Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 2800511 C 2022/03/01

(11)(21) 2 800 511

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 2011/05/27
(87) Date publication PCT/PCT Publication Date: 2011/12/01
(45) Date de délivrance/Issue Date: 2022/03/01
(85) Entrée phase nationale/National Entry: 2012/11/22
(86) N° demande PCT/PCT Application No.: US 2011/038358
(87) N° publication PCT/PCT Publication No.: 2011/150342
(30) Priorité/Priority: 2010/05/27 (US61/349,145)

(51) Cl.Int./Int.Cl. *C07K 14/33* (2006.01),
C12N 15/09 (2006.01), *C12N 15/75* (2006.01)

(72) Inventeurs/Inventors:
SCHOLL, DEAN M., US;
GEBHART, DANA M., US;
WILLIAMS, STEVEN R., US;
GOVONI, GREGORY R., US;
MARTIN, DAVID W., US

(73) Propriétaire/Owner:
PYLUM BIOSCIENCES, INC., US

(74) Agent: SMART & BIGGAR LLP

(54) Titre : DIFFOCINES ET LEURS PROCEDES D'UTILISATION
(54) Title: DIFFOCINS AND METHODS OF USE THEREOF

(57) Abrégé/Abstract:
This disclosure relates to the discovery and isolation of the entire cluster of genes encoding R-type high molecular weight bacteriocins that specifically kill Clostridium difficile bacteria, dangerous human pathogens. Also disclosed are methods of producing the R-type bacteriocins in innocuous producer cells that, unlike C. difficile, do not die in the presence of oxygen. Disclosed also is the specific gene of the isolated gene cluster that determines the killing spectrum of the R-type bacteriocin and the demonstration that the killing spectra of diffocins can be altered by engineering orf1374 of the diffocin genetic locus. This invention offers a potent bactericidal agent and a means to make it in order to kill selectively C. difficile bacteria in the environment of the gastrointestinal tract where they can cause great harm and even death of the infected patient or farm animal.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 December 2011 (01.12.2011)

PCT

(10) International Publication Number
WO 2011/150342 A1

(51) **International Patent Classification**:
C07K 14/33 (2006.01) ***C12N 15/75*** (2006.01)
C12N 15/09 (2006.01)

(21) **International Application Number**:
PCT/US2011/038358

(22) **International Filing Date**:
27 May 2011 (27.05.2011)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
61/349,145 27 May 2010 (27.05.2010) US

(71) **Applicant** *(for all designated States except US)*: **AVIDBIOTICS CORP.** [US/US]; 300 Utah Avenue, Suite 150, South San Francisco, California 94080 (US).

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: **SCHOLL, Dean M.** [US/US]; 368 Susie Way #4, South San Francisco, California 94080 (US). **GEBHART, Dana M.** [US/US]; 210 Lisbon Street, San Francisco, California 94112 (US). **WILLIAMS, Steven R.** [US/US]; 3094 Market Street, San Francisco, California 94114 (US). **GOVONI, Gregory R.** [US/US]; 1425 St. Francis Way, San Carlos, California 94070 (US). **MARTIN, David W.** [US/US]; 348 Magee Ave., Mill Valley, California 94941 (US).

(74) **Agents**: **GERSTIN, Karin** et al.; Sughrue Mion, PLLC, 4250 Executive Square, Suite 900, La Jolla, California 92037 (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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, 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, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2011/150342 A1

(54) **Title**: DIFFOCINS AND METHODS OF USE THEREOF

(57) **Abstract**: This disclosure relates to the discovery and isolation of the entire cluster of genes encoding R-type high molecular weight bacteriocins that specifically kill *Clostridium difficile* bacteria, dangerous human pathogens. Also disclosed are methods of producing the R-type bacteriocins in innocuous producer cells that, unlike C. *difficile*, do not die in the presence of oxygen. Disclosed also is the specific gene of the isolated gene cluster that determines the killing spectrum of the R-type bacteriocin and the demonstration that the killing spectra of diffocins can be altered by engineering orf1374 of the diffocin genetic locus. This invention offers a potent bactericidal agent and a means to make it in order to kill selectively C. *difficile* bacteria in the environment of the gastrointestinal tract where they can cause great harm and even death of the infected patient or farm animal.

DIFFOCINS AND METHODS OF USE THEREOF

[01] This description contains a sequence listing in electronic form in ASCII text format. A copy of the sequence listing in electronic form is available from the Canadian Intellectual Property Office.

FIELD OF THE INVENTION

[02] This application relates generally to the identification and isolation of a cluster of genes sufficient to produce a bacteriocin, and more specifically, an R-type high molecular weight bacteriocin that specifically kills *Clostridium difficile*, and methods to alter its bactericidal specificity, produce and use the same.

BACKGROUND OF THE INVENTION

[03] *Clostridium difficile* is an obligate anaerobic, spore-forming, gram positive bacterium that is a notorious pathogen for humans and other mammals (Bartlett et al., 1977; Bartlett et al., 1979; Keel et al., 2007; Sunenshine & McDonald, 2006). At low densities *C. difficile* can reside innocuously in the mammalian gastrointestinal (GI) tract, but upon expansion, frequently as the result of administered antibiotics reducing the commensal bacteria, *C. difficile* bacteria produce sufficient exotoxins to cause a range of diseases from a mild diarrheal disease to a characteristic pseudo-membranous colitis, which is life-threatening, particularly to older humans and others with significant co-morbidities (Bartlett, 2002).

[04] Because spores formed by this pathogen disseminate widely and are difficult to eradicate or inactivate in hospitals and chronic care facilities, the probability of patients being colonized by *C. difficile* increases sharply upon their entering such a facility (Bartlett, 2007). In fact, a relatively new strain of *C. difficile* that is a hypervirulent toxigenic bacterial strain of *C. difficile*, BI/NAP1/027, which causes severe disease in massive outbreak settings,

has recently been well documented (Spigaglia et al., 2002; Pépin et al., 2004; McDonald et al., 2005; Muto et al., 2005; Loo et al., 2005; Belmares et al., 2009). The incidence of *C. difficile* associated disease (CDAD) in children, previously at low risk, has also increased substantially (Benson et al., 2007; Zilberberg et al., 2010).

[05] Eliminating the pathogen prophylactically in asymptomatic carrier or colonized subjects by administering antibiotics is strongly contraindicated because of the high risk of inducing *C. difficile* associated disease.

[06] R-type bacteriocins made by gram negative bacteria have been described and have been deployed by such bacteria to kill other competitive gram negative strains, even in some circumstances other species or genera of gram negative bacteria (Kageyama et al., 1964; Kageyama et al., 1964a; Kingsbury, D, 1966; Blackwell and Law, 1981; Blackwell et al., 1982; Campagnari et al., 1994; Strauch et al., 2001; Jabrane et al., 2002). The fusion of base plate attachment regions (BPAR) of R-type pyocins to heterologous receptor binding domains (RBD), resulting in the creation of novel R-type pyocins with novel bactericidal specificities for gram negative bacteria has been described (Williams et al., 2008; Scholl et al., 2009).

[07] Other high-molecular-weight bacteriocins or R-type bacteriocins have been described in gram-positive bacteria (Coetzee et al., 1968; Thompson and Pattee, 1981; Zink et al., 1995). However, much less is known about R-type high molecular weight bacteriocin structures produced by gram positive bacteria. But while such have been described, none has been characterized at a genetic level or manipulated in a manner supportive or necessary for developing a useful agent. High molecular weight bacteriocins have been described for 2 *Clostridium* species, *botulinum* and *perfringens* (Ellison and Kautter, 1970; Anastasio et al., 1971; Nieves et al., 1981). None has been described that is produced by *C. difficile* or kills *C. difficile.*

SUMMARY OF THE INVENTION

[08] This invention is based on the isolation of the entire genetic locus or gene cluster encoding the *C. difficile*-specific R-type bacteriocins (herein termed "diffocins") that are bactericidal against other strains of *C. difficile*; the expression of the diffocin gene cluster and production of diffocins in aerobic bacteria; and the discovery that the open reading frame (ORF) 1374 of the diffocin gene cluster determines the bactericidal spectrum of that diffocin against *C. difficile* strains. This invention provides a practical means of altering the specificity of diffocins by genetic engineering to produce novel diffocins and of manufacturing and administering directly or indirectly diffocins to eliminate *C. difficile* from the gastrointestinal (GI) tract of colonized animals, including humans. The administration of diffocins can treat or prevent the development of *C. difficile* infection and associated disease without harming, as do traditional antibiotics, the commensal GI bacteria so necessary for good health.

[09] In accordance with the present invention, there are provided isolated nucleic acid molecules encoding R-type high molecular weight (hmw) bacteriocins. In one embodiment there are provided, isolated nucleic acid molecules encoding R-type high molecular weight (hmw) bacteriocins, wherein the nucleic acid molecule is from a genome of a strain of *Clostridium difficile*, and wherein the R-type hmw bacteriocin comprises a polypeptide that is at least 80% identical to a polypeptide selected from the group consisting of SEQ ID NOs: 4-16, 18, 19, and 66-80, and the R-type hmw bacteriocin has a receptor binding domain (RBD) that binds a receptor of at least one other strain of *C. difficile* and therefore has bactericidal activity against the other strain or strains of *C. difficile*. In particular embodiments, the nucleic acid molecule is from a genome of a strain of *Clostridium difficile* selected from the group consisting of Cd4, Cd16, Cd19108, Cd19123, Cd19126, Cd19145, and ATCC Accession No. 43593. In some embodiments, the strain is

Cd16 and the nucleic acid molecule includes SEQ ID NO:1 or the strain is Cd4 and the nucleic acid molecule includes SEQ ID NO:61

[10] In another embodiment of the invention, there are provided isolated R-type bacteriocins encoded by a nucleic acid molecule of the invention. In one aspect, the R-type bacteriocins are expressed in an aerobic producer bacterium.

[11] In another embodiment of the invention, there are provided isolated R-type high molecular weight (hmw) bacteriocins having bactericidal activity, wherein the R-type hmw bacteriocin includes a base plate attachment region (BPAR) of a first strain of a first species of bacteria of genus *Clostridium*, and a receptor binding domain (RBD) from a second strain of the first species, or from a second species of the genus *Clostridium* or of a bacteriophage that infects a *Clostridium* species, or a modified form of an RBD, wherein the bacteriocin has bactericidal activity against at least one strain of *Clostridium difficile*. In particular embodiments, the BPAR is from a first strain of *Clostridium difficile* and the RBD is from a second strain of *Clostridium difficile* or from a bacteriophage that infects *Clostridium difficile*. In some embodiments, the BPAR is at least 80% identical to the corresponding segment of one or more of SEQ ID NOs: 16, 54-56 and 78. In other embodiments, the RBD is at least 80% identical to the corresponding segment of one or more of SEQ ID NOs: 17, and 49-56. In certain embodiments, the BPAR is at least 80% identical to a polypeptide containing 50 or more contiguous amino acids of SEQ ID NO:16 or 78 or containing 50 or more contiguous amino acids of SEQ ID NOs: 54-56.

[12] In another embodiment of the invention, there are provided expression cassettes containing a nucleic acid molecule of the invention. Expression cassettes may be contained within an expression vector, such as a plasmid, or may be contained within the chromosome of a producer cell. In some embodiments, the expression cassette contains a heterologous promoter operably linked to the nucleic acid molecule encoding the R-type

bacteriocin. The promoter may be inducible, repressible, or constitutively active. In one aspect, the promoter is inducible; in another aspect the promoter is repressible. In some embodiments, the promoter is induced by adding or removing a small molecule inducer, repressor, or de-repressor. In one aspect, the promoter is induced by a small molecule inducer or de-repressor. In a particular aspect, the expression of the cassette is regulated by an operably linked *recA* gene encoding a constitutively active RecA protein and under the control of an heterologous promoter responsive to a small molecule inducer or de-repressor.

[13] In still another embodiment of the invention, there are provided producer cells containing the expression cassettes of the invention. The expression cassette may be contained in an episomal expression vector within the producer cell. Alternatively, the producer cells may contain within their chromosome a nucleic acid molecule or expression cassette of the invention. In certain embodiments, the producer cell is a non-pathogenic and not obligate anaerobic bacterium. In some embodiments, the non-pathogenic and not obligate anaerobic bacterium is a species from a genus of bacteria selected from the group consisting of *Bacillus, Lactobacillus*, and *Listeria*. In certain embodiments, the non-pathogenic and not obligate anaerobic bacterium is from the genus *Bacillus*. In some aspects, the bacterium is *Bacillus subtilis*. In a particular aspect, the *B. subtilis* lacks the PBSX gene cluster. In another embodiment the producer cell is an obligate anaerobic but non-pathogenic bacterium.

[14] In yet another embodiment of the invention, there are provided methods of producing an R-type hmw bacteriocin of the invention. The method includes exposing a producer cell containing a nucleic acid sequence of the invention operably linked to an inducible or derepressible promoter sensitive to an inducing or repressing agent, to the agent in a concentration effective to induce expression of the R-type bacteriocin, and purifying the expressed R-type bacteriocin. In some embodiments, the nucleic acid molecule encoding the R-type bacteriocin is heterologous to the genome of the producer cell. In particular aspects,

the nucleic acid molecule is contained within the producer cell's chromosome or is contained in an extrachromosomal expression vector within the producer cell. In certain embodiments, the producer cell is a non-pathogenic and not obligate anaerobic bacterium. In some embodiments, the non-pathogenic and not obligate anaerobic bacterium is a species from a genus of bacteria selected from the group consisting of *Bacillus, Lactobacillus*, and *Listeria*. In certain embodiments, the non-pathogenic and not obligate anaerobic bacterium is from the genus *Bacillus*. In some aspects, the bacterium is *Bacillus subtilis*. In a particular aspect, the *B. subtilis* does not lyse when induced to produce the R-type bacteriocin. In a further aspect, the *B. subtilis* lacks the PBSX gene cluster.

[15] In a further embodiment of the invention, there are provided methods of killing a pathogenic bacterium and use of an R-type bacteriocin of the invention for such killing. This involves contacting the pathogenic bacterium with an R-type bacteriocin of the invention, whereby the R-type bacteriocin binds and kills the pathogenic bacterium. In one aspect, the pathogenic bacterium is *Clostridium difficile*. In one aspect, the *Clostridium difficile* is in an animal and a bactericidal amount of the R-type bacteriocin is for administration to the animal.

[16] In another embodiment of the invention there are provided methods of treating or preventing a disease-causing infection of *Clostridium difficile* in an animal and use of an R-type bacteriocin or producer cell of this invention for such treating or preventing. This may involve administration of a bactericidal amount of an R-type bacteriocin of the invention directly to an animal in need thereof, administration of the agent indirectly by administering producer cells, or administration of spores of *C. difficile* bacteria which produce natural diffocins but have been genetically modified to not produce toxins. In particular embodiments, an infection of *Clostridium difficile* in an animal is treated by administration to an animal in need thereof an amount of a producer cell of the invention to produce a bactericidal amount of the bacteriocin, thereby treating the infection. In one aspect, the nucleic acid encoding the

bacteriocin, thereby treating the infection. In one aspect, the nucleic acid encoding the bacteriocin is under the control of a *lac* promoter and the animal is administered lactose. In some embodiments, the animal is a mammal. In one aspect, the mammal is a human.

[16A] Various embodiments of the claimed invention relate to an isolated nucleic acid molecule encoding a non-natural R-type bacteriocin with bactericidal activity, wherein the non-natural R-type bacteriocin is constructed from a natural R-type bacteriocin encoded by a genome a first strain of *C. difficile*, wherein in the non-natural R-type bacteriocin, the receptor binding domain (RBD) present in the natural R-type bacteriocin encoded by the genome of the first strain of *C. difficile* has been replaced with a different RBD, wherein the different RBD is: (1) an RBD of an R-type bacteriocin of a *C. difficile* strain different from the first *C. difficile* strain; or (2) an RBD of a bacteriophage that infects *C. difficile*, wherein the non-natural R-type bacteriocin encoded by said nucleic acid molecule comprises polypeptides that are at least 80% identical to the polypeptides consisting of SEQ ID NOs: 4-15 or consisting of SEQ ID NOs: 66-77, wherein the nucleic acid molecule further comprises a heterologous sequence encoding the RBD of the R-type bacteriocin of the *C. difficile* strain different from the first *C. difficile* strain, or encoding the RBD of the bacteriophage that infects *C. difficile*, and wherein the non-natural R-type bacteriocin encoded by the isolated nucleic acid molecule has a bactericidal spectrum that is different from the natural R-type bacteriocin encoded by the genome of the first *C. difficile* strain.

[16B] Various embodiments of the claimed invention also relate to an isolated non-natural R-type bacteriocin having bactericidal activity against at least one strain of *C. difficile*, wherein the non-natural R-type bacteriocin is constructed from a natural R-type bacteriocin encoded by a genome of a first strain of *C. difficile*, wherein the non-natural R-type bacteriocin

Date Reçue/Date Received 2021-02-08

comprises a base plate attachment region (BPAR) from the natural R-type bacteriocin encoded by the genome of the first strain of *C. difficile,* wherein in the non-natural R-type bacteriocin, a receptor binding domain (RBD) present in the natural R-type bacteriocin encoded by the genome of the first strain of *C. difficile* has been replaced with a different RBD, wherein the different RBD is: (1) an RBD of an R-type bacteriocin of a *C. difficile* strain different from the first *C. difficile* strain; or (2) an RBD of a bacteriophage that infects *C. difficile,* and wherein the BPAR comprises a polypeptide of 50 or more contiguous amino acids at least 80% identical to SEQ ID NO: 16 or 78.

BRIEF DESCRIPTION OF THE DRAWINGS

[17] Figure 1 provides the nucleic acid or amino acid sequences of SEQ ID NOs:1-80.

[18] Figure 2. Bactericidal activity of diffocins on clinical *C. difficile* isolates. The indicator strain numbers are shown along the top of the matrices; the asterisk-marked strains are NAP1/027/BI strains. The identities of the *C. difficil*e sources of the diffocins tested are shown along the left borders. Strains were acquired from LC Fortier (4 and 16), ATCC (43593), or RM Alden Research Lab, Culver City, CA.

[19] Figure 3 shows a scanning electron micrograph of dif16. Note the flower-like tail fiber appendages.

[20] Figure 4 shows a photograph of the results of a spot test of dif4 on the target strain Cd19135. Purified diffocin was serially diluted 10-fold, and 5 µl aliquots of the dilutions were spotted on a lawn of target *C. difficile* bacteria. After anaerobic incubation at 37°C overnight, diffocin killing was indicated by the clearing of growth on the lawn.

Date Reçue/Date Received 2021-02-08

[21] Figure 5 shows a photograph of a silver stained SDS-PAGE of both filtered ("F") and unfiltered ("UF") preparations of dif4 ("4") and dif16 ("16"). The arrows indicate bands that were excised and identified by mass spectrometry.

[22] Figure 6 shows a schematic of the diffocin gene cluster from Cd630 and Cd16. The locus consisted of the ORFs that encode structural proteins and structural assembly proteins typical of a *Myoviridae* phage tail apparatus, ORF1362 to ORF1375, indicated below the map. Flanking these genes were genes that encode putative phage-like regulatory proteins. The bottom arrows indicate the direction in which the ORFs are transcribed and the putative functions of the ORFs are indicated above the map.

[23] Figure 7 is an image of spot tests of dif16 produced by *Bacillus subtilis* BDR123-488 (sector 1) and BDR123-491 (sector 2) tested on a lawn of C. difficile strain 19099.

[24] Figure 8 provides the results of a ClustalW analysis of partial amino acid sequences (SEQ ID NOs:81-86, respectively, in order of appearance) encoded by the ORF 1374 gene from each of 5 strains of *C. difficile* that produced active diffocins. Under each row of aligned sequences the "*" represents amino acid identities at that position encoded by all 5 genes, the ":" represents highly similar amino acids at that position encoded by all 5 genes, and the "." represents somewhat similar amino acids at that position encoded by all 5 genes; and the blank at a given position in the sequence alignment represents no amino acid similarity encoded by all 5 genes.

[25] Figure 9 is an image of spot tests of "Dif4" produced by BDR123-580 and "Dif 4-16". The latter was produced by *Bacillus subtilis* BDR123-587 in which the orf1374 of dif4 (SEQ ID NO:49) was switched to orf1374 of dif16 (SEQ ID NO:17). These two diffocins produced by *B. subtilis* were both tested on lawns of strains 19137 and 19145, as indicated. The specificity of killing by "Dif 4-16" has been switched from that of dif4 to that of dif16.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

[26] As used herein, an "R-type high molecular weight (hmw) bacteriocin" is also known as simply an "R-type bacteriocin" and includes R-type pyocins, diffocins, monocins, enterocoliticins, meningocins, or other high molecular weight (hmw) bacteriocins related structurally or genetically to the *myoviridae* family of bacteriophages. An R-type bacteriocin

includes modified versions of R-type pyocins, diffocins, enterocoliticins, monocins, and meningocins (Williams et al. 2008; Strauch et al., 2001; Kingsbury, 1966; Zink et al. 1995).

[27] The term "diffocin," as used herein refers to an R-type high molecular weight bacteriocin isolated from or derived from *Clostridium difficile* and includes native particles obtained from *C. difficile* as well as particles obtained through expression of the diffocin gene cluster in a non-natural producer cell. A diffocin may also be an engineered particle comprised of polypeptides encoded by genes derived from one or more strains of *C. difficile*, and 80% or more identical to one or more polypeptides of SEQ ID NOs: 2-23, 49 and 62-80.

[28] An R-type bacteriocin of the invention may be thermolabile, mild acid resistant, trypsin resistant, sedimentable by centrifugation at about 65,000 x g, and resolvable by electron microscopy (Kageyama et al., 1962; Bradley, 1967; Daw et al., 1996; Jabrane et al., 2002; Fortier et al., 2007). In many cases, an engineered R-type bacteriocin disclosed herein has one or more, in any combination, of these properties. An additional property common to the R-type bacteriocins disclosed herein is that they do not contain nucleic acid and thus are replication deficient such that they cannot reproduce themselves after or during the killing of a target bacterium, as can many bacteriophages. They are purely proteins, not organisms.

[29] R-type bacteriocins disclosed herein are complex molecules comprising multiple protein, or polypeptide, subunits and resemble the tail structures of bacteriophages of the *myoviridae* family. In naturally occurring R-type bacteriocins, the subunit structures are encoded by the bacterial genome such as that of *C. difficile* or *P. aeruginosa* and form R-type bacteriocins to serve as natural defenses against other bacteria (Kageyama, 1975). A sensitive, target bacterium can typically be killed by a single R-type bacteriocin molecule (Kageyama et al., 1964; Kageyama et al., 1964a; Morse et al., 1980; Strauch et al., 2001).

[30] A "target bacterium" or "target bacteria" refers to a bacterium or bacteria that are bound by an R-type bacteriocin of the disclosure and/or whose growth, survival, or replication is inhibited thereby. In some embodiments, the target bacterium is from the genus *Clostridum*. In particular embodiments, the bacterium is *Clostridium difficile*. In one aspect, more than one strain of *C. difficile* is targeted. Exemplary strains of *C. difficile* include but are not limited to NAP1/BI/ribotype 027, as well as those listed in Figure 2.

[31] The term "growth inhibition" or variations thereof refers to the slowing or stopping of the rate of a bacterial cell's division or cessation of bacterial cell division, or to death of the bacterium or bacteria.

[32] As used herein, a "nucleic acid" or a "nucleic acid molecule" typically refers to deoxyribonucleotide or ribonucleotides polymers (pure or mixed) in single-or double-stranded form. The term may encompass nucleic acids containing nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding, structural, or functional properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-0-methyl ribonucleotides, and peptide-nucleic acids (PNAs). The term nucleic acid may, in some contexts, be used interchangeably with gene, cDNA, mRNA, oligonucleotide, and polynucleotide.

[33] A particular nucleic acid sequence also encompasses conservatively modified variants thereof (such as degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third ("wobble") position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues. Thus, a

nucleic acid sequence encoding a protein sequence disclosed herein also encompasses modified variants thereof as described herein.

[34] The term "segment" as used herein in reference to an amino acid sequence refers to a contiguous sequence of amino acids that may be 10, 12, 15, 20, 25, 50, or 100 amino acid residues in length.

[35] As used herein, the term "heterologous," when used with reference to portions of a protein or nucleic acid sequence, indicates that the sequence comprises two or more subsequences that are not usually found in the same relationship to each other in nature. In one example, the heterologous sequences are from different species of bacteria. In another example, heterologous sequences are from different strains of the same species of bacteria. In one aspect, the heterologous sequences are from different strains of *C. difficile*. In another aspect the heterologous sequences are from a bacterium and a bacteriophage or from a bacterium and a synthetic, non-natural sequence of DNA.

[36] The terms "polypeptide", "peptide", and "protein" are typically used interchangeably herein to refer to a polymer of amino acid residues. Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission.

[37] Virulence factors are those molecules that contribute to the pathogenicity of an organism but not necessarily its general viability. Upon the loss of a virulence factor the organism is less pathogenic but not necessarily less viable. Virulence factors may have any one of numerous functions, for example, regulating gene expression, providing adhesion or mobility, providing a toxin, injecting a toxin, pumping out antibiotic agents, or forming protective coatings including biofilms.

[38] Fitness factors are those molecules that contribute to the organism's general viability, growth rate or competitiveness in its environment. Upon the loss of a fitness factor,

the organism is less viable or competitive and because of this compromise, indirectly less pathogenic. Fitness factors may also possess any one of numerous functions, for example, acquiring nutrients, ions or water, forming components or protectants of cell membranes or cell walls, replicating, repairing or mutagenizing nucleic acids, providing defense from or offense towards environmental or competitive insults.

[39] The term "producer cell" as used herein refers to a cell that is capable of producing or expressing a diffocin-encoding nucleic acid molecule and which does not naturally contain such a nucleic acid molecule. The producer cell may be capable of surviving and growing in the presence of oxygen and is transformed with a vector containing nucleic acid molecule encoding the diffocin, which may be integrated into the chromosome of the producer cell or may be episomal. The producer cell may be a gram positive bacterium. In certain embodiments, the producer cell may be a bacterium from the genus *Bacillus*, *Lactobacillus*, *Clostridium*, or *Listeria*. In some embodiments, the bacterium is a species from the genus *Bacillus* selected from the group consisting of *subtilis*, *amyloliquefaciens*, and *megaterium*. In one aspect, the bacterium is *Bacillus subtilis*. In a particular aspect, the producer cell is a *B. subtilis* strain that lacks the PBSX gene cluster. In other embodiments, the bacterium is a species from the genus *Lactobacillus* selected from the group consisting of *acidophilus*, *casei*, and *bulgaricus*. In still other embodiments, the bacterium is *Listeria innocua*. In another embodiment, the non-pathogenic producer cell may be *Escherichia coli* or of the genus *Clostridium*

DETAILED DESCRIPTION OF MODES OF PRACTICING THE DISCLOSURE

R-type bacteriocins isolated from *C. difficile* have bactericidal activity.

[40] To test for bactericidal activity, lysates of two *C. difficile* strains, Cd4 and Cd16 (from LC Fortier), were made by growing the cells to mid-log phase under strict

anaerobic conditions throughout and then exposing the culture to 3 μg/ml of mitomycin C. After the bacterial cells lysed, particles in the lysates were concentrated and purified by high speed centrifugation (see Example 1). These preparations were shown by electron microscopy to contain concentrated headless phage-like particles (Figure 3). After concentration and purification, the preparations were assessed for bactericidal activity by a spot plate method whereby samples were applied to an overlay lawn of target *C. difficile* bacteria, again under strict anaerobic conditions. Since most phage tail-like bacteriocins typically target strains different from the producing bacteria, target strains consisting initially of a panel of 29 *C. difficile* clinical isolates were tested. After overnight incubation, the plates were examined for the presence of killing activity as indicated by a clearing in the lawn of *C. difficile* bacteria where the purified materials were spotted. Figure 4 shows a typical spot assay of diffocin from Cd4 (dif 4) spotted onto *C. difficile* strain, Cd19135.

[41] Dif4 (diffocin from Cd4) and dif16 (diffocin from Cd16) demonstrated different bactericidal spectra based on spot assays (Figure 2). Dif4 showed bactericidal activity on 10 of 29 *C. difficile* clinical isolate strains, whereas dif16 had activity against 8 of 29 strains. There was some overlap of strains susceptible to both diffocins, but the two diffocins had distinct bactericidal spectra. Diffocins from Cd108 and Cd 43593 had very similar killing spectra and killed at least 9 of the 10 NAP1/027/BI hypervirulent strains in the panel of 29 *C. difficile* clinical isolates. When tested further, dif43593 killed all 18 additional, independent NAP1/027/BI isolates. Thus, dif43593 kills all 28 tested strains of NAP1/027/BI *C. difficile,* Figure 2.

Identification of the diffocin locus.

[42] Diffocin particles, isolated and purified from strain Cd4, were denatured, and the protein components were separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) and detected by silver stain (Figure 5). Two individual bands,

~200 kD and ~40 kD, were excised and analyzed by mass spectrometry. Peptides from the 40 kD band were found to match the predicted products from two *C. difficile* open reading frames (ORF) encoded in several of the *C. difficile* strains for which the complete genomes have been sequenced, including the reference strain Cd630 (GenBank Acc. No. NC_009089.1). The first of these was ORF 1363 (SEQ ID NO:5), a 39,192 Dalton phage-like protein. The second predominant polypeptide in this band corresponded to ORF 1371 (SEQ ID NO:14), a 39,565 Dalton phage-like baseplate protein. Since these proteins were coincidentally very close in molecular weight, they migrated in the same SDS PAGE band. The 200 kD band yielded a dominant polypeptide that corresponded to a portion of the C-terminus of ORF 1374 (SEQ ID NO:17), just downstream of the two other phage-like ORFs.

[43] Since these ORFs mapped in very close proximity within the Cd630 genome, the surrounding region was analyzed. A prophage-like element was found between bases 1574593 and 1596384, which includes ORFs 1360A-1379 (Figure 6 and SEQ ID NO:1-23). Of particular note was that the structural genes encoded in this region corresponded only to the components of the tail structure of a typical *Myoviridae* phage; no genes for capsid, capsid assembly protein, or portal proteins were found. Also absent were any ORFs which encoded putative DNA replication or DNA packaging machinery. Thus, this locus was consistent with that of an R-type (*Myoviridae* phage tail-like) bacteriocin. Flanking the structural genes were several ORFs that encode putative phage-like regulatory proteins. Since several ORFs from this locus encoded polypeptides found in diffocin particles and the fact that the gene cluster resembled that of an R-type bacteriocin, we assigned the diffocin locus to this region.

[44] The structural genes were all encoded on one strand and transcribed in the same direction. The organization of the genes resembled that of the genes of R-type pyocins and many *Myoviridae* phages. Several of the structural proteins displayed sequence similarity

to known *C. difficile* bacteriophages including phages Φ119, and ΦC2 (Goh et al., 2007; Govind et al., 2006). *C. difficile* strain Cd630 was also known to encode two intact prophages, both of which were known to be inducible (Goh et al., 2007). Several of the diffocin ORFs had sequence similarity to prophage 1 and 2 of *C. difficile* strain Cd630, suggesting that these *C. difficile Myoviridae* phages and the diffocins share common ancestry.

[45] Of particular note was ORF 1374 (SEQ ID NO:17) of the diffocin. This gene encoded a large polypeptide that is located just downstream of ORF 1373, SEQ ID NO:16, a location in the cluster that indicated that ORF 1374 was part of the R-type bacteriocin tail fiber, that is, a receptor binding domain (RBD). Since electron micrographs of what have been designated as diffocins revealed a large flower like structure in the tail fiber region, it was apparent that this structure comprised a large protein. ORF 1373 (SEQ ID NO:16) encoded the base plate attachment region, BPAR, of the diffocin tail fiber and appeared to be a truncated form of a the analogous ORF 1373 in phages of *C. difficile* which encoded the RBD as well as the BPAR, given that such phages lack an ORF 1374. Thus, the tail fibers of naturally occurring diffocins are comprised of two proteins, ORF1373 and ORF1374, which form a jointed tail fiber, whereas bacteriophages of *C. difficile* have a tail fiber comprised of a single protein, a somewhat longer ORF1373 which provides the BPAR and the RBD functions. With this knowledge the nucleic acids encoding either ORF1373 (SEQ ID NO:16) or ORF1374 (SEQ ID NO:17) were deployed herein as substrates from which to engineer new RBD specificity functions.

Diffocins are wide-spread among *C. difficile* isolates.

[46] A series of independent clinical isolates of *C. difficile* were tested for the ability to produce diffocin particles that might have different bactericidal spectra. Clinical isolates Cd19123, Cd19145, Cd19126, Cd19108 (from RM Alden Research Laboratory,

Culver City, CA), and ATCC Cd43593 were induced with mitomycin C under strict anaerobic conditions, followed by purification and concentration of particles that were produced. The purified materials were then spotted separately on lawns of other *C. difficile* strains in the clinical collection to detect bactericidal activity. The results showed that each of these isolates produced particles (diffocins termed dif123, dif145, dif126, dif108, and dif43593 from clinical isolates Cd19123, Cd19145, Cd19126, Cd19108, and ATCC Cd43593, respectively) with differing bactericidal spectra (Figure 2.). Dif108 and dif43593 had very similar and broad spectra, whereas strain Cd19145 produced particles with a narrow bactericidal spectrum. Some isolates that were tested for induction, including Cd19113 and Cd19150, lysed after exposure to mitomycin but did not produce detectable diffocin particles.

[47] It was recognized herein that the genome sequenced laboratory strain Cd630 encoded a diffocin locus; and, in fact, this genome was used herein to identify the diffocin genes. However, while strain Cd630 lysed after induction with mitomycin C, no detectable diffocin particles were produced. Instead, small amounts of prophage 1 and prophage 2 were produced.

Identification and cloning of a diffocin cluster.

[48] Cd630 did not produce detectable diffocin particles upon induction of lysis by mitomycin C. Accordingly, the diffocin gene cluster from Cd16 was cloned, instead. A draft genome sequence of Cd16 was first obtained. A diffocin locus, analogous to that of Cd630, was identified and annotated. The entire diffocin gene cluster from producer strain Cd16 was cloned (SEQ ID NO:1) into a BAC and subsequently demonstrated to include all the genes necessary to produce an active diffocin in a non-pathogenic microorganism and one that is not an obligate anaerobe. *C. difficile* is both pathogenic and an obligate anaerobic bacterium, making it an impractical production cell for natural diffocins or diffocins modified by recombinant DNA engineering. The isolated diffocin cluster can also be engineered by

recombinant DNA technology to have its expression regulated by a heterologous promoter responding to a non-toxic small molecule inducer or de-repressor added to the culture or induction medium of the producer microorganism. Such small molecules include but are not limited to tetracycline, anhydrotetracycline, lactose, arabinose, xylose and their non-metabolized analogs, such as IPTG to replace lactose.

DIFFOCINS

[49] Diffocins are R-type hmw bacteriocins isolated from *Clostridium difficile*, which are bactericidal against other strains of *C. difficile*. Diffocins may be isolated from *C. difficile* strains grown under anaerobic conditions in the presence of mitomycin C. In some embodiments, the diffocin is from *C. difficile* clinical isolate Cd4, Cd16, Cd19123, Cd19145, Cd19126, Cd19108, or ATCC Cd43593 (termed dif 4, dif 16, dif123, dif145, dif126, dif108, and dif43593, respectively). In one aspect, the diffocin is from Cd4; in another aspect the diffocin is from Cd16.

[50] In another embodiment of the invention, there are provided isolated nucleic acid molecules encoding diffocins derived from the genome of the genus *Clostridium* bacteria. In one aspect, the nucleic acid molecule contains the gene cluster of SEQ ID NO:1. In another aspect, the nucleic acid molecule contains the gene cluster of SEQ ID NO:61. In other embodiments, the nucleic acid molecule encodes a diffocin that includes one or more polypeptides selected from the group consisting of SEQ ID NOs: 2-23; 49, 62-80. In still other embodiments, the nucleic acid molecule encodes a diffocin that includes one or more polypeptides selected from the group consisting of SEQ ID NOs: 4-16, 18, 19, and 66-80. In one aspect, the nucleic acid molecule encodes the polypeptides of SEQ ID NOs: 2-23. In another aspect, the nucleic acid molecule encodes the polypeptides of SEQ ID NOs: 49 and 62-80.

[51] Also provided are variant diffocins. Variant diffocins include those diffocins having an amino acid sequence that are at least 80% identical to a polypeptide selected from the group consisting of SEQ ID NOs: 4-16, 18, 19, and 66-80. In other embodiments, the variant diffocin has an amino acid sequence that is at least 85%, 88%, 89%, 90%, 95%, 96%, 97%, 98%, or even 99% identical to a polypeptide selected from the group consisting of SEQ ID NOs: 4-16, 18, 19, and 66-80.

[52] In some embodiments, the variant diffocin may include a heterologous base plate attachment region (BPAR), wherein the BPAR is at least 80% identical to the corresponding segment of one or more of SEQ ID NOs: 16, 54-56, and 78. In another embodiment, the BPAR is at least 85% identical to the corresponding segment of one or more of SEQ ID NOs: 16, 54-56, and 78. In another embodiment, the BPAR is at least 89% identical to the corresponding segment of one or more of SEQ ID NOs: 16, 54-56, and 78. In another embodiment, the BPAR is at least 90% identical to the corresponding segment of one or more of SEQ ID NOs: 16, 54-56, and 78. In still another embodiment, the BPAR is at least 95% identical to the corresponding segment of one or more of SEQ ID NOs: 16, 54-56, and 78. In yet another embodiment, the BPAR is at least 98% identical to the corresponding segment of one or more of SEQ ID NOs: 16, 54-56, and 78. In a further embodiment, the BPAR is at least 99% identical to the corresponding segment of one or more of SEQ ID NOs: 16, 54-56, and 78.

[53] In further embodiments, the variant diffocin may include a heterologous receptor binding domain (RBD), wherein the RBD is at least 80% identical to the corresponding segment of one or more of SEQ ID NO:17 and 49-53, or at least 80% identical to a polypeptide containing the receptor binding domain (RBD) region of a polypeptide selected from the group consisting of SEQ ID NOs:54-56. In another embodiment, the RBD is at least 85% identical to the corresponding segment of one or more of SEQ ID NO:17 and

49-53, or at least 85% identical to a polypeptide containing the receptor binding domain (RBD) region of a polypeptide selected from the group consisting of SEQ ID NOs:54-56. In another embodiment, the RBD is at least 89% identical to the corresponding segment of one or more of SEQ ID NO:17 and 49-53, or at least 89% identical to a polypeptide containing the receptor binding domain (RBD) region of a polypeptide selected from the group consisting of SEQ ID NOs:54-56. In another embodiment, the RBD is at least 90% identical to the corresponding segment of one or more of SEQ ID NO:17 and 49-53, or at least 90% identical to a polypeptide containing the receptor binding domain (RBD) region of a polypeptide selected from the group consisting of SEQ ID NOs:54-56. In another embodiment, the RBD is at least 95% identical to the corresponding segment of one or more of SEQ ID NO:17 and 49-53, or at least 95% identical to a polypeptide containing the receptor binding domain (RBD) region of a polypeptide selected from the group consisting of SEQ ID NOs:54-56. In another embodiment, the RBD is at least 98% identical to the corresponding segment of one or more of SEQ ID NO:17 and 49-53, or at least 98% identical to a polypeptide containing the receptor binding domain (RBD) region of a polypeptide selected from the group consisting of SEQ ID NOs:54-56. In another embodiment, the RBD is at least 99% identical to the corresponding segment of one or more of SEQ ID NO:17 and 49-53, or at least 99% identical to a polypeptide containing the receptor binding domain (RBD) region of a polypeptide selected from the group consisting of SEQ ID NOs:54-56. In some embodiments, the receptor binding domain (RBD) region comprises amino acid residue 51 to the carboxy-terminal residue of SEQ ID NOs:54, 55, or 56.

[54] In another embodiment, diffocins can be engineered to have altered bactericidal spectra by fusing phage tail RBD to the product of diffocin ORF 1373. While ORF 1374 encodes the primary spectra determinant or RBD of natural diffocins, this very large protein is complexed with the ORF 1373 protein, and ORF 1373 protein provides the

BPAR, i.e., it attaches the RBD of ORF 1374 protein to the diffocin baseplate structure. ORF 1373 is analogous to, and shares amino acid sequence identity with, the tail fiber genes of *myoviridae* bacteriophages such as ΦCD2 (SEQ ID NO:54), ΦCD119 (SEQ ID NO:55), and ΦCD27 (SEQ ID NO:56) as well as with the tail fibers of R-type pyocins. The ORF 1373 (e.g., SEQ ID NOs:16 or 78) of diffocins shares significant sequence identity, particularly in the first 160 amino acids at the N-terminal portion or BPAR, with the tail fibers of the *C. difficile* myoviridae phage, ΦCD2 (SEQ ID NO:54). The phage tail fibers are, however, longer than diffocin ORF 1373 protein and contain a C-terminal RBD for recognizing their bacterial targets. Diffocins' ORF 1373 proteins do not contain this latter domain, the RBD function of which has been replaced by a separate polypeptide, encoded by ORF1374. Thus, ORF 1374 can be deleted altogether from the diffocin cluster and an RBD of a phage tail fiber, such as that of ΦCD2, can be fused to the diffocin BPAR, encoded by ORF 1373, thereby generating a diffocin that has a phage tail fiber-like protein and accordingly, a bactericidal spectrum related to the host range of the donor phage. Importantly, because the regions of amino acid sequence homology between the *C. difficile* phage tail fibers and the ORF 1373 protein enable successful functional fusions between the two, one can select host-range variants from mutagenized or non-mutagenized *C. difficile* phages that can then be sources of novel RBD's for creating modified diffocins with novel bactericidal spectra.

[55] In one embodiment of an engineered diffocin, there is provided a diffocin in which the RBD has been replaced with an RBD from another strain of *C. difficile* or with an RBD from a bacteriophage that infects *C. difficile*. In one example, the nucleic acid molecule comprises a sequence encoding SEQ ID NO:16 or 78 but does not contain the corresponding native RBD (i.e., the sequence encoding SEQ ID NO:17 or 49, respectively); instead, the native RBD is replaced with a heterologous sequence encoding an RBD. In particular embodiments, the nucleic acid molecule contains a heterologous sequence encoding a

receptor binding domain (RBD) of an R-type bacteriocin of a different strain of *C. difficile*. In one aspect, the nucleic acid molecule contains a sequence encoding SEQ ID NO:16 or 78 and a sequence encoding a polypeptide selected from the group consisting of SEQ ID NOs: 49-53 or the receptor binding region of a polypeptide selected from the group consisting of SEQ ID NOs: 54-56. In another aspect, the nucleic acid molecule is comprised of a sequence encoding SEQ ID NOs:2-16 and 18-23 or SEQ ID NOs: 62-80, and a heterologous sequence encoding an RBD from a polypeptide selected from the group consisting of SEQ ID NOs: 17 and 49-56.

[56] In other embodiments of an engineered diffocin, the RBD of the diffocin may be replaced with a modified form of a native RBD. A "native RBD" refers to a RBD having an amino acid sequence that is identical to a RBD isolated or cloned from a strain of *C. difficile* or from a bacteriophage that infects *C. difficile*. Exemplary native RBDs from a number of *C. difficile* strains include SEQ ID NOs: 17 and 49-53. Exemplary native RBDs from bacteriophages that infect *C. difficile* include SEQ ID NOs: 54-56 (e.g., amino acid residue 51 to the carboxy terminal residue). In some embodiments, a modified RBD includes a change in the amino acid sequence of the RBD relative to a native RBD. Non-limiting examples of a change in amino acid sequence include substitution, insertion (or addition), or deletion of one or more amino acids. In further embodiments, a diffocin includes a substitution with, or insertion of, an RBD derived from an organism that diversifies the structure by deploying a Diversity Generating Retroelement (DGR), as described in published Patent Application US 2006-0121450, published Jun. 8, 2006.

[57] In some embodiments, the modified form has a bactericidal spectrum that is different from the corresponding unmodified or native RBD. In particular embodiments, the modified form is at least 80% identical the native RBD. In other embodiments, the RBD has

an amino acid sequence that is at least 85%, 88%, 89%, 90%, 95%, 96%, 97%, 98%, or even 99% identical to a polypeptide selected from the group consisting of SEQ ID NOs: 17 and 49-53 or the receptor binding region of a polypeptide selected from the group consisting of SEQ ID NOs: 54-56 and the modified RBD has a bactericidal spectrum that is different from the corresponding unmodified or native RBD.

TARGET BACTERIA

[58] *Clostridium difficile* strains isolated from patients vary widely by pulse gel electrophoresis and in their pathogenicity. The BI/NAP1 or ribotype 027 strains that hyperproduce toxins are particularly virulent as a result of their having lost the function of gene tcdC that negatively regulates the expression level of toxin A and toxin B (McDonald et al., 2005).

[59] In fact, *C. difficile* strains harboring a specific mutant allele of the *tcdC* gene have been shown to spread epidemically within and among healthcare facilities. These epidemic and highly virulent strains are especially important target bacteria that could be eliminated prophylactically from the GI tract of carrier patients by oral application of diffocins prior to or shortly after the commencement of traditional antibiotic therapy. *C. difficile* bacteria that produce wild-type levels of toxins A and B are important target pathogens as well since they are also potentially lethal, particularly to patients older than 50 years or with co-morbidities (Bartlett JG, 2002).

[60] Targeting surface accessible virulence or fitness factors such as S-layer proteins, prevalent on *C. difficile* strains, whether hyperproducers or not, offer an attractive means of forcing such pathogens to compromise their virulence or fitness if they emerge as resistant to the targeted R-type bacteriocin. Because of the high specificity of the RBD of diffocins, organisms other than *C. difficile* are not targets, a distinct and powerful advantage

of diffocins since they will not cause collateral damage to commensal bacteria of the GI tract—bacteria necessary for normal GI function and good health.

[61] An "infection" refers to growth of bacteria, such as in a subject or tissue or non-bacterial cell, wherein the bacteria actually or potentially could cause disease or a symptom in the subject, tissue or non-bacterial cell. Treatment of an infection may include prophylactic treatment with substances, materials, producer cells, or the spores of detoxified *C. difficile* bacteria capable of producing diffocins such as dif43593. Non-limiting examples of treated objects include donated organs, tissues, and cells; medical equipment, like a respirator or dialysis machine; or wounds, such as those during or after surgery. Other uses include the removal of target bacteria which may cause problems upon further growth. In additional embodiments, an hmw bacteriocin is used to treat food, plants or harvested parts of plants with bacterial infections or contaminations, or to treat environmental occurrences of the target bacteria, such as in a hospital or commercial setting.

[62] As described herein, an anti-bacterial R-type bacteriocin may be used to inhibit growth, survival, or replication of a particular bacterium. The bacterium may be a pathogenic or environmentally deleterious strain, or may be treated in a prophylactic manner. A pathogenic microorganism generally causes disease, sometimes only in particular circumstances.

PREPARATION AND USE OF DIFFOCINS

[63] Diffocins are particles of approximately 10 million daltons and thus can be isolated and purified by differential centrifugation, differential filtration, aqueous two-phase separations, polyethylene glycol (PEG) precipitation and/or ion exchange chromatography to create biopharmaceutical grade oral antibacterial agents. R-type bacteriocins have been found to be stable to freezing-thawing and can be spray dried to create stable formulations.

[64] In some embodiments of the invention, there are provided methods of producing an R-type hmw bacteriocin. The method includes exposing a producer cell to a nucleic acid sequence encoding an R-type hmw bacteriocin operably linked to an inducible promoter sensitive to an inducing agent in a concentration that brings about expression of the R-type bacteriocin, and purifying the expressed R-type bacteriocin. In one aspect, the R-type high molecular weight (hmw) bacteriocin contains one or more polypeptides selected from the group consisting of SEQ ID NOs:2-23 or SEQ ID NOs:49 and 62-80. The nucleic acid molecule is heterologous to the natural nucleic acid of the producer cell and may be contained in the producer cell's chromosome or may be contained in an episomal expression vector.

[65] As targeted, potent antibacterial agents, diffocins will be used to remove, or decolonize, C. *difficile* from the lower GI tract of humans and other animals so as to prevent CDAD. Animals and humans treated with broad spectrum antibiotics are at high risk to develop potentially lethal CDAD if they have been colonized by C. *difficile*. Decolonization is a particularly attractive utility of diffocins because of their sparing of the healthy GI microbiota. In addition, diffocins can be administered directly or indirectly via administered producer cells or spores of detoxified *C. difficile* bacteria capable of producing diffocins to reduce the pathogen load in acute CDAD and/or to reduce the high incidence or recurrence or relapse of CDAD after successful treatment by other modalities.

MODES OF ADMINISTRATION

[66] R-type bacteriocins are inactivated by pH 4.0 or lower, the acidity of a normally functioning, fed stomach and upper duodenum. However, diffocins must transit the upper GI tract to reach the targeted bacterial pathogen colonizing predominately the lower GI tract. Thus, diffocins can be formulated by one or several known methods that protect a vulnerable agent from the acid and proteases of the upper GI tract and deliver such agent in

an active state to the distal upper GI tract or lower GI tract. In addition, animals can be treated with antihistamines such as cimetidine or proton pump inhibitors to prevent stomach acidification before oral administration of R-type bacteriocins. Thus, oral administration of properly formulated diffocins, producer cells capable of producing diffocins, or spores of diffocin-producing detoxified *C. difficile* bacteria to humans or animals with normal stomachs or to those in whom the acidification has been pharmaceutically prevented will enable delivery to the colonized portion of the intestine and thereby enable efficacy. Based on bowel transit time, the frequency of per oral administration directly or indirectly of diffocins to decolonize asymptomatic persons or animals may be every 6, every 12, every 18, every 24 hours, weekly, or monthly. Diffocins may also be administered to patients with CDAD, or recently "cured" of CDAD, at frequencies the same or greater. Particularly for management of active CDAD, diffocins may be formulated for and administered directly or indirectly per rectum by suppository, enema or colonic perfusion.

[67] An engineered diffocin of the disclosure may be administered to any subject afflicted with, diagnosed as afflicted with, or suspected of being afflicted with, an infection or contamination by bacteria susceptible to the diffocin. Non-limiting examples of such a subject include animal (mammalian, reptilian, amphibian, avian, and fish) species as well as insects, plants and fungi. Representative, and non-limiting, examples of mammalian species include humans; non-human primates; agriculturally relevant species such as cattle, pigs, goats, and sheep; rodents, such as mice and rats; mammals for companionship, display, or show, such as dogs, cats, guinea pigs, rabbits, and horses; and mammals for work, such as dogs and horses. Representative, and non-limiting, examples of avian species include chickens, ducks, geese, and birds for companionship or show, such as parrots and parakeets. An animal subject treated with an engineered diffocin of the disclosure may also be a quadruped, a biped, an aquatic animal, a vertebrate, or an invertebrate, including insects.

[68] In some embodiments, the subject to be treated is a human child or other young animal which has yet to reach maturity. Thus the disclosure includes the treatment of pediatric conditions comprising infection with bacteria or other microorganism susceptible to a diffocin of the disclosure.

[69] The disclosure also provides for the treatment or prevention of an opportunistic infection, such as that resulting from an undesirable growth of bacteria that are present in the microbial flora of a human subject or a non-human animal. An opportunistic infection may be the result of an immunosuppressed condition in a subject or the result of antibiotic treatment that alter the commensal flora of the genitourinary (GU) or gastrointestinal (GI) tract. Thus the disclosure also provides for the treatment or prophylaxis of immunosuppressed subjects and subjects exposed to other pharmaceutical agents. A diffocin with its anti-bacterial activity may be used in combination with another anti-bacterial or anti-microbial agent, such as an antibiotic or anti-fungal agent as non-limiting examples. An "anti-microbial agent" is an agent or compound that can be used to inhibit the growth of, or to kill, single celled organisms. Anti-microbial agents include antibiotics, chemotherapeutic agents, antibodies (with or without complement), chemical inhibitors of DNA, RNA, protein, lipid, or cell wall synthesis or functions.

[70] In some embodiments, diffocins, producer cells, or spores of detoxified *C. difficile* bacteria capable of producing diffocins are formulated with a "pharmaceutically acceptable" excipient, enteric coating or carrier. Such a component is one that is suitable for use with humans, animals, and/or plants without undue adverse side effects. Non-limiting examples of adverse side effects include toxicity, irritation, and/or allergic response. The excipient or carrier is typically one that is commensurate with a reasonable benefit/risk ratio. Non-limiting pharmaceutically suitable carriers include sterile aqueous or non-aqueous solutions, suspensions, and emulsions. Examples include, but are not limited to, standard

pharmaceutical excipients such as a phosphate buffered saline solution, water, emulsions such as oil/water emulsion, and various types of wetting agents. Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyloleate. Aqueous carriers include water, alcoholic/aqueous solutions, emulsions or suspensions, including saline and buffered media. Parenteral vehicles include sodium chloride solution, Ringer's dextrose, dextrose and sodium chloride, lactated Ringer's or fixed oils. Intravenous vehicles include fluid and nutrient replenishers, electrolyte replenishers (such as those based on Ringer's dextrose), and the like.

[71] Additional formulations and pharmaceutical compositions disclosed herein comprise an isolated diffocin specific for a bacterial pathogen; a mixture of two, three, five, ten, or twenty or more different diffocins, producer cells or spores of detoxified *C. difficile* bacteria capable of producing diffocins that target the same bacterial pathogen; and a mixture of two, three, five, ten, or twenty or more that target different bacterial pathogens or different strains of the same bacterial pathogen.

[72] Optionally, a composition comprising a diffocin or producer cells of the disclosure may also be spray dried or lyophilized using means well known in the art. Subsequent reconstitution and use may be practiced as known in the field.

[73] A diffocin is typically used in an amount or concentration that is "safe and effective", which refers to a quantity that is sufficient to produce a desired therapeutic response without undue adverse side effects like those described above. A diffocin may also be used in an amount or concentration that is "therapeutically effective", which refers to an amount effective to yield a desired therapeutic response, such as, but not limited to, an amount effective to slow the rate of bacterial cell division, or to cause cessation of bacterial cell division, or to cause death or decrease rate of population growth of the bacteria. The safe and effective amount or therapeutically or prophylactically effective amount will vary with

various factors but may be readily determined by the skilled practitioner without undue experimentation. Non-limiting examples of factors include the particular condition being treated, the physical condition of the subject, the type of subject being treated, the duration of the treatment, the nature of concurrent therapy (if any), and the specific formulations employed.

[74] Having now generally described the inventive subject matter, the same will be more readily understood through reference to the following examples which are provided by way of illustration, and are not intended to be limiting of the disclosure, unless specified.

Examples

1. Determination of the bactericidal activity of diffocins.

[75] *C. difficile* cultures were grown under strict anaerobic conditions in a Forma™ Scientific environmental chamber with an atmosphere of 10% CO_2, 10% H_2, 80% N_2 . All media, buffers, and plates were reduced in this atmosphere for at least 24 hours prior to use. Cultures were streaked on *C. difficile* selective agar plates (BD Diagnostics, BBL Cat. 222228), and incubated at 37°C for two days. These plates as stocks were then stored anaerobically at ambient temperature.

[76] To induce diffocins, *C. difficile* bacteria were grown in liquid cultures using Brucella medium (Difco) at 37°C with no shaking. At an OD_{600} of approximately 0.2, mitomycin C was added to a final concentration of 3μg/ml. Cultures were then incubated for 3-16 hours. Bacterial lysis was detected by a visual clearing of the culture.

[77] Cultures were removed from the anaerobic chamber, and cellular debris was removed by centrifugation at 5,000 x g. The supernatants were then passed through a 0.2 μm cellulose acetate syringe filter. The filtrate was centrifuged at 90,000 x g for 2 hours to pellet the diffocin particles. The pellets were resuspended in 10 mM Tris pH 7.5, 50 mM NaCl, 3% mannitol in 1/50 original culture volume.

[78] Target strains were grown in *Brucella* broth overnight at 37°C. Culture volumes of 100 µl were added to 5 ml of tempered, reduced, *Brucella* overlay agar (0.5% agar), poured onto a *Brucella* agar plate (1.5% agar) and allowed to set. Samples of 5 µl of the diffocin preparations were spotted onto the plates and allowed to air dry (about 30 min). The plates were then incubated anaerobically at 37°C overnight. Bactericidal activity was determined by a clearing, or lack of bacterial growth, at the position or spot where a sample was applied to the lawn.

2. Cloning of the Cd16 diffocin locus.

[79] A draft genome sequence of *C. difficile* strain Cd16 was obtained by 454 instrument sequence analysis of genomic DNA. The entire dif16 locus or cluster (SEQ ID NO:1) was identified by comparison to strain Cd630 (see above).

[80] **Preparing a backbone BAC vector.** The starting vector was pETcoco1 (Novagen) . This was modified to remove the two XhoI sites with primers AV1419 (SEQ ID NO:24) and AV1420 (SEQ ID NO:25), which have BbsI ends. To do so, a specific region was amplified from pETcoco1 DNA with these primers, and subsequently the PCR product was cut with BbsI and ligated back into the larger pETcoco1 vector fragment that was previously cut with XhoI. This ligation destroyed the two XhoI sites of pETcoco1. This latter plasmid was then further modified by a similar strategy to destroy the EcoRI sites using primers AV1416 (SEQ ID NO:26) and AV1245 (SEQ ID NO:27). The resulting vector was termed **SW251**.

[81] **Preparing a pUC19 vector to accept fragments of the diffocin cluster.** The polylinker of pUC19 (New England BioLabs) was modified by digesting with EcoRI and HindIII and ligating in oligos AV1372 (SEQ ID NO:28), AV1373 (SEQ ID NO:29), AV1374 (SEQ ID NO:30), and AV1375 (SEQ ID NO:31). This changed the polylinker to NotI-NheI-KpnI-XhoI-EcoRV-BstBI-BbsI-EcoRI-NsiI-SphI-BamHI-AscI. This was termed **SW232.**

[82] **Cloning the diffocin cluster into SW232.** Three fragments of the diffocin cluster, SEQ ID NO 1, were individually amplified by PCR from Cd16 DNA. The 5' fragment (SEQ ID NO: 32) was amplified with primers 1368 (SEQ ID NO:35) and 1289 (SEQ ID NO:36), which had Not1 and Xho1 ends, respectively. The middle fragment (SEQ ID NO:33) was amplified with primers AV1288 (SEQ ID NO:37) and AV1366 (SEQ ID NO:38), which had Xho1 and EcoR1 ends, respectively. The 3' fragment, SEQ ID NO:34, was amplified with primers AV1367 (SEQ ID NO:39) and AV1300 (SEQ ID NO:40), which had EcoR1 and BamH1 ends, respectively. These three PCR fragments were separately cloned into SW232, and termed SW241, SW242, and SW243 for the 5', middle, and 3' portions, respectively.

[83] **Cloning the diffocin cluster into the BAC SW251.** The three fragments of the diffocin cluster (in SW241, SW242, and SW243), each having been expanded by cloning in *E. coli* and purified, were excised from the SW241, SW242, and SW243 vectors. SW241 was digested with NotI and XhoI, SW242 was digested with XhoI and EcoRI, and SW243 was digested with EcoRI and AscI. (Note that this AscI sites was part of the modified SW232 polylinker described above.)

[84] These three fragments were assembled into SW251 that was first digested with NotI and AscI. The resulting plasmid was termed **DG461** and contained the entire dif16 cluster. It was amplified in *E. coli*.

[85] **Making a diffocin integration vector for expression in *B. subtilis*.** The *B. subtilis* integration vector, pDR111, which included portions of the *amyE* gene flanking a cloning/promoter region and a spectinomycin-resistance gene, was used.

[86] The pDR111 polylinker was modified by digesting the vector with HindIII and SphI and ligating in oligos DG1 (SEQ ID NO:41) and DG2 (SEQ ID NO:42). This added NotI and AscI sites to pDR111. The region containing the entire *amyE* front and back region

with the modified polylinker was then amplified using primers DG9 (SEQ ID NO:43) and DG10 (SEQ lD NO:44), which both have BsaI ends. This fragment was ligated into the NotI and AscI sites of SW251 (resulting in destruction of the two sites) and created **DG487**. Note that there were new NotI and AscI sites introduced into DG487 by the modified poly linker of the pDR111-derived insert.

[87] After expansion of DG487 in *E. coli* the NotI/AscI fragment containing the diffocin cluster from DG461 was then excised and cloned into the NotI/AscI site of DG487. This new construct was named **DG488** and was the vector used to introduce the entire diffocin gene cluster into *B. subtilis* (below).

3. Expression of a diffocin gene cluster in *Bacillus subtilis*.

[88] The natural diffocin producer is C. *difficile*, an obligate anaerobe. If exposed to even traces of oxygen, C. *difficile* bacteria sporulate and die promptly. The ability to generate even trace amounts of diffocins from cultured C. *difficile* is difficult and taxing, and certainly the production of quantities of diffocins useful for prophylactic or therapeutic applications is not practical under the required strict anaerobic conditions. Accordingly, the entire diffocin gene cluster from C. *difficile* was first identified and then isolated by molecular cloning and introduced the cluster into an aerobic gram positive bacterium, *Bacillus subtilis*, for further engineering and production.

[89] The *Bacillus subtilis* integration vector, DG488, was made as described in Example 2 and contained the entire 22,827 base diffocin locus (SEQ ID NO:1). This vector was used to recombine the diffocin locus into the *Bacillus subtilis* genome.

[90] The recipient *Bacillus subtilis* strain was BDR123, which had a chloramphenicol resistance marker inserted within the *amyE* gene. When this strain was transformed with DG488, recombination occurred between the front and back *amyE* sequences within the vector and the genomic *amyE* sequences. This resulted in insertion, into

the BDR123 genome, of all of the sequences between the front and back *amyE* regions of DG488 including the diffocin locus and the spectinomycin resistance gene. Successful recombinants were spectinomycin resistant but had become chloramphenicol-sensitive due to loss of that genomic marker as a result of the recombination event. This *B. subtilis* strain was termed BDR123-488.

[91] Since the entire diffocin locus was inserted into the DG488 vector (Example 2), it therefore was also inserted in its entirety into BDR123-488. This inserted diffocin locus included all of the regulatory genes that were required for normal expression in *Clostridium*, and the structural diffocin particle genes were under control of these genes and/or regulatory elements. Because *Bacillus* and *Clostridium* are related bacteria, it was predicted that these diffocin regulatory elements would function in the *Bacillus* background and, as in their natural state, would be induced by DNA damage through a RecA-mediated mechanism. This in fact was the case, and diffocin particle production was induced in BDR123-488 by contact with the DNA damaging agent mitomycin C and killed strain Cd19099, Figure 7.

[92] Diffocin regulatory genes (ORFs 1359, 1360, 1361) were located in the 5' region of the locus in relation to the structural genes. There were also regulatory genes (ORFs 1377 (SEQ ID NO:20), 1378 (SEQ ID NO:21), and 1379 (SEQ ID NO:23)) located downstream, 3' in relation to the structural genes. To eliminate these latter regulatory genes DG491 was generated from DG488 in a single three way ligation. One PCR fragment was made from DG488 by PCR amplification with primers DG13 (SEQ ID NO:45) and DG14 (SEQ ID NO:46), and the other was made by PCR amplification with primers DG15 (SEQ ID NO:47) and DG16 (SEQ ID NO:48). Both PCR fragments were digested with AscI and SphI. DG488 was then digested with SphI, and the two digested PCR fragments were ligated into the large vector fragment from SphI-digested DG488 to produce DG491. DG491 was transformed into BDR123 (BDR123-491) to generate recombinant *B. subtilis* that contained

the diffocin gene cluster lacking orfs1377 (SEQ ID NO:20), 1378 (SEQ ID NO:21), and 1379 (SEQ ID NO:23). The modified diffocin cluster lacking these ORFs expressed active diffocins upon exposure to mitomycin C (Figure 7), as did the wild type diffocin cluster in BDR123-488.

4. Characterization of bactericidal spectrum-determining sequences from multiple diffocins.

[93] A comparison of the Cd16 diffocin locus (SEQ ID NO:1) with that of Cd630 as well as other Cd strains that have been sequenced (QCD-66c26; QCD-23m63; QCD-32g58; QCD-63q42) and Cd4 (SEQ ID NO:61) showed that, with one exception, all of the open reading frames (SEQ ID NOs:1 and 61) shared 89-100% amino acid sequence identity. The exception was ORF 1374. This exceptional sequence was variable among all the sequenced diffocins and although similar in size, shared as little as 30% sequence identity. The position of ORF1374 within the diffocin cluster was consistent with that of a receptor binding domain. The sequences of the ORF1374s of the active diffocins that were isolated were determined and it was found that they too were highly variable in sequence (SEQ ID NO:17, 49-53). A comparison of these sequences is shown in Figure 8. Furthermore, the spectra of the isolated diffocins (Figure 2) reflected the similarities or dissimilarities of the ORF1374 amino acid sequences, Figure 8. For example, the sequences of ORF1374 of dif16 (SEQ ID NO:17) and dif126 (SEQ ID NO:52) differed by only 1 amino acid and their bactericidal spectra were nearly identical. Whereas the sequences of ORF1374 of dif16 (SEQ ID NO:17) and dif108 (SEQ ID NO:50) differed by 188 amino acids, and their bactericidal spectra were very dissimilar with little overlap. For this reason and because ORF1374 was the only variable protein in the gene cluster, it was concluded that ORF1374 was the target recognition determinant and responsible for the unique spectrum of each particular diffocin.

5. Cloning and Expression of Dif4 in *B. subtilis*.

[94] The diffocin 4 locus was cloned from Cd4 by methods similar to those for diffocin 16. However, some modifications were required due to the absence of an EcoR1 site within the dif4 gene cluster, SEQ ID NO:61. Plasmid SW251 (see Example 2 above) was modified to have an XhoI site in the polylinker using oligos DG211, SEQ ID NO:57 and DG212, SEQ ID NO:58, to introduce NotI and AscI sites, respectively. This created vector DG577.

[95] The diffocin cluster from Cd4 DNA was amplified in three fragments. The first used primers DG210 (SEQ ID NO:59) and AV1288 (SEQ ID NO:37) to introduce XhoI and NcoI sites. The second used primers DG209 (SEQ ID NO:60) and DG15 (SEQ ID NO:47) to introduce NcoI and AscI sites. These two were cloned into DG577, previously cut with XhoI/AscI to create DG578. The third fragment was amplified using AV1368 (SEQ ID NO:35) and AV1289 (SEQ ID NO:36) to introduce XhoI and NotI sites and cloned into DG578, previously cut with XhoI and NotI to create **DG579**. This latter construct containing the dif4 cluster (SEQ ID NO:61) was the equivalent of DG491 for dif16, i.e. it lacked orf1377, orf1378 and orf1379, the unnecessary presumed regulatory sequences downstream of the structural genes for diffocin. The integration vector for introducing the dif4 cluster into *B. subtilis* was made by taking the NotI AscI fragment from DG579 and cloning it into DG487 (see Example 2 above). This constructed plasmid was **DG580**.

[96] To express dif4 in *B. subtilis*, DG580, which contained the dif4 locus without orf1377-1379 (SEQ ID NO:61), was recombined into the *Bacillus subtilis* genome. The recipient *Bacillus subtilis* strain was BDR123, which had a chloramphenicol resistance marker inserted within the *amyE* gene. When this strain was transformed with DG580, recombination occurred between the front and back *amyE* sequences within the vector and the genomic *amyE* sequences. This resulted in insertion, into the BDR123 genome, of all of

the sequences between the front and back *amyE* regions of DG580 including the diffocin locus and the spectinomycin resistance gene. Successful recombinants were spectinomycin resistant but had become chloramphenicol-sensitive due to loss of that genomic marker as a result of the recombination event. This *B. subtilis* strain was termed BDR123-580.

[97] This integrated dif4 locus included all of the regulatory genes required for normal diffocin expression in *C. difficile*, and as expected and shown previously for dif16 in *B. subtilis*, dif4 particle production was induced in BDR123-580 by contact with mitomycin C, Figure 9. Thus, the present examples provide the cloning of the genetic loci for both dif4 and dif16 and the expression of each in *B. subtilis*, an example of a non-pathogenic aerobic production bacterium.

6. Orf1374 Determines the Bactericidal Spectra of Diffocins.

[98] Orf 1374 encodes a large predicted polypeptide (~200 kDa) that was shown by mass spectrometry to be part of the purified diffocin structure. When comparing the gene clusters of diffocins 16 and diffocin 4, most of the gene products, particularly those that are predicted to be structural components, are nearly identical at the amino acid level. The major amino acid sequence difference between the two clusters is orf1374. For this reason and other reasons discussed below, it was speculated that this gene product confers the target specificity of the diffocins. To test this, Orf 1374 of dif 4 (i.e., the sequence encoding SEQ ID NO:49) was replaced in DG580 with Orf 1374 from Cd16 (i.e., the sequence encoding SEQ ID NO:17) to create DG587. DG587 was integrated into the genome of *B. subtilis* BDR123 to make a BDR123-587 recombinant, as provided above for dif16 and dif4. The resulting BDR123-587 was exposed to mitomycin and the lysate treated so as to prepare diffocins. The resulting diffocin particles had bactericidal activity against *C. difficile* strain 19145, which was sensitive to diffocin 16, and had lost the ability to kill strain 19137, which was sensitive to dif4 (Figure 9).

This experiment was further refined. Construction of DG587 resulted in Orf1373 being a chimera of Cd4 and Cd16. A construct was made such that Orf 1373 of DG587 was restored to be 100% identical to the original Cd4 1373, SEQ ID NO:78, thus creating a construct that was a clean replacement of only Orf 1374, SEQ ID NO:17. This construct was termed DG603. This construct was integrated into the *B. subtilis* BDR123 genome and induced with mitomycin C as described above. The resulting diffocin particles had bactericidal activity against strains 19099 and 19145 and lost the ability to kill 19137. Thus, the bactericidal sprectra of diffocins was determined by the protein encoded by orf1374, and changing orf1374 changed the bactericidal spectrum of diffocin, as demonstrated herein.

7. Producer cell without PBSX that does not lyse when RecA is activated.

[99] The PBSX prophage is ubiquitous in wild type *Bacillus subtilis*. The prophage when induced is defective in that it possesses a stunted head structure and contains only small, random fragments of DNA. It is under the control of RecA, thus it is induced by DNA damaging agents, e.g. mitomycin C, and other forms of severe stress to the bacterium. When induced it causes lysis of the bacterium and releases PBSX particles. In order to avoid contamination of culture medium with PBSX particles and to eliminate lysis of the *Bacillus subtilis* producer bacteria when the expression of diffocins is regulated by modifying *recA* or *dinR/lexA* activity, the PBSX gene cluster was eliminated from *Bacillus subtilis* BDR11 bacteria.

[100] The PBSX knockout was constructed by following the procedure outlined in Liu *et al.* Briefly, using the primers and overlapped extension PCR techniques used in the Liu paper, the *araR* gene of parental strain BDR11 was deleted and replaced with the neomycin/kanamycin-resistance gene under the *Bacillus* arabinose promoter, P_{araA}-neo^R, to make strain BDG2. This deletion of the *araR* gene was confirmed by PCR and by the conferral of resistance to kanamycin.

[101] Next, a DNA construct was made to delete the PBSX locus itself. To make this construct, the following five PCR products were spliced by overlapped extension PCR into one large product: 1kb of sequence 5' of the *xylB* gene, amplified from BDR11; 1 kb of sequence 3' of the *xylA* gene, amplified from BDR11; a chloramphenicol resistance gene, *cat*, amplified from plasmid pJW034; *araR*, amplified from BDR11; and finally, 1 kb of sequence containing the *xylB* gene, amplified from BDR11. The overlapped extension PCR product was cloned into the XmaI and SpeI sites of pUC19. This construct was then linearized with SacII and transformed into strain BDG2 bacteria, which were plated onto LB agar plates supplemented with 5 μg chloramphenicol /ml. Colonies were picked from this plate and patched onto LB agar plates supplemented with either 5 μg chloramphenicol /ml or 20 μg kanamycin /ml. Strains that were chloramphenicol resistant and kanamycin sensitive were grown for 4 hours in LB broth with no antibiotic selection and then plated onto LB agar plates supplemented with 20 μg kanamycin /ml. The colonies that grew on these plates were tested by colony PCR for the presence of PBSX genes. The deletion of the PBSX gene cluster was confirmed in strain BDG 9 by sequencing PCR products that spanned the site of PBSX genes in wt strain BD123. Further analysis showed that unlike *Bacillus subtilis* strains BD123 or BDG2, BDG9 did not lyse or produce PBSX particles in the presence of 3 μg mitomycin C /ml.

[102] The PBSX deletion strain, BDG9, was transformed with plasmid DG580, to create BDG27. Integration of the Cd4 diffocin cluster was confirmed by spectinomycin resistance. BDG27 was grown and induced with mitomycin C as described above. After 16 hours cells were harvested and lysed with BugBuster™ (Novagen) to break open the cells since, without PBSX, we expected the diffocins to accumulate intracellularly. After lysing the cells with BugBuster™, debris was removed by centrifugation, and the supernatant was tested for bactericidal activity against strain 19137. The diffocin produced by BDG27 showed activity

against Cd19137 but not Cd19099, thus demonstrating that diffocin 4 was produced in this non-lytic, PBSX deleted strain.

[103] The term "comprising", which is used interchangeably with "including," "containing," or "characterized by," is inclusive or open-ended language and does not exclude additional, unrecited elements or method steps. The phrase "consisting of" excludes any element, step, or ingredient not specified in the claim. The phrase "consisting essentially of" limits the scope of a claim to the specified materials or steps and those that do not materially affect the basic and novel characteristics of the claimed invention. The present disclosure contemplates embodiments of the invention compositions and methods corresponding to the scope of each of these phrases. Thus, a composition or method comprising recited elements or steps contemplates particular embodiments in which the composition or method consists essentially of or consists of those elements or steps.

[104] <deleted>

[105] Having now fully described this invention, it will be appreciated by those skilled in the art that the same can be performed within a wide range of equivalent parameters, concentrations, and conditions without departing from the spirit and scope of the invention and without undue experimentation.

[106] While this invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications. This application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the

present disclosure as come within known or customary practice within the art to which the invention pertains and as may be applied to the essential features hereinbefore set forth.

[107] Although the invention has been described with reference to the above examples, it will be understood that modifications and variations are encompassed within the scope of the invention as defined by the following claims.

REFERENCES

[108] Anastasio, KL, JA Soucheck, and H Sugiyama, 1971. Boticinogeny and Actions of the Bacteriocin. J. of Bacteriology 107: 143-149.

[109] Bartlett JG, Onderdonk AB, Cisneros RL, Kasper DL. 1977. Clindamycin-associated colitis due to a toxin-producing species of Clostridium in hamsters. J Infect Dis. 136:701-705.

[110] Bartlett JG, Chang T, Taylor NS, Onderdonk AB. 1979. Colitis induced by Clostridium difficile. Rev Infect Dis 1:370-8.

[111] Bartlett JG. 2002. Antibiotic-associated Diarrhea. N Engl J Med 346: 334-9.

[112] Bartlett JG, 2007. Clostridium difficile: Old and New Observations. J Clin Gastroenterol. 41 Suppl 1:S24-9.

[113] Benson L, Song X, Campos J, Singh N. Changing epidemiology of Clostridium difficile–associated disease in children. Infect Control Hosp Epidemiol. 2007;28:1233–5.

[114] Blackwell, C. C. and J. A. Law. 1981. Typing of non-serogroupable *Neisseria meningitidis* by means of sensitivity to R-type pyocins of *Pseudomonas aeruginosa*;

[115] Blackwell, C. C., F. P. Winstanley, and W. A. Telfer-Brunton. 1982. Sensitivity of thermophilic campylobacters to R-type pyocines of *Pseudomonas aeruginosa*. J. Med. Microbiol. 15:247–251.

[116] Bradley. Bacteriocins. Bacteriol. Revs. 31:230-314, 1967.

[117] Campagnari, A. A., R. Karalus, M. Apicella, W. Melaugh, A. J. Lesse, and B. W. Gibson. 1994. Use of pyocin to select a *Haemophilus ducreyi* variant defective in lipooligosaccharide biosynthesis. Infect. Immun. 62:2379–2386.

[118] Coetzee, H. L., H. C. De Klerk, J. N. Coetzee, and J. A. Smit. 1968. Bacteriophage- tail-like particles associated with intra-species killing of *Proteus vulgaris*. J. Gen. Virol. 2:29–36.

[119] Daw, MA, and FR Falkiner, 1996. Bacteriocins: nature, function and structure Review Article. Micron 27:467-479.

[120] Ellison, JS and JA Kautter, 1970. Purification and Some Properties of Two Boticins. J. of Bacteriology,104: 19-26.

[121] Filiatrault, M. J., R. S. Munson, Jr., and A. A. Campagnari. 2001. Genetic analysis of a pyocin-resistant lipooligosaccharide (LOS) mutant of *Haemophilus ducreyi*: restoration of full-length LOS restores pyocin sensitivity. J. inhibition Bacteriol. 183:5756–5761.

[122] Fortier, LC and S Moineau, 2007. Morphological and genetic diversity of temperate phages in *Clostridium difficile. Appl Environ Microbiol.* **73:**7358-7366.

[123] Goh, S, PF Ong, KP Song, TV Riley and BJ Chang, 2007. The complete genome sequence of *Clostridium difficile* phage phiC2 and comparisons to phiCD119 and inducible prophages of CD630. *Microbiology*, 153**:** 676-685.

[124] Govind, R, JA Fralick, and RD Rolfe, 2006. Genomic organization and molecular characterization of *Clostridium difficile* bacteriophage phiCD119. *J. Bacteriol.* 188:2568-2577.

[125] Jabrane, A., A. Sabri, P. Compe`re, P. Jacques, I. Vandenberghe, J. Van Beeumen, and P. Thenart. 2002. Characterization of serracin P, a phagetail-like bacteriocin,

and its activity against *Erwinia amylovora*, the fire blight pathogen. Appl. Environ. Microbiol. 68:5704–5710.

[126] Kageyama et al. Life Sciences 9:471-476, 1962.

[127] Kageyama, M., K. Ikeda, and F. Egami. 1964. Studies of a pyocin. III. Biological properties of the pyocin. J. Biochem. 55:59–64.

[128] Kageyama, M., K. Ikeda, and F. Egami. 1964a. Studies of a pyocin. I. Physical and chemical properties. J. Biochem. 55:49–53.

[129] Kageyama, M. 1975. Bacteriocins and bacteriophages in Pseudomonas aeruginosa, p. 291–305. In T. Mitsuhashi and H. Hashimoto (ed.), Microbial drug resistance. University of Tokyo Press, Tokyo, Japan.

[130] Keel, K, JS Brazier, KW Post, S Weese and JG Songer, 2007. Prevalence of PCR Ribotypes Among Clostridium Difficile Isolates from Pigs, Calves, and Other Species J. Clinical Microbiology, 45: 1963-1964.

[131] Kingsbury, D, 1966. Bacteriocin production by strains of *Neisseria meningitidis*." J Bacteriol. 91:1696-9.

[132] Krogh, S,, M O'Reilly, N Nolan and KM Devine, 1996. The phage-like element PBSX and part of the *skin* element, which are resident at different locations on the *Bacillus subtilis* chromosome, are highly homologous. Microbiology 142: 2031-2040

[133] Liu S, Endo K, Ara K, Ozaki K, Ogasawara N. 2008. Introduction of marker-free deletions in Bacillus subtilis using the AraR repressor and the ara promoter. Microbiology. 154: 2562-70.

[134] Loo VG, Poirier L, Miller MA, Oughton M, Libman MD, Michaud S, et al. A predominantly clonal multi-institutional outbreak of *Clostridium difficile*–associated diarrhea with high morbidity and mortality. N Engl J Med. 2005;353:2442–9.

[135] McDonald LC, Killgore GE, Thompson A, Owens RC Jr, Kazakova SV, SambolSP, et al. An epidemic, toxin gene-variant strain of *Clostridium difficile*. N Engl J Med. 2005;353:2433–41.

[136] Morse, S. A., B. V. Jones, and P. G. Lysko. 1980. Pyocin of *Neisseria gonorrhoeae*: mechanism of action. Antimicrob. Agents Chemother. 18:416–423.

[137] Muto CA, Pokrywka M, Shutt K, Mendelsohn MB, Nouri K, Posey K, et al. A large outbreak of *Clostridium difficile*–associated disease with an unexpected proportion of deaths and colectomies at a teaching hospital following increased fluoroquinolone use. Infect Control Hosp Epidemiol. 2005; 26:273–80.

[138] Nieves, BM, F Gil and FJ Castillo, 1981. Growth inhibition activity and bacteriophage and bacteriocinlike particles associated with different species of *Clostridium*. Can. J. Microbiol. 27: 216-225.

[139] Pépin J, Valiquette L, Alary ME, Villemure P, Pelletier A, Forget K, et al. *Clostridium difficile*–associated diarrhea in a region of Quebec from 1991 to 2003: a changing pattern of disease severity. CMAJ. 2004;171:466–72.

[140] Scholl, D, and DW Martin, Jr., 2008. Antibacterial efficacy of R-type pyocins towards Pseudomonas aeruginosa in a mouse peritonitis model. Antimicrob. Agents Chemother. 52:1647–1652.

[141] Scholl, D, M Cooley, SR Williams, D Gebhart, D Martin, A Bates, and R Mandrell, 2009. An Engineered R-Type Pyocin Is a Highly Specific and Sensitive Bactericidal Agent for the Food-Borne Pathogen *Escherichia coli* O157:H7. Antimicrob. Agents Chemother. 53: 3074–3080.

[142] Strauch, E., H. Kaspar, C. Schaudinn, P. Dersch, K. Madela, C. Gewinner, S. Hertwig, J. Wecke, and B. Appel. 2001. Characterization of enterocoliticin, a phage tail-like

bacteriocin, and its effect on pathogenic *Yersinia enterocolitica* strains. Appl. Environ. Microbiol. 67:5634–5642;

[143] Sunenshine, RH & LC McDonald, 2006. Clostridium difficile-associated disease: New challenges from an established pathogen, Cleveland Clinic J. of Medicine, 73: 187.

[144] Williams, S., D. Gebhart, D. W. Martin, and D. Scholl. 2008. Re-targeting R-type pyocins to generate novel bactericidal protein complexes. Appl. Environ. Microbiol. 74:3868–3876.

[145] Wood, HE, MT Dawson, KM Devine, D J McConell, 1990. Characterization of PBSX, a Defective Prophage of Bacillus subtilis. J Bacteriology 172: 2667-2674.

[146] Zilberberg, MD, Tillotson, GS and McDonald, LC., 2010 *Clostridium difficile* Infections Among Hospitalized Children, United States, 1997–2006. Emerging Infect Dis 16: 604-609.

[147] Zink, R., M. J. Loessner, and S. Schere. 1995. Characterization of cryptic prophages (monocins) in *Listeria* and sequence analysis of a holin/endolysin gene. Microbiology 141:2577–2584.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule encoding a non-natural R-type bacteriocin with bactericidal activity,

wherein the non-natural R-type bacteriocin is constructed from a natural R-type bacteriocin encoded by a genome a first strain of *C. difficile*, wherein in the non-natural R-type bacteriocin, the receptor binding domain (RBD) present in the natural R-type bacteriocin encoded by the genome of the first strain of *C. difficile* has been replaced with a different RBD, wherein the different RBD is: (1) an RBD of an R-type bacteriocin of a *C. difficile* strain different from the first *C. difficile* strain; or (2) an RBD of a bacteriophage that infects *C. difficile*,

wherein the non-natural R-type bacteriocin encoded by said nucleic acid molecule comprises polypeptides that are at least 80% identical to the polypeptides consisting of SEQ ID NOs: 4-15 or consisting of SEQ ID NOs: 66-77,

wherein the nucleic acid molecule further comprises a heterologous sequence encoding the RBD of the R-type bacteriocin of the *C. difficile* strain different from the first *C. difficile* strain, or encoding the RBD of the bacteriophage that infects *C. difficile*,

and wherein the non-natural R-type bacteriocin encoded by the isolated nucleic acid molecule has a bactericidal spectrum that is different from the natural R-type bacteriocin encoded by the genome of the first *C. difficile* strain.

2. The isolated nucleic acid molecule of claim 1, wherein the first strain of *C. difficile* is Cd16 and the nucleic acid molecule comprises SEQ ID NO: 1, or wherein the first strain of *C. difficile* is Cd4 and the nucleic acid molecule comprises SEQ ID NO: 61.

3. The isolated nucleic acid molecule of claim 1, wherein the non-natural R-type bacteriocin comprises a polypeptide of SEQ ID NO: 16 or SEQ ID NO: 78.

4. The isolated nucleic acid molecule of claim 1, wherein the RBD encoded by said heterologous sequence comprises a polypeptide at least 80% identical to a polypeptide selected from the group consisting of SEQ ID NO: 17 and 49-53, or comprises a polypeptide at least 80% identical to a polypeptide selected from the group consisting of SEQ ID NOs: 54-56.

Date Reçue/Date Received 2021-02-08

5. The isolated nucleic acid molecule of claim 1, wherein the non-natural R-type bacteriocin comprises the polypeptides of SEQ ID NOs: 4-16, 18, and 19; or comprises the polypeptides of SEQ ID NOs: 66-80.

6. The isolated nucleic acid molecule of claim 1, wherein the RBD encoded by said heterologous sequence is a native RBD selected from the group consisting of SEQ ID NOs: 17 and 49-53, an RBD region of a polypeptide selected from the group consisting of SEQ ID NOs: 54-56, or a sequence at least 80% identical to an RBD region of a polypeptide selected from the group consisting of SEQ ID NOs: 54-56.

7. An isolated non-natural R-type bacteriocin encoded by a nucleic acid molecule of any one of claims 1 to 6.

8. An expression cassette comprising a nucleic acid molecule of any one of claims 1 to 6.

9. The expression cassette of claim 8, wherein the nucleic acid molecule is operably linked to a heterologous promoter.

10. The expression cassette of claim 9, wherein the promoter is inducible or repressible.

11. The expression cassette of claim 10, wherein the promoter is induced by a small molecule inducer or de-repressor.

12. The expression cassette of claim 9, further comprising a *recA* gene encoding a constitutively active RecA protein and under the control of a heterologous promoter responsive to a small molecule inducer or de-repressor.

13. A producer cell comprising the expression cassette of any one of claims 9 to 12.

14. A producer cell comprising a nucleic acid molecule of any one of claims 1 to 6 within a chromosome of said producer cell.

Date Reçue/Date Received 2021-02-08

15. The producer cell of claim 13 or claim 14, wherein the cell is a non-pathogenic and aerobic bacterium.

16. The producer cell of claim 15, wherein the non-pathogenic and aerobic bacterium is a species from a genus of bacteria selected from the group consisting of *Bacillus, Lactobacillus*, and *Listeria*.

17. The producer cell of claim 16, wherein the bacterium is *B. subtilis*.

18. The producer cell of claim 17, wherein the bacterium lacks a PBSX gene cluster.

19. An isolated non-natural R-type bacteriocin having bactericidal activity against at least one strain of *C. difficile*,

wherein the non-natural R-type bacteriocin is constructed from a natural R-type bacteriocin encoded by a genome of a first strain of *C. difficile*,

wherein the non-natural R-type bacteriocin comprises a base plate attachment region (BPAR) from the natural R-type bacteriocin encoded by the genome of the first strain of *C. difficile,*

wherein in the non-natural R-type bacteriocin, a receptor binding domain (RBD) present in the natural R-type bacteriocin encoded by the genome of the first strain of *C. difficile* has been replaced with a different RBD, wherein the different RBD is: (1) an RBD of an R-type bacteriocin of a *C. difficile* strain different from the first *C. difficile* strain; or (2) an RBD of a bacteriophage that infects *C. difficile,*

and wherein the BPAR comprises a polypeptide of 50 or more contiguous amino acids at least 80% identical to SEQ ID NO: 16 or 78.

20. The isolated non-natural R-type bacteriocin of claim 19, wherein said different RBD is at least 80% identical to a polypeptide selected from the group consisting of SEQ ID NO: 17 and 49-53, or at least 80% identical to a polypeptide selected from the group consisting of SEQ ID NOs: 54-56.

Date Reçue/Date Received 2021-02-08

21. A method of producing a non-natural R-type bacteriocin, comprising exposing a producer cell comprising a nucleic acid molecule according to any one of claims 1-6, operably linked to an inducible promoter, to an inducing agent in a concentration effective to induce expression of the non-natural R-type bacteriocin, and purifying the expressed R-type bacteriocin.

22. The method of claim 21, wherein the nucleic acid molecule encoding the non-natural R-type bacteriocin is heterologous to the genome of the producer cell, and wherein the nucleic acid molecule is contained within a chromosome of the producer cell, or is contained in an extrachromosomal expression vector within the producer cell.

23. The method of claim 21 or claim 22, wherein the producer cell is a nonpathogenic and aerobic bacterium.

24. The method of claim 23, wherein the non-pathogenic aerobic producer cell is a species from a genus of bacteria selected from the group consisting of *Bacillus, Lactobacillus*, and *Listeria*.

25. The method of claim 22, wherein the producer cell is *B. subtilis*.

26. The method of claim 25, wherein the producer cell does not lyse when induced to produce the non-natural R-type bacteriocin.

27. The method of claim 25 or 26, wherein the producer cell lacks a PBSX gene cluster.

28. Use of the non-natural R-type bacteriocin of claim 7, 19, or 20 for killing *C. difficile.*

29. Use of the non-natural R-type bacteriocin of any one of claims 7, 19, or 20 for killing *C. difficile* in an animal.

30. The use of claim 29, wherein the animal is a mammal.

31. The use of claim 30, wherein the mammal is a human.

Date Reçue/Date Received 2021-02-08

32. Use of the producer cell of any one of claims 13 to 18 for treating a *C. difficile* infection in an animal.

33. The use of claim 32, wherein the nucleic acid molecule encoding the non-natural bacteriocin is under the control of a *lac* promoter.

Date Reçue/Date Received 2021-02-08

Figure 1. SEQ ID NO:1-61

SEQ ID NO:1 >Dif16 locus

GGCCGCAATACCCACTACACCTTCgTCATCTTTAAATTTAAGAGTTTTTACTATTGAATAAT
AAAGGTATATTCCAGTAAAAATAATCTTTAAATACAAGAAAAATAAACTCTTTGGGTATATT
AAAAAGCTAAAAAGTGTAAATATAAAAGCAAGTAGAGTACTTATCCTGTAAAAGAAATCTAT
TTGTGTAATGTCTTTATATTTTATCATAAACACCGAATATAAAATGATGAAAATAATTGCGA
CGATTGCATATATGGTAAATAACATATTTTCAAGAGTACCATTTGAAATTACTATCCACTTA
TACCACATAATTGGCCAAAATAATAGTGCTAAGAACTTAAAATAATTATCAAACAACTTTTC
TTTATACATTCATCAAACAACCTTTCTTAACAAAAGCATATATTTGTTTTTAGAATTTTAAA
TAATATGATATCATTATTATATATTAATATTGAATTTATAGAAACCAAAATTTGTTAAAATA
AATATATAGATTTTACTGTTAAGCCAGTTAAAATTACTACTATTTTTATTATGAAATTGGAT
CAAATATGTAGAAATACGGCAAATTAGTTAATATTAAATATTTATTATTTCCAAGTTGTAAA
GACTGTTTTTTTAATGATAAAAATTCTAATCTTTTTTGAAAGAAAGTAATATCCACATTAAG
TATGTCTGCCATTTCATAAACGCAAGTGATGCCAGAATTAATTATGTTTATTATATCTTCTT
CAGTAATTAAGAACTCACAAGCCCATTTTAAGGCTTTATTTTCGCACTTATCTATAATAATT
TTTGTATAATAATCGTTATAAGAGGATACATAGTATCCAAGGCTAGTGAAATGATGTCCAAG
TTCTTCAGCTAAGATGGATGTCAATTTTTTTGAGTTTTGTTTTAAATTACTGAGTAATGATA
TAATTTTAATACCATGTTTGTTTATATATAGCCCTTCTAAATCACCTGCAATATAAGTGGTA
TAATGAATTATTATCTCTTCTTGAGAAGCTAATTCAAAAAGCTTATCCAAATTATTCATAAA
AATCCCCCTAAAATAGAATGTATGTTTGCCTTTAAATTATATTAAAAGAGCAGAAAAATAGA
CTGCTCATCATATGGTTTATTTTTTTTTATATTTATTTAGTAAAAATTCTATATAATCATTA
AGTTGTTCTTGTGCTTCTTCAGGTAACTCTTCATGTGGATTTTTTCTATGTGCAGCTACTGT
ATCAATATTTTCCTTAACTAAGGTTCTTCCAAGAAGGTAATCAACTGATACATTAAATACAT
CAGCCAATTTGTTTAAAATGTGTTCATCAGGAAATCTGTTTTCTGTTTCATAGTACCCTAAG
ACTCTTTGGGAAACGCCTACTTTTTCTCCAAGTTCTCTTTGAGTCAATCCAAATTCCTTTCT
AAGTTCTCTTAATCTTTTGGCAAACATTATAACACCACCTTATGTATAGATTATAACAAATT
GTTCTAAAAAATAAAACTAATAAAATATAAAAGAATATTTTTTCTAAAATCTATTGATAAAG
AACAAATGATTCTATATAATCTAAGTGTGGAAGAACAAAATATTCTTAATGGTAATGGAGGT
ATAAAACAATGTTTAAAAATAACTTGAAATATTATAGAAAATGCAAAGGTATGACACAAATT
CAACTTGCCAGAAAGGCTGGAATTACAAATGACTATATATCTCAAATAGAAAGAGGTATAAA
AAATCCTGGTCTTCTTATGGCTAAGAAGATTTCTAGTATTTTAGAACAAAATATAGAAGAAG
TTTTTTTTATACAGTTATAGAACAATATGTTCTTGAAAGTTGTGAGATTAGTAAAAAACTGT
GCACTAAAGAGATTATTGTAAATTTGAAGCTAATAATAAGTATATAAAAAAGGAGAAGTACT
ATGGAAAACAAAAAAGATATATTATTTAAAGAAACAGATGAAAGATTACATAATTATAAGTA
TTTGGATATAAAGATAAAGAATATTAATTTGGACATAAAAAGATGTGAGAATGAATACTCTG
GATGTGGAGCAATGGTATATACAGAAAAGACTAGTAACACATATAACATAAGTTCTTCTGTG
GAAAATGAGGTGTTAAAAAGAGAGGAAAGATTAAGAAAATTAAAAATGGAAAAAGAAGATAT
AGAAATAGAAAAAGAGAAGATAGAAAATGCTCTAACATGTCTAAATGATATAGAAATGGAAT
TTTTTAATCTTTTTTATAATAGTAAGACAAAAAACAATATGACATATATTTCTATGAAACTA
CACTTAGATAGAACATCTTGCTACAATTTAAAGAAAAAAATGATATTTAAATTGAGTGAGAT
ATTATAAAAAATATGACAACTTTACAACACTTTATATACACTATTGCAACACTAGGCAATAA
AATATGTGAGATAATGTTATTGTGAAAGAAATCCATATTGAAGGAGGTGATAAATTGAAAAG
AATAATATTACCTAAAAATATAGAAGATATTTGACAGGAATAAATGAGATGTATATTTAAAA
ATGACTTATATCATTTATAGTAAGATTATCAGATTAAGCAAGAATATTTAGTGATAGTGTGG
TGATTATTTGCTTAAATACAAGGAAATATTAGAAACAATTATTGAGATTCTCAAAAAAAACT
TTACTGAAAGTATTTTTATTGATGATGAAAGTGTGCAAGGCTCTGAAGGGTCTTGTTTTTTT
GTAAGTATACTATCAGTTATTTGTACACCTATAATGTTAAATACGAATAATAAAGATATTGT
TATCTCTATAAAATACTTACCAAAACCACAGTCAAAGAGTATTAGAATGTATGAAATTTCAG
ATGAATTAAATAAGTTATTCAACAGAAATATAAAGGTAACAGACAGAAAATTAAATATAACA

AAGCTAGAACAAAGTATTAAAAAAGAAGAGTCAATTTATGTATTGAACTTTACAATTACACT
AAATTATCTGGATAGTGTATATGAAGAAGATGTAGTATATGAAAATATGGAAGAAATCAATT
TAAATTTAGGAGAGTGATAGTATGGCTATAGGATTACCAAGTATCAACATATCATTTAAGGA
GCTAGCTACAACTGTTAAAGAACGTTCAGCTAGAGGAATAATTGCAATGGTGCTTAAAGATG
CTAAGGCACTAGGTCTTAATGAAATACATGAAAAAGAGGATATACCAGTTGATTTATCTGCT
GAAAATAAAGAGTATATAAATTTAGCTTTGATGGGAAATGTTAACACTCCAAATAAATTATT
AGTTTATGTAATAGAAGGAGAAGCAGATATTCAAACTGCATTAGATTTTTTAGAGACTAAGG
AATTTAATTATCTATGTATGCCAAAAGCAGTAGAAGCTGATAAGACTGCTATAAAAAATTGG
ATAATTAAACTTAGAGATATAGATAAGGTTAAGGTTAAAGCTGTATTAGGAAAAGTTGTAGG
AAATCATGAAGGGATAATTAATTTTACTACAGAAGATGTGTTAGTTGGAGAAAAGAAATACA
GTGTTGATGAGTTTACAAGTAGGGTGGCTGGACTTATAGCAGGAACACCTTTAAGTCAATCA
GTAACTTATACTAAGCTTAGTGATGTAGTTGATATACCTAAGATGACGAAAGTTGATGCAGA
ATCAAGGGTTAATAAAGGAGAGCTTATACTTATTAAGGAAGCAGGGGCTATAAGAATTGCAA
GAGGAGTAAATTCTTTAACTGAGTTAACAGAAGAAAAAGGAGAAATGTTCCAGAAAATAAAA
ATAGTTGACACTTTAGATATTATACATAGTGACATAAGAAAGGTGATAATAGATGACTATAT
AGGAAAGGTTACTAACAGTTATGACAACAAATGTTTATTGATAGTAGCTATAAAAAGTTATT
TAGAAGAATTAGAAAAGTCAGCACTTATAGAATCTGATTCTACTGTTGAAATAGATTTTGAA
GCACAAAAATCGTATTTAAAATCAAAAGGAGTAGATTTATCTTATATGACATTACAAGAAAT
AAAAGAAGCTAACACAGGTTCTAAAGTATTTTTAAAAGCAAAAATAAAAGTACTTGATGCTA
TGGAAGATATAGATTTATCAATAGAAATATAGGAGGATTATTAATATGGCAAATATGGAAGC
TAGAAATGTAATGAGTGGTACTTGGGGAGAACTTTGGCTTGATGGAAACAAAGTAGCAGAAG
TAAAGAAGTTTCAAGCAAAGATGGAATTTACAAAAGAGGATATTATAATAGCAGGTCAAATG
GGTACTGATACAAAGTATATGGGATATAAAGGAAAAGGTTCAATAACTCTATACCATGTTAG
TTCAAGAATGCACAAGTTAATTGGAGAAAAGATAAAGAGAGGTTCTGAACCTAGATTTGTTG
CTATATCTAAATTAAATGACCCAGATTCTTATGGAGCAGAAAGAATAGCAGTAAAAAATATA
GCATTTGATGATTTAACTTTAGCTGATTGGGAGGTTGGAGTAAAAGGAGAGATAGAAGCACC
TTTCACATTTACTGAGTATGATTTTCTTGATATAATTTAGTTTTATATTTGGTTTTATACTG
ATATTTAGTAGATATATACTTAATAAATTTAGGTAGTTAATAAGTAAAAAAGTTAGTTGATT
GAATTTGATTGATAAAGGAGCAAATAATAATGAATGAAAATGGATTATCAAAAAATATAAAC
ATAGTAGATTTACTTTTAAATGCAGATACAGAAAACTTAGAAAGACCAAGTACTATAGTTGA
ACTTAAGAGATTATCAACTATATTTGGGCAGGAATTTAAAGTAATGTGTAGAGCTTTAACAA
TAAGTAAAGATGAAGAGATACAAAATACTTGTCTTAAAATTGATGAAAATATGAAAACGGAT
ATAGACTTACCGGAGATGCAGATGCTTACAATTATAGAAGGTGTTTGTGATTTGGATGGAAA
GCTTTTATTTAAAAATAAGGAACTAATGGATAAATTTAAGGCTCCAACACCAAAAGAATTGG
CAAGAAAACTATTATTACCAGGTGAAATTACCAACCTATATAGAATACTTCAAGATGTTATG
GGTTATGGTAAAAATGCAGTGATAGAAGAGGTAAAAAACTAATAGGGACGGATACCAAGACT
ACAATAATGTACTATTATTGGAAGAAAAAAGGTATAAGACCGTCCCTTTTTTATGCAATGGA
TAAAGGCGAATTAAAGCTTATTGAAGCTTTTTTCGCCTTAGAAATTGAGGAAGAAGTTGAAA
AAATGAAACATGGATATGGAGTGTGTCCTTTGACAGGAGGTGGTATGTAATGGGAAATGTGA
GAGAAGAAGGTATAAATATGTATCTTACAGATAATTACACACCAAAAATGAACCAAATTATA
TCAGTAACTGATAATTTTAGGAGAGCAACTGTGGCTGTTTCACTTTCCACTAATGTAATGGC
TAGTAGCATAAAAAATTCTATTGGAAGTGCAAGTAGTAGAGTAAACAGTTTAAATTCCTCGT
TAAGAAAAGTTCAAACTACTGCTAGTAGTGTAAGTTCAACTATGGCAAAATTAAGTTCTAGC
ATAAATGCTGTTTCAGGAGTTATTGGAAGTTTAAATGGAAGTATTATGAGACTAGCAATAAC
TATAGCTATGATTATTGATTATTTTAATAAGTTGATTCAAAAGAAAAATGAGTTTAATTCAA
ATATTATGATTATATTAATATTTAAAGCTAAAAGTGATGAAGTAGAAAAAACTAAAAATAAA
TTACTTGGAAATTTAAAAAAGATTGGTGGCAAGATTTGGAATATCGTAATAAAAGCAAAAGA
TATGACTAAGAGAGTGATAAGTAGTATCTTGGGAAAATTAAAACGAGTAGAGAAACGTCCTT
ATCAAGGAAGTATTAATCTTAAAGATATGGTAAGTAGTGCTATGGCTAGAATTTTGCCTAAG
TTAATGTTGTTTAAAAATACTTTTTGGAGTGGTGTAATAGCTATAAAAGATATGGCAAGTAG
CATTATAAGTAAAGTATTTCCCAAATTGAGATTGTTTGCAGGTAAGGTATGGAGTGGTGCAA

TAGCTGTAAAGGATATGGCAAGTGGAATACTTGGTTCGATAAAAGGGAAGATATCTGATTTG
ACAAATGGTGCTACTATAGGTGTCGCTGTGAAAAAGGGTGTTGACTTACTTGGTCAGGAACA
AAATCAGAAAGTTGTTCTAGAAAGTGTAATGAAAAGAAATACTGGAAAAACTAGCCAAAAAG
ATGTTGATAAGTATTATGACAGTTTAGTAAATATGGCAAATGATACGCCTTTTGACCCTGAA
GATGTTGTTGCAATGGGAACTAAAGCTAAAATGATTAGTAATATTACTGGTGGCAAAAAAGA
AAAAGATATAACTCAAGCTATGGTAGATGTTAGAGCTTTAAATATGAATACAAGTAGTGAAC
AAGATGTATCAGCAGCTTTCTTAAGTGCAGCAAAAGGAAATATGGAATCTCTTAATACTCTG
GTAGGAGAAAATTATAAAACTTTTGATGAAGCATTGGAAGGCATAAGTGTAAAGCAGATGGG
GTTAGCTAAAGAAATGAGTAATACAATACCAGGTATAATATCAGGAGCTCAAACAAGCATTA
ACAATGGTTTGAAGAGTATTGTTAAACCTTTTGATGATATTTTAGGTCAAGGACTAAAGAAA
ATAAAAACTTTTATAGAAAGTGGATTAGGGAATTTAGCTGGCTTATCTGAAAAAATGGCTGG
TAAAATAGGCAATGTAATGAATGGTAAGATAATTATTGGCAACAAATATGACCAGATGCAAT
CTAGAAGTGTAAAAAATGGAAAAGAGTTTTCTGATTCTACTCAATATCGAATTTCTAATGAG
GCTGAAAAGCGTAAAATGATGGTTGAAAATAAGCAAGAACGTTTTGAAAATCATGCAGCAAC
AATGATAGGGAATGCACCAAAAGCAATTGTTAACGCAGGAAGTACACTATTACAAAATATTG
ATTTTACAGCATTAATAGATTCACTACTTCCAGTAGTAAACTTAGTAAATAATTTACTAGAT
AGTATAAACAATAAATCACCAATTGCACAAGGATTAATAAGTATATTTGGTACAATAGTAAC
TACAGCATTCCAACTAATCGGACCTGTAGTTGAAGCTGTTAGTCCTATTATCACAAGAATTT
TTACTTTTTTAGGTGAATATGCACCTCAAATAAACAATTTTATAGAGACACTGGGTGTTATT
TGGAAAACTGTATGGGAGACCTTAGGACCTCTGTTGGAAACTGGATGGAAAATTATAGAGCC
AATATTGGGAGCTTTTTTTAACATATTAGATAAAGTATGTAAAATAGTTAAAGATATATGCA
AATGGTGGCAAACTATGATTAATAAGATAAAAAATGGAAGCATCACAGGAACAGTTTTAAAT
CTAGTGGAAAAGAGTAAAAAAAATTACAAAGATAATCCATATGCTGGAACAAAGGCTGGTGA
TTCTGGTAAAGCTTATTCAAGTAAGAAAGGTAATAATGCATTTGGATTGAACTATGTTCCTT
ATAATGACTATCAAACCAGACTCCATGAAGGTGAAATGGTTTTAACTAAACAAGAAGCAAAT
CAATATAGAAGCAGAAAAAATGGTGGAAATATAAACATAGCTAAGTTAGCTGATACAATAGT
GATTAGAGAAGAAGCTGATATAGAAAAGATAACATCAAAATTAGTTGCAAGTATCCAATTGG
CACAGTTAGGGGGTGTCTTATAATGGAAATGTGGCTTAGACAAGCAGAAGATAGATTTAGAT
TTCCAGTATTTCCATCTTCCTTTAGTATTAATGGAAAAGCTGCTGTAAACTCTTCTAGTATA
CTCAAAATAGGTGAAGTAGCAACTTTTGGTGGTGTAGCTCTTAAAAGCATTTCAATATCAAG
TTTTTTTCCAAATAAAGACTACACTTTCTGTGACTATACAGGTTTTCCATCACCATATGATT
GTGTAAATAAGATAGAAAAATGGATGAAGGAAGGTTTTATATTAAGATTTACAATTACGGAA
ACAAATATAAATATGGAAGTCATAATTGAAGGGTTTAGTTATGAAGAAAGAGATGGGAC
TCGAGATGTATATTTTACATTAGATTTAAAAGAGTATAAAAGAATAAAGATACCAAAAGTAA
CTCCAAAACAATAACTATTATAGATAATAAGTTGTAAGTAACTGCTGATAGAATTAAATGAA
AAGGCAGGTGATTTTTTATTATTAAGATTTGGGTACACATAAAAAACGGAAGTATATATGAC
ATAACTGACATAGTAGACAAGGTATCATGGTCAGGTGATTATAAATCTCCATCAAGGACACT
AGAGTTTTCAATAATACAATCATCATTTGATGTAAATTTCCAACAAATCGATATACCAATAG
CTAGTACAGTCTGTTTCTATGTAGATGAGAAAGAACTCTTTAGAGGAATGATAATTAATAGG
TCTAAAGATTCAAGCAGTAATGAAATTAGTTTTGTATCTAAAGATATGGGATTTTTACTTAC
ACAAAGTGAAGTGTCATACAATTTTAAAGATAAGTTAGTTGAAGACATAGCAAAGCAAGTAT
TTGCTGAAAATAGGCTTTCAGTTGGAACAATAGCAAAGACCAATGTCAAGTATACAAAGATG
TTTATAGGAGTAAATGGTTATGACACAATAATGAGTGCATATACAGAGGCAAGTAAAAAGAC
AAAGAAAAAGTATATGATAGAGGCTAATTTAGATAAGTTTAATGTTATTGAAAAAGGAACTG
TTACATTAAGTGTTATGTTTGAAGAGGGATTTAATATTATAAATACCACCTTTTCGGAGAGC
ATGGAAAATGTAAAAAATAAAGTAATAGTGGTAGACCAGTATGGAAGCAAGATTAGCGAAAA
AATAGATAATGAAATTTTTAAGGAAGTAAATGTAATAATGCAAAAAGTAATTCAGCAACAAG
AAAATCAAGATGTAGATATTGATAGCGAGTTTAATGGGATAGAAAAAAGCTGTTCTCTTAAa
GGTTATGGAGATGTAAGTTGTATAACTGGTAGAGGAGTAAAAGTTAAAGATTCTTATACAAA
GCTTGTAGGACTATTTTATATAGATACAGACAAACATACTTGGCAAAATGGAGAATATCAAA
TTGAGCTTGAACTTAATTTTCAAAATCTTATGGATGAAAAGTCAGCAGGACAGGATGAACCT

AAGGAAGAAAGTAATTTAGGGGGAGAAGATTATGCAGGAGGAAAAGAGTTTACAGCAGAATT
TACAGCTTACTGTCCTAGAAAAGAAGAAGGTGGAGATACAGATTGTAGAAAGAAAAAACTTG
ACCCATCTAAAAAACTTGCGCTGCTCCTATGGTTGGTAAATATGAGCAAACTTATTATACAA
AAGAGTTTTTAAATAAACATCCTTTATTGAACTATGGAGATGAAATACAGGTAATTACAGGA
GTTTCTGGTCGTGATGGAGTCTATAAAGTAAATGACGTAGGACCTGCAATAACTATAGAAAA
AAATGGAACATACCATATAGATATTTTATTTGGAAATGTTGAAGAAGCTAGTAAATTTGGAA
GAAGAAAAGGAAAAATTATTATTGGTGGTTATTCTGGTAATGTATCTGATAAAGCTAAAATA
GTAATATCAGAGGCAAAAAAACATCTAGGTAAACCTTATAAATGGGGTGGAAATGGACCAAG
TAGTTTTGACTGTTCTGGTTTAATGGTCTACTGTTTTAAAAAAGTTAATGTTAGTTTGCCAA
GAACGTCAAATCAACAATCTAAAAAAGGCAAGAAAGTAGAACAAAAAAATCTTCAAGCAGGA
GATTTAGTATTTTTTCATAATCCAGTCAGCCATGTTGGATTATATATAGGTAATGGAGAATT
TTTACATGCTCCACAAAAAGGTGATGTAGTTAAAATAAGTAAGTTAAGTAGTAGAAGAGATT
TTAATACAGCTAGGAGAGTATTATAAAAGGATGGTGATATAATGGCTAATCCAATAAATGAA
TTTATAGGAATAATAAGAGAAGAAGGAAAGTATCATAATCAACCTTCTTTTTTATTGGAAAA
TTAAAAGTAAATTACCAGATTTAAAAATAGAGACAAATAACATCATATTAGAAAAAGAAGAT
ATTTTGATAGATAGTTGGATGATTGATAGACAGCTAGAAACATTTGACACAGAAACAAATCA
AGAACACCAGCATGAAGTAAAAAATCCTTTTATAGATAACTTTGAATCTGGGGATATGGTAA
TAATGTTTAGAATAGGCGAAAAATTTGCTGTTGTAAGTAAGTTGGTGAGCTTATAATGAGTA
CAATATTTCCTTTTATAGGTGTCCCAGAGGATTATATCTTACCTAAAACAGAAGAATTGCCA
ATCTTTCGTGAAGTGGCATGGGATTTTGAAAAAGATGAACCTATTTTAGAAAAAGGTGACTT
TAAAATAATTGAAAAAAAAAGAAGCCTTAAAAGTTTGGATATACAAGTGTATAAAGACAAAT
AGATATGAACATGAGATATACTCTTTAGAATATGGGACAGAGCTTTCAGAACTAATAGGACA
AAAATATACAAAAGGTCTTACAGAAAGTGAAGCTAGTAGATTCATAAAAGAGGCCCTTCTAA
TAAATCCATATATATTAGAAGTAAACGTAAAAAGTGCTAACTTTAACAGAGACGTATTGAGT
GCAAATGTAAAAGTATCCACTATCTATGGGGAGGTGGAAATAAATGTATAGTGACCAGACAT
ATGAAGTAATAAAAAATAGAACTCTTGAAAATATTAATCTTGATATTTATAAAGGAGAAGGT
TCTTTTCTAAACAACATGGTATCTGGAAATAATCTAGAACTTTCGAAGATATATCTAGAACT
TTCAAAGATACATAAAATGGCTTTTATACAAGACACATATAACCAGTTTCTTGATAAAAGAG
TCAATGAATTTGGTGTATATAGAAAGTTAGGTACAGAGTCAAATGGAGAAGTTGAATTTATT
GGAGAGAAAGGAACTGTAATAAATAATGGCACAATAATATCATATAGAGATTTACTATTTGT
AGTAATAAAAGATGTAACTATTGGTAGTGAAGAAGGTGACAATAGCCCAGTTCAAGCTCTGG
AAGTTGGTAAGAAATATAATTTACCTACAAATTGTGAATTTAAACTAGTTGATAATATATCT
GGAGTAACAAAGATTACTAACACAAGAAGTTTTGAAGGTGGTACAGATATAGAGACAGATGA
AGAACTAAAAGAAAGATTTTATAAAATCCAAAGAAATCAAGCTACAAGTGGAAATAAAGCTC
ACTATGAAGAATGGGCTTTGGAAGTAGATGGAGTCTATAATGTTAAGGTTTATCCAAGATGG
GATGGTCCAGGAACAGTTAAGGTCTTGATATTTGGGGAAAATAATCAAGCTGTTGATACAGA
AACGATTGAAAGGTGTCAGCAACATATAGATGAAGAGAAGCCTATTGGACCAACTATAACAG
TTGTGACACCATTACCAATAGAAATAAGTATAAGTGCAGTAATGAAACTAGAAGATGGATAT
ACATTAGACAATGTAAAAGAATCTTTCCTAGAAAGTATAAATACATACTTTAGAGATATTAG
AGGAGAGATAATCTATACAAAAGTCATGGGAATACTTATAAATACTACTGGTGTACACGATT
TAAGTAATCTACTTATAAATGGAAGTACAGATAATATAACTATTAATGAAGATAAAATACCT
AGTGTAACAACTGTTAATTTTAGTGAGGTGGAAAATCAATGAAGCTAATTGATAAACTACCA
TCATTTGATAGAAATTACATTGTAGAGGAGATACAAGGTGCATACGATACAGAATTAAATAT
TCTTAAAGAAGATATTGATGATACCTTTAACCAATTATTTGTTGATACAGCGACATGGGGAT
TAGATATGTGGGAAGACATACTCTGCATTGaAAAAAAAGAACTTGATTTTGACACAAGACGT
AGCAATATAAAAGCTAAAATGAGAAGCAGAGGTACTAGTACTATTGAAGTTATAAAAAGTAT
ATGTGAGGCATATACAAAATCAGAAACAGATATAAAAGTTTATAGTGATGAATTTACATTCG
TATTGAGTTTTATAGCAAATAACTGTGACTATAAAACTCTTTTAGATTGTAGCGAGATGATT
GAAAGAGTAAAACCTGCTCACTTATTACACTATTTAGAACCAATAATACTAGATAAAAGTAT
GGTCTATTGTGGTGGAGGTATGGTATGTAGTGAAGAGGTAAAAGTTCATCCATACTTTGAAC
CAATTATAAAATGTAGTGCTGTTGTAAACTGTGGAGCTGGAATGTTAAGTAGAGAAGAAATA

AAGGTTTATCCTTTAAGCATTAAATGCATTGAAAATAATTGTAAGATTAATATAGCTATTGC
AAATGATACAGGCGTAGAAAATGTAGTAGTTTATCCTAAATCGGAGGTGGTATAATTGGAAG
AAAAATTTTATATAATATTAACCAAAATTGGTAGAGAAAAAATAGCAAATGCAACTGCACTA
GGAGAGCTTGTTGGATTAACCAAGTTTCAAGTTGGAGATAGTAATGGAGAATATTATGAGCC
AACAGAGGAACAAACTGCTTTAAAGAATGTAGTTTGGGAAGGAAATATAAATTCTCTAAGAA
TTGATGAAAAAAATCCTAATTGGATAGTTATAGAGACTATTTTACCAGGAACAGTTGGTGGA
TTTATGATAAGAGAAGCTGCTGTTCTGGATAATGAGAATAATATAATAGCTATWGGTAAGTA
TCCAGAGACGTATAAGCCACGTGCTGAAGATGGCAGTATTAAAGATTTGGTTGTAAAAATGA
TTTTACAATTGTCCAATACTTCAAATGTTACATTAGAAGTAGACCCGACGTTGGTTTTTGTA
ACTCAAAAGGATATTCAAGATTTAGATGATAAGTTTGATAAAAATATAAAAGAAATAAAAGT
AAAAATTGGAGATACAGATATATTAACTACAGATTCTAAAGATTTATCAGGAGCTATAAATG
AGGTAGTTAAAAAAATAGAAAATATATCTTTTGATGATGTTATAAGTGGTCAAATACAAACT
GATATATCAGTATTAAAAAATAGCTATAACAAATTATCTGAAAAAGTGCTAGATATATTAAT
ATACCTAGAATTAGAGTCAGAAGTAACTGTAGATGAGGCTGGTTATTGGTATGATACATTAG
CAAATGGAAATAACATAGTAGCTATAGAAGGGCTTAAGTTAGATTTAAATAGAAAATGTATA
ACAGGTGAAATTGGTAATGTGATTTTTAGAGATGTAGTATTACCATTTAGTGCAAATAGAGT
TAGATATATACATGATATGGATAATAACTTTGTTGAGACAAAATCTAGTAACACTTATTTAA
AAGAACAAAAAGATATAACTCTAAGTAAATATTCATATGAAATAAGATAAATAAAGGAGGTA
GTACTAATAATGAAGCAAAATAAACTTTTACAGCGTGGTGCTTATTTTAATGATAAGAACAT
ATTGATTGATGATTTTGATAAAAGATATAATGATTATGATTTTGTAGAATTTTTTACTGGTA
TAAGTAATAGTACCTTTGGTTTAAAATCAGATGGTAATTTATATGCTTGTGGCGATAATACA
GGTTTTCAACTAGGACTTGGAAAAGATTCGTCAGAGAGAAGGATGTTTAGTAAAGTAAAAAT
TGATAATGTAAAATATGTATCTTGTGGTTCAAAACACAGTGTAGCAGTAACTAAAGATGGAT
TTGCATATGGAGCAGGAACAAGTAATGTAGGTCAATTAGGTGTAATTGAGTCTACAGTATAT
TATGAATTTACTAAGCTACCAATAGATGATGTAAAAACTGTTGCATGTGGTTATGACTTTAC
ATTTGTGCTTAAAAATGATGGAACATTATATTCAGCAGGTTTAAACTCAAGTGGTCAACTTG
GACTAGGTGATACTAACAATAGAGCTACTTTTACTAAAGTAAATATAGATAGTGTGAAAGAT
GTAGTGACTTATAATCAATCTGTATTTATCATAAAAATGGATGGGACAGCACATGCATGTGG
ATTAAATTCAAATGGGCAGTTGGGAATTAATAGTACTTTAAATAAAAGTGTATTTAATAAAA
TAGAAGGTATGGATAATGTAAAACAGATAGCGTGTGGTAGTAGTCATACAATTCTTATTAAG
AATGATGGAACTATGTATACTACAGGCTATAATGGAGTTGGTCAGCTTGGTACAGGAAATAA
TAATAATTCAATTGTATTTACTCTTTCTAGTATAAATAATGTTAAGTATGCTTCTTGTGGAA
ATAATCATACTATGATATTAAAATACGATAATACACTGTTTAGTACAGGACAAAACAATTAT
GGTCAACTAGCCAATGCCAATAAAGATGTAGCATCAAGAAATACTTTTGCTAAGGTTAATGT
AGAAAATATAAAAGATATTAAATGTGGTTCTCAATTTAATTTTTtAATAAATGGTTCAAAAG
AGATATTTGTATCTGGCTGTAATTTAGCAGGTCAACTTGGTTCATTTTTTCATACAACTTTT
CTGTATGAGTTTTCAAATGTGCAATCTTCAAATTTAGATAATTATTCAGGTTTATTGGTTAA
TGATGATTATTTATATGTTACAAAGGACAATAGTGAATTTTTAAATGTAAAGTTAAGTGATA
ATTTTCAAGATTATAAGAAGATAGAGTTAACAGATAGCAATATGTTTATTGTTATGAATGAT
GGTACATTGTATGCTTGTGGTTTAAATAATTATGGACAGTTAGGATTGGGAGATACTGTTAA
CAGGTCAGTTATGACTAAGGTGGATATAGATAATGTTTTGGATATAAAAGGAAACGGAAACT
CAACTTTTGTGCTTAAGAATAATGGAACATTATATTCATGTGGTTTAAATAGTAATGGACAA
TTGGGTTTAAGAGATGAAGTTAATAGAAATATATTTACAAAAATAGAAATAGAGAATGTAAA
GGATTTTTGTGTAGGAAGCAATTATGTCATAGCTTTAAATCACTCAAAAGAAGTATATGGAT
GGGGAAATAATCCTTATAATAATATAGAAAAAACTTCTAATTATCCATATAAGCAGGGAATA
AGTAATATTGAAAAGATAGCAGCATATGATTATTCTGTATATATGATAAACAGTGAAGGGAA
ACTATATGTTTCTGGATACAATTATAATTATCAATTAGGTAAAGGAAATAATAGTAACCAAA
GCAAAGCATTAGTATCTCAATGTAGAACAAATTCAACATCTTCTACATCAAATGGACTTAGA
ACGTTACCTAAAATAACTAATGTTTTtCCTTTTTATGATGGTTGTGCAATAATTGACGAAGG
AGGTTATGTTTATTTAACAGGATATCATGGATATTTAAGAACATTAAATAGCAGTCCAAGTA
TATCTGATTATTCAAGATATGGAACTTTTATTGAGGCTACAAATTCAAATCATAATACTTAT

TTTATACAAGAGACTGATTTTAGTGGAATTGAAAAAGTAATAGGGATGTCAAATAATATATT
ATTTTTTAAGAAAGGAAGTTCATATATTACTGGATATCCAAAAACATTTGGCTCAACCATTA
CTGGACATAGAAGTTATACTAGTATTAATTCTGAGAGTTCTAATTTAGGAAGTAATTTTATA
ATATATCATAGTAATTCCAAGTTATATGGAAAAGGGATTGCTAATAGTGGGCAATTTGGGAA
TTCAACAAATATAGATGGCACAAGTAACTATGATACAGGATTAAAAGACATAAAAGATATAA
TTGTAAAAGGAAATACTGTAGTAGTAGTAGATAAAAATAACAATATATATGTAACAGGAATG
AATCAGAATAACAAACTTGGGATAGGGGAATATAACAACGAACCAGTAAAAAAATTCACAAA
TATAACTGAACAATCAAACTCATTTATATTTATGGATGATATAAAAGAAATTACAACATCAA
GAAATACAATGTTTATAGTAAAAAATGATGGAACAGCCTATGCCACAGGAAATAATAGTTCT
GGACAATTAGGATTAGGTGACACAATAAATAGAAATAAGTTCACTCAGATAAACCTTGATAA
TATAAAGAAAATATCAACAAGTATAGATGGTAACACAACATTTGCAATTAGAAATGATGGAA
CACTATACTCCACAGGATTAAATACCAAAGGACAACTGGGATTAGGTGATATAGTAAATAGA
AATACATTTACCAAAGTAAACATCCAAAATGTAAGAGATGTTGTTTTAGGGACTACTCACTC
GCATGCAATCAAAGATGATAACACATTATATTCATGTGGAGaAAACACTCATGGGCAACTGG
GCTTAGGAAGCGAAAGCAACCATCCAGACGTATTGACATTTACTGTAAACAATATAACTAAT
GTAAGAGATGTGTACTGCTCAGATACAACAACATTTATTGTAAAGGACACAAACATTGCATA
TTGTTGTGGATACAATAATAATTCACAACTAGGTATGGGAAATACTACTGACCAGTATAGTT
TTATAAAGTGTATGGAAAATGTAAAAGAAGTTATACCAAATGAAATAAATACCTATATAATA
ACAATCTATAATACTGCATATAGTACAGGTTTAAATACTGATTATTGCTTAGGTCTAAATAG
TAATAGCAATCAAAGTTCATTTTCTGAAATTCCAATTTCAAATGTAGTAAAAGTAGCTCCAA
ACAGAAATAATGCAGTACTTTTACTTACAAGTGAAGGGGATGTATATACTGCAGGCAAATGT
AGTAATGGTTCAGGTACAGGAAGTGAGACTCCAGAGAAGATTAAAAAAATAGCATCAAAGGC
AAAGGATATTGGAATGAATTATAGATGTGGACATTATGTAAGTGATAATGGAGACCTATATG
GTACAGGTTTTAATAATAATGGACAATTAGGTGTTGGTGATGTAACAAAAAGAGATACATTT
ATAAAAACCAATACAAGAGTAAAGAAAATACTTCCTTTAGAATATGCAAATATAGCAATAAA
AGATACTAATGATATATATATTTGTGGATTAAATAACTATGGACAATTAGGTGTTGGAAATA
GATACGATAGTAGAAATAATGATAATAGAATATTTAATTATAAGCATATGAATTTTGTAATG
GGTGATTTGACATCTATTAAAAACAGACATAACTTTATACTTCTAAACAATAAGATAGTGAT
ACCTACCACAAAAGACATAGATTATGGTTTAGTATTAGGAAATTTATACAAAGGAGACCTTT
ATACTGAGCTTCCATATGAAGATATAAAAGAAGTATCTATTTCTAAGACTCATATTATTATA
TTACTTAATGATGGAACAATGTATGGATGTGGTACAAACTACCATGGAGAATTATTGCAAGA
CTTGTCTATAAATCAAGTGGATGAATTTGTGCAGATTAATGTATCAGATGTAAAGCATGTTT
CATGTGGAGATAACTTTACTTATTTTATAAAATCTGATGATAGTCTTTGGTCTATTGGTAAA
AATTCCGAATATCAATTAGGTATAGGTCACAATAATCCAGTTACTGAATTACAAAGAATTAC
AACTATATCTAGCTGTAAAGAAGTACATTGTGGTAAAAACTATACATTAGTAGTAACTACAG
GTAATGAATTATTTGTACAAGGATATAATGATAAGGGAGCTTTAGGATTAGGAAGCGATAGT
GAAAATACTATAATTAAGTTCTTTACAAAAGCACTAACAGACATAAGAGAAATAAAATCTTA
TGGAAGTGACCATATATTAGTACTTAAAAATGATAATTCAGTATGGGTTACTGGAAAAAATA
GGGATGTATATAAAATTGAACAACCAGTAGAATTTTTAAAAGAATTTACTATAGTACCTATT
TCTGAAGATGTAAATACAGTAAAGGATGTACTTGCAACAGACAATACATTATATATTATATC
AGAAGTAGGAACGACAAATGCTGCTATAGAAATTACTGAAAAATCAATTTCATCAATTAAGA
TAAAAATACAAGACCCTAATAAAGATATAAGTAGAATAGAAATGCTTATAAATGGTGAAAGT
GTAAAATCTGTAAGTGATTTAACTACTGAAAAAATATCCTTTGAAGTACCACCAGATAAAAT
TAAAATAGGAGAGAATAAGATACTATTTAGAGCTTATTGTAAAGGTGATGATTTATATGCAT
CTTTATTTATTTTTAAAGAGAGTACTGGAAATTCTATAATTAAAGATTCTTATGTTATGATA
GGTAATAGAATGTACAAGGTAGTTAATACAACATCTAATGAACAAGATATTACAATTACACT
AGATAGAGGACTTGAAGAAGATTTAAATCTTGGAGACCCTATATATCAATTAATAAATAAAA
CTAAAGTTCAAGTAAAAATAAATAAATCTGACTTATTCAAAGACATGAAACTAGTTGAAATC
AAAAAATCAGACTCAAGTTATCAAGAAATCTATGAATTAGAAGAAGCCAACATAAAAAGTGC
TCAGCCTAAAATCATAGTAGAAAAAGGAGATAAATGGACAGCTATAAAACGTCCATCTATGA
TTTTTAGATATGATGCTGAAAACAACGAGCCACAAGCTTAAAATGGAGGTGTAAAAATTGTT

TAAATTCGATAAAAATAAAATAGAACAAATCAAACAAGGTAGAAAAGTAGAAATGCAGTATA
AAGACATTTCAGACATAAGTATAGGTCAAGCAAAGCAAGATGATGATATAACAAATAATTTT
ATAGCAAATGCAGAAATATATGAGATGTTGTTAAGTCAAAGTTCTGTCAATGAAGCAAGTAA
TATAAGCACTTTTAGTGTAAGAAAATCTGGAGGTGAGAGTGGAATGGTAGAAGTATATGTAG
CTTTAATTTTAAGAGGCAGAAAAACAATAGAAGAAGTACCAGCAGTAATTAGAGAGCAAGTT
AGAATTAGATGTAAAGAATTAGAAATACCAGTTGAATAGTAAATTTAGAATAACTATGTATT
AGTTATtTTTTTTTATGTAAAGTACAAGGTCTTAACTTTAATAAGTAAGCCTTGTACTTATTT
TTTGTTATATTAGAAATTGTATATATATTTATTATTTATTCAATCTATAAATTAAACCTACA
ATTTAAAGTACAGAAGATTAAATTGATAATCCTGAAAATATAATATTGCATGATGTAAGAAT
ATAACAAAAATTAAAGCTATAAGTATAAAAAATTTAGACAATAGGAGGCTATAATGGATAAA
TTAATAACCGAATTGAGTAGTCTGGGGGCAATAGGTATACTATGTGCTCTATTATTTAAAAA
TACTATGCAGGAGAAAAAAGAAGATAGAGACATGTATAAAAAAACTGTAGAAAATTTTATAG
AATTATCTACACAACAACAAGAAATAAACAAAAATATACTTGTTCAAATGGGAATAATGAAA
ACAGATGTAGAGGAAATTAAGGAAGATGTTACTGATATAAAAGGTATGTTACAAAACGGTGT
ATAACATGAAAGTAGCAGTAGCACCAGATTATATATTATTAGGAAAAGATAAAGTAGTATTG
TAGATAGTGCCCTATTTTATTGAGAAGGATTTTATATTTTAAAATATTAATTAAAAAAAGTA
ATAAAAATAATATATAAAAATAACATATAAAAATTCAAAAAGGAGTTAAGCTTAAATTTGAT
TAGAAAAAATCAATTTTAAGACAACTCCTTTTTTTtATTAAATTATTGTCTATTAACCAAAA
TAGCTATTTTAGCATCTGGATTATAACTTATCTGAACCATTTGATTTTTCTTAACATGTTCA
AGGTCTTCACCACCATAAGCTATTTGTAACTTAACTGGTAACTTACCTTGTTTTATAATAGC
AACGTACTCTTTTTTACCTTTTTCTCTAAACTAATCAAATTGCCAACATAAGGTTTAAAGTT
CTGATACTTTTTACTAGAATTTCTTATGTAGAAGAAAGCACCAACAGCAATAACTAAATTTA
TGCCAAGTGTAACCCAAGAATTGATTTTAAGCATAGCTCCAGCGATTATTATCACGAACATT
AAAACGATAGGTAATATAGCTTTCTTAAGAAGCAATTTACCCATTATTTCATTAGCCTTTTT
CTCAGGGCCACTCATAGTTTTTGATCTAGCAAATGATTGCGCGAATTTGTCTCTTAAGCCCA
TTTTATCCTCCTAATTTTAATAAATATTTAGTTATAATAACGAGATATTACTTGAAACTAAA
AATTTACTACATTTATATTATGTTTGACTTTTGTATAAATAATTACATTCAAGTAAAGCAAA
ATATACTAATTATTTTATCATAAAATTATAAAAAAGAAAATAAATGAAATAAAAATATTAGA
ACAAAGAAATGATGTAAAATCGTATCAAAAGCAACATAAAAATTATTTATCTATTTTCTCAT
CTTTATTTTTGTTATACTCAATTTTTCCTAAATCCTTCTCTTTTTCATATTCATGAAGTTTT
AATTCAATCATACCTTCTATTTTGGCTTTATCATAATCATTTAACTTTCTAAAGTTGTTTAA
AAGCTTTATTTCATTAGAGTTTATGCTATTAAGTGGATAGTTTGAGGAGGAATCGCAAATTA
AATCTGATTTATGTGATAGATTATCTCCATTTAATAGCCAGTCTACTGAAACATTAAATATC
TCAGCTATAGATTTTAATATTTCATAATTTGGTTTTCTAATGTTTCTCTCAAATTTACTTAA
GTTGTCACAGCCTAACATTTCTTCTAGTTCATATTGTTTAAGGTTTTTTGCTTTTCTCAAAT
AAACAATTCTTTCTCCTAAAGTATCCATAAACACTCTCCATTCAATTAATGTCAAAAAGACT
TTTTAAGATGTAAATAGTTTCAAATTAAAGGTCAAAATGACATAAAAACCATTGACTTAAGG
TCAAAATGACTTTATAATTAACTTAATGATACGAATTTACATCCTAATTTTAGCACAAAGTA
ATCAAAAAATCTTATTTAGTATTAAATAAATTTATATACTTAATATGTGTACATATTAAAAA
TATATACTAAATAGAGGGGGTGCGTAAGCTAAAGTAATATAAAAGTAAATATAAATCACTTA
GAAAGGAAGTTGATAAATGGATGCTCGAAAAAAATGGATACCTTTTTTGGGAGTGCAAGTCA
AGCAAAGACTTATTGAATTAAATATGACTCAAAGGGAATTAGCGAAGAAAATAGGTGTTAAT
GAAAACTATTTGTCAGCTATTTTAAATGGAAGAAGAACAGGTAAAAAATATAAATCATCAAT
TTATCAATTACTTAATATAGAATATTCAGAAGATGATTAATAAATAGTATATAAAGTAGGTG
AATATTCTTGTGTGCAAATTGGATTCAGATGGGGTTATAGAGTGTTGTAGAGCAATTGATGA
TTTTATTACAGCACTTAGTAATATAAAAAGCTTAAATATGGAAAGATTAAATACTTTAACTA
AATATTCTAGTACATGTTCAATCCTTCTTAAAGAGGGGAATTATGAAGGATGTACAATTGTG
TATAGAAAGATGTTGGAAGAATTAAAAACATGAGTAATGCATTTCTTAGGAATATAAATTAT
ACATAGAAATGTATTATATTTTTCAAAGTACTTAAACTAAAATATGGATAAGATAATCTAAA
TATTATAAATGTGCTTGAAATTAGACTATACTTGTTTTTAAATAATCCAATATCCATATTTT
AGTAATATACTACAAAAAAAGAAGGTTAATAGATGATGTAAAATCGTATCAAATTATGTATG

TTTAAACCATTTTATCTTCATTATTATTAGAGGAATGCTTTTTTAAGTCTTTATATTCAGAT
ATCTTAAGTTCAAGTATTCCTTCTATTTTTATTTTATCACGTTCGTTTAGTTGTCTGTATAG
ATTTAATATCATCATTTCATCATTAGTAACATGTAAGTAATCTTCTTTATCTTCTTTTACAC
TACTATTGACATTTACCTTCTCTTTACCATAGAGAAGCCAGTCAGTCGTAACATTAAAATAA
TCAGCTATTGACATTAGTATATCACAATTAGGTTTTCTATCTCCTGTTTCATACTTGCCTAA
GTTTTCAAATTTTAAAATATCCATAAGTTTGCGCTGAGTAAGTTTTTTGGAGTTTCTCAAAT
AAGCAATTCTTTTTCCTAAAGTATCCACAAAATACACTCCTTTCTTTTTATGAGTAATGTCT
AAATGACATTTGAAATTAAAAATATATAAATTTATAATATAAAACTACTAAATTAAAGTCTA
AATGACATTTTGCTTAAATTAATATGCTCATAATATGATTTTAACATATTATAGTTGAAAAT
ATATGGTTTATTTTGATTTGTATATATAACAATAGATTTAATTGTTATAAAAATGTAAAGGG
GTGTATGAATAGATTGTATAAATTTATTTCGATAAACTAAGATTGCTTTTTGATTGTCTGTA
AAAGAGAAAAAGATTAAGATAAAAATAGTATTATATTGTAATTTATATTAATCAATTACAAA
GATTTTATGAATTTATTCTTTAGGGTAAAATATTTAAGAATAAGATAAATTTACAATATAAT
ACTATAACACTCTTTTATCTAGTTTTATTTTCTTTATAGAACAATAATATTATAAATGCTAG
TAGATTTACACAGAATACTGTTATATACATCTGTTTGAATCCTGAGTTTAGAGTAGATTGTA
GTGtGGATCCGG

SEQ ID NO:2 > orf 1360A (100% 630)
MFKNNLKYYRKCKGMTQIQLARKAGITNDYISQIERGIKNPGLLMAKKISSILEQNIEEVFF
IQL

SEQ ID NO:3 > orf 1361 (100% 630)
MENKKDILFKETDERLHNYKYLDIKIKNINLDIKRCENEYSGCGAMVYTEKTSNTYNISSSV
ENEVLKREERLRKLKMEKEDIEIEKEKIENALTCLNDIEMEFFNLFYNSKTKNNMTYISMKL
HLDRTSCYNLKKKMIFKLSEIL

SEQ ID NO:4 > orf 1362
NFTESIFIDDESVQGSEGSCFFVSILSVICTPIMLNTNNKDIVISIKYLPKPQSKSIRMYEI
SDELNKLFNRNIKVTDRKLNITKLEQSIKKEESIYVLNFTITLNYLDSVYEEDVVYENMEEI
NLNLGE

SEQ ID NO:5 > orf 1363 (99+% 630)
MAIGLPSINISFKELATTVKERSARGIIAMVLKDAKALGLNEIHEKEDIPVDLSAENKEYIN
LALMGNVNTPNKLLVYVIEGEADIQTALDFLETKEFNYLCMPKAVEADKTAIKNWIIKLRDI
DKVKVKAVLGKVVGNHEGIINFTTEDVLVGEKKYSVDEFTSRVAGLIAGTPLSQSVTYTKLS
DVVDIPKMTKVDAESRVNKGELILIKEAGAIRIARGVNSLTELTEEKGEMFQKIKIVDTLDI
IHSDIRKVIIDDYIGKVTNSYDNKCLLIVAIKSYLEELEKSALIESDSTVEIDFEAQKSYLK
SKGVDLSYMTLQEIKEANTGSKVFLKAKIKVLDAMEDIDLSIEI

SEQ ID NO:6 > orf 1364 (100% 630)
MANMEARNVMSGTWGELWLDGNKVAEVKKFQAKMEFTKEDIIIAGQMGTDTKYMGYKGKGSI
TLYHVSSRMHKLIGEKIKRGSEPRFVAISKLNDPDSYGAERIAVKNIAFDDLTLADWEVGVK
GEIEAPFTFTEYDFLDII

SEQ ID NO:7 > orf 1365 (98+% 630)
MNENGLSKNINIVDLLLNADTENLERPSTIVELKRLSTIFGQEFKVMCRALTISKDEEIQNT
CLKIDENMKTDIDLPEMQMLTIIEGVCDLDGKLLFKNKELMDKFKAPTPKELARKLLLPGEI
TNLYRILQDVMGYGKNAVIEEVKN

SEQ ID NO:8 > orf 1365A (100% 630)
MYYYWKKKGIRPSLFYAMDKGELKLIEAFFALEIEEEVEKMKHGYGVCPLTGGGM

SEQ ID NO:9 > orf 1366 (98% 630) Tape Measure
MGNVREEGINMYLTDNYTPKMNQIISVTDNFRRATVAVSLSTNVMASSIKNSIGSASSRVNS
LNSSLRKVQTTASSVSSTMAKLSSSINAVSGVIGSLNGSIMRLAITIAMIIDYFNKLIQKKN
EFNSNIMIILIFKAKSDEVEKTKNKLLGNLKKIGGKIWNIVIKAKDMTKRVISSILGKLKRV
EKRPYQGSINLKDMVSSAMARILPKLMLFKNTFWSGVIAIKDMASSIISKVFPKLRLFAGKV
WSGAIAVKDMASGILGSIKGKISDLTNGATIGVAVKKGVDLLGQEQNQKVVLESVMKRNTGK
TSQKDVDKYYDSLVNMANDTPFDPEDVVAMGTKAKMISNITGGKKEKDITQAMVDVRALNMN
TSSEQDVSAAFLSAAKGNMESLNTLVGENYKTFDEALEGISVKQMGLAKEMSNTIPGIISGA
QTSINNGLKSIVKPFDDILGQGLKKIKTFIESGLGNLAGLSEKMAGKIGNVMNGKIIIGNKY
DQMQSRSVKNGKEFSDSTQYRISNEAEKRKMMVENKQERFENHAATMIGNAPKAIVNAGSTL
LQNIDFTALIDSLLPVVNLVNNLLDSINNKSPIAQGLISIFGTIVTTAFQLIGPVVEAVSPI
ITRIFTFLGEYAPQINNFIETLGVIWKTVWETLGPLLETGWKIIEPILGAFFNILDKVCKIV
KDICKWWQTMINKIKNGSITGTVLNLVEKSKKNYKDNP
YAGTKAGDSGKAYSSKKGNNAFGLNYVPYNDYQTRLHEGEMVLTKQEANQYRSRKNGGNINI
AKLADTIVIREEADIEKITSKLVASIQLAQLGGVL

SEQ ID NO:10 > orf 1367 (100% 630) Baseplate
MEMWLRQAEDRFRFPVFPSSFSINGKAAVNSSSILKIGEVATFGGVALKSISISSFFPNKDY
TFCDYTGFPSPYDCVNKIEKWMKEGFILRFTITETNINMEVIIEGFSYEERDGTRDVYFTLD
LKEYKRIKIPKVTPKQ

SEQ ID NO:11 > orf 1368 (100% 630) Baseplate
MIINRSKDSSSNEISFVSKDMGFLLTQSEVSYNFKDKLVEDIAKQVFAENRLSVGTIAKTNV
KYTKMFIGVNGYDTIMSAYTEASKKTKKKYMIEANLDKFNVIEKGTVTLSVMFEEGFNIINT
TFSESMENVKNKVIVVDQYGSKISEKIDNEIFKEVNVIMQKVIQQQENQDVDIDSEFNGIEK
SCSLKGYGDVSCITGRGVKVKDSYTKLVGLFYIDTDKHTWQNGEYQIELELNFQNLMDEKSA
GQDEPKEESNLGGEDYAGGKEFTAEFTAYCPRKEEGGDTDCRKKKLDPSKKTCAAPMVGKYE
QTYYTKEFLNKHPLLNYGDEIQVITGVSGRDGVYKVNDVGPAITIEKNGTYHIDILFGNVEE
ASKFGRRKGKIIIGGYSGNVSDKAKIVISEAKKHLGKPYKWGGNGPSSFDCSGLMVYCFKKV
NVSLPRTSNQQSKKGKKVEQKNLQAGDLVFFHNPVSHVGLYIGNGEFLHAPQKGDVVKISKL
SSRRDFNTARRVL

SEQ ID NO:12 > orf 1369 (100% 630)
MANPINEFIGIIREEGKYHNQPSFFIGKIKSKLPDLKIETNNIILEKEDILIDSWMIDRQLE
TFDTETNQEHQHEVKNPFIDNFESGDMVIMFRIGEKFAVVSKLVSL

SEQ ID NO:13 > orf 1370 (99+% 630)
MSTIFPFIGVPEDYILPKTEELPIFREVAWDFEKDEPILEKGDFKIIEKKEALKVWIYKCIK
TNRYEHEIYSLEYGTELSELIGQKYTKGLTESEASRFIKEALLINPYILEVNVKSANFNRDV
LSANVKVSTIYGEVEINV

SEQ ID NO:14 > orf 1371 (100% 630) Tail
MYSDQTYEVIKNRTLENINLDIYKGEGSFLNNMVSGNNLELSKIYLELSKIHKMAFIQDTYN
QFLDKRVNEFGVYRKLGTESNGEVEFIGEKGTVINNGTIISYRDLLFVVIKDVTIGSEEGDN
SPVQALEVGKKYNLPTNCEFKLVDNISGVTKITNTRSFEGGTDIETDEELKERFYKIQRNQA
TSGNKAHYEEWALEVDGVYNVKVYPRWDGPGTVKVLIFGENNQAVDTETIERCQQHIDEEKP
IGPTITVVTPLPIEISISAVMKLEDGYTLDNVKESFLESINTYFRDIRGEIIYTKVMGILIN
TTGVHDLSNLLINGSTDNITINEDKIPSVTTVNFSEVENQ

SEQ ID NO:15 > orf 1372 (100% 630)
MKLIDKLPSFDRNYIVEEIQGAYDTELNILKEDIDDTFNQLFVDTATWGLDMWEDILCIEKK
ELDFDTRRSNIKAKMRSRGTSTIEVIKSICEAYTKSETDIKVYSDEFTFVLSFIANNCDYKT
LLDCSEMIERVKPAHLLHYLEPIILDKSMVYCGGGMVCSEEVKVHPYFEPIIKCSAVVNCGA
GMLSREEIKVYPLSIKCIENNCKINIAIANDTGVENVVVYPKSEVV

SEQ ID NO:16 > orf 1373 (99% 630) Tail fiber
MEEKFYIILTKIGREKIANATALGELVGLTKFQVGDSNGEYYEPTEEQTALKNVVWEGNINS
LRIDEKNPNWIVIETILPGTVGGFMIREAAVLDNENNIIAIGKYPETYKPRAEDGSIKDLVV
KMILQLSNTSNVTLEVDPTLVFVTQKDIQDLLDDKFDKNIKEIKVKIGDTDILTTDSKDLSGA
INEVVKKIENISFDDVISGQIQTDISVLKNSYNKLSEKVLDILIYLELESEVTVDEAGYWYD
TLANGNNIVAIEGLKLDLNRKCITGEIGNVIFRDVVLPFSANRVRYIHDMDNNFVETKSSNT
YLKEQKDITLSKYSYEIR

SEQ ID NO:17 > orf 1374 of #16 receptor binding-variable
MKQNKLLQRGAYFNDKNILIDDFDKRYNDYDFVEFFTGISNSTFGLKSDGNLYACGDNTGFQ
LGLGKDSSERRMFSKVKIDNVKYVSCGSKHSVAVTKDGFAYGAGTSNVGQLGVIESTVYYEF
TKLPIDDVKTVACGYDFTFVLKNDGTLYSAGLNSSGQLGLGDTNNRATFTKVNIDSVKDVVT
YNQSVFIIKMDGTAHACGLNSNGQLGINSTLNKSVFNKIEGMDNVKQIACGSSHTILIKNDG
TMYTTGYNGVGQLGTGNNNNSIVFTLSSINNVKYASCGNNHTMILKYDNTLFSTGQNNYGQL
ANANKDVASRNTFAKVNVENIKDIKCGSQFNFLINGSKEIFVSGCNLAGQLGSFFHTTFLYE
FSNVQSSNLDNYSGLLVNDDYLYVTKDNSEFLNVKLSDNFQDYKKIELTDSNMFIVMNDGTL
YACGLNNYGQLGLGDTVNRSVMTKVDIDNVLDIKGNGNSTFVLKNNGTLYSCGLNSNGQLGL
RDEVNRNIFTKIEIENVKDFCVGSNYVIALNHSKEVYGWGNNPYNNIEKTSNYPYKQGISNI
EKIAAYDYSVYMINSEGKLYVSGYNYNYQLGKGNNSNQSKALVSQCRTNSTSSTSNGLRTLP
KITNVFPFYDGCAIIDEGGYVYLTGYHGYLRTLNSSPSISDYSRYGTFIEATNSNHNTYFIQ
ETDFSGIEKVIGMSNNILFFKKGSSYITGYPKTFGSTITGHRSYTSINSESSNLGSNFIIYH
SNSKLYGKGIANSGQFGNSTNIDGTSNYDTGLKDIKDIIVKGNTVVVVDKNNNIYVTGMNQN
NKLGIGEYNNEPVKKFTNITEQSNSFIFMDDIKEITTSRNTMFIVKNDGTAYATGNNSSGQL
GLGDTINRNKFTQINLDNIKKISTSIDGNTTFAIRNDGTLYSTGLNTKGQLGLGDIVNRNTF
TKVNIQNVRDVVLGTTHSHAIKDDNTLYSCGENTHGQLGLGSESNHPDVLTFTVNNITNVRD
VYCSDTTTFIVKDTNIAYCCGYNNNSQLGMGNTTDQYSFIKCMENVKEVIPNEINTYIITIY
NTAYSTGLNTDYCLGLNSNSNQSSFSEIPISNVVKVAPNRNNAVLLLTSEGDVYTAGKCSNG
SGTGSETPEKIKKIASKAKDIGMNYRCGHYVSDNGDLYGTGFNNNGQLGVGDVTKRDTFIKT
NTRVKKILPLEYANIAIKDTNDIYICGLNNYGQLGVGNRYDSRNNDNRIFNYKHMNFVMGDL
TSIKNRHNFILLNNKIVIPTTKDIDYGLVLGNLYKGDLYTELPYEDIKEVSISKTHIIILLN
DGTMYGCGTNYHGELLQDLSINQVDEFVQINVSDVKHVSCGDNFTYFIKSDDSLWSIGKNSE
YQLGIGHNNPVTELQRITTISSCKEVHCGKNYTLVVTTGNELFVQGYNDKGALGLGSDSENT
IIKFFTKALTDIREIKSYGSDHILVLKNDNSVWVTGKNRDVYKIEQPVEFLKEFTIVPISED
VNTVKDVLATDNTLYIISEVGTTNAAIEITEKSISSIKIKIQDPNKDISRIEMLINGESVKS
VSDLTTEKISFEVPPDKIKIGENKILFRAYCKGDDLYASLFIFKESTGNSIIKDSYVMIGNR
MYKVVNTTSNEQDITITLDRGLEEDLNLGDPIYQLINKTKVQVKINKSDLFKDMKLVEIKKS
DSSYQEIYELEEANIKSAQPKIIVEKGDKWTAIKRPSMIFRYDAENNEPQA

SEQ ID NO:18 >orf1375
MQYKDISDISIGQAKQDDDITNNFIANAEIYEMLLSQSSVNEASNISTFSVRKSGGESGMVE
VYVALILRGRKTIEEVPAVIREQVRIRCKELEIPVE

SEQ ID NO:19 >orf 1376 (100% 630)
MDKLITELSSLGAIGILCALLFKNTMQEKKEDRDMYKKTVENFIELSTQQQEINKNILVQMG
IMKTDVEEIKEDVTDIKGMLQNGV

SEQ ID NO:20 >1377 (100% 630)
MGLRDKFAQSFARSKTMSGPEKKANEIMGKLLLKKAILPIVLMFVIIIAGAMLKINSWVTLG
INLVIAVGAFFYIRNSSKKYQNFKPYVGNLISLEKKGKKEYVAIIKQGKLPVKLQIAYGGED
LEHVKKNQMVQISYNPDAKIAILVNRQ

SEQ ID NO:21 >AV1378 (100% 630)
MDTLGERIVYLRKAKNLKQYELEEMLGCDNLSKFERNIRKPNYEILKSIAEIFNVSVDWLLN
GDNLSHKSDLICDSSSNYPLNSINSNEIKLLNNFRKLNDYDKAKIEGMIELKLHEYEKEKDL
GKIEYNKNKDEKIDK

SEQ ID NO:22 >1378A (100% 630) regulatory
MDARKKWIPFLGVQVKQRLIELNMTQRELAKKIGVNENYLSAILNGRRTGKKYKSSIYQLLN
IEYSEDD

SEQ ID NO:23 >AV1379
VDTLGKRIAYLRNSKKLTQRKLMDILKFENLGKYETGDRKPNCDILMSIADYFNVTTDWLLY
GKEKVNVNSSVKEDKEDYLHVTNDEMMILNLYRQLNERDKIKIEGILELKISEYKDLKKHSS
NNNEDKMV

SEQ ID NO:24 >AV1419
tttcttGAAGACcAtcgaAgcaccaccaccaccaccactg

SEQ ID NO:25 >AV1420
ttttttGAAGACAAtcgaAgggcttcgccctgtcgctcgac

SEQ ID NO:26 >AV1416
ttccttGAAGACCTaattTggggcaatcccgcaaggag

SEQ ID NO:27 >AV1245
ccccttGAAGACCCaattTcgtatggcaatgaaagacgg

SEQ ID NO:28 >AV1372
aattGCGGCCGCagctcGCTAGCggtacCTCGAGgatatcttcgaaGAAGACACATCCg

SEQ ID NO:29 >AV1373
aattccgggatGCATGCctctaGGATCCGGCGCGCC

SEQ ID NO:30 >AV1374
agctGGCGCGCCGGATCCtagagGCATGCatcccggaattcGGATGTGTCTTCtt

SEQ ID NO:31 >AV1375
cgaagatatcCTCGAGgtaccGCTAGCgagctGCGGCCGC

SEQ ID NO:32 5′ fragment
GGCCGCAATACCCACTACACCTTCgTCATCTTTAAATTTAAGAGTTTTTACTATTGAATAAT
AAAGGTATATTCCAGTAAAAATAATCTTTAAATACAAGAAAAATAAACTCTTTGGGTATATT
AAAAAGCTAAAAAGTGTAAATATAAAAGCAAGTAGAGTACTTATCCTGTAAAAGAAATCTAT
TTGTGTAATGTCTTTATATTTTATCATAAACACCGAATATAAAATGATGAAAATAATTGCGA
CGATTGCATATATGGTAAATAACATATTTTCAAGAGTACCATTTGAAATTACTATCCACTTA
TACCACATAATTGGCCAAAATAATAGTGCTAAGAACTTAAAATAATTATCAAACAACTTTTC

TTTATACATTCATCAAACAACCTTTCTTAACAAAAGCATATATTTGTTTTTAGAATTTTAAA
TAATATGATATCATTATTATATATTAATATTGAATTTATAGAAACCAAAATTTGTTAAAATA
AATATATAGATTTTACTGTTAAGCCAGTTAAAATTACTACTATTTTTATTATGAAATTGGAT
CAAATATGTAGAAATACGGCAAATTAGTTAATATTAAATATTTATTATTTCCAAGTTGTAAA
GACTGTTTTTTTAATGATAAAAATTCTAATCTTTTTTGAAAGAAAGTAATATCCACATTAAG
TATGTCTGCCATTTCATAAACGCAAGTGATGCCAGAATTAATTATGTTTATTATATCTTCTT
CAGTAATTAAGAACTCACAAGCCCATTTTAAGGCTTTATTTTCGCACTTATCTATAATAATT
TTTGTATAATAATCGTTATAAGAGGATACATAGTATCCAAGGCTAGTGAAATGATGTCCAAG
TTCTTCAGCTAAGATGGATGTCAATTTTTTTGAGTTTTGTTTTAAATTACTGAGTAATGATA
TAATTTTAATACCATGTTTGTTTATATATAGCCCTTCTAAATCACCTGCAATATAAGTGGTA
TAATGAATTATTATCTCTTCTTGAGAAGCTAATTCAAAAAGCTTATCCAAATTATTCATAAA
AATCCCCCTAAAATAGAATGTATGTTTGCCTTTAAATTATATTAAAAGAGCAGAAAAATAGA
CTGCTCATCATATGGTTTATTTTTTTTTATATTTATTTAGTAAAAATTCTATATAATCATTA
AGTTGTTCTTGTGCTTCTTCAGGTAACTCTTCATGTGGATTTTTTCTATGTGCAGCTACTGT
ATCAATATTTTCCTTAACTAAGGTTCTTCCAAGAAGGTAATCAACTGATACATTAAATACAT
CAGCCAATTTGTTTAAAATGTGTTCATCAGGAAATCTGTTTTCTGTTTCATAGTACCCTAAG
ACTCTTTGGGAAACGCCTACTTTTTCTCCAAGTTCTCTTTGAGTCAATCCAAATTCCTTTCT
AAGTTCTCTTAATCTTTTGGCAAACATTATAACACCACCTTATGTATAGATTATAACAAATT
GTTCTAAAAAATAAAACTAATAAAATATAAAAGAATATTTTTTCTAAAATCTATTGATAAAG
AACAAATGATTCTATATAATCTAAGTGTGGAAGAACAAAATATTCTTAATGGTAATGGAGGT
ATAAAACAATGTTTAAAAATAACTTGAAATATTATAGAAAATGCAAAGGTATGACACAAATT
CAACTTGCCAGAAAGGCTGGAATTACAAATGACTATATATCTCAAATAGAAAGAGGTATAAA
AAATCCTGGTCTTCTTATGGCTAAGAAGATTTCTAGTATTTTAGAACAAAATATAGAAGAAG
TTTTTTTTATACAGTTATAGAACAATATGTTCTTGAAAGTTGTGAGATTAGTAAAAAACTGT
GCACTAAAGAGATTATTGTAAATTTGAAGCTAATAATAAGTATATAAAAAAGGAGAAGTACT
ATGGAAAACAAAAAAGATATATTATTTAAAGAAACAGATGAAAGATTACATAATTATAAGTA
TTTGGATATAAAGATAAAGAATATTAATTTGGACATAAAAAGATGTGAGAATGAATACTCTG
GATGTGGAGCAATGGTATATACAGAAAAGACTAGTAACACATATAACATAAGTTCTTCTGTG
GAAAATGAGGTGTTAAAAAGAGAGGAAAGATTAAGAAAATTAAAAATGGAAAAAGAAGATAT
AGAAATAGAAAAAGAGAAGATAGAAAATGCTCTAACATGTCTAAATGATATAGAAATGGAAT
TTTTTAATCTTTTTTATAATAGTAAGACAAAAAACAATATGACATATATTTCTATGAAACTA
CACTTAGATAGAACATCTTGCTACAATTTAAAGAAAAAAATGATATTTAAATTGAGTGAGAT
ATTATAAAAAATATGACAACTTTACAACACTTTATATACACTATTGCAACACTAGGCAATAA
AATATGTGAGATAATGTTATTGTGAAAGAAATCCATATTGAAGGAGGTGATAAATTGAAAAG
AATAATATTACCTAAAAATATAGAAGATATTTGACAGGAATAAATGAGATGTATATTTAAAA
ATGACTTATATCATTTATAGTAAGATTATCAGATTAAGCAAGAATATTTAGTGATAGTGTGG
TGATTATTTGCTTAAATACAAGGAAATATTAGAAACAATTATTGAGATTCTCAAAAAAAACT
TTACTGAAAGTATTTTTATTGATGATGAAAGTGTGCAAGGCTCTGAAGGGTCTTGTTTTTTT
GTAAGTATACTATCAGTTATTTGTACACCTATAATGTTAAATACGAATAATAAAGATATTGT
TATCTCTATAAAATACTTACCAAAACCACAGTCAAAGAGTATTAGAATGTATGAAATTTCAG
ATGAATTAAATAAGTTATTCAACAGAAATATAAAGGTAACAGACAGAAAATTAAATATAACA
AAGCTAGAACAAAGTATTAAAAAAGAAGAGTCAATTTATGTATTGAACTTTACAATTACACT
AAATTATCTGGATAGTGTATATGAAGAAGATGTAGTATATGAAAATATGGAAGAAATCAATT
TAAATTTAGGAGAGTGATAGTATGGCTATAGGATTACCAAGTATCAACATATCATTTAAGGA
GCTAGCTACAACTGTTAAAGAACGTTCAGCTAGAGGAATAATTGCAATGGTGCTTAAAGATG
CTAAGGCACTAGGTCTTAATGAAATACATGAAAAAGAGGATATACCAGTTGATTTATCTGCT
GAAAATAAAGAGTATATAAATTTAGCTTTGATGGGAAATGTTAACACTCCAAATAAATTATT
AGTTTATGTAATAGAAGGAGAAGCAGATATTCAAACTGCATTAGATTTTTTAGAGACTAAGG
AATTTAATTATCTATGTATGCCAAAAGCAGTAGAAGCTGATAAGACTGCTATAAAAAATTGG
ATAATTAAACTTAGAGATATAGATAAGGTTAAGGTTAAAGCTGTATTAGGAAAAGTTGTAGG
AAATCATGAAGGGATAATTAATTTTACTACAGAAGATGTGTTAGTTGGAGAAAAGAAATACA

GTGTTGATGAGTTTACAAGTAGGGTGGCTGGACTTATAGCAGGAACACCTTTAAGTCAATCA
GTAACTTATACTAAGCTTAGTGATGTAGTTGATATACCTAAGATGACGAAAGTTGATGCAGA
ATCAAGGGTTAATAAAGGAGAGCTTATACTTATTAAGGAAGCAGGGGCTATAAGAATTGCAA
GAGGAGTAAATTCTTTAACTGAGTTAACAGAAGAAAAAGGAGAAATGTTCCAGAAAATAAAA
ATAGTTGACACTTTAGATATTATACATAGTGACATAAGAAAGGTGATAATAGATGACTATAT
AGGAAAGGTTACTAACAGTTATGACAACAAATGTTTATTGATAGTAGCTATAAAAAGTTATT
TAGAAGAATTAGAAAAGTCAGCACTTATAGAATCTGATTCTACTGTTGAAATAGATTTTGAA
GCACAAAAATCGTATTTAAAATCAAAAGGAGTAGATTTATCTTATATGACATTACAAGAAAT
AAAAGAAGCTAACACAGGTTCTAAAGTATTTTTAAAAGCAAAAATAAAAGTACTTGATGCTA
TGGAAGATATAGATTTATCAATAGAAATATAGGAGGATTATTAATATGGCAAATATGGAAGC
TAGAAATGTAATGAGTGGTACTTGGGGAGAACTTTGGCTTGATGGAAACAAAGTAGCAGAAG
TAAAGAAGTTTCAAGCAAAGATGGAATTTACAAAAGAGGATATTATAATAGCAGGTCAAATG
GGTACTGATACAAAGTATATGGGATATAAAGGAAAAGGTTCAATAACTCTATACCATGTTAG
TTCAAGAATGCACAAGTTAATTGGAGAAAAGATAAAGAGAGGTTCTGAACCTAGATTTGTTG
CTATATCTAAATTAAATGACCCAGATTCTTATGGAGCAGAAAGAATAGCAGTAAAAAATATA
GCATTTGATGATTTAACTTTAGCTGATTGGGAGGTTGGAGTAAAAGGAGAGATAGAAGCACC
TTTCACATTTACTGAGTATGATTTTCTTGATATAATTTAGTTTTATATTTGGTTTTATACTG
ATATTTAGTAGATATATACTTAATAAATTTAGGTAGTTAATAAGTAAAAAAGTTAGTTGATT
GAATTTGATTGATAAAGGAGCAAATAATAATGAATGAAAATGGATTATCAAAAAATATAAAC
ATAGTAGATTTACTTTTAAATGCAGATACAGAAAACTTAGAAAGACCAAGTACTATAGTTGA
ACTTAAGAGATTATCAACTATATTTGGGCAGGAATTTAAAGTAATGTGTAGAGCTTTAACAA
TAAGTAAAGATGAAGAGATACAAAATACTTGTCTTAAAATTGATGAAAATATGAAAACGGAT
ATAGACTTACCGGAGATGCAGATGCTTACAATTATAGAAGGTGTTTGTGATTTGGATGGAAA
GCTTTTATTTAAAAATAAGGAACTAATGGATAAATTTAAGGCTCCAACACCAAAAGAATTGG
CAAGAAAACTATTATTACCAGGTGAAATTACCAACCTATATAGAATACTTCAAGATGTTATG
GGTTATGGTAAAAATGCAGTGATAGAAGAGGTAAAAAACTAATAGGGACGGATACCAAGACT
ACAATAATGTACTATTATTGGAAGAAAAAAGGTATAAGACCGTCCCTTTTTTATGCAATGGA
TAAAGGCGAATTAAAGCTTATTGAAGCTTTTTTCGCCTTAGAAATTGAGGAAGAAGTTGAAA
AAATGAAACATGGATATGGAGTGTGTCCTTTGACAGGAGGTGGTATGTAATGGGAAATGTGA
GAGAAGAAGGTATAAATATGTATCTTACAGATAATTACACACCAAAAATGAACCAAATTATA
TCAGTAACTGATAATTTTAGGAGAGCAACTGTGGCTGTTTCACTTTCCACTAATGTAATGGC
TAGTAGCATAAAAAATTCTATTGGAAGTGCAAGTAGTAGAGTAAACAGTTTAAATTCCTCGT
TAAGAAAAGTTCAAACTACTGCTAGTAGTGTAAGTTCAACTATGGCAAAATTAAGTTCTAGC
ATAAATGCTGTTTCAGGAGTTATTGGAAGTTTAAATGGAAGTATTATGAGACTAGCAATAAC
TATAGCTATGATTATTGATTATTTTAATAAGTTGATTCAAAAGAAAAATGAGTTTAATTCAA
ATATTATGATTATATTAATATTTAAAGCTAAAAGTGATGAAGTAGAAAAAACTAAAAATAAA
TTACTTGGAAATTTAAAAAAGATTGGTGGCAAGATTTGGAATATCGTAATAAAAGCAAAAGA
TATGACTAAGAGAGTGATAAGTAGTATCTTGGGAAAATTAAAACGAGTAGAGAAACGTCCTT
ATCAAGGAAGTATTAATCTTAAAGATATGGTAAGTAGTGCTATGGCTAGAATTTTGCCTAAG
TTAATGTTGTTTAAAAATACTTTTTGGAGTGGTGTAATAGCTATAAAAGATATGGCAAGTAG
CATTATAAGTAAAGTATTTCCCAAATTGAGATTGTTTGCAGGTAAGGTATGGAGTGGTGCAA
TAGCTGTAAAGGATATGGCAAGTGGAATACTTGGTTCGATAAAAGGGAAGATATCTGATTTG
ACAAATGGTGCTACTATAGGTGTCGCTGTGAAAAAGGGTGTTGACTTACTTGGTCAGGAACA
AAATCAGAAAGTTGTTCTAGAAAGTGTAATGAAAAGAAATACTGGAAAAACTAGCCAAAAAG
ATGTTGATAAGTATTATGACAGTTTAGTAAATATGGCAAATGATACGCCTTTTGACCCTGAA
GATGTTGTTGCAATGGGAACTAAAGCTAAAATGATTAGTAATATTACTGGTGGCAAAAAAGA
AAAAGATATAACTCAAGCTATGGTAGATGTTAGAGCTTTAAATATGAATACAAGTAGTGAAC
AAGATGTATCAGCAGCTTTCTTAAGTGCAGCAAAAGGAAATATGGAATCTCTTAATACTCTG
GTAGGAGAAAATTATAAAACTTTTGATGAAGCATTGGAAGGCATAAGTGTAAAGCAGATGGG
GTTAGCTAAAGAAATGAGTAATACAATACCAGGTATAATATCAGGAGCTCAAACAAGCATTA
ACAATGGTTTGAAGAGTATTGTTAAACCTTTTGATGATATTTTAGGTCAAGGACTAAAGAAA

ATAAAAACTTTTATAGAAAGTGGATTAGGGAATTTAGCTGGCTTATCTGAAAAAATGGCTGG
TAAAATAGGCAATGTAATGAATGGTAAGATAATTATTGGCAACAAATATGACCAGATGCAAT
CTAGAAGTGTAAAAAATGGAAAAGAGTTTTCTGATTCTACTCAATATCGAATTTCTAATGAG
GCTGAAAAGCGTAAAATGATGGTTGAAAATAAGCAAGAACGTTTTGAAAATCATGCAGCAAC
AATGATAGGGAATGCACCAAAAGCAATTGTTAACGCAGGAAGTACACTATTACAAAATATTG
ATTTTACAGCATTAATAGATTCACTACTTCCAGTAGTAAACTTAGTAAATAATTTACTAGAT
AGTATAAACAATAAATCACCAATTGCACAAGGATTAATAAGTATATTTGGTACAATAGTAAC
TACAGCATTCCAACTAATCGGACCTGTAGTTGAAGCTGTTAGTCCTATTATCACAAGAATTT
TTACTTTTTTAGGTGAATATGCACCTCAAATAAACAATTTTATAGAGACACTGGGTGTTATT
TGGAAAACTGTATGGGAGACCTTAGGACCTCTGTTGGAAACTGGATGGAAAATTATAGAGCC
AATATTGGGAGCTTTTTTTAACATATTAGATAAAGTATGTAAAATAGTTAAAGATATATGCA
AATGGTGGCAAACTATGATTAATAAGATAAAAAATGGAAGCATCACAGGAACAGTTTTAAAT
CTAGTGGAAAAGAGTAAAAAAAATTACAAAGATAATCCATATGCTGGAACAAAGGCTGGTGA
TTCTGGTAAAGCTTATTCAAGTAAGAAAGGTAATAATGCATTTGGATTGAACTATGTTCCTT
ATAATGACTATCAAACCAGACTCCATGAAGGTGAAATGGTTTTAACTAAACAAGAAGCAAAT
CAATATAGAAGCAGAAAAAATGGTGGAAATATAAACATAGCTAAGTTAGCTGATACAATAGT
GATTAGAGAAGAAGCTGATATAGAAAAGATAACATCAAAATTAGTTGCAAGTATCCAATTGG
CACAGTTAGGGGGTGTCTTATAATGGAAATGTGGCTTAGACAAGCAGAAGATAGATTTAGAT
TTCCAGTATTTCCATCTTCCTTTAGTATTAATGGAAAAGCTGCTGTAAACTCTTCTAGTATA
CTCAAAATAGGTGAAGTAGCAACTTTTGGTGGTGTAGCTCTTAAAAGCATTTCAATATCAAG
TTTTTTTCCAAATAAAGACTACACTTTCTGTGACTATACAGGTTTTCCATCACCATATGATT
GTGTAAATAAGATAGAAAAATGGATGAAGGAAGGTTTTATATTAAGATTTACAATTACGGAA
ACAAATATAAATATGGAAGTCATAATTGAAGGGTTTAGTTATGAAGAAAGAGATGGGAC

SEQ ID NO:33 Middle fragment
TCGAGATGTATATTTTACATTAGATTTAAAAGAGTATAAAAGAATAAAGATACCAAAAGTAA
CTCCAAAACAATAACTATTATAGATAATAAGTTGTAAGTAACTGCTGATAGAATTAAATGAA
AAGGCAGGTGATTTTTTATTATTAAGATTTGGGTACACATAAAAAACGGAAGTATATATGAC
ATAACTGACATAGTAGACAAGGTATCATGGTCAGGTGATTATAAATCTCCATCAAGGACACT
AGAGTTTTCAATAATACAATCATCATTTGATGTAAATTTCCAACAAATCGATATACCAATAG
CTAGTACAGTCTGTTTCTATGTAGATGAGAAAGAACTCTTTAGAGGAATGATAATTAATAGG
TCTAAAGATTCAAGCAGTAATGAAATTAGTTTTGTATCTAAAGATATGGGATTTTTACTTAC
ACAAAGTGAAGTGTCATACAATTTTAAAGATAAGTTAGTTGAAGACATAGCAAAGCAAGTAT
TTGCTGAAAATAGGCTTTCAGTTGGAACAATAGCAAAGACCAATGTCAAGTATACAAAGATG
TTTATAGGAGTAAATGGTTATGACACAATAATGAGTGCATATACAGAGGCAAGTAAAAAGAC
AAAGAAAAAGTATATGATAGAGGCTAATTTAGATAAGTTTAATGTTATTGAAAAAGGAACTG
TTACATTAAGTGTTATGTTTGAAGAGGGATTTAATATTATAAATACCACCTTTTCGGAGAGC
ATGGAAAATGTAAAAAATAAAGTAATAGTGGTAGACCAGTATGGAAGCAAGATTAGCGAAAA
AATAGATAATGAAATTTTTAAGGAAGTAAATGTAATAATGCAAAAAGTAATTCAGCAACAAG
AAAATCAAGATGTAGATATTGATAGCGAGTTTAATGGGATAGAAAAAAGCTGTTCTCTTAAa
GGTTATGGAGATGTAAGTTGTATAACTGGTAGAGGAGTAAAAGTTAAAGATTCTTATACAAA
GCTTGTAGGACTATTTTATATAGATACAGACAAACATACTTGGCAAAATGGAGAATATCAAA
TTGAGCTTGAACTTAATTTTCAAAATCTTATGGATGAAAAGTCAGCAGGACAGGATGAACCT
AAGGAAGAAAGTAATTTAGGGGGAGAAGATTATGCAGGAGGAAAAGAGTTTACAGCAGAATT
TACAGCTTACTGTCCTAGAAAAGAAGAAGGTGGAGATACAGATTGTAGAAAGAAAAAACTTG
ACCCATCTAAAAAACTTGCGCTGCTCCTATGGTTGGTAAATATGAGCAAACTTATTATACAA
AAGAGTTTTTAAATAAACATCCTTTATTGAACTATGGAGATGAAATACAGGTAATTACAGGA
GTTTCTGGTCGTGATGGAGTCTATAAAGTAAATGACGTAGGACCTGCAATAACTATAGAAAA
AAATGGAACATACCATATAGATATTTTATTTGGAAATGTTGAAGAAGCTAGTAAATTTGGAA
GAAGAAAAGGAAAAATTATTATTGGTGGTTATTCTGGTAATGTATCTGATAAAGCTAAAATA
GTAATATCAGAGGCAAAAAAACATCTAGGTAAACCTTATAAATGGGGTGGAAATGGACCAAG

TAGTTTTGACTGTTCTGGTTTAATGGTCTACTGTTTTAAAAAAGTTAATGTTAGTTTGCCAA
GAACGTCAAATCAACAATCTAAAAAAGGCAAGAAAGTAGAACAAAAAAATCTTCAAGCAGGA
GATTTAGTATTTTTTCATAATCCAGTCAGCCATGTTGGATTATATATAGGTAATGGAGAATT
TTTACATGCTCCACAAAAAGGTGATGTAGTTAAAATAAGTAAGTTAAGTAGTAGAAGAGATT
TTAATACAGCTAGGAGAGTATTATAAAAGGATGGTGATATAATGGCTAATCCAATAAATGAA
TTTATAGGAATAATAAGAGAAGAAGGAAAGTATCATAATCAACCTTCTTTTTTATTGGAAAA
TTAAAAGTAAATTACCAGATTTAAAAATAGAGACAAATAACATCATATTAGAAAAAGAAGAT
ATTTTGATAGATAGTTGGATGATTGATAGACAGCTAGAAACATTTGACACAGAAACAAATCA
AGAACACCAGCATGAAGTAAAAAATCCTTTTATAGATAACTTTGAATCTGGGGATATGGTAA
TAATGTTTAGAATAGGCGAAAAATTTGCTGTTGTAAGTAAGTTGGTGAGCTTATAATGAGTA
CAATATTTCCTTTTATAGGTGTCCCAGAGGATTATATCTTACCTAAAACAGAAGAATTGCCA
ATCTTTCGTGAAGTGGCATGGGATTTTGAAAAAGATGAACCTATTTTAGAAAAAGGTGACTT
TAAAATAATTGAAAAAAAAAGAAGCCTTAAAAGTTTGGATATACAAGTGTATAAAGACAAAT
AGATATGAACATGAGATATACTCTTTAGAATATGGGACAGAGCTTTCAGAACTAATAGGACA
AAAATATACAAAAGGTCTTACAGAAAGTGAAGCTAGTAGATTCATAAAAGAGGCCCTTCTAA
TAAATCCATATATATTAGAAGTAAACGTAAAAAGTGCTAACTTTAACAGAGACGTATTGAGT
GCAAATGTAAAAGTATCCACTATCTATGGGGAGGTGGAAATAAATGTATAGTGACCAGACAT
ATGAAGTAATAAAAAATAGAACTCTTGAAAATATTAATCTTGATATTTATAAAGGAGAAGGT
TCTTTTCTAAACAACATGGTATCTGGAAATAATCTAGAACTTTCGAAGATATATCTAGAACT
TTCAAAGATACATAAAATGGCTTTTATACAAGACACATATAACCAGTTTCTTGATAAAAGAG
TCAATGAATTTGGTGTATATAGAAAGTTAGGTACAGAGTCAAATGGAGAAGTTGAATTTATT
GGAGAGAAAGGAACTGTAATAAATAATGGCACAATAATATCATATAGAGATTTACTATTTGT
AGTAATAAAAGATGTAACTATTGGTAGTGAAGAAGGTGACAATAGCCCAGTTCAAGCTCTGG
AAGTTGGTAAGAAATATAATTTACCTACAAATTGTGAATTTAAACTAGTTGATAATATATCT
GGAGTAACAAAGATTACTAACACAAGAAGTTTTGAAGGTGGTACAGATATAGAGACAGATGA
AGAACTAAAAGAAAGATTTTATAAAATCCAAAGAAATCAAGCTACAAGTGGAAATAAAGCTC
ACTATGAAGAATGGGCTTTGGAAGTAGATGGAGTCTATAATGTTAAGGTTTATCCAAGATGG
GATGGTCCAGGAACAGTTAAGGTCTTGATATTTGGGGAAAATAATCAAGCTGTTGATACAGA
AACGATTGAAAGGTGTCAGCAACATATAGATGAAGAGAAGCCTATTGGACCAACTATAACAG
TTGTGACACCATTACCAATAGAAATAAGTATAAGTGCAGTAATGAAACTAGAAGATGGATAT
ACATTAGACAATGTAAAAGAATCTTTCCTAGAAAGTATAAATACATACTTTAGAGATATTAG
AGGAGAGATAATCTATACAAAAGTCATGGGAATACTTATAAATACTACTGGTGTACACGATT
TAAGTAATCTACTTATAAATGGAAGTACAGATAATATAACTATTAATGAAGATAAAATACCT
AGTGTAACAACTGTTAATTTTAGTGAGGTGGAAAATCAATGAAGCTAATTGATAAACTACCA
TCATTTGATAGAAATTACATTGTAGAGGAGATACAAGGTGCATACGATACAGAATTAAATAT
TCTTAAAGAAGATATTGATGATACCTTTAACCAATTATTTGTTGATACAGCGACATGGGGAT
TAGATATGTGGGAAGACATACTCTGCATTGaAAAAAAAGAACTTGATTTTGACACAAGACGT
AGCAATATAAAAGCTAAAATGAGAAGCAGAGGTACTAGTACTATTGAAGTTATAAAAAGTAT
ATGTGAGGCATATACAAAATCAGAAACAGATATAAAAGTTTATAGTGATGAATTTACATTCG
TATTGAGTTTTATAGCAAATAACTGTGACTATAAAACTCTTTTAGATTGTAGCGAGATGATT
GAAAGAGTAAAACCTGCTCACTTATTACACTATTTAGAACCAATAATACTAGATAAAAGTAT
GGTCTATTGTGGTGGAGGTATGGTATGTAGTGAAGAGGTAAAAGTTCATCCATACTTTGAAC
CAATTATAAAATGTAGTGCTGTTGTAAACTGTGGAGCTGGAATGTTAAGTAGAGAAGAAATA
AAGGTTTATCCTTTAAGCATTAAATGCATTGAAAATAATTGTAAGATTAATATAGCTATTGC
AAATGATACAGGCGTAGAAAATGTAGTAGTTTATCCTAAATCGGAGGTGGTATAATTGGAAG
AAAAATTTTATATAATATTAACCAAAATTGGTAGAGAAAAAATAGCAAATGCAACTGCACTA
GGAGAGCTTGTTGGATTAACCAAGTTTCAAGTTGGAGATAGTAATGGAGAATATTATGAGCC
AACAGAGGAACAAACTGCTTTAAAGAATGTAGTTTGGGAAGGAAATATAAATTCTCTAAGAA
TTGATGAAAAAAATCCTAATTGGATAGTTATAGAGACTATTTTACCAGGAACAGTTGGTGGA
TTTATGATAAGAGAAGCTGCTGTTCTGGATAATGAGAATAATATAATAGCTATWGGTAAGTA
TCCAGAGACGTATAAGCCACGTGCTGAAGATGGCAGTATTAAAGATTTGGTTGTAAAAATGA

TTTTACAATTGTCCAATACTTCAAATGTTACATTAGAAGTAGACCCGACGTTGGTTTTTGTA
ACTCAAAAGGATATTCAAGATTTAGATGATAAGTTTGATAAAAATATAAAAGAAATAAAAGT
AAAAATTGGAGATACAGATATATTAACTACAGATTCTAAAGATTTATCAGGAGCTATAAATG
AGGTAGTTAAAAAAATAGAAAATATATCTTTTGATGATGTTATAAGTGGTCAAATACAAACT
GATATATCAGTATTAAAAAATAGCTATAACAAATTATCTGAAAAAGTGCTAGATATATTAAT
ATACCTAGAATTAGAGTCAGAAGTAACTGTAGATGAGGCTGGTTATTGGTATGATACATTAG
CAAATGGAAATAACATAGTAGCTATAGAAGGGCTTAAGTTAGATTTAAATAGAAAATGTATA
ACAGGTGAAATTGGTAATGTGATTTTTAGAGATGTAGTATTACCATTTAGTGCAAATAGAGT
TAGATATATACATGATATGGATAATAACTTTGTTGAGACAAAATCTAGTAACACTTATTTAA
AAGAACAAAAAGATATAACTCTAAGTAAATATTCATATGAAATAAGATAAATAAAGGAGGTA
GTACTAATAATGAAGCAAAATAAACTTTTACAGCGTGGTGCTTATTTTAATGATAAGAACAT
ATTGATTGATGATTTTGATAAAAGATATAATGATTATGATTTTGTAGAATTTTTTACTGGTA
TAAGTAATAGTACCTTTGGTTTAAAATCAGATGGTAATTTATATGCTTGTGGCGATAATACA
GGTTTTCAACTAGGACTTGGAAAAGATTCGTCAGAGAGAAGGATGTTTAGTAAAGTAAAAAT
TGATAATGTAAAATATGTATCTTGTGGTTCAAAACACAGTGTAGCAGTAACTAAAGATGGAT
TTGCATATGGAGCAGGAACAAGTAATGTAGGTCAATTAGGTGTAATTGAGTCTACAGTATAT
TATGAATTTACTAAGCTACCAATAGATGATGTAAAAACTGTTGCATGTGGTTATGACTTTAC
ATTTGTGCTTAAAAATGATGGAACATTATATTCAGCAGGTTTAAACTCAAGTGGTCAACTTG
GACTAGGTGATACTAACAATAGAGCTACTTTTACTAAAGTAAATATAGATAGTGTGAAAGAT
GTAGTGACTTATAATCAATCTGTATTTATCATAAAAATGGATGGGACAGCACATGCATGTGG
ATTAAATTCAAATGGGCAGTTGGGAATTAATAGTACTTTAAATAAAAGTGTATTTAATAAAA
TAGAAGGTATGGATAATGTAAAACAGATAGCGTGTGGTAGTAGTCATACAATTCTTATTAAG
AATGATGGAACTATGTATACTACAGGCTATAATGGAGTTGGTCAGCTTGGTACAGGAAATAA
TAATAATTCAATTGTATTTACTCTTTCTAGTATAAATAATGTTAAGTATGCTTCTTGTGGAA
ATAATCATACTATGATATTAAAATACGATAATACACTGTTTAGTACAGGACAAAACAATTAT
GGTCAACTAGCCAATGCCAATAAAGATGTAGCATCAAGAAATACTTTTGCTAAGGTTAATGT
AGAAAATATAAAAGATATTAAATGTGGTTCTCAATTTAATTTTTtAATAAATGGTTCAAAAG
AGATATTTGTATCTGGCTGTAATTTAGCAGGTCAACTTGGTTCATTTTTTCATACAACTTTT
CTGTATGAGTTTTCAAATGTGCAATCTTCAAATTTAGATAATTATTCAGGTTTATTGGTTAA
TGATGATTATTTATATGTTACAAAGGACAATAGTGAATTTTTAAATGTAAAGTTAAGTGATA
ATTTTCAAGATTATAAGAAGATAGAGTTAACAGATAGCAATATGTTTATTGTTATGAATGAT
GGTACATTGTATGCTTGTGGTTTAAATAATTATGGACAGTTAGGATTGGGAGATACTGTTAA
CAGGTCAGTTATGACTAAGGTGGATATAGATAATGTTTTGGATATAAAAGGAAACGGAAACT
CAACTTTTGTGCTTAAGAATAATGGAACATTATATTCATGTGGTTTAAATAGTAATGGACAA
TTGGGTTTAAGAGATGAAGTTAATAGAAATATATTTACAAAAATAGAAATAGAGAATGTAAA
GGATTTTTGTGTAGGAAGCAATTATGTCATAGCTTTAAATCACTCAAAAGAAGTATATGGAT
GGGGAAATAATCCTTATAATAATATAGAAAAAACTTCTAATTATCCATATAAGCAGGGAATA
AGTAATATTGAAAAGATAGCAGCATATGATTATTCTGTATATATGATAAACAGTGAAGGGAA
ACTATATGTTTCTGGATACAATTATAATTATCAATTAGGTAAAGGAAATAATAGTAACCAAA
GCAAAGCATTAGTATCTCAATGTAGAACAAATTCAACATCTTCTACATCAAATGGACTTAGA
ACGTTACCTAAAATAACTAATGTTTTtCCTTTTTATGATGGTTGTGCAATAATTGACGAAGG
AGGTTATGTTTATTTAACAGGATATCATGGATATTTAAGAACATTAAATAGCAGTCCAAGTA
TATCTGATTATTCAAGATATGGAACTTTTATTGAGGCTACAAATTCAAATCATAATACTTAT
TTTATACAAGAGACTGATTTTAGTGGAATTGAAAAAGTAATAGGGATGTCAAATAATATATT
ATTTTTTAAGAAAGGAAGTTCATATATTACTGGATATCCAAAAACATTTGGCTCAACCATTA
CTGGACATAGAAGTTATACTAGTATTAATTCTGAGAGTTCTAATTTAGGAAGTAATTTTATA
ATATATCATAGTAATTCCAAGTTATATGGAAAAGGGATTGCTAATAGTGGGCAATTTGGG

SEQ ID NO:34 3' fragment

AATTCAACAAATATAGATGGCACAAGTAACTATGATACAGGATTAAAAGACATAAAAGATAT
AATTGTAAAAGGAAATACTGTAGTAGTAGTAGATAAAAATAACAATATATATGTAACAGGAA
TGAATCAGAATAACAAACTTGGGATAGGGGAATATAACAACGAACCAGTAAAAAAATTCACA
AATATAACTGAACAATCAAACTCATTTATATTTATGGATGATATAAAAGAAATTACAACATC
AAGAAATACAATGTTTATAGTAAAAAATGATGGAACAGCCTATGCCACAGGAAATAATAGTT
CTGGACAATTAGGATTAGGTGACACAATAAATAGAAATAAGTTCACTCAGATAAACCTTGAT
AATATAAAGAAAATATCAACAAGTATAGATGGTAACACAACATTTGCAATTAGAAATGATGG
AACACTATACTCCACAGGATTAAATACCAAAGGACAACTGGGATTAGGTGATATAGTAAATA
GAAATACATTTACCAAAGTAAACATCCAAAATGTAAGAGATGTTGTTTTAGGGACTACTCAC
TCGCATGCAATCAAAGATGATAACACATTATATTCATGTGGAGaAAACACTCATGGGCAACT
GGGCTTAGGAAGCGAAAGCAACCATCCAGACGTATTGACATTTACTGTAAACAATATAACTA
ATGTAAGAGATGTGTACTGCTCAGATACAACAACATTTATTGTAAAGGACACAAACATTGCA
TATTGTTGTGGATACAATAATAATTCACAACTAGGTATGGGAAATACTACTGACCAGTATAG
TTTTATAAAGTGTATGGAAAATGTAAAAGAAGTTATACCAAATGAAATAAATACCTATATAA
TAACAATCTATAATACTGCATATAGTACAGGTTTAAATACTGATTATTGCTTAGGTCTAAAT
AGTAATAGCAATCAAAGTTCATTTTCTGAAATTCCAATTTCAAATGTAGTAAAAGTAGCTCC
AAACAGAAATAATGCAGTACTTTTACTTACAAGTGAAGGGGATGTATATACTGCAGGCAAAT
GTAGTAATGGTTCAGGTACAGGAAGTGAGACTCCAGAGAAGATTAAAAAAATAGCATCAAAG
GCAAAGGATATTGGAATGAATTATAGATGTGGACATTATGTAAGTGATAATGGAGACCTATA
TGGTACAGGTTTTAATAATAATGGACAATTAGGTGTTGGTGATGTAACAAAAAGAGATACAT
TTATAAAAACCAATACAAGAGTAAAGAAAATACTTCCTTTAGAATATGCAAATATAGCAATA
AAAGATACTAATGATATATATATTTGTGGATTAAATAACTATGGACAATTAGGTGTTGGAAA
TAGATACGATAGTAGAAATAATGATAATAGAATATTTAATTATAAGCATATGAATTTTGTAA
TGGGTGATTTGACATCTATTAAAAACAGACATAACTTTATACTTCTAAACAATAAGATAGTG
ATACCTACCACAAAAGACATAGATTATGGTTTAGTATTAGGAAATTTATACAAAGGAGACCT
TTATACTGAGCTTCCATATGAAGATATAAAAGAAGTATCTATTTCTAAGACTCATATTATTA
TATTACTTAATGATGGAACAATGTATGGATGTGGTACAAACTACCATGGAGAATTATTGCAA
GACTTGTCTATAAATCAAGTGGATGAATTTGTGCAGATTAATGTATCAGATGTAAAGCATGT
TTCATGTGGAGATAACTTTACTTATTTTATAAAATCTGATGATAGTCTTTGGTCTATTGGTA
AAAATTCCGAATATCAATTAGGTATAGGTCACAATAATCCAGTTACTGAATTACAAAGAATT
ACAACTATATCTAGCTGTAAAGAAGTACATTGTGGTAAAAACTATACATTAGTAGTAACTAC
AGGTAATGAATTATTTGTACAAGGATATAATGATAAGGGAGCTTTAGGATTAGGAAGCGATA
GTGAAAATACTATAATTAAGTTCTTTACAAAAGCACTAACAGACATAAGAGAAATAAAATCT
TATGGAAGTGACCATATATTAGTACTTAAAAATGATAATTCAGTATGGGTTACTGGAAAAAA
TAGGGATGTATATAAAATTGAACAACCAGTAGAATTTTTAAAAGAATTTACTATAGTACCTA
TTTCTGAAGATGTAAATACAGTAAAGGATGTACTTGCAACAGACAATACATTATATATTATA
TCAGAAGTAGGAACGACAAATGCTGCTATAGAAATTACTGAAAAATCAATTTCATCAATTAA
GATAAAAATACAAGACCCTAATAAAGATATAAGTAGAATAGAAATGCTTATAAATGGTGAAA
GTGTAAAATCTGTAAGTGATTTAACTACTGAAAAAATATCCTTTGAAGTACCACCAGATAAA
ATTAAAATAGGAGAGAATAAGATACTATTTAGAGCTTATTGTAAAGGTGATGATTTATATGC
ATCTTTATTTATTTTTAAAGAGAGTACTGGAAATTCTATAATTAAAGATTCTTATGTTATGA
TAGGTAATAGAATGTACAAGGTAGTTAATACAACATCTAATGAACAAGATATTACAATTACA
CTAGATAGAGGACTTGAAGAAGATTTAAATCTTGGAGACCCTATATATCAATTAATAAATAA
AACTAAAGTTCAAGTAAAAATAAATAAATCTGACTTATTCAAAGACATGAAACTAGTTGAAA
TCAAAAAATCAGACTCAAGTTATCAAGAAATCTATGAATTAGAAGAAGCCAACATAAAAAGT
GCTCAGCCTAAAATCATAGTAGAAAAAGGAGATAAATGGACAGCTATAAAACGTCCATCTAT
GATTTTTAGATATGATGCTGAAAACAACGAGCCACAAGCTTAAAATGGAGGTGTAAAAATTG
TTTAAATTCGATAAAAATAAAATAGAACAAATCAAACAAGGTAGAAAAGTAGAAATGCAGTA
TAAAGACATTTCAGACATAAGTATAGGTCAAGCAAAGCAAGATGATGATATAACAAATAATT
TTATAGCAAATGCAGAAATATATGAGATGTTGTTAAGTCAAAGTTCTGTCAATGAAGCAAGT

AATATAAGCACTTTTAGTGTAAGAAAATCTGGAGGTGAGAGTGGAATGGTAGAAGTATATGT
AGCTTTAATTTTAAGAGGCAGAAAAACAATAGAAGAAGTACCAGCAGTAATTAGAGAGCAAG
TTAGAATTAGATGTAAAGAATTAGAAATACCAGTTGAATAGTAAATTTAGAATAACTATGTA
TTAGTTATTTTTTTTATGTAAAGTACAAGGTCTTAACTTTAATAAGTAAGCCTTGTACTTAT
TTTTTGTTATATTAGAAATTGTATATATATTTATTATTTATTCAATCTATAAATTAAACCTA
CAATTTAAAGTACAGAAGATTAAATTGATAATCCTGAAAATATAATATTGCATGATGTAAGA
ATATAACAAAAATTAAAGCTATAAGTATAAAAAATTTAGACAATAGGAGGCTATAATGGATA
AATTAATAACCGAATTGAGTAGTCTGGGGGCAATAGGTATACTATGTGCTCTATTATTTAAA
AATACTATGCAGGAGAAAAAAGAAGATAGAGACATGTATAAAAAAACTGTAGAAAATTTTAT
AGAATTATCTACACAACAACAAGAAATAAACAAAAATATACTTGTTCAAATGGGAATAATGA
AAACAGATGTAGAGGAAATTAAGGAAGATGTTACTGATATAAAAGGTATGTTACAAAACGGT
GTATAACATGAAAGTAGCAGTAGCACCAGATTATATATTATTAGGAAAAGATAAAGTAGTAT
TGTAGATAGTGCCCTATTTTATTGAGAAGGATTTTATATTTTAAAATATTAATTAAAAAAAG
TAATAAAAATAATATATAAAAATAACATATAAAAATTCAAAAAGGAGTTAAGCTTAAATTTG
ATTAGAAAAAATCAATTTTAAGACAACTCCTTTTTTTtATTAAATTATTGTCTATTAACCAA
AATAGCTATTTTAGCATCTGGATTATAACTTATCTGAACCATTTGATTTTTCTTAACATGTT
CAAGGTCTTCACCACCATAAGCTATTTGTAACTTAACTGGTAACTTACCTTGTTTTATAATA
GCAACGTACTCTTTTTTACCTTTTTCTCTAAACTAATCAAATTGCCAACATAAGGTTTAAAG
TTCTGATACTTTTTACTAGAATTTCTTATGTAGAAGAAAGCACCAACAGCAATAACTAAATT
TATGCCAAGTGTAACCCAAGAATTGATTTTAAGCATAGCTCCAGCGATTATTATCACGAACA
TTAAAACGATAGGTAATATAGCTTTCTTAAGAAGCAATTTACCCATTATTTCATTAGCCTTT
TTCTCAGGGCCACTCATAGTTTTTGATCTAGCAAATGATTGCGCGAATTTGTCTCTTAAGCC
CATTTTATCCTCCTAATTTTAATAAATATTTAGTTATAATAACGAGATATTACTTGAAACTA
AAAATTTACTACATTTATATTATGTTTGACTTTTGTATAAATAATTACATTCAAGTAAAGCA
AAATATACTAATTATTTTATCATAAAATTATAAAAAAGAAAATAAATGAAATAAAAATATTA
GAACAAAGAAATGATGTAAAATCGTATCAAAAGCAACATAAAAATTATTTATCTATTTTCTC
ATCTTTATTTTTGTTATACTCAATTTTTCCTAAATCCTTCTCTTTTTCATATTCATGAAGTT
TTAATTCAATCATACCTTCTATTTTGGCTTTATCATAATCATTTAACTTTCTAAAGTTGTTT
AAAAGCTTTATTTCATTAGAGTTTATGCTATTAAGTGGATAGTTTGAGGAGGAATCGCAAAT
TAAATCTGATTTATGTGATAGATTATCTCCATTTAATAGCCAGTCTACTGAAACATTAAATA
TCTCAGCTATAGATTTTAATATTTCATAATTTGGTTTTCTAATGTTTCTCTCAAATTTACTT
AAGTTGTCACAGCCTAACATTTCTTCTAGTTCATATTGTTTAAGGTTTTTTGCTTTTCTCAA
ATAAACAATTCTTTCTCCTAAAGTATCCATAAACACTCTCCATTCAATTAATGTCAAAAAGA
CTTTTTAAGATGTAAATAGTTTCAAATTAAAGGTCAAAATGACATAAAAACCATTGACTTAA
GGTCAAAATGACTTTATAATTAACTTAATGATACGAATTTACATCCTAATTTTAGCACAAAG
TAATCAAAAAATCTTATTTAGTATTAAATAAATTTATATACTTAATATGTGTACATATTAAA
AATATATACTAAATAGAGGGGGTGCGTAAGCTAAAGTAATATAAAAGTAAATATAAATCACT
TAGAAAGGAAGTTGATAAATGGATGCTCGAAAAAAATGGATACCTTTTTTGGGAGTGCAAGT
CAAGCAAAGACTTATTGAATTAAATATGACTCAAAGGGAATTAGCGAAGAAAATAGGTGTTA
ATGAAAACTATTTGTCAGCTATTTTAAATGGAAGAAGAACAGGTAAAAAATATAAATCATCA
ATTTATCAATTACTTAATATAGAATATTCAGAAGATGATTAATAAATAGTATATAAAGTAGG
TGAATATTCTTGTGTGCAAATTGGATTCAGATGGGGTTATAGAGTGTTGTAGAGCAATTGAT
GATTTTATTACAGCACTTAGTAATATAAAAAGCTTAAATATGGAAAGATTAAATACTTTAAC
TAAATATTCTAGTACATGTTCAATCCTTCTTAAAGAGGGGAATTATGAAGGATGTACAATTG
TGTATAGAAAGATGTTGGAAGAATTAAAAACATGAGTAATGCATTTCTTAGGAATATAAATT
ATACATAGAAATGTATTATATTTTTCAAAGTACTTAAACTAAAATATGGATAAGATAATCTA
AATATTATAAATGTGCTTGAAATTAGACTATACTTGTTTTTAAATAATCCAATATCCATATT
TTAGTAATATACTACAAAAAAAGAAGGTTAATAGATGATGTAAAATCGTATCAAATTATGTA
TGTTTAAACCATTTTATCTTCATTATTATTAGAGGAATGCTTTTTTAAGTCTTTATATTCAG
ATATCTTAAGTTCAAGTATTCCTTCTATTTTTATTTTATCACGTTCGTTTAGTTGTCTGTAT
AGATTTAATATCATCATTTCATCATTAGTAACATGTAAGTAATCTTCTTTATCTTCTTTTAC

ACTACTATTGACATTTACCTTCTCTTTACCATAGAGAAGCCAGTCAGTCGTAACATTAAAAT
AATCAGCTATTGACATTAGTATATCACAATTAGGTTTTCTATCTCCTGTTTCATACTTGCCT
AAGTTTTCAAATTTTAAAATATCCATAAGTTTGCGCTGAGTAAGTTTTTTGGAGTTTCTCAA
ATAAGCAATTCTTTTTCCTAAAGTATCCACAAAATACACTCCTTTCTTTTTATGAGTAATGT
CTAAATGACATTTGAAATTAAAAATATATAAATTTATAATATAAAACTACTAAATTAAAGTC
TAAATGACATTTTGCTTAAATTAATATGCTCATAATATGATTTTAACATATTATAGTTGAAA
ATATATGGTTTATTTTGATTTGTATATATAACAATAGATTTAATTGTTATAAAAATGTAAAG
GGGTGTATGAATAGATTGTATAAATTTATTTCGATAAACTAAGATTGCTTTTTGATTGTCTG
TAAAAGAGAAAAAGATTAAGATAAAAATAGTATTATATTGTAATTTATATTAATCAATTACA
AAGATTTTATGAATTTATTCTTTAGGGTAAAATATTTAAGAATAAGATAAATTTACAATATA
ATACTATAACACTCTTTTATCTAGTTTTATTTTCTTTATAGAACAATAATATTATAAATGCT
AGTAGATTTACACAGAATACTGTTATATACATCTGTTTGAATCCTGAGTTTAGAGTAGATTG
TAGTGtGGATCCGG

SEQ ID NO:35 >AV1368
ttttttGCGGCCGCaatacccactacaccttcgtc

SEQ ID NO:36 >AV1289
Tatacatctcgagtcccatctctttc

SEQ ID NO:37 >AV1288
gaagaaagagatgggactcgagatg

SEQ ID NO:38 >AV1366
CTTGTGCCATCTATATTTGTTG

SEQ ID NO:39 >AV1367
GGAAAAGGGATTGCTAATAGTG

SEQ ID NO:40 >AV1300
tcccccGGATCCacactacaatctactctaaactcagg

SEQ ID NO:41 >DG1
GGCGCGCCACTAGTACCGGTGCCATGGCGGCCGC

SEQ ID NO:42 >DG2
AGCTGCGGCCGCCATGGCACCGGTACTAGTGGCGCGCCCATG

SEQ ID NO:43 >DG9
ttccttggtctcAcgcgAACAAAATTCTCCAGTCTTC

SEQ ID NO:44 >DG10
ttccttggtctcAggccGTCGCGACTAAGAAAATGCC

SEQ ID NO:45 >DG13
gtgagcggataacaattccc

SEQ ID NO:46 >DG14
AGATTGTAGTGtGGATCCGG

SEQ ID NO:47 >DG15
tccttcggcgcgccTCAAATTTAAGCTTAACTCC

SEQ ID NO:48 >DG16
TTTAGGGACTACTCACTCGC

SEQ ID NO:49 >CD4 Orf1374
MKRTKLLQRGNFFGDKNMVVDEFDEGYDNYDFINFFTGCCNYTFGLKNNNILYGCGDNSNFQ
LGLGEDNTTRKLFTKIPNISTNIKKVACGESHAVILTSDGELLVAGINTDGQMGLGLEKVGK
TVSTFEKVPEIKGVKDIACGLQSTYLLYNDGTLYVAGNNLYGQLGLGTNGASANVNTFTKVD
VDNVKAVFSYNKSAFIIKNDNKCYSTGFNNQGQLGLGDKNNRDLFSLVSINDVKTIACGSEH
TVLMTYNNDIYGCGKEKCFGNALQSSLFTKIEEVNIKTIACGHGNTMLIDNKGTLKVAGNND
IYQLGIANYSENIDNSFIDLKNIVAKNIFIGLSHSILIDSNNDSYCTGDNTYGQLGSFFDDM
HIVEFKKMDSEKYSYSNYINLIKSEDKLTLLKEEMEIKDIELPLDIHSVRDVVFSPYCTLVI
LGNGDVYGLGNNRYKGMGSDLPSQLNELTKLSISNVKSIVASKNISGGIFYIKNDDTCYYSG
PNSNSIAGVLPSNSDVFKKISIDNVKKVVINTDLSNWFSLIVTNNKQIYTSGKSSSYVNGLS
NALISQYTEISLSNVTDAYSSYNATFIVVDEKKVYATGINTNYLLGFSTSDGSNVNLGLLSD
WYYINISGSSYSRVSCTNNITKINNIIIYEYVTVFCTNIGSFLTGYHGTSWTKPTDSSYRVQ
YQGISYAGYLDSYIYNYYPTRCTQSSSSTTFAYLYNGESSSNLKNVNPDNLLISGGSSYIHQ
YGRNYLNNQSSNNIAASNINSGPITSDKAIFLYKALLYLSSNTLYGFGNISESAKELDVSDT
QDGYNATNYKKVMKNIKNIFIPPYDLSRDKTRFAILTDKSLFICGYNSKGTHGISVNSSLNL
NNKINYNKKNSSSEISSNIQEIYSHSKSTYLLTNNNMLYSVGLNDVGQLGVGDEINRKVFTK
INIDNIKSINVNRFTDNSKHAFAIKNDNTCYAVGLNNSGQLGIGDNVNRNIFTKINVENVKY
VAVYGNTSLLLTNDGLLYGAGNNGKGQLGLGDTTSRNIFTRIPINGVRDVYLCNDVSIIVKN
DNTCYVCGLVNGYFGFTEGSISTFTKINIENVKSVVTAGSEATFFITNDNMIYTTGKKERVF
FSTETNDIKGIRVINNIINAKKIVVNGYTSAILTNDNKLFVGGLSGYGSIANNNNTNSVEDV
KDVFVTANNTLYIDNNNNLISSGRDTYGISDESYRDMSVPYYKVSIKKDVDTVFSSYNTIFI
KDIYGKFYSSTRDNRYNHLGIHHRYDNDKNEALEGSLHSYFKTDNTSDKIVFNKKNEKLVMF
NDKYIKTNNKYINYKNIFKDNFKYTSIILPFEVSDIDISKTHSLAVAKDGKLYGIGSNSYKE
INQTLEDIELLTLTEVNISDVKKVACGDNYSYIIKTDNTLWSYGKNTEYQLGVGHNNDVREL
QKVTGLPSVKDISIYNSMTLVLTNEGELYAQGYNTNGLFGLGESEKDKIIRTFTKVLTNVKE
IKSHNDDHILVIKNDNSLWITGKNKSMYKISISITDLYEFTKIPIPEHLNDILDIELSDDTI
YMITKVDTSKASIEIVEKSISQVRVVVQDPNNVIEKLEMFINDELISTKTNLEINSIIFEIP
QNKIVLGENKILIKASSPTGDLYSSMFIFKSETGLKVKKDSILMINNKVYSIINITENNTDL
IVTLNEGLKDDMMENNPIYQLINKTKVQVKINKSDLFKDMKLVEIKKSDSSYQEIYELEEAN
IKSAQPKIIVEKGDKWTAIKRPSMIFRYDAENNEPQA

SEQ ID NO:50 >CD108 Orf1374

MKQNKLLQRGAYFNDKNILIDDFDKRYNDYDFVEFFTGISNSTFGLKSDGNLYACGDNTGFQ
LGLGKDSSERRMFSKVKIDNVKYVSCGSKHSVAVTKDGFAYGAGTSNVGQLGVIESTVYYEF
TKLPIDDVKTVACGYDFTFVLKNDGTLYSAGLNSSGQLGLGDTNNRVTFTKVNIDSVKDVVT
YNQSVFIIKMDGTAHACGLNSNGQLGINSTLNKSVFNKIEGMDNVKQIACGSSHTILIKNDG
TMYTTGSNGYGQLGTGNNNNSIVFTLSSINNVKYASCGNNHTMILKYDNTLFSTGQNNYGQL
ANANKDVASRNTFVKVNVENIKDIKCGSQFNFLINGSKEIFVSGCNLAGQLGSFFHTTFLYE
FSKVQSSNLDNYSGLLVNDDYLYVTKDNSEFLNVKLSDNFQDYKKIELTDNNMFIVMNDGTL
YACGLNNYGQLGLGDTVNRSVMTKVDIDNVLDIKGNGNSTFVLKNNGTLYSCGYNSSGILGL
KDNTNRNIFTKIEIENIKEFCVESNYIVALNHSKELYGWGNQSYIVYGDNRNYPYKDTRVSN
VEKIATWSDTLYILDSTGATKTIGYSYNGSGGYPAPSSSSTYREGGYINKNTSYRTLEFYNT
SKTKLVNLFAFYNGCVFVDENGLAYCIGENNINFRGGSTTNENNSLRFINNSGVYYTNTDGT
DYTCYQWTYKLIRCSIFDSPQNIIGNSKNILYLSKNNSTFKCTGNCITYGINSQNWYSYFSD
SSNGAIALGNEFILKNYSGECLLKGYGKATNGEFGNSTNISSISNYDTGLKDIKDIIVKNNT
VVVVDKNNNIYVTGANQFNKLGIGEYNNQPIRKFTNITEQSNSFIFMDDIKEITTSRNTMFI
VKNDGTAYATGNNSSGQLGLGDTINRNKFTQINLDNIKKISTSIDGNTTFAIRNDGTLYSTG
LNTKGQLGLGDIVNRNTFTKVNIQNVRDVVLGTTHSHAIKDDNTLYSCGENTHGQLGLGSES
NHPDVLTFTVNNITNVRDVYCSDTTTFIVKDTNIAYCCGYNNNSQLGMGNTTDQYSFIKCME
NVKEVIPNEINTYIITIYNTAYSTGLNTDYCLGLNSNSNQSSFSEIPISNVVKVAPNRNNAV
LLLTSEGDVYTAGKCSNGSGTGSETPEKIKKIASKAKDIGMNYRCGHYVSDNGDLYGTGFND
CGQLGVGNVTKRDTFIKTNTRVKKILPLEYANIAIKDTNDIYICGLNNYGQLGVGNRYDSRN
NDNRIFNYKHMNFVMGDLTSIKNRHNFILLNNKIVIPTTKDIDYGLVLGNLYKGDLYTELPY
EDIKEVSISKTHIIILLNDGTMYGCGTNYHGELLQDLSINQVDEFVQINVSDVKHVSCGDNF
TYFIKSDDSLWSIGKNSEYQLGIGHNNPVTELQRITTISSCKEVHCGKNYTLVVTTSNELFV
QGYNDKGALGLGSDSENTIIKFFTKALTDIREIKSYGSDHILVLKNDNSVWVTGKNRDVYKI
EQPVEFLKEFTIVPISEDVNTVKDVLATDNTLYIISEVGTTNAAIEITEKSISSIKIKIQDP
NKDISRIEMLINGESVKSVSDLITEKISFEVPPDKIKIGENKILFRAYCKGDDLYASLFIFK
ESTGNSIIKDSYVMIGNRMYKVVNTTSNEQDITITLDRGLEEDLNLGDPIYQLINKTKVQVK
INKSDLFKDMKLVEIKKSDSSYQEIYELEEANIKSAQPKIIVEKGDKWTAIKRPSMIFRYDA
ENNEPQA

SEQ ID NO:51 >CD123 Orf1374

MKRTKLLQRGNFFGDKNMVVDEFDEGYDNYDFINFFTGCCNYTFGLKNNNILYGCGDNSNFQ
LGLGEDNTTRKLFTKIPNISTNIKKVACGESHAVILTSDGELLVAGINTDGQMGLGLEKVGK
TVSTFEKVPEIKGVKDIACGLQSTYLLYNDGTLYVAGNNLYGQLGLGTNGASANVNTFTKVD
VDNVKAVFSYNKSAFIIKNDNKCYSTGFNNQGQLGLGDKNNRDLFSLVSINDVKTIACGSEH
TVLMTYNNDIYGCGKEKCFGNALQSSLFTKIEEVNIKTIACGHGNTMLIDNKGTLKVAGNND
IYQLGIANYSENIDNSFIDLKNIVAKNIFIGLSHSILIDSNNDSYCTGDNTYGQLGSFFDDM
HIVEFKKMDSEKYSYSNYINLIKSEDKLTLLKEEMEIKDIELPLDIHSVRDVVFSPYCTLVI
LGNGDVYGLGNNRYKGMGSDLPSQLNELTKLSISNVKSIVASKNISGGIFYIKNDDTCYYSG
PNSNSIAGVLPSNSDVFKKISIDNVKKVVINTDLSNWFSLIVTNNKQIYTSGKSSSYVNGLS
NALISQYTEISLSNVTDAYSSYNATFIVVDEKKVYATGINTNYLLGFSTSDGSNVNLGLLSD
WYYINISGSSYSRVSCTNNITKINNIIIYEYVTVFCTNIGSFLTGYHGTSWTKPTDSSYRVQ
YQGISYAGYLDSYIYNYYPTRCTQSSSSTTFAYLYNGESSSNLKNVNPDNLLISGGSSYIHQ
YGRNYLNNQSSNNIAASNINSGPITSDKAIFLYKALLYLSSNTLYGFGNISESAKELDVSDT
QDGYNATNYKKVMKNIKNIFIPPYDLSRDKTRFAILTDKSLFICGYNSKGTHGISVNSSLNL
NNKINYHKKNSSSEISSNIQEIYSHSKSTYLLTNNNMLYSVGLNDVGQLGVGDEINRKVFTK
INIDNIKSINVNRFTDNSKHAFAIKNDNTCYAVGLNNSGQLGIGDNVNRNIFTKINVENVKY
VAVYGNTSLLLTNDGLLYGAGNNGKGQLGLGDTTSRNIFTRIPINGVRDVYLCNDVSIIVKN
DNTCYVCGLVNGYFGFTEGSISTFTKINIENVKSVVTAGSEATFFITNDNMIYTTGKKERVF
FSTETNDIKGIRVINNIINAKKIVVNGYTSAILTNDNKLFVGGLSGYGSIANNNNTNSVEDV

KDVFVTANNTLYIDNNNNLISSGRDTYGISDESYRDMSVPYYKVSIKKDVDTVFSSYNTIFI
KDIYGKFYSSTRDNRYNHLGIHHRYDNDKNEALEGSLHSYFKTDNTSDKIVFNKKNEKLVMF
NDKYIKTNNKYINYKNIFKDNFKYTSIILPFEVSDIDISKTHSLAVAKDGKLYGIGSNSYKE
INQTLEDIELLTLTEVNISDVKKVACGDNYSYIIKTDNTLWSYGKNTEYQLGVGHNNDVREL
QKVTGLPSVKDISIYNSMTLVLTNEGELYAQGYNTNGLFGLGESEKDKIIRTFTKVLTNVKE
IKSHNDDHILVIKNDNSLWITGKNKSMYKISISITDLYEFTKIPIPEHLNDILDIELSDDTI
YMITKVDTSKASIEIVEKSISQVRVVVQDPNNVIEKLEMFINDELISTKTNLEINSIIFEIP
QNKIVLGENKILIKASSPTGDLYSSMFIFKSETGLKVKKDSILMINNKVYSIINITENNTDL
IVTLNEGLKDDMMENNPIYQLINKTKVQVKINKSDLFKDMKLVEIKKSDSSYQEIYELEEAN
IKSAQPKIIVEKGDKWTAIKRPSMIFRYDAENNEPQA

SEQ ID NO:52 >CD126 Orf1374
MKQNKLLQRGAYFNDKNILIDDFDKRYNDYDFVEFFTGISNSTFGLKSDGNLYACGDNTGFP
LGLGKDSSERRMFSKVKIDNVKYVSCGSKHSVAVTKDGFAYGAGTSNVGQLGVIESTVYYEF
TKLPIDDVKTVACGYDFTFVLKNDGTLYSAGLNSSGQLGLGDTNNRATFTKVNIDSVKDVVT
YNQSVFIIKMDGTAHACGLNSNGQLGINSTLNKSVFNKIEGMDNVKQIACGSSHTILIKNDG
TMYTTGYNGVGQLGTGNNNNSIVFTLSSINNVKYASCGNNHTMILKYDNTLFSTGQNNYGQL
ANANKDVASRNTFAKVNVENIKDIKCGSQFNFLINGSKEIFVSGCNLAGQLGSFFHTTFLYE
FSNVQSSNLDNYSGLLVNDDYLYVTKDNSEFLNVKLSDNFQDYKKIELTDSNMFIVMNDGTL
YACGLNNYGQLGLGDTVNRSVMTKVDIDNVLDIKGNGNSTFVLKNNGTLYSCGLNSNGQLGL
RDEVNRNIFTKIEIENVKDFCVGSNYVIALNHSKEVYGWGNNPYNNIEKTSNYPYKQGISNI
EKIAAYDYSVYMINSEGKLYVSGYNYNYQLGKGNNSNQSKALVSQCRTNSTSSTSNGLRTLP
KITNVFPFYDGCAIIDEGGYVYLTGYHGYLRTLNSSPSISDYSRYGTFIEATNSNHNTYFIQ
ETDFSGIEKVIGMSNNILFFKKGSSYITGYPKTFGSTITGHRSYTSINSESSNLGSNFIIYH
SNSKLYGKGIANSGQFGNSTNIDGTSNYDTGLKDIKDIIVKGNTVVVVDKNNNIYVTGMNQN
NKLGIGEYNNEPVKKFTNITEQSNSFIFMDDIKEITTSRNTMFIVKNDGTAYATGNNSSGQL
GLGDTINRNKFTQINLDNIKKISTSIDGNTTFAIRNDGTLYSTGLNTKGQLGLGDIVNRNTF
TKVNIQNVRDVVLGTTHSHAIKDDNTLYSCGENTHGQLGLGSESNHPDVLTFTVNNITNVRD
VYCSDTTTFIVKDTNIAYCCGYNNNSQLGMGNTTDQYSFIKCMENVKEVIPNEINTYIITIY
NTAYSTGLNTDYCLGLNSNSNQSSFSEIPISNVVKVAPNRNNAVLLLTSEGDVYTAGKCSNG
SGTGSETPEKIKKIASKAKDIGMNYRCGHYVSDNGDLYGTGFNNNGQLGVGDVTKRDTFIKT
NTRVKKILPLEYANIAIKDTNDIYICGLNNYGQLGVGNRYDSRNNDNRIFNYKHMNFVMGDL
TSIKNRHNFILLNNKIVIPTTKDIDYGLVLGNLYKGDLYTELPYEDIKEVSISKTHIIILLN
DGTMYGCGTNYHGELLQDLSINQVDEFVQINVSDVKHVSCGDNFTYFIKSDDSLWSIGKNSE
YQLGIGHNNPVTELQRITTISSCKEVHCGKNYTLVVTTGNELFVQGYNDKGALGLGSDSENT
IIKFFTKALTDIREIKSYGSDHILVLKNDNSVWVTGKNRDVYKIEQPVEFLKEFTIVPISED
VNTVKDVLATDNTLYIISEVGTTNAAIEITEKSISSIKIKIQDPNKDISRIEMLINGESVKS
VSDLTTEKISFEVPPDKIKIGENKILFRAYCKGDDLYASLFIFKESTGNSIIKDSYVMIGNR
MYKVVNTTSNEQDITITLDRGLEEDLNLGDPIYQLINKTKVQVKINKSDLFKDMKLVEIKKS
DSSYQEIYELEEANIKSAQPKIIVEKGDKWTAIKRPSMIFRYDAENNEPQA

SEQ ID NO:53 >43593 Orf1374
MKQNKLLQRGAYFNDKNILIDDFDKRYNDYDFVEFFTGISNSTFGLKSDGNLYACGNNTGFP
LGLGKDSSERRMFSKVKIDNVKYVSCGSKHSVAVTKDGFAYGAGTSNVGQLGVIESTVYYEF
TKLPIDDVKTVACGYDFTFVLKNDGTLYSAGLNSSGQLGLGDTNNRATFTKVNIDSVKDVVT
YNQSVFIIKMDGTAHACGLNSNGQLGINSTLNKSVFNKIEGMDNVKQIACGSSHTILIKNDG
TMYTTGYNGVGQLGTGNNNNSIVFTLSSINNVKYASCGNNHTMILKYDNTLFSTGQNTYGQL
ANANKDVASRNTFAKVNVENIKDIKCGSQFNFLINGSKEIFVSGCNLAGQLGSFFHTTFLYE
FSKVQSSNLDNYSGLLVNDDYLYVTKDNSEFLNVKLSDNFQDYKKIELTDNNMFIVMNDGTL
YACGLNNYGQLGLGDTVNRSVMTKVDIDNVLDIKGNGNSTFVLKNNGTLYSCGYNSSGILGL
KDNTNRNIFTKIEIENVKAFCVESNYIVVLNHSKELYGWGNESYIVYGNSRNYPYKDTRVSN

VEKIATWSDTLYILDSTGATKTIGYSYNGSGGYPAPSSSSTYRDGGYINKNTSYRTLEFYNT
SKTKLVNLFAFYNGCVFVDENGLAYCIGENNINFRGNSTTNENNSLRFINNSGVYYTNTDGT
DYTCYQWTYKLIRCSIFDSPQNIIGNSKNILYLSKNNSTFKCTGNCITYGINSQNWYSYFSD
SSNGAIALGNEFILKNYSGECLLKGYGKATNGEFGNSTNISSISNYDTGLKDIKDIIVKNNT
VVVVDKNNNIYVTGANQFNKLGIGEYNNQPIKKFTNITEQSNSFIFMDDIKEITTSRNTMFI
VKNDGTAYATGNNSSGQLGLGDTINRNKFTQINLDNIKKISTSIDGNTTFAIRNDGTLYSTG
LNTKGQLGLGDIVNRNTFTKVNIQNVRDVVLGTTHSHAIKDDNTLYSCGENTHGQLGLGSES
NHPDVLTFTVNNITNVRDVYCSDTTTFIVKDTNIAYCCGYNNNSQLGMGNTTDQYSFIKCME
NVKEVIPNEINTYIITIYNTAYSTGLNTDYCLGLNSNSNQSSFSEIPISNVVKVAPNRNNAV
LLLTSEGDVYTAGKCSNGSGTGSETPEKIKKIASKAKDIGMNYRCGHYVSDNGDLYGTGFND
CGQLGVGDVTKRDTFIKTNTRVKKILPLEYANIAIKDTNDIYICGLNNYGQLGVGNRYDSRN
NDNRIFNYKHMNFVMGDLTSIKNRHNFILLNNKIVIPTTKDIDYGLVLGNLYKGDLYTELPY
EDIKEVSISKTHIIILLNDGTMYGCGTNYHGELLQDLSINQVDEFVQINVSDVKHVSCGDNF
TYFIKSDDSLWSIGKNSEYQLGIGHNNPVTELQRITTISSCKEVHCGKNYTLVVTTGNELFV
QGYNDKGALGLGSDSENTIIKFFTKALTDIREIKSYGSDHILVLKNDNSVWVTGKNRDVYKI
EQPVEFLKEFTIIPISEDVNTVKDVLATDNTLYIISEVGTTNAAIEITEKSISSIKIKIQDP
NKDISRIEMLINGESVKSVSDLITEKISFEVPPDKIKIGENKILFRAYCKGDDLYASLFIFK
ESTGNSIIKDSYVMIGNRMYKVVNTTSNEQDITITLDRGLEEDLNLGDPIYQLINKTKVQVK
INKSDLFKDMKLVEIKKSDSSYQEIYELEEANIKSAQPKIIVEKGDKWTAIKRPSMIFRYDA
ENNEPQA

SEQ ID NO:54 >ΦC2 Orf1373
MAIDKSYYTIITDVGKAKIANASVTGNKVGFVKIQLGDGGGSEYTPTESQTALKNVVWEGNI
GNTTTDETAPNCIILESLIPSSVGGFMIREIGYLDDENNLIAISKYKECYKPSIEQGAVVDM
KVKTVLIVSNVNNIELKIDPTIIFATLKDIQDLETKIGTVNTKIDTTKTELTSNIETTKTEL
NTRIDTENEKQNIKIDQLIAGGSNVASTQIITIDDWVEDAENGFKATVTHSLLTQRIVVNII
DATTKENVVTNFKIIDDNSIEIRSEVKVELNVYVINGNAETHFINATVDDNRVSEMTTYSSK
KIEDRLVNIEEKVNGGLSNIATSVNELITYC

SEQ ID NO:55 >ΦCD119 Orf1373
MAEQQYFTLVTDIGKAAIANASVTGEKVDFAKIKVGDGGGSSYTPNESQTALKNVVWESTLE
HAQVDKDNPNWVVIQKFIPGDVGGFEIREVGLFDSKDQLLAVSSYPTTYKPESRFGDCKRTI
NKSNISCI

SEQ ID NO:56 >ΦCD27 Orf1373
MPNELNFNNEIEEYLITTPAHANEFNNRQQKLLDNDKYLNNKIDTTKTELNTRIDTENEKQN
IKIDQLIAGGSNVAYTQRVAIDDWVEDAENGFKATVTHSLLTQRIVVNIIDATTKENVVTNF
KIIDDNSIEIRSETRSELNVYVINGNAETHFINATVDDNRVSEMTTYSSKKIEDRLVNIEEK
VNGGLSNIATSVNELITYC

SEQ ID NO:57 >DG211
ggccgcctcgaggg

SEQ ID NO:58 >DG212
cgcgccctcgaggc

SEQ ID NO:59 >DG210
TGAAGTACCATGGTATCCAG

SEQ ID NO:60 >DG209
ACTGGATACCATGGTACTTC

SEQ ID NO:61 >Dif4 Locus without orf1377-1379

GCAATACCCACTACACCTTCGTCATCTTTAAATTTAAGAGTTTTTACTATTGAATAATAAAG
GTATATTCCAGTAAAAATAATCTTTAAATACAAGAAAAATAAACTCTTTGGGTATATTAAAA
AGCTAAAAAGTGTAAATATAAAAGCAAGTAGAGTACTTATCCTGTAAAAGAAATCTATTTGT
GTAATGTCTTTATATTTTATCATAAACACCGAATATAAAATGATGAAAACAATTGCGACGAT
TGCATATATGGTAAATAACATATTTTCAAGAGTACCATTTGAAATTACTATCCACTTATACC
ACATAATTGGCCAAAATAATAGTGCTAAGAACTTAAAATAATTATCAAACAACTTTTCTTTA
TACATTCATCAAACAACCTTTCTTAACAAAAGCATATATTTGTTTTTAGAATTTTAAATAAT
ATGATATCATTATTATATATTAATATTGAATTTATAGAAACCAAAATTTGTTAAAATAAATA
TATAGATTTTACTGTTAAGCCAGTTAAAATTACTACTATTTTTATTATGAAATTGGATCAAA
TATGTAGAAATACGGCAAATTAGTTAATATTAAATATTTATTATTTCCAAGTTGTAAAGACT
GTTTTTTTAATGATAGAAATTCTAATCTTTTTTGAAAGAAAGTAATATCCACATTAAGTATG
TCTGCCATTTCATAAACGCAAGTGATGCCAGAGTTAATTATGTTTATTATATCTTCTTCAGT
AATTAAGAACTCACAAGCCCATTTTAAGGCTTTATTTTCGCACTTATCTATAATAATTTTTG
TATAATAATCGTTATAAGAGGATACATAGTATCCAAGGCTAGTGAAATGATGTCCAAGTTCT
TCAGCTAAGATGGATGTCAATTTTTTTGAGTTTTGTTTTAAATTACTGAGTAATGATATAAT
TTTAATACCATGTTTGTTTATATATAGTCCTTCTAAATCACCTGCAATATAAGTGGTATAAT
GAATTATTATCTCTTCTTGAGAAGCTAATTCAAAAAGCTTATCCAAATTATTCATAAAAATC
CCCCTAAAATAGAATGTATGTTTGCCTTTAAATTATATTAAAAGAGCAGAAAAATAGACTGC
TCATCATATGGTTTATTTTTTTTTATATTTATTTAGTAAAAATTCTATATAATCATTAAGTT
GTTCTTGTGCTTCTTCAGGTAACTCTTCATGTGGATTTTTTCTATGTGCAGCTACTGTATCA
ATATTTTCCTTAACTAAGGTTCTTCCAAGAAGGTAATCAACTGATACATTAAATACATCAGC
CAATTTATTTAAAATGTGTTCATCAGGAAATCTGTTTTCTGTTTCATAGTACCCTAAGACTC
TTTGGGAAACGCCTACTTTTTCTCCAAGTTCTCTTTGAGTCAATCCAAATTCCTTTCTAAGT
TCTCTTAATCTTTTGGCAAACATTATAACACCACCTTATGTATAGATTATAACAAATTGTTC
TAAAAAATAAAACTAATAAAATATAAAAGAATATTTTTTTCTAAAATCTATTGATAAAGAAC
AAATAATTCTATATAATCTAAGTGAGGAAGAACAAAATATTCTTAATAGTAATGGAGGTATA
AAACAATGTTTAAAAATAACTTGAAATATTATAGAAAATGCAAAGGTATGACACAAATTCAA
CTTGCCAGAAAGGCTGGAATTACAAATGATTATATATCTCAAATAGAAAGAGGTATAAAAAA
TCCTGGACTTCTTATGGCTAAGAAGATTTCTAGTATTTTAGAACAAAATATAGAAGAAGTTT
TTTTTATACAGTTATAGAACAATATGTTCTTGAAAGTTGTGAGATTAGTAAAAAACTGTGCA
CTAAAGAGATTATTGTAAATTTGAAGCTAATAATAAGTATATAAAAAAGGGGAAGTACTATG
GAAAACAAAAAAGATATATTATTTAAAGAAACAGATAAAAGATTACATAATTATAAGTATTT
GGATATAAAGATAAAGAATATTAACTTGGACATAAAAAGATGTGAGAATGAATACTCTGGAT
GTGGAGCAATGGTATATACAGAAAAGACTAGTAACACATATAACATAAGCTCTTCTGTGGAA
AATGAGGTGTTAAAAAGAGAGGAAAGATTAAGAAAATTAAAAATGGAAAAAGAAGATATAGA
AATAGAAAAAGAGAAGATAGAAAATGCTCTAACGTGTCTAAATGATATAGAAATGGAATTTT
TTAATCTTTTTTATAATAGTAAGACAAAAAACAATATGACATATATTTCTATGAAACTACAC
TTAGATAGAACATCTTGCTACAATTTAAAGAAAAAGATGATATTTAAATTGAGTGAGATATT
ATAAAAAATAGGACAATTTTACAACACTTTATATACACCATTGCAACAATAGGCAATAAAAT
ATGTGAGATAATGTTATTGTGAAAGAAATCCATATTGAAGGAGGTGATAGATTGAAAAGAAT
AATATTACCTAAAAATATAGAAGATACTTGACAGGAATAAATGAGATATATATTTAAAAATG
ACTTATATCATTTATAGTAAGATTATCAGATTAAGCAAGAATATTTAGTGATAGTGTGGTGA
TTATTTGCTTAAATACAAGGAAATATTAGAAACAATTATTGAGATTCTCAAAAAAAACTTTA
CTGAAAGTATTTTTATTGATGATGAAAGTGTGCAAGGCTCTGAAGGGTCTTGTTTTTTTGTA
AGTATACTATCAGTTATTTGTACACCTGTAATGTTAAATACGAATAACAAAGATATTGTTAT
CTCTATAAAATACTTACCAAAACCACAGTCAAAGAGTATTAGAATGTATGAAATTTCAGATG
AATTAAATAAGCTATTTAACAGAAATATAAAGGTAACAGACAGAAAATTAAATATAACAAAG
CTAGAACAAAGTATTAAAAAAGAAGAGTCAATTTATGTATTGAACTTTACATTTACACTAAA
CTATCTGGATAGTGTATATGAAGAAGATGTAGTATATGAAAATATGAAAGAAATCAATTTAA

ATTTAGGAGAGTGATAGTATGGCTATAGGATTACCAAGTATCAACATATCATTTAAGGAGCT
AGCTACAACTGTTAAAGAACGTTCAGCTAGAGGAATAATTGCAATGGTACTTAAAGATGCTA
AGGCACTAGGTCTTAATGAAATACATGAAAAAGAGGATATACCAGTTGATTTATCTGCTGAA
AATAAAGAATATATAAATTTAGCTTTGATGGGAAATGTTAACACTCCAAATAAATTATTAGT
TTATGTAATAGAAGGAGAAGCAGATATTCAAACTGCATTAGATTTTTTAGAGACTAAGGAAT
TTAATTATCTATGTATGCCAAAAGCAGTAGAAGCTGATAAGACTGCTATAAAAAATTGGATA
ATTAAACTTAGAGATATAGATAAGGTTAAGGTTAAAGCTGTATTAGGAAAAGTTGTAGGAAA
TCATGAAGGGATAATTAATTTTACTACAGAAGATGTGTTAGTTGGAGAAAAGAAATACAGTG
TTGATGAGTTTACAAGCAGGGTGGCTGGACTTATAGCAGGTACACCTTTAAGTCAATCAGTA
ACTTATACTAAACTTAGTGATGTAGTCGATATACCTAAGATGACGAAAGTTGATGCAGAATC
AAGGGTTAATAAAGGAGAGCTTATACTTATTAAAGAAGCAGGAGCTATAAGAATTGCTAGAG
GAGTAAATTCTTTAACTGAATTAACAGCAGAAAAAGGAGAAATGTTCCAGAAAATAAAAATA
GTTGACACTTTAGATATTATACATAGTGACATAAGAAAGGTGATAATAGATGACTATATAGG
AAAGGTTACTAACAGTTATGACAACAAATGTTTATTGATAGTAGCTATAAAAAGTTATTTAG
AAGAATTAGAAAAATCAGCACTTATAGAATCTGATTCTACTGTTGAAATAGATTTTGAAGCA
CAAAAATCGTATTTAAAATCAAAAGGAGTAGATTTATCTTATATGACATTACAAGAAATAAA
AGAAGCTAACACAGGTTCTAAAGTATTTTTAAAAGCAAAAATAAAAGTACTTGATGCTATGG
AAGATATAGATTTATCAATAGAAATATAGGAGGATTATTAATATGGCAAATATGGAAGCTAG
AAATGTAATGAGTGGTACTTGGGGAGAACTTTGGCTTGATGGAAACAAAGTAGCAGAAGTAA
AGAAATTTCAAGCAAAGATGGAATTTACAAAAGAAGATATTATAATAGCAGGTCAAATGGGT
ACTGATACAAAGTATATGGGATATAAAGGAAAAGGCTCAATAACTCTATACCACGTTAGTTC
AAGAATGCACAAGTTAATTGGAGAAAAGATAAAGAGAGGTTCTGAACCTAGATTTGTTGCTA
TATCAAAATTAAATGACCCAGATTCTTATGGAGCAGAAAGAATAGCAGTAAAAAATATAGCA
TTTGATGATTTAACTTTAGCTGATTGGGAGGTTGGAGTAAAAGGAGAGATAGAAGCTCCTTT
CACATTTACTGAGTATGATTTTCTTGATATAATTTAGTTTTATATTTAGTTTTATACTGATA
TTTAGTAAGTATATACTTAATAAATTCAGATAGTTAATAAGTAAAAAAGTTAGTTGATTGAA
TTTGATTGATAAAGGAGCAAATAATAATGAGTGAAAATGGATTATCAAAAAATATAAACATA
GTAGATTTACTTTTAAATTCAGATACAGAAAACTTAGAAAGACCAAGTACTATAGTTGAACT
TAAGAGATTATCAACTATATTTGGGCAGGAATTTAAAGTAATGTGTAGAGCTTTAACAATAA
GTAAAGATGAAGAAATACAAAATACTTGTCTTAAAATTGATGAAAATATGAAAACGGATATA
GACTTACCGGAGATGCAGATGCTTACAATTATAGAAGGTGTTTGTGATTTGGATGGAAAGCT
TTTATTTAAAAATAAGGAGCTAATGGATAAATTTAAGGCTCCAACACCAAAAGAATTAGCAA
GAAAATTATTATTACCAGGTGAAATTACCAACTTATATAGAATACTTCAAGATGTTATGGGT
TATGGTAAAAATGCAGTGATAGAAGAGGTAAAAAACTAATAGGGACGGATACCAGGACTACA
ATAATGTACTATTATTGGAAGAAAAAAGGTATAAGACCGTCCCTTTTTTATGCAATGGATAA
AGGCGAATTAAAGCTTATTGAAGCTTTTTTCGCCTTAGAAATTGAGGAAGAAGTTGAAAAAA
TGAAACATGGATATGGAGTGTGTCCTTTGACAGGAGGTGGTATGTAATGGGAAATGTGAGAG
AAGAAGGTATAAATATGTACCTTACAGATAATTACACACCAAAAATGAATCAAATTATATCA
GTAACTGATAATTTTAGGAGAGCAACTGTGGCTGTTTCACTTTCCACTAATGTAATGGCTAG
TAGCATAAAAAATTCTATTGGAAGTGCAAGTAATAGAGTAAACAGTTTAAATTCCTCGTTAA
GAAAAGTTCAAACTACTGCTAGTAGTGTAAGTTCAACTATGACAAAATTAAGTTCTAGCATA
AATGCTGTTTCAGGAGTTATTGGAAGTTTAAATGGAAGTATTATGAGACTAGCAATAACTAT
AGCTATGATTATTGATTATTTTAATAAGTTGATTCAAAAGAAAAATGAGTTTAATTCAAATA
TTATGATTATATTAATATTTAAAGCTAAAAGTGATGAAGTAGAAAAAACTAAAAATAAATTA
CTTGGAAATTTAAAAAAGATTGGTGGCAAGATTTGGAATATCGTAATAAAAGCAAAAGATAT
GACTAAGAGAGTGATAAGTAGTATCTTGGGAAAATTAAAACAAGTAGAGAAACGTCCTTATC
AAGGAAGTATTAATCTTAAAGATATGGTGAGTAGTGCTATGGGTAGAATTTTGCCTAAGTTA
ATGTTGTTTAAAAATACTTTTTGGAGTGGTGTAATAGCTATAAAAGATATGGCAAGTGGCAT
TATAAGTAAAGTATTTCCCAAATTGAGATTATTTGCAGGTAAGGTATGGAGTGGTGCAATAG
CTGTAAAGGATATGGCAAGTGGAATACTTGGTTCGATAAAAGGGAAGATATCTGATTTGACA
AATGGTGCTACTATAGGTGTCGCTGTGAAAAAGGGTGTTGATTTACTTGGTCAGGAACAAAA

TCAGAAAGTTGTTCTAGAAAGTGTAATGAAAAGAAATACTGGAAAAGTTAATCAAATAGATG
TTGATGATTATTATGGCAGTTTAGTAAGAATGGCAAATGATACGCCTTTTGACCCTGAAGAT
GTTGTTGCAATGGGAACTAAAGCTAAAATGATTAGTAATATTACTGGTGGCAAAAAAGAAAA
AGATATAACTCAAGCTATGGTAGATGTTAGAGCTTTAAATATGAATACAAGTAGTGAACAAG
ATGTATCAGCAGCTTTCTTAAGTGCAGCAAAAGGAAACATGGAATCTCTTAATACTCTGGTA
GGAGAAAATTATAAAACTTTTGATGAAGCATTGGAAGGCATAAGTGTAAAGCAGATGGGGTT
AGCTAAAGAAATGAGTAATACAATACCAGGTATAATATCAGGAGCTCAAACAAGCATTAACA
ATGGCTTGAAGAGTATTGTTAAACCTTTTGATGATATTTTAGGTCAAGGACTAAAGAAAATA
AAAACTTTTATAGAAAGTGGATTAGGCAATTTAGCTGGCTTATCTGAAAAAATGGCTGGTAA
AATAGGCAATGTAATGAATGGTAAGATAATTATTGGCAACAAATATGACCAGATGCAATCTA
GAAGTGTAAAAAATGGAAAAGAGTTTTCTGATTCTACTCAATATCGAATTTCTAATGAGGCT
GAAAAGCGTAAAATGATGGTTGAAAATAAGCAAGAACGTTTTGAAAATCATGCAGCAACAAT
GATAGGGAATGCACCAAAAGCAATTGTTAACGCAGGAAGTACACTATTACAAAATATCGATT
TTACAGCATTAATAGATTCATTACTTCCAGTAGTAAACTTAGTAAATAATTTACTAGATAGT
ATAAACAATAAATCACCAATTGCACAAGGATTAATAAGTATATTTGGTACAATAGTAACTAC
AGCATTCCAACTAATCGGACCTGTAGTTGAAGCTGTTAGTCCTATTATCACAAGAATTTTTA
CTTTTTTAGGTGAATATGCACCTCAAATAAACAATTTTATAGAGACACTGGGTGTTATTTGG
AAAACTGTATGGGAGACCTTAGGACCTCTGTTGGAAACTGGATGGAAAATTATAGAGCCAAT
ATTGGGAGCTTTTTTTAACATATTAGATAAAGTATGTAAAATAGTTAAAGATATATGTAAAT
GGTGGCAAACTATGATTAATAAGATAAAAAATGGAAGCATCACAGGAACAGTTTTAAATCTA
GTGGAAAAGAGTAAAAAAAATTACAAAGATAATCCATATGCTGGAACAAAGGCTGGTGATTC
TGGTAAAGCTTATTCAGGTAAGAAAGGTAATAATGCATTTGGATTGAACTATGTTCCTTATA
ATGACTATCAAACCAGACTCCATGAAGGTGAAATGGTTTTAACTAAACAAGAAGCAAATCAA
TATAGAAGCAGAAAAAATGGTGGAAATATAAACATAGCTAAGTTAGCTGATACAATAGTGAT
TAGAGAAGAAGCTGATATAGAAAAGATAACATCAAAATTAGTTGCAAGTATCCAATTGGCAC
AGTTAGGGGGTGTCTTATAATGGAAATGTGGCTTAGACAAGCTGAAGATAGATTTAGATTTC
CAGTATTTCCATCTTCCTTTAGTATTAATGGAAAAGCTGCTGTAAACTCTTCTAGTATACTC
AAAATAGGTGAAATAGCAACTTTTGGTGGTGTAGCTCTTAAAAGCATTTCAATATCAAGTTT
TTTTCCAAATAAAGACTATACTTTCTGTGACTATACAGGTTTTCCATCACCATATGATTGTG
TAAATAAGATAGAAAAATGGATGAAGGAAGGTTTTATATTAAGATTTACAATTACGGAAACA
AATATAAATATGGAAGTCATAATTGAAGGGTTTAGTTATGAAGAAAGAGATGGGACTCGAGA
TGTATATTTTACATTAGATTTAAAAGAGTATAAAAGAATAAAGATACCAAAAGTAACTCCAA
AACAATAACTATTATAGATAATAAGTTATAAATAACTGCTGATAGAATTAAATGAAAAGGCA
GGTGATTTTTTATTATTAAGATTTGGGTACACATAAAAAACGGAAGTATATATGACATAACT
GACATAGTAGACAAGGTATCATGGTCAGGTGATTATAAATCTCCATCAAGGACACTAGAGTT
TTCAATAATACAATCATCATTTGATGTAAATTTCCAACAAATCGATATACCAATAGCTAGTA
CAGTCTGTTTCTATGTAGATGAGAAAGAACTCTTTAGAGGAATGATAATTAATAGGTCTAAA
GATTCAAGCAGTAATGAAATTAGTTTTGTATCTAAAGATATGGGATTTTTACTTACACAAAG
TGAAGTGTCATACAATTTTAAAGATAAGTTAGTTGAAGACATAGCAAAGCAAGTATTTGCTG
AAAATAGGCTTTCAGTTGGAATAATAGCAAAGACCAATGTCAAGTATACAAAGATGTTTATA
GGAGTAAATGGTTATGACACAATAATGAGTGCATATACAGAAGCAAGTAAAAAGACAAAGAA
AAAGTATATGATAGAGGCCAATTTAGATAAGTTTAATGTTATTGAAAAAGGAACTGTTACAT
TAAGTGTTATGTTTGAAGAGGGATTTAATATTATAAATACCACCTTTTCGGAGAGCATGGAA
AATGTAAAAAATAAAGTAATAGTGGTAGACCAGTATGGAAGCAAGATTAGCGAAAAAATAGA
TAATGAAATTTTTAAGGAAGTAAATGTAATAATGCAAAAAGTAATTCAGCAACAAGAAAATC
AAGATGTAGATATTGATAGCGAGTTTAATGGGATAGAAAAAAGCTGTTCTCTTAAAGGTTAT
GGAGATGTAAGTTGTATAACTGGTAGAGGAGTAAAAGTTAAAGATTCTTATACAAAGCTTGT
AGGACTATTTTATATAGATACAGACAAACATACTTGGCAAAATGGAGAATATCAAATTGAGC
TTGAACTTAATTTTCAAAATCTTATGGATGAAAAGTCAGCAGGACAGGATGAACCTAAGGAA
GAAAGTAATTTAGGGGGAGAAGATTATGCAGGAGGAAAAGAGTTTACAGCAGAATTTACAGC
TTACTGTCCTAGAAAAGAAGAAGGTGGAGATACAGATTGTAGAAAGAAAAAACTTGACCCAT

CTAAAAAAACTTGCGCTGCTCCTATGGTTGGTAAATATGAGCAAACTTATTATACAAAAGAG
TTTTTAAATAAACATCCTTTATTAAACTATGGAGATGAAATACAGGTAATTACAGGAGTTTC
TGGTCGTGATGGAGTCTATAAAGTAAATGACGTAGGACCTGCAATAACTATAGAAAAGAATG
GAACATACCATATAGATATTTTATTTGGAAATGTTGAAGAAGCTAGTAAATTTGGAAGAAGA
AAAGGAAAAATTATTATTGGTGGTTATTCTGGTAATGTATCTGATAAAGCTAAAATAGTAAT
ATCAGAAGCAAAAAAACATCTAGGTAAACCTTATAAATGGGGTGGAAATGGACCAAGTAGTT
TTGACTGTTCTGGTTTAATGGTCTATTGTTTTAAAAAAGTTAATGTTAGTTTGCCAAGAACG
TCAAATCAACAATCTAAAAAAGGCAAGAAAGTAGAACAAAAAAATCTTCAAGCAGGAGATTT
AGTATTTTTTCATAATCCAGTCAGCCATGTTGGATTATATATAGGTAATGGAGAATTTTTAC
ATGCTCCACAAAAAGGTGATGTAGTTAAAATAAGTAAGTTAAGTAGTAGAAGAGATTTTAAC
ACAGCTAGGAGAGTATTATAAAAGGATGGTGATATAATGGCTAATCCAATAAATGAATTTAT
AGGAATAATAAGAGAAGAAGGAAAGTATCATAATCAACCTTCTTTTTTTATTGGAAAAATTA
AAAGTAAATTACCAGATTTAAAAATAGAGACAAATAACATCATATTAGAAAAAGAAGATATT
TTGATAGATAGTTGGATGATTGATAGACAGCTAGAAACATTTGACACAGAAACAAATCAAGA
ACACCAGCATGAAGTAAAAAATCCTTTTATAGATAACTTTGAATCTGGGGATATGGTAATAA
TGTTTAGAATAGGCGAAAAATTTGCTGTTGTAAGTAAGTTGGTGAGCTTATAATGAGTACAA
TATTTCCTTTTATAGGTGTCCCAGAGGATTATATCTTACCTAAAACAGAAGAATTGCCAATC
TTTCGTGAAGTGGCATGGGATTTTGAAAAAGATGAACCTATTTTAGAAAAAGGTGACTTTAA
AATAATTGAAAAAAAAGAAGCCTTAAAAGTTTGGATATACAAGTGTATAAAGACAAATAGAT
ATGAACATGAGATATACTCTTTAGAATATGGGACAGAGCTTTCAGAACTAATAGGACAAAAA
TATACAAAAGGTCTTACAGAAAGTGAAGCTAGTAGATTCATAAAAGAGGCCCTTCTAATAAA
TCCATATATATTAGAAGTAAACGTAAAAAGTGCTAACTTTAACAGAGACATATTGAGTGCAA
ATGTAAAAGTATCCACTATCTATGGGGAGGTGGAAATAAATGTATAGTGACCAGACATATGA
AGTAATAAAAAATAGAACTCTTGAAAATATTAATCTTGATATTTATAAAGGAGAAGGTTCTT
TTCTAAACAACATGGTATCTGGAAATAATCTAGAACTTTCGAAGATATATCTAGAACTTTCA
AAGATGCATAAAATGGCTTTTATACAAGACACATATAACCAGTTTCTTGATAAAAGAGTCAA
TGAATTTGGTGTATATAGAAAGTTAGGTACAGAGTCAAATGGAGAAGTTGAATTTATTGGAG
AGAAAGGTACTGTAATAAATAATGGCACAATAATATCATATAGAGATTTACTATTTGTAGTA
ATAAAAGATGTAACTATTGGTAGTGAAGAAGGTGACAATAGCCCAGTTCAAGCTCTGGAAGT
TGGTAAGAAATATAATTTACCTACAAATTGTGAATTTAAACTAGTTGATAATATATCTGGAG
TAACAAAGATTACTAACACAAGAAGTTTTGAAGGTGGTACAGATATAGAGACAGATGAAGAA
CTAAAAGAAAGATTTTATAAAATCCAAAGAAATCAAGCTACAAGTGGAAATAAAGCTCACTA
TGAAGAATGGGCTTTGGAAGTAGATGGAGTCTATAATGTTAAGGTTTATCCAAGATGGGATG
GTCCGGGAACAGTTAAGGTCTTGATATTTGGGAAAAATAATCAAGCTGTTGATACAGAAACA
ATTGAAAGGTGTCAGCAACATATAGATGAAGAGAAGCCTATTGGACCAACTATAACAGTTGT
GACACCATTACCAATAGAAATAAGTATAAGTGCAGTAATGAAACTAGAAGATGGATATACAT
TAGACAATGTAAAAGAATCTTTCCTAGAAAGTATAAATACATACTTTAGAGATATTAGAGGA
GAGATAATCTATACAAAAGTAATGGGAATACTTATAAATACTACTGGTGTACACGATTTAAG
TAACCTACTTATAAATGGAAGTACAGATAATATAACTATTAATGAAGATAAAATACCTAGTG
TAACAACTGTTAATTTTAGTGAGGTGGAAAATCAATGAAGCTAATTGATAAACTACCATCAT
TTGATAGAAATTACATTGTAGAGGAGATACAAGGTGCATACGATACAGAATTAAATATTCTT
AAAGAAGATATTGATGATACCTTTAACCAATTATTTGTTGACACTGCAACATGGGGATTAGA
TATGTGGGAAGACATACTCTGCATTGAAAAAAAAGAACTTGATTTTGACACAAGACGTAGCA
ATATAAAAGCTAAAATGAGAAGCAGAGGTACTAGTACTATTGAAGTTATAAAAAGTATATGT
GAGGCATATACAAAATCAGAAACAGATATAAAAGTTTATAGTGATGAATTTACATTCGTATT
GAGTTTTATAGCAAATAACTGTGACTATAAAACTCTTTTAGATTGTAGCGATATGATTGAAA
GAGTAAAACCTGCTCACTTATTACACTATTTAGAACCAATAATACTAGATAAAAGTATGGTC
TATTGTGGTGGAGGTATGGTATGTAGTGAAGAGGTAAAAGTTCATCCATACTTTGAACCAAT
TATAAAATGTAGTGCTGTTGTAAACTGTGGAGCTGGAATGATAAGTAGAGAAGAAATAAAGG
TTTATCCTTTAAGCATTAAATGCATTGAAAATAATTGTAAGATTAATATAGCTATTGCAAAT
GATACAGGTGTAGAAAATGTAGTAGTTTATCCTAAATCGGAGGTGGTATAATTGGAAGAAAA

ATTTTATATAATATTAACCAAAATTGGTAGAGAAAAAATAGCAAATGCAACTGCACTAGGAG
AGCTTGTTGGATTAACCAAGTTTCAAGTTGGAGATAGTAATGGAGAATATTATGAGCCAACA
GAGGAACAAACTGCTTTAAAGAATGTAGTTTGGGAAGGAAATATAAATTCTCTAAGAATTGA
TGAAAAAAATCCTAATTGGATAGTTATAGAGACTATTTTACCAGGAACAGTTGGTGGATTTA
TGATAAGAGAAGCTGCTGTTCTGGATAATGAGAATAATATAATAGCTATAGGTAAGTATCCA
GAGACGTATAAGCCACGTGCTGAAGATGGCAGTATTAAAGATTTGGTTGTAAAAATGATTTT
ACAATTGTCCAATACTTCAAATGTTACATTAGAAGTAGACCCGACGTTGGTTTTTGTAACTC
AAAAGGATATTCAAGATTTAGATGATAAGTTTGATAAAAATATAAAAGAAATAAAAGTAAAA
ATTGGCGAAGAACTCTTATCTACAGAAGCTAAAAACTTATCAGGAGCTATAAATGAGGTAGT
AGAAAAAATTAAAAATATATCTATTGATGATGTAATAGGAGGTCAAATACAAACTGAACTAT
CTGTATTAAAAAATAGTTACAATAAATTATCTGAAAAAGTATTAGATATCTTAATATACTTA
GAATTAGAGTCAGAAATAGATGTAGATGAAGCTGGATATTGGTATGATACCTTAACTAATGC
TAAAAACATAATAGCTATAGAAGGCCTTAAGTTAGATTTAAATAGAAAGTGTATAACTGGAG
AACTTGGTAGTGTTACATTTAAGAATGTGGTGCTACCATTTAATGCAAATAGAGTTAGATAT
ATACATGAAATGGATAATAACTTTGTTGAAACAAAATCTAATAGGGCATATTCAATTGGTCA
GACAGATATAACTTTAAATAAATATTCGTATGAAATAAGATAATTAGGAGGTTTTTATAATG
AAAAGAACTAAACTACTTCAAAGAGGTAATTTCTTTGGCGATAAAAATATGGTAGTTGATGA
ATTTGATGAAGGGTATGATAATTATGACTTTATTAATTTTTTTACTGGATGTTGTAACTATA
CATTTGGTCTAAAAAATAATAATATCTTGTATGGATGTGGAGATAATAGTAACTTTCAACTT
GGATTGGGAGAAGACAATACAACAAGAAAATTATTTACGAAAATACCAAATATATCTACCAA
TATTAAAAAAGTTGCATGTGGAGAATCTCATGCAGTTATACTTACTTCAGATGGAGAATTAC
TTGTCGCAGGTATAAATACAGATGGTCAAATGGGATTGGGATTAGAAAAAGTAGGGAAAACA
GTTTCTACATTTGAGAAGGTTCCAGAAATAAAAGGCGTAAAGGATATTGCATGTGGACTTCA
ATCAACATATCTTTTATACAATGATGGAACTTTATATGTTGCTGGAAATAATTTGTATGGTC
AATTAGGTCTAGGAACTAATGGAGCATCTGCAAATGTAAATACATTTACAAAAGTAGATGTT
GACAATGTAAAGGCTGTATTTTCATATAATAAATCAGCTTTTATAATAAAGAATGACAATAA
ATGCTATTCTACTGGTTTTAATAATCAAGGTCAACTAGGTTTAGGAGATAAGAATAATAGAG
ATTTATTTAGTTTAGTTTCTATTAATGATGTTAAGACTATAGCTTGTGGTTCTGAACACACT
GTGTTAATGACGTATAATAATGATATATATGGTTGTGGAAAGGAAAAATGTTTTGGAAATGC
ACTTCAATCATCACTATTTACTAAGATAGAAGAAGTAAATATAAAAACTATTGCATGTGGTC
ATGGTAACACTATGCTTATAGATAACAAAGGTACTTTAAAGGTTGCTGGAAATAATGATATA
TATCAGTTAGGTATAGCAAATTACTCTGAGAATATAGATAATTCATTTATAGATTTAAAAAA
TATTGTAGCTAAGAATATTTTCATTGGTTTATCACATAGCATACTAATTGATTCAAATAATG
ATTCATATTGTACAGGAGATAATACTTATGGACAATTAGGTTCGTTTTTTGATGATATGCAC
ATTGTAGAATTTAAGAAAATGGATAGTGAAAAATATAGTTATAGTAATTATATAAATTTAAT
TAAATCTGAGGATAAATTAACTTTATTAAAAGAAGAAATGGAAATAAAGGATATTGAACTTC
CACTAGATATACATTCTGTAAGAGATGTCGTTTTTAGTCCTTATTGTACTCTGGTTATTTTA
GGGAATGGAGATGTATATGGTCTAGGAAATAATAGATACAAAGGAATGGGTTCTGACTTACC
AAGTCAATTAAATGAGTTGACAAAATTAAGTATCTCTAATGTAAAGTCTATAGTAGCATCAA
AAAATATTTCTGGAGGAATATTCTACATTAAAAATGATGATACTTGTTATTATTCTGGACCA
AATAGTAACTCAATAGCAGGTGTTCTTCCTTCTAATTCAGATGTATTTAAGAAAATATCTAT
AGATAATGTAAAAAAAGTTGTTATAAATACTGATTTATCAAACTGGTTTTCATTAATTGTAA
CTAATAATAAGCAAATATACACTTCTGGAAAGAGTTCAAGTTATGTTAATGGACTTAGTAAT
GCATTAATAAGTCAATATACTGAGATTAGCCTTAGTAATGTAACTGATGCTTATAGTTCATA
TAATGCAACATTTATTGTAGTTGATGAAAAAAAGGTATATGCAACTGGTATAAATACAAATT
ACCTGTTAGGTTTTAGTACTTCTGATGGATCTAATGTAAATCTAGGTTTATTAAGTGATTGG
TATTATATAAATATATCAGGGTCAAGTTATAGTAGAGTTTCATGCACGAATAATATTACTAA
AATTAATAATATTATCATATATGAGTATGTAACTGTATTTTGTACAAACATTGGATCTTTTC
TAACTGGATACCATGGTACTTCATGGACAAAACCAACTGATTCAAGCTATAGAGTTCAATAT
CAGGGAATTTCATATGCAGGATATCTTGATTCTTATATATATAATTATTATCCTACAAGATG
TACACAATCATCATCTTCTACAACTTTTGCTTATTTATATAATGGGGAATCGTCAAGTAATT

TAAAAAATGTCAATCCAGATAATTTACTTATTTCTGGAGGTTCATCTTATATACATCAATAT
GGAAGGAATTATCTTAACAATCAATCATCTAATAATATTGCAGCATCTAATATAAATTCAGG
TCCTATTACCTCTGATAAAGCCATATTTTTATATAAAGCTCTATTGTATTTATCTTCTAACA
CGCTATATGGTTTTGGGAATATATCTGAAAGTGCAAAAGAACTAGATGTTTCAGATACACAA
GATGGATATAATGCCACTAATTATAAAAAGGTAATGAAAAATATAAAAAATATATTTATACC
TCCTTATGATTTAAGTAGAGATAAAACTAGATTTGCAATATTAACTGATAAGAGCTTATTTA
TATGTGGATATAACTCTAAGGGTACGCATGGTATATCAGTTAATAGTAGTTTAAATTTAAAT
AATAAGATAAATTACAATAAAAAGAATAGCAGTAGTGAAATATCTTCTAATATACAAGAAAT
ATATAGCCATTCAAAGTCTACATATTTATTAACTAATAATAATATGCTTTACAGTGTTGGTT
TAAATGATGTAGGTCAATTAGGAGTGGGAGATGAGATAAATAGAAAGGTATTTACTAAAATA
AATATTGATAATATAAAATCTATAAATGTAAATAGATTTACTGACAATAGTAAACATGCATT
TGCGATAAAAAATGATAATACCTGTTATGCTGTTGGTTTAAATAATTCTGGTCAGTTAGGAA
TAGGAGATAATGTAAATAGAAATATATTTACTAAAATAAATGTTGAAAATGTAAAATATGTA
GCTGTATATGGAAACACATCTCTATTATTAACTAATGATGGTCTTTTATATGGAGCAGGTAA
TAATGGAAAAGGACAGTTAGGATTGGGTGATACTACAAGTAGGAATATATTTACACGTATAC
CTATAAATGGTGTTAGAGATGTATATCTATGTAATGATGTATCAATCATTGTTAAAAATGAT
AATACATGCTATGTATGTGGACTTGTAAATGGCTATTTTGGGTTTACTGAAGGAAGTATAAG
TACATTTACAAAAATAAATATTGAGAATGTAAAATCTGTTGTGACAGCAGGAAGTGAAGCTA
CATTTTTTATAACAAATGACAATATGATTTATACTACAGGGAAAAAAGAGAGGGTATTCTTT
TCAACAGAGACTAATGATATAAAGGGGATACGAGTAATTAATAATATTATAAATGCAAAAAA
AATAGTAGTTAATGGATATACTTCAGCCATTTTAACAAATGACAATAAACTATTTGTTGGAG
GTCTTAGTGGATATGGAAGTATAGCAAATAATAATAATACAAATAGTGTGGAAGATGTTAAA
GATGTTTTTGTAACAGCTAATAATACACTTTATATAGATAATAATAACAATTTGATATCATC
AGGTAGAGATACTTATGGTATATCTGATGAATCTTATAGGGATATGTCAGTTCCATATTATA
AAGTATCTATAAAGAAAGATGTTGATACTGTATTTTCTAGTTACAATACTATATTTATTAAA
GATATATATGGAAAATTTTATTCTTCAACAAGAGATAATAGATATAATCATTTAGGTATTCA
CCATAGATATGATAATGATAAAAATGAAGCTCTTGAAGGTTCCCTACATTCATATTTTAAAA
CAGATAACACATCAGATAAAATAGTTTTTAATAAGAAAAATGAAAAGCTAGTAATGTTTAAT
GATAAGTATATAAAAACAAATAATAAGTATATAAATTATAAAAACATATTTAAAGATAATTT
TAAGTATACTTCAATAATATTGCCATTTGAGGTATCTGATATTGATATATCAAAAACACATT
CATTGGCTGTTGCTAAGGATGGCAAGTTATATGGAATAGGAAGTAATTCATATAAAGAAATT
AATCAAACCCTTGAAGATATAGAATTATTAACTCTTACTGAAGTAAATATATCAGATGTCAA
AAAAGTTGCTTGTGGAGATAACTACTCCTATATTATTAAGACAGATAATACTCTATGGTCAT
ATGGAAAGAATACTGAGTACCAATTGGGAGTTGGCCACAATAATGATGTAAGAGAGTTACAA
AAGGTTACTGGATTACCTTCTGTTAAAGATATAAGTATATATAACTCAATGACACTTGTTTT
AACTAATGAGGGAGAGTTGTACGCTCAAGGGTACAATACAAATGGATTATTTGGACTAGGAG
AAAGTGAAAAAGATAAGATAATAAGAACTTTTACTAAAGTATTAACTAATGTTAAAGAAATT
AAGTCACATAATGATGACCACATACTAGTAATTAAAAATGATAATAGTCTATGGATAACTGG
TAAAAATAAATCTATGTATAAAATATCTATATCAATTACTGATTTATATGAATTTACTAAAA
TACCAATTCCTGAACATCTAAATGATATTTTAGATATAGAGCTTTCAGATGATACAATATAC
ATGATAACAAAAGTAGATACAAGTAAAGCATCTATAGAAATAGTTGAAAAATCAATATCTCA
AGTGAGAGTTGTAGTACAAGACCCTAATAATGTTATAGAAAAACTTGAAATGTTTATAAATG
ATGAATTAATATCTACTAAGACTAATTTGGAAATAAATAGCATTATATTTGAGATACCACAA
AATAAAATAGTATTAGGAGAAAATAAGATACTGATTAAAGCCAGTAGTCCTACAGGCGATTT
ATATTCAAGTATGTTTATATTTAAATCAGAAACAGGGCTTAAAGTAAAAAAGGATTCTATTT
TAATGATAAACAATAAAGTATATTCAATCATAAACATTACTGAAAATAACACTGACTTAATA
GTAACATTAAATGAGGGATTAAAGGATGATATGATGGAAAACAATCCTATATATCAATTAAT
AAATAAAACTAAAGTTCAAGTAAAAATAAATAAATCTGACTTATTCAAAGACATGAAACTAG
TTGAAATCAAAAAATCAGACTCAAGTTACCAAGAAATCTATGAATTAGAAGAAGCCAACATA
AAAAGTGCTCAGCCTAAAATCATAGTAGAAAAAGGAGATAAATGGACAGCTATAAAACGTCC
ATCTATGATTTTTAGATATGATGCTGAAAACAACGAGCCACAAGCTTAAAATGGAGGTGTGA

AAATTGTTTAAATTCGATAAAAATAAAATAGAACAAATCAAACAAGGTAGAAAAGTAGAAAT
GCAGTATAAAGACATTTCAGACATAAGTATAGGTCAAGTAAAGCAAGATGATGATATAACAA
ATAATTTTATAGCAAATGTAGAAATATATGAGATGTTGTTAAATCAAAGTTCTGTCAATGAA
GCAAGTAATATAAGCACTTTTAGTGTAAGAAAATCTGGAGGTGAGAGTGGAATGGTAGAAGT
ATATGTAGCTTTAATTTTAAGAGGCAAAAAAACAATAGAAGAAGTACCAGCAGTAATTAGAG
AGCAAGTTAGAATTAGATGTAAAGAATTAGAAATACCAGTTGAATAGTAAATTTAGAATAAC
TATGTATTAGTTATTTTTTTTATGTAAAGTACAAGGTCTTAACTTTAATAAGTAAGCCTTGT
ACTTATTTTTTGTTACATTAGAACTTGTATATATATTTATTATTTATTCAATCTATAAATTA
CACCTACAATTTAAAGTACAGAAGATTAAATTGATAATCCTGAAAATATAATATTGCATGAT
GTAAGAATACAACAAAAATTAAAGCTATAAGTATAAAAAATTTAGACAATAGGAGGCTATAA
TGGATAAATTAATAACCGAATTGAGTAGTCTAGGGGCAATAGGTATACTATGTGCTCTATTA
TTTAAAAATACTATGCAGGAGAAAAAAGAAGATAGAGACATGTATAAAAAAACTGTAGAAAA
TTTTATAGAATTATCTACACAACAACAAGAAATAAACAAAAATATACTTGTTCAAATGGGTA
TAATGAAAACAGATGTAGAGGAAATTAAGGAAGATGTTACTGATATAAAAGGTATGTTACAA
AATGGTGTATAACATGAAAGAGTAGCACCAGATTATATATTGTTAGGAAAAGATAAAGTAGT
ATTGTAGATAGTTCACTATTTTATTGAGAAGGATTTAATATTTAAAATATTAATTAAAAAAA
GTAATAAAAATAACATATAAAAATTAAAAAAGGAGTTAAGCTTAAATTTGAGGCGCG

SEQ ID NO:62 >CD4-1359
MNNLDKLFELASQEEIIIHYTTYIAGDLEGLYINKHGIKIISLLSNLKQNSKKLTSILAEEL
GHHFTSLGYYVSSYNDYYTKIIIDKCENKALKWACEFLITEEDIINIINSGITCVYEMADIL
NVDITFFQKRLEFLSLKKQSLQLGNNKYLILTNLPYFYIFDPIS

SEQ ID NO:63 >CD4-1360
MFAKRLRELRKEFGLTQRELGEKVGVSQRVLGYYETENRFPDEHILNKLADVFNVSVDYLLG
RTLVKENIDTVAAHRKNPHEELPEEAQEQLNDYIEFLLNKYKKK

SEQ ID NO:64 >CD4-1360A
MFKNNLKYYRKCKGMTQIQLARKAGITNDYISQIERGIKNPGLLMAKKISSILEQNIEEVFF
IQL

SEQ ID NO:65 >CD4-1361
MENKKDILFKETDKRLHNYKYLDIKIKNINLDIKRCENEYSGCGAMVYTEKTSNTYNISSSV
ENEVLKREERLRKLKMEKEDIEIEKEKIENALTCLNDIEMEFFNLFYNSKTKNNMTYISMKL
HLDRTSCYNLKKKMIFKLSEIL

SEQ ID NO:66 >CD4-1362
LLKYKEILETIIEILKKNFTESIFIDDESVQGSEGSCFFVSILSVICTPVMLNTNNKDIVIS
IKYLPKPQSKSIRMYEISDELNKLFNRNIKVTDRKLNITKLEQSIKKEESIYVLNFTFTLNY
LDSVYEEDVVYENMKEINLNLGE

SEQ ID NO:67 >CD4-1363
MAIGLPSINISFKELATTVKERSARGIIAMVLKDAKALGLNEIHEKEDIPVDLSAENKEYIN
LALMGNVNTPNKLLVYVIEGEADIQTALDFLETKEFNYLCMPKAVEADKTAIKNWIIKLRDI
DKVKVKAVLGKVVGNHEGIINFTTEDVLVGEKKYSVDEFTSRVAGLIAGTPLSQSVTYTKLS
DVVDIPKMTKVDAESRVNKGELILIKEAGAIRIARGVNSLTELTAEKGEMFQKIKIVDTLDI
IHSDIRKVIIDDYIGKVTNSYDNKCLLIVAIKSYLEELEKSALIESDSTVEIDFEAQKSYLK
SKGVDLSYMTLQEIKEANTGSKVFLKAKIKVLDAMEDIDLSIEI

SEQ ID NO:68 >CD4-1364
MANMEARNVMSGTWGELWLDGNKVAEVKKFQAKMEFTKEDIIIAGQMGTDTKYMGYKGKGSI
TLYHVSSRMHKLIGEKIKRGSEPRFVAISKLNDPDSYGAERIAVKNIAFDDLTLADWEVGVK
GEIEAPFTFTEYDFLDII

SEQ ID NO:69 >CD4-1365
MSENGLSKNINIVDLLLNSDTENLERPSTIVELKRLSTIFGQEFKVMCRALTISKDEEIQNT
CLKIDENMKTDIDLPEMQMLTIIEGVCDLDGKLLFKNKELMDKFKAPTPKELARKLLLPGEI
TNLYRILQDVMGYGKNAVIEEVKN

SEQ ID NO:70 >CD4-1365A
MYYYWKKKGIRPSLFYAMDKGELKLIEAFFALEIEEEVEKMKHGYGVCPLTGGGM

SEQ ID NO:71 >CD4-1366
MGNVREEGINMYLTDNYTPKMNQIISVTDNFRRATVAVSLSTNVMASSIKNSIGSASNRVNS
LNSSLRKVQTTASSVSSTMTKLSSSINAVSGVIGSLNGSIMRLAITIAMIIDYFNKLIQKKN
EFNSNIMIILIFKAKSDEVEKTKNKLLGNLKKIGGKIWNIVIKAKDMTKRVISSILGKLKQV
EKRPYQGSINLKDMVSSAMGRILPKLMLFKNTFWSGVIAIKDMASGIISKVFPKLRLFAGKV
WSGAIAVKDMASGILGSIKGKISDLTNGATIGVAVKKGVDLLGQEQNQKVVLESVMKRNTGK
VNQIDVDDYYGSLVRMANDTPFDPEDVVAMGTKAKMISNITGGKKEKDITQAMVDVRALNMN
TSSEQDVSAAFLSAAKGNMESLNTLVGENYKTFDEALEGISVKQMGLAKEMSNTIPGIISGA
QTSINNGLKSIVKPFDDILGQGLKKIKTFIESGLGNLAGLSEKMAGKIGNVMNGKIIIGNKY
DQMQSRSVKNGKEFSDSTQYRISNEAEKRKMMVENKQERFENHAATMIGNAPKAIVNAGSTL
LQNIDFTALIDSLLPVVNLVNNLLDSINNKSPIAQGLISIFGTIVTTAFQLIGPVVEAVSPI
ITRIFTFLGEYAPQINNFIETLGVIWKTVWETLGPLLETGWKIIEPILGAFFNILDKVCKIV
KDICKWWQTMINKIKNGSITGTVLNLVEKSKKNYKDNPYAGTKAGDSGKAYSGKKGNNAFGL
NYVPYNDYQTRLHEGEMVLTKQEANQYRSRKNGGNINIAKLADTIVIREEADIEKITSKLVA
SIQLAQLGGVL

SEQ ID NO:72 >CD4-1367
MEMWLRQAEDRFRFPVFPSSFSINGKAAVNSSSILKIGEIATFGGVALKSISISSFFPNKDY
TFCDYTGFPSPYDCVNKIEKWMKEGFILRFTITETNINMEVIIEGFSYEERDGTRDVYFTLD
LKEYKRIKIPKVTPKQ

SEQ ID NO:73 >CD4-1368
MIINRSKDSSSNEISFVSKDMGFLLTQSEVSYNFKDKLVEDIAKQVFAENRLSVGIIAKTNV
KYTKMFIGVNGYDTIMSAYTEASKKTKKKYMIEANLDKFNVIEKGTVTLSVMFEEGFNIINT
TFSESMENVKNKVIVVDQYGSKISEKIDNEIFKEVNVIMQKVIQQQENQDVDIDSEFNGIEK
SCSLKGYGDVSCITGRGVKVKDSYTKLVGLFYIDTDKHTWQNGEYQIELELNFQNLMDEKSA
GQDEPKEESNLGGEDYAGGKEFTAEFTAYCPRKEEGGDTDCRKKKLDPSKKTCAAPMVGKYE
QTYYTKEFLNKHPLLNYGDEIQVITGVSGRDGVYKVNDVGPAITIEKNGTYHIDILFGNVEE
ASKFGRRKGKIIIGGYSGNVSDKAKIVISEAKKHLGKPYKWGGNGPSSFDCSGLMVYCFKKV
NVSLPRTSNQQSKKGKKVEQKNLQAGDLVFFHNPVSHVGLYIGNGEFLHAPQKGDVVKISKL
SSRRDFNTARRVL

SEQ ID NO:74 >CD4-1369
MANPINEFIGIIREEGKYHNQPSFFIGKIKSKLPDLKIETNNIILEKEDILIDSWMIDRQLE
TFDTETNQEHQHEVKNPFIDNFESGDMVIMFRIGEKFAVVSKLVSL

SEQ ID NO:75 >CD4-1370
MSTIFPFIGVPEDYILPKTEELPIFREVAWDFEKDEPILEKGDFKIIEKKEALKVWIYKCIK
TNRYEHEIYSLEYGTELSELIGQKYTKGLTESEASRFIKEALLINPYILEVNVKSANFNRDI
LSANVKVSTIYGEVEINV

SEQ ID NO:76 >CD4-1371
MYSDQTYEVIKNRTLENINLDIYKGEGSFLNNMVSGNNLELSKIYLELSKMHKMAFIQDTYN
QFLDKRVNEFGVYRKLGTESNGEVEFIGEKGTVINNGTIISYRDLLFVVIKDVTIGSEEGDN
SPVQALEVGKKYNLPTNCEFKLVDNISGVTKITNTRSFEGGTDIETDEELKERFYKIQRNQA
TSGNKAHYEEWALEVDGVYNVKVYPRWDGPGTVKVLIFGKNNQAVDTETIERCQQHIDEEKP
IGPTITVVTPLPIEISISAVMKLEDGYTLDNVKESFLESINTYFRDIRGEIIYTKVMGILIN
TTGVHDLSNLLINGSTDNITINEDKIPSVTTVNFSEVENQ

SEQ ID NO:77 >CD4-1372
MKLIDKLPSFDRNYIVEEIQGAYDTELNILKEDIDDTFNQLFVDTATWGLDMWEDILCIEKK
ELDFDTRRSNIKAKMRSRGTSTIEVIKSICEAYTKSETDIKVYSDEFTFVLSFIANNCDYKT
LLDCSDMIERVKPAHLLHYLEPIILDKSMVYCGGGMVCSEEVKVHPYFEPIIKCSAVVNCGA
GMISREEIKVYPLSIKCIENNCKINIAIANDTGVENVVVYPKSEVV

SEQ ID NO:78 >CD4-1373
LEEKFYIILTKIGREKIANATALGELVGLTKFQVGDSNGEYYEPTEEQTALKNVVWEGNINS
LRIDEKNPNWIVIETILPGTVGGFMIREAAVLDNENNIIAIGKYPETYKPRAEDGSIKDLVV
KMILQLSNTSNVTLEVDPTLVFVTQKDIQDLDDKFDKNIKEIKVKIGEELLSTEAKNLSGAI
NEVVEKIKNISIDDVIGGQIQTELSVLKNSYNKLSEKVLDILIYLELESEIDVDEAGYWYDT
LTNAKNIIAIEGLKLDLNRKCITGELGSVTFKNVVLPFNANRVRYIHEMDNNFVETKSNRAY
SIGQTDITLNKYSYEIR

SEQ ID NO:79 >CD4-1375
MQYKDISDISIGQVKQDDDITNNFIANVEIYEMLLNQSSVNEASNISTFSVRKSGGESGMVE
VYVALILRGKKTIEEVPAVIREQVRIRCKELEIPVE

SEQ ID NO:80 >CD4-1376
MDKLITELSSLGAIGILCALLFKNTMQEKKEDRDMYKKTVENFIELSTQQQEINKNILVQMG
IMKTDVEEIKEDVTDIKGMLQNGV

Figure 2.

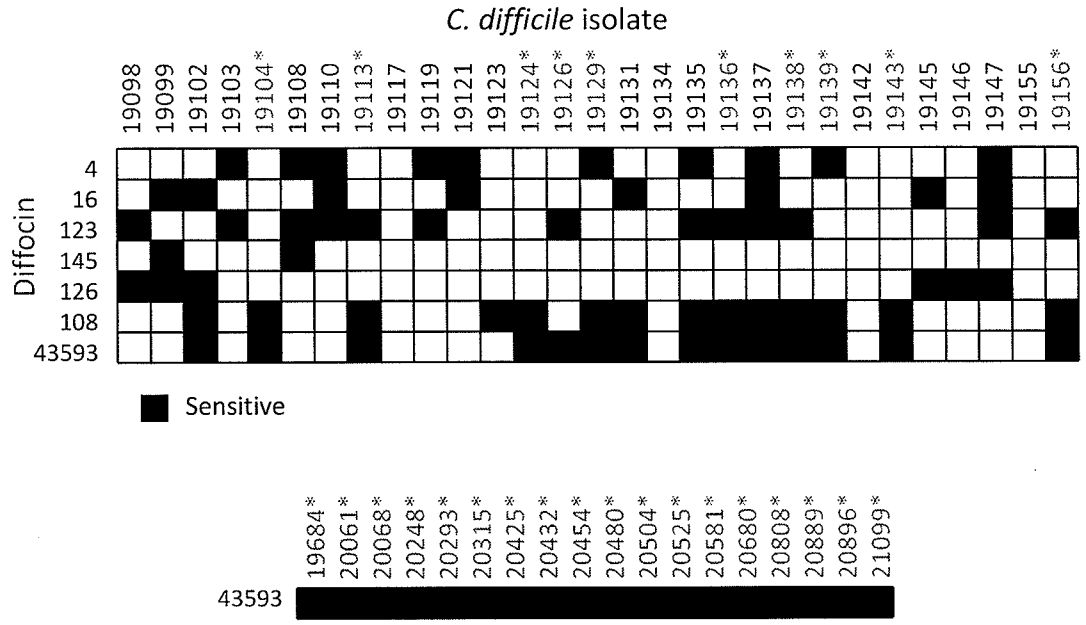
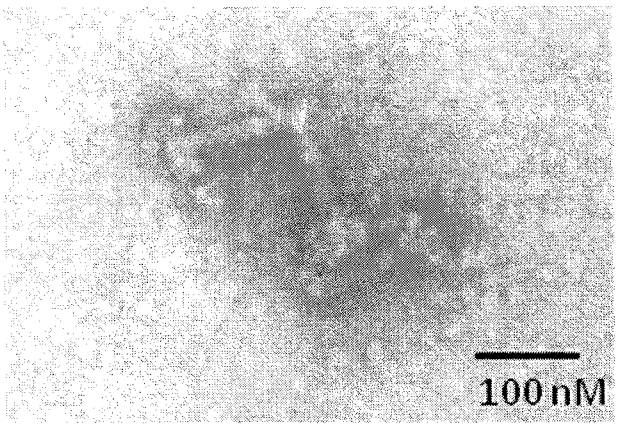


Figure 3.



Diffocin 16

Figure 4.

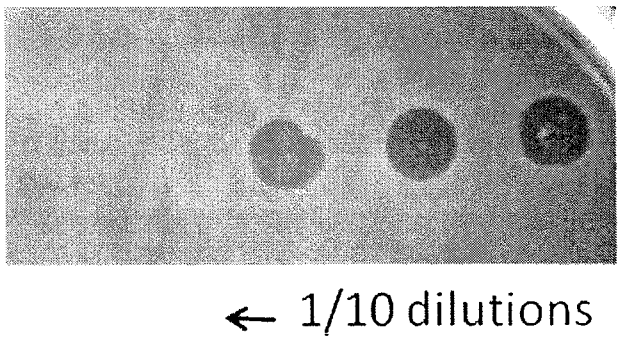


Figure 5.

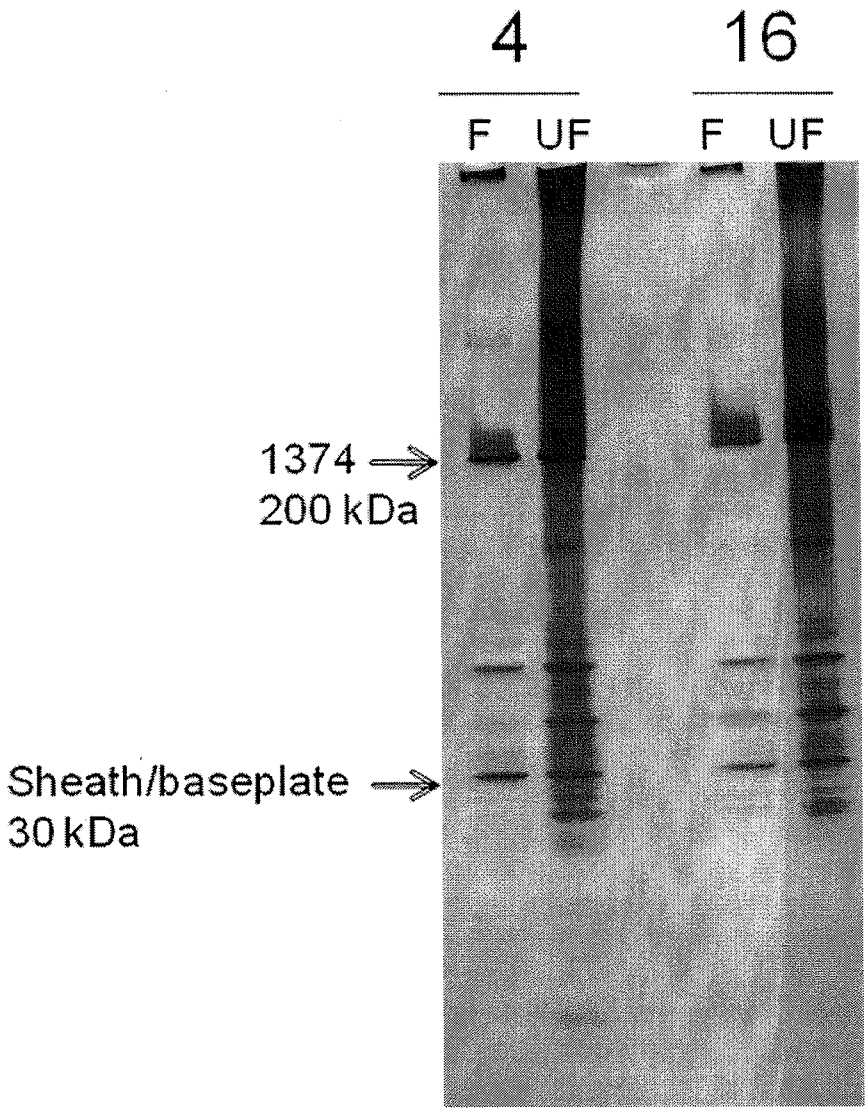


Figure 6.

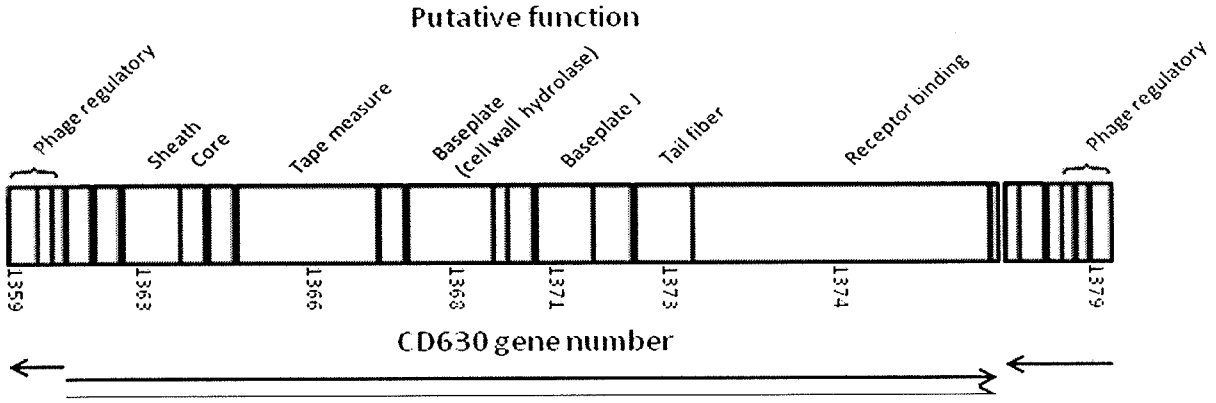


Figure 7.

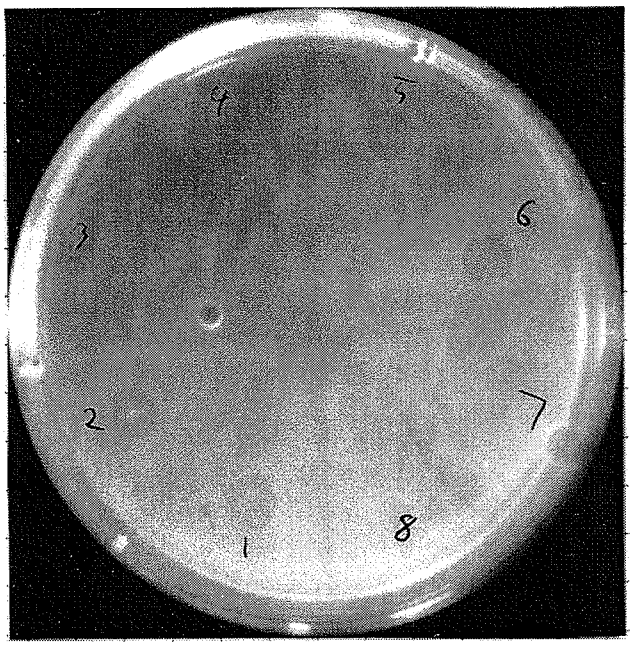


Figure 8.

```
CD108           MKQNKLLQRGAYFNDKNILIDDFDKRYNDYDFVEFFTGISNSTFGLKSDGNLYACGDNTG 60
43593           MKQNKLLQRGAYFNDKNILIDDFDKRYNDYDFVEFFTGISNSTFGLKSDGNLYACGNNTG 60
CD16            MKQNKLLQRGAYFNDKNILIDDFDKRYNDYDFVEFFTGISNSTFGLKSDGNLYACGDNTG 60
CD126           MKQNKLLQRGAYFNDKNILIDDFDKRYNDYDFVEFFTGISNSTFGLKSDGNLYACGDNTG 60
CD4             MKRTKLLQRGNFFGDKNMVVDEFDEGYDNYDFINFFTGCCNYTFGLKNNNILYGCGDNSN 60
CD123           MKRTKLLQRGNFFGDKNMVVDEFDEGYDNYDFINFFTGCCNYTFGLKNNNILYGCGDNSN 60
                **:.****** :*.***:::*:**: *::***::**** .* *****.:. **.**:*:.

CD108           FQLGLGKDSSERRMFSKVK--IDNVKYVSCGSKHSVAVTKDGFAYGAGTSNVGQLGVIES 118
43593           FPLGLGKDSSERRMFSKVK--IDNVKYVSCGSKHSVAVTKDGFAYGAGTSNVGQLGVIES 118
CD16            FQLGLGKDSSERRMFSKVK--IDNVKYVSCGSKHSVAVTKDGFAYGAGTSNVGQLGVIES 118
CD126           FPLGLGKDSSERRMFSKVK--IDNVKYVSCGSKHSVAVTKDGFAYGAGTSNVGQLGVIES 118
CD4             FQLGLGEDNTTRKLFTKIPNISTNIKKVACGESHAVILTSDGELLVAGINTDGQMGLGLE 120
CD123           FQLGLGEDNTTRKLFTKIPNISTNIKKVACGESHAVILTSDGELLVAGINTDGQMGLGLE 120
                * ****:*.: *::*:*:     *:* *:**..*:* :*.**    ** .. **:*:  .

CD108           TV---YYEFTKLP-IDDVKTVACGYDFTFVLKNDGTLYSAGLNSSGQLGLG---DTNNRV 171
43593           TV---YYEFTKLP-IDDVKTVACGYDFTFVLKNDGTLYSAGLNSSGQLGLG---DTNNRA 171
CD16            TV---YYEFTKLP-IDDVKTVACGYDFTFVLKNDGTLYSAGLNSSGQLGLG---DTNNRA 171
CD126           TV---YYEFTKLP-IDDVKTVACGYDFTFVLKNDGTLYSAGLNSSGQLGLG---DTNNRA 171
CD4             KVGKTVSTFEKVPEIKGVKDIACGLQSTYLLYNDGTLYVAGNNLYGQLGLGTNGASANVN 180
CD123           KVGKTVSTFEKVPEIKGVKDIACGLQSTYLLYNDGTLYVAGNNLYGQLGLGTNGASANVN 180
                .*      * *:* *..** :*** : *::* ****** ** *  ******    : *

CD108           TFTKVNIDSVKDVVTYNQSVFIIKMDGTAHACGLNSNGQLGINSTLNKSVFNKIEGMDNV 231
43593           TFTKVNIDSVKDVVTYNQSVFIIKMDGTAHACGLNSNGQLGINSTLNKSVFNKIEGMDNV 231
CD16            TFTKVNIDSVKDVVTYNQSVFIIKMDGTAHACGLNSNGQLGINSTLNKSVFNKIEGMDNV 231
CD126           TFTKVNIDSVKDVVTYNQSVFIIKMDGTAHACGLNSNGQLGINSTLNKSVFNKIEGMDNV 231
CD4             TFTKVDVDNVKAVFSYNKSAFIIKNDNKCYSTGFNNQGQLGLGDKNNRDLFSLVS-INDV 239
CD123           TFTKVDVDNVKAVFSYNKSAFIIKNDNKCYSTGFNNQGQLGLGDKNNRDLFSLVS-INDV 239
                *****::*.** *.:**:*.**** *...:: *:*.:****:... *:.:*. :. :::*

CD108           KQIACGSSHTILIKNDGTMYTTGSNGYGQLGTGNNNNSIVFTLSSINNVKYASCGNNHTM 291
43593           KQIACGSSHTILIKNDGTMYTTGYNGVGQLGTGNNNNSIVFTLSSINNVKYASCGNNHTM 291
CD16            KQIACGSSHTILIKNDGTMYTTGYNGVGQLGTGNNNNSIVFTLSSINNVKYASCGNNHTM 291
CD126           KQIACGSSHTILIKNDGTMYTTGYNGVGQLGTGNNNNSIVFTLSSINNVKYASCGNNHTM 291
CD4             KTIACGSEHTVLMTYNNDIYGCGK----EKCFGNALQSSLFTKIEEVNIKTIACGHGNTM 295
CD123           KTIACGSEHTVLMTYNNDIYGCGK----EKCFGNALQSSLFTKIEEVNIKTIACGHGNTM 295
                * *****.**:*:. :. :*  *     :   **  :* :**  .  *:*  :**:.:**

CD108           ILKYDNTLFSTGQNNYGQLANANKDVASRNTFVKVNVENIKDIKCGSQFNFLINGSKEIF 351
43593           ILKYDNTLFSTGQNTYGQLANANKDVASRNTFAKVNVENIKDIKCGSQFNFLINGSKEIF 351
CD16            ILKYDNTLFSTGQNNYGQLANANKDVASRNTFAKVNVENIKDIKCGSQFNFLINGSKEIF 351
CD126           ILKYDNTLFSTGQNNYGQLANANKDVASRNTFAKVNVENIKDIKCGSQFNFLINGSKEIF 351
CD4             LIDNKGTLKVAGNNDIYQLGIANYSENIDNSFIDLKNIVAKNIFIGLSHSILIDSNNDSY 355
CD123           LIDNKGTLKVAGNNDIYQLGIANYSENIDNSFIDLKNIVAKNIFIGLSHSILIDSNNDSY 355
                ::. ..**  :*:*   **. ** .    *:* .::    *:*  * ...:**:..:: :

CD108           VSGCNLAGQLGSFFHTTFLYEFSKVQSS--NLDNYSGLLVNDDYLYVTKDNSEFLNVKLS 409
43593           VSGCNLAGQLGSFFHTTFLYEFSKVQSS--NLDNYSGLLVNDDYLYVTKDNSEFLNVKLS 409
CD16            VSGCNLAGQLGSFFHTTFLYEFSNVQSS--NLDNYSGLLVNDDYLYVTKDNSEFLNVKLS 409
CD126           VSGCNLAGQLGSFFHTTFLYEFSNVQSS--NLDNYSGLLVNDDYLYVTKDNSEFLNVKLS 409
CD4             CTGDNTYGQLGSFFDDMHIVEFKKMDSEKYSYSNYINLIKSEDKLTLLKEEMEIKDIELP 415
CD123           CTGDNTYGQLGSFFDDMHIVEFKKMDSEKYSYSNYINLIKSEDKLTLLKEEMEIKDIELP 415
                 :* *  *******.  .: **.:::*.  . .** .*: .:* * : *:: *: :::*.

CD108           DNFQDYKKIELTDNNMFIVMNDGTLYACGLNNYGQLGLGDTVNRSVMTKVDIDNVLDIKG 469
43593           DNFQDYKKIELTDNNMFIVMNDGTLYACGLNNYGQLGLGDTVNRSVMTKVDIDNVLDIKG 469
CD16            DNFQDYKKIELTDSNMFIVMNDGTLYACGLNNYGQLGLGDTVNRSVMTKVDIDNVLDIKG 469
CD126           DNFQDYKKIELTDSNMFIVMNDGTLYACGLNNYGQLGLGDTVNRSVMTKVDIDNVLDIKG 469
CD4             LDIHSVRDVVFSPYCTLVILGNGDVYGLGNNRYKGMGSDLPSQLNELTKLSISNVKSIVA 475
CD123           LDIHSVRDVVFSPYCTLVILGNGDVYGLGNNRYKGMGSDLPSQLNELTKLSISNVKSIVA 475
                 :::. :.: ::    ::::.:* :*. * *.*  :* . . : . :**:.*.** .* .
```

```
CD108           NGNST---FVLKNNGTLYSCGYNSSGILGLKDNTNRNIFTKIEIENIKEFCVESN----Y 522
43593           NGNST---FVLKNNGTLYSCGYNSSGILGLKDNTNRNIFTKIEIENVKAFCVESN----Y 522
CD16            NGNST---FVLKNNGTLYSCGLNSNGQLGLRDEVNRNIFTKIEIENVKDFCVGSN----Y 522
CD126           NGNST---FVLKNNGTLYSCGLNSNGQLGLRDEVNRNIFTKIEIENVKDFCVGSN----Y 522
CD4             SKNISGGIFYIKNDDTCYYSGPNSNSIAGVLP-SNSDVFKKISIDNVKKVVINTDLSNWF 534
CD123           SKNISGGIFYIKNDDTCYYSGPNSNSIAGVLP-SNSDVFKKISIDNVKKVVINTDLSNWF 534
                . * :   * :**:.* * .* **..  *:    * ::*.**.*:*:* . : ::    :

CD108           IVALNHSKELYGWGNQS--YIVYGDNRNYPYKDTRVSNVEKIATWSDTLYILDSTGATKT 580
43593           IVVLNHSKELYGWGNES--YIVYGNSRNYPYKDTRVSNVEKIATWSDTLYILDSTGATKT 580
CD16            VIALNHSKEVYGWGNNP--YNNIEKTSNYPYKQG-ISNIEKIAAYDYSVYMINSEGKLYV 579
CD126           VIALNHSKEVYGWGNNP--YNNIEKTSNYPYKQG-ISNIEKIAAYDYSVYMINSEGKLYV 579
CD4             SLIVTNNKQIYTSGKSSSYVNGLSNALISQYTEISLSNVTDAYSSYNATFIVVDEKKVYA 594
CD123           SLIVTNNKQIYTSGKSSSYVNGLSNALISQYTEISLSNVTDAYSSYNATFIVVDEKKVYA 594
                 : :.:.*::*  *:..       .     *.:  :**: .  :   : ::: .     .

CD108           IGYSYNGSGGYPAPSSSSTYR----EGGYINKNTSYRTLEFYNTSKTKLVNLFAFYNGCV 636
43593           IGYSYNGSGGYPAPSSSSTYR----DGGYINKNTSYRTLEFYNTSKTKLVNLFAFYNGCV 636
CD16            SGYNYNYQLGKGNNSNQSKAL----VSQCRTNSTSSTSNGLR--TLPKITNVFPFYDGCA 633
CD126           SGYNYNYQLGKGNNSNQSKAL----VSQCRTNSTSSTSNGLR--TLPKITNVFPFYDGCA 633
CD4             TGINTNYLLGFSTSDGSNVNLGLLSDWYYINISGSSYSRVSCTNNITKINNIIIYEYVTV 654
CD123           TGINTNYLLGFSTSDGSNVNLGLLSDWYYINISGSSYSRVSCTNNITKINNIIIYEYVTV 654
                 * . *   *    ....            . . *  :      . .*: *:: :    .

CD108           FVDENGLAYCIGENNINFRGGSTTNENNSLRFINNSGVYYT--------NTDGTDYTCYQ 688
43593           FVDENGLAYCIGENNINFRGNSTTNENNSLRFINNSGVYYT--------NTDGTDYTCYQ 688
CD16            IIDEGGYVYLTG-----YHGYLRT--LNSSPSISDYSRYGT--------FIEATNSNHNT 678
CD126           IIDEGGYVYLTG-----YHGYLRT--LNSSPSISDYSRYGT--------FIEATNSNHNT 678
CD4             FCTNIG-SFLTG-----YHGTSWTKPTDSSYRVQYQGISYAGYLDSYIYNYYPTRCTQSS 708
CD123           FCTNIG-SFLTG-----YHGTSWTKPTDSSYRVQYQGISYAGYLDSYIYNYYPTRCTQSS 708
                :  : *  :  *     ::*   *   :*   :.  .   :            *  .

CD108           WTYKLIRCSIFDSPQNIIGNSKNILYLSKNNSTFKCTGNCITYGINSQNWYSYFS----- 743
43593           WTYKLIRCSIFDSPQNIIGNSKNILYLSKNNSTFKCTGNCITYGINSQNWYSYFS----- 743
CD16            Y---FIQETDFSGIEKVIGMSNNILFFKKGSS--YITGYPKTFGSTITGHRSYTS----- 728
CD126           Y---FIQETDFSGIEKVIGMSNNILFFKKGSS--YITGYPKTFGSTITGHRSYTS----- 728
CD4             SSTTFAYLYNGESSSNLKNVNPDNLLISGGSS--YIHQYGRNYLNNQSSNNIAASNINSG 766
CD123           SSTTFAYLYNGESSSNLKNVNPDNLLISGGSS--YIHQYGRNYLNNQSSNNIAASNINSG 766
                    :      .. .:: . . : * :. ..*         .:  .  .     *

CD108           -DSSNGAIALGNEFILKNYSGECLLKGYG----KATNGEFGNSTNISSISNYDTGLKDIK 798
43593           -DSSNGAIALGNEFILKNYSGECLLKGYG----KATNGEFGNSTNISSISNYDTGLKDIK 798
CD16            -INSE-SSNLGSNFIIYHSN--SKLYGKG----IANSGQFGNSTNIDGTSNYDTGLKDIK 780
CD126           -INSE-SSNLGSNFIIYHSN--SKLYGKG----IANSGQFGNSTNIDGTSNYDTGLKDIK 780
CD4             PITSDKAIFLYKALLYLSSN---TLYGFGNISESAKELDVSDTQDGYNATNYKKVMKNIK 823
CD123           PITSDKAIFLYKALLYLSSN---TLYGFGNISESAKELDVSDTQDGYNATNYKKVMKNIK 823
                  .*: :  * . ::    .    * * *     *.. :..:: :  . :**.. :*:**
```

```
CD108           DIIV------KNNTVVVVDKNNNIYVTGANQFNKLGIGEYNNQPIRKFTNITEQSNSFIF 852
43593           DIIV------KNNTVVVVDKNNNIYVTGANQFNKLGIGEYNNQPIKKFTNITEQSNSFIF 852
CD16            DIIV------KGNTVVVVDKNNNIYVTGMNQNNKLGIGEYNNEPVKKFTNITEQSNSFIF 834
CD126           DIIV------KGNTVVVVDKNNNIYVTGMNQNNKLGIGEYNNEPVKKFTNITEQSNSFIF 834
CD4             NIFIPPYDLSRDKTRFAILTDKSLFICGYNSKGTHGISVNSSLNLNNKINYNKKNSSSEI 883
CD123           NIFIPPYDLSRDKTRFAILTDKSLFICGYNSKGTHGISVNSSLNLNNKINYHKKNSSSEI 883
                :*::      :.:* ..: .::.::: * *. .. **.  ..  :.:  *  ::..*  :

CD108           MDDIKEITTSRNTMFIVKNDGTAYATGNNSSGQLGLGDTINRNKFTQINLDNIKKIST-- 910
43593           MDDIKEITTSRNTMFIVKNDGTAYATGNNSSGQLGLGDTINRNKFTQINLDNIKKIST-- 910
CD16            MDDIKEITTSRNTMFIVKNDGTAYATGNNSSGQLGLGDTINRNKFTQINLDNIKKIST-- 892
CD126           MDDIKEITTSRNTMFIVKNDGTAYATGNNSSGQLGLGDTINRNKFTQINLDNIKKIST-- 892
CD4             SSNIQEIYSHSKSTYLLTNNNMLYSVGLNDVGQLGVGDEINRKVFTKINIDNIKSINVNR 943
CD123           SSNIQEIYSHSKSTYLLTNNNMLYSVGLNDVGQLGVGDEINRKVFTKINIDNIKSINVNR 943
                 .:*:** :  :: :::.*:.  *:.* *. ****:** ***: **:**:****.*..

CD108           -SIDGNTTFAIRNDGTLYSTGLNTKGQLGLGDIVNRNTFTKVNIQNVRDVVLGTTHSHAI 969
43593           -SIDGNTTFAIRNDGTLYSTGLNTKGQLGLGDIVNRNTFTKVNIQNVRDVVLGTTHSHAI 969
CD16            -SIDGNTTFAIRNDGTLYSTGLNTKGQLGLGDIVNRNTFTKVNIQNVRDVVLGTTHSHAI 951
CD126           -SIDGNTTFAIRNDGTLYSTGLNTKGQLGLGDIVNRNTFTKVNIQNVRDVVLGTTHSHAI 951
CD4             FTDNSKHAFAIKNDNTCYAVGLNNSGQLGIGDNVNRNIFTKINVENVKYVAVYGNTSLLL 1003
CD123           FTDNSKHAFAIKNDNTCYAVGLNNSGQLGIGDNVNRNIFTKINVENVKYVAVYGNTSLLL 1003
                 : :.: :***:**.* *:.***..****:** **** ***:*::**: *.:  . *  :

CD108           KDDNTLYSCGENTHGQLGLGSESNHPDVLTFTVNNITNVRDVYCSDTTTFIVKDTNIAYC 1029
43593           KDDNTLYSCGENTHGQLGLGSESNHPDVLTFTVNNITNVRDVYCSDTTTFIVKDTNIAYC 1029
CD16            KDDNTLYSCGENTHGQLGLGSESNHPDVLTFTVNNITNVRDVYCSDTTTFIVKDTNIAYC 1011
CD126           KDDNTLYSCGENTHGQLGLGSESNHPDVLTFTVNNITNVRDVYCSDTTTFIVKDTNIAYC 1011
CD4             TNDGLLYGAGNNGKGQLGLGDTTSRN---IFTRIPINGVRDVYLCNDVSIIVKNDNTCYV 1060
CD123           TNDGLLYGAGNNGKGQLGLGDTTSRN---IFTRIPINGVRDVYLCNDVSIIVKNDNTCYV 1060
                .:*. **..*:* :******. :.:     **   *..***** .: .::***: * .*

CD108           CGYNNNSQLGMGNTTDQYSFIKCMENVKEVIPN-EINTYIITIYNTAYSTGLNTDYCLGL 1088
43593           CGYNNNSQLGMGNTTDQYSFIKCMENVKEVIPN-EINTYIITIYNTAYSTGLNTDYCLGL 1088
CD16            CGYNNNSQLGMGNTTDQYSFIKCMENVKEVIPN-EINTYIITIYNTAYSTGLNTDYCLGL 1070
CD126           CGYNNNSQLGMGNTTDQYSFIKCMENVKEVIPN-EINTYIITIYNTAYSTGLNTDYCLGL 1070
CD4             CGLVNGYFGFTEGSISTFTKIN-IENVKSVVTAGSEATFFITNDNMIYTTGKKERVFFST 1119
CD123           CGLVNGYFGFTEGSISTFTKIN-IENVKSVVTAGSEATFFITNDNMIYTTGKKERVFFST 1119
                **  *.      .: . :: *: :****.*:.  .  *::**  *  *:** :    :.

CD108           NSNSNQSSFSEIPISNVVKVAPNRNNAVLLLTSEGDVYTAGKCSNGSGTGSETPEKIKKI 1148
43593           NSNSNQSSFSEIPISNVVKVAPNRNNAVLLLTSEGDVYTAGKCSNGSGTGSETPEKIKKI 1148
CD16            NSNSNQSSFSEIPISNVVKVAPNRNNAVLLLTSEGDVYTAGKCSNGSGTGSETPEKIKKI 1130
CD126           NSNSNQSSFSEIPISNVVKVAPNRNNAVLLLTSEGDVYTAGKCSNGSGTGSETPEKIKKI 1130
CD4             ETNDIKGIRVINNIINAKKIVVN-GYTSAILTNDNKLFVGG----LSGYGSIANNNNTNS 1174
CD123           ETNDIKGIRVINNIINAKKIVVN-GYTSAILTNDNKLFVGG----LSGYGSIANNNNTNS 1174
                ::*. :.      * *. *:. * . :  :**.:..::..*     ** ** : :: .:

CD108           ASKAKDIGMNYRCGHYVSDNGDLYGTGFNDCGQLGVGNVTKRDTFIKTNTR-VKKILPLE 1207
43593           ASKAKDIGMNYRCGHYVSDNGDLYGTGFNDCGQLGVGDVTKRDTFIKTNTR-VKKILPLE 1207
CD16            ASKAKDIGMNYRCGHYVSDNGDLYGTGFNNNGQLGVGDVTKRDTFIKTNTR-VKKILPLE 1189
CD126           ASKAKDIGMNYRCGHYVSDNGDLYGTGFNNNGQLGVGDVTKRDTFIKTNTR-VKKILPLE 1189
CD4             VEDVKDVFVTANNTLYIDNNNNLISSGRDTYGISDESYRDMSVPYYKVSIKKDVDTVFSS 1234
CD123           VEDVKDVFVTANNTLYIDNNNNLISSGRDTYGISDESYRDMSVPYYKVSIKKDVDTVFSS 1234
                ....**: :. .   *:.:*.:* .:* :  *  . .      .: *.. :   . :  .
```

```
CD108           YANIAIKDTN-DIYICGLNN-YGQLGVGNRYDSRNN---DNRIFNYKHMNFVMGDLTSIK 1262
43593           YANIAIKDTN-DIYICGLNN-YGQLGVGNRYDSRNN---DNRIFNYKHMNFVMGDLTSIK 1262
CD16            YANIAIKDTN-DIYICGLNN-YGQLGVGNRYDSRNN---DNRIFNYKHMNFVMGDLTSIK 1244
CD126           YANIAIKDTN-DIYICGLNN-YGQLGVGNRYDSRNN---DNRIFNYKHMNFVMGDLTSIK 1244
CD4             YNTIFIKDIYGKFYSSTRDNRYNHLGIHHRYDNDKNEALEGSLHSYFKTDNTSDKIVFNK 1294
CD123           YNTIFIKDIYGKFYSSTRDNRYNHLGIHHRYDNDKNEALEGSLHSYFKTDNTSDKIVFNK 1294
                * .* ***   .:* .  :* *.:**: :***. :*   :. :..* : : . ..:.  *

CD108           NRHNFILLNNKIVIPTTKDIDYGLVLGNLYKGDLYTELPYEDIKEVSISKTHIIILLNDG 1322
43593           NRHNFILLNNKIVIPTTKDIDYGLVLGNLYKGDLYTELPYEDIKEVSISKTHIIILLNDG 1322
CD16            NRHNFILLNNKIVIPTTKDIDYGLVLGNLYKGDLYTELPYEDIKEVSISKTHIIILLNDG 1304
CD126           NRHNFILLNNKIVIPTTKDIDYGLVLGNLYKGDLYTELPYEDIKEVSISKTHIIILLNDG 1304
CD4             KNEKLVMFNDKYIKTNNKYINYKNIFKDNFK--YTSIILPFEVSDIDISKTHSLAVAKDG 1352
CD123           KNEKLVMFNDKYIKTNNKYINYKNIFKDNFK--YTSIILPFEVSDIDISKTHSLAVAKDG 1352
                :..::::::*:* : ...* *:*  :: : :*    : :   ::.::.***** : : :**

CD108           TMYGCGTNYHGELLQDLSINQVDEFVQINVSDVKHVSCGDNFTYFIKSDDSLWSIGKNSE 1382
43593           TMYGCGTNYHGELLQDLSINQVDEFVQINVSDVKHVSCGDNFTYFIKSDDSLWSIGKNSE 1382
CD16            TMYGCGTNYHGELLQDLSINQVDEFVQINVSDVKHVSCGDNFTYFIKSDDSLWSIGKNSE 1364
CD126           TMYGCGTNYHGELLQDLSINQVDEFVQINVSDVKHVSCGDNFTYFIKSDDSLWSIGKNSE 1364
CD4             KLYGIGSNSYKEINQTLEDIELLTLTEVNISDVKKVACGDNYSYIIKTDNTLWSYGKNTE 1412
CD123           KLYGIGSNSYKEINQTLEDIELLTLTEVNISDVKKVACGDNYSYIIKTDNTLWSYGKNTE 1412
                .:** *:* : *: * *.  ::  :.::*:****:*:****::*:**:*::*** ***:*

CD108           YQLGIGHNNPVTELQRITTISSCKEVHCGKNYTLVVTTSNELFVQGYNDKGALGLGSDSE 1442
43593           YQLGIGHNNPVTELQRITTISSCKEVHCGKNYTLVVTTGNELFVQGYNDKGALGLGSDSE 1442
CD16            YQLGIGHNNPVTELQRITTISSCKEVHCGKNYTLVVTTGNELFVQGYNDKGALGLGSDSE 1424
CD126           YQLGIGHNNPVTELQRITTISSCKEVHCGKNYTLVVTTGNELFVQGYNDKGALGLGSDSE 1424
CD4             YQLGVGHNNDVRELQKVTGLPSVKDISIYNSMTLVLTNEGELYAQGYNTNGLFGLGESEK 1472
CD123           YQLGVGHNNDVRELQKVTGLPSVKDISIYNSMTLVLTNEGELYAQGYNTNGLFGLGESEK 1472
                ****:**** * ***::* :.* *::   :. ***:*. .**:.**** :* :***...:

CD108           NTIIKFFTKALTDIREIKSYGSDHILVLKNDNSVWVTGKNRDVYKIEQPVEFLKEFTIVP 1502
43593           NTIIKFFTKALTDIREIKSYGSDHILVLKNDNSVWVTGKNRDVYKIEQPVEFLKEFTIIP 1502
CD16            NTIIKFFTKALTDIREIKSYGSDHILVLKNDNSVWVTGKNRDVYKIEQPVEFLKEFTIVP 1484
CD126           NTIIKFFTKALTDIREIKSYGSDHILVLKNDNSVWVTGKNRDVYKIEQPVEFLKEFTIVP 1484
CD4             DKIIRTFTKVLTNVKEIKSHNDDHILVIKNDNSLWITGKNKSMYKISISITDLYEFTKIP 1532
CD123           DKIIRTFTKVLTNVKEIKSHNDDHILVIKNDNSLWITGKNKSMYKISISITDLYEFTKIP 1532
                :.**: ***.**:::****:..*****:*****:*:****:.:***. .:  * *** :*

CD108           ISEDVNTVKDVLATDNTLYIISEVGTTNAAIEITEKSISSIKIKIQDPNKDISRIEMLIN 1562
43593           ISEDVNTVKDVLATDNTLYIISEVGTTNAAIEITEKSISSIKIKIQDPNKDISRIEMLIN 1562
CD16            ISEDVNTVKDVLATDNTLYIISEVGTTNAAIEITEKSISSIKIKIQDPNKDISRIEMLIN 1544
CD126           ISEDVNTVKDVLATDNTLYIISEVGTTNAAIEITEKSISSIKIKIQDPNKDISRIEMLIN 1544
CD4             IPEHLNDILDIELSDDTIYMITKVDTSKASIEIVEKSISQVRVVVQDPNNVIEKLEMFIN 1592
CD123           IPEHLNDILDIELSDDTIYMITKVDTSKASIEIVEKSISQVRVVVQDPNNVIEKLEMFIN 1592
                *.*.:* : *:  :*:*:*:*::*.*::*:***.*****.::: :****: *.::**:**

CD108           GESVKSVSDLITEKISFEVPPDKIKIGENKILFRAYCKGDDLYASLFIFKESTGNSIIKD 1622
43593           GESVKSVSDLITEKISFEVPPDKIKIGENKILFRAYCKGDDLYASLFIFKESTGNSIIKD 1622
CD16            GESVKSVSDLTTEKISFEVPPDKIKIGENKILFRAYCKGDDLYASLFIFKESTGNSIIKD 1604
CD126           GESVKSVSDLTTEKISFEVPPDKIKIGENKILFRAYCKGDDLYASLFIFKESTGNSIIKD 1604
CD4             DELISTKTNLEINSIIFEIPQNKIVLGENKILIKASSPTGDLYSSMFIFKSETGLKVKKD 1652
CD123           DELISTKTNLEINSIIFEIPQNKIVLGENKILIKASSPTGDLYSSMFIFKSETGLKVKKD 1652
                .* :.: ::*  :.* **:* :** :******::* .  .***:*:****..** .: **
```

```
CD108           SYVMIGNRMYKVVNTTSNEQDITITLDRGLEEDLNLGDPIYQLINKTKVQVKINKSDLFK 1682
43593           SYVMIGNRMYKVVNTTSNEQDITITLDRGLEEDLNLGDPIYQLINKTKVQVKINKSDLFK 1682
CD16            SYVMIGNRMYKVVNTTSNEQDITITLDRGLEEDLNLGDPIYQLINKTKVQVKINKSDLFK 1664
CD126           SYVMIGNRMYKVVNTTSNEQDITITLDRGLEEDLNLGDPIYQLINKTKVQVKINKSDLFK 1664
CD4             SILMINNKVYSIINITENNTDLIVTLNEGLKDDMMENNPIYQLINKTKVQVKINKSDLFK 1712
CD123           SILMINNKVYSIINITENNTDLIVTLNEGLKDDMMENNPIYQLINKTKVQVKINKSDLFK 1712
                * :**.*::*.::* *.*: *: :**:.**::*:  .:**********************

CD108           DMKLVEIKKSDSSYQEIYELEEANIKSAQPKIIVEKGDKWTAIKRPSMIFRYDAENNEPQ 1742
43593           DMKLVEIKKSDSSYQEIYELEEANIKSAQPKIIVEKGDKWTAIKRPSMIFRYDAENNEPQ 1742
CD16            DMKLVEIKKSDSSYQEIYELEEANIKSAQPKIIVEKGDKWTAIKRPSMIFRYDAENNEPQ 1724
CD126           DMKLVEIKKSDSSYQEIYELEEANIKSAQPKIIVEKGDKWTAIKRPSMIFRYDAENNEPQ 1724
CD4             DMKLVEIKKSDSSYQEIYELEEANIKSAQPKIIVEKGDKWTAIKRPSMIFRYDAENNEPQ 1772
CD123           DMKLVEIKKSDSSYQEIYELEEANIKSAQPKIIVEKGDKWTAIKRPSMIFRYDAENNEPQ 1772
                ************************************************************
```

Figure 9.

