu Lys Leu
        595                 600                 605

Lys Glu Lys Gly Lys Gly Ile His Gly Leu Arg Asn Phe Ile Thr Asn
    610                 615                 620
```

```
Asn Val Ile Glu Ser Ser Arg Phe Val Tyr Leu Ile Lys Tyr Ala Asn
625                 630                 635                 640


Ala Gln Lys Ile Arg Glu Val Ala Lys Asp Glu Lys Val Val Met Phe
                645                 650                 655


Val Leu Gly Gly Ile Pro Asp Thr Gln Ile Glu Arg Tyr Tyr Lys Ser
            660                 665                 670


Cys Val Glu Phe Pro Asp Met Asn Ser Ser Leu Glu Ala Lys Arg Ser
        675                 680                 685


Glu Leu Ala Arg Met Ile Lys Asn Ile Ser Phe Asp Asp Phe Lys Asn
    690                 695                 700


Val Lys Gln Gln Ala Lys Gly Arg Glu Asn Val Ala Lys Glu Arg Ala
705                 710                 715                 720


Lys Ala Val Ile Gly Leu Tyr Leu Thr Val Met Tyr Leu Leu Val Lys
                725                 730                 735


Asn Leu Val Asn Val Asn Ala Arg Tyr Val Ile Ala Ile His Cys Leu
            740                 745                 750


Glu Arg Asp Phe Gly Leu Tyr Lys Glu Ile Ile Pro Glu Leu Ala Ser
        755                 760                 765


Lys Asn Leu Lys Asn Asp Tyr Arg Ile Leu Ser Gln Thr Leu Cys Glu
    770                 775                 780


Leu Cys Asp Asp Arg Asn Glu Ser Ser Asn Leu Phe Leu Lys Lys Asn
785                 790                 795                 800


Lys Arg Leu Arg Lys Cys Val Glu Val Asp Ile Asn Asn Ala Asp Ser
                805                 810                 815


Ser Met Thr Arg Lys Tyr Arg Asn Cys Ile Ala His Leu Thr Val Val
            820                 825                 830
```

Arg Glu Leu Lys Glu Tyr Ile Gly Asp Ile Arg Thr Val Asp Ser Tyr
835 840 845

Phe Ser Ile Tyr His Tyr Val Met Gln Arg Cys Ile Thr Lys Arg Gly
850 855 860

Asp Asp Thr Lys Gln Glu Glu Lys Ile Lys Tyr Glu Asp Asp Leu Leu
865 870 875 880

Lys Asn His Gly Tyr Thr Lys Asp Phe Val Lys Ala Leu Asn Ser Pro
885 890 895

Phe Gly Tyr Asn Ile Pro Arg Phe Lys Asn Leu Ser Ile Glu Gln Leu
900 905 910

Phe Asp Arg Asn Glu Tyr Leu Thr Glu Lys
915 920

<210> 110
<211> 531
<212> PRT
<213> Artificial Sequence

<220>
<223> gut_metagenome_P11E90k2120140920_c43000123

<400> 110

Phe Arg Phe Lys Asp Lys Gln Tyr Asp Ser Val Arg Ser Lys Met Tyr
1 5 10 15

Lys Leu Met Asp Phe Leu Leu Phe Cys Asn Tyr Tyr Arg Asn Asp Val
20 25 30

Val Ala Gly Glu Ala Leu Val Arg Lys Leu Arg Phe Ser Met Thr Asp
35 40 45

Asp Glu Lys Glu Gly Ile Tyr Ala Asp Glu Ala Ala Lys Leu Trp Gly
50 55 60

Lys Phe Arg Asn Asp Phe Glu Asn Ile Ala Asp His Met Asn Gly Asp
65 70 75 80

```
Val Ile Lys Glu Leu Gly Lys Ala Asp Met Asp Phe Asp Glu Lys Ile
                85                  90                  95

Leu Asp Ser Glu Lys Lys Asn Ala Ser Asp Leu Leu Tyr Phe Ser Lys
            100                 105                 110

Met Ile Tyr Met Leu Thr Tyr Phe Leu Asp Gly Lys Glu Ile Asn Asp
        115                 120                 125

Leu Leu Thr Thr Leu Ile Ser Lys Phe Asp Asn Ile Lys Glu Phe Leu
    130                 135                 140

Lys Ile Met Lys Ser Ser Ala Val Asp Val Glu Cys Glu Leu Thr Ala
145                 150                 155                 160

Gly Tyr Lys Leu Phe Asn Asp Ser Gln Arg Ile Thr Asn Glu Leu Phe
                165                 170                 175

Ile Val Lys Asn Ile Ala Ser Met Arg Lys Pro Ala Ala Ser Ala Arg
            180                 185                 190

Leu Thr Met Phe Arg Asp Ala Leu Thr Ile Leu Gly Ile Asp Asp Lys
        195                 200                 205

Ile Thr Asp Asp Arg Ile Ser Glu Ile Leu Lys Leu Lys Glu Lys Gly
    210                 215                 220

Lys Gly Ile His Gly Leu Arg Asn Phe Ile Thr Asn Asn Val Ile Glu
225                 230                 235                 240

Ser Ser Arg Phe Val Tyr Leu Ile Lys Tyr Ala Asn Ala Gln Lys Ile
                245                 250                 255

Arg Glu Val Ala Lys Asn Glu Lys Val Val Met Phe Val Leu Gly Gly
            260                 265                 270

Ile Pro Asp Thr Gln Ile Glu Arg Tyr Tyr Lys Ser Cys Val Glu Phe
        275                 280                 285
```

```
Pro Asp Met Asn Ser Ser Leu Glu Ala Lys Arg Ser Glu Leu Ala Arg
    290                 295                 300

Met Ile Lys Asn Ile Ser Phe Asp Asp Phe Lys Asn Val Lys Gln Gln
305                 310                 315                 320

Ala Lys Gly Arg Glu Asn Val Ala Lys Glu Arg Ala Lys Ala Val Ile
                325                 330                 335

Gly Leu Tyr Leu Thr Val Met Tyr Leu Leu Val Lys Asn Leu Val Asn
            340                 345                 350

Val Asn Ala Arg Tyr Val Ile Ala Ile His Cys Leu Glu Arg Asp Phe
        355                 360                 365

Gly Leu Tyr Lys Glu Ile Ile Pro Glu Leu Ala Ser Lys Asn Leu Lys
    370                 375                 380

Asn Asp Tyr Arg Ile Leu Ser Gln Thr Leu Cys Glu Leu Cys Asp Lys
385                 390                 395                 400

Ser Pro Asn Leu Phe Leu Lys Lys Asn Glu Arg Leu Arg Lys Cys Val
                405                 410                 415

Glu Val Asp Ile Asn Asn Ala Asp Ser Ser Met Thr Arg Lys Tyr Arg
            420                 425                 430

Asn Cys Ile Ala His Leu Thr Val Val Arg Glu Leu Lys Glu Tyr Ile
        435                 440                 445

Gly Asp Ile Arg Thr Val Asp Ser Tyr Phe Ser Ile Tyr His Tyr Val
    450                 455                 460

Met Gln Arg Cys Ile Thr Lys Arg Glu Asn Asp Thr Lys Gln Glu Asp
465                 470                 475                 480

Lys Ile Lys Tyr Glu Asp Asp Leu Leu Lys Asn His Gly Tyr Thr Lys
                485                 490                 495
```

```
Asp Phe Val Lys Ala Leu Asn Ser Pro Phe Gly Tyr Asn Ile Pro Arg
            500                 505                 510


Phe Lys Asn Leu Ser Ile Glu Gln Leu Phe Asp Arg Asn Glu Tyr Leu
        515                 520                 525


Thr Glu Lys
    530


<210>  111
<211>  925
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  M6_gut_metagenome_P13E90k2120140920_c7000009

<400>  111

Met Ala Lys Lys Asn Lys Met Lys Pro Arg Glu Leu Arg Glu Ala Gln
1               5                   10                  15


Lys Lys Ala Arg Gln Leu Lys Ala Ala Glu Ile Lys Asn Asn Ala Val
            20                  25                  30


Pro Ala Ile Ala Ala Met Pro Ala Ala Glu Ala Ala Ala Pro Ala Val
        35                  40                  45


Glu Lys Lys Lys Ser Ser Val Lys Ala Ala Gly Met Lys Ser Ile Leu
    50                  55                  60


Val Ser Glu Asn Lys Met Tyr Ile Thr Ser Phe Gly Lys Gly Asn Ser
65                  70                  75                  80


Ala Val Leu Glu Tyr Glu Val Asp Lys Val Asp Asn Asn Asn Tyr Asn
                85                  90                  95


Lys Thr Gln Leu Ser Ser Lys Asp Asn Ser Asn Ile Glu Leu Gly Asp
            100                 105                 110
```

```
Val Asn Glu Val Asn Ile Thr Phe Ser Ser Lys Arg Gly Asn Glu Ser
        115                 120                 125

Gly Val Glu Ile Asn Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser
    130                 135                 140

Ser Pro Val Arg Trp Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys
145                 150                 155                 160

Arg Phe Phe Gly Lys Thr Phe Asp Asp Asn Ile His Ile Gln Leu Ile
                165                 170                 175

Tyr Asn Ile Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn
            180                 185                 190

Ile Val Tyr Ala Leu Asn Asn Met Leu Gly Ile Lys Gly Ser Glu Ser
        195                 200                 205

Tyr Asp Asp Phe Met Gly Tyr Leu Ser Ala Arg Asn Thr Tyr Glu Val
    210                 215                 220

Phe Thr His Pro Asp Lys Ser Asn Leu Ser Asp Lys Val Lys Gly Asn
225                 230                 235                 240

Ile Lys Lys Ser Leu Ser Lys Phe Asn Asp Leu Leu Lys Thr Lys Arg
                245                 250                 255

Leu Gly Tyr Phe Gly Leu Glu Glu Pro Lys Thr Lys Asp Thr Arg Val
            260                 265                 270

Ser Gln Ala Tyr Lys Lys Arg Val Tyr His Met Leu Ala Ile Val Gly
        275                 280                 285

Gln Ile Arg Gln Cys Val Phe His Asp Lys Ser Gly Ala Lys Arg Phe
    290                 295                 300

Asp Leu Tyr Ser Phe Ile Asn Asn Ile Asp Pro Glu Tyr Arg Asp Thr
305                 310                 315                 320
```

```
Leu Asp Tyr Leu Val Glu Glu Arg Leu Lys Ser Ile Asn Lys Asp Phe
                325                 330                 335

Ile Glu Gly Asn Lys Val Asn Ile Ser Leu Leu Ile Asp Met Met Lys
            340                 345                 350

Gly Tyr Glu Ala Asp Asp Ile Ile Arg Leu Tyr Tyr Asp Phe Ile Val
        355                 360                 365

Leu Lys Ser Gln Lys Asn Leu Gly Phe Ser Ile Lys Lys Leu Arg Glu
    370                 375                 380

Lys Met Leu Glu Glu Tyr Gly Tyr Arg Phe Lys Asp Lys Gln Tyr Asp
385                 390                 395                 400

Ser Val Arg Ser Lys Met Tyr Lys Leu Met Asp Phe Leu Leu Phe Cys
                405                 410                 415

Asn Tyr Tyr Arg Asn Asp Val Val Ala Gly Glu Ala Leu Val Arg Lys
            420                 425                 430

Leu Arg Phe Ser Met Thr Asp Asp Glu Lys Glu Gly Ile Tyr Ala Asp
        435                 440                 445

Glu Ala Ser Lys Leu Trp Gly Lys Phe Arg Asn Asp Phe Glu Asn Ile
    450                 455                 460

Ala Asp His Met Asn Gly Asp Val Ile Lys Glu Leu Gly Lys Ala Asp
465                 470                 475                 480

Met Asp Phe Asp Glu Lys Ile Leu Asp Ser Glu Lys Lys Asn Ala Ser
                485                 490                 495

Asp Leu Leu Tyr Phe Ser Lys Met Ile Tyr Met Leu Thr Tyr Phe Leu
            500                 505                 510

Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Ser Lys Phe
        515                 520                 525
```

```
Asp Asn Ile Lys Glu Phe Leu Lys Ile Met Lys Ser Ser Ala Val Asp
    530                 535                 540

Val Glu Cys Glu Leu Thr Ala Gly Tyr Lys Leu Phe Asn Asp Ser Gln
545                 550                 555                 560

Arg Ile Thr Asn Glu Leu Phe Ile Val Lys Asn Ile Ala Ser Met Arg
                565                 570                 575

Lys Pro Ala Ala Ser Ala Lys Leu Thr Met Phe Arg Asp Ala Leu Thr
            580                 585                 590

Ile Leu Gly Ile Asp Asp Lys Ile Thr Asp Asp Arg Ile Ser Glu Ile
        595                 600                 605

Leu Lys Leu Lys Glu Lys Gly Lys Gly Ile His Gly Leu Arg Asn Phe
    610                 615                 620

Ile Thr Asn Asn Val Ile Glu Ser Ser Arg Phe Val Tyr Leu Ile Lys
625                 630                 635                 640

Tyr Ala Asn Ala Gln Lys Ile Arg Glu Val Ala Lys Asn Glu Lys Val
                645                 650                 655

Val Met Phe Val Leu Gly Gly Ile Pro Asp Thr Gln Ile Glu Arg Tyr
            660                 665                 670

Tyr Lys Ser Cys Val Glu Phe Pro Asp Met Asn Ser Ser Leu Glu Ala
        675                 680                 685

Lys Arg Ser Glu Leu Ala Arg Met Ile Lys Asn Ile Arg Phe Asp Asp
    690                 695                 700

Phe Lys Asn Val Lys Gln Gln Ala Lys Gly Arg Glu Asn Val Ala Lys
705                 710                 715                 720

Glu Arg Ala Lys Ala Val Ile Gly Leu Tyr Leu Thr Val Met Tyr Leu
                725                 730                 735
```

```
Leu Val Lys Asn Leu Val Asn Val Asn Ala Arg Tyr Val Ile Ala Ile
            740                 745                 750


His Cys Leu Glu Arg Asp Phe Gly Leu Tyr Lys Glu Ile Ile Pro Glu
        755                 760                 765


Leu Ala Ser Lys Asn Leu Lys Asn Asp Tyr Arg Ile Leu Ser Gln Thr
    770                 775                 780


Leu Cys Glu Leu Cys Asp Asp Arg Asp Glu Ser Pro Asn Leu Phe Leu
785                 790                 795                 800


Lys Lys Asn Lys Arg Leu Arg Lys Cys Val Glu Val Asp Ile Asn Asn
                805                 810                 815


Ala Asp Ser Ser Met Thr Arg Lys Tyr Arg Asn Cys Ile Ala His Leu
            820                 825                 830


Thr Val Val Arg Glu Leu Lys Glu Tyr Ile Gly Asp Ile Arg Thr Val
        835                 840                 845


Asp Ser Tyr Phe Ser Ile Tyr His Tyr Val Met Gln Arg Cys Ile Thr
    850                 855                 860


Lys Arg Glu Asp Asp Thr Lys Gln Glu Glu Lys Ile Lys Tyr Glu Asp
865                 870                 875                 880


Asp Leu Leu Lys Asn His Gly Tyr Thr Lys Asp Phe Val Lys Ala Leu
                885                 890                 895


Asn Ser Pro Phe Gly Tyr Asn Ile Pro Arg Phe Lys Asn Leu Ser Ile
            900                 905                 910


Glu Gln Leu Phe Asp Arg Asn Glu Tyr Leu Thr Glu Lys
        915                 920                 925


<210>  112
<211>  921
<212>  PRT
<213>  Artificial Sequence
```

<220>
<223> M19_gut_metagenome_P17E90k2120140920

<400> 112

```
Met Ala Lys Lys Asn Lys Met Lys Pro Arg Glu Arg Arg Glu Ala Gln
1               5                  10                  15


Lys Lys Ala Arg Gln Leu Lys Ala Ala Glu Ile Asn Asn Asn Ala Val
            20                  25                  30


Pro Ala Ile Ala Ala Met His Ala Ala Glu Val Ile Ala Pro Ala Ala
        35                  40                  45


Glu Lys Lys Lys Ser Ser Val Lys Ala Ala Gly Met Lys Ser Ile Leu
    50                  55                  60


Val Ser Glu Asn Lys Met Tyr Ile Thr Ser Phe Gly Lys Gly Asn Ser
65                  70                  75                  80


Ala Val Leu Glu Tyr Glu Val Asp Asn Asn Asp Tyr Asn Gln Thr Gln
                85                  90                  95


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


Val Asn Ile Thr Phe Ser Ser Lys His Gly Phe Glu Ser Gly Val Glu
        115                 120                 125


Ile Asn Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser Ser Pro Val
    130                 135                 140


Arg Gly Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys Arg Phe Phe
145                 150                 155                 160


Gly Lys Thr Phe Asp Asp Asn Ile His Ile Gln Leu Ile Tyr Asn Ile
                165                 170                 175


Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn Ile Val Tyr
            180                 185                 190
```

```
Ala Leu Asn Asn Met Leu Gly Glu Gly Asp Glu Ser Asn Tyr Asp Phe
        195                 200                 205

Met Gly Tyr Leu Ser Thr Phe Asn Thr Tyr Lys Val Phe Thr Asn Pro
    210                 215                 220

Asn Gly Ser Thr Leu Ser Asp Asp Lys Lys Glu Asn Ile Arg Lys Ser
225                 230                 235                 240

Leu Ser Lys Phe Asn Ala Leu Leu Lys Thr Lys Arg Leu Gly Tyr Phe
                245                 250                 255

Gly Leu Glu Glu Pro Lys Thr Lys Asp Thr Arg Ala Ser Glu Ala Tyr
            260                 265                 270

Lys Lys Arg Val Tyr His Met Leu Ala Ile Val Gly Gln Ile Arg Gln
        275                 280                 285

Cys Val Phe His Asp Lys Ser Gly Ala Lys Arg Phe Asp Leu Tyr Ser
    290                 295                 300

Phe Ile Asn Asn Ile Tyr Pro Glu Tyr Arg Asp Thr Leu Asp Tyr Leu
305                 310                 315                 320

Val Glu Glu Arg Leu Lys Ser Ile Asn Lys Asp Phe Ile Gln Gly Asn
                325                 330                 335

Lys Val Asn Ile Ser Leu Leu Ile Asp Met Met Lys Gly Tyr Glu Ala
            340                 345                 350

Asp Asp Ile Ile Arg Leu Tyr Tyr Asp Phe Ile Val Leu Lys Ser Gln
        355                 360                 365

Lys Asn Leu Gly Phe Ser Ile Lys Lys Leu Arg Glu Lys Met Leu Asp
    370                 375                 380

Glu Tyr Gly Phe Arg Phe Lys Asp Lys Gln Tyr Asp Ser Val Arg Ser
385                 390                 395                 400
```

```
Lys Met Tyr Lys Leu Met Asp Phe Leu Leu Phe Cys Asn Tyr Tyr Arg
                405                 410                 415

Asn Asp Val Val Ala Gly Glu Ala Leu Val Arg Lys Leu Arg Phe Ser
            420                 425                 430

Met Thr Asp Asp Glu Lys Glu Gly Ile Tyr Ala Asp Glu Ala Ala Lys
        435                 440                 445

Leu Trp Gly Lys Phe Arg Asn Asp Phe Glu Asn Ile Ala Asp His Met
    450                 455                 460

Asn Gly Asp Val Ile Lys Glu Leu Gly Lys Ala Asp Met Asp Phe Asp
465                 470                 475                 480

Glu Lys Ile Leu Asp Ser Glu Lys Lys Asn Ala Ser Asp Leu Leu Tyr
                485                 490                 495

Phe Ser Lys Met Ile Tyr Met Leu Thr Tyr Phe Leu Asp Gly Lys Glu
            500                 505                 510

Ile Asn Asp Leu Leu Thr Thr Leu Ile Ser Lys Phe Asp Asn Ile Lys
        515                 520                 525

Glu Phe Leu Lys Ile Met Lys Ser Ser Ala Val Asp Val Glu Cys Glu
    530                 535                 540

Leu Thr Ala Gly Tyr Lys Leu Phe Asn Asp Ser Gln Arg Ile Thr Asn
545                 550                 555                 560

Glu Leu Phe Ile Val Lys Asn Ile Ala Ser Met Arg Lys Pro Ala Ala
                565                 570                 575

Ser Ala Lys Leu Thr Met Phe Arg Asp Ala Leu Thr Ile Leu Gly Ile
            580                 585                 590

Asp Asp Lys Ile Thr Asp Asp Arg Ile Ser Glu Ile Leu Lys Leu Lys
        595                 600                 605
```

```
Glu Lys Gly Lys Gly Ile His Gly Leu Arg Asn Phe Ile Thr Asn Asn
    610                 615                 620

Val Ile Glu Ser Ser Arg Phe Val Tyr Leu Ile Lys Tyr Ala Asn Ala
625                 630                 635                 640

Gln Lys Ile Arg Glu Val Ala Lys Asn Glu Lys Val Val Met Phe Val
                645                 650                 655

Leu Gly Gly Ile Pro Asp Thr Gln Ile Glu Arg Tyr Tyr Lys Ser Cys
            660                 665                 670

Val Glu Phe Pro Asp Met Asn Ser Ser Leu Glu Ala Lys Arg Ser Glu
        675                 680                 685

Leu Ala Arg Met Ile Lys Asn Ile Ser Phe Asp Asp Phe Lys Asn Val
    690                 695                 700

Lys Gln Gln Ala Lys Gly Arg Glu Asn Val Ala Lys Glu Arg Ala Lys
705                 710                 715                 720

Ala Val Ile Gly Leu Tyr Leu Thr Val Met Tyr Leu Leu Val Lys Asn
                725                 730                 735

Leu Val Asn Val Asn Ala Arg Tyr Val Ile Ala Ile His Cys Leu Glu
            740                 745                 750

Arg Asp Phe Gly Leu Tyr Lys Glu Ile Ile Pro Glu Leu Ala Ser Lys
        755                 760                 765

Asn Leu Lys Asn Asp Tyr Arg Ile Leu Ser Gln Thr Leu Cys Glu Leu
    770                 775                 780

Cys Asp Asp Arg Asp Glu Ser Pro Asn Leu Phe Leu Lys Lys Asn Lys
785                 790                 795                 800

Arg Leu Arg Lys Cys Val Glu Val Asp Ile Asn Asn Ala Asp Ser Ser
                805                 810                 815
```

```
Met Thr Arg Lys Tyr Arg Asn Cys Ile Ala His Leu Thr Val Val Arg
            820                 825                 830

Glu Leu Lys Lys Tyr Ile Gly Asp Ile Arg Thr Val Asp Ser Tyr Phe
        835                 840                 845

Ser Ile Tyr His Tyr Val Met Gln Arg Cys Ile Thr Lys Arg Glu Asp
    850                 855                 860

Asp Thr Lys Gln Glu Glu Lys Ile Lys Tyr Glu Asp Asp Leu Leu Lys
865                 870                 875                 880

Asn His Gly Tyr Thr Lys Asp Phe Val Lys Ala Leu Asn Ser Pro Phe
                885                 890                 895

Gly Tyr Asn Ile Pro Arg Phe Lys Asn Leu Ser Ile Glu Gln Leu Phe
            900                 905                 910

Asp Arg Asn Glu Tyr Leu Thr Glu Lys
        915                 920
```

<210> 113
<211> 234
<212> PRT
<213> Artificial Sequence

<220>
<223> gut_metagenome_P17E0k2120140920,_c87000043

<400> 113

```
Met Tyr Phe Ser Lys Met Ile Tyr Met Leu Thr Tyr Phe Leu Asp Gly
1               5                   10                  15

Lys Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Ser Lys Phe Asp Asn
            20                  25                  30

Ile Lys Glu Phe Leu Lys Ile Met Lys Ser Ser Ala Val Asp Val Glu
        35                  40                  45
```

```
Cys Glu Leu Thr Ala Gly Tyr Lys Leu Phe Asn Asp Ser Gln Arg Ile
    50                  55                  60

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

Ala Ala Ser Ala Lys Leu Thr Met Phe Arg Asp Ala Leu Thr Ile Leu
                85                  90                  95

Gly Ile Asp Asp Lys Ile Thr Asp Asp Arg Ile Ser Glu Ile Leu Lys
            100                 105                 110

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

Asn Asn Val Ile Glu Ser Ser Arg Phe Val Tyr Leu Ile Lys Tyr Ala
    130                 135                 140

Asn Ala Gln Lys Ile Arg Glu Val Ala Lys Asn Glu Lys Val Val Met
145                 150                 155                 160

Phe Val Leu Gly Gly Ile Pro Asp Thr Gln Ile Glu Arg Tyr Tyr Lys
                165                 170                 175

Ser Cys Val Glu Phe Pro Asp Met Asn Ser Ser Leu Glu Ala Lys Arg
            180                 185                 190

Ser Glu Leu Ala Arg Met Ile Lys Asn Ile Ser Phe Asp Asp Phe Lys
        195                 200                 205

Asn Val Lys Gln Gln Ala Lys Gly Arg Glu Asn Val Ala Lys Glu Arg
    210                 215                 220

Ala Lys Ala Val Ile Gly Leu Tyr Leu Thr
225                 230
```

<210> 114
<211> 2859
<212> DNA
<213> Artificial Sequence

<220>
<223> Human codon-optimized Eubacterium siraeum nCas1

<400> 114
ggcaagaaga tccatgccag ggacctcagg gagcaaagaa agaccgatag gacagaaaag 60

ttcgccgatc agaacaagaa gagggaggct gagagagctg tgcccaaaaa agacgccgcc 120

gtctccgtga agagcgtgtc cagcgtcagc agcaagaagg acaacgtcac caagagcatg 180

gccaaggccg ccggagtgaa gtccgtcttt gctgtgggaa acaccgtgta catgacctcc 240

ttcggcaggg gcaacgatgc cgtgctggag cagaaaatcg tggacacctc ccatgagccc 300

ctgaatatcg acgatcccgc ctatcaactc aacgtggtga ccatgaatgg atacagcgtc 360

acaggccata gaggagagac agtgtccgcc gtgacagata atcccctgag gagattcaac 420

ggcaggaaga aggacgaacc cgagcagagc gtccctacag acatgctctg cctgaagccc 480

accctcgaga aaaaattttt tggaaaggag tttgatgaca acattcatat tcagctgatc 540

tataatatcc tggatatcga gaaaatcctg gccgtgtact ccaccaacgc catctacgct 600

ctcaacaaca tgtccgccga cgaaaacatc gagaactccg acttcttcat gaaaagaaca 660

acagacgaga cctttgacga cttcgagaag aagaaggaga gcaccaacag cagagagaaa 720

gctgactttg acgcctttga aaagttcatc ggaaactata ggctggccta tttcgctgac 780

gctttctacg tgaataaaaa gaaccccaag ggaaaggcca agaatgtcct gagggaggac 840

aaggagctct acagcgtgct gacactgatc ggaaagctga gacactggtg tgtccacagc 900

gaggaaggaa gggctgaatt ctggctgtac aagctggacg agctcaaaga cgatttcaag 960

aacgtcctgg atgtggtgta taacagaccc gtggaggaga tcaacaacag gttcatcgaa 1020

aataacaaag tcaacatcca gatcctgggc tccgtgtata agaacaccga catcgccgag 1080

ctggtgaggt cctactatga gttcctgatc accaaaaaat acaagaatat gggcttttcc 1140

atcaagaagc tcagggagag catgctggag ggcaagggct acgctgataa ggagtacgat 1200

agcgtgagaa ataagctgta ccagatgacc gatttcatcc tctacaccgg ctacatcaac 1260

gaagatagcg atagggctga cgacctggtg aacacactga ggagctccct caaggaggac 1320

gacaaaacaa ccgtctactg caaggaggcc gattacctgt ggaaaaagta cagggaaagc 1380

attagagaag tcgccgatgc cctggatggc gacaacatta agaagctgtc caagagcaat 1440

attgagattc aggaggataa gctgagaaaa tgtttcatta gctatgctga tagcgtgagc 1500

gagtttacca agctgatcta cctcctcaca agattcctga gcggcaagga gatcaatgac 1560

ctggtgacca ccctgatcaa caagttcgat aatattagat cctttctgga gattatggat 1620

gaactgggcc tggacaggac atttaccgcc gagtacagct tcttcgaagg ctccacaaag 1680

tacctggccg agctcgtgga gctgaatagc tttgtgaagt cctgctcctt cgacatcaat 1740

gccaagagga ccatgtacag ggacgccctg gatatcctgg gcatcgagtc cgacaagacc 1800

gaggaggata ttgagaagat gattgataat atcctgcaga tcgacgccaa cggagataaa 1860

aagctgaaga aaaacaacgg actgagaaac ttcatcgctt ccaacgtcat cgacagcaat 1920

agattcaaat acctggtcag gtacggcaac cccaagaaga ttagagagac agccaagtgc 1980

aagcccgctg tgaggttcgt gctgaacgaa attcccgacg cccagatcga aaggtactat 2040

gaggcttgct gccccaagaa caccgccctc tgcagcgcca ataagagaag ggagaagctg 2100

gccgacatga tcgctgagat caaattcgag aattttagcg atgccggcaa ttaccagaag 2160

gctaacgtca catccaggac cagcgaggcc gagatcaaga gaaaaaacca ggccattatt 2220

aggctctatc tgacagtgat gtatattatg ctgaaaaacc tggtgaacgt gaatgccagg 2280

tacgtcattg ccttccactg cgtggaaagg gacaccaagc tgtacgccga aagcggcctg 2340

gaagtgggca acatcgaaaa gaacaaaaca aacctcacca tggccgtgat gggagtgaaa 2400

ctcgaaaacg gaattatcaa gacagagttc gataagagct tcgccgagaa cgccgctaac 2460

agatacctga ggaacgccag gtggtataaa ctgatcctgg acaacctgaa aaaatccgag 2520

agagccgtgg tgaatgagtt cagaaatacc gtgtgccacc tgaatgccat caggaacatc 2580

aatatcaaca tcaaggaaat taaggaggtg gagaattatt tcgctctgta ccactacctg 2640

atccagaagc acctggagaa caggttcgcc gataagaagg tggaaaggga tacaggcgat 2700

ttcatcagca aactggagga gcacaagacc tattgcaaag actttgtgaa ggcttactgc 2760

acccccttg gctacaacct ggtgagatat aagaatctca ccattgacgg cctcttcgac 2820

aagaactacc ctggcaagga tgactccgat gagcagaag 2859

<210> 115
<211> 2862

<212> DNA
<213> Artificial Sequence

<220>
<223> Human codon-optimized Eubacterium siraeum nCas1 HEPN-/-

<400> 115
ggcaagaaga tccatgccag ggacctcagg gagcaaagaa agaccgatag gacagaaaag 60

ttcgccgatc agaacaagaa gagggaggct gagagagctg tgcccaaaaa agacgccgcc 120

gtctccgtga agagcgtgtc cagcgtcagc agcaagaagg acaacgtcac caagagcatg 180

gccaaggccg ccggagtgaa gtccgtcttt gctgtgggaa acaccgtgta catgacctcc 240

ttcggcaggg gcaacgatgc cgtgctggag cagaaaatcg tggacacctc ccatgagccc 300

ctgaatatcg acgatcccgc ctatcaactc aacgtggtga ccatgaatgg atacagcgtc 360

acaggccata gaggagagac agtgtccgcc gtgacagata atcccctgag gagattcaac 420

ggcaggaaga aggacgaacc cgagcagagc gtccctacag acatgctctg cctgaagccc 480

accctcgaga aaaaattttt tggaaaggag tttgatgaca acattcatat tcagctgatc 540

tataatatcc tggatatcga gaaaatcctg gccgtgtact ccaccaacgc catctacgct 600

ctcaacaaca tgtccgccga cgaaaacatc gagaactccg acttcttcat gaaaagaaca 660

acagacgaga cctttgacga cttcgagaag aagaaggaga gcaccaacag cagagagaaa 720

gctgactttg acgcctttga aaagttcatc ggaaactata ggctggccta tttcgctgac 780

gctttctacg tgaataaaaa gaaccccaag ggaaaggcca agaatgtcct gagggaggac 840

aaggagctct acagcgtgct gacactgatc ggaaagctgg cacactggtg tgtcgccagc 900

gaggaaggaa gggctgaatt ctggctgtac aagctggacg agctcaaaga cgatttcaag 960

aacgtcctgg atgtggtgta taacagaccc gtggaggaga tcaacaacag gttcatcgaa 1020

aataacaaag tcaacatcca gatcctgggc tccgtgtata agaacaccga catcgccgag 1080

ctggtgaggt cctactatga gttcctgatc accaaaaaat acaagaatat gggcttttcc 1140

atcaagaagc tcagggagag catgctggag ggcaagggct acgctgataa ggagtacgat 1200

agcgtgagaa ataagctgta ccagatgacc gatttcatcc tctacaccgg ctacatcaac 1260

gaagatagcg atagggctga cgacctggtg aacacactga ggagctccct caaggaggac 1320

gacaaaacaa ccgtctactg caaggaggcc gattacctgt ggaaaaagta cagggaaagc 1380

attagagaag tcgccgatgc cctggatggc gacaacatta agaagctgtc caagagcaat 1440

attgagattc aggaggataa gctgagaaaa tgtttcatta gctatgctga tagcgtgagc 1500

gagtttacca agctgatcta cctcctcaca agattcctga gcggcaagga gatcaatgac 1560

ctggtgacca ccctgatcaa caagttcgat aatattagat cctttctgga gattatggat 1620

gaactgggcc tggacaggac atttaccgcc gagtacagct tcttcgaagg ctccacaaag 1680

tacctggccg agctcgtgga gctgaatagc tttgtgaagt cctgctcctt cgacatcaat 1740

gccaagagga ccatgtacag ggacgccctg gatatcctgg gcatcgagtc cgacaagacc 1800

gaggaggata ttgagaagat gattgataat atcctgcaga tcgacgccaa cggagataaa 1860

aagctgaaga aaaacaacgg actgagaaac ttcatcgctt ccaacgtcat cgacagcaat 1920

agattcaaat acctggtcag gtacggcaac cccaagaaga ttagagagac agccaagtgc 1980

aagcccgctg tgaggttcgt gctgaacgaa attcccgacg cccagatcga aaggtactat 2040

gaggcttgct gccccaagaa caccgccctc tgcagcgcca ataagagaag ggagaagctg 2100

gccgacatga tcgctgagat caaattcgag aattttagcg atgccggcaa ttaccagaag 2160

gctaacgtca catccaggac cagcgaggcc gagatcaaga gaaaaaacca ggccattatt 2220

aggctctatc tgacagtgat gtatattatg ctgaaaaacc tggtgaacgt gaatgccagg 2280

tacgtcattg ccttccactg cgtggaaagg gacaccaagc tgtacgccga aagcggcctg 2340

gaagtgggca acatcgaaaa gaacaaaaca aacctcacca tggccgtgat gggagtgaaa 2400

ctcgaaaacg gaattatcaa gacagagttc gataagagct tcgccgagaa cgccgctaac 2460

agatacctga ggaacgccag gtggtataaa ctgatcctgg acaacctgaa aaaatccgag 2520

agagccgtgg tgaatgagtt cgcaaatacc gtgtgcgccc tgaatgccat caggaacatc 2580

aatatcaaca tcaaggaaat taaggaggtg gagaattatt tcgctctgta ccactacctg 2640

atccagaagc acctggagaa caggttcgcc gataagaagg tggaaaggga tacaggcgat 2700

ttcatcagca aactggagga gcacaagacc tattgcaaag actttgtgaa ggcttactgc 2760

acccccttg gctacaacct ggtgagatat aagaatctca ccattgacgg cctcttcgac 2820

aagaactacc ctggcaagga tgactccgat gagcagaagt ga 2862

<210> 116
<211> 2892
<212> DNA
<213> Artificial Sequence

<220>
<223> Human codon-optimized Eubacterium siraeum nCas1 with N-terminal NLS

<400> 116
agccccaaga agaagagaaa ggtggaggcc agcggcaaga agatccatgc cagggacctc 60

agggagcaaa gaaagaccga taggacagaa aagttcgccg atcagaacaa gaagagggag 120

gctgagagag ctgtgcccaa aaaagacgcc gccgtctccg tgaagagcgt gtccagcgtc 180

agcagcaaga aggacaacgt caccaagagc atggccaagg ccgccggagt gaagtccgtc 240

tttgctgtgg gaaacaccgt gtacatgacc tccttcggca ggggcaacga tgccgtgctg 300

gagcagaaaa tcgtggacac ctcccatgag cccctgaata tcgacgatcc cgcctatcaa 360

ctcaacgtgg tgaccatgaa tggatacagc gtcacaggcc atagaggaga gacagtgtcc 420

gccgtgacag ataatccccct gaggagattc aacggcagga agaaggacga acccgagcag 480

agcgtcccta cagacatgct ctgcctgaag cccaccctcg agaaaaaatt ttttggaaag 540

gagtttgatg acaacattca tattcagctg atctataata tcctggatat cgagaaaatc 600

ctggccgtgt actccaccaa cgccatctac gctctcaaca acatgtccgc cgacgaaaac 660

atcgagaact ccgacttctt catgaaaaga acaacagacg agacctttga cgacttcgag 720

aagaagaagg agagcaccaa cagcagagag aaagctgact ttgacgcctt tgaaaagttc 780

atcggaaact ataggctggc ctatttcgct gacgctttct acgtgaataa aaagaacccc 840

aagggaaagg ccaagaatgt cctgagggag gacaaggagc tctacagcgt gctgacactg 900

atcggaaagc tgagacactg gtgtgtccac agcgaggaag gaagggctga attctggctg 960

tacaagctgg acgagctcaa agacgatttc aagaacgtcc tggatgtggt gtataacaga 1020

cccgtggagg agatcaacaa caggttcatc gaaaataaca aagtcaacat ccagatcctg 1080

ggctccgtgt ataagaacac cgacatcgcc gagctggtga ggtcctacta tgagttcctg 1140

atcaccaaaa aatacaagaa tatgggcttt tccatcaaga agctcaggga gagcatgctg 1200

gagggcaagg gctacgctga taaggagtac gatagcgtga gaaataagct gtaccagatg 1260

accgatttca tcctctacac cggctacatc aacgaagata gcgatagggc tgacgacctg 1320

gtgaacacac tgaggagctc cctcaaggag gacgacaaaa caaccgtcta ctgcaaggag 1380

gccgattacc tgtggaaaaa gtacagggaa agcattagag aagtcgccga tgccctggat 1440

ggcgacaaca ttaagaagct gtccaagagc aatattgaga ttcaggagga taagctgaga 1500

aaatgtttca ttagctatgc tgatagcgtg agcgagttta ccaagctgat ctacctcctc 1560

acaagattcc tgagcggcaa ggagatcaat gacctggtga ccaccctgat caacaagttc 1620

gataatatta gatcctttct ggagattatg gatgaactgg gcctggacag gacatttacc 1680

gccgagtaca gcttcttcga aggctccaca aagtacctgg ccgagctcgt ggagctgaat 1740

agctttgtga agtcctgctc cttcgacatc aatgccaaga ggaccatgta cagggacgcc 1800

ctggatatcc tgggcatcga gtccgacaag accgaggagg atattgagaa gatgattgat 1860

aatatcctgc agatcgacgc caacggagat aaaaagctga agaaaaacaa cggactgaga 1920

aacttcatcg cttccaacgt catcgacagc aatagattca aatacctggt caggtacggc 1980

aaccccaaga agattagaga gacagccaag tgcaagcccg ctgtgaggtt cgtgctgaac 2040

gaaattcccg acgcccagat cgaaaggtac tatgaggctt gctgccccaa gaacaccgcc 2100

ctctgcagcg ccaataagag aagggagaag ctggccgaca tgatcgctga gatcaaattc 2160

gagaatttta gcgatgccgg caattaccag aaggctaacg tcacatccag gaccagcgag 2220

gccgagatca agagaaaaaa ccaggccatt attaggctct atctgacagt gatgtatatt 2280

atgctgaaaa acctggtgaa cgtgaatgcc aggtacgtca ttgccttcca ctgcgtggaa 2340

agggacacca agctgtacgc cgaaagcggc ctggaagtgg gcaacatcga aaagaacaaa 2400

acaaacctca ccatggccgt gatgggagtg aaactcgaaa acggaattat caagacagag 2460

ttcgataaga gcttcgccga gaacgccgct aacagatacc tgaggaacgc caggtggtat 2520

aaactgatcc tggacaacct gaaaaaatcc gagagagccg tggtgaatga gttcagaaat 2580

accgtgtgcc acctgaatgc catcaggaac atcaatatca acatcaagga aattaaggag 2640

gtggagaatt atttcgctct gtaccactac ctgatccaga agcacctgga gaacaggttc 2700

gccgataaga aggtggaaag ggatacaggc gatttcatca gcaaactgga ggagcacaag 2760

acctattgca aagactttgt gaaggcttac tgcaccccct ttggctacaa cctggtgaga 2820

tataagaatc tcaccattga cggcctcttc gacaagaact accctggcaa ggatgactcc 2880

gatgagcaga ag 2892

<210> 117
<211> 2931
<212> DNA
<213> Artificial Sequence

<220>
<223> Human codon-optimized Eubacterium siraeum nCas1 with N- and C-terminal NLS tags

<400> 117
agccccaaga agaagagaaa ggtggaggcc agcggcaaga agatccatgc cagggacctc 60

agggagcaaa gaaagaccga taggacagaa aagttcgccg atcagaacaa gaagagggag 120

gctgagagag ctgtgcccaa aaaagacgcc gccgtctccg tgaagagcgt gtccagcgtc 180

agcagcaaga aggacaacgt caccaagagc atggccaagg ccgccggagt gaagtccgtc 240

tttgctgtgg gaaacaccgt gtacatgacc tccttcggca ggggcaacga tgccgtgctg 300

gagcagaaaa tcgtggacac ctcccatgag cccctgaata tcgacgatcc cgcctatcaa 360

ctcaacgtgg tgaccatgaa tggatacagc gtcacaggcc atagaggaga gacagtgtcc 420

gccgtgacag ataatcccct gaggagattc aacggcagga agaaggacga acccgagcag 480

agcgtcccta cagacatgct ctgcctgaag cccaccctcg agaaaaaatt ttttggaaag 540

gagtttgatg acaacattca tattcagctg atctataata tcctggatat cgagaaaatc 600

ctggccgtgt actccaccaa cgccatctac gctctcaaca acatgtccgc cgacgaaaac 660

atcgagaact ccgacttctt catgaaaaga acaacagacg agacctttga cgacttcgag 720

aagaagaagg agagcaccaa cagcagagag aaagctgact ttgacgcctt tgaaaagttc 780

atcggaaact ataggctggc ctatttcgct gacgctttct acgtgaataa aaagaacccc 840

aagggaaagg ccaagaatgt cctgagggag gacaaggagc tctacagcgt gctgacactg 900

atcggaaagc tgagacactg gtgtgtccac agcgaggaag gaagggctga attctggctg 960

tacaagctgg acgagctcaa agacgatttc aagaacgtcc tggatgtggt gtataacaga 1020

cccgtggagg agatcaacaa caggttcatc gaaaataaca aagtcaacat ccagatcctg 1080

ggctccgtgt ataagaacac cgacatcgcc gagctggtga ggtcctacta tgagttcctg 1140

atcaccaaaa aatacaagaa tatgggcttt tccatcaaga agctcaggga gagcatgctg 1200

gagggcaagg gctacgctga taaggagtac gatagcgtga gaaataagct gtaccagatg 1260

accgatttca tcctctacac cggctacatc aacgaagata gcgatagggc tgacgacctg 1320

gtgaacacac tgaggagctc cctcaaggag gacgacaaaa caaccgtcta ctgcaaggag 1380

gccgattacc tgtggaaaaa gtacagggaa agcattagag aagtcgccga tgccctggat 1440

ggcgacaaca ttaagaagct gtccaagagc aatattgaga ttcaggagga taagctgaga 1500

aaatgtttca ttagctatgc tgatagcgtg agcgagttta ccaagctgat ctacctcctc 1560

acaagattcc tgagcggcaa ggagatcaat gacctggtga ccaccctgat caacaagttc 1620

gataatatta gatcctttct ggagattatg gatgaactgg gcctggacag gacatttacc 1680

gccgagtaca gcttcttcga aggctccaca aagtacctgg ccgagctcgt ggagctgaat 1740

agctttgtga agtcctgctc cttcgacatc aatgccaaga ggaccatgta cagggacgcc 1800

ctggatatcc tgggcatcga gtccgacaag accgaggagg atattgagaa gatgattgat 1860

aatatcctgc agatcgacgc caacggagat aaaaagctga agaaaaacaa cggactgaga 1920

aacttcatcg cttccaacgt catcgacagc aatagattca aatacctggt caggtacggc 1980

aaccccaaga agattagaga gacagccaag tgcaagcccg ctgtgaggtt cgtgctgaac 2040

gaaattcccg acgcccagat cgaaaggtac tatgaggctt gctgccccaa gaacaccgcc 2100

ctctgcagcg ccaataagag aagggagaag ctggccgaca tgatcgctga gatcaaattc 2160

gagaatttta gcgatgccgg caattaccag aaggctaacg tcacatccag gaccagcgag 2220

gccgagatca agagaaaaaa ccaggccatt attaggctct atctgacagt gatgtatatt 2280

atgctgaaaa acctggtgaa cgtgaatgcc aggtacgtca ttgccttcca ctgcgtggaa 2340

agggacacca agctgtacgc cgaaagcggc ctggaagtgg gcaacatcga aaagaacaaa 2400

acaaacctca ccatggccgt gatgggagtg aaactcgaaa acggaattat caagacagag 2460

ttcgataaga gcttcgccga gaacgccgct aacagatacc tgaggaacgc caggtggtat 2520

aaactgatcc tggacaacct gaaaaaatcc gagagagccg tggtgaatga gttcagaaat 2580

accgtgtgcc acctgaatgc catcaggaac atcaatatca acatcaagga aattaaggag 2640

gtggagaatt atttcgctct gtaccactac ctgatccaga agcacctgga gaacaggttc 2700

gccgataaga aggtggaaag ggatacaggc gatttcatca gcaaactgga ggagcacaag 2760

acctattgca aagactttgt gaaggcttac tgcaccccct ttggctacaa cctggtgaga 2820

tataagaatc tcaccattga cggcctcttc gacaagaact accctggcaa ggatgactcc 2880

gatgagcaga agggatccgg acctaagaaa aagaggaagg tggcggccgc t 2931

<210> 118
<211> 2754
<212> DNA
<213> Artificial Sequence

<220>
<223> Human codon-optimized uncultured Ruminococcus sp. nCas1

<400> 118
gccaagaaga acaagatgaa acccagggag ctgagagagg ctcagaagaa ggctaggcaa 60

ctgaaagccg ccgagatcaa taacaatgcc gcccctgcca tcgccgctat gcccgctgcc 120

gaagtcatcg ctcccgccgc cgaaaagaag aagagctccg tgaaagccgc tggaatgaag 180

tccattctgg tgtccgagaa caaaatgtat atcacctcct tcggcaaagg caattccgct 240

gtgctggagt atgaggtgga caacaacgac tacaatcaga cccagctcag cagcaaggat 300

aactccaaca tccagctggg cggcgtgaat gaagtgaata tcaccttcag cagcaagcac 360

ggatttgagt ccggagtcga gattaacaca agcaacccta cccacaggtc cggagaaagc 420

agccctgtca gaggagacat gctgggactg aaaagcgagc tggagaagag attctttgga 480

aagacctttg atgacaatat ccacatccag ctcatttaca acatcctgga tatcgagaaa 540

atcctggccg tgtatgtgac caacatcgtg tacgctctga acaacatgct cggcgtcaag 600

ggcagcgagt cccacgacga ctttattggc tatctgtcca ccaacaatat ttacgatgtc 660

ttcattgacc ccgacaacag cagcctctcc gacgacaaaa aagccaacgt caggaagagc 720

ctcagcaaat tcaacgccct gctgaagacc aagaggctcg gatactttgg cctggaggaa 780

cccaagacca aggacaacag agtgagccag gcctataaga aaagggtgta ccacatgctg 840

gccatcgtgg gccagatcag acagtgcgtg ttccatgaca agtccggcgc caagagattc 900

gacctgtact ccttcattaa caacatcgac cctgagtaca gggatacact ggattacctc 960

gtggaggaga ggctgaagag catcaacaaa gactttattg aggacaacaa agtcaacatc 1020

agcctgctga tcgatatgat gaagggctac gaagctgacg acatcatcag gctgtactat 1080

gatttcattg tcctgaagag ccagaagaac ctgggattct ccattaagaa actgagagaa 1140

aagatgctcg acgagtacgg cttcaggttc aaggacaagc agtatgacag cgtgagatcc 1200

aagatgtaca agctgatgga ttttctcctg ttttgcaact attacaggaa cgacatcgcc 1260

gccggcgaat ccctggtcag gaaactgaga ttcagcatga ccgacgacga gaaggaaggc 1320

atctacgccg acgaagccgc taagctgtgg ggaaaattta gaaatgactt cgagaacatc 1380

gctgaccaca tgaacggcga tgtgatcaag gagctcggca aggctgacat ggatttcgac 1440

gagaagatcc tggattccga gaagaagaat gccagcgatc tgctgtactt cagcaagatg 1500

atttacatgc tcacatattt tctcgatggc aaggagatca acgatctcct gaccacactg 1560

atcagcaagt tcgacaacat taaagagttc ctcaaaatta tgaaatccag cgccgtggat 1620

gtcgagtgcg aactcaccgc cggctacaag ctcttcaacg acagccagag gatcaccaat 1680

gagctgttca ttgtgaaaaa tatcgctagc atgaggaaac ccgccgcctc cgctaaactg 1740

acaatgttta gagatgccct caccatcctg ggcatcgacg ataagatcac agacgacaga 1800

attagcggaa tcctgaagct caaggaaaag ggcaagggaa tccatggcct cagaaatttc 1860

attaccaata acgtgatcga aagcagcagg ttcgtgtatc tgatcaagta cgccaatgcc 1920

caaaagatca gggaggtggc caagaacgaa aaggtggtga tgttcgtgct gggcggcatt 1980

cctgacacac agatcgaaag gtactacaag agctgtgtgg agttccccga tatgaactcc 2040

agcctgggcg tgaagagatc cgaactggct aggatgatca agaatattag cttcgacgac 2100

ttcaagaacg tgaagcaaca ggccaaagga agggagaacg tcgccaaaga gagagccaag 2160

gccgtgatcg gactctacct gaccgtcatg tacctcctcg tgaagaacct ggtgaacgtg 2220

aacgctagat acgtcatcgc catccattgc ctcgagaggg acttcggact gtacaaggag 2280

atcattcccg agctggccag caaaaatctc aagaacgatt acagaatcct gagccaaacc 2340

ctgtgcgagc tgtgtgacaa gtcccccaac ctgtttctga agaagaatga aaggctgaga 2400

aagtgcgtcg aagtcgacat caacaacgct gacagctcca tgaccaggaa atacaggaac 2460

tgtatcgccc atctgaccgt ggtgagggaa ctcaaggagt acatcggaga tatctgcacc 2520

gtggactcct acttcagcat ttatcactac gtgatgcaga ggtgcattac caagagagaa 2580

aatgacacaa agcaggagga gaagattaag tacgaggacg atctgctgaa aaaccacgga 2640

tacaccaagg atttcgtcaa ggccctcaat agccccttcg gctacaatat ccctaggttt 2700

aaaaacctga gcatcgagca gctgtttgac agaaatgaat acctgacaga gaag 2754

<210> 119
<211> 2757
<212> DNA
<213> Artificial Sequence

<220>
<223> Human codon-optimized uncultured Ruminococcus sp. nCas1 HEPN-/-

<400> 119
atggccaaga agaacaagat gaaacccagg gagctgagag aggctcagaa gaaggctagg 60

caactgaaag ccgccgagat caataacaat gccgcccctg ccatcgccgc tatgcccgct 120

gccgaagtca tcgctcccgc cgccgaaaag aagaagagct ccgtgaaagc cgctggaatg 180

aagtccattc tggtgtccga gaacaaaatg tatatcacct ccttcggcaa aggcaattcc 240

gctgtgctgg agtatgaggt ggacaacaac gactacaatc agacccagct cagcagcaag 300

gataactcca acatccagct gggcggcgtg aatgaagtga atatcacctt cagcagcaag 360

cacggatttg agtccggagt cgagattaac acaagcaacc ctacccacag gtccggagaa 420

agcagccctg tcagaggaga catgctggga ctgaaaagcg agctggagaa gagattcttt 480

ggaaagacct ttgatgacaa tatccacatc cagctcattt acaacatcct ggatatcgag 540

aaaatcctgg ccgtgtatgt gaccaacatc gtgtacgctc tgaacaacat gctcggcgtc 600

aagggcagcg agtcccacga cgactttatt ggctatctgt ccaccaacaa tatttacgat 660

gtcttcattg accccgacaa cagcagcctc tccgacgaca aaaaagccaa cgtcaggaag 720

agcctcagca aattcaacgc cctgctgaag accaagaggc tcggatactt tggcctggag 780

gaacccaaga ccaaggacaa cagagtgagc caggcctata agaaaagggt gtaccacatg 840

ctggccatcg tgggccagat cgcacagtgc gtgttcgctg acaagtccgg cgccaagaga 900

ttcgacctgt actccttcat taacaacatc gaccctgagt acagggatac actggattac 960

ctcgtggagg agaggctgaa gagcatcaac aaagacttta ttgaggacaa caaagtcaac 1020

atcagcctgc tgatcgatat gatgaagggc tacgaagctg acgacatcat caggctgtac 1080

tatgatttca ttgtcctgaa gagccagaag aacctgggat tctccattaa gaaactgaga 1140

gaaaagatgc tcgacgagta cggcttcagg ttcaaggaca agcagtatga cagcgtgaga 1200

tccaagatgt acaagctgat ggattttctc ctgttttgca actattacag gaacgacatc 1260

gccgccggcg aatccctggt caggaaactg agattcagca tgaccgacga cgagaaggaa 1320

ggcatctacg ccgacgaagc cgctaagctg tggggaaaat ttagaaatga cttcgagaac 1380

atcgctgacc acatgaacgg cgatgtgatc aaggagctcg gcaaggctga catggatttc 1440

gacgagaaga tcctggattc cgagaagaag aatgccagcg atctgctgta cttcagcaag 1500

atgatttaca tgctcacata ttttctcgat ggcaaggaga tcaacgatct cctgaccaca 1560

ctgatcagca agttcgacaa cattaaagag ttcctcaaaa ttatgaaatc cagcgccgtg 1620

gatgtcgagt gcgaactcac cgccggctac aagctcttca acgacagcca gaggatcacc 1680

aatgagctgt tcattgtgaa aaatatcgct agcatgagga aacccgccgc ctccgctaaa 1740

ctgacaatgt ttagagatgc cctcaccatc ctgggcatcg acgataagat cacagacgac 1800

agaattagcg gaatcctgaa gctcaaggaa aagggcaagg gaatccatgg cctcagaaat 1860

ttcattacca ataacgtgat cgaaagcagc aggttcgtgt atctgatcaa gtacgccaat 1920

gcccaaaaga tcagggaggt ggccaagaac gaaaaggtgg tgatgttcgt gctgggcggc 1980

attcctgaca cacagatcga aaggtactac aagagctgtg tggagttccc cgatatgaac 2040

tccagcctgg gcgtgaagag atccgaactg gctaggatga tcaagaatat tagcttcgac 2100

gacttcaaga acgtgaagca acaggccaaa ggaagggaga acgtcgccaa agagagagcc 2160

aaggccgtga tcggactcta cctgaccgtc atgtacctcc tcgtgaagaa cctggtgaac 2220

gtgaacgcta gatacgtcat cgccatccat tgcctcgaga gggacttcgg actgtacaag 2280

gagatcattc ccgagctggc cagcaaaaat ctcaagaacg attacagaat cctgagccaa 2340

accctgtgcg agctgtgtga caagtccccc aacctgtttc tgaagaagaa tgaaaggctg 2400

agaaagtgcg tcgaagtcga catcaacaac gctgacagct ccatgaccag gaaatacgcg 2460

aactgtatcg ccgctctgac cgtggtgagg gaactcaagg agtacatcgg agatatctgc 2520

accgtggact cctacttcag catttatcac tacgtgatgc agaggtgcat taccaagaga 2580

gaaaatgaca caaagcagga ggagaagatt aagtacgagg acgatctgct gaaaaaccac 2640

ggatacacca aggatttcgt caaggccctc aatagcccct tcggctacaa tatccctagg 2700

tttaaaaacc tgagcatcga gcagctgttt gacagaaatg aatacctgac agagaag 2757

<210> 120
<211> 2787
<212> DNA
<213> Artificial Sequence

<220>
<223> Human codon-optimized uncultured Ruminococcus sp. nCas1 with N-terminal NLS

<400> 120
agccccaaga agaagagaaa ggtggaggcc agcgccaaga agaacaagat gaaacccagg 60

gagctgagag aggctcagaa gaaggctagg caactgaaag ccgccgagat caataacaat 120

gccgcccctg ccatcgccgc tatgcccgct gccgaagtca tcgctcccgc cgccgaaaag 180

aagaagagct ccgtgaaagc cgctggaatg aagtccattc tggtgtccga gaacaaaatg 240

tatatcacct ccttcggcaa aggcaattcc gctgtgctgg agtatgaggt ggacaacaac 300

gactacaatc agacccagct cagcagcaag gataactcca acatccagct gggcggcgtg 360

aatgaagtga atatcacctt cagcagcaag cacggatttg agtccggagt cgagattaac 420

acaagcaacc ctacccacag gtccggagaa agcagccctg tcagaggaga catgctggga 480

ctgaaaagcg agctggagaa gagattcttt ggaaagacct ttgatgacaa tatccacatc 540

cagctcattt acaacatcct ggatatcgag aaaatcctgg ccgtgtatgt gaccaacatc 600

gtgtacgctc tgaacaacat gctcggcgtc aagggcagcg agtcccacga cgactttatt 660

ggctatctgt ccaccaacaa tatttacgat gtcttcattg accccgacaa cagcagcctc 720

tccgacgaca aaaaagccaa cgtcaggaag agcctcagca aattcaacgc cctgctgaag 780

accaagaggc tcggatactt tggcctggag gaacccaaga ccaaggacaa cagagtgagc 840

caggcctata agaaaagggt gtaccacatg ctggccatcg tgggccagat cagacagtgc 900

gtgttccatg acaagtccgg cgccaagaga ttcgacctgt actccttcat taacaacatc 960

gaccctgagt acagggatac actggattac ctcgtggagg agaggctgaa gagcatcaac 1020

aaagacttta ttgaggacaa caaagtcaac atcagcctgc tgatcgatat gatgaagggc 1080

tacgaagctg acgacatcat caggctgtac tatgatttca ttgtcctgaa gagccagaag 1140

aacctgggat tctccattaa gaaactgaga gaaaagatgc tcgacgagta cggcttcagg 1200

ttcaaggaca agcagtatga cagcgtgaga tccaagatgt acaagctgat ggattttctc 1260

ctgttttgca actattacag gaacgacatc gccgccggcg aatccctggt caggaaactg 1320

agattcagca tgaccgacga cgagaaggaa ggcatctacg ccgacgaagc cgctaagctg 1380

tggggaaaat ttagaaatga cttcgagaac atcgctgacc acatgaacgg cgatgtgatc 1440

aaggagctcg gcaaggctga catggatttc gacgagaaga tcctggattc cgagaagaag 1500

aatgccagcg atctgctgta cttcagcaag atgatttaca tgctcacata ttttctcgat 1560

ggcaaggaga tcaacgatct cctgaccaca ctgatcagca agttcgacaa cattaaagag 1620

ttcctcaaaa ttatgaaatc cagcgccgtg gatgtcgagt gcgaactcac cgccggctac 1680

aagctcttca acgacagcca gaggatcacc aatgagctgt tcattgtgaa aaatatcgct 1740

agcatgagga aacccgccgc ctccgctaaa ctgacaatgt ttagagatgc cctcaccatc 1800

ctgggcatcg acgataagat cacagacgac agaattagcg gaatcctgaa gctcaaggaa 1860

aagggcaagg gaatccatgg cctcagaaat ttcattacca ataacgtgat cgaaagcagc 1920

aggttcgtgt atctgatcaa gtacgccaat gcccaaaaga tcagggaggt ggccaagaac 1980

gaaaaggtgg tgatgttcgt gctgggcggc attcctgaca cacagatcga aaggtactac 2040

aagagctgtg tggagttccc cgatatgaac tccagcctgg gcgtgaagag atccgaactg 2100

gctaggatga tcaagaatat tagcttcgac gacttcaaga acgtgaagca acaggccaaa 2160

ggaagggaga acgtcgccaa agagagagcc aaggccgtga tcggactcta cctgaccgtc 2220

atgtacctcc tcgtgaagaa cctggtgaac gtgaacgcta gatacgtcat cgccatccat 2280

tgcctcgaga gggacttcgg actgtacaag gagatcattc ccgagctggc cagcaaaaat 2340

ctcaagaacg attacagaat cctgagccaa accctgtgcg agctgtgtga caagtccccc 2400

aacctgtttc tgaagaagaa tgaaaggctg agaaagtgcg tcgaagtcga catcaacaac 2460

gctgacagct ccatgaccag gaaatacagg aactgtatcg cccatctgac cgtggtgagg 2520

gaactcaagg agtacatcgg agatatctgc accgtggact cctacttcag catttatcac 2580

tacgtgatgc agaggtgcat taccaagaga gaaaatgaca caaagcagga ggagaagatt 2640

aagtacgagg acgatctgct gaaaaaccac ggatacacca aggatttcgt caaggccctc 2700

aatagcccct tcggctacaa tatccctagg tttaaaaacc tgagcatcga gcagctgttt 2760

gacagaaatg aatacctgac agagaag 2787

<210> 121
<211> 2826
<212> DNA
<213> Artificial Sequence

<220>
<223> Human codon-optimized uncultured Ruminococcus sp. nCas1 with N- and C-terminal NLS tags

<400> 121
agccccaaga agaagagaaa ggtggaggcc agcgccaaga agaacaagat gaaacccagg 60

gagctgagag aggctcagaa gaaggctagg caactgaaag ccgccgagat caataacaat 120

gccgcccctg ccatcgccgc tatgcccgct gccgaagtca tcgctcccgc cgccgaaaag 180

aagaagagct ccgtgaaagc cgctggaatg aagtccattc tggtgtccga gaacaaaatg 240

tatatcacct ccttcggcaa aggcaattcc gctgtgctgg agtatgaggt ggacaacaac 300

gactacaatc agacccagct cagcagcaag gataactcca acatccagct gggcggcgtg 360

aatgaagtga atatcacctt cagcagcaag cacggatttg agtccggagt cgagattaac 420

acaagcaacc ctacccacag gtccggagaa agcagccctg tcagaggaga catgctggga 480

ctgaaaagcg agctggagaa gagattcttt ggaaagacct ttgatgacaa tatccacatc 540

cagctcattt acaacatcct ggatatcgag aaaatcctgg ccgtgtatgt gaccaacatc 600

gtgtacgctc tgaacaacat gctcggcgtc aagggcagcg agtcccacga cgactttatt 660

ggctatctgt ccaccaacaa tatttacgat gtcttcattg accccgacaa cagcagcctc 720

tccgacgaca aaaaagccaa cgtcaggaag agcctcagca aattcaacgc cctgctgaag 780

accaagaggc tcggatactt tggcctggag gaacccaaga ccaaggacaa cagagtgagc 840

caggcctata agaaaagggt gtaccacatg ctggccatcg tgggccagat cagacagtgc 900

gtgttccatg acaagtccgg cgccaagaga ttcgacctgt actccttcat taacaacatc 960

gaccctgagt acagggatac actggattac ctcgtggagg agaggctgaa gagcatcaac 1020

aaagacttta ttgaggacaa caaagtcaac atcagcctgc tgatcgatat gatgaagggc 1080

tacgaagctg acgacatcat caggctgtac tatgatttca ttgtcctgaa gagccagaag 1140

aacctgggat tctccattaa gaaactgaga gaaaagatgc tcgacgagta cggcttcagg 1200

ttcaaggaca agcagtatga cagcgtgaga tccaagatgt acaagctgat ggattttctc 1260

ctgttttgca actattacag gaacgacatc gccgccggcg aatccctggt caggaaactg 1320

agattcagca tgaccgacga cgagaaggaa ggcatctacg ccgacgaagc cgctaagctg 1380

tggggaaaat ttagaaatga cttcgagaac atcgctgacc acatgaacgg cgatgtgatc 1440

aaggagctcg gcaaggctga catggatttc gacgagaaga tcctggattc cgagaagaag 1500

aatgccagcg atctgctgta cttcagcaag atgatttaca tgctcacata ttttctcgat 1560

ggcaaggaga tcaacgatct cctgaccaca ctgatcagca agttcgacaa cattaaagag 1620

ttcctcaaaa ttatgaaatc cagcgccgtg gatgtcgagt gcgaactcac cgccggctac 1680

aagctcttca acgacagcca gaggatcacc aatgagctgt tcattgtgaa aaatatcgct 1740

agcatgagga aacccgccgc ctccgctaaa ctgacaatgt ttagagatgc cctcaccatc 1800

ctgggcatcg acgataagat cacagacgac agaattagcg gaatcctgaa gctcaaggaa 1860

aagggcaagg gaatccatgg cctcagaaat ttcattacca ataacgtgat cgaaagcagc 1920

aggttcgtgt atctgatcaa gtacgccaat gcccaaaaga tcagggaggt ggccaagaac 1980

gaaaaggtgg tgatgttcgt gctgggcggc attcctgaca cacagatcga aaggtactac 2040

aagagctgtg tggagttccc cgatatgaac tccagcctgg gcgtgaagag atccgaactg 2100

gctaggatga tcaagaatat tagcttcgac gacttcaaga acgtgaagca acaggccaaa 2160

ggaagggaga acgtcgccaa agagagagcc aaggccgtga tcggactcta cctgaccgtc 2220

atgtacctcc tcgtgaagaa cctggtgaac gtgaacgcta gatacgtcat cgccatccat 2280

tgcctcgaga gggacttcgg actgtacaag gagatcattc ccgagctggc cagcaaaaat 2340

ctcaagaacg attacagaat cctgagccaa accctgtgcg agctgtgtga caagtccccc 2400

aacctgtttc tgaagaagaa tgaaaggctg agaaagtgcg tcgaagtcga catcaacaac 2460

gctgacagct ccatgaccag gaaatacagg aactgtatcg cccatctgac cgtggtgagg 2520

gaactcaagg agtacatcgg agatatctgc accgtggact cctacttcag catttatcac 2580

tacgtgatgc agaggtgcat taccaagaga gaaaatgaca caaagcagga ggagaagatt 2640

aagtacgagg acgatctgct gaaaaaccac ggatacacca aggatttcgt caaggccctc 2700

aatagcccct tcggctacaa tatccctagg tttaaaaacc tgagcatcga gcagctgttt 2760

gacagaaatg aatacctgac agagaaggga tccggaccta agaaaaagag gaaggtggcg 2820

gccgct 2826

<210> 122
<211> 2803
<212> DNA
<213> Artificial Sequence

<220>
<223> Human codon-optimized Ruminococcus flavefaciens FD1 nCas1

<400> 122
aagaaaaaaa tgagcctcag agagaagagg gaagccgaga aacaagccaa aaaagccgcc 60

tacagcgctg cctccaaaaa cacagatagc aaacctgctg agaagaaggc cgagacaccc 120

aagcctgctg agattatttc cgacaactcc agaaacaaaa ccgccgtcaa ggccgccggc 180

ctgaaatcca caattatctc cggcgataag ctgtacatga ccagcttcgg aaagggcaac 240

gccgccgtga tcgagcagaa aattgatatt aatgattact ccttctccgc catgaaggat 300

acccccagcc tggaagtgga caaggccgag agcaaggaga tctccttctc cagccaccac 360

ccttttgtca aaaatgataa actgaccacc tataatcccc tgtacggcgg aaaagataac 420

cccgaaaagc ccgtgggcag agatatgctc ggcctcaaag acaagctcga ggagaggtac 480

ttcggatgca ccttcaatga caacctgcac attcagatta tctataacat cctcgacatt 540

gagaaaattc tggccgtgca ctccgccaac attaccacag ctctggacca catggtggac 600

gaggacgatg agaagtacct caatagcgat tacattggct acatgaacac catcaacacc 660

tatgatgtgt tcatggaccc ctccaaaaat tcctccctct cccccaagga caggaagaac 720

attgataaca gcagggccaa gtttgagaag ctcctgtcca caaagaggct cggctacttt 780

ggctttgact acgatgccaa cggcaaggat aagaagaaaa acgaggagat caagaaaagg 840

ctctatcacc tgacagcctt tgccggccag ctcaggcagt ggtccttcca tagcgccggc 900

aattatccca gaacctggct ctacaagctc gatagcctgg acaaagagta tctcgacacc 960

ctggaccatt attttgataa gagattcaac gatattaatg acgatttcgt caccaagaac 1020

gccacaaacc tgtacatcct gaaggaggtc ttccccgagg ctaatttcaa ggatatcgcc 1080

gatctctact acgatttcat tgtcatcaaa agccataaga atatgggatt ttccatcaag 1140

aagctgaggg agaagatgct ggagtgcgat ggagctgata gaatcaagga acaggacatg 1200

gactccgtga ggagcaagct gtataagctg atcgattttt gcattttcaa atactaccac 1260

gagttccctg agctgagcga gaagaacgtc gacatcctca gggctgccgt gagcgatacc 1320

aaaaaggaca acctgtatag cgacgaggcc gctaggctct ggagcatctt caaggaaaag 1380

ttcctgggct tctgcgacaa gattgtggtg tgggtgaccg gcgagcacga gaaagacatc 1440

acctccgtga ttgacaagga cgcttacagg aacaggtcca acgtgagcta cttttccaag 1500

ctgatgtacg ccatgtgttt cttcctggat ggaaaggaaa ttaacgatct gctgaccaca 1560

ctgatcaaca aattcgacaa cattgccaac cagatcaaga ccgccaagga gctcggcatt 1620

aacaccgcct tcgtgaaaaa ctacgatttt ttcaaccaca gcgaaaagta cgtggacgag 1680

ctgaacattg tgaaaaacat cgccaggatg aaaaagccct ccagcaacgc caagaaggcc 1740

atgtaccatg acgccctgac aatcctgggc atccccgagg atatggacga gaaggccctg 1800

gatgaggagc tcgatctcat cctggaaaag aagaccgatc ccgtgaccgg aaagcccctc 1860

aaaggcaaaa atcccctgag gaattttatc gctaataatg tgattgagaa tagcaggttc 1920

atttacctga ttaaattttg caaccctgag aatgtcagga aaatcgtgaa caacacaaag 1980

gtgaccgagt tcgtgctcaa gaggatcccc gatgcccaga tcgagaggta ctacaagagc 2040

tgtaccgatt ccgagatgaa ccctcccaca gaaaagaaga tcaccgagct ggccggcaag 2100

ctgaaagata tgaattttgg aaacttcagg aacgtgaggc agtccgccaa ggaaaacatg 2160

gagaaagaga ggttcaaggc cgtcattggc ctgtacctca ccgtcgtcta cagagtggtc 2220

aaaaacctgg tggacgtcaa tagcaggtac atcatggcct tccatagcct ggagagagat 2280

tcccagctgt acaacgtgag cgtcgataat gattacctgg ccctcaccga caccctggtc 2340

aaggagggcg acaacagcag gtccaggtac ctggccggaa ataaaaggct gagggattgc 2400

gtgaagcaag acatcgacaa tgccaaaaag tggttcgtct ccgacaagta caactccatt 2460

acaaagtaca ggaacaacgt ggcccacctg accgccgtga gaaactgcgc cgagttcatc 2520

ggcgacatta ccaagatcga ctcctacttc gccctctacc actacctcat tcagagacag 2580

ctggctaagg gcctcgatca tgagaggtcc ggattcgaca ggaattaccc ccagtacgcc 2640

cctctgttca aatggcacac ctacgtgaag gatgtcgtca aggccctgaa cgccccttc 2700

ggatacaata tccctaggtt caagaatctg agcatcgacg ccctctttga taggaatgaa 2760

atcaagaaga acgatggcga gaagaagtcc gacgactgag cct 2803

<210> 123
<211> 2799
<212> DNA
<213> Artificial Sequence

<220>
<223> Human codon-optimized Ruminococcus flavefaciens FD1 nCas1 HEPN-/-

<400> 123
aagaaaaaaa tgagcctcag agagaagagg gaagccgaga aacaagccaa aaaagccgcc 60

tacagcgctg cctccaaaaa cacagatagc aaacctgctg agaagaaggc cgagacaccc 120

aagcctgctg agattatttc cgacaactcc agaaacaaaa ccgccgtcaa ggccgccggc 180

ctgaaatcca caattatctc cggcgataag ctgtacatga ccagcttcgg aaagggcaac 240

gccgccgtga tcgagcagaa aattgatatt aatgattact ccttctccgc catgaaggat 300

acccccagcc tggaagtgga caaggccgag agcaaggaga tctccttctc cagccaccac 360

ccttttgtca aaaatgataa actgaccacc tataatcccc tgtacggcgg aaaagataac 420

cccgaaaagc ccgtgggcag agatatgctc ggcctcaaag acaagctcga ggagaggtac 480

ttcggatgca ccttcaatga caacctgcac attcagatta tctataacat cctcgacatt 540

gagaaaattc tggccgtgca ctccgccaac attaccacag ctctggacca catggtggac 600

gaggacgatg agaagtacct caatagcgat tacattggct acatgaacac catcaacacc 660

tatgatgtgt tcatggaccc ctccaaaaat tcctccctct cccccaagga caggaagaac 720

attgataaca gcagggccaa gtttgagaag ctcctgtcca caaagaggct cggctacttt 780

ggctttgact acgatgccaa cggcaaggat aagaagaaaa acgaggagat caagaaaagg 840

ctctatcacc tgacagcctt tgccggccag ctcgcgcagt ggtccttcgc tagcgccggc 900

aattatccca gaacctggct ctacaagctc gatagcctgg acaaagagta tctcgacacc 960

ctggaccatt attttgataa gagattcaac gatattaatg acgatttcgt caccaagaac 1020

gccacaaacc tgtacatcct gaaggaggtc ttccccgagg ctaatttcaa ggatatcgcc 1080

gatctctact acgatttcat tgtcatcaaa agccataaga atatgggatt ttccatcaag 1140

aagctgaggg agaagatgct ggagtgcgat ggagctgata gaatcaagga acaggacatg 1200

gactccgtga ggagcaagct gtataagctg atcgattttt gcattttcaa atactaccac 1260

gagttccctg agctgagcga gaagaacgtc gacatcctca gggctgccgt gagcgatacc 1320

aaaaaggaca acctgtatag cgacgaggcc gctaggctct ggagcatctt caaggaaaag 1380

ttcctgggct tctgcgacaa gattgtggtg tgggtgaccg gcgagcacga gaaagacatc 1440

acctccgtga ttgacaagga cgcttacagg aacaggtcca acgtgagcta cttttccaag 1500

ctgatgtacg ccatgtgttt cttcctggat ggaaaggaaa ttaacgatct gctgaccaca 1560

ctgatcaaca aattcgacaa cattgccaac cagatcaaga ccgccaagga gctcggcatt 1620

aacaccgcct tcgtgaaaaa ctacgatttt ttcaaccaca gcgaaaagta cgtggacgag 1680

ctgaacattg tgaaaaacat cgccaggatg aaaaagccct ccagcaacgc caagaaggcc 1740

atgtaccatg acgccctgac aatcctgggc atccccgagg atatggacga gaaggccctg 1800

gatgaggagc tcgatctcat cctggaaaag aagaccgatc ccgtgaccgg aaagcccctc 1860

aaaggcaaaa atcccctgag gaattttatc gctaataatg tgattgagaa tagcaggttc 1920

atttacctga ttaaattttg caaccctgag aatgtcagga aaatcgtgaa caacacaaag 1980

gtgaccgagt tcgtgctcaa gaggatcccc gatgcccaga tcgagaggta ctacaagagc 2040

tgtaccgatt ccgagatgaa ccctcccaca gaaaagaaga tcaccgagct ggccggcaag 2100

ctgaaagata tgaattttgg aaacttcagg aacgtgaggc agtccgccaa ggaaaacatg 2160

gagaaagaga ggttcaaggc cgtcattggc ctgtacctca ccgtcgtcta cagagtggtc 2220

aaaaacctgg tggacgtcaa tagcaggtac atcatggcct tccatagcct ggagagagat 2280

tcccagctgt acaacgtgag cgtcgataat gattacctgg ccctcaccga caccctggtc 2340

aaggagggcg acaacagcag gtccaggtac ctggccggaa ataaaaggct gagggattgc 2400

gtgaagcaag acatcgacaa tgccaaaaag tggttcgtct ccgacaagta caactccatt 2460

acaaagtacg cgaacaacgt ggccgccctg accgccgtga gaaactgcgc cgagttcatc 2520

ggcgacatta ccaagatcga ctcctacttc gccctctacc actacctcat tcagagacag 2580

ctggctaagg gcctcgatca tgagaggtcc ggattcgaca ggaattaccc ccagtacgcc 2640

cctctgttca aatggcacac ctacgtgaag gatgtcgtca aggccctgaa cgccccttc 2700

ggatacaata tccctaggtt caagaatctg agcatcgacg ccctctttga taggaatgaa 2760

atcaagaaga acgatggcga gaagaagtcc gacgactga 2799

<210> 124
<211> 2757
<212> DNA
<213> Artificial Sequence

<220>
<223> Ruminococcus bicirculans nCas1 sequence

<400> 124
atggcaaaaa agaataaaat gaagcctaga gagctgcgtg aggctcagaa aaaagccaga 60

cagctcaaag cggctgagat aaataataac gctgttcctg caatcgctgc catgcctgct 120

gcagaggctg ctgcacctgc ggcagagaag aaaaaatcct ccgtaaaggc ggcaggaatg 180

aagtctattc ttgtcagtga aaataaaatg tacataacct ctttcggcaa gggcaattct 240

gcggtgcttg aatatgaggt ggataataat gactacaaca aaactcagct ttcctcaaag 300

gacaacagta atatcgagct ctgtgatgta ggcaaagtaa acatcacttt ttcgagcaga 360

cgtggctttg agagcggtgt ggagataaac acttcaaacc ctactcacag aagcggtgaa 420

agctcgtctg taagagggga tatgctgggg cttaaatcgg agcttgaaaa gcgctttttc 480

ggcaagaatt ttgatgataa tatacatatc cagcttattt acaacattct ggatatcgaa 540

aagatacttg cagtgtatgt gacgaatatc gtttatgcac tgaacaatat gcttggggaa 600

ggcgatgaga gcaattacga tttcatgggg tatctttcca catttaacac ttataaagtt 660

tttactaatc ctaatggcag cactttatcc gacgataaga aagagaatat cagaaaaagt 720

cttagcaaat tcaatgccct gctgaaaact aagcgtcttg gctatttcgg ccttgaagag 780

ccaaagacaa aggatacaag agcttcggaa gcatacaaaa agcgtgttta tcatatgctt 840

gcaattgtgg ggcagataag acagtgtgtt tttcatgata aatcgggtgc aaaaagattt 900

gacctttaca gttttattaa caatattgat cccgaataca gagaaaccct tgactatctt 960

gtagatgaga gatttgattc tataaataag ggctttatcc agggcaacaa ggtcaatatc 1020

agcttgctta ttgatatgat gaaaggctat gaggctgatg atatcatacg cctttattac 1080

gatttcattg tgcttaaatc tcagaaaaat ctcggttttt ctatcaaaaa gcttcgtgag 1140

aaaatgctgg acgaatacgg cttcagattt aaggacaagc aatatgactc tgtgcgctca 1200

aagatgtaca agcttatgga ttttctgctt ttctgcaact actacagaaa tgacattgcc 1260

gcaggcgaat ctcttgtgcg caaactgcgt ttttcaatga ccgatgatga aaaagagggg 1320

atatatgctg atgaagcggc aaagctttgg ggcaaattca ggaatgattt tgaaaatatc 1380

gccgaccaca tgaacggtga cgttatcaag gagcttggca aggctgacat ggattttgat 1440

gagaaaattc ttgacagcga aaagaagaat gcgtctgacc ttttgtattt ctccaaaatg 1500

atatatatgc tcacatattt tcttgacggc aaggagataa acgaccttct tacaacgctt 1560

atcagcaagt ttgataacat caaggagttt ttgaagataa tgaaaagctc tgctgttgat 1620

gttgagtgtg aacttacggc gggctacaag ctgttcaatg acagccagag gataaccaac 1680

gagcttttta tcgtaaagaa cattgcttcc atgagaaagc ctgcggcttc ggcgaagctt 1740

acgatgttcc gtgacgcact gactatactc ggtatagacg acaagatcac ggacgatagg 1800

ataagcgaga ttctaaaact taaagaaaaa ggcaagggca tacatggcct gagaaatttc 1860

ataacaaaca atgttatcga gtcctctcgg tttgtatacc ttatcaagta tgcgaacgct 1920

cagaagataa gagaagtggc taagaatgag aaagttgtca tgtttgttct tgggggtatc 1980

cctgacacgc agatagagcg ttattacaag agttgtgtgg aatttcctga catgaacagt 2040

tctttgggag taaagcgcag tgagcttgcg agaatgataa agaacatcag ctttgatgat 2100

ttcaaaaatg tgaaacagca ggcaaagggc agagaaaacg tggctaagga gagggcaaag 2160

gctgttatcg ggctttatct tacggtcatg tatctgctgg tgaaaaatct tgtgaatgtc 2220

aatgcaaggt atgttattgc gatacactgc cttgaacgtg attttgggct gtataaggag 2280

ataattcctg agttggcttc aaagaacttg aaaaatgact acaggatact ttcacagacg 2340

ctttgtgaac tttgtgataa gtcgccgaat ttgttcttga aaaagaacga gcggctgcgc 2400

aagtgcgttg aagttgatat caataatgca gacagcagca tgacaagaaa ataccgcaac 2460

tgtattgctc atcttactgt agttcgtgaa ctgaaagaat acataggaga tatttgtaca 2520

gtggattctt acttctccat ttatcattat gttatgcagc gctgtatcac gaaaagggaa 2580

aatgacacaa agcaagaaga gaaaataaag tatgaggacg atcttttaaa aaatcacggc 2640

tatacgaaag actttgtaaa ggctctcaac tcgccgtttg gatacaacat tccgaggttt 2700

aaaaatcttt caattgagca gttgtttgac agaaatgaat atcttactga aaagtag 2757

<210> 125
<211> 2865
<212> DNA
<213> Artificial Sequence

<220>
<223> Eubacterium siraeum nCas1 sequence

<400> 125
atgggtaaga aaatacacgc acgagatctc agagaacaaa gaaagaccga tagaacggaa 60

aaatttgcag atcagaacaa aaaacgtgaa gcagagaggg cagttccgaa aaaagacgca 120

gccgtttctg taaaatcagt ttcttctgtt tcatcaaaaa aagacaatgt aacaaaatct 180

atggctaaag ccgcaggcgt gaagtcggtt tttgctgtag gaaatactgt ttatatgact 240

tcattcggca gaggaaacga tgctgtactt gagcagaaaa tagtcgatac atcgcacgaa 300

ccgctgaata ttgacgatcc tgcatatcag ttgaacgttg tcacaatgaa cggttattcg 360

gttaccggtc acagaggtga aacggtatct gccgtaacgg ataatccgct gcgccgtttt 420

aacggaagaa agaaagatga accggaacag tctgtgccta cggatatgct gtgcctgaaa 480

ccgactcttg aaaagaaatt cttcggcaaa gaattcgatg ataatataca tatccagctt 540

atttacaata ttcttgacat tgaaaaaata ctggcggttt attcgaccaa cgctatttac 600

gcattgaata atatgagtgc tgacgaaaat atcgaaaaca gcgatttctt catgaaacgt 660

accaccgatg aaacctttga cgattttgaa aagaaaaagg agagtacaaa cagtcgagag 720

aaagccgatt ttgacgcatt tgaaaaattc atcggcaatt acaggctggc ttattttgcc 780

gatgcatttt atgtaaataa aaagaatccc aaaggtaaag caaaaaatgt tctgcgtgag 840

gataaagaac tttactccgt gctcactctg atcggtaaac tgcgtcattg gtgtgttcac 900

agtgaggagg gcagagcaga attctggctg tataagctcg atgaacttaa agatgatttc 960

aaaaatgtac tcgacgttgt ttataaccgt cctgttgaag aaataaacaa ccgctttata 1020

gaaaacaata aggtaaacat acagatactg ggctcggtat acaagaacac cgatattgcc 1080

gaacttgtaa ggtcatatta cgaatttctt atcacaaaga agtataaaaa tatgggcttt 1140

tcaataaaga agctccgtga gagtatgctc gaaggtaaag gttacgccga taaagaatat 1200

gattctgtaa ggaataagct gtatcagatg acggatttca tcttatacac aggatatatc 1260

aacgaagaca gcgatagagc cgacgatctt gtgaacactt tgagaagttc gctcaaagag 1320

gatgataaga caaccgtata ttgcaaggaa gcggattatc tgtggaaaaa ataccgtgaa 1380

tccataagag aggttgccga tgcgcttgat ggcgataaca ttaaaaagct gagcaaatcg 1440

aatattgaaa ttcaggaaga caagctgaga aaatgtttta tcagctatgc cgacagcgta 1500

tcggaattta ccaagcttat ttatctgctg acaagatttt taagcggtaa ggagatcaac 1560

gatcttgtca caacgctgat aaacaagttt gacaatatca gaagcttcct tgaaataatg 1620

gacgagcttg ggcttgacag gaccttcacc gccgagtaca gcttctttga aggcagtaca 1680

aagtatcttg ccgagcttgt cgagcttaac agctttgtga aatcgtgttc gtttgatata 1740

aacgcaaaaa gaacaatgta tcgcgatgcg ctggatattc tcggcattga atcggataag 1800

accgaagaag atattgagaa gatgatcgat aatatccttc agatcgacgc aaacggtgat 1860

aaaaagctca agaaaaacaa cggtctgaga aatttcattg caagtaacgt tatagattca 1920

aaccgattca agtaccttgt gcggtacgga aatccaaaga agattcgtga aacggcaaaa 1980

tgcaagcccg ctgtaaggtt tgtgctgaat gagatcccgg acgcacagat cgaaagatat 2040

tatgaggctt gttgcccaaa aaatacagct ttatgctctg caaataagag acgtgagaaa 2100

ctggctgata tgatagctga aataaagttt gagaattttt cggatgccgg caattatcag 2160

aaagcaaatg tcacatcaag aacgtctgaa gctgaaatca agcggaagaa tcaggctata 2220

atccgtcttt atcttaccgt tatgtacatt atgctgaaga accttgtaaa tgtgaacgcc 2280

agatacgtta tcgctttcca ttgcgttgaa agggatacga agctgtatgc ggaaagcggt 2340

ctggaagtcg gtaatataga aaaaaacaag acaaatctta ctatggctgt aatgggagtc 2400

aagctcgaaa acggaatcat aaaaacggaa tttgacaaga gctttgcaga aaatgccgca 2460

aacagatatc tcaggaatgc acgctggtac aagctgatac tggataattt aaagaagtcg 2520

gaaagagcgg ttgtcaatga gttcagaaat actgtctgcc atctgaatgc gataaggaat 2580

atcaatatca atatcaagga aataaaagag gtcgagaact actttgctct gtaccactac 2640

ctcattcaga aacatctcga aaatcgtttt gccgataaaa aagtagaaag agacaccggc 2700

gattttataa gcaagctcga agaacacaag acttactgca aggactttgt aaaagcatat 2760

tgtacgcctt tcggatataa ccttgtgaga tataaaaacc ttacgataga cgggctgttt 2820

gataagaatt accccggaaa agacgattct gatgaacaga aataa 2865

<210> 126
<211> 2799
<212> DNA
<213> Artificial Sequence

<220>
<223> Ruminococcus flavefaciens FD1 nCas1 sequence

<400> 126
atgaaaaaga aaatgtctct ccgtgaaaag cgtgaagccg agaaacaggc taaaaaagct 60

gcatattcag cagcttcaaa aaatacagat tctaagcctg cggaaaagaa agcagaaact 120

ccaaagcctg cggagattat ttccgataat tccagaaata agaccgctgt aaaggcggct 180

ggtctgaaat caacaattat cagcggcgat aagctgtata tgacatcttt cggcaagggt 240

aacgctgctg ttattgagca gaaaatagat atcaatgatt attctttttc agctatgaaa 300

gatactccgt cgcttgaagt tgataaagca gaatcaaaag agatctcttt ttcaagtcac 360

catccttttg taaagaatga taagctgaca acatataacc ctttatacgg cggcaaggat 420

aaccccgaaa agcctgtcgg cagggatatg ctcggcttaa aagataagct tgaagaacgc 480

tatttcggat gtacattcaa tgataatctt cacatccaga ttatctataa catacttgac 540

atcgagaaga ttttagctgt tcattctgca aatatcacaa ctgcgcttga ccacatggtt 600

gatgaagacg atgaaaaata tcttaacagc gattatatcg gctacatgaa taccataaat 660

acatatgacg tgtttatgga tccttcaaag aattcttcat taagccctaa agatagaaag 720

aatattgaca acagccgtgc aaaatttgag aaactgcttt caactaagcg ccttggctat 780

tttggatttg actatgatgc aaacggtaag gacaagaaaa agaacgagga aataaaaaag 840

cgtttatatc atctcacagc ttttgcaggt cagctccgtc agtggagttt tcatagtgct 900

ggcaattatc cgagaacatg gctttacaag ctcgattcac tggataagga atatcttgat 960

actcttgacc attacttcga taaacgtttt aacgatataa acgatgattt cgtaactaag 1020

aatgctacca atctctatat tctgaaagaa gtatttcccg aagcaaactt caaggatatt 1080

gccgatcttt attacgattt catagttata aagtcgcaca aaaatatggg attctccata 1140

aaaaagctga gggagaagat gcttgaatgt gatggtgcag acaggataaa agaacaggat 1200

atggactctg ttcgctcaaa gctgtataag ctcatagact tttgcatttt caagtattat 1260

cacgaatttc ctgaacttag tgaaaagaat gtggatatac tcagagcggc tgtatccgat 1320

acaaaaaaag ataaccttta ttctgatgag gctgcacgtt tatggagcat atttaaagaa 1380

aaattcctcg gcttctgtga taagatagtt gtatgggtaa caggagagca tgagaaagat 1440

atcacatccg ttattgataa ggatgcttac aggaacagga gcaatgtttc atatttctca 1500

aagctgatgt atgcaatgtg ctttttcctt gacggaaaag agataaatga ccttctcact 1560

actcttatca acaaattcga taatatcgct aaccagataa aaacagccaa agaacttggc 1620

attaatactg cttttgtaaa gaattacgat ttcttcaatc acagcgagaa atatgtcgat 1680

gaactgaaca tcgtcaagaa tattgcaaga atgaagaagc cttcaagtaa tgccaaaaaa 1740

gctatgtatc atgatgcgct tactattctc ggaatacctg aggatatgga tgaaaaagct 1800

cttgatgagg aactggattt aattcttgaa aaaaagacag acccagtaac tggcaagcca 1860

ctgaaaggta agaatccttt acgtaatttt atcgcaaaca atgtgataga gaattcaaga 1920

ttcatatatc ttatcaagtt ctgcaatcct gagaatgtac gtaaaatcgt gaataataca 1980

aaggtcactg agtttgtgtt aaagcgtatt cccgatgctc agatcgaacg ctattataag 2040

tcgtgtacag attctgaaat gaatccgcct actgaaaaga agatcaccga acttgctggt 2100

aagttaaagg atatgaactt tggcaacttc cgaaatgtga gacagtctgc taaagagaat 2160

atggagaagg agcgcttcaa agctgttata gggctttatc tcacggtagt atatcgtgtt 2220

gtcaagaatc ttgttgatgt aaactcacga tatatcatgg cttttcattc gcttgaacgt 2280

gattcacaac tgtataacgt atctgttgat aatgattatc ttgcacttac cgatactctt 2340

gttaaggagg gagataattc cagaagcaga tatcttgcag gcaacaagcg tctgagagat 2400

tgtgtgaagc aggatatcga taatgcaaaa aagtggtttg ttagtgataa gtacaatagc 2460

ataaccaagt acaggaataa cgttgcccat cttaccgctg tacgtaactg cgctgaattc 2520

atcggagata taacgaagat agactcctat tttgcattgt atcattatct cattcagaga 2580

cagcttgcga aaggtcttga ccatgagcga agtggctttg acagaaacta tccacagtat 2640

gcaccgctgt ttaagtggca tacgtatgta aaggatgttg tcaaggctct gaatgctcca 2700

tttggctaca atatccctcg tttcaagaat ctcagcatag atgcactttt tgaccgcaac 2760

gaaataaaga agaatgacgg cgagaaaaaa tccgatgat 2799

<210> 127
<211> 2835
<212> DNA
<213> Artificial Sequence

<220>
<223> Ruminococcus albus nCas1 sequence

<400> 127
atggcaaaaa aatcgaaagg tatgagcctt agagaaaaac gtgaacttga aaagcagaaa 60

aggatacaaa aggcagctgt gaattcagtt aatgatacac ctgaaaaaac agaagaagca 120

aatgtcgtat ctgtaaatgt caggacatcg gctgagaata agcatagtaa aaaatctgct 180

gccaaagctt tgggactgaa atccgggctg gttatcggtg atgagctgta ccttacttca 240

ttcggcagag gtaacgaagc aaagcttgaa aagaagatat ccggtgacac tgtcgaaaaa 300

cttggcattg gtgcttttga agtcgccgaa cgtgacgaat caacgcttac cctcgaaagt 360

ggcaggataa aggacaagac cgccagaccc aaagacccca gacatataac cgtcgataca 420

caaggtaaat tcaaggaaga tatgcttggg atacgcagtg tactggagaa aaagatattt 480

ggcaaaacat ttgatgataa tatccatgtt cagcttgcgt acaatatcct ggatgtcgaa 540

aagataatgg cacagtatgt cagcgatatc gtatatatgc tgcataatac tgataaaaca 600

gaaagaaacg ataatcttat ggggtatatg agcatcagga atacctataa gacattttgt 660

gatacgtcaa atcttcccga tgatacaaaa caaaaagttg aaaatcagaa gagagagttt 720

gacaagatca taaaaagcgg cagacttggg tatttcggcg aagcttttat ggtaaacagc 780

ggcaatagta ccaagcttag acccgagaaa gagatatatc atatctttgc gcttatggcg 840

agcctgaggc agagttactt tcacggatat gtaaaagata ccgattatca gggaaccaca 900

tgggcatata ctcttgagga caagctgaaa ggtccgagcc atgagttcag ggaaaccatt 960

gataagatat ttgatgaggg attcagcaag atcagcaagg actttggcaa gatgaacaag 1020

gtcaaccttc agatacttga acagatgatt ggtgaactgt atggcagtat agaacgacaa 1080

aacctcactt gcgattacta tgacttcatt caactgaaaa agcataagta tcttggattt 1140

tctataaagc gtcttagaga gaccatgctt gaaacaacac cggctgaatg ttataaagct 1200

gaatgctata acagcgagcg tcaaaagctg tataagctga tagatttcct gatatatgat 1260

ctttactata accgtaagcc tgcacgcatc gaagaaatcg tggacaagct gagggaatct 1320

gtgaacgacg aagagaaaga atccatatat tcagttgagg cgaagtatgt ctatgaatca 1380

cttagcaaag ttctggataa atcgctgaaa aacagtgtgt ctggtgaaac gataaaggat 1440

ctccaaaaga gatatgatga cgaaacagca aacaggatct gggatatctc acagcacagt 1500

ataagtggaa atgtcaactg tttctgcaag ctaatttata ttatgaccct gatgcttgac 1560

ggcaaggaga taaatgatct gctgacaacg ctggtaaaca agttcgataa catagcatca 1620

tttatagatg ttatggacga acttggcttg gagcatagtt ttacagataa ctataaaatg 1680

tttgccgaca gcaaggctat atgccttgat ctgcagttca taaacagttt tgcacgtatg 1740

tcaaagatcg atgatgagaa gtcaaaaaga cagcttttcc gtgatgcgct tgtcatactg 1800

gatatcggta ataaagatga gacttggata aataattatc tggattctga tattttcaaa 1860

ctggacaaag aaggtaacaa gttaaagggc gcaaggcatg atttcaggaa ctttatagcc 1920

aataatgtta taaagtcatc acgtttcaaa tacctagtaa aatacagcag tgccgatggt 1980

atgataaagc tgaaaacgaa tgaaaagctg ataggctttg ttctggataa gcttccagaa 2040

acgcagatag accgctacta tgaatcatgc ggacttgaca atgcggtagt agataagaaa 2100

gtcaggatag aaaagctatc ggggcttatc agagatatga agttcgatga tttcagcggt 2160

gtcaaaacct caaacaaagc aggagataat gacaaacagg ataaggcgaa atatcaggcg 2220

ataataagcc tgtacctcat ggtgctgtat cagatagtca agaacatgat atatgtcaac 2280

tcacgttatg ttatcgcttt ccattgtctt gaacgtgact ttggtatgta tggaaaagat 2340

tttggaaagt attatcaagg ctgccgaaaa cttacagatc attttattga agaaaagtac 2400

atgaaagagg gtaaacttgg ctgcaataaa aaagtcggca gatatctgaa aaataatatt 2460

tcctgctgca ctgatggact gataaatacc taccgtaatc aggttgatca ctttgcagtg 2520

gtaaggaaga taggcaacta tgcggcatat atcaagagta tcggttcgtg gtttgaactt 2580

tatcactatg taatacagag gatagttttt gacgaataca gatttgcact taacaacact 2640

gaaagcaact ataagaacag catcatcaag caccatacct actgtaagga tatggtcaag 2700

gcactgaaca cacctttcgg ttatgacctg ccgagataca agaatctttc tatcggtgat 2760

ctgtttgatc gcaataatta tctgaataaa acaaaagagt caatagatgc aaatagctct 2820

attgacagtc agtga 2835

<210> 128
<211> 2904
<212> DNA
<213> Artificial Sequence

<220>
<223> Ruminococcus flavefaciens XPD nCas1 sequence

<400> 128
atgatcgaaa agaagaagtc atttgcaaag ggcatgggag taaaatcaac acttgtatcc 60

ggttcaaagg tatacatgac gacgttcgca gaaggaagcg atgccagact tgaaaagatc 120

gttgaaggcg attctatcag atctgtcaac gaaggagaag cgttctcagc tgaaatggct 180

gataagaatg caggctacaa gatcggtaac gcaaagttca gccacccaaa gggctatgct 240

gtagttgcaa acaacccctt atacaccgga ccggtacagc aggatatgct cggtctgaag 300

gaaacgcttg aaaagagata ttttggagag tctgccgacg gaaatgataa tatctgtatt 360

caggtcatcc ataatatcct cgatatcgaa aagatcctcg ctgaatatat aaccaatgct 420

gcttatgcgg taaacaatat ttccggtctt gataaggata tcatcggttt tggtaagttc 480

agtacggtct atacttatga tgagttcaag gatcctgaac atcacagagc agctttcaac 540

aataacgata agttaattaa tgccatcaag gcacagtatg atgaatttga caatttcctt 600

gataatcctc gtctcggcta ctttggacag gctttttca gtaaggaagg cagaaattac 660

attatcaatt acggcaacga gtgttatgat attcttgctt tactcagcgg attgcgtcac 720

tgggtagtac ataataatga ggaagaatca aggatttccc gtacatggct ttataatctc 780

gacaagaatc ttgacaacga atatatctct actctcaatt atctgtatga tagaattaca 840

aacgaattaa caaattcctt ctcaaagaat agtgcagcca acgtaaacta tatcgctgaa 900

acccttggta ttaatcctgc tgaatttgca gagcagtatt tcagattcag tatcatgaag 960

gaacagaaga atctcggttt caatattact aagctgagag aagtaatgct tgacagaaag 1020

gatatgtctg agatccgtaa aaatcataag gtctttgatt caatccgtac taaggtctat 1080

actatgatgg atttcgttat ctacagatat tacattgaag aggatgcaaa ggttgctgct 1140

gccaacaagt ctctgccgga taacgaaaaa agcctcagtg aaaaggatat ctttgttata 1200

aatctcagag gaagctttaa cgatgatcag aaggatgccc tttattatga tgaggccaat 1260

cgtatttgga gaaagctcga aaacattatg cacaatatca aggaattcag aggcaataag 1320

acacgtgaat acaagaagaa ggatgctcca agactcccca gaattcttcc tgccggaagg 1380

gatgtttccg cgttctcaaa gttgatgtac gctcttacca tgttccttga tggtaaggag 1440

atcaatgatc ttctcaccac gctcatcaat aagttcgata acatccagag tttcctcaag 1500

gtaatgcctc ttatcggagt gaatgcaaag tttgttgagg aatatgcctt cttcaaggac 1560

agcgcaaaga ttgctgacga actcaggctg attaagagct ttgccagaat gggagaacct 1620

atcgcagatg caagacgtgc tatgtatatc gatgctatca ggattctcgg aacaaacctc 1680

agctatgatg agcttaaggc ccttgccgat acttttcgc ttgatgaaaa cggcaacaag 1740

cttaagaagg gcaagcacgg catgagaaac ttcatcatta ataatgtaat cagtaacaag 1800

cgcttccatt atctcattcg ttacggtgat cctgcacatc tccatgagat cgccaagaat 1860

gaagctgttg taaagttcgt cctcggcagg atagctgata tccagaagaa gcagggacag 1920

aacggaaaga atcagatcga caggtactat gagacctgta tcggcaagga caagggcaag 1980

tctgtctccg aaaaggttga tgccctcaca aagattatca ccggtatgaa ctacgatcag 2040

ttcgataaga agagaagcgt tattgaggat actggaagag aaaacgctga gagagaaaag 2100

ttcaagaaga tcatcagcct ctatcttact gtcatttatc acatccttaa gaatattgtt 2160

aatatcaatg cgcgttacgt tatcggcttc cattgcgttg agcgtgatgc acagctctat 2220

aaggaaaagg gctatgatat caacctcaag aagctcgaag aaaaggggtt ttcatcagtc 2280

acaaagctgt gtgcaggtat tgatgagact gctcctgaca agcgtaagga tgttgaaaag 2340

gaaatggctg agcgtgcaaa ggaatctatc gatagccttg aatctgcaaa tcctaagctt 2400

tacgcaaact atatcaagta ttctgacgag aagaaggctg aggaatttac tagacagatc 2460

aaccgtgaga aggcaaagac cgctctgaat gcatatctca gaaatactaa gtggaatgtg 2520

ataatcaggg aagatcttct tagaatcgat aataagacat gtacgctctt tagaaataag 2580

gccgttcatc ttgaagttgc aagatatgtt catgcatata tcaacgatat tgccgaagta 2640

aacagctatt tccagcttta tcattacatc atgcagagaa tcatcatgaa cgaaagatat 2700

gaaaagtctt ctggaaaggt aagcgaatac ttcgatgctg tgaacgatga aaagaagtac 2760

aacgacaggc ttctgaagct gttgtgcgtt ccatttggtt actgcatccc gagattcaag 2820

aatctctcca ttgaagcttt gttcgacagg aacgaagcag ctaagtttga caaggaaaag 2880

aagaaagtat caggtaattc atag 2904

<210> 129
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Consensus DR

<400> 129
gaactacacc cgtgcaaaaa tgcaggggtc taaaac 36

<210> 130
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Consensus DR

<400> 130
ctactacact ggtgcaaatt tgcactagtc taaaac 36

<210> 131
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Consensus DR

<400> 131
caaguaaacc ccuaccaacu ggucgggguu ugaaac 36

<210> 132
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Consensus DR

<400> 132
cuacuacacu ggugcgaauu ugcacuaguc uaaaac 36

<210> 133
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Consensus DR

<400> 133
gaacuauagu agugugaauu uacacuacuc uaaaac 36

<210> 134
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Consensus DR

<400> 134
cuacuauacu agugugauuu uacacuaguc uaaaac 36

<210> 135
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Consensus DR

<400> 135
gaccaacacc ucugcaaaac ugcagggguc uaaaac 36

<210> 136
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Consensus DR

<400> 136
gcacuacacc ccccugaaac augagggguc uaaaac 36

<210> 137
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Consensus DR

<400> 137
gcactacacc cccctgaaac atgaggggtc taaaac 36

<210> 138
<211> 966
<212> PRT
<213> Artificial Sequence

<220>
<223> 50% consensus sequence for 7 full-length nCas1 orthologues

<220>
<221> misc_feature
<222> (2)..(4)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (13)..(13)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (22)..(23)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (25)..(25)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (32)..(33)

<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (39)..(40)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (43)..(44)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (46)..(56)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (60)..(60)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (69)..(70)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (72)..(72)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (74)..(76)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (83)..(83)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (86)..(86)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (88)..(88)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (94)..(95)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (97)..(108)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (110)..(135)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (139)..(147)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (149)..(150)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (157)..(158)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (162)..(162)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (167)..(167)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (191)..(192)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (201)..(209)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature

<222> (211)..(211)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (213)..(214)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (216)..(217)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (220)..(221)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (225)..(225)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (227)..(228)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (230)..(242)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (244)..(249)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (251)..(251)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (253)..(257)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (264)..(264)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (267)..(270)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (273)..(280)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (283)..(283)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (285)..(285)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (290)..(292)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (296)..(298)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (300)..(310)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (314)..(314)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (316)..(318)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (320)..(321)
<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc_feature
<222> (323)..(325)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (329)..(331)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (333)..(337)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (340)..(341)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (343)..(345)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (350)..(350)
<223> Xaa can be Ile or Leu

<220>
<221> misc_feature
<222> (351)..(352)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (354)..(357)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (359)..(364)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (365)..(365)
<223> Xaa can be Ile or Leu

<220>
<221> misc_feature
<222> (366)..(368)

<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (373)..(375)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (377)..(378)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (391)..(391)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (394)..(397)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (399)..(400)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (401)..(401)
<223> Xaa can be Ile or Leu

<220>
<221> misc_feature
<222> (402)..(404)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (411)..(411)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (415)..(415)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (417)..(417)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (420)..(420)
<223> Xaa can be Ile or Leu

<220>
<221> misc_feature
<222> (423)..(424)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (426)..(427)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (429)..(433)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (435)..(436)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (438)..(439)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (442)..(442)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (444)..(445)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (447)..(447)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (450)..(452)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature

<222> (454)..(455)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (458)..(459)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (462)..(477)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (479)..(480)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (481)..(481)
<223> Xaa can be Ile or Leu

<220>
<221> misc_feature
<222> (483)..(485)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (487)..(487)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (489)..(489)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (491)..(506)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (509)..(510)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (517)..(517)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (520)..(520)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (542)..(542)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (546)..(546)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (549)..(549)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (554)..(556)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (558)..(560)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (562)..(563)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (565)..(566)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (568)..(569)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (571)..(572)
<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc_feature
<222> (575)..(577)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (580)..(580)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (584)..(584)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (587)..(589)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (593)..(593)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (600)..(600)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (605)..(612)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (614)..(616)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (618)..(623)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (624)..(624)
<223> Xaa can be Ile or Leu

<220>
<221> misc_feature
<222> (626)..(629)

<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (631)..(632)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (634)..(636)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (648)..(648)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (650)..(650)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (653)..(653)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (656)..(656)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (659)..(659)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (662)..(664)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (666)..(666)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (667)..(667)
<223> Xaa can be Ile or Leu

<220>
<221> misc_feature
<222> (669)..(669)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (673)..(675)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (679)..(680)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (684)..(684)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (694)..(696)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (698)..(706)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (708)..(709)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (711)..(713)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (715)..(718)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (720)..(720)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature

<222> (723)..(723)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (726)..(726)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (728)..(733)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (735)..(738)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (742)..(742)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (745)..(745)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (753)..(753)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (755)..(756)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (778)..(779)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (782)..(797)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (800)..(812)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (816)..(852)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (856)..(856)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (858)..(858)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (861)..(861)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (863)..(863)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (865)..(865)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (866)..(866)
<223> Xaa can be Ile or Leu

<220>
<221> misc_feature
<222> (867)..(868)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (871)..(871)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (874)..(875)
<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc_feature
<222> (877)..(877)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (886)..(887)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (890)..(890)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (892)..(897)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (899)..(912)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (914)..(914)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (917)..(917)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (920)..(920)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (926)..(926)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (932)..(932)
<223> Xaa can be Ile or Leu

<220>
<221> misc_feature
<222> (941)..(942)

<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (949)..(954)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (956)..(957)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (959)..(963)
<223> Xaa can be any naturally occurring amino acid

<400> 138

Met Xaa Xaa Xaa Lys Lys Lys Met Ser Leu Arg Glu Xaa Arg Glu Ala
1   5   10   15

Glu Lys Gln Ala Arg Xaa Xaa Lys Xaa Ala Ala Ala Ser Lys Asn Xaa
            20   25   30

Xaa Ala Val Pro Ala Lys Xaa Xaa Ala Ala Xaa Xaa Pro Xaa Xaa Xaa
        35   40   45

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Lys Lys Ser Xaa Ala Lys Ala Ala
    50   55   60

Gly Leu Lys Ser Xaa Xaa Val Xaa Gly Xaa Xaa Xaa Tyr Met Thr Ser
65   70   75   80

Phe Gly Xaa Gly Asn Xaa Ala Xaa Leu Glu Lys Lys Ile Xaa Xaa Asx
                85   90   95

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Glu Xaa Xaa Xaa
            100   105   110

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
        115   120   125

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Asn Pro Leu Xaa Xaa Xaa Xaa Xaa Xaa

130 135 140

Xaa Xaa Xaa Val Xaa Xaa Asp Met Leu Gly Leu Lys Xaa Xaa Leu Glu
145 150 155 160

Lys Xaa Tyr Phe Gly Lys Xaa Phe Asp Asp Asn Ile His Ile Gln Leu
165 170 175

Ile Tyr Asn Ile Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Xaa Xaa
180 185 190

Asn Ile Val Tyr Ala Leu Asn Asn Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
195 200 205

Xaa Asn Xaa Asp Xaa Xaa Gly Xaa Xaa Ser Thr Xaa Xaa Thr Tyr Asp
210 215 220

Xaa Phe Xaa Xaa Pro Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
225 230 235 240

Xaa Xaa Ile Xaa Xaa Xaa Xaa Xaa Xaa Phe Xaa Lys Xaa Xaa Xaa Xaa
245 250 255

Xaa Arg Leu Gly Tyr Phe Gly Xaa Ala Phe Xaa Xaa Xaa Xaa Lys Asx
260 265 270

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Lys Glu Xaa Tyr Xaa Ile Leu Ala
275 280 285

Leu Xaa Xaa Xaa Leu Arg Gln Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa
290 295 300

Xaa Xaa Xaa Xaa Xaa Xaa Trp Leu Tyr Xaa Leu Xaa Xaa Xaa Leu Xaa
305 310 315 320

Xaa Glu Xaa Xaa Xaa Thr Leu Asp Xaa Xaa Xaa Asp Xaa Xaa Xaa Xaa
325 330 335

Xaa Ile Asn Xaa Xaa Phe Xaa Xaa Xaa Asn Lys Val Asn Xaa Xaa Xaa

340 345 350

Leu Xaa Xaa Xaa Xaa Gly Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
355 360 365

Tyr Tyr Asp Phe Xaa Xaa Xaa Lys Xaa Xaa Lys Asn Leu Gly Phe Ser
370 375 380

Ile Lys Lys Leu Arg Glu Xaa Met Leu Xaa Xaa Xaa Xaa Ala Xaa Xaa
385 390 395 400

Xaa Xaa Xaa Xaa Glx Tyr Asp Ser Val Arg Xaa Lys Leu Tyr Xaa Leu
405 410 415

Xaa Asp Phe Xaa Ile Tyr Xaa Xaa Tyr Xaa Xaa Glu Xaa Xaa Xaa Xaa
420 425 430

Xaa Glu Xaa Xaa Val Xaa Xaa Leu Arg Xaa Ser Xaa Xaa Asp Xaa Glu
435 440 445

Lys Xaa Xaa Xaa Tyr Xaa Xaa Glu Ala Xaa Xaa Leu Trp Xaa Xaa Xaa
450 455 460

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Gly Xaa Xaa
465 470 475 480

Xaa Lys Xaa Xaa Xaa Lys Xaa Asx Xaa Asp Xaa Xaa Xaa Xaa Xaa Xaa
485 490 495

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Asx Val Xaa Xaa Phe Ser
500 505 510

Lys Leu Ile Tyr Xaa Leu Thr Xaa Phe Leu Asp Gly Lys Glu Ile Asn
515 520 525

Asp Leu Leu Thr Thr Leu Ile Asn Lys Phe Asp Asn Ile Xaa Ser Phe
530 535 540

Leu Xaa Val Met Xaa Glu Leu Gly Leu Xaa Xaa Xaa Phe Xaa Xaa Xaa

545 550 555 560

Tyr Xaa Xaa Phe Xaa Xaa Ser Xaa Xaa Ile Xaa Xaa Glu Leu Xaa Xaa
565 570 575

Xaa Lys Asn Xaa Ala Arg Met Xaa Lys Pro Xaa Xaa Xaa Ala Lys Arg
580 585 590

Xaa Met Tyr Arg Asp Ala Leu Xaa Ile Leu Gly Ile Xaa Xaa Xaa Xaa
595 600 605

Xaa Xaa Xaa Xaa Ile Xaa Xaa Xaa Leu Xaa Xaa Xaa Xaa Xaa Xaa Xaa
610 615 620

Asp Xaa Xaa Xaa Xaa Lys Xaa Xaa Lys Xaa Xaa Xaa Gly Leu Arg Asn
625 630 635 640

Phe Ile Ala Asn Asn Val Ile Xaa Ser Xaa Arg Phe Xaa Tyr Leu Xaa
645 650 655

Arg Tyr Xaa Asn Pro Xaa Xaa Xaa Arg Xaa Xaa Ala Xaa Asn Glu Lys
660 665 670

Xaa Xaa Xaa Phe Val Leu Xaa Xaa Ile Pro Asp Xaa Gln Ile Glu Arg
675 680 685

Tyr Tyr Glu Ser Cys Xaa Xaa Xaa Asx Xaa Xaa Xaa Xaa Xaa Xaa Xaa
690 695 700

Xaa Xaa Ile Xaa Xaa Leu Xaa Xaa Xaa Ile Xaa Xaa Xaa Xaa Phe Xaa
705 710 715 720

Asx Phe Xaa Asx Val Xaa Gln Xaa Xaa Xaa Xaa Xaa Xaa Asn Xaa Xaa
725 730 735

Xaa Xaa Lys Glu Lys Xaa Lys Ala Xaa Ile Gly Leu Tyr Leu Thr Val
740 745 750

Xaa Tyr Xaa Xaa Val Lys Asn Leu Val Asn Val Asn Ala Arg Tyr Val

```
        755                 760                 765

Ile Ala Phe His Cys Leu Glu Arg Asp Xaa Xaa Leu Tyr Xaa Xaa Xaa
    770                 775                 780

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Leu Thr Xaa
785                 790                 795                 800

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Tyr Leu Lys Xaa
                805                 810                 815

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
            820                 825                 830

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
        835                 840                 845

Xaa Xaa Xaa Xaa Tyr Arg Asn Xaa Val Xaa His Leu Xaa Ala Xaa Arg
    850                 855                 860

Xaa Xaa Xaa Xaa Tyr Ile Xaa Asp Ile Xaa Xaa Val Xaa Ser Tyr Phe
865                 870                 875                 880

Ala Leu Tyr His Tyr Xaa Xaa Gln Arg Xaa Leu Xaa Xaa Xaa Xaa Xaa
                885                 890                 895

Xaa Lys Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
            900                 905                 910

His Xaa Thr Tyr Xaa Lys Asp Xaa Val Lys Ala Leu Asn Xaa Pro Phe
        915                 920                 925

Gly Tyr Asn Xaa Pro Arg Tyr Lys Asn Leu Ser Ile Xaa Xaa Leu Phe
    930                 935                 940

Asp Arg Asn Glu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Glu Xaa Xaa Lys Xaa Xaa
945                 950                 955                 960

Xaa Xaa Xaa Asp Ser Gln
```

965

<210> 139
<211> 2718
<212> DNA
<213> Artificial Sequence

<220>
<223> Gut metagenome P1E0 nCas1 sequence

<400> 139
atggaacgtg aggtcaaaaa gccccccaaa aaatccttgg caaaggcggc gggactgaag 60

tccaccttcg tcatttcgcc gcaggagaag gagctggcca tgaccgcctt cgggcggggc 120

aacgacgcgc tgctgcaaaa gcgcatcgtg gacggcgtgg tgcgggatgt ggccggagag 180

aagcagcagt ttcaggtgca gcggcaggac gagagccgtt tccgactgca aaacagccgc 240

cttgccgacc gcacggtaac ggcggacgac ccgctgcacc gcgcggagac gccccgccgc 300

cagccgctgg gcgcaggtat ggatcagctg cgccgcaagg caatactgga gcagaaatac 360

ttcggcagga cgtttgacga caacatccac atccaactga tctacaacat attggacatc 420

cacaagatgc tggctgtccc cgccaaccac atcgtccaca cgctgaatct gctgggcgga 480

tacggcgaga cggactttgt ggggatgctg cccgccgggc tgccgtatga caagctgagg 540

gtcgtgaaaa aaaagaacgg cgacaccgtg gacatcaagg cggatatcgc cgcctatgcc 600

aagcggccgc agctggccta tctgggtgcg gcgttctatg acgtgacccc cggcaagagc 660

aaaagggacg cggcgcgggg cagagtcaag cgcgagcagg acgtgtatgc cattctctcg 720

ctgatgagcc ttctgcggca gttctgcgcc cacgacagtg tgcggatatg gggacagaac 780

accaccgccg ccctgtatca tttgcaggcg ctgccgcagg acatgaagga tctgctggat 840

gacggctggc gcagggcgct gggcggcgtc aacgaccatt tcctcgacac caacaaggtg 900

aacctgctga ccctgttcga gtactacggc gccgagacga agcaggcgcg ggtggcgctg 960

acgcaggact tctaccgctt cgtggtgctg aaggagcaga agaacatggg cttctccctg 1020

aggcgtctgc gggaggagct gctgaagctg ccggacgccg cctacctcac cggtcaggag 1080

tacgactcgg tgcggcagaa gctgtatatg ctgctggact tcctgctgtg ccgcctctac 1140

gcgcaggagc gggccgaccg ctgcgaggag ctggtgtccg cgctgcgctg cgccttgtcg 1200

gatgaggaaa aggacaccgt ctatcaggcg gaggccgccg cgctgtggca ggcgctgggc 1260

gacacgctgc gccggaagct gctgccgctg ctgaaaggca aaaagctgca agacaaggac 1320

aagaagaagt cggatgagct ggggctgtcc cgtgacgtgc tggatggtgt gctgttccgc 1380

cccgcacagc agggcagccg cgccaacgcc gactatttct gccggctgat gcacctgtcc 1440

acatggttca tggacggcaa ggagatcaac accctgctga ccaccctcat cagcaagctt 1500

gagaatatcg acagcctgcg gagcgtgctg gagagcatgg gactggcgta cagctttgta 1560

ccggcctacg ccatgtttga ccacagccgc tatatcgccg gtcagctgcg ggtggtcaac 1620

aacatcgcgc ggatgcgtaa gcctgccatc ggcgccaagc gggagatgta ccgcgccgcc 1680

gtggtgctgc tgggcgtgga cagcccggag gcagccgcgg ccatcaccga cgacctgctg 1740

cagatcgacc cggaaacggg gaaggtgcgc ccccgcagcg acagcgcacg ggacacgggt 1800

ctgcggaact tcatcgccaa caacgtggtg gaatcccgcc gcttcaccta cctgctgcgc 1860

tacatgacgc cggagcaggc gcgcgtgctg gcgcagaacg aaaagctcat cgccttcgtg 1920

ctgtccaccg tgcccgacac ccagctggag cgctactgcc gcacctgcgg gcgggaggac 1980

atcaccggcc gtccggcgca gatacgttac ctgacggcgc agatcatggg cgtccggtat 2040

gagagcttca ccgatgtgga gcagcgcggg agaggggata accccaaaaa ggagcggtac 2100

aaggctctca tcggcctgta cctgacggtg ctgtatctgg cggtgaagaa tatggtcaac 2160

tgcaacgccc gctacgtcat cgccttctac tgccgtgacc gtgacacagc gctctatcag 2220

aaggaggtct gctggtacga tctggaggag gacaagaaaa gcggcaagca gcggcaggtg 2280

gaggactaca ccgccctgac ccgctatttt gtgtcgcagg gctatctgaa ccgccatgcc 2340

tgcggctatc tgcgaagcaa tatgaacggc atcagcaaca gcctgctgac cgcctaccgc 2400

aacgcggtgg atcacctgaa cgcgatcccg ccgctgggca gcctgtgccg ggacatcggc 2460

cgggtggaca gctactttgc cctgtaccac tacgccgtgc agcagtatct gaacggcaga 2520

tattacagaa agacgccccg cgagcaggag ctgtttgccg ccatggcgca gcataggaca 2580

tggtgcagcg atctggtgaa ggcactgaac acccccttg gctacaacct cgcccgctac 2640

aagaacctgt ccattgacgg tctgtttgac cgcgagggcg accatgtggt gcgggaggac 2700

ggcgaaaaac cggcggag 2718

<210> 140
<211> 2835
<212> DNA
<213> Artificial Sequence

<220>
<223> Anaerobic digester nCas1 sequence

<400> 140
aataacaaaa gaaaaaccaa agctaaggca gctggactca aatccgtttt tttcgaccaa 60

aagcaagctg ttttgacgac ttttgcaaaa ggcaataatt ctcaaatcga aaagaaggtc 120

gtaaatagcg aggttaagga tttgcgacaa ccacctgctt ttgatctcga acttaaagag 180

aagacttttt atatcagcgg taaaaacaac atcaatacct caagagaaaa cccattggct 240

tccgcgtcgc ttcctctaag caagcgacaa cgaattagag cagaaagaat taaacgggcg 300

agagaggaga atcgccccta ccataacgta aagagagtcg gcgaagatga cttacgggcg 360

aaagccgatt tagagaagca ctatttcgga aaagaatact ctgacaattt aaaaattcaa 420

ataatctaca acattctcga tataaataag attatttccc cctatattaa tgatattgtt 480

tattctatga ataatcttgc tcgtaacgac gagtatattg atggcaagat tgatgttatc 540

ggcagtttaa gctcgacgac cgattactct tcatttatga gtccaaataa ggatttagaa 600

aaagaaaaaa aattctcttt tcatcgcgaa aactacaaaa aatttgtcga agcctctaag 660

ccatatatgc gatattatgg aaaagttttt attcgcgatg tcaaaaagag caaactcagc 720

accggcaagg gcgagaaaat cgaagtcatg tatcgctctg acgaggaaat tttcacaatc 780

ttccaaatat taagttatgt tcgtcaaagc atcatgcata atgatatagg caataagtca 840

tcgattttag caatcgaaaa atatcctgcc cgctttgttg ggttcttatc ggacttgctt 900

aaaacaaaaa caaatgatgt caatagaatg ttcatcgata ataatagcca gactaatttc 960

tgggtacttt tttctatctt tggtcttcaa gatcatacga gcggcgcgga caaaatttgt 1020

cgtaattttt atgatttcgt cattaaagcc gatagcaaaa acttaggttt ttcacttaaa 1080

aaaatacggg aattaatgct cgatttgcca aatgccaata tgcttagaga tcatcaattt 1140

gacactgttc gcagcaagtt ttacacactg cttgatttca tcatttacca gcactattta 1200

gaagaaaagt cgcgaataga caatatggtt gagaagttaa gaatgacttt aaaagaggaa 1260

gaaaaggaag tgctttatgc tgccgaggca aaaatagtct ggaacgccat cggggcgaaa 1320

gtcataaaca agcttgttcc gatgatgaat ggcgacgcgc taaaagagat taaaaggaaa 1380

aacagagacc gcaaattgcc tcaaagcgtt atcgcaacgg tccaagttaa ttcagacgct 1440

aacgtattta gtggcctgat atacttcctt accctctttt tagacggcaa agaaattaat 1500

gaaatggtca gcaaccttat taccaaattc gaaaacattg atagtttact ccatgttgat 1560

cgcgaaattt ataaaagcga tgaaaaggat ctcgatttag aaattgaaaa actcgccctt 1620

ttctttaaag gtgtcgtgcg accgaacgct aaaacggata ccggcgccgg agaaatatct 1680

aaatcatttt cgatatttca atccgccgaa agaattatcg aggagctgaa gtttatcaaa 1740

aatgtgacgc ggatggataa tgaaattttc ccatcagaag gagtgttttt agatgcggca 1800

aatgtcttgg gagttagagg cgacgacttc gatttttcaa acgaatttgt cggtgatgat 1860

ttgcattctg acgccaataa aaaaattatc aataagatca acggcaccaa agaagatcgc 1920

aatttgcgca acttcataat aaataatgtt gtcaagtcgc ggcgctttca atatatcgcc 1980

cgccacatga acacacacta tgttaaacaa ctcgccaaca acgaaactct caaccgcttt 2040

gtcttaaata aaatgggtga tgctaaaatc attaaccgct actacgaaag catcagtgga 2100

aacactccaa atatcgaggt tagaagccag atcgattatt tggttaaacg tcttcggagt 2160

ttcagttttg aagatctcaa cgacgtaaaa caaaaagtga ggcccggcac aaatgaaagc 2220

atagaaaagg aaaagaaaaa agcgctagtg ggtctttgct taaccattca atacttggta 2280

tataagaatc tcgtcaatat aaacgctcgt tatacgactg ctttctattg tcttgaacgg 2340

gacagcaaac ttaaaggctt tggagttgat gtttggcgag attttgaaag ctatacagcc 2400

ctgacaaatc actttattaa agagggctat cttcctgtcc gcaaggcaga aattttacga 2460

gcgaatttaa aacatcttga ttgcgaagat ggatttaaat attatcgcaa ccaagtcaca 2520

caccttaacg ccatccgcgt agcatacaaa tatattaacg aaattaaaag tgttcactca 2580

tatttcgcgc tctatcatta tattatgcaa cgccacctat atgattcttt gcaggcaaag 2640

gcaaaagatt ctagtggttt tgttattgac gccttaaaaa agagttttga gcacaagatt 2700

tactcgaagg atcttttgca cgttcttcat tcgcccttcg gttacaatac ggcccgatac 2760

aaaaacttaa gtatcgaggc attatttgac aaaaacgaaa gccggcccga ggtgaatcct 2820

ttatcgacca acgat 2835

<210> 141
<211> 2769
<212> DNA
<213> Artificial Sequence

<220>
<223> Ruminococcus sp. CAG:57 nCas1 sequence

<400> 141
atggcaaaaa agaataaaat gaagcctaga gagctgcgtg aggctcagaa aaaagccaga 60

cagctcaaag cggctgagat aaataataac gctgctcctg cgatcgctgc catgcctgct 120

gcagaggtca ttgcacctgt ggcagagaag aaaaaatcct ccgtaaaggc ggcaggaatg 180

aagtctattc ttgtcagcga aaataaaatg tacataacct ctttcggcaa gggcaattct 240

gctgtgcttg aatatgaggt ggacaataat gactacaaca aaactcagct ttcttcaaag 300

gacaacagca atatcgagct tggtgatgta aacgaggtaa acatcacttt ttcaagcaag 360

catggctttg ggagcggagt ggagataaat acttcaaacc ctactcacag aagcggtgaa 420

agctcgcctg taagagggga tatgctgggg cttaaatcgg agcttgaaaa gcgctttttc 480

ggcaaaactt ttgatgataa tatacatatc cagcttattt acaacattct ggatatcgaa 540

aagatacttg cggtgtatgt aacgaatatc gtttatgcgc tgaacaatat gcttggtata 600

aaggattctg aaagttatga tgattttatg gggtatcttt ctgcaagaaa tacttatgaa 660

gtttttactc accctgacaa aagcaatctt tccgataagg taaagggtaa tatcaagaaa 720

agccttagca agtttaatga cttgctgaaa actaagcgcc ttggctattt cggccttgaa 780

gagccaaaga caaaagacac aagagcttcg gaagcataca aaaagcgtgt ttatcatatg 840

cttgcaattg tggggcagat aagacagtgt gtttttcatg ataaatcggg tgcaaaaaga 900

tttgaccttt acagttttat taacaatatt gatcccgaat acagagatac tcttgactat 960

cttgttgagg agcgtttaaa gtccataaac aaggacttta tcgagggtaa caaggtcaat 1020

atcagcctgc ttattgatat gatgaaaggc tatgaggctg atgatatcat acgcctttat 1080

tacgatttca ttgtgcttaa atctcagaaa aatctcggct tttctatcaa aaagcttcgt 1140

gagaaaatgc tggaggaata cggtttcaga tttaaggaca agcaatatga ctctgtgcgc 1200

tcaaagatgt acaagcttat ggatttcctg cttttctgca actactacag aaatgacgtt 1260

gccgcaggcg aagctcttgt gcgtaaactg cgtttttcaa tgaccgatga tgaaaaagag 1320

gggatatatg ctgatgaagc ggcaaagctt tggggcaaat tcaggaatga ttttgaaaat 1380

atcgccgacc acatgaacgg tgacgttatc aaggagcttg gcaaggctga catggatttt 1440

gatgagaaaa ttcttgacag tgaaaagaag aatgcgtctg acctttttgta tttctccaaa 1500

atgatatata tgctcacata ttttcttgac ggcaaggaga taaacgatct tcttacaacg 1560

cttatcagca agtttgataa catcaaggag tttttgaaga taatgaaaag ctctgctgtt 1620

gatgttgagt gtgagcttac ggcgggctac aagctgttca atgacagcca gaggataacc 1680

aacgagcttt ttatcgtaaa gaacattgct tccatgagaa agcctgcggc ttcagcgaag 1740

cttacgatgt tccgtgacgc actgactata ctcggtatag acgacaatat cacggacgat 1800

aggataagcg agattctaaa acttaaagaa aaaggcaagg gcatacatgg tctgagaaat 1860

tttataacaa acaatgttat cgagtcctct cggtttgtat accttatcaa gtatgcgaac 1920

gctcagaaga taagagaagt ggctaaggat gagaaagttg tcatgtttgt tcttgggggt 1980

atccctgaca cgcagataga gcgttattac aagagttgtg tggagtttcc tgacatgaat 2040

agttctttgg aagcaaagcg cagtgagctt gcgagaatga taaagaacat cagctttgat 2100

gatttcaaaa atgtgaaaca gcaggcaaag ggcagagaaa acgtggctaa ggagagggca 2160

aaggctgtta tcgggcttta tcttacggtc atgtatctgc tggtgaaaaa tcttgtgaat 2220

gtcaatgcaa ggtatgttat tgcgatacac tgccttgaac gtgattttgg gctgtataag 2280

gagataattc ctgagttggc ttcaaagaac ttgaaaaatg actacaggat actttcacag 2340

acgctttgtg aactttgtga tgatcgtaat gagtcgtcga atttgttctt gaaaaagaac 2400

aagcggctgc gcaagtgcgt tgaagttgat atcaataatg cagacagcag catgacaaga 2460

aaataccgca actgtattgc tcatcttact gtagttcgtg aactgaaaga atacatagga 2520

gatattcgta cagtggattc ttacttctcc atttatcatt atgttatgca gcgttgtatc 2580

acgaaaaggg gagatgacac aaagcaagaa gagaaaataa agtatgagga cgatctttta 2640

aaaaatcacg gctatacgaa agactttgta aaggctctca actcgccgtt tggatacaac 2700

attccgaggt ttaaaaatct ttcaattgag cagttgtttg acagaaatga atatcttact 2760

gaaaagtag 2769

<210> 142
<211> 2715
<212> DNA
<213> Artificial Sequence

<220>
<223> Human codon-optimized Gut metagenome P1E0 nCas1

<400> 142
gaaagggagg tgaagaagcc tcccaagaaa tccctggcca aagccgccgg actgaagtcc 60

acattcgtca tcagccccca ggaaaaggaa ctcgctatga ccgcctttgg cagaggcaac 120

gatgccctgc tgcaaaagag aatcgtggac ggcgtggtcc gtgatgttgc cggagagaaa 180

cagcagttcc aggtccagag gcaagacgag agcagattca gactccaaaa ctccagactg 240

gccgatagga ccgtcaccgc tgacgacccc ctccacaggg ctgaaacacc caggagacag 300

cctctgggcg ctggcatgga tcagctgagg agaaaagcca ttctggagca aaagtacttt 360

ggcaggacat tcgacgacaa catccacatc cagctgatct acaacatcct ggacatccac 420

aagatgctgg ctgtgcctgc taaccatatc gtgcacaccc tgaacctcct gggaggctac 480

ggcgaaactg attttgtggg aatgctgcct gccggactcc cctacgataa gctcagggtc 540

gtgaagaaaa aaaatggcga taccgtcgac atcaaggccg acattgctgc ctatgccaaa 600

agaccccagc tggcttacct gggagctgcc ttctatgacg tgacccctgg aaaaagcaag 660

cgtgatgctg ccaggggcag ggtgaagagg gaacaggacg tctatgccat tctgtccctc 720

atgtccctgc tgagacaatt ctgcgcccac gatagcgtca ggatctgggg ccagaatacc 780

accgccgccc tgtaccacct ccaggccctg ccccaggata tgaaggacct gctggacgac 840

ggctggagga gagctctggg aggcgtgaat gaccatttcc tcgacaccaa caaggtgaac 900

ctcctcaccc tcttcgagta ctatggcgcc gaaaccaagc aggctagggt ggccctgaca 960

caggactttt acaggttcgt ggtgctgaag gagcagaaga acatgggctt tagcctcagg 1020

aggctgaggg aagagctgct gaagctgccc gacgccgctt acctgaccgg acaggagtac 1080

gactccgtga ggcagaaact gtacatgctc ctggacttcc tgctgtgcag gctgtacgcc 1140

caggagagag ccgatagatg cgaagagctg gtgagcgccc tcagatgtgc tctgagcgac 1200

gaggagaagg atacagtcta ccaagctgaa gccgctgccc tgtggcaagc cctgggcgac 1260

acactcagga ggaaactcct ccccctcctg aagggcaaga aactgcaaga caaggataag 1320

aaaaagtccg acgaactcgg cctgtccagg gatgtgctcg atggcgtgct ctttagacct 1380

gcccagcagg gcagcagagc caacgccgac tacttttgta ggctcatgca cctgagcacc 1440

tggtttatgg acggcaagga gatcaatacc ctgctgacca ccctgatctc caagctggag 1500

aatatcgata gcctgaggtc cgtcctggag tccatgggcc tggcctatag ctttgtccct 1560

gcctacgcca tgtttgacca cagcagatac atcgctggcc agctgagggt ggtgaacaac 1620

atcgccagaa tgaggaaacc cgccattgga gccaaaagag agatgtacag ggccgctgtc 1680

gtgctgctcg gagtggacag ccccgaagct gctgccgcca ttacagacga tctgctgcag 1740

atcgatcccg aaactggaaa ggttcgtcct agatccgact ccgctagaga caccggactg 1800

aggaacttca tcgctaacaa cgtcgtggag agcaggaggt tcacctacct cctgagatac 1860

atgacccctg aacaggccag ggtcctggct caaaacgaga agctgattgc cttcgtgctc 1920

agcacagtgc ccgacaccca gctggagagg tattgtagga cctgtggaag ggaggacatc 1980

acaggcaggc ccgcccagat taggtatctg accgctcaga tcatgggcgt gaggtacgaa 2040

agctttaccg acgtggaaca gaggggaagg ggcgataatc ctaaaaagga gagatacaag 2100

gccctgatcg gcctctacct gacagtgctg tatctggccg tcaaaaatat ggtgaattgt 2160

aacgccaggt acgtgatcgc cttctactgc agggataggg acaccgccct ctaccagaag 2220

gaagtgtgtt ggtatgacct ggaggaggac aagaagtccg gaaagcaaag acaggtcgag 2280

gactatacag ccctgaccag atatttcgtg tcccagggct acctgaacag acacgcctgc 2340

ggctatctca ggtccaacat gaacggcatc agcaacagcc tgctgaccgc ctacagaaat 2400

gccgtggacc acctgaacgc cattcctcct ctgggctccc tgtgtaggga catcggcaga 2460

gtggattcct acttcgctct gtaccactac gccgtgcaac agtacctcaa tggaaggtat 2520

tacagaaaaa cacccagaga acaggaactc ttcgccgcca tggctcagca caggacctgg 2580

tgttccgatc tcgtgaaggc cctgaacacc cctttcggat acaacctggc caggtacaag 2640

aacctctcca tcgacggact gttcgacagg gagggcgatc atgtggtgag agaggatggc 2700

gaaaaacctg ccgaa 2715

<210> 143
<211> 2835
<212> DNA
<213> Artificial Sequence

<220>
<223> Human codon-optimized Anaerobic Digester nCas1

<400> 143
aacaacaaga gaaaaacaaa ggctaaagcc gccggcctga agagcgtttt tttcgatcag 60

aagcaggccg tcctgaccac attcgctaaa ggcaacaaca gccagatcga gaagaaggtg 120

gtcaatagcg aagtgaagga cctgagacag ccccctgcct ttgacctcga actgaaagaa 180

aaaacttttt acatcagcgg aaagaataat atcaacacca gcagagagaa ccctctggcc 240

agcgccagcc tgcctctgtc caaaaggcag aggattaggg ctgagaggat caagagggcc 300

agagaggaga acaggcccta ccacaacgtc aaaagggtcg gcgaggatga tctgagggcc 360

aaagccgacc tggagaagca ttattttgga aaggaataca gcgataatct caagatccag 420

atcatttata acattctgga catcaataag atcatcagcc cctacatcaa tgacatcgtg 480

tactccatga acaatctggc caggaatgat gagtatatcg acggcaagat cgatgtgatc 540

ggttctctgt cctccaccac agactatagc agcttcatga gccccaataa ggatctggaa 600

aaggagaaga aattcagctt tcacagagaa aactataaaa aatttgtcga ggcctccaag 660

ccctacatga gatattacgg caaagtcttt atcagggacg tgaaaaaatc caagctgtcc 720

accggcaagg gcgaaaagat cgaggtgatg tacaggagcg atgaagagat tttcaccatc 780

ttccaaatcc tgtcctacgt gaggcagagc atcatgcata atgacattgg caacaagagc 840

agcatcctcg ccatcgaaaa ataccccgcc agattcgtgg gcttcctgtc cgacctctta 900

aaaactaaga ccaacgacgt gaacaggatg ttcatcgata acaatagcca gacaaatttc 960

tgggtcctgt tctccatctt cggcctccag gaccatacaa gcggcgctga taagatctgc 1020

aggaacttct atgatttcgt gatcaaggcc gacagcaaga acctgggatt ctccctgaaa 1080

aagatcaggg agctcatgct ggacctccct aacgctaaca tgctcaggga tcaccagttc 1140

gacacagtga ggagcaagtt ctataccctc ctcgacttta tcatttacca gcattacctg 1200

gaagagaaga gcaggatcga caacatggtc gagaagctca ggatgaccct gaaagaagag 1260

gagaaggagg tgctctacgc tgccgaagcc aagatcgtgt ggaatgccat tggcgccaag 1320

gtcatcaaca agctggtgcc tatgatgaac ggtgatgctc tgaaggagat taaaagaaag 1380

aacagggata ggaagctgcc ccagtccgtg atcgctaccg tccaggtgaa tagcgatgcc 1440

aatgtgtttt ccggcctcat ctacttcctc accctcttcc tggacggcaa agagattaac 1500

gagatggtga gcaatctcat caccaagttc gagaatatcg actcctgct gcatgtggat 1560

agggagatct ataagtccga cgagaaggac ctcgacctcg aaatcgagaa actggctctg 1620

tttttcaaag gcgtggtcag gcccaacgcc aaaacagaca caggagctgg cgaaatcagc 1680

aagagcttca gcattttcca gagcgccgag aggatcatcg aggagctgaa atttatcaag 1740

aatgtgacaa ggatggacaa cgagatcttt cccagcgagg gagtgttcct ggatgctgcc 1800

aatgtgctgg gcgtgagagg agatgacttc gacttttcca acgagttcgt cggcgatgac 1860

ctccactccg atgccaataa aaagatcatc aacaaaatca atggcaccaa ggaggacagg 1920

aacctcagaa acttcattat taataatgtg gtgaagagca ggaggttcca atacatcgcc 1980

aggcacatga atacccacta tgtgaagcag ctggccaaca acgagacact gaacaggttc 2040

gtgctcaaca aaatgggcga cgccaagatt atcaacaggt actatgagtc catcagcggc 2100

aacaccccca atatcgaggt cagatcccag attgactacc tcgtcaaaag actcaggagc 2160

ttctccttcg aggacctcaa tgacgtgaag caaaaagtca gacccggcac caacgaatcc 2220

atcgagaagg aaaagaaaaa ggctctcgtc ggactctgcc tgacaatcca atatctggtg 2280

tacaaaaacc tggtcaacat caacgctagg tacaccacag ccttctactg cctggaaagg 2340

gattccaagc tgaagggctt tggcgtggat gtgtggaggg atttcgagag ctacaccgct 2400

ctgaccaacc acttcatcaa ggaaggctat ctccccgtca ggaaggccga gattctgaga 2460

gccaatctga agcatctgga ctgcgaggac ggcttcaagt attacagaaa ccaggtcacc 2520

catctgaacg ccatcagagt ggcttataag tatattaacg aaattaagtc cgtgcacagc 2580

tactttgccc tgtaccacta catcatgcag aggcacctct atgactccct ccaagccaaa 2640

gccaaggata gctccggctt cgtcattgac gccctgaaga agtccttcga acataaaatc 2700

tacagcaaag acctcctgca cgtgctccat tccccctttcg gctacaacac cgccaggtac 2760

aagaacctga gcatcgaggc tctgttcgac aagaatgaga gcaggcctga agtcaacccc 2820

ctctccacca acgac 2835

<210> 144
<211> 2898
<212> DNA
<213> Artificial Sequence

<220>
<223> Human codon-optimized Ruminococcus flavefaciens XPD nCas1

<400> 144
atcgaaaaaa aaaagtcctt cgccaagggc atgggcgtga agtccacact cgtgtccggc 60

tccaaagtgt acatgacaac cttcgccgaa ggcagcgacg ccaggctgga aaagatcgtg 120

gagggcgaca gcatcaggag cgtgaatgag ggcgaggcct tcagcgctga aatggccgat 180

aaaaacgccg gctataagat cggcaacgcc aaattcagcc atcctaaggg ctacgccgtg 240

gtggctaaca accctctgta tacaggaccc gtccagcagg atatgctcgg cctgaaggaa 300

actctggaaa agaggtactt cggcgagagc gctgatggca atgacaatat ttgtatccag 360

gtgatccata acatcctgga cattgaaaaa atcctcgccg aatacattac caacgccgcc 420

tacgccgtca acaatatctc cggcctggat aaggacatta ttggattcgg caagttctcc 480

acagtgtata cctacgacga attcaaagac cccgagcacc atagggccgc tttcaacaat 540

aacgataagc tcatcaacgc catcaaggcc cagtatgacg agttcgacaa cttcctcgat 600

aaccccagac tcggctattt cggccaggcc tttttcagca aggagggcag aaattacatc 660

atcaattacg gcaacgaatg ctatgacatt ctggccctcc tgagcggact gaggcactgg 720

gtggtccata acaacgaaga agagtccagg atctccagga cctggctcta caacctcgat 780

aagaacctcg acaacgaata catctccacc ctcaactacc tctacgacag gatcaccaat 840

gagctgacca actccttctc caagaactcc gccgccaacg tgaactatat tgccgaaact 900

ctgggaatca accctgccga attcgccgaa caatatttca gattcagcat tatgaaagag 960

cagaaaaacc tcggattcaa tatcaccaag ctcagggaag tgatgctgga caggaaggat 1020

atgtccgaga tcaggaaaaa tcataaggtg ttcgactcca tcaggaccaa ggtctacacc 1080

atgatggact ttgtgattta taggtattac atcgaagagg atgccaaggt ggctgccgcc 1140

aataagtccc tccccgataa tgagaagtcc ctgagcgaga aggatatctt tgtgattaac 1200

ctgaggggct ccttcaacga cgaccagaag gatgccctct actacgatga agctaataga 1260

atttggagaa agctcgaaaa tatcatgcac aacatcaagg aatttagggg aaacaagaca 1320

agagagtata agaagaagga cgcccctaga ctgcccagaa tcctgcccgc tggccgtgat 1380

gtttccgcct tcagcaaact catgtatgcc ctgaccatgt tcctggatgg caaggagatc 1440

aacgacctcc tgaccaccct gattaataaa ttcgataaca tccagagctt cctgaaggtg 1500

atgcctctca tcggagtcaa cgctaagttc gtggaggaat acgccttttt caaagactcc 1560

gccaagatcg ccgatgagct gaggctgatc aagtccttcg ctagaatggg agaacctatt 1620

gccgatgcca ggagggccat gtatatcgac gccatccgta ttttaggaac caacctgtcc 1680

tatgatgagc tcaaggccct cgccgacacc ttttccctgg acgagaacgg aaacaagctc 1740

aagaaaggca agcacggcat gagaaatttc attattaata acgtgatcag caataaaagg 1800

ttccactacc tgatcagata cggtgatcct gcccacctcc atgagatcgc caaaaacgag 1860

gccgtggtga agttcgtgct cggcaggatc gctgacatcc agaaaaaaca gggccagaac 1920

ggcaagaacc agatcgacag gtactacgaa acttgtatcg gaaaggataa gggcaagagc 1980

gtgagcgaaa aggtggacgc tctcacaaag atcatcaccg gaatgaacta cgaccaattc 2040

gacaagaaaa ggagcgtcat tgaggacacc ggcagggaaa acgccgagag ggagaagttt 2100

aaaaagatca tcagcctgta cctcaccgtg atctaccaca tcctcaagaa tattgtcaat 2160

atcaacgcca ggtacgtcat cggattccat tgcgtcgagc gtgatgctca actgtacaag 2220

gagaaaggct acgacatcaa tctcaagaaa ctggaagaga agggattcag ctccgtcacc 2280

aagctctgcg ctggcattga tgaaactgcc cccgataaga gaaaggacgt ggaaaaggag 2340

atggctgaaa gagccaagga gagcattgac agcctcgaga gcgccaaccc caagctgtat 2400

gccaattaca tcaaatacag cgacgagaag aaagccgagg agttcaccag gcagattaac 2460

agggagaagg ccaaaaccgc cctgaacgcc tacctgagga acaccaagtg gaatgtgatc 2520

atcagggagg acctcctgag aattgacaac aagacatgta ccctgttcag aaacaaggcc 2580

gtccacctgg aagtggccag gtatgtccac gcctatatca acgacattgc cgaggtcaat 2640

tcctacttcc aactgtacca ttacatcatg cagagaatta tcatgaatga gaggtacgag 2700

aaaagcagcg gaaaggtgtc cgagtacttc gacgctgtga atgacgagaa gaagtacaac 2760

gataggctcc tgaaactgct gtgtgtgcct ttcggctact gtatccccag gtttaagaac 2820

ctgagcatcg aggccctgtt cgataggaac gaggccgcca agttcgacaa ggagaaaaag 2880

aaggtgtccg gcaattcc 2898

<210> 145
<211> 2829
<212> DNA
<213> Artificial Sequence

<220>
<223> Human codon-optimized Ruminococcus albus nCas1

<400> 145
gccaagaagt ccaagggaat gtccctcagg gagaagaggg agctggagaa gcaaaagagg 60

attcagaagg ccgccgtgaa tagcgtgaac gataccctg aaaaaactga ggaggctaat 120

gtggtcagcg tgaacgtgag aaccagcgcc gagaacaagc actccaaaaa gagcgctgcc 180

aaagccctgg gcctgaaaag cggcctcgtg attggagatg agctctacct gaccagcttc 240

ggaaggggca acgaagccaa gctggagaag aagattagcg gcgacaccgt ggagaaactg 300

ggaattggcg cctttgaggt cgccgaacgt gatgaaagca cactcacact cgagagcggc 360

aggatcaaag acaaaaccgc caggcctaag gaccccagac atattaccgt ggacacccag 420

ggaaagttta aggaggacat gctgggcatc aggtccgtgc tcgaaaagaa gatctttgga 480

aaaaccttcg acgataacat ccacgtccag ctggcctaca acatcctgga cgtcgagaag 540

atcatggctc agtatgtgag cgacatcgtc tacatgctgc acaacaccga caagaccgag 600

aggaacgata acctgatggg atacatgtcc attagaaata cctacaaaac cttctgcgac 660

acatccaatc tgcccgacga cacaaagcag aaggtggaga atcagaaaag ggaattcgac 720

aagatcatta agagcggcag actgggctac ttcggcgaag ctttcatggt caacagcggc 780

aactccacca agctgaggcc cgagaaagag atttatcata tcttcgccct catggccagc 840

ctgaggcaga gctacttcca cggatacgtc aaggataccg actatcaggg cacaacatgg 900

gcctacacac tggaggataa gctgaagggc cccagccatg aatttaggga gacaatcgac 960

aagatctttg acgagggatt cagcaagatc tccaaggact tcggaaaaat gaacaaagtg 1020

aatctccaaa tcctggagca gatgatcggc gagctgtacg gaagcatcga aaggcagaac 1080

ctgacctgcg attactacga ttttattcaa ctcaagaagc acaagtatct cggctttagc 1140

atcaagaggc tgagagagac aatgctcgag acaacccctg ccgagtgtta taaggctgag 1200

tgctacaata gcgagaggca aaagctctac aagctgatcg acttcctcat ctacgacctc 1260

tactataaca ggaagcccgc cagaatcgaa gagatcgtcg acaagctgag agagagcgtc 1320

aacgacgagg agaaggagag catctatagc gtggaggcca agtacgtgta tgagtccctg 1380

agcaaggtgc tggataagag cctgaagaat tccgtgtccg gcgaaactat caaggatctg 1440

cagaagaggt acgacgacga aactgccaac agaatttggg acatctccca gcactccatt 1500

tccggcaatg tcaactgctt ctgcaagctg atctacatta tgaccctcat gctggatgga 1560

aaggagatca acgacctcct caccaccctg gtgaacaagt ttgacaatat tgcttccttt 1620

attgacgtga tggacgaact cggcctggaa cacagcttta cagacaacta caagatgttt 1680

gccgacagca aggccatctg cctggacctg caattcatca atagctttgc tagaatgagc 1740

aagatcgacg acgagaaaag caagagacag ctgttcaggg atgctctggt gatcctcgat 1800

atcggaaaca aggacgaaac ttggattaac aactacctcg actccgacat cttcaagctg 1860

gacaaggaag gcaacaagct gaagggagcc aggcatgact ttaggaattt tatcgccaat 1920

aatgtcatta agtcctccag attcaagtac ctcgtcaaat acagctccgc cgacggcatg 1980

attaagttaa aaactaacga aaagctgatt ggcttcgtgc tcgacaagct ccccgaaact 2040

caaatcgata ggtactacga atcctgcggc ctcgacaatg ccgtcgtgga caagaaggtg 2100

aggatcgaga aactgtccgg actcatcagg gacatgaagt tcgacgattt ttccggcgtc 2160

aagacctcca ataaggccgg cgacaacgat aagcaggaca aggccaagta ccaggccatt 2220

atctccctgt acctgatggt gctctaccag attgtgaaga atatgatcta cgtcaacagc 2280

aggtatgtga ttgccttcca ctgcctggag agggacttcg gaatgtacgg caaggatttc 2340

ggcaagtact accagggctg taggaagctc accgatcatt tcatcgagga gaaatacatg 2400

aaggagggca agctgggctg taacaagaaa gtcggcaggt acctgaagaa taacatcagc 2460

tgttgcaccg acggcctgat taacacctac aggaaccagg tcgatcattt cgccgtcgtc 2520

agaaaaatcg gaaattacgc cgcctacatc aaatccattg gcagctggtt cgagctctat 2580

cattacgtga ttcagaggat cgtttttgac gaatatagat tcgccctcaa caacaccgag 2640

tccaactata agaacagcat cattaagcac catacatact gcaaggacat ggtgaaggcc 2700

ctgaataccc ctttcggcta cgacctgccc agatataaaa acctgagcat tggtgattta 2760

ttcgacagga acaactatct gaacaaaaca aaggagtcca tcgatgccaa ctccagcatt 2820

gatagccag 2829

<210> 146
<211> 330
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial sequence

<400> 146
ctactacact ggtgcgaatt tgcactagtc taaaacaagc tttttcttca ttgtctacaa 60

atgcaactac tacactagtg cgaatttgca ctagtctaaa actctatgta gcttttcttg 120

taaaacatat ttctactaca ctggtgcgat tttgcactag tctaaaactc atctgccttc 180

tgcatatcgg acacttgact actacaccgg tgcgaatttg cactagtcta aaacttacac 240

ctccttatgc gattttatcg tgcgctacta cactggtgcg aatttgcact agtctaaaac 300

ttcatcttca agaccgagca gggacatcat 330

<210> 147
<211> 939
<212> PRT
<213> Artificial Sequence

<220>
<223> Metagenomic hit (no protein accession): contig
emb|OBVH01003037.1, human gut metagenome sequence (also found in
WGS contigs emb|OBXZ01000094.1| and emb|OBJF01000033.1|)

<400> 147

Met Ala Lys Lys Lys Arg Ile Thr Ala Lys Glu Arg Lys Gln Asn His
1 5 10 15

Arg Glu Leu Leu Met Lys Lys Ala Asp Ser Asn Ala Glu Lys Glu Lys
20 25 30

Ala Lys Lys Pro Val Val Glu Asn Lys Pro Asp Thr Ala Ile Ser Lys

35 40 45

Asp Asn Thr Pro Lys Pro Asn Lys Glu Ile Lys Lys Ser Lys Ala Lys
50 55 60

Leu Ala Gly Val Lys Trp Val Ile Lys Ala Asn Asp Asp Val Ala Tyr
65 70 75 80

Ile Ser Ser Phe Gly Lys Gly Asn Asn Ser Val Leu Glu Lys Arg Ile
85 90 95

Met Gly Asp Val Ser Ser Asn Val Asn Lys Asp Ser His Met Tyr Val
100 105 110

Asn Pro Lys Tyr Thr Lys Lys Asn Tyr Glu Ile Lys Asn Gly Phe Ser
115 120 125

Ser Gly Ser Ser Leu Val Thr Tyr Pro Asn Lys Pro Asp Lys Asn Ser
130 135 140

Gly Met Asp Ala Leu Cys Leu Lys Pro Tyr Phe Glu Lys Asp Phe Phe
145 150 155 160

Gly His Ile Phe Thr Asp Asn Met His Ile Gln Ala Ile Tyr Asn Ile
165 170 175

Phe Asp Ile Glu Lys Ile Leu Ala Lys His Ile Thr Asn Ile Ile Tyr
180 185 190

Thr Val Asn Ser Phe Asp Arg Asn Tyr Asn Gln Ser Gly Asn Asp Thr
195 200 205

Ile Gly Phe Gly Leu Asn Tyr Arg Val Pro Tyr Ser Glu Tyr Gly Gly
210 215 220

Gly Lys Asp Ser Asn Gly Glu Pro Lys Asn Gln Ser Lys Trp Glu Lys
225 230 235 240

Arg Asp Asn Phe Ile Lys Phe Tyr Asn Glu Ser Lys Pro His Leu Gly

245 250 255

Tyr Tyr Glu Asn Ile Phe Tyr Asp His Gly Glu Pro Ile Ser Glu Glu
260 265 270

Lys Phe Tyr Asn Tyr Leu Asn Ile Leu Asn Phe Ile Arg Asn Asn Thr
275 280 285

Phe His Tyr Lys Asp Asp Asp Ile Glu Leu Tyr Ser Glu Asn Tyr Ser
290 295 300

Glu Glu Phe Val Phe Ile Asn Cys Leu Asn Lys Phe Val Lys Asn Lys
305 310 315 320

Phe Lys Asn Val Asn Lys Asn Phe Ile Ser Asn Glu Lys Asn Asn Leu
325 330 335

Tyr Ile Ile Leu Asn Ala Tyr Gly Lys Asp Thr Glu Asn Val Glu Val
340 345 350

Val Lys Lys Tyr Ser Lys Glu Leu Tyr Lys Leu Ser Val Leu Lys Thr
355 360 365

Asn Lys Asn Leu Gly Val Asn Val Lys Lys Leu Arg Glu Ser Ala Ile
370 375 380

Glu Tyr Gly Tyr Cys Pro Leu Pro Tyr Asp Lys Glu Lys Glu Val Ala
385 390 395 400

Lys Leu Ser Ser Val Lys His Lys Leu Tyr Lys Thr Tyr Asp Phe Val
405 410 415

Ile Thr His Tyr Leu Asn Ser Asn Asp Lys Leu Leu Leu Glu Ile Val
420 425 430

Glu Thr Leu Arg Leu Ser Lys Asn Asp Asp Glu Lys Glu Asn Val Tyr
435 440 445

Lys Lys Tyr Ala Glu Lys Leu Phe Lys Ala Asp Asp Val Ile Asn Pro

```
    450                 455                 460

Ile Lys Ala Ile Ser Lys Leu Phe Ala Arg Lys Gly Asn Lys Leu Phe
465                 470                 475                 480

Lys Glu Lys Ile Ile Ile Lys Lys Glu Tyr Ile Glu Asp Val Ser Ile
                485                 490                 495

Asp Lys Asn Ile Tyr Asp Phe Thr Lys Val Ile Phe Phe Met Thr Cys
            500                 505                 510

Phe Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr Asn Ile Ile Ser
        515                 520                 525

Lys Leu Gln Val Ile Glu Asp His Asn Asn Val Ile Lys Phe Ile Ser
    530                 535                 540

Asn Asn Lys Asp Ala Val Tyr Lys Asp Tyr Ser Asp Lys Tyr Ala Ile
545                 550                 555                 560

Phe Arg Asn Ala Gly Lys Ile Ala Thr Glu Leu Glu Ala Ile Lys Ser
                565                 570                 575

Ile Ala Arg Met Glu Asn Lys Ile Glu Asn Ala Pro Gln Glu Pro Leu
            580                 585                 590

Leu Lys Asp Ala Leu Leu Ser Leu Gly Val Ser Asp Asp Thr Lys Val
        595                 600                 605

Leu Glu Asn Thr Tyr Asn Lys Tyr Phe Asp Ser Lys Glu Lys Thr Asp
    610                 615                 620

Lys Gln Ser Gln Lys Val Ser Thr Phe Leu Met Asn Asn Val Ile Asn
625                 630                 635                 640

Asn Asn Arg Phe Lys Tyr Val Ile Lys Tyr Ile Asn Pro Ala Asp Ile
                645                 650                 655

Asn Gly Leu Ala Lys Asn Arg Tyr Leu Val Lys Phe Val Leu Ser Lys
```

660 665 670

Ile Pro Glu Glu Gln Ile Asp Ser Tyr Tyr Lys Leu Phe Ser Asn Glu
675 680 685

Glu Glu Pro Gly Cys Glu Glu Lys Ile Lys Leu Leu Thr Lys Lys Ile
690 695 700

Ser Lys Leu Asn Phe Gln Thr Leu Phe Glu Asn Asn Lys Ile Pro Asn
705 710 715 720

Val Glu Lys Glu Lys Lys Lys Ala Ile Ile Thr Leu Tyr Phe Thr Ile
725 730 735

Val Tyr Ile Leu Val Lys Asn Leu Val Asn Ile Asn Gly Leu Tyr Thr
740 745 750

Leu Ala Leu Tyr Phe Val Glu Arg Asp Gly Tyr Phe Tyr Lys Asp Ile
755 760 765

Cys Gly Lys Lys Asp Lys Lys Lys Ser Tyr Asn Asp Val Asp Tyr Leu
770 775 780

Leu Leu Pro Glu Ile Phe Ser Gly Ser Lys Tyr Arg Glu Glu Thr Lys
785 790 795 800

Asn Leu Lys Leu Pro Lys Glu Lys Asp Arg Asp Ile Met Lys Lys Tyr
805 810 815

Leu Pro Asn Asp Lys Asp Arg Glu Lys Tyr Asn Lys Phe Phe Thr Ala
820 825 830

Tyr Arg Asn Asn Ile Val His Leu Asn Ile Ile Ala Lys Leu Ser Glu
835 840 845

Leu Thr Lys Asn Ile Asp Lys Asp Ile Asn Ser Tyr Phe Asp Ile Tyr
850 855 860

His Tyr Cys Thr Gln Arg Val Met Phe Asn Tyr Cys Lys Glu Lys Asn

865 870 875 880

Asp Val Val Leu Ala Lys Met Lys Asp Leu Ala His Ile Lys Ser Asp
885 890 895

Cys Asn Glu Phe Ser Ser Lys His Thr Tyr Pro Phe Ser Ser Ala Val
900 905 910

Leu Arg Phe Met Asn Leu Pro Phe Ala Tyr Asn Val Pro Arg Phe Lys
915 920 925

Asn Leu Ser Tyr Lys Lys Phe Phe Asp Lys Gln
930 935

<210> 148
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR sequence

<400> 148
gattgaaagg attgtaaatt tgcaaggtct taaaac 36

<210> 149
<211> 958
<212> PRT
<213> Artificial Sequence

<220>
<223> Metagenomic hit (no protein accession): contig tpg|DBYI01000091.1| (Uncultivated Ruminococcus flavefaciens UBA1190 assembled from bovine gut metagenome)

<400> 149

Met Lys Lys Lys Ile Lys Ala Arg Asp Leu Arg Glu Ala Lys Lys Gln
1 5 10 15

Glu Lys Leu Ala Ala Phe Ser Ala Lys Ala Asn Thr Val Tyr Glu Asn
20 25 30

Glu Asp Lys Asn Val Glu Ala Phe Pro Glu Ala Leu Asn Leu Arg Ser

```
        35                  40                  45

Ile Lys Lys Ser Met Asn Lys Ala Ala Gly Leu Lys Ser Thr Leu Ile
    50                  55                  60

Asp Gly Lys Ser Leu Tyr Leu Thr Ala Phe Gly Lys Gly Asn Asn Ala
65                  70                  75                  80

Val Val Glu His Met Ile Ala Thr Asp Asp Ser Tyr Ser Leu Lys Thr
                85                  90                  95

Leu Glu Asn Glu Pro Ser Leu Lys Val Lys Ala Ala Asp Glu Leu Lys
            100                 105                 110

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

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

Gly Gln Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys Arg Tyr Phe
145                 150                 155                 160

Gly Lys Ile Phe Asp Asp Asn Leu His Ile Gln Ile Ile His Asn Ile
                165                 170                 175

Leu Asp Ile Glu Lys Ile Ile Ala Val Tyr Ala Thr Asn Ile Thr Ala
            180                 185                 190

Ala Ile Asp His Met Val Asp Asp Asp Asn Glu Gln Tyr Leu Gln Gly
        195                 200                 205

Asp Phe Ile Gly Tyr Met Asn Thr Leu Asn Thr Tyr Glu Val Phe Met
    210                 215                 220

Glu Pro Ser Lys Asn Pro Arg Leu Asp Ser Asn Ala Arg Lys Asn Ile
225                 230                 235                 240

Glu Asn Ser Arg Glu Lys Phe Glu Tyr Leu Leu Asp Thr Gln Arg Leu
```

245 250 255

Gly Tyr Leu Ser Leu Glu Tyr Asp Lys Arg Ser Lys Asp Lys Arg Lys
260 265 270

Ser Glu Glu Ile Lys Lys Arg Leu Tyr His Leu Val Ala Phe Ala Gly
275 280 285

Gln Leu Arg Gln Trp Ser Phe His Ser Val Glu Gly Leu Pro Arg Thr
290 295 300

Trp Ile Tyr Gln Leu Asp Asn Pro Lys Leu Ala Gln Glu Tyr Arg Asp
305 310 315 320

Thr Leu Asp Tyr Phe Phe Asn Glu Arg Phe Asp Ala Ile Asn Lys Asp
325 330 335

Phe Ile Glu Thr Asn Asn Ile Asn Leu His Ile Leu Lys Glu Val Phe
340 345 350

Pro Ala Glu Asp Phe Gln Lys Leu Ala Ala Leu Tyr Tyr Asp Phe Ile
355 360 365

Val Lys Lys Thr Phe Lys Asn Ile Gly Phe Ser Ile Lys Asn Leu Arg
370 375 380

Glu Gln Met Leu Glu Cys Asp Glu Ala Glu Lys Ile Arg Ser Lys Asp
385 390 395 400

Met Asn Ser Val Arg Ser Lys Leu Tyr Lys Leu Phe Asp Phe Cys Ile
405 410 415

Phe Tyr Gln Tyr Phe Ile Asp Glu Glu Arg Ser Arg Glu Asn Val Asn
420 425 430

Tyr Leu Arg Ser Thr Leu Asn Asp Glu Gln Lys Asp Ala Phe Tyr Glu
435 440 445

Glu Glu Gly Lys Arg Leu Trp Ser Glu Asn Arg Lys Lys Phe Ile Tyr

450 455 460

Phe Cys Asp Asn Ile Asn Lys Trp Val Lys Asn Asp Tyr Ser Asp Glu
465 470 475 480

Val Ala Lys Cys Ile Asp Leu Asn Glu Phe Arg Val Asn Ser Asn Val
485 490 495

Ser Tyr Phe Ser Lys Leu Leu Tyr Ala Met Ser Phe Phe Leu Asp Gly
500 505 510

Lys Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Asn Lys Phe Asp Asn
515 520 525

Ile Arg Ser Phe Ile Asp Thr Ala Asn Phe Leu Asn Ile Asp Val Lys
530 535 540

Phe Thr Lys Asp Tyr Asp Phe Phe Asn Ile Ile Cys Asp Tyr Ala Gly
545 550 555 560

Glu Leu Asn Ile Ile Lys Asn Ile Ala Arg Met Lys Lys Pro Ser Pro
565 570 575

Ser Ala Lys Lys Asn Met Tyr Arg Asp Ala Leu Thr Ile Leu Gly Ile
580 585 590

Pro Thr Glu Met Ser Asp Glu Gln Leu Asp Ala Glu Ile Asp Lys Ile
595 600 605

Leu Glu Lys Lys Ile Asn Pro Val Thr Gly Lys Thr Glu Lys Gly Lys
610 615 620

Asn Pro Phe Arg Asn Phe Ile Ala Asn Asn Val Ile Glu Asn Lys Arg
625 630 635 640

Phe Ile Tyr Val Ile Lys Phe Cys Asn Pro Lys Asn Val Arg Lys Leu
645 650 655

Val Asn Asn Thr Lys Val Thr Glu Phe Val Leu Lys Arg Met Pro Glu

660 665 670

Thr Gln Ile Asp Arg Tyr Phe Glu Ser Cys Ile Glu Gly Asn Leu Asn
675 680 685

Pro Thr Thr Glu Lys Lys Ile Glu Lys Leu Ala Glu Met Ile Lys Asn
690 695 700

Ile Lys Phe Glu Glu Phe Arg Asn Val Lys Gln Lys Val Arg Asp Asn
705 710 715 720

Ser Gln Glu Ala Val Glu Lys Glu Arg Phe Lys Ala Ile Ile Gly Leu
725 730 735

Tyr Leu Thr Val Ile Tyr Leu Leu Val Lys Asn Leu Val Asn Val Asn
740 745 750

Ser Arg Tyr Val Met Ala Phe His Cys Leu Glu Arg Asp Ala Lys Leu
755 760 765

Tyr Gly Val Gln Asn Ile Gly Gly Asp Tyr Leu Ala Leu Thr Ala Lys
770 775 780

Leu Cys Ala Glu Gly Asp Asp Tyr Gly Lys Lys Leu Ser Glu Ala Lys
785 790 795 800

Gln Asn Ile Asn Gln Asp Lys Val Gln Met Pro Lys Asn Tyr Phe Leu
805 810 815

Ala Arg Asn Lys Arg Trp Arg Glu Ala Ile Glu Gln Asp Ile Asp Asn
820 825 830

Ala Lys Lys Trp Phe Ile Gly Glu Lys Phe Asn Asn Val Lys Asn Tyr
835 840 845

Arg Asn Asn Val Ala His Leu Thr Ala Ile Arg Asn Cys Ala Glu Phe
850 855 860

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

865 870 875 880

Leu Ile Gln Arg Gln Leu Ala Gly Arg Leu Asp Pro Asn His Pro Gly
885 890 895

Phe Glu Lys Asn Tyr Pro Gln Tyr Ala Pro Leu Phe Lys Trp Asn Thr
900 905 910

Tyr Val Lys Asp Met Val Lys Ala Leu Asn Ser Pro Phe Gly Tyr Asn
915 920 925

Ile Pro Arg Phe Lys Asp Leu Ser Ile Asp Ala Leu Phe Asp Arg Asn
930 935 940

Glu Met Lys Glu Glu Thr Asp Asp Glu Lys Lys Ile Gln Thr
945 950 955

<210> 150
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR 1

<400> 150
gaactatacc cctaccaaat ggtcggggtc tgaaac 36

<210> 151
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> DR2

<400> 151
tgaactacac ccctgccaag tggttggggt ctgaaac 37

<210> 152
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR3

<400> 152
gaactacacc cctaccaagt ggttggggtc tgaaac 36

<210> 153
<211> 877
<212> PRT
<213> Artificial Sequence

<220>
<223> Metagenomic hit (no protein accession): contig tpg|DJXD01000002.1| (uncultivated Ruminococcus assembly, UBA7013, from sheep gut metagenome)

<400> 153

Met Lys Lys Gln Lys Ser Lys Lys Thr Val Ser Lys Thr Ser Gly Leu
1 5 10 15

Lys Glu Ala Leu Ser Val Gln Gly Thr Val Ile Met Thr Ser Phe Gly
20 25 30

Lys Gly Asn Met Ala Asn Leu Ser Tyr Lys Ile Pro Ser Ser Gln Lys
35 40 45

Pro Gln Asn Leu Asn Ser Ser Ala Gly Leu Lys Asn Val Glu Val Ser
50 55 60

Gly Lys Lys Ile Lys Phe Gln Gly Arg His Pro Lys Ile Ala Thr Thr
65 70 75 80

Asp Asn Pro Leu Phe Lys Pro Gln Pro Gly Met Asp Leu Leu Cys Leu
85 90 95

Lys Asp Lys Leu Glu Met His Tyr Phe Gly Lys Thr Phe Asp Asp Asn
100 105 110

Ile His Ile Gln Leu Ile Tyr Gln Ile Leu Asp Ile Glu Lys Ile Leu
115 120 125

Ala Val His Val Asn Asn Ile Val Phe Thr Leu Asp Asn Val Leu His

```
    130                 135                 140

Pro Gln Lys Glu Glu Leu Thr Glu Asp Phe Ile Gly Ala Gly Gly Trp
145                 150                 155                 160

Arg Ile Asn Leu Asp Tyr Gln Thr Leu Arg Gly Gln Thr Asn Lys Tyr
                165                 170                 175

Asp Arg Phe Lys Asn Tyr Ile Lys Arg Lys Glu Leu Leu Tyr Phe Gly
            180                 185                 190

Glu Ala Phe Tyr His Glu Asn Glu Arg Arg Tyr Glu Glu Asp Ile Phe
        195                 200                 205

Ala Ile Leu Thr Leu Leu Ser Ala Leu Arg Gln Phe Cys Phe His Ser
    210                 215                 220

Asp Leu Ser Ser Asp Glu Ser Asp His Val Asn Ser Phe Trp Leu Tyr
225                 230                 235                 240

Gln Leu Glu Asp Gln Leu Ser Asp Glu Phe Lys Glu Thr Leu Ser Ile
                245                 250                 255

Leu Trp Glu Glu Val Thr Glu Arg Ile Asp Ser Glu Phe Leu Lys Thr
            260                 265                 270

Asn Thr Val Asn Leu His Ile Leu Cys His Val Phe Pro Lys Glu Ser
        275                 280                 285

Lys Glu Thr Ile Val Arg Ala Tyr Tyr Glu Phe Leu Ile Lys Lys Ser
    290                 295                 300

Phe Lys Asn Met Gly Phe Ser Ile Lys Lys Leu Arg Glu Ile Met Leu
305                 310                 315                 320

Glu Gln Ser Asp Leu Lys Ser Phe Lys Glu Asp Lys Tyr Asn Ser Val
                325                 330                 335

Arg Ala Lys Leu Tyr Lys Leu Phe Asp Phe Ile Ile Thr Tyr Tyr Tyr
```

340 345 350

Asp His His Ala Phe Glu Lys Glu Ala Leu Val Ser Ser Leu Arg Ser
355 360 365

Ser Leu Thr Glu Glu Asn Lys Glu Glu Ile Tyr Ile Lys Thr Ala Arg
370 375 380

Thr Leu Ala Ser Ala Leu Gly Ala Asp Phe Lys Lys Ala Ala Ala Asp
385 390 395 400

Val Asn Ala Lys Asn Ile Arg Asp Tyr Gln Lys Lys Ala Asn Asp Tyr
405 410 415

Arg Ile Ser Phe Glu Asp Ile Lys Ile Gly Asn Thr Gly Ile Gly Tyr
420 425 430

Phe Ser Glu Leu Ile Tyr Met Leu Thr Leu Leu Leu Asp Gly Lys Glu
435 440 445

Ile Asn Asp Leu Leu Thr Thr Leu Ile Asn Lys Phe Asp Asn Ile Ile
450 455 460

Ser Phe Ile Asp Ile Leu Lys Lys Leu Asn Leu Glu Phe Lys Phe Lys
465 470 475 480

Pro Glu Tyr Ala Asp Phe Phe Asn Met Thr Asn Cys Arg Tyr Thr Leu
485 490 495

Glu Glu Leu Arg Val Ile Asn Ser Ile Ala Arg Met Gln Lys Pro Ser
500 505 510

Ala Asp Ala Arg Lys Ile Met Tyr Arg Asp Ala Leu Arg Ile Leu Gly
515 520 525

Met Asp Asn Arg Pro Asp Glu Glu Ile Asp Arg Glu Leu Glu Arg Thr
530 535 540

Met Pro Val Gly Ala Asp Gly Lys Phe Ile Lys Gly Lys Gln Gly Phe

545 550 555 560

Arg Asn Phe Ile Ala Ser Asn Val Ile Glu Ser Ser Arg Phe His Tyr
565 570 575

Leu Val Arg Tyr Asn Asn Pro His Lys Thr Arg Thr Leu Val Lys Asn
580 585 590

Pro Asn Val Val Lys Phe Val Leu Glu Gly Ile Pro Glu Thr Gln Ile
595 600 605

Lys Arg Tyr Phe Asp Val Cys Lys Gly Gln Glu Ile Pro Pro Thr Ser
610 615 620

Asp Lys Ser Ala Gln Ile Asp Val Leu Ala Arg Ile Ile Ser Ser Val
625 630 635 640

Asp Tyr Lys Ile Phe Glu Asp Val Pro Gln Ser Ala Lys Ile Asn Lys
645 650 655

Asp Asp Pro Ser Arg Asn Phe Ser Asp Ala Leu Lys Lys Gln Arg Tyr
660 665 670

Gln Ala Ile Val Ser Leu Tyr Leu Thr Val Met Tyr Leu Ile Thr Lys
675 680 685

Asn Leu Val Tyr Val Asn Ser Arg Tyr Val Ile Ala Phe His Cys Leu
690 695 700

Glu Arg Asp Ala Phe Leu His Gly Val Thr Leu Pro Lys Met Asn Lys
705 710 715 720

Lys Ile Val Tyr Ser Gln Leu Thr Thr His Leu Leu Thr Asp Lys Asn
725 730 735

Tyr Thr Thr Tyr Gly His Leu Lys Asn Gln Lys Gly His Arg Lys Trp
740 745 750

Tyr Val Leu Val Lys Asn Asn Leu Gln Asn Ser Asp Ile Thr Ala Val

755 760 765

Ser Ser Phe Arg Asn Ile Val Ala His Ile Ser Val Val Arg Asn Ser
770 775 780

Asn Glu Tyr Ile Ser Gly Ile Gly Glu Leu His Ser Tyr Phe Glu Leu
785 790 795 800

Tyr His Tyr Leu Val Gln Ser Met Ile Ala Lys Asn Asn Trp Tyr Asp
805 810 815

Thr Ser His Gln Pro Lys Thr Ala Glu Tyr Leu Asn Asn Leu Lys Lys
820 825 830

His His Thr Tyr Cys Lys Asp Phe Val Lys Ala Tyr Cys Ile Pro Phe
835 840 845

Gly Tyr Val Val Pro Arg Tyr Lys Asn Leu Thr Ile Asn Glu Leu Phe
850 855 860

Asp Arg Asn Asn Pro Asn Pro Glu Pro Lys Glu Glu Val
865 870 875

<210> 154
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 154
caactacaac cccgtaaaaa tacggggttc tgaaac 36

<210> 155
<211> 984
<212> PRT
<213> Artificial Sequence

<220>
<223> Metagenomic hit (no protein accession): contig OGZC01000639.1
(human gut metagenome assembly)

<400> 155

```
Met Lys Lys Lys Asn Ile Arg Ala Thr Arg Glu Ala Leu Lys Ala Gln
1               5                  10                  15

Lys Ile Lys Lys Ser Gln Glu Asn Glu Ala Leu Lys Lys Gln Lys Leu
            20                  25                  30

Ala Glu Glu Ala Ala Gln Lys Arg Arg Glu Glu Leu Glu Lys Lys Asn
        35                  40                  45

Leu Ala Gln Trp Glu Glu Thr Ser Ala Glu Gly Arg Arg Ser Arg Val
    50                  55                  60

Lys Ala Val Gly Val Lys Ser Val Phe Val Val Gly Asp Asp Leu Tyr
65                  70                  75                  80

Leu Ala Thr Phe Gly Asn Gly Asn Glu Thr Val Leu Glu Lys Lys Ile
                85                  90                  95

Thr Pro Asp Gly Lys Ile Thr Thr Phe Pro Glu Glu Glu Thr Phe Thr
            100                 105                 110

Ala Lys Leu Lys Phe Ala Gln Thr Glu Pro Thr Val Ala Thr Ser Ile
        115                 120                 125

Gly Ile Ser Asn Gly Arg Ile Val Leu Pro Glu Ile Ser Val Asp Asn
    130                 135                 140

Pro Leu His Thr Thr Met Gln Lys Asn Thr Ile Lys Arg Ser Ala Gly
145                 150                 155                 160

Glu Asp Ile Leu Gln Leu Lys Asp Val Leu Glu Asn Arg Tyr Phe Asp
                165                 170                 175

Arg Ser Phe Asn Asp Asp Leu His Ile Arg Leu Ile Tyr Asn Ile Leu
            180                 185                 190

Asp Ile Glu Lys Ile Leu Ala Glu Tyr Thr Thr Asn Ala Val Phe Ala
        195                 200                 205
```

```
Ile Asp Asn Val Ser Gly Cys Ser Asp Asp Phe Leu Ser Asn Phe Ser
    210                 215                 220

Thr Arg Asn Gln Trp Asp Glu Phe Gln Asn Pro Glu Gln His Arg Glu
225                 230                 235                 240

His Phe Gly Asn Lys Asp Asn Val Ile Cys Ser Val Lys Lys Gln Gln
                245                 250                 255

Asp Leu Phe Phe Asn Phe Phe Lys Asn Asn Arg Ile Gly Tyr Phe Gly
            260                 265                 270

Lys Ala Phe Phe His Ala Glu Ser Glu Arg Lys Ile Val Lys Lys Thr
        275                 280                 285

Glu Lys Glu Val Tyr His Ile Leu Thr Leu Ile Gly Ser Leu Arg Gln
    290                 295                 300

Trp Ile Thr His Ser Thr Glu Gly Gly Ile Ser Arg Leu Trp Leu Tyr
305                 310                 315                 320

Gln Leu Glu Asp Ala Leu Ser Arg Glu Tyr Gln Glu Thr Met Asn Asn
                325                 330                 335

Cys Tyr Asn Ser Thr Ile Tyr Gly Leu Gln Lys Asp Phe Glu Lys Thr
            340                 345                 350

Asn Ala Pro Asn Leu Asn Phe Leu Ala Glu Ile Leu Gly Lys Asn Ala
        355                 360                 365

Ser Glu Leu Ala Glu Pro Tyr Phe Arg Phe Ile Ile Thr Lys Glu Tyr
    370                 375                 380

Lys Asn Leu Gly Phe Ser Ile Lys Thr Leu Arg Glu Met Leu Leu Asp
385                 390                 395                 400

Gln Pro Asp Leu Gln Glu Ile Arg Glu Asn His Asn Val Tyr Asp Ser
                405                 410                 415
```

```
Ile Arg Ser Lys Leu Tyr Lys Met Ile Asp Phe Val Leu Val Tyr Ala
            420                 425                 430

Tyr Ser Asn Glu Arg Lys Ser Lys Ala Asp Ala Leu Ala Ser Asn Leu
        435                 440                 445

Arg Ser Ala Ile Thr Glu Asp Ala Lys Lys Arg Ile Tyr Gln Asn Glu
    450                 455                 460

Ala Asp Gln Leu Trp Thr Ser Tyr Gln Glu Leu Phe Lys Arg Ile Arg
465                 470                 475                 480

Gly Phe Lys Gly Ala Gln Val Lys Glu Tyr Ser Ser Lys Asn Met Pro
                485                 490                 495

Ile Pro Ile Gln Lys Gln Ile Gln Asn Ile Leu Lys Pro Ala Glu Gln
            500                 505                 510

Val Thr Tyr Phe Thr Lys Leu Met Tyr Leu Leu Thr Met Phe Leu Asp
        515                 520                 525

Gly Lys Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Asn Lys Phe Asp
    530                 535                 540

Asn Ile Ser Ser Leu Leu Lys Thr Met Glu Gln Leu Glu Leu Gln Thr
545                 550                 555                 560

Thr Phe Lys Glu Asp Tyr Thr Phe Phe Gln Gln Ser Ser Arg Leu Cys
                565                 570                 575

Lys Glu Ile Thr Gln Leu Lys Ser Phe Ala Arg Met Gly Asn Pro Ile
            580                 585                 590

Ser Asn Leu Lys Glu Val Met Met Val Asp Ala Ile Gln Ile Leu Gly
        595                 600                 605

Thr Glu Lys Ser Glu Gln Glu Leu Gln Ser Met Ala Cys Phe Phe Phe
    610                 615                 620
```

```
Arg Asp Lys Asn Gly Lys Lys Leu Asn Thr Gly Glu His Gly Met Arg
625                 630                 635                 640

Asn Phe Ile Gly Asn Asn Val Ile Ser Asn Thr Arg Phe Gln Tyr Leu
                645                 650                 655

Ile Arg Tyr Gly Asn Pro Gln Lys Leu His Thr Leu Ser Gln Asn Glu
            660                 665                 670

Thr Val Val Arg Phe Val Leu Ser Arg Ile Ala Lys Asn Gln Arg Val
        675                 680                 685

Gln Gly Met Asn Gly Lys Asn Gln Ile Asp Arg Tyr Tyr Glu Thr Cys
    690                 695                 700

Gly Gly Thr Asn Ser Trp Ser Val Ser Glu Glu Glu Lys Ile Asn Phe
705                 710                 715                 720

Leu Cys Lys Ile Leu Thr Asn Met Ser Tyr Asp Gln Phe Gln Asp Val
                725                 730                 735

Lys Gln Ser Gly Ala Glu Ile Thr Ala Glu Glu Lys Arg Lys Lys Glu
            740                 745                 750

Arg Tyr Lys Ala Ile Ile Ser Leu Tyr Leu Thr Val Leu Tyr Gln Leu
        755                 760                 765

Ile Lys Asn Leu Val Asn Ile Asn Ala Arg Tyr Ile Ile Ala Phe His
    770                 775                 780

Cys Leu Glu Arg Asp Ala Ile Leu Tyr Ser Ser Lys Phe Asn Thr Ser
785                 790                 795                 800

Ile Asn Leu Lys Lys Arg Tyr Thr Ala Leu Thr Glu Met Ile Leu Gly
                805                 810                 815

Tyr Glu Thr Asp Glu Lys Ala Arg Arg Lys Asp Thr Arg Thr Val Tyr
            820                 825                 830
```

```
Glu Lys Ala Glu Ala Ala Lys Asn Arg His Leu Lys Asn Val Lys Trp
        835                 840                 845

Asn Cys Lys Thr Arg Glu Asn Leu Glu Asn Ala Asp Lys Asn Ala Ile
    850                 855                 860

Val Ala Phe Arg Asn Ile Val Ala His Leu Trp Ile Ile Arg Asp Ala
865                 870                 875                 880

Asp Arg Phe Ile Thr Gly Met Gly Ala Met Lys Arg Tyr Phe Asp Cys
                885                 890                 895

Tyr His Tyr Leu Leu Gln Arg Glu Leu Gly Tyr Ile Leu Glu Lys Ser
            900                 905                 910

Asn Gln Gly Ser Glu Tyr Thr Lys Lys Ser Leu Glu Lys Val Gln Gln
        915                 920                 925

Tyr His Ser Tyr Cys Lys Asp Phe Leu His Met Leu Cys Leu Pro Phe
    930                 935                 940

Ala Tyr Cys Ile Pro Arg Tyr Lys Asn Leu Ser Ile Ala Glu Leu Phe
945                 950                 955                 960

Asp Arg His Glu Pro Glu Ala Glu Pro Lys Glu Glu Ala Ser Ser Val
                965                 970                 975

Asn Asn Ser Gln Phe Ile Thr Thr
            980


<210>  156
<211>  33
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  DR 1

<400>  156
actacaatcc atggttatcg tggatactaa aac                                    33
```

<210> 157
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> DR2

<400> 157
actacaatcc acgtttatcg tggatactaa aac 33

<210> 158
<211> 978
<212> PRT
<213> Artificial Sequence

<220>
<223> Metagenomic hit (no protein accession): contig emb|OHBM01000764.1
(human gut metagenome assembly)

<220>
<221> misc_feature
<222> (1)..(222)
<223> Xaa can be any naturally occurring amino acid

<400> 158

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1 5 10 15

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
20 25 30

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
35 40 45

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
50 55 60

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
65 70 75 80

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa

85 90 95

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
100 105 110

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
115 120 125

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
130 135 140

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
145 150 155 160

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
165 170 175

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
180 185 190

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
195 200 205

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa His Pro
210 215 220

Leu Gln Lys Arg Tyr Arg Tyr Leu Thr Ser Thr Asn Leu Lys Ser Phe
225 230 235 240

Glu Thr Tyr Lys Asn Asn Leu Val Asn Lys Lys Lys Phe Asp Leu Asp
245 250 255

Arg Val Lys Lys Ile Pro Gln Leu Ala Tyr Phe Gly Ser Ala Phe Tyr
260 265 270

Asn Thr Pro Glu Asp Thr Ser Ala Lys Ile Thr Lys Thr Lys Ile Lys
275 280 285

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

```
    290                 295                 300

Asn Phe Ser Ala His Tyr Leu Asp Arg Asn Arg Ala Lys Ser Ser Asp
305                 310                 315                 320

Ala Glu Asp Phe Asp Gly Thr Ser Val Ile Met Tyr Asn Leu Asp Asn
                325                 330                 335

Glu Glu Leu Tyr Lys Lys Leu Tyr Asn Lys Lys Val His Met Ala Leu
            340                 345                 350

Thr Gly Met Lys Lys Val Leu Asp Ala Asn Phe Asn Lys Lys Val Glu
        355                 360                 365

His Leu Asn Asn Ser Phe Ile Lys Asn Ser Ala Lys Asp Phe Val Ile
    370                 375                 380

Leu Cys Glu Val Leu Gly Ile Lys Ser Arg Asp Glu Lys Thr Lys Phe
385                 390                 395                 400

Val Lys Asp Tyr Tyr Asp Phe Val Val Arg Lys Asn Tyr Lys His Leu
                405                 410                 415

Gly Phe Ser Val Lys Glu Leu Arg Glu Leu Leu Phe Ala Asn His Asp
            420                 425                 430

Ser Asn Lys Tyr Ile Lys Glu Phe Asp Lys Ile Ser Asn Lys Lys Phe
        435                 440                 445

Asp Ser Val Arg Ser Arg Leu Asn Arg Leu Ala Asp Tyr Ile Ile Tyr
    450                 455                 460

Asp Tyr Tyr Asn Lys Asn Asn Ala Lys Val Ser Asp Leu Val Lys Tyr
465                 470                 475                 480

Leu Arg Ala Ala Ala Asp Asp Glu Gln Lys Lys Lys Ile Tyr Leu Asn
                485                 490                 495

Glu Ser Ile Asn Leu Val Lys Ser Gly Ile Leu Glu Arg Ile Lys Lys
```

500 505 510

Ile Leu Pro Lys Leu Asn Gly Lys Ile Ile Gly Asn Met Gln Pro Asp
515 520 525

Ser Thr Ile Thr Ala Ser Met Leu His Asn Thr Gly Lys Asp Trp His
530 535 540

Pro Ile Ser Glu Asn Ala His Tyr Phe Thr Lys Trp Ile Tyr Thr Leu
545 550 555 560

Thr Leu Phe Met Asp Gly Lys Glu Ile Asn Asp Leu Val Thr Thr Leu
565 570 575

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

Gln Ser Val Cys Thr His Phe Ser Glu Glu Arg Lys Met Phe Ile Asp
595 600 605

Ser Ala Glu Ile Cys Ser Glu Leu Ser Ala Met Asn Ser Phe Ala Arg
610 615 620

Met Glu Ala Pro Gly Ala Ser Ser Lys Arg Ala Met Phe Val Glu Ala
625 630 635 640

Ala Arg Ile Leu Gly Asp Asn Arg Ser Lys Glu Glu Leu Glu Glu Tyr
645 650 655

Phe Asp Thr Leu Phe Asp Lys Ser Ala Ser Lys Lys Glu Lys Gly Phe
660 665 670

Arg Asn Phe Ile Arg Asn Asn Val Val Asp Ser Asn Arg Phe Lys Tyr
675 680 685

Leu Thr Arg Tyr Thr Asp Thr Ser Ser Val Lys Ala Phe Ser Asn Asn
690 695 700

Lys Ala Leu Val Lys Phe Ala Ile Lys Asp Ile Pro Gln Glu Gln Ile

705 710 715 720

Leu Arg Tyr Tyr Asn Ser Cys Phe Gly Ala Ser Glu Arg Tyr Tyr Asn
725 730 735

Asp Gly Met Ser Asp Lys Leu Val Glu Ala Ile Gly Lys Ile Asn Leu
740 745 750

Met Gln Phe Asn Gly Val Ile Gln Gln Ala Asp Arg Asn Met Leu Pro
755 760 765

Glu Glu Lys Lys Lys Ala Asn Ala Gln Lys Glu Lys Tyr Lys Ser Ile
770 775 780

Ile Arg Leu Tyr Leu Thr Val Cys Tyr Leu Phe Phe Lys Asn Leu Val
785 790 795 800

Tyr Val Asn Ser Arg Tyr Tyr Ser Ala Phe Tyr Asn Leu Glu Lys Asp
805 810 815

Arg Ser Leu Phe Glu Ile Asn Gly Glu Leu Lys Pro Thr Gly Lys Phe
820 825 830

Asp Glu Gly His Tyr Thr Gly Leu Val Lys Leu Phe Ile Asp Asn Gly
835 840 845

Trp Ile Asn Pro Arg Ala Ser Ala Tyr Leu Thr Val Asn Leu Ala Asn
850 855 860

Ser Asp Glu Thr Ala Ile Arg Thr Phe Arg Asn Thr Ala Glu His Leu
865 870 875 880

Glu Ala Leu Arg Asn Ala Asp Lys Tyr Leu Asn Asp Leu Lys Gln Phe
885 890 895

Asp Ser Tyr Phe Glu Ile Tyr His Tyr Ile Thr Gln Arg Asn Ile Lys
900 905 910

Glu Lys Cys Glu Met Leu Lys Glu Gln Thr Val Lys Tyr Asn Asn Asp

915 920 925

Leu Leu Lys Tyr His Gly Tyr Ser Lys Asp Phe Val Lys Ala Leu Cys
930 935 940

Val Pro Phe Gly Tyr Asn Leu Pro Arg Phe Lys Asn Leu Ser Ile Asp
945 950 955 960

Ala Leu Phe Asp Lys Asn Asp Lys Arg Glu Lys Leu Lys Lys Gly Phe
965 970 975

Glu Asp

<210> 159
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 159
gaactacagc cctccttaaa tgggggttc taaaac 36

<210> 160
<211> 1023
<212> PRT
<213> Artificial Sequence

<220>
<223> Metagenomic hit (no protein accession): contig emb|OHCP01000044.1
(human gut metagenome assembly)

<400> 160

Met Ala Lys Lys Ile Thr Ala Lys Gln Lys Arg Glu Glu Lys Glu Arg
1 5 10 15

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

Glu Thr Lys Glu Glu Ile Lys Thr Gly Glu Ile Gln Asp Asn Asn Arg
35 40 45

```
Lys Arg Ser Arg Gln Lys Ser Gln Ala Lys Ala Met Gly Leu Lys Ala
    50                  55                  60

Val Leu Ser Phe Asp Asn Lys Ile Ala Ile Ala Ser Phe Val Ser Ser
65                  70                  75                  80

Lys Asn Ala Lys Ser Ser His Ile Glu Arg Ile Thr Asp Lys Glu Gly
                85                  90                  95

Thr Thr Ile Ser Val Asn Ser Lys Met Phe Glu Ser Ser Val Asn Lys
            100                 105                 110

Arg Asp Ile Asn Ile Glu Lys Arg Ile Thr Ile Glu Glu Pro Gln Gln
        115                 120                 125

Asp Gly Thr Ile Lys Lys Glu Glu Lys Gly Val Lys Ser Thr Thr Cys
    130                 135                 140

Asn Pro Tyr Phe Lys Val Gly Gly Lys Asp Tyr Ile Gly Ile Lys Glu
145                 150                 155                 160

Ile Ala Glu Glu His Phe Phe Gly Arg Ala Phe Pro Asn Glu Asn Leu
                165                 170                 175

Arg Val Gln Ile Ala Tyr Asn Ile Phe Asp Val Gln Lys Ile Leu Gly
            180                 185                 190

Thr Phe Val Asn Asn Ile Ile Tyr Ser Phe Tyr Asn Leu Ser Arg Asp
        195                 200                 205

Glu Val Gln Ser Asp Asn Asp Val Ile Gly Met Leu Tyr Ser Ile Ser
    210                 215                 220

Asp Tyr Asp Arg Gln Lys Glu Thr Glu Thr Phe Leu Gln Ala Lys Ser
225                 230                 235                 240

Leu Leu Lys Gln Thr Glu Ala Tyr Tyr Ala Tyr Phe Asp Asp Val Phe
                245                 250                 255
```

```
Lys Lys Asn Lys Lys Pro Asp Lys Asn Lys Glu Gly Asp Asn Ser Lys
            260                 265                 270

Gln Tyr Gln Glu Asn Leu Arg His Asn Phe Asn Ile Leu Arg Val Leu
        275                 280                 285

Ser Phe Leu Arg Gln Ile Cys Met His Ala Glu Val His Val Ser Asp
    290                 295                 300

Asp Glu Gly Cys Thr Arg Thr Gln Asn Tyr Thr Asp Ser Leu Glu Ala
305                 310                 315                 320

Leu Phe Asn Ile Ser Lys Ala Phe Gly Lys Lys Met Pro Glu Leu Lys
                325                 330                 335

Thr Leu Ile Asp Asn Ile Tyr Ser Lys Gly Ile Asn Ala Ile Asn Asp
            340                 345                 350

Glu Phe Val Lys Asn Gly Lys Asn Asn Leu Tyr Ile Leu Ser Lys Val
        355                 360                 365

Tyr Pro Asn Glu Lys Arg Glu Val Leu Leu Arg Glu Tyr Tyr Asn Phe
    370                 375                 380

Val Val Cys Lys Glu Gly Ser Asn Ile Gly Ile Ser Thr Arg Lys Leu
385                 390                 395                 400

Lys Glu Thr Met Ile Ala Gln Asn Met Pro Ser Leu Lys Glu Glu Asn
                405                 410                 415

Thr Tyr Arg Asn Lys Leu Tyr Thr Val Met Asn Phe Ile Leu Val Arg
            420                 425                 430

Glu Leu Lys Asn Cys Ala Thr Ile Arg Glu Gln Met Ile Lys Glu Leu
        435                 440                 445

Arg Ala Asn Met Asp Glu Glu Glu Gly Arg Asp Arg Ile Tyr Ser Lys
    450                 455                 460
```

```
Tyr Ala Lys Glu Ile Tyr Leu Tyr Val Lys Asp Lys Leu Lys Leu Met
465                 470                 475                 480

Leu Asn Val Phe Lys Glu Glu Ala Glu Gly Ile Ile Ile Pro Gly Lys
                485                 490                 495

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

Glu Ser Phe Cys Leu Thr Thr Lys Asn Thr Glu Asp Ile Thr Lys Val
        515                 520                 525

Ile Tyr Phe Leu Cys Lys Phe Leu Asp Gly Lys Glu Ile Asn Glu Leu
    530                 535                 540

Cys Cys Ala Met Met Asn Lys Leu Asp Gly Ile Ser Asp Leu Ile Glu
545                 550                 555                 560

Thr Ala Lys Gln Cys Gly Glu Asp Val Glu Phe Val Asp Gln Phe Lys
                565                 570                 575

Cys Leu Ser Lys Cys Ala Thr Met Ser Asn Gln Ile Arg Ile Val Lys
            580                 585                 590

Asn Ile Ser Arg Met Lys Lys Glu Met Thr Ile Asp Asn Asp Thr Ile
        595                 600                 605

Phe Leu Asp Ala Leu Glu Leu Leu Gly Arg Lys Ile Glu Lys Tyr Gln
    610                 615                 620

Lys Asp Lys Asn Gly Asp Tyr Val Lys Asp Glu Lys Gly Lys Lys Val
625                 630                 635                 640

Tyr Thr Lys Asp Tyr Asn Asn Phe Gln Asp Met Phe Phe Glu Gly Lys
                645                 650                 655

Asn His Arg Val Arg Asn Phe Val Ser Asn Asn Val Ile Lys Ser Lys
            660                 665                 670
```

```
Trp Phe Ser Tyr Val Val Arg Tyr Asn Lys Pro Ala Glu Cys Gln Ala
        675                 680                 685

Leu Met Arg Asn Ser Lys Leu Val Lys Phe Ala Leu Asp Glu Leu Pro
    690                 695                 700

Asp Ser Gln Ile Glu Lys Tyr Tyr Ile Ser Val Phe Gly Glu Lys Ser
705                 710                 715                 720

Ser Ser Ser Asn Glu Glu Met Arg Arg Glu Leu Leu Lys Lys Leu Cys
                725                 730                 735

Asp Phe Ser Val Arg Gly Phe Leu Asp Glu Ile Val Leu Leu Ser Glu
            740                 745                 750

Asp Glu Met Lys Gln Lys Asp Lys Phe Ser Glu Lys Glu Lys Lys Lys
        755                 760                 765

Ser Leu Ile Arg Leu Tyr Leu Thr Ile Val Tyr Leu Ile Thr Lys Ser
    770                 775                 780

Met Val Lys Ile Asn Thr Arg Phe Ser Ile Ala Cys Ala Thr Tyr Glu
785                 790                 795                 800

Arg Asp Tyr Ile Leu Leu Cys Gln Ser Glu Lys Ala Glu Arg Ala Trp
                805                 810                 815

Glu Lys Gly Ala Thr Ala Phe Ala Leu Thr Arg Lys Phe Leu Asn His
            820                 825                 830

Asp Lys Pro Thr Phe Glu Gln Tyr Tyr Thr Arg Glu Arg Glu Ile Ser
        835                 840                 845

Ala Met Pro Gln Glu Lys Arg Lys Glu Leu Arg Lys Glu Asn Asp Gln
    850                 855                 860

Leu Leu Lys Lys Thr His Tyr Ser Lys His Ala Tyr Cys Tyr Ile Val
865                 870                 875                 880
```

```
Asp Asn Val Asn Asn Leu Thr Gly Ala Val Ala Asn Asp Asn Gly Arg
                885                 890                 895


Gly Leu Pro Cys Leu Ser Glu Lys Asn Asp Asn Ala Asn Leu Phe Leu
            900                 905                 910


Glu Met Arg Asn Lys Ile Val His Leu Asn Val Val His Asp Met Val
        915                 920                 925


Lys Tyr Ile Asn Glu Ile Lys Asn Ile Thr Ser Tyr Tyr Ala Phe Phe
    930                 935                 940


Cys Tyr Val Leu Gln Arg Met Ile Ile Gly Asn Asn Ser Asn Glu Gln
945                 950                 955                 960


Asn Lys Phe Lys Ala Lys Tyr Ser Lys Thr Leu Gln Glu Phe Gly Thr
                965                 970                 975


Tyr Ser Lys Asp Leu Met Trp Val Leu Asn Leu Pro Phe Ala Tyr Asn
            980                 985                 990


Leu Pro Arg Tyr Lys Asn Leu Ser  Asn Glu Gln Leu Phe  Tyr Asp Glu
        995                 1000                1005


Glu Glu  Arg Met Glu Lys Ile  Val Gly Arg Lys Asn  Asp Ser Arg
    1010                1015                1020


<210>  161
<211>  36
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  DR

<400>  161
gtactaaagc ccgctagtat agacgggttc taagac                                  36


<210>  162
<211>  926
```

<212> PRT
<213> Artificial Sequence

<220>
<223> Metagenomic hit (no protein accession): contig emb|OGDF01008514.1| (human gut metagenome assembly)

<400> 162

```
Met Thr Glu Thr Lys Pro Lys Arg Glu Asp Ile Ala Lys Thr Pro Ala
1               5                  10                  15

Ala Lys Ser Arg Ser Lys Ala Ala Gly Leu Lys Ser Thr Phe Ala Val
            20                  25                  30

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

Pro Glu Lys Leu Ile Thr Glu Lys Ala Val Ser Glu Ile Asn Thr Val
    50                  55                  60

Lys Pro Arg Phe Ser Val Glu Lys Pro Ala Thr Ser Tyr Ser Ser Ser
65                  70                  75                  80

Phe Gly Ile Lys Ser His Ile Ser Ala Thr Ala Asp Asn Pro Leu Ala
                85                  90                  95

Gly Arg Ala Pro Val Gly Glu Asp Ala Ile His Ala Lys Glu Val Leu
            100                 105                 110

Glu Gln Arg Val Phe Gly Lys Thr Phe Ser Asp Asp Asn Ile His Ile
        115                 120                 125

Gln Leu Ile Tyr Asn Ile Leu Asp Ile Arg Lys Ile Leu Ser Thr Tyr
    130                 135                 140

Ala Asn Asn Val Val Phe Thr Ile Asn Ser Met Arg Arg Leu Asp Glu
145                 150                 155                 160

Tyr Asp Arg Glu Gln Asp Tyr Leu Gly Tyr Leu Tyr Thr Gly Asn Ser
                165                 170                 175
```

```
Tyr Glu Arg Leu Leu Asp Ile Ala Asp Lys Tyr Ala Val Asp Gly Glu
            180                 185                 190

Asp Trp Arg Asn Thr Ala Ala Gly Ile Ser Asn Asp Phe Glu Lys Lys
        195                 200                 205

Gln Phe Gln Thr Ile Asn Gly Phe Trp Asp Leu Leu Asp Met Ile Glu
    210                 215                 220

Pro Tyr Met Cys Tyr Phe Ser Glu Ala Phe Phe Cys Glu Thr Thr Val
225                 230                 235                 240

Lys Asp Pro Asp Ser Gly Arg Ile Val Pro Cys Leu Glu Gln Arg Ser
                245                 250                 255

Asp Gly Asp Ile Tyr Asn Ile Leu Arg Ile Leu Ser Ile Val Arg Gln
            260                 265                 270

Thr Cys Met His Asp Asn Ala Ser Met Arg Thr Val Met Phe Thr Leu
        275                 280                 285

Gly Gln Asn Ser Val Arg Asp Arg Lys Asn Gly Phe Asp Glu Leu Ala
    290                 295                 300

Glu Leu Leu Asp Tyr Leu Tyr Asp Glu Lys Ile Asp Ile Val Asn Arg
305                 310                 315                 320

Asp Phe Leu Arg Asn Gln Lys Asn Asn Ile Glu Leu Leu Ser Arg Ile
                325                 330                 335

Tyr Gly Ser Ser Ala Asp Ser Pro Glu Arg Asp Arg Leu Val Gln Asn
            340                 345                 350

Phe Tyr Asp Phe Arg Val Leu Ser Gln Asp Lys Asn Leu Gly Phe Ser
        355                 360                 365

Ile Lys Lys Leu Arg Glu Lys Leu Leu Asp Ser Pro Ala Leu Ser Val
    370                 375                 380
```

```
Val Arg Ser Lys Lys Tyr Asp Thr Met Arg Ser Lys Ile Tyr Ser Leu
385                 390                 395                 400


Ile Asp Phe Met Ile Tyr Arg Lys Phe Ser Glu Asn His Val Ala Val
                405                 410                 415


Asp Asp Phe Val Glu Glu Leu Arg Ser Leu Leu Thr Glu Asp Glu Lys
            420                 425                 430


Glu Ser Ala Tyr Ser Arg Trp Ala Glu Thr Leu Ile Asn Asp Gly Phe
        435                 440                 445


Ala Gln Glu Ile Leu Val Lys Leu Leu Pro Gln Thr Asp Pro Ala Val
    450                 455                 460


Ile Gly Lys Ile Lys Gly Lys Lys Leu Leu Asn Asp Ser Ile Ala Gly
465                 470                 475                 480


Ile Lys Leu Lys Lys Asp Ala Ser Phe Phe Thr Lys Ile Ile Asn Val
                485                 490                 495


Leu Cys Met Phe Gln Asp Gly Lys Glu Ile Asn Glu Leu Val Ser Ser
            500                 505                 510


Leu Val Asn Lys Phe Ala Asn Ile Gln Ser Phe Val Asp Val Met Arg
        515                 520                 525


Ser Gln Gly Ile Asp Ser Gly Phe Thr Ala Asp Tyr Ala Met Phe Ala
    530                 535                 540


Glu Ser Gly Arg Ile Ser Arg Glu Leu His Ile Leu Lys Gly Ile Ala
545                 550                 555                 560


Arg Met Gln His Ser Ile Ala Gly Leu Gly Asp Val Lys Ile Tyr Gly
                565                 570                 575


Ser Asp Asp Lys Phe His Gly Val Ser Arg Arg Val Tyr Thr Asp Ala
            580                 585                 590
```

```
Ala Tyr Ile Leu Gly Phe Gly Glu Arg Ser Glu Asp Asn Asp Gly Tyr
        595                 600                 605

Val Asp Asp Tyr Val Ser Ser Lys Leu Leu Gly Gly Ala Asp Lys Asn
    610                 615                 620

Leu Arg Asn Phe Ile Thr Asn Asn Val Ile Lys Asn Arg Arg Phe Leu
625                 630                 635                 640

Tyr Thr Val Arg Tyr Met Asn Pro Lys Arg Ala Lys Lys Leu Val Gln
                645                 650                 655

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

Ile Asp Arg Tyr Tyr Lys Ser Cys Ile Glu Lys Arg Ser Phe Asn Pro
        675                 680                 685

Asp Leu Asn Glu Lys Ile Ala Ala Leu Ser Glu Met Ile Thr Thr Leu
    690                 695                 700

Lys Ile Asp Asp Phe Glu Asp Val Lys Gln Asn Pro Glu Lys Asn Ala
705                 710                 715                 720

Asn Tyr Glu Ala Lys Lys Asn Gln Arg Ile Ser Lys Glu Arg Tyr Lys
                725                 730                 735

Ala Cys Ile Gly Leu Tyr Leu Thr Val Leu Tyr Leu Ile Cys Lys Asn
            740                 745                 750

Leu Val Lys Ile Asn Ala Arg Tyr Ser Ile Ala Ile Gly Cys Leu Glu
        755                 760                 765

Arg Asp Thr Gln Leu His Gly Val Asp Phe Lys Gly Ala Ala Tyr Met
    770                 775                 780

Thr Arg Asp Val Phe Ile Ala Lys Gly Trp Ile Asn Pro Lys Lys Pro
785                 790                 795                 800
```

Thr Val Lys Ser Ile Lys Glu Gln Tyr Ala Phe Leu Thr Pro Tyr Ile
805 810 815

Phe Thr Thr Tyr Arg Asn Met Ile Ala His Leu Ala Ala Val Thr Asn
820 825 830

Ala Tyr Lys Tyr Ile Pro Gln Met Asp Arg Phe Lys Ser Trp Phe His
835 840 845

Leu Tyr His Thr Val Ile Gln His Ser Leu Ile Gln Gln Tyr Glu Tyr
850 855 860

Asp Arg Asp Tyr Gly Arg Lys Gly Ala Pro Val Val Ser Glu Arg Val
865 870 875 880

Leu Gln Leu Leu Glu Gln Cys Arg Glu His Ser Asn Tyr Ser Arg Asp
885 890 895

Leu Leu His Ile Leu Asn Leu Pro Phe Gly Tyr Asn Leu Pro Arg Tyr
900 905 910

Leu Asn Leu Ser Ser Glu Lys Tyr Phe Asp Ala Asn Ala Ile
915 920 925

<210> 163
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 163
ctactacact ggtgcgaatt tgcactagtc taaaac 36

<210> 164
<211> 1030
<212> PRT
<213> Artificial Sequence

<220>
<223> Metagenomic hit (no protein accession): contig emb|OGPN01002610.1

(human gut metagenome assembly)

<400> 164

```
Met Ala Lys Lys Ile Thr Ala Lys Gln Lys Arg Glu Glu Lys Glu Arg
1               5                  10                  15

Leu Asn Lys Gln Lys Trp Ala Lys Gln Asp Thr Pro Val Val Pro Lys
            20                  25                  30

Ser Lys Thr Glu Glu Lys Pro Val Ala Ala Ser Asp Asp Lys Leu Leu
        35                  40                  45

Lys Thr Thr Gln Val Lys Lys Val Gln Thr Lys Ser Lys Ala Lys Ala
    50                  55                  60

Met Gly Leu Lys Thr Val Leu Ser Phe Asp Asp Lys Ile Ala Ile Ala
65                  70                  75                  80

Ser Phe Val Asn Asp Lys Lys Thr Lys Leu Pro His Ile Glu Arg Ile
                85                  90                  95

Thr Asp Lys Ser Gly Thr Thr Ile His Glu Asn Ala Arg Met Phe Asp
            100                 105                 110

Ser Ser Val Asp Glu Gln Asn Val Asn Ile Glu Lys Arg Met Thr Ile
        115                 120                 125

Glu Glu Lys Gln Asn Asp Gly Thr Phe Lys Lys Asp Glu Lys Asp Val
    130                 135                 140

Lys Ala Thr Ile Cys Asn Pro Tyr Phe Lys Thr Cys Gly Lys Asp Tyr
145                 150                 155                 160

Ile Gly Ile Lys Asp Val Ala Glu Lys Tyr Phe Phe Gly Lys Thr Phe
                165                 170                 175

Pro Asn Glu Asn Leu Arg Val Gln Ile Ala Tyr Asn Val Phe Asp Ile
            180                 185                 190
```

```
Gln Lys Ile Leu Gly Thr Tyr Val Asn Asn Ile Ile Tyr Ser Phe Tyr
        195                 200                 205

Asn Leu Arg Arg Asp Gly Lys Ser Asp Val Asp Ile Ile Gly Ser Leu
    210                 215                 220

Tyr Ala Phe Ala Asp Phe Asp Asn Gln Leu Lys Asp Lys Pro Ala Phe
225                 230                 235                 240

Arg Glu Ala Lys Asp Leu Leu Lys Asn Thr Glu Ala Tyr Phe Ser Tyr
                245                 250                 255

Phe Gly Asp Val Phe Lys Lys Ser Lys Lys Gly Lys Lys Asp Glu Asn
            260                 265                 270

Asn Glu Asp Tyr Glu Lys Asn Leu Arg His Asn Phe Asn Val Leu Arg
        275                 280                 285

Val Leu Ser Phe Leu Arg Gln Ile Cys Thr His Ala Tyr Val Lys Cys
    290                 295                 300

Thr Gly Gly Ala Lys Asn Asn Gly Asp Ser Thr Lys Val Glu Ala Glu
305                 310                 315                 320

Ser Leu Asp Ala Leu Phe Asn Ile Thr Glu Tyr Phe Ala Lys Thr Ala
                325                 330                 335

Pro Glu Leu Ser Lys Thr Ile Asn Glu Ile Tyr Lys Glu Gly Ile Asp
            340                 345                 350

Arg Ile Asn Asn Asp Phe Val Thr Asn Gly Lys Asn Asn Leu Tyr Ile
        355                 360                 365

Leu Ser Lys Val Tyr Pro Asp Met Gln Arg Asn Glu Leu Val Lys Lys
    370                 375                 380

Tyr Tyr Gln Phe Val Val Cys Lys Glu Gly Asn Asn Val Gly Ile Asn
385                 390                 395                 400
```

```
Thr Arg Lys Leu Lys Glu Ser Ile Ile Ser Gln His Pro Trp Ile Thr
                405                 410                 415

Thr Pro Gln Asp Asn Asn Lys Ala Asn Asp Tyr Glu Ser Cys Arg His
            420                 425                 430

Lys Leu Tyr Thr Ile Met Cys Phe Ile Leu Val Ala Glu Leu Asp Ala
        435                 440                 445

His Glu Ser Ile Arg Asp Asn Met Val Ala Glu Leu Arg Ala Asn Met
    450                 455                 460

Asp Gly Asp Asp Gly Arg Asp Ala Ile Tyr Glu Lys Tyr Ala Lys Asp
465                 470                 475                 480

Ile Tyr His Ile Val Lys Asp Lys Leu Leu Ala Met Gln Lys Val Phe
                485                 490                 495

Asp Glu Glu Leu Val Pro Val Lys Val Glu Gly Lys Asn Asp Pro Gln
            500                 505                 510

Gln Phe Thr His Gly Lys Leu Gly Lys Lys Glu Ile Glu Ser Phe Cys
        515                 520                 525

Leu Ser Asp Lys Asn Thr Ser Asp Ile Ala Lys Val Val Tyr Phe Leu
    530                 535                 540

Cys Asn Phe Leu Asp Gly Lys Glu Ile Asn Glu Leu Cys Cys Ala Met
545                 550                 555                 560

Met Asn Lys Phe Asp Gly Ile Gly Asp Leu Ile Asp Thr Ala Lys Gln
                565                 570                 575

Cys Gly Glu Glu Val Lys Phe Ile Glu Glu Phe Ala Cys Leu Ser Asn
            580                 585                 590

Cys Arg Lys Ile Thr Asn Asp Ile Arg Val Ala Lys Ser Ile Ser Lys
        595                 600                 605
```

```
Met Lys Asn Lys Val Asn Ile Asp Asn Asp Ile Ile Tyr Leu Asp Ala
    610                 615                 620

Ile Glu Leu Leu Gly Arg Lys Ile Glu Lys Tyr Gln Lys Asp Glu Asn
625                 630                 635                 640

Gly Lys Ile Leu Leu Gly Thr Asp Gly Lys Arg Leu Tyr Thr Gln Glu
                645                 650                 655

Tyr Lys Tyr Phe Asn Asp Met Phe Phe Asn Ala Gly Asn His Lys Val
            660                 665                 670

Arg Asn Phe Ile Ala Asn Asn Val Met Gln Ser Lys Trp Phe Phe Tyr
        675                 680                 685

Val Val Arg Tyr Asn Lys Pro Ala Glu Cys Gln Ile Ile Met Arg Asn
    690                 695                 700

Lys Thr Leu Val Lys Phe Thr Leu Asp Asp Leu Pro Asp Met Gln Ile
705                 710                 715                 720

Gln Arg Tyr Tyr Ser Ser Val Phe Gly Asp Asn Asn Met Pro Ala Val
                725                 730                 735

Asp Glu Met Arg Lys Arg Leu Leu Asp Lys Ile Asn Gln Phe Ser Val
            740                 745                 750

Arg Gly Phe Leu Asp Glu Leu Asp Glu Ile Val Leu Met Ser Asp Glu
        755                 760                 765

Glu Ser Lys Arg Asn Lys Ser Ser Glu Lys Glu Gln Lys Lys Ser Leu
    770                 775                 780

Ile Arg Leu Tyr Leu Thr Ile Ala Tyr Leu Ile Thr Lys Ser Met Val
785                 790                 795                 800

Lys Ile Asn Thr Arg Phe Ser Ile Ala Cys Ala Met Tyr Glu Arg Asp
                805                 810                 815
```

```
Tyr Ala Leu Leu Cys Gln Ser Glu Met Lys Gly Gly Pro Trp Asp Gly
            820                 825                 830

Gly Ala Gln Ala Leu Ala Val Thr Arg Lys Phe Leu Asn His Asp Arg
        835                 840                 845

Glu Val Phe Asp Arg Tyr Cys Ala Arg Glu Ala Glu Ile Ala Arg Leu
    850                 855                 860

Pro Ser Glu Glu Arg Lys Pro Leu Arg Lys Ala Asn Asp Lys Leu Leu
865                 870                 875                 880

Lys Gln Thr His Tyr Thr Asn His Ser Tyr Thr Tyr Ile Val Asn Asn
                885                 890                 895

Leu Asn Ser Phe Thr Asp Ile Asp Tyr Cys Ala Lys Asp Val Gly Leu
            900                 905                 910

Pro Ala Pro Asn Asp Lys Asn Asp Asn Ala Ser Ile Leu Gly Glu Met
        915                 920                 925

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

Ile Glu Glu Leu Lys Asp Ile Ser Ser Tyr Tyr Ala Phe Tyr Cys Tyr
945                 950                 955                 960

Val Leu Gln Arg Arg Leu Val Gly Lys Asp Pro Asn Cys Gln Asn Lys
                965                 970                 975

Phe Lys Ala Lys Tyr Ala Lys Glu Leu Asn Asp Tyr Gly Thr Tyr Asn
            980                 985                 990

Lys Asn Leu Met Trp Met Leu Asn  Leu Pro Phe Ala Tyr  Asn Leu Pro
        995                 1000                1005

Arg Tyr  Lys Asn Leu Ser Ser  Glu Phe Leu Phe Tyr  Asp Met Glu
    1010                1015                1020
```

Tyr Asn Lys Lys Asp Asp Glu
1025 1030

<210> 165
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 165
gaacccgtct atactagcgg gctttagtac aaaac 35

<210> 166
<211> 1001
<212> PRT
<213> Artificial Sequence

<220>
<223> Metagenic hit (no protein accession): contig NFIR01000008.1
(Eubacterium sp. An3, from chicken gut metagenome)

<400> 166

Met Ala Lys Lys Leu Arg Pro Lys Glu Leu Arg Glu Lys Arg Arg Met
1 5 10 15

Ala Glu Lys Glu Glu His Lys Lys Gln Glu Lys Leu Arg Lys Glu Gln
20 25 30

Glu Glu Leu Arg Lys Lys Gln Glu Lys Gln Arg Glu Asp Gln Lys Glu
35 40 45

Leu Glu Lys Ile Lys Lys Glu Glu Gly Gly Glu Gly Glu Lys Lys Lys
50 55 60

Ser Gly Ala Lys Ala Leu Gly Leu Lys Ser Thr Phe Ile Leu Asp Arg
65 70 75 80

Asp Glu Gln Lys Met Leu Met Thr Ser Phe Gly Arg Gly Asn Lys Ala
85 90 95

Val Arg Asp Lys Tyr Ile Ile Gly Asp Lys Val Ser Asp Ile Asp Asp

100 105 110

Ser Trp Glu Asn Lys Lys Ala Ala Leu Ser Val Glu Val Cys Gly Lys
115 120 125

Ser Phe Asn Ile Ser Lys Lys Glu Asn Asp Asp Cys Glu Pro Val Lys
130 135 140

Val Asn Asn Pro Val Leu Ser Gly Asn Lys Lys Asp Asp Asp Leu Ile
145 150 155 160

His Cys Arg Lys Asn Leu Glu Glu Met Tyr Phe Gly Gln Gln Phe Lys
165 170 175

Asp Asn Ile His Ile Gln Leu Ile Tyr Asn Ile Leu Asp Ile Glu Lys
180 185 190

Ile Leu Ala Val Gln Ile Asn Asn Ile Val Phe Ile Leu Asn Asn Leu
195 200 205

Leu Arg Trp Ser Gly Glu Glu Glu Phe Asp Leu Ile Gly Ser Leu Gly
210 215 220

Val Asn His Thr Tyr Glu Glu Phe Arg Gly Arg Asn Lys Asn Tyr Gly
225 230 235 240

Lys Phe Ser Glu Leu Ile Lys Gln Ser Gln Met Arg Tyr Phe Gly Ser
245 250 255

Thr Phe Cys Leu Phe Asn Glu Asn Glu Glu Arg Ile Thr Ser Glu Asn
260 265 270

Lys Lys Glu Trp Lys Arg Phe Glu Lys Lys Cys Tyr His Leu Leu Ala
275 280 285

Val Leu Gly Met Met Arg Gln Ala Thr Ala His Gly Asp Ser Lys Arg
290 295 300

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

305 310 315 320

Arg Gly Cys Arg Pro Glu Ala Arg Lys Glu Leu Asp Glu Leu Tyr Lys
325 330 335

Lys Lys Ile His Glu Met Asn Gln Gly Phe Leu Lys Asn Ser Lys Ser
340 345 350

Asp Ile Leu Met Leu Phe Arg Ile Tyr Asn Ala Glu Ser Lys Glu Ala
355 360 365

Lys Arg Lys Leu Ala Gln Glu Tyr Tyr Glu Phe Ile Met Leu Lys Ser
370 375 380

Tyr Lys Asn Thr Gly Phe Ser Ile Lys His Leu Arg Glu Thr Met Ile
385 390 395 400

Asp Lys Met Asp Glu Asp Lys Lys Glu Lys Leu Lys Asp Asp Lys Tyr
405 410 415

Asn Pro Ile Arg Arg Lys Ile Tyr Arg Ile Met Asp Phe Met Ile Tyr
420 425 430

Gln Tyr Tyr Gln Glu Pro Glu His Gln Glu Glu Ala Glu Glu Leu Val
435 440 445

Arg Lys Leu Arg Asn Ala Glu Ile Glu Ala Lys Lys Glu Leu Ala Tyr
450 455 460

Arg Lys Glu Ala Glu Lys Leu Lys Lys Glu Leu Glu Lys Ile Ile Phe
465 470 475 480

Asn Ser Val Leu Pro Ser Cys Asp Arg Ile Leu Ser Glu Met Asp Glu
485 490 495

Arg Arg Asn Lys Lys Val Asn Gln Glu Ser Ser Asp Thr Asp Lys Glu
500 505 510

Glu Pro Leu Asp Ser Glu Ile Ala Glu Gly Ile Thr Phe Ile Lys Glu

515 520 525

Thr Ala His Ser Phe Ser Glu Met Ile Tyr Leu Leu Thr Val Phe Leu
530 535 540

Asp Gly Lys Glu Ile Asn Ile Leu Leu Thr Gln Leu Ile His Cys Phe
545 550 555 560

Asp Asn Ile Ser Ser Phe Met Asp Thr Met Glu Glu Glu Asn Leu Leu
565 570 575

Thr Lys Leu Lys Glu Asp Tyr Glu Ile Phe Glu Glu Ser Lys Glu Ile
580 585 590

Ser Arg Glu Leu Arg Ile Ile Asn Ser Phe Ala Arg Met Thr Glu Pro
595 600 605

Val Pro Lys Thr Glu Arg Ile Met Phe Ile Glu Ala Ala Gln Ile Leu
610 615 620

Gly Tyr Ser Asn Gly Glu Lys Glu Leu Glu Gly Tyr Val Asp Ala Leu
625 630 635 640

Leu Asp Thr Lys Asn Lys Thr Asn Asp Lys Lys Lys Lys Gly Phe Val
645 650 655

Arg Tyr Ile Trp Asn Asn Val Ile Lys Ser Thr Arg Phe Arg Tyr Leu
660 665 670

Val Arg Tyr Ala Asp Pro Lys Lys Val Arg Ala Phe Ala Ala Asn Lys
675 680 685

Lys Val Val Ala Phe Val Leu Lys Asp Ile Pro Asp Asp Gln Ile Arg
690 695 700

Ala Tyr Tyr Asn Ser Cys Phe Arg Gln Asn Ser Asp Ser Ser Ser Asn
705 710 715 720

Asn Ser Asn Ala Ser Trp Asp Ala Asp Ser Asn Lys Arg Asp Ile Ser

725 730 735

Val Ser Asp Met Arg Lys Ala Leu Thr Glu Lys Ile Thr Gly Leu Asn
740 745 750

Phe Gly Asp Phe Glu Glu Glu Ser Lys Lys Gly Ile Arg Lys Glu Glu
755 760 765

Ser Asp Lys Asn Ile Ile Arg Leu Tyr Leu Thr Val Leu Tyr Leu Val
770 775 780

Gln Lys Asn Leu Ile Tyr Val Asn Ser Arg Tyr Phe Leu Ala Phe His
785 790 795 800

Cys Ala Glu Arg Asp Glu Met Leu Tyr Asn Gly Glu Thr Ile Asp Asn
805 810 815

Asn Lys Glu Lys Gly Ser Glu Lys Asp Trp Arg Lys Phe Ala Lys Gln
820 825 830

Phe Ile Met Glu His Ser Pro Lys Lys Lys Val Lys Asp Tyr Leu Ala
835 840 845

Lys Asn Phe Glu Tyr Ser Asn Lys Trp Ser Leu Lys Glu Phe Arg Asn
850 855 860

Ser Val Gln His Leu Asn Val Ile Arg Asp Ala His Lys Tyr Ile Lys
865 870 875 880

Tyr Ile Asn Asp Asn Lys Asp Val Gln Ser Tyr Phe Ala Leu Tyr His
885 890 895

Tyr Leu Val Gln Arg Tyr Ile Ser Glu Arg Ala Ala Asn Arg Thr Asp
900 905 910

Lys Glu Ser Leu Ser Glu Gly Arg Leu Gln Tyr Tyr Leu Ser Gln Val
915 920 925

Lys Glu Tyr Arg Thr Tyr Cys Lys Asp Phe Val Lys Ala Leu Asn Val

930 935 940

Pro Phe Ala Tyr Asn Leu Pro Arg Tyr Lys Asn Leu Ser Ile Asp Glu
945 950 955 960

Leu Phe Asp Arg Asn Asn Tyr Leu Pro Asn Lys Ala Lys Lys Trp Ile
965 970 975

Pro Glu Lys Lys Glu Asn Gly Glu Tyr Val Met Glu Asp Cys Gly Asn
980 985 990

Lys Asp Ala Gly Gln Val Glu Asn Ala
995 1000

<210> 167
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 167
gaactacact ctggctgaaa gtcagggtct aaaac 35

<210> 168
<211> 1006
<212> PRT
<213> Artificial Sequence

<220>
<223> Metagenomic hit (no protein accession): contig NFLV01000009.1
(Eubacterium sp. An11, from chicken gut metagenome)

<400> 168

Met Ser Lys Lys Gln Arg Pro Lys Asp Ile Arg Lys Arg Gln Glu Glu
1 5 10 15

Glu Lys Arg Glu Lys Tyr Lys Lys Gln Glu Glu Leu Arg Lys Lys Gln
20 25 30

Glu Glu Leu Arg Lys Glu Gln Glu Gln Arg Arg Glu Asp Gln Lys Glu
35 40 45

```
Leu Glu Lys Ile Lys Lys Glu Val Gly Glu Glu Gly Glu Lys Lys Lys
    50                  55                  60

Ser Arg Ala Lys Ala Leu Gly Leu Lys Ser Thr Phe Ile Leu Asp Arg
65                  70                  75                  80

Asp Glu Gln Lys Val Leu Met Thr Ser Phe Gly Gln Gly Asn Lys Ala
                85                  90                  95

Val Arg Asp Lys Tyr Ile Ile Gly Asp Lys Val Ser Asp Ile Asn Asp
            100                 105                 110

Asp Arg Lys Asn Lys Lys Ala Ala Leu Leu Val Glu Val Cys Gly Lys
        115                 120                 125

Ser Phe Asn Ile Ser Lys Lys Glu Asn Asp Asp Cys Asp Pro Val Lys
    130                 135                 140

Val Asn Asn Pro Val Val Ser Arg Asn Lys Lys Asp Asp Asp Leu Ile
145                 150                 155                 160

His Cys Arg Lys Lys Leu Glu Glu Leu Tyr Phe Gly Glu Gln Phe Lys
                165                 170                 175

Asp Asn Ile His Ile Gln Leu Ile Tyr Asn Ile Leu Asp Ile Glu Lys
            180                 185                 190

Ile Leu Ala Val Gln Val Asn Asn Ile Val Phe Ala Leu Asn Asn Leu
        195                 200                 205

Leu Ser Trp Ser Gly Glu Glu Lys Phe Asp Leu Ile Gly Tyr Leu Gly
    210                 215                 220

Val Asn Asp Thr Tyr Glu Lys Phe Arg Asp Ala Lys Gly Lys Arg Lys
225                 230                 235                 240

Gly Leu Tyr Glu Lys Phe Ser Thr Leu Ile Glu Lys Lys Arg Met Arg
                245                 250                 255
```

```
Tyr Phe Gly Ser Thr Phe Tyr Pro Leu Asn Glu Lys Gly Glu Glu Ile
            260                 265                 270

Thr Ser Asn Asp Lys Lys Glu Trp Glu Gln Phe Glu Lys Lys Cys Tyr
        275                 280                 285

His Leu Leu Ala Val Leu Gly Met Met Arg Gln Ala Thr Ala His Gly
    290                 295                 300

Asp Ser Lys Arg Arg Ala Glu Ile Tyr Lys Leu Gly Lys Glu Phe Asp
305                 310                 315                 320

Lys Ser Glu Ala Arg Gly Cys Arg Gln Glu Ala Arg Lys Glu Leu Asp
                325                 330                 335

Asp Leu Tyr Arg Lys Lys Ile His Glu Met Asn Gln Ser Phe Leu Lys
            340                 345                 350

Asn Ser Lys Arg Asp Ile Leu Met Leu Phe Arg Ile Tyr Asp Ala Glu
        355                 360                 365

Ser Lys Glu Ala Lys Arg Lys Leu Ala Gln Glu Tyr Tyr Glu Phe Ile
    370                 375                 380

Met Leu Lys Ser Tyr Lys Asn Thr Gly Phe Ser Ile Lys His Leu Arg
385                 390                 395                 400

Glu Thr Val Ile Asp Lys Met Asp Glu Asp Ile Lys Glu Lys Ile Lys
                405                 410                 415

Asp Asp Lys Tyr Asn Pro Ile Arg Arg Lys Leu Tyr Arg Ile Met Asp
            420                 425                 430

Phe Val Ile Tyr Gln Tyr Tyr Gln Glu Ser Glu Gln Gln Glu Glu Ala
        435                 440                 445

Met Glu Leu Val Arg Lys Leu Arg Asn Ala Glu Thr Lys Val Glu Lys
    450                 455                 460
```

```
Glu Leu Thr Tyr Arg Lys Glu Ala Glu Lys Leu Lys Glu Glu Leu Glu
465                 470                 475                 480

Lys Ile Ile Arg Asn Ser Ile Leu Ser Val Cys Asp Arg Ile Leu Ala
                485                 490                 495

Glu Met Asn Glu Lys Arg His Lys Lys Val Asn Gln Glu Ser Ser Asp
            500                 505                 510

Thr Asp Ser Glu Glu Pro Leu Asp Pro Glu Ile Ser Glu Gly Ile Thr
        515                 520                 525

Phe Ile Lys Glu Thr Ala His Ser Phe Ser Glu Met Ile Tyr Leu Leu
    530                 535                 540

Thr Val Phe Leu Asp Gly Lys Glu Ile Asn Ile Leu Leu Thr Gln Leu
545                 550                 555                 560

Ile His Cys Phe Asp Asn Ile Ser Ser Phe Met Asp Thr Met Lys Glu
                565                 570                 575

Glu Asn Leu Leu Thr Lys Leu Lys Glu Asp Tyr Glu Ile Phe Glu Glu
            580                 585                 590

Ser Lys Glu Ile Ser Lys Glu Leu Arg Ile Ile Asn Ser Phe Ala Arg
        595                 600                 605

Met Thr Glu Pro Val Pro Lys Thr Glu Lys Thr Met Phe Ile Asp Ala
    610                 615                 620

Ala Gln Ile Leu Gly Tyr Ser Asn Asp Glu Lys Glu Leu Glu Gly Tyr
625                 630                 635                 640

Val Asp Ala Leu Leu Asp Thr Lys Asn Lys Thr Lys Asp Lys Glu Arg
                645                 650                 655

Lys Gly Phe Glu Lys Tyr Ile Trp Asn Asn Val Ile Lys Ser Thr Arg
            660                 665                 670
```

```
Phe Arg Tyr Leu Val Arg Tyr Ala Asp Pro Lys Lys Val Arg Ala Phe
        675                 680                 685

Ala Ala Asn Lys Lys Val Val Ala Phe Val Leu Lys Asp Ile Pro Asp
    690                 695                 700

Glu Gln Ile Lys Ala Tyr Tyr Asn Ser Cys Phe Ser Gln Asn Ser Asp
705                 710                 715                 720

Ser Ser Ser Asn Met Ser Ile Ala Phe Gln Asp Gly Asp Ser Asn Lys
                725                 730                 735

Lys Gly Thr Ser Val His Asp Met Met Arg Lys Ala Leu Thr Glu Lys
            740                 745                 750

Ile Thr Gly Leu Asn Phe Gly Asp Phe Glu Glu Glu Ser Lys Lys Gly
        755                 760                 765

Ile Arg Arg Glu Glu Ser Asp Lys Asn Ile Ile Arg Leu Tyr Leu Thr
    770                 775                 780

Val Leu Tyr Leu Val Gln Lys Asn Leu Ile Tyr Val Asn Ser Arg Tyr
785                 790                 795                 800

Phe Leu Ala Phe His Cys Ala Glu Arg Asp Glu Val Leu Tyr Asn Gly
                805                 810                 815

Glu Thr Ile Asp Asn Asn Lys Glu Lys Gly Ser Glu Lys Asp Trp Lys
            820                 825                 830

Lys Phe Ala Lys Glu Phe Ile Ile Glu His Pro Pro Lys Lys Lys Val
        835                 840                 845

Lys Asp Tyr Leu Ala Lys Asn Phe Glu Tyr Ser Asn Lys Trp Ser Leu
    850                 855                 860

Arg Val Phe Arg Asn Ser Val Gln His Leu Asn Val Ile Arg Asp Ala
865                 870                 875                 880
```

Tyr Lys Tyr Ile Lys Cys Ile Asp Asp Asn Lys Asp Val Gln Ser Tyr
885 890 895

Phe Ala Leu Tyr His Tyr Leu Val Gln Arg Tyr Ile Ser Glu Met Ala
900 905 910

Glu Asn Leu Thr Asp Lys Gly Glu Leu Ser Glu Gly Arg Leu Gln Tyr
915 920 925

Tyr Leu Ser Gln Val Glu Asn Tyr Arg Thr Tyr Cys Lys Asp Phe Val
930 935 940

Lys Ala Leu Asn Val Pro Phe Ala Tyr Asn Leu Pro Arg Tyr Lys Asn
945 950 955 960

Leu Ser Ile Asp Glu Leu Phe Asp Arg Asn Asn Tyr Leu Pro Asn Lys
965 970 975

Ala Lys Lys Trp Ile Ser Glu Lys Lys Glu Asn Gly Glu Tyr Val Met
980 985 990

Glu Asp Cys Gly Asn Lys Gly Ala Gly Gln Val Glu Asn Ala
995 1000 1005

<210> 169
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 169
gaactacacc ctggctgaaa gtcagggtct aaaac 35

<210> 170
<211> 1027
<212> PRT
<213> Artificial Sequence

<220>

<223> Metagenomic hit (no protein accession): from contig emb|OBLI01020244 and emb|OBLI01038679 (from pig gut metagenome)

<400> 170

```
Met Ala Lys Lys Ile Thr Ala Lys Gln Arg Arg Glu Glu Arg Glu Arg
1               5                  10                  15

Gln Asn Lys Gln Lys Trp Ala Lys Lys Gln Ala Asp Ala Thr Ala Val
            20                  25                  30

Phe Glu Cys Glu Ala Asp Ile Lys Pro Ala Asp Ser Lys Asp Glu Asp
        35                  40                  45

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

Ala Met Gly Leu Lys Thr Val Leu Gly Phe Asp Asn Lys Ile Ala Ile
65                  70                  75                  80

Ala Ser Phe Met Ser Ser Lys Asp Ser Lys Ser Ser His Ile Glu Arg
                85                  90                  95

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

Asp Ser Asn Val Asp Glu Cys Ser Ile Asn Leu Glu Lys Arg Met Thr
        115                 120                 125

Val Glu Glu Arg Gln Lys Asp Gly Thr Ile Lys Lys Asp Glu Lys Asp
    130                 135                 140

Val Lys Ser Thr Ile Cys Asn Pro Tyr Ser Asn Glu Cys Gly Lys Asp
145                 150                 155                 160

Tyr Ile Gly Ile Lys Ser Val Ala Glu Glu Leu Phe Phe Gly Arg Thr
                165                 170                 175

Phe Pro Asn Asp Asn Leu Arg Val Gln Ile Ala Tyr Asn Ile Phe Asp
            180                 185                 190
```

```
Ile Gln Lys Ile Leu Gly Thr Tyr Ile Asn Asn Ile Ile Tyr Ser Phe
        195                 200                 205

Tyr Asn Leu Ser Arg Asp Glu Ser Gln Ser Asp Asn Asp Val Ile Gly
    210                 215                 220

Thr Leu Tyr Met Leu Lys Asp Phe Asp Gly Gln Lys Glu Thr Asp Thr
225                 230                 235                 240

Phe Arg Gln Ala Arg Ala Leu Leu Glu Arg Thr Glu Ala Tyr Tyr Ser
                245                 250                 255

Tyr Phe Asp Asn Val Phe Lys Lys Ile Asp Lys Asn Lys Lys Lys Ser
            260                 265                 270

Asp Asp Cys Lys Arg Glu Arg Asn Glu Ile Leu Arg Tyr Asn Phe Asn
        275                 280                 285

Val Leu Arg Val Leu Ser Phe Leu Arg Gln Ile Cys Ala His Ala Gln
    290                 295                 300

Val Lys Ile Ser Asn Glu His Asp Arg Glu Lys Gly Gly Gly Leu Val
305                 310                 315                 320

Asp Ser Leu Asp Ala Leu Phe Asn Ile Ser Arg Phe Phe Asp Ala Val
                325                 330                 335

Ala Pro Glu Leu Asn Glu Val Ile Asn Ser Val Tyr Ser Lys Gly Ile
            340                 345                 350

Asp Asp Ile Asn Asp Asn Phe Val Lys Asn Gly Lys Asn Asn Phe Tyr
        355                 360                 365

Ile Leu Ser Lys Ile Tyr Pro Glu Val Ala Arg Glu Asp Leu Leu Arg
    370                 375                 380

Glu Tyr Tyr Tyr Phe Val Val Ser Lys Glu Gly Asn Asn Ile Gly Ile
385                 390                 395                 400
```

```
Ser Thr Lys Lys Leu Lys Glu Ala Ile Ile Val Gln Asp Met Ser Tyr
                405                 410                 415

Ile Lys Ser Glu Asp Tyr Asp Thr Tyr Arg Asn Lys Leu Tyr Thr Val
            420                 425                 430

Leu Cys Phe Ile Leu Val Lys Glu Leu Asn Glu Arg Thr Thr Ile Arg
        435                 440                 445

Glu Gln Met Val Ala Asp Leu Arg Ala Asn Met Asn Gly Asp Ile Gly
    450                 455                 460

Arg Glu Asp Ile Tyr Ser Lys Tyr Ala Lys Ile Ile Tyr Ala Gln Val
465                 470                 475                 480

Lys Pro Arg Phe Asp Thr Met Lys Ser Ala Phe Glu Glu Glu Ala Lys
                485                 490                 495

Asp Val Ile Val Pro Asp Lys Lys Lys Pro Val Lys Phe Ser His Gly
            500                 505                 510

Lys Leu Asp Lys Asn Glu Ile Glu Arg Phe Cys Ile Thr Ser Ala Asn
        515                 520                 525

Thr Asp Ser Val Ala Lys Ile Ile Tyr Phe Leu Cys Lys Phe Leu Asp
    530                 535                 540

Gly Lys Glu Ile Asn Glu Leu Cys Cys Ala Met Met Asn Lys Leu Asp
545                 550                 555                 560

Gly Ile Asn Asp Leu Ile Glu Thr Ala Glu Gln Cys Gly Ala Lys Val
                565                 570                 575

Glu Phe Val Asp Lys Phe Ser Val Leu Ser Asn Cys Glu Thr Ile Ser
            580                 585                 590

Asp Gln Ile Arg Ile Val Lys Ser Ile Ser Lys Met Lys Lys Glu Ile
        595                 600                 605
```

```
Ala Ile Asp Asn Asp Thr Ile Phe Leu Asp Ala Leu Glu Leu Leu Gly
    610                 615                 620

Arg Lys Ile Asp Lys Tyr Lys Lys Asp Ala Thr Gly Lys Tyr Leu Lys
625                 630                 635                 640

Asp Glu Asn Gly Lys Tyr Leu Tyr Ser Lys Glu Tyr Asp Asp Phe Gln
                645                 650                 655

Tyr Met Phe Phe Lys Asp Ser His Arg Val Arg Asn Phe Ile Ser Asn
            660                 665                 670

Ser Val Ile Lys Ser Lys Trp Phe Ser Tyr Ile Val Arg Tyr Asn Gln
        675                 680                 685

Pro Ser Glu Cys Arg Ala Ile Met Lys Asn Lys Thr Leu Val Lys Phe
    690                 695                 700

Ala Leu Asp Glu Leu Pro Asp Leu Gln Ile Gln Arg Tyr Phe Val Ala
705                 710                 715                 720

Leu Tyr Gly Asp Glu Asp Leu Pro Ser Tyr Gly Glu Met Arg Lys Ile
                725                 730                 735

Leu Leu Lys Lys Leu His Asp Phe Ser Ile Lys Gly Phe Leu Asp Glu
            740                 745                 750

Ile Val Leu Leu Ser Asp Leu Asp Met Glu Ser Gln Asp Lys Tyr Cys
        755                 760                 765

Glu Lys Glu Gln Lys Lys Ser Leu Phe Arg Leu Tyr Leu Thr Ile Ala
    770                 775                 780

Tyr Leu Ile Thr Lys Ser Met Val Lys Ile Asn Thr Arg Phe Ser Ile
785                 790                 795                 800

Ala Cys Ala Thr Tyr Glu Arg Asp Tyr Ala Leu Leu Cys Ala Ser Asn
                805                 810                 815
```

```
Lys Gln Glu Arg Ala Trp Ser Ser Gly Ala Thr Ala Leu Ala Leu Thr
            820                 825                 830

Arg Arg Phe Leu Asn Gln Asp Lys Leu Ile Phe Glu Lys His Tyr Ala
        835                 840                 845

Arg Glu Gly Glu Ile Ser Lys Leu Pro Lys Glu Glu Arg Lys Ala Met
    850                 855                 860

Arg Lys Val Asn Asp Gln Leu Leu Lys Arg Thr His Phe Ser Lys His
865                 870                 875                 880

Ser Tyr Cys Tyr Ile Val Asp Asn Val Asn Arg Leu Thr Gly Gly Glu
                885                 890                 895

Cys Arg Thr Asp Lys Arg Val Leu Pro Val Leu Asn Glu Lys Asn Asp
            900                 905                 910

Asn Ala Gly Ile Leu Leu Asp Phe Arg Lys Thr Ile Ala His Leu Asn
        915                 920                 925

Val Val His Lys Met Val Asp Tyr Val Asp Glu Ile Lys Gly Ile Thr
    930                 935                 940

Ser Tyr Tyr Ala Phe Phe Cys Tyr Val Leu Gln Arg Met Leu Val Gly
945                 950                 955                 960

Asn Asn Leu Asn Glu Lys Asn Ala Ile Lys Glu Lys Tyr Ser Ala Thr
                965                 970                 975

Val Lys Ser Phe Gly Thr Tyr Ser Lys Asp Phe Met Trp Leu Ile Asn
            980                 985                 990

Leu Pro Phe Ala Tyr Asn Leu Pro  Arg Tyr Lys Asn Leu  Ser Asn Glu
        995                 1000                1005

Gln Leu  Phe Tyr Asp Glu Glu  Glu Arg Asn Glu Thr  Glu Glu Gln
    1010                1015                1020
```

Ile Asp Arg Leu
    1025

<210> 171
<211> 39
<212> PRT
<213> Artificial Sequence

<220>
<223> Motif 1

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> Xaa can be Ile, Met, or Val

<220>
<221> misc_feature
<222> (4)..(4)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (5)..(5)
<223> Xaa can be Asn or Ser

<220>
<221> misc_feature
<222> (6)..(7)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (9)..(9)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (11)..(11)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (12)..(12)
<223> Xaa can be Ala, Ile, Leu, or Val

<220>
<221> MISC_FEATURE
<222> (13)..(13)

<223> Xaa can be Lys or Arg

<220>
<221> MISC_FEATURE
<222> (14)..(14)
<223> Xaa can be Phe, His, or Tyr

<220>
<221> misc_feature
<222> (15)..(16)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (17)..(17)
<223> Xaa can be Ala, Pro, or Thr

<220>
<221> misc_feature
<222> (18)..(25)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (26)..(26)
<223> Xaa can be Cys or Asn

<220>
<221> misc_feature
<222> (27)..(31)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (32)..(32)
<223> Xaa can be Phe or Leu

<220>
<221> misc_feature
<222> (33)..(33)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (35)..(36)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (37)..(37)
<223> Xaa can be Ile, Leu, Met, or Val

<220>
<221> MISC_FEATURE
<222> (38)..(38)
<223> Xaa can be Ala, Gly, or Pro

<220>
<221> MISC_FEATURE
<222> (39)..(39)
<223> Xaa can be Asp, Glu, or Lys

<400> 171

Asn Val Xaa Xaa Xaa Xaa Xaa Phe Xaa Tyr Xaa Xaa Xaa Xaa Xaa Xaa
1 5 10 15

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
20 25 30

Xaa Leu Xaa Xaa Xaa Xaa Xaa
35

<210> 172
<211> 33
<212> PRT
<213> Artificial Sequence

<220>
<223> Motif 2

<220>
<221> misc_feature
<222> (2)..(2)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> Xaa can be Phe or Leu

<220>
<221> misc_feature
<222> (4)..(4)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (5)..(5)
<223> Xaa can be Ile or Val

<220>
<221> misc_feature
<222> (6)..(6)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (8)..(10)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (12)..(12)
<223> Xaa can be Asn or Ser

<220>
<221> MISC_FEATURE
<222> (13)..(13)
<223> Xaa can be Ile, Leu, or Met

<220>
<221> MISC_FEATURE
<222> (14)..(14)
<223> Xaa can be Ile or Val

<220>
<221> misc_feature
<222> (15)..(16)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (18)..(18)
<223> Xaa can be Ala, Gly, Ser, or Thr

<220>
<221> misc_feature
<222> (19)..(19)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (20)..(20)
<223> Xaa can be Phe or Tyr

<220>
<221> misc_feature
<222> (21)..(22)
<223> Xaa can be any naturally occurring amino acid

<220>

<221> MISC_FEATURE
<222> (23)..(23)
<223> Xaa can be Ala or Gly

<220>
<221> misc_feature
<222> (24)..(27)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (28)..(28)
<223> Xaa can be Asp or Glu

<220>
<221> misc_feature
<222> (31)..(32)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (33)..(33)
<223> Xaa can be Phe, Leu, or Met

<400> 172

Leu Xaa Xaa Xaa Xaa Xaa Tyr Xaa Xaa Xaa Lys Xaa Xaa Xaa Xaa Xaa
1 5 10 15

Asn Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Arg Asp Xaa Xaa
20 25 30

Xaa

<210> 173
<211> 19
<212> PRT
<213> Artificial Sequence

<220>
<223> Motif 3

<220>
<221> misc_feature
<222> (5)..(5)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (6)..(6)
<223> Xaa can be Leu or Met

<220>
<221> misc_feature
<222> (7)..(7)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (8)..(8)
<223> Xaa can be Cys, Ser, or Thr

<220>
<221> MISC_FEATURE
<222> (9)..(9)
<223> Xaa can be Ala, Asn, Gln, Ser, or Thr

<220>
<221> MISC_FEATURE
<222> (10)..(10)
<223> Xaa can be Ile, Leu, or Met

<220>
<221> MISC_FEATURE
<222> (11)..(11)
<223> Xaa can be Ile, Met, or Val

<220>
<221> misc_feature
<222> (12)..(13)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (14)..(14)
<223> Xaa can be Phe or Leu

<220>
<221> MISC_FEATURE
<222> (15)..(15)
<223> Xaa can be Ala, Asp, Glu, or Gln

<220>
<221> MISC_FEATURE
<222> (16)..(16)
<223> Xaa can be Gly, Asn, or Val

<220>
<221> misc_feature

<222> (18)..(18)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (19)..(19)
<223> Xaa can be Asp, Glu, Asn, or Ser

<400> 173

Lys Glu Ile Asn Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1 5 10 15

Ile Xaa Xaa

<210> 174
<211> 18
<212> PRT
<213> Artificial Sequence

<220>
<223> Motif 4

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> Xaa can be Ala or Gly

<220>
<221> MISC_FEATURE
<222> (5)..(5)
<223> Xaa can be Cys, Asp, Asn, or Val

<220>
<221> misc_feature
<222> (6)..(6)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (7)..(7)
<223> Xaa can be Ala, Pro, or Val

<220>
<221> MISC_FEATURE
<222> (9)..(9)
<223> Xaa can be Phe or Tyr

<220>
<221> misc_feature
<222> (10)..(10)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (11)..(11)
<223> Xaa can be Asp or Asn

<220>
<221> MISC_FEATURE
<222> (13)..(13)
<223> Xaa can be Ser or Thr

<220>
<221> misc_feature
<222> (14)..(17)
<223> Xaa can be any naturally occurring amino acid

<400> 174

Pro Phe Xaa Tyr Xaa Xaa Xaa Arg Xaa Xaa Xaa Leu Xaa Xaa Xaa Xaa
1 5 10 15

Xaa Phe

<210> 175
<211> 979
<212> PRT
<213> Artificial Sequence

<220>
<223> Metagenomic hit (no protein accession): contig OJMM01002900 human gut metagenome sequence

<400> 175

Met Gly Lys His Leu Asn Ala Lys Gln Arg Lys Leu Glu Lys Lys Leu
1 5 10 15

Lys Asn Gln Gln Lys Asp Met Met Tyr Thr Lys Ser Thr Asp Ala Val
20 25 30

Ser Val Pro Thr Leu Lys Ala Ala Pro Thr Lys Ala Glu Met Ser Gln
35 40 45

```
Asp Thr Ala Glu Ala Ser Thr Leu Ile Thr Pro Gly Thr Leu Lys Thr
    50                  55                  60

Lys Ala Lys Ala Met Gly Leu Lys Ser Thr Leu Val Phe Asp Asp Lys
65                  70                  75                  80

Ile Val Val Thr Ser Phe Leu Asn Ser Lys Thr Glu Glu Asn Glu Lys
                85                  90                  95

Cys Ala His Ile Glu Lys Ile Ala Asp Cys Asn Gly Gln Thr Ile Val
            100                 105                 110

Glu Arg Pro Arg Met Phe Asn Thr Ser Ile Asn Ala Lys Lys Val Asp
        115                 120                 125

Leu Ser Lys Asp Asn Asp Glu Thr Asn Tyr Pro Asn Pro Ala Phe Glu
    130                 135                 140

Asp Cys Gly Arg Asp Tyr Ile Asn Val Lys Ser Ala Leu Glu Lys Arg
145                 150                 155                 160

Val Phe Gly Lys Thr Tyr Asn Lys Asp Asn Leu His Val Gln Ile Ala
                165                 170                 175

Tyr Asn Ile Phe Asp Ile Lys Lys Ile Ile Gly Ala Tyr Ile Asn Asn
            180                 185                 190

Ile Ile Tyr Ile Phe Tyr Asn Leu Gly Arg Glu Glu Tyr Asp Ala Lys
        195                 200                 205

Lys Asp Ile Ile Gly Thr Gln Asp Ser Ala Tyr Ile Ser Lys Ile Leu
    210                 215                 220

Asn Asn Thr Ser Ala Tyr Phe Thr Tyr Phe Asp Gly Val Phe Lys Gln
225                 230                 235                 240

Ile Thr Asp Arg Asp Ser Asn Lys Asp Arg Glu Ile Lys Asn Ser Tyr
                245                 250                 255
```

```
Asn Ala Leu Val Leu Lys Val Leu Tyr Tyr Leu Arg Gln Phe Cys Met
            260                 265                 270

His Gly Asn Thr Tyr Thr Lys Arg Asn Glu Glu Ser Phe Leu Ser Asp
        275                 280                 285

Thr Ala Leu Tyr Asn Ala Lys Glu Phe Phe Ala Lys Ala Asp Pro Gln
    290                 295                 300

Ile Asn Glu Leu Ile Asp Ala Val Tyr Ala Asp Gly Ile Lys Thr Ile
305                 310                 315                 320

Asn Ser Asp Phe Met Ala His Ala Lys Asn Asn Met Tyr Ile Ile Cys
                325                 330                 335

Glu Val Tyr Lys Asn Glu Ala Glu Asp Ser Leu Met Lys Glu Tyr Tyr
            340                 345                 350

Asp Phe Val Val Arg Lys Glu Gly Asn Asn Leu Gly Phe Asn Thr Arg
        355                 360                 365

Gln Leu Arg Glu Ile Leu Ile Asp Lys Tyr Val Gly Asn Leu Arg Gly
    370                 375                 380

Lys Lys Tyr Asn Thr Phe Arg Asn Lys Leu Tyr Thr Val Leu Gly Phe
385                 390                 395                 400

Ile Leu Val Lys Glu Ile Lys Arg Asn Pro Lys Ile Gln Asp Ser Phe
                405                 410                 415

Ile Ala Lys Leu Arg Ala Asn Gln Asn Gly Asp Glu Gly Lys Leu Asn
            420                 425                 430

Ile Tyr Asn Glu Phe Ala Pro Lys Ile Trp Ser Val Val Ser Ser Lys
        435                 440                 445

Leu Asn Ser Ala Ile Thr Cys Phe Asp Glu Glu Ser Leu Ser Lys Phe
    450                 455                 460
```

```
Lys Gly Tyr Lys Asp Ile Asp Glu Ser Leu Ile Ser Arg Tyr Gly Ile
465                 470                 475                 480

Thr Val Ala Asn Thr Asp Thr Leu Val Lys Ile Leu Tyr Phe Leu Cys
                485                 490                 495

Lys Phe Leu Asp Gly Lys Glu Ile Asn Glu Leu Cys Cys Ala Met Ile
            500                 505                 510

Asn Lys Phe Asp Asn Ile Asn Asp Leu Ile Lys Thr Ala Ala Gln Cys
        515                 520                 525

Gly Glu Asp Ile Glu Phe Val Lys Glu Tyr Lys Leu Phe Ile Asn Ser
    530                 535                 540

Asn Asp Leu Ser Asp Gln Ile Arg Ile Val Lys Ser Ile Ser Lys Met
545                 550                 555                 560

Lys Pro Glu Leu Ser Lys Ile Gly Glu Ala Leu Ile Leu Asp Ala Ile
                565                 570                 575

Asp Ile Leu Gly Tyr Lys Ile Asn Lys Tyr Lys Tyr Asp Ala Ala Gly
            580                 585                 590

Asn Arg Leu Val Asp Ser Asn Asn Lys Pro Val Tyr Ser Glu Glu Tyr
        595                 600                 605

Cys Ala Phe Lys Lys Asp Phe Phe Glu Thr Cys Glu Leu Asp Glu Phe
    610                 615                 620

Gly Arg Val Lys Tyr Asn Lys Lys Gly Lys Pro Val Ile Asn His Arg
625                 630                 635                 640

Arg Arg Asn Phe Ile Ile Asn Asn Val Leu Ser Ser Lys Trp Phe Phe
                645                 650                 655

Tyr Val Ala Lys Tyr Asn Arg Pro Ser Glu Cys Gln Lys Phe Met Lys
            660                 665                 670
```

```
Ser Lys Lys Leu Ile Ala Leu Val Leu Lys Asp Val Pro Glu Thr Gln
        675                 680                 685

Ile Ala Arg Tyr Tyr Gln Ser Val Thr Gly Gly Arg Thr Gln Ala Asn
    690                 695                 700

Ser Glu Ala Met Arg Met Thr Leu Ile Lys Leu Leu His Glu Phe Ser
705                 710                 715                 720

Ile Lys Asn Val Leu Ser Asp Val Gly Thr Met Thr Ala Ser Glu Asn
                725                 730                 735

Lys Arg Gln Ile Glu Asn Ser Arg Lys Glu Arg Met Lys Ala Ile Val
            740                 745                 750

Lys Leu Tyr Leu Thr Val Val Tyr Leu Ile Ala Lys Ser Leu Val Lys
        755                 760                 765

Val Asn Thr Arg Phe Ser Ile Ala Phe Ser Ala Tyr Glu Arg Asp Val
    770                 775                 780

Ser Leu Leu Ala Asp Glu Asn Glu Leu Ile Ala Leu Ala Asn Asn Glu
785                 790                 795                 800

Asp Asp Lys Trp Lys Lys Gly Asn Tyr Val Phe Ala Leu Thr Lys His
                805                 810                 815

Phe Trp Asp Asn Asp Glu Pro Tyr Phe Asp Lys Tyr Asn Asn Ala Leu
            820                 825                 830

Gln Gln Ile Arg Ser Ile Val Asp Pro Asn Glu Arg Arg Leu Ala Tyr
        835                 840                 845

Arg Ala Asn Asp Lys Val Val Lys His Thr His Phe Asn Leu His Ser
    850                 855                 860

Tyr Lys Tyr Val Lys His Asn Tyr Glu Glu Ile Ser Lys Ala Ser Lys
865                 870                 875                 880
```

Ile Ile Thr Ala Tyr Arg Asn Asn Val Gln His Leu Asn Val Met Asn
885 890 895

Ser Ile Thr Lys Tyr Leu Gly Asp Ile Ser Glu Val Thr Ser Tyr Tyr
900 905 910

Ser Leu Tyr Cys Tyr Thr Leu Gln Arg Leu Leu Leu Asp Asp Asn Asn
915 920 925

Asn Asp Lys Phe Ala Ser Ile Lys Gly Asn Leu Arg Lys Phe Gly Ile
930 935 940

Tyr Asn Lys Asp Phe Met Trp Leu Leu Asn Ile Pro Phe Ala Tyr Asn
945 950 955 960

Leu Pro Arg Tyr Lys Asn Leu Ser Asn Glu Glu Ile Phe Tyr Asp Glu
965 970 975

Leu Gln Lys

<210> 176
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Direct repeat

<400> 176
gtacaatagc cctctcgtag ttgagggctc tgagac 36

<210> 177
<211> 1009
<212> PRT
<213> Artificial Sequence

<220>
<223> Metagenomic hit (no protein accession): contig ODAI011611274.1 gut metagenome sequence

<400> 177

```
Met Lys Lys Lys Ile Ser Leu Lys Glu Gln Arg Asn Thr Lys Lys Ala
1               5                  10                  15

Glu Asn Lys Leu Lys Tyr Gln Lys Ala Gln Ala Glu Arg Ala Ala Ala
            20                  25                  30

Ala Gln Gln Thr Ala Ala Gly Ala Glu Ser Glu Glu Asn Pro Cys Phe
        35                  40                  45

Asp Val Val Lys Asp Thr Lys Arg Lys Ala Leu Asn Pro Leu His Val
    50                  55                  60

Glu Ile Glu Ala Pro Ser Ala Lys Lys Ser Ser Val Lys Ala Asn Gly
65                  70                  75                  80

Leu Lys Ser Leu Leu Leu Thr Asp Gly Lys Thr Val Met Thr Ser Phe
                85                  90                  95

Gly Arg Gly Ser Glu Ala Asn Val Glu Lys Arg Phe Asp Glu Thr Gly
            100                 105                 110

Thr Lys Thr Phe Asp Arg Asp Pro Glu Leu Phe Ser Ala Lys Pro Leu
        115                 120                 125

Glu Thr Gly Tyr Arg Ile Gln Arg Phe Asn Ala Ser Pro Lys Asp Ala
    130                 135                 140

Gly Leu Ala Tyr Arg Pro Ala Gly Val Arg Pro Asp Gln Ile Gly Ala
145                 150                 155                 160

Lys Ala Ala Leu Glu Lys Arg Tyr Phe Gly Lys Glu Thr Pro Gly Asp
                165                 170                 175

Asn Ile His Val Gln Ile Ala Tyr Gln Ile Gln Asp Ile Glu Lys Leu
            180                 185                 190

Leu Ala Val Tyr Ile Ser Asn Ile Ile Tyr Ala Val Asn Asn Val Thr
        195                 200                 205
```

```
Gly Val Ser Ala Met Lys Asp Ser Lys Gly Arg Pro Val Asp Leu Leu
    210                 215                 220

Gly Asp Tyr Gly Ile Leu Gly Glu Glu Gly Leu Thr Lys Arg Leu Gln
225                 230                 235                 240

Arg Ile Pro Glu Gln Ala Asp Glu Glu Ala Lys Ala Leu Gln Ala Phe
                245                 250                 255

Leu Cys Ser Glu Arg Leu Ser Tyr Phe Gly Lys Glu Phe Cys Leu Val
            260                 265                 270

Arg Asn Ser Pro Lys Gln Pro Asp Lys Glu Glu Lys Arg Gln Tyr Lys
        275                 280                 285

Leu Met Arg Val Leu Cys Leu Leu Gly Glu Leu Arg Gln Phe Leu Val
    290                 295                 300

His Gly Lys Lys Lys Glu Lys Glu Phe Ala Trp Leu Tyr Arg Leu Asp
305                 310                 315                 320

Arg Gln Leu Ser Gln Glu Tyr Arg Lys Leu Leu Gly Glu Phe Tyr Asp
                325                 330                 335

Ala Gln Val Asp Lys Val Asn Lys Ser Phe Leu Thr Asn Ser Thr Val
            340                 345                 350

Asn Leu Glu Val Leu Phe Arg Ala Leu Lys Thr Gly Thr Asp Pro Glu
        355                 360                 365

Arg Lys Thr Val Thr Gln Glu Tyr Tyr Gln Phe Thr Val Arg Lys Glu
    370                 375                 380

Asp Gly Asn Leu Gly Phe Ser Leu Lys Thr Leu Arg Glu Ile Leu Leu
385                 390                 395                 400

Ser Ala Tyr Lys His Glu Val Arg Asp Lys Glu Tyr Asp Ser Ile Arg
                405                 410                 415
```

```
His Lys Leu Tyr Gln Leu Phe Ser Phe Ala Leu Tyr His Tyr Tyr Lys
            420                 425                 430

Thr Gly Val Gly Ala Glu Arg Arg Glu Ala Phe Val Ala Lys Leu Arg
        435                 440                 445

Ala Val Met Thr Ala Glu Ala Lys Gln Arg Ala Tyr Ala Asp Glu Ala
    450                 455                 460

Ala Glu Ile Trp Asn Asp Glu Gly Ser Gly Ile Arg Ala Ala Phe Leu
465                 470                 475                 480

Glu Ile Leu Glu Ala Val Asp Phe Gly Ser Ala Val Lys Gly Ile Lys
                485                 490                 495

Ala Arg Ser Ser Val Ala Gly Asp Lys Arg Phe Ala Glu Trp Leu Glu
            500                 505                 510

Glu Val Arg Ile Arg Pro Glu Gly Val Ser Cys Phe Thr Lys Leu Met
        515                 520                 525

Tyr Leu Leu Thr Arg Phe Leu Asp Gly Lys Glu Ile Asn Glu Leu Leu
    530                 535                 540

Thr Gly Leu Ile Asn Lys Leu Glu Asn Ile Gln Ser Phe Leu Asp Val
545                 550                 555                 560

Met Gln Gln Glu His Ala Glu Thr Gly Leu Ser Asp Ala Phe Ser Phe
                565                 570                 575

Phe Glu Tyr Ser Gly Glu Ile Ala Ala Glu Leu Arg Met Thr Arg Ser
            580                 585                 590

Phe Ala Arg Met Ala Ala Ala Asp Pro Glu Ala Lys Arg Phe Met Val
        595                 600                 605

Val Asp Gly Ala Lys Leu Leu Gly Phe Asn Pro Lys Asp Thr Glu Ser
    610                 615                 620
```

```
Glu Asp Glu Gly Ile Ile Arg Ala Ile Tyr Gly Asp Ala Cys Ala Glu
625                 630                 635                 640


Tyr Leu Gln Phe Ser Glu Glu Glu Lys Glu Ala Phe Tyr Val Gln Glu
                645                 650                 655


Gly Leu Tyr Gly Lys Glu Arg Glu Lys Phe Ser Pro Tyr Ala Tyr Phe
            660                 665                 670


His Thr Asp Thr Ser Leu Arg Asn Phe Ile Ala Lys Asn Val Val Glu
        675                 680                 685


Ser Ala Arg Phe Arg Tyr Val Ile Arg Tyr Val Ser Pro Glu Ile Ala
    690                 695                 700


Arg Lys Tyr Ala Arg Gln Glu Ala Leu Val Arg Phe Ala Leu His Arg
705                 710                 715                 720


Val Pro Leu Leu Gln Leu Arg Arg Tyr Tyr Gln Ser Cys Cys Gly Pro
                725                 730                 735


Lys Lys Asp Pro Asp Ala Ala Glu Cys Val Asp Phe Leu Ala Gly Val
            740                 745                 750


Val Asn Arg Val Asp Phe Ala Asn Phe Thr Asp Val Arg Thr Gly Asp
        755                 760                 765


Ser Ser Lys Ser Glu Gln Glu Lys Lys Gln Lys Tyr Gln Ala Ile Val
    770                 775                 780


Gly Leu Tyr Leu Thr Val Val Tyr Trp Ile Val Lys Asn Leu Val Asn
785                 790                 795                 800


Val Asn Ser Arg Tyr Val Met Ala Phe His Ile Leu Glu Arg Asp Thr
                805                 810                 815


Val Leu Leu Glu Gly Lys Arg Leu Phe Val Gly Gly Met Lys Ala Glu
            820                 825                 830
```

```
Asp Pro Phe Leu Leu Thr Asp Gly Tyr Val Ser Arg Gln Asp Ala Tyr
        835                 840                 845

Val Arg Lys Arg Ile Gly Glu Asn Lys Arg Ala Asn Arg His Gly Leu
    850                 855                 860

Asn Cys Val Leu Glu Asn Arg Asn Ala Leu Gly Ser Asp Pro Ala Ser
865                 870                 875                 880

Thr Asp Ala Ala Ala Ser Leu Ile Trp Ser Tyr Arg Asn Ala Ala Ala
                885                 890                 895

His Leu Thr Ala Val Ala Ala Ala Gln Glu Tyr Val Ser Glu Leu Arg
            900                 905                 910

Glu Ile His Ser Tyr Phe Glu Val Tyr His Tyr Ala Met Gln Arg Tyr
        915                 920                 925

Leu Lys Ser Gly Ala Glu Phe Ala Glu Leu Val Ser Lys Asn Gly Pro
    930                 935                 940

Ala Ser Gly Lys Ile Ala Ala Trp Ala Asn Ala Val Asp Arg Cys His
945                 950                 955                 960

Ser Phe Cys Lys Asp Trp Leu Trp Leu Leu Asn Val Pro Phe Ala Tyr
                965                 970                 975

Asn Pro Ala Arg Tyr Lys Asn Leu Ser Ile Ala Asn Leu Phe Asp Lys
            980                 985                 990

Asn Glu Ala Ala Pro Val Thr Glu  Asp Ala Ser Glu Gln  Lys Glu Asp
        995                 1000                1005

Glu
```

<210> 178
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Direct repeat

<400> 178
gtactacacc cctgcagttt tgcaggggtc tgaaac 36

<210> 179
<211> 961
<212> PRT
<213> Artificial Sequence

<220>
<223> Metagenomic hit (no protein accession): contig OIZX01000427.1

<400> 179

Met Ala Lys Lys Lys Lys Thr Ala Arg Gln Leu Arg Glu Glu Met Gln
1   5   10   15

Gln Gln Arg Lys Gln Ala Ile Gln Lys Gln Gln Glu Gln Arg Gln Glu
            20   25   30

Lys Ala Ala Ala Ala Arg Glu Thr Ala Ala Pro Glu Gln Pro Ala Ala
        35   40   45

Ala Pro Val Pro Lys Arg Gln Arg Lys Ser Leu Ala Lys Ala Ala Gly
    50   55   60

Leu Lys Ser Asn Phe Ile Leu Asp Pro Gln Arg Arg Thr Thr Val Met
65   70   75   80

Thr Ala Phe Gly Gln Gly Ser Thr Ala Ile Leu Glu Lys Gln Ile Val
                85   90   95

Asp Arg Ala Ile Ser Asp Leu Gln Pro Val Gln Gln Phe Gln Val Glu
            100   105   110

Pro Ala Ser Ala Ala Lys Tyr Arg Leu Lys Asn Ser Arg Val Arg Phe
        115   120   125

Pro Asn Val Thr Ala Asp Asp Pro Leu Tyr Arg Arg Lys Asp Gly Gly
    130   135   140

```
Phe Val Pro Gly Met Asp Ala Leu Arg Arg Lys Asn Val Leu Glu Gln
145                 150                 155                 160

Arg Phe Phe Gly Lys Ser Phe Ala Asp Asn Ile His Ile Gln Met Ile
                165                 170                 175

Tyr Ser Ile Leu Asp Ile His Lys Ile Leu Ala Ala Ala Ser Gly His
            180                 185                 190

Ile Val His Leu Leu Asn Ile Val Asn Gly Ser Lys Asp Arg Asp Phe
        195                 200                 205

Ile Gly Met Leu Ala Ala His Val Leu Tyr Asn Glu Leu Asn Glu Glu
    210                 215                 220

Ala Lys Arg Ser Ile Ala Asp Phe Cys Lys Ser Pro Arg Leu Ile Tyr
225                 230                 235                 240

Tyr Ser Ala Ala Phe Tyr Glu Thr Leu Asp Asn Gly Lys Ser Glu Arg
                245                 250                 255

Arg Ser Asn Glu Asp Ile Phe Asn Ile Leu Ala Leu Met Thr Cys Leu
            260                 265                 270

Arg Asn Phe Ser Ser His His Ser Ile Ala Ile Lys Val Lys Asp Tyr
        275                 280                 285

Ser Ala Ala Gly Leu Tyr Asn Leu Arg Arg Leu Gly Pro Asp Met Lys
    290                 295                 300

Lys Met Leu Asp Thr Phe Tyr Thr Glu Ala Phe Ile Gln Leu Asn Gln
305                 310                 315                 320

Ser Phe Gln Asp His Asn Thr Thr Asn Leu Thr Cys Leu Phe Asp Ile
                325                 330                 335

Leu Asn Ile Ser Asp Ser Ala Arg Gln Lys Gln Leu Ala Glu Glu Phe
            340                 345                 350
```

```
Tyr Arg Tyr Val Val Phe Lys Glu Gln Lys Asn Leu Gly Phe Ser Val
        355                 360                 365

Arg Lys Leu Arg Glu Glu Met Leu Leu Leu Pro Asp Ala Ala Val Ile
    370                 375                 380

Ala Asp Lys Arg Tyr Asp Thr Cys Arg Ser Lys Leu Tyr Asn Leu Met
385                 390                 395                 400

Asp Phe Leu Ile Leu Arg Val Tyr Arg Thr Gly Arg Ala Asp Arg Cys
                405                 410                 415

Asp Lys Leu Pro Glu Ala Leu Arg Ala Ala Leu Thr Asp Glu Glu Lys
            420                 425                 430

Ala Val Val Tyr His Lys Glu Ala Leu Ser Leu Trp Asn Glu Met Arg
        435                 440                 445

Thr Leu Ile Leu Asp Gly Leu Leu Pro Gln Met Thr Pro Glu Asn Leu
    450                 455                 460

Ser Arg Leu Ser Gly Gln Lys Arg Lys Gly Glu Leu Ser Leu Asp Asp
465                 470                 475                 480

Ala Met Leu Lys Glu Cys Leu Tyr Glu Pro Gly Pro Val Pro Glu Asp
                485                 490                 495

Ala Ala Pro Glu Glu Ala Asn Ala Glu Tyr Phe Cys Arg Met Ile Tyr
            500                 505                 510

Leu Ala Thr Leu Phe Met Asp Gly Lys Glu Ile Asn Thr Leu Leu Thr
        515                 520                 525

Thr Leu Ile Ser Lys Phe Glu Asn Ile Ala Ala Phe Leu Gln Thr Met
    530                 535                 540

Glu Gln Leu Asn Ile Glu Ala Glu Leu Gly Pro Glu Tyr Ala Met Phe
545                 550                 555                 560
```

```
Thr Arg Ser Arg Ala Val Ala Glu Gln Leu Arg Val Ile Asn Ser Phe
                565                 570                 575

Ala Leu Met Lys Lys Pro Gln Val Asn Ala Lys Gln Gln Leu Tyr Arg
            580                 585                 590

Ala Ala Val Thr Leu Leu Gly Thr Glu Asp Pro Asp Gly Val Thr Asp
        595                 600                 605

Glu Met Leu Cys Ile Asp Pro Val Thr Gly Lys Met Leu Pro Pro Asn
    610                 615                 620

Gln Arg His His Gly Asp Thr Gly Leu Arg Asn Phe Ile Ala Asn Asn
625                 630                 635                 640

Val Val Glu Ser Arg Arg Phe Gln Tyr Leu Ile Arg Tyr Ser Asp Pro
                645                 650                 655

Ala Gln Leu His Gln Leu Ala Ser Asn Lys Lys Leu Val Arg Phe Val
            660                 665                 670

Leu Ser Ser Ile Pro Asp Thr Gln Ile Asn Arg Tyr Tyr Glu Thr Cys
        675                 680                 685

Gly Gln Thr Arg Leu Ala Gly Arg Ala Ala Lys Val Glu Phe Leu Thr
    690                 695                 700

Asp Met Ile Ala Ala Ile Arg Phe Asp Gln Phe Arg Asp Val Asn Gln
705                 710                 715                 720

Lys Glu Arg Gly Ala Asn Thr Gln Lys Glu Arg Tyr Lys Ala Met Leu
                725                 730                 735

Gly Leu Tyr Gln Thr Val Leu Tyr Leu Ala Val Lys Asn Leu Val Asn
            740                 745                 750

Ile Asn Ala Arg Tyr Val Met Ala Phe His Cys Val Glu Arg Asp Met
        755                 760                 765
```

```
Phe Leu Tyr Asp Gly Glu Leu Thr Asp Pro Lys Gly Glu Ser Val Ser
    770                 775                 780

Ala Phe Leu Ala Val Asn Gly Lys Lys Gly Val Gln Pro Gln Tyr Leu
785                 790                 795                 800

Leu Leu Thr Gln Leu Phe Ile Arg Arg Asp Tyr Leu Lys Arg Ser Ala
                805                 810                 815

Cys Glu Gln Ile Gln His Asn Met Glu Asn Ile Ser Asp Arg Leu Leu
            820                 825                 830

Arg Glu Tyr Arg Asn Ala Val Ala His Leu Asn Val Ile Ala His Leu
        835                 840                 845

Ala Asp Tyr Ser Ala Asp Met Arg Glu Ile Thr Ser Tyr Tyr Gly Leu
    850                 855                 860

Tyr His Tyr Leu Met Gln Arg His Leu Phe Lys Arg His Ala Trp Gln
865                 870                 875                 880

Ile Arg Gln Pro Glu Arg Pro Thr Glu Glu Glu Gln Lys Leu Ile Glu
                885                 890                 895

Gln Glu Gln Lys Gln Leu Ala Trp Glu Lys Ala Leu Phe Asp Lys Thr
            900                 905                 910

Leu Gln Tyr His Ser Tyr Asn Lys Asp Leu Val Lys Ala Leu Asn Ala
        915                 920                 925

Pro Phe Gly Tyr Asn Leu Ala Arg Tyr Lys Asn Leu Ser Ile Glu Pro
    930                 935                 940

Leu Phe Ser Lys Glu Ala Ala Pro Ala Ala Glu Ile Lys Ala Thr His
945                 950                 955                 960

Ala
```

<210> 180
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Direct repeat

<400> 180
actatagccc tgccggaaat gacagggttc tacaac 36

<210> 181
<211> 963
<212> PRT
<213> Artificial Sequence

<220>
<223> Metagenomic hit (no protein accession): contig emb|OCVV012889144.1|

<400> 181

Met Ser Lys Asn Lys Glu Ser Tyr Ala Lys Gly Met Gly Leu Lys Ser
1 5 10 15

Ala Leu Val Ser Gly Ser Lys Val Tyr Met Thr Ser Phe Glu Gly Gly
20 25 30

Asn Asp Ala Lys Leu Glu Lys Val Val Glu Asn Ser Glu Ile Val Ser
35 40 45

Leu Ala Glu Lys Glu Ser Phe Ser Ala Glu Ile Phe Lys Lys Asn Ile
50 55 60

Gly Cys Lys Ile Glu Asn Lys Lys Phe Lys His Pro Lys Arg Tyr Asp
65 70 75 80

Val Ile Ala Asp Asn Pro Leu Tyr Lys Gly Ser Val Arg Gln Asp Met
85 90 95

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

```
Gly Thr Asp Asn Val Cys Ile Gln Val Ile His Asn Ile Leu Asp Ile
        115                 120                 125

Glu Lys Ile Leu Ala Glu Tyr Ile Thr Asn Ala Val Tyr Ser Phe Asp
    130                 135                 140

Asn Ile Ala Gly Phe Gly Glu Asp Ile Ile Gly Met Gly Gly Phe Lys
145                 150                 155                 160

Pro Ile Tyr Thr Tyr Lys Gln Phe Lys Glu Pro Asp Lys Tyr Asn Lys
                165                 170                 175

Lys Phe Asp Asp Ile Leu Asn Asn Ser Arg Leu Gly Tyr Tyr Gly Lys
            180                 185                 190

Ala Phe Phe Glu Lys Asn Asp Leu Lys His Asn Pro Asn Lys Lys Lys
        195                 200                 205

Arg Asp Lys Asn Pro Tyr Ile Leu Lys Tyr Asp Asn Glu Cys Tyr Tyr
    210                 215                 220

Ile Ile Ala Leu Leu Ser Gly Leu Arg His Trp Asn Ile His Ser His
225                 230                 235                 240

Ala Lys Asp Asp Leu Val Ser Tyr Arg Trp Leu Tyr Asn Leu Asp Ser
                245                 250                 255

Ile Leu Asn Arg Glu Tyr Ile Ser Thr Leu Asn Tyr Leu Tyr Asp Asp
            260                 265                 270

Ile Ala Asp Glu Leu Thr Glu Ser Phe Ser Lys Asn Ser Ser Ala Asn
        275                 280                 285

Val Asn Tyr Ile Ala Glu Thr Leu Asn Ile Asp Pro Ser Glu Phe Ala
    290                 295                 300

Gln Gln Tyr Phe Arg Phe Ser Ile Met Lys Glu Gln Lys Asn Met Gly
305                 310                 315                 320
```

```
Phe Asn Val Ser Lys Leu Arg Glu Ile Met Leu Asp Arg Lys Glu Leu
                325                 330                 335

Ser Asp Ile Arg Asp Asn His Arg Val Phe Asp Ser Ile Arg Ser Lys
            340                 345                 350

Leu Tyr Thr Met Met Asp Phe Val Ile Tyr Arg Tyr Tyr Ile Glu Glu
        355                 360                 365

Ala Ala Lys Thr Glu Ala Glu Asn Arg Asn Leu Pro Glu Asn Glu Lys
    370                 375                 380

Lys Ile Ser Glu Lys Asp Phe Phe Val Ile Asn Leu Arg Gly Ser Phe
385                 390                 395                 400

Asp Glu Asn Gln Lys Glu Lys Leu Tyr Ile Glu Glu Ala Lys Arg Leu
                405                 410                 415

Trp Glu Lys Leu Lys Asp Ile Met Leu Lys Ile Lys Glu Phe Arg Gly
            420                 425                 430

Glu Lys Val Lys Glu Tyr Lys Lys Glu Val Pro Arg Ile Glu Arg Ile
        435                 440                 445

Leu Pro Gln Gly Lys Asp Ile Ser Ala Phe Ser Lys Leu Met Tyr Met
    450                 455                 460

Leu Ser Met Phe Leu Asp Gly Lys Glu Ile Asn Asp Leu Phe Thr Thr
465                 470                 475                 480

Leu Ile Asn Lys Phe Asp Asn Ile Gln Cys Phe Leu Lys Ile Met Pro
                485                 490                 495

Leu Ile Gly Val Asn Val Lys Phe Thr Glu Asp Tyr Ala Phe Phe Asn
            500                 505                 510

Lys Ser Glu Lys Ile Ala Asp Glu Leu Arg Leu Ile Lys Ser Phe Ala
        515                 520                 525
```

```
Arg Met Gly Glu Pro Val Ala Asn Thr Lys Arg Val Met Met Ile Asp
    530                 535                 540

Ala Val Lys Ile Leu Gly Thr Asp Leu Ser Asp Asp Glu Leu Val Asn
545                 550                 555                 560

Met Ala Asp Ser Phe Phe Asn Asp Ser Glu Gly Asn Leu Leu Lys Lys
                565                 570                 575

Gly Lys His Gly Met Arg Asn Phe Ile Ile Asn Asn Val Ile Glu Asn
            580                 585                 590

Lys Arg Phe His Tyr Leu Ile Arg Tyr Gly Asp Pro Ala His Leu His
        595                 600                 605

Glu Ile Ala Lys Asn Glu Ala Val Val Arg Phe Val Leu Gly Arg Ile
    610                 615                 620

Ala Asp Ile Gln Lys Lys Gln Gly Lys Asp Gly Lys Asn Gln Ile Asp
625                 630                 635                 640

Arg Tyr Tyr Glu Ile Cys Val Gly Lys Asp Arg Gly Lys Thr Val Ser
                645                 650                 655

Gln Lys Leu Asp Glu Leu Thr Lys Ile Ile Val Asn Met Asn Tyr Asp
            660                 665                 670

Gln Phe Glu Gln Lys Lys Glu Ile Ile Glu Asn Thr Arg Thr Lys Thr
        675                 680                 685

Lys Tyr Met Arg Tyr Asp Glu Glu Gln Lys Leu Arg Lys Asp Asn Ala
    690                 695                 700

Glu Arg Glu Lys Phe Lys Lys Ile Ile Ser Leu Tyr Leu Thr Val Leu
705                 710                 715                 720

Tyr Tyr Val Val Lys Asn Ile Val Asn Ile Asn Ser Arg Tyr Val Ile
                725                 730                 735
```

```
Gly Phe His Cys Leu Glu Arg Asp Ala Gln Leu Tyr Ser Glu Lys Tyr
            740                 745                 750

Asn Lys Glu Ile Asp Lys Lys Ile Ser Ser Leu Thr Lys Phe Cys Leu
        755                 760                 765

Gly Glu Ser Asp Glu Lys Lys Leu Lys Ala Leu Ala Lys Lys Ser Leu
    770                 775                 780

Glu Glu Leu Lys Thr Thr Asn Ser Lys Leu Tyr Glu Asn Tyr Ile Lys
785                 790                 795                 800

Tyr Ser Asp Glu Arg Lys Ala Glu Glu Ala Lys Arg Gln Ile Asn Arg
                805                 810                 815

Glu Arg Ala Lys Thr Ala Met Asn Ala His Leu Arg Asn Thr Lys Trp
            820                 825                 830

Asn Asp Ile Met Tyr Gly Gln Leu Lys Asp Leu Ala Asp Ser Lys Ser
        835                 840                 845

Arg Ile Cys Ser Glu Phe Arg Asn Lys Ala Ala His Leu Glu Val Ala
    850                 855                 860

Arg Tyr Ala His Met Tyr Ile Asn Asp Ile Ser Glu Val Lys Ser Tyr
865                 870                 875                 880

Phe Arg Leu Tyr His Tyr Ile Met Gln Arg Arg Ile Ile Asp Val Ile
                885                 890                 895

Glu Asn Asn Pro Lys Ala Lys Tyr Glu Gly Lys Val Lys Val Tyr Phe
            900                 905                 910

Glu Asp Val Lys Lys Asn Lys Lys Tyr Asn Lys Asn Leu Leu Lys Leu
        915                 920                 925

Met Cys Val Pro Phe Gly Tyr Cys Ile Pro Arg Phe Lys Asn Leu Ser
    930                 935                 940
```

Ile Glu Gln Met Phe Asp Met Asn Glu Thr Asp Asn Ser Asp Lys Lys
945 950 955 960

Lys Glu Lys

<210> 182
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Direct repeat

<400> 182
catatacaac cccgtaaagt tacggggttc tgaaac 36

<210> 183
<211> 911
<212> PRT
<213> Artificial Sequence

<220>
<223> Metagenomic hit (no protein accession): contig OCTW011587266.1

<400> 183

Met Lys Gln Asn Asp Arg Glu Asn Asn Asn Lys Ile Lys Lys Ser Ala
1 5 10 15

Ala Lys Ala Val Gly Val Lys Ser Leu Ala Arg Leu Ser Asp Gly Ser
20 25 30

Thr Val Val Ser Ser Phe Gly Lys Gly Ala Ala Ala Glu Leu Glu Ser
35 40 45

Leu Ile Thr Gly Gly Glu Ile Arg Lys Leu Ser Asp Lys Ala Ile Leu
50 55 60

Glu Ile Thr Asp Asp Thr Gln Asn Lys Asn Ala Tyr Asn Val Lys Ser
65 70 75 80

Ser Arg Ile Pro Asn Leu Thr Ala Arg Thr Asp Lys Leu Ser Asp Lys

85 90 95

Ser Gly Met Asp Asp Leu Gly Phe Lys Arg Glu Leu Glu Leu Glu Val
100 105 110

Phe Gly Gln Cys Phe Asp Asp Ser Ile His Ile Gln Ile Ala His Ala
115 120 125

Val Phe Asp Ile Gln Lys Ser Leu Ala Ala Val Ile Pro Asn Val Leu
130 135 140

Tyr Thr Leu Asn Asn Leu Asp Arg Ser Tyr Ser Thr Asp Asn Thr Ser
145 150 155 160

Asp Lys Lys Asp Ile Ile Gly Asn Thr Leu Asn Tyr Gln His Ser Tyr
165 170 175

Glu Ser Phe Asn Val Glu Lys Arg Gly Glu Phe Thr Glu Tyr Tyr Asn
180 185 190

Ala Ala Lys Asp Arg Phe Ser Tyr Phe Pro Asp Ile Leu Cys Val Leu
195 200 205

Glu Lys Val Asn Gly Lys Asp Arg Tyr Gln Pro Lys Ser Glu Lys Asp
210 215 220

Ala Phe Asn Val Leu Ser Ser Val Asn Met Leu Arg Asn Ser Leu Phe
225 230 235 240

His Phe Ala Pro Lys Ser Asn Asp Gly Lys Ala Arg Ile Ala Val Phe
245 250 255

Lys Asn Gln Phe Asp Ser Asp Phe Ser His Ile Thr Ser Thr Val Asn
260 265 270

Lys Ile Tyr Ser Ala Lys Ile Ala Gly Val Asn Glu Asn Phe Leu Asn
275 280 285

Asn Glu Gly Asn Asn Leu Tyr Ile Ile Leu Lys Ala Thr Asn Trp Asp

290 295 300

Ile Lys Lys Ile Val Pro Gln Leu Tyr Arg Phe Ser Val Leu Lys Ser
305 310 315 320

Asp Lys Asn Met Gly Phe Asn Met Arg Lys Leu Arg Glu Phe Ala Val
325 330 335

Glu Ser Lys Asn Ile Asp Leu Ser Arg Leu Asn Asp Lys Phe Leu Thr
340 345 350

Asn Asn Arg Lys Lys Leu Tyr Lys Val Ile Asp Phe Ile Ile Tyr Tyr
355 360 365

His Leu Asn Lys Val Leu Lys Asp Ser Phe Val Asp Asp Phe Val Ala
370 375 380

Ala Leu Arg Ala Ser Gln Ser Glu Glu Glu Lys Glu Lys Leu Tyr Ala
385 390 395 400

Gln Tyr Ser Glu Arg Leu Phe Ala Asp Glu Gly Leu Lys Ser Ala Ile
405 410 415

Lys Lys Ala Val Asp Met Ile Ser Asp Thr Lys Ser Asn Ile Phe Lys
420 425 430

Met Lys Thr Pro Leu Asp Lys Ala Leu Ile Glu Asn Ile Lys Val Asn
435 440 445

Ser Asp Ala Ser Asp Phe Cys Lys Leu Ile Tyr Val Phe Thr Arg Phe
450 455 460

Leu Asp Gly Lys Glu Ile Asn Ile Leu Leu Asn Ser Leu Ile Lys Lys
465 470 475 480

Phe Gln Asp Ile His Ser Phe Asn Thr Thr Val Lys Lys Leu Ser Glu
485 490 495

Asn Asn Leu Ile Ile Asn Ala Asp Tyr Val Asp Asp Tyr Ser Leu Phe

500 505 510

Glu Gln Ser Gly Thr Val Ala Arg Glu Leu Met Leu Ile Lys Ser Ile
515 520 525

Ser Lys Met Asp Phe Gly Leu Asp Asn Ile Asn Leu Ser Phe Met Tyr
530 535 540

Asp Asp Ala Leu Arg Thr Leu Gly Val Ser Asp Glu Asn Leu Pro Glu
545 550 555 560

Val Lys Arg Glu Tyr Phe Gly Lys Thr Lys Asn Leu Ser Ala Tyr Ile
565 570 575

Arg Asn Asn Val Leu Glu Asn Arg Arg Phe Lys Tyr Val Ile Lys Tyr
580 585 590

Ile His Pro Ser Asp Val Gln Lys Ile Ala Cys Asn Lys Ala Ile Ala
595 600 605

Gly Phe Val Leu Asn Arg Met Pro Asp Thr Gln Ile Lys Arg Tyr Tyr
610 615 620

Asp Ser Leu Ile Asn Lys Gly Ala Thr Asp Ile Gln Ala Gln Ala Lys
625 630 635 640

Ala Leu Leu Asp Cys Ile Thr Gly Ile Ser Phe Asp Ala Ile Lys Asp
645 650 655

Asp Lys His Leu His Lys Ser Lys Glu Lys Ser Pro Gln Arg Ser Ala
660 665 670

Asp Arg Glu Arg Lys Lys Ala Met Leu Thr Leu Tyr Tyr Thr Ile Val
675 680 685

Tyr Ile Phe Val Lys Gln Met Leu His Ile Asn Ser Leu Tyr Thr Ile
690 695 700

Gly Phe Phe Tyr Leu Glu Arg Asp Gln Arg Phe Ile Tyr Ser Arg Ala

705 710 715 720

Lys Lys Glu Asn Lys Asn Pro Ser Lys Asn Ser Tyr Leu Asn Asp Phe
725 730 735

Arg Ser Val Thr Ala Tyr Phe Ile Pro Ser Glu Ile Met Lys Arg Ile
740 745 750

Glu Lys Asn Glu Asn Lys Gly Phe Leu Glu Asp Phe Glu Ala Leu Trp
755 760 765

Asn Ser Cys Gly Lys Thr Ser Arg Leu Arg Lys Glu Asp Val Leu Leu
770 775 780

Tyr Ala Arg Tyr Ile Ser Pro Asp His Ala Leu Lys Asn Tyr Lys Met
785 790 795 800

Ile Leu Asn Ser Tyr Arg Asn Lys Ile Ala His Ile Asn Val Ile Met
805 810 815

Ser Ala Gly Lys Tyr Thr Gly Gly Ile Lys Arg Met Asp Ser Tyr Phe
820 825 830

Ser Val Phe Gln His Leu Val Gln Cys Asp Ile Leu Ser Asn Pro Asn
835 840 845

Asn Lys Gly Lys Cys Phe Glu Ser Glu Ser Leu Lys Pro Leu Leu Leu
850 855 860

Asp Met Lys Phe Asp Gly Thr Asp Glu Lys Leu Tyr Ser Lys Arg Leu
865 870 875 880

Thr Arg Ala Leu Asn Ile Pro Phe Gly Tyr Asn Val Pro Arg Tyr Lys
885 890 895

Asn Leu Thr Phe Glu Lys Ile Tyr Leu Lys Ser Ser Ile Asn Glu
900 905 910

<210> 184

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

<220>
<223> Direct repeat

<400> 184
cttatacaac acccattttc acagtgggtc tataac 36

<210> 185
<211> 904
<212> PRT
<213> Artificial Sequence

<220>
<223> Metagenomic hit (no protein accession): contig emb|OGNF01009141.1

<400> 185

Met Ala Asp Ile Asp Lys Lys Lys Ser Ser Ala Lys Ala Ala Gly Leu
1 5 10 15

Lys Ser Thr Phe Val Leu Glu Asn Asn Lys Leu Leu Met Thr Ser Phe
20 25 30

Gly Asn Gly Asn Lys Ala Val Ile Glu Lys Ile Ile Asp Glu Lys Val
35 40 45

Asp Ser Ile Asn Glu Pro Glu Val Phe Ser Val Thr Pro Cys Asp Lys
50 55 60

Lys Phe Glu Leu Gln Pro Ala Lys Arg Gly Leu Ala Ala Asp Ser Leu
65 70 75 80

Val Asp Asn Pro Leu Lys Ser Lys Lys Thr Ala Gly Asp Asp Ala Ile
85 90 95

His Ser Arg Lys Phe Leu Glu Arg Gln Phe Phe Asp Gly Asn Thr Phe
100 105 110

Asn Asp Asn Ile His Ile Gln Leu Ile Tyr Asn Ile Leu Asp Ile Glu
115 120 125

```
Lys Ile Leu Ser Val His Val Asn Asp Ile Val Tyr Ser Val Asn Asn
    130                 135                 140

Ile Leu Ser Arg Gly Glu Gly Met Glu Tyr Asn Asp Tyr Ile Gly Thr
145                 150                 155                 160

Leu Asn Leu Lys Ser Phe Glu Thr Tyr Lys Asn Asn Leu Val Asn Lys
                165                 170                 175

Lys Lys Phe Asp Leu Asp Arg Val Lys Lys Ile Pro Gln Leu Ala Tyr
            180                 185                 190

Phe Gly Ser Ala Phe Tyr Asn Thr Pro Glu Asp Thr Ser Ala Lys Ile
        195                 200                 205

Thr Lys Thr Lys Ile Lys Ser Asn Glu Glu Ile Tyr Tyr Thr Phe Met
    210                 215                 220

Leu Leu Ser Thr Ala Arg Asn Phe Ser Ala His Tyr Leu Asp Arg Asn
225                 230                 235                 240

Arg Ala Lys Ser Ser Asp Ala Glu Asp Phe Asp Gly Thr Ser Val Ile
                245                 250                 255

Met Tyr Asn Leu Asp Asn Glu Glu Leu Tyr Lys Lys Leu Tyr Asn Lys
            260                 265                 270

Lys Val His Met Ala Leu Thr Gly Met Lys Lys Val Leu Asp Ala Asn
        275                 280                 285

Phe Asn Lys Lys Val Glu His Leu Asn Asn Ser Phe Ile Lys Asn Ser
    290                 295                 300

Ala Lys Asp Phe Val Ile Leu Cys Glu Val Leu Gly Ile Lys Ser Arg
305                 310                 315                 320

Asp Glu Lys Thr Lys Phe Val Lys Asp Tyr Tyr Asp Phe Val Val Arg
                325                 330                 335
```

```
Lys Asn Tyr Lys His Leu Gly Phe Ser Val Lys Glu Leu Arg Glu Leu
            340                 345                 350

Leu Phe Ala Asn His Asp Ser Asn Lys Tyr Ile Lys Glu Phe Asp Lys
        355                 360                 365

Ile Ser Asn Lys Lys Phe Asp Ser Val Arg Ser Arg Leu Asn Arg Leu
    370                 375                 380

Ala Asp Tyr Ile Ile Tyr Asp Tyr Tyr Asn Lys Asn Asn Ala Lys Val
385                 390                 395                 400

Ser Asp Leu Val Lys Tyr Leu Arg Ala Ala Ala Asp Asp Glu Gln Lys
                405                 410                 415

Lys Lys Ile Tyr Leu Asn Glu Ser Ile Asn Leu Val Lys Ser Gly Ile
            420                 425                 430

Leu Glu Arg Ile Lys Lys Ile Leu Pro Lys Leu Asn Gly Lys Ile Ile
        435                 440                 445

Gly Asn Met Gln Pro Asp Ser Thr Ile Thr Ala Ser Met Leu His Asn
    450                 455                 460

Thr Gly Lys Asp Trp His Pro Ile Ser Glu Asn Ala His Tyr Phe Thr
465                 470                 475                 480

Lys Trp Ile Tyr Thr Leu Thr Leu Phe Met Asp Gly Lys Glu Ile Asn
                485                 490                 495

Asp Leu Val Thr Thr Leu Ile Asn Lys Phe Asp Asn Ile Ala Ser Phe
            500                 505                 510

Ile Glu Val Leu Lys Ser Gln Ser Val Cys Thr His Phe Ser Glu Glu
        515                 520                 525

Arg Lys Met Phe Ile Asp Ser Ala Glu Ile Cys Ser Glu Leu Ser Ala
    530                 535                 540
```

```
Met Asn Ser Phe Ala Arg Met Glu Ala Pro Gly Ala Ser Ser Lys Arg
545                 550                 555                 560

Ala Met Phe Val Glu Ala Ala Arg Ile Leu Gly Asp Asn Arg Ser Lys
                565                 570                 575

Glu Glu Leu Glu Glu Tyr Phe Asp Thr Leu Phe Asp Lys Ser Ala Ser
            580                 585                 590

Lys Lys Glu Lys Gly Phe Arg Asn Phe Ile Arg Asn Asn Val Val Asp
        595                 600                 605

Ser Asn Arg Phe Lys Tyr Leu Thr Arg Tyr Thr Asp Thr Ser Ser Val
    610                 615                 620

Lys Ala Phe Ser Asn Asn Lys Ala Leu Val Lys Phe Ala Ile Lys Asp
625                 630                 635                 640

Ile Pro Gln Glu Gln Ile Leu Arg Tyr Tyr Asn Ser Cys Phe Gly Ala
                645                 650                 655

Ser Glu Arg Tyr Tyr Asn Asp Gly Met Ser Asp Lys Leu Val Glu Ala
            660                 665                 670

Ile Gly Lys Ile Asn Leu Met Gln Phe Asn Gly Val Ile Gln Gln Ala
        675                 680                 685

Asp Arg Asn Met Leu Pro Glu Glu Lys Lys Lys Ala Asn Ala Gln Lys
    690                 695                 700

Glu Lys Tyr Lys Ser Ile Ile Arg Leu Tyr Leu Thr Val Cys Tyr Leu
705                 710                 715                 720

Phe Phe Lys Asn Leu Val Tyr Val Asn Ser Arg Tyr Tyr Ser Ala Phe
                725                 730                 735

Tyr Asn Leu Glu Lys Asp Arg Ser Leu Phe Glu Ile Asn Gly Glu Leu
            740                 745                 750
```

```
Lys Pro Thr Gly Lys Phe Asp Glu Gly His Tyr Thr Gly Leu Val Lys
        755                 760                 765

Leu Phe Ile Asp Asn Gly Trp Ile Asn Pro Arg Ala Ser Ala Tyr Leu
    770                 775                 780

Thr Val Asn Leu Ala Asn Ser Asp Glu Thr Ala Ile Arg Thr Phe Arg
785                 790                 795                 800

Asn Thr Ala Glu His Leu Glu Ala Leu Arg Asn Ala Asp Lys Tyr Leu
                805                 810                 815

Asn Asp Leu Lys Gln Phe Asp Ser Tyr Phe Glu Ile Tyr His Tyr Ile
            820                 825                 830

Thr Gln Arg Asn Ile Lys Glu Lys Cys Glu Met Leu Lys Glu Gln Thr
        835                 840                 845

Val Lys Tyr Asn Asn Asp Leu Leu Lys Tyr His Gly Tyr Ser Lys Asp
    850                 855                 860

Phe Val Lys Ala Leu Cys Val Pro Phe Gly Tyr Asn Leu Pro Arg Phe
865                 870                 875                 880

Lys Asn Leu Ser Ile Asp Ala Leu Phe Asp Lys Asn Asp Lys Arg Glu
                885                 890                 895

Lys Leu Lys Lys Gly Phe Glu Asp
            900
```

<210> 186
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Direct repeat

<400> 186
gaactacagc cctccttaaa tgggggttc taaaac 36

<210> 187
<211> 933
<212> PRT
<213> Artificial Sequence

<220>
<223> Metagenomic hit (no protein accession): contig emb|OIEN01002196.1|

<400> 187

```
Met Glu Arg Gln Lys Arg Lys Met Lys Ser Lys Ser Lys Met Ala Gly
1               5                  10                  15

Val Lys Ser Val Phe Val Ile Gly Asp Glu Leu Leu Met Thr Ser Phe
            20                  25                  30

Gly Asp Gly Asp Asp Ala Val Leu Glu Lys Asp Ile Asp Glu Asn Gly
        35                  40                  45

Val Val Asn Asp Cys Arg Asn Pro Ala Ala Tyr Asp Ala Val Tyr Gly
    50                  55                  60

Thr Asp Ser Ile Arg Val Lys Lys Thr Asn Asn Asn Ile Arg Ala Lys
65                  70                  75                  80

Val Asn Asn Pro Leu Ala Lys Ser Asn Ile Arg Ser Glu Glu Ser Ala
                85                  90                  95

Leu Phe Arg Thr Arg Val Asn Glu Tyr Lys Arg Glu Gln Lys Asp Lys
            100                 105                 110

Tyr Glu Thr Leu Phe Phe Gly Lys Thr Phe Asp Asp Asn Ile His Ile
        115                 120                 125

Gln Leu Ile Ser Lys Ile Leu Asp Ile Glu Lys Thr Phe Ser Val Val
    130                 135                 140

Ile Gly Asn Ile Val Tyr Ala Ile Asn Asn Leu Ser Leu Glu Gln Ser
145                 150                 155                 160
```

```
Ile Asp Arg Pro Ile Asp Ile Phe Gly Asp Lys Asn Thr Gln Gly Ile
                165                 170                 175

Ser Leu Arg Glu Asp Asn Asp Tyr Leu Lys Thr Met Leu Pro Arg Cys
            180                 185                 190

Glu Tyr Leu Phe His Asn Ile Leu Asn Ser Asp Ser Asp Asn Asn Ser
        195                 200                 205

Lys Met Asn Tyr Asn Lys Val Asn Lys Gly Lys Glu Glu Lys Asp Asn
    210                 215                 220

Arg Asn Asn Glu Asn Ile Glu Lys Leu Lys Lys Ala Leu Glu Val Ile
225                 230                 235                 240

Lys Ile Ile Arg Val Asp Ser Phe His Gly Val Asp Gly Ile Lys Gly
                245                 250                 255

Asp Gln Lys Phe Pro Arg Ser Lys Tyr Asn Leu Ala Val Asn Tyr Asn
            260                 265                 270

Glu Glu Ile Gln Lys Thr Ile Ser Glu Pro Phe Asn Arg Lys Val Glu
        275                 280                 285

Glu Val Gln Gln Asp Phe Tyr Arg Asn Ser Cys Val Asn Ile Asp Phe
    290                 295                 300

Leu Lys Glu Ile Met Tyr Gly Ser Asn Tyr Thr Asp Arg Gly Ser Asp
305                 310                 315                 320

Ser Leu Glu Cys Ser Tyr Phe Asn Phe Ala Ile Leu Lys Gln Asn Lys
                325                 330                 335

Asn Met Gly Phe Ser Ile Thr Ser Ile Arg Glu Cys Leu Leu Asp Leu
            340                 345                 350

Tyr Glu Leu Asn Phe Glu Ser Met Gln Asn Leu Arg Pro Arg Ala Asn
        355                 360                 365
```

```
Ser Phe Cys Asp Phe Leu Ile Tyr Asp Tyr Tyr Cys Lys Asn Glu Ser
    370                 375                 380

Glu Arg Ala Asn Leu Val Asp Cys Leu Arg Ser Ala Ala Ser Glu Glu
385                 390                 395                 400

Glu Lys Lys Asn Ile Tyr Phe Gln Thr Ala Glu Arg Val Lys Glu Lys
                405                 410                 415

Phe Arg Asn Ala Phe Asn Arg Ile Ser Arg Phe Asp Ala Ser Tyr Ile
            420                 425                 430

Lys Asn Ser Arg Glu Lys Asn Leu Ser Gly Gly Ser Ser Leu Pro Lys
        435                 440                 445

Tyr Ser Phe Ile Glu Gly Phe Thr Lys Arg Ser Lys Lys Ile Asn Asp
    450                 455                 460

Asn Asp Glu Lys Asn Ala Asp Leu Phe Cys Asn Met Leu Tyr Tyr Leu
465                 470                 475                 480

Ala Gln Phe Leu Asp Gly Lys Glu Ile Asn Ile Phe Leu Thr Ser Ile
                485                 490                 495

His Asn Ile Phe Gln Asn Ile Asp Ser Phe Leu Lys Val Met Lys Glu
            500                 505                 510

Lys Gly Met Glu Cys Lys Phe Gln Lys Asp Phe Lys Met Phe Ser His
        515                 520                 525

Ala Gly His Val Ala Lys Lys Ile Glu Ile Val Ile Ser Leu Ala Lys
    530                 535                 540

Met Lys Lys Thr Leu Asp Phe Tyr Asn Ala Gln Ala Leu Lys Asp Ala
545                 550                 555                 560

Val Thr Ile Leu Gly Val Ser Lys Lys His Gln Tyr Leu Asp Met Asn
                565                 570                 575
```

```
Ser Tyr Leu Asp Phe Tyr Met Phe Asp Asn Arg Ser Gly Ala Thr Gly
            580                 585                 590

Lys Asn Ala Gly Lys Asp His Asn Leu Arg Asn Phe Leu Val Ser Asn
        595                 600                 605

Val Ile Arg Ser Arg Lys Phe Asn Tyr Leu Ser Arg Tyr Ser Asn Leu
    610                 615                 620

Ala Glu Val Lys Lys Leu Ala Gln Asn Pro Ser Leu Val Gln Phe Val
625                 630                 635                 640

Leu Ser Arg Ile Glu Pro Ser Leu Ile Cys Arg Tyr Tyr Glu Ser Ser
                645                 650                 655

Gln Gly Ile Ser Ser Glu Gly Ile Thr Ile Asp Glu Gln Ile Lys Lys
            660                 665                 670

Leu Thr Gly Ile Ile Val Asp Met Asn Ile Asp Ser Phe Glu Asn Ile
        675                 680                 685

Asn Asn Gly Glu Ile Gly Met Arg Tyr Ser Lys Ala Thr Pro Gln Ser
    690                 695                 700

Ile Glu Arg Arg Asn Gln Met Arg Val Cys Val Gly Leu Tyr Leu Asn
705                 710                 715                 720

Val Leu Tyr Gln Ile Glu Lys Asn Leu Met Asn Val Asn Ala Arg Tyr
                725                 730                 735

Val Leu Ala Phe Ala Phe Ala Glu Arg Asp Ala Leu Met Leu Asn Phe
            740                 745                 750

Thr Leu Glu Glu Cys Lys Lys Asn Lys Lys Arg Ser Ser Gly Gly Phe
        755                 760                 765

Ser Phe Ile Glu Met Thr Gln Phe Phe Ile Asp Lys Lys Leu Phe Lys
    770                 775                 780
```

```
Val Ala Thr Glu Ala Ile Lys Lys Asn Val Leu Lys Tyr Asn Gly Asn
785                 790                 795                 800


Pro Glu Ser Leu Asn His Ile Pro Gly Glu Tyr Ile Cys Lys Asn Met
                805                 810                 815


Glu Gly Tyr His Glu Asn Thr Val Arg Asn Phe Arg Asn Met Val Ala
            820                 825                 830


His Leu Thr Ala Val Ala Arg Val Pro Leu Tyr Ile Ser Glu Val Thr
        835                 840                 845


Gln Ile Asp Ser Tyr Tyr Ala Leu Tyr His Tyr Cys Met Gln Met Asn
    850                 855                 860


Ile Leu Gln Gly Ile Glu Gln Ser Gly Lys Ile Leu Asp Asn Ile Lys
865                 870                 875                 880


Leu Lys Asn Ala Leu Glu Asn Ala Arg Val His Arg Thr Tyr Ser Lys
                885                 890                 895


Asp Ala Val Lys Tyr Leu Cys Leu Pro Phe Ala Tyr Asn Ile Ser Arg
            900                 905                 910


Tyr Lys Ala Leu Thr Ile Lys Asp Leu Phe Asp Trp Thr Glu Tyr Ser
        915                 920                 925


Cys Lys Lys Asp Glu
    930


<210>  188
<211>  36
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  Direct repeat

<400>  188
gccccttgac cttacgaaat ggtaaggttc caaaac                                  36
```

<210> 189
<211> 1034
<212> PRT
<213> Artificial Sequence

<220>
<223> Metagenomic hit (no protein accession): contig e-k87_11092736

<400> 189

Met Lys Arg Gln Lys Thr Phe Ala Lys Arg Ile Gly Ile Lys Ser Thr
1 5 10 15

Val Ala Tyr Gly Gln Gly Lys Tyr Ala Ile Thr Thr Phe Gly Lys Gly
20 25 30

Ser Lys Ala Glu Ile Ala Val Arg Ser Ala Asp Pro Pro Glu Glu Thr
35 40 45

Leu Pro Thr Glu Ser Asp Ala Thr Leu Ser Ile His Ala Lys Phe Ala
50 55 60

Lys Ala Gly Arg Asp Gly Arg Glu Phe Lys Cys Gly Asp Val Asp Glu
65 70 75 80

Thr Arg Ile His Thr Ser Arg Ser Glu Tyr Glu Ser Leu Ile Ser Asn
85 90 95

Pro Ala Glu Ser Pro Arg Glu Asp Tyr Leu Gly Leu Lys Gly Thr Leu
100 105 110

Glu Arg Lys Phe Phe Gly Asp Glu Tyr Pro Lys Asp Asn Leu Arg Ile
115 120 125

Gln Ile Ile Tyr Ser Ile Leu Asp Ile Gln Lys Ile Leu Gly Leu Tyr
130 135 140

Val Glu Asp Ile Leu His Phe Val Asp Gly Leu Gln Asp Glu Pro Glu
145 150 155 160

Asp Leu Val Gly Leu Gly Leu Gly Asp Glu Lys Met Gln Lys Leu Leu

165 170 175

Ser Lys Ala Leu Pro Tyr Met Gly Phe Phe Gly Ser Thr Asp Val Phe
180 185 190

Lys Val Thr Lys Lys Arg Glu Glu Arg Ala Ala Ala Asp Glu His Asn
195 200 205

Ala Lys Val Phe Arg Ala Leu Gly Ala Ile Arg Gln Lys Leu Ala His
210 215 220

Phe Lys Trp Lys Glu Ser Leu Ala Ile Phe Gly Ala Asn Ala Asn Met
225 230 235 240

Pro Ile Arg Phe Phe Gln Gly Ala Thr Gly Gly Arg Gln Leu Trp Asn
245 250 255

Asp Val Ile Ala Pro Leu Trp Lys Lys Arg Ile Glu Arg Val Arg Lys
260 265 270

Ser Phe Leu Ser Asn Ser Ala Lys Asn Leu Trp Val Leu Tyr Gln Val
275 280 285

Phe Lys Asp Asp Thr Asp Glu Lys Lys Lys Ala Arg Ala Arg Gln Tyr
290 295 300

Tyr His Phe Ser Val Leu Lys Glu Gly Lys Asn Leu Gly Phe Asn Leu
305 310 315 320

Thr Lys Thr Arg Glu Tyr Phe Leu Asp Lys Phe Phe Pro Ile Phe His
325 330 335

Ser Ser Ala Pro Asp Val Lys Arg Lys Val Asp Thr Phe Arg Ser Lys
340 345 350

Phe Tyr Ala Ile Leu Asp Phe Ile Ile Tyr Glu Ala Ser Val Ser Val
355 360 365

Ala Asn Ser Gly Gln Met Gly Lys Val Ala Pro Trp Lys Gly Ala Ile

370 375 380

Asp Asn Ala Leu Val Lys Leu Arg Glu Ala Pro Asp Glu Glu Ala Lys
385 390 395 400

Glu Lys Ile Tyr Asn Val Leu Ala Ala Ser Ile Arg Asn Asp Ser Leu
405 410 415

Phe Leu Arg Leu Lys Ser Ala Cys Asp Lys Phe Gly Ala Glu Gln Asn
420 425 430

Arg Pro Val Phe Pro Asn Glu Leu Arg Asn Asn Arg Asp Ile Arg Asn
435 440 445

Val Arg Ser Glu Trp Leu Glu Ala Thr Gln Asp Val Asp Ala Ala Ala
450 455 460

Phe Val Gln Leu Ile Ala Phe Leu Cys Asn Phe Leu Glu Gly Lys Glu
465 470 475 480

Ile Asn Glu Leu Val Thr Ala Leu Ile Lys Lys Phe Glu Gly Ile Gln
485 490 495

Ala Leu Ile Asp Leu Leu Arg Asn Leu Glu Gly Val Asp Ser Ile Arg
500 505 510

Phe Glu Asn Glu Phe Ala Leu Phe Asn Asp Asp Lys Gly Asn Met Ala
515 520 525

Gly Arg Ile Ala Arg Gln Leu Arg Leu Leu Ala Ser Val Gly Lys Met
530 535 540

Lys Pro Asp Met Thr Asp Ala Lys Arg Val Leu Tyr Lys Ser Ala Leu
545 550 555 560

Glu Ile Leu Gly Ala Pro Pro Asp Glu Val Ser Asp Glu Trp Leu Ala
565 570 575

Glu Asn Ile Leu Leu Asp Lys Ser Asn Asn Asp Tyr Gln Lys Ala Lys

580 585 590

Lys Thr Val Asn Pro Phe Arg Asn Tyr Ile Ala Lys Asn Val Ile Thr
595 600 605

Ser Arg Ser Phe Tyr Tyr Leu Val Arg Tyr Ala Lys Pro Thr Ala Val
610 615 620

Arg Lys Leu Met Ser Asn Pro Lys Ile Val Arg Tyr Val Leu Lys Arg
625 630 635 640

Leu Pro Glu Lys Gln Val Ala Ser Tyr Tyr Ser Ala Ile Trp Thr Gln
645 650 655

Ser Glu Ser Asn Ser Asn Glu Met Val Lys Leu Ile Glu Met Ile Asp
660 665 670

Arg Leu Thr Thr Glu Ile Ala Gly Phe Ser Phe Ala Val Leu Lys Asp
675 680 685

Lys Lys Asp Ser Ile Val Ser Ala Ser Arg Glu Ser Arg Ala Val Asn
690 695 700

Leu Glu Val Glu Arg Leu Lys Lys Leu Thr Thr Leu Tyr Met Ser Ile
705 710 715 720

Ala Tyr Ile Ala Val Lys Ser Leu Val Lys Val Asn Ala Arg Tyr Phe
725 730 735

Ile Ala Tyr Ser Ala Leu Glu Arg Asp Leu Tyr Phe Phe Asn Glu Lys
740 745 750

Tyr Gly Glu Glu Phe Arg Leu His Phe Ile Pro Tyr Glu Leu Asn Gly
755 760 765

Lys Thr Cys Gln Phe Glu Tyr Leu Ala Ile Leu Lys Tyr Tyr Leu Ala
770 775 780

Arg Asp Glu Glu Thr Leu Lys Arg Lys Cys Glu Ile Cys Glu Glu Ile

785 790 795 800

Lys Val Gly Cys Glu Lys His Lys Lys Asn Ala Asn Pro Pro Tyr Glu
805 810 815

Tyr Asp Gln Glu Trp Ile Asp Lys Lys Lys Ala Leu Asn Ser Glu Arg
820 825 830

Lys Ala Cys Glu Arg Arg Leu His Phe Ser Thr His Trp Ala Gln Tyr
835 840 845

Ala Thr Lys Arg Asp Glu Asn Met Ala Lys His Pro Gln Lys Trp Tyr
850 855 860

Asp Ile Leu Ala Ser His Tyr Asp Glu Leu Leu Ala Leu Gln Ala Thr
865 870 875 880

Gly Trp Leu Ala Thr Gln Ala Arg Asn Asp Ala Glu His Leu Asn Pro
885 890 895

Val Asn Glu Phe Asp Val Tyr Ile Glu Asp Leu Arg Arg Tyr Pro Glu
900 905 910

Gly Thr Pro Lys Asn Lys Asp Tyr His Ile Gly Ser Tyr Phe Glu Ile
915 920 925

Tyr His Tyr Ile Arg Gln Arg Ala Tyr Leu Glu Glu Val Leu Ala Lys
930 935 940

Arg Lys Glu Tyr Arg Asp Ser Gly Ser Phe Thr Asp Glu Gln Leu Asp
945 950 955 960

Lys Leu Gln Lys Ile Leu Asp Asp Ile Arg Ala Arg Gly Ser Tyr Asp
965 970 975

Lys Asn Leu Leu Lys Leu Glu Tyr Leu Pro Phe Ala Tyr Asn Leu Pro
980 985 990

Arg Tyr Lys Asn Leu Thr Thr Glu Ala Leu Phe Asp Asp Asp Ser Val

995 1000 1005

Ser Gly Lys Lys Arg Val Ala Glu Trp Arg Glu Arg Glu Lys Thr
1010 1015 1020

Arg Glu Ala Glu Arg Glu Gln Arg Arg Gln Arg
1025 1030

<210> 190
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Direct repeat 1

<400> 190
gtgagaagtc tccttatggg gagatgctac 30

<210> 191
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Direct repeat 2

<400> 191
gtagcatctc cccataagga gacttctcac 30

<210> 192
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Direct repeat 3

<400> 192
gctagggctt gtaggaattc cttgctatgg tacaat 36

<210> 193
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Direct repeat 4

<400> 193
attgtaccat agcaaggaat tcctacaagc cctagc 36

<210> 194
<211> 116
<212> PRT
<213> Artificial Sequence

<220>
<223> Gut_metagenome_contig6893000291

<400> 194

Met Tyr Pro Phe Ile Asn Phe Ala Ile Phe Arg Tyr Tyr Asn Val Tyr
1 5 10 15

His Lys Glu Ala Asp Tyr Leu Val Asp Ser Leu Arg Ala Ser Met Asn
            20 25 30

Glu Val Glu Lys Asp Ala Val Tyr Arg Arg Glu Ala Val Arg Ile Trp
        35 40 45

Pro Arg Leu Lys Lys Leu Ile Leu Asn His Ile Leu Pro Glu Met Thr
    50 55 60

Gly Ser Lys Ile Lys Glu Lys Asp Ala Lys Pro Asp Pro Glu Met Asp
65 70 75 80

Ala Ser Met Ile Asp Ser Val Lys Ile Gly Thr Glu Ser Thr Leu Phe
                85 90 95

Thr Lys Ile Ile Phe Leu Thr Thr Leu Phe Leu Asn Gly Lys Glu Thr
            100 105 110

Asn Asp Leu Ile
        115

<210> 195
<211> 30
<212> PRT

<213> Artificial Sequence

<220>
<223> Artificial sequence

<220>
<221> misc_feature
<222> (2)..(6)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (7)..(7)
<223> Xaa can be Phe, Trp, or Tyr

<220>
<221> misc_feature
<222> (8)..(10)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (12)..(12)
<223> Xaa can be Asn, Gln, or Ser

<220>
<221> MISC_FEATURE
<222> (13)..(13)
<223> Xaa can be Ile, Leu, or Met

<220>
<221> MISC_FEATURE
<222> (14)..(14)
<223> Xaa can be Ile, Leu, Met, or Val

<220>
<221> misc_feature
<222> (15)..(16)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (18)..(19)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (20)..(20)
<223> Xaa can be Phe, Trp, or Tyr

<220>

<221> misc_feature
<222> (21)..(22)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (23)..(23)
<223> Xaa can be Ala or Gly

<220>
<221> misc_feature
<222> (24)..(27)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (28)..(28)
<223> Xaa can be Asp or Glu

<220>
<221> misc_feature
<222> (29)..(29)
<223> Xaa can be any naturally occurring amino acid

<400> 195

Leu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Lys Xaa Xaa Xaa Xaa Xaa
1 5 10 15

Asn Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Asp
20 25 30

<210> 196
<211> 18
<212> PRT
<213> Artificial Sequence

<220>
<223> Artificial sequence

<220>
<221> MISC_FEATURE
<222> (1)..(1)
<223> Xaa can be Phe, Trp, or Tyr

<220>
<221> MISC_FEATURE
<222> (2)..(2)
<223> Xaa can be Ile, Leu, or Val

<220>
<221> misc_feature
<222> (3)..(4)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (5)..(5)
<223> Xaa can be Asn, Gln, or Ser

<220>
<221> MISC_FEATURE
<222> (6)..(6)
<223> Xaa can be Ile, Leu, or Val

<220>
<221> misc_feature
<222> (7)..(8)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (9)..(9)
<223> Xaa can be Asp, Asn, Ser, or Thr

<220>
<221> misc_feature
<222> (10)..(11)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (13)..(13)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (15)..(16)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (17)..(17)
<223> Xaa can be His, Lys, or Arg

<220>
<221> MISC_FEATURE
<222> (18)..(18)
<223> Xaa can be Phe, His, or Tyr

<400> 196

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Phe Xaa Tyr Xaa Xaa
1 5 10 15

Xaa Xaa

<210> 197
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Artificial sequence

<220>
<221> MISC_FEATURE
<222> (2)..(2)
<223> Xaa can be Cys, Asp, Asn, Ser, or Val

<220>
<221> misc_feature
<222> (3)..(4)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (6)..(6)
<223> Xaa can be Phe, Trp, or Tyr

<220>
<221> misc_feature
<222> (7)..(7)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (8)..(8)
<223> Xaa can be Ala, Asp, Asn, or Thr

<220>
<221> MISC_FEATURE
<222> (9)..(9)
<223> Xaa can be Leu or Met

<220>
<221> MISC_FEATURE
<222> (10)..(10)
<223> Xaa can be Ser or Thr

<220>
<221> misc_feature
<222> (11)..(14)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (15)..(15)
<223> Xaa can be Phe, Trp, or Tyr

<400> 197

Tyr Xaa Xaa Xaa Arg Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1   5   10   15

<210> 198
<211> 940
<212> PRT
<213> Artificial Sequence

<220>
<223> Ga0224415_10007274

<400> 198

Met Ile Glu Lys Lys Lys Ser Tyr Ala Lys Gly Met Gly Leu Lys Ser
1   5   10   15

Thr Ile Val Ser Gly Ser Lys Val Tyr Met Thr Thr Phe Gly Asp Gly
            20   25   30

Ser Glu Ala Arg Leu Glu Lys Val Val Glu Asn Asp Ser Ile Lys Thr
        35   40   45

Leu His Glu Gly Glu Ala Phe Ser Ala Glu Met Thr Asp Asn Asn Ala
    50   55   60

Gly Tyr Lys Ile Gly Asn Val Lys Phe Ser His Pro Asn Gly Tyr Asp
65   70   75   80

Val Val Ala Asn Asn Pro Phe Tyr Thr Gly Pro Val Gln Gln Asp Met
                85   90   95

Leu Gly Leu Lys Glu Ile Leu Glu Arg Arg Tyr Phe Gly Ser Ser Thr

100 105 110

Asp Gly Asn Asn Thr Ile Cys Ile Gln Ile Ile His Asn Ile Leu Asp
115 120 125

Ile Glu Lys Ile Leu Ala Glu Tyr Ile Thr Asn Ala Val Tyr Ala Thr
130 135 140

Asn Asn Ile Ile Asp Pro Asp Asn Asp Val Ile Gly Gly Lys Lys Phe
145 150 155 160

Thr Ser Ile Lys Thr Phe Ala Gln Phe Ser Ala Ser Asp Ser Ser Asn
165 170 175

Asp Phe Glu Gln Phe Leu Lys Asn Pro Arg Leu Gly Tyr Leu Gly Lys
180 185 190

Ala Phe Phe Tyr Lys Asp Gly Lys Arg Asn Asn Arg Gln Lys Asp Pro
195 200 205

Ile Glu Cys Tyr His Leu Leu Ala Leu Leu Cys Gly Leu Arg Asn Trp
210 215 220

Val Val His Asn Asn Glu Glu Lys Asp Leu Ile Lys Tyr Thr Trp Leu
225 230 235 240

Tyr Asn Leu Asp Lys Tyr Leu Asp Ala Glu Tyr Ile Thr Thr Leu Asn
245 250 255

Tyr Met Tyr Asn Asp Ile Gly Asp Glu Leu Thr Asp Ser Phe Ser Lys
260 265 270

Asn Ser Ala Ala Asn Ile Asn Tyr Ile Ala Glu Thr Leu Gly Ile Asp
275 280 285

Pro Lys Thr Phe Ala Glu Gln Tyr Phe Arg Phe Ser Ile Met Lys Glu
290 295 300

Gln Lys Asn Leu Gly Phe Asn Leu Thr Lys Leu Arg Glu Val Met Leu

305 310 315 320

Asp Arg Lys Asp Met Ser Glu Ile Arg Glu Asn His Asn Asp Phe Asp
325 330 335

Ser Ile Arg Ala Lys Val Tyr Thr Met Met Asp Phe Val Ile Tyr Arg
340 345 350

Tyr Tyr Ile Glu Glu Ala Ala Lys Val Asn Ala Ala Asn Lys Ser Leu
355 360 365

Pro Asp Asn Glu Lys Ser Leu Ser Glu Lys Asp Ile Phe Val Ile Ser
370 375 380

Leu Arg Gly Ser Phe Asn Glu Asp Gln Lys Asp Arg Leu Tyr Tyr Asp
385 390 395 400

Glu Ala Gln Arg Leu Trp Ser Lys Val Gly Lys Leu Met Leu Lys Ile
405 410 415

Lys Lys Phe Arg Gly Lys Asp Thr Arg Lys Tyr Lys Asn Met Gly Thr
420 425 430

Pro Arg Ile Arg Arg Leu Ile Pro Glu Gly Arg Asp Ile Ser Thr Phe
435 440 445

Ser Lys Leu Met Tyr Ala Leu Thr Met Phe Leu Asp Gly Lys Glu Ile
450 455 460

Asn Asp Leu Leu Thr Thr Leu Ile Asn Lys Phe Asp Asn Ile Gln Ser
465 470 475 480

Phe Leu Lys Val Met Pro Leu Ile Gly Val Asn Ala Lys Phe Ala Glu
485 490 495

Glu Tyr Ser Phe Phe Asn Asn Ser Glu Lys Ile Ala Asp Glu Leu Arg
500 505 510

Leu Ile Lys Ser Phe Ala Arg Met Gly Glu Pro Val Ala Asp Ala Arg

515 520 525

Arg Ala Met Tyr Ile Asp Ala Ile Arg Ile Leu Gly Thr Asp Leu Ser
530 535 540

Asp Asp Glu Leu Lys Ala Leu Ala Asp Ser Phe Ser Leu Asp Glu Asn
545 550 555 560

Gly Asn Lys Leu Gly Lys Gly Lys His Gly Met Arg Asn Phe Ile Ile
565 570 575

Asn Asn Val Ile Thr Asn Lys Arg Phe His Tyr Leu Ile Arg Tyr Gly
580 585 590

Asn Pro Val His Leu His Glu Ile Ala Lys Asn Glu Ala Val Val Lys
595 600 605

Phe Val Leu Gly Arg Ile Ala Asp Ile Gln Lys Lys Gln Gly Gln Asn
610 615 620

Gly Lys Asn Gln Ile Asp Arg Tyr Tyr Glu Thr Cys Ile Gly Lys Asp
625 630 635 640

Ser Ser Lys Ser Val Ala Glu Lys Val Asn Ala Leu Thr Lys Ile Ile
645 650 655

Thr Gly Met Asn Tyr Asp Gln Phe Asp Ser Arg Arg Asn Val Ile Glu
660 665 670

Asn Thr Gly Ala Gly Asn Ala Glu Arg Glu Lys Tyr Lys Lys Ile Ile
675 680 685

Ser Leu Tyr Leu Thr Val Ile Tyr His Ile Leu Lys Asn Ile Val Asn
690 695 700

Ile Asn Ser Arg Tyr Val Ile Gly Phe His Cys Val Glu Arg Asp Ala
705 710 715 720

Gln Leu Tyr Lys Glu Lys Gly Tyr Asp Ile Asn Leu Lys Lys Leu Lys

725 730 735

Asp Lys Gly Phe Thr Ser Val Thr Lys Leu Cys Ala Gly Ile Asp Glu
740 745 750

Glu Cys Lys Asp Val Glu Lys Glu Met Thr Glu Arg Ala Lys Ala Ser
755 760 765

Phe Ala Ala Leu Glu Thr Ala Asn Pro Lys Leu Tyr Ala Thr Tyr Ile
770 775 780

Asn Tyr Ser Asp Glu Glu Lys Asn Ala Glu Leu Arg Lys Gln Ile Asn
785 790 795 800

Arg Glu Lys Ala Lys Thr Ala Leu Asn Ala His Leu Arg Asn Thr Lys
805 810 815

Trp Asn Val Ile Ile Arg Glu Asp Leu Leu Arg Arg Asp Asn Lys Ala
820 825 830

Cys Lys Ile Phe Arg Asn Lys Val Ala His Leu Glu Ala Ile Arg Tyr
835 840 845

Ala His Leu Tyr Ile Asn Asp Ile Ala Glu Val Asn Ser Tyr Phe Gln
850 855 860

Phe Tyr His Tyr Ile Met Gln Arg Arg Ile Met Ala Glu Arg Tyr Asp
865 870 875 880

Lys Ser Ser Gly Lys Val Arg Glu Tyr Phe Asp Ala Val Asn Asn Glu
885 890 895

Lys Lys Tyr Asn Asp Arg Leu Leu Lys Leu Leu Cys Val Pro Phe Gly
900 905 910

Tyr Cys Ile Pro Arg Phe Lys Asn Leu Ser Ile Glu Ala Leu Phe Asp
915 920 925

Met Asn Glu Ala Val Lys Phe Asp Lys Glu Lys Lys

930 935 940

<210> 199
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 199
caagtaaacc cctaccaagt ggtcggggtt tgaaac 36

<210> 200
<211> 1057
<212> PRT
<213> Artificial Sequence

<220>
<223> EMG_10003641

<400> 200

Met Glu Glu Thr Lys Val Thr Lys Glu Thr Thr Ile Glu Lys Gln Ser
1 5 10 15

Thr Lys Arg His Lys Gln Lys Ser Lys Lys Thr Ala Thr Lys Met Ser
20 25 30

Gly Leu Lys Ser Ala Leu Val Ile Asn Asn His Glu Met Leu Leu Thr
35 40 45

Ser Phe Gly Lys Gly Asn Asn Ala Ile Ala Glu Lys Arg Tyr Ile Leu
50 55 60

Asp Gly Asp Ile Glu Thr Ile Asn Asn Lys Asn Lys Lys Phe Asp Ala
65 70 75 80

Asn Asn Asp Ser Lys Val Val Val Ile Lys Gly Ile Ser Asn Pro Asn
85 90 95

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

```
Pro Asn Arg Thr Ser Gly Asn Asp Met Ile Gly Ile Arg Arg Met Leu
        115                 120                 125

Glu Arg Lys Tyr Phe Val His Asn Glu Glu Asn Lys Glu Phe Gln Asp
    130                 135                 140

Asn Ile Arg Ile Gln Ile Ala Tyr Cys Ile Leu Asp Ile Glu Lys Ile
145                 150                 155                 160

Leu Met Pro His Ile Asn Asn Ile Cys Phe Glu Ile Asn Asn Met Leu
                165                 170                 175

Arg Leu Glu Gly Tyr Gln Glu Asp Ser Phe Met Gly Ser Phe Asn Leu
            180                 185                 190

Tyr Lys Pro Tyr Asp Ala Phe Ile Ala Thr Thr Asp Asp Lys Glu Ser
        195                 200                 205

Ser Arg Arg Asp Asn Phe Ala Lys Leu Met Thr Ser Lys Gln Val Arg
    210                 215                 220

Tyr Leu Gly Asn Ala Leu Tyr Ser Asp Ser Leu Ser Asn Leu Thr Lys
225                 230                 235                 240

Asp Glu Ile Leu Asp Gly Lys Arg Ser Lys Glu Leu Lys Lys Tyr Tyr
                245                 250                 255

Gln Glu Leu Cys Leu Leu Gly Met Val Arg Gln Ser Met Ile His Ser
            260                 265                 270

Asn Gln Phe Asn Ser Ser Ile Tyr Thr Leu Asp Ser Ser Tyr Asp Ser
        275                 280                 285

Thr Met Asn Thr Ala Glu Leu Leu Gly Lys Gly Asp Asp Ser Ser Leu
    290                 295                 300

Val Ala Leu Ala Thr Asp Ala Arg Val Glu Ala Arg Ala Ile Leu Asp
305                 310                 315                 320
```

```
Glu Ile Tyr Lys Lys Gly Val Asp Ser Ile Asn Asn Ser Phe Leu Ser
                325                 330                 335

Asn Ser Ile Asn Asp Leu Glu Asn Leu Phe Lys Ile Tyr Lys Cys Asp
            340                 345                 350

Ser Ser Glu Lys Lys Thr Glu Leu Ile Lys Gln Tyr Tyr Asp Phe Cys
        355                 360                 365

Ile Arg Lys Pro Gln Met Asn Met Gly Phe Ser Ile Thr Thr Ile Arg
    370                 375                 380

Glu Gly Met Phe Thr Arg Cys Ser Glu Ala Asn Thr Leu Leu Leu Cys
385                 390                 395                 400

Asp Glu Gly Ser Thr Val Lys Leu Asn Val His Asp Thr Met Lys Ser
                405                 410                 415

Lys Phe Tyr Lys Asn Leu Asp Phe Met Ile Tyr Lys Tyr Tyr Lys Tyr
            420                 425                 430

Glu Asn Pro Glu Lys Gly Glu Lys Leu Ile Glu Asp Leu Arg Ser Lys
        435                 440                 445

Ile Lys Gly Lys Lys Lys Glu Asp Glu Asp Lys Lys Gln Arg Tyr Ala
    450                 455                 460

Glu Glu Ser Ala Cys Ile Leu Lys Ala Lys Arg Asp Ile Ile Lys Lys
465                 470                 475                 480

Asp Leu Thr Glu Ala Ala Asn Lys Asp Leu Phe Ala Asp Leu Val Lys
                485                 490                 495

Ser Asn Lys Asn Glu Lys Gln Lys Phe Lys Asn Glu Tyr Glu Glu Leu
            500                 505                 510

Leu Lys Pro Phe Met Ile Pro Val Lys Val Asp Tyr Phe Ser Glu Leu
        515                 520                 525
```

```
Ile Tyr Leu Val Thr Arg Phe Leu Ser Gly Lys Glu Ile Asn Asp Leu
    530                 535                 540

Leu Thr Gln Leu Ile Asn Lys Phe Glu Asn Ile Ala Ala Phe Ile Arg
545                 550                 555                 560

Met Tyr Gln Asn Asp Gln Gly Lys Leu Glu Phe Thr Ala Asn Tyr Lys
                565                 570                 575

Met Phe Glu Ile Asp Pro Gln Lys Asp Ile Pro Lys Asp Gly Lys Arg
            580                 585                 590

Val Leu Ser Gly Ser Ala Lys Ile Ala Tyr Tyr Leu Arg Thr Ile Asn
        595                 600                 605

Tyr Ile Ala Arg Met Glu Ser Phe Glu Ile Lys Ser Asp Lys Thr Ala
    610                 615                 620

Ile Asn Asp Ala Ile Ser Leu Leu Gly Tyr Asn Ser Asn Glu His Arg
625                 630                 635                 640

Asp Glu Phe Ile Thr Tyr Thr Met Ala Lys His Val Val Asp Lys Tyr
                645                 650                 655

Gln Asn Thr Asp Tyr Gln Lys Ile Val Lys Asp Phe Leu Ser Ala Asn
            660                 665                 670

Lys Thr Leu Asp Cys Lys Ser Lys Asn Met Gln Ala Phe Val Ser Glu
        675                 680                 685

Leu Lys Asn Ala His Leu Ser Glu Asn Tyr Glu Gln Arg Glu Lys Glu
    690                 695                 700

Ile Tyr Glu Leu Ala Asp Thr Asn Leu Pro Ala Tyr Phe Ser Glu Glu
705                 710                 715                 720

Asp Lys Glu Lys Leu Ala Arg Tyr Ile Val His Ser Asp Gly Thr Tyr
                725                 730                 735
```

```
Lys Lys Phe Leu Lys Glu Ser Phe Tyr Ala Ile Glu Glu Leu Pro Asn
            740                 745                 750

Glu Gly Phe Arg Asn Phe Ile Ser Asn Asn Val Ile Asn Ser Arg Arg
        755                 760                 765

Phe Asn Tyr Ile Met Arg Phe Cys Asn Pro Glu Lys Ile Ala Asn Ile
    770                 775                 780

Gly Lys Asn Lys Val Leu Ile Ser Phe Ala Leu Ser Ser Leu Ala Glu
785                 790                 795                 800

Lys Thr Asp Met Ile Ala Lys Tyr Tyr Arg Val Phe Cys Asp Arg Ile
                805                 810                 815

Asp Asp Gln Lys Thr Met Glu Asp Tyr Leu Val Asn Lys Leu Thr Lys
            820                 825                 830

Ile Ser Tyr Thr Glu Phe Leu Asn Val Asn Gln Lys Ala Asn Ala Glu
        835                 840                 845

Lys Asn Lys Glu Lys Asp Arg Ser Gln Lys Leu Ile Gly Leu Tyr Ile
    850                 855                 860

Thr Leu Leu Tyr Glu Ile Val Lys Asn Leu Val Asn Ile Asn Ser Arg
865                 870                 875                 880

Tyr Asn Ile Ala Phe Gln Arg Cys Asp Asn Asp Ser Ile Met Ile Leu
                885                 890                 895

Gln Gly Gln Tyr Asp Glu Arg Ala Val Gln Glu Ser Lys Leu Thr Lys
            900                 905                 910

Lys Phe Ile Ser Asn Gln Lys Leu Asn Ser Tyr Ser Cys Arg Tyr Leu
        915                 920                 925

Thr His Asn Ile Ser Gln Leu Asp Arg Cys Asn Asp Phe Ile Arg Gln
    930                 935                 940
```

Tyr Arg Asn Lys Val Ala His Leu Glu Val Val Ser Asn Ile Asp Glu
945 950 955 960

Tyr Leu Ser Gly Ile Lys His Ile Glu Ser Tyr Tyr Ala Leu Tyr His
965 970 975

Tyr Leu Met Gln Lys Cys Leu Leu Lys Asn Tyr Arg Ile Glu Asp His
980 985 990

Ser Gln Asn Glu Tyr Lys Asn Leu Asn Asp Phe Ser Ser Lys Leu Asp
995 1000 1005

Lys His Gly Thr Tyr Val Lys Asp Phe Val Lys Ala Leu Asn Val
1010 1015 1020

Pro Phe Gly Tyr Asn Leu Pro Arg Tyr Lys Asn Leu Ser Ile Asp
1025 1030 1035

Glu Leu Phe Asp Arg Asn Lys Leu Lys Thr Gly Gly Thr Ile Glu
1040 1045 1050

Met Lys Gly Glu
1055

<210> 201
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 201
gaactacacc cctgcagaaa tgctggggtc tgaaac 36

<210> 202
<211> 973
<212> PRT
<213> Artificial Sequence

<220>
<223> Ga0129306_1000735

<400> 202

Met Gln Lys Gln Arg Glu Gln Gln Thr Val Thr Asp Glu Ser Glu Arg
1 5 10 15

Lys Lys Lys Pro Leu Lys Ser Gly Ala Lys Ala Ala Gly Leu Lys Ser
20 25 30

Val Phe Val Leu Ser Glu Gly Lys Glu Leu Leu Thr Ser Phe Gly Arg
35 40 45

Gly Asn Glu Ala Val Pro Glu Lys Arg Val Thr Gly Gly Thr Ile Ala
50 55 60

Asn Ala Arg Thr Asp Asn Lys Glu Ala Phe Ser Ala Ala Leu Gln Asn
65 70 75 80

Lys Arg Phe Glu Val Phe Gly Arg Thr Ala Gly Ser Ser Asp Asp Pro
85 90 95

Leu Ala Val Ser Arg Ala Pro Gly Gln Asp Leu Ile Gly Ala Lys Thr
100 105 110

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

Met Gln Val Ile Tyr Ala Ile Gln Asp Ile Asn Lys Ile Leu Ala Val
130 135 140

His Ala Asn Asn Ile Val Tyr Thr Leu Asn Asn Leu Asp Arg Glu Ala
145 150 155 160

Asp Pro Glu Thr Asp Asp Phe Ile Gly Ser Gly Tyr Leu Thr Leu Lys
165 170 175

Asn Thr Phe Glu Thr Tyr Cys Asp Pro Ala Ala Leu Asn Glu Arg Glu
180 185 190

Arg Glu Lys Val Thr Val Ser Lys Gln His Phe Asp Ala Phe Met Gln

195 200 205

Asn Pro Arg Leu Ala Tyr Tyr Gly Asn Ala Phe Phe Arg Lys Leu Ser
210 215 220

Lys Ala Glu Arg Leu Ala Arg Gly Arg Glu Ile Phe Asp Lys Glu Ser
225 230 235 240

Pro Glu Arg Arg Gln Glu Ile Leu Gly Ser Arg Gly Lys Asn Lys Ser
245 250 255

Val Asp Asp Glu Ile Arg Ala Leu Ala Pro Glu Trp Val Lys Arg Glu
260 265 270

Glu Arg Asp Val Tyr Ser Glu Leu Val Leu Met Ser Glu Leu Arg Gln
275 280 285

Ser Cys Phe His Gly Gln Gln Lys Asn Ser Ala Arg Ile Phe Arg Leu
290 295 300

Asp Asn Asp Leu Gly Pro Gly Val Asp Gly Ala Arg Glu Leu Leu Asp
305 310 315 320

Arg Leu Tyr Ala Glu Lys Ile Asn Asp Leu Arg Ser Phe Asp Lys Thr
325 330 335

Ser Ala Ser Ser Asn Phe Arg Leu Leu Phe Asn Ala Tyr His Ala Asp
340 345 350

Asn Glu Lys Lys Lys Glu Leu Ala Gln Glu Phe Tyr Arg Phe Ser Val
355 360 365

Leu Lys Val Ser Lys Asn Thr Gly Phe Ser Ile Arg Thr Leu Arg Glu
370 375 380

Lys Ile Ile Glu Asp His Ala Ala Gln Tyr Arg Asp Lys Ile Tyr Asp
385 390 395 400

Ser Met Arg Lys Lys Leu Phe Ser Thr Phe Asp Phe Phe Leu Trp Arg

405 410 415

Phe Tyr Glu Glu Arg Glu Asp Glu Ala Glu Glu Leu Arg Ala Cys Leu
420 425 430

Arg Ala Ala Arg Ser Asp Glu Glu Lys Glu Gln Ile Tyr Ala Glu Ala
435 440 445

Ala Ala Ser Cys Trp Pro Ser Val Lys Pro Phe Val Glu Ser Val Ala
450 455 460

Ala Thr Leu Cys Asp Val Val Lys Gly Arg Thr Lys Leu Asn Lys Leu
465 470 475 480

Lys Leu Ser Ala Asp Glu Ser Thr Leu Val Arg Asn Ala Ile Asp Gly
485 490 495

Val Arg Ile Ser Pro Arg Ala Ser Tyr Phe Thr Lys Leu Ile Tyr Leu
500 505 510

Met Thr Leu Phe Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr Thr
515 520 525

Leu Ile His Ala Phe Glu Asn Ile Asp Ser Phe Leu Ser Val Leu Gly
530 535 540

Ser Glu Arg Leu Glu Arg Thr Phe Asp Ala Asn Tyr Arg Ile Phe Ala
545 550 555 560

Asp Ser Gly Val Ile Ala Gln Glu Leu Arg Ala Val Asn Ser Phe Ala
565 570 575

Arg Met Thr Thr Glu Pro Phe Asn Ser Lys Leu Val Met Phe Glu Asp
580 585 590

Ala Ala Gln Leu Phe Gly Met Ser Gly Gly Leu Val Glu His Ala Glu
595 600 605

Glu Leu Arg Glu Tyr Leu Asp Asn Lys Met Leu Asp Lys Thr Lys Leu

```
    610                 615                 620

Arg Leu Leu Pro Asp Gly Lys Val Asp Thr Gly Phe Arg Asn Phe Ile
625                 630                 635                 640

Ile Ser Asn Val Thr Glu Ser Arg Arg Phe Arg Tyr Leu Val Arg Tyr
                645                 650                 655

Cys Glu Pro Arg Ala Val Arg Asp Tyr Met Ser Cys Arg Pro Leu Ile
            660                 665                 670

Arg Leu Thr Leu Arg Asp Met Pro Asp Thr Ile Leu Arg Arg Tyr Tyr
        675                 680                 685

Glu Gln Ser Val Gly Ala Ala Thr Val Asp Arg Glu Arg Ile Leu Asp
    690                 695                 700

Thr Leu Ala Asp Lys Leu Leu Ser Leu Arg Phe Thr Asp Phe Glu Asn
705                 710                 715                 720

Val Asn Gln Arg Ala Asn Ala Glu Arg Asn Arg Glu Lys Gln Lys Met
                725                 730                 735

Met Gly Ile Ile Ser Leu Tyr Leu Asn Val Ala Tyr Gln Ile Val Lys
            740                 745                 750

Asn Leu Val Tyr Val Asn Ala Arg Tyr Thr Met Ala Tyr His Cys Ala
        755                 760                 765

Glu Arg Asp Thr Glu Leu Leu Leu Asn Ala Ala Gly Glu Gly Asn Leu
    770                 775                 780

Leu Arg Arg Asp Arg Ser Trp Pro Ala Arg Leu His Leu Pro Arg Arg
785                 790                 795                 800

Ala Leu Ala Arg Arg Arg Asp Arg Val Glu Val Met Glu Arg Asp Val
                805                 810                 815

Ala Arg Gly Pro Glu Ala Tyr Asn Arg Asp Glu Trp Leu Gly Leu Val
```

820 825 830

Arg Thr Leu Arg Arg Glu Lys Arg Val Cys Asp Asn Leu His Asn Asn
835 840 845

Tyr Ala Tyr Leu Cys Gly Ala Asp Ala Glu Pro Gly Asp Ala Ser Leu
850 855 860

Ser Leu Leu Phe Val Tyr Arg Asn Lys Ala Ala His Leu Ser Val Leu
865 870 875 880

Asn Lys Gly Gly Arg Leu Ser Gly Asp Leu Lys Glu Ala Lys Ser Trp
885 890 895

Phe Tyr Val Tyr His Phe Leu Met Gln Arg Val Leu Glu Glu Glu Phe
900 905 910

Arg Asn Thr Gln Ala Leu Pro Glu Arg Leu Arg Glu Leu Leu Met Met
915 920 925

Ala Glu Arg Tyr Arg Gly Cys Ser Lys Asp Leu Ile Lys Val Leu Asn
930 935 940

Leu Thr Phe Ala Tyr Asn Leu Pro Arg Tyr Lys Asn Leu Ser Ile Asp
945 950 955 960

Gly Arg Phe Asp Lys Asn His Pro Asp Pro Ser Asp Glu
965 970

<210> 203
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 203
ccactacacc cctgtcgaac aacgggggtc taaaac 36

<210> 204

<211> 854
<212> PRT
<213> Artificial Sequence

<220>
<223> Ga0129317_1008067

<400> 204

Met Lys Lys Gln Lys Lys Ser Leu Val Lys Ala Ala Gly Leu Lys Ser
1 5 10 15

Ala Phe Val Val Gly Asp Ser Val Tyr Leu Thr Ser Phe Gly Lys Gly
20 25 30

Asn Ala Ala Arg Leu Asp Thr Lys Ile Asn Pro Asp Asn Ser Thr Glu
35 40 45

Arg Tyr Val Ser Asp Ser Glu Lys His Thr Leu Lys Ile Asn Ser Ile
50 55 60

Thr Asp Thr Glu Leu Arg Leu Ser Gly Pro Phe Pro Lys Gln Ala Glu
65 70 75 80

Ala Lys Asn Pro Thr His Lys Lys Asp Asn Glu Gln Lys Asn Thr Arg
85 90 95

Gln Asp Met Leu Gly Leu Lys Ser Thr Leu Glu Lys Phe Tyr Phe Gly
100 105 110

Ser Thr Phe Asp Asp Asn Ile His Ile Gln Ile Ile His Asn Ile Gln
115 120 125

Asp Ile Ala Lys Ile Leu Ala Ala His Ser Asn Asn Ala Gly Tyr Ala
130 135 140

Leu Asp Asn Met Leu Ala Tyr Gln Gly Val Glu Phe Ser Asp Met Ile
145 150 155 160

Gly Tyr Met Gly Thr Ser Arg Thr Phe Asp Asn Tyr Asp Pro Asn His
165 170 175

```
Lys Asn Asn Lys Asp Phe Phe Arg Phe Leu Lys Leu Pro Arg Leu Gly
            180                 185                 190

Tyr Phe Gly Ser Ala Phe Tyr Ser Gln Lys Gly Lys Asp Phe Glu Lys
        195                 200                 205

Arg Ser Asp Glu Glu Val Tyr Asn Ile Cys Ala Leu Met Gly Gln Ile
    210                 215                 220

Arg Gln Cys Cys Phe His Gly Lys Gln Glu Lys Tyr Gln Leu Lys Trp
225                 230                 235                 240

Leu Tyr Asn Phe His Asn Phe Lys Ser Asn Lys Pro Phe Leu Asp Thr
                245                 250                 255

Leu Asp Lys His Phe Asp Glu Met Ile Asp Arg Ile Asn Lys Asn Phe
            260                 265                 270

Ile Lys Asn Asn Thr Pro Asp Leu Ile Ile Leu Ser Gly Leu Tyr Pro
        275                 280                 285

Asp Met Ala Lys Lys Glu Leu Val Arg Leu Phe Tyr Asp Phe Thr Thr
    290                 295                 300

Val Lys Glu Tyr Lys Asn Met Gly Phe Ser Val Lys Lys Leu Arg Glu
305                 310                 315                 320

Lys Met Leu Glu Ser Glu Glu Ala Ser Asp Phe Arg Asp Lys Asp Tyr
                325                 330                 335

Asp Ser Val Arg Arg Lys Leu Tyr Lys Leu Met Asp Phe Cys Ile Tyr
            340                 345                 350

Tyr Leu Tyr Tyr Ser Asp Ser Glu Arg Asn Glu Asn Leu Val Ser Arg
        355                 360                 365

Leu Arg Glu Ser Leu Thr Asp Glu Asn Lys Asp Ile Ile Tyr Ser Lys
    370                 375                 380
```

```
Glu Ala Lys Ile Val Trp Asn Glu Leu Arg Lys Lys Phe Ser Thr Ile
385                 390                 395                 400


Leu Asp Asn Val Lys Gly Ser Asn Ile Lys Lys Leu Glu Asn Val Lys
                405                 410                 415


Glu Lys Phe Ile Ser Glu Asp Glu Phe Asp Asp Ile Lys Leu Asp Ile
            420                 425                 430


Asp Ile Ser Tyr Phe Ser Lys Leu Met Tyr Val Met Cys Tyr Phe Leu
        435                 440                 445


Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr Thr Leu Val Ser Lys Phe
    450                 455                 460


Asp Asn Ile Gly Ser Ile Ile Glu Ala Ala Thr Gln Ile Gly Ile Asn
465                 470                 475                 480


Ile Glu Phe Ile Asp Asp Phe Lys Phe Phe Asp Arg Ser Lys Asp Ile
                485                 490                 495


Ser Val Glu Leu Asn Ile Ile Arg Asn Phe Ala Arg Met Gln Ala Pro
            500                 505                 510


Val Pro Asn Ala Lys Arg Ala Met Gln Glu Asp Ala Ile Arg Ile Leu
        515                 520                 525


Gly Gly Ser Glu Glu Asp Ile Phe Ser Ile Leu Asp Asp Met Thr Gly
    530                 535                 540


Tyr Asp Lys Ser Gly Lys Lys Leu Ala Gln Ser Lys Lys Gly Phe Arg
545                 550                 555                 560


Asn Phe Ile Ile Asn Asn Val Val Glu Ser Ser Arg Phe Lys Tyr Ile
                565                 570                 575


Val Arg Tyr Ser Asn Pro Gln Lys Ile Arg Lys Leu Ala Asn Asn Ser
            580                 585                 590
```

```
Val Val Val Gly Phe Val Leu Gly Lys Leu Pro Asp Ala Gln Ile Glu
        595                 600                 605

Ser Tyr Phe Asn Ser Cys Leu Pro Asn Arg Val Tyr Ser Thr Pro Asp
    610                 615                 620

Lys Ala Arg Glu Ser Leu Arg Asp Met Leu His Asn Ile Ser Phe Asn
625                 630                 635                 640

Asp Phe Ala Asp Val Lys Gln Asp Asp Arg Arg Ala Thr Pro Glu Glu
                645                 650                 655

Lys Val Glu Lys Glu Arg Tyr Lys Ala Ile Ile Gly Leu Tyr Leu Thr
            660                 665                 670

Val Met Tyr His Leu Val Lys Asn Leu Val Tyr Val Asn Ser Arg Tyr
        675                 680                 685

Val Met Ala Phe His Cys Leu Glu Arg Asp Ala Met His Tyr Asp Val
    690                 695                 700

Ser Leu Asp Asn Tyr Arg Asp Leu Ile Arg His Leu Ile Ser Glu Gly
705                 710                 715                 720

Asp Ser Ser Cys Asn His Phe Ile Ser His Asn Arg Arg Met Arg Asp
                725                 730                 735

Cys Ile Glu Glu Asn Val Lys Asn Ser Glu Gln Leu Ile Phe Gly Lys
            740                 745                 750

Glu Asp Ala Val Ile Arg Phe Arg Asn Asn Val Ala His Leu Ser Ala
        755                 760                 765

Ile Arg Asn Ala Asn Glu Tyr Ile Gly Asp Ile Arg Glu Ile Thr Ser
    770                 775                 780

Tyr Phe Ala Leu Tyr His Tyr Leu Met Gln Arg Lys Leu Ile Asp Asp
785                 790                 795                 800
```

Cys Lys Val Asn Asp Thr Ala His Lys Tyr Phe Glu Gln Leu Thr Lys
805 810 815

Tyr Lys Thr Tyr Val Met Asp Met Val Lys Ala Leu Cys Ser Pro Phe
820 825 830

Gly Tyr Asn Leu Pro Arg Phe Lys Asn Leu Ser Ile Glu Gly Lys Phe
835 840 845

Asp Met His Glu Ser Lys
850

<210> 205
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 205
gaactacact actgtaattt tgcagtagtc taaaac 36

<210> 206
<211> 965
<212> PRT
<213> Artificial Sequence

<220>
<223> Ga0224415_10048792

<400> 206

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

Lys Ser Thr Leu Val Ser Asp Ser Lys Val Tyr Leu Thr Thr Phe Ala
20 25 30

Asp Gly Ser Asn Ala Lys Leu Glu Lys Cys Val Glu Asn Asn Lys Ile
35 40 45

Ile Cys Ile Ser Asn Asp Lys Glu Ala Phe Ala Ala Ser Ile Ala Asn

```
    50                  55                  60

Lys Asn Val Gly Tyr Lys Ile Lys Asn Asp Glu Lys Phe Arg His Pro
65                  70                  75                  80

Lys Gly Tyr Asp Ile Ile Ser Asn Asn Pro Leu Leu His Asn Asn Ser
                85                  90                  95

Val Gln Gln Asp Met Leu Gly Leu Lys Asn Val Leu Glu Lys Arg Tyr
            100                 105                 110

Phe Gly Lys Ser Ser Gly Gly Asp Asn Asn Leu Cys Ile Gln Ile Ile
        115                 120                 125

His Asn Ile Ile Asp Ile Glu Lys Ile Leu Ser Glu Tyr Ile Pro Asn
    130                 135                 140

Val Val Tyr Ala Phe Asn Asn Ile Ala Gly Phe Lys Asp Glu His Asn
145                 150                 155                 160

Asn Ile Ile Asp Ile Ile Gly Thr Gln Thr Tyr Asn Ser Ser Tyr Thr
                165                 170                 175

Tyr Ala Asp Phe Ser Lys Asp Lys Ser Asp Lys Lys Tyr Ile Glu Phe
            180                 185                 190

Gln Lys Leu Leu Lys Asn Lys Arg Leu Gly Tyr Trp Gly Lys Ala Phe
        195                 200                 205

Phe Thr Gly Gln Gly Asn Asn Ala Lys Val Arg Gln Glu Asn Gln Cys
    210                 215                 220

Phe His Ile Ile Ala Leu Leu Ile Ser Leu Arg Asn Trp Ala Thr His
225                 230                 235                 240

Ser Asn Glu Leu Asp Lys His Thr Lys Arg Thr Trp Leu Tyr Lys Leu
                245                 250                 255

Asp Asp Thr Asn Ile Leu Asn Ala Glu Tyr Val Lys Thr Leu Asn Tyr
```

260 265 270

Leu Tyr Asp Thr Ile Ala Asp Glu Leu Thr Lys Ser Phe Ser Lys Asn
275 280 285

Gly Ala Val Asn Val Asn Tyr Leu Ala Lys Lys Tyr Asn Ile Lys Asp
290 295 300

Asp Leu Pro Gly Phe Ser Glu Gln Tyr Phe Arg Phe Ser Ile Met Lys
305 310 315 320

Glu Gln Lys Asn Leu Gly Phe Asn Ile Ser Lys Leu Arg Glu Asn Met
325 330 335

Leu Asp Phe Lys Asp Met Ser Val Ile Arg Asp Asp His Asn Arg Tyr
340 345 350

Asp Lys Asp Arg Ser Lys Ile Tyr Thr Met Met Asp Phe Val Ile Tyr
355 360 365

Arg Tyr Tyr Ile Asp Asn Asn Asn Asp Ser Ile Asp Phe Ile Asn Lys
370 375 380

Leu Arg Ser Ser Ile Asp Glu Lys Ser Lys Glu Lys Leu Tyr Asn Glu
385 390 395 400

Glu Ala Asn Arg Leu Trp Asn Lys Leu Lys Glu Tyr Met Leu Tyr Ile
405 410 415

Lys Glu Phe Asn Gly Lys Leu Ala Ser Arg Thr Pro Asp Arg Asp Gly
420 425 430

Asn Ile Ser Glu Phe Val Glu Ser Leu Pro Lys Ile His Arg Leu Leu
435 440 445

Pro Arg Gly Gln Lys Ile Ser Asn Phe Ser Lys Leu Met Tyr Leu Leu
450 455 460

Thr Met Phe Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr Thr Leu

465 470 475 480

Ile Asn Lys Phe Glu Asn Ile Gln Gly Phe Leu Asp Ile Met Pro Glu
485 490 495

Ile Asn Val Asn Ala Lys Phe Glu Pro Glu Tyr Val Phe Phe Asn Lys
500 505 510

Ser His Glu Ile Ala Gly Glu Leu Lys Leu Ile Lys Gly Phe Ala Gln
515 520 525

Met Gly Glu Pro Ala Ala Thr Leu Lys Leu Glu Met Thr Ala Asp Ala
530 535 540

Ile Lys Ile Leu Gly Thr Glu Lys Glu Asp Ala Glu Leu Ile Lys Leu
545 550 555 560

Ala Glu Ser Leu Phe Lys Asp Glu Asn Gly Lys Leu Leu Gly Asn Lys
565 570 575

Gln His Gly Met Arg Asn Phe Ile Gly Asn Asn Val Ile Lys Ser Lys
580 585 590

Arg Phe His Tyr Leu Ile Arg Tyr Gly Asp Pro Ala His Leu His Lys
595 600 605

Ile Ala Thr Asn Lys Asn Val Val Arg Phe Val Leu Gly Arg Ile Ala
610 615 620

Asp Met Gln Lys Lys Gln Gly Gln Lys Gly Lys Asn Gln Ile Asp Arg
625 630 635 640

Tyr Tyr Glu Val Cys Val Gly Asn Lys Asp Ile Lys Lys Thr Ile Glu
645 650 655

Glu Lys Ile Asp Ala Leu Thr Asp Ile Ile Val Asn Met Asn Tyr Asp
660 665 670

Gln Phe Glu Lys Lys Lys Ala Val Ile Glu Asn Gln Asn Arg Gly Lys

675 680 685

Thr Phe Glu Glu Lys Asn Lys Tyr Lys Arg Asp Asn Ala Glu Arg Glu
690 695 700

Lys Phe Lys Lys Ile Ile Ser Leu Tyr Leu Thr Val Ile Tyr His Ile
705 710 715 720

Leu Lys Asn Ile Val Asn Val Asn Ser Arg Tyr Ile Leu Gly Phe His
725 730 735

Cys Leu Glu Arg Asp Lys Gln Leu Tyr Ile Glu Lys Tyr Asn Lys Asp
740 745 750

Lys Leu Asp Gly Phe Val Ala Leu Thr Lys Phe Cys Leu Gly Asp Glu
755 760 765

Glu Arg Phe Glu Asp Leu Lys Ala Lys Ala Gln Ala Ser Ile Gln Ala
770 775 780

Leu Glu Thr Ala Asn Pro Lys Leu Tyr Ala Lys Tyr Met Asn Tyr Ser
785 790 795 800

Asp Glu Glu Lys Lys Glu Glu Phe Lys Lys Gln Leu Asn Arg Glu Arg
805 810 815

Val Lys Asn Ala Arg Asn Ala Tyr Leu Lys Asn Ile Lys Asn Tyr Ile
820 825 830

Met Ile Arg Leu Gln Leu Arg Asp Gln Thr Asp Ser Ser Gly Tyr Leu
835 840 845

Cys Gly Glu Phe Arg Asp Lys Val Ala His Leu Glu Val Ala Arg His
850 855 860

Ala His Glu Tyr Ile Gly Asn Ile Lys Glu Val Asn Ser Tyr Phe Gln
865 870 875 880

Leu Tyr His Tyr Ile Met Gln Cys Arg Leu Tyr Asp Val Leu Lys Asn

885 890 895

Asn Thr Lys Ala Glu Ala Met Val Lys Gly Lys Ala Lys Glu Tyr Phe
900 905 910

Glu Ala Leu Glu Lys Glu Gly Thr Tyr Asn Asp Lys Leu Leu Lys Ile
915 920 925

Ala Cys Val Pro Phe Gly Tyr Cys Ile Pro Arg Tyr Lys Asn Leu Ser
930 935 940

Met Glu Glu Leu Phe Asp Met Asn Glu Glu Lys Lys Phe Lys Lys Lys
945 950 955 960

Ala Pro Glu Asn Thr
965

<210> 207
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 207
caagtaaacc cctaccaaat gtaaggggtt tgaaac 36

<210> 208
<211> 1022
<212> PRT
<213> Artificial Sequence

<220>
<223> 160582958 _gene49834

<400> 208

Met Lys Asn Ser Val Thr Phe Lys Leu Ile Gln Ala Gln Glu Asn Lys
1 5 10 15

Glu Ala Ala Arg Lys Lys Ala Lys Asp Ile Ala Glu Gln Ala Arg Ile
20 25 30

```
Ala Lys Arg Asn Gly Val Val Lys Lys Glu Glu Asn Arg Ile Asn Arg
        35                  40                  45

Ile Gln Ile Glu Ile Gln Thr Gln Lys Lys Ser Asn Thr Gln Asn Ala
    50                  55                  60

Tyr His Leu Lys Ser Leu Ala Lys Ala Ala Gly Val Lys Ser Val Phe
65                  70                  75                  80

Ala Ile Gly Asn Asp Leu Leu Met Thr Gly Phe Gly Pro Gly Asn Asp
                85                  90                  95

Ala Thr Ile Glu Lys Arg Val Phe Gln Asn Arg Ala Ile Glu Thr Leu
            100                 105                 110

Ser Ser Pro Glu Gln Tyr Ser Ala Glu Phe Gln Asn Lys Gln Phe Lys
        115                 120                 125

Ile Lys Gly Asn Ile Lys Val Leu Asn His Ser Thr Gln Lys Met Glu
    130                 135                 140

Glu Ile Gln Thr Glu Leu Gln Asp Asn Tyr Asn Arg Pro His Phe Asp
145                 150                 155                 160

Leu Leu Gly Cys Lys Asn Val Leu Glu Gln Lys Tyr Phe Gly Arg Thr
                165                 170                 175

Phe Ser Asp Asn Ile His Val Gln Ile Ala Tyr Asn Ile Met Asp Ile
            180                 185                 190

Glu Lys Leu Leu Thr Pro Tyr Ile Asn Asn Ile Ile Tyr Thr Leu Asn
        195                 200                 205

Glu Leu Met Arg Asp Asn Ser Lys Asp Asp Phe Phe Gly Cys Asp Ser
    210                 215                 220

His Phe Ser Val Ala Tyr Leu Tyr Asp Glu Leu Lys Ala Gly Tyr Ser
225                 230                 235                 240
```

```
Asp Arg Leu Lys Thr Lys Pro Asn Leu Ser Lys Asn Ile Asp Arg Ile
                245                 250                 255

Trp Asn Asn Phe Cys Asn Tyr Met Asn Ser Asp Ser Gly Asn Thr Glu
            260                 265                 270

Ala Arg Leu Ala Tyr Phe Gly Glu Leu Phe Tyr Lys Pro Lys Glu Thr
        275                 280                 285

Gly Asp Ala Lys Ser Asp Tyr Lys Thr His Leu Ser Asn Asn Gln Lys
    290                 295                 300

Glu Glu Trp Glu Leu Lys Ser Asp Lys Glu Val Tyr Asn Ile Phe Ala
305                 310                 315                 320

Ile Leu Cys Asp Leu Arg His Phe Cys Thr His Gly Glu Ser Ile Thr
                325                 330                 335

Pro Ser Gly Lys Pro Phe Pro Tyr Asn Leu Glu Lys Asn Leu Phe Pro
            340                 345                 350

Glu Ala Lys Gln Val Leu Asn Ser Leu Phe Glu Glu Lys Ala Glu Ser
        355                 360                 365

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

Ile Leu Leu Lys Val Phe Glu Lys Glu Gln Ala Ser Gln Lys Glu Gln
385                 390                 395                 400

Gln Ala Leu Leu Lys Glu Tyr Tyr Asp Phe Lys Val Gln Lys Thr Tyr
                405                 410                 415

Lys Asn Met Gly Phe Ser Ile Lys Lys Leu Arg Glu Ala Ile Met Glu
            420                 425                 430

Ile Pro Asp Ala Ala Lys Phe Lys Asp Asp Leu Tyr Ser Ser Leu Arg
        435                 440                 445
```

```
His Lys Leu Tyr Gly Leu Phe Asp Phe Ile Leu Val Lys His Phe Leu
    450                 455                 460

Asp Thr Ser Asp Ser Glu Asn Leu Gln Asn Asn Asp Ile Phe Arg Gln
465                 470                 475                 480

Leu Arg Ala Cys Arg Cys Glu Glu Glu Lys Asp Gln Val Tyr Arg Ser
                485                 490                 495

Ile Ala Val Lys Val Trp Glu Lys Val Lys Lys Lys Glu Leu Asn Met
            500                 505                 510

Phe Lys Gln Val Val Val Ile Pro Ser Leu Ser Lys Asp Glu Leu Lys
        515                 520                 525

Gln Met Glu Met Thr Lys Asn Thr Glu Leu Leu Ser Ser Ile Glu Thr
    530                 535                 540

Ile Ser Thr Gln Ala Ser Leu Phe Ser Glu Met Ile Phe Met Met Thr
545                 550                 555                 560

Tyr Leu Leu Asp Gly Lys Glu Ile Asn Leu Leu Cys Thr Ser Leu Ile
                565                 570                 575

Glu Lys Phe Glu Asn Ile Ala Ser Phe Asn Glu Val Leu Lys Ser Pro
            580                 585                 590

Gln Ile Gly Tyr Glu Thr Lys Tyr Thr Glu Gly Tyr Ala Phe Phe Lys
        595                 600                 605

Asn Ala Asp Lys Thr Ala Lys Glu Leu Arg Gln Val Asn Asn Met Ala
    610                 615                 620

Arg Met Thr Lys Pro Leu Gly Gly Val Asn Thr Lys Cys Val Met Tyr
625                 630                 635                 640

Asn Glu Ala Ala Lys Ile Leu Gly Ala Lys Pro Met Ser Lys Ala Glu
                645                 650                 655
```

```
Leu Glu Ser Val Phe Asn Leu Asp Asn His Asp Tyr Thr Tyr Ser Pro
            660                 665                 670

Ser Gly Lys Lys Ile Pro Asn Lys Asn Phe Arg Asn Phe Ile Ile Asn
        675                 680                 685

Asn Val Ile Thr Ser Arg Arg Phe Leu Tyr Leu Ile Arg Tyr Gly Asn
    690                 695                 700

Pro Glu Lys Ile Arg Lys Ile Ala Ile Asn Pro Ser Ile Ile Ser Phe
705                 710                 715                 720

Val Leu Lys Gln Ile Pro Asp Glu Gln Ile Lys Arg Tyr Tyr Pro Pro
                725                 730                 735

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

Gly Lys Met Leu Gln Ser Val Asn Phe Glu Gln Phe Ser Arg Val Asn
        755                 760                 765

Asn Lys Gln Asn Ala Lys Gln Asn Pro Asn Gly Glu Lys Ala Arg Leu
    770                 775                 780

Gln Ala Cys Val Arg Leu Tyr Leu Thr Val Pro Tyr Leu Phe Ile Lys
785                 790                 795                 800

Asn Met Val Asn Ile Asn Ala Arg Tyr Val Leu Ala Phe His Cys Leu
                805                 810                 815

Glu Arg Asp His Ala Leu Cys Phe Asn Ser Arg Lys Leu Asn Asp Asp
            820                 825                 830

Ser Tyr Asn Glu Met Ala Asn Lys Phe Gln Met Val Arg Lys Ala Lys
        835                 840                 845

Lys Glu Gln Tyr Glu Lys Glu Tyr Lys Cys Lys Lys Gln Glu Thr Gly
    850                 855                 860
```

```
Thr Ala His Thr Lys Lys Ile Glu Lys Leu Asn Gln Gln Ile Ala Tyr
865                 870                 875                 880


Ile Asp Lys Asp Ile Lys Asn Met His Ser Tyr Thr Cys Arg Asn Tyr
                885                 890                 895


Arg Asn Leu Val Ala His Leu Asn Val Val Ser Lys Leu Gln Asn Tyr
            900                 905                 910


Val Ser Glu Leu Pro Asn Asp Tyr Gln Ile Thr Ser Tyr Phe Ser Phe
        915                 920                 925


Tyr His Tyr Cys Met Gln Leu Gly Leu Met Glu Lys Val Ser Ser Lys
    930                 935                 940


Asn Ile Pro Leu Val Glu Ser Leu Lys Asn Glu Ala Asn Asp Ala Gln
945                 950                 955                 960


Ser Tyr Ser Ala Lys Lys Thr Leu Glu Tyr Phe Asp Leu Ile Glu Lys
                965                 970                 975


Asn Arg Thr Tyr Cys Lys Asp Phe Leu Lys Ala Leu Asn Ala Pro Phe
            980                 985                 990


Ser Tyr Asn Leu Pro Arg Phe Lys  Asn Leu Ser Ile Glu  Ala Leu Phe
        995                 1000                1005


Asp Lys  Asn Ile Val Tyr Glu  Gln Ala Asp Leu Lys  Lys Glu
    1010                1015                1020


<210>  209
<211>  36
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  DR

<400>  209
gaactacacc cctctgttct tgtaggggtc taacac                                  36
```

<210> 210
<211> 998
<212> PRT
<213> Artificial Sequence

<220>
<223> 250twins_35838_GL0110300

<400> 210

```
Met Gly Asn Lys Gln Arg Val Ser Ala Gln Lys Arg Arg Glu Asn Ala
1               5                   10                  15


Lys Leu Cys Asn Gln Gln Lys Ala Arg Gln Ala Glu Ser Gln Arg Asp
            20                  25                  30


Lys Ile Lys Asn Met Asn Val Glu Lys Met Lys Asn Ile Asn Thr Asn
        35                  40                  45


Asp Ile Lys His Thr Lys Thr Thr Ala Lys Lys Leu Gly Leu Lys Ser
    50                  55                  60


Thr Ile Ile Ala Asp Lys Lys Ile Ile Leu Thr Ser Phe Ile Asn Glu
65                  70                  75                  80


Gln Ser Ser Lys Thr Ala Asn Ile Glu Lys Val Ala Gly Phe Lys Gly
                85                  90                  95


Asp Thr Ile Asp Thr Ile Ser Tyr Thr Pro Arg Met Phe Arg Ser Glu
            100                 105                 110


Ile Asn Pro Gly Glu Ile Val Ile Ser Lys Gly Asp Asp Leu Ser Glu
        115                 120                 125


Phe Ala Asn Pro Ala Asn Phe Pro Ile Gly Arg Asp Tyr Val Lys Ile
    130                 135                 140


Arg Ser Ala Leu Glu Lys Gln Tyr Phe Gly Lys Glu Phe Pro Glu Asp
145                 150                 155                 160


Asn Leu His Val Gln Ile Ala Tyr Asn Val Ala Asp Ile Lys Lys Ile
```

```
                165                 170                 175

Leu Ser Val Tyr Ile Asn Asn Ile Ile Tyr Met Phe Tyr Asn Leu Ala
            180                 185                 190

Arg Ser Glu Glu Tyr Asp Ile Phe Tyr Asn Ser Gln Ser Glu Asn Ser
        195                 200                 205

Gly Arg Asp Cys Asp Val Ile Gly Ser Leu Tyr Tyr Gln Ala Ser Tyr
    210                 215                 220

Arg Asn Gln Asp Ala Asn Arg Phe Glu Lys Asp Gly Lys Lys Lys Ala
225                 230                 235                 240

Ile Asp Ser Leu Leu Asp Asp Thr Arg Ala Tyr Tyr Thr Tyr Phe Asp
                245                 250                 255

Gly Leu Phe Ser Val Pro Lys Arg Glu Asp Asp Gly Lys Ile Lys Glu
            260                 265                 270

Ser Glu Lys Glu Lys Ala Lys Asp Gln Asn Phe Asp Val Leu Arg Leu
        275                 280                 285

Leu Ser Val Gly Arg Gln Leu Thr Phe His Ser Asp Lys Ser Asn Asn
    290                 295                 300

Glu Ala Tyr Leu Phe Asp Leu Ser Lys Leu Thr Arg Ala Ala Gln Asp
305                 310                 315                 320

Glu Asn Arg Arg Gln Asp Ile Gln Ser Leu Leu Asn Ile Leu Asn Ser
                325                 330                 335

Thr Cys Arg Ser Asn Leu Glu Gly Val Asn Gly Asp Phe Val Lys His
            340                 345                 350

Ala Lys Asn Asn Leu Tyr Val Leu Asn Gln Leu Tyr Pro Ser Leu Lys
        355                 360                 365

Ala Asn Asp Leu Ile Gly Glu Tyr Tyr Asn Phe Ile Val Lys Lys Glu
```

```
    370                 375                 380

Asn Arg Asn Ile Gly Ile Arg Leu Ile Thr Val Arg Glu Leu Ile Ile
385                 390                 395                 400

Glu His Asn Tyr Thr Asn Leu Lys Asp Ser Lys Tyr Asp Thr Tyr Arg
                405                 410                 415

Asn Lys Ile Tyr Thr Val Leu Asn Phe Ile Leu Phe Arg Glu Ile Gln
            420                 425                 430

Glu Asn Ser Ile Ala Ile Lys Asn Phe Arg Glu Lys Leu Arg Ser Thr
        435                 440                 445

Glu Lys Ala Glu Gln Pro Ala Leu Tyr Gln Ala Phe Ala Asn Lys Ile
    450                 455                 460

Tyr Pro Met Val Gln Ala Lys Phe Ala Lys Ala Ile Asp Leu Phe Glu
465                 470                 475                 480

Glu Gln Tyr Lys Thr Lys Phe Lys Ser Glu Phe Lys Gly Gly Ile Ser
                485                 490                 495

Ile Glu Asn Met Gln Gln Gln Asn Ile Leu Leu Gln Thr Glu Asn Ile
            500                 505                 510

Asp Tyr Phe Ser Lys Tyr Val Leu Phe Leu Thr Lys Phe Leu Asp Gly
        515                 520                 525

Lys Glu Ile Asn Glu Leu Leu Cys Ala Leu Ile Asn Lys Phe Asp Asn
    530                 535                 540

Ile Ala Asp Leu Leu Asp Ile Ser Lys Gln Ile Gly Thr Pro Val Val
545                 550                 555                 560

Phe Cys Ala Asp Tyr Glu Ser Leu Asn Asp Ala Ala Lys Ile Ala Glu
                565                 570                 575

Asn Ile Arg Leu Ile Lys Asn Ile Ala His Leu Arg Pro Ala Ile Gln
```

580 585 590

Glu Ala Gln Ser Ser Lys Asp Asn Ala Asp Ala Ala Gly Thr Pro Ala
595 600 605

Thr Leu Leu Ile Asp Ala Tyr Asn Met Leu Asn Thr Asp Ile Gln Leu
610 615 620

Val Tyr Gly Glu Ala Ala Tyr Glu Glu Leu Arg Lys Asp Leu Phe Glu
625 630 635 640

Arg Lys Asn Gly Thr Lys Tyr Asn Lys Lys Gly Lys Lys Val Asp Val
645 650 655

Tyr Asp His Lys Phe Arg Asn Phe Leu Ile Asn Asn Val Ile Lys Ser
660 665 670

Lys Trp Phe Phe Tyr Ile Ala Lys Tyr Val Lys Pro Ala Asp Cys Ala
675 680 685

Lys Met Met Ser Asn Lys Lys Met Ile Glu Phe Ala Leu Arg Asp Leu
690 695 700

Pro Glu Thr Gln Ile Lys Arg Tyr Tyr Tyr Thr Ile Thr Gly Asn Glu
705 710 715 720

Ala Leu Gly Asp Ala Glu Ser Leu Lys Gly Val Ile Ile Glu Gln Leu
725 730 735

His Ala Phe Ser Ile Lys Asn Thr Leu Leu Ser Ile Lys Asn Met Gly
740 745 750

Glu Gly Glu Tyr Lys Ile Gln Gln Ile Gly Ser Ser Lys Glu Lys Leu
755 760 765

Lys Ala Ile Val Asn Leu Tyr Leu Thr Val Ala Tyr Leu Leu Thr Lys
770 775 780

Ser Leu Val Lys Val Asn Ile Arg Phe Ser Ile Ala Phe Gly Cys Leu

785 790 795 800

Glu Arg Asp Leu Val Leu Gln Lys Lys Ser Glu Lys Lys Phe Asp Ala
805 810 815

Ile Ile Asn Glu Ile Leu Leu Glu Asp Asp Lys Ile Arg Lys Glu Cys
820 825 830

Asp Lys Glu Arg Ala Gln Ala Lys Thr Leu Pro Arg Glu Leu Ala Gln
835 840 845

Glu Arg Phe Ala Gln Ile Lys Arg Arg Glu Ser Gly Cys Tyr Phe Lys
850 855 860

Ser Tyr His Val Tyr Asp Tyr Leu Ser Lys Asn Ser Asn Glu Phe Lys
865 870 875 880

Gln Asn His Ile Asp Phe Ala Val Thr Ser Tyr Arg Asn Asn Val Glu
885 890 895

His Leu Asn Val Val His Cys Met Thr Lys Tyr Phe Ser Glu Val Lys
900 905 910

Asp Val Lys Ser Tyr Tyr Gly Val Tyr Cys Tyr Ile Met Gln Arg Met
915 920 925

Leu Cys Asp Glu Leu Ile Ile Lys Asn Gln Asp Lys Pro Asp Val Arg
930 935 940

Gln Thr Phe Glu Glu Tyr Asn Arg Leu Leu Lys Asp His Gly Thr Tyr
945 950 955 960

Ser Lys Asn Leu Met Trp Leu Leu Asn Phe Pro Phe Ala Tyr Asn Leu
965 970 975

Ala Arg Tyr Lys Asn Leu Ser Asn Glu Asp Leu Phe Asn Ala Lys Asn
980 985 990

Asn Asp Gln Lys Ser Lys

995

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

<220>
<223> DR

<400> 211
gtacaatagc ccgttgaaat agatgggttc taagac 36

<210> 212
<211> 1020
<212> PRT
<213> Artificial Sequence

<220>
<223> 250twins_36050_GL0158985

<400> 212

Met Lys Lys Lys His Gln Ser Ala Ala Glu Lys Arg Gln Val Lys Lys
1 5 10 15

Leu Lys Asn Gln Glu Lys Ala Gln Lys Tyr Ala Ser Glu Pro Ser Pro
20 25 30

Leu Gln Ser Asp Thr Ala Gly Val Glu Cys Ser Gln Lys Lys Thr Val
35 40 45

Val Ser His Ile Ala Ser Ser Lys Thr Leu Ala Lys Ala Met Gly Leu
50 55 60

Lys Ser Thr Leu Val Met Gly Asp Lys Leu Val Ile Thr Ser Phe Ala
65 70 75 80

Ala Ser Lys Ala Val Gly Gly Ala Gly Tyr Lys Ser Ala Asn Ile Glu
85 90 95

Lys Ile Thr Asp Leu Gln Gly Arg Val Ile Glu Glu His Glu Arg Met
100 105 110

```
Phe Ser Ala Asp Val Gly Glu Lys Asn Ile Glu Leu Ser Lys Asn Asp
        115                 120                 125

Cys His Thr Asn Val Asn Asn Pro Val Val Thr Asn Ile Gly Lys Asp
    130                 135                 140

Tyr Ile Gly Leu Lys Ser Arg Leu Glu Gln Glu Phe Phe Gly Lys Thr
145                 150                 155                 160

Phe Glu Asn Asp Asn Leu His Val Gln Leu Ala Tyr Asn Ile Leu Asp
                165                 170                 175

Ile Lys Lys Ile Leu Gly Thr Tyr Val Asn Asn Ile Ile Tyr Ile Phe
            180                 185                 190

Tyr Asn Leu Asn Arg Ala Gly Thr Gly Arg Asp Glu Arg Met Tyr Asp
        195                 200                 205

Asp Leu Ile Gly Thr Leu Tyr Ala Tyr Lys Pro Met Glu Ala Gln Gln
    210                 215                 220

Thr Tyr Leu Leu Lys Gly Asp Lys Asp Met Arg Arg Phe Glu Glu Val
225                 230                 235                 240

Lys Gln Leu Leu Gln Asn Thr Ser Ala Tyr Tyr Val Tyr Tyr Gly Thr
                245                 250                 255

Leu Phe Glu Lys Val Lys Ala Lys Ser Lys Lys Glu Gln Arg Ala Lys
            260                 265                 270

Glu Ala Glu Ile Asp Ala Cys Thr Ala His Asn Tyr Asp Val Leu Arg
        275                 280                 285

Leu Leu Ser Leu Met Arg Gln Leu Cys Met His Ser Val Ala Gly Thr
    290                 295                 300

Ala Phe Lys Leu Ala Glu Ser Ala Leu Phe Asn Ile Glu Asp Val Leu
305                 310                 315                 320
```

```
Ser Ala Asp Leu Lys Glu Ile Leu Asp Glu Ala Phe Ser Gly Ala Val
                325                 330                 335

Asn Lys Leu Asn Asp Gly Phe Val Gln His Ser Gly Asn Asn Leu Tyr
            340                 345                 350

Val Leu Gln Gln Leu Tyr Pro Asn Glu Thr Ile Glu Arg Ile Ala Glu
        355                 360                 365

Lys Tyr Tyr Arg Leu Thr Val Arg Lys Glu Asp Leu Asn Met Gly Val
    370                 375                 380

Asn Ile Lys Lys Leu Arg Glu Leu Ile Val Gly Gln Tyr Phe Pro Glu
385                 390                 395                 400

Val Leu Asp Lys Glu Tyr Asp Leu Ser Lys Asn Gly Asp Ser Val Val
                405                 410                 415

Thr Tyr Arg Ser Lys Ile Tyr Thr Val Met Asn Tyr Ile Leu Leu Tyr
            420                 425                 430

Tyr Leu Glu Asp His Asp Ser Ser Arg Glu Ser Met Val Glu Ala Leu
        435                 440                 445

Arg Gln Asn Arg Glu Gly Asp Glu Gly Lys Glu Glu Ile Tyr Arg Gln
    450                 455                 460

Phe Ala Lys Lys Val Trp Asn Gly Val Ser Gly Leu Phe Gly Val Cys
465                 470                 475                 480

Leu Asn Leu Phe Lys Thr Glu Lys Arg Asn Lys Phe Arg Ser Lys Val
                485                 490                 495

Ala Leu Pro Asp Val Ser Gly Ala Ala Tyr Met Leu Ser Ser Glu Asn
            500                 505                 510

Ile Asp Tyr Phe Val Lys Met Leu Phe Phe Val Cys Lys Phe Leu Asp
        515                 520                 525
```

```
Gly Lys Glu Ile Asn Glu Leu Leu Cys Ala Leu Ile Asn Lys Phe Asp
    530                 535                 540

Asn Ile Ala Asp Ile Leu Asp Ala Ala Ala Gln Cys Gly Ser Ser Val
545                 550                 555                 560

Trp Phe Val Asp Ser Tyr Arg Phe Phe Glu Arg Ser Arg Arg Ile Ser
                565                 570                 575

Ala Gln Ile Arg Ile Val Lys Asn Ile Ala Ser Lys Asp Phe Lys Lys
            580                 585                 590

Ser Lys Lys Asp Ser Asp Glu Ser Tyr Pro Glu Gln Leu Tyr Leu Asp
        595                 600                 605

Ala Leu Ala Leu Leu Gly Asp Val Ile Ser Lys Tyr Lys Gln Asn Arg
    610                 615                 620

Asp Gly Ser Val Val Ile Asp Asp Gln Gly Asn Ala Val Leu Thr Glu
625                 630                 635                 640

Gln Tyr Lys Arg Phe Arg Tyr Glu Phe Phe Glu Glu Ile Lys Arg Asp
                645                 650                 655

Glu Ser Gly Gly Ile Lys Tyr Lys Lys Ser Gly Lys Pro Glu Tyr Asn
            660                 665                 670

His Gln Arg Arg Asn Phe Ile Leu Asn Asn Val Leu Lys Ser Lys Trp
        675                 680                 685

Phe Phe Tyr Val Val Lys Tyr Asn Arg Pro Ser Ser Cys Arg Glu Leu
    690                 695                 700

Met Lys Asn Lys Glu Ile Leu Arg Phe Val Leu Arg Asp Ile Pro Asp
705                 710                 715                 720

Ser Gln Val Arg Arg Tyr Phe Lys Ala Val Gln Gly Glu Glu Ala Tyr
                725                 730                 735
```

```
Ala Ser Ala Glu Ala Met Arg Thr Arg Leu Val Asp Ala Leu Ser Gln
            740                 745                 750

Phe Ser Val Thr Ala Cys Leu Asp Glu Val Gly Gly Met Thr Asp Lys
        755                 760                 765

Glu Phe Ala Ser Gln Arg Ala Val Asp Ser Lys Glu Lys Leu Arg Ala
    770                 775                 780

Ile Ile Arg Leu Tyr Leu Thr Val Ala Tyr Leu Ile Thr Lys Ser Met
785                 790                 795                 800

Val Lys Val Asn Thr Arg Phe Ser Ile Ala Phe Ser Val Leu Glu Arg
                805                 810                 815

Asp Tyr Tyr Leu Leu Ile Asp Gly Lys Lys Lys Ser Ser Asp Tyr Thr
            820                 825                 830

Gly Glu Asp Met Leu Ala Leu Thr Arg Lys Phe Val Gly Glu Asp Ala
        835                 840                 845

Gly Leu Tyr Arg Glu Trp Lys Glu Lys Asn Ala Glu Ala Lys Asp Lys
    850                 855                 860

Tyr Phe Asp Lys Ala Glu Arg Lys Lys Val Leu Arg Gln Asn Asp Lys
865                 870                 875                 880

Met Ile Arg Lys Met His Phe Thr Pro His Ser Leu Asn Tyr Val Gln
                885                 890                 895

Lys Asn Leu Glu Ser Val Gln Ser Asn Gly Leu Ala Ala Val Ile Lys
            900                 905                 910

Glu Tyr Arg Asn Ala Val Ala His Leu Asn Ile Ile Asn Arg Leu Asp
        915                 920                 925

Glu Tyr Ile Gly Ser Ala Arg Ala Asp Ser Tyr Tyr Ser Leu Tyr Cys
    930                 935                 940
```

Tyr Cys Leu Gln Met Tyr Leu Ser Lys Asn Phe Ser Val Gly Tyr Leu
945 950 955 960

Ile Asn Val Gln Lys Gln Leu Glu Glu His His Thr Tyr Met Lys Asp
965 970 975

Leu Met Trp Leu Leu Asn Ile Pro Phe Ala Tyr Asn Leu Ala Arg Tyr
980 985 990

Lys Asn Leu Ser Asn Glu Lys Leu Phe Tyr Asp Glu Glu Ala Ala Ala
995 1000 1005

Glu Lys Ala Asp Lys Ala Glu Asn Glu Arg Gly Glu
1010 1015 1020

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

<220>
<223> DR

<400> 213
gtacaatagc cctgcagtaa ggcagggttc taagac 36

<210> 214
<211> 903
<212> PRT
<213> Artificial Sequence

<220>
<223> 31009_GL0034153

<400> 214

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

Ala Lys Ala Leu Gly Val Lys Ser Ala Phe Leu Lys Asp Asp Asp Val
20 25 30

Ile Met Thr Ser Phe Gly Arg Gly Asn Asp Ala Val Val Glu Lys Ile

```
        35                  40                  45

Ile Lys Asn Asn Glu Val Gln Asn Ile Asn Ile Asp Lys Pro Val Tyr
    50                  55                  60

Asp Val Val Gly Val Glu Lys Glu Thr Gly Ile Ile Asn Val Val Gly
65                  70                  75                  80

Thr Lys Arg Gly Ile Asn Thr Ile Asn Lys Lys Lys Gln Tyr Ala Asn
                85                  90                  95

Thr Gln Val Ser Gln Lys Arg Met Gly Met Asp Leu Ile Gln Arg Lys
            100                 105                 110

Asp Lys Leu Glu Lys Val Phe Phe Asn Glu Asn Phe Asn Asp Asn Ile
        115                 120                 125

His Ile Gln Leu Ile Tyr Asn Ile Leu Asp Ile Glu Lys Ile Leu Ser
    130                 135                 140

Val Tyr Val Asn Asn Ile Val Tyr Ala Leu Gly Asn Leu Glu Arg Asp
145                 150                 155                 160

Asn Ser Asp Asp Glu Lys Asp Leu Ile Gly Tyr Ala Ser Thr Leu Asn
                165                 170                 175

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

Ile Asp Lys Phe Ile Glu Asn Gly Asn Arg Leu Gly Tyr Phe Gly Asn
        195                 200                 205

Val Phe Tyr Lys Asn Lys Asn Glu Leu Lys Asp Lys Lys Asp Ile Phe
    210                 215                 220

Asp Met Leu Cys Leu Leu Gly Ser Leu Arg Gln Phe Cys Phe His Phe
225                 230                 235                 240

Lys Gly Ala Lys Phe Glu Gly Val Glu Gln Ile Ser Asn Ser Trp Leu
```

245 250 255

Tyr Asn Leu Ser Gln Leu His Asn Glu Phe Lys Ser Thr Leu Asp Ser
260 265 270

Phe Tyr Asn Glu Lys Ile Glu Asp Ile Asn Gly Asn Phe Ile Glu Lys
275 280 285

Asn Lys Val Asn Ile Gln Ile Leu Cys Asn Asp Leu Lys Leu Asp Lys
290 295 300

Asn Lys Leu Val Gln Asp Tyr Tyr Asn Phe Leu Ile Ser Lys Asn His
305 310 315 320

Lys Asn Met Gly Phe Ser Ile Lys Lys Ile Arg Glu Tyr Ile Leu Asp
325 330 335

Ile Asp Tyr Ala Ser Tyr Ile Lys Glu Gln Gly Tyr Asp Ser Val Arg
340 345 350

Ser Lys Leu Phe Thr Leu Ile Asp Phe Ile Ile Tyr Asp Leu Tyr Val
355 360 365

Asn Lys Glu Lys Asp Arg Ile Asp Asp Ile Val Ser Lys Leu Arg Gly
370 375 380

Ala Ile Tyr Asp Tyr Lys Lys Asp Asp Ile Tyr Lys Ala Glu Ala Asp
385 390 395 400

Trp Val Trp Lys Lys Leu Asn Arg Gln Ile Lys Ala Leu Thr Lys Leu
405 410 415

Leu Ser Asn Arg Arg Ile Phe Ser Gln Asp Asn Asp Ile Thr Asn Trp
420 425 430

Asn Gly Val Glu Phe Glu Lys Ile Gly Glu Gln Lys Leu Glu Asp Asn
435 440 445

Ala Asp Tyr Phe Cys Lys Leu Met Tyr Phe Val Thr Leu Phe Leu Asp

```
    450                 455                 460

Gly Lys Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Asn Lys Phe Asp
465                 470                 475                 480

Asn Ile Arg Ser Phe Ile Glu Val Met Glu Glu Met Ser Leu Lys Cys
                485                 490                 495

Glu Phe Met Glu Glu Tyr Lys Phe Phe Asn Glu Ser Lys Asp Ile Cys
            500                 505                 510

Leu Arg Leu Lys Glu Ile His Ser Phe Ala Arg Met Lys Arg Pro Leu
        515                 520                 525

Asp Asn Lys Asn Val Glu Lys Glu Met Tyr Lys Asp Ala Ile Ser Ile
    530                 535                 540

Leu Met Lys Asp Ser Ser Ile Glu Asn Ser Asn Ile Asp Val Ile Phe
545                 550                 555                 560

Asn Thr Tyr Ile Tyr Gln Lys Asp Asn Lys Asn Ala Lys Lys Asp Phe
                565                 570                 575

Arg Asn Phe Ile Ile Lys Asn Val Ile Lys Ser Asn Arg Phe Ile Tyr
            580                 585                 590

Leu Ile Lys Tyr Ser Asn Pro Thr Asn Val Arg Lys Phe Ala Ser Asn
        595                 600                 605

Thr Asn Val Ile Lys Phe Val Leu Gly Thr Leu Pro Asp Lys Gln Leu
    610                 615                 620

Glu Arg Tyr Tyr Lys Ser Cys Gly Leu Ser Asp Gly Ile Gly Lys Thr
625                 630                 635                 640

Glu Trp Val Arg Glu Leu Ser Glu Ile Ile Thr Gly Ile Asp Tyr Ser
                645                 650                 655

Asp Phe Leu Asn Val Lys Gln Arg Ala Lys Ser Lys Glu Glu Lys Glu
```

660 665 670

Asp Lys Ala Lys Lys Gln Ala Ile Val Thr Leu Tyr Leu Thr Ile Leu
675 680 685

Tyr Ile Leu Thr Lys Asn Leu Val Asn Val Asn Ser Arg Tyr Val Ile
690 695 700

Ala Val His Cys Ile Glu Arg Asp Ser Glu Leu Tyr Lys Thr Glu Leu
705 710 715 720

Tyr Glu Lys Leu Lys Pro Glu Gln Pro Glu Lys Leu Tyr Lys Ile Lys
725 730 735

Pro Asp Lys Pro Leu Gln Pro Ala Gln Tyr Phe Leu Phe Thr Lys Tyr
740 745 750

Leu Leu Gln Asn Asp Phe Phe Gly Glu Tyr Thr Ser Lys Lys Asn Gln
755 760 765

Tyr Ile Gln Lys Asn Met Lys Ile Tyr Ser Lys Ile Asn Lys Asn Phe
770 775 780

Asn Val Phe Val Arg Tyr Arg Asn Asn Ile Ala His Leu Asn Ser Ile
785 790 795 800

Lys Asn Ala Thr Lys Tyr Val Asn Asp Ile Gly Thr Phe Thr Ser Tyr
805 810 815

Phe Gln Leu Tyr His Tyr Ile Met Gln Cys Asp Leu Lys Asp Trp Val
820 825 830

Lys Lys Glu Val Glu Glu Lys Lys Ile Ser Asp Lys Tyr Tyr Asn Tyr
835 840 845

Leu Glu Lys Tyr Tyr Lys Asn Asn Asn Ile Tyr Leu Lys Asp Phe Val
850 855 860

Lys Ala Ile Asn Val Pro Phe Ala Tyr Asn Tyr Ser Arg Phe Lys Asn

865 870 875 880

Leu Ser Ile Asp Lys Leu Phe Asp Lys Asn Asp Pro Arg Lys Lys Lys
885 890 895

Trp Ile Val Ile Phe Asn Phe
900

<210> 215
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 215
gaactacacc cttgttaaat aacatgggtc taaaac 36

<210> 216
<211> 951
<212> PRT
<213> Artificial Sequence

<220>
<223> 530373_GL0023589

<400> 216

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

Glu Asn Cys Ser Ile Ala Lys Ala Asn Gly Leu Lys Ser Thr Phe Lys
20 25 30

Leu Ala Glu Asn Thr Tyr Tyr Met Thr Ser Phe Gly Lys Asn Asn Thr
35 40 45

Ala Asp Pro Glu Lys Asn Ile Val Arg Met Asn Val Glu Asn Ile Arg
50 55 60

Asn Thr Phe Ser Ala Thr Pro Ile Ser Glu Arg Ile Val Asp Ile Thr
65 70 75 80

```
Gly Glu Thr Lys Lys Thr Ala Gln Val Val Leu Pro Ala Cys Ser Asp
                85                  90                  95

Asn Gln Leu His Ala Lys Asp Thr Ile Glu Gln Met Tyr Phe Gly Lys
            100                 105                 110

Thr Tyr Thr Asp Asn Ile His Ile Gln Ile Ala Tyr Asn Ile Met Asp
        115                 120                 125

Met Lys Lys Ile Phe Gly Val Tyr Ser Asn Asn Ile Ile Tyr Ser Ile
    130                 135                 140

Asn Asn Leu Leu Glu Glu Arg Glu Gly Ser Asp Phe Leu Gly Val Phe
145                 150                 155                 160

Tyr Thr Glu Asn Ser Tyr Gln Ala Met Asn Phe Ala Asp Ala Ile Ile
                165                 170                 175

Lys Ser Lys Leu Leu Thr Lys Asn Ser Lys Tyr Leu Asn Leu Asp Met
            180                 185                 190

Asn Asn Val Trp Phe Asn Lys Glu Val Ser Glu Leu Ser Ser Phe Arg
        195                 200                 205

Lys Tyr Val Tyr Ser Arg Ser Tyr Ala Asn Lys Thr Ser Thr Ala Lys
    210                 215                 220

Lys Ile Cys Ile Cys Leu Gln Glu Asp Phe His Phe Phe Asp Tyr Ala
225                 230                 235                 240

Thr Val Lys Lys Ala Ala Asp Gln Phe Thr Leu Leu Lys Arg Ile Gln
                245                 250                 255

Arg Ser Asn Tyr Phe Phe Asp Met Leu Cys Asp Ser Ala Gly Asn Tyr
            260                 265                 270

Gln Cys Lys Asn Ala Gln Ile Leu Leu Ser Ile Leu Gly Asn Leu Arg
        275                 280                 285
```

```
Gln Glu Ser Phe His Glu Asn Arg Ser Thr Ser Ala Trp Met Phe Asn
    290                 295                 300

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

Ile Asp Ser Arg Thr Asn His Ile Asn Asn His Phe Ile Glu Asn Ser
                325                 330                 335

Asn Lys Asn Leu Asn Ala Ile Phe Arg Leu Phe Pro Glu Lys Ser Gln
            340                 345                 350

Ala Asp Leu Ile Val Glu Tyr Tyr Asp Phe Ser Val Arg Lys Ala Tyr
        355                 360                 365

Lys Asn Leu Gly Phe Ser Val Lys Leu Leu Arg Glu Met Leu Leu Lys
    370                 375                 380

Thr Gln Gln Glu Leu Ser Pro Leu Asn Asp Lys Ala Tyr Asp Ser Ile
385                 390                 395                 400

Arg Ser Lys Leu Tyr Thr Met Phe Asp Phe Val Leu Tyr Lys Leu Tyr
                405                 410                 415

Val Asp Asn Glu Glu Gln Ile Asp Gln Phe Val Arg Lys Leu Arg Ala
            420                 425                 430

Ser Thr Ser Asn Val Glu Lys Glu Ala Leu Tyr Gln Gln Glu Ser Glu
        435                 440                 445

Arg Thr Trp Asn Glu Ile Arg Gly Gln Ile Ile Asp Asn Val Ile Pro
    450                 455                 460

Phe Ile Asp Ser Ile Gln Lys Gln Glu Asn Glu Asp Lys Leu Thr Lys
465                 470                 475                 480

Arg Ile Asp Gln Ser Ile Ile Asp Lys Ala Val Pro His Phe Val Asn
                485                 490                 495
```

```
Ser Arg Asp Leu Ser Tyr Phe Ala Lys Val Val Tyr Val Val Ala Ser
            500                 505                 510

Phe Leu Asp Ser Lys Glu Ile Asn Met Phe Phe Asp Ala Leu Ile Asn
        515                 520                 525

Glu Leu Glu Asn Ile Ala Ser Phe Asn Asp Val Leu Glu Gln Leu Asp
    530                 535                 540

Met Pro Leu Asp Tyr Gln Glu Lys Tyr Arg Phe Phe Glu Arg Ser Arg
545                 550                 555                 560

Glu Glu Ala Ala Gly Leu Arg Phe Ile Lys Ser Ile Ala Lys Met Asn
                565                 570                 575

Lys Thr Arg Val Ala Val Lys Asn Asp Thr Lys Thr Val Lys Leu Lys
            580                 585                 590

Gln Tyr Cys Asp Ala Gly Ala Val Leu Gly Glu Ile Asn Ala Glu Lys
        595                 600                 605

Ile Lys Ala Ala Phe Gly Leu Gly Asp Lys Phe Asn Thr Glu Glu Ala
    610                 615                 620

Arg Val Asp His Thr Phe Arg Asn Phe Ile Ile Asn Asn Val Ile Asn
625                 630                 635                 640

Ser Asn Arg Phe Asn Tyr Val Met Arg Phe Met Asn Pro Lys Ser Ala
                645                 650                 655

Arg Arg Ile Met Gln Cys Glu Ser Leu Val Ala Phe Ser Leu Lys Asp
            660                 665                 670

Ile Pro Asp Thr Gln Ile Glu Arg Tyr Cys Lys Ser Val Asn Ile Pro
        675                 680                 685

Phe Asp Ser Ala Asn Pro Asn Tyr Ala His Met Lys Glu Gln Leu Thr
    690                 695                 700
```

```
Tyr Lys Leu Leu Gln Val Gln Phe Cys Asn Phe Ser Lys Val Ser Asn
705                 710                 715                 720

Ser Ser His Gln Ala Val Glu Lys Glu Arg Leu Lys Ala Leu Val Gly
                725                 730                 735

Leu Tyr Leu Ser Val Leu Tyr Arg Ile Val Lys Ser Ile Val Arg Ile
            740                 745                 750

Asn Thr Ser Tyr Thr Ile Ala Phe Ala Ala Tyr Glu Arg Asp Cys Val
        755                 760                 765

Ile Leu Asn Arg Lys Met Gly Lys Asn Leu Tyr Gly Leu Lys Thr Asp
    770                 775                 780

Pro Met Gly Ile Thr Asn Asp Phe Met His Met Pro Thr Val Ala Pro
785                 790                 795                 800

Ser Asn Pro His Asn Asp Phe Ile Asn Glu Thr Lys Leu Ser Lys Arg
                805                 810                 815

Val Cys Ala Leu Val Asn Lys Asn Arg Thr Leu Cys Thr Asn Arg Thr
            820                 825                 830

Tyr Gln Val Tyr Arg Asn Ile Ile Met His Leu Asn Ala Val Ser Ala
        835                 840                 845

Phe Pro Gln Tyr Leu Glu Gly Leu Lys Val Val His Ser Tyr Phe Asp
    850                 855                 860

Leu Tyr His Tyr Thr Val Met Met Leu Leu His Arg Gln Asn Glu Lys
865                 870                 875                 880

Glu Lys Lys Arg Asn Glu Lys Glu Lys Lys Gln Asn Glu Lys Glu Asn
                885                 890                 895

Asn Met Ile Leu Ser Thr Ala Ala Leu Glu Lys Tyr Asp Met Val Arg
            900                 905                 910
```

Lys Tyr Gln Ser Ala Asn Lys Asp Phe Ile Tyr Ala Ile Asn Ser Pro
915 920 925

Leu Ala Tyr Asn Ala Ala Arg Tyr Ile Asn Leu Ser Cys Arg Thr Lys
930 935 940

Phe Glu Glu Gly Phe Gly Lys
945 950

<210> 217
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 217
acttaaatcc ccctgtaaat gcgggggttc taaaac 36

<210> 218
<211> 928
<212> PRT
<213> Artificial Sequence

<220>
<223> BMZ-11B_GL0037771

<400> 218

Met Ser Lys Asp Lys Lys Thr Lys Ala Lys Arg Met Gly Val Lys Ala
1 5 10 15

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

Gly Asn Arg Ser Lys Ile Glu Phe Thr Glu Gly Tyr His Gly Arg Ala
35 40 45

Leu Glu Thr Pro Lys His Phe Gly Lys Arg Gly Phe Glu Val Arg Lys
50 55 60

Ile Asp Glu Asn Val Asp Leu Tyr Gly Asp Leu Glu Glu Gly Lys Thr

65 70 75 80

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

Leu Lys Leu Lys Gly Thr Leu Glu Lys Arg Phe Phe Gly Arg Glu Phe
100 105 110

Pro His Asp Asn Ile Arg Ile Gln Leu Ile Tyr Asn Ile Leu Asp Ile
115 120 125

Tyr Lys Ile Leu Gly Met Asn Val Ala Asp Ile Leu Tyr Ala Leu Gly
130 135 140

Asn Met Gln Asp Thr Glu Leu Asp Ile Asp Met Phe Gly Gln Ser Leu
145 150 155 160

Asn Asn Pro Glu Lys Ala Lys Asp Val Val Gln Arg Met Lys Pro Tyr
165 170 175

Met Gly Phe Phe Gly Gly Ile Phe Arg Thr Gln Lys Lys Glu Asp Arg
180 185 190

Lys Lys Asp Ser Asn Leu Ser Glu Glu Glu Lys Glu Glu Lys Lys Lys
195 200 205

Lys Gln Glu Glu Leu Leu Gln Lys Asp Leu Gln His Asn Met Asp Val
210 215 220

Ala Arg Cys Ile Ser Ala Leu Arg His Ala Thr Ala His Asn Lys Pro
225 230 235 240

Ala Thr Ala His Asp Lys Gln Asp Glu Tyr Pro Trp Phe Lys Ser Ser
245 250 255

Asp Ile Tyr Glu Thr Lys Ile Phe Lys Ala Gly Met Trp Ser Val Ile
260 265 270

Glu Ala Asp Tyr Lys Lys Lys Ile Gln Asp Val Asn Lys Gln Phe Phe

```
        275                 280                 285

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

Arg Asp Val Lys Gln Lys Lys Arg Ile Ser Asp Glu Phe Tyr Arg Phe
305                 310                 315                 320

Thr Ile Arg Lys Asp Gly Lys Asn Leu Gly Met Asn Leu Val Lys Ile
                325                 330                 335

Arg Glu Ile Ile Ile Asp Arg Tyr Ala Ser Gly Leu Arg Asp Lys Lys
            340                 345                 350

His Asp Pro His Arg Gln Lys Ile Asn Val Ile Ala Asp Phe Leu Ile
        355                 360                 365

Phe Arg Ala Leu Ser Gln Asn Gln Gly Ile Ile Asp Lys Thr Val Ser
    370                 375                 380

Ser Leu Arg Leu Thr Lys Asp Glu Glu Glu Lys Asp His Val Tyr Gln
385                 390                 395                 400

Asn Ala Ala Glu Leu Val Trp Gly Met Val Ser Asn Cys Leu Thr Pro
                405                 410                 415

Tyr Phe Asn Asp Pro Lys Asn Lys Tyr Ile Leu Lys Tyr Lys Asp Ala
            420                 425                 430

Lys Thr Pro Gly Asp Phe Glu Asp Trp Ile Thr Ser Lys Ile Ser Glu
        435                 440                 445

Asp Asp Gly Glu Pro Phe Val Lys Val Leu Ser Phe Leu Cys Asn Phe
    450                 455                 460

Leu Glu Gly Lys Glu Ile Asn Glu Leu Leu Thr Ala Tyr Ile His Lys
465                 470                 475                 480

Phe Glu Cys Ile Gln Asp Phe Leu Asn Val Ile Ser Ser Leu Gly Glu
```

485 490 495

Asn Val Gln Phe Gln Pro Arg Phe Ala Leu Phe Asn Asn Ala Ser Phe
500 505 510

Ala Gln Asn Val Ala Val Gln Leu Arg Ile Leu Ala Ser Ile Gly Lys
515 520 525

Met Lys Pro Asp Leu Thr Glu Ala Lys Arg Pro Leu Tyr Lys Ala Ala
530 535 540

Ile Arg Met Leu Cys Pro Pro Glu Lys Trp Glu Lys Tyr Thr Ser Asp
545 550 555 560

Glu Trp Leu Glu Lys Asn Met Leu Leu Asn Ser Glu Asp Arg Lys Asn
565 570 575

Asp Lys Lys Lys Lys Gln Val Asn Pro Phe Arg Asn Phe Ile Ala Gly
580 585 590

Asn Val Ile Glu Ser Arg Arg Phe Met Tyr Leu Val Arg Tyr Ser Lys
595 600 605

Pro Lys Ala Val Arg Ala Ile Met Gln Asn Arg Ser Ile Val Asn Tyr
610 615 620

Val Leu His Arg Leu Pro Ser Glu Gln Ile Lys Thr Tyr Ser Arg Val
625 630 635 640

Phe Pro Glu Asp Phe Ser Asp Thr Glu Ala Glu Ile Asp Phe Leu Val
645 650 655

Asn Lys Leu Ser Glu Phe Ser Phe Glu Thr Phe Ile Ser Asn Gln Gln
660 665 670

Thr Ile Leu Asn Asn Ser Lys Arg Gly Phe Ser Pro Asn Arg Arg Glu
675 680 685

Thr Ala Glu Glu Ile Glu Arg Leu Lys Ala Ile Thr Gly Leu Tyr Leu

690 695 700

Ser Val Ala Tyr Ile Ala Ile Lys Asn Ile Val Lys Ala Asn Ala Arg
705 710 715 720

Tyr Tyr Ile Ala Phe Ala Val Phe Glu Arg Asp Lys Glu Leu Val Lys
725 730 735

Ala Lys Asp Ala Arg Ile Gln Thr Thr Ile Pro Asn Thr Lys Phe Thr
740 745 750

Asn Tyr Phe Cys Leu Thr Gln Tyr Tyr Leu Asp Arg Asp Glu Glu Lys
755 760 765

Lys Phe Pro Gly Asp Pro Arg Asp Lys Glu Ala Phe Phe Glu His Leu
770 775 780

Arg Lys Thr Lys Arg His Phe Ser Lys Gln Trp Arg Glu Trp Leu Asn
785 790 795 800

Glu Lys Ile Ala Asp Ala Lys Ser Ser Gln Ala Thr Gly Leu Leu Leu
805 810 815

Arg Glu Ala Arg Asn Asp Val Glu His Leu Asn Val Leu Arg Ala Ile
820 825 830

Pro Asp Tyr Ile Gln Asp Phe Arg His Gly Glu Lys Gly Glu Thr Ala
835 840 845

Met Asn Ser Tyr Phe Glu Leu Tyr His Tyr Leu Met Gln Arg Leu Met
850 855 860

Leu Lys Asn Thr Glu Leu Asp Leu Ser His Trp Ser Gly Trp Ile Met
865 870 875 880

Arg Ser Gly Arg Pro Asp Arg Asp Leu Ile Gln Ile Ala Phe Val Ser
885 890 895

Leu Ala Tyr Asn Leu Pro Arg Tyr Arg Asn Leu Thr Lys Glu His His

900 905 910

Phe Asp Asp Thr Val Leu Gln Lys Ile Arg Glu Lys Glu Ser Leu Asp
915 920 925

<210> 219
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 219
ggtcaacagt gtggggaatt acccgcactc tacaac 36

<210> 220
<211> 940
<212> PRT
<213> Artificial Sequence

<220>
<223> BMZ-11B_GL0037915

<400> 220

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

Ser Val Ile Ala His Asp Gly Gly Lys Leu Ala Val Thr Thr Phe Gly
20 25 30

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

Phe Asp Glu Glu Lys Lys Ser Glu Leu Gly Gly Phe Asn Val Thr Gly
50 55 60

Val Asp Glu Asn Thr Val Asp Ala Glu Arg Gly Ile Pro Gly Glu Gly
65 70 75 80

Leu Ile Glu Ala Ser Val Ala Asn Pro Val Asn Asp Leu Ile Asp Glu
85 90 95

```
Ser Arg Lys Glu Tyr Gly Gln Asp Tyr Leu Met Leu Lys Asp Gln Leu
            100                 105                 110

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

Gln Ile Ile Tyr Asn Ile Leu Asp Met Gln Lys Ala Phe Gly Leu Tyr
    130                 135                 140

Val Asn Asp Val Leu Tyr Ala Val Asn Asn Leu Gln Asn Thr Gly Arg
145                 150                 155                 160

Leu Leu Thr Asp Asp Asp Trp Thr Ser Asp Val Ile Gly Leu Ala Leu
                165                 170                 175

Gly Lys Gly Arg Gly Lys Glu Ile Ile Glu Ser Met Ser Asp Ala Leu
            180                 185                 190

Gly Phe Phe Gly Ser Ala Phe Lys Pro Ile Pro Lys Lys Pro Gly Asp
        195                 200                 205

Lys Thr Ser Gly Lys Tyr Ala Gly Gly Lys Leu Lys Glu Arg Asp Ala
    210                 215                 220

Ala Ile Ala Glu Asp Val Ala Val Leu Arg Val Leu Cys Gly Leu Arg
225                 230                 235                 240

Gln Ile Ser Ala His Phe Arg Ser Glu Lys Lys Gly Glu Ser Leu Trp
                245                 250                 255

Ala Ser Asp Leu Gly Val Glu Lys Tyr Phe Lys Asn Glu Met Ser Val
            260                 265                 270

Val Asp Ala Asn Tyr Lys Ala Arg Ala Lys Lys Ile Asp Ala Asn Phe
        275                 280                 285

Met Ala Thr Ser Arg Ala Asn Leu Val Leu Leu Phe Arg Ile Leu Asn
    290                 295                 300
```

```
Ala Lys Thr Lys Glu Glu Lys Glu Lys Ile Ala Val Asp Tyr Tyr Lys
305                 310                 315                 320

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

Ala Arg Glu Ile Leu Leu Asp Lys His Tyr Pro Asp Ala Lys Asn Lys
            340                 345                 350

Gln Phe Asp Ser Tyr Arg Gly Lys Leu Tyr Thr Ile Thr Asp Phe Ile
        355                 360                 365

Ile Tyr Arg Met Ala Ala Glu Asp Gly Ala Phe Leu Glu Glu Ser Val
    370                 375                 380

Gln Lys Leu Arg Ser Thr Pro Ser Glu Glu Glu Lys Glu Lys Val Tyr
385                 390                 395                 400

Asp Val Leu Ala Glu Thr Ile Trp Asn Ala Thr Lys Asp Asp Val Lys
                405                 410                 415

Lys Val Phe Asp Ile Arg Gln Gln Asp Ile Lys Asn Ala Ala Ala Gln
            420                 425                 430

Ser Leu Asp Asp Val Asn Leu Asp Ser Val Thr Val Glu Lys Glu Lys
        435                 440                 445

Glu Phe Pro Val Val Lys Leu Leu Ala Phe Leu Cys Asn Phe Leu Asp
    450                 455                 460

Gly Lys Glu Ile Asn Glu Leu Leu Ser Thr Phe Val Asn Arg Leu Glu
465                 470                 475                 480

Asn Val Gln Ala Leu Ile Asp Leu Ala Lys Asn Leu Gly Glu Thr Val
                485                 490                 495

Asn Phe Ser Lys Asn Tyr Ala Val Phe Asn Asp Cys Gly Ser Gly Ala
            500                 505                 510
```

```
Ala Gly Arg Ile Ala Glu Arg Leu Arg Leu Leu Ala Ser Ile Gly Lys
        515                 520                 525

Met Asp Ala Gly Leu Asp Ser Leu Lys Pro Leu Phe Ile Arg Phe Ala
    530                 535                 540

Ile Gln Ile Leu Gly Gly Ala Arg Ile Ser Asp Glu Asp Val Asp Ala
545                 550                 555                 560

Tyr Ile Lys Glu Asn Ile Leu Val Arg Lys Asp Asp Pro Glu Lys Lys
                565                 570                 575

Arg Ser Asn Thr Phe Arg Asn Phe Ile Thr Asn Asn Val Ile Lys Ser
            580                 585                 590

Arg Arg Phe Val Tyr Leu Val Arg Tyr Thr Lys Ala Asp Ala Val Arg
        595                 600                 605

Ala Val Met Lys Asn Gly Lys Ile Val Arg Tyr Val Leu Ala Arg Ile
    610                 615                 620

Pro Glu Thr Gln Val Asp Ser Tyr Tyr Lys Ser Val Val Asp Ser Ser
625                 630                 635                 640

Asp Ala Gly Asn Gly Lys Pro Ile Lys Glu Lys Ile Lys Ala Leu Ala
                645                 650                 655

Glu Lys Leu Gly Glu Phe Ser Phe Asp Leu Ile Ala Glu Asn Arg Asp
            660                 665                 670

Gly Ile Ile Asp Asn Ala Lys Leu Asp Glu Glu Lys Lys Asn Val Glu
        675                 680                 685

Ile Glu Arg Leu Lys Ala Val Thr Gly Leu Tyr Leu Thr Val Ala Phe
    690                 695                 700

Ile Ala Ile Lys Asn Leu Val Lys Ala Asn Ala Arg Tyr Tyr Ile Ala
705                 710                 715                 720
```

```
Tyr Ser Ala Phe Glu Arg Asp Ser Glu Leu Leu Lys Glu Lys Leu Gly
                725                 730                 735

Glu Glu Glu Phe Tyr Lys Leu Cys Ile Glu Phe Lys Asn Lys Lys Gly
            740                 745                 750

Glu Thr Gly Tyr Asn Asn Ala Phe Ala Leu Leu Lys Tyr Tyr Ile Asp
        755                 760                 765

Arg Asp Glu Asp Ile Trp Arg Lys Glu Glu Ala Thr His Arg Gly Arg
    770                 775                 780

Asp Glu Asn Gly Lys Leu Asp Tyr Asp Pro Asp Pro Arg Arg Lys His
785                 790                 795                 800

Tyr Pro Lys Arg Arg Trp Leu Asn Ile Leu Lys Asp Asn Leu Ala Glu
                805                 810                 815

Met Glu Arg Ile Asp Pro Thr Gly Met Leu Ala Ala Gln Met Arg Asn
            820                 825                 830

Asp Ala Met His Leu Asn Ala Met Phe Lys Ile Ala Lys Phe Val Gly
        835                 840                 845

Asp Phe Arg Lys Gly Ser Ser His Glu Met Arg Ser Tyr Phe Glu Leu
    850                 855                 860

Phe His Phe Leu Glu Gln Lys Thr Leu Leu Asp Ser Lys Glu Asp Phe
865                 870                 875                 880

Val Gln Gly Met Arg Glu Ser Tyr Gly Asp Arg Ile Arg Val Glu Pro
                885                 890                 895

Cys Val Asp Leu Ile Lys Tyr Leu Tyr Ile Thr Leu Gly Tyr Ser Leu
            900                 905                 910

Pro Arg Tyr Lys Thr Met Thr Val Ser Gly Leu Phe Asp Pro Asp Ser
        915                 920                 925
```

Glu Asp Ala Lys Glu Arg Ala Lys Gly Lys Arg Ala
930 935 940

<210> 221
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 221
ggtcaacagt gtggggaatt acccgcactc tacaac 36

<210> 222
<211> 923
<212> PRT
<213> Artificial Sequence

<220>
<223> BMZ-11B_GL0069617

<400> 222

Met Pro His Asn Lys Lys Ser Thr Ile Lys Arg Ile Gly Val Lys Ser
1 5 10 15

Ala Phe Ser His Gly Asn Ser Arg Leu Thr Ile Thr Gln Phe Gly Arg
20 25 30

Gly Asn Thr Ala Asp Ile Ala Val Glu Thr Asp Ser Arg Gly Arg Ser
35 40 45

Leu Ser Leu Pro His Arg Val Glu Lys Ser Phe Thr Val Asn Lys Ile
50 55 60

Asp Asp Glu Ile Asp Val Gln Arg Ser Ser Leu Leu Gly Thr Leu Leu
65 70 75 80

Asn Pro Ala Glu Asn Asp Pro Val Asp Tyr Leu Gly Leu Lys Ser Thr
85 90 95

Leu Glu Lys Glu Phe Phe Gly Gln Glu Phe Pro Asn Asp Ser Ile Arg

100 105 110

Ile Gln Ile Ile His Asn Ile Leu Asp Ile Gln Lys Ile Leu Gly Ile
115 120 125

Tyr Ile Asn Asp Val Leu Tyr Ser Ile Ala Asn Leu Gln Asn Val Glu
130 135 140

Thr Val Glu Asp Cys Leu Gly Gln Ser Leu Ser Lys Gly Gln Asp Glu
145 150 155 160

Lys Leu Lys Lys Ile Leu Asn Ser Ile Thr Pro Tyr Leu Gly Phe Leu
165 170 175

Gly Ala Gly Phe Lys Gln Ser Glu Arg Pro Lys His Gln Lys Asn Ser
180 185 190

Arg Pro Asp Pro Arg Gln Lys Ile Glu Asp Pro Asn Trp Ala Tyr Asn
195 200 205

Lys Ser Ala Leu Arg Ile Leu Ser Thr Leu Arg Asn Lys Leu Ala His
210 215 220

Phe Glu Asp Ser Ser Ile Leu Phe Cys Asp Asn Leu Lys Gly Gln Phe
225 230 235 240

Leu Gly Ser Asn Gly Asn Trp Ile Thr Ile Glu Thr His Tyr Asp Glu
245 250 255

Met Ile Thr Arg Ile Asn Glu Ser Phe Phe Asn Asn Ser Arg Met Asn
260 265 270

Leu Lys Ile Leu Tyr Leu Val Leu Asn Ala Ala Thr Leu Glu Asp Lys
275 280 285

Lys Lys Ile Thr Glu Glu Tyr Tyr Arg Phe Ser Ile Leu Lys Glu Gly
290 295 300

Lys Asn Leu Gly Val Asn Met Arg Lys Leu Arg Glu Met Met Val Asp

305 310 315 320

Met Phe Tyr Ser Glu Ile Lys Asp Lys Gln Tyr Asp Ser Tyr Arg Gln
325 330 335

Lys Ile Tyr Thr Ile Leu Asp Phe Leu Leu Phe Arg Glu Ile Lys Asp
340 345 350

Ser Glu Glu Leu Glu Thr Met Val Ala Asp Leu Arg Glu Ser Glu Ser
355 360 365

Asp Glu Lys Lys Glu Glu Leu Tyr His Asp Tyr Ala Glu Leu Phe Trp
370 375 380

Asp Gly Val Lys Asp Leu Ile Val Pro Phe Phe Ala Ala Phe Asp Gly
385 390 395 400

Asp Phe Thr Val Phe Thr Thr Glu Lys Val Pro Ala Asp Thr Ser Lys
405 410 415

Glu Pro Phe Leu Ser Ser Ala Asn Thr Glu Pro Leu Val Lys Leu Leu
420 425 430

Ala Phe Leu Cys Asn Phe Trp Asp Gly Lys Glu Val Asn Glu Val Leu
435 440 445

Ser Ala Tyr Ile His Lys Phe Glu Ser Ile Gln Ala Phe Ile Asp Val
450 455 460

Leu Glu Ser Lys Glu Leu Asn Glu Lys Val Val Phe Ser Ala Lys Lys
465 470 475 480

Gly Arg Asn Phe Ser Ala Phe Asn Glu His Gly Phe Ala Gly Arg Val
485 490 495

Ala Asp Gln Leu Arg Val Leu Ala Ser Ile Gly Lys Met Lys Pro Asp
500 505 510

Leu Glu Gly Ala Lys Arg Ala Leu Phe Arg Asp Ala Val Glu Thr Leu

```
        515                 520                 525

Gly Ile Val Glu Asn Ser Asn Trp Leu Asn Glu Asp Gly Thr Ile Ser
    530                 535                 540

Asn Asp Trp Leu Glu Thr Tyr Val Leu Pro Gly Asp Lys Val Thr Glu
545                 550                 555                 560

Glu Glu Lys Arg Ala Ile Lys Pro Phe Arg Asn Phe Ile Ala Asn Asn
                565                 570                 575

Val Ile Glu Ser Arg Arg Phe Met Tyr Leu Ala Arg Tyr Ala Lys Pro
            580                 585                 590

Lys Ser Val Arg Glu Leu Met Lys His Glu Glu Ile Ile Arg Tyr Val
        595                 600                 605

Leu Thr Arg Leu Pro Glu Lys Gln Ile Asp Thr Tyr Phe Ser Asn Ile
    610                 615                 620

Ser Glu Gly Asp Glu Ser Glu Ile Ser Arg Lys Ile Asp Thr Leu Thr
625                 630                 635                 640

Lys Ala Leu Thr Gly Phe Ser Phe Asn Ser Leu Phe Ile Asn Arg Asp
                645                 650                 655

Lys Ile Val Glu Asn Thr Lys Leu Asp Thr Ser Asn Lys Asn Val Lys
            660                 665                 670

Ile Glu Arg Leu Lys Ala Leu Thr Gly Leu Tyr Leu Thr Val Ala Phe
        675                 680                 685

Val Ala Val Lys Asn Leu Val Lys Ala Asn Ala Arg Tyr Phe Ile Ala
    690                 695                 700

Phe Ser Ala Phe Glu Arg Asp Tyr Gly Phe Met Lys Ser His Asp Pro
705                 710                 715                 720

Asp Val Ala Gln Ser Cys Leu Ala Tyr Glu Glu Ile Met Asp Gly Lys
```

725 730 735

Ser Lys Thr Lys Tyr Asn Glu Ile Tyr Ala Leu Thr Glu Tyr Tyr Leu
740 745 750

Lys Lys Glu Glu Glu Asn Asp Tyr His Pro Glu Pro Gly Gln Pro Phe
755 760 765

Asp Lys Glu Ala Cys Arg Lys His Leu Asp Ser Ile Arg Arg His Phe
770 775 780

Thr Lys Lys Trp Arg Asp Ile Phe Arg Lys Glu Ile Asp Glu Ala Val
785 790 795 800

Lys Ile His Phe Thr Gly Tyr Leu Gly Val Trp Ala Arg Asn Gln Ser
805 810 815

Ala His Leu Asn Val Leu Thr Gln Ala Leu Pro Gln Phe Ile Gly Gln
820 825 830

Phe Ala Arg Asn Ser Lys Thr Gly Lys Thr Ala Arg Met Thr Ser Tyr
835 840 845

Phe Gln Leu Tyr His Tyr Leu Leu Gln Arg Leu Leu Cys Glu Asp Leu
850 855 860

Asn Phe Asp Val Pro Glu Glu Trp Lys Asn Lys Ile Gln Glu Thr Gly
865 870 875 880

Leu Pro Cys Asn Asp Leu Ile His Ile Cys Phe Val Pro Leu Gly Tyr
885 890 895

Asn Leu Pro Arg Tyr Lys Asn Leu Thr Ile Glu Cys Leu Phe Asp Glu
900 905 910

Asp Ser Val Ser Ala Lys Glu Gly Lys Lys Ala
915 920

<210> 223

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

<220>
<223> DR

<400> 223
atacagcagc cctacagatt cgtagggttc tgtgac 36

<210> 224
<211> 993
<212> PRT
<213> Artificial Sequence

<220>
<223> DLF014_GL0011914

<400> 224

Met Ala Lys Lys Lys Asn Ala Lys Gln Arg Arg Glu Glu Glu Lys Asn
1 5 10 15

Arg Leu Lys Ala Lys Ile Glu Lys Phe Lys Ile Lys Tyr Ala Glu Gln
20 25 30

Glu Thr Val Asp Lys Ile Thr Glu Glu Ile Ala Gln Ser Asp Lys Ile
35 40 45

Met Arg Thr Glu Ser Ile Leu Val Gln His Thr Lys Lys Ser Leu Ala
50 55 60

Lys Ala Ser Gly Val Lys Ser Val Phe Ile Lys Asp Asp Glu Val Ile
65 70 75 80

Met Thr Ser Phe Gly Arg Gly Asn Asp Ala Val Val Glu Lys Ile Ile
85 90 95

Lys Asp Asn Asn Val Asp Glu Lys Asn Thr Lys Pro Ala Tyr Glu Ile
100 105 110

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

```
Cys Lys Asn Ile Glu Ser Ala Asn Thr Glu Leu Pro Pro Glu Lys Ile
    130                 135                 140

Gly Met Asp Leu Ile Arg Cys Lys Asp Lys Leu Glu Glu Ile Tyr Phe
145                 150                 155                 160

Gly Arg Thr Phe Asn Asp Asn Ile His Ile Gln Leu Ile Tyr Asn Ile
                165                 170                 175

Leu Asp Ile Glu Lys Ile Leu Ser Val Tyr Ile Asn Asn Ala Val Tyr
            180                 185                 190

Ala Leu Gly Asn Leu Glu Arg Lys Glu Asn Asp Lys Glu Lys Asp Leu
        195                 200                 205

Val Gly Tyr Leu Ser Ala Arg Ala Gly Tyr Asn Lys Val Ile Asn Lys
    210                 215                 220

Ala Ile Asn Asp Glu Lys Ile Lys Asp Arg Ile Glu Asp Leu Lys Glu
225                 230                 235                 240

Phe Ile Glu Asn Gly Asn Arg Leu Gly Tyr Phe Gly Asp Ile Phe Phe
                245                 250                 255

Lys Lys Glu Glu Val Ser Lys Asn Gly Lys Lys Pro Lys Lys Val Glu
            260                 265                 270

Val Leu Lys Ser Glu Lys Glu Ile Phe Asp Ile Leu Ala Leu Leu Gly
        275                 280                 285

Ser Leu Arg Gln Phe Cys Phe His Tyr Asp Glu Ala Val Phe Glu Asn
    290                 295                 300

Glu Asp Gly Val Ile Asp Lys Glu Tyr Asn Asp Asn Ser Trp Leu Tyr
305                 310                 315                 320

Asn Leu Glu Lys Leu Glu Pro Glu Phe Lys Asp Thr Leu Asp Tyr Phe
                325                 330                 335
```

```
Tyr Asn Arg Lys Ile Asn Asp Ile Asn Lys Gly Phe Ile Lys Thr Asn
            340                 345                 350

Gln Ile Asn Leu His Ile Ile Cys Asp Glu Leu Gly Leu Asn Trp Asp
        355                 360                 365

Lys Glu Gln Val Val Gly Asp Tyr Tyr Asp Phe Leu Ile Ala Lys Lys
    370                 375                 380

Tyr Lys Asn Ile Gly Phe Ser Ile Lys Lys Ile Arg Glu His Met Phe
385                 390                 395                 400

Asp Ile Tyr Glu Ala Phe Asp Ile Lys Asp Lys Arg Phe Asp Ser Val
                405                 410                 415

Arg Ser Ile Met Tyr Lys Ile Ile Asp Phe Val Ile Tyr Tyr Ser Phe
            420                 425                 430

Ile His Tyr Lys Gln Asp Glu Val Asp Glu Ile Val Asn Tyr Leu Arg
        435                 440                 445

Ser Thr Leu Ser Glu Glu Gln Lys Lys Glu Lys Tyr Ile Lys Leu Ala
    450                 455                 460

Gly Asn Ile Trp Ala Glu Tyr Lys Thr Gln Phe Asn Lys Leu Lys Asp
465                 470                 475                 480

Ile Leu Leu Ala Asn Ile Lys Glu Phe Ser Asp Ala Lys Glu Asp Asp
                485                 490                 495

Val Tyr Tyr Glu Glu Phe Lys Asn Phe Lys Phe Asn Glu Val Gly Lys
            500                 505                 510

Glu Lys Leu Gly Glu Asn Ala Glu Tyr Phe Cys Lys Leu Met Tyr Leu
        515                 520                 525

Leu Thr Leu Phe Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr Thr
    530                 535                 540
```

```
Leu Ile Asn Lys Phe Asp Asn Ile Arg Ser Phe Val Glu Leu Met Lys
545                 550                 555                 560

Glu Glu Asn Ile Glu Cys Ile Phe Asp Lys Lys Phe Ala Phe Phe Asp
                565                 570                 575

Glu Ser Lys Asn Val Cys Ser Thr Leu Arg Glu Ile Asn Ser Phe Ala
            580                 585                 590

Arg Met Gln Lys Pro Leu Asn Asn Val Ser Val Lys Arg Glu Met Tyr
        595                 600                 605

Lys Asp Ala Ile Lys Ile Leu Leu Lys Asp Thr Arg Ala Glu Glu Gln
    610                 615                 620

Asn Ile Asp Arg Ile Leu Asn Glu Tyr Ile Pro Ser Lys Glu Asp Lys
625                 630                 635                 640

Gly Ile Lys Lys Asp Phe Arg Asn Phe Ile Ile Lys Asn Ile Ile Lys
                645                 650                 655

Ser Asn Arg Phe Met Tyr Leu Val Lys Tyr Ser Asn Pro Thr Asp Val
            660                 665                 670

Arg Asn Leu Ala Ser Asn Arg Asn Val Val Lys Phe Val Leu Asn Ala
        675                 680                 685

Ile Pro Asp Thr Gln Ile Asp Arg Tyr Tyr Lys Ser Cys Ile Gln His
    690                 695                 700

Tyr Ser Asn Glu Val Asp Lys Glu Thr Gln Ile Gln Glu Leu Thr Asp
705                 710                 715                 720

Ile Ile Thr Gly Ile Asp Tyr Leu Asp Phe Leu Asp Val Arg Gln Ser
                725                 730                 735

Tyr Arg Asn Glu Asp Lys Ser Lys Lys Gln Ala Val Val Thr Leu Tyr
            740                 745                 750
```

```
Leu Thr Ile Leu Tyr Ile Leu Thr Lys Asn Leu Val Asn Val Asn Ser
        755                 760                 765

Arg Tyr Val Ile Ala Leu His Cys Leu Glu Arg Asp Ser Lys Leu Tyr
    770                 775                 780

Gly Tyr Thr Lys Leu Ser Lys Lys Asn Asn Lys Lys Leu Glu Lys Ile
785                 790                 795                 800

Lys Pro Pro Lys Tyr His Glu Leu Thr Lys Tyr Phe Ile Asp Asn Arg
                805                 810                 815

Tyr Phe Asp Arg Lys Lys Lys Asn Lys Lys Tyr Gly Glu Tyr Thr Ser
            820                 825                 830

Lys Lys Ile Ser Thr Tyr Ile Glu Thr Asn Met Lys Asn Phe Ile Gly
        835                 840                 845

Cys Glu Gln Phe Arg Thr Pro Glu Arg Tyr Arg Asp Thr Gln Ile Asp
    850                 855                 860

Met Phe Val Lys Tyr Arg Asn Ser Ile Ala His Leu Asn Ala Val Arg
865                 870                 875                 880

Lys Ala Ser Lys Tyr Ile Asn Asp Ile Arg Gly Phe Asp Thr Tyr Phe
                885                 890                 895

Glu Leu Tyr His Tyr Ile Met Gln Arg Asp Leu Lys Asp Ile Leu Glu
            900                 905                 910

Lys Ile Ala Glu Gly Ala Ala Lys Glu Thr Glu Glu Gly Gln Lys Gln
        915                 920                 925

Asp Glu Leu Leu Lys Thr Tyr Phe Asp Asn Leu Asp Lys Tyr Gly Thr
    930                 935                 940

Tyr Val Lys Asp Phe Val Lys Ala Leu Asn Ile Pro Phe Ala Tyr Asn
945                 950                 955                 960
```

Tyr Pro Arg Tyr Lys Asn Leu Ser Ile Asp Glu Leu Phe Asp Lys Asn
965 970 975

Asn Thr Arg Lys Thr Ile Lys Ile Gln Asp Lys Lys Leu Glu Glu Arg
980 985 990

Thr

<210> 225
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 225
gaactactac ctcacagatt tgtggggttc agaaac 36

<210> 226
<211> 924
<212> PRT
<213> Artificial Sequence

<220>
<223> EYZ-362B_GL0088915

<400> 226

Met Lys Lys Lys Leu Asn Ile Arg Glu Gln Arg Glu Ala Lys Lys Ala
1 5 10 15

Ala Ala Lys Gln Ile Tyr Ala Ala Lys Ala Ser Ala Ala Gln Lys Lys
20 25 30

Ala Asp Glu Thr Val Val Ala Ala Gln Pro Ala Gly Asp Thr Ala Asn
35 40 45

Ser Lys Pro Ile Lys Ser Arg Asn Lys Ala Ala Gly Leu Lys Ser Thr
50 55 60

Phe Val Val Gly Asp Ser Leu Tyr Met Thr Ser Phe Gly Lys Gly Asn

65 70 75 80

Leu Ser Val Pro Glu Tyr Arg Val Asp Thr Gly Asp Ala Tyr Arg Val
85 90 95

Asp Pro Ile Ser Glu His Ser Glu Leu His Val Leu Gly Ala Asp Asp
100 105 110

Leu Ala Val Ser Phe Ser Gly Thr Arg Arg Met Lys Asp Val Ser Glu
115 120 125

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

Lys Asn Gly Thr Asp Ile Leu Gly Leu Lys Gly Ile Leu Glu Gln Arg
145 150 155 160

Tyr Tyr Gly Arg Asn Phe Asp Asp Asn Val His Ile Gln Ile Ile Tyr
165 170 175

Asn Ile Leu Asp Ile Glu Lys Asn Leu Ala Ile Tyr Val Thr Asn Val
180 185 190

Val Ala Ala Leu Asp His Met Cys Asp Glu Thr Leu Asp Thr Glu His
195 200 205

Glu Asp Phe Ile Gly Tyr Met Asn Thr Met Asn Ser Phe Asp Val Phe
210 215 220

Met His Pro Ser Gln Ser Ser Arg Ala Asp Ile Tyr Lys Lys Ile Asn
225 230 235 240

Asp Ser Arg Arg Arg Phe Glu Asn Leu Leu Arg Thr Lys Arg Leu Gly
245 250 255

Tyr Phe Gly Phe Ala Tyr Asn Pro Lys Asp Arg Asn Gln Lys Lys Arg
260 265 270

Leu Tyr Tyr Leu Met Ser Met Ile Gly Gln Leu Arg Gln Phe Ala Phe

```
        275                 280                 285

His Lys Asp Ser Pro Phe Asn Arg Val Trp Leu Tyr Gln Leu Asp Glu
    290                 295                 300

Lys Leu Ser Lys Glu His Lys Asp Thr Leu Asn Tyr Phe Phe Asp Gln
305                 310                 315                 320

Arg Tyr Asp Glu Ile His Glu Asn Phe Asn Lys Thr Asn Lys Val Asn
                325                 330                 335

Leu Glu Leu Leu Ala Glu Ala Phe Pro Ser Glu Asp Tyr Ala Thr Leu
            340                 345                 350

Ala Asn Gln Tyr Tyr Asp Phe Thr Val Phe Lys Ala His Lys Asn Leu
        355                 360                 365

Gly Phe Ser Ile Lys Lys Leu Arg Glu Lys Leu Ile Glu Asn Ala Leu
    370                 375                 380

Arg Asp Glu Lys Ile Asp Ser Phe Arg Ser Lys Leu Tyr Arg Leu Ile
385                 390                 395                 400

Asp Phe Cys Leu Phe Arg Ile Tyr Leu Lys Asp Pro Asp Arg Ala Glu
                405                 410                 415

Lys Asn Val Arg Phe Leu Arg Ser Cys Ile Asn Glu Gln Glu Lys Glu
            420                 425                 430

Gln Phe Tyr Thr Gln Glu Ala Asn Ala Leu Lys Ser Glu Asn Pro Gly
        435                 440                 445

Leu Val Asn Asp Phe Cys Met Gln Leu Lys Asp Arg Tyr Ser Ser Gly
    450                 455                 460

Ser Asn Ser Thr Glu Asn Ala Glu Lys Lys Glu Asp Phe Pro Val Phe
465                 470                 475                 480

Pro His Cys His Asp Lys Gln Lys Val Ser Asn Val Ser Leu Phe Ala
```

485 490 495

Lys Leu Met Tyr Ala Phe Cys Leu Phe Leu Asp Gly Lys Glu Ile Asn
500 505 510

Glu Leu Leu Ser Ala Leu Leu Asn Lys Phe Asp Thr Ile Ala Cys Met
515 520 525

Thr Lys Cys Thr Lys Asp Leu Gly Ile Tyr Ala Pro Phe Ser Lys Asp
530 535 540

Phe Asn Phe Phe Glu Asp Ser Glu Gln Arg Ala Glu Glu Met Asn Leu
545 550 555 560

Val Lys Ser Ile Ala Arg Met Thr Lys Pro Asp Ile Gln Ala Lys Glu
565 570 575

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

Ser Glu Glu Ala Leu Gln Lys Glu Ile His Gln Tyr Phe Gln Lys Gly
595 600 605

Ala Lys His Asp Phe Arg Asn Phe Ile Ile Asn Asn Val Ile Asn Ser
610 615 620

Thr Arg Phe Ile Tyr Val Ile Lys Phe Cys Asn Pro Lys Thr Ala Gly
625 630 635 640

Gln Ile Ile Arg Asn Glu Asn Ile Leu Arg Phe Val Leu Gln Arg Met
645 650 655

Pro Gln Glu Gln Ile Glu Arg Tyr Tyr His Ala Cys Thr Lys Asn Pro
660 665 670

Ser Leu Cys Thr His Asp Glu Lys Ile Gln Cys Leu Ser Asn Ile Ile
675 680 685

Met Glu Val Arg Tyr Lys Lys Phe Cys Asn Val Ala Gln Lys Ser Arg

690 695 700

Asp Thr Thr Lys Gln Met Glu Lys Glu Arg Phe Lys Ala Ile Ile Arg
705 710 715 720

Leu Tyr Leu Thr Ala Ile Tyr Leu Val Val Lys Asn Leu Val Asn Val
725 730 735

Asn Ala Arg Tyr Val Leu Ala Phe His Cys Leu Glu Arg Asp Ala Asp
740 745 750

Leu Phe Gly Leu Gln Ile Lys Gln Thr Leu Pro Asp Arg Lys Lys Lys
755 760 765

Ser Asp Tyr Leu Lys Leu Ala Val Asp Gln Cys Gly Asp Asn Pro Glu
770 775 780

Asn Lys Pro Lys Gly Tyr Leu Ala Arg Asn Lys Arg Leu Cys Gln Cys
785 790 795 800

Val Arg Glu Asp Ile Asp Asn Ala Lys His Leu Asn Gly Asp Ile Thr
805 810 815

Asp Tyr Arg Asn Cys Val Ala His Leu Asn Val Val Arg Asn Cys Asp
820 825 830

Gln Tyr Leu Glu Asn Ile Asp Thr Val His Ser Tyr Phe Ser Leu Tyr
835 840 845

Gln Tyr Leu Val Gln His Tyr Leu Leu Lys Arg Phe Glu Lys Asp Lys
850 855 860

Tyr Arg Cys Lys Asp Asp Leu Ala Ile His Tyr Tyr Ser Gln Val Lys
865 870 875 880

Gln Tyr Arg Ser Tyr Val Lys Asp Leu Val Lys Ala Leu Asn Ala Pro
885 890 895

Phe Gly Tyr Asn Ile Pro Arg Phe Lys Asn Leu Thr Val Glu Gln Leu

900 905 910

Phe Asp Arg Asn Ala Asp Pro Glu Thr Ile Thr Ala
915 920

<210> 227
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR1

<400> 227
gaactatagt agtgtgaatt tacactactc taaaac 36

<210> 228
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR2

<400> 228
ctactacact ggtgcaaatt tgcactagtc taaaac 36

<210> 229
<211> 1054
<212> PRT
<213> Artificial Sequence

<220>
<223> Ga0099364_10024192

<400> 229

Met Ser Gln Ser Thr Lys Thr Lys Ala Lys Arg Met Gly Val Lys Ser
1 5 10 15

Val Leu Ala His Gly Lys Asp Glu Lys Gly His Ile Lys Leu Ala Ile
20 25 30

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

```
Glu Lys Gly Ser Asn Leu Ala Lys Thr Tyr Lys Glu Arg Asn Ile Thr
    50                  55                  60

Ala Asn Lys Ile Val Ser Glu Gly Ile Gln Thr Ser Gly Thr Ile Ala
65                  70                  75                  80

Gly Glu Gly His Ala Thr Phe Leu Asn Asn Pro Ala Glu His Val Gly
                85                  90                  95

Thr Asp Tyr Leu Lys Leu Lys Glu Thr Leu Glu Met Glu Phe Phe Gly
            100                 105                 110

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

Leu Asp Ile Gln Lys Leu Leu Gly Ile Tyr Ile Thr Asp Ile Ile Tyr
    130                 135                 140

Cys Ile Asn Asn Leu Arg Asp Glu Thr His Leu Asp His Glu Ser Asp
145                 150                 155                 160

Ile Val Gly Leu Ser Met Ser Asn Thr Asn Val Asn Leu Ala Leu Asn
                165                 170                 175

Gln Met Arg Pro Tyr Phe Gly Phe Phe Gly Glu Ala Phe Arg Pro Val
            180                 185                 190

Gly Asp Asp Lys Val Lys Glu Ile Thr Leu Ser Asp Glu Val Arg Lys
        195                 200                 205

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

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

Met Gly Lys Phe Lys Ser Lys Asp Ala Phe Glu Thr Ala Lys Lys Lys
                245                 250                 255
```

```
Tyr Asn Arg Ile Val Ala Asp Glu Thr Asn Ala Lys Thr Leu Arg Ile
            260                 265                 270

Leu Gly Ala Met Arg Gln Ile Thr Ala His Phe Lys Asp Gln Ala Thr
        275                 280                 285

Leu Phe Met Ser Asp Val Glu Leu Pro Lys Ile Leu Lys Lys Glu Phe
    290                 295                 300

Ser Lys Ala Asp Trp Gln Thr Val Glu Asp Tyr Tyr Ala Lys Leu Val
305                 310                 315                 320

Asp Arg Ile Asn Glu Gly Phe Cys Lys Asn Ala Ala Thr Asn Val His
                325                 330                 335

Phe Leu Thr Glu Leu Leu Pro Glu Glu Ser Lys Lys Gln Leu Thr Glu
            340                 345                 350

Asp Tyr Phe Arg Phe Ala Ile Leu Lys Glu Gly Lys Asn Leu Gly Val
        355                 360                 365

Asn Met Lys Arg Leu Arg Glu Val Met Phe Ala Leu Phe Val Pro Glu
    370                 375                 380

Leu Thr Ala Pro Glu Thr Lys Lys Arg Tyr Asp Ser Tyr Arg Ala Lys
385                 390                 395                 400

Ile Tyr Gly Leu Thr Asp Phe Leu Leu Phe Lys His Ile His Asn Thr
                405                 410                 415

Lys Gln Leu Glu Glu Trp Val Ala Val Leu Arg Glu Thr Ser Asn Glu
            420                 425                 430

Asp Ala Lys Glu Asn Leu Tyr Asp Glu Phe Ala Arg Thr Ala Trp Asn
        435                 440                 445

Thr Val Gly Asp Ser Ala Lys Gln Leu Ile Glu Asn Met Gln Ser Tyr
    450                 455                 460
```

```
Phe Thr Lys Lys Glu Lys Glu Ile Thr Lys Thr Ala Gln Pro Val Leu
465                 470                 475                 480

Ser Thr Ser Ser Ile Ala His Thr Ser Lys Lys Ile Thr Gln Phe Ser
                485                 490                 495

Ser Phe Ala Lys Leu Leu Ala Phe Leu Cys Asn Phe Trp Glu Gly Lys
            500                 505                 510

Glu Ile Asn Glu Leu Leu Ser Ala Tyr Ile His Lys Phe Glu Asn Ile
        515                 520                 525

Gln Glu Phe Ile Asn Leu Leu Glu Lys Leu Glu Gly Lys Lys Pro Gln
    530                 535                 540

Phe Thr Glu Asn Tyr Ala Leu Phe Asn Glu Ala Ala Gly Gln Arg Ala
545                 550                 555                 560

Gly Glu Ile Ala Gln Asn Leu Arg Ile Leu Ala Ser Ile Gly Lys Met
                565                 570                 575

Lys Pro Asp Leu Gly Asp Ala Lys Arg Gln Leu Tyr Lys Ala Ala Ile
            580                 585                 590

Glu Met Leu Gly Ile Asp Thr Glu Glu Tyr Ile Ser Asp Glu Trp Leu
        595                 600                 605

Glu Pro Asn Met Leu Leu Ala Gln Pro Pro Lys Glu Pro Lys Lys Asp
    610                 615                 620

Asn Glu Lys Tyr Arg Lys Glu Pro His Lys Tyr Ser Tyr Glu Lys Asp
625                 630                 635                 640

Met Glu Thr Tyr Arg Lys Lys Leu Arg Glu Tyr Glu Glu Thr Trp Arg
                645                 650                 655

Ser Leu Ile Asp Tyr Glu Tyr Leu Met Pro Glu Thr Asn Pro Phe Arg
            660                 665                 670
```

```
Asn Phe Val Ala Lys Gln Val Ile Glu Ser Arg Arg Phe Met Tyr Leu
        675                 680                 685

Val Arg Tyr Thr Lys Pro Lys Thr Val Arg Ala Leu Met Ser Asn Arg
    690                 695                 700

Ala Ile Val His Tyr Val Leu Ser Arg Ile Ala Asp Ile Gln Asp His
705                 710                 715                 720

His Met Thr Glu Ser Gln Ile Asp Arg Tyr Tyr Gln Asn Leu Pro Gln
                725                 730                 735

Tyr Asn Glu Gln Gln His Lys Asn Val Ser Leu Glu Thr Lys Ile Asp
            740                 745                 750

Ala Leu Ala Asp Tyr Leu Cys Lys Tyr Thr Phe Glu Lys Asn Val Leu
        755                 760                 765

Lys Gln Lys Asn Gly Ile Val Leu Asn Thr Lys Ser Ala Thr Lys Asn
    770                 775                 780

Val Glu Ile Glu His Leu Lys Ala Leu Thr Gly Leu Tyr Leu Thr Val
785                 790                 795                 800

Ala Tyr Ile Ala Val Lys Asn Leu Val Lys Ala Asn Ala Arg Tyr Tyr
                805                 810                 815

Ile Ala Phe Ser Ile Phe Glu Arg Asp Tyr Ala Leu Phe Glu Lys Lys
            820                 825                 830

Leu Gly Lys Asp Thr Leu Glu Lys Tyr Val Lys Pro Phe Lys Tyr Ile
        835                 840                 845

Asp Lys Gly Glu Glu Lys Glu Gly Lys Asn Asn Phe Phe Ala Leu Thr
    850                 855                 860

Glu Tyr Leu Leu Asp Lys Asp Asn Ser Leu Arg Tyr Gln Trp Asn Asn
865                 870                 875                 880
```

```
Asp Leu Ser Asp Glu Glu Asn Lys Gln Ala Leu Arg Lys His Leu Asp
                885                 890                 895


Lys Lys Glu Ile Arg Ser Gln Arg His Phe Ser Gln Tyr Trp Leu Asp
            900                 905                 910


Ile Phe Ala Arg Gln Ile Glu Asn Ala Lys Lys Thr Ser Glu Ser Gly
        915                 920                 925


Tyr Leu Leu Thr Ala Ala Arg Asn Cys Ala Leu His Leu Asn Val Leu
    930                 935                 940


Thr Ala Leu Pro Glu Phe Val Gly Glu Phe Arg Lys Thr Gly Asp Lys
945                 950                 955                 960


Met Thr Ser Tyr Phe Glu Leu Tyr His Phe Leu Leu Gln Lys Leu Met
                965                 970                 975


Leu Ala Glu Ala Gly Leu Asn Leu Asp Glu Tyr Arg Glu Arg Ile Asp
            980                 985                 990


Thr Tyr Gln Thr Ala Cys Lys Asp  Leu Ile Asn Ile Thr  Tyr Val Ser
        995                 1000                1005


Leu Gly  Tyr Asn Leu Pro Arg  Tyr Lys Asn Leu Thr  Cys Glu Pro
    1010                1015                1020


Leu Phe  Asp Glu Glu Ser Ala  Thr Gly Lys Glu Arg  Gln Thr Arg
    1025                1030                1035


Leu Asp  Glu Lys Ser Lys Glu  Lys Lys Gln Arg Lys  Gly Gly Gln
    1040                1045                1050


Lys



<210>  230
<211>  36
<212>  DNA
```

<213> Artificial Sequence

<220>
<223> DR

<400> 230
gtgcagtagc cttacagatt cgtagggttc tgagac 36

<210> 231
<211> 971
<212> PRT
<213> Artificial Sequence

<220>
<223> Ga0187910_10006931

<400> 231

Met Ile Glu Lys Lys Lys Ser Tyr Ala Lys Gly Met Gly Leu Lys Ser
1 5 10 15

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

Asn Asp Ala Arg Leu Glu Lys Val Val Glu Asn Asn Ala Ile Ser Cys
        35 40 45

Leu Val Asp Lys Lys Glu Ala Phe Val Ala Glu Ile Thr Asp Lys Asn
    50 55 60

Ala Gly Tyr Lys Ile Ile Asn Lys Lys Phe Gly His Pro Lys Gly Tyr
65 70 75 80

Asp Val Val Ala Asn Asn Pro Leu Tyr Thr Gly Pro Val Gln Gln Asp
                85 90 95

Met Leu Gly Leu Lys Glu Thr Leu Glu Lys Arg Tyr Phe Gly Ser Ser
            100 105 110

Val Ser Gly Asn Asp Asn Ile Cys Ile Gln Val Ile His Asn Ile Leu
        115 120 125

Asp Ile Glu Lys Ile Leu Ala Glu Tyr Ile Thr Asn Ala Ala Tyr Ala

130 135 140

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

Phe Ser Thr Val Tyr Thr Phe Asp Glu Phe Ala Glu Pro Asp Arg His
165 170 175

Lys Glu Arg Phe Ile Lys Asp Gly Lys Leu Asp Thr Lys Leu Ile Asn
180 185 190

Gln Leu Lys Asn Gln Tyr Asp Glu Phe Asp Ala Phe Leu Asp Asp Thr
195 200 205

Arg Phe Gly Tyr Phe Gly Lys Ala Phe Phe Cys Lys Glu Gly Asp Lys
210 215 220

Tyr Leu Asn Lys Gln Asp Asn Glu Arg Tyr His Ile Leu Ala Leu Leu
225 230 235 240

Ser Gly Leu Arg Asn Trp Val Val His Asn Asn Glu Val Glu Ser Lys
245 250 255

Ile Asp Arg Lys Trp Leu Tyr Asn Leu Asp Lys Asn Leu Asp Lys Glu
260 265 270

Tyr Ile Thr Thr Leu Asp Tyr Met Tyr Ser Asp Ile Ala Asp Glu Leu
275 280 285

Thr Lys Ser Phe Ser Lys Asn Ser Ala Ala Asn Val Asn Tyr Ile Ala
290 295 300

Glu Ile Leu Asn Ile Asp Ser Lys Thr Phe Ala Glu Gln Tyr Phe Arg
305 310 315 320

Phe Ser Ile Met Lys Glu Gln Lys Asn Leu Gly Phe Thr Leu Thr Lys
325 330 335

Leu Arg Glu Cys Met Leu Asp Arg Glu Glu Leu Ser Asp Ile Arg Asp

340 345 350

Asn His Lys Val Phe Asp Ser Ile Arg Ser Lys Leu Tyr Thr Met Met
355 360 365

Asp Phe Val Ile Tyr Arg Tyr Tyr Ile Glu Glu Ala Lys Lys Ile Glu
370 375 380

Asn Glu Asn Lys Thr Leu Ser Asp Asp Lys Lys Lys Leu Ser Glu Lys
385 390 395 400

Asp Ile Phe Ile Ile Ser Leu Arg Gly Ser Phe Ser Glu Glu Gln Lys
405 410 415

Asp Lys Leu Tyr Ser Asp Glu Ala Glu Arg Leu Trp Ala Lys Leu Gly
420 425 430

Lys Leu Met Leu Glu Ile Lys Lys Phe Arg Gly Gln Met Thr Arg Asp
435 440 445

Tyr Lys Lys Ser Asp Thr Pro Thr Leu Asn Arg Ile Leu Pro Glu Ser
450 455 460

Glu Asp Val Ser Thr Phe Ser Lys Leu Met Tyr Ala Leu Thr Met Phe
465 470 475 480

Leu Asp Gly Lys Glu Ile Asn Glu Leu Leu Thr Thr Leu Ile Asn Lys
485 490 495

Phe Asp Asn Ile Gln Ser Met Leu Lys Ile Met Pro Leu Ile Gly Val
500 505 510

Asn Ala Lys Phe Ser Ser Asp Tyr Ala Phe Phe Asn Asn Ser Glu Lys
515 520 525

Ile Ala Asp Glu Leu Lys Leu Ile Lys Ser Phe Ala Arg Met Gly Glu
530 535 540

Pro Val Ala Asn Ala Lys Arg Asp Met Met Ile Asp Ala Ile Lys Ile

545 550 555 560

Leu Gly Thr Asp Leu Asp Asp Asn Glu Ile Lys Lys Leu Ala Asp Ser
565 570 575

Phe Phe Lys Asp Ser Asn Gly Lys Leu Leu His Lys Gly Lys His Gly
580 585 590

Met Arg Asn Phe Ile Ile Asn Asn Val Val Asn Asn Lys Arg Phe His
595 600 605

Tyr Ile Ile Arg Tyr Gly Asp Pro Ala His Leu His Glu Ile Ala Lys
610 615 620

Asn Glu Val Val Val Arg Phe Val Leu Gly Arg Ile Ala Asp Ile Gln
625 630 635 640

Lys Lys Gln Gly Lys Gly Gly Lys Asn Gln Ile Asp Arg Tyr Tyr Glu
645 650 655

Ile Cys Ile Gly Asn Gly Tyr Gly Lys Ser Val Ser Glu Lys Ile Asp
660 665 670

Ala Leu Thr Lys Val Ile Ile Asn Met Asn Tyr Asp Gln Phe Glu Ala
675 680 685

Lys Arg Lys Val Ile Glu Asn Thr Gly Arg Asp Asn Ala Glu Arg Glu
690 695 700

Lys Tyr Lys Lys Ile Ile Ser Leu Tyr Leu Thr Val Ile Tyr Gln Ile
705 710 715 720

Leu Lys Asn Leu Val Asn Val Asn Ser Arg Tyr Val Ile Gly Phe His
725 730 735

Cys Val Glu Arg Asp Ala Gln Leu Tyr Lys Glu Lys Gly Tyr Asp Ile
740 745 750

Asn Thr Asn Asn Leu Glu Ser Lys Gly Phe Thr Ser Val Thr Lys Leu

```
        755                 760                 765

Cys Val Gly Ile Ala Asp Asp Asp Pro Val Lys Tyr Lys Asn Val Glu
    770                 775                 780

Ile Glu Leu Lys Glu Arg Ala Leu Ala Ser Phe Asp Ala Leu Glu Lys
785                 790                 795                 800

Glu Asn Pro Glu Leu Tyr Glu Lys Tyr Asn Met Tyr Ser Glu Lys Gln
                805                 810                 815

Lys Glu Ala Glu Leu Glu Lys Gln Ile Asn Arg Glu Lys Ala Lys Thr
            820                 825                 830

Ala Leu Asn Ala His Leu Arg Asn Thr Lys Trp Asn Val Ile Ile Arg
        835                 840                 845

Glu Asn Ile Arg Asn Thr Glu Lys Asp Ala Cys Lys Gln Phe Arg Asn
    850                 855                 860

Lys Ala Asp His Leu Glu Val Ala Arg Tyr Ala Tyr Lys Tyr Ile Asn
865                 870                 875                 880

Asp Ile Ser Glu Val Asn Ser Tyr Phe Gln Leu Tyr His Tyr Ile Met
                885                 890                 895

Gln Arg Ile Ile Ile Asp Ser Ser Gly Asn Asn Ala Asn Gly Met Ile
            900                 905                 910

Lys Lys Tyr Tyr Glu Ser Val Ile Ser Asp Lys Lys Tyr Asn Asp Arg
        915                 920                 925

Leu Leu Lys Leu Leu Cys Val Pro Phe Gly Tyr Cys Ile Pro Arg Phe
    930                 935                 940

Lys Asn Leu Ser Ile Glu Ala Leu Phe Asp Lys Asn Glu Ala Ala Lys
945                 950                 955                 960

Tyr Asp Lys Ile Lys Lys Lys Val Ala Val Arg
```

965 970

<210> 232
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 232
catgtaaacc cctaacaaat ggtaggggtt tgaaac 36

<210> 233
<211> 1105
<212> PRT
<213> Artificial Sequence

<220>
<223> Ga0187910_10015336

<400> 233

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

Arg Phe Glu Ile Tyr Glu Asn Pro Lys Pro Lys Lys Gln Asn Gln Arg
20 25 30

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

Ala Lys Ile Gln Lys Asn Glu Phe Lys Thr Phe Asp Gln Phe Thr Lys
50 55 60

Glu Leu Gln Glu Thr Gln Lys Val Asn Gly Glu Thr Lys Lys Glu His
65 70 75 80

Ile Thr Lys Asn Lys His Thr Asn Val Lys Ala Ala Gly Val Lys Ser
85 90 95

Val Phe Ala Val Asp Asp Gly Asn Val Leu Ile Thr Ser Phe Gly Arg
100 105 110

```
Gly Asn Ala Ala Asp Ile Glu Thr Leu Lys Ser Asp Asp Asp Lys Thr
        115                 120                 125


Ile Asn Leu Thr Glu Thr Glu Asn Gln Lys Lys Tyr Val Val Thr Asn
    130                 135                 140


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


Ser Ile Ile Pro Gly Glu Thr Gln Ile Gly Phe Lys Ser Ile Leu Glu
                165                 170                 175


Lys His Phe Phe Gly Arg Thr Phe Asn Asp Asn Ile His Ile Gln Ile
            180                 185                 190


Ile His Asn Ile Leu Asp Ile Lys Lys Ile Leu Ala Val His Thr Asn
        195                 200                 205


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


Ser Ala Glu Lys Pro Ile Asp Met Ile Gly Ala Gly Gly Ile Ser Thr
225                 230                 235                 240


Ser Lys Glu Tyr Glu Gln Tyr Cys Ser Glu Lys Ser Asp Tyr Glu Asp
                245                 250                 255


Asn Phe Leu Lys Gln Leu Ile Asn Asn Glu Arg Ile Ala Tyr Phe Gly
            260                 265                 270


Asn Ala Phe Phe Lys Asp Glu Gly Asn Lys Lys Val Met Arg Thr Glu
        275                 280                 285


Lys Glu Ile Tyr Tyr Ile Leu Gly Met Leu Asn Glu Val Arg Asn Val
    290                 295                 300


Ser Thr His Phe Thr Glu Glu Asp Asn Arg Asp Trp Ala Lys Ala Asn
305                 310                 315                 320
```

```
Leu Tyr Asn Leu Ser Asn Arg Leu Lys Val Gly Ser Lys Glu Val Leu
                325                 330                 335

Asn Gln Leu Tyr Lys Glu Lys Ile Asp Lys Ile Asp Ala Asn Gly Phe
            340                 345                 350

Val Asn Lys Gly Cys Lys Arg Asp Phe Ser Ile Leu Phe Lys Ser Leu
        355                 360                 365

Asn Leu Thr Thr Asp Lys Asp Lys Gly Glu Leu Val Val Gly Phe Tyr
    370                 375                 380

Asp Phe Ser Ile Arg Lys Asn Tyr Lys Asn Ile Gly Phe Ser Ile Lys
385                 390                 395                 400

Thr Leu Arg Glu Tyr Met Leu Lys Ile Ser Asn Ser Thr Leu Cys Ala
                405                 410                 415

Asp Thr Ile Ser Asn Asn Ala Ile Arg Pro Lys Ala Tyr Lys Leu Tyr
            420                 425                 430

Asp Phe Ile Ile Trp His Tyr Tyr Met Asn Lys Pro Asp Lys Ile Asn
        435                 440                 445

Asp Phe Val Glu Lys Leu Arg Thr Gln Asn Lys Asn Asp Glu Lys Ile
    450                 455                 460

Lys Leu Tyr Tyr Asp Glu Ala Val Cys Leu Leu Ser Glu Leu Gly Arg
465                 470                 475                 480

Glu Ile His Thr Met Thr Ser Cys Val His Asn Ile Glu Asn Thr Ser
                485                 490                 495

Tyr Glu Ile Thr Asp Lys Lys Gln Lys Glu Tyr Tyr Lys Met Gln Ile
            500                 505                 510

Asn Ser Leu Asn Ser Ala Asp Lys Val Ser Asp Phe Ser Lys Val Ile
        515                 520                 525
```

```
Tyr Leu Val Thr Leu Phe Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu
    530                 535                 540

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

Leu Glu Lys Gln Ser Gly Lys Lys Val Glu Phe Val Glu Asn Tyr Ser
                565                 570                 575

Phe Phe Asn Ser Ser Asn Leu Leu Lys Glu Lys Thr Leu Asn Lys Ser
            580                 585                 590

Glu Asn Tyr Thr Cys Lys Ile Val Glu Glu Leu Arg Glu Ile Asn Ser
        595                 600                 605

Phe Ala Arg Met Thr Gly Asp Cys Lys Ile Arg Lys Ser Ala Phe Glu
    610                 615                 620

Asp Ala Ser Gln Leu Leu Gly Tyr His Asp Lys Thr Val Asn Asn Leu
625                 630                 635                 640

Phe Glu Val Leu Arg Leu Lys Glu Leu Glu Ser Lys Asp Trp Lys Lys
                645                 650                 655

Arg Thr Asp Asp Glu Gln Gln Glu Tyr Asp Arg Leu Leu Asn Lys His
            660                 665                 670

His Tyr Phe Lys Ser Gly Lys Lys Leu Pro Asp Thr Gly Leu Arg Asn
        675                 680                 685

Phe Ile Ile Asn Asn Val Ile Glu Ser Arg Arg Phe Asn Tyr Ile Val
    690                 695                 700

Arg Tyr Ala Asp Pro Lys Lys Ile Arg Lys Cys Thr Glu Asn Asn Glu
705                 710                 715                 720

Leu Leu Lys Phe Ala Phe Lys Asp Val Pro Asp Ser Gln Val Asp Arg
                725                 730                 735
```

```
Tyr Tyr Asn Ile Cys Val Thr Asn Lys Ile Thr Asn Ala Thr Arg Glu
            740                 745                 750

Glu Lys Ile Glu Arg Leu Val Asp Ile Ile Lys Ser Met Asn Leu Ser
        755                 760                 765

Lys Val Ala Thr Val Lys Gln Arg Asp Lys Gln Asp Asn Val Glu Lys
    770                 775                 780

Gln Lys Gln Leu Ala Ile Met Ser Leu Tyr Leu Asn Ile Leu Tyr Gln
785                 790                 795                 800

Ile Ala Lys Asn Leu Val Tyr Val Asn Ser Arg Tyr Val Met Ala Phe
                805                 810                 815

His Ser Leu Glu Arg Asp Ser Gln Met Leu Phe Asp Ala Tyr Tyr Asp
            820                 825                 830

Val Lys Arg Gly Tyr Cys Asp Leu Ser Thr Val Leu Leu Phe Gly Val
        835                 840                 845

Asp Asp Leu Gln Asn Arg Asn Arg Gly Ser Tyr Lys Tyr Leu Arg Asp
    850                 855                 860

Asn Arg Arg Ser Asn Lys Asp Val Ile Glu Thr Phe Gly Asp Phe Lys
865                 870                 875                 880

Gly Lys Val Ser Lys Val Val Glu Lys Lys Asn Gln Gly Leu Thr Asn
                885                 890                 895

Glu Ile Tyr Asp Ser Leu Cys Asn Val Ala Gly Thr Thr Lys Thr Glu
            900                 905                 910

Val Gln Asn Glu Ile Lys Ser Ile Leu Lys Ser Asn Gly Leu Asp Glu
        915                 920                 925

Ser Ala Ser Ser Tyr Leu Ser His Lys Leu Val Asn Lys Val His Ser
    930                 935                 940
```

```
Tyr Lys Tyr Leu Lys Gln Asn Leu Asp Cys Ala Asp Asn Thr Met Ile
945                 950                 955                 960


Asn Gln Phe Arg Asn Asn Val Ala His Leu Asn Thr Ile Arg Asn Met
                965                 970                 975


Asp Gly Ile Glu Asn Val Thr Gly Ile Thr Ser Tyr Phe Gln Ile Tyr
            980                 985                 990


His Tyr Leu Met Gln Lys Ala Leu  Tyr Lys Glu Phe Lys  Lys Cys Arg
        995                 1000                1005


Glu Asn  Ala Val Arg Lys Trp  Ile Pro Tyr Ile Thr  Glu Asn Ala
    1010                1015                1020


Glu Pro  Lys Tyr Val Tyr Trp  Asn Lys Lys Glu Gln  Gln Glu Val
    1025                1030                1035


Glu Val  Ser Phe Asn Pro Lys  Ile Phe Gly Tyr Met  Glu Asn Ile
    1040                1045                1050


Lys Asn  His Ser Asn Thr Tyr  Cys Lys Asp Phe Val  Lys Ala Leu
    1055                1060                1065


Cys Ala  Pro Phe Ala Tyr Asn  Leu Pro Arg Phe Lys  Asn Leu Ser
    1070                1075                1080


Ile Glu  Glu Leu Phe Asp Met  His Glu Leu Ser Glu  Glu Pro Lys
    1085                1090                1095


Glu Ser  Met Lys Leu Thr Asp
    1100                1105


<210>  234
<211>  36
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  DR
```

<400> 234
gaacgacgtc actacacacc gagaggtgtc taaaac 36

<210> 235
<211> 927
<212> PRT
<213> Artificial Sequence

<220>
<223> Ga0187910_10040531

<400> 235

Met Ala Lys Lys Lys Lys Ala Lys Gln Arg Arg Glu Glu Gln Glu Ala
1 5 10 15

Ala Arg Met Asn Lys Ile Gln Ser Ala Val Lys Ala Lys Ala Glu Thr
20 25 30

Ala Pro Ala Val Ser Ser Ala Phe Val Glu Lys Arg Lys Asp Lys Gln
35 40 45

Ser Lys Lys Thr Phe Ala Lys Ala Ser Gly Leu Lys Ser Thr Leu Ala
50 55 60

Val Asp Asn Ser Ala Val Met Thr Val Phe Gly Arg Gly Asn Glu Ala
65 70 75 80

Lys Leu Asp His Arg Ile Asn Ala Asp Leu Gln Ser Glu Ser Leu His
85 90 95

Pro Gln Ala Ala Leu Lys Asn Val His Ala Pro Asn Lys Gln Lys Ile
100 105 110

His Phe Ile Gly Arg Met Gln Asp Met Asn Leu Thr Ala Asp His Pro
115 120 125

Leu His Ser His Asp Gly Glu Arg Ala Val Gly Ala Asp Leu Leu Cys
130 135 140

Ala Lys Asp Lys Leu Glu Gln Leu Tyr Phe Gly Arg Thr Phe Asn Asp

145 150 155 160

Asn Ile His Ile Gln Leu Ile Tyr Gln Ile Leu Asp Ile Gln Lys Ile
165 170 175

Leu Ala Leu His Ala Asn Asn Ile Ile Phe Ala Leu Asp Asn Leu Leu
180 185 190

His Lys Lys Asn Asp Glu Leu Ser Asp Asp Phe Val Gly Met Gly Arg
195 200 205

Met Arg Ala Thr Ile Gly Tyr Asp Ala Phe Arg Asn Ser Thr Asn Gln
210 215 220

Lys Val Gln Glu Thr Tyr Arg Glu Phe Gln Glu Phe Val Arg Arg Lys
225 230 235 240

Glu Leu Leu Tyr Phe Gly Ser Ala Phe Tyr Asn Gly Asp Thr Arg Arg
245 250 255

Asp Glu Lys Val Ile Tyr His Ile Leu Ser Leu Ala Ala Ser Val Arg
260 265 270

Gln Phe Cys Phe His Asn Asp Tyr Thr Ser Asp Asp Gly Lys Gly Phe
275 280 285

Ile Lys Ala Asp Trp Met Tyr Arg Leu Glu Glu Ala Leu Pro Ala Glu
290 295 300

Tyr Lys Asp Thr Leu Asp Ala Leu Tyr Leu Glu Gly Val Glu Gly Leu
305 310 315 320

Asp Gln Ser Phe Leu Lys Asn Asn Thr Val Asn Ile Gln Ile Leu Cys
325 330 335

Ser Ile Phe Asn His Asp Asp Pro Asn Lys Ile Ala Glu Glu Tyr Tyr
340 345 350

Gly Phe Leu Met Thr Lys Glu Tyr Lys Asn Met Gly Phe Ser Ile Lys

```
        355                 360                 365

Lys Leu Arg Glu Cys Met Leu Glu Leu Pro Glu Leu Ser Gly Tyr Lys
    370                 375                 380

Glu Asp Gln Tyr Asn Ser Val Arg Ser Lys Leu Tyr Lys Leu Phe Asp
385                 390                 395                 400

Phe Ile Ile Ala His Tyr Phe Arg Lys His Pro Glu Lys Gly Glu Glu
                405                 410                 415

Met Val Asp Cys Leu Arg Leu Cys Met Thr Glu Asp Glu Lys Asp Ser
            420                 425                 430

His Tyr Glu Gly Thr Ala Lys Lys Leu Val Arg Glu Leu Ala Tyr Asp
        435                 440                 445

Met Gln Glu Ala Ala Glu Gln Ala Asn Gly Ser Asn Ile Thr Gln Met
    450                 455                 460

Gln Lys Asn Glu Gln Gln Gly Lys Thr Lys Gly Met Phe Ala Ile Arg
465                 470                 475                 480

Asp Glu Ile Arg Val Ser Arg Lys Pro Val Ser Tyr Phe Ser Lys Val
                485                 490                 495

Ile Tyr Val Met Thr Leu Leu Leu Asp Gly Lys Glu Ile Asn Asp Leu
            500                 505                 510

Leu Thr Thr Leu Ile Asn Lys Phe Glu Asn Ile Val Ser Phe Glu Asp
        515                 520                 525

Val Leu Arg Gln Leu Asn Val Asp Cys Thr Phe Lys Pro Glu Phe Ala
    530                 535                 540

Phe Phe Gly Tyr Asp Arg Cys Arg Asn Ile Ser Gly Glu Leu Arg Leu
545                 550                 555                 560

Ile Asn Ser Phe Ala Arg Met Gln Lys Pro Ser Ala Lys Ala Lys His
```

565 570 575

Val Met Tyr Arg Asp Ala Leu Arg Ile Leu Gly Leu Asp Asn Gly Met
580 585 590

Ser Glu Glu Ala Leu Asp Gln Glu Val Arg Arg Ile Leu Gln Ile Gly
595 600 605

Ala Asp Gly Lys Pro Ile Lys Asn Ala Asn Lys Gly Phe Arg Asn Phe
610 615 620

Ile Ala Ser Asn Val Ile Glu Ser Ser Arg Phe Arg Tyr Leu Val Arg
625 630 635 640

Tyr Asn Asn Pro His Lys Thr Arg Met Ile Ala Gln Asn Glu Ala Ile
645 650 655

Val Arg Phe Val Leu Ser Glu Ile Pro Asp Glu Gln Ile Arg Arg Tyr
660 665 670

Tyr Asp Val Cys Arg Asp Pro Lys Leu Pro Arg Ser Ser Ser Arg Glu
675 680 685

Ala Gln Val Asp Ile Leu Thr Gly Ile Ile Thr Asp Val Asn Tyr Arg
690 695 700

Ile Phe Glu Asp Val Pro Gln Ser Lys Lys Ile Asn Lys Asp Arg Pro
705 710 715 720

Asp Ala Asn Asp Arg Met Thr Leu Lys Lys Gln Arg Tyr Gln Ala Ile
725 730 735

Val Ser Leu Tyr Leu Thr Val Met Tyr Leu Val Thr Lys Asn Leu Val
740 745 750

Tyr Val Asn Ser Arg Tyr Val Met Ala Phe His Ala Leu Glu Arg Asp
755 760 765

Ala Tyr Leu Tyr Gly Ile Thr Asn Ile Lys Gly Asp Tyr Arg Lys Leu

```
    770                 775                 780

Thr Asp Asn Leu Leu Ala Asp Glu Asn Tyr Lys Lys Phe Gly His Phe
785                 790                 795                 800

Lys Asn Lys Lys Trp Arg Gly Ile Ala Glu Gln Asn Leu Arg Asn Ser
                805                 810                 815

Asp Val Pro Val Ile Lys Ser Phe Arg Asn Met Ala Ala His Ile Ser
            820                 825                 830

Val Ile Arg Asn Ile Asp Leu Tyr Ile Gly Asp Ile Gln Lys Val Asp
        835                 840                 845

Ser Tyr Phe Ala Leu Tyr His Phe Leu Met Gln Lys Leu Ile Gln Arg
    850                 855                 860

Val Val Pro Glu Asn Thr Lys Gly Leu Ser Asp Gln Thr Lys Lys Tyr
865                 870                 875                 880

Tyr Asp Ala Leu Glu Gln Tyr Asn Thr Tyr Cys Lys Asp Phe Val Lys
                885                 890                 895

Ala Tyr Cys Thr Pro Phe Ala Tyr Val Thr Pro Arg Tyr Lys Asn Leu
            900                 905                 910

Thr Ile Asp Gly Leu Phe Asp Arg Asn Arg Pro Gly Glu Asp Lys
        915                 920                 925


<210>  236
<211>  36
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  DR

<400>  236
caactactac cctgccaaat ggcagggttc agaaac                                  36


<210>  237
```

<211> 900
<212> PRT
<213> Artificial Sequence

<220>
<223> Ga0187911_10069260

<400> 237

```
Met Ser Thr Lys Lys Arg Phe Arg Tyr Ser Val Ala Ala Lys Ala Ala
1               5                   10                  15


Gly Leu Lys Ser Ser Leu Ala Val Asp Thr Asp Arg Thr Val Met Thr
            20                  25                  30


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


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


Ile Asn Asp Lys Lys Tyr Ala Leu Thr Gly His His Ala Gly Val His
65                  70                  75                  80


Ala Leu Val Asp Gln Pro Gln Asn Arg Ser Asp Ala Val His Ile Arg
                85                  90                  95


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


His Val Gln Ile Ala Tyr Asn Ile Leu Asp Ile Thr Lys Ile Leu Thr
        115                 120                 125


Val Tyr Ala Asn Asn Val Val Tyr Ala Leu Asn Asn Leu Val His Ala
    130                 135                 140


Asp Asp Asp Thr Gln Ala Asp Glu Leu Asp Ser Leu Gly Asn Phe Ser
145                 150                 155                 160


Ala Gly Thr Ser Tyr Ala Lys Ser Lys Ser Lys Ser Lys Ser Lys Gln
                165                 170                 175
```

```
Gln Asp Phe Val Glu Leu Phe Ile Lys Lys Lys Glu Ile His Gly Tyr
            180                 185                 190

Phe Gly Asp Thr Phe Ala Phe Leu Asp Lys Arg Ile Ala Asp Ala Asp
        195                 200                 205

Lys Glu Lys Gln Val Tyr Ala Met Leu Ala Cys Leu Gly Ser Leu Arg
    210                 215                 220

Gln Ala Cys Ser His Tyr Arg Ile Arg Tyr Ser Val Asn Gly Lys Asn
225                 230                 235                 240

Val Asp Ala Asp Ala Asp Thr Trp Leu Phe Ser Ser Ala Gln Leu Asp
                245                 250                 255

Gln Thr Asp Pro Leu Phe Ser Glu Met Leu Asn Arg Ile Tyr Ser His
            260                 265                 270

Lys Ile Lys Thr Val Asn Gln Asn Phe Phe Glu Asn Asn Arg Lys Ala
        275                 280                 285

Asn Phe Pro Ile Leu Lys Lys Met Tyr Pro Glu Thr Thr Leu Lys Val
    290                 295                 300

Leu Met Asn Glu Tyr Tyr Asp Phe Ser Ile Arg Lys Gly Tyr Lys Asn
305                 310                 315                 320

Phe Gly Phe Ser Ile Lys Ser Leu Arg Glu Ala Leu Leu Ser Pro Gln
                325                 330                 335

Tyr Glu Ser Leu Ile Gly Val Gln Ile Lys Asp Asn Lys Glu Tyr Asp
            340                 345                 350

Thr Val Arg Ser Lys Leu Tyr Gln Leu Phe Asp Phe Ala Leu Thr Arg
        355                 360                 365

Tyr Phe Asn Gln His Pro Asp Met Val Asp Ala Phe Val Val Glu Leu
    370                 375                 380
```

```
Arg Ser Leu Ala Lys Asp Glu Asp Ala Lys Asn Ala Val Tyr Glu Lys
385                 390                 395                 400

Tyr Ala Lys Ala Val Trp Asn Asp Val Lys Gln Pro Ile Ala Val Met
                405                 410                 415

Leu Ser Tyr Met Asn Gly Ser Ala Ile Lys Asn Ile Lys Ala Phe Glu
            420                 425                 430

Leu Lys Pro Asp Gln Lys Glu Leu Asn Gly Ile Met Asn Ser Asn Ala
        435                 440                 445

Leu Asp Val Pro His Phe Cys Lys Leu Val Tyr Phe Leu Thr Arg Phe
    450                 455                 460

Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr Thr Leu Val Asn Lys
465                 470                 475                 480

Phe Asp Asn Ile His Ser Phe Asn Gln Val Leu Thr Ala Leu Gly Leu
                485                 490                 495

Ser Ala Ser Tyr Glu Ala Asp Tyr Lys Ile Phe Glu Asp Ser Gly Arg
            500                 505                 510

Val Val Glu Tyr Leu Arg Glu Ile Asn Ser Phe Ala Arg Met Thr Val
        515                 520                 525

Asp Met Glu Lys Ile Lys Arg Ser Ala Tyr Lys Lys Ala Leu Leu Ile
    530                 535                 540

Leu Gly Ser Ser Lys Tyr Ser Asp Glu Asp Leu Asp Ala Arg Val Asp
545                 550                 555                 560

Glu Met Leu Gly Val Asp Tyr Asn Gln Asn Gly Glu Lys Ile Lys Val
                565                 570                 575

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

```
Ser Ser Arg Phe His Tyr Leu Ile Arg Tyr Cys His Pro Arg Lys Ile
        595                 600                 605

Arg Asn Leu Ala Gly Asn Ala Ala Leu Ile Glu Tyr Gln Leu Arg Arg
    610                 615                 620

Leu Pro Glu Leu Gln Ile Leu Arg Tyr Tyr Glu Ala Cys Thr Glu Pro
625                 630                 635                 640

Ile Lys Arg Thr Ala Arg Thr Met Asp Glu Lys Ile Gly Thr Leu Ile
                645                 650                 655

Asp Leu Ile Val Lys Met Asp Phe Ser Gln Phe Glu Asp Val Gln Gln
            660                 665                 670

Asn Asp Arg Val Arg Val Phe Ser Asp Ala Glu Lys Lys Glu Lys Ile
        675                 680                 685

Arg Lys Met Arg Glu Lys Gln Arg Tyr Gln Ser Ile Ile Ser Leu Tyr
    690                 695                 700

Leu Thr Met Leu Tyr Leu Ile Val Lys Asn Leu Val Asn Ile Asn Ala
705                 710                 715                 720

Arg Tyr Val Met Ala Phe Gln Ala Trp Glu Arg Asp Asn Tyr Leu Leu
                725                 730                 735

Leu Gln Leu Ser Gly Lys Glu Ala Glu Ala Glu Tyr Leu Asn Leu Thr
            740                 745                 750

Arg His Phe Ile Glu Pro Leu Asp Gly Ala Lys Pro Tyr Leu Lys Lys
        755                 760                 765

Arg Pro Val Glu Tyr Leu Lys Lys Asp Met Ala Met Val Gly Asn Ser
    770                 775                 780

Ser Ile Arg His Phe Arg Asn Ala Thr Val His Leu Asn Val Ile Met
785                 790                 795                 800
```

Glu Ala His Arg Tyr Thr Lys Asp Ile Lys Tyr Ile Gly Ser Tyr Tyr
805 810 815

Ala Leu Tyr His Tyr Ile Leu Gln Arg His Leu Leu Asp Lys Ile Glu
820 825 830

Glu Asp Ser Tyr Ala Glu Lys Thr Val Ser Glu Lys Leu Trp Glu Ser
835 840 845

Gln Ile Ser Gln Tyr Gly Thr Tyr Ser Lys Asp Phe Val Lys Ala Leu
850 855 860

Cys Cys Pro Phe Gly Tyr Asn Leu Pro Arg Phe Lys Asn Leu Ser Ile
865 870 875 880

Glu Gln Leu Phe Asp Arg Asn Glu Ser Lys Glu Ile Thr Asp Ala Thr
885 890 895

Ala Pro Arg Gln
900

<210> 238
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 238
gaactacagc cctgtgaaat aacggggttc taaaac 36

<210> 239
<211> 991
<212> PRT
<213> Artificial Sequence

<220>
<223> MH0288_GL0082219

<400> 239

Met Lys Lys Met Thr Ala Lys Glu Arg Arg Glu Gln Glu Lys Gln Lys

1 5 10 15

Arg Ala Glu Val Ile Arg Gly Lys Gln Glu Ala Asn Arg Lys Lys Ala
20 25 30

Gln Ala Ala Ala Thr Gln Asp Ala Thr Ala Pro Val His His Lys Asn
35 40 45

Ala Lys Ser Leu Ala Lys Ala Ala Gly Leu Lys Ser Thr Phe Thr Leu
50 55 60

Pro Thr Gly Glu Leu Leu Met Thr Ser Phe Gly Arg Gly Ser Glu Ala
65 70 75 80

Leu Ile Gln Lys Arg Ile Met Lys Asp Val Ile Ser Asp Val Gln Glu
85 90 95

Gly Asp Pro Ala Phe Ser Val Ile Pro Met Pro Lys Gln Phe Glu Ile
100 105 110

Gln Gly Arg Ala Ile Asp Val Gln Lys Glu Gly Ser Ser Val Pro Glu
115 120 125

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

Leu Ile Gly Ala Lys Pro Ala Leu Glu Lys Arg Tyr Phe Gly Lys Thr
145 150 155 160

Phe Thr Asp Asn Ile His Ile Gln Leu Ile Tyr Arg Ile Leu Asp Ile
165 170 175

Lys Lys Val Leu Ala Ile His Ala Asn Asn Ile Val Tyr Thr Leu Asp
180 185 190

Asn Leu Ser Arg Arg Glu Gly Thr Glu Gly Thr Asp Phe Ile Gly Ile
195 200 205

Gly Gly Leu Thr Leu Lys Asn Ser Tyr Ala Ala Tyr Cys Glu Pro Arg

```
    210                 215                 220

Thr Phe Phe Asp Glu Gly Lys Glu Lys Asn Ile Ile Glu Asn Ile Glu
225                 230                 235                 240

Lys Ser His Gly Glu Phe Leu Glu Tyr Met Lys Asn Pro Arg Leu Ala
                245                 250                 255

Tyr Tyr Gly Glu Ala Phe Tyr Arg Arg Leu Ser Asn Asn Glu Arg Ser
            260                 265                 270

Lys Leu Gly Met Glu Arg Tyr Glu Arg Met Ser Asp Ala Gln Arg Asp
        275                 280                 285

Ala Phe Asp Arg Lys Tyr Pro Ile Asp Phe Cys Asp Gly Glu Leu His
    290                 295                 300

Phe Arg Glu Ile Glu Lys Leu Ala Pro Glu Tyr Val Arg Lys Ser Glu
305                 310                 315                 320

Lys Glu Ile Tyr Tyr Ile Leu Ala Met Leu Ser Glu Leu Arg Gln Cys
                325                 330                 335

Val Phe His Asp Ser Ala Ala Ser Arg Ser Trp Leu Tyr Asn Leu Glu
            340                 345                 350

Lys Val Leu Pro Ala Glu Ala Met Gln Val Leu Asp Gly Val Tyr Ser
        355                 360                 365

Gln Lys Val Ala Ala Leu Asn Asn Phe Ser Lys Thr Ser Ser Lys Ser
    370                 375                 380

Asn Phe Arg Leu Leu Phe Asn Ala Leu His Ala Asp Thr Asp Ala Gln
385                 390                 395                 400

Lys Lys Gln Ile Ala Glu Asp Phe Tyr Lys Phe Ser Val Arg Lys Thr
                405                 410                 415

Tyr Lys Asn Ile Gly Phe Ser Ile Lys Leu Leu Arg Glu Ser Met Leu
```

```
            420                 425                 430

Asp Thr Phe Ala Pro Glu Tyr Lys Glu Lys Ser Tyr Asp Ser Val Arg
        435                 440                 445

Gly Lys Leu Tyr Gln Leu Val Asp Phe Leu Ile Trp Asn Ile Tyr Asp
    450                 455                 460

Lys Asn Glu Lys Met Ala Asp Ala Leu Val Ala Gln Leu Arg Ala Glu
465                 470                 475                 480

Val Lys Arg Asp Arg Asp Glu Lys Ser Pro Val Tyr Asp Arg Glu Ala
                485                 490                 495

Ala Arg Leu Trp Thr Met Ile Arg Glu Ala Val Ala Gln Thr Ala Lys
            500                 505                 510

Thr Leu Arg Ala Val Thr Leu Ser Asn Asn Thr Pro Gly Ser Ile Lys
        515                 520                 525

Ile Lys Glu Leu Ala Leu Ser Lys Tyr Asp Arg Asp Leu Val Glu Asn
    530                 535                 540

Ala Ile Asp Ser Ala Val Asp Glu Arg Gly Arg Arg Leu Lys Ile Thr
545                 550                 555                 560

Asp Glu Ala Ser Tyr Phe Ser Lys Leu Ile Tyr Leu Leu Thr Leu Phe
                565                 570                 575

Leu Asp Gly Lys Glu Ile Asn Asp Leu Ile Thr Thr Leu Val Asn Ser
            580                 585                 590

Phe Asp Glu Ile Asp Ser Phe Cys Cys Val Leu Gln Lys Ala Lys Leu
        595                 600                 605

Asp Arg Gly Phe Thr Glu Gly Tyr Arg Phe Phe Ala His Ser Ala Glu
    610                 615                 620

Ile Ser Gln Glu Leu Arg Ala Ile Asn Ser Phe Ala Arg Met Thr Thr
```

625 630 635 640

Glu Thr Val Gln Gly Lys Pro Cys Met Phe Glu Asp Ala Ala Asp Leu
645 650 655

Leu Gly Met Asp Thr Asp Leu Glu Glu Val Arg Glu Ala Leu Glu Ala
660 665 670

Phe Tyr Asn Lys Glu Lys Met Ser Leu Lys Trp Thr Lys Asp Gly Gln
675 680 685

Glu Ala Asp Thr Gly Phe Arg Asn Phe Ile Ile Ser Asn Val Leu Lys
690 695 700

Ser Ser Arg Phe Arg Tyr Val Val Arg Tyr Gly Asn Pro Lys Lys Leu
705 710 715 720

Arg Glu Val Met Gln Asn Glu Ala Val Val Lys Phe Val Leu Asn Asp
725 730 735

Leu Pro Glu Glu Gln Ile Asn Arg Tyr Ser Glu Asn Ile Leu Gly Leu
740 745 750

Asp Ala Ala Ala Lys Glu Lys Lys Val Ser Glu Leu Ala Arg Phe Leu
755 760 765

Arg Ala Leu Ser Leu Glu Asp Phe Arg Asp Val Glu Gln Ser Ala Asn
770 775 780

Ala Ala Arg Asn Ala Arg Lys Ala Arg Tyr Gln Gly Ala Met Ser Leu
785 790 795 800

Tyr Leu Asn Val Ile Tyr Gln Ile Ile Lys Asn Met Val Tyr Ile Asn
805 810 815

Ser Arg Tyr Thr Leu Ala Phe Ala Lys Tyr Glu His Asp Ala Ser Leu
820 825 830

Leu Leu Asp Asn Trp Val Tyr Asp Pro Tyr Lys Thr Val Phe Ser Asp

835 840 845

Leu Thr Glu Asp Ser Leu Gln Lys Gly Lys Leu Asn Arg Arg Ala Ala
850 855 860

Lys Tyr Leu Gln Thr Asn Leu Glu His Ser Asp Asn Thr Leu Ile Arg
865 870 875 880

Gln Tyr Arg Asn Ser Val Ala His Leu Asn Ala Val Arg Asn Met Asn
885 890 895

Leu Tyr Ile Gly Asp Leu Arg Glu Ile Thr Ser Tyr Phe Gly Ile Tyr
900 905 910

His Tyr Ile Met Gln Arg Asp Leu Ile Thr Arg Cys Gln Lys Val Lys
915 920 925

Met Lys Asn Pro Val Val Val Gly Tyr Phe Asp Lys Val Arg Thr Tyr
930 935 940

Arg Thr Cys Cys Lys Asp Leu Val Lys Ala Leu Asn Ile Thr Phe Gly
945 950 955 960

Tyr Asn Leu Pro Arg Tyr Lys Asn Leu Ser Ile Glu Gly Leu Phe Asp
965 970 975

Lys Asn Arg Pro Gly Asp Lys Lys Glu Ser Ala Leu Ile Ala Glu
980 985 990

<210> 240
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 240
ccactacacc cctgtcgaac aacgggggtc taaaac 36

<210> 241

<211> 925
<212> PRT
<213> Artificial Sequence

<220>
<223> O2.UC29-0_GL0096317

<400> 241

Met Glu Arg Gln Lys Arg Arg Lys Lys Ser Cys Ala Lys Met Ala Gly
1 5 10 15

Val Lys Ser Thr Phe Val Ile Asp Ser Asp Leu Leu Met Thr Ser Phe
20 25 30

Gly Glu Arg Asp Lys Ala Val Arg Glu Lys Tyr Ile Ala Gln Gly Asn
35 40 45

Val Lys Ser Leu Arg Lys Gln Glu Gln Ala Phe Asp Ala Gln Tyr Gly
50 55 60

Lys Glu Lys Ile Phe Val Thr Asn Ala Asp Asn Lys Leu His Val Ala
65 70 75 80

Ile Asn Asn Pro Leu Cys Gly Asn Lys Lys Gln Ser Pro Glu Ala Gln
85 90 95

Glu Tyr Ser Arg Arg Val Leu Glu His Lys Lys Ala Ile Asn Gln Lys
100 105 110

Tyr Glu Thr Ile Val Phe Gly Lys Pro Ser Asn Asp Asn Ile His Ile
115 120 125

Gln Val Val Ser Lys Ile Met Asp Ile Glu Lys Thr Phe Ser Phe Val
130 135 140

Met Gly Asn Ile Val Tyr Ala Ile Asn Asn Leu Asp Pro Glu Gln Asn
145 150 155 160

Ile Asp Asn Pro Ile Asp Tyr Met Ser Asn Lys Ala Arg Thr Asp Glu
165 170 175

```
Asn Leu Lys Arg Val Leu Arg Arg Cys Met Tyr Leu Phe Asp Leu Gly
            180                 185                 190


Phe Glu Tyr Ala Glu Pro Ser Pro Tyr Lys Lys Ser Glu Gln Gln Ile
        195                 200                 205


Glu Gln Glu Arg Glu Glu Arg Glu Lys Glu Asn Leu Glu Thr Leu Lys
    210                 215                 220


Lys Ala Leu Asp Val Ile Ser Val Met Arg Thr Asp Ser Phe His Gly
225                 230                 235                 240


Tyr Asp Met Cys Asn Ser Thr Ser Ala Ile Val Lys Phe Pro Ser Lys
                245                 250                 255


Lys Phe Asn Leu Asp Gln Lys Gly Lys Asn Leu Ile Tyr Arg Asn Ala
            260                 265                 270


Phe Lys Tyr Lys Thr Asp Thr Ile Lys Lys Gln Tyr Ala Gln Asn Ser
        275                 280                 285


Lys Val Asn Ile Asn Phe Leu Ser Tyr Val Leu Tyr Gly Asn Gly Cys
    290                 295                 300


Thr Glu Glu Gln Lys Lys Glu Leu Ser Arg Glu Tyr Phe Asn Phe Ala
305                 310                 315                 320


Ile Leu Lys Gln Tyr Lys Asn Leu Gly Phe Ser Ile Thr Ser Leu Arg
                325                 330                 335


Glu Gln Leu Leu Ser Leu Tyr Asp Ile Asp Phe Glu Cys Lys Arg Asn
            340                 345                 350


Leu Arg Ser Arg Ala Asn Ser Phe Cys Asp Tyr Val Leu Tyr Tyr Arg
        355                 360                 365


Tyr Ala Val Gln Phe Pro Glu Arg Ile Val Asp Leu Val Asp Lys Leu
    370                 375                 380
```

```
Arg Cys Ser Ala Asn Glu Lys Glu Lys Thr Asn Val Tyr Ile Glu Glu
385                 390                 395                 400


Ala Asn Arg Ile Lys Glu Asp Leu Arg Ile Ala Phe Asn Arg Ile Asn
                405                 410                 415


Asn Phe Asp Ala Lys Ser Ile Lys Leu Tyr Ser Thr Lys Asn Asn Val
            420                 425                 430


Asp Asp Gln Pro Met Leu Pro Phe Glu Gln Val Ser Lys Asp Ile Lys
        435                 440                 445


Gly Gly Asp Ala Ala Arg Glu Lys Ala Asp Ala Phe Cys Asp Met Thr
    450                 455                 460


Tyr Tyr Leu Thr Gln Phe Leu Asp Gly Lys Glu Ile Asn Val Phe Leu
465                 470                 475                 480


Thr Thr Leu His Asn Leu Phe Glu Asn Ile Ala Ser Phe Leu Asp Val
                485                 490                 495


Met His Lys Gln Asn Ile Ser Tyr Asp Phe Thr Glu Gln Tyr Val Met
            500                 505                 510


Phe Tyr Asp Ser Ala Tyr Ile Val Lys Lys Leu Glu Thr Val Ile Ser
        515                 520                 525


Val Ser Arg Met Lys Lys Ala Leu Asp Phe Tyr Asn Gly Gln Ala Leu
    530                 535                 540


Arg Asp Ala Ile Arg Val Leu Gly Ile Ser Glu Arg His Val Asp Met
545                 550                 555                 560


Asp Met Lys Glu Tyr Val Asp Lys Tyr Leu Phe Asp Asp Cys Lys Asp
                565                 570                 575


Val Lys His Ala Asp His Asn Leu Arg Asn Phe Ile Val Thr Asn Val
            580                 585                 590
```

```
Ile Lys Ser Arg Lys Phe Asn Tyr Leu Ala Arg Tyr Ser Asp Met Ala
        595                 600                 605

Thr Val Ser Lys Leu Ser Lys Asn Lys Glu Val Val Lys Phe Val Leu
    610                 615                 620

Lys Arg Val Asp Pro Lys Gln Ile Glu Arg Tyr Tyr Arg Ser Val Cys
625                 630                 635                 640

Asp Val Arg Asn Ile Ser Ala Glu Asp Glu Ile Glu Lys Leu Val Asp
                645                 650                 655

Ile Ile Tyr Asn Phe Asn Ile Asp Glu Leu Glu Gln Val Asn Asn Ala
            660                 665                 670

Glu Ile Gly Thr Gly His Ser Gly Arg Lys Val Ala Ser Lys Arg Ser
        675                 680                 685

Val Gln Glu Arg Asn Gln Ala Arg Ala Cys Val Ser Leu Tyr Leu Asn
    690                 695                 700

Val Leu Tyr Gln Ile Val Lys Asn Leu Met Asn Val Asn Ser Arg Tyr
705                 710                 715                 720

Ser Ile Ala Phe Ala Met Thr Glu Arg Asp Met Ile Val Glu Gly Lys
                725                 730                 735

Gly Ile Glu Ala Ile Glu Thr Phe Lys Asp Asp Pro Asp Ala Gly Lys
            740                 745                 750

Lys Ser Lys Lys Gln Ile Ala Tyr Arg Thr Lys Tyr Ile Ala Leu Thr
        755                 760                 765

Glu Glu Tyr Leu Asp Lys Gly Tyr Ile Lys Cys Ala Thr Glu Arg Ile
    770                 775                 780

Lys Glu Asn Ile Lys Lys Tyr Gly Gly Asp Pro Lys Ser Ile Asp His
785                 790                 795                 800
```

Ile Pro Gly Glu Tyr Ile Lys Lys Asn Leu Glu Gly Tyr Cys Ala Gly
805 810 815

Ser Thr Asn Thr Phe Arg Asn Ala Val Ala His Leu Ala Leu Met Ala
820 825 830

Arg Val Ser Glu Tyr Ala Ser Asp Ile Lys Asn Val Asn Ser Tyr Phe
835 840 845

Ser Leu Tyr His Tyr Cys Met Gln Lys His Ile Leu Tyr Cys Ala Arg
850 855 860

Arg Phe Asn Lys Gln Gly Asn Glu Lys Met Ala Ala Ala Leu Asn Asp
865 870 875 880

Ala Glu Ile His Arg Ser Tyr Ser Lys Asp Ala Val Lys Phe Leu Cys
885 890 895

Leu Pro Phe Ala Tyr Asn Leu Pro Arg Tyr Lys Ala Leu Thr Val Glu
900 905 910

Gly Leu Phe Asp Trp Asn Glu Tyr Gly Gln Asp Asp Asp
915 920 925

<210> 242
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 242
gtactaagac cttactaaat ggtaaggttc caaaac 36

<210> 243
<211> 954
<212> PRT
<213> Artificial Sequence

<220>
<223> PIG-014_GL0226364

<400> 243

```
Met Thr Lys Pro Glu Lys Lys Gln Ile Lys Gln Ala Thr Leu Thr Lys
1               5                  10                  15


Arg Met Gly Leu Lys Ser Thr Leu Val Phe Asp Asn Lys Ile Thr Val
            20                  25                  30


Thr Ser Phe Ala Asn Ser Ser Asn Ser Glu Lys Ala Ser Ala His Ile
        35                  40                  45


Glu Lys Arg Thr Asp Phe Trp Gly Lys Thr Ile Met Gln Asp Glu Lys
    50                  55                  60


Met Phe Arg Thr Asp Val Glu Lys Gln Gln Ile Tyr Ile Glu Lys Glu
65                  70                  75                  80


Phe Pro Asp Glu Asn Asn Lys Ile Ile Thr Gln Thr Ala Met Phe Glu
                85                  90                  95


Asn Pro Ala Leu Arg Asn Cys Gly Lys Asp Tyr Ile Gly Met Lys Asn
            100                 105                 110


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


His Ile Gln Ile Ala Tyr Asn Ile Leu Asp Ile Lys Lys Ile Ile Cys
    130                 135                 140


Pro Tyr Ile Asn Asn Val Val Asn Ile Phe Tyr Asn Leu Gly Arg Gly
145                 150                 155                 160


Lys Leu Lys Asp Asp Lys Asn Asp Leu Leu Gly Lys Gln Ile Gly Pro
                165                 170                 175


Lys Ser Glu Glu Met Val Glu Asp Leu Leu Lys Asn Ser Ser Ala Tyr
            180                 185                 190


Phe Val Tyr Phe Gly Asp Val Phe Lys Gln Val Lys Leu Ser Lys Glu
```

195 200 205

Gln Gln Glu Lys Ser Glu Glu Glu Lys Arg Arg Ile Trp Gln Glu Leu
210 215 220

Trp Asn Lys Asn Phe Asp His Asn Lys Asn Ile Val Asn Ser Leu Ile
225 230 235 240

Lys Phe Arg Gln Met Cys Phe His Gln Asn Tyr Glu Asn Ser Asp Lys
245 250 255

Glu Val Ser Ser Thr Ala Leu Phe Asn Ile Glu Gln Phe Leu Asn Lys
260 265 270

Glu Glu Lys Ala Ser Leu Asp Glu Ile Tyr Thr Gly Ala Ile Asp Lys
275 280 285

Val Asn Cys Asp Phe Asn Lys Asn Ser Ser Asn Asn Leu Tyr Ile Leu
290 295 300

Glu Gln Ile Tyr Cys Asn Asp Lys Asn Leu Gly Gln Arg Tyr Tyr Asp
305 310 315 320

Phe Val Val Lys Lys Ser Gln Cys Asn Leu Gly Leu Asp Thr Lys Leu
325 330 335

Ile Arg Glu Leu Leu Ile Arg Lys Leu Ser Asp Glu Gly Leu Asp Ile
340 345 350

Asn Asp Ser Arg Tyr Asp Thr Tyr Arg Asn Lys Ile Tyr Thr Val Phe
355 360 365

Gly Phe Ile Leu Tyr Tyr Glu Leu Thr Lys Asn Glu Thr Phe Cys Gln
370 375 380

Glu Ile Val Ala Lys Leu Arg Arg Phe Gly Gly Asp Glu Ile Glu Lys
385 390 395 400

Asn Lys Ile Tyr Asn Glu Tyr Phe Leu Arg Lys Glu Ala Val Tyr Asn

405 410 415

Gln Ile Ala Glu Lys Phe Lys Gln Ser Val Cys Trp Phe Lys Lys Glu
420 425 430

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

Ser Val Gln Lys Ile Glu Asp Gln Asn Ala Ser Val Phe Ala Lys Met
450 455 460

Val Tyr Leu Ile Ser Arg Phe Leu Asp Gly Lys Glu Ile Asn Glu Leu
465 470 475 480

Cys Thr Lys Leu Leu Asn Lys Phe Lys Asn Ile Asn Glu Leu Val Glu
485 490 495

Thr Ala Lys Arg Cys Gly Glu Thr Ile Glu Phe Cys Pro Glu Tyr Lys
500 505 510

Met Phe Lys Asn Ser Ala Lys Leu Ala Glu Glu Met Arg Phe Val Gln
515 520 525

Ser Val Ser Lys Met Lys Lys Leu Thr Lys Lys Glu Asp Lys Ser Ile
530 535 540

Lys Pro Gln Val Ile Asp Asp Ala Leu Glu Leu Leu Gly Leu Lys Ile
545 550 555 560

Glu Lys Tyr Lys Lys Asp Asp Gln Gly Asn Phe Val Leu Asp Asp Gln
565 570 575

Asn His Asn Val Trp Thr Asp Glu Tyr Lys Ser Tyr Arg Lys Glu Cys
580 585 590

Phe Thr Pro Lys Asn Ser Asp Ser Met Asn Asp Ala Glu Lys Ser Phe
595 600 605

Val Ala Lys Gly His Gln Lys Arg Asn Phe Ile Ile Ser Asn Val Val

610 615 620

Leu Asn Lys Arg Phe Leu Tyr Val Ala Lys Tyr Gly Ser Pro Thr Ala
625 630 635 640

Cys Asn Gln Met Met Lys Asn Gln Lys Ile Val Arg Met Ala Leu Glu
645 650 655

Ser Ile Ser Glu Asp Gln Leu Arg Lys Tyr Ala Arg Arg Val Tyr Asp
660 665 670

Asn Lys Glu His Phe Gln Asn Gln Thr Ala Asp Glu Ile Arg Glu Asn
675 680 685

Leu Ala Lys Lys Leu Lys Asn Phe Asn Ile Gly Ala Glu Leu Lys Thr
690 695 700

Val Gly Asn Leu Ser Glu Lys Asp Phe Arg Ser Ser Lys Asn Ala Leu
705 710 715 720

Lys Asp Lys Ser Lys Ser Leu Ile Gly Leu Tyr Leu Thr Val Ala Tyr
725 730 735

Leu Ile Thr Lys Asn Met Ile Lys Ile Asn Thr Arg Phe Ser Ile Ala
740 745 750

Phe Gly Ala Tyr Glu Arg Asp Met Glu Phe Leu Leu Gln Gln Lys Lys
755 760 765

Glu Lys Gly Lys Arg Phe Asp Trp Leu Ala Leu Thr Gln Lys Phe Leu
770 775 780

Ser Glu Asp Lys Asn Tyr Tyr Phe Glu His Asp Leu Arg Arg Lys Glu
785 790 795 800

Ile Asn Lys Ile Ser Asp Arg Asp Glu Arg Lys Lys Gln His Arg Glu
805 810 815

Asn Asp Lys Tyr Tyr Asp Ser Cys Met His Phe Ser Arg His Ser Phe

820 825 830

Glu Tyr Ile Ser Lys Asn Phe Ala Glu Val Gln Ser His Pro Asp Pro
835 840 845

Leu Lys Leu Phe Ile Lys Tyr Arg Asn Asn Val Met His Leu Asn Ile
850 855 860

Ile Arg Asp Leu Asn Asp Tyr Leu Pro Asp Phe Lys Arg Gly Lys Ala
865 870 875 880

Phe Glu Lys Thr Asp Thr Phe Tyr Asp Leu Tyr Cys Tyr Cys Leu Gln
885 890 895

Lys Met Leu Leu Ser Glu Val Asp Gly Phe Ser Ala Ile Asn Glu Asn
900 905 910

Ile Asp Lys Tyr Cys Thr Ala Ser Lys Asp Phe Met Asn Leu Leu Asn
915 920 925

Ile Ile Phe Ala Tyr Asn Val Pro Arg Tyr Lys Asn Leu Ser Asn Lys
930 935 940

Glu Leu Phe Tyr Arg Leu Asp Ser Glu Gln
945 950

<210> 244
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 244
gtacaaaagc cctccagttt ggacgggttc taagac 36

<210> 245
<211> 941
<212> PRT
<213> Artificial Sequence

<220>
<223> PIG-018_GL0023397

<400> 245

```
Met Ala Lys Lys Leu Ser Pro Arg Glu Leu Arg Glu Gln Arg Glu Arg
1               5                  10                  15


Glu Arg Gln Ala Lys Phe Ala Lys Gln Val Gln Glu Arg Glu Ala Glu
            20                  25                  30


Lys Ala Lys Arg Glu Lys Gln Val Ser Val Ser Asp Thr Val Gly Ala
        35                  40                  45


Ala Ala Ala Val Gln Asn Glu Asn Gly Lys Lys Ser Met Ala Lys Ala
    50                  55                  60


Ala Gly Val Lys Ser Val Phe Ala Val Gly Asp Thr Val Tyr Met Thr
65                  70                  75                  80


Ser Phe Gly Arg Gly Asn Asp Ala Val Leu Glu Lys Lys Ile Val Asp
                85                  90                  95


Glu Ser His Glu Ser Leu Asn Lys Ala Asn Pro Ala Tyr Glu Ile Thr
            100                 105                 110


Asp Ile Ser Ser Val Asn Tyr Ser Val Glu Gly His Arg Gly Gln Ser
        115                 120                 125


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


Val Pro Thr Asp Met Leu Cys Leu Lys Pro Thr Leu Glu Arg Val Tyr
145                 150                 155                 160


Phe Gly Lys Thr Phe Asp Asp Asn Leu His Ile Gln Leu Ile Tyr Asn
                165                 170                 175


Ile Leu Asp Ile Glu Lys Ile Leu Ala Ile Tyr Ser Thr Asn Ala Val
            180                 185                 190
```

```
Tyr Ala Leu Asn Asn Met Ile Ala Asp Glu Asp Asp Glu Ser Ser Asp
        195                 200                 205

Leu Phe Gln Asn Met Thr Thr Asp Tyr Thr Tyr Asp Lys Phe Arg Gly
    210                 215                 220

Arg Ser Glu Arg Ile Val Lys Ser Phe Asn Asp Phe Val Ser Asn Pro
225                 230                 235                 240

Arg Leu Ala Tyr Tyr Ser Asp Ala Phe Tyr Glu Lys Lys Gly Lys Gly
                245                 250                 255

Ser Lys Arg Arg Ser Glu Lys Glu Ile Tyr Ser Met Leu Ala Met Leu
            260                 265                 270

Gly Lys Leu Arg His Trp Cys Val His Ser Glu Ser Gly Glu Ala Glu
        275                 280                 285

Phe Trp Leu Tyr Lys Leu Asp Ser Leu Asn Gly Glu Phe Ile Asp Val
    290                 295                 300

Leu Asn Thr Leu Tyr Asn Arg Ser Val Asn Glu Ile Asn Arg Gly Phe
305                 310                 315                 320

Val Asp Thr Asn Lys Val Asn Ile Gln Ile Leu Gln Ala Val Cys Lys
                325                 330                 335

Asn Met Asp Leu Ile Ser Leu Ile Lys Ala Tyr Tyr Glu Phe Leu Ile
            340                 345                 350

Thr Lys Lys Tyr Lys Asn Met Gly Phe Ser Ile Lys Lys Leu Arg Glu
        355                 360                 365

Met Met Leu Ala Glu Thr Ala Tyr Thr Ser Asn Pro Asp Asp Arg Leu
    370                 375                 380

Tyr Asp Lys Glu Ile Asn Ser Val Arg Asn Lys Leu Tyr Gln Met Thr
385                 390                 395                 400
```

```
Asp Phe Ile Ile Tyr Leu Glu Tyr Arg Asp Asn Lys Ser Ile Ile Asp
                405                 410                 415

Thr Leu Val Ser Lys Leu Arg Ser Ser Val Thr Asp Glu Ala Lys Glu
            420                 425                 430

Ala Val Tyr Arg Glu Glu Ala Ala Asn Leu Arg Gly Arg Leu Leu Asp
        435                 440                 445

Ala Val Ala Leu Ile Ala Lys Ser Leu Asn Gly Val Asn Ile Lys Lys
    450                 455                 460

Leu Gln Gln Asn Thr Leu Glu Ile Pro Asp Ser Lys Leu Glu Asn Cys
465                 470                 475                 480

Phe Ile Lys Ser Ala Glu Lys Val Ser Ala Phe Thr Lys Leu Asn Tyr
                485                 490                 495

Leu Leu Thr Arg Phe Ile Asn Gly Lys Glu Ile Asn Asp Leu Leu Thr
            500                 505                 510

Thr Leu Ile Asn Lys Phe Asp Asn Ile Arg Ser Phe Leu Glu Thr Met
        515                 520                 525

Asp Glu Leu Gly Leu Glu Arg Thr Phe Thr Glu Asn Tyr Arg Phe Phe
    530                 535                 540

Glu Asp Ser Gln Lys Tyr Arg Asp Glu Leu Ile Glu Leu Asn Ser Phe
545                 550                 555                 560

Ala Lys Ser Cys Ser Phe Asp Ile Ser Ala Lys Arg Ile Met Tyr Arg
                565                 570                 575

Asp Ala Leu Asp Ile Leu Gly Ile Gln Ala Glu Leu Ser Glu Asp Glu
            580                 585                 590

Ile Gly Arg Met Ile Asp Asn Ile Leu Gln Leu Asp Ala Asn Gly Glu
        595                 600                 605
```

```
Arg Ile Ser Lys Asn Gln Arg Glu Asn Gly Leu Arg Asn Phe Ile Ala
    610                 615                 620

Ser Asn Val Ile Asp Ser Asn Arg Phe Lys Tyr Leu Val Arg Tyr Gly
625                 630                 635                 640

Asn Pro Lys Lys Ile Arg Glu Thr Ala Lys Cys Pro Pro Ala Val Lys
                645                 650                 655

Phe Val Leu Asp Gly Ile Pro Asp Ser Gln Ile Glu Arg Tyr Tyr Asn
            660                 665                 670

Ser Cys Phe Pro Glu Asp Thr Ala Ile Cys Ser Ala Asp Lys Lys Arg
        675                 680                 685

Arg Glu Leu Ala Glu Met Ile Ser Asn Ile Ser Phe Asp Trp Phe Lys
    690                 695                 700

Asp Ala Gly Asn Val Gln Arg Val Asn Ala Glu Asp Lys Ser Arg Asn
705                 710                 715                 720

Ala Glu Ile Lys Arg Lys Asn Gln Ala Ile Ile Arg Leu Tyr Leu Thr
                725                 730                 735

Val Met Tyr Leu Met Leu Lys Asn Leu Val Asn Val Asn Ser Arg Tyr
            740                 745                 750

Val Ile Gly Phe His Cys Val Glu Arg Asp Ala Lys Leu Tyr Lys Leu
        755                 760                 765

Ser Gly Leu Thr Ile Glu Gly Asn Thr Asp Lys Asn Arg Thr Leu Leu
    770                 775                 780

Thr Ala Thr Ile Leu Gly Ile Thr Ser Ala Asp Arg Ala Pro Ser Gly
785                 790                 795                 800

Glu Ile Val Ala Asp Cys Asp Leu Gly Arg Ala Glu Thr Ala Ala Asn
                805                 810                 815
```

Arg His Leu Arg Lys Lys Arg Trp Tyr Val Gln Ile Phe Asn Asn Leu
820 825 830

Lys Gln Ala Asp Arg Asn Val Ile Asn Glu Phe Arg Asn Ser Val Cys
835 840 845

His Leu Asn Ala Ile Arg Asn Ile Asn Thr Asn Ile Glu Gly Ile Gly
850 855 860

Phe Val Lys Asn Tyr Phe Ala Leu Tyr His Tyr Leu Ile Gln Arg His
865 870 875 880

Ile Ser Ser Lys Arg Ile Ser Phe Asn Ser Phe Ser Ala Glu Tyr Ile
885 890 895

Ser Asn Leu Glu Lys Tyr His Thr Pro Ser Glu Tyr Phe Ile Lys Ala
900 905 910

Tyr Cys Thr Pro Phe Ala Tyr Asn Leu Val Arg Tyr Lys Asn Leu Thr
915 920 925

Ile Asp Gly Leu Phe Asp Arg Asn Ala Pro Ser Glu Lys
930 935 940

<210> 246
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 246
gaactacacc cctgttaaat tgcaggggtc taaaac 36

<210> 247
<211> 948
<212> PRT
<213> Artificial Sequence

<220>
<223> PIG-025_GL0099734

<400> 247

```
Met Ala Lys Lys Met Asn Ala Lys Glu Gln Arg Ala Ala Ile Gln Ala
1               5                  10                  15

Glu Lys Asn Lys Lys Ile Gln Met Ala Asn Ala Ala Arg Glu Lys Gln
            20                  25                  30

Ala Glu Glu Ala Arg Leu Lys Ala Glu Ala Glu Arg Ile Glu Lys Glu
        35                  40                  45

Glu Asn Lys Lys Arg Glu Ala Glu Leu Lys Arg Phe Lys Leu Asp Gln
    50                  55                  60

Glu Thr Arg Gly Lys Leu Pro Lys Asn Glu Arg Lys Ser Leu Ala Lys
65                  70                  75                  80

Ala Ala Gly Leu Lys Ser Thr Phe Ala Val Gly Thr Asp Leu Tyr Met
                85                  90                  95

Thr Ser Phe Gly Lys Gly Asn Asp Ala Ile Val Glu Lys Lys Ile Ser
            100                 105                 110

Asn Lys Gln Val Thr Gln Ile Asn Thr Glu Lys Glu Asn Phe Thr Val
        115                 120                 125

Asp Thr Asp Gly Ile Thr Asp Thr Ile Val Pro Val Gln Ser Lys Arg
    130                 135                 140

Ile Ala Cys Leu Asn Thr Gln Ala Asp Asn Pro Leu Tyr Arg Lys Asp
145                 150                 155                 160

Ser Glu Asn Gln Val Gln Pro Asp Lys Leu Leu Leu Lys Asp Thr Leu
                165                 170                 175

Glu Lys Val Tyr Phe Gly Lys Thr Phe Asn Asp Thr Leu His Ile Gln
            180                 185                 190

Leu Ile Tyr Asn Ile Leu Asp Ile Gln Lys Ile Leu Thr Val Tyr Ser
```

195 200 205

Val Asn Thr Val Tyr Thr Leu Asn Asn Leu Phe Gly Lys Glu Ile Ser
210 215 220

Glu Arg Gly Asp Phe Ile Gly Ala Leu Ser Phe Gln Gln Thr Tyr Asp
225 230 235 240

Asp Tyr Thr Ser Asp Ser Lys Asn Lys Glu Ile Phe Glu Asp Phe Tyr
245 250 255

Ser Lys Arg Ser Leu Gly Tyr Tyr Gly His Ile Phe Phe Lys Asp Lys
260 265 270

Lys Lys Arg Ser Lys Arg Glu Ile Tyr Asp Ile Leu Ala Leu Ile Ala
275 280 285

Thr Ile Arg Gln Trp Cys Ile His Ser Glu Glu Asp Lys Lys Thr Trp
290 295 300

Ile Tyr Asn Ala Glu Asn Val Leu Asp Ser Glu Tyr Leu Asp Ile Leu
305 310 315 320

Asp Asp Val Tyr Gly Gly Leu Val Glu Lys Ile Asn Lys Asn Phe Ile
325 330 335

Lys Asp Asn Lys Val Asn Leu Ile Leu Leu Ser Asp Ile Leu Asp Met
340 345 350

Lys Ser Lys Asp Asp Phe Glu Glu Leu Ile Arg Gln Tyr Tyr Arg Phe
355 360 365

Ile Ile Thr Lys Glu Gln Lys Leu Leu Gly Phe Ser Ile Thr Lys Ile
370 375 380

Arg Glu Ala Met Leu Glu Asp Thr Val Phe Lys Thr Asp Lys Lys Tyr
385 390 395 400

Asp Ser Ile Arg Ser Lys Leu Tyr Lys Leu Ile Asp Phe Met Met Phe

405 410 415

Tyr Lys Tyr Thr Thr Val Asp Ser Glu Lys Glu His Ala Asp Ser Ile
420 425 430

Val Glu Lys Leu Arg Ala Thr Leu Ser Gly Glu Ser Lys Glu Glu Ile
435 440 445

Tyr Lys Ala Glu Ala Asn Arg Ile Trp Asn Leu Tyr Glu Asp Ile Ile
450 455 460

Met Asn Lys Ile Glu Pro Ala Leu Asp Glu Gly Val Ile Ser Gly Leu
465 470 475 480

Lys Lys Asn Lys Ser Tyr Asp Asn Met Ser Ile Glu Gly Ile Val Ala
485 490 495

Glu Glu Thr Ala Asn Val Ser Tyr Phe Ser Lys Ile Ile Tyr Leu Leu
500 505 510

Ala Gln Phe Ile Asp Gly Lys Glu Val Asn Asp Leu Thr Thr Thr Leu
515 520 525

Ile Asn Lys Phe Asp Asn Ile Arg Ser Phe Ile Asp Thr Ala Asn Gln
530 535 540

Ile Gly Leu Glu Cys Asn Leu Leu Asp Glu Tyr Gln Phe Phe Asn Thr
545 550 555 560

Ser Asp Thr Val Lys Lys Glu Leu His Val Ile Lys Asn Leu Thr Gly
565 570 575

Met Gly Tyr Tyr Asp Val Ser Ile Lys His Gln Met Tyr Lys Asp Ala
580 585 590

Ile Asp Ile Leu Gly Leu His Asp Asn Leu Ser Asp Glu Glu Ile Asn
595 600 605

Gln Leu Ile Asp Thr Asn Ile Leu Phe Ile Gly Ala Asp Gly Lys Pro

610 615 620

Val Lys Met Ala Lys Gly Lys Arg Gly Phe Arg Asn Phe Ile Ile Ser
625 630 635 640

Asn Val Leu Lys Ser Ser Arg Phe Arg Tyr Leu Val Lys Tyr Cys Asn
645 650 655

Pro Lys Lys Ile Arg Lys Ile Ala Asn Asn Glu Lys Ile Val Lys Phe
660 665 670

Val Leu Gly Arg Ile Thr Glu Thr Gln Ile Glu Arg Tyr Tyr Tyr Ser
675 680 685

Cys Asn Pro Glu Leu Lys Arg Ala Ala Tyr Pro Gly Arg Glu Ala Ala
690 695 700

Ile Asp Asp Leu Thr Lys Ile Ile Ile Asn Met Lys Phe Asp Asp Phe
705 710 715 720

Lys Asn Val Lys Gln Ser Ala Asn Val Glu Asp Ser Ser Lys Glu Ala
725 730 735

Gln Glu Lys Met Lys Tyr Gln Thr Ile Ile Ser Leu Tyr Leu Thr Val
740 745 750

Cys Tyr His Leu Val Lys Asn Leu Val Asn Ile Asn Ala Arg Tyr Ala
755 760 765

Met Ala Phe His Ser Leu Glu Arg Asp Ala Arg Leu Tyr Gly Ile Tyr
770 775 780

Gly Ser Glu Lys Asp Tyr Val Glu Lys Leu Lys Lys Glu Gln Ala Val
785 790 795 800

Leu Val Glu Lys Leu Leu Ser Asp Gly Pro Glu Thr Ala Gly Asn Leu
805 810 815

His Leu Arg Ser Lys Lys Trp Asn Arg Leu Met Arg Val Asn Leu Glu

820 825 830

Asn Tyr Ser Ala Ala Ala Ser Gln Asn Phe Arg Asn Ala Val Ala His
835 840 845

Leu Asn Pro Ile Arg Asn Ala Asp Val Leu Ile Glu Gly Ile Glu Asn
850 855 860

Ile Thr Ser Tyr Phe Asp Ile Tyr His Tyr Ile Val Gln Arg Ser Val
865 870 875 880

Leu Asn Arg Thr Asn Lys Thr Pro Lys Met Lys Glu Tyr Glu Glu Leu
885 890 895

Ile Asn Lys Tyr His Thr Tyr Asn Lys Asp Phe Val Lys Ala Leu Cys
900 905 910

Ala Pro Phe Gly Tyr Asn Val Val Arg Phe Lys Ala Leu Thr Ile Tyr
915 920 925

Glu Met Phe Asp Arg Asn Tyr Pro Glu Glu Lys Arg Ala Glu Glu Ala
930 935 940

Arg Lys Ser Cys
945

<210> 248
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 248
taatatgccc tatatcaatt tataggggtt cagaac 36

<210> 249
<211> 894
<212> PRT
<213> Artificial Sequence

<220>
<223> PIG-028_GL0185479

<400> 249

```
Met Ser Glu Ser Asp Lys Ile Lys Ile Lys Ser Lys Lys Thr Arg Thr
1               5                  10                  15


Lys Arg Ala Gly Leu Lys Ser Ser Phe Val Ser Glu Gly Lys Gly Leu
            20                  25                  30


Met Thr Ala Phe Gly Lys Gly Asn Asn Ala Val Ile Glu Lys Arg Ile
        35                  40                  45


Ser Gly Asp Leu Ile Glu Asn Val Ser Asn Lys Pro Arg Phe Ser Val
    50                  55                  60


Glu Phe Gly Asn Lys Glu Tyr Ile Ile Asn Asp Lys Lys Leu Gly Ile
65                  70                  75                  80


Val Ser Asp Asp Pro Arg Glu Ser Gln Lys Pro Val Gly Glu Asp Gly
                85                  90                  95


Ile His Cys Lys Arg Ala Leu Glu Glu Arg Tyr Phe Gly Lys Thr Phe
            100                 105                 110


Asn Asp Asn Ile His Ile Gln Leu Ile Tyr Ser Ile Leu Asp Leu Glu
        115                 120                 125


Lys Ile Leu Ala Val His Ser Asn Asn Ile Val Tyr Thr Phe Asp Asn
    130                 135                 140


Leu Arg Arg Lys Glu Thr Asp Glu Tyr Asp Asp Phe Ile Gly Tyr Met
145                 150                 155                 160


Ser Thr Val Asn Ser Tyr Glu Glu Phe Ile Ser Pro Glu Lys Leu Leu
                165                 170                 175


Pro Gly Lys Glu Gln Ser Val Tyr Asp Asn Ile Asn Arg Ser Gln Glu
            180                 185                 190
```

```
Thr Phe Leu Glu Tyr Arg Lys Asn Pro Arg Leu Gly Tyr Phe Gly Lys
        195                 200                 205

Val Phe Tyr Glu Lys Ser Thr Asp Ala Lys Asp Lys Asn Lys Lys Lys
    210                 215                 220

Met Arg Leu Ile Ser Glu Lys Glu Met Tyr Tyr Ile Phe Ala Leu Leu
225                 230                 235                 240

Gly Gly Leu Arg Gln Phe Cys Thr His Asp Lys Asn Arg Ala Arg Asn
                245                 250                 255

Trp Leu Tyr Ser Leu Asp Thr Met Leu Asn Asp Asp Ala Lys Ala Val
            260                 265                 270

Leu Asp Lys Tyr Tyr Val Ser Thr Val Ser Ser Leu Asn Lys His Phe
        275                 280                 285

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

Asn Ile Thr Asp Ile Glu Leu Lys Lys Asn Ile Ala Lys Glu Tyr Tyr
305                 310                 315                 320

Glu Phe Val Cys Arg Lys Asn Phe Lys Asn Ile Gly Phe Ser Ile Lys
                325                 330                 335

Lys Leu Arg Glu Tyr Leu Leu Glu Lys Pro Glu Phe Arg Asn Leu Lys
            340                 345                 350

Asp Asp Lys Tyr Asn Ser Val Arg Ser Lys Leu Tyr Arg Leu Ile Asp
        355                 360                 365

Phe Ile Ile Tyr Leu Asn Tyr Lys Asp Ala Lys Phe Lys Asn Asp Asn
    370                 375                 380

Asp Glu Lys Ile Glu Arg Phe Val Ser Lys Leu Arg Lys Cys Pro Asp
385                 390                 395                 400
```

```
Glu Ala Ser Lys Glu Thr Val Tyr Gln Thr Glu Ala Asp Ile Ile Lys
                405                 410                 415

Lys Ser Ile Ser Gly Ser Ile Asn Ser Ile Lys Arg Asp Met Met Leu
            420                 425                 430

Val Lys Ser Lys Tyr Gly Lys Leu Glu Leu Asp Thr Gln Thr Val Asn
        435                 440                 445

Gln Leu Lys Arg Ala Val Asp Asp Ile Ser Ile Asn Thr Lys Gly Asp
    450                 455                 460

Tyr Phe Thr Lys Leu Met Tyr Leu Ile Ser Leu Leu Ile Asp Gly Lys
465                 470                 475                 480

Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile His Gln Phe Glu Asn Ile
                485                 490                 495

Ala Ser Phe Ile Asp Val Met Lys Lys Glu Asn Ile Glu Cys Glu Phe
            500                 505                 510

Thr Glu Lys Phe Ala Phe Phe Lys Asn Ser Glu Ile Ile Ala His Glu
        515                 520                 525

Leu Arg Thr Val Asn Ser Phe Ala Arg Met Thr Gly Phe Met Glu Asn
    530                 535                 540

Asp Glu Ser Thr Lys Phe Ile Met Phe Arg Glu Ala Ala Asp Ile Leu
545                 550                 555                 560

Gly Thr Ser Tyr Thr Asp Asp Glu Leu Thr Asp His Leu Asn Leu Lys
                565                 570                 575

Lys Thr Arg Leu Ile Ala Arg Ser Lys Ala Asp Met Gly Leu Arg Asn
            580                 585                 590

Phe Ile Ile Ser Asn Val Ile Lys Ser Ser Arg Phe Lys Tyr Leu Ile
        595                 600                 605
```

```
Arg Tyr Ser Asn Pro Lys Lys Val Lys Ala Ile Ala Glu Asn Lys Lys
    610                 615                 620

Leu Val Ser Phe Val Leu Asp Lys Ile Pro Asp Thr Gln Ile Ile Arg
625                 630                 635                 640

Tyr Cys Asn Thr Cys Gly Phe Lys Val Gly Lys Arg Gln Glu Asn Ile
                645                 650                 655

Asn Thr Leu Ala Glu Ala Ile Thr Tyr Phe Ser Phe Ser Asn Phe Glu
            660                 665                 670

Lys Val Lys Gln Ser Ala Asn Val Glu Asn Asp Thr Pro Asp Ala Ile
        675                 680                 685

Asp Lys Ala Gln Lys Gln Gly Leu Ile Ser Leu Tyr Leu Thr Val Leu
    690                 695                 700

Tyr Leu Phe Ile Lys Asn Met Val Tyr Val Asn Ser Arg Tyr Val Leu
705                 710                 715                 720

Ala Phe His Cys Leu Glu Arg Asp Ser Gln Leu Leu Lys Ile Thr Ser
                725                 730                 735

Gly Tyr Asn Asn Leu Ser Tyr Asn Asn Leu Thr Asn His Phe Ile Glu
            740                 745                 750

Lys Glu Tyr Leu Asn Ser His Ala Arg Asn Tyr Leu Arg Val Asn Ile
        755                 760                 765

Ser Asn Ser Asp Glu Trp Ser Ile Val Lys Phe Arg Asn Gly Val Ala
    770                 775                 780

His Leu Asn Ser Val Arg Asn Ile Asp Thr Ser Ile Lys Gly Leu Arg
785                 790                 795                 800

Glu Phe Lys Ser Tyr Phe Glu Ile Tyr Ala Tyr Thr Val Gln Lys Ser
                805                 810                 815
```

Ile Lys Asn Tyr Tyr Asp Ala Glu Ser Pro Val Gln Ile Asn Pro Lys
820 825 830

Thr Ile Glu Tyr Met Glu Ser Val Asp Lys Tyr Gly Ser Tyr Thr Lys
835 840 845

Asp Phe Val Lys Ala Leu Asn Thr Pro Phe Gly Tyr Cys Leu Ala Arg
850 855 860

Tyr Lys Asn Leu Ser Ile Glu Pro Leu Phe Asp Arg Asn Thr Val Tyr
865 870 875 880

Asn Lys Lys Asp Val Lys Glu Glu Phe Ala Val Asn Ser Lys
885 890

<210> 250
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 250
gaactatagc cctccgattt cggagggttc tgaaac 36

<210> 251
<211> 980
<212> PRT
<213> Artificial Sequence

<220>
<223> Ga0224422_10645759

<400> 251

Met Asp Lys Tyr Lys Leu Leu Arg Asn Gln Lys Phe Asp Tyr Val Ile
1 5 10 15

Lys Leu Asp Asn Ile Asp Asn Ile Arg Pro Ile Leu Cys Lys Glu Ile
20 25 30

Val Asn Lys Ile Leu Lys Lys Leu Lys Leu Asp Phe Phe Phe Ser Thr

```
        35                  40                  45

Lys Glu Lys Glu Ile His Phe Thr Ala Asp Asp Ile Ile Thr Lys Lys
    50                  55                  60

Tyr Lys Lys Ile Glu Leu Val Gly Asp Asp Lys Glu Val Phe Pro Ile
65                  70                  75                  80

Tyr Val Arg Leu Arg Glu Val Pro Leu Lys Gly Lys Ala Lys Arg Leu
                85                  90                  95

Gly Leu Lys Ser Ile Phe Val Glu Asp Asn Lys Leu Ile Leu Thr Thr
            100                 105                 110

Ile Gly Lys Lys Asp Asp Val Leu Ile Glu Lys Glu Ile Leu Asn Asn
        115                 120                 125

Ser Lys Asn Glu Val Ile Ser His Ile Asn Pro Glu Asn Tyr Ala Val
    130                 135                 140

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

Ala Asp Ile Asp Asn Pro Ser His Ser Lys Val Glu Glu Thr Ile Ser
                165                 170                 175

Ile Lys Lys Asp Tyr Leu Glu Arg Lys Asp Val Ile Glu Lys Lys Phe
            180                 185                 190

Phe Asn Asn Thr Phe Asn Asp Asn Ile His Ile Gln Leu Ala Tyr Asn
        195                 200                 205

Ile Met Asp Ile Asn Lys Ile Phe Asn Ile Tyr Ser Asn Glu Val Val
    210                 215                 220

Phe Thr Met Glu Asn Leu Ala His Ser Asp Asp Ala Lys Asp Phe Ile
225                 230                 235                 240

Gly Gly Ile Asp Tyr Lys Lys Asp Val Ile Gln Asn Ala Lys Phe Asn
```

245 250 255

Asp Tyr Phe Ala Ser Ile Phe Asn Ser Leu Arg Tyr Phe Ser Gln Ala
260 265 270

Phe Pro Asn Lys Thr Phe Asp Lys Ile Lys Ile Asp Gly Glu Lys Asp
275 280 285

Tyr Asn Arg Glu Glu Ile Lys Lys Asn Thr Lys Ile Ile Leu Arg Ile
290 295 300

Leu Ser Met Phe Arg Gln Phe Ser Ile His Gly Tyr Thr Asp Leu Phe
305 310 315 320

Asp Met Glu Lys Asp Pro Lys Ile Arg Asp Leu Leu Asp Tyr Phe Thr
325 330 335

Ala Arg Tyr Asp Glu Val Tyr Asn Asn Leu Ile Arg Asn Phe Ile Lys
340 345 350

Thr Ser Lys Thr Asn Ile Ala Ile Val Phe Glu Ala Leu Asn Leu Pro
355 360 365

Ile Asn Asp Asn Gly Leu Lys Lys Glu Ile Leu Asn Asp Tyr Phe Asp
370 375 380

Phe Ser Ile Thr Lys Ser Tyr Val Asn Met Gly Phe Asn Phe Lys Arg
385 390 395 400

Leu Lys Glu Val Leu Ile Ser Lys Leu Gly Leu Asn Val Ser Asp Ser
405 410 415

Asn Lys Gln Ala Thr Tyr Asn Val Ile Asn Tyr Leu Ile Tyr Arg His
420 425 430

Leu Lys Asp Asn Pro Ile Val Asp Lys Tyr Val Lys Cys Leu Val Ala
435 440 445

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

```
    450                 455                 460

Glu Ile Lys Leu Pro Leu Lys Asn Ile Ile Thr Gln Glu Leu Val Asp
465                 470                 475                 480

Asn Gln Lys Arg Asn Gln Lys Tyr Lys Tyr Phe Asn Ile Asp Gly Ser
                485                 490                 495

Trp Leu Asn Thr Asn Phe Asn Asn Ile Asp Pro Phe Cys Lys Ala Ile
            500                 505                 510

Tyr Phe Leu Thr Ile Phe Leu Asp Thr Lys Glu Ile Asn Met Leu Val
        515                 520                 525

Thr Ser Ile Val Asn Lys Leu Arg Asn Ile Asn Gly Leu Val Lys Val
    530                 535                 540

Leu Asn Asn Lys Ser Asp Leu Ile Tyr Asp Tyr Arg Ile Phe Asp Lys
545                 550                 555                 560

Gly Asn Ala Gln Lys Ile Ala Asp Gln Leu Ser Val Val Leu Ser Ile
                565                 570                 575

Ala Arg Met Gln Lys Asn Phe Asn Lys Lys Phe Lys Ser Asp Asn Lys
            580                 585                 590

Asn Ser Ala Asn Lys Pro Ser Ser Asp Met Ile Tyr Asp Ala Tyr Cys
        595                 600                 605

Leu Phe Gly Tyr Lys Glu Ser Lys Glu Thr Phe Asp Glu Glu Leu Glu
    610                 615                 620

Ala Asp Thr Ser Lys Pro Ser Lys Gly Ile Ala Lys Phe Gln Asn Lys
625                 630                 635                 640

Ile Thr Ser Ala Ile Val Asn Ser Asn Lys Phe Arg Phe Ile Ile Lys
                645                 650                 655

Tyr Ala Asn Pro Arg Asn Cys Tyr Ser Ile Met Lys Asn Thr Arg Val
```

660 665 670

Val Ser Phe Val Leu Lys Glu Ile Pro Gln Ser Gln Ile Asp Lys Tyr
675 680 685

Ile Ile Lys Lys Glu Lys Tyr Asn Ser Leu Asn Tyr Asp Ala Lys Ile
690 695 700

Asn Ala Leu Ala Lys Glu Leu Ala Asp Val Ser Ile Tyr Gly Met Arg
705 710 715 720

Lys Ala Tyr Asn Ile Arg Ala Lys Val Glu Ala Asn Asn Leu Glu Leu
725 730 735

Pro Lys Leu Phe Met Thr Ile Ile Tyr Gln Val Ile Lys Ser Leu Val
740 745 750

Arg Ile Asn Thr Tyr Tyr Asn Ile Ala Phe Ser Cys Leu Glu Arg Asp
755 760 765

Lys Lys Leu Phe Asn Val Pro Asp Asp Tyr Cys Asp Tyr Ser Asp Leu
770 775 780

Met Leu Leu Asp Ala Phe Val Tyr Gly Thr Asn Glu Phe Glu Val Ser
785 790 795 800

Lys Asp Gly Lys Cys Glu Cys Ile Thr Ile Asp Val Pro Leu Asn Lys
805 810 815

Tyr Glu Arg Lys Val Asp Gly Lys Ile Ile Ser Val Gln Gly Thr Lys
820 825 830

Leu Gly Ser Ser Lys Tyr Ser Trp Leu Lys Asp Asp Ile Asp Val Phe
835 840 845

Lys Glu Ile Phe Tyr Met Asn Asn Lys Pro Thr Glu Ile Glu Ser Cys
850 855 860

Ser Val Ile Tyr Arg Ala Cys Tyr Arg Asn Thr Ile Ala His Leu Asn

865 870 875 880

Ala Ile Ser Asn Met Thr Ser Tyr Leu Ala Ser Phe Val Gly Arg Asn
885 890 895

Asn Arg Met Asn Ser Tyr Phe Glu Leu Tyr Gln Tyr Ile Ile Gln Met
900 905 910

Asn Phe Ile Asn Lys Tyr Lys Asp Asn Tyr Ser Asn Glu Phe Ile Asp
915 920 925

Lys Gln Ser Glu Ile Leu Lys Lys Tyr Lys Thr Tyr Ser Lys Asp Leu
930 935 940

Leu His Ile Leu Asn Leu Pro Leu Ala Tyr Asn Met Ala Arg Tyr Lys
945 950 955 960

Asn Leu Thr Ile Glu Asp Ile Phe Tyr Asp Lys Phe Ile Asn Tyr Ile
965 970 975

Lys Lys Ser Glu
980

<210> 252
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> DR

<400> 252
ggacaataac ctgcgaattt tggcaggttc tatgac 36

<210> 253
<211> 996
<212> PRT
<213> Artificial Sequence

<220>
<223> ODAI chimera

<400> 253

```
Met Leu Lys Lys Cys Ser Val Cys Gly Lys Lys Phe Glu Thr Asn Tyr
1               5                  10                  15

Gln Gly Gly Lys Pro Val Cys Arg Asn Cys Arg Ser Asn Lys Ser Lys
            20                  25                  30

Pro Val Asn Pro Lys Ser Ser Arg Lys Pro Glu Ser Ile Asp Lys Ala
        35                  40                  45

Ile Glu Ile Leu Thr Glu Lys Leu Ile Ser Asp Cys Lys Lys Asn Asp
    50                  55                  60

Cys Ala Ser Glu Cys Gly Lys Gly Arg Arg Lys Ser Lys Ala Lys Ala
65                  70                  75                  80

Ala Gly Leu Lys Ser Ala Leu Asp Leu Gly Asn Gly Gln Ala Leu Ile
                85                  90                  95

Thr Ser Phe Gly Lys Gly Asn Glu Ala Val Pro Glu Lys Arg Val Glu
            100                 105                 110

Gly Asp Lys Ile Ile Ser Leu Arg Glu Gly Glu Ala Phe Ala Leu Ala
        115                 120                 125

Lys Glu Pro Val Asn His Lys Leu Val Leu Lys Ser Gly Arg Ile Lys
    130                 135                 140

Ala Glu Gly Gly Pro Val Glu Phe Asp Gly Pro Gln Lys Asn Ser Ser
145                 150                 155                 160

Met Ser Ala Val Arg Leu Asp Glu Tyr Asn Asn Gln Ile Glu Arg Lys
                165                 170                 175

Phe Phe Arg Gly Asn Lys Val Leu Gln Asp Asn Ile Arg Ile Gln Ile
            180                 185                 190

Ala His Asn Ile Met Asp Ile Glu Lys Ile Leu Ala Val His Ile Asn
        195                 200                 205
```

```
Asn Ile Val Tyr Ser Val Asn Asn Leu Phe Arg Val Gln Gly Asn Glu
    210                 215                 220

Gly Asp Asp Phe Leu Gly Tyr Arg Gly Leu Asp Lys Asn Tyr Ser Asp
225                 230                 235                 240

Phe Cys Lys Asn Lys Pro Glu Asp Lys Glu Arg Phe Phe Asn Phe Ile
                245                 250                 255

Gly Lys Gly Asp Gly Lys Val Thr Asn Pro Arg Leu Ser Tyr Tyr Gly
            260                 265                 270

Ala Ala Phe Ala Glu Lys Val Gln Pro Thr Leu Lys Tyr Arg Glu Ala
        275                 280                 285

Glu Asp Ile Tyr Tyr Ile Ile Cys Met Leu Gly Glu Ala Arg Gln Met
    290                 295                 300

Phe Ala His Gly Asn Asp Asp Arg Thr Lys Thr Phe Leu Tyr Asn Ile
305                 310                 315                 320

Val Gly Tyr Arg Asn Gly Thr Glu Ile Asn Ser Lys Phe Thr Ala Val
                325                 330                 335

Leu Asp Gly Ile Tyr Ser Lys Lys Val Glu Glu Leu Asn Gly Asn Phe
            340                 345                 350

Ile Lys Asn Ser Lys Lys Asp Leu Thr Ile Leu Phe Ser Ala Phe Gly
        355                 360                 365

Cys Gln Asp Glu Ala Ser Lys Lys Lys Leu Ile Glu Lys Tyr Tyr Asp
    370                 375                 380

Phe Ile Val Lys Lys Thr Asn Lys Asn Met Gly Phe Ser Val Thr Lys
385                 390                 395                 400

Leu Arg Glu Lys Ile Ile Glu Gly Tyr Arg Phe Arg Gly Asn Pro Val
                405                 410                 415
```

```
Thr Gly Thr Asp Tyr Asp Ser Val Arg His Lys Leu Tyr Arg Thr Phe
            420                 425                 430

Asp Phe Ile Ile Cys Ser Tyr Tyr Asp Asp His Ala Lys Glu Lys Asp
        435                 440                 445

Thr Leu Val Ser Ser Leu Arg Ala Ser Leu Asn Asp Thr Lys Lys Glu
    450                 455                 460

Asn Leu Tyr Ala Gly Glu Ala Lys Thr Ile Ala Glu Ser Ile Lys Asp
465                 470                 475                 480

Arg Met Asp Thr Val Cys Arg Met Tyr Asn Lys Gly Glu Gly Ile Lys
                485                 490                 495

His Ile Thr Val Asp Arg Phe Ile Ala Gly Ala Gly Ala Gly Lys Leu
            500                 505                 510

Ile Asp Lys Cys Ser Ile Ser Thr Glu Cys Asp Tyr Phe Cys Lys Phe
        515                 520                 525

Ile Tyr Leu Met Thr Leu Phe Ile Asp Gly Lys Glu Ile Asn Asp Leu
    530                 535                 540

Leu Thr Thr Leu Ile Asn Lys Phe Asp Asn Ile Asp Ser Pro Ser Val
545                 550                 555                 560

Asp Lys Ala Val Asn Asn Gln Asn Arg Val Glu Phe Gly Asp Lys Tyr
                565                 570                 575

Asp Met Phe Asn Lys Ala Gly Glu Ile Cys Glu Ser Leu Arg Ser Ile
            580                 585                 590

Asn Ser Phe Ala Arg Met Glu Ser Pro Ala Pro Ala Ala Lys Arg Thr
        595                 600                 605

Met Phe Val Asp Ala Ile Arg Met Phe Gly Phe Ala Asp Val Ser Asp
    610                 615                 620
```

```
Asn Asp Ala Glu Lys Phe Tyr Lys Glu Asn Ile Glu Pro Phe Ile Tyr
625                 630                 635                 640

Asp Arg Asn Gly Asp Ile Glu Asp Glu Trp Lys Lys Gly Lys Leu Val
                645                 650                 655

His Lys Lys Asp Glu Lys Lys Lys Gly Val Tyr Asn Phe Ile Ala Thr
            660                 665                 670

Asn Val Ile Glu Ser Asn Arg Phe Gln Tyr Leu Val Arg Tyr Ala Asn
        675                 680                 685

Thr Glu Lys Ile Ala Ala Leu Ala Arg Asn Glu Thr Ile Val Arg Phe
    690                 695                 700

Val Ile Lys Lys Leu Pro Ala Ala Gln Ile Val Arg Tyr Tyr Ile Ala
705                 710                 715                 720

Cys Asn Arg Ile Arg Glu Asn Ile Asn Ser Asn Val Leu Thr Ser Asp
                725                 730                 735

Asp Met Ala Glu Ser Leu Val Lys Leu Val Thr Gly Met Asp Phe Gly
            740                 745                 750

Lys Ile Lys Gly Val Arg Asn Asp Asp Arg Ser Gly Asp Lys Glu Met
        755                 760                 765

Lys Lys Ala Ile Val Thr Leu Tyr Leu Thr Ile Leu Tyr His Leu Thr
    770                 775                 780

Lys Asn Leu Val Tyr Val Asn Ser Arg Tyr Phe Met Ala Phe His Cys
785                 790                 795                 800

Leu Glu Arg Asp Met Asn Leu Phe Gly Glu Gly Gln Lys Gln Ala Leu
                805                 810                 815

Arg Asn Thr Gly Asp Tyr Pro Ala Asp Asn Ala Asp Leu Arg Arg Arg
            820                 825                 830
```

```
Asp Ile Lys Tyr Cys Val Leu Thr Lys Ala His Ile Lys Gln Glu Glu
        835                 840                 845


Val Lys Ser Arg Arg Asn Asp Thr Asp Ser Lys Ser Gln Gln Lys Ala
    850                 855                 860


Lys Ser Arg Arg Arg Arg Trp Cys Leu Leu Leu Lys Glu Asp Ile Asp
865                 870                 875                 880


Lys Leu Asn Gly Cys Glu Ser Leu Arg Ile Leu Tyr Arg Asn Ala Val
                885                 890                 895


Ala His Leu Asn Ile Pro Arg Asn Ala Ile Asp Tyr Ile Gly Gly Ile
            900                 905                 910


Lys Lys Ala Asp Ser Tyr Phe Ala Ile Tyr His Tyr Met Met Gln Gln
        915                 920                 925


Tyr Leu Ser Lys Lys His Gly Leu Pro Glu Gly Leu Gly Cys Thr His
    930                 935                 940


Gly Val Tyr Cys Lys Asp Leu Val Lys Ala Leu Asn Ala Pro Phe Gly
945                 950                 955                 960


Tyr Asn Leu Pro Arg Phe Lys Asn Leu Ser Val Glu Ala Leu Phe Asp
                965                 970                 975


Met Asn Glu Glu Pro Ala Pro Glu Gly Lys Glu Arg Ser Lys Asn Ala
            980                 985                 990


Asp Gly Cys Asn
        995


<210>  254
<211>  36
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  DR
```

<400> 254
gaactacagc cctccttaaa tgggggttc taaaac 36

<210> 255
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> HEPN motif

<220>
<221> misc_feature
<222> (2)..(5)
<223> Xaa can be any naturally occurring amino acid

<400> 255

Arg Xaa Xaa Xaa Xaa His
1 5

<210> 256
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<223> N-terminal NLS

<400> 256

Ser Pro Lys Lys Lys Arg Lys Val Glu Ala Ser
1 5 10

<210> 257
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> N-terminal NLS

<400> 257
agccccaaga agaagagaaa ggtggaggcc agc 33

<210> 258

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

<220>
<223> C-terminal NLS

<400> 258

Gly Pro Lys Lys Lys Arg Lys Val Ala Ala Ala
1 5 10

<210> 259
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> dCas9 target

<400> 259
uuaucgucug uuuguggaug 20

<210> 260
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> C-terminal NLS

<400> 260
ggacctaaga aaaagaggaa ggtggcggcc gct 33

<210> 261
<211> 271
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial Eubacterium siraeum nCas1 array targeting ccdB

<400> 261
gaacuacacc cgugcaaaaa ugcagggguc uaaaacuaac ggcucucucu uuuauaggug 60

uaaaccgaac uacacccgug caaaaaugca ggggucuaaa accuuuaucu gacagcagac 120

gugcacuggc cagaacuaca cccgugcaaa aaugcagggg ucuaaaacca ucaugcgcca 180

gcuuucaucc ccgauaugga acuacacccg ugcaaaaaug caggggucua aaacuaaugg 240

cguuuuugau gucauuuucg cgguccgcug a 271

<210> 262
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial Eubacterium siraeum nCas1 array targeting ccdB - full 36 nt direct repeat

<400> 262
gaacuacacc cgugcaaaaa ugcagggguc uaaaac 36

<210> 263
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial Eubacterium siraeum nCas1 array targeting ccdB - spacer 1

<400> 263
uaacggcucu cucuuuuaua gguguaaacc 30

<210> 264
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial Eubacterium siraeum nCas1 array targeting ccdB - spacer 2

<400> 264
cuuuaucuga cagcagacgu gcacuggcca 30

<210> 265
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial Eubacterium siraeum nCas1 array targeting ccdB - spacer 3

<400> 265
caucaugcgc cagcuuucau ccccgauaug 30

<210> 266
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial Eubacterium siraeum nCas1 array targeting ccdB - spacer 4

<400> 266
uaauggcguu uuugauguca uuuucgcggu ccgcuga 37

<210> 267
<211> 268
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial uncultured Ruminoccus sp. nCas1 array targeting ccdB

<400> 267
cuacuacacu ggugcaaauu ugcacuaguc uaaaacuaac ggcucucucu uuuauaggug 60

uaaacccuac uacacuggug caaauuugca cuagucuaaa accuuuaucu gacagcagac 120

gugcacuggc cacuacuaca cuggugcaaa uuugcacuag ucuaaaacca ucaugcgcca 180

gcuuucaucc ccgauaugcu acuacacugg ugcaaauuug cacuagucua aaacuaaugg 240

cguuuuugau gucauuuucg cgguccgc 268

<210> 268
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial uncultured Ruminoccus sp. nCas1 array targeting ccdB - full 36 nt direct repeat

<400> 268
cuacuacacu ggugcaaauu ugcacuaguc uaaaac 36

<210> 269

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

<220>
<223> Artificial uncultured Ruminoccus sp. nCas1 array targeting ccdB - spacer 1

<400> 269
uaacggcucu cucuuuuaua gguguaaacc 30

<210> 270
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Arificial uncultured Ruminoccus sp. nCas1 array targeting ccdB - spacer 2

<400> 270
cuuuaucuga cagcagacgu gcacuggcca 30

<210> 271
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial uncultured Ruminoccus sp. nCas1 array targeting ccdB - spacer 3

<400> 271
caucaugcgc cagcuuucau ccccgauaug 30

<210> 272
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Artificial uncultured Ruminoccus sp. nCas1 array targeting ccdB - spacer 4

<400> 272
uaauggcguu uuugauguca uuuucgcggu ccgcuga 37

<210> 273

<211> 303
<212> DNA
<213> Artificial Sequence

<220>
<223> Target RNA (ccdB sequence)

<400> 273
augcaguuua agguuuacac cuauaaaaga gagagccguu aucgucuguu uguggaugua 60

cagagugaua uuauugacac gcccgggcga cggaugguga ucccccuggc cagugcacgu 120

cugcugucag auaaagucuc ccgugaacuu uacccggugg ugcauaucgg ggaugaaagc 180

uggcgcauga ugaccaccga uauggccagu gugccggucu ccguuaucgg ggaagaagug 240

gcugaucuca gccaccgcga aaaugacauc aaaaacgcca uuaaccugau guuuugggga 300

aua 303

<210> 274
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Target RNA (ccdB sequence) - spacer 1

<400> 274
gguuuacacc uauaaaagag agagccguua 30

<210> 275
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Target RNA (ccdB sequence) - spacer 2

<400> 275
uggccagugc acgucugcug ucagauaaag 30

<210> 276
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Target RNA (ccdB sequence) - spacer 3

<400> 276
cauaucgggg augaaagcug gcgcaugaug 30

<210> 277
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Target RNA (ccdB sequence)

<400> 277
accgcgaaaa ugacaucaaa aacgccauua 30

<210> 278
<211> 915
<212> PRT
<213> Artificial Sequence

<220>
<223> Mutated Cas13d sequence, NLS-Ga_0531(trunc)-NLS-HA

<400> 278

Met Ser Pro Lys Lys Lys Arg Lys Val Glu Ala Ser Lys Gln Ser Lys
1 5 10 15

Lys Thr Phe Ala Lys Ala Ser Gly Leu Lys Ser Thr Leu Ala Val Asp
20 25 30

Asn Ser Ala Val Met Thr Val Phe Gly Arg Gly Asn Glu Ala Lys Leu
35 40 45

Asp His Arg Ile Asn Ala Asp Leu Gln Ser Glu Ser Leu His Pro Gln
50 55 60

Ala Ala Leu Lys Asn Val His Ala Pro Asn Lys Gln Lys Ile His Phe
65 70 75 80

Ile Gly Arg Met Gln Asp Met Asn Leu Thr Ala Asp His Pro Leu His
85 90 95

Ser His Asp Gly Glu Arg Ala Val Gly Ala Asp Leu Leu Cys Ala Lys

100 105 110

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

His Ile Gln Leu Ile Tyr Gln Ile Leu Asp Ile Gln Lys Ile Leu Ala
130 135 140

Leu His Ala Asn Asn Ile Ile Phe Ala Leu Asp Asn Leu Leu His Lys
145 150 155 160

Lys Asn Asp Glu Leu Ser Asp Asp Phe Val Gly Met Gly Arg Met Arg
165 170 175

Ala Thr Ile Gly Tyr Asp Ala Phe Arg Asn Ser Thr Asn Gln Lys Val
180 185 190

Gln Glu Thr Tyr Arg Glu Phe Gln Glu Phe Val Arg Arg Lys Glu Leu
195 200 205

Leu Tyr Phe Gly Ser Ala Phe Tyr Asn Gly Asp Thr Arg Arg Asp Glu
210 215 220

Lys Val Ile Tyr His Ile Leu Ser Leu Ala Ala Ser Val Arg Gln Phe
225 230 235 240

Cys Phe His Asn Asp Tyr Thr Ser Asp Asp Gly Lys Gly Phe Ile Lys
245 250 255

Ala Asp Trp Met Tyr Arg Leu Glu Glu Ala Leu Pro Ala Glu Tyr Lys
260 265 270

Asp Thr Leu Asp Ala Leu Tyr Leu Glu Gly Val Glu Gly Leu Asp Gln
275 280 285

Ser Phe Leu Lys Asn Asn Thr Val Asn Ile Gln Ile Leu Cys Ser Ile
290 295 300

Phe Asn His Asp Asp Pro Asn Lys Ile Ala Glu Glu Tyr Tyr Gly Phe

305 310 315 320

Leu Met Thr Lys Glu Tyr Lys Asn Met Gly Phe Ser Ile Lys Lys Leu
325 330 335

Arg Glu Cys Met Leu Glu Leu Pro Glu Leu Ser Gly Tyr Lys Glu Asp
340 345 350

Gln Tyr Asn Ser Val Arg Ser Lys Leu Tyr Lys Leu Phe Asp Phe Ile
355 360 365

Ile Ala His Tyr Phe Arg Lys His Pro Glu Lys Gly Glu Glu Met Val
370 375 380

Asp Cys Leu Arg Leu Cys Met Thr Glu Asp Glu Lys Asp Ser His Tyr
385 390 395 400

Glu Gly Thr Ala Lys Lys Leu Val Arg Glu Leu Ala Tyr Asp Met Gln
405 410 415

Glu Ala Ala Glu Gln Ala Asn Gly Ser Asn Ile Thr Gln Met Gln Lys
420 425 430

Asn Glu Gln Gln Gly Lys Thr Lys Gly Met Phe Ala Ile Arg Asp Glu
435 440 445

Ile Arg Val Ser Arg Lys Pro Val Ser Tyr Phe Ser Lys Val Ile Tyr
450 455 460

Val Met Thr Leu Leu Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr
465 470 475 480

Thr Leu Ile Asn Lys Phe Glu Asn Ile Val Ser Phe Glu Asp Val Leu
485 490 495

Arg Gln Leu Asn Val Asp Cys Thr Phe Lys Pro Glu Phe Ala Phe Phe
500 505 510

Gly Tyr Asp Arg Cys Arg Asn Ile Ser Gly Glu Leu Arg Leu Ile Asn

515 520 525

Ser Phe Ala Arg Met Gln Lys Pro Ser Ala Lys Ala Lys His Val Met
530 535 540

Tyr Arg Asp Ala Leu Arg Ile Leu Gly Leu Asp Asn Gly Met Ser Glu
545 550 555 560

Glu Ala Leu Asp Gln Glu Val Arg Arg Ile Leu Gln Ile Gly Ala Asp
565 570 575

Gly Lys Pro Ile Lys Asn Ala Asn Lys Gly Phe Arg Asn Phe Ile Ala
580 585 590

Ser Asn Val Ile Glu Ser Ser Arg Phe Arg Tyr Leu Val Arg Tyr Asn
595 600 605

Asn Pro His Lys Thr Arg Met Ile Ala Gln Asn Glu Ala Ile Val Arg
610 615 620

Phe Val Leu Ser Glu Ile Pro Asp Glu Gln Ile Arg Arg Tyr Tyr Asp
625 630 635 640

Val Cys Arg Asp Pro Lys Leu Pro Arg Ser Ser Ser Arg Glu Ala Gln
645 650 655

Val Asp Ile Leu Thr Gly Ile Ile Thr Asp Val Asn Tyr Arg Ile Phe
660 665 670

Glu Asp Val Pro Gln Ser Lys Lys Ile Asn Lys Asp Arg Pro Asp Ala
675 680 685

Asn Asp Arg Met Thr Leu Lys Lys Gln Arg Tyr Gln Ala Ile Val Ser
690 695 700

Leu Tyr Leu Thr Val Met Tyr Leu Val Thr Lys Asn Leu Val Tyr Val
705 710 715 720

Asn Ser Arg Tyr Val Met Ala Phe His Ala Leu Glu Arg Asp Ala Tyr

725 730 735

Leu Tyr Gly Ile Thr Asn Ile Lys Gly Asp Tyr Arg Lys Leu Thr Asp
740 745 750

Asn Leu Leu Ala Asp Glu Asn Tyr Lys Lys Phe Gly His Phe Lys Asn
755 760 765

Lys Lys Trp Arg Gly Ile Ala Glu Gln Asn Leu Arg Asn Ser Asp Val
770 775 780

Pro Val Ile Lys Ser Phe Arg Asn Met Ala Ala His Ile Ser Val Ile
785 790 795 800

Arg Asn Ile Asp Leu Tyr Ile Gly Asp Ile Gln Lys Val Asp Ser Tyr
805 810 815

Phe Ala Leu Tyr His Phe Leu Met Gln Lys Leu Ile Gln Arg Val Val
820 825 830

Pro Glu Asn Thr Lys Gly Leu Ser Asp Gln Thr Lys Lys Tyr Tyr Asp
835 840 845

Ala Leu Glu Gln Tyr Asn Thr Tyr Cys Lys Asp Phe Val Lys Ala Tyr
850 855 860

Cys Thr Pro Phe Ala Tyr Val Thr Pro Arg Tyr Lys Asn Leu Thr Ile
865 870 875 880

Asp Gly Leu Phe Asp Arg Asn Arg Pro Gly Glu Asp Lys Gly Ser Gly
885 890 895

Pro Lys Lys Lys Arg Lys Val Ala Ala Ala Tyr Pro Tyr Asp Val Pro
900 905 910

Asp Tyr Ala
915

<210> 279

<211> 919
<212> PRT
<213> Artificial Sequence

<220>
<223> Mutated Cas13d sequence, NES-Ga_0531(trunc)-NES-HA

<400> 279

Met Leu Gln Leu Pro Pro Leu Glu Arg Leu Thr Leu Gly Ser Gly Lys
1 5 10 15

Gln Ser Lys Lys Thr Phe Ala Lys Ala Ser Gly Leu Lys Ser Thr Leu
20 25 30

Ala Val Asp Asn Ser Ala Val Met Thr Val Phe Gly Arg Gly Asn Glu
35 40 45

Ala Lys Leu Asp His Arg Ile Asn Ala Asp Leu Gln Ser Glu Ser Leu
50 55 60

His Pro Gln Ala Ala Leu Lys Asn Val His Ala Pro Asn Lys Gln Lys
65 70 75 80

Ile His Phe Ile Gly Arg Met Gln Asp Met Asn Leu Thr Ala Asp His
85 90 95

Pro Leu His Ser His Asp Gly Glu Arg Ala Val Gly Ala Asp Leu Leu
100 105 110

Cys Ala Lys Asp Lys Leu Glu Gln Leu Tyr Phe Gly Arg Thr Phe Asn
115 120 125

Asp Asn Ile His Ile Gln Leu Ile Tyr Gln Ile Leu Asp Ile Gln Lys
130 135 140

Ile Leu Ala Leu His Ala Asn Asn Ile Ile Phe Ala Leu Asp Asn Leu
145 150 155 160

Leu His Lys Lys Asn Asp Glu Leu Ser Asp Asp Phe Val Gly Met Gly
165 170 175

```
Arg Met Arg Ala Thr Ile Gly Tyr Asp Ala Phe Arg Asn Ser Thr Asn
            180                 185                 190

Gln Lys Val Gln Glu Thr Tyr Arg Glu Phe Gln Glu Phe Val Arg Arg
        195                 200                 205

Lys Glu Leu Leu Tyr Phe Gly Ser Ala Phe Tyr Asn Gly Asp Thr Arg
    210                 215                 220

Arg Asp Glu Lys Val Ile Tyr His Ile Leu Ser Leu Ala Ala Ser Val
225                 230                 235                 240

Arg Gln Phe Cys Phe His Asn Asp Tyr Thr Ser Asp Asp Gly Lys Gly
                245                 250                 255

Phe Ile Lys Ala Asp Trp Met Tyr Arg Leu Glu Glu Ala Leu Pro Ala
            260                 265                 270

Glu Tyr Lys Asp Thr Leu Asp Ala Leu Tyr Leu Glu Gly Val Glu Gly
        275                 280                 285

Leu Asp Gln Ser Phe Leu Lys Asn Asn Thr Val Asn Ile Gln Ile Leu
    290                 295                 300

Cys Ser Ile Phe Asn His Asp Asp Pro Asn Lys Ile Ala Glu Glu Tyr
305                 310                 315                 320

Tyr Gly Phe Leu Met Thr Lys Glu Tyr Lys Asn Met Gly Phe Ser Ile
                325                 330                 335

Lys Lys Leu Arg Glu Cys Met Leu Glu Leu Pro Glu Leu Ser Gly Tyr
            340                 345                 350

Lys Glu Asp Gln Tyr Asn Ser Val Arg Ser Lys Leu Tyr Lys Leu Phe
        355                 360                 365

Asp Phe Ile Ile Ala His Tyr Phe Arg Lys His Pro Glu Lys Gly Glu
    370                 375                 380
```

```
Glu Met Val Asp Cys Leu Arg Leu Cys Met Thr Glu Asp Glu Lys Asp
385                 390                 395                 400

Ser His Tyr Glu Gly Thr Ala Lys Lys Leu Val Arg Glu Leu Ala Tyr
                405                 410                 415

Asp Met Gln Glu Ala Ala Glu Gln Ala Asn Gly Ser Asn Ile Thr Gln
            420                 425                 430

Met Gln Lys Asn Glu Gln Gln Gly Lys Thr Lys Gly Met Phe Ala Ile
        435                 440                 445

Arg Asp Glu Ile Arg Val Ser Arg Lys Pro Val Ser Tyr Phe Ser Lys
    450                 455                 460

Val Ile Tyr Val Met Thr Leu Leu Leu Asp Gly Lys Glu Ile Asn Asp
465                 470                 475                 480

Leu Leu Thr Thr Leu Ile Asn Lys Phe Glu Asn Ile Val Ser Phe Glu
                485                 490                 495

Asp Val Leu Arg Gln Leu Asn Val Asp Cys Thr Phe Lys Pro Glu Phe
            500                 505                 510

Ala Phe Phe Gly Tyr Asp Arg Cys Arg Asn Ile Ser Gly Glu Leu Arg
        515                 520                 525

Leu Ile Asn Ser Phe Ala Arg Met Gln Lys Pro Ser Ala Lys Ala Lys
    530                 535                 540

His Val Met Tyr Arg Asp Ala Leu Arg Ile Leu Gly Leu Asp Asn Gly
545                 550                 555                 560

Met Ser Glu Glu Ala Leu Asp Gln Glu Val Arg Arg Ile Leu Gln Ile
                565                 570                 575

Gly Ala Asp Gly Lys Pro Ile Lys Asn Ala Asn Lys Gly Phe Arg Asn
            580                 585                 590
```

```
Phe Ile Ala Ser Asn Val Ile Glu Ser Ser Arg Phe Arg Tyr Leu Val
        595                 600                 605

Arg Tyr Asn Asn Pro His Lys Thr Arg Met Ile Ala Gln Asn Glu Ala
    610                 615                 620

Ile Val Arg Phe Val Leu Ser Glu Ile Pro Asp Glu Gln Ile Arg Arg
625                 630                 635                 640

Tyr Tyr Asp Val Cys Arg Asp Pro Lys Leu Pro Arg Ser Ser Ser Arg
                645                 650                 655

Glu Ala Gln Val Asp Ile Leu Thr Gly Ile Ile Thr Asp Val Asn Tyr
            660                 665                 670

Arg Ile Phe Glu Asp Val Pro Gln Ser Lys Lys Ile Asn Lys Asp Arg
        675                 680                 685

Pro Asp Ala Asn Asp Arg Met Thr Leu Lys Lys Gln Arg Tyr Gln Ala
    690                 695                 700

Ile Val Ser Leu Tyr Leu Thr Val Met Tyr Leu Val Thr Lys Asn Leu
705                 710                 715                 720

Val Tyr Val Asn Ser Arg Tyr Val Met Ala Phe His Ala Leu Glu Arg
                725                 730                 735

Asp Ala Tyr Leu Tyr Gly Ile Thr Asn Ile Lys Gly Asp Tyr Arg Lys
            740                 745                 750

Leu Thr Asp Asn Leu Leu Ala Asp Glu Asn Tyr Lys Lys Phe Gly His
        755                 760                 765

Phe Lys Asn Lys Lys Trp Arg Gly Ile Ala Glu Gln Asn Leu Arg Asn
    770                 775                 780

Ser Asp Val Pro Val Ile Lys Ser Phe Arg Asn Met Ala Ala His Ile
785                 790                 795                 800
```

```
Ser Val Ile Arg Asn Ile Asp Leu Tyr Ile Gly Asp Ile Gln Lys Val
                805                 810                 815


Asp Ser Tyr Phe Ala Leu Tyr His Phe Leu Met Gln Lys Leu Ile Gln
            820                 825                 830


Arg Val Val Pro Glu Asn Thr Lys Gly Leu Ser Asp Gln Thr Lys Lys
        835                 840                 845


Tyr Tyr Asp Ala Leu Glu Gln Tyr Asn Thr Tyr Cys Lys Asp Phe Val
    850                 855                 860


Lys Ala Tyr Cys Thr Pro Phe Ala Tyr Val Thr Pro Arg Tyr Lys Asn
865                 870                 875                 880


Leu Thr Ile Asp Gly Leu Phe Asp Arg Asn Arg Pro Gly Glu Asp Lys
                885                 890                 895


Gly Ser Gly Leu Gln Leu Pro Pro Leu Glu Arg Leu Thr Leu Tyr Pro
            900                 905                 910


Tyr Asp Val Pro Asp Tyr Ala
        915


<210>  280
<211>  1000
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Mutated Cas13d sequence, NLS-RfxCas13d-NLS-HA

<400>  280

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


Lys Ser Phe Ala Lys Gly Met Gly Val Lys Ser Thr Leu Val Ser Gly
            20                  25                  30


Ser Lys Val Tyr Met Thr Thr Phe Ala Glu Gly Ser Asp Ala Arg Leu
```

```
        35                  40                  45

Glu Lys Ile Val Glu Gly Asp Ser Ile Arg Ser Val Asn Glu Gly Glu
    50                  55                  60

Ala Phe Ser Ala Glu Met Ala Asp Lys Asn Ala Gly Tyr Lys Ile Gly
65                  70                  75                  80

Asn Ala Lys Phe Ser His Pro Lys Gly Tyr Ala Val Val Ala Asn Asn
                85                  90                  95

Pro Leu Tyr Thr Gly Pro Val Gln Gln Asp Met Leu Gly Leu Lys Glu
            100                 105                 110

Thr Leu Glu Lys Arg Tyr Phe Gly Glu Ser Ala Asp Gly Asn Asp Asn
        115                 120                 125

Ile Cys Ile Gln Val Ile His Asn Ile Leu Asp Ile Glu Lys Ile Leu
    130                 135                 140

Ala Glu Tyr Ile Thr Asn Ala Ala Tyr Ala Val Asn Asn Ile Ser Gly
145                 150                 155                 160

Leu Asp Lys Asp Ile Ile Gly Phe Gly Lys Phe Ser Thr Val Tyr Thr
                165                 170                 175

Tyr Asp Glu Phe Lys Asp Pro Glu His His Arg Ala Ala Phe Asn Asn
            180                 185                 190

Asn Asp Lys Leu Ile Asn Ala Ile Lys Ala Gln Tyr Asp Glu Phe Asp
        195                 200                 205

Asn Phe Leu Asp Asn Pro Arg Leu Gly Tyr Phe Gly Gln Ala Phe Phe
    210                 215                 220

Ser Lys Glu Gly Arg Asn Tyr Ile Ile Asn Tyr Gly Asn Glu Cys Tyr
225                 230                 235                 240

Asp Ile Leu Ala Leu Leu Ser Gly Leu Arg His Trp Val Val His Asn
```

245 250 255

Asn Glu Glu Glu Ser Arg Ile Ser Arg Thr Trp Leu Tyr Asn Leu Asp
260 265 270

Lys Asn Leu Asp Asn Glu Tyr Ile Ser Thr Leu Asn Tyr Leu Tyr Asp
275 280 285

Arg Ile Thr Asn Glu Leu Thr Asn Ser Phe Ser Lys Asn Ser Ala Ala
290 295 300

Asn Val Asn Tyr Ile Ala Glu Thr Leu Gly Ile Asn Pro Ala Glu Phe
305 310 315 320

Ala Glu Gln Tyr Phe Arg Phe Ser Ile Met Lys Glu Gln Lys Asn Leu
325 330 335

Gly Phe Asn Ile Thr Lys Leu Arg Glu Val Met Leu Asp Arg Lys Asp
340 345 350

Met Ser Glu Ile Arg Lys Asn His Lys Val Phe Asp Ser Ile Arg Thr
355 360 365

Lys Val Tyr Thr Met Met Asp Phe Val Ile Tyr Arg Tyr Tyr Ile Glu
370 375 380

Glu Asp Ala Lys Val Ala Ala Ala Asn Lys Ser Leu Pro Asp Asn Glu
385 390 395 400

Lys Ser Leu Ser Glu Lys Asp Ile Phe Val Ile Asn Leu Arg Gly Ser
405 410 415

Phe Asn Asp Asp Gln Lys Asp Ala Leu Tyr Tyr Asp Glu Ala Asn Arg
420 425 430

Ile Trp Arg Lys Leu Glu Asn Ile Met His Asn Ile Lys Glu Phe Arg
435 440 445

Gly Asn Lys Thr Arg Glu Tyr Lys Lys Lys Asp Ala Pro Arg Leu Pro

```
    450                 455                 460

Arg Ile Leu Pro Ala Gly Arg Asp Val Ser Ala Phe Ser Lys Leu Met
465                 470                 475                 480

Tyr Ala Leu Thr Met Phe Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu
                485                 490                 495

Thr Thr Leu Ile Asn Lys Phe Asp Asn Ile Gln Ser Phe Leu Lys Val
            500                 505                 510

Met Pro Leu Ile Gly Val Asn Ala Lys Phe Val Glu Glu Tyr Ala Phe
        515                 520                 525

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

Phe Ala Arg Met Gly Glu Pro Ile Ala Asp Ala Arg Arg Ala Met Tyr
545                 550                 555                 560

Ile Asp Ala Ile Arg Ile Leu Gly Thr Asn Leu Ser Tyr Asp Glu Leu
                565                 570                 575

Lys Ala Leu Ala Asp Thr Phe Ser Leu Asp Glu Asn Gly Asn Lys Leu
            580                 585                 590

Lys Lys Gly Lys His Gly Met Arg Asn Phe Ile Ile Asn Asn Val Ile
        595                 600                 605

Ser Asn Lys Arg Phe His Tyr Leu Ile Arg Tyr Gly Asp Pro Ala His
    610                 615                 620

Leu His Glu Ile Ala Lys Asn Glu Ala Val Val Lys Phe Val Leu Gly
625                 630                 635                 640

Arg Ile Ala Asp Ile Gln Lys Lys Gln Gly Gln Asn Gly Lys Asn Gln
                645                 650                 655

Ile Asp Arg Tyr Tyr Glu Thr Cys Ile Gly Lys Asp Lys Gly Lys Ser
```

660 665 670

Val Ser Glu Lys Val Asp Ala Leu Thr Lys Ile Ile Thr Gly Met Asn
675 680 685

Tyr Asp Gln Phe Asp Lys Lys Arg Ser Val Ile Glu Asp Thr Gly Arg
690 695 700

Glu Asn Ala Glu Arg Glu Lys Phe Lys Lys Ile Ile Ser Leu Tyr Leu
705 710 715 720

Thr Val Ile Tyr His Ile Leu Lys Asn Ile Val Asn Ile Asn Ala Arg
725 730 735

Tyr Val Ile Gly Phe His Cys Val Glu Arg Asp Ala Gln Leu Tyr Lys
740 745 750

Glu Lys Gly Tyr Asp Ile Asn Leu Lys Lys Leu Glu Glu Lys Gly Phe
755 760 765

Ser Ser Val Thr Lys Leu Cys Ala Gly Ile Asp Glu Thr Ala Pro Asp
770 775 780

Lys Arg Lys Asp Val Glu Lys Glu Met Ala Glu Arg Ala Lys Glu Ser
785 790 795 800

Ile Asp Ser Leu Glu Ser Ala Asn Pro Lys Leu Tyr Ala Asn Tyr Ile
805 810 815

Lys Tyr Ser Asp Glu Lys Lys Ala Glu Glu Phe Thr Arg Gln Ile Asn
820 825 830

Arg Glu Lys Ala Lys Thr Ala Leu Asn Ala Tyr Leu Arg Asn Thr Lys
835 840 845

Trp Asn Val Ile Ile Arg Glu Asp Leu Leu Arg Ile Asp Asn Lys Thr
850 855 860

Cys Thr Leu Phe Arg Asn Lys Ala Val His Leu Glu Val Ala Arg Tyr

865 870 875 880

Val His Ala Tyr Ile Asn Asp Ile Ala Glu Val Asn Ser Tyr Phe Gln
885 890 895

Leu Tyr His Tyr Ile Met Gln Arg Ile Ile Met Asn Glu Arg Tyr Glu
900 905 910

Lys Ser Ser Gly Lys Val Ser Glu Tyr Phe Asp Ala Val Asn Asp Glu
915 920 925

Lys Lys Tyr Asn Asp Arg Leu Leu Lys Leu Leu Cys Val Pro Phe Gly
930 935 940

Tyr Cys Ile Pro Arg Phe Lys Asn Leu Ser Ile Glu Ala Leu Phe Asp
945 950 955 960

Arg Asn Glu Ala Ala Lys Phe Asp Lys Glu Lys Lys Lys Val Ser Gly
965 970 975

Asn Ser Gly Ser Gly Pro Lys Lys Lys Arg Lys Val Ala Ala Ala Tyr
980 985 990

Pro Tyr Asp Val Pro Asp Tyr Ala
995 1000

<210> 281
<211> 980
<212> PRT
<213> Artificial Sequence

<220>
<223> Mutated Cas13d sequence, NLS-RfxCas13d(del5)-NLS-HA

<400> 281

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

Lys Ser Phe Ala Lys Gly Met Gly Val Lys Ser Thr Leu Val Ser Gly
20 25 30

```
Ser Lys Val Tyr Met Thr Thr Phe Ala Glu Gly Ser Asp Ala Arg Leu
        35                  40                  45

Glu Lys Ile Val Glu Gly Asp Ser Ile Arg Ser Val Asn Glu Gly Glu
    50                  55                  60

Ala Phe Ser Ala Glu Met Ala Asp Lys Asn Ala Gly Tyr Lys Ile Gly
65                  70                  75                  80

Asn Ala Lys Phe Ser His Pro Lys Gly Tyr Ala Val Val Ala Asn Asn
                85                  90                  95

Pro Leu Tyr Thr Gly Pro Val Gln Gln Asp Met Leu Gly Leu Lys Glu
            100                 105                 110

Thr Leu Glu Lys Arg Tyr Phe Gly Glu Ser Ala Asp Gly Asn Asp Asn
        115                 120                 125

Ile Cys Ile Gln Val Ile His Asn Ile Leu Asp Ile Glu Lys Ile Leu
    130                 135                 140

Ala Glu Tyr Ile Thr Asn Ala Ala Tyr Ala Val Asn Asn Ile Ser Gly
145                 150                 155                 160

Leu Asp Lys Asp Ile Ile Gly Phe Gly Lys Phe Ser Thr Val Tyr Thr
                165                 170                 175

Tyr Asp Glu Phe Lys Asp Pro Glu His His Arg Ala Ala Phe Asn Asn
            180                 185                 190

Asn Asp Lys Leu Ile Asn Ala Ile Lys Ala Gln Tyr Asp Glu Phe Asp
        195                 200                 205

Asn Phe Leu Asp Asn Pro Arg Leu Gly Tyr Phe Gly Gln Ala Phe Phe
    210                 215                 220

Ser Lys Glu Gly Arg Asn Tyr Ile Ile Asn Tyr Gly Asn Glu Cys Tyr
225                 230                 235                 240
```

```
Asp Ile Leu Ala Leu Leu Ser Gly Leu Arg His Trp Val Val His Asn
                245                 250                 255

Asn Glu Glu Glu Ser Arg Ile Ser Arg Thr Trp Leu Tyr Asn Leu Asp
            260                 265                 270

Lys Asn Leu Asp Asn Glu Tyr Ile Ser Thr Leu Asn Tyr Leu Tyr Asp
        275                 280                 285

Arg Ile Thr Asn Glu Leu Thr Asn Ser Phe Ser Lys Asn Ser Ala Ala
    290                 295                 300

Asn Val Asn Tyr Ile Ala Glu Thr Leu Gly Ile Asn Pro Ala Glu Phe
305                 310                 315                 320

Ala Glu Gln Tyr Phe Arg Phe Ser Ile Met Lys Glu Gln Lys Asn Leu
                325                 330                 335

Gly Phe Asn Ile Thr Lys Leu Arg Glu Val Met Leu Asp Arg Lys Asp
            340                 345                 350

Met Ser Glu Ile Arg Lys Asn His Lys Val Phe Asp Ser Ile Arg Thr
        355                 360                 365

Lys Val Tyr Thr Met Met Asp Phe Val Ile Tyr Arg Tyr Tyr Ile Glu
    370                 375                 380

Glu Asp Ala Lys Val Ala Ala Ala Asn Lys Ser Leu Pro Asp Asn Glu
385                 390                 395                 400

Lys Ser Leu Ser Glu Lys Asp Ile Phe Val Ile Asn Leu Arg Gly Ser
                405                 410                 415

Phe Asn Asp Asp Gln Lys Asp Ala Leu Tyr Tyr Asp Glu Ala Asn Arg
            420                 425                 430

Ile Trp Arg Lys Leu Glu Asn Ile Met His Asn Ile Lys Glu Phe Arg
        435                 440                 445
```

```
Gly Asn Lys Thr Arg Glu Tyr Lys Lys Lys Asp Ala Pro Arg Leu Pro
    450                 455                 460

Arg Ile Leu Pro Ala Gly Arg Asp Val Ser Ala Phe Ser Lys Leu Met
465                 470                 475                 480

Tyr Ala Leu Thr Met Phe Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu
                485                 490                 495

Thr Thr Leu Ile Asn Lys Phe Asp Asn Ile Gln Ser Phe Leu Lys Val
            500                 505                 510

Met Pro Leu Ile Gly Val Asn Ala Lys Phe Val Glu Glu Tyr Ala Phe
        515                 520                 525

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

Phe Ala Arg Met Gly Glu Pro Ile Ala Asp Ala Arg Arg Ala Met Tyr
545                 550                 555                 560

Ile Asp Ala Ile Arg Ile Leu Gly Gly Gly Gly Gly Ser Gly Gly Gly
                565                 570                 575

Gly Ser Met Arg Asn Phe Ile Ile Asn Asn Val Ile Ser Asn Lys Arg
            580                 585                 590

Phe His Tyr Leu Ile Arg Tyr Gly Asp Pro Ala His Leu His Glu Ile
        595                 600                 605

Ala Lys Asn Glu Ala Val Val Lys Phe Val Leu Gly Arg Ile Ala Asp
    610                 615                 620

Ile Gln Lys Lys Gln Gly Gln Asn Gly Lys Asn Gln Ile Asp Arg Tyr
625                 630                 635                 640

Tyr Glu Thr Cys Ile Gly Lys Asp Lys Gly Lys Ser Val Ser Glu Lys
                645                 650                 655
```

```
Val Asp Ala Leu Thr Lys Ile Ile Thr Gly Met Asn Tyr Asp Gln Phe
            660                 665                 670

Asp Lys Lys Arg Ser Val Ile Glu Asp Thr Gly Arg Glu Asn Ala Glu
        675                 680                 685

Arg Glu Lys Phe Lys Lys Ile Ile Ser Leu Tyr Leu Thr Val Ile Tyr
    690                 695                 700

His Ile Leu Lys Asn Ile Val Asn Ile Asn Ala Arg Tyr Val Ile Gly
705                 710                 715                 720

Phe His Cys Val Glu Arg Asp Ala Gln Leu Tyr Lys Glu Lys Gly Tyr
                725                 730                 735

Asp Ile Asn Leu Lys Lys Leu Glu Glu Lys Gly Phe Ser Ser Val Thr
            740                 745                 750

Lys Leu Cys Ala Gly Ile Asp Glu Thr Ala Pro Asp Lys Arg Lys Asp
        755                 760                 765

Val Glu Lys Glu Met Ala Glu Arg Ala Lys Glu Ser Ile Asp Ser Leu
    770                 775                 780

Glu Ser Ala Asn Pro Lys Leu Tyr Ala Asn Tyr Ile Lys Tyr Ser Asp
785                 790                 795                 800

Glu Lys Lys Ala Glu Glu Phe Thr Arg Gln Ile Asn Arg Glu Lys Ala
                805                 810                 815

Lys Thr Ala Leu Asn Ala Tyr Leu Arg Asn Thr Lys Trp Asn Val Ile
            820                 825                 830

Ile Arg Glu Asp Leu Leu Arg Ile Asp Asn Lys Thr Cys Thr Leu Phe
        835                 840                 845

Arg Asn Lys Ala Val His Leu Glu Val Ala Arg Tyr Val His Ala Tyr
    850                 855                 860
```

```
Ile Asn Asp Ile Ala Glu Val Asn Ser Tyr Phe Gln Leu Tyr His Tyr
865                 870                 875                 880


Ile Met Gln Arg Ile Ile Met Asn Glu Arg Tyr Glu Lys Ser Ser Gly
                885                 890                 895


Lys Val Ser Glu Tyr Phe Asp Ala Val Asn Asp Glu Lys Lys Tyr Asn
            900                 905                 910


Asp Arg Leu Leu Lys Leu Leu Cys Val Pro Phe Gly Tyr Cys Ile Pro
        915                 920                 925


Arg Phe Lys Asn Leu Ser Ile Glu Ala Leu Phe Asp Arg Asn Glu Ala
    930                 935                 940


Ala Lys Phe Asp Lys Glu Lys Lys Lys Val Ser Gly Asn Ser Gly Ser
945                 950                 955                 960


Gly Pro Lys Lys Lys Arg Lys Val Ala Ala Ala Tyr Pro Tyr Asp Val
                965                 970                 975


Pro Asp Tyr Ala
            980


<210>  282
<211>  965
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Mutated Cas13d sequence, NLS-RfxCas13d(del5.12)-NLS-HA

<400>  282

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


Lys Ser Phe Ala Lys Gly Met Gly Val Lys Ser Thr Leu Val Ser Gly
            20                  25                  30


Ser Lys Val Tyr Met Thr Thr Phe Ala Glu Gly Ser Asp Ala Arg Leu
```

35 40 45

Glu Lys Ile Val Glu Gly Asp Ser Ile Arg Ser Val Asn Glu Gly Glu
50 55 60

Ala Phe Ser Ala Glu Met Ala Asp Lys Asn Ala Gly Tyr Lys Ile Gly
65 70 75 80

Asn Ala Lys Phe Ser His Pro Lys Gly Tyr Ala Val Val Ala Asn Asn
85 90 95

Pro Leu Tyr Thr Gly Pro Val Gln Gln Asp Met Leu Gly Leu Lys Glu
100 105 110

Thr Leu Glu Lys Arg Tyr Phe Gly Glu Ser Ala Asp Gly Asn Asp Asn
115 120 125

Ile Cys Ile Gln Val Ile His Asn Ile Leu Asp Ile Glu Lys Ile Leu
130 135 140

Ala Glu Tyr Ile Thr Asn Ala Ala Tyr Ala Val Asn Asn Ile Ser Gly
145 150 155 160

Leu Asp Lys Asp Ile Ile Gly Phe Gly Lys Phe Ser Thr Val Tyr Thr
165 170 175

Tyr Asp Glu Phe Lys Asp Pro Glu His His Arg Ala Ala Phe Asn Asn
180 185 190

Asn Asp Lys Leu Ile Asn Ala Ile Lys Ala Gln Tyr Asp Glu Phe Asp
195 200 205

Asn Phe Leu Asp Asn Pro Arg Leu Gly Tyr Phe Gly Gln Ala Phe Phe
210 215 220

Ser Lys Glu Gly Arg Asn Tyr Ile Ile Asn Tyr Gly Asn Glu Cys Tyr
225 230 235 240

Asp Ile Leu Ala Leu Leu Ser Gly Leu Arg His Trp Val Val His Asn

245 250 255

Asn Glu Glu Glu Ser Arg Ile Ser Arg Thr Trp Leu Tyr Asn Leu Asp
260 265 270

Lys Asn Leu Asp Asn Glu Tyr Ile Ser Thr Leu Asn Tyr Leu Tyr Asp
275 280 285

Arg Ile Thr Asn Glu Leu Thr Asn Ser Phe Ser Lys Asn Ser Ala Ala
290 295 300

Asn Val Asn Tyr Ile Ala Glu Thr Leu Gly Ile Asn Pro Ala Glu Phe
305 310 315 320

Ala Glu Gln Tyr Phe Arg Phe Ser Ile Met Lys Glu Gln Lys Asn Leu
325 330 335

Gly Phe Asn Ile Thr Lys Leu Arg Glu Val Met Leu Asp Arg Lys Asp
340 345 350

Met Ser Glu Ile Arg Lys Asn His Lys Val Phe Asp Ser Ile Arg Thr
355 360 365

Lys Val Tyr Thr Met Met Asp Phe Val Ile Tyr Arg Tyr Tyr Ile Glu
370 375 380

Glu Asp Ala Lys Val Ala Ala Ala Asn Lys Ser Leu Pro Asp Asn Glu
385 390 395 400

Lys Ser Leu Ser Glu Lys Asp Ile Phe Val Ile Asn Leu Arg Gly Ser
405 410 415

Phe Asn Asp Asp Gln Lys Asp Ala Leu Tyr Tyr Asp Glu Ala Asn Arg
420 425 430

Ile Trp Arg Lys Leu Glu Asn Ile Met His Asn Ile Lys Glu Phe Arg
435 440 445

Gly Asn Lys Thr Arg Glu Tyr Lys Lys Lys Asp Ala Pro Arg Leu Pro

```
    450                 455                 460

Arg Ile Leu Pro Ala Gly Arg Asp Val Ser Ala Phe Ser Lys Leu Met
465                 470                 475                 480

Tyr Ala Leu Thr Met Phe Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu
                485                 490                 495

Thr Thr Leu Ile Asn Lys Phe Asp Asn Ile Gln Ser Phe Leu Lys Val
            500                 505                 510

Met Pro Leu Ile Gly Val Asn Ala Lys Phe Val Glu Glu Tyr Ala Phe
        515                 520                 525

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

Phe Ala Arg Met Gly Glu Pro Ile Ala Asp Ala Arg Arg Ala Met Tyr
545                 550                 555                 560

Ile Asp Ala Ile Arg Ile Leu Gly Gly Gly Gly Gly Ser Gly Gly Gly
                565                 570                 575

Gly Ser Met Arg Asn Phe Ile Ile Asn Asn Val Ile Ser Asn Lys Arg
            580                 585                 590

Phe His Tyr Leu Ile Arg Tyr Gly Asp Pro Ala His Leu His Glu Ile
        595                 600                 605

Ala Lys Asn Glu Ala Val Val Lys Phe Val Leu Gly Arg Ile Ala Asp
    610                 615                 620

Ile Gln Lys Lys Gln Gly Gln Asn Gly Lys Asn Gln Ile Asp Arg Tyr
625                 630                 635                 640

Tyr Glu Thr Cys Ile Gly Lys Asp Lys Gly Lys Ser Val Ser Glu Lys
                645                 650                 655

Val Asp Ala Leu Thr Lys Ile Ile Thr Gly Met Asn Tyr Asp Gln Phe
```

660 665 670

Asp Lys Lys Arg Ser Val Ile Glu Asp Thr Gly Arg Glu Asn Ala Glu
675 680 685

Arg Glu Lys Phe Lys Lys Ile Ile Ser Leu Tyr Leu Thr Val Ile Tyr
690 695 700

His Ile Leu Lys Asn Ile Val Asn Ile Asn Ala Arg Tyr Val Ile Gly
705 710 715 720

Phe His Cys Val Glu Arg Asp Ala Gln Leu Tyr Lys Glu Lys Gly Tyr
725 730 735

Asp Ile Asn Leu Lys Lys Leu Glu Glu Lys Gly Phe Ser Ser Val Thr
740 745 750

Lys Leu Cys Ala Gly Ile Asp Glu Thr Ala Pro Asp Lys Arg Lys Asp
755 760 765

Val Glu Lys Glu Met Ala Glu Arg Ala Lys Glu Ser Ile Asp Ser Leu
770 775 780

Glu Ser Ala Asn Pro Lys Leu Tyr Ala Asn Tyr Ile Lys Tyr Ser Asp
785 790 795 800

Glu Lys Lys Ala Glu Glu Phe Thr Arg Gln Ile Asn Arg Glu Lys Ala
805 810 815

Lys Thr Ala Leu Asn Ala Tyr Leu Arg Asn Thr Lys Trp Asn Val Ile
820 825 830

Ile Arg Glu Asp Leu Leu Arg Ile Asp Asn Lys Thr Cys Thr Leu Phe
835 840 845

Arg Asn Lys Ala Val His Leu Glu Val Ala Arg Tyr Val His Ala Tyr
850 855 860

Ile Asn Asp Ile Ala Glu Val Asn Ser Tyr Phe Gln Leu Tyr His Tyr

865 870 875 880

Ile Met Gln Arg Ile Ile Met Asn Glu Arg Tyr Glu Lys Ser Ser Gly
885 890 895

Lys Val Ser Glu Tyr Phe Asp Ala Val Asn Asp Glu Lys Lys Tyr Asn
900 905 910

Asp Arg Leu Leu Lys Leu Leu Cys Val Pro Phe Gly Tyr Cys Ile Pro
915 920 925

Arg Phe Lys Asn Leu Ser Ile Glu Ala Leu Phe Asp Arg Asn Glu Gly
930 935 940

Ser Gly Pro Lys Lys Lys Arg Lys Val Ala Ala Ala Tyr Pro Tyr Asp
945 950 955 960

Val Pro Asp Tyr Ala
965

<210> 283
<211> 963
<212> PRT
<213> Artificial Sequence

<220>
<223> Mutated Cas13d sequence, NLS-RfxCas13d(del5.13)-NLS-HA

<400> 283

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

Lys Ser Phe Ala Lys Gly Met Gly Val Lys Ser Thr Leu Val Ser Gly
20 25 30

Ser Lys Val Tyr Met Thr Thr Phe Ala Glu Gly Ser Asp Ala Arg Leu
35 40 45

Glu Lys Ile Val Glu Gly Asp Ser Ile Arg Ser Val Asn Glu Gly Glu
50 55 60

```
Ala Phe Ser Ala Glu Met Ala Asp Lys Asn Ala Gly Tyr Lys Ile Gly
65                  70                  75                  80

Asn Ala Lys Phe Ser His Pro Lys Gly Tyr Ala Val Val Ala Asn Asn
                85                  90                  95

Pro Leu Tyr Thr Gly Pro Val Gln Gln Asp Met Leu Gly Leu Lys Glu
            100                 105                 110

Thr Leu Glu Lys Arg Tyr Phe Gly Glu Ser Ala Asp Gly Asn Asp Asn
        115                 120                 125

Ile Cys Ile Gln Val Ile His Asn Ile Leu Asp Ile Glu Lys Ile Leu
    130                 135                 140

Ala Glu Tyr Ile Thr Asn Ala Ala Tyr Ala Val Asn Asn Ile Ser Gly
145                 150                 155                 160

Leu Asp Lys Asp Ile Ile Gly Phe Gly Lys Phe Ser Thr Val Tyr Thr
                165                 170                 175

Tyr Asp Glu Phe Lys Asp Pro Glu His His Arg Ala Ala Phe Asn Asn
            180                 185                 190

Asn Asp Lys Leu Ile Asn Ala Ile Lys Ala Gln Tyr Asp Glu Phe Asp
        195                 200                 205

Asn Phe Leu Asp Asn Pro Arg Leu Gly Tyr Phe Gly Gln Ala Phe Phe
    210                 215                 220

Ser Lys Glu Gly Arg Asn Tyr Ile Ile Asn Tyr Gly Asn Glu Cys Tyr
225                 230                 235                 240

Asp Ile Leu Ala Leu Leu Ser Gly Leu Arg His Trp Val Val His Asn
                245                 250                 255

Asn Glu Glu Glu Ser Arg Ile Ser Arg Thr Trp Leu Tyr Asn Leu Asp
            260                 265                 270
```

```
Lys Asn Leu Asp Asn Glu Tyr Ile Ser Thr Leu Asn Tyr Leu Tyr Asp
        275                 280                 285

Arg Ile Thr Asn Glu Leu Thr Asn Ser Phe Ser Lys Asn Ser Ala Ala
    290                 295                 300

Asn Val Asn Tyr Ile Ala Glu Thr Leu Gly Ile Asn Pro Ala Glu Phe
305                 310                 315                 320

Ala Glu Gln Tyr Phe Arg Phe Ser Ile Met Lys Glu Gln Lys Asn Leu
                325                 330                 335

Gly Phe Asn Ile Thr Lys Leu Arg Glu Val Met Leu Asp Arg Lys Asp
            340                 345                 350

Met Ser Glu Ile Arg Lys Asn His Lys Val Phe Asp Ser Ile Arg Thr
        355                 360                 365

Lys Val Tyr Thr Met Met Asp Phe Val Ile Tyr Arg Tyr Tyr Ile Glu
    370                 375                 380

Glu Asp Ser Glu Lys Asp Ile Phe Val Ile Asn Leu Arg Gly Ser Phe
385                 390                 395                 400

Asn Asp Asp Gln Lys Asp Ala Leu Tyr Tyr Asp Glu Ala Asn Arg Ile
                405                 410                 415

Trp Arg Lys Leu Glu Asn Ile Met His Asn Ile Lys Glu Phe Arg Gly
            420                 425                 430

Asn Lys Thr Arg Glu Tyr Lys Lys Lys Asp Ala Pro Arg Leu Pro Arg
        435                 440                 445

Ile Leu Pro Ala Gly Arg Asp Val Ser Ala Phe Ser Lys Leu Met Tyr
    450                 455                 460

Ala Leu Thr Met Phe Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr
465                 470                 475                 480
```

```
Thr Leu Ile Asn Lys Phe Asp Asn Ile Gln Ser Phe Leu Lys Val Met
                485                 490                 495

Pro Leu Ile Gly Val Asn Ala Lys Phe Val Glu Glu Tyr Ala Phe Phe
            500                 505                 510

Lys Asp Ser Ala Lys Ile Ala Asp Glu Leu Arg Leu Ile Lys Ser Phe
        515                 520                 525

Ala Arg Met Gly Glu Pro Ile Ala Asp Ala Arg Arg Ala Met Tyr Ile
    530                 535                 540

Asp Ala Ile Arg Ile Leu Gly Gly Gly Gly Gly Ser Gly Gly Gly Gly
545                 550                 555                 560

Ser Met Arg Asn Phe Ile Ile Asn Asn Val Ile Ser Asn Lys Arg Phe
                565                 570                 575

His Tyr Leu Ile Arg Tyr Gly Asp Pro Ala His Leu His Glu Ile Ala
            580                 585                 590

Lys Asn Glu Ala Val Val Lys Phe Val Leu Gly Arg Ile Ala Asp Ile
        595                 600                 605

Gln Lys Lys Gln Gly Gln Asn Gly Lys Asn Gln Ile Asp Arg Tyr Tyr
    610                 615                 620

Glu Thr Cys Ile Gly Lys Asp Lys Gly Lys Ser Val Ser Glu Lys Val
625                 630                 635                 640

Asp Ala Leu Thr Lys Ile Ile Thr Gly Met Asn Tyr Asp Gln Phe Asp
                645                 650                 655

Lys Lys Arg Ser Val Ile Glu Asp Thr Gly Arg Glu Asn Ala Glu Arg
            660                 665                 670

Glu Lys Phe Lys Lys Ile Ile Ser Leu Tyr Leu Thr Val Ile Tyr His
        675                 680                 685
```

```
Ile Leu Lys Asn Ile Val Asn Ile Asn Ala Arg Tyr Val Ile Gly Phe
    690                 695                 700

His Cys Val Glu Arg Asp Ala Gln Leu Tyr Lys Glu Lys Gly Tyr Asp
705                 710                 715                 720

Ile Asn Leu Lys Lys Leu Glu Glu Lys Gly Phe Ser Ser Val Thr Lys
                725                 730                 735

Leu Cys Ala Gly Ile Asp Glu Thr Ala Pro Asp Lys Arg Lys Asp Val
            740                 745                 750

Glu Lys Glu Met Ala Glu Arg Ala Lys Glu Ser Ile Asp Ser Leu Glu
        755                 760                 765

Ser Ala Asn Pro Lys Leu Tyr Ala Asn Tyr Ile Lys Tyr Ser Asp Glu
    770                 775                 780

Lys Lys Ala Glu Glu Phe Thr Arg Gln Ile Asn Arg Glu Lys Ala Lys
785                 790                 795                 800

Thr Ala Leu Asn Ala Tyr Leu Arg Asn Thr Lys Trp Asn Val Ile Ile
                805                 810                 815

Arg Glu Asp Leu Leu Arg Ile Asp Asn Lys Thr Cys Thr Leu Phe Arg
            820                 825                 830

Asn Lys Ala Val His Leu Glu Val Ala Arg Tyr Val His Ala Tyr Ile
        835                 840                 845

Asn Asp Ile Ala Glu Val Asn Ser Tyr Phe Gln Leu Tyr His Tyr Ile
    850                 855                 860

Met Gln Arg Ile Ile Met Asn Glu Arg Tyr Glu Lys Ser Ser Gly Lys
865                 870                 875                 880

Val Ser Glu Tyr Phe Asp Ala Val Asn Asp Glu Lys Lys Tyr Asn Asp
                885                 890                 895
```

Arg Leu Leu Lys Leu Leu Cys Val Pro Phe Gly Tyr Cys Ile Pro Arg
900 905 910

Phe Lys Asn Leu Ser Ile Glu Ala Leu Phe Asp Arg Asn Glu Ala Ala
915 920 925

Lys Phe Asp Lys Glu Lys Lys Lys Val Ser Gly Asn Ser Gly Ser Gly
930 935 940

Pro Lys Lys Lys Arg Lys Val Ala Ala Ala Tyr Pro Tyr Asp Val Pro
945 950 955 960

Asp Tyr Ala

<210> 284
<211> 948
<212> PRT
<213> Artificial Sequence

<220>
<223> Mutated Cas13d sequence, NLS-RfxCas13d(del5.12+5.13)-NLS-HA

<400> 284

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

Lys Ser Phe Ala Lys Gly Met Gly Val Lys Ser Thr Leu Val Ser Gly
20 25 30

Ser Lys Val Tyr Met Thr Thr Phe Ala Glu Gly Ser Asp Ala Arg Leu
35 40 45

Glu Lys Ile Val Glu Gly Asp Ser Ile Arg Ser Val Asn Glu Gly Glu
50 55 60

Ala Phe Ser Ala Glu Met Ala Asp Lys Asn Ala Gly Tyr Lys Ile Gly
65 70 75 80

Asn Ala Lys Phe Ser His Pro Lys Gly Tyr Ala Val Val Ala Asn Asn

```
                85                  90                  95

Pro Leu Tyr Thr Gly Pro Val Gln Gln Asp Met Leu Gly Leu Lys Glu
            100                 105                 110

Thr Leu Glu Lys Arg Tyr Phe Gly Glu Ser Ala Asp Gly Asn Asp Asn
        115                 120                 125

Ile Cys Ile Gln Val Ile His Asn Ile Leu Asp Ile Glu Lys Ile Leu
    130                 135                 140

Ala Glu Tyr Ile Thr Asn Ala Ala Tyr Ala Val Asn Asn Ile Ser Gly
145                 150                 155                 160

Leu Asp Lys Asp Ile Ile Gly Phe Gly Lys Phe Ser Thr Val Tyr Thr
                165                 170                 175

Tyr Asp Glu Phe Lys Asp Pro Glu His His Arg Ala Ala Phe Asn Asn
            180                 185                 190

Asn Asp Lys Leu Ile Asn Ala Ile Lys Ala Gln Tyr Asp Glu Phe Asp
        195                 200                 205

Asn Phe Leu Asp Asn Pro Arg Leu Gly Tyr Phe Gly Gln Ala Phe Phe
    210                 215                 220

Ser Lys Glu Gly Arg Asn Tyr Ile Ile Asn Tyr Gly Asn Glu Cys Tyr
225                 230                 235                 240

Asp Ile Leu Ala Leu Leu Ser Gly Leu Arg His Trp Val Val His Asn
                245                 250                 255

Asn Glu Glu Glu Ser Arg Ile Ser Arg Thr Trp Leu Tyr Asn Leu Asp
            260                 265                 270

Lys Asn Leu Asp Asn Glu Tyr Ile Ser Thr Leu Asn Tyr Leu Tyr Asp
        275                 280                 285

Arg Ile Thr Asn Glu Leu Thr Asn Ser Phe Ser Lys Asn Ser Ala Ala
```

```
    290                 295                 300

Asn Val Asn Tyr Ile Ala Glu Thr Leu Gly Ile Asn Pro Ala Glu Phe
305                 310                 315                 320

Ala Glu Gln Tyr Phe Arg Phe Ser Ile Met Lys Glu Gln Lys Asn Leu
                325                 330                 335

Gly Phe Asn Ile Thr Lys Leu Arg Glu Val Met Leu Asp Arg Lys Asp
            340                 345                 350

Met Ser Glu Ile Arg Lys Asn His Lys Val Phe Asp Ser Ile Arg Thr
        355                 360                 365

Lys Val Tyr Thr Met Met Asp Phe Val Ile Tyr Arg Tyr Tyr Ile Glu
    370                 375                 380

Glu Asp Ser Glu Lys Asp Ile Phe Val Ile Asn Leu Arg Gly Ser Phe
385                 390                 395                 400

Asn Asp Asp Gln Lys Asp Ala Leu Tyr Tyr Asp Glu Ala Asn Arg Ile
                405                 410                 415

Trp Arg Lys Leu Glu Asn Ile Met His Asn Ile Lys Glu Phe Arg Gly
            420                 425                 430

Asn Lys Thr Arg Glu Tyr Lys Lys Lys Asp Ala Pro Arg Leu Pro Arg
        435                 440                 445

Ile Leu Pro Ala Gly Arg Asp Val Ser Ala Phe Ser Lys Leu Met Tyr
    450                 455                 460

Ala Leu Thr Met Phe Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr
465                 470                 475                 480

Thr Leu Ile Asn Lys Phe Asp Asn Ile Gln Ser Phe Leu Lys Val Met
                485                 490                 495

Pro Leu Ile Gly Val Asn Ala Lys Phe Val Glu Glu Tyr Ala Phe Phe
```

500 505 510

Lys Asp Ser Ala Lys Ile Ala Asp Glu Leu Arg Leu Ile Lys Ser Phe
515 520 525

Ala Arg Met Gly Glu Pro Ile Ala Asp Ala Arg Arg Ala Met Tyr Ile
530 535 540

Asp Ala Ile Arg Ile Leu Gly Gly Gly Gly Gly Ser Gly Gly Gly Gly
545 550 555 560

Ser Met Arg Asn Phe Ile Ile Asn Asn Val Ile Ser Asn Lys Arg Phe
565 570 575

His Tyr Leu Ile Arg Tyr Gly Asp Pro Ala His Leu His Glu Ile Ala
580 585 590

Lys Asn Glu Ala Val Val Lys Phe Val Leu Gly Arg Ile Ala Asp Ile
595 600 605

Gln Lys Lys Gln Gly Gln Asn Gly Lys Asn Gln Ile Asp Arg Tyr Tyr
610 615 620

Glu Thr Cys Ile Gly Lys Asp Lys Gly Lys Ser Val Ser Glu Lys Val
625 630 635 640

Asp Ala Leu Thr Lys Ile Ile Thr Gly Met Asn Tyr Asp Gln Phe Asp
645 650 655

Lys Lys Arg Ser Val Ile Glu Asp Thr Gly Arg Glu Asn Ala Glu Arg
660 665 670

Glu Lys Phe Lys Lys Ile Ile Ser Leu Tyr Leu Thr Val Ile Tyr His
675 680 685

Ile Leu Lys Asn Ile Val Asn Ile Asn Ala Arg Tyr Val Ile Gly Phe
690 695 700

His Cys Val Glu Arg Asp Ala Gln Leu Tyr Lys Glu Lys Gly Tyr Asp

705 710 715 720

Ile Asn Leu Lys Lys Leu Glu Glu Lys Gly Phe Ser Ser Val Thr Lys
725 730 735

Leu Cys Ala Gly Ile Asp Glu Thr Ala Pro Asp Lys Arg Lys Asp Val
740 745 750

Glu Lys Glu Met Ala Glu Arg Ala Lys Glu Ser Ile Asp Ser Leu Glu
755 760 765

Ser Ala Asn Pro Lys Leu Tyr Ala Asn Tyr Ile Lys Tyr Ser Asp Glu
770 775 780

Lys Lys Ala Glu Glu Phe Thr Arg Gln Ile Asn Arg Glu Lys Ala Lys
785 790 795 800

Thr Ala Leu Asn Ala Tyr Leu Arg Asn Thr Lys Trp Asn Val Ile Ile
805 810 815

Arg Glu Asp Leu Leu Arg Ile Asp Asn Lys Thr Cys Thr Leu Phe Arg
820 825 830

Asn Lys Ala Val His Leu Glu Val Ala Arg Tyr Val His Ala Tyr Ile
835 840 845

Asn Asp Ile Ala Glu Val Asn Ser Tyr Phe Gln Leu Tyr His Tyr Ile
850 855 860

Met Gln Arg Ile Ile Met Asn Glu Arg Tyr Glu Lys Ser Ser Gly Lys
865 870 875 880

Val Ser Glu Tyr Phe Asp Ala Val Asn Asp Glu Lys Lys Tyr Asn Asp
885 890 895

Arg Leu Leu Lys Leu Leu Cys Val Pro Phe Gly Tyr Cys Ile Pro Arg
900 905 910

Phe Lys Asn Leu Ser Ile Glu Ala Leu Phe Asp Arg Asn Glu Gly Ser

915 920 925

Gly Pro Lys Lys Lys Arg Lys Val Ala Ala Ala Tyr Pro Tyr Asp Val
930 935 940

Pro Asp Tyr Ala
945

<210> 285
<211> 983
<212> PRT
<213> Artificial Sequence

<220>
<223> Mutated Cas13d sequence, NLS-RfxCas13d(del13)-NLS-HA

<400> 285

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

Lys Ser Phe Ala Lys Gly Met Gly Val Lys Ser Thr Leu Val Ser Gly
20 25 30

Ser Lys Val Tyr Met Thr Thr Phe Ala Glu Gly Ser Asp Ala Arg Leu
35 40 45

Glu Lys Ile Val Glu Gly Asp Ser Ile Arg Ser Val Asn Glu Gly Glu
50 55 60

Ala Phe Ser Ala Glu Met Ala Asp Lys Asn Ala Gly Tyr Lys Ile Gly
65 70 75 80

Asn Ala Lys Phe Ser His Pro Lys Gly Tyr Ala Val Val Ala Asn Asn
85 90 95

Pro Leu Tyr Thr Gly Pro Val Gln Gln Asp Met Leu Gly Leu Lys Glu
100 105 110

Thr Leu Glu Lys Arg Tyr Phe Gly Glu Ser Ala Asp Gly Asn Asp Asn
115 120 125

```
Ile Cys Ile Gln Val Ile His Asn Ile Leu Asp Ile Glu Lys Ile Leu
    130                 135                 140

Ala Glu Tyr Ile Thr Asn Ala Ala Tyr Ala Val Asn Asn Ile Ser Gly
145                 150                 155                 160

Leu Asp Lys Asp Ile Ile Gly Phe Gly Lys Phe Ser Thr Val Tyr Thr
                165                 170                 175

Tyr Asp Glu Phe Lys Asp Pro Glu His His Arg Ala Ala Phe Asn Asn
            180                 185                 190

Asn Asp Lys Leu Ile Asn Ala Ile Lys Ala Gln Tyr Asp Glu Phe Asp
        195                 200                 205

Asn Phe Leu Asp Asn Pro Arg Leu Gly Tyr Phe Gly Gln Ala Phe Phe
    210                 215                 220

Ser Lys Glu Gly Arg Asn Tyr Ile Ile Asn Tyr Gly Asn Glu Cys Tyr
225                 230                 235                 240

Asp Ile Leu Ala Leu Leu Ser Gly Leu Arg His Trp Val Val His Asn
                245                 250                 255

Asn Glu Glu Glu Ser Arg Ile Ser Arg Thr Trp Leu Tyr Asn Leu Asp
            260                 265                 270

Lys Asn Leu Asp Asn Glu Tyr Ile Ser Thr Leu Asn Tyr Leu Tyr Asp
        275                 280                 285

Arg Ile Thr Asn Glu Leu Thr Asn Ser Phe Ser Lys Asn Ser Ala Ala
    290                 295                 300

Asn Val Asn Tyr Ile Ala Glu Thr Leu Gly Ile Asn Pro Ala Glu Phe
305                 310                 315                 320

Ala Glu Gln Tyr Phe Arg Phe Ser Ile Met Lys Glu Gln Lys Asn Leu
                325                 330                 335
```

```
Gly Phe Asn Ile Thr Lys Leu Arg Glu Val Met Leu Asp Arg Lys Asp
            340                 345                 350

Met Ser Glu Ile Arg Lys Asn His Lys Val Phe Asp Ser Ile Arg Thr
        355                 360                 365

Lys Val Tyr Thr Met Met Asp Phe Val Ile Tyr Arg Tyr Tyr Ile Glu
    370                 375                 380

Glu Asp Ser Glu Lys Asp Ile Phe Val Ile Asn Leu Arg Gly Ser Phe
385                 390                 395                 400

Asn Asp Asp Gln Lys Asp Ala Leu Tyr Tyr Asp Glu Ala Asn Arg Ile
                405                 410                 415

Trp Arg Lys Leu Glu Asn Ile Met His Asn Ile Lys Glu Phe Arg Gly
            420                 425                 430

Asn Lys Thr Arg Glu Tyr Lys Lys Lys Asp Ala Pro Arg Leu Pro Arg
        435                 440                 445

Ile Leu Pro Ala Gly Arg Asp Val Ser Ala Phe Ser Lys Leu Met Tyr
    450                 455                 460

Ala Leu Thr Met Phe Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr
465                 470                 475                 480

Thr Leu Ile Asn Lys Phe Asp Asn Ile Gln Ser Phe Leu Lys Val Met
                485                 490                 495

Pro Leu Ile Gly Val Asn Ala Lys Phe Val Glu Glu Tyr Ala Phe Phe
            500                 505                 510

Lys Asp Ser Ala Lys Ile Ala Asp Glu Leu Arg Leu Ile Lys Ser Phe
        515                 520                 525

Ala Arg Met Gly Glu Pro Ile Ala Asp Ala Arg Arg Ala Met Tyr Ile
    530                 535                 540
```

```
Asp Ala Ile Arg Ile Leu Gly Thr Asn Leu Ser Tyr Asp Glu Leu Lys
545                 550                 555                 560

Ala Leu Ala Asp Thr Phe Ser Leu Asp Glu Asn Gly Asn Lys Leu Lys
                565                 570                 575

Lys Gly Lys His Gly Met Arg Asn Phe Ile Ile Asn Asn Val Ile Ser
            580                 585                 590

Asn Lys Arg Phe His Tyr Leu Ile Arg Tyr Gly Asp Pro Ala His Leu
        595                 600                 605

His Glu Ile Ala Lys Asn Glu Ala Val Val Lys Phe Val Leu Gly Arg
    610                 615                 620

Ile Ala Asp Ile Gln Lys Lys Gln Gly Gln Asn Gly Lys Asn Gln Ile
625                 630                 635                 640

Asp Arg Tyr Tyr Glu Thr Cys Ile Gly Lys Asp Lys Gly Lys Ser Val
                645                 650                 655

Ser Glu Lys Val Asp Ala Leu Thr Lys Ile Ile Thr Gly Met Asn Tyr
            660                 665                 670

Asp Gln Phe Asp Lys Lys Arg Ser Val Ile Glu Asp Thr Gly Arg Glu
        675                 680                 685

Asn Ala Glu Arg Glu Lys Phe Lys Lys Ile Ile Ser Leu Tyr Leu Thr
    690                 695                 700

Val Ile Tyr His Ile Leu Lys Asn Ile Val Asn Ile Asn Ala Arg Tyr
705                 710                 715                 720

Val Ile Gly Phe His Cys Val Glu Arg Asp Ala Gln Leu Tyr Lys Glu
                725                 730                 735

Lys Gly Tyr Asp Ile Asn Leu Lys Lys Leu Glu Glu Lys Gly Phe Ser
            740                 745                 750
```

```
Ser Val Thr Lys Leu Cys Ala Gly Ile Asp Glu Thr Ala Pro Asp Lys
        755                 760                 765

Arg Lys Asp Val Glu Lys Glu Met Ala Glu Arg Ala Lys Glu Ser Ile
    770                 775                 780

Asp Ser Leu Glu Ser Ala Asn Pro Lys Leu Tyr Ala Asn Tyr Ile Lys
785                 790                 795                 800

Tyr Ser Asp Glu Lys Lys Ala Glu Glu Phe Thr Arg Gln Ile Asn Arg
                805                 810                 815

Glu Lys Ala Lys Thr Ala Leu Asn Ala Tyr Leu Arg Asn Thr Lys Trp
            820                 825                 830

Asn Val Ile Ile Arg Glu Asp Leu Leu Arg Ile Asp Asn Lys Thr Cys
        835                 840                 845

Thr Leu Phe Arg Asn Lys Ala Val His Leu Glu Val Ala Arg Tyr Val
    850                 855                 860

His Ala Tyr Ile Asn Asp Ile Ala Glu Val Asn Ser Tyr Phe Gln Leu
865                 870                 875                 880

Tyr His Tyr Ile Met Gln Arg Ile Ile Met Asn Glu Arg Tyr Glu Lys
                885                 890                 895

Ser Ser Gly Lys Val Ser Glu Tyr Phe Asp Ala Val Asn Asp Glu Lys
            900                 905                 910

Lys Tyr Asn Asp Arg Leu Leu Lys Leu Leu Cys Val Pro Phe Gly Tyr
        915                 920                 925

Cys Ile Pro Arg Phe Lys Asn Leu Ser Ile Glu Ala Leu Phe Asp Arg
    930                 935                 940

Asn Glu Ala Ala Lys Phe Asp Lys Glu Lys Lys Lys Val Ser Gly Asn
945                 950                 955                 960
```

Ser Gly Ser Gly Pro Lys Lys Lys Arg Lys Val Ala Ala Ala Tyr Pro
965 970 975

Tyr Asp Val Pro Asp Tyr Ala
980

<210> 286
<211> 1378
<212> PRT
<213> Artificial Sequence

<220>
<223> Effector sequence used to edit ADAR2 expression

<400> 286

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

Thr Leu Val Ser Gly Ser Lys Val Tyr Met Thr Thr Phe Ala Glu Gly
20 25 30

Ser Asp Ala Arg Leu Glu Lys Ile Val Glu Gly Asp Ser Ile Arg Ser
35 40 45

Val Asn Glu Gly Glu Ala Phe Ser Ala Glu Met Ala Asp Lys Asn Ala
50 55 60

Gly Tyr Lys Ile Gly Asn Ala Lys Phe Ser His Pro Lys Gly Tyr Ala
65 70 75 80

Val Val Ala Asn Asn Pro Leu Tyr Thr Gly Pro Val Gln Gln Asp Met
85 90 95

Leu Gly Leu Lys Glu Thr Leu Glu Lys Arg Tyr Phe Gly Glu Ser Ala
100 105 110

Asp Gly Asn Asp Asn Ile Cys Ile Gln Val Ile His Asn Ile Leu Asp
115 120 125

Ile Glu Lys Ile Leu Ala Glu Tyr Ile Thr Asn Ala Ala Tyr Ala Val

130 135 140

Asn Asn Ile Ser Gly Leu Asp Lys Asp Ile Ile Gly Phe Gly Lys Phe
145 150 155 160

Ser Thr Val Tyr Thr Tyr Asp Glu Phe Lys Asp Pro Glu His His Arg
165 170 175

Ala Ala Phe Asn Asn Asn Asp Lys Leu Ile Asn Ala Ile Lys Ala Gln
180 185 190

Tyr Asp Glu Phe Asp Asn Phe Leu Asp Asn Pro Arg Leu Gly Tyr Phe
195 200 205

Gly Gln Ala Phe Phe Ser Lys Glu Gly Arg Asn Tyr Ile Ile Asn Tyr
210 215 220

Gly Asn Glu Cys Tyr Asp Ile Leu Ala Leu Leu Ser Gly Leu Arg His
225 230 235 240

Trp Val Val His Asn Asn Glu Glu Glu Ser Arg Ile Ser Arg Thr Trp
245 250 255

Leu Tyr Asn Leu Asp Lys Asn Leu Asp Asn Glu Tyr Ile Ser Thr Leu
260 265 270

Asn Tyr Leu Tyr Asp Arg Ile Thr Asn Glu Leu Thr Asn Ser Phe Ser
275 280 285

Lys Asn Ser Ala Ala Asn Val Asn Tyr Ile Ala Glu Thr Leu Gly Ile
290 295 300

Asn Pro Ala Glu Phe Ala Glu Gln Tyr Phe Arg Phe Ser Ile Met Lys
305 310 315 320

Glu Gln Lys Asn Leu Gly Phe Asn Ile Thr Lys Leu Arg Glu Val Met
325 330 335

Leu Asp Arg Lys Asp Met Ser Glu Ile Arg Lys Asn His Lys Val Phe

340 345 350

Asp Ser Ile Arg Thr Lys Val Tyr Thr Met Met Asp Phe Val Ile Tyr
355 360 365

Arg Tyr Tyr Ile Glu Glu Asp Ala Lys Val Ala Ala Ala Asn Lys Ser
370 375 380

Leu Pro Asp Asn Glu Lys Ser Leu Ser Glu Lys Asp Ile Phe Val Ile
385 390 395 400

Asn Leu Arg Gly Ser Phe Asn Asp Asp Gln Lys Asp Ala Leu Tyr Tyr
405 410 415

Asp Glu Ala Asn Arg Ile Trp Arg Lys Leu Glu Asn Ile Met His Asn
420 425 430

Ile Lys Glu Phe Arg Gly Asn Lys Thr Arg Glu Tyr Lys Lys Lys Asp
435 440 445

Ala Pro Arg Leu Pro Arg Ile Leu Pro Ala Gly Arg Asp Val Ser Ala
450 455 460

Phe Ser Lys Leu Met Tyr Ala Leu Thr Met Phe Leu Asp Gly Lys Glu
465 470 475 480

Ile Asn Asp Leu Leu Thr Thr Leu Ile Asn Lys Phe Asp Asn Ile Gln
485 490 495

Ser Phe Leu Lys Val Met Pro Leu Ile Gly Val Asn Ala Lys Phe Val
500 505 510

Glu Glu Tyr Ala Phe Phe Lys Asp Ser Ala Lys Ile Ala Asp Glu Leu
515 520 525

Arg Leu Ile Lys Ser Phe Ala Arg Met Gly Glu Pro Ile Ala Asp Ala
530 535 540

Arg Arg Ala Met Tyr Ile Asp Ala Ile Arg Ile Leu Gly Thr Asn Leu

545 550 555 560

Ser Tyr Asp Glu Leu Lys Ala Leu Ala Asp Thr Phe Ser Leu Asp Glu
565 570 575

Asn Gly Asn Lys Leu Lys Lys Gly Lys His Gly Met Arg Asn Phe Ile
580 585 590

Ile Asn Asn Val Ile Ser Asn Lys Arg Phe His Tyr Leu Ile Arg Tyr
595 600 605

Gly Asp Pro Ala His Leu His Glu Ile Ala Lys Asn Glu Ala Val Val
610 615 620

Lys Phe Val Leu Gly Arg Ile Ala Asp Ile Gln Lys Lys Gln Gly Gln
625 630 635 640

Asn Gly Lys Asn Gln Ile Asp Arg Tyr Tyr Glu Thr Cys Ile Gly Lys
645 650 655

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

Ile Thr Gly Met Asn Tyr Asp Gln Phe Asp Lys Lys Arg Ser Val Ile
675 680 685

Glu Asp Thr Gly Arg Glu Asn Ala Glu Arg Glu Lys Phe Lys Lys Ile
690 695 700

Ile Ser Leu Tyr Leu Thr Val Ile Tyr His Ile Leu Lys Asn Ile Val
705 710 715 720

Asn Ile Asn Ala Arg Tyr Val Ile Gly Phe His Cys Val Glu Arg Asp
725 730 735

Ala Gln Leu Tyr Lys Glu Lys Gly Tyr Asp Ile Asn Leu Lys Lys Leu
740 745 750

Glu Glu Lys Gly Phe Ser Ser Val Thr Lys Leu Cys Ala Gly Ile Asp

```
        755                 760                 765

Glu Thr Ala Pro Asp Lys Arg Lys Asp Val Glu Lys Glu Met Ala Glu
    770                 775                 780

Arg Ala Lys Glu Ser Ile Asp Ser Leu Glu Ser Ala Asn Pro Lys Leu
785                 790                 795                 800

Tyr Ala Asn Tyr Ile Lys Tyr Ser Asp Glu Lys Lys Ala Glu Glu Phe
                805                 810                 815

Thr Arg Gln Ile Asn Arg Glu Lys Ala Lys Thr Ala Leu Asn Ala Tyr
            820                 825                 830

Leu Arg Asn Thr Lys Trp Asn Val Ile Ile Arg Glu Asp Leu Leu Arg
        835                 840                 845

Ile Asp Asn Lys Thr Cys Thr Leu Phe Arg Asn Lys Ala Val Ala Leu
    850                 855                 860

Glu Val Ala Arg Tyr Val His Ala Tyr Ile Asn Asp Ile Ala Glu Val
865                 870                 875                 880

Asn Ser Tyr Phe Gln Leu Tyr His Tyr Ile Met Gln Arg Ile Ile Met
                885                 890                 895

Asn Glu Arg Tyr Glu Lys Ser Ser Gly Lys Val Ser Glu Tyr Phe Asp
            900                 905                 910

Ala Val Asn Asp Glu Lys Lys Tyr Asn Asp Arg Leu Leu Lys Leu Leu
        915                 920                 925

Cys Val Pro Phe Gly Tyr Cys Ile Pro Arg Phe Lys Asn Leu Ser Ile
    930                 935                 940

Glu Ala Leu Phe Asp Arg Asn Glu Ala Ala Lys Phe Asp Lys Glu Lys
945                 950                 955                 960

Lys Lys Val Ser Gly Asn Ser Ser Arg Gly Ser Gly Pro Lys Lys Lys
```

965 970 975

Arg Lys Val Ala Ala Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Gly
980 985 990

Ser Gln Leu His Leu Pro Gln Val Leu Ala Asp Ala Val Ser Arg Leu
995 1000 1005

Val Leu Gly Lys Phe Gly Asp Leu Thr Asp Asn Phe Ser Ser Pro
1010 1015 1020

His Ala Arg Arg Lys Val Leu Ala Gly Val Val Met Thr Thr Gly
1025 1030 1035

Thr Asp Val Lys Asp Ala Lys Val Ile Ser Val Ser Thr Gly Thr
1040 1045 1050

Lys Cys Ile Asn Gly Glu Tyr Met Ser Asp Arg Gly Leu Ala Leu
1055 1060 1065

Asn Asp Cys His Ala Glu Ile Ile Ser Arg Arg Ser Leu Leu Arg
1070 1075 1080

Phe Leu Tyr Thr Gln Leu Glu Leu Tyr Leu Asn Asn Lys Asp Asp
1085 1090 1095

Gln Lys Arg Ser Ile Phe Gln Lys Ser Glu Arg Gly Gly Phe Arg
1100 1105 1110

Leu Lys Glu Asn Val Gln Phe His Leu Tyr Ile Ser Thr Ser Pro
1115 1120 1125

Cys Gly Asp Ala Arg Ile Phe Ser Pro His Glu Pro Ile Leu Glu
1130 1135 1140

Glu Pro Ala Asp Arg His Pro Asn Arg Lys Ala Arg Gly Gln Leu
1145 1150 1155

Arg Thr Lys Ile Glu Ser Gly Gln Gly Thr Ile Pro Val Arg Ser

1160 1165 1170

Asn Ala Ser Ile Gln Thr Trp Asp Gly Val Leu Gln Gly Glu Arg
1175 1180 1185

Leu Leu Thr Met Ser Cys Ser Asp Lys Ile Ala Arg Trp Asn Val
1190 1195 1200

Val Gly Ile Gln Gly Ser Leu Leu Ser Ile Phe Val Glu Pro Ile
1205 1210 1215

Tyr Phe Ser Ser Ile Ile Leu Gly Ser Leu Tyr His Gly Asp His
1220 1225 1230

Leu Ser Arg Ala Met Tyr Gln Arg Ile Ser Asn Ile Glu Asp Leu
1235 1240 1245

Pro Pro Leu Tyr Thr Leu Asn Lys Pro Leu Leu Ser Gly Ile Ser
1250 1255 1260

Asn Ala Glu Ala Arg Gln Pro Gly Lys Ala Pro Asn Phe Ser Val
1265 1270 1275

Asn Trp Thr Val Gly Asp Ser Ala Ile Glu Val Ile Asn Ala Thr
1280 1285 1290

Thr Gly Lys Asp Glu Leu Gly Arg Ala Ser Arg Leu Cys Lys His
1295 1300 1305

Ala Leu Tyr Cys Arg Trp Met Arg Val His Gly Lys Val Pro Ser
1310 1315 1320

His Leu Leu Arg Ser Lys Ile Thr Lys Pro Asn Val Tyr His Glu
1325 1330 1335

Ser Lys Leu Ala Ala Lys Glu Tyr Gln Ala Ala Lys Ala Arg Leu
1340 1345 1350

Phe Thr Ala Phe Ile Lys Ala Gly Leu Gly Ala Trp Val Glu Lys

1355 1360 1365

Pro Thr Glu Gln Asp Gln Phe Ser Leu Thr
1370 1375

<210> 287
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<223> HIV NES

<400> 287

Leu Gln Leu Pro Pro Leu Glu Arg Leu Thr Leu
1 5 10

<210> 288
<211> 26
<212> PRT
<213> Artificial Sequence

<220>
<223> Cas sequence

<220>
<221> MISC_FEATURE
<222> (2)..(2)
<223> Xaa can be Asp or Glu

<220>
<221> misc_feature
<222> (8)..(8)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (9)..(9)
<223> Xaa can be Leu or Met

<220>
<221> misc_feature
<222> (10)..(10)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE

<222> (11)..(11)
<223> Xaa can be Cys, Ser, or Thr

<220>
<221> MISC_FEATURE
<222> (12)..(12)
<223> Xaa can be Ala, Asn, Ser, or Thr

<220>
<221> misc_feature
<222> (15)..(15)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (18)..(18)
<223> Xaa can be Asp or Glu

<220>
<221> MISC_FEATURE
<222> (19)..(19)
<223> Xaa can be Gly or Asn

<220>
<221> misc_feature
<222> (21)..(21)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (22)..(22)
<223> Xaa can be Ala, Asp, or Ser

<220>
<221> MISC_FEATURE
<222> (23)..(23)
<223> Xaa can be Phe, Ile, or Leu

<220>
<221> misc_feature
<222> (24)..(24)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (25)..(25)
<223> Xaa can be Asp, Glu, His, or Lys

<220>
<221> MISC_FEATURE
<222> (26)..(26)
<223> Xaa can be Ala, Ile, Leu, or Val

<400> 288

Leu Xaa Gly Lys Glu Ile Asn Xaa Xaa Xaa Xaa Xaa Leu Ile Xaa Lys
1 5 10 15

Phe Xaa Xaa Ile Xaa Xaa Xaa Xaa Xaa Xaa
20 25

<210> 289
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Cas sequence

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> Xaa can be Phe or Tyr

<220>
<221> MISC_FEATURE
<222> (5)..(5)
<223> Xaa can be Ala, Ile, or Leu

<220>
<221> misc_feature
<222> (6)..(6)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (9)..(9)
<223> Xaa can be Ile, Leu, or Val

<220>
<221> misc_feature
<222> (10)..(10)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (11)..(11)
<223> Xaa can be Asn or Ser

<220>
<221> misc_feature

<222> (12)..(13)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (15)..(15)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (17)..(17)
<223> Xaa can be Ile, Leu, or Val

<220>
<221> MISC_FEATURE
<222> (18)..(18)
<223> Xaa can be Ala, Ile, or Val

<220>
<221> MISC_FEATURE
<222> (19)..(19)
<223> Xaa can be Lys or Arg

<220>
<221> MISC_FEATURE
<222> (20)..(20)
<223> Xaa can be His or Tyr

<220>
<221> misc_feature
<222> (21)..(22)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (23)..(23)
<223> Xaa can be Pro or Tyr

<220>
<221> misc_feature
<222> (24)..(26)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (27)..(27)
<223> Xaa can be His, Lys, or Arg

<220>
<221> misc_feature
<222> (28)..(28)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (29)..(29)
<223> Xaa can be Ile or Leu

<220>
<221> misc_feature
<222> (30)..(31)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (33)..(34)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (35)..(35)
<223> Xaa can be Ile, Leu, or Val

<220>
<221> misc_feature
<222> (36)..(37)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (38)..(38)
<223> Xaa can be Phe or Tyr

<220>
<221> misc_feature
<222> (41)..(42)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (43)..(43)
<223> Xaa can be Ile, Leu, or Met

<220>
<221> MISC_FEATURE
<222> (44)..(44)
<223> Xaa can be Ala, Gly, or Pro

<220>
<221> MISC_FEATURE
<222> (45)..(45)
<223> Xaa can be Asp or Glu

<220>

<221> misc_feature
<222> (46)..(49)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (50)..(50)
<223> Xaa can be Asn, Gln, Arg, or Ser

<400> 289

Arg Asn Xaa Ile Xaa Xaa Asn Val Xaa Xaa Xaa Xaa Xaa Phe Xaa Tyr
1 5 10 15

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Asn
20 25 30

Xaa Xaa Xaa Xaa Xaa Xaa Val Leu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
35 40 45

Xaa Xaa
50

<210> 290
<211> 33
<212> PRT
<213> Artificial Sequence

<220>
<223> Cas sequence

<220>
<221> misc_feature
<222> (2)..(2)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> Xaa can be Leu or Met

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> Xaa can be Ser or Thr

<220>

<221> MISC_FEATURE
<222> (5)..(5)
<223> Xaa can be Ile or Val

<220>
<221> misc_feature
<222> (6)..(6)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (8)..(10)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (12)..(12)
<223> Xaa can be Asn or Ser

<220>
<221> MISC_FEATURE
<222> (13)..(13)
<223> Xaa can be Ile, Leu, or Met

<220>
<221> misc_feature
<222> (15)..(15)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (16)..(16)
<223> Xaa can be Ile or Val

<220>
<221> MISC_FEATURE
<222> (18)..(18)
<223> Xaa can be Ala, Ser, or Thr

<220>
<221> MISC_FEATURE
<222> (20)..(20)
<223> Xaa can be Phe or Tyr

<220>
<221> misc_feature
<222> (21)..(22)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (23)..(23)

<223> Xaa can be Ala or Gly

<220>
<221> MISC_FEATURE
<222> (24)..(24)
<223> Xaa can be Phe or Tyr

<220>
<221> misc_feature
<222> (25)..(25)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (26)..(26)
<223> Xaa can be Ala, Cys, or Val

<220>
<221> MISC_FEATURE
<222> (27)..(27)
<223> Xaa can be Leu or Val

<220>
<221> misc_feature
<222> (31)..(32)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (33)..(33)
<223> Xaa can be Phe or Leu

<400> 290

Leu Xaa Xaa Xaa Xaa Xaa Tyr Xaa Xaa Xaa Lys Xaa Xaa Val Xaa Xaa
1   5   10   15

Asn Xaa Arg Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Glu Arg Asp Xaa Xaa
            20   25   30

Xaa

<210> 291
<211> 21
<212> PRT
<213> Artificial Sequence

<220>

<223> Cas sequence

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> Xaa can be Ala, Gly, or Ser

<220>
<221> MISC_FEATURE
<222> (5)..(5)
<223> Xaa can be Cys, Asn, or Val

<220>
<221> misc_feature
<222> (6)..(6)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (7)..(7)
<223> Xaa can be Ala or Pro

<220>
<221> MISC_FEATURE
<222> (9)..(9)
<223> Xaa can be Phe or Tyr

<220>
<221> MISC_FEATURE
<222> (13)..(13)
<223> Xaa can be Ser or Thr

<220>
<221> misc_feature
<222> (14)..(14)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (15)..(15)
<223> Xaa can be Glu or Asn

<220>
<221> misc_feature
<222> (16)..(16)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> misc_feature
<222> (19)..(19)
<223> Xaa can be any naturally occurring amino acid

<220>
<221> MISC_FEATURE
<222> (20)..(20)
<223> Xaa can be Asp, Lys, or Arg

<220>
<221> MISC_FEATURE
<222> (21)..(21)
<223> Xaa can be Asp, Glu, or Asn

<400> 291

Pro Phe Xaa Tyr Xaa Xaa Xaa Arg Xaa Lys Asn Leu Xaa Xaa Xaa Xaa
1   5   10   15

Leu Phe Xaa Xaa Xaa
            20

<210> 292
<211> 1057
<212> PRT
<213> Artificial Sequence

<220>
<223> MH_4866

<400> 292

Met Glu Glu Asn Lys Val Arg Phe Thr Asn Thr Ser Asn Val Lys Arg
1   5   10   15

Pro Lys Ser Leu Leu Lys Gln Met Gly His Lys Ser Thr Phe Leu Leu
            20   25   30

Asp Asp Thr Leu Ile Val Thr Ser Val Ala Lys Gly Lys Asn Glu Asp
        35   40   45

Gln Lys Gly Lys Glu Gln Ala Ala Ser Lys Ser Lys Val Ala Pro Gln
    50   55   60

Ser His Ile Glu Leu Leu Ser Thr Ile Asp Gly Glu Val Leu Gly Gly
65   70   75   80

Arg Lys Glu Leu Leu Ala Thr Thr Val Ser Arg Gln Glu Val Thr Ile

85 90 95

Ser Gly Lys Ile Tyr Asn Leu Asn Thr Lys Thr Pro Glu Ser Thr Thr
100 105 110

Val Thr Met Ala Asn Pro Ala Val Leu Leu Gln Lys Asn Ser Ser Thr
115 120 125

Thr Gln Val Ala Arg Lys Thr Asp Tyr Leu Gly Ile Lys Ser Ala Leu
130 135 140

Glu Lys Glu Phe Phe Asn Asp Phe Ser Asn Ser Gly Phe Asn Asp Asn
145 150 155 160

Val Arg Val Gln Ile Ala Tyr Asn Ile Leu Asp Ile Arg Lys Ile Leu
165 170 175

Ala Thr Thr Ile Asn Asn Ile Val Ser Met Phe Tyr Asn Leu Asp Arg
180 185 190

Thr Thr Gly Asp Ile Ala Asp Glu Thr Lys Lys Ser Thr Asn Thr Asp
195 200 205

Ile Ile Gly Met Ile Lys Thr Tyr Leu Lys Tyr Asp Gln Ala Lys Asn
210 215 220

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

Leu Ile Asn Thr Lys Ala Tyr His Thr Tyr Phe Asp Arg Leu Phe Thr
245 250 255

Thr Pro Thr Arg Gln Ala Phe Lys Ala Gly Lys Leu Lys Asp Thr Asn
260 265 270

Asp Lys Lys Ser Asn Leu Pro Cys Glu Phe Ser Lys Leu Ser Glu Lys
275 280 285

Asp Gln Glu Glu Phe Lys Asn Glu Ala Ile Thr His Asn Tyr Asp Val

```
    290                 295                 300

Leu Arg Leu Leu Gly Leu Gly Arg Gln Leu Val Val His Glu Lys Tyr
305                 310                 315                 320

Lys Ser Asp Glu His Asn Lys Asp Asn Ser Ile Glu Ser Leu Phe Asn
                325                 330                 335

Ile Asp Lys Ile Ile Arg Asn Ser Asn Asp Thr Met Asp Leu Leu Lys
            340                 345                 350

Arg Leu Asp Asn Ile Tyr Leu Glu Ser Ala Glu Gln Ile Asp Asn Ser
        355                 360                 365

Phe Lys Lys Ser Ser Ala Thr Asn Cys Leu Ile Leu Glu Glu Leu Tyr
    370                 375                 380

Lys Arg Ile Gly Lys Ala Lys Thr Gln Ala Glu Leu Tyr Lys Asp Tyr
385                 390                 395                 400

Tyr Arg Tyr Ala Ile Leu Lys Glu Asp Gln Cys Leu Gly Ile Ser Phe
                405                 410                 415

Asn Arg Ile Lys Glu Ala Val Leu Glu Lys Leu Lys Glu Glu Gln Arg
            420                 425                 430

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

Cys Asn Phe Val Leu Trp His Tyr Met Val Asp Leu Arg Asp Arg Asn
    450                 455                 460

Glu Thr Arg Ala Gly His Gln Gln Thr Asp Arg Asn Glu Ile Glu Glu
465                 470                 475                 480

Leu Lys Ser Lys Leu Arg Ala Cys Cys Asp Asn Asn Gly Asp Gly Gly
                485                 490                 495

Glu Asn Gln Lys Tyr Ala Ile Tyr Lys Glu Phe Val Asn Phe His Trp
```

```
            500                 505                 510

Thr Asp Leu Arg Thr Ser Ile Asn Thr Leu Tyr Asp Leu Cys Val Lys
        515                 520                 525

Tyr Lys Tyr Thr Ala Lys Asn Tyr Arg Ile Asp Ser Lys Ile Gln Ile
    530                 535                 540

Asp Asn Gln Leu Ser Ser Ser Ala Asn Asn Val Lys Thr Phe Val Lys
545                 550                 555                 560

Tyr Ile Tyr Phe Leu Ser Glu Phe Leu Asp Gly Lys Glu Val Asn Val
                565                 570                 575

Phe Phe Asn Glu Leu Val Glu Lys Phe Ser Asn Ile Ala Glu Ile Val
            580                 585                 590

Thr Thr Val Asn Asp Asn Ile Asn Asn Ser Ser Lys Lys Ser Pro Asn
        595                 600                 605

Ile Ser Phe Leu Pro Gln Tyr Ser Leu Phe Gln Glu Ala Gly Thr Val
    610                 615                 620

Ala Asp Gln Leu Arg Asn Ala Arg Asn Ile Lys His Ala Thr Leu Arg
625                 630                 635                 640

Ala Arg Ile Asn Asp Lys Val Ala Asp Lys Asp Ser Lys Met Ser Pro
                645                 650                 655

Ser Lys Gln Thr Ile Ser Asp Ala Leu Tyr Met Leu Ile Asp Glu Asn
            660                 665                 670

Asp Thr Ile Glu Asn Tyr Lys Gln Leu Ile Trp Asp Glu Lys Glu Ser
        675                 680                 685

Gly Thr Trp Val Lys Lys Ala Thr Gly Asn Asn Lys Val Arg Asn Phe
    690                 695                 700

Ile Ile Asn Asn Ile Ile Met Ser Lys Trp Phe Met Tyr Ile Ser Lys
```

705 710 715 720

Tyr Ser Asp Pro Arg Asn Cys Ser Lys Ile Met His Asn Lys Glu Ala
725 730 735

Val Val Phe Ala Met Ser Ser Ile Pro Asp Ala Gln Ile Glu Lys Tyr
740 745 750

Tyr Lys Arg Val Thr Ala Lys Ser Val Gln Tyr Gly Asn Lys Thr Ala
755 760 765

Glu Glu Met Arg Ser Ile Val Ile Asp Ala Leu Met Asn Phe Ser Leu
770 775 780

Lys Pro Val Leu Thr Gly Val Ala Asn Asn Asp Tyr Ser Asn Glu Asn
785 790 795 800

Arg Glu Arg Asp Ile Gly Ile Ile Gln Leu Tyr Leu Thr Val Ala Tyr
805 810 815

Ile Val Thr Lys Ser Leu Val Asn Val Asn Lys Arg Phe Leu Leu Ala
820 825 830

Phe Ser Cys Leu Glu Arg Asp Val Pro Leu Ile Leu Asn Gly Lys Ser
835 840 845

Phe Lys Asp Ala Asn Ile Leu Val Leu Ser Glu Lys Phe Ile Ala Glu
850 855 860

Asp Gln Glu Ile Glu Gln Lys Tyr Lys Leu Ala Lys Lys Asn Arg Ile
865 870 875 880

Thr Pro Asn Met Ser Lys Glu Asp Val Lys Asn Val Phe Lys Glu Leu
885 890 895

Asn Ala Glu Tyr Lys Ser Lys His Tyr Ala Glu Ser Leu Thr Arg Pro
900 905 910

Arg Asp Pro Asp Ser Lys Lys Asp Lys Ser Cys Ala Tyr Leu Val Asn

915 920 925

Asn Leu Asn Glu Ala Lys Lys Ile Glu Asn Val Leu Met Asp Phe Lys
930 935 940

Asn Lys Val Asp His Leu Asn Val Val Lys Ser Ala Pro Glu Tyr Met
945 950 955 960

Ser Asn Ile Lys Asn Ile Asp Ser Phe Tyr Gly Ile Phe Cys Phe Ala
965 970 975

Leu Gln Asn Ile Leu Ala Asp Lys Met Ile Lys Trp Ser Val Asn Glu
980 985 990

Lys Asn Ser Lys Glu Asp Val Glu Ile Ile Asn Ser Tyr Gly Gln Glu
995 1000 1005

Leu Lys Lys Lys Leu Asp Ser Arg His Thr Phe Asp Arg Asp Ala
1010 1015 1020

Met Trp Ile Ile Asn Met Pro Phe Ala Tyr Asn Phe Pro Arg Tyr
1025 1030 1035

Lys Asn Leu Ser Ser Glu Tyr Leu Phe Tyr Asp Lys Glu Tyr Glu
1040 1045 1050

Lys Thr Lys Lys
1055

<210> 293
<211> 945
<212> PRT
<213> Artificial Sequence

<220>
<223> 037_-_emb|OIZA01000315.1| compressed

<400> 293

Met Ala Lys Lys Lys Arg Met Thr Ala Lys Glu Arg Lys Gln Asn His
1 5 10 15

```
Arg Glu Ser Leu Met Lys Lys Ala Asp Ser Asn Ala Glu Lys Glu Lys
            20                  25                  30

Ala Lys Lys Pro Val Val Glu Asn Lys Pro Asp Thr Ala Ile Ser Lys
        35                  40                  45

Asp Asn Lys Pro Lys Pro Asn Lys Glu Ile Lys Lys Ser Lys Ala Lys
    50                  55                  60

Leu Ala Gly Val Lys Trp Ile Ile Lys Ala Asn Asp Asp Val Thr Tyr
65                  70                  75                  80

Ile Ser Ser Phe Gly Lys Gly Asn Asn Ser Val Leu Glu Lys Arg Ile
                85                  90                  95

Ile Gly Asp Val Ser Gly Asp Val Asn Lys Asp Ser His Met Tyr Val
            100                 105                 110

Asn Pro Lys Tyr Thr Lys Lys Asn Tyr Glu Ile Lys Asn Gly Phe Ser
        115                 120                 125

Ser Gly Ser Ser Leu Val Thr Tyr Pro Asn Lys Pro Asp Lys Asn Ser
    130                 135                 140

Gly Met Asp Ala Leu Cys Leu Lys Pro Tyr Phe Glu Ile Asp Phe Phe
145                 150                 155                 160

Gly His Ile Phe Thr Asp Asn Met His Ile Gln Ala Ile Tyr Asn Ile
                165                 170                 175

Phe Asp Ile Glu Lys Ile Leu Ala Lys His Ile Thr Asn Ile Ile Tyr
            180                 185                 190

Thr Val Asn Ser Phe Asp Arg Asn Tyr Asn Gln Ser Gly Asn Asp Thr
        195                 200                 205

Ile Gly Phe Asp Leu Asn Tyr Arg Val Pro Tyr Leu Glu Tyr Gly Gly
    210                 215                 220
```

```
Gly Lys Asp Ser Asn Gly Lys Pro Asn Lys Ile Ser Ala Trp Lys Lys
225                 230                 235                 240


Arg Glu Asn Phe Ile Asn Phe Tyr Asn Glu Ala Lys Pro His Leu Gly
                245                 250                 255


Tyr Tyr Glu Asn Ile Phe Tyr Asp His Gly Glu Pro Ile Ser Glu Glu
            260                 265                 270


Lys Phe Tyr Asn Tyr Leu Asn Ile Leu Asn Phe Ile Arg Asn Asn Thr
        275                 280                 285


Phe His Tyr Lys Asp Asn Asp Ile Glu Leu Tyr Ser Glu Asn Tyr Ser
    290                 295                 300


Glu Glu Tyr Val Phe Ile Asn Cys Leu Asn Glu Phe Val Lys Asn Lys
305                 310                 315                 320


Phe Lys Asn Val Asn Lys Asn Phe Ile Ser Asn Glu Lys Asn Asn Leu
                325                 330                 335


Tyr Ile Ile Leu Lys Ala Tyr Gly Glu Asp Thr Glu Asn Val Glu Val
            340                 345                 350


Val Lys Lys Tyr Ser Lys Glu Leu Tyr Lys Leu Ser Val Leu Lys Thr
        355                 360                 365


Asn Lys Asn Leu Gly Val Asn Val Lys Lys Leu Arg Glu Ser Ala Ile
    370                 375                 380


Glu Tyr Gly Tyr Cys Pro Leu Pro Tyr Asp Lys Glu Lys Glu Val Ala
385                 390                 395                 400


Lys Leu Ser Ser Val Lys His Lys Leu Tyr Lys Thr Tyr Asp Phe Val
                405                 410                 415


Ile Thr His Tyr Leu Asn Ser Asn Asp Lys Leu Leu Leu Glu Ile Val
            420                 425                 430
```

```
Glu Thr Leu Arg Leu Ser Lys Asn Asp Asp Glu Lys Glu Asn Val Tyr
        435                 440                 445

Lys Lys Tyr Ala Glu Lys Leu Phe Lys Ala Asp Asp Val Ile Asn Pro
    450                 455                 460

Ile Lys Ala Ile Ser Lys Leu Phe Ala Glu Lys Gly Asn Lys Leu Phe
465                 470                 475                 480

Lys Glu Lys Ile Ile Ile Lys Lys Glu Tyr Ile Glu Asp Val Ser Ile
                485                 490                 495

Asp Lys Asn Ile Tyr Asp Phe Thr Lys Val Ile Phe Phe Met Thr Cys
            500                 505                 510

Phe Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu Thr Asn Ile Ile Ser
        515                 520                 525

Lys Leu Gln Val Ile Glu Asp His Asn Asn Val Ile Lys Phe Ile Phe
    530                 535                 540

His Asn Lys Asp Ala Val Tyr Lys Asp Tyr Ser Asp Lys Tyr Ala Ile
545                 550                 555                 560

Phe Arg Asn Ala Gly Lys Ile Ala Thr Glu Leu Glu Ala Ile Lys Ser
                565                 570                 575

Ile Ala Arg Met Glu Asn Lys Ile Glu Asn Ala Pro Lys Glu Pro Leu
            580                 585                 590

Leu Lys Asp Ala Leu Leu Ala Leu Gly Val Ser Ser Asn Asp Phe Asp
        595                 600                 605

Glu Lys Tyr Glu Lys Tyr Phe Lys Thr Asp Val Asp Ala Asp Lys Asp
    610                 615                 620

His Gln Lys Val Ser Thr Phe Leu Met Asn Asn Val Ile Asn Asn Ser
625                 630                 635                 640
```

```
Arg Phe Lys Tyr Val Val Lys Tyr Ile Asn Pro Ala Asp Ile Asn Gly
                645                 650                 655

Leu Ala Lys Asn Arg Tyr Leu Val Lys Phe Val Leu Ser Lys Ile Pro
            660                 665                 670

Glu Glu Gln Ile Asp Ser Tyr Tyr Lys Leu Phe Ser Asn Glu Glu Glu
        675                 680                 685

Pro Ser Cys Glu Glu Lys Ile Lys Leu Leu Thr Lys Lys Ile Ser Lys
    690                 695                 700

Leu Asn Phe Gln Thr Leu Phe Glu Asn Asn Lys Ile Pro Asn Val Glu
705                 710                 715                 720

Lys Glu Arg Lys Lys Ala Ile Ile Thr Leu Tyr Phe Thr Ile Val Tyr
                725                 730                 735

Ile Leu Val Lys Asn Leu Val Asn Ile Asn Gly Leu Tyr Thr Leu Ala
            740                 745                 750

Leu Tyr Phe Val Glu Arg Asp Gly Phe Phe Tyr Lys Gln Ile Cys Glu
        755                 760                 765

Lys Lys Leu Ile Glu Thr Leu Lys Lys Lys Asp Lys Lys Gln Leu Tyr
    770                 775                 780

Asn Asp Asp Asp Tyr Leu Leu Leu Pro Glu Ile Phe Ser Gly Ser Lys
785                 790                 795                 800

Tyr Arg Glu Glu Thr Lys Asn Leu Lys Leu Pro Lys Glu Lys Asp Arg
                805                 810                 815

Asp Ile Met Lys Lys Tyr Leu Pro Asn Asp Lys Asp Arg Glu Glu Tyr
            820                 825                 830

Asn Lys Phe Phe Lys Gln Tyr Arg Asn Asn Ile Val His Leu Lys Ile
        835                 840                 845
```

Ile Ala Lys Leu Ser Glu Leu Thr Lys Asn Ile Asp Lys Asp Ile Asn
    850   855   860

Ser Tyr Phe Asp Ile Tyr His Tyr Cys Thr Gln Arg Val Met Phe Asp
865   870   875   880

Tyr Cys Lys Lys Asn Asn Asn Val Val Leu Ala Lys Met Lys Asp Leu
                885   890   895

Ala His Ile Lys Ser Asp Cys Asp Glu Phe Ser Ser Lys His Thr Tyr
            900   905   910

Pro Tyr Ser Ser Ala Val Leu Arg Phe Met Asn Leu Pro Phe Ala Tyr
        915   920   925

Asn Val Pro Arg Phe Lys Asn Leu Ser Tyr Lys Lys Phe Phe Asp Lys
    930   935   940

Gln
945

<210> 294
<211> 924
<212> PRT
<213> Artificial Sequence

<220>
<223> PIG-022_GL0026351 compressed

<400> 294

Met Ala Glu Ser Tyr Lys Ile Lys Ser Lys Lys Thr Arg Thr Lys Arg
1   5   10   15

Ala Gly Leu Lys Ser Ser Phe Val Asn Glu Asn Lys Gly Leu Met Thr
            20   25   30

Ala Phe Gly Lys Gly Asn Asn Ala Val Ile Glu Lys Arg Ile Ser Gly
        35   40   45

Asn Leu Ile Glu Asn Val Ser Asn Lys Pro Arg Phe Ser Val Glu Phe

50 55 60

Gly Asn Lys Glu Tyr Ile Ile Asn Asp Lys Lys Leu Gly Ile Val Ser
65 70 75 80

Asp Asp Pro Arg Glu Ser Gln Lys Pro Val Gly Glu Asp Gly Ile His
85 90 95

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

Asn Ile His Ile Gln Leu Ile Tyr Ser Ile Leu Asp Leu Glu Lys Ile
115 120 125

Leu Ala Val His Ser Asn Asn Ile Val Tyr Thr Phe Asp Asn Leu Arg
130 135 140

Arg Lys Glu Thr Asp Glu Tyr Asp Asp Phe Ile Gly Tyr Met Ser Thr
145 150 155 160

Val Asn Ser Tyr Glu Glu Phe Met Asn Pro Glu Lys Leu Leu Pro Gly
165 170 175

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

Leu Glu Tyr Arg Lys Asn Pro Arg Leu Gly Tyr Phe Gly Lys Val Phe
195 200 205

Tyr Glu Lys Val Thr Asp Pro Lys Asn Lys Ser Lys Lys Lys Met Arg
210 215 220

Leu Ile Ser Glu Lys Glu Met Tyr Tyr Ile Phe Ala Leu Leu Gly Gly
225 230 235 240

Leu Arg Gln Phe Cys Thr His Asp Lys Asn Arg Ala Arg Asn Trp Leu
245 250 255

Tyr Ser Leu Asp Thr Met Leu Asn Asp Asp Ala Lys Ala Val Leu Asp

```
            260                 265                 270

Lys Tyr Tyr Val Ser Thr Val Ser Ser Leu Asn Lys His Phe Val Asp
        275                 280                 285

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

Thr Asp Ile Glu Leu Lys Lys Asn Ile Ala Lys Glu Tyr Tyr Glu Phe
305                 310                 315                 320

Val Cys Arg Lys Asn Phe Lys Asn Ile Gly Phe Ser Ile Lys Lys Leu
                325                 330                 335

Arg Glu Tyr Leu Leu Glu Lys Pro Glu Phe Cys Asn Leu Lys Glu Asp
            340                 345                 350

Lys Tyr Asn Ser Val Arg Ser Lys Leu Tyr Arg Leu Ile Asp Phe Ile
        355                 360                 365

Ile Tyr Leu Asn Tyr Lys Asp Ala Lys Phe Glu Asn Asp Asn Asp Glu
    370                 375                 380

Lys Ile Glu Arg Phe Val Ser Lys Leu Arg Lys Cys Pro Asp Glu Ala
385                 390                 395                 400

Ser Lys Glu Thr Val Tyr Gln Thr Glu Ser Asp Ile Ile Lys Lys Ser
                405                 410                 415

Ile Tyr Gly Ser Ile Asn Ser Ile Lys Arg Gly Met Met Ser Val Lys
            420                 425                 430

Ser Lys Tyr Gly Lys Leu Glu Leu Asp Lys Gln Thr Ile Asn Gln Leu
        435                 440                 445

Lys Arg Ala Val Asp Asp Ile Ser Ile Asn Thr Lys Gly Asp Tyr Phe
    450                 455                 460

Thr Lys Leu Met Tyr Leu Ile Ser Leu Leu Ile Asp Gly Lys Glu Ile
```

465 470 475 480

Asn Asp Leu Leu Thr Thr Leu Ile His Gln Phe Glu Asn Ile Ala Ser
485 490 495

Phe Ile Asp Val Met Lys Lys Glu Asn Ile Asn Tyr Glu Phe Ala Glu
500 505 510

Lys Tyr Lys Phe Phe Asn Thr Ser Lys Asn Ile Ala Cys Glu Leu Arg
515 520 525

Thr Val Asn Ser Phe Ala Arg Met Thr Gly Phe Met Glu Asn Asp Glu
530 535 540

Ser Thr Lys Phe Ile Met Phe Arg Glu Ala Ala Asp Ile Leu Gly Thr
545 550 555 560

Ser Tyr Thr Asp Asp Glu Leu Thr Asp His Leu Asn Leu Lys Lys Thr
565 570 575

Arg Leu Asn Ala Arg Ser Lys Ala Asp Thr Gly Leu Arg Asn Phe Ile
580 585 590

Ile Ser Asn Val Ile Lys Ser Ser Arg Phe Lys Tyr Leu Ile Arg Tyr
595 600 605

Ser Asn Pro Lys Lys Val Lys Ala Ile Ala Glu Asn Lys Lys Leu Val
610 615 620

Arg Phe Val Leu Asp Lys Ile Pro Asp Pro Gln Ile Ile Arg Tyr Cys
625 630 635 640

Asn Thr Cys Gly Phe Lys Val Gly Glu Arg Gln Glu Asn Ile Asn Thr
645 650 655

Leu Ala Lys Ala Ile Thr Asn Phe Ser Phe Ser Ala Phe Glu Lys Val
660 665 670

Lys Gln Ser Ala Asn Val Glu Asn Asp Thr Pro Asp Ala Ile Asp Lys

675 680 685

Ala Gln Lys Gln Gly Leu Ile Ser Leu Tyr Leu Thr Val Leu Tyr Leu
690 695 700

Phe Ile Lys Asn Met Val Tyr Val Asn Ser Arg Tyr Val Leu Ala Phe
705 710 715 720

His Cys Leu Glu Arg Asp Ser Gln Leu Leu Lys Ile Thr Ser Gly Tyr
725 730 735

Asn Asn Ser Ser Tyr Asn Asn Leu Thr Asn Tyr Phe Ile Asp Asn Lys
740 745 750

Tyr Leu Asn Pro His Ala Ser Arg Tyr Leu Lys Val Asn Ile Ser Asn
755 760 765

Ala Asp Glu Trp Ser Ile Val Lys Phe Arg Asn Gly Thr Ala His Leu
770 775 780

Asn Ser Val Arg Asn Ile Asp Thr Cys Ile Glu Gly Leu Lys Asp Phe
785 790 795 800

Lys Ser Tyr Phe Glu Ile Tyr Ala Tyr Thr Val Gln Lys Ser Ile Lys
805 810 815

Lys Tyr Phe Asp Ala Glu Phe Ser Glu Asp Arg Asn Ile Glu Asp Gly
820 825 830

Lys Ile Ser Leu Cys Asn Gly Val Tyr Ile His Glu Arg Gln Lys Asn
835 840 845

Gly Lys Pro Lys Ile Thr Val Lys Ile Asn Pro Lys Thr Ile Glu Tyr
850 855 860

Met Glu Ser Val Asp Lys Tyr Gly Ser Tyr Thr Lys Asp Phe Val Lys
865 870 875 880

Ala Leu Asn Thr Pro Phe Gly Tyr Cys Leu Ala Arg Tyr Lys Asn Leu

885 890 895

Ser Ile Glu Pro Leu Phe Asp Arg Asn Thr Val Tyr Asn Lys Glu Asp
900 905 910

Val Thr Glu Glu Phe Ser Val Asn Ser Lys Glu Gln
915 920

<210> 295
<211> 1022
<212> PRT
<213> Artificial Sequence

<220>
<223> PIG-046_GL0077813 compressed

<400> 295

Met Ala Lys Lys Ile Thr Ala Lys Gln Arg Arg Glu Glu Arg Glu Arg
1 5 10 15

Gln Asn Lys Gln Lys Trp Ala Lys Lys Gln Asp Asp Ala Thr Ala Val
20 25 30

Phe Glu Cys Glu Ala Asn Ile Lys Pro Ala Asp Ser Lys Asp Glu Asp
35 40 45

Cys Ala Asn Thr Ser Ile Lys Arg Lys Lys Thr Gln Ala Lys Ala Met
50 55 60

Gly Leu Lys Thr Val Leu Gly Phe Asp Asn Lys Ile Ala Ile Ala Ser
65 70 75 80

Phe Thr Ser Ser Lys Asp Ser Lys Ser Ser His Ile Glu Arg Ile Thr
85 90 95

Asp Pro Asn Gly Lys Thr Ile Arg Glu Asp Val Arg Met Phe Asp Ser
100 105 110

Asn Val Asp Glu Cys Ser Ile Asn Leu Glu Lys Arg Met Thr Val Glu
115 120 125

```
Glu Arg Gln Lys Asp Gly Thr Ile Lys Lys Asp Glu Lys Asp Val Lys
    130                 135                 140

Ser Thr Ile Cys Asn Pro Tyr Ser Asn Glu Cys Gly Lys Asp Tyr Ile
145                 150                 155                 160

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

Asn Asp Asn Leu Arg Val Gln Ile Ala Tyr Asn Ile Phe Asp Ile Gln
            180                 185                 190

Lys Ile Phe Gly Thr Tyr Ile Asn Asn Ile Ile Tyr Ser Phe Tyr Asn
        195                 200                 205

Leu Ser Arg Asp Glu Phe Gln Ser Asp Asn Asp Val Ile Gly Thr Leu
    210                 215                 220

Tyr Met Leu Lys Asp Phe Asp Gly Gln Lys Glu Thr Asp Thr Phe Arg
225                 230                 235                 240

Gln Ala Arg Ala Leu Leu Glu Arg Thr Glu Ala Tyr Tyr Ser Tyr Phe
                245                 250                 255

Asp Asn Val Phe Lys Lys Ile Asp Lys Lys Lys Lys Arg Asn Asp Asp
            260                 265                 270

Tyr Glu Gly Lys Arg Asn Glu Ile Leu Arg Tyr Asn Phe Asn Val Leu
        275                 280                 285

Arg Val Leu Ser Phe Leu Arg Gln Ile Cys Ala His Ala Lys Val Lys
    290                 295                 300

Val Ser Ser Asp Asn Asp Arg Glu Lys Gly Gly Val Phe Val Asp Ser
305                 310                 315                 320

Leu Asp Ala Leu Phe Asn Ile Ser Ser Phe Phe Asp Ala Val Ala Pro
                325                 330                 335
```

```
Glu Leu Asn Glu Val Ile Asn Ser Val Tyr Ser Lys Gly Ile Tyr Asp
            340                 345                 350

Ile Asn Asp Asn Phe Val Lys Asn Asp Lys Asn Asn Phe Tyr Ile Leu
        355                 360                 365

Ser Lys Ile Tyr Pro Glu Val Ala Arg Glu Asn Leu Leu Lys Glu Tyr
    370                 375                 380

Tyr Tyr Phe Val Val Ser Lys Glu Gly Asn Asn Ile Gly Ile Ser Thr
385                 390                 395                 400

Lys Lys Leu Lys Glu Ala Ile Ile Ala Gln Asp Met Ser Tyr Ile Lys
                405                 410                 415

Ser Lys Asp Tyr Asp Thr Tyr Arg Asn Lys Leu Tyr Thr Val Leu Cys
            420                 425                 430

Phe Ile Leu Val Lys Glu Leu Asn Glu Arg Thr Thr Ile Arg Glu Gln
        435                 440                 445

Met Val Ala Asp Leu Arg Ala Asn Met Asn Gly Asp Ile Gly Arg Glu
    450                 455                 460

Asp Ile Tyr Ser Lys Tyr Ala Lys Ile Ile Tyr Ala Gln Val Lys Pro
465                 470                 475                 480

Arg Phe Asp Gly Met Lys Asn Val Phe Glu Glu Glu Ala Lys Gly Val
                485                 490                 495

Ile Val Pro Asp Lys Glu Glu Ala Val Lys Phe Ser Asn Gly Ala Leu
            500                 505                 510

Asp Lys Ser Glu Ile Glu Arg Phe Cys Ile Thr Ser Ala Asn Thr Asp
        515                 520                 525

Ser Val Ala Lys Ile Ile Tyr Phe Leu Cys Lys Phe Leu Asp Gly Lys
    530                 535                 540
```

```
Glu Ile Asn Glu Phe Cys Cys Ala Met Met Asn Lys Leu Asp Gly Ile
545                 550                 555                 560

Asn Asp Leu Met Glu Thr Ala Glu Gln Cys Gly Ala Lys Ile Glu Phe
                565                 570                 575

Val Asp Asn Phe Asn Val Leu Ser Lys Cys Gly Thr Ile Ser Gly Gln
            580                 585                 590

Ile Arg Val Val Lys Ser Ile Ser Lys Met Lys Lys Glu Ile Asp Val
        595                 600                 605

Asp Asn Glu Thr Ile Phe Leu Asp Ala Phe Glu Leu Leu Gly Lys Arg
    610                 615                 620

Ile Glu Lys Tyr Lys Lys Asp Asp Ser Gly Asn Val Leu Arg Asn Asp
625                 630                 635                 640

Asp Gly Lys Val Leu Arg Thr Lys Glu Tyr Asn Asn Ile Arg Lys Met
                645                 650                 655

Leu Cys Lys Lys Asp Arg Arg Leu Asn Phe Ile Ala Asn Ser Ile Ile
            660                 665                 670

Lys Asn Lys Trp Phe Ser Tyr Ile Val Arg Tyr Asn Gln Pro Ser Glu
        675                 680                 685

Cys Gln Ala Ile Met Lys Asn Lys Thr Leu Val Lys Phe Ala Leu Asp
    690                 695                 700

Glu Leu Pro Asp Leu Gln Ile Gln Arg Tyr Phe Val Ala Leu Tyr Gly
705                 710                 715                 720

Asp Glu Asp Leu Pro Ser Tyr Gly Glu Met Arg Lys Ile Leu Leu Lys
                725                 730                 735

Lys Leu His Asp Phe Ser Ile Arg Asp Phe Asp Glu Glu Ile Gly Ser
            740                 745                 750
```

```
Leu Phe Asp Glu Met Lys Arg Lys Asp Lys Phe Cys Glu Ser Asp Arg
        755                 760                 765

Lys Lys Cys Leu Phe Arg Leu Tyr Leu Thr Ile Ala Tyr Leu Ile Thr
    770                 775                 780

Lys Ser Met Val Lys Ile Asn Thr Arg Phe Ser Ile Ala Cys Ala Thr
785                 790                 795                 800

Tyr Glu Arg Asp Tyr Ala Leu Leu Cys Ala Ser Asn Lys Gln Glu Arg
                805                 810                 815

Ala Trp Ser Ser Gly Ala Thr Ala Leu Ala Leu Thr Arg Lys Phe Leu
            820                 825                 830

Asn Gln Asp Lys Leu Ile Phe Glu Lys His Tyr Ala Arg Glu Gly Glu
        835                 840                 845

Ile Ser Lys Leu Pro Lys Glu Glu Arg Lys Ala Met Arg Lys Val Asn
    850                 855                 860

Asp Gln Leu Leu Lys Gln Thr His Phe Ser Lys His Ser Tyr Cys Tyr
865                 870                 875                 880

Ile Val Asp Asn Val Asn Arg Leu Thr Gly Gly Glu Cys Arg Thr Asp
                885                 890                 895

Lys Arg Val Leu Pro Val Leu Asn Glu Lys Asn Asp Asn Ala Gly Ile
            900                 905                 910

Leu Ser Asp Phe Arg Lys Thr Ile Ala His Leu Asn Val Val His Lys
        915                 920                 925

Met Val Asp Tyr Val Ala Glu Ile Lys Gly Ile Thr Ser Tyr Tyr Ala
    930                 935                 940

Phe Phe Cys Tyr Val Leu Gln Arg Met Pro Val Gly Asn Lys Asn Glu
945                 950                 955                 960
```

```
Lys Asn Ala Leu Lys Glu Lys Tyr Ser Glu Thr Ala Lys Ser Tyr Gly
                965                 970                 975


Thr Tyr Ser Lys Asp Phe Met Trp Leu Ile Asn Leu Pro Phe Ala Tyr
            980                 985                 990


Asn Leu Pro Arg Tyr Lys Asn Leu  Ser Asn Glu Gln Leu  Phe Tyr Asp
        995                 1000                1005


Glu Glu  Glu Arg Asn Glu Thr  Glu Gly Gln Ile Asp  Arg Leu
    1010                1015                1020


<210>  296
<211>  1027
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  pig_chimera compressed

<400>  296

Met Ala Lys Lys Ile Thr Ala Lys Gln Arg Arg Glu Glu Arg Glu Arg
1               5                   10                  15


Gln Asn Lys Gln Lys Trp Ala Lys Lys Gln Ala Asp Ala Thr Ala Val
            20                  25                  30


Phe Glu Cys Glu Ala Asp Ile Lys Pro Ala Asp Ser Lys Asp Glu Asp
        35                  40                  45


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


Ala Met Gly Leu Lys Thr Val Leu Gly Phe Asp Asn Lys Ile Ala Ile
65                  70                  75                  80


Ala Ser Phe Met Ser Ser Lys Asp Ser Lys Ser Ser His Ile Glu Arg
                85                  90                  95


Ile Thr Asp Pro Asn Gly Lys Thr Ile Arg Glu Asp Val Arg Met Phe
```

100 105 110

Asp Ser Asn Val Asp Glu Cys Ser Ile Asn Leu Glu Lys Arg Met Thr
115 120 125

Val Glu Glu Arg Gln Lys Asp Gly Thr Ile Lys Lys Asp Glu Lys Asp
130 135 140

Val Lys Ser Thr Ile Cys Asn Pro Tyr Ser Asn Glu Cys Gly Lys Asp
145 150 155 160

Tyr Ile Gly Ile Lys Ser Val Ala Glu Glu Leu Phe Phe Gly Arg Thr
165 170 175

Phe Pro Asn Asp Asn Leu Arg Val Gln Ile Ala Tyr Asn Ile Phe Asp
180 185 190

Ile Gln Lys Ile Leu Gly Thr Tyr Ile Asn Asn Ile Ile Tyr Ser Phe
195 200 205

Tyr Asn Leu Ser Arg Asp Glu Ser Gln Ser Asp Asn Asp Val Ile Gly
210 215 220

Thr Leu Tyr Met Leu Lys Asp Phe Asp Gly Gln Lys Glu Thr Asp Thr
225 230 235 240

Phe Arg Gln Ala Arg Ala Leu Leu Glu Arg Thr Glu Ala Tyr Tyr Ser
245 250 255

Tyr Phe Asp Asn Val Phe Lys Lys Ile Asp Lys Asn Lys Lys Lys Ser
260 265 270

Asp Asp Cys Lys Arg Glu Arg Asn Glu Ile Leu Arg Tyr Asn Phe Asn
275 280 285

Val Leu Arg Val Leu Ser Phe Leu Arg Gln Ile Cys Ala His Ala Gln
290 295 300

Val Lys Ile Ser Asn Glu His Asp Arg Glu Lys Gly Gly Gly Leu Val

305 310 315 320

Asp Ser Leu Asp Ala Leu Phe Asn Ile Ser Arg Phe Phe Asp Ala Val
325 330 335

Ala Pro Glu Leu Asn Glu Val Ile Asn Ser Val Tyr Ser Lys Gly Ile
340 345 350

Asp Asp Ile Asn Asp Asn Phe Val Lys Asn Gly Lys Asn Asn Phe Tyr
355 360 365

Ile Leu Ser Lys Ile Tyr Pro Glu Val Ala Arg Glu Asp Leu Leu Arg
370 375 380

Glu Tyr Tyr Tyr Phe Val Val Ser Lys Glu Gly Asn Asn Ile Gly Ile
385 390 395 400

Ser Thr Lys Lys Leu Lys Glu Ala Ile Ile Val Gln Asp Met Ser Tyr
405 410 415

Ile Lys Ser Glu Asp Tyr Asp Thr Tyr Arg Asn Lys Leu Tyr Thr Val
420 425 430

Leu Cys Phe Ile Leu Val Lys Glu Leu Asn Glu Arg Thr Thr Ile Arg
435 440 445

Glu Gln Met Val Ala Asp Leu Arg Ala Asn Met Asn Gly Asp Ile Gly
450 455 460

Arg Glu Asp Ile Tyr Ser Lys Tyr Ala Lys Ile Ile Tyr Ala Gln Val
465 470 475 480

Lys Pro Arg Phe Asp Thr Met Lys Ser Ala Phe Glu Glu Glu Ala Lys
485 490 495

Asp Val Ile Val Pro Asp Lys Lys Lys Pro Val Lys Phe Ser His Gly
500 505 510

Lys Leu Asp Lys Asn Glu Ile Glu Arg Phe Cys Ile Thr Ser Ala Asn

515 520 525

Thr Asp Ser Val Ala Lys Ile Ile Tyr Phe Leu Cys Lys Phe Leu Asp
530 535 540

Gly Lys Glu Ile Asn Glu Leu Cys Cys Ala Met Met Asn Lys Leu Asp
545 550 555 560

Gly Ile Asn Asp Leu Ile Glu Thr Ala Glu Gln Cys Gly Ala Lys Val
565 570 575

Glu Phe Val Asp Lys Phe Ser Val Leu Ser Asn Cys Glu Thr Ile Ser
580 585 590

Asp Gln Ile Arg Ile Val Lys Ser Ile Ser Lys Met Lys Lys Glu Ile
595 600 605

Ala Ile Asp Asn Asp Thr Ile Phe Leu Asp Ala Leu Glu Leu Leu Gly
610 615 620

Arg Lys Ile Asp Lys Tyr Lys Lys Asp Ala Thr Gly Lys Tyr Leu Lys
625 630 635 640

Asp Glu Asn Gly Lys Tyr Leu Tyr Ser Lys Glu Tyr Asp Asp Phe Gln
645 650 655

Tyr Met Phe Phe Lys Asp Ser His Arg Val Arg Asn Phe Ile Ser Asn
660 665 670

Ser Val Ile Lys Ser Lys Trp Phe Ser Tyr Ile Val Arg Tyr Asn Gln
675 680 685

Pro Ser Glu Cys Arg Ala Ile Met Lys Asn Lys Thr Leu Val Lys Phe
690 695 700

Ala Leu Asp Glu Leu Pro Asp Leu Gln Ile Gln Arg Tyr Phe Val Ala
705 710 715 720

Leu Tyr Gly Asp Glu Asp Leu Pro Ser Tyr Gly Glu Met Arg Lys Ile

725 730 735

Leu Leu Lys Lys Leu His Asp Phe Ser Ile Lys Gly Phe Leu Asp Glu
740 745 750

Ile Val Leu Leu Ser Asp Leu Asp Met Glu Ser Gln Asp Lys Tyr Cys
755 760 765

Glu Lys Glu Gln Lys Lys Ser Leu Phe Arg Leu Tyr Leu Thr Ile Ala
770 775 780

Tyr Leu Ile Thr Lys Ser Met Val Lys Ile Asn Thr Arg Phe Ser Ile
785 790 795 800

Ala Cys Ala Thr Tyr Glu Arg Asp Tyr Ala Leu Leu Cys Ala Ser Asn
805 810 815

Lys Gln Glu Arg Ala Trp Ser Ser Gly Ala Thr Ala Leu Ala Leu Thr
820 825 830

Arg Arg Phe Leu Asn Gln Asp Lys Leu Ile Phe Glu Lys His Tyr Ala
835 840 845

Arg Glu Gly Glu Ile Ser Lys Leu Pro Lys Glu Glu Arg Lys Ala Met
850 855 860

Arg Lys Val Asn Asp Gln Leu Leu Lys Arg Thr His Phe Ser Lys His
865 870 875 880

Ser Tyr Cys Tyr Ile Val Asp Asn Val Asn Arg Leu Thr Gly Gly Glu
885 890 895

Cys Arg Thr Asp Lys Arg Val Leu Pro Val Leu Asn Glu Lys Asn Asp
900 905 910

Asn Ala Gly Ile Leu Leu Asp Phe Arg Lys Thr Ile Ala His Leu Asn
915 920 925

Val Val His Lys Met Val Asp Tyr Val Asp Glu Ile Lys Gly Ile Thr

```
    930                 935                 940

Ser Tyr Tyr Ala Phe Phe Cys Tyr Val Leu Gln Arg Met Leu Val Gly
945                 950                 955                 960

Asn Asn Leu Asn Glu Lys Asn Ala Ile Lys Glu Lys Tyr Ser Ala Thr
                965                 970                 975

Val Lys Ser Phe Gly Thr Tyr Ser Lys Asp Phe Met Trp Leu Ile Asn
            980                 985                 990

Leu Pro Phe Ala Tyr Asn Leu Pro  Arg Tyr Lys Asn Leu  Ser Asn Glu
        995                 1000                1005

Gln Leu  Phe Tyr Asp Glu Glu  Glu Arg Asn Glu Thr  Glu Glu Gln
    1010                1015                1020

Ile Asp  Arg Leu
    1025
```