

(19) World Intellectual Property
Organization

International Bureau

(43) International Publication Date
24 December 2020 (24.12.2020)(10) International Publication Number
WO 2020/257610 A8

(51) International Patent Classification:

A61K 35/74 (2015.01) C12N 15/70 (2006.01)
A61K 38/44 (2006.01) C12N 9/10 (2006.01)

(21) International Application Number:

PCT/US2020/038675

(22) International Filing Date:

19 June 2020 (19.06.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/865,129 21 June 2019 (21.06.2019) US
62/864,875 21 June 2019 (21.06.2019) US

(71) Applicants: **SYNOLOGIC OPERATING COMPANY, INC.** [US/US]; 301 Binney Street, Suite 402, Cambridge, MA 02142 (US). **GINKGO BIOWORKS, INC.** [US/US]; 27 Drydock Avenue, 8th Floor, Boston, MA 02210 (US).

(72) Inventors; and

(71) Applicants: **CARLIN, Dylan, Alexander** [US/US]; 49 Saint John Street, Apt. 1, Jamaica Plain, MA 02130 (US). **STONE, Laura, K.** [US/US]; 54 Kelly Road, Apt. 2R, Cambridge, MA 02139 (US). **TUCKER, Alex, C.** [US/US]; 17 Pond Street, Unit 2, Dorchester, MA 02125 (US). **BOYLE, Patrick, M.** [US/US]; 37 Forest Hills Street, Unit 1, Boston, MA 02130 (US).

(72) Inventors: **LI, Ning**; 322 Cambridge Street, Winchester, MA 01890 (US). **GAO, Jian-Rong**; 18 Roberts Street, Malden, MA 02148 (US). **MORRISON, Philippa Jane Reeder**; 73 School Street, Arlington, MA 02476 (US). **BERGERON, Christopher, George**; 314 Walton Street, Fitchburg, MA 01420 (US).

(74) Agent: **CLARKE, Marcie, B.** et al.; McCarter & English, LLP, 265 Franklin Street, Boston, MA 02110 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ,

CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

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

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))

(48) Date of publication of this corrected version:

28 January 2021 (28.01.2021)

(15) Information about Correction:

see Notice of 28 January 2021 (28.01.2021)

(54) Title: OPTIMIZED BACTERIA ENGINEERED TO TREAT DISORDERS INVOLVING THE CATABOLISM OF LEUCINE, ISOLEUCINE, AND/OR VALINE

(57) Abstract: The present disclosure provides recombinant bacterial cells that have been engineered with genetic circuitry which allow the recombinant bacterial cells to sense a patient's internal environment and respond by turning an engineered metabolic pathway on or off. When turned on, the recombinant bacterial cells complete all of the steps in a metabolic pathway to achieve a therapeutic effect in a host subject. These recombinant bacterial cells are designed to drive therapeutic effects throughout the body of a host from a point of origin of the microbiome. Specifically, the present disclosure provides recombinant bacterial cells comprising a heterologous gene encoding an improved leucine catabolism enzyme with higher activity and/or specificity for leucine over other branched chain amino acids, such as isoleucine or valine. The disclosure further provides pharmaceutical compositions comprising the recombinant bacteria, and methods for treating disorders involving the catabolism of leucine, isoleucine, and/or valine using the pharmaceutical compositions disclosed herein.

OPTIMIZED BACTERIA ENGINEERED TO TREAT DISORDERS INVOLVING THE CATABOLISM OF LEUCINE, ISOLEUCINE, AND/OR VALINE

Related Applications

[01] This application claims priority to U.S. Provisional Application No. 62/864,875, filed on June 21, 2019, and U.S. Provisional Application No. 62/865,129, filed on June 21, 2019. The entire contents of each of the foregoing applications are hereby incorporated by reference.

Sequence Listing

[02] The instant application contains a sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on June 18, 2020, is named 126046-04220_SL.txt and is 876,737 bytes in size.

Background

[03] The branched chain amino acids (BCAAs), *e.g.*, leucine, isoleucine, and valine, play an important role in the metabolism of living organisms. Transamination of branched chain amino acids gives rise to their corresponding branched chain α -keto acids (BCKAs) (α -keto- β -methylvalerate, α -ketoisocaproate, and α -ketoisovalerate), which undergo further oxidative decarboxylation to produce acyl-CoA derivatives that enter the TCA cycle. Branched chain amino acids provide a nonspecific carbon source of oxidation for production of energy and also act as a precursor for muscle protein synthesis (Monirujjaman and Ferdouse, *Advances in Molec. Biol.*, 2014, Article ID 36976, 6 pages, 2014).

[04] Enzyme deficiencies or mutations which lead to the toxic accumulation of branched chain amino acids and their corresponding alpha-keto acids in the blood, cerebrospinal fluid, and tissues result in the development of metabolic disorders associated with the abnormal catabolism of branched chain amino acids in subjects, such as maple syrup urine disease (MSUD), isovaleric acidemia, propionic acidemia, methylmalonic acidemia, and diabetes ketoacidosis. Clinical manifestations of the disease vary depending on the degree of enzyme deficiency and include neurological dysfunction, seizures and death (Homanics *et al.* 2009).

[05] Branched chain amino acids, such as leucine, or their corresponding alpha-keto acids, have also been linked to mTor activation (see, for example, Harlan *et al.*, *Cell Metabolism*, 17:599-606, 2013) which is, in turn, associated with diseases such as cancer, obesity, type 2 diabetes, neurodegeneration, autism, Alzheimer's disease, Lymphangioliomyomatosis (LAM), transplant rejection, glycogen storage disease, obesity, tuberous sclerosis, hypertension, cardiovascular disease, hypothalamic activation, musculoskeletal disease, Parkinson's disease, Huntington's disease,

psoriasis, rheumatoid arthritis, lupus, multiple sclerosis, Leigh's syndrome, and Friedrich's ataxia (see Laplante and Sabatini, *Cell*, 149(2):74-293, 2012).

[06] Currently available treatments for disorders involving the catabolism of branched chain amino acids are inadequate for the long term management of the disorders and have severe limitations (Svkvorak, *J. Inherit. Metab. Dis.*, 32(2):229-246, 2009). A low protein/BCAA-restricted diet, with micronutrient and vitamin supplementation, as necessary, is the widely accepted long-term disease management strategy for many such disorders (Homanics *et al.*, *BMC Med. Genet.*, 7:33, 2006). However, BCAA-intake restrictions can be particularly problematic since branched chain amino acids can only be acquired through diet and are necessary for metabolic activities including protein synthesis and branched-chain fatty acid synthesis (Skvorak, 2009). Thus, even with proper monitoring and patient compliance, branched chain amino acid dietary restrictions result in a high incidence of mental retardation and mortality (Skvorak, 2009; Homanics *et al.*, 2009). A few cases of MSUD have been treated by liver transplantation (Popescu and Dima, *Liver Transpl.*, 1:22-28, 2012). However, the limited availability of donor organs, the costs associated with the transplantation itself, and the undesirable effects associated with continued immunosuppressant therapy limit the practicality of liver transplantation for treatment of disorders involving the catabolism of a branched chain amino acid (Homanics *et al.*, 2012; Popescu and Dima, 2012). Therefore, there is significant unmet need for effective, reliable, and/or long-term treatment for disorders involving the catabolism of branched chain amino acids.

Summary

[07] The present disclosure relates to compositions and therapeutic methods for reducing one or more excess branched chain amino acids, and/or an accumulated metabolite(s) thereof, for example, by converting the one or more excess branched chain amino acid(s) or accumulated metabolite(s) into alternate by product(s). In certain aspects, the disclosure relates to genetically engineered microorganisms, *e.g.*, bacteria, yeast or viruses, that have been optimized to reduce one or more excess branched chain amino acids, and/or an accumulated metabolite(s) thereof, particularly in low-oxygen conditions, such as in the mammalian gut. In certain aspects, the compositions and methods disclosed herein may be used to treat disorders associated with excess branched chain amino acids and/or an accumulated metabolite(s) thereof, *e.g.*, MSUD.

[08] Specifically, the present disclosure provides recombinant microorganisms that have been engineered with optimized genetic circuitry, which allows the recombinant microorganism to import and/or metabolize one or more branched chain amino acid and/or one or more metabolite(s) thereof at an improved rate. In some embodiments, the engineered microorganism is capable of sensing a patient's internal environment, *e.g.*, the gut, and responding by turning an engineered metabolic pathway on or off. When turned on, the engineered microorganism, *e.g.*, bacterial, yeast or

virus cell, expresses one or more enzymes in a metabolic pathway to achieve a therapeutic effect in a host subject.

[09] In certain aspects, the present disclosure provides engineered bacterial cells, pharmaceutical compositions thereof, and methods of modulating BCAA(s) and/or metabolite(s) thereof and treating diseases associated with the catabolism of branched chain amino acids. Specifically, the engineered bacteria disclosed herein have been modified to comprise optimized gene sequence(s) encoding one or more enzymes involved in branched chain amino acid catabolism, as well as other circuitry, *e.g.*, to regulate gene expression, including, for example, sequences for one or more inducible promoter(s), sequences for ribosome binding sites, sequences for importing one or more BCAA(s) and/or metabolite(s) thereof into the bacterial cell (*e.g.*, transporter sequence(s)), sequences for the secretion or non-secretion of BCAA(s), metabolites or by-products (*e.g.*, exporter(s) or exporter knockouts), and circuitry to guarantee the safety and non-colonization of the subject that is administered the recombinant bacteria, such as auxotrophies, kill switches, *etc.* These engineered bacteria are safe and well tolerated and augment the innate activities of the subject's microbiome to achieve a therapeutic effect.

[010] In one aspect, disclosed herein is a recombinant bacterium capable of consuming leucine at a rate of at least about $0.5 \mu\text{mol}/10^9 \text{ CFU/h}$ *in vitro*. In one embodiment, the bacterium is capable of consuming leucine at a rate of at least about $0.75 \mu\text{mol}/10^9 \text{ CFU/h}$. In one embodiment, the bacterium is capable of consuming leucine at a rate of at least about $1 \mu\text{mol}/10^9 \text{ CFU/h}$. In one embodiment, the bacterium is capable of consuming leucine at a rate of at least about $1.25 \mu\text{mol}/10^9 \text{ CFU/h}$. In one embodiment, the bacterium is capable of consuming leucine at a rate of at least about $1.5 \mu\text{mol}/10^9 \text{ CFU/h}$. In one embodiment, the bacterium is capable of consuming leucine at a rate of at least about $1.75 \mu\text{mol}/10^9 \text{ CFU/h}$. In one embodiment, the bacterium is capable of consuming leucine at a rate of at least about $2 \mu\text{mol}/10^9 \text{ CFU/h}$. In one embodiment, the bacterium is capable of consuming leucine at a rate of at least about $2.25 \mu\text{mol}/10^9 \text{ CFU/h}$. In one embodiment, the bacterium is capable of consuming leucine at a rate of at least about $2.5 \mu\text{mol}/10^9 \text{ CFU/h}$. In one embodiment, the bacterium is capable of consuming leucine at a rate of at least about $2.75 \mu\text{mol}/10^9 \text{ CFU/h}$.

[011] In one embodiment, the bacterium is capable of consuming leucine at a rate of about 0.5 to about $2.75 \mu\text{mol}/10^9 \text{ CFU/h}$. In one embodiment, the bacterium is capable of consuming leucine at a rate of about 0.75 to about $2.75 \mu\text{mol}/10^9 \text{ CFU/h}$. In one embodiment, the bacterium is capable of consuming leucine at a rate of about 1 to about $2.75 \mu\text{mol}/10^9 \text{ CFU/h}$. In one embodiment, the bacterium is capable of consuming leucine at a rate of about 1.25 to about $2.75 \mu\text{mol}/10^9 \text{ CFU/h}$. In one embodiment, the bacterium is capable of consuming leucine at a rate of about 1.75 to about $2.75 \mu\text{mol}/10^9 \text{ CFU/h}$. In one embodiment, the bacterium is capable of consuming leucine at a rate of about 2 to about $2.75 \mu\text{mol}/10^9 \text{ CFU/h}$. In one embodiment, the bacterium is capable of consuming leucine at a rate of about 2.25 to about $2.75 \mu\text{mol}/10^9 \text{ CFU/h}$. In one embodiment, the bacterium is capable of consuming leucine at a rate of about 2.5 to about $2.75 \mu\text{mol}/10^9 \text{ CFU/h}$. In one

embodiment, the bacterium is capable of consuming leucine at a rate of about 0.5 to about 2 $\mu\text{mol}/10^9$ CFU/h. In one embodiment, the bacterium is capable of consuming leucine at a rate of about 1 to about 2 $\mu\text{mol}/10^9$ CFU/h. In one embodiment, the bacterium is capable of consuming leucine at a rate of about 0.5 to about 2.5 $\mu\text{mol}/10^9$ CFU/h.

[012] In one embodiment, the bacterium comprises at least one gene sequence encoding at least one leucine catabolism enzyme operably linked to at least one directly or indirectly inducible promoter that is not associated with a gene encoding the at least one leucine catabolism enzyme in nature. In one embodiment, the promoter is selected from the group consisting of an FNRS promoter, a P_{tet} promoter, a P_{cmt} promoter, an FNRS promoter, a P_{cl857} promoter, and a P_{BAD} promoter. In one embodiment, the at least one gene sequence encoding the at least one leucine catabolism enzyme is *leuDH*, *kivD*, and/or *adh2*.

[013] In one embodiment, the bacterium further comprises at least one gene sequence encoding at least one transporter capable of transporting leucine. In one embodiment, the at least one gene sequence encoding the at least one transporter is *brnQ*.

[014] In one embodiment, the bacterium comprises an operon, wherein the operon comprises *leuDH*, *kivD*, *adh2*, and *brnQ*. In one embodiment, the operon comprises a sequence having at least 90% identity to a sequence of any one of SEQ ID NOs: 25-48. In one embodiment, the operon comprises a sequence having at least 95% identity to a sequence of any one of SEQ ID NOs: 25-48. In one embodiment, the operon comprises a sequence comprising a sequence of any one of SEQ ID NOs: 25-48. In one embodiment, the operon comprises a sequence consisting of any one of SEQ ID NOs: 25-48.

[015] In one embodiment, the *leuDH* gene encodes a LeuDH protein which comprises a sequence having at least 90% identity to a sequence of any one of SEQ ID NOs: 49-72. In one embodiment, the *leuDH* gene encodes a LeuDH protein which comprises a sequence having at least 95% identity to a sequence of any one of SEQ ID NOs: 49-72. In one embodiment, the *leuDH* gene encodes a LeuDH protein which comprises a sequence of any one of SEQ ID NOs: 49-72. In one embodiment, the *leuDH* gene encodes a LeuDH protein which consists of a sequence of any one of SEQ ID NOs: 49-72.

[016] In one embodiment, the *kivD* gene encodes a KivD protein which comprises a sequence having at least 90% identity to a sequence of any one of SEQ ID NOs: 73-96. In one embodiment, the *kivD* gene encodes a KivD protein which comprises a sequence having at least 95% identity to a sequence of any one of SEQ ID NOs: 73-96. In one embodiment, the *kivD* gene encodes a KivD protein which comprises a sequence of any one of SEQ ID NOs: 73-96. In one embodiment, the *kivD* gene encodes a KivD protein which consists of a sequence of any one of SEQ ID NOs: 73-96.

[017] In one embodiment, the *adh2* gene encodes an Adh2 protein which comprises a sequence having at least 90% identity to a sequence of any one of SEQ ID NOs: 97-120. In one

embodiment, the *adh2* gene encodes an Adh2 protein which comprises a sequence having at least 95% identity to a sequence of any one of SEQ ID NOs: 97-120. In one embodiment, the *adh2* gene encodes an Adh2 protein which comprises a sequence of any one of SEQ ID NOs: 97-120. In one embodiment, the *adh2* gene encodes an Adh2 protein which consists of a sequence of any one of SEQ ID NOs: 97-120.

[018] In one embodiment, the *brnQ* gene encodes a BrnQ protein which comprises a sequence having at least 90% identity to a sequence of any one of SEQ ID NOs: 121-144. In one embodiment, the *brnQ* gene encodes a BrnQ protein which comprises a sequence having at least 95% identity to a sequence of any one of SEQ ID NOs: 121-144. In one embodiment, the *brnQ* gene encodes a BrnQ protein which comprises a sequence of any one of SEQ ID NOs: 121-144. In one embodiment, the *brnQ* gene encodes a BrnQ protein which consists of a sequence of any one of SEQ ID NOs: 121-144.

[019] In another aspect, disclosed herein is a recombinant bacterium comprising an operon, wherein the operon comprises an optimized *leuD*H gene which encodes a LeuDH protein, an optimized *kivD* gene which encodes a KivD protein, and an optimized *adh2* gene which encodes an Adh2 protein, and an optimized *brnQ* gene which encodes a BrnQ protein, wherein the operon comprises a sequence having at least 90% identity to a sequence of any one of SEQ ID NOs: 25-48, and wherein the operon is operably linked to an inducible promoter that is not associated with a *leuD*H gene, a *kivD* gene, a *adh2* gene, or a *brnQ* gene in nature.

[020] In one embodiment, the operon comprises a sequence having at least 95% identity to a sequence of any one of SEQ ID NOs: 25-48. In one embodiment, the operon comprises a sequence of any one of SEQ ID NOs: 25-48. In one embodiment, the operon consists of a sequence of any one of SEQ ID NOs: 25-48.

[021] In one embodiment, the LeuDH protein comprises a sequence having at least 90% identity to a sequence of any one of SEQ ID NOs: 49-72. In one embodiment, the LeuDH protein comprises a sequence having at least 95% identity to a sequence of any one of SEQ ID NOs: 49-72. In one embodiment, the LeuDH protein comprises a sequence of any one of SEQ ID NOs: 49-72. In one embodiment, the LeuDH protein consists of a sequence of any one of SEQ ID NOs: 49-72.

[022] In one embodiment, the KivD protein comprises a sequence having at least 90% identity to a sequence of any one of SEQ ID NOs: 73-96. In one embodiment, the KivD protein comprises a sequence having at least 95% identity to a sequence of any one of SEQ ID NOs: 73-96. In one embodiment, the KivD protein comprises a sequence of any one of SEQ ID NOs: 73-96. In one embodiment, the KivD protein consists of a sequence of any one of SEQ ID NOs: 73-96.

[023] In one embodiment, the Adh2 protein comprises a sequence having at least 90% identity to a sequence of any one of SEQ ID NOs: 97-120. In one embodiment, the Adh2 protein comprises a sequence having at least 95% identity to a sequence of any one of SEQ ID NOs: 97-120. In one embodiment, the Adh2 protein comprises a sequence of any one of SEQ ID NOs: 97-120. In

one embodiment, the Adh2 protein comprises a sequence consisting of any one of SEQ ID NOs: 97-120.

[024] In one embodiment, the BrnQ protein comprises a sequence having at least 90% identity to a sequence of any one of SEQ ID NOs: 121-144. In one embodiment, the BrnQ protein comprises a sequence having at least 95% identity to a sequence of any one of SEQ ID NOs: 121-144. In one embodiment, the BrnQ protein comprises a sequence of any one of SEQ ID NOs: 121-144. In one embodiment, the BrnQ protein consists of a sequence of any one of SEQ ID NOs: 121-144.

[025] In one embodiment, the at least one gene sequence encoding the at least one leucine catabolism enzyme and the inducible promoter, are integrated into the chromosome of the bacterium. In one embodiment, the at least one gene sequence encoding the at least one leucine catabolism enzyme and the inducible promoter, are located in a plasmid in the bacterium. In another embodiment, the operon and the inducible promoter are integrated into the chromosome of the bacterium. In another embodiment, the operon and the inducible promoter are located in a plasmid in the bacterium.

[026] In one embodiment, the promoter is selected from the group consisting of an FNRS promoter, a P_{tet} promoter, a P_{cmt} promoter, an FNRS promoter, a P_{cl857} promoter, and a P_{BAD} promoter.

[027] In one embodiment, the bacterium further comprises a second *brnQ* gene which encodes a second BrnQ protein, wherein the second *brnQ* gene is operably linked to an inducible promoter that is not associated with a *brnQ* gene in nature, and wherein the second *brnQ* gene is inserted into the *malE/K* or *malP/T* locus on a chromosome of the bacterium.

[028] In one embodiment, the bacterium comprises a genetic modification in *leuE* that reduces leucine export from the bacterium. In one embodiment, the bacterium comprises a genetic modification in *ilvC* that reduces endogenous biosynthesis of leucine in the bacterium. In one embodiment, the bacterium further comprises a *livKHMGF* gene sequence encoding at least one transporter capable of transporting leucine. In one embodiment, the *livKHMGF* gene sequence is operably linked to at least one promoter that is not associated with a *livKHMGF* gene in nature.

[029] In one embodiment, the bacterium is capable of consuming leucine at a rate of at least about 0.5 $\mu\text{mol}/10^9\text{CFU/h}$ *in vitro*. In one embodiment, the bacterium is capable of consuming leucine at a rate of at least about 1 $\mu\text{mol}/10^9\text{CFU/h}$ *in vitro*. In one embodiment, the bacterium is capable of consuming leucine at a rate of at least about 1.5 $\mu\text{mol}/10^9\text{CFU/h}$ *in vitro*. In one embodiment, the bacterium is capable of consuming leucine at a rate of at least about 2.0 $\mu\text{mol}/10^9\text{CFU/h}$ *in vitro*. In one embodiment, the bacterium is capable of consuming leucine at a rate of at least about 2.5 $\mu\text{mol}/10^9\text{CFU/h}$ *in vitro*. In one embodiment, the bacterium is capable of consuming leucine at a rate of at least about 2.75 $\mu\text{mol}/10^9\text{CFU/h}$ *in vitro*.

[030] In one embodiment, the bacterium is capable of producing isopentanol at a rate of at least about 0.2 $\mu\text{mol}/10^9\text{CFU/h}$ *in vitro*. In one embodiment, the bacterium is capable of producing isopentanol at a rate of at least about 0.3 $\mu\text{mol}/10^9\text{CFU/h}$ *in vitro*. In one embodiment, the bacterium

is capable of producing isopentanol at a rate of at least about 0.4 $\mu\text{mol}/10^9\text{CFU/h}$ *in vitro*. In one embodiment, the bacterium is capable of producing isopentanol at a rate of at least about 0.5 $\mu\text{mol}/10^9\text{CFU/h}$ *in vitro*. In one embodiment, the bacterium is capable of producing isopentanol at a rate of at least about 0.2 to 0.5 $\mu\text{mol}/10^9\text{CFU/h}$ *in vitro*.

[031] In one embodiment, the bacterium exhibits preferentially consumes leucine over valine and isoleucine. In one embodiment, the bacterium exhibits a leucine/valine activity ratio of at least about 1.1 to at least about 2.75. In one embodiment, the bacterium exhibits a leucine/isoleucine activity ratio of at least about 1.1 to at least about 5.

[032] In one embodiment, the bacterium is a probiotic bacterium selected from the group consisting of *Bacteroides*, *Bifidobacterium*, *Clostridium*, *Escherichia*, *Lactobacillus*, and *Lactococcus*. In one embodiment, the bacterium is *Escherichia coli* strain Nissle.

[033] In one aspect, disclosed herein is a pharmaceutically acceptable composition comprising the recombinant bacterium of disclosed herein, and a pharmaceutically acceptable carrier.

[034] In one aspect, disclosed herein is a method of reducing the level of leucine in a subject, the method comprising a step of administering to the subject a pharmaceutically acceptable composition.

[035] In one aspect, disclosed herein is a method of treating a disease associated with excess leucine and/or a metabolic disorder involving the abnormal catabolism of leucine in a subject, the method comprising a step of administering to the subject a pharmaceutically acceptable composition.

[036] In one embodiment, the subject has, or is suspected of having maple syrup urine disease (MSUD).

Brief Description of the Drawings

[037] **Figure 1** depicts graphs showing sequence similar networks for LeuD_H (**Figure 1A**), KivD (**Figure 1B**), and ADH (**Figure 1C**) demonstrating diversity of each enzyme metagenomic library. Each spot in **Figure 1(A-C)** represents a single amino acid sequence available in sequence databases, and the distance between spots demonstrating relatedness. Each sequence similarity network has a color key with information regarding the annotation or source of the enzyme. For LeuD_H enzymes, color indicates the annotation of the enzyme. For KivD and Adh, the color of each spot represents the phylogenetic clade from which the enzyme was sourced. In each network, the seed enzyme sequence from the prototype pathway (SYN1980: comprising ΔleuE , ΔilvC , $\text{lacZ}::\text{tetR-Ptet-livKHMgf}$, $\text{tetR-Ptet-leuD}_H(\text{Bc})$ - kivD - adh2 - brnQ - rrnB ter (in low-copy pSC101 plasmid) is indicated by a red dot.

[038] **Figure 2** depicts a graph showing the LeuD_H enzyme activity measured in high throughput. The top 220 LeuD_H enzymes from the primary screen were re-screened with increased

biological replication (n=4) to validate enzyme activity. Activities are reported as fold-improvement over the *B. cereus* LeuDH activity.

[039] **Figure 3** depicts a graph showing the activity specificity for Leu over Val and Ile among the top ~200 LeuDH enzymes screened for the activity on Leu, Val, and Ile. Activity of LeuDH enzymes on Leu are reported relative the *B. cereus* LeuDH activity. Specificity was measured as the ratio of activity on Leu relative to Val/Leu. In the left panel, enzyme activity on Leu is reported relative to the Leu/Val specificity. In the right panel, enzyme activity is reported relative to the Leu/Ile specificity. Rationally engineered active site variants are shown in green. Sourced LeuDH enzymes are shown in blue-gray. The negative control is shown in gray. Enzymes exhibited up to ~2.7-fold preference for Leu over Val, and up to a 5-fold preference for Leu over Ile. The positive control *B. cereus* LeuDH, shown in red, exhibited equal preference for Leu, Val, and Ile.

[040] **Figure 4** depicts a graph showing the specificity comparison for Leu/Ile to Leu/Val among the top ~200 LeuDH enzymes screened for the activity on Leu, Val, and Ile. Specificity was measured as the ratio of activity on Leu relative to Val/Leu. Rationally engineered active site variants are shown in green. Sourced LeuDH enzymes are shown in blue-gray. The negative control is shown in gray. The positive control *B. cereus* LeuDH, shown in red, showed equal preference for Leu, Val, and Ile when measured in this assay.

[041] **Figure 5** depicts a graph showing the KivD enzyme activity measured in high throughput. The top 55 KivD enzymes from the primary screen were re-screened for activity with increased biological replication (n=4). Activities are reported as fold-improvement over the *S. aureus* KivD activity (which was equated to background).

[042] **Figure 6** depicts a graph showing the Adh enzyme activity measured in high throughput. The top 55 KivD enzymes from the primary screen were re-screened for activity with increased biological replication (n=4). Activities are reported as fold-improvement over the *S. cerevisiae* ADH2 activity (which was equated to background).

[043] **Figure 7** depicts a graph showing the consumption of Leu, Ile, and Val within the reaction mixture for each LeuDH enzyme. A total of 21 LeuDH enzymes were screened in cell lysate assays similar to the HTP screen, except that the reaction mixture contained Leu, Val, and Ile at 1:1:1 molar ratio. At least 10 LeuDH enzymes showed improved preference for Leu over Val and Ile when compared to the parent *B. subtilis* LeuDH. For nearly all LeuDH enzymes, least preference was shown for valine.

[044] **Figure 8** depicts a graph showing a schematic view of the optimized pathways. Each selected pathway enzyme (6 LeuDH enzymes, 3 KivD enzymes, and 3 Adh enzymes) were each paired with 3 RBSs. The RBS-enzyme pairs were combined in a partial combinatorial library, maintaining the gene order of the prototype pathway and keeping the plasmid-encoded BrnQ intact. In the schematic above, unique enzyme sequences for each family are indicated by a unique color. Unique RBSs are indicated with unique grayscale shading.

[045] **Figure 9** depicts graphs showing pathways with optimized enzymes screened for Leu consumption (top panel) and isopentanol production (bottom panel, same strain order). Strains were assayed in biological triplicate and rank-ordered based on the best performance of any single replicate. The top ~100 leucine consuming strains are shown here. The control strains are SYN1980 (comprising Δ leuE, Δ ilvC, lacZ:tetR-Ptet-livKHMGE, tetR-Ptet-leuDH(Bc)-kivD-adh2-brnQ-rnB ter (in low-copy pSC101 plasmid)), SYN1980 Δ brnQ, and SYN1980 Δ leuDH.

[046] **Figure 10** depicts graphs showing targeted metabolomics of top 10 Leu consuming strains demonstrating improved Leu consumption and relieved KivD bottleneck. Strains harboring the top 10 improved pathways identified in **Figure 9** were characterized for pathway intermediates in the Leu consuming assay. Reactions were sampled at 2-hour intervals over 4 hours.

[047] **Figure 11** depicts a graph showing the leucine consumption activities of engineered strains SYN5721-SYN5744 optimized for BCAA consumption (comprising Δ leuE, Δ ilvC, lacZ:tetR-Ptet-livKHMGE, pSC101-Ptet-optimized operon-brnQ-ampR). SYN5721-SYN5744 were compared against engineered strain SYN1980 (comprising Δ leuE, Δ ilvC, lacZ:tetR-Ptet-livKHMGE, tetR-Ptet-leuDH(Bc)-kivD-adh2-brnQ-rnB ter (in low-copy pSC101 plasmid). Strains were assayed by adding 4 mM leucine to minimum media and sampling at 0, 1, or 2 hour time points after anaerobic incubation.

[048] **Figure 12** depicts a graph showing engineered strains generated with different host strains to demonstrate the host strain effects on leucine consumption activity. See **Table 1** for detailed strain description. Strains were assayed by adding 8 mM leucine to minimum media and sampling at 0, 2, or 4 hour time points after anaerobic incubation.

[049] **Figure 13** depicts a graph showing engineered strains generated with different inducible promoters operatively linked to the optimized operon to demonstrate the promoter effects on leucine consumption activity. See **Table 1** for detailed strain description. Strains were assayed by adding 8 mM leucine to minimum media and sampling at 0, 1.5, or 3 hour time points after anaerobic incubation.

[050] **Figure 14** depicts graphs showing the ability of engineered strain SYN5941 optimized for BCAA consumption (comprising pSC101-P_{tet}-optimized operon-brnQ-ampR) to consume BCAA in the in vitro gastrointestinal simulation (IVS) model system. SYN5941 was compared against engineered strain SYN1980 (comprising Δ leuE, Δ ilvC, lacZ:tetR-Ptet-livKHMGE, tetR-Ptet-leuDH(Bc)-kivD-adh2-brnQ-rnB ter (in low-copy pSC101 plasmid) and the wild-type strain SYN001 (*i.e.*, EcN). **Figure 14A** depicts a higher ability for SYN5941 than SYN1980 and SYN001 to consume leucine (left panel) and produce isopentanol (right panel) in the presence of 5 mM leucine in IVS mimicking the low oxygen tension condition of the human upper gastrointestinal tract. **Figure 14B** depicts a higher depicts higher abilities for SYN5941 (left panel) than SYN1980 (right panel) to consume leucine, isoleucine and valine in the presence of a mixture of leucine, isoleucine and valine each at 5 mM.

[051] **Figures 15A, 15B, and 15C** depict graphs showing the ability of leucine consumption of engineered strain SYN5941 in various conditions. **Figure 15A**, **Figure 15B**, and **Figure 15C** show leucine consumption in increasing concentrations of leucine, increasing pH, and increasing oxygen. For testing the leucine consumption rate under various substrate concentration, 0, 0.63, 1.25, 2.5, 5, 10, 15 or 20 mM of leucine was added into the reaction media. For testing the leucine consumption rate under various pH values, the reaction media was pH at 3, 4, 5, 6, or 7 when performed the assay. For testing the leucine consumption rate under various oxygen concentration, the reaction was carried out at 0, 7 or 21% oxygen by having the reaction in anaerobic chamber, microaerobic chamber or ambient.

[052] **Figure 16** depicts graphs showing the ability of engineered strain SYN5941 optimized for BCAA consumption (comprising pSC101-P_{tet}-optimized operon-brnQ-ampR) to decrease plasma BCAA levels *in vivo* in the intermediate MSUD (iMSUD) animal model fed with a high-protein diet, wherein the leucine levels were 1647 ± 265.8 μ M in iMSUD males (left panels) and 1165 ± 151.9 μ M in iMSUD females (right panels). SYN5941 was compared against engineered strain SYN1980 (comprising Δ leuE, Δ ilvC, lacZ:tetR-Ptet-livKHMGE, tetR-Ptet-leuD_H(Bc)-kivD-adh2-brnQ-rrnB ter (in low-copy pSC101 plasmid). **Figure 16A**, **Figure 16B**, and **Figure 16C** show plasma leucine, valine and isoleucine concentrations on day 5 of the study. SYN5941 lowered the levels of leucine, isoleucine, and valine more significantly in males than SYN1980. Data were collected from 11 males and 15 females and presented as mean $\pm$ SEM. Statistical significance determined as $p < 0.05$ using one-way ANOVA followed by Tukey's multiple comparison test. * $p < 0.05$.

[053] **Figure 17** depicts graphs showing the ability of engineered strain SYN5941 optimized for BCAA consumption (comprising pSC101-P_{tet}-optimized operon-brnQ-ampR) to decrease plasma BCAA levels *in vivo* in the intermediate MSUD (iMSUD) animal model fed with a high-protein diet, wherein the leucine levels were 2428 ± 485.1 μ M in iMSUD males (left panels) and 1626 ± 150.1 μ M in iMSUD females (right panels). SYN5941 was compared against engineered strain SYN1980 (comprising Δ leuE, Δ ilvC, lacZ:tetR-Ptet-livKHMGE, tetR-Ptet-leuD_H(Bc)-kivD-adh2-brnQ-rrnB ter (in low-copy pSC101 plasmid). **Figure 17A**, **Figure 17B**, and **Figure 17C** show plasma leucine, valine and isoleucine concentrations on day 5 of the study. Data were collected from 11 males and 15 females and presented as mean $\pm$ SEM. Statistical significance determined as $p < 0.05$ using one-way ANOVA followed by Tukey's multiple comparison test. * $p < 0.05$.

[054] **Figure 18** depicts graphs showing the ability of engineered strain SYN5941 optimized for BCAA consumption (comprising pSC101-P_{tet}-optimized operon-brnQ-ampR) to decrease plasma BCAA levels *in vivo* in healthy non-human primates. SYN5941 was compared against engineered strain SYN1980 (comprising Δ leuE, Δ ilvC, lacZ:tetR-Ptet-livKHMGE, tetR-Ptet-leuD_H(Bc)-kivD-adh2-brnQ-rrnB ter (in low-copy pSC101 plasmid). An orally administered vehicle (15% glycerol in PBS) served as control. **Figure 18A**, **Figure 18B**, and **Figure 18C** show plasma

leucine, valine and isoleucine concentrations as a function of time. **Figure 18D**, **Figure 18E**, and **Figure 18F** show area under the curves. SYN5941 lowered the levels of leucine more significantly than SYN1980. Data collected from n=10-12 animals and presented as mean $\pm$ SEM. Statistical significance determined as $p < 0.05$ using two-way ANOVA (top) or one-way ANOVA (bottom) followed by Tukey's multiple comparison test.

[055] **Figure 19** depicts a schematic view of the intermediate MSUD (iMSUD) animal model design.

[056] **Figure 20** depicts a graph showing a comparison of the rate of Leu consumption over time between top Leu consuming strains (5941, 5942 and 5943) and a prototype strain (1980). 8 mM leucine was added to minimum media and samples were taken at 0, 2, and 4 hour time points after anaerobic incubation.

[057] **Figure 21** depicts a graph showing extracellular profiles of the isopentanol pathway intermediates for SYN5941 assayed in Ambr15 bioreactors (n=2). Error bars reflect standard deviation across the duplicate bioreactors. The "Sum" shown in black bars represents the aggregate total concentration of the intermediates shown. Leu = Leucine, Acid = 2-oxoisocaproate, Aldehyde = isovaleraldehyde, Alcohol = isopentanol.

Detailed Description

[058] The disclosure includes optimized, engineered, and programmed microorganisms, *e.g.*, bacteria, yeast, and viruses, pharmaceutical compositions thereof, and methods of modulating and treating disorders involving the catabolism of a branched chain amino acid, such as leucine, valine, and isoleucine. In some embodiments, the microorganism, *e.g.*, bacterium, yeast, or virus, has been engineered to comprise heterologous gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s), such as a leucine catabolism enzyme.

[059] In some embodiments, the engineered microorganism, *e.g.*, engineered bacterium, comprises optimized heterologous gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) and is capable of reducing the level of one or more branched chain amino acids and/or corresponding alpha-keto acid(s) and/or other corresponding metabolite(s). For example, the engineered bacterium, may comprise a BCAA transporter, such as LivKHMgf and/or BrnQ. In some embodiments, the engineered bacterium comprises heterologous gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) and is capable of metabolizing one or more branched chain amino acids and/or corresponding alpha-keto acid(s) and/or other corresponding metabolite(s). In some embodiments, the engineered bacterium comprises heterologous gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) and is capable of transporting one or more branched chain amino acids and/or corresponding alpha-keto acid(s) and/or other corresponding metabolite(s) into the bacterium. In some embodiments, the engineered bacterium comprises heterologous gene sequence(s) encoding one or more branched chain amino acid

catabolism enzyme(s) and is capable of reducing the level of and/or metabolizing one or more branched chain amino acids and/or corresponding alpha-keto acid(s) and/or other corresponding metabolite(s) in low-oxygen environments, *e.g.*, the gut. In some embodiments, the engineered bacteria convert the branched chain amino acid(s) and/or corresponding alpha-keto acid(s) and/or other corresponding metabolite(s) to a non-toxic or low toxicity metabolite, *e.g.*, isovaleraldehyde, isobutyraldehyde, 2-methylbutyraldehyde, isovaleric acid, isobutyric acid, 2-methylbutyric acid, isopentanol, isobutanol, and 2-methylbutanol. In some embodiments, the engineered bacterium comprises a genetic modification that reduces export of a branched chain amino acid from the bacterial cell, for example, the bacterial cell may comprise a knockout or knock-down of a gene that encodes a BCAA exporter, such as *leuE* (which encodes a leucine exporter). In some embodiments, the engineered bacterium comprises gene sequence(s) or gene cassette(s) encoding one or more transporters of a branched chain amino acid, which transporters function to import one or more BCAA(s) into the bacterial cell. In some embodiments, the bacterium has been engineered to comprise a genetic modification that reduces or inhibits endogenous production of one or more branched chain amino acids and/or one or more corresponding alpha-keto acids or other metabolite(s), for example, the bacterium may comprise a knockout or knock-down of a gene that encodes a molecule required for BCAA synthesis, such as *ilvC* (which encodes keto acid reductoisomerase). In some embodiments, the bacterium has been engineered to comprise an auxotroph, including, for example, a BCAA auxotrophy, such as *ilvC* (which is required for BCAA synthesis and requires the cell to import isoleucine and valine to survive) or other auxotrophy, as provided herein and known in the art, *e.g.*, *thyA* auxotrophy. In some embodiments, the bacterium has been engineered to comprise a kill-switch, such as any of the kill-switches provided herein and known in the art. In some embodiments, the bacterium has been engineered to comprise antibiotic resistance, such as any of the antibiotic resistance provided herein and known in the art. In any of these embodiments, the gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s), transporter(s), and other molecules can be integrated into the bacterial chromosome and/or can be present on a plasmid(s) (low copy and/or high copy). In any of these embodiments, the gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s), transporter(s), and other molecules can be under the control of an inducible or constitutive promoter. Exemplary inducible promoters described herein include oxygen level-dependent promoters (*e.g.*, FNR-inducible promoter), promoters induced by inflammation or an inflammatory response (RNS, ROS promoters), and promoters induced by a metabolite that may or may not be naturally present (*e.g.*, can be exogenously added) in the gut, *e.g.*, arabinose and tetracycline.

[060] Thus, the recombinant bacterial cells and pharmaceutical compositions comprising the bacterial cells disclosed herein may be used to catabolize branched chain amino acids, *e.g.*, leucine, isoleucine, valine, and/or their corresponding alpha-keto acid counterparts, to modify, ameliorate, treat and/or prevent conditions associated with disorders involving the catabolism of a

branched chain amino acid. In one embodiment, the disorder involving the catabolism of a branched chain amino acid is a metabolic disorder involving the abnormal catabolism of a branched chain amino acid, including but not limited to maple syrup urine disease (MSUD), isovaleric acidemia, propionic acidemia, methylmalonic acidemia, or diabetes ketoacidosis, 3-MCC Deficiency, 3-Methylglutaconyl-CoA hydratase Deficiency, HMG-CoA Lyase Deficiency, Acetyl-CoA Carboxylase Deficiency, Malonyl-CoA Decarboxylase Deficiency, short-branched chain acylCoA dehydrogenase deficiency, 2-methyl-3-hydroxybutyric acidemia, beta-ketothiolase deficiency, isobutyryl-CoA dehydrogenase deficiency, HIBCH deficiency), and 3-Hydroxyisobutyric aciduria. In another embodiment, the disorder involving the catabolism of a branched chain amino acid is a disorder caused by the activation of mTor, for example, cancer, obesity, type 2 diabetes, neurodegeneration, autism, Alzheimer's disease, Lymphangioliomyomatosis (LAM), transplant rejection, glycogen storage disease, obesity, tuberous sclerosis, hypertension, cardiovascular disease, hypothalamic activation, musculoskeletal disease, Parkinson's disease, Huntington's disease, psoriasis, rheumatoid arthritis, lupus, multiple sclerosis, Leigh's syndrome, and Friedrich's ataxia.

Definitions

[061] In order that the disclosure may be more readily understood, certain terms are first defined. These definitions should be read in light of the remainder of the disclosure and as understood by a person of ordinary skill in the art. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by a person of ordinary skill in the art. Additional definitions are set forth throughout the detailed description.

[062] As used herein, the term "recombinant microorganism" refers to a microorganism, *e.g.*, bacterial, yeast or viral cell, or bacteria, yeast or virus, that has been genetically modified from its native state. Thus, a "recombinant bacterial cell" or "recombinant bacteria" refers to a bacterial cell or bacteria that have been genetically modified from their native state. For instance, a recombinant bacterial cell may have nucleotide insertions, nucleotide deletions, nucleotide rearrangements, and nucleotide modifications introduced into their DNA. These genetic modifications may be present in the chromosome of the bacteria or bacterial cell, or on a plasmid in the bacteria or bacterial cell. Recombinant bacterial cells disclosed herein may comprise exogenous nucleotide sequences on plasmids. Alternatively, recombinant bacterial cells may comprise exogenous nucleotide sequences stably incorporated into their chromosome.

[063] A "programmed or engineered microorganism" refers to a microorganism, *e.g.*, bacterial, yeast, or viral cell, or bacteria, yeast, or virus, that has been genetically modified from its native state to perform a specific function. Thus, a "programmed or engineered bacterial cell" or "programmed or engineered bacteria" refers to a bacterial cell or bacteria that has been genetically modified from its native state to perform a specific function, *e.g.*, to metabolize a branched chain amino acid. In certain embodiments, the programmed or engineered bacterial cell has been modified to express one or more proteins, for example, one or more proteins that have a therapeutic activity or

serve a therapeutic purpose. The programmed or engineered bacterial cell may additionally have the ability to stop growing or to destroy itself once the protein(s) of interest have been expressed.

[064] As used herein, the term “gene” refers to a nucleic acid fragment that encodes a protein or fragment thereof, optionally including regulatory sequences preceding (5’ non-coding sequences) and following (3’ non-coding sequences) the coding sequence. In one embodiment, a “gene” does not include regulatory sequences preceding and following the coding sequence. A “native gene” refers to a gene as found in nature, optionally with its own regulatory sequences preceding and following the coding sequence. A “chimeric gene” refers to any gene that is not a native gene, optionally comprising regulatory sequences preceding and following the coding sequence, wherein the coding sequences and/or the regulatory sequences, in whole or in part, are not found together in nature. Thus, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory and coding sequences that are derived from the same source, but arranged differently than is found in nature.

[065] As used herein, the term “gene sequence” is meant to refer to a genetic sequence, *e.g.*, a nucleic acid sequence. The gene sequence or genetic sequence is meant to include a complete gene sequence or a partial gene sequence. The gene sequence or genetic sequence is meant to include sequence that encodes a protein or polypeptide and is also meant to include genetic sequence that does not encode a protein or polypeptide, *e.g.*, a regulatory sequence, leader sequence, signal sequence, or other non-protein coding sequence.

[066] As used herein, a “heterologous” gene or “heterologous sequence” refers to a nucleotide sequence that is not normally found in a given cell in nature. As used herein, a heterologous sequence encompasses a nucleic acid sequence that is exogenously introduced into a given cell and can be a native sequence (naturally found or expressed in the cell) or non-native sequence (not naturally found or expressed in the cell) and can be a natural or wild-type sequence or a variant, non-natural, or synthetic sequence. “Heterologous gene” includes a native gene, or fragment thereof, that has been introduced into the host cell in a form that is different from the corresponding native gene. For example, a heterologous gene may include a native coding sequence that is a portion of a chimeric gene to include non-native regulatory regions that is reintroduced into the host cell. A heterologous gene may also include a native gene, or fragment thereof, introduced into a non-native host cell. Thus, a heterologous gene may be foreign or native to the recipient cell; a nucleic acid sequence that is naturally found in a given cell but expresses an unnatural amount of the nucleic acid and/or the polypeptide which it encodes; and/or two or more nucleic acid sequences that are not found in the same relationship to each other in nature. As used herein, the term “endogenous gene” refers to a native gene in its natural location in the genome of an organism. As used herein, the term “transgene” refers to a gene that has been introduced into the host organism, *e.g.*, host bacterial cell, genome.

[067] As used herein, a “non-native” nucleic acid sequence refers to a nucleic acid sequence not normally present in a microorganism, *e.g.*, an extra copy of an endogenous sequence, or a heterologous sequence such as a sequence from a different species, strain, or substrain of bacteria, yeast, or virus, or a sequence that is modified and/or mutated as compared to the unmodified sequence from bacteria, yeast, or virus of the same subtype. In some embodiments, the non-native nucleic acid sequence is a synthetic, non-naturally occurring sequence (see, *e.g.*, Purcell *et al.*, 2013). The non-native nucleic acid sequence may be a regulatory region, a promoter, a gene, and/or one or more genes in gene cassette. In some embodiments, “non-native” refers to two or more nucleic acid sequences that are not found in the same relationship to each other in nature. The non-native nucleic acid sequence may be present on a plasmid or chromosome. In some embodiments, the genetically engineered microorganism of the disclosure comprises a gene that is operably linked to a promoter that is not associated with said gene in nature. For example, in some embodiments, the genetically engineered bacteria disclosed herein comprise a gene that is operably linked to a directly or indirectly inducible promoter that is not associated with said gene in nature, *e.g.*, an FNR responsive promoter (or other promoter disclosed herein) operably linked to a gene encoding a branched chain amino acid catabolism enzyme. In some embodiments, the genetically engineered yeast or virus of the disclosure comprises a gene that is operably linked to a directly or indirectly inducible promoter that is not associated with said gene in nature, *e.g.*, a promoter operably linked to a gene encoding a branched chain amino acid catabolism enzyme.

[068] As used herein, the term “coding region” refers to a nucleotide sequence that codes for a specific amino acid sequence. The term “regulatory sequence” refers to a nucleotide sequence located upstream (5′ non-coding sequences), within, or downstream (3′ non-coding sequences) of a coding sequence, and which influences the transcription, RNA processing, RNA stability, or translation of the associated coding sequence. Examples of regulatory sequences include, but are not limited to, promoters, translation leader sequences, effector binding sites, signal sequences, and stem-loop structures. In one embodiment, the regulatory sequence comprises a promoter, *e.g.*, an FNR responsive promoter or another promoter disclosed herein.

[069] “Operably linked” refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. A regulatory element is operably linked with a coding sequence when it is capable of affecting the expression of the gene coding sequence, regardless of the distance between the regulatory element and the coding sequence. More specifically, operably linked refers to a nucleic acid sequence, *e.g.*, a gene encoding a branched chain amino acid catabolism enzyme, that is joined to a regulatory sequence in a manner which allows expression of the nucleic acid sequence, *e.g.*, the gene encoding the branched chain amino acid catabolism enzyme. In other words, the regulatory sequence acts in *cis*. In one embodiment, a gene may be “directly linked” to a regulatory sequence in a manner which allows expression of the gene. In another embodiment, a gene may be “indirectly linked” to a regulatory sequence in a manner which

allows expression of the gene. In one embodiment, two or more genes may be directly or indirectly linked to a regulatory sequence in a manner which allows expression of the two or more genes.

A regulatory region or sequence is a nucleic acid that can direct transcription of a gene of interest and may comprise promoter sequences, enhancer sequences, response elements, protein recognition sites, inducible elements, promoter control elements, protein binding sequences, 5' and 3' untranslated regions, transcriptional start sites, termination sequences, polyadenylation sequences, and introns.

[070] A “promoter” as used herein, refers to a nucleotide sequence that is capable of controlling the expression of a coding sequence or gene. Promoters are generally located 5' of the sequence that they regulate. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from promoters found in nature, and/or comprise synthetic nucleotide segments. Those skilled in the art will readily ascertain that different promoters may regulate expression of a coding sequence or gene in response to a particular stimulus, *e.g.*, in a cell- or tissue-specific manner, in response to different environmental or physiological conditions, or in response to specific compounds. Prokaryotic promoters are typically classified into two classes: inducible and constitutive. A “constitutive promoter” refers to a promoter that allows for continual transcription of the coding sequence or gene under its control.

[071] “Constitutive promoter” refers to a promoter that is capable of facilitating continuous transcription of a coding sequence or gene under its control and/or to which it is operably linked. Constitutive promoters and variants are well known in the art and include, but are not limited to, P_{tac} promoter, BBa_J23100, a constitutive *Escherichia coli* σ^S promoter (*e.g.*, an osmY promoter (International Genetically Engineered Machine (iGEM) Registry of Standard Biological Parts Name BBa_J45992; BBa_J45993)), a constitutive *Escherichia coli* σ^{32} promoter (*e.g.*, htpG heat shock promoter (BBa_J45504)), a constitutive *Escherichia coli* σ^{70} promoter (*e.g.*, lacq promoter (BBa_J54200; BBa_J56015), *E. coli* CreABCD phosphate sensing operon promoter (BBa_J64951), GlnRS promoter (BBa_K088007), lacZ promoter (BBa_K119000; BBa_K119001); M13K07 gene I promoter (BBa_M13101); M13K07 gene II promoter (BBa_M13102), M13K07 gene III promoter (BBa_M13103), M13K07 gene IV promoter (BBa_M13104), M13K07 gene V promoter (BBa_M13105), M13K07 gene VI promoter (BBa_M13106), M13K07 gene VIII promoter (BBa_M13108), M13110 (BBa_M13110)), a constitutive *Bacillus subtilis* σ^A promoter (*e.g.*, promoter veg (BBa_K143013), promoter 43 (BBa_K143013), P_{liaG} (BBa_K823000), P_{lepA} (BBa_K823002), P_{veg} (BBa_K823003)), a constitutive *Bacillus subtilis* σ^B promoter (*e.g.*, promoter ctc (BBa_K143010), promoter gsiB (BBa_K143011)), a *Salmonella* promoter (*e.g.*, Pspv2 from *Salmonella* (BBa_K112706), Pspv from *Salmonella* (BBa_K112707)), a bacteriophage T7 promoter (*e.g.*, T7 promoter (BBa_I712074; BBa_I719005; BBa_J34814; BBa_J64997; BBa_K113010; BBa_K113011; BBa_K113012; BBa_R0085; BBa_R0180; BBa_R0181; BBa_R0182; BBa_R0183; BBa_Z0251; BBa_Z0252; BBa_Z0253)), and a bacteriophage SP6 promoter (*e.g.*, SP6 promoter (BBa_J64998)).

[072] An “inducible promoter” refers to a regulatory region that is operably linked to one or more genes, wherein expression of the gene(s) is increased in the presence of an inducer of said regulatory region. An “inducible promoter” refers to a promoter that initiates increased levels of transcription of the coding sequence or gene under its control in response to a stimulus or an exogenous environmental condition. A “directly inducible promoter” refers to a regulatory region, wherein the regulatory region is operably linked to a gene encoding a protein or polypeptide, where, in the presence of an inducer of said regulatory region, the protein or polypeptide is expressed. An “indirectly inducible promoter” refers to a regulatory system comprising two or more regulatory regions, for example, a first regulatory region that is operably linked to a first gene encoding a first protein, polypeptide, or factor, *e.g.*, a transcriptional regulator, which is capable of regulating a second regulatory region that is operably linked to a second gene, the second regulatory region may be activated or repressed, thereby activating or repressing expression of the second gene. Both a directly inducible promoter and an indirectly inducible promoter are encompassed by “inducible promoter.” Exemplary inducible promoters described herein include oxygen level-dependent promoters (*e.g.*, FNR-inducible promoter), promoters induced by inflammation or an inflammatory response (RNS, ROS promoters), and promoters induced by a metabolite that may or may not be naturally present (*e.g.*, can be exogenously added) in the gut, *e.g.*, arabinose and tetracycline. Examples of inducible promoters include, but are not limited to, an FNR responsive promoter, a P_{araC} promoter, a P_{araBAD} promoter, and a P_{TetR} promoter, each of which are described in more detail herein. Examples of other inducible promoters are provided herein below.

[073] As used herein, “stably maintained” or “stable” bacterium is used to refer to a bacterial host cell carrying non-native genetic material, *e.g.*, a gene encoding a branched chain amino acid catabolism enzyme, which is incorporated into the host genome or propagated on a self-replicating extra-chromosomal plasmid, such that the non-native genetic material is retained, expressed, and propagated. The stable bacterium is capable of survival and/or growth *in vitro*, *e.g.*, in medium, and/or *in vivo*, *e.g.*, in the gut. For example, the stable bacterium may be a genetically engineered bacterium comprising a gene encoding a branched chain amino acid catabolism enzyme, in which the plasmid or chromosome carrying the gene is stably maintained in the bacterium, such that branched chain amino acid catabolism enzyme can be expressed in the bacterium, and the bacterium is capable of survival and/or growth *in vitro* and/or *in vivo*. In some embodiments, copy number affects the stability of expression of the non-native genetic material. In some embodiments, copy number affects the level of expression of the non-native genetic material.

[074] As used herein, the term “expression” refers to the transcription and stable accumulation of sense (mRNA) or anti-sense RNA derived from a nucleic acid, and/or to translation of an mRNA into a polypeptide.

[075] As used herein, the term “plasmid” or “vector” refers to an extrachromosomal nucleic acid, *e.g.*, DNA, construct that is not integrated into a bacterial cell’s genome. Plasmids are usually

circular and capable of autonomous replication. Plasmids may be low-copy, medium-copy, or high-copy, as is well known in the art. Plasmids may optionally comprise a selectable marker, such as an antibiotic resistance gene, which helps select for bacterial cells containing the plasmid and which ensures that the plasmid is retained in the bacterial cell. A plasmid disclosed herein may comprise a nucleic acid sequence encoding a heterologous gene, *e.g.*, a gene encoding a branched chain amino acid catabolism enzyme.

[076] As used herein, the term “transform” or “transformation” refers to the transfer of a nucleic acid fragment into a host bacterial cell, resulting in genetically-stable inheritance. Host bacterial cells comprising the transformed nucleic acid fragment are referred to as “recombinant” or “transgenic” or “transformed” organisms.

[077] The term “genetic modification,” as used herein, refers to any genetic change. Exemplary genetic modifications include those that increase, decrease, or abolish the expression of a gene, including, for example, modifications of native chromosomal or extrachromosomal genetic material. Exemplary genetic modifications also include the introduction of at least one plasmid, modification, mutation, base deletion, base addition, base substitution, and/or codon modification of chromosomal or extrachromosomal genetic sequence(s), gene over-expression, gene amplification, gene suppression, promoter modification or substitution, gene addition (either single or multi-copy), antisense expression or suppression, or any other change to the genetic elements of a host cell, whether the change produces a change in phenotype or not. Genetic modification can include the introduction of a plasmid, *e.g.*, a plasmid comprising a branched chain amino acid catabolism enzyme operably linked to a promoter, into a bacterial cell. Genetic modification can also involve a targeted replacement in the chromosome, *e.g.*, to replace a native gene promoter with an inducible promoter, regulated promoter, strong promoter, or constitutive promoter. Genetic modification can also involve gene amplification, *e.g.*, introduction of at least one additional copy of a native gene into the chromosome of the cell. Alternatively, chromosomal genetic modification can involve a genetic mutation.

[078] As used herein, the term “genetic mutation” refers to a change or changes in a nucleotide sequence of a gene or related regulatory region that alters the nucleotide sequence as compared to its native or wild-type sequence. Mutations include, for example, substitutions, additions, and deletions, in whole or in part, within the wild-type sequence. Such substitutions, additions, or deletions can be single nucleotide changes (*e.g.*, one or more point mutations), or can be two or more nucleotide changes, which may result in substantial changes to the sequence. Mutations can occur within the coding region of the gene as well as within the non-coding and regulatory sequence of the gene. The term “genetic mutation” is intended to include silent and conservative mutations within a coding region as well as changes which alter the amino acid sequence of the polypeptide encoded by the gene. A genetic mutation in a gene coding sequence may, for example, increase, decrease, or otherwise alter the activity (*e.g.*, enzymatic activity) of the gene’s polypeptide

product. A genetic mutation in a regulatory sequence may increase, decrease, or otherwise alter the expression of sequences operably linked to the altered regulatory sequence.

[079] Specifically, the term “genetic modification that reduces export of a branched chain amino acid from the bacterial cell” refers to a genetic modification that reduces the rate of export or quantity of export of a branched chain amino acid from the bacterial cell, as compared to the rate of export or quantity of export of the branched chain amino acid from a bacterial cell not having said modification, *e.g.*, a wild-type bacterial cell. In one embodiment, a recombinant bacterial cell having a genetic modification that reduces export of a branched chain amino acid from the bacterial cell comprises a genetic mutation in a native gene, *e.g.*, a *leuE* gene. In another embodiment, a recombinant bacterial cell having a genetic modification that reduces export of a branched chain amino acid from the bacterial cell comprises a genetic mutation in a native promoter, *e.g.*, a *leuE* promoter, which reduces or inhibits transcription of the *leuE* gene. In another embodiment, a recombinant bacterial cell having a genetic modification that reduces export of a branched chain amino acid from the bacterial cell comprises a genetic mutation leading to overexpression of a repressor of an exporter of a branched chain amino acid. In another embodiment, a recombinant bacterial cell having a genetic modification that reduces export of a branched chain amino acid from the bacterial cell comprises a genetic mutation which reduces or inhibits translation of the gene encoding the exporter, *e.g.*, the *leuE* gene.

[080] Moreover, the term “genetic modification that increases import of a branched chain amino acid into the bacterial cell” refers to a genetic modification that increases the uptake rate or increases the uptake quantity of a branched chain amino acid into the cytosol of the bacterial cell, as compared to the uptake rate or uptake quantity of the branched chain amino acid into the cytosol of a bacterial cell not having said modification, *e.g.*, a wild-type bacterial cell. In some embodiments, an engineered bacterial cell having a genetic modification that increases import of a branched chain amino acid into the bacterial cell refers to a bacterial cell comprising heterologous gene sequence (native or non-native) encoding one or more importer(s) (transporter(s)) of a branched chain amino acid. In some embodiments, the genetically engineered bacteria comprising genetic modification that increases import of a branched chain amino acid into the bacterial cell comprise gene sequence(s) encoding a BCAA transporter or other amino acid transporter that transports one or more BCAA(s) into the bacterial cell, for example a transporter that is capable of transporting leucine, valine, and/or isoleucine into a bacterial cell. The transporter can be any transporter that assists or allows import of a BCAA into the cell. In certain embodiments, the BCAA transporter is a leucine transporter, *e.g.*, a high-affinity leucine transporter, *e.g.*, LivKHMGE. In certain embodiments, the engineered bacterial cell contains gene sequence of one or more of *livK*, *livH*, *livM*, *livG*, and *livF* genes. In certain embodiments, the BCAA transporter is a BCAA transporter, *e.g.*, a low affinity BCAA transporter, *e.g.*, BmQ. In certain embodiments, the engineered bacterial cell contains gene sequence encoding *brnQ* gene. In some embodiments, the engineered bacteria comprise more than one copy of gene

sequence encoding a BCAA transporter. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding more than one BCAA transporter, *e.g.*, two or more different BCAA transporters.

[081] As used herein, the term “transporter” is meant to refer to a mechanism, *e.g.*, protein, proteins, or protein complex, for importing a molecule, *e.g.*, amino acid, peptide (di-peptide, tri-peptide, polypeptide, etc.), toxin, metabolite, substrate, as well as other biomolecules into the microorganism from the extracellular milieu.

[082] As used herein, the term “polypeptide of interest” or “polypeptides of interest”, “protein of interest”, “proteins of interest”, “payload”, “payloads” includes, but is not limited to, any or a plurality of any of the branched chain amino acid catabolism enzymes, and/or branched chain amino acid transporters, and/or branched chain amino acid binding proteins and/or branched chain amino acid exporters described herein. As used herein, the term “gene of interest” or “gene sequence of interest” includes any or a plurality of any of the gene(s) an/or gene sequence(s) and or gene cassette(s) encoding one or more branched chain amino acid catabolism enzymes, branched chain amino acid transporters, branched chain amino acid binding proteins, and branched chain amino acid exporters described herein.

[083] The term “branched chain amino acid” or “BCAA,” as used herein, refers to an amino acid which comprises a branched side chain. Leucine, isoleucine, and valine are naturally occurring amino acids comprising a branched side chain. However, non-naturally occurring, usual, and/or modified amino acids comprising a branched side chain are also encompassed by the term branched chain amino acid.

[084] Conversion of a branched chain amino acid to its corresponding alpha-keto acid is the first step in branched chain amino acid catabolism and is reversible when catalyzed by a leucine dehydrogenase (leuDh) or a branched chain amino acid aminotransferase (ilvE), or irreversible when catalyzed by an amino acid oxidase (also referred as amino acid deaminase, *e.g.*, L-AAD). As used herein, the terms “alpha-keto acid” or “ α -keto acid” refers to the molecules which are produced after deamination of a branched chain amino acid, and include the naturally occurring alpha-keto acids: α -ketoisocaproic acid (KIC) (also known as 4-methyl-2-oxopentanoate), α -ketoisovaleric acid (KIV) (also known as 2-oxoisopentanoate), and α -keto-beta-methylvaleric acid (KMV) (also known as 3-methyl-2-oxopentanoate). However, non-naturally occurring, unusual, or modified alpha-keto acids are also encompassed by the term “alpha-keto acid.”

[085] Conversion of a branched chain alpha-keto acid to its corresponding branched acyl-CoA derivative is the second step in branched chain amino acid catabolism and is irreversible. As used herein, the term “acyl-CoA derivative” refers to the molecules which are produced after dehydrogenation of a branched chain alpha-keto acid, and include the naturally occurring acyl-CoA derivatives, propionyl-CoA and acetyl-CoA. However, non-naturally occurring, unusual, or modified acyl-CoA derivatives are also encompassed by the term “acyl-CoA derivative.”

[086] In an alternative BCAA catabolism pathway, known as the “Ehrlich pathway”, a branched chain alpha-keto acid is irreversibly decarboxylated to its corresponding branched chain amino acid-derived aldehyde. This irreversible catabolic conversion of a branched chain alpha-keto acid, *e.g.*, alpha-keto acids α -ketoisocaproic acid (KIC), α -ketoisovaleric acid (KIV), or α -keto-beta-methylvaleric acid (KMV) to its corresponding branched chain amino acid-derived aldehyde, *e.g.*, isovaleraldehyde, isobutyraldehyde, or 2-methylbutyraldehyde, is catalyzed by ketoacid decarboxylase (KivD). As used herein, the term “branched chain amino acid-derived aldehyde” refers to the molecules which are produced after decarboxylation of a branched chain alpha-keto acid, and include the naturally occurring aldehydes isovaleraldehyde, 2-methyl butyraldehyde, and isobutyraldehyde. However, non-naturally occurring, unusual, or modified aldehydes are also encompassed by the term “aldehyde.” BCAA-derived aldehydes can then be converted to alcohols (*e.g.*, isopentanol, isobutanol, 2-methylbutanol) by an alcohol dehydrogenase, *e.g.*, Adh2 or Ygh.D. Alternatively, BCAA-derived aldehydes can be converted to their respective carboxylic acids (*e.g.*, isovalerate, isobutyrate, and 2-methylbutyrate) by an aldehyde dehydrogenase, *e.g.*, PadA.

[087] As used herein, the phrase “exogenous environmental condition” or “exogenous environment signal” refers to settings, circumstances, stimuli, or biological molecules under which a promoter described herein is directly or indirectly induced. The phrase “exogenous environmental conditions” is meant to refer to the environmental conditions external to the engineered microorganism, but endogenous or native to the host subject environment. Thus, “exogenous” and “endogenous” may be used interchangeably to refer to environmental conditions in which the environmental conditions are endogenous to a mammalian body, but external or exogenous to an intact microorganism cell. In some embodiments, the exogenous environmental conditions are specific to the gut of a mammal. In some embodiments, the exogenous environmental conditions are specific to the upper gastrointestinal tract of a mammal. In some embodiments, the exogenous environmental conditions are specific to the lower gastrointestinal tract of a mammal. In some embodiments, the exogenous environmental conditions are specific to the small intestine of a mammal. In some embodiments, the exogenous environmental conditions are low-oxygen, microaerobic, or anaerobic conditions, such as the environment of the mammalian gut. In some embodiments, exogenous environmental conditions are molecules or metabolites that are specific to the mammalian gut, *e.g.*, propionate. In some embodiments, the exogenous environmental condition is a tissue-specific or disease-specific metabolite or molecule(s). In some embodiments, the exogenous environmental condition is specific to a branched chain amino acid catabolic enzyme disease, *e.g.*, MSUD. In some embodiments, the exogenous environmental condition is a low-pH environment. In some embodiments, the genetically engineered microorganism of the disclosure comprises a pH-dependent promoter. In some embodiments, the genetically engineered microorganism of the disclosure comprise an oxygen level-dependent promoter. In some aspects, bacteria have evolved transcription factors that are capable of sensing oxygen levels. Different signaling pathways may be

triggered by different oxygen levels and occur with different kinetics. An “oxygen level-dependent promoter” or “oxygen level-dependent regulatory region” refers to a nucleic acid sequence to which one or more oxygen level-sensing transcription factors is capable of binding, wherein the binding and/or activation of the corresponding transcription factor activates downstream gene expression.

[088] Examples of oxygen level-dependent transcription factors include, but are not limited to, FNR (fumarate and nitrate reductase), ANR, and DNR. Corresponding FNR-responsive promoters, ANR (anaerobic nitrate respiration)-responsive promoters, and DNR (dissimilatory nitrate respiration regulator)-responsive promoters are known in the art (see, *e.g.*, Castiglione *et al.*, 2009; Eiglmeier *et al.*, 1989; Galimand *et al.*, 1991; Hasegawa *et al.*, 1998; Hoeren *et al.*, 1993; Salmon *et al.*, 2003), and non-limiting examples are shown in Table 1.

[089] In a non-limiting example, a promoter (PfnrS) was derived from the *E. coli* Nissle fumarate and nitrate reductase gene S (fnrS) that is known to be highly expressed under conditions of low or no environmental oxygen (Durand and Storz, 2010; Boysen *et al.*, 2010). The PfnrS promoter is activated under anaerobic conditions by the global transcriptional regulator FNR that is naturally found in Nissle. Under anaerobic conditions, FNR forms a dimer and binds to specific sequences in the promoters of specific genes under its control, thereby activating their expression. However, under aerobic conditions, oxygen reacts with iron-sulfur clusters in FNR dimers and converts them to an inactive form. In this way, the PfnrS inducible promoter is adopted to modulate the expression of proteins or RNA. PfnrS is used interchangeably in this application as FNRS, fnrs, FNR, P-FNRS promoter and other such related designations to indicate the promoter PfnrS.

Table 1. Examples of transcription factors and responsive genes and regulatory regions

Transcription Factor	Examples of responsive genes, promoters, and/or regulatory regions:
FNR	<i>nirB</i> , <i>ydfZ</i> , <i>pdhR</i> , <i>focA</i> , <i>ndH</i> , <i>hlyE</i> , <i>narK</i> , <i>narX</i> , <i>narG</i> , <i>yfiD</i> , <i>tdcD</i>
ANR	<i>arcDABC</i>
DNR	<i>norB</i> , <i>norC</i>

[090] In some embodiments, the exogenous environmental conditions are the presence or absence of reactive oxygen species (ROS). In other embodiments, the exogenous environmental conditions are the presence or absence of reactive nitrogen species (RNS). In some embodiments, exogenous environmental conditions are biological molecules that are involved in the inflammatory response, for example, molecules present in an inflammatory disorder of the gut. In some embodiments, the exogenous environmental conditions or signals exist naturally or are naturally absent in the environment in which the recombinant bacterial cell resides. In some embodiments, the exogenous environmental conditions or signals are artificially created, for example, by the creation or removal of biological conditions and/or the administration or removal of biological molecules.

[091] In some embodiments, the exogenous environmental condition(s) and/or signal(s) stimulates the activity of an inducible promoter. In some embodiments, the exogenous environmental condition(s) and/or signal(s) that serves to activate the inducible promoter is not naturally present within the gut of a mammal. In some embodiments, the inducible promoter is stimulated by a molecule or metabolite that is administered in combination with the pharmaceutical composition of the disclosure, for example, tetracycline, arabinose, or any biological molecule that serves to activate an inducible promoter. In some embodiments, the exogenous environmental condition(s) and/or signal(s) is added to culture media comprising a recombinant bacterial cell of the disclosure. In some embodiments, the exogenous environmental condition that serves to activate the inducible promoter is naturally present within the gut of a mammal (for example, low oxygen or anaerobic conditions, or biological molecules involved in an inflammatory response). In some embodiments, the loss of exposure to an exogenous environmental condition (for example, *in vivo*) inhibits the activity of an inducible promoter, as the exogenous environmental condition is not present to induce the promoter (for example, an aerobic environment outside the gut). “Gut” refers to the organs, glands, tracts, and systems that are responsible for the transfer and digestion of food, absorption of nutrients, and excretion of waste. In humans, the gut comprises the gastrointestinal (GI) tract, which starts at the mouth and ends at the anus, and additionally comprises the esophagus, stomach, small intestine, and large intestine. The gut also comprises accessory organs and glands, such as the spleen, liver, gallbladder, and pancreas. The upper gastrointestinal tract comprises the esophagus, stomach, and duodenum of the small intestine. The lower gastrointestinal tract comprises the remainder of the small intestine, *i.e.*, the jejunum and ileum, and all of the large intestine, *i.e.*, the cecum, colon, rectum, and anal canal. Bacteria can be found throughout the gut, *e.g.*, in the gastrointestinal tract, and particularly in the intestines.

[092] As used herein, the term “low oxygen” is meant to refer to a level, amount, or concentration of oxygen (O₂) that is lower than the level, amount, or concentration of oxygen that is present in the atmosphere (*e.g.*, <21% O₂; <160 torr O₂). Thus, the term “low oxygen condition or conditions” or “low oxygen environment” refers to conditions or environments containing lower levels of oxygen than are present in the atmosphere. In some embodiments, the term “low oxygen” is meant to refer to the level, amount, or concentration of oxygen (O₂) found in a mammalian gut, *e.g.*, lumen, stomach, small intestine, duodenum, jejunum, ileum, large intestine, cecum, colon, distal sigmoid colon, rectum, and anal canal. In some embodiments, the term “low oxygen” is meant to refer to a level, amount, or concentration of O₂ that is 0-60 mmHg O₂ (0-60 torr O₂) (*e.g.*, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, and 60 mmHg O₂), including any and all incremental fraction(s) thereof (*e.g.*, 0.2 mmHg O₂, 0.5 mmHg O₂, 0.75 mmHg O₂, 1.25 mmHg O₂, 2.175 mmHg O₂, 3.45 mmHg O₂, 3.75 mmHg O₂, 4.5 mmHg O₂, 6.8 mmHg O₂, 11.35 mmHg O₂, 46.3 mmHg O₂, 58.75 mmHg, etc., which exemplary fractions are listed

here for illustrative purposes and not meant to be limiting in any way). In some embodiments, “low oxygen” refers to about 60 mmHg O₂ or less (*e.g.*, 0 to about 60 mmHg O₂). The term “low oxygen” may also refer to a range of O₂ levels, amounts, or concentrations between 0-60 mmHg O₂ (inclusive), *e.g.*, 0-5 mmHg O₂, < 1.5 mmHg O₂, 6-10 mmHg, < 8 mmHg, 47-60 mmHg, etc. which listed exemplary ranges are listed here for illustrative purposes and not meant to be limiting in any way. See, for example, Albenberg *et al.*, *Gastroenterology*, 147(5): 1055-1063 (2014); Bergofsky *et al.*, *J Clin. Invest.*, 41(11): 1971- 1980 (1962); Crompton *et al.*, *J Exp. Biol.*, 43: 473-478 (1965); He *et al.*, *PNAS (USA)*, 96: 4586-4591 (1999); McKeown, *Br. J. Radiol.*, 87:20130676 (2014) (doi: 10.1259/brj.20130676), each of which discusses the oxygen levels found in the mammalian gut of various species and each of which are incorporated by reference herewith in their entireties. In some embodiments, the term “low oxygen” is meant to refer to the level, amount, or concentration of oxygen (O₂) found in a mammalian organ or tissue other than the gut, *e.g.*, urogenital tract, tumor tissue, etc. in which oxygen is present at a reduced level, *e.g.*, at a hypoxic or anoxic level. In some embodiments, “low oxygen” is meant to refer to the level, amount, or concentration of oxygen (O₂) present in partially aerobic, semi aerobic, microaerobic, nanoaerobic, microoxic, hypoxic, anoxic, and/or anaerobic conditions. For example, Table 2 summarizes the amount of oxygen present in various organs and tissues. In some embodiments, the level, amount, or concentration of oxygen (O₂) is expressed as the amount of dissolved oxygen (“DO”) which refers to the level of free, non-compound oxygen (O₂) present in liquids and is typically reported in milligrams per liter (mg/L), parts per million (ppm; 1mg/L = 1 ppm), or in micromoles (umole) (1 umole O₂ = 0.022391 mg/L O₂). Fondriest Environmental, Inc., “Dissolved Oxygen”, *Fundamentals of Environmental Measurements*, 19 Nov 2013. In some embodiments, the term “low oxygen” is meant to refer to a level, amount, or concentration of oxygen (O₂) that is about 6.0 mg/L DO or less, *e.g.*, 6.0 mg/L, 5.0 mg/L, 4.0 mg/L, 3.0 mg/L, 2.0 mg/L, 1.0 mg/L, or 0 mg/L, and any fraction therein, *e.g.*, 3.25 mg/L, 2.5 mg/L, 1.75 mg/L, 1.5 mg/L, 1.25 mg/L, 0.9 mg/L, 0.8 mg/L, 0.7 mg/L, 0.6 mg/L, 0.5 mg/L, 0.4 mg/L, 0.3 mg/L, 0.2 mg/L and 0.1 mg/L DO, which exemplary fractions are listed here for illustrative purposes and not meant to be limiting in any way. The level of oxygen in a liquid or solution may also be reported as a percentage of air saturation or as a percentage of oxygen saturation (the ratio of the concentration of dissolved oxygen (O₂) in the solution to the maximum amount of oxygen that will dissolve in the solution at a certain temperature, pressure, and salinity under stable equilibrium). Well-aerated solutions (*e.g.*, solutions subjected to mixing and/or stirring) without oxygen producers or consumers are 100% air saturated. In some embodiments, the term “low oxygen” is meant to refer to 40% air saturation or less, *e.g.*, 40%, 39%, 38%, 37%, 36%, 35%, 34%, 33%, 32%, 31%, 30%, 29%, 28%, 27%, 26%, 25%, 24%, 23%, 22%, 21%, 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, and 0% air saturation, including any and all incremental fraction(s) thereof (*e.g.*, 30.25%, 22.70%, 15.5%, 7.7%, 5.0%, 2.8%, 2.0%, 1.65%, 1.0%, 0.9%, 0.8%, 0.75%, 0.68%, 0.5%, 0.44%, 0.3%, 0.25%, 0.2%, 0.1%, 0.08%, 0.075%, 0.058%, 0.04%.

0.032%, 0.025%, 0.01%, etc.) and any range of air saturation levels between 0-40%, inclusive (*e.g.*, 0-5%, 0.05 – 0.1%, 0.1-0.2%, 0.1-0.5%, 0.5 – 2.0%, 0-10%, 5-10%, 10-15%, 15-20%, 20-25%, 25-30%, etc.). The exemplary fractions and ranges listed here are for illustrative purposes and not meant to be limiting in any way. In some embodiments, the term “low oxygen” is meant to refer to 9% O₂ saturation or less, *e.g.*, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0%, O₂ saturation, including any and all incremental fraction(s) thereof (*e.g.*, 6.5%, 5.0%, 2.2%, 1.7%, 1.4%, 0.9%, 0.8%, 0.75%, 0.68%, 0.5%, 0.44%, 0.3%, 0.25%, 0.2%, 0.1%, 0.08%, 0.075%, 0.058%, 0.04%, 0.032%, 0.025%, 0.01%, etc.) and any range of O₂ saturation levels between 0-9%, inclusive (*e.g.*, 0-5%, 0.05 – 0.1%, 0.1-0.2%, 0.1-0.5%, 0.5 – 2.0%, 0-8%, 5-7%, 0.3-4.2% O₂, etc.). The exemplary fractions and ranges listed here are for illustrative purposes and not meant to be limiting in any way.

Table 2.

Compartment	Oxygen Tension
stomach	~60 torr (<i>e.g.</i> , 58 +/- 15 torr)
duodenum and first part of jejunum	~30 torr (<i>e.g.</i> , 32 +/- 8 torr); ~20% oxygen in ambient air
Ileum (mid- small intestine)	~10 torr; ~6% oxygen in ambient air (<i>e.g.</i> , 11 +/- 3 torr)
Distal sigmoid colon	~ 3 torr (<i>e.g.</i> , 3 +/- 1 torr)
colon	<2torr
Lumen of cecum	<1 torr
tumor	<32 torr (most tumors are <15 torr)

[093] “Microorganism” refers to an organism or microbe of microscopic, submicroscopic, or ultramicroscopic size that typically consists of a single cell. Examples of microorganisms include bacteria, viruses, parasites, fungi, certain algae, yeast, and protozoa. In some aspects, the microorganism is engineered (“engineered microorganism”) to produce one or more therapeutic molecules, *e.g.*, lysosomal enzyme(s). In certain embodiments, the engineered microorganism is an engineered bacterium. In certain embodiments, the engineered microorganism is an engineered yeast or virus.

[094] “Non-pathogenic bacteria” refer to bacteria that are not capable of causing disease or harmful responses in a host. In some embodiments, non-pathogenic bacteria are Gram-negative bacteria. In some embodiments, non-pathogenic bacteria are Gram-positive bacteria. In some embodiments, non-pathogenic bacteria do not contain lipopolysaccharides (LPS). In some embodiments, non-pathogenic bacteria are commensal bacteria. Examples of non-pathogenic bacteria include, but are not limited to certain strains belonging to the genus *Bacillus*, *Bacteroides*, *Bifidobacterium*, *Brevibacteria*, *Clostridium*, *Enterococcus*, *Escherichia coli*, *Lactobacillus*, *Lactococcus*, *Saccharomyces*, and *Staphylococcus*, *e.g.*, *Bacillus coagulans*, *Bacillus subtilis*,

Bacteroides fragilis, *Bacteroides subtilis*, *Bacteroides thetaiotaomicron*, *Bifidobacterium bifidum*, *Bifidobacterium infantis*, *Bifidobacterium lactis*, *Bifidobacterium longum*, *Clostridium butyricum*, *Enterococcus faecium*, *Escherichia coli*, *Escherichia coli* Nissle, *Lactobacillus acidophilus*, *Lactobacillus bulgaricus*, *Lactobacillus casei*, *Lactobacillus johnsonii*, *Lactobacillus paracasei*, *Lactobacillus plantarum*, *Lactobacillus reuteri*, *Lactobacillus rhamnosus*, *Lactococcus lactis* and *Saccharomyces boulardii* (Sonnenborn *et al.*, 2009; Dinleyici *et al.*, 2014; U.S. Patent No. 6,835,376; U.S. Patent No. 6,203,797; U.S. Patent No. 5,589,168; U.S. Patent No. 7,731,976). Non-pathogenic bacteria also include commensal bacteria, which are present in the indigenous microbiota of the gut. In one embodiment, the disclosure further includes non-pathogenic *Saccharomyces*, such as *Saccharomyces boulardii*. Naturally pathogenic bacteria may be genetically engineered to reduce or eliminate pathogenicity.

[095] “Probiotic” is used to refer to live, non-pathogenic microorganisms, *e.g.*, bacteria, which can confer health benefits to a host organism that contains an appropriate amount of the microorganism. In some embodiments, the host organism is a mammal. In some embodiments, the host organism is a human. In some embodiments, the probiotic bacteria are Gram-negative bacteria. In some embodiments, the probiotic bacteria are Gram-positive bacteria. Some species, strains, and/or subtypes of non-pathogenic bacteria are currently recognized as probiotic bacteria. Examples of probiotic bacteria include, but are not limited to, certain strains belonging to the genus *Bifidobacteria*, *Escherichia Coli*, *Lactobacillus*, and *Saccharomyces e.g.*, *Bifidobacterium bifidum*, *Enterococcus faecium*, *Escherichia coli* strain Nissle, *Lactobacillus acidophilus*, *Lactobacillus bulgaricus*, *Lactobacillus paracasei*, and *Lactobacillus plantarum*, and *Saccharomyces boulardii* (Dinleyici *et al.*, 2014; U.S. Patent No. 5,589,168; U.S. Patent No. 6,203,797; U.S. Patent 6,835,376). The probiotic may be a variant or a mutant strain of bacterium (Arthur *et al.*, 2012; Cuevas-Ramos *et al.*, 2010; Olier *et al.*, 2012; Nougayrede *et al.*, 2006). Non-pathogenic bacteria may be genetically engineered to enhance or improve desired biological properties, *e.g.*, survivability. Non-pathogenic bacteria may be genetically engineered to provide probiotic properties. Probiotic bacteria may be genetically engineered to enhance or improve probiotic properties.

[096] As used herein, the term “auxotroph” or “auxotrophic” refers to an organism that requires a specific factor, *e.g.*, an amino acid, a sugar, or other nutrient) to support its growth. An “auxotrophic modification” is a genetic modification that causes the organism to die in the absence of an exogenously added nutrient essential for survival or growth because it is unable to produce said nutrient. As used herein, the term “essential gene” refers to a gene which is necessary to for cell growth and/or survival. Essential genes are described in more detail *infra* and include, but are not limited to, DNA synthesis genes (such as *thyA*), cell wall synthesis genes (such as *dapA*), and amino acid genes (such as *serA* and *metA*).

[097] As used herein, the terms “modulate” and “treat” and their cognates refer to an amelioration of a disease, disorder, and/or condition, or at least one discernible symptom thereof. In

another embodiment, “modulate” and “treat” refer to an amelioration of at least one measurable physical parameter, not necessarily discernible by the patient. In another embodiment, “modulate” and “treat” refer to inhibiting the progression of a disease, disorder, and/or condition, either physically (*e.g.*, stabilization of a discernible symptom), physiologically (*e.g.*, stabilization of a physical parameter), or both. In another embodiment, “modulate” and “treat” refer to slowing the progression or reversing the progression of a disease, disorder, and/or condition. As used herein, “prevent” and its cognates refer to delaying the onset or reducing the risk of acquiring a given disease, disorder and/or condition or a symptom associated with such disease, disorder, and/or condition.

[098] Those in need of treatment may include individuals already having a particular medical disease, as well as those at risk of having, or who may ultimately acquire the disease. The need for treatment is assessed, for example, by the presence of one or more risk factors associated with the development of a disease, the presence or progression of a disease, or likely receptiveness to treatment of a subject having the disease. Diseases associated with the catabolism of a branched chain amino acid, *e.g.*, MSUD, may be caused by inborn genetic mutations for which there are no known cures. Diseases can also be secondary to other conditions, *e.g.*, liver diseases. Treating diseases involving the catabolism of a branched chain amino acid, such as MSUD, may encompass reducing normal levels of branched chain amino acids, reducing excess levels of branched chain amino acids, or eliminating branched chain amino acids, *e.g.*, leucine, and does not necessarily encompass the elimination of the underlying disease.

[099] As used herein, the term “catabolism” refers to the breakdown of a molecule into a smaller unit. As used herein, the term “branched chain amino acid catabolism” refers to the conversion of a branched chain amino acid, such as leucine, isoleucine, or valine, into a corresponding metabolite, for example a corresponding alpha keto acid, acyl-CoA derivative, aldehyde, alcohol, or other metabolite, such as any of the BCAA metabolites disclosed herein; or the conversion of a branched chain alpha keto acid into its corresponding acyl-CoA derivative, aldehydes, alcohols or other metabolites, such as any of the BCAA metabolites disclosed herein. The “branched chain amino acid catabolism” refers to both native and non-native conversion of a branched chain amino acid, such as leucine, isoleucine, or valine, into a corresponding metabolite. Thus, the term additionally covers catabolism of BCAA that may not occur in nature and is artificially induced as a result of genetic engineering. In one embodiment, “abnormal catabolism” refers to a decrease in the rate or the level of conversion of a branched chain amino acid or its corresponding alpha-keto acid to a corresponding metabolite, leading to the accumulation of the branched chain amino acid, accumulation of the branched chain alpha-keto acid, and/or accumulation of a BCAA metabolite that is toxic or that accumulates to a toxic level in a subject. In one embodiment, the branched chain amino acid, branched chain alpha-keto acid, or other metabolite thereof, accumulates to a toxic level in the blood or the brain of a subject, leading to the development of a disease or disorder associated with the abnormal catabolism of the branched chain amino acid in the subject. In one embodiment, “abnormal

leucine catabolism” refers to a level of greater than 4 mg/dL of leucine in the plasma of a subject. In another embodiment, “normal leucine catabolism” refers to a level of less than 4 mg/dL of leucine in the plasma of a subject.

[0100] As used herein, the term “disorder involving the abnormal catabolism of a branched amino acid” or “disease involving the abnormal catabolism of a branched amino acid” or “branched chain amino acid disease” or “disease associated with excess branched chain amino acid” refers to a disease or disorder wherein the catabolism of a branched chain amino acid or a branched chain alpha-keto acid is abnormal. Such diseases are genetic disorders that result from deficiency in at least one of the enzymes required to catabolize a branched chain amino acid, *e.g.*, leucine, isoleucine, or valine. As a result, individuals suffering from branched chain amino acid disease have accumulated branched chain amino acids in their cells and tissues. Examples of branched chain amino acid diseases include, but are not limited to, MSUD, isovaleric acidemia, 3-MCC deficiency, 3-methylglutaconyl-CoA hydrolase deficiency, HMG-CoA lyase deficiency, Acetyl CoA carboxylase deficiency, malonylCoA decarboxylase deficiency, short branched chain acyl-CoA dehydrogenase, 2-methyl-3-hydroxybutyric acidemia, beta-ketothiolase deficiency, isobutyl-CoA dehydrogenase deficiency, HIBCH deficiency, 3-hydroxyisobutyric aciduria, propionic acidemia, methylmalonic acidemia, as well as those diseases resulting from mTor activation, including but not limited to cancer, obesity, type 2 diabetes, neurodegeneration, autism, Alzheimer’s disease, Lymphangiomyomatosis (LAM), transplant rejection, glycogen storage disease, obesity, tuberous sclerosis, hypertension, cardiovascular disease, hypothalamic activation, musculoskeletal disease, Parkinson’s disease, Huntington’s disease, psoriasis, rheumatoid arthritis, lupus, multiple sclerosis, Leigh’s syndrome, and Friedrich’s ataxia. In one embodiment, a “disease or disorder involving the catabolism of a branched chain amino acid” is a metabolic disease or disorder involving the abnormal catabolism of a branched chain amino acid. In another embodiment, a “disease or disorder involving the catabolism of a branched chain amino acid” is a disease or disorder caused by the activation of mTor. In one embodiment, the activation of mTor is abnormal.

[0101] In one embodiment, “abnormal catabolism” refers to a decrease in the rate or the level of conversion of the branched chain amino acid or its alpha-keto acid counterpart, leading to the accumulation of the branched chain amino acid or alpha-keto acid in a subject. In one embodiment, accumulation of the branched chain amino acid in the blood or the brain of a subject becomes toxic and leads to the development of a disease or disorder associated with the abnormal catabolism of the branched chain amino acid in the subject.

[0102] As used herein, the term “disease caused by the activation of mTor” or “disorder caused by the activation of mTor” refers to a disease or a disorder wherein the levels of branched chain amino acid may be normal, and wherein the branched chain amino acid causes the activation of mTor at a level higher than the normal level of mTor activity. In another embodiment, the subject having a disorder caused by the activation of mTor may have higher levels of a branched chain amino

acid than normal. Diseases caused by the activation of mTor are known in the art. See, for example, Laplante and Sabatini, *Cell*, 149(2):74-293, 2012. As used herein, the term “disease caused by the activation of mTor” includes cancer, obesity, type 2 diabetes, neurodegeneration, autism, Alzheimer’s disease, Lymphangi leiomyomatosis (LAM), transplant rejection, glycogen storage disease, obesity, tuberous sclerosis, hypertension, cardiovascular disease, hypothalamic activation, musculoskeletal disease, Parkinson’s disease, Huntington’s disease, psoriasis, rheumatoid arthritis, lupus, multiple sclerosis, Leigh’s syndrome, and Friedrich’s ataxia. In one embodiment, the subject has normal levels of a branched chain amino acid, such as leucine, before administration of the engineered bacteria of the present disclosure. In another embodiment, the subject has decreased levels of the branched chain amino acid after the administration of the engineered bacteria of the present disclosure, thereby decreasing the levels of mTor or the activity of mTor, thereby treating the disorder in the subject. In one embodiment, the pharmaceutical composition disclosed herein decreases the activity of mTor by at least about 2-fold, 3-fold, 4-fold, or 5-fold in the subject.

[0103] As used herein, the term “anabolism” refers the conversion of a branched chain alpha-keto acid or an acyl-CoA derivative or other metabolite into its corresponding branched chain amino acid, such as leucine, isoleucine, or valine or alpha-keto acid, respectively.

[0104] As used herein a "pharmaceutical composition" refers to a preparation of genetically engineered microorganism of the disclosure, *e.g.*, genetically engineered bacteria yeast or virus, with other components such as a physiologically suitable carrier and/or excipient.

[0105] The phrases "physiologically acceptable carrier" and "pharmaceutically acceptable carrier" which may be used interchangeably refer to a carrier or a diluent that does not cause significant irritation to an organism and does not abrogate the biological activity and properties of the administered bacterial or viral compound. An adjuvant is included under these phrases.

[0106] The term "excipient" refers to an inert substance added to a pharmaceutical composition to further facilitate administration of an active ingredient. Examples include, but are not limited to, calcium bicarbonate, sodium bicarbonate calcium phosphate, various sugars and types of starch, cellulose derivatives, gelatin, vegetable oils, polyethylene glycols, and surfactants, including, for example, polysorbate 20.

[0107] The terms “therapeutically effective dose” and “therapeutically effective amount” are used to refer to an amount of a compound that results in prevention, delay of onset of symptoms, or amelioration of symptoms of a condition, *e.g.*, lysosomal storage disease (LSD). A therapeutically effective amount may, for example, be sufficient to treat, prevent, reduce the severity, delay the onset, and/or reduce the risk of occurrence of one or more symptoms of a disorder associated with lysosomal storage disease. A therapeutically effective amount, as well as a therapeutically effective frequency of administration, can be determined by methods known in the art and discussed below.

[0108] As used herein, the term "bacteriostatic" or "cytostatic" refers to a molecule or protein which is capable of arresting, retarding, or inhibiting the growth, division, multiplication or replication of recombinant bacterial cell of the disclosure.

[0109] As used herein, the term "bactericidal" refers to a molecule or protein which is capable of killing the recombinant bacterial cell of the disclosure.

[0110] As used herein, the term "toxin" refers to a protein, enzyme, or polypeptide fragment thereof, or other molecule which is capable of arresting, retarding, or inhibiting the growth, division, multiplication or replication of the recombinant bacterial cell of the disclosure, or which is capable of killing the recombinant bacterial cell of the disclosure. The term "toxin" is intended to include bacteriostatic proteins and bactericidal proteins. The term "toxin" is intended to include, but not limited to, lytic proteins, bacteriocins (*e.g.*, microcins and colicins), gyrase inhibitors, polymerase inhibitors, transcription inhibitors, translation inhibitors, DNases, and RNases. The term "anti-toxin" or "antitoxin," as used herein, refers to a protein or enzyme which is capable of inhibiting the activity of a toxin. The term anti-toxin is intended to include, but not limited to, immunity modulators, and inhibitors of toxin expression. Examples of toxins and antitoxins are known in the art and described in more detail *infra*.

[0111] As used herein, the term "branched chain amino acid catabolic or catabolism enzyme" or "BCAA catabolic or catabolism enzyme" or "branched chain or BCAA amino acid metabolic enzyme" refers to any enzyme that is capable of metabolizing a branched chain amino acid or capable of reducing accumulated branched chain amino acid or that can lessen, ameliorate, or prevent one or more branched chain amino acid diseases or disease symptoms. Examples of branched chain amino acid metabolic enzymes include, but are not limited to, leucine dehydrogenase (*e.g.*, LeuDH), branched chain amino acid aminotransferase (*e.g.*, IlvE), branched chain α -ketoacid dehydrogenase (*e.g.*, KivD), L-Amino acid deaminase (*e.g.*, L-AAD), alcohol dehydrogenase (*e.g.*, Adh2, YqhD), and aldehyde dehydrogenase (*e.g.*, PadA), and any other enzymes that catabolizes BCAA. Functional deficiencies in these proteins result in the accumulation of BCAA or its corresponding α -keto acid in cells and tissues. BCAA metabolic enzymes of the present disclosure include both wild-type or modified BCAA metabolic enzymes and can be produced using recombinant and synthetic methods or purified from nature sources. BCAA metabolic enzymes include full-length polypeptides and functional fragments thereof, as well as homologs and variants thereof. BCAA metabolic enzymes include polypeptides that have been modified from the wild-type sequence, including, for example, polypeptides having one or more amino acid deletions, insertions, and/or substitutions and may include, for example, fusion polypeptides and polypeptides having additional sequence, *e.g.*, regulatory peptide sequence, linker peptide sequence, and other peptide sequence.

[0112] As used herein, "payload" refers to one or more molecules of interest to be produced by a genetically engineered microorganism, such as a bacterium, yeast, or a virus. In some

embodiments, the payload is a therapeutic payload, *e.g.*, a branched chain amino acid catabolic enzyme or a BCAA transporter polypeptide. In some embodiments, the payload is a regulatory molecule, *e.g.*, a transcriptional regulator such as FNR. In some embodiments, the payload comprises a regulatory element, such as a promoter or a repressor. In some embodiments, the payload comprises an inducible promoter, such as from FNRS. In some embodiments, the payload comprises a repressor element, such as a kill switch. In some embodiments, the payload comprises an antibiotic resistance gene or genes. In some embodiments, the payload is encoded by a gene, multiple genes, gene cassette, or an operon. In alternate embodiments, the payload is produced by a biosynthetic or biochemical pathway, wherein the biosynthetic or biochemical pathway may optionally be endogenous to the microorganism. In alternate embodiments, the payload is produced by a biosynthetic or biochemical pathway, wherein the biosynthetic or biochemical pathway is not endogenous to the microorganism. In some embodiments, the genetically engineered microorganism comprises two or more payloads.

[0113] As used herein, the term “conventional branched chain amino acid or BCAA disease treatment” or “conventional branched chain amino acid or BCAA disease therapy” refers to treatment or therapy that is currently accepted, considered current standard of care, and/or used by most healthcare professionals for treating a disease or disorder associated with BCAA. It is different from alternative or complementary therapies, which are not as widely used.

[0114] As used herein, the term “polypeptide” includes “polypeptide” as well as “polypeptides,” and refers to a molecule composed of amino acid monomers linearly linked by amide bonds (*i.e.*, peptide bonds). The term “polypeptide” refers to any chain or chains of two or more amino acids, and does not refer to a specific length of the product. Thus, “peptides,” “dipeptides,” “tripeptides,” “oligopeptides,” “protein,” “amino acid chain,” or any other term used to refer to a chain or chains of two or more amino acids, are included within the definition of “polypeptide,” and the term “polypeptide” may be used instead of, or interchangeably with any of these terms. The term “polypeptide” is also intended to refer to the products of post-expression modifications of the polypeptide, including but not limited to glycosylation, acetylation, phosphorylation, amidation, derivatization, proteolytic cleavage, or modification by non-naturally occurring amino acids. A polypeptide may be derived from a natural biological source or produced by recombinant technology. In other embodiments, the polypeptide is produced by the genetically engineered bacteria, yeast, or virus of the current invention. A polypeptide of the invention may be of a size of about 3 or more, 5 or more, 10 or more, 20 or more, 25 or more, 50 or more, 75 or more, 100 or more, 200 or more, 500 or more, 1,000 or more, or 2,000 or more amino acids. Polypeptides may have a defined three-dimensional structure, although they do not necessarily have such structure. Polypeptides with a defined three-dimensional structure are referred to as folded, and polypeptides, which do not possess a defined three-dimensional structure, but rather can adopt a large number of different conformations, are referred to as unfolded. The term “peptide” or “polypeptide” may refer to an amino acid sequence

that corresponds to a protein or a portion of a protein or may refer to an amino acid sequence that corresponds with non-protein sequence, *e.g.*, a sequence selected from a regulatory peptide sequence, leader peptide sequence, signal peptide sequence, linker peptide sequence, and other peptide sequence.

[0115] An “isolated” polypeptide or a fragment, variant, or derivative thereof refers to a polypeptide that is not in its natural milieu. No particular level of purification is required. Recombinantly produced polypeptides and proteins expressed in host cells, including but not limited to bacterial or mammalian cells, are considered isolated for purposes of the invention, as are native or recombinant polypeptides which have been separated, fractionated, or partially or substantially purified by any suitable technique. Recombinant peptides, polypeptides or proteins refer to peptides, polypeptides or proteins produced by recombinant DNA techniques, *i.e.* produced from cells, microbial or mammalian, transformed by an exogenous recombinant DNA expression construct encoding the polypeptide. Proteins or peptides expressed in most bacterial cultures will typically be free of glycan. Fragments, derivatives, analogs or variants of the foregoing polypeptides, and any combination thereof are also included as polypeptides. The terms “fragment,” “variant,” “derivative” and “analog” include polypeptides having an amino acid sequence sufficiently similar to the amino acid sequence of the original peptide and include any polypeptides, which retain at least one or more properties of the corresponding original polypeptide. Fragments of polypeptides of the present invention include proteolytic fragments, as well as deletion fragments. Fragments also include specific antibody or bioactive fragments or immunologically active fragments derived from any polypeptides described herein. Variants may occur naturally or be non-naturally occurring. Non-naturally occurring variants may be produced using mutagenesis methods known in the art. Variant polypeptides may comprise conservative or non-conservative amino acid substitutions, deletions or additions.

[0116] Polypeptides also include fusion proteins. As used herein, the term “variant” includes a fusion protein, which comprises a sequence of the original peptide or sufficiently similar to the original peptide. As used herein, the term “fusion protein” refers to a chimeric protein comprising amino acid sequences of two or more different proteins. Typically, fusion proteins result from well known *in vitro* recombination techniques. Fusion proteins may have a similar structural function (but not necessarily to the same extent), and/or similar regulatory function (but not necessarily to the same extent), and/or similar biochemical function (but not necessarily to the same extent) and/or immunological activity (but not necessarily to the same extent) as the individual original proteins which are the components of the fusion proteins. “Derivatives” include but are not limited to peptides, which contain one or more naturally occurring amino acid derivatives of the twenty standard amino acids. “Similarity” between two peptides is determined by comparing the amino acid sequence of one peptide to the sequence of a second peptide. An amino acid of one peptide is similar to the corresponding amino acid of a second peptide if it is identical or a conservative amino acid

substitution. Conservative substitutions include those described in Dayhoff, M. O., ed., *The Atlas of Protein Sequence and Structure* 5, National Biomedical Research Foundation, Washington, D.C. (1978), and in Argos, *EMBO J.* 8 (1989), 779-785. For example, amino acids belonging to one of the following groups represent conservative changes or substitutions: Ala, Pro, Gly, Gln, Asn, Ser, Thr, Cys, Ser, Tyr, Thr, Val, Ile, Leu, Met, Ala, Phe, Lys, Arg, His, Phe, Tyr, Trp, His, Asp, and Glu.

[0117] As used herein, the term “sufficiently similar” means a first amino acid sequence that contains a sufficient or minimum number of identical or equivalent amino acid residues relative to a second amino acid sequence such that the first and second amino acid sequences have a common structural domain and/or common functional activity. For example, amino acid sequences that comprise a common structural domain that is at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 100%, identical are defined herein as sufficiently similar. Preferably, variants will be sufficiently similar to the amino acid sequence of the peptides of the invention. Such variants generally retain the functional activity of the peptides of the present invention. Variants include peptides that differ in amino acid sequence from the native and wild-type peptide, respectively, by way of one or more amino acid deletion(s), addition(s), and/or substitution(s). These may be naturally occurring variants as well as artificially designed ones.

[0118] As used herein the term “linker”, “linker peptide” or “peptide linkers” or “linker” refers to synthetic or non-native or non-naturally-occurring amino acid sequences that connect or link two polypeptide sequences, *e.g.*, that link two polypeptide domains. As used herein the term “synthetic” refers to amino acid sequences that are not naturally occurring. Exemplary linkers are described herein. Additional exemplary linkers are provided in US 20140079701, the contents of which are herein incorporated by reference in its entirety.

[0119] As used herein the term “codon-optimized” refers to the modification of codons in the gene or coding regions of a nucleic acid molecule to reflect the typical codon usage of the host organism without altering the polypeptide encoded by the nucleic acid molecule. Such optimization includes replacing at least one, or more than one, or a significant number, of codons with one or more codons that are more frequently used in the genes of the host organism. A “codon-optimized sequence” refers to a sequence, which was modified from an existing coding sequence, or designed, for example, to improve translation in an expression host cell or organism of a transcript RNA molecule transcribed from the coding sequence, or to improve transcription of a coding sequence. Codon optimization includes, but is not limited to, processes including selecting codons for the coding sequence to suit the codon preference of the expression host organism. Many organisms display a bias or preference for use of particular codons to code for insertion of a particular amino acid in a growing polypeptide chain. Codon preference or codon bias, differences in codon usage between

organisms, is allowed by the degeneracy of the genetic code, and is well documented among many organisms. Codon bias often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, inter alia, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization.

[0120] As used herein, the terms “secretion system” or “secretion protein” refers to a native or non-native secretion mechanism capable of secreting or exporting a biomolecule, *e.g.*, polypeptide from the microbial, *e.g.*, bacterial cytoplasm. The secretion system may comprise a single protein or may comprise two or more proteins assembled in a complex *e.g.* HlyBD. Non-limiting examples of secretion systems for gram negative bacteria include the modified type III flagellar, type I (*e.g.*, hemolysin secretion system), type II, type IV, type V, type VI, and type VII secretion systems, resistance-nodulation-division (RND) multi-drug efflux pumps, various single membrane secretion systems. Non-limiting examples of secretion systems for gram positive bacteria include Sec and TAT secretion systems. In some embodiments, the polypeptide to be secreted include a “secretion tag” of either RNA or peptide origin to direct the polypeptide to specific secretion systems. In some embodiments, the secretion system is able to remove this tag before secreting the polypeptide from the engineered bacteria. For example, in Type V auto-secretion-mediated secretion the N-terminal peptide secretion tag is removed upon translocation of the “passenger” peptide from the cytoplasm into the periplasmic compartment by the native Sec system. Further, once the auto-secretor is translocated across the outer membrane the C-terminal secretion tag can be removed by either an autocatalytic or protease-catalyzed *e.g.*, OmpT cleavage thereby releasing the lysosomal enzyme(s) into the extracellular milieu. In some embodiments, the secretion system involves the generation of a “leaky” or de-stabilized outer membrane, which may be accomplished by deleting or mutagenizing genes responsible for tethering the outer membrane to the rigid peptidoglycan skeleton, including for example, *lpp*, *ompC*, *ompA*, *ompF*, *tolA*, *tolB*, *pal*, *degS*, *degP*, and *nlpl*. *Lpp* functions as the primary ‘staple’ of the bacterial cell wall to the peptidoglycan. *TolA-PAL* and *OmpA* complexes function similarly to *Lpp* and are other deletion targets to generate a leaky phenotype. Additionally, leaky phenotypes have been observed when periplasmic proteases, such as *degS*, *degP* or *nlpl*, are deactivated. Thus, in some embodiments, the engineered bacteria have one or more deleted or mutated membrane genes, *e.g.*, selected from *lpp*, *ompA*, *ompA*, *ompF*, *tolA*, *tolB*, and *pal* genes. In some embodiments, the engineered bacteria have one or more deleted or mutated periplasmic protease genes, *e.g.*, selected from *degS*, *degP*, and *nlpl*. In some embodiments, the engineered bacteria have one or more deleted or mutated gene(s), selected from *lpp*, *ompA*, *ompA*, *ompF*, *tolA*, *tolB*, *pal*, *degS*, *degP*, and *nlpl* genes.

[0121] The articles “a” and “an,” as used herein, should be understood to mean “at least one,” unless clearly indicated to the contrary.

[0122] The phrase “and/or,” when used between elements in a list, is intended to mean either (1) that only a single listed element is present, or (2) that more than one element of the list is present. For example, “A, B, and/or C” indicates that the selection may be A alone; B alone; C alone; A and B; A and C; B and C; or A, B, and C. The phrase “and/or” may be used interchangeably with “at least one of” or “one or more of” the elements in a list.

[0123] Ranges provided herein are understood to be shorthand for all of the values within the range. For example, a range of 1 to 50 is understood to include any number, combination of numbers, or sub-range from the group consisting 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50.

Recombinant Bacteria

[0124] The genetically engineered microorganisms, or programmed microorganisms, such as genetically engineered bacteria, of the disclosure are capable of producing one or more enzymes for metabolizing a branched amino acid and/or a metabolite thereof. In some aspects, the disclosure provides a bacterial cell that comprises one or more optimized heterologous gene sequence(s) encoding a branched chain amino acid catabolic enzyme or other protein that results in a decrease in BCAA levels, *e.g.*, leucine levels.

[0125] In certain embodiments, the genetically engineered bacteria are obligate anaerobic bacteria. In certain embodiments, the genetically engineered bacteria are facultative anaerobic bacteria. In certain embodiments, the genetically engineered bacteria are aerobic bacteria. In some embodiments, the genetically engineered bacteria are Gram-positive bacteria. In some embodiments, the genetically engineered bacteria are Gram-positive bacteria and lack LPS. In some embodiments, the genetically engineered bacteria are Gram-negative bacteria. In some embodiments, the genetically engineered bacteria are Gram-positive and obligate anaerobic bacteria. In some embodiments, the genetically engineered bacteria are Gram-positive and facultative anaerobic bacteria. In some embodiments, the genetically engineered bacteria are non-pathogenic bacteria. In some embodiments, the genetically engineered bacteria are commensal bacteria. In some embodiments, the genetically engineered bacteria are probiotic bacteria. In some embodiments, the genetically engineered bacteria are naturally pathogenic bacteria that are modified or mutated to reduce or eliminate pathogenicity. Exemplary bacteria include, but are not limited to, *Bacillus*, *Bacteroides*, *Bifidobacterium*, *Brevibacteria*, *Caulobacter*, *Clostridium*, *Enterococcus*, *Escherichia coli*, *Lactobacillus*, *Lactococcus*, *Listeria*, *Mycobacterium*, *Saccharomyces*, *Salmonella*, *Staphylococcus*, *Streptococcus*, *Vibrio*, *Bacillus coagulans*, *Bacillus subtilis*, *Bacteroides fragilis*, *Bacteroides subtilis*, *Bacteroides thetaiotaomicron*, *Bifidobacterium adolescentis*, *Bifidobacterium bifidum*, *Bifidobacterium breve*

UCC2003, *Bifidobacterium infantis*, *Bifidobacterium lactis*, *Bifidobacterium longum*, *Clostridium acetobutylicum*, *Clostridium butyricum*, *Clostridium butyricum M-55*, *Clostridium cochlearum*, *Clostridium felsineum*, *Clostridium histolyticum*, *Clostridium multifementans*, *Clostridium novyi-NT*, *Clostridium paraputrificum*, *Clostridium pasteurianum*, *Clostridium pectinovorum*, *Clostridium perfringens*, *Clostridium roseum*, *Clostridium sporogenes*, *Clostridium tertium*, *Clostridium tetani*, *Clostridium tyrobutyricum*, *Corynebacterium parvum*, *Escherichia coli MG1655*, *Escherichia coli Nissle 1917*, *Listeria monocytogenes*, *Mycobacterium bovis*, *Salmonella choleraesuis*, *Salmonella typhimurium*, and *Vibrio cholera*. In certain embodiments, the genetically engineered bacteria are selected from the group consisting of *Enterococcus faecium*, *Lactobacillus acidophilus*, *Lactobacillus bulgaricus*, *Lactobacillus casei*, *Lactobacillus johnsonii*, *Lactobacillus paracasei*, *Lactobacillus plantarum*, *Lactobacillus reuteri*, *Lactobacillus rhamnosus*, *Lactococcus lactis*, and *Saccharomyces boulardii*. In certain embodiments, the genetically engineered bacteria are selected from *Bacteroides fragilis*, *Bacteroides thetaiotaomicron*, *Bacteroides subtilis*, *Bifidobacterium bifidum*, *Bifidobacterium infantis*, *Bifidobacterium lactis*, *Clostridium butyricum*, *Escherichia coli*, *Escherichia coli Nissle*, *Lactobacillus acidophilus*, *Lactobacillus plantarum*, *Lactobacillus reuteri*, and *Lactococcus lactis* bacterial cell. In one embodiment, the bacterial cell is a *Bacteroides fragilis* bacterial cell. In one embodiment, the bacterial cell is a *Bacteroides thetaiotaomicron* bacterial cell. In one embodiment, the bacterial cell is a *Bacteroides subtilis* bacterial cell. In one embodiment, the bacterial cell is a *Bifidobacterium bifidum* bacterial cell. In one embodiment, the bacterial cell is a *Bifidobacterium infantis* bacterial cell. In one embodiment, the bacterial cell is a *Bifidobacterium lactis* bacterial cell. In one embodiment, the bacterial cell is a *Clostridium butyricum* bacterial cell. In one embodiment, the bacterial cell is an *Escherichia coli* bacterial cell. In one embodiment, the bacterial cell is a *Lactobacillus acidophilus* bacterial cell. In one embodiment, the bacterial cell is a *Lactobacillus plantarum* bacterial cell. In one embodiment, the bacterial cell is a *Lactobacillus reuteri* bacterial cell. In one embodiment, the bacterial cell is a *Lactococcus lactis* bacterial cell.

[0126] In some embodiments, the genetically engineered bacteria are *Escherichia coli* strain Nissle 1917 (*E. coli* Nissle), a Gram-negative bacterium of the Enterobacteriaceae family that has evolved into one of the best characterized probiotics (Ukena *et al.*, 2007). The strain is characterized by its complete harmlessness (Schultz, 2008), and has GRAS (generally recognized as safe) status (Reister *et al.*, 2014, emphasis added). Genomic sequencing confirmed that *E. coli* Nissle lacks prominent virulence factors (e.g., *E. coli* α -hemolysin, P-fimbrial adhesins) (Schultz, 2008). In addition, it has been shown that *E. coli* Nissle does not carry pathogenic adhesion factors, does not produce any enterotoxins or cytotoxins, is not invasive, and not uropathogenic (Sonnenborn *et al.*, 2009). As early as in 1917, *E. coli* Nissle was packaged into medicinal capsules, called Mutaflor, for therapeutic use. *E. coli* Nissle has since been used to treat ulcerative colitis in humans *in vivo* (Rembacken *et al.*, 1999), to treat inflammatory bowel disease, Crohn's disease, and pouchitis in humans *in vivo* (Schultz, 2008), and to inhibit enteroinvasive *Salmonella*, *Legionella*, *Yersinia*, and

Shigella in vitro (Altenhoefer *et al.*, 2004). It is commonly accepted that *E. coli* Nissle's therapeutic efficacy and safety have convincingly been proven (Ukena *et al.*, 2007).

[0127] One of ordinary skill in the art would appreciate that the genetic modifications disclosed herein may be adapted for other species, strains, and subtypes of bacteria. Furthermore, genes from one or more different species can be introduced into one another, *e.g.*, the *kivD* gene from *Lactococcus lactis* can be expressed in *Escherichia coli*. In one embodiment, the recombinant bacterial cell does not colonize the subject having the disorder. Unmodified *E. coli* Nissle and the genetically engineered bacteria of the invention may be destroyed, *e.g.*, by defense factors in the gut or blood serum (Sonnenborn *et al.*, 2009). In some embodiments, the residence time is calculated for a human subject. In some embodiments, residence time *in vivo* is calculated for the genetically engineered bacteria of the invention.

[0128] In some embodiments, the bacterial cell is a genetically engineered bacterial cell. In another embodiment, the bacterial cell is a recombinant bacterial cell. In some embodiments, the disclosure comprises a colony of bacterial cells disclosed herein.

[0129] In another aspect, the disclosure provides a recombinant bacterial culture which comprises bacterial cells disclosed herein. In one aspect, the disclosure provides a recombinant bacterial culture which reduces levels of a branched chain amino acid, *e.g.*, leucine, in the media of the culture. In one embodiment, the levels of the branched chain amino acid, *e.g.*, leucine, are reduced by about 50%, about 75%, or about 100% in the media of the cell culture. In another embodiment, the levels of the branched chain amino acid, *e.g.*, leucine, are reduced by about two-fold, three-fold, four-fold, five-fold, six-fold, seven-fold, eight-fold, nine-fold, or ten-fold in the media of the cell culture. In one embodiment, the levels of the branched chain amino acid, *e.g.*, leucine, are reduced below the limit of detection in the media of the cell culture.

[0130] In some embodiments of the above described genetically engineered bacteria, the gene encoding a branched chain amino acid catabolism enzyme is present on a plasmid in the bacterium and operatively linked on the plasmid to a promoter that is induced under low-oxygen or anaerobic conditions, such as any of the promoters disclosed herein. In other embodiments, the gene encoding a branched chain amino acid catabolism enzyme is present in the bacterial chromosome and is operatively linked in the chromosome to the promoter that is induced under low-oxygen or anaerobic conditions, such as any of the promoters disclosed herein. In some embodiments of the above described genetically engineered bacteria, the gene encoding a branched chain amino acid catabolic enzyme is present on a plasmid in the bacterium and operatively linked on the plasmid to the promoter that is induced under inflammatory conditions, such as any of the promoters disclosed herein. In other embodiments, the gene encoding a branched chain amino acid catabolic enzyme is present in the bacterial chromosome and is operatively linked in the chromosome to the promoter that is induced under inflammatory conditions, such as any of the promoters disclosed herein.

[0131] In some embodiments, the genetically engineered bacteria comprising gene sequence encoding a branched chain amino acid catabolic enzyme is an auxotroph. In one embodiment, the genetically engineered bacteria is an auxotroph selected from a *cysE*, *glnA*, *ilvD*, *leuB*, *lysA*, *serA*, *metA*, *glyA*, *hisB*, *ilvA*, *pheA*, *proA*, *thrC*, *trpC*, *tyrA*, *thyA*, *uraA*, *dapA*, *dapB*, *dapD*, *dapE*, *dapF*, *flhD*, *metB*, *metC*, *proAB*, and *thi1* auxotroph. In some embodiments, the engineered bacteria have more than one auxotrophy, for example, they may be a Δ *thyA* and Δ *dapA* auxotroph. In some embodiments, the genetically engineered bacteria comprising gene sequence encoding a branched chain amino acid catabolic enzyme lacks functional *ilvC* gene sequence, e.g., is a *ilvC* auxotroph. *IlvC* encodes keto acid reductoisomerase, which enzyme is required for BCAA synthesis. Knock out of *ilvC* creates an auxotroph and requires the bacterial cell to import isoleucine and valine to survive.

[0132] In some embodiments, the genetically engineered bacteria comprising gene sequence encoding a branched chain amino acid catabolism enzyme further comprise gene sequence(s) encoding a BCAA transporter or other amino acid transporter that transports one or more BCAA(s) into the bacterial cell, for example a transporter that is capable of transporting leucine, valine, and/or isoleucine into a bacterial cell. In certain embodiments, the BCAA transporter is a leucine transporter, e.g., a high-affinity leucine transporter. In certain embodiments, the bacterial cell contains *livK*, *livH*, *livM*, *livG*, and *livF* genes. In certain embodiments, the BCAA transporter is a BCAA transporter, e.g., a low affinity BCAA transporter. In certain embodiments, the bacterial cell contains gene sequence encoding *brnQ* gene.

[0133] In some embodiments, the genetically engineered bacteria comprising gene sequence encoding a branched chain amino acid catabolism enzyme further comprise gene sequence(s) encoding a secretion protein or protein complex for secreting a biomolecule, such as any of the secretion systems disclosed herein.

[0134] In some embodiments, the genetically engineered bacteria comprising gene sequence encoding a branched chain amino acid catabolism enzyme further comprise gene sequence(s) encoding one or more antibiotic gene(s), such as any of the antibiotic genes disclosed herein.

[0135] In some embodiments, the genetically engineered bacteria comprising a branched chain amino acid catabolism enzyme further comprise a kill-switch circuit, such as any of the kill-switch circuits provided herein. For example, in some embodiments, the genetically engineered bacteria further comprise one or more genes encoding one or more recombinase(s) under the control of an inducible promoter, and an inverted toxin sequence. In some embodiments, the genetically engineered bacteria further comprise one or more genes encoding an antitoxin. In some embodiments, the engineered bacteria further comprise one or more genes encoding one or more recombinase(s) under the control of an inducible promoter and one or more inverted excision genes, wherein the excision gene(s) encode an enzyme that deletes an essential gene. In some embodiments, the genetically engineered bacteria further comprise one or more genes encoding an antitoxin. In some embodiments, the engineered bacteria further comprise one or more genes encoding a toxin

under the control of a promoter having a TetR repressor binding site and a gene encoding the TetR under the control of an inducible promoter that is induced by arabinose, such as P_{araBAD} . In some embodiments, the genetically engineered bacteria further comprise one or more genes encoding an antitoxin.

[0136] In some embodiments, the genetically engineered bacteria are an auxotroph comprising gene sequence encoding a branched chain amino acid catabolism enzyme and further comprises a kill-switch circuit, such as any of the kill-switch circuits described herein.

[0137] In some embodiments of the above described genetically engineered bacteria, the gene encoding a branched chain amino acid catabolism enzyme is present on a plasmid in the bacterium. In some embodiments, the gene encoding a branched chain amino acid catabolism enzyme is present in the bacterial chromosome. In some embodiments, the gene sequence(s) encoding a BCAA transporter or other amino acid transporter that transports one or more BCAA(s) into the bacterial cell, for example a transporter that is capable of transporting leucine, valine, and/or isoleucine into a bacterial cell, is present on a plasmid in the bacterium. In some embodiments, the gene sequence(s) encoding a BCAA transporter or other amino acid transporter that transports one or more BCAA(s) into the bacterial cell, for example a transporter that is capable of transporting leucine, valine, and/or isoleucine into a bacterial cell, is present in the bacterial chromosome. In some embodiments, the gene sequence encoding a secretion protein or protein complex for secreting a biomolecule, such as any of the secretion systems disclosed herein, is present on a plasmid in the bacterium. In some embodiments, the gene sequence encoding a secretion protein or protein complex for secreting a biomolecule, such as any of the secretion systems disclosed herein, is present in the bacterial chromosome. In some embodiments, the gene sequence(s) encoding an antibiotic resistance gene is present on a plasmid in the bacterium. In some embodiments, the gene sequence(s) encoding an antibiotic resistance gene is present in the bacterial chromosome.

Branched Chain Amino Acid Catabolism Enzymes

[0138] As used herein, the term “branched chain amino acid catabolic or catabolism enzyme” refers to an enzyme involved in the catabolism of a branched chain amino acid to its corresponding α -keto acid counterpart; or the catabolism of an alpha-keto acid to its corresponding aldehyde, acyl-CoA, alcohol, carboxylic acid, or other metabolite counterpart. In some embodiments, the present disclosure provides an engineered bacterium comprising gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s). In some embodiments, the branched chain amino acid catabolism enzyme is used to convert a branched chain amino acid, *e.g.*, leucine, valine, isoleucine, to its corresponding α -keto-acid, *e.g.*, α -ketoisocaproate, α -keto- β -methylvalerate, and α -ketoisovalerate. In some embodiments, wherein a branched chain amino acid catabolism enzyme is used to convert a branched chain amino acid, *e.g.*, leucine, valine, isoleucine, to its corresponding α -keto-acid, the engineered bacteria further comprise one or more branched chain amino acid

catabolism enzyme(s) to convert an α -keto-acid to its corresponding acetyl CoA, *e.g.*, isovaleryl-CoA, α -methylbutyryl-CoA, and isobutyryl-CoA. In some embodiments, wherein a branched chain amino acid catabolism enzyme is used to convert a branched chain amino acid, *e.g.*, leucine, valine, isoleucine, to its corresponding α -keto-acid, the engineered bacteria further comprise one or more branched chain amino acid catabolism enzyme(s) to convert an α -keto-acid to its corresponding aldehyde, *e.g.*, isovaleraldehyde, isobutyraldehyde, and 2-methylbutyraldehyde. In some embodiments, the engineered bacteria may further comprise an alcohol dehydrogenase enzyme in order to convert the branched chain amino acid-derived aldehyde (*e.g.*, isovaleraldehyde, isobutyraldehyde, 2-methylbutyraldehyde) to its respective alcohol. In some embodiments, the engineered bacteria may further comprise an aldehyde dehydrogenase enzyme in order to convert the branched chain amino acid-derived aldehyde (*e.g.*, isovaleraldehyde, isobutyraldehyde, 2-methylbutyraldehyde) to its respective carboxylic acid.

[0139] Enzymes involved in the catabolism of branched chain amino acids are well known to those of skill in the art. For example, in bacteria, leucine dehydrogenase (LeuDh), branched chain amino acid transferase (IlvE), amino acid oxidase (also known as amino acid deaminase) (L-AAD), as well as other known enzymes, can be used to convert a BCAA to its corresponding α -keto acid, *e.g.*, ketoisocaproate (KIC), ketoisovalerate (KIV), and ketomethylvalerate (KMV). Leucine dehydrogenases, branched chain amino acid transamination enzymes (EC 2.6.1.42), and L-amino acid deaminases (L-AAD), which oxidatively deaminate branched chain amino acids into their respective α -keto acid, are known (Baker *et al.*, *Structure*, 3(7):693-705, 1995; Peng *et al.*, *J. Bact.*, 139(2):339-45, 1979; and Kline *et al.*, *J. Bact.*, 130(2):951-3, 1977). In bacteria, branched chain keto acid dehydrogenases (“BCKDs”) are enzyme complexes that oxidatively decarboxylate all three branched chain keto acids into their respective acyl-CoA derivatives. Thus, in one embodiment, the branched chain amino acid catabolism enzyme is a branched chain keto acid dehydrogenase (BCKD). Moreover, in mammals, dehydrogenases specific for 2-ketoisovalerate (EC 1.2.4.4) and 2-keto-3-methylvalerate and 2-keto-isocaproate (EC 1.2.4.3) have been identified (see, for example, Massey *et al.*, *Bacteriol Rev.*, 40(1):42-54, 1976). In bacteria, branched chain keto acid dehydrogenases (“BCKDs”) are enzyme complexes that oxidatively decarboxylate all three branched chain keto acids into their respective acyl-CoA derivatives. Also, for example, in bacteria, α -ketoisovalerate decarboxylase (KivD) enzymes are capable of converting α -keto acids into aldehydes (*e.g.*, isovaleraldehyde, isobutyraldehyde, 2-methylbutyraldehyde). Specifically, the α -ketoisovalerate decarboxylase enzyme KivD is capable of metabolizing valine by converting α -ketoisovalerate to isobutyraldehyde (see, for example, de la Plaza *et al.*, *FEMS Microbiol. Lett.* 2004, 238(2):367-374). KivD is capable of metabolizing leucine by converting α -ketoisocaproate (KIC) to isovaleraldehyde. KivD is also capable of metabolizing isoleucine by converting α -ketomethylvalerate (KMV) to 2-methylbutyraldehyde. In addition, enzymes for converting isovaleraldehyde, isobutyraldehyde, and 2-methylbutyraldehyde to their respective alcohols or carboxylic acids are known and available. For

example, alcohol dehydrogenases (*e.g.*, Adh2, YqhD) can convert isovaleraldehyde, isobutyraldehyde, 2-methylbutyraldehyde to isopentanol, isobutanol, and 2-methylbutanol, respectively. Aldehyde dehydrogenases (*e.g.*, PadA) can convert isovaleraldehyde, isobutyraldehyde, 2-methylbutyraldehyde to isovalerate, isobutyrate, and 2-methylbutyrate, respectively.

[0140] In some embodiments, the branched chain amino acid catabolism enzyme increases the rate of branched chain amino acid catabolism. In some embodiments, the branched chain amino acid catabolism enzyme decreases the level of one or more branched chain amino acids, *e.g.*, leucine, isoleucine, and/or valine, in a cell, tissue, or organism. In some embodiments, the branched chain amino acid catabolism enzyme decreases the level of alpha-keto acid derived from BCAA in a cell, tissue, or organism. In some embodiments, the branched chain amino acid catabolism enzyme decreases the level of branched chain amino acid as compared to the level of its corresponding alpha-keto acid in a cell, tissue, or organism. In other embodiments, the branched chain amino acid catabolism enzyme increases the level of alpha-keto acid as compared to the level of its corresponding branched chain amino acid in a cell, tissue, or organism. In some embodiments, the branched chain amino acid catabolism enzyme decreases the level of the branched chain amino acid as compared to the level of its corresponding Acyl-CoA derivative in a cell, tissue, or organism. In some embodiments, the branched chain amino acid catabolism enzyme increases the level of the Acyl-CoA derivative as compared to the level of the branched chain amino acid in a cell, tissue, or organism. In some embodiments, the branched chain amino acid catabolism enzyme decreases the level of alpha-keto aldehyde derived from BCAA, *e.g.*, isovaleraldehyde, isobutyraldehyde, and 2-methylbutyraldehyde, in a cell, tissue, or organism. In some embodiments, the branched chain amino acid catabolism enzyme decreases the level of branched chain amino acid as compared to the level of its corresponding alpha-keto aldehyde, *e.g.*, isovaleraldehyde, isobutyraldehyde, and 2-methylbutyraldehyde, in a cell, tissue, or organism. In other embodiments, the branched chain amino acid catabolism enzyme increases the level of alpha-keto aldehyde, *e.g.*, isovaleraldehyde, isobutyraldehyde, and 2-methylbutyraldehyde, as compared to the level of its corresponding branched chain amino acid in a cell, tissue, or organism. In some embodiments, the branched chain amino acid catabolism enzyme decreases the level of a corresponding downstream metabolite, *e.g.*, isovalerate, isobutyrate, 2-methylbutyrate, isopentanol, isobutanol, and 2-methylbutanol, in a cell, tissue, or organism. In some embodiments, the branched chain amino acid catabolism enzyme decreases the level of branched chain amino acid as compared to the level of a corresponding downstream metabolite, *e.g.*, isovalerate, isobutyrate, 2-methylbutyrate, isopentanol, isobutanol, and 2-methylbutanol, in a cell, tissue, or organism. In other embodiments, the branched chain amino acid catabolism enzyme increases the level of a downstream metabolite, *e.g.*, isovalerate, isobutyrate, 2-methylbutyrate, isopentanol, isobutanol, and 2-methylbutanol, as compared to the level of its corresponding branched chain amino acid in a cell, tissue, or organism.

[0141] In some embodiments, the branched chain amino acid catabolism enzyme is a leucine catabolism enzyme. In other embodiments, the branched chain amino acid catabolism enzyme is an isoleucine catabolism enzyme. In other embodiments, the branched chain amino acid catabolism enzyme is a valine catabolism enzyme. In some embodiments, the branched chain amino acid catabolism enzyme is involved in the catabolism of leucine, isoleucine, and valine. In another embodiment, the branched chain amino acid catabolism enzyme is involved in the catabolism of leucine and valine, isoleucine and valine, or leucine and isoleucine. In some embodiments, the branched chain amino acid catabolism enzyme converts leucine, isoleucine, and/or valine into its corresponding α -keto acid. In certain specific embodiments, the present disclosure provides an engineered bacterium comprising gene sequence(s) encoding one or more catabolism enzymes selected from leucine dehydrogenase (LeuDH), BCAA aminotransferase (IlvE), and/or amino acid oxidase (L-AAD).

[0142] In some embodiments, the branched chain amino acid catabolism enzyme is an alpha-ketoisocaproic acid (KIC) catabolism enzyme. In other embodiments, the branched chain amino acid catabolism enzyme is an alpha-ketoisovaleric acid (KIV) catabolism enzyme. In other embodiments, the branched chain amino acid catabolism enzyme is an alpha-keto-beta-methylvaleric acid (KMV) catabolism enzyme. In other embodiments, the branched chain amino acid catabolism enzyme is involved in the catabolism of alpha-ketoisocaproic acid (KIC), alpha-ketoisovaleric acid (KIV), and alpha-keto-beta-methylvaleric acid (KMV). In other embodiments, the branched chain amino acid catabolism enzyme is involved in the catabolism of KIC and KIV, KIC and KMV, or KIV and KMV. In some embodiments, the branched chain amino acid catabolism enzyme converts alpha-ketoisocaproic acid (KIC), alpha-ketoisovaleric acid (KIV), and/or alpha-keto-beta-methylvaleric acid (KMV) into its corresponding aldehyde, *e.g.*, isovaleraldehyde, isobutyraldehyde, and/or 2-methylbutyraldehyde. In some embodiments, the present disclosure provides an engineered bacterium comprising gene sequence(s) encoding KivD.

[0143] In one embodiment, the branched chain amino acid catabolism enzyme is an isovaleraldehyde catabolism enzyme. In another embodiment, the branched chain amino acid catabolism enzyme is an isobutyraldehyde catabolism enzyme. In another embodiment, the branched chain amino acid catabolism enzyme is 2-methylbutyraldehyde catabolism enzyme. In another embodiment, the branched chain amino acid catabolism enzyme is involved in the catabolism of isovaleraldehyde, isobutyraldehyde, and 2-methylbutyraldehyde. In another embodiment, the branched chain amino acid catabolism enzyme is involved in the catabolism of isovaleraldehyde and isobutyraldehyde, isovaleraldehyde and 2-methylbutyraldehyde, or isobutyraldehyde and 2-methylbutyraldehyde. In some embodiments, the present disclosure provides an engineered bacterium comprising gene sequence(s) encoding one or more alcohol dehydrogenase(s), *e.g.*, Ahd2.

[0144] In some embodiments, the present disclosure provides an engineered bacterium comprising gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s)

involved in the catabolism of leucine, isoleucine, and/or valine, and further comprises gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) involved in the catabolism of KIC, KIV, and/or KMV. In some embodiments, the present disclosure provides an engineered bacteria comprising gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) involved in the catabolism of leucine, isoleucine, and/or valine, further comprises gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) involved in the catabolism of KIC, KIV, and/or KMV, and further comprises gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) involved in the catabolism of isovaleraldehyde, isobutyraldehyde, and/or 2-methylbutyraldehyde. In some embodiments, the present disclosure provides an engineered bacterium comprising gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) selected from LeuDH, KivD, and Adh2.

[0145] In some embodiments, the present disclosure provides an engineered bacterium comprising gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) involved in the conversion of leucine, isoleucine, and/or valine to KIC, KIV, and/or KMV, respectively. In some embodiments, the present disclosure provides an engineered bacterium comprising gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) involved in the conversion of KIC, KIV, and/or KMV to isovaleraldehyde, isobutyraldehyde, and/or 2-methylbutyraldehyde, respectively. In some embodiments, the present disclosure provides an engineered bacterium comprising gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) involved in the conversion of isovaleraldehyde, isobutyraldehyde, and/or 2-methylbutyraldehyde to isovalerate, isobutyrate, and/or 2-methylbutyrate, respectively. In some embodiments, the present disclosure provides an engineered bacterium comprising gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) involved in the conversion of isovaleraldehyde, isobutyraldehyde, and/or 2-methylbutyraldehyde to isopentanol, isobutanol, and/or 2-methylbutanol respectively.

[0146] In some embodiments, the present disclosure provides an engineered bacteria comprising gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) involved in the conversion of leucine, isoleucine, and/or valine to KIC, KIV, and/or KMV, respectively, and further comprises gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) involved in the conversion of KIC, KIV, and/or KMV to isovaleraldehyde, isobutyraldehyde, and/or 2-methylbutyraldehyde, respectively. In some embodiments, the present disclosure provides an engineered bacterium comprising gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) involved in the conversion of KIC, KIV, and/or KMV to isovaleraldehyde, isobutyraldehyde, and/or 2-methylbutyraldehyde, respectively, and further comprises gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) involved in the conversion of isovaleraldehyde, isobutyraldehyde, and/or 2-methylbutyraldehyde to isovalerate, isobutyrate, and/or 2-methylbutyrate, respectively. In some embodiments, the present

disclosure provides an engineered bacteria comprising gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) involved in the conversion of KIC, KIV, and/or KMV to isovaleraldehyde, isobutyraldehyde, and/or 2-methylbutyraldehyde, respectively, and further comprises gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) involved in the conversion of isovaleraldehyde, isobutyraldehyde, and/or 2-methylbutyraldehyde to isopentanol, isobutanol, and/or 2-methylbutanol respectively.

[0147] In some embodiments, the present disclosure provides an engineered bacteria comprising gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) involved in the conversion of leucine, isoleucine, and/or valine to KIC, KIV, and/or KMV, respectively, further comprises gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) involved in the conversion of KIC, KIV, and/or KMV to isovaleraldehyde, isobutyraldehyde, and/or 2-methylbutyraldehyde, respectively, and further comprises gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) involved in the conversion of isovaleraldehyde, isobutyraldehyde, and/or 2-methylbutyraldehyde to isovalerate, isobutyrate, and/or 2-methylbutyrate, respectively. In some embodiments, the present disclosure provides an engineered bacteria comprising gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) involved in the conversion of leucine, isoleucine, and/or valine to KIC, KIV, and/or KMV, respectively, further comprises gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) involved in the conversion of KIC, KIV, and/or KMV to isovaleraldehyde, isobutyraldehyde, and/or 2-methylbutyraldehyde, respectively, and further comprises gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) involved in the conversion of isovaleraldehyde, isobutyraldehyde, and/or 2-methylbutyraldehyde to isopentanol, isobutanol, and/or 2-methylbutanol, respectively. In some embodiments, the present disclosure provides an engineered bacterium comprising gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) selected from LeuDH, KivD, and Adh2.

[0148] Enzymes involved in the catabolism of a branched chain amino acid may be expressed or modified in the bacteria disclosed herein to enhance catabolism of a branched chain amino acid, *e.g.*, leucine. Specifically, when a branched chain amino acid catabolism enzyme is expressed in the engineered bacteria disclosed herein, the engineered bacteria are able to convert (deaminate) more branched chain amino acids (*e.g.*, leucine, valine, isoleucine) into their respective alpha-keto acids (KIC, KIV, KMV) and/or convert more BCAA alpha-keto acids (*e.g.*, KIC, KIV, KMV) into respective BCAA-derived aldehydes (*e.g.*, isovaleraldehyde, isobutyraldehyde, 2-methylbutyraldehyde) and/or convert more BCAA-derived aldehydes into respective alcohols (*e.g.*, isopentanol, isobutanol, 2-methylbutanol) and/or convert more BCAA-derived aldehydes into respective carboxylic acids (isovalerate, isobutyrate, 2-methylbutyrate), and/or convert (decarboxylate) more branched chain alpha-keto acids into their respective acyl-CoA derivatives when the catabolism enzyme(s) is expressed, in comparison with unmodified bacteria of the same

bacterial subtype under the same conditions. Thus, for example, the genetically engineered bacteria comprising gene sequence encoding a branched chain amino acid catabolism enzyme can catabolize the branched chain amino acid, *e.g.*, leucine, and/or its corresponding alpha-keto acid, *e.g.*, alpha-ketoisocaproate, to treat diseases associated with catabolism of branched chain amino acids, such as Maple Syrup Urine Disease (MSUD) and others described herein.

[0149] In some embodiments, the engineered bacteria comprise gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) and gene sequence(s) encoding one or more transporter(s) capable of importing a BCAA or metabolite thereof. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain amino acid catabolism enzyme and gene sequence(s) encoding two or more copies of a transporter capable of importing a BCAA or metabolite thereof. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain amino acid catabolism enzyme and gene sequence(s) encoding two or more different transporter(s) capable of importing a BCAA or metabolite thereof. In certain embodiments, the transporter is a leucine transporter. In certain embodiments, the transporter is a valine transporter. In certain embodiments, the transporter is an isoleucine transporter. In certain embodiments, the transporter is a branched chain amino acid transporter, *e.g.*, capable of importing leucine, isoleucine, and valine. In certain specific embodiments, the transporter is selected from LivKHMGGF and BmQ.

[0150] In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain amino acid catabolism enzyme and gene sequence(s) encoding one or more BCAA binding proteins, *e.g.*, a BCAA binding protein that assists in bringing BCAA(s) into the bacterial cell. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain amino acid catabolism enzyme, gene sequence(s) encoding one or more transporter(s) capable of importing one or more BCAAs, and gene sequence(s) encoding one or more BCAA binding proteins, *e.g.*, a BCAA binding protein that assists in bringing BCAA(s) into the bacterial cell. In any of these embodiments, the engineered bacteria comprise gene sequence(s) encoding two or more copies of a BCAA binding protein. In any of these embodiments, the engineered bacteria comprise gene sequence(s) encoding two or more different BCAA binding proteins. In certain embodiments, the BCAA binding protein is LivJ.

[0151] In any of the embodiments described above and herein, the engineered bacteria may further comprise one or more genetic modification(s) that reduces export of a branched chain amino acid from the bacteria, *e.g.*, a deletion or mutation in at least one gene associated with the export of a BCAA, *e.g.*, deletion or mutation in *leuE* gene and/or its promoter (which reduces or eliminates the export of leucine). In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain amino acid catabolism enzyme and at least one genetic modification that reduces export of a branched chain amino acid. In certain specific embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain amino acid

catabolism enzyme, gene sequence(s) encoding one or more transporter(s) capable of importing a BCAA, and at least one genetic modification that reduces export of a branched chain amino acid. In certain specific embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain amino acid catabolism enzyme, gene sequence(s) encoding one or more transporter(s) capable of importing a BCAA, gene sequence(s) encoding one or more BCAA binding proteins, and at least one genetic modification that reduces export of a branched chain amino acid. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain amino acid catabolism enzyme, gene sequence(s) encoding one or more BCAA binding proteins, and at least one genetic modification that reduces export of a branched chain amino acid. In any of these embodiments, the genetic modification may be a deletion or mutation in one or more gene(s) that allow or assist in the export of a BCAA. In any of these embodiment, the genetic modification may be a deletion or mutation in a *leuE* gene and/or its promoter.

[0152] In any of the embodiments described above and herein, the engineered bacteria comprise gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s), and at least one genetic modification that reduces endogenous biosynthesis of a branched chain amino acid, for example, a deletion or mutation in at least one gene required for BCAA synthesis, *e.g.*, deletion or mutation in *ilvC* gene and/or its promoter, which gene is required for BCAA synthesis and whose absence creates an auxotroph requiring the bacterial cell to import leucine. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s), gene sequence(s) encoding one or more transporter(s) capable of importing a BCAA, and at least one genetic modification that reduces endogenous biosynthesis of a branched chain amino acid, for example, a deletion or mutation in at least one gene required for BCAA synthesis. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain amino acid catabolism enzyme, gene sequence(s) encoding one or more BCAA binding proteins, and at least one genetic modification that reduces endogenous biosynthesis of a branched chain amino acid. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain amino acid catabolism enzyme, at least one genetic modification that reduces export of a branched chain amino acid, and at least one genetic modification that reduces endogenous biosynthesis of a branched chain amino acid. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain amino acid catabolism enzyme, gene sequence(s) encoding one or more transporter(s) capable of importing a BCAA, gene sequence(s) encoding one or more BCAA binding proteins, and at least one genetic modification that reduces endogenous biosynthesis of a branched chain amino acid. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain amino acid catabolism enzyme, gene sequence(s) encoding one or more transporter(s) capable of importing a BCAA, at least one genetic modification that reduces export of a branched chain amino acid, and at least one genetic modification that reduces endogenous

biosynthesis of a branched chain amino acid. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain amino acid catabolism enzyme, gene sequence(s) encoding one or more BCAA binding proteins, at least one genetic modification that reduces export of a branched chain amino acid, and at least one genetic modification that reduces endogenous biosynthesis of a branched chain amino acid. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain amino acid catabolism enzyme, gene sequence(s) encoding one or more transporter(s) capable of importing a BCAA, gene sequence(s) encoding one or more BCAA binding proteins, at least one genetic modification that reduces export of a branched chain amino acid, and at least one genetic modification that reduces endogenous biosynthesis of a branched chain amino acid. In any of these embodiments, the at least one genetic modification that reduces endogenous biosynthesis of a branched chain amino acid can be a deletion or mutation in at least one gene required for BCAA synthesis, *e.g.*, deletion or mutation in *ilvC* gene and/or its promoter.

[0153] In any of the embodiments described above and herein, the gene sequence(s) encoding at least one branched chain amino acid catabolism enzyme, and/or gene sequence(s) encoding one or more transporter(s) capable of importing a BCAA, and/or gene sequence(s) encoding one or more BCAA binding proteins, and/or other sequence can be present in the bacterial chromosome. In any of the embodiments described above and herein, the gene sequence(s) encoding at least one branched chain amino acid catabolism enzyme, and/or gene sequence(s) encoding one or more transporter(s) capable of importing a BCAA, and/or gene sequence(s) encoding one or more BCAA binding proteins, and/or other sequence can be present in one or more plasmids.

[0154] In some embodiments, the engineered bacteria comprise gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) in which the one or more enzymes are from a different organism, *e.g.*, a different species of bacteria. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) in which the one or more enzymes are native to the bacterium. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) in which one or more of the enzymes are native and one or more of the enzymes are from a different organism, *e.g.*, a different species of bacteria. In other embodiments, the bacterial cell comprises more than one copy of a native gene encoding a branched chain amino acid catabolism enzyme. In other embodiments, the bacterial cell comprises more than one copy of a non-native gene encoding a branched chain amino acid catabolism enzyme. In other embodiments, the bacterial cell comprises at least one, two, three, four, five, six or more copies of a gene encoding a branched chain amino acid catabolism enzyme, which can be native or non-native. In other embodiments, the bacterial cell comprises multiple copies of a gene encoding a branched chain amino acid catabolism enzyme. In some embodiments, the bacterial cell comprises gene sequence(s) encoding multiple copies of two or more different branched chain amino acid catabolism enzymes.

[0155] In some embodiments, the engineered bacteria comprise gene sequence(s) encoding one or more branched chain amino acid transporters in which the one or more transporters are from a different organism, *e.g.*, a different species of bacteria. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding one or more branched chain amino acid transporter(s) in which the one or more transporters are native to the bacterium. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding one or more branched chain amino acid transporters(s) in which one or more of the transporters are native and one or more of the transporters are from a different organism, *e.g.*, a different species of bacteria. In other embodiments, the bacterial cell comprises more than one copy of a native gene encoding a branched chain amino acid transporter. In other embodiments, the bacterial cell comprises more than one copy of a non-native gene encoding a branched chain amino acid transporter. In other embodiments, the bacterial cell comprises at least one, two, three, four, five, six or more copies of a gene encoding a branched chain amino acid transporter, which can be native or non-native. In other embodiments, the bacterial cell comprises multiple copies of a gene encoding a branched chain amino acid transporters. In some embodiments, the bacterial cell comprises gene sequence(s) encoding multiple copies of two or more different branched chain amino acid transporters.

[0156] In some embodiments, the engineered bacteria comprise gene sequence(s) encoding one or more branched chain amino acid binding protein(s) in which the one or more binding protein(s) are from a different organism, *e.g.*, a different species of bacteria. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding one or more branched chain amino acid binding protein(s) in which the one or more binding protein(s) are native to the bacterium. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding one or more branched chain amino acid binding protein(s) in which one or more of the binding protein(s) are native and one or more of the binding protein(s) are from a different organism, *e.g.*, a different species of bacteria. In other embodiments, the bacterial cell comprises more than one copy of a native gene encoding a branched chain amino acid binding protein. In other embodiments, the bacterial cell comprises more than one copy of a non-native gene encoding a branched chain amino acid binding protein. In other embodiments, the bacterial cell comprises at least one, two, three, four, five, six or more copies of a gene encoding a branched chain amino acid binding protein, which can be native or non-native. In other embodiments, the bacterial cell comprises multiple copies of a gene encoding a branched chain amino acid binding protein. In some embodiments, the bacterial cell comprises gene sequence(s) encoding multiple copies of two or more different branched chain amino acid binding proteins.

[0157] Branched chain amino acid catabolism enzymes are known in the art. In some embodiments, the branched chain amino acid catabolism enzyme is encoded by a gene encoding a branched chain amino acid catabolism enzyme derived from a bacterial species. In some embodiments, a branched chain amino acid catabolism enzyme is encoded by a gene encoding a branched chain amino acid catabolism enzyme derived from a non-bacterial species. In some

embodiments, a branched chain amino acid catabolism enzyme is encoded by a gene derived from a eukaryotic species, *e.g.*, a yeast species or a plant species. In one embodiment, a branched chain amino acid catabolism enzyme is encoded by a gene derived from a mammalian species, *e.g.*, a human. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain amino acid catabolism enzyme derived from a bacterial species and at least one branched chain amino acid catabolism enzyme derived from a non-bacterial species. In one embodiment, the gene encoding the branched chain amino acid catabolism enzyme is derived from an organism of the genus or species that includes, but is not limited to, *Acetobacter*, *Azospirillum*, *Bacillus*, *Bacteroides*, *Bifidobacterium*, *Brevibacteria*, *Burkholderia*, *Citrobacter*, *Clostridium*, *Corynebacterium*, *Cronobacter*, *Enterobacter*, *Enterococcus*, *Erwinia*, *Helicobacter*, *Klebsiella*, *Lactobacillus*, *Lactococcus*, *Leishmania*, *Listeria*, *Macrococcus*, *Mycobacterium*, *Nakamurella*, *Nasonia*, *Nostoc*, *Pantoea*, *Pectobacterium*, *Pseudomonas*, *Psychrobacter*, *Ralstonia*, *Saccharomyces*, *Salmonella*, *Sarcina*, *Serratia*, *Staphylococcus*, and *Yersinia*, *e.g.*, *Acetobacter radioresistens*, *Acetobacter baumannii*, *Acetobacter calcoaceticus*, *Azospirillum brasilense*, *Bacillus anthracis*, *Bacillus cereus*, *Bacillus coagulans*, *Bacillus megaterium*, *Bacillus subtilis*, *Bacillus thuringiensis*, *Bacteroides fragilis*, *Bacteroides subtilis*, *Bacteroides thetaiotaomicron*, *Bifidobacterium bifidum*, *Bifidobacterium infantis*, *Bifidobacterium lactis*, *Bifidobacterium longum*, *Burkholderia xenovorans*, *Citrobacter youngae*, *Citrobacter koseri*, *Citrobacter rodentium*, *Clostridium acetobutylicum*, *Clostridium butyricum*, *Corynebacterium aurimucosum*, *Corynebacterium kroppenstedtii*, *Corynebacterium striatum*, *Cronobacter sakazakii*, *Cronobacter turicensis*, *Enterobacter cloacae*, *Enterobacter cancerogenus*, *Enterococcus faecium*, *Erwinia amylovora*, *Erwinia pyrifoliae*, *Erwinia tasmaniensis*, *Helicobacter mustelae*, *Klebsiella pneumonia*, *Klebsiella variicola*, *Lactobacillus acidophilus*, *Lactobacillus bulgaricus*, *Lactobacillus casei*, *Lactobacillus johnsonii*, *Lactobacillus paracasei*, *Lactobacillus plantarum*, *Lactobacillus reuteri*, *Lactobacillus rhamnosus*, *Lactococcus lactis*, *Leishmania infantum*, *Leishmania major*, *Leishmania braziliensis*, *Listeria grayi*, *Macrococcus caseolyticus*, *Mycobacterium avium*, *Mycobacterium intracellulare*, *Mycobacterium kansasii*, *Mycobacterium leprae*, *Mycobacterium marinum*, *Mycobacterium smegmatis*, *Mycobacterium tuberculosis*, *Mycobacterium ulcerans*, *Nakamurella multipartita*, *Nasonia vitipennis*, *Nostoc punctiforme*, *Pantoea ananatis*, *Pantoea agglomerans*, *Pectobacterium atrosepticum*, *Pectobacterium carotovorum*, *Pseudomonas aeruginosa*, *Psychrobacter articus*, *Psychrobacter cryohalolentis*, *Ralstonia eutropha*, *Saccharomyces boulardii*, *Salmonella enterica*, *Sarcina ventriculi*, *Serratia odorifera*, *Serratia proteamaculans*, *Staphylococcus aerus*, *Staphylococcus capitis*, *Staphylococcus carnosus*, *Staphylococcus epidermidis*, *Staphylococcus hominis*, *Staphylococcus haemolyticus*, *Staphylococcus lugdunensis*, *Staphylococcus saprophyticus*, *Staphylococcus warneri*, *Yersinia enterocolitica*, *Yersinia mollaretii*, *Yersinia kristensenii*, *Yersinia rohdei*, and *Yersinia aldovae*.

[0158] In some embodiments, the gene encoding a branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence has been codon-optimized

for use in the host organism. In one embodiment, the gene encoding a branched chain amino acid catabolism enzyme has been codon-optimized for use in *Escherichia coli*.

[0159] In some embodiments, the branched chain amino acid catabolism enzyme converts a branched chain amino acid to its corresponding alpha-keto acid. Such enzymes include, for example, LeuDH. In other embodiments, the branched chain amino acid catabolism enzyme converts a branched chain keto acid to its corresponding aldehyde. For example, in some embodiments, the branched chain amino acid catabolism enzyme is an α -ketoisovalerate decarboxylase (KivD).

[0160] In other embodiments, the branched chain amino acid catabolism enzyme converts an aldehyde into its corresponding alcohol. For example, the branched chain amino acid catabolism enzyme may be an alcohol dehydrogenase. In one embodiment, the alcohol dehydrogenase is Adh2.

[0161] In some embodiments, the engineered bacteria comprise gene sequence(s) encoding one or more transporter(s) capable of importing a BCAA or metabolite thereof. In certain embodiments, the transporter is a leucine transporter. In certain embodiments, the transporter is a valine transporter. In certain embodiments, the transporter is an isoleucine transporter. In certain embodiments, the transporter is a branched chain amino acid transporter, *e.g.*, capable of importing leucine, isoleucine, and valine. The term “BCAA transporter” is meant to refer to a transporter that specifically transports leucine, isoleucine, or valine, and also to a transporter that is able to transport any BCAA, including for example, the ability to transport leucine, isoleucine, and valine. For example, in some embodiments, the transporter is LivKHMGE. In some embodiments, the transporter is BrnQ.

[0162] The present disclosure further comprises genes encoding functional fragments of a branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence or functional variants of a branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence. As used herein, the term “functional fragment thereof” or “functional variant thereof” of a branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence refers to fragment or variant sequence having qualitative biological activity in common with the wild-type branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence from which the fragment or variant was derived. For example, a functional fragment or a functional variant of a mutated branched chain amino acid catabolism enzyme is one which retains essentially the same ability to catabolize a branched chain amino acid and/or its corresponding alpha-keto acid or aldehyde or other metabolite as the branched chain amino acid catabolism enzyme from which the functional fragment or functional variant was derived. For example, a polypeptide having branched chain amino acid catabolism enzyme activity may be truncated at the N-terminus or C-terminus and the retention of branched chain amino acid catabolism enzyme activity assessed using assays known to those of skill in the art, including the exemplary assays provided herein. In one embodiment, the recombinant bacterial cell disclosed herein comprises a heterologous gene encoding a branched chain amino acid

catabolism enzyme functional variant. In another embodiment, the recombinant bacterial cell disclosed herein comprises a heterologous gene encoding a branched chain amino acid catabolism enzyme functional fragment.

[0163] The present disclosure encompasses genes encoding a branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence comprising amino acids in its sequence that are substantially the same as an amino acid sequence described herein. Amino acid sequences that are substantially the same as the sequences described herein include sequences comprising conservative amino acid substitutions, as well as amino acid deletions and/or insertions. A conservative amino acid substitution refers to the replacement of a first amino acid by a second amino acid that has chemical and/or physical properties (*e.g.*, charge, structure, polarity, hydrophobicity/hydrophilicity) that are similar to those of the first amino acid. Conservative substitutions include replacement of one amino acid by another within the following groups: lysine (K), arginine (R) and histidine (H); aspartate (D) and glutamate (E); asparagine (N), glutamine (Q), serine (S), threonine (T), tyrosine (Y), K, R, H, D and E; alanine (A), valine (V), leucine (L), isoleucine (I), proline (P), phenylalanine (F), tryptophan (W), methionine (M), cysteine (C) and glycine (G); F, W and Y; C, S and T. Similarly contemplated is replacing a basic amino acid with another basic amino acid (*e.g.*, replacement among Lys, Arg, His), replacing an acidic amino acid with another acidic amino acid (*e.g.*, replacement among Asp and Glu), replacing a neutral amino acid with another neutral amino acid (*e.g.*, replacement among Ala, Gly, Ser, Met, Thr, Leu, Ile, Asn, Gln, Phe, Cys, Pro, Trp, Tyr, Val).

[0164] The present disclosure encompasses branched chain amino acid catabolism enzymes, BCAA transporters, BCAA binding proteins, and/or other sequences which have a certain percent identity to a gene or protein sequence described herein. For example, the disclosure encompasses branched chain amino acid catabolism enzymes, BCAA transporters, BCAA binding proteins, and/or other sequences having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to a nucleic acid sequence or amino acid sequence disclosed herein. As used herein, the term “percent (%) sequence identity” or “percent (%) identity,” also including “homology,” is defined as the percentage of amino acid residues or nucleotides in a candidate sequence that are identical with the amino acid residues or nucleotides in the reference sequences after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Optimal alignment of the sequences for comparison may be produced, besides manually, by means of the local homology algorithm of Smith and Waterman, 1981, *Adv. App. Math.* 2, 482, by means of the local homology algorithm of Needleman and Wunsch, 1970, *J. Mol. Biol.* 48, 443, by means of the similarity search method of Pearson and Lipman, 1988, *Proc. Natl. Acad. Sci. USA* 85, 2444, or by means of computer programs which use these algorithms (GAP, BESTFIT, FASTA,

BLAST P, BLAST N and TFASTA in Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Drive, Madison, Wis.).

[0165] Assays for testing the activity of a branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence, or a functional variant, or a functional fragment thereof are well known to one of ordinary skill in the art. For example, branched chain amino acid catabolism, BCAA transporter, BCAA binding protein, and/or other sequence can be assessed by expressing the protein, functional variant, or fragment thereof, in a recombinant bacterial cell that lacks endogenous branched chain amino acid catabolism enzyme activity. Branched chain amino acid catabolism can be assessed using the coupled enzymatic assay method as described by Zhang *et al.* (see, for example, Zhang *et al.*, *Proc. Natl. Acad. Sci.*, 105(52):20653-58, 2008). Furthermore, catabolism of branched chain amino acids can also be assessed *in vitro* by measuring the disappearance of alpha-ketoisovalerate as described by de la Plaza (see, for example, de la Plaza *et al.*, *FEMS Microbiol. Letters*, 2004, 238(2):367-374). BCAA as well as the branched chain keto acid can be quantified by liquid chromatography tandem mass spectrometry (LC-MS/MS), as described herein.

[0166] In some embodiments, the gene encoding a branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence is mutagenized; mutants exhibiting increased activity are selected; and the mutagenized gene encoding the branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence is isolated and inserted into the bacterial cell. In some embodiments, the gene encoding an α -ketoisovalerate decarboxylase, *e.g.*, *kivD*, is mutagenized; mutants exhibiting decreased activity are selected; and the mutagenized gene encoding the α -ketoisovalerate decarboxylase, *e.g.*, *kivD*, is isolated and inserted into the bacterial cell. The gene comprising the modifications described herein may be present on a plasmid or chromosome.

[0167] In some embodiments, the gene encoding the branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence is directly operably linked to a first promoter. In other embodiments, the gene encoding the branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence is indirectly operably linked to a first promoter. In some embodiment, the gene encoding the branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence is operably linked to a promoter that is not its native promoter.

[0168] In some embodiments, the gene encoding the branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence is expressed under the control of a constitutive promoter. In other embodiments, the gene encoding the branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence is expressed under the control of an inducible promoter. In some embodiments, the gene encoding the branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or

other sequence is expressed under the control of a promoter that is directly or indirectly induced by exogenous environmental conditions. In some embodiments, the gene encoding the branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence is expressed under the control of a promoter that is directly or indirectly induced by low-oxygen or anaerobic conditions, wherein expression of the gene encoding the branched chain amino acid catabolism enzyme is activated under low-oxygen or anaerobic environments, such as the environment of the mammalian gut. In some embodiments, the gene encoding the branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence is expressed under the control of a promoter that is directly or indirectly induced by inflammatory conditions. Exemplary inducible promoters described herein include oxygen level-dependent promoters (*e.g.*, FNR-inducible promoter), promoters induced by inflammation or an inflammatory response (RNS, ROS promoters), and promoters induced by a metabolite that may or may not be naturally present (*e.g.*, can be exogenously added) in the gut, *e.g.*, arabinose and tetracycline. Examples of inducible promoters include, but are not limited to, an FNR responsive promoter, a P_{araC} promoter, a P_{araBAD} promoter, and a P_{TetR} promoter, each of which are described in more detail herein. Examples of other inducible promoters are provided herein below.

[0169] The gene encoding the branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence may be present on a plasmid or chromosome in the bacterial cell. In some embodiments, the gene encoding the branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence is located on a plasmid in the bacterial cell. In other embodiments, the gene encoding the branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence is located in the chromosome of the bacterial cell. In other embodiments, a native copy of the gene encoding the branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence is located in the chromosome of the bacterial cell, and a gene encoding a branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence from the same or a different species of bacteria is located on a plasmid in the bacterial cell. In other embodiments, a native copy of the gene encoding the branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence is located on a plasmid in the bacterial cell, and a gene encoding the branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence from a different species of bacteria is located on a plasmid in the bacterial cell. In other embodiments, a native copy of the gene encoding the branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence is located in the chromosome of the bacterial cell, and a gene encoding the branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence from a different species of bacteria is located in the chromosome of the bacterial cell.

[0170] In some embodiments, the gene encoding the branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence is expressed on a low-copy plasmid. In some embodiments, the gene encoding the branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence is expressed on a high-copy plasmid. In some embodiments, the high-copy plasmid may be useful for increasing expression of the branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, and/or other sequence, thereby increasing the catabolism of the branched chain amino acid, *e.g.*, leucine.

[0171] In some embodiments, the engineered bacteria convert the branched chain amino acid(s) and/or corresponding alpha-keto acid(s) and/or other corresponding metabolite(s) to a non-toxic or low toxicity metabolite, *e.g.*, isovaleraldehyde, isobutyraldehyde, 2-methylbutyraldehyde, isovaleric acid, isobutyric acid, 2-methylbutyric acid, isopentanol, isobutanol, and 2-methylbutanol. **Table 3** chart showing that the products of BCAA degradation by the engineered bacteria have very low oral toxicity.

Table 3. Toxicity of BCAA degradation products

Compound	Oral LD50 (rat) (mg/kg)	Oral NOAEL* (rat)(mg/kg/d)
Isovaleraldehyde	5740	N/D
Isobutyraldehyde	3730	N/D
2-methylbutyraldehyde	6884	N/D
Isovaleric acid	2500	N/D
Isobutyric acid	2230	N/D
2-methylbutyric acid	1750	N/D
Isopentanol	>5000	1250
Isobutanol	3350	>1450
2-methylbutanol	4170	N/D

* No-Observed-Adverse-Effect

A. Optimized Branched chain ketoacid decarboxylase

[0172] In one embodiment, the branched chain amino acid catabolism enzyme is an optimized branched chain ketoacid decarboxylase, including but not limited to, KivD. In a non-limiting example, KivD is from *Lactococcus lactis*. Thus, in some embodiments, the engineered bacteria comprise gene sequence(s) encoding one or more copies of a branched chain ketoacid decarboxylase. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding one, two, three, four, five, six, or more copies of a branched chain ketoacid decarboxylase. The one or more copies of a branched chain ketoacid decarboxylase can be one or more copies of the same gene or can be different genes encoding α -ketoisovalerate decarboxylase, *e.g.*, gene(s) from a different species or otherwise having a different gene sequence. The one or more copies of a

branched chain ketoacid decarboxylase can be present in the bacterial chromosome or can be present in one or more plasmids.

[0173] As used herein “ α -ketoisovalerate decarboxylase” or “alpha-ketoisovalerate decarboxylase” or “branched-chain α -keto acid decarboxylase” or “ α -ketoacid decarboxylase” or “branched chain ketoacid decarboxylase” or “2-ketoisovalerate decarboxylase” (referred to herein also as KivD or ketoisovalerate decarboxylase) refers to any polypeptide having enzymatic activity that catalyzes the conversion of a branched chain alpha-keto acid (BCKA), such as α -ketoisovalerate (2-oxoisopentanoate), α -ketomethylvalerate (3-methyl-2-oxopentanoate), or α -ketoisocaproate (4-methyl-2-oxopentanoate), to its corresponding aldehyde, such as isobutyraldehyde, 2-methylbutyraldehyde, or isovaleraldehyde, and carbon dioxide. Branched chain ketoacid decarboxylase enzymes are available from many microorganism sources, including those disclosed herein. Branched chain ketoacid decarboxylase employs the co-factor thiamine diphosphate (also known as thiamine pyrophosphate or “TPP” or “TDP”). Thiamine is the vitamin form of the co-factor which, when transported into a cell, is converted to thiamine diphosphate. Alpha-ketoisovalerate decarboxylase also employs Mg^{2+} . Branched chain ketoacid decarboxylase may be a homotetramer.

[0174] The bacterial cells disclosed herein may comprise a heterologous gene encoding a branched chain ketoacid decarboxylase enzyme and are capable of converting α -keto acids into aldehydes. For example, the branched chain ketoacid decarboxylase enzyme KivD is capable of metabolizing leucine (see, for example, de la Plaza *et al.*, *FEMS Microbiol. Lett.* 2004, 238(2):367-374), and a cytosolically active KivD should generally exhibit the ability to convert ketoisovalerate, ketomethylvalerate, and ketoisocaproate to isobutyraldehyde, 2-methylbutyraldehyde, and isovaleraldehyde.

[0175] Multiple distinct a branched chain ketoacid decarboxylase proteins are known in the art (see, *e.g.*, US Pat. Appl. Publ. No. 2013/0203138, the entire contents of which are incorporated herein by reference). In some embodiments, branched chain ketoacid decarboxylase is encoded by a branched chain ketoacid decarboxylase gene derived from a bacterial species. In some embodiments, a branched chain ketoacid decarboxylase is encoded by a branched chain ketoacid decarboxylase gene derived from a non-bacterial species. In some embodiments, a branched chain ketoacid decarboxylase is encoded by a branched chain ketoacid decarboxylase gene derived from a eukaryotic species, *e.g.*, a yeast species or a plant species. In some embodiments, a branched chain ketoacid decarboxylase is encoded by a branched chain ketoacid decarboxylase gene derived from a mammalian species. In one embodiment, the a branched chain ketoacid decarboxylase gene is derived from an organism of the genus or species that includes, but is not limited to, *Acetivibrio*, *Azospirillum*, *Bacillus*, *Bacteroides*, *Bifidobacterium*, *Brevibacterium*, *Burkholderia*, *Citrobacter*, *Clostridium*, *Corynebacterium*, *Cronobacter*, *Enterobacter*, *Enterococcus*, *Erwinia*, *Helicobacter*, *Klebsiella*, *Lactobacillus*, *Lactococcus*, *Leishmania*, *Listeria*, *Macrococcus*, *Mycobacterium*, *Nakamurella*, *Nasonia*, *Nostoc*, *Pantoea*, *Pectobacterium*, *Psychrobacter*, *Ralstonia*, *Saccharomyces*, *Salmonella*,

Sarcina, *Serratia*, *Staphylococcus*, and *Yersinia*, e.g., *Acetobacter radioresistens*, *Acetobacter baumannii*, *Acetobacter calcoaceticus*, *Azospirillum brasilense*, *Bacillus anthracis*, *Bacillus cereus*, *Bacillus coagulans*, *Bacillus megaterium*, *Bacillus subtilis*, *Bacillus thuringiensis*, *Bacteroides fragilis*, *Bacteroides subtilis*, *Bacteroides thetaiotaomicron*, *Bifidobacterium bifidum*, *Bifidobacterium infantis*, *Bifidobacterium lactis*, *Bifidobacterium longum*, *Burkholderia xenovorans*, *Citrobacter youngae*, *Citrobacter koseri*, *Citrobacter rodentium*, *Clostridium acetobutylicum*, *Clostridium butyricum*, *Corynebacterium aurimucosum*, *Corynebacterium kroppenstedtii*, *Corynebacterium striatum*, *Cronobacter sakazakii*, *Cronobacter turicensis*, *Enterobacter cloacae*, *Enterobacter cancerogenus*, *Enterococcus faecium*, *Erwinia amylovora*, *Erwinia pyrifoliae*, *Erwinia tasmaniensis*, *Helicobacter mustelae*, *Klebsiella pneumonia*, *Klebsiella variicola*, *Lactobacillus acidophilus*, *Lactobacillus bulgaricus*, *Lactobacillus casei*, *Lactobacillus johnsonii*, *Lactobacillus paracasei*, *Lactobacillus plantarum*, *Lactobacillus reuteri*, *Lactobacillus rhamnosus*, *Lactococcus lactis*, *Leishmania infantum*, *Leishmania major*, *Leishmania braziliensis*, *Listeria grayi*, *Macrococcus caseolyticus*, *Mycobacterium avium*, *Mycobacterium intracellulare*, *Mycobacterium kansasii*, *Mycobacterium leprae*, *Mycobacterium marinum*, *Mycobacterium smegmatis*, *Mycobacterium tuberculosis*, *Mycobacterium ulcerans*, *Nakamurella multipartita*, *Nasonia vitipennis*, *Nostoc punctiforme*, *Pantoea ananatis*, *Pantoea agglomerans*, *Pectobacterium atrosepticum*, *Pectobacterium carotovorum*, *Psychrobacter articus*, *Psychrobacter cryohalolentis*, *Ralstonia eutropha*, *Saccharomyces boulardii*, *Salmonella enterica*, *Sarcina ventriculi*, *Serratia odorifera*, *Serratia proteamaculans*, *Staphylococcus aerus*, *Staphylococcus capitis*, *Staphylococcys carnosus*, *Staphylococcus epidermidis*, *Staphylococcus hominis*, *Staphylococcus haemolyticus*, *Staphylococcus lugdunensis*, *Staphylococcus saprophyticus*, *Staphylococcus warneri*, *Yersinia enterocolitica*, *Yersinia mollaretii*, *Yersinia kristensenii*, *Yersinia rohdei*, and *Yersinia aldovae*. In some embodiments, the branched chain ketoacid decarboxylase is encoded by an a branched chain ketoacid decarboxylase gene derived from *Lactococcus lactis*, e.g., IFPL730. In another embodiment, the branched chain ketoacid decarboxylase, e.g., *kivD* gene, is derived from *Enterobacter cloacae* (Accession No. P23234.1), *Mycobacterium smegmatis* (Accession No. A0R480.1), *Mycobacterium tuberculosis* (Accession NO. O53865.1), *Mycobacterium avium* (Accession No. Q742Q2.1), *Azospirillum brasilense* (Accession No. P51852.1), or *Bacillus subtilis* (see Oku et al., *J. Biol. Chem.* 263: 18386-96, 1988).

[0176] In one embodiment, the branched chain ketoacid decarboxylase gene has been codon-optimized for use in the recombinant bacterial cell. In one embodiment, the branched chain ketoacid decarboxylase gene has been codon-optimized for use in *Escherichia coli*.

[0177] In one embodiment, the branched chain ketoacid decarboxylase gene is a *kivD* gene. In another embodiment, the *kivD* gene is a *Lactococcus lactis* *kivD* gene or *kivD* gene derived from *Lactococcus lactis*. When a branched chain ketoacid decarboxylase is expressed in the recombinant bacterial cells disclosed herein, the bacterial cells catabolize more branched chain amino acid, e.g.,

leucine, than unmodified bacteria of the same bacterial subtype under the same conditions (*e.g.*, culture or environmental conditions). Thus, the genetically engineered bacteria comprising a heterologous gene encoding a branched chain ketoacid decarboxylase may be used to catabolize excess branched chain amino acids, *e.g.*, leucine, to treat a disease associated with the catabolism of a branched chain amino acid, including Maple Syrup Urine Disease (MSUD).

[0178] The present disclosure further comprises genes encoding functional fragments of a branched chain ketoacid decarboxylase or functional variants of a branched chain ketoacid decarboxylase gene. As used herein, the term “functional fragment thereof” or “functional variant thereof” of a branched chain ketoacid decarboxylase gene relates to a sequence having qualitative biological activity in common with the wild-type branched chain ketoacid decarboxylase from which the fragment or variant was derived. For example, a functional fragment or a functional variant of a mutated branched chain ketoacid decarboxylase protein is one which retains essentially the same ability to catabolize BCKAs as a branched chain ketoacid decarboxylase protein from which the functional fragment or functional variant was derived. For example, a polypeptide having branched chain ketoacid decarboxylase activity may be truncated at the N-terminus or C-terminus and the retention of branched chain ketoacid decarboxylase activity assessed using assays known to those of skill in the art, including the exemplary assays provided herein. In one embodiment, the recombinant bacterial cell disclosed herein comprises a heterologous gene encoding a branched chain ketoacid decarboxylase functional variant. In another embodiment, the recombinant bacterial cell disclosed herein comprises a heterologous gene encoding a branched chain ketoacid decarboxylase functional fragment.

[0179] Assays for testing the activity of a branched chain ketoacid decarboxylase, a branched chain ketoacid decarboxylase functional variant, or a branched chain ketoacid decarboxylase functional fragment are well known to one of ordinary skill in the art. For example, branched chain ketoacid decarboxylase activity can be assessed by expressing the protein, functional variant, or fragment thereof, in a recombinant bacterial cell that lacks endogenous branched chain ketoacid decarboxylase activity. Also, branched chain ketoacid decarboxylase activity can be assessed using the coupled enzymatic assay method as described by Zhang *et al.* (see, for example, Zhang *et al.*, *Proc. Natl. Acad. Sci.*, 105(52):20653-58, 2008). Alpha-ketoisovalerate decarboxylase activity can also be assessed *in vitro* by measuring the disappearance of alpha-ketoisovalerate as described by de la Plaza (see, for example, de la Plaza *et al.*, *FEMS Microbiol. Letters*, 2004, 238(2):367-374).

[0180] In some embodiments, the gene encoding a branched chain ketoacid decarboxylase, *e.g.*, *kivD*, is mutagenized; mutants exhibiting increased activity are selected; and the mutagenized gene encoding the α -ketoisovalerate decarboxylase, *e.g.*, *kivD*, is isolated and inserted into the bacterial cell. The gene comprising the modifications described herein may be present on a plasmid or chromosome.

[0181] Accordingly, in some embodiments, the *kivD* gene has at least about 80% identity with the entire sequence of any one of SEQ ID NOs: 73-96. Accordingly, in one embodiment, the *kivD* gene has at least about 90% identity with the entire sequence of any one of SEQ ID NOs: 73-96. Accordingly, in one embodiment, the *kivD* gene has at least about 95% identity with the entire sequence of any one of SEQ ID NOs: 73-96. Accordingly, in one embodiment, the *kivD* gene has at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity with the entire sequence of any one of SEQ ID NOs: 73-96. In another embodiment, the *kivD* gene comprises the sequence of any one of SEQ ID NOs: 73-96. In yet another embodiment, the *kivD* gene consists of any one of SEQ ID NOs: 73-96.

[0182] In other embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain ketoacid decarboxylase enzyme and further comprise gene sequence encoding one or more polypeptides selected from other branched chain amino acid catabolism enzyme(s), BCAA transporter(s), and BCAA binding protein(s). Thus, in some embodiments, the at least one branched chain ketoacid decarboxylase enzyme is coexpressed with an additional branched chain amino acid catabolism enzyme, *e.g.*, a branched chain amino acid dehydrogenase, amino acid oxidase (also known as amino acid deaminase), and/or aminotransferase. In some embodiments, the at least one α -ketoisovalerate decarboxylase gene is coexpressed with a leucine dehydrogenase, *e.g.*, (*leuDH* or *ldh*), described in more detail below. In other embodiments, the at least one branched chain ketoacid decarboxylase gene is coexpressed with a branched chain amino acid aminotransferase, *e.g.*, *ilvE*, described in more detail below. In other embodiments, the at least one branched chain ketoacid decarboxylase gene is coexpressed with an amino acid deaminase, *e.g.*, *L-AAD*, described in more detail below. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain ketoacid decarboxylase enzyme and gene sequence(s) encoding one or more branched chain amino acid dehydrogenase(s) (*e.g.*, *leuDH*). In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain ketoacid decarboxylase enzyme and gene sequence(s) encoding one or more aminotransferase(s) (*e.g.*, *ilvE*).

[0183] In some embodiments, the at least one branched chain ketoacid decarboxylase enzyme is coexpressed with an alcohol dehydrogenase, *e.g.*, *adh2*, described in more detail below. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain ketoacid decarboxylase enzyme and gene sequence(s) encoding one or more alcohol dehydrogenase(s) (*e.g.*, *adh2*).

[0184] In some embodiments, the at least one α -ketoisovalerate decarboxylase gene is coexpressed with a leucine dehydrogenase, *e.g.*, *leuDH*, and an alcohol dehydrogenase, *e.g.*, *adh2*.

[0185] In some embodiments, the at least one branched chain ketoacid decarboxylase enzyme is coexpressed with one or more BCAA transporter(s), for example, a high affinity leucine transporter, *e.g.*, LivKHMgf and/or low affinity BCAA transporter, *e.g.*, BrnQ. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched

chain ketoacid decarboxylase enzyme and gene sequence(s) encoding one or more BCAA transporter(s) (*e.g.*, *livKHMGE*, *brnQ*).

[0186] In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain ketoacid decarboxylase enzyme and genetic modification that reduces export of a branched chain amino acid, *e.g.*, a genetic mutation in a *leuE* gene or promoter thereof. In one embodiment, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain ketoacid decarboxylase enzyme and a genetic modification that reduces or eliminates branched chain amino acid synthesis, *e.g.*, a genetic mutation in a *ilvC* gene or promoter thereof.

[0187] In some embodiments, the gene sequence(s) encoding the one or more branched chain ketoacid decarboxylase enzyme(s) is expressed under the control of a constitutive promoter. In some embodiments, the gene sequence(s) encoding the one or more branched chain ketoacid decarboxylase enzyme(s) is expressed under the control of an inducible promoter. In some embodiments, the gene sequence(s) encoding the one or more branched chain ketoacid decarboxylase enzyme(s) is expressed under the control of a promoter that is directly or indirectly induced by exogenous environmental conditions. In some embodiments, the gene sequence(s) encoding the one or more branched chain ketoacid decarboxylase enzyme(s) is expressed under the control of a promoter that is directly or indirectly induced by low-oxygen or anaerobic conditions, wherein expression of the gene encoding the branched chain ketoacid decarboxylase enzyme is activated under low-oxygen or anaerobic environments, such as the environment of the mammalian gut. In some embodiments, the gene sequence(s) encoding the one or more branched chain ketoacid decarboxylase enzyme(s) is expressed under the control of a promoter that is directly or indirectly induced by inflammatory conditions. Exemplary inducible promoters described herein include oxygen level-dependent promoters (*e.g.*, FNR-inducible promoter), promoters induced by inflammation or an inflammatory response (RNS, ROS promoters), and promoters induced by a metabolite that may or may not be naturally present (*e.g.*, can be exogenously added) in the gut, *e.g.*, arabinose and tetracycline. Examples of inducible promoters include, but are not limited to, an FNR responsive promoter, a P_{araC} promoter, a P_{araBAD} promoter, and a P_{TetR} promoter, each of which are described in more detail herein.

B. Optimized Branched Chain Amino Acid Deamination Enzymes

[0188] In one embodiment, the branched chain amino acid catabolism enzyme is a branched chain amino acid deamination enzyme. Thus, in some embodiments, the engineered bacteria comprise gene sequence(s) encoding one or more copies of a branched chain amino acid deamination enzyme. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding one, two, three, four, five, six, or more copies of a branched chain amino acid deamination enzyme. The one or more copies of branched chain amino acid deamination enzyme can be one or more copies of the same gene or can be different genes encoding branched chain amino acid deamination enzyme, *e.g.*, gene(s) from a different species or otherwise having a different gene sequence. The one or more

copies of branched chain amino acid deamination enzyme can be present in the bacterial chromosome or can be present in one or more plasmids. As used herein, the term “branched chain amino acid deamination enzyme” refers to an enzyme involved in the deamination, or the removal of an amine group, of a branched chain amino acid, which produces a corresponding branched chain alpha-keto acid (*e.g.*, α -ketoisocaproate, α -keto- β -methylvalerate, α -ketoisovalerate). Enzymes involved in the deamination of branched chain amino acids are well known to those of skill in the art. For example, in bacteria, leucine dehydrogenase (LeuDH, *leuDH*), *e.g.*, derived from *Pseudomonas aeruginosa* PA01, is capable of catalyzing the reversible deamination of branched chain amino acids, such as leucine, into their corresponding keto-acid counterpart (Baker *et al.*, *Structure*, 3(7):693-705, 1995).

[0189] In one embodiment, the branched chain amino acid deamination enzyme increases the rate of branched chain amino acid deamination in the cell. In one embodiment, the branched chain amino acid deamination enzyme decreases the level of branched chain amino acid in the cell as compared to the level of its corresponding alpha-keto acid in the cell. In another embodiment, the branched chain amino acid deamination enzyme increases the level of alpha-keto acid in the cell as compared to the level of its corresponding branched chain amino acid in the cell.

[0190] In one embodiment, the branched chain amino acid deamination enzyme is a leucine deamination enzyme. In another embodiment, the branched chain amino acid deamination enzyme is an isoleucine deamination enzyme. In another embodiment, the branched chain amino acid deamination enzyme is a valine deamination enzyme. In another embodiment, the branched chain amino acid deamination enzyme is involved in the deamination of leucine, isoleucine, and valine. In another embodiment, the branched chain amino acid deamination enzyme is involved in the deamination of leucine and valine, isoleucine and valine, or leucine and isoleucine. In some embodiments, the branched chain amino acid deamination enzyme is encoded by a branched chain amino acid deamination enzyme gene derived from a bacterial species. In some embodiments, the branched chain amino acid deamination enzyme is encoded by a branched chain amino acid deamination enzyme gene derived from a non-bacterial species. In some embodiments, the branched chain amino acid deamination enzyme is encoded by a branched chain amino acid deamination enzyme gene derived from a eukaryotic species, *e.g.*, a yeast species or a plant species. In another embodiment, the branched chain amino acid deamination enzyme is encoded by a branched chain amino acid deamination enzyme gene derived from a mammalian species, *e.g.*, human.

[0191] In some embodiments, the branched chain amino acid deamination enzyme is a branched chain amino acid dehydrogenase. Thus, in some embodiments, the engineered bacteria comprise gene sequence(s) encoding one or more copies of a branched chain amino acid dehydrogenase. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding one, two, three, four, five, six, or more copies of a branched chain amino acid dehydrogenase. The one or more copies of branched chain amino acid dehydrogenase can be one or more copies of the same gene or can be different genes encoding branched chain amino acid dehydrogenase, *e.g.*, gene(s)

from a different species or otherwise having a different gene sequence. The one or more copies of branched chain amino acid dehydrogenase can be present in the bacterial chromosome or can be present in one or more plasmids.

[0192] In some embodiments, the branched chain amino acid deamination enzyme is leucine dehydrogenase (“*leuDH*” or “*leuDH*”). As used herein “leucine dehydrogenase” refers to any polypeptide having enzymatic activity that deaminates leucine to its corresponding ketoacid, alpha-ketoisocaproate (KIC), deaminates valine to its corresponding ketoacid, ketoisovalerate (KIV), and deaminates isoleucine to its corresponding ketoacid, ketomethylvalerate (KMV). In some embodiments, the bacterial cells disclosed herein comprise a heterologous gene encoding a leucine dehydrogenase enzyme and are capable of converting leucine, valine, and/or isoleucine to their respective α -keto acids. For example, the leucine dehydrogenase enzyme *LeuDH* is capable of metabolizing leucine and a cytosolically active *LeuDH* should generally exhibit the ability to convert valine, isoleucine, and leucine to ketoisovalerate, ketomethylvalerate, and ketoisocaproate, respectively. Leucine dehydrogenase employs the co-factor NAD⁺. In some embodiments, *leuDH* encodes an octamer.

[0193] Multiple distinct leucine dehydrogenases (EC 1.4.1.9) are known in the art and are available from many microorganism sources, including those disclosed herein, as well as from eukaryotic sources (see, for example, Baker *et al.*, *Structure*, 3(7):693-705, 1995). In some embodiments, the branched chain amino acid deamination enzyme is encoded by at least one gene encoding a branched chain amino acid deamination enzyme derived from a bacterial species. In some embodiments, the branched chain amino acid deamination enzyme is encoded by at least one gene encoding a branched chain amino acid deamination enzyme derived from a non-bacterial species. In some embodiments, the branched chain amino acid deamination enzyme is encoded by at least one gene derived from a eukaryotic species, *e.g.*, a yeast species or a plant species. In another embodiment, the branched chain amino acid deamination enzyme is encoded by at least one gene derived from a mammalian species, *e.g.*, human.

[0194] In one embodiment, the engineered bacteria comprise gene sequence(s) encoding one or more branched chain amino acid dehydrogenase(s), *e.g.*, leucine dehydrogenase enzyme(s). In some embodiments, the branched chain amino acid dehydrogenase, *e.g.*, leucine dehydrogenase enzyme is derived from an organism of the genus or species that includes, but is not limited to, *Bacillus*, *Brevibacillus*, *Geobacillus*, *Lysinibacillus*, *Moorella*, *Natrialba*, *Pseudomonas*, *Sporosarcinia*, and *Thermoactinomyces*. In some embodiments, the *leuDH* gene is encoded by a gene derived from *Bacillus caldolyticus*, *Bacillus cereus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus mycoides*, *Bacillus niger*, *Bacillus pumilus*, *Bacillus subtilis*, *Brevibacillus brevis*, *Geobacillus stearothermophilus*, *Lysinibacillus sphaeriscus*, *Moorella Thermoacetica*, *Natrialba magadii*, *Sporosarcina psychrophila*, *Thermoactinomyces intermedius*, *Pseudomonas aeruginosa*, or *Pseudomonas resinovorans*. In some embodiments, the branched chain amino acid dehydrogenase

enzyme, *e.g.*, leucine dehydrogenase, is encoded by at least one gene derived from *Pseudomonas aeruginosa* PA01. In some embodiments, the branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase, is encoded by at least one gene derived from *Bacillus cereus*. In some embodiments, the at least one gene encoding the branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase, has been codon-optimized for use in the recombinant bacterial cell. In one embodiment, the at least one gene encoding the branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase, has been codon-optimized for use in *Escherichia coli*.

[0195] When a branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase, is expressed in the recombinant bacterial cells disclosed herein, the bacterial cells catabolize more branched chain amino acid, *e.g.*, leucine, isoleucine, and valine, than unmodified bacteria of the same bacterial subtype under the same conditions (*e.g.*, culture or environmental conditions). Thus, the genetically engineered bacteria comprising at least one heterologous gene encoding a branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase, may be used to catabolize excess branched chain amino acids, *e.g.*, leucine, valine, and isoleucine, to treat a disease associated with the deamination of a branched chain amino acid, including Maple Syrup Urine Disease (MSUD) as well as other disease provided herein.

[0196] The present disclosure further comprises genes encoding functional fragments of a branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase, or functional variants of branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase. The present disclosure encompasses genes encoding a branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase, comprising amino acids in its sequence that are substantially the same as an amino acid sequence described herein. As used herein, the term “functional fragment thereof” or “functional variant thereof” of a branched chain amino acid dehydrogenase enzyme, *e.g.*, a leucine dehydrogenase, gene relates to a sequence having qualitative biological activity in common with the wild-type branched chain amino acid dehydrogenase, *e.g.*, a leucine dehydrogenase, from which the fragment or variant was derived. For example, a functional fragment or a functional variant of a mutated branched chain amino acid dehydrogenase protein, *e.g.*, a leucine dehydrogenase, is one which retains essentially the same ability to catabolize BCAAs as the branched chain amino acid dehydrogenase protein, *e.g.*, a leucine dehydrogenase, from which the functional fragment or functional variant was derived. For example, a polypeptide having branched chain amino acid dehydrogenase enzyme, *e.g.*, a leucine dehydrogenase, activity may be truncated at the N-terminus or C-terminus and the retention of branched chain amino acid dehydrogenase enzyme, *e.g.*, a leucine dehydrogenase, activity assessed using assays known to those of skill in the art, including the exemplary assays provided herein. In one embodiment, the recombinant bacterial cell disclosed herein comprises a heterologous gene encoding an α branched chain amino acid dehydrogenase enzyme, *e.g.*, a leucine dehydrogenase, functional variant. In another embodiment, the recombinant

bacterial cell disclosed herein comprises a heterologous gene encoding a branched chain amino acid dehydrogenase enzyme, *e.g.*, a leucine dehydrogenase, functional fragment.

[0197] Assays for testing the activity of a branched chain amino acid dehydrogenase enzyme functional variant or functional fragment, *e.g.*, a leucine dehydrogenase functional variant or a leucine dehydrogenase functional fragment are well known to one of ordinary skill in the art. For example, leucine dehydrogenase activity can be assessed by expressing the protein, functional variant, or fragment thereof, in a recombinant bacterial cell that lacks endogenous leucine dehydrogenase activity. Also, activity can be assessed using the enzymatic assay methods as described by Soda *et al.* (*Biochem. Biophys. Res. Commun.*, 44:931, 1971), and Ohshima *et al.* (*J. Biol. Chem.*, 253:5719, 1978), the entire contents of each of which are expressly incorporated herein by reference.

[0198] In some embodiments, the at least one gene encoding a branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase is mutagenized, mutants exhibiting increased activity are selected, and the mutagenized gene(s) encoding the branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase, are isolated and inserted into the bacterial cell. In some embodiments, the at least one gene encoding a branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase is mutagenized, mutants exhibiting decreased activity are selected, and the mutagenized gene(s) encoding the branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase, are isolated and inserted into the bacterial cell. The gene comprising the modifications described herein may be present on a plasmid or chromosome.

[0199] In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase, and further comprise gene sequence encoding one or more polypeptides selected from other branched chain amino acid catabolism enzyme(s), BCAA transporter(s), and BCAA binding protein(s). Thus, in some embodiments, the at least one branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase, is coexpressed with an additional branched chain amino acid deamination enzyme, *e.g.*, branched chain amino acid dehydrogenase, a branched chain aminotransferase, and/or amino acid oxidase (also known as amino acid deaminase). For example, in some embodiments, the branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase, is coexpressed with one or more other branched chain amino acid dehydrogenase enzyme(s).

[0200] In some embodiments, the branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase, is coexpressed with one or more other branched chain amino acid catabolism enzyme(s), for example, a ketoacid decarboxylase, such as *kivD*. In some embodiments, the branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase, is coexpressed with one or more other branched chain amino acid catabolism enzyme(s), for example, a branched chain alcohol dehydrogenase, such as *adh2*. In other embodiments, the branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase, is coexpressed with a second branched chain amino acid catabolism enzyme, for example, *kivD*, and a branched chain alcohol dehydrogenase, for example,

adh2. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase enzyme, and gene sequence(s) encoding one or more keto-acid decarboxylase(s) (*e.g.*, *kivD*). In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase enzyme, and gene sequence(s) encoding one or more alcohol dehydrogenase(s) (*e.g.*, *adh2*).

[0201] In some embodiments, the at least one branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase enzyme is coexpressed with one or more BCAA transporter(s), for example, a high affinity leucine transporter, *e.g.*, LivKHMGF and/or the low affinity BCAA transporter BrnQ. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase enzyme and gene sequence(s) encoding one or more BCAA transporter(s) (*e.g.*, *livKHMGF* and/or *brnQ*).

[0202] In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase enzyme, and genetic modification that reduces export of a branched chain amino acid, *e.g.*, a genetic mutation in a *leuE* gene or promoter thereof. In one embodiment, the engineered bacteria comprise gene sequence(s) encoding at least one branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase enzyme, and a genetic modification that reduces or eliminates branched chain amino acid synthesis, *e.g.*, a genetic mutation in a *ilvC* gene or promoter thereof.

[0203] In some embodiments, the gene sequence(s) encoding the one or more branched chain amino acid dehydrogenase enzyme(s), *e.g.*, leucine dehydrogenase enzyme(s) is expressed under the control of a constitutive promoter. In some embodiments, the gene sequence(s) encoding the one or more branched chain amino acid dehydrogenase enzyme(s), *e.g.*, leucine dehydrogenase enzyme(s) is expressed under the control of an inducible promoter. In some embodiments, the gene sequence(s) encoding the one or more branched chain amino acid dehydrogenase enzyme(s), *e.g.*, leucine dehydrogenase enzyme(s) is expressed under the control of a promoter that is directly or indirectly induced by exogenous environmental conditions. In some embodiments, the gene sequence(s) encoding the one or more branched chain amino acid dehydrogenase enzyme(s), *e.g.*, leucine dehydrogenase enzyme(s) is expressed under the control of a promoter that is directly or indirectly induced by low-oxygen or anaerobic conditions, wherein expression of the gene encoding the branched chain amino acid dehydrogenase enzyme, *e.g.*, leucine dehydrogenase enzyme is activated under low-oxygen or anaerobic environments, such as the environment of the mammalian gut. In some embodiments, the gene sequence(s) encoding the one or more branched chain amino acid dehydrogenase enzyme(s), *e.g.*, leucine dehydrogenase enzyme(s) is expressed under the control of a promoter that is directly or indirectly induced by inflammatory conditions. Exemplary inducible promoters described herein include oxygen level-dependent promoters (*e.g.*, FNR-inducible

promoter), promoters induced by inflammation or an inflammatory response (RNS, ROS promoters), and promoters induced by a metabolite that may or may not be naturally present (*e.g.*, can be exogenously added) in the gut, *e.g.*, arabinose and tetracycline. Examples of inducible promoters include, but are not limited to, an FNR responsive promoter, a P_{araC} promoter, a P_{araBAD} promoter, and a P_{TetR} promoter, each of which are described in more detail herein.

[0204] In some embodiments, the branched chain amino acid dehydrogenase is leucine dehydrogenase. The present disclosure further comprises genes encoding functional fragments of leucine dehydrogenase, or functional variants of leucine dehydrogenase. The present disclosure encompasses genes encoding leucine dehydrogenase, comprising amino acids in its sequence that are substantially the same as an amino acid sequence described herein.

[0205] In some embodiments, the at least one *leuDH* gene has at least about 80% identity with any one of SEQ ID NOs:49-72. Accordingly, in one embodiment, the at least one *leuDH* gene has at least about 90% identity with any one of SEQ ID NOs:49-72. Accordingly, in one embodiment, the at least one *leuDH* gene has at least about 95% identity with any one of SEQ ID NOs:49-72. Accordingly, in one embodiment, the at least one *leuDH* gene has at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity with any one of SEQ ID NOs:49-72. In another embodiment, the at least one *leuDH* gene comprises any one of SEQ ID NOs:49-72. In yet another embodiment the at least one *leuDH* gene consists of any one of SEQ ID NOs:49-72.

C. Optimized Alcohol Dehydrogenase Enzymes

[0206] In some embodiments, wherein a branched chain amino acid catabolism enzyme is used to convert a ketoacid to its corresponding aldehyde, the recombinant bacterial cells may further comprise an alcohol dehydrogenase enzyme in order to convert the branched chain amino acid-derived aldehyde to its respective alcohol. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding one or more copies of an alcohol dehydrogenase. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding one, two, three, four, five, six, or more copies of an alcohol dehydrogenase. The one or more copies of an alcohol dehydrogenase can be one or more copies of the same gene or can be different genes encoding alcohol dehydrogenase, *e.g.*, gene(s) from a different species or otherwise having a different gene sequence. The one or more copies of alcohol dehydrogenase can be present in the bacterial chromosome or can be present in one or more plasmids. As used herein, "alcohol dehydrogenase" refers to any polypeptide having enzymatic activity that catalyzes the conversion of a branched chain amino acid-derived aldehyde, *e.g.*, isovaleraldehyde, isobutyraldehyde, and 2-methylbutyraldehyde, into its respective alcohol, *e.g.*, isopentanol, isobutanol, and 2-methylbutanol.

[0207] In general, alcohol dehydrogenases (EC 1.1.1.1) belong to a group of dehydrogenase enzymes that facilitate the interconversion between alcohols and aldehydes or ketones with the

reduction of nicotinamide adenine dinucleotide (NAD⁺ to NADH). Multiple distinct alcohol dehydrogenases are known in the art and are available from many microorganism sources, including those disclosed herein, as well as eukaryotic and plant sources (see, for example, Bennetzen *et al.*, *J. Biol. Chem.*, 257(6):3018-25, 1982 and Teng *et al.*, *Human Genetics*, 53(1):87-90, 1979, the entire contents of each of which are expressly incorporated herein by reference).

[0208] In some embodiments, the alcohol dehydrogenase is encoded by at least one gene derived from a bacterial species. In some embodiments, the alcohol dehydrogenase is encoded by at least one gene derived from a non-bacterial species. In some embodiments, the alcohol dehydrogenase is encoded by at least one gene derived from a eukaryotic species, *e.g.*, a yeast species or a plant species. In another embodiment, the alcohol dehydrogenase is encoded by at least one gene derived from a mammalian species, *e.g.*, human.

[0209] In one embodiment, the at least one gene encoding the alcohol dehydrogenase is derived from an organism of the genus or species that includes, but is not limited to, *Acetobacter*, *Azospirillum*, *Bacillus*, *Bacteroides*, *Bifidobacterium*, *Brevibacteria*, *Burkholderia*, *Citrobacter*, *Clostridium*, *Corynebacterium*, *Cronobacter*, *Enterobacter*, *Enterococcus*, *Erwinia*, *Helicobacter*, *Klebsiella*, *Lactobacillus*, *Lactococcus*, *Leishmania*, *Listeria*, *Macrococcus*, *Mycobacterium*, *Nakamurella*, *Nasonia*, *Nostoc*, *Pantoea*, *Pectobacterium*, *Proteus*, *Pseudomonas*, *Psychrobacter*, *Ralstonia*, *Saccharomyces*, *Salmonella*, *Sarcina*, *Serratia*, *Staphylococcus*, *Streptococcus*, and *Yersinia*, *e.g.*, *Acetobacter radioresistens*, *Acetobacter baumannii*, *Acetobacter calcoaceticus*, *Azospirillum brasilense*, *Bacillus anthracis*, *Bacillus cereus*, *Bacillus coagulans*, *Bacillus megaterium*, *Bacillus subtilis*, *Bacillus thuringiensis*, *Bacteroides fragilis*, *Bacteroides subtilis*, *Bacteroides thetaiotaomicron*, *Bifidobacterium bifidum*, *Bifidobacterium infantis*, *Bifidobacterium lactis*, *Bifidobacterium longum*, *Burkholderia xenovorans*, *Citrobacter youngae*, *Citrobacter koseri*, *Citrobacter rodentium*, *Clostridium acetobutylicum*, *Clostridium butyricum*, *Corynebacterium aurimucosum*, *Corynebacterium kroppenstedtii*, *Corynebacterium striatum*, *Cronobacter sakazakii*, *Cronobacter turicensis*, *Enterobacter cloacae*, *Enterobacter cancerogenus*, *Enterococcus faecium*, *Enterococcus faecalis*, *Erwinia amylovora*, *Erwinia pyrifoliae*, *Erwinia tasmaniensis*, *Helicobacter mustelae*, *Klebsiella pneumonia*, *Klebsiella variicola*, *Lactobacillus acidophilus*, *Lactobacillus bulgaricus*, *Lactobacillus casei*, *Lactobacillus johnsonii*, *Lactobacillus paracasei*, *Lactobacillus plantarum*, *Lactobacillus reuteri*, *Lactobacillus rhamnosus*, *Lactococcus lactis*, *Leishmania infantum*, *Leishmania major*, *Leishmania braziliensis*, *Listeria grayi*, *Macrococcus caseolyticus*, *Mycobacterium avium*, *Mycobacterium intracellulare*, *Mycobacterium kansasii*, *Mycobacterium leprae*, *Mycobacterium marinum*, *Mycobacterium smegmatis*, *Mycobacterium tuberculosis*, *Mycobacterium ulcerans*, *Nakamurella multipartita*, *Nasonia vitipennis*, *Nostoc punctiforme*, *Pantoea ananatis*, *Pantoea agglomerans*, *Pectobacterium atrosepticum*, *Pectobacterium carotovorum*, *Pseudomonas putida*, *Pseudomonas aeruginosa*, *Psychrobacter articus*, *Proteus vulgaris*, *Psychrobacter cryohalolentis*, *Ralstonia eutropha*, *Saccharomyces boulardii*, *Salmonella enterica*, *Sarcina*

ventriculi, *Serratia odorifera*, *Serratia proteamaculans*, *Staphylococcus aerus*, *Staphylococcus capitis*, *Staphylococcys carnosus*, *Staphylococcus epidermidis*, *Staphylococcus hominis*, *Staphylococcus haemolyticus*, *Staphylococcus lugdunensis*, *Staphylococcus saprophyticus*, *Staphylococcus warneri*, *Streptococcus faecalis*, *Yersinia enterocolitica*, *Yersinia mollaretii*, *Yersinia kristensenii*, *Yersinia rohdei*, and *Yersinia aldovae*. In some embodiments, the alcohol dehydrogenase is *adh2*. In some embodiments, the alcohol dehydrogenase is encoded by at least one gene derived from *Saccharomyces cerevisiae*. In another embodiment, the alcohol dehydrogenase is encoded by at least one gene derived from *E. coli*. In another embodiment, the alcohol dehydrogenase is encoded by at least one gene derived from *Oryza sativa*. In another embodiment, the alcohol dehydrogenase is encoded by at least one gene derived from *Penicillium brasilianum*. In another embodiment, the alcohol dehydrogenase is encoded by at least one gene derived from *Bifidobacterium longum*.

[0210] Accordingly, in some embodiments, the *adh2* gene has at least about 80% identity with the entire sequence of any one of SEQ ID NOs: 97-120. Accordingly, in one embodiment, the *adh2* gene has at least about 90% identity with the entire sequence of any one of SEQ ID NOs: 97-120. Accordingly, in one embodiment, the *adh2* gene has at least about 95% identity with the entire sequence of any one of SEQ ID NOs: 97-120. Accordingly, in one embodiment, the *adh2* gene has at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity with the entire sequence of any one of SEQ ID NOs: 97-120. In another embodiment, the *adh2* gene comprises the sequence of any one of SEQ ID NOs: 97-120. In yet another embodiment, the *adh2* gene consists of any one of SEQ ID NOs: 97-120.

[0211] In some embodiments, the at least one gene encoding the alcohol dehydrogenase has been codon-optimized for use in the recombinant bacterial cell. In some embodiments, the at least one gene encoding the alcohol dehydrogenase has been codon-optimized for use in *Escherichia coli*.

[0212] When an alcohol dehydrogenase is expressed in the recombinant bacterial cells disclosed herein, the bacterial cells convert more branched chain amino acid-derived aldehydes to their respective alcohols than unmodified bacteria of the same bacterial subtype under the same conditions (*e.g.*, culture or environmental conditions). Thus, the genetically engineered bacteria comprising at least one heterologous gene encoding an alcohol dehydrogenase may be used to catabolize excess branched chain amino acid-derived aldehydes, *e.g.*, isovaleraldehyde, to treat a disease associated with a branched chain amino acid, including Maple Syrup Urine Disease (MSUD).

[0213] The present disclosure encompasses genes encoding an alcohol dehydrogenase comprising amino acids in its sequence that are substantially the same as an amino acid sequence described herein. The present disclosure further comprises genes encoding functional fragments of an alcohol dehydrogenase or functional variants of an alcohol dehydrogenase. As used herein, the term “functional fragment thereof” or “functional variant thereof” of an alcohol dehydrogenase gene relates to a sequence having qualitative biological activity in common with the wild-type alcohol

dehydrogenase, from which the fragment or variant was derived. For example, a functional fragment or a functional variant of a mutated alcohol dehydrogenase is one which retains essentially the same ability to catabolize BCAAs as alcohol dehydrogenase from which the functional fragment or functional variant was derived. For example, a polypeptide having alcohol dehydrogenase activity may be truncated at the N-terminus or C-terminus and the retention alcohol dehydrogenase activity assessed using assays known to those of skill in the art, including the exemplary assays provided herein. In one embodiment, the recombinant bacterial cell disclosed herein comprises a heterologous gene encoding an alcohol dehydrogenase functional variant. In another embodiment, the recombinant bacterial cell disclosed herein comprises a heterologous gene encoding alcohol dehydrogenase functional fragment.

[0214] In some embodiments, the at least one gene encoding an alcohol dehydrogenase is mutagenized, mutants exhibiting increased activity are selected, and the mutagenized gene(s) encoding the alcohol dehydrogenase are isolated and inserted into the bacterial cell. In some embodiments, the at least one gene encoding the alcohol dehydrogenase is mutagenized, mutants exhibiting decreased activity are selected, and the mutagenized gene(s) encoding the alcohol dehydrogenase are isolated and inserted into the bacterial cell. The gene comprising the modifications described herein may be present on a plasmid or chromosome.

[0215] Assays for testing the activity of an alcohol dehydrogenase, an alcohol dehydrogenase functional variant, or an alcohol dehydrogenase functional fragment are well known to one of ordinary skill in the art. For example, alcohol dehydrogenase activity can be assessed by expressing the protein, functional variant, or fragment thereof, in a recombinant bacterial cell that lacks endogenous alcohol dehydrogenase activity. Also, activity can be assessed using the enzymatic assay methods as described by Kagi *et al.* (*J. Biol. Chem.*, 235:3188-92, 1960), and Walker (*Biochem. Educ.*, 20(1):42-43, 1992).

[0216] In some embodiments, the alcohol dehydrogenase is co-expressed with an additional branched chain amino acid catabolism enzyme. In some embodiments, the alcohol dehydrogenase is co-expressed with one or more keto-acid decarboxylase enzyme(s), *e.g.*, *kivD*. In some embodiments, the alcohol dehydrogenase is co-expressed with one or more keto-acid decarboxylase enzymes, *e.g.*, *kivD* and one or more other alcohol dehydrogenase enzymes.

[0217] In some embodiments, the at least one alcohol dehydrogenase enzyme is coexpressed with one or more BCAA transporter(s), for example, a high affinity leucine transporter, *e.g.*, LivKHMGF and/or a low affinity BCAA transporter, *e.g.*, BrnQ.

[0218] In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one alcohol dehydrogenase enzyme and genetic modification that reduces export of a branched chain amino acid, *e.g.*, a genetic mutation in a *leuE* gene or promoter thereof. In one embodiment, the engineered bacteria comprise gene sequence(s) encoding at least one alcohol

dehydrogenase enzyme and a genetic modification that reduces or eliminates branched chain amino acid synthesis, *e.g.*, a genetic mutation in a *ilvC* gene or promoter thereof.

[0219] In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one alcohol dehydrogenase enzyme and further comprise gene sequence encoding one or more polypeptides selected from other branched chain amino acid catabolism enzyme(s), BCAA transporter(s), and BCAA binding protein(s). In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one alcohol dehydrogenase and gene sequence(s) encoding one or more branched chain amino acid dehydrogenase(s) (*e.g.*, *leuDH*). In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one alcohol dehydrogenase and gene sequence(s) encoding one or more keto-acid decarboxylase enzyme(s), *e.g.*, *kivD*. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one alcohol dehydrogenase and gene sequence(s) encoding one or more other alcohol dehydrogenase(s) (*e.g.*, *adh2*).

[0220] In some embodiments, the engineered bacteria comprise gene sequence(s) encoding at least one alcohol dehydrogenase and genetic modification that reduces export of a branched chain amino acid, *e.g.*, a genetic mutation in a *leuE* gene or promoter thereof. In one embodiment, the engineered bacteria comprise gene sequence(s) encoding at least one alcohol dehydrogenase and a genetic modification that reduces or eliminates branched chain amino acid synthesis, *e.g.*, a genetic mutation in a *ilvC* gene or promoter thereof.

[0221] In one embodiment, the at least one gene encoding the branched chain amino acid alcohol dehydrogenase comprises the *adh2* gene. In one embodiment, the *adh2* gene has at least about 80% identity with the sequence of any one of SEQ ID NOs:97-120. In one embodiment, the *adh2* gene has at least about 90% identity with the sequence of any one of SEQ ID NOs:97-120. In one embodiment, the *adh2* gene has at least about 95% identity with the sequence of any one of SEQ ID NOs:97-120. In one embodiment, the *adh2* gene has at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity with the sequence of any one of SEQ ID NOs:97-120. In another embodiment, the *adh2* gene comprises the sequence of any one of SEQ ID NOs:97-120. In yet another embodiment, the *adh2* gene consists of the sequence of any one of SEQ ID NOs:97-120.

[0222] In any of these embodiments, the alcohol dehydrogenase is coexpressed with one or more branched chain amino acid catabolism enzymes, *e.g.*, *leuDH*, and/or *kivD*, each of which are described in more detail herein. In other embodiments, the alcohol dehydrogenase is further coexpressed with a transporter of a branched chain amino acid, *e.g.*, *brnQ*, and/or a binding protein of a BCAA.

[0223] In some embodiments, the gene sequence(s) encoding the one or more alcohol dehydrogenase is expressed under the control of a constitutive promoter. In some embodiments, the gene sequence(s) encoding the one or more alcohol dehydrogenase is expressed under the control of

an inducible promoter. In some embodiments, the gene sequence(s) encoding the one or more alcohol dehydrogenase is expressed under the control of a promoter that is directly or indirectly induced by exogenous environmental conditions. In some embodiments, the gene sequence(s) encoding the one or more alcohol dehydrogenase is expressed under the control of a promoter that is directly or indirectly induced by low-oxygen or anaerobic conditions, wherein expression of the gene encoding the alcohol dehydrogenase is activated under low-oxygen or anaerobic environments, such as the environment of the mammalian gut. In some embodiments, the gene sequence(s) encoding the one or more alcohol dehydrogenase is expressed under the control of a promoter that is directly or indirectly induced by inflammatory conditions. Exemplary inducible promoters described herein include oxygen level-dependent promoters (*e.g.*, FNR-inducible promoter), promoters induced by inflammation or an inflammatory response (RNS, ROS promoters), and promoters induced by a metabolite that may or may not be naturally present (*e.g.*, can be exogenously added) in the gut, *e.g.*, arabinose and tetracycline. Examples of inducible promoters include, but are not limited to, an FNR responsive promoter, a P_{araC} promoter, a P_{araBAD} promoter, and a P_{TetR} promoter, each of which are described in more detail herein.

Transporter (Importer) of a Branched Chain Amino Acid

[0224] In some embodiments, a recombinant bacterial cell disclosed herein comprising gene sequence(s) encoding at least one branched chain amino acid catabolism enzyme (*e.g.*, in some embodiments expressed on a high-copy plasmid) does not increase branched chain amino acid catabolism or decrease branched chain amino acid levels in the absence of a heterologous transporter (importer) of the branched chain amino acid, *e.g.*, leucine, and additional copies of a native importer of the branched chain amino acid, *e.g.*, *livKHMGF*. It has been surprisingly discovered that in some embodiments, the rate-limiting step of branched chain amino acid catabolism, *e.g.*, leucine catabolism, is *not* expression of a branched chain amino acid catabolism enzyme, but rather availability of the branched chain amino acid, *e.g.*, leucine. Thus, in some embodiments, it may be advantageous to increase branched chain amino acid transport, *e.g.*, leucine transport, into the cell, thereby enhancing branched chain amino acid catabolism. Surprisingly, in conjunction with overexpression of a transporter of a branched chain amino acid, *e.g.*, LivKHMGF, even low copy number plasmids comprising a gene encoding at least one branched chain amino acid catabolism enzyme are capable of almost completely eliminating a branched chain amino acid, *e.g.*, leucine, from a sample. Furthermore, in some embodiments that incorporate a transporter of a branched chain amino acid into the recombinant bacterial cell, there may be additional advantages to using a low-copy plasmid comprising the gene encoding the branched chain amino acid catabolism enzyme in conjunction in order to enhance the stability of expression of the branched chain amino acid catabolism enzyme, while maintaining high branched chain amino acid catabolism and to reduce negative selection pressure on the transformed bacterium. In alternate embodiments, the branched chain amino acid transporter is used in conjunction with a high-copy plasmid. In alternate

embodiments, the gene(s) at least one BCAA catabolism enzyme is integrated in the bacterial chromosome.

[0225] In some embodiments, in which the engineered bacterial cell comprises gene sequence encoding a branched amino acid transporter, the bacterial cell comprises gene sequence encoding one branched chain amino acid catabolism enzyme. In other embodiments, in which the engineered bacterial cell comprises gene sequence encoding a branched amino acid transporter, the bacterial cell comprises gene sequence(s) encoding two branched chain amino acid catabolism enzymes. In other embodiments, in which the engineered bacterial cell comprises gene sequence encoding a branched amino acid transporter, the bacterial cell comprises gene sequence(s) encoding three or more branched chain amino acid catabolism enzymes. In other embodiments, in which the engineered bacterial cell comprises gene sequence encoding a branched amino acid transporter, the bacterial cell comprises gene sequence(s) encoding four, five, six or more branched chain amino acid catabolism enzymes.

[0226] In some embodiments, the branched chain amino acid catabolism enzyme converts a branched chain amino acid, *e.g.*, leucine, valine, isoleucine, into its corresponding branched chain alpha-keto acid counterpart. In other embodiments, the branched chain amino acid catabolism enzyme converts a branched chain alpha-keto acid, *e.g.*, alpha-ketoisocaproate, alpha-keto- beta-methylvalerate, alpha-ketoisovalerate into its corresponding aldehyde. In other embodiments, the branched chain amino acid catabolism enzyme converts a branched chain alpha-keto acid, *e.g.*, alpha-ketoisocaproate, alpha keto- beta- methylvalerate, alpha-ketoisovalerate into its corresponding acetyl-CoA, *e.g.*, isovaleryl-CoA, alpha-methylbutyryl-CoA, isobutyl-CoA. In other embodiments, the branched chain amino acid catabolism enzyme converts a branched chain aldehyde to its corresponding alcohol. In another embodiment, the branched chain amino acid catabolism enzyme converts a branched chain aldehyde to its corresponding carboxylic acid.

[0227] In some embodiments, the engineered bacteria comprising gene sequence(s) encoding one or more BCAA transporter(s), further comprise gene sequence(s) encoding one or more BCAA catabolism enzymes, *e.g.*, *leuD*H. In some embodiments, the engineered bacteria comprising gene sequence(s) encoding one or more BCAA transporter(s), further comprise gene sequence(s) encoding one or more BCAA catabolism enzymes, *e.g.*, *leuD*H, and gene sequence(s) encoding one or more keto acid decarboxylase enzyme(s), *e.g.*, *kivD*. In any of these embodiments, the engineered bacteria also comprise a genetic modification that reduces export of a branched chain amino acid, *e.g.*, a genetic mutation in a *leuE* gene or promoter thereof. In any of these embodiments, the bacteria also comprise a genetic modification that reduces or eliminates branched chain amino acid synthesis, *e.g.*, a genetic mutation in a *ilvC* gene or promoter thereof.

[0228] In one embodiment, the bacterial cell comprises gene sequence encoding one or more transporter(s) of branched chain amino acids, gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s), and at least one genetic modification that reduces export of a

branched chain amino acid, *e.g.*, genetic modification in a *leuE* gene or promoter thereof. In one embodiment, the bacterial cell comprises gene sequence encoding one or more transporter(s) of branched chain amino acids, gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s), and at least one genetic modification that reduces or eliminates branched chain amino acid synthesis, *e.g.*, a genetic mutation in a *ilvC* gene or promoter thereof. In one embodiment, the bacterial cell comprises gene sequence encoding one or more transporter(s) of branched chain amino acids, gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s), at least one genetic modification that reduces export of a branched chain amino acid., and at least one genetic modification that reduces or eliminates branched chain amino acid synthesis.

[0229] The uptake of branched chain amino acids into bacterial cells is mediated by proteins well known to those of skill in the art. For example, two well characterized BCAA transport systems have been characterized in several bacteria, including *Escherichia coli*. BCAAs are transported by two systems into bacterial cells (*i.e.*, imported), the osmotic-shock-sensitive systems designated LIV-I and LS (leucine-specific), and by an osmotic-shock resistant system, BrnQ, formerly known as LIV-II (see Adams *et al.*, *J. Biol. Chem.* 265:11436-43 (1990); Anderson and Oxender, *J. Bacteriol.* 130:384-92 (1977); Anderson and Oxender, *J. Bacteriol.* 136:168-74 (1978); Haney *et al.*, *J. Bacteriol.* 174:108-15 (1992); Landick and Oxender, *J. Biol. Chem.* 260:8257-61 (1985); Nazos *et al.*, *J. Bacteriol.* 166:565-73 (1986); Nazos *et al.*, *J. Bacteriol.* 163:1196-202 (1985); Oxender *et al.*, *Proc. Natl. Acad. Sci. USA* 77:1412-16 (1980); Quay *et al.*, *J. Bacteriol.* 129:1257-65 (1977); Rahmanian *et al.*, *J. Bacteriol.* 116:1258-66 (1973); Wood, *J. Biol. Chem.* 250:4477-85 (1975); Guardiola *et al.*, *J. Bacteriol.* 117:393-405 (1974); Guardiola and Iaccarino, *J. Bacteriol.* 108:1034-44 (1971); Ohnishi *et al.*, *Jpn. J. Genet.* 63:343-57 (1988); Yamato and Anraku, *J. Bacteriol.* 144:36-44 (1980); and Yamato *et al.*, *J. Bacteriol.* 138:24-32 (1979)). Transport by the BrnQ system is mediated by a single membrane protein. Transport mediated by the LIV-I system is dependent on the substrate binding protein LivJ (also known as LIV-BP), while transport mediated by LS system is mediated by the substrate binding protein LivK (also known as LS-BP). LivJ is encoded by the *livJ* gene, and binds isoleucine, leucine and valine with K_d values of $\sim 10^{-6}$ and $\sim 10^{-7}$ M, while LivK is encoded by the *livK* gene, and binds leucine with a K_d value of $\sim 10^{-6}$ M (See Landick and Oxender, *J. Biol. Chem.* 260:8257-61 (1985)). Both LivJ and LivK interact with the inner membrane components LivHMGF to enable ATP-hydrolysis-coupled transport of their substrates into the cell, forming the LIV-I and LS transport systems, respectively. The LIV-I system transports leucine, isoleucine and valine, and to a lesser extent serine threonine and alanine, whereas the LS system only transports leucine. The six genes encoding the *E. coli* LIV-I and LS systems are organized into two transcriptional units, with *livKHMFG* transcribed as a single operon, and *livJ* transcribed separately. LivKHMFG is an ABC transporter comprised of five subunits, including LivK, which is a periplasmic amino acid binding protein, LivHM, which are membrane subunits, and LivGF, which are ATP-binding subunits. The *Escherichia coli* *liv* genes can be grouped according to protein function, with the *livJ* and *livK* genes

encoding periplasmic binding proteins with the binding affinities described above, the *livH* and *livM* genes encoding inner membrane permeases, and the *livG* and *livF* genes encoding cytoplasmic ATPases.

[0230] BrnQ is a branched chain amino acid transporter highly similar to the *Salmonella typhimurium* BrnQ branched chain amino acid transporter (Ohnishi *et al.*, Cloning and nucleotide sequence of the *brnQ* gene, the structural gene for a membrane-associated component of the LIV-II transport system for branched-chain amino acids in *Salmonella typhimurium*. Jpn J Genet. 1988 Aug;63(4):343-57) and corresponds to the Liv-II branched chain amino acid transport system in *E. coli*, which has been shown to transport leucine, valine, and isoleucine (Guardiola *et al.*, Mutations affecting the different transport systems for isoleucine, leucine, and valine in *Escherichia coli* K-12. J Bacteriol. 1974 Feb;117(2):393-405), Guardiola and Oxender, Genetic separation of high- and low-affinity transport systems for branched-chain amino acids in *Escherichia coli* K-12. J Bacteriol. 1978 Oct;136(1):168-74., Anderson and Oxender, Genetic separation of high- and low-affinity transport systems for branched-chain amino acids in *Escherichia coli* K-12 J Bacteriol. 1978 Oct;136(1):168-74.78). BrnQ is a member of the LIVCS family of branched chain amino acid transporters and likely functions as a sodium/branched chain amino acid symporter.

[0231] Branched chain amino acid transporters, *e.g.*, leucine importers, may be expressed or modified in the bacteria disclosed herein in order to enhance branched chain amino acid, *e.g.*, leucine, transport into the cell. For example, the gene sequence(s) for endogenous transporter(s) may be modified (*e.g.*, codon-optimized and/or expressed by a strong promoter) to overexpress the transporter and/or additional copies of the transporter may be added. Alternatively, or additionally, gene sequence(s) for one or more non-endogenous or non-native transporters may be expressed in the bacterial cell. Specifically, when the transporter of a branched chain amino acid is expressed in the recombinant bacterial cells disclosed herein, the bacterial cells import more branched chain amino acids into the cell when the transporter is expressed than unmodified bacteria of the same bacterial subtype under the same conditions (not expressing the transporter). Thus, in some embodiments, the engineered bacteria comprise gene sequence(s) encoding one or more transporter(s) of a branched chain amino acid, *e.g.*, leucine, which may be used to import branched chain amino acids, *e.g.*, leucine, into the bacteria so that any gene encoding a branched chain amino acid catabolism enzyme expressed in the bacteria catabolize the branched chain amino acid, *e.g.*, leucine, to treat diseases associated with the catabolism of branched chain amino acids, such as Maple Syrup Urine Disease (MSUD). In one embodiment, the bacterial cell comprises gene sequence(s) encoding one or more transporter(s) of branched chain amino acids. In one embodiment, the bacterial cell comprises gene sequence(s) encoding one or more transporter(s) of branched chain amino acids and a gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s). In one embodiment, the bacterial cell comprises gene sequence(s) encoding a one or more transporter(s) of branched chain amino acids and a genetic modification that reduces export of a branched chain amino

acid, *e.g.*, a genetic mutation in a *leuE* gene or promoter thereof. In one embodiment, the bacterial cell comprises gene sequence(s) encoding one or more transporter(s) of branched chain amino acids and a genetic modification that reduces or eliminates branched chain amino acid synthesis, *e.g.*, a genetic mutation in a *ilvC* gene or promoter thereof. In one embodiment, the bacterial cell comprises gene sequence encoding one or more transporter(s) of branched chain amino acids, gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s), and at least one genetic modification that reduces export of a branched chain amino acid. In one embodiment, the bacterial cell comprises gene sequence encoding one or more transporter(s) of branched chain amino acids, gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s), and at least one genetic modification that reduces or eliminates branched chain amino acid synthesis. In one embodiment, the bacterial cell comprises gene sequence encoding one or more transporter(s) of branched chain amino acids, gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s), at least one genetic modification that reduces export of a branched chain amino acid, and at least one genetic modification that reduces or eliminates branched chain amino acid synthesis. In any of these embodiments, the engineered bacterial cell may further comprise gene sequence encoding *livJ*, which brings BCAA into the bacterial cell. In any of these embodiments, the transporter may be a native transporter, *e.g.*, the bacteria may comprise additional copies of the native transporter. In any of these embodiments, the transporter may be a non-native transporter. In any of these embodiments, the transporter may be LivKHMGE. In any of these embodiments, the transporter may be BrnQ. In any of these embodiments, the bacterial cell may comprise gene sequence(s) encoding LivKHMGE and BrnQ.

[0232] In some embodiments, the engineered bacteria comprise gene sequence(s) encoding one or more BCAA catabolism enzyme(s) and gene sequence(s) encoding one or more BCAA transporters, in which the gene sequence(s) encoding one or more BCAA catabolism enzyme(s) and the gene sequence(s) encoding one or more transporter(s) are operably linked to different copies of the same promoter. In some embodiments, the engineered bacteria comprise gene sequence(s) encoding one or more BCAA catabolism enzyme(s) and gene sequence(s) encoding one or more BCAA transporters, in which the gene sequence(s) encoding one or more BCAA catabolism enzyme(s) and the gene sequence(s) encoding one or more transporter(s) are operably linked to different promoters. Thus, in some embodiments, the disclosure provides a bacterial cell that comprises gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) operably linked to a first promoter and gene sequence encoding one or more transporter(s) of a branched chain amino acid, *e.g.*, leucine. In some embodiments, the disclosure provides a bacterial cell that comprises gene sequence(s) encoding one or more transporters of a branched chain amino acid operably linked to the first promoter. In other embodiments, the disclosure provides a bacterial cell comprising gene sequence(s) encoding one or more branched chain amino acid catabolism enzyme(s) operably linked to a first promoter and gene sequence(s) encoding one or more transporter(s) of a branched chain

amino acid operably linked to a second promoter. In one embodiment, the first promoter and the second promoter are separate copies of the same promoter. In another embodiment, the first promoter and the second promoter are different promoters. In one embodiment, the first promoter and the second promoter are inducible promoters. In another embodiment, the first promoter is an inducible promoter and the second promoter is a constitutive promoter. In some embodiments, the gene sequence(s) encoding the one or more BCAA catabolism enzymes and the gene sequence(s) encoding the one or more transporters is expressed under the control of a constitutive promoter. In some embodiments, the gene sequence(s) encoding the one or more BCAA catabolism enzymes and the gene sequence(s) encoding the one or more BCAA transporters is expressed under the control of an inducible promoter. In some embodiments, the gene sequence(s) encoding the one or more BCAA transporters is expressed under the control of an inducible promoter that is directly or indirectly induced by exogenous environmental conditions. In some embodiments, the gene sequence(s) encoding the one or more BCAA catabolism enzymes and the gene sequence(s) encoding the one or more BCAA transporters is expressed under the control of an inducible promoter that is directly or indirectly induced by exogenous environmental conditions. In some embodiments, the gene sequence(s) encoding the one or more BCAA transporters is expressed under the control of an inducible promoter that is directly or indirectly induced by low-oxygen or anaerobic conditions, wherein expression of the gene encoding the one or more transporters is activated under low-oxygen or anaerobic environments, such as the environment of the mammalian gut. In some embodiments, the gene sequence(s) encoding the one or more BCAA catabolism enzymes and the gene sequence(s) encoding the one or more BCAA transporters is expressed under the control of an inducible promoter that is directly or indirectly induced by low-oxygen or anaerobic conditions, wherein expression of the gene(s) encoding the one or more BCAA catabolism enzymes and expression of the gene(s) encoding the one or more BCAA transporters is activated under low-oxygen or anaerobic environments, such as the environment of the mammalian gut. In some embodiments, the gene sequence(s) encoding the one or more BCAA transporters is expressed under the control of an inducible promoter that is directly or indirectly induced by inflammatory conditions. In some embodiments, the gene sequence(s) encoding the one or more BCAA catabolism enzymes and the gene sequence(s) encoding the one or more BCAA transporters is expressed under the control of an inducible promoter that is directly or indirectly induced by inflammatory conditions. Exemplary inducible promoters described herein include oxygen level-dependent promoters (*e.g.*, FNR-inducible promoter), promoters induced by inflammation or an inflammatory response (RNS, ROS promoters), and promoters induced by a metabolite that may or may not be naturally present (*e.g.*, can be exogenously added) in the gut, *e.g.*, arabinose and tetracycline. Examples of inducible promoters include, but are not limited to, an FNR responsive promoter, a P_{araC} promoter, a P_{araBAD} promoter, and a P_{TetR} promoter, each of which are described in more detail herein.

[0233] In one embodiment, the bacterial cell comprises at least one gene encoding a transporter of a branched chain amino acid from a different organism, *e.g.*, a different species of bacteria. In one embodiment, the bacterial cell comprises at least one native gene encoding a transporter of a branched chain amino acid. In some embodiments, the at least one native gene encoding a transporter of a branched chain amino acid is not modified. In another embodiment, the bacterial cell comprises more than one copy of at least one native gene encoding a transporter of a branched chain amino acid. In yet another embodiment, the bacterial cell comprises a copy of at least one gene encoding a native transporter of a branched chain amino acid, as well as at least one copy of at least one heterologous gene encoding a transporter of a branched chain amino acid. The heterologous gene sequence may encode an additional copy or copies of the native transporter, may encode one or more copies of a non-native transporter, and/or may encode one or more copies of a homologous or different transporter from a different bacterial species. In one embodiment, the bacterial cell comprises at least one, two, three, four, five, or six copies of the at least one heterologous gene encoding a transporter of a branched chain amino acid. In one embodiment, the bacterial cell comprises multiple copies of the at least one heterologous gene encoding a transporter of a branched chain amino acid. In one embodiment, the bacterial cell comprises gene sequence(s) encoding two or more different transporters of a branched chain amino acid. In one embodiment, the gene sequence(s) encoding two or more different transporters of a branched chain amino acid is under the control of one or more inducible promoters. In one embodiment, the gene sequence(s) encoding two or more different transporters of a branched chain amino acid is under the control of one or more constitutive promoters. In one embodiment, the gene sequence(s) encoding two or more different transporters of a branched chain amino acid is under the control of at least one inducible promoter and at least one constitutive promoter. In any of these embodiments, the gene sequence(s) encoding the one or more BCAA transporter(s) may be present on one or more plasmids. In any of these embodiments, the gene sequence(s) encoding the one or more BCAA transporter(s) may be present in the bacterial chromosome.

[0234] In one embodiment, the transporter of a branched chain amino acid imports a branched chain amino acid into the bacterial cell. In one embodiment, the transporter of a branched chain amino acid imports leucine into the bacterial cell. In one embodiment, the transporter of a branched chain amino acid imports isoleucine into the bacterial cell. In one embodiment, the transporter of a branched chain amino acid imports valine into the bacterial cell. In one embodiment, the transporter of a branched chain amino acid imports one or more of leucine, isoleucine, and valine into the bacterial cell.

[0235] In some embodiments, the transporter of a branched chain amino acid is encoded by a transporter of a branched chain amino acid gene derived from a bacterial genus or species, including but not limited to, *Bacillus*, *Campylobacter*, *Clostridium*, *Escherichia*, *Lactobacillus*, *Pseudomonas*, *Salmonella*, *Staphylococcus*, *Bacillus subtilis*, *Campylobacter jejuni*, *Clostridium perfringens*,

Escherichia coli, *Lactobacillus delbrueckii*, *Pseudomonas aeruginosa*, *Salmonella typhimurium*, or *Staphylococcus aureus*. In some embodiments, the bacterial species is *Escherichia coli*. In some embodiments, the bacterial species is *Escherichia coli* strain Nissle.

[0236] Multiple distinct transporters of branched chain amino acids are known in the art. In one embodiment, the at least one gene encoding a transporter of a branched chain amino acid is the *brnQ* gene. In one embodiment, the at least one gene encoding a transporter of a branched chain amino acid is the *livKHMGF* operon. In another embodiment, the *livKHMGF* operon is an *Escherichia coli livKHMGF* operon. In any of these embodiments, the bacterial cell may comprise more than one copy of any of one or more of these gene sequences. In any of these embodiments, the bacterial cell may over-express any one or more of these gene sequences. In any of these embodiments, the bacterial cell may further comprise gene sequence(s) encoding one or more additional BCAA transporters, *e.g.*, BrnQ transporter.

[0237] The present disclosure further provides genes encoding functional fragments of a transporter of a branched chain amino acid or functional variants of an importer of a branched chain amino acid. As used herein, the term “functional fragment thereof” or “functional variant thereof” of a transporter of a branched chain amino acid relates to an element having qualitative biological activity in common with the wild-type transporter of a branched chain amino acid from which the fragment or variant was derived. For example, a functional fragment or a functional variant of a mutated transporter of a branched chain amino acid protein is one which retains essentially the same ability to import leucine into the bacterial cell as does the importer protein from which the functional fragment or functional variant was derived. In one embodiment, the recombinant bacterial cell disclosed herein comprises at least one heterologous gene encoding a functional fragment of a transporter of branched chain amino acid. In another embodiment, the recombinant bacterial cell disclosed herein comprises a heterologous gene encoding a functional variant of a transporter of branched chain amino acid.

[0238] Assays for testing the activity of an importer of a branched chain amino acid, a functional variant of a transporter of a branched chain amino acid, or a functional fragment of a transporter of a branched chain amino acid are well known to one of ordinary skill in the art. For example, import of a branched chain amino acid may be determined using the methods as described in Haney *et al.*, *J. Bact.*, 174(1):108-15, 1992; Rahmanian *et al.*, *J. Bact.*, 116(3):1258-66, 1973; and Ribardo and Hendrixson, *J. Bact.*, 173(22):6233-43, 2011, the entire contents of each of which are expressly incorporated by reference herein.

[0239] In one embodiment, the genes encoding the transporter of a branched chain amino acid have been codon-optimized for use in the host organism. In one embodiment, the genes encoding the importer of a branched chain amino acid have been codon-optimized for use in *Escherichia coli*.

[0240] The present disclosure also encompasses genes encoding a transporter of a branched chain amino acid, *e.g.*, a transporter of leucine, comprising amino acids in its sequence that are substantially the same as an amino acid sequence described herein. Amino acid sequences that are substantially the same as the sequences described herein include sequences comprising conservative amino acid substitutions, as well as amino acid deletions and/or insertions.

[0241] In some embodiments, the at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMGF*, is mutagenized; mutants exhibiting increased branched chain amino acid, *e.g.*, leucine, transport are selected; and the mutagenized at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMGF*, is isolated and inserted into the bacterial cell. In some embodiments, the at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMGF*, is mutagenized; mutants exhibiting decreased branched chain amino acid, *e.g.*, leucine, transport are selected; and the mutagenized at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMGF*, is isolated and inserted into the bacterial cell. The importer modifications described herein may be present on a plasmid or chromosome.

[0242] In one embodiment, the *livKHMGF* operon has at least about 80% identity with the uppercase sequence of SEQ ID NO:145. Accordingly, in one embodiment, the *livKHMGF* operon has at least about 90% identity with the uppercase sequence of SEQ ID NO:145. Accordingly, in one embodiment, the *livKHMGF* operon has at least about 95% identity with the uppercase sequence of SEQ ID NO:145. Accordingly, in one embodiment, the *livKHMGF* operon has at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity with the uppercase sequence of SEQ ID NO:145. In another embodiment, the *livKHMGF* operon comprises the uppercase sequence of SEQ ID NO:145. In yet another embodiment the *livKHMGF* operon consists of the uppercase sequence of SEQ ID NO:145.

[0243] In some embodiments, the *brnQ* gene has at least about 80% identity with the entire sequence of any one of SEQ ID NOs: 121-144. Accordingly, in one embodiment, the *brnQ* gene has at least about 90% identity with the entire sequence of any one of SEQ ID NOs: 121-144. Accordingly, in one embodiment, the *brnQ* gene has at least about 95% identity with the entire sequence of any one of SEQ ID NOs: 121-144. Accordingly, in one embodiment, the *brnQ* gene has at least about 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity with the entire sequence of any one of SEQ ID NOs: 121-144. In another embodiment, the *brnQ* gene comprises the sequence of any one of SEQ ID NOs: 121-144. In yet another embodiment, the *brnQ* gene consists of any one of SEQ ID NOs: 121-144.

[0244] In some embodiments, the bacterial cell comprises a heterologous gene encoding a branched chain amino acid catabolism enzyme operably linked to a first promoter and at least one heterologous gene encoding a transporter of a branched chain amino acid. In some embodiments, the at least one heterologous gene encoding a transporter of a branched chain amino acid is operably linked to the first promoter. In other embodiments, the at least one heterologous gene encoding a

transporter of a branched chain amino acid is operably linked to a second promoter. In one embodiment, the at least one gene encoding a transporter of a branched chain amino acid is directly operably linked to the second promoter. In another embodiment, the at least one gene encoding a transporter of a branched chain amino acid is indirectly operably linked to the second promoter.

[0245] In some embodiments, expression of at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMGF* and/or *brnQ*, is controlled by a different promoter than the promoter that controls expression of the gene encoding the branched chain amino acid catabolism enzyme. In some embodiments, expression of the at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMGF* and/or *brnQ*, is controlled by the same promoter that controls expression of the branched chain amino acid catabolism enzyme. In some embodiments, at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMGF* and/or *brnQ*, and the branched chain amino acid catabolism enzyme are divergently transcribed from a promoter region. In some embodiments, expression of each of genes encoding the at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMGF* and/or *brnQ*, and the gene encoding the branched chain amino acid catabolism enzyme is controlled by different promoters.

[0246] In one embodiment, the at least one gene encoding a transporter of a branched chain amino acid is not operably linked to its native promoter. In some embodiments, the at least one gene encoding the transporter of a branched chain amino acid, *e.g.*, *livKHMGF*, is controlled by its native promoter. In some embodiments, the at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMGF* and/or *brnQ*, is controlled by an inducible promoter. In some embodiments, the at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMGF* and/or *brnQ*, is controlled by a promoter that is stronger than its native promoter. In some embodiments, the at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMGF*, is controlled by a constitutive promoter.

[0247] In another embodiment, the promoter is an inducible promoter. Inducible promoters are described in more detail *infra*.

[0248] In one embodiment, the at least one gene encoding a transporter of a branched chain amino acid is located on a plasmid in the bacterial cell. In another embodiment, the at least one gene encoding a transporter of a branched chain amino acid is in the chromosome of the bacterial cell. In yet another embodiment, a native copy of the at least one gene encoding a transporter of a branched chain amino acid is located in the chromosome of the bacterial cell, and a copy of at least one gene encoding a transporter of a branched chain amino acid from a different species of bacteria is located on a plasmid in the bacterial cell. In yet another embodiment, a native copy of the at least one gene encoding a transporter of a branched chain amino acid is located on a plasmid in the bacterial cell, and a copy of at least one gene encoding a transporter of a branched chain amino acid from a different species of bacteria is located on a plasmid in the bacterial cell. In yet another embodiment, a native copy of the at least one gene encoding a transporter of a branched chain amino acid is located in the

chromosome of the bacterial cell, and a copy of the at least one gene encoding an importer of a branched chain amino acid from a different species of bacteria is located in the chromosome of the bacterial cell.

[0249] In some embodiments, the at least one native gene encoding a transporter, *e.g.*, *livKHMGF* and/or *brnQ*, in the bacterial cell is not modified, and one or more additional copies of the native transporter, *e.g.*, *livKHMGF* and/or *brnQ*, are inserted into the genome. In some embodiments, the at least one native gene encoding a transporter, *e.g.*, *livKHMGF* and/or *brnQ*, in the bacterial cell is not modified, and one or more additional copies of the native transporter, *e.g.*, *livKHMGF* and/or *brnQ*, are present on a plasmid, *e.g.*, a high copy or low copy plasmid. In one embodiment, the one or more additional copies of the native a transporter, *e.g.*, *livKHMGF* and/or *brnQ*, that is inserted into the genome are under the control of the same inducible promoter that controls expression of the gene encoding the branched chain amino acid catabolism enzyme, *e.g.*, the FNR responsive promoter, or a different inducible promoter than the one that controls expression of the branched chain amino acid catabolism enzyme, or a constitutive promoter. In alternate embodiments, the at least one native gene encoding the transporter, *e.g.*, *livKHMGF* and/or *brnQ*, is not modified, and one or more additional copies of the transporter, *e.g.*, *livKHMGF*, from a different bacterial species is inserted into the genome of the bacterial cell. In one embodiment, the one or more additional copies of the transporter inserted into the genome of the bacterial cell are under the control of the same inducible promoter that controls expression of the gene encoding the branched chain amino acid catabolism enzyme, *e.g.*, the FNR responsive promoter, or a different inducible promoter than the one that controls expression of the gene encoding the branched chain amino acid catabolism enzyme, or a constitutive promoter.

[0250] In some embodiments, at least one native gene encoding the transporter, *e.g.*, *livKHMGF* and/or *brnQ*, in the genetically modified bacteria is not modified, and one or more additional copies of at least one native gene encoding the transporter, *e.g.*, *livKHMGF* and/or *brnQ*, are present in the bacterial cell on a plasmid. In one embodiment, the at least one native gene encoding the transporter *e.g.*, *livKHMGF* and/or *brnQ*, present in the bacterial cell on a plasmid is under the control of the same inducible promoter that controls expression of the gene encoding the branched chain amino acid catabolism enzyme, *e.g.*, the FNR responsive promoter, or a different inducible promoter than the one that controls expression of the gene encoding the branched chain amino acid catabolism enzyme, or a constitutive promoter. In alternate embodiments, the at least one native gene encoding the transporter, *e.g.*, *livKHMGF* and/or *brnQ*, is not modified, and a copy of at least one gene encoding the transporter, *e.g.*, *livKHMGF* and/or *brnQ*, from a different bacterial species is present in the bacteria on a plasmid. In one embodiment, the copy of at least one gene encoding the transporter, *e.g.*, *livKHMGF* and/or *brnQ*, from a different bacterial species is under the control of the same inducible promoter that controls expression of the gene encoding the branched chain amino acid catabolism enzyme, *e.g.*, the FNR responsive promoter, or a different inducible

promoter than the one that controls expression of the gene encoding the branched chain amino acid catabolism enzyme, or a constitutive promoter.

[0251] In some embodiments, the bacterium is *E. coli* Nissle, and the at least one native gene encoding the transporter, *e.g.*, *livKHMGF* and/or *brnQ*, in *E. coli* Nissle is not modified; one or more additional copies at least one native gene encoding the transporter, *e.g.*, *livKHMGF* and/or *brnQ*, from *E. coli* Nissle is inserted into the *E. coli* Nissle genome under the control of the same inducible promoter that controls expression of the gene encoding the branched chain amino acid catabolism enzyme, *e.g.*, the FNR responsive promoter, or a different inducible promoter than the one that controls expression of the gene encoding the branched chain amino acid catabolism enzyme, or a constitutive promoter. In an alternate embodiment, the at least one native gene encoding the a transporter, *e.g.*, *livKHMGF* and/or *brnQ* in *E. coli* Nissle is not modified, and a copy of at least one gene encoding the transporter, *e.g.*, *livKHMGF* and/or *brnQ*, from a different bacterial species is inserted into the *E. coli* Nissle genome under the control of the same inducible promoter that controls expression of the gene encoding the branched chain amino acid catabolism enzyme, *e.g.*, the FNR responsive promoter, or a different inducible promoter than the one that controls expression of the gene encoding the branched chain amino acid catabolism enzyme, or a constitutive promoter.

[0252] In some embodiments, the bacterial cell is *E. coli* Nissle, and the at least one native gene encoding the transporter, *e.g.*, *livKHMGF* and/or *brnQ*, in *E. coli* Nissle is not modified; one or more additional copies the at least one native gene encoding the transporter, *e.g.*, *livKHMGF* and/or *brnQ*, *E. coli* Nissle is present in the bacterium on a plasmid and under the control of the same inducible promoter that controls expression of the gene encoding the branched chain amino acid catabolism enzyme, *e.g.*, the FNR responsive promoter, or a different inducible promoter than the one that controls expression of the gene encoding the branched chain amino acid catabolism enzyme, or a constitutive promoter. In an alternate embodiment, the at least one native gene encoding the transporter, *e.g.*, *livKHMGF*, in *E. coli* Nissle is not modified, and a copy of at least one native gene encoding the transporter, *e.g.*, *livKHMGF* and/or *brnQ*, from a different bacterial species of are present in the bacterium on a plasmid and under the control of the same inducible promoter that controls expression of the gene encoding the branched chain amino acid catabolism enzyme, *e.g.*, the FNR responsive promoter, or a different inducible promoter than the one that controls expression of the gene encoding the branched chain amino acid catabolism enzyme, or a constitutive promoter.

[0253] In one embodiment, when the transporter of a branched chain amino acid is expressed in the recombinant bacterial cells disclosed herein, the bacterial cells import 10% more branched chain amino acids, *e.g.*, leucine, into the bacterial cell when the transporter is expressed as compared to unmodified bacteria of the same bacterial subtype under the same conditions. In another embodiment, when the transporter of a branched chain amino acid is expressed in the recombinant bacterial cells disclosed herein, the bacterial cells import 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or 100% more branched chain amino acids, *e.g.*, leucine, into the bacterial cell when the

transporter is expressed as compared with unmodified bacteria of the same bacterial subtype under the same conditions. In yet another embodiment, when the transporter of a branched chain amino acid is expressed in the recombinant bacterial cells disclosed herein, the bacterial cells import two-fold more branched chain amino acids, *e.g.*, leucine, into the cell when the transporter is expressed than unmodified bacteria of the same bacterial subtype under the same conditions. In yet another embodiment, when the transporter of a branched chain amino acid is expressed in the recombinant bacterial cells disclosed herein, the bacterial cells import three-fold, four-fold, five-fold, six-fold, seven-fold, eight-fold, nine-fold, or ten-fold more branched chain amino acids, *e.g.*, leucine, into the cell when a transporter is expressed as compared with unmodified bacteria of the same bacterial subtype under the same conditions.

Exporter of a Branched Chain Amino Acid

[0254] The bacterial cells disclosed herein may comprise a genetic modification that inhibits or decreases the export of a branched chain amino acid and/or its corresponding alpha keto acid or other metabolite from the bacterial cell. Knocking-out or reducing export of one or more branched chain amino acids from a bacterial cell allows the bacterial cell to more efficiently retain and catabolize exogenous branched chain amino acids and/or their alpha-keto acid counterparts or other metabolite counterparts in order to treat the diseases and disorders described herein. Any of the bacterial cells disclosed herein comprising gene sequence encoding one or more BCAA catabolism enzymes and/or one or more BCAA transporters may further a genetic modification that inhibits or decreases the export of a branched chain amino acid and/or its corresponding alpha keto acid or other metabolite from the bacterial cell.

[0255] The export of branched chain amino acids from bacterial cells is mediated by proteins well known to those of skill in the art. For example, one branched chain amino acid exporter, the leucine exporter LeuE has been characterized in *Escherichia coli* (Kutukova *et al.*, *FEBS Letters* 579:4629-34 (2005); incorporated herein by reference). LeuE is encoded by the *leuE* gene in *Escherichia coli* (also known as *yeaS*). Additionally, a two-gene encoded exporter of the branched chain amino acids isoleucine, valine and leucine, denominated BrnFE was identified in the bacteria *Corynebacterium glutamicum* (Kennerknecht *et al.*, *J. Bacteriol.* 184:3947-56 (2002); incorporated herein by reference). The BrnFE system is encoded by the *Corynebacterium glutamicum* genes *brnF* and *brnE*, and homologues of said genes have been identified in several organisms, including *Agrobacterium tumefaciens*, *Achaeglobus fulgidus*, *Bacillus subtilis*, *Deinococcus radiodurans*, *Escherichia coli*, *Haemophilus influenzae*, *Helicobacter pylori*, *Lactococcus lactis*, *Streptococcus pneumoniae*, and *Vibrio cholerae* (see Kennerknecht *et al.*, 2002).

[0256] The bacterial cells disclosed herein comprise a genetic modification that reduces export of a branched chain amino acid from the bacterial cell. Multiple distinct exporters of branched chain amino acids, *e.g.*, leucine, are known in the art. In one embodiment, the recombinant bacterial cell disclosed herein comprises a genetic modification that reduces export of a branched chain amino

acid from the bacterial cell, wherein the endogenous gene encoding an exporter of a branched chain amino acid is a *leuE* gene.

[0257] In one embodiment, the recombinant bacterial cell disclosed herein comprises a genetic modification that reduces export of a branched chain amino acid from the bacterial cell and a heterologous gene encoding a branched chain amino acid catabolism enzyme. When the recombinant bacterial cells disclosed herein comprise a genetic modification that reduces export of a branched chain amino acid, the bacterial cells retain more branched chain amino acids, *e.g.*, leucine, in the bacterial cell than unmodified bacteria of the same bacterial subtype under the same conditions. Thus, the recombinant bacteria comprising a genetic modification that reduces export of a branched chain amino acid may be used to retain more branched chain amino acids in the bacterial cell so that any branched chain amino acid catabolism enzyme expressed in the organism, *e.g.*, co-expressed α -ketoisovalerate decarboxylase or co-expressed branched chain keto dehydrogenase, can catabolize the branched chain amino acids, *e.g.*, *leucine*, to treat diseases associated with the catabolism of branched chain amino acids, including Maple Syrup Urine Disease (MSUD). In one embodiment, the recombinant bacteria further comprise a heterologous gene encoding an importer of a branched chain amino acid, *e.g.*, a *livKHMGF* and/or *brnQ* gene.

[0258] In one embodiment, the genetic modification reduces export of a branched chain amino acid, *e.g.*, leucine, from the bacterial cell. In one embodiment, the bacterial cell is from a bacterial genus or species that includes but is not limited to, *Bacillus*, *Bacteroides*, *Bifidobacterium*, *Brevibacteria*, *Clostridium*, *Enterococcus*, *Escherichia*, *Lactobacillus*, *Lactococcus*, and *Staphylococcus*, *e.g.*, *Bacillus coagulans*, *Bacillus subtilis*, *Bacteroides fragilis*, *Bacteroides subtilis*, *Bacteroides thetaiotaomicron*, *Bifidobacterium bifidum*, *Bifidobacterium infantis*, *Bifidobacterium lactis*, *Bifidobacterium longum*, *Clostridium butyricum*, *Enterococcus faecium*, *Escherichia coli*, *Lactobacillus acidophilus*, *Lactobacillus bulgaricus*, *Lactobacillus casei*, *Lactobacillus johnsonii*, *Lactobacillus paracasei*, *Lactobacillus plantarum*, *Lactobacillus reuteri*, *Lactobacillus rhamnosus*, *Lactococcus lactis*. In another embodiment, the bacterial cell is an *Escherichia coli* bacterial cell. In another embodiment, the bacterial cell is an *Escherichia coli* strain Nissle bacterial cell.

[0259] In one embodiment, the genetic modification is a mutation in an endogenous gene encoding an exporter of a branched chain amino acid. In one embodiment, the genetic mutation is a deletion of the endogenous gene encoding an exporter, *e.g.*, *leuE*, of a branched chain amino acid. In another embodiment, the genetic mutation results in an exporter having reduced activity as compared to a wild-type exporter protein. In one embodiment, the activity of the exporter is reduced at least 50%, at least 75%, or at least 100%. In another embodiment, the activity of the exporter is reduced at least two-fold, three-fold, four-fold, or five-fold. In another embodiment, the genetic mutation results in an exporter, *e.g.*, LeuE, having no activity, *i.e.*, results in an exporter, *e.g.*, LeuE, which cannot export a branched chain amino acid, *e.g.*, lysine, from the bacterial cell.

[0260] It is routine for one of ordinary skill in the art to make mutations in a gene of interest. Mutations include substitutions, insertions, deletions, and/or truncations of one or more specific amino acid residues or of one or more specific nucleotides or codons in the polypeptide or polynucleotide of the exporter of a branched chain amino acid. Mutagenesis and directed evolution methods are well known in the art for creating variants. See, *e.g.*, U.S. Pat. No. 7,783,428; U.S. Pat. No. 6,586,182; U.S. Pat. No. 6,117,679; and Ling, *et al.*, 1999, "Approaches to DNA mutagenesis: an overview," *Anal. Biochem.*, 254(2):157-78; Smith, 1985, "In vitro mutagenesis," *Ann. Rev. Genet.*, 19:423-462; Carter, 1986, "Site-directed mutagenesis," *Biochem. J.*, 237:1-7; and Minshull, *et al.*, 1999, "Protein evolution by molecular breeding," *Current Opinion in Chemical Biology*, 3:284-290. For example, the lambda red system can be used to knock-out genes in *E. coli* (see, for example, Datta *et al.*, *Gene*, 379:109-115 (2006)).

[0261] The term "inactivated" as applied to a gene refers to any genetic modification that decreases or eliminates the expression of the gene and/or the functional activity of the corresponding gene product (mRNA and/or protein). The term "inactivated" encompasses complete or partial inactivation, suppression, deletion, interruption, blockage, promoter alterations, antisense RNA, dsRNA, or down-regulation of a gene. This can be accomplished, for example, by gene "knockout," inactivation, mutation (*e.g.*, insertion, deletion, point, or frameshift mutations that disrupt the expression or activity of the gene product), or by use of inhibitory RNAs (*e.g.*, sense, antisense, or RNAi technology). A deletion may encompass all or part of a gene's coding sequence. The term "knockout" refers to the deletion of most (at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%) or all (100%) of the coding sequence of a gene. In some embodiments, any number of nucleotides can be deleted, from a single base to an entire piece of a chromosome.

[0262] Assays for testing the activity of an exporter of a branched chain amino acid, *e.g.*, leucine, are well known to one of ordinary skill in the art. For example, export of a branched chain amino acid, such as leucine, may be determined using the methods described by Haney *et al.*, *J. Bact.*, 174(1):108-15, 1992; Rahmanian *et al.*, *J. Bact.*, 116(3):1258-66, 1973; and Ribardo and Hendrixson, *J. Bact.*, 173(22):6233-43, 2011, the entire contents of which are expressly incorporated herein by reference.

[0263] In another embodiment, the genetic modification is a mutation in a promoter of an endogenous gene encoding an exporter of a branched chain amino acid. In one embodiment, the genetic mutation results in decreased expression of the *leuE* gene. In one embodiment, *leuE* gene expression is reduced by about 50%, 75%, or 100%. In another embodiment, *leuE* gene expression is reduced about two-fold, three-fold, four-fold, or five-fold. In another embodiment, the genetic mutation completely inhibits expression of the *leuE* gene.

[0264] Assays for testing the level of expression of a gene, such as an exporter of a branched chain amino acid, *e.g.*, *leuE*, are well known to one of ordinary skill in the art. For example, reverse-

transcriptase polymerase chain reaction may be used to detect the level of mRNA expression of a gene. Alternatively, Western blots using antibodies directed against a protein may be used to determine the level of expression of the protein.

[0265] In another embodiment, the genetic modification is an overexpression of a repressor of an exporter of a branched chain amino acid. In one embodiment, the overexpression of the repressor of the exporter is caused by a mutation which renders the promoter of the repressor constitutively active. In another embodiment, the overexpression of the repressor of the exporter is caused by the insertion of an inducible promoter in front of the repressor so that the expression of the repressor can be induced. Inducible promoters are described in more detail herein.

Reduction of Endogenous Bacterial Branched Chain Amino Acid Production

[0266] The bacterial cells disclosed herein may comprise a genetic modification that inhibits or decreases the biosynthesis of a branched chain amino acid and/or its corresponding alpha keto acid or other metabolite in the bacterial cell. Knocking-out or reducing production of endogenous branched chain amino acids in a bacterial cell allows the bacterial cell to more efficiently take up and catabolize exogenous branched chain amino acids and/or their alpha-keto acid counterparts or other metabolite counterparts in order to treat the diseases and disorders described herein. Knock-out or knock down of a gene encoding an enzyme required for branched chain amino acid biosynthesis creates an auxotroph, which requires the cell to import the branched chain amino acid or a metabolite to survive. Any of the bacterial cells disclosed herein comprising gene sequence encoding one or more BCAA catabolism enzymes and/or one or more BCAA transporters may further a genetic modification that inhibits or decreases the biosynthesis of a branched chain amino acid and/or its corresponding alpha keto acid or other metabolite in the bacterial cell.

[0267] As used herein, the term “branched chain amino acid biosynthesis” enzyme refers to an enzyme involved in the biosynthesis of a branched chain amino acid and/or its corresponding alpha-keto acid or other metabolite. Multiple distinct genes involved in biosynthetic pathways of branched chain amino acids, *e.g.*, isoleucine, leucine, and valine, are known in the art. For example, the *ilvC* gene encodes a keto-acid reductoisomerase enzyme that catalyzes the conversion of acetohydroxy acids into dihydroxy valerates, which leads to the synthesis of the essential branched side chain amino acids valine and isoleucine (EC 1.1.1.86) and has been characterized in *Escherichia coli* (Wek and Hatfield, *J. Biol. Chem.* 261:2441-50 (1986), the entire contents of which are expressly incorporated herein by reference). Additionally, homologues of *ilvC* have been identified in several organisms, including *Candida albicans*, *Oryza sativa*, *Saccharomyces cerevisiae*, *Pseudomonas aeruginosa*, *Corynebacterium glutamicum*, and *Spinacia oleracea*.

[0268] In one embodiment, the genetic modification is a mutation in an endogenous gene encoding a protein that is involved in the biosynthesis of a branched chain amino acid or an alpha keto acid, *e.g.*, *ilvC*. *ilvC* is an acetohydroxy acid isomeroreductase that is required for branched chain amino acid synthesis. In one embodiment, the genetic mutation is a deletion of an endogenous gene

encoding a protein that is involved in the biosynthesis of a branched chain amino acid or an alpha-keto acid, or other BCAA metabolite, *e.g.*, *ilvC*. In another embodiment, the genetic mutation results in an enzyme having reduced activity as compared to a wild-type enzyme. In one embodiment, the activity of the enzyme is reduced at least 50%, at least 75%, or at least 100%. In another embodiment, the activity of the enzyme is reduced at least two-fold, three-fold, four-fold, or five-fold. In another embodiment, the genetic mutation results in an enzyme, *e.g.*, *IlvC*, having no activity, *i.e.*, results in an enzyme, *e.g.*, *IlvC*, which cannot catalyze the conversion of acetohydroxy acids into dihydroxy valerates, thereby inhibiting the endogenous synthesis of the branched chain amino acids valine and isoleucine in the recombinant bacterial cell.

[0269] It is routine for one of ordinary skill in the art to make mutations in a gene of interest. Mutations include substitutions, insertions, deletions, and/or truncations of one or more specific amino acid residues or of one or more specific nucleotides or codons in the polypeptide or polynucleotide of the exporter of a branched chain amino acid. Mutagenesis and directed evolution methods are well known in the art for creating variants. See, *e.g.*, U.S. Pat. No. 7,783,428; U.S. Pat. No. 6,586,182; U.S. Pat. No. 6,117,679; and Ling, *et al.*, 1999, "Approaches to DNA mutagenesis: an overview," *Anal. Biochem.*, 254(2):157-78; Smith, 1985, "In vitro mutagenesis," *Ann. Rev. Genet.*, 19:423-462; Carter, 1986, "Site-directed mutagenesis," *Biochem. J.*, 237:1-7; and Minshull, *et al.*, 1999, "Protein evolution by molecular breeding," *Current Opinion in Chemical Biology*, 3:284-290. For example, the lambda red system can be used to knock-out genes in *E. coli* (see, for example, Datta *et al.*, *Gene*, 379:109-115 (2006)).

[0270] The term "inactivated," as applied to a gene, refers to any genetic modification that decreases or eliminates the expression of the gene and/or the functional activity of the corresponding gene product (mRNA and/or protein). The term "inactivated" encompasses complete or partial inactivation, suppression, deletion, interruption, blockage, promoter alterations, antisense RNA, dsRNA, or down-regulation of a gene. This can be accomplished, for example, by gene "knockout," inactivation, mutation (*e.g.*, insertion, deletion, point, or frameshift mutations that disrupt the expression or activity of the gene product), or by use of inhibitory RNAs (*e.g.*, sense, antisense, or RNAi technology). A deletion may encompass all or part of a gene's coding sequence. The term "knockout" refers to the deletion of most (at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%) or all (100%) of the coding sequence of a gene. In some embodiments, any number of nucleotides can be deleted, from a single base to an entire piece of a chromosome.

[0271] Assays for testing the activity of enzymes involved in the biosynthesis of branched chain amino acids and/or alpha-keto acids, and/or other BCAA metabolite *e.g.*, *ilvC*, are well known to one of ordinary skill in the art. For example, the activity of a ketol-acid reductoisomerase enzyme may be determined using the methods described by Durner *et al.*, *Plant Physiol.*, 103:903-910, 1993, the entire contents of which are expressly incorporated herein by reference.

[0272] In another embodiment, the genetic modification is a mutation in a promoter of an endogenous gene encoding the branched chain amino acid biosynthesis enzyme. In one embodiment, the genetic mutation results in decreased expression of the branched chain amino acid biosynthesis enzyme gene. In one embodiment, gene expression is reduced by about 50%, 75%, or 100%. In another embodiment, gene expression is reduced about two-fold, three-fold, four-fold, or five-fold. In another embodiment, the genetic mutation completely inhibits expression of the gene.

[0273] Assays for testing the level of expression of a gene, such as *ilvC*, are well known to one of ordinary skill in the art. For example, reverse-transcriptase polymerase chain reaction may be used to detect the level of mRNA expression of a gene. Alternatively, Western blots using antibodies directed against a protein may be used to determine the level of expression of the protein.

[0274] In another embodiment, the genetic modification is an overexpression of a repressor of a branched chain amino acid biosynthesis gene. In one embodiment, the overexpression of the repressor of the gene is caused by a mutation which renders the promoter of the repressor constitutively active. In another embodiment, the overexpression of the repressor of the branched chain amino acid biosynthesis gene is caused by the insertion of an inducible promoter in front of the repressor so that the expression of the repressor can be induced. Inducible promoters are described in more detail herein.

Inducible Promoters

[0275] In some embodiments, the bacterial cell comprises a stably maintained plasmid or chromosome carrying the one or more optimized genes, *e.g.*, *kivD*, *leuDH*, *adh2*, and *brnQ*, such that the optimized enzymes can be expressed in the host cell, and the host cell is capable of survival and/or growth *in vitro*, *e.g.*, in medium, and/or *in vivo*, *e.g.*, in the gut.

[0276] In some embodiments, the gene encoding the branched chain amino acid catabolism enzyme is present on a plasmid and operably linked to a directly or indirectly inducible promoter. In some embodiments, the gene encoding the branched chain amino acid catabolism enzyme is present on a plasmid and operably linked to a promoter that is induced under low-oxygen or anaerobic conditions. In some embodiments, the gene encoding the branched chain amino acid catabolism enzyme is present on a chromosome and operably linked to a directly or indirectly inducible promoter. In some embodiments, the gene encoding the branched chain amino acid catabolism enzyme is present in the chromosome and operably linked to a promoter that is induced under low-oxygen or anaerobic conditions. In some embodiments, the gene encoding the branched chain amino acid catabolism enzyme is present on a plasmid and operably linked to a promoter that is induced by exposure to tetracycline or arabinose.

[0277] In some embodiments, the bacterial cell comprises a stably maintained plasmid or chromosome carrying the at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMgf* and/or *brnQ*, such that the transporter, *e.g.*, *LivKHMgf* and/or *BrnQ*, can be

expressed in the host cell, and the host cell is capable of survival and/or growth *in vitro*, *e.g.*, in medium, and/or *in vivo*, *e.g.*, in the gut. In some embodiments, bacterial cell comprises two or more distinct copies of the at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMGF* and/or *brnQ*. In some embodiments, the genetically engineered bacteria comprise multiple copies of the same at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMGF* and/or *brnQ*. In some embodiments, the at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMGF* and/or *brnQ*, is present on a plasmid and operably linked to a directly or indirectly inducible promoter. In some embodiments, the at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMGF* and/or *brnQ*, is present on a plasmid and operably linked to a promoter that is induced under low-oxygen or anaerobic conditions. In some embodiments, the at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMGF* and/or *brnQ*, is present on a chromosome and operably linked to a directly or indirectly inducible promoter. In some embodiments, the at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMGF* and/or *brnQ*, is present in the chromosome and operably linked to a promoter that is induced under low-oxygen or anaerobic conditions. In some embodiments, the at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMGF* and/or *brnQ*, is present on a plasmid and operably linked to a promoter that is induced by exposure to tetracycline.

[0278] In some embodiments, the promoter that is operably linked to the gene encoding the branched chain amino acid catabolism enzyme and the promoter that is operably linked to the gene encoding the transporter of a branched chain amino acid, *e.g.*, *livKHMGF* and/or *brnQ*, is directly induced by exogenous environmental conditions. In some embodiments, the promoter that is operably linked to the gene encoding the branched chain amino acid catabolism enzyme and the promoter that is operably linked to the gene encoding the transporter of a branched chain amino acid, *e.g.*, *livKHMGF* and/or *brnQ*, is indirectly induced by exogenous environmental conditions. In some embodiments, the promoter is directly or indirectly induced by exogenous environmental conditions specific to the gut of a mammal. In some embodiments, the promoter is directly or indirectly induced by exogenous environmental conditions specific to the small intestine of a mammal. In some embodiments, the promoter is directly or indirectly induced by low-oxygen or anaerobic conditions such as the environment of the mammalian gut. In some embodiments, the promoter is directly or indirectly induced by molecules or metabolites that are specific to the gut of a mammal, *e.g.*, propionate. In some embodiments, the promoter is directly or indirectly induced by a molecule that is co-administered with the bacterial cell.

[0279] In certain embodiments, the bacterial cell comprises optimized genes, *e.g.*, *kivD*, *leuDH*, *adh2*, and *brnQ*, which are expressed under the control of the fumarate and nitrate reductase regulator (FNR) promoter. In certain embodiments, the bacterial cell comprises at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMGF*, is expressed under the control of the fumarate and nitrate reductase regulator (FNR) promoter. In *E. coli*, FNR is a major

transcriptional activator that controls the switch from aerobic to anaerobic metabolism (Unden *et al.*, 1997). In the anaerobic state, FNR dimerizes into an active DNA binding protein that activates hundreds of genes responsible for adapting to anaerobic growth. In the aerobic state, FNR is prevented from dimerizing by oxygen and is inactive.

[0280] FNR responsive promoters include, but are not limited to, the FNR responsive promoters listed in the chart, below. Underlined sequences are predicted ribosome binding sites, and bolded sequences are restriction sites used for cloning. FNR responsive promoters are set forth as SEQ ID NO: 146-162.

[0281] In some embodiments, multiple distinct FNR nucleic acid sequences are inserted in the genetically engineered bacteria. In alternate embodiments, the genetically engineered bacteria comprise a gene encoding a branched chain amino acid catabolism enzyme, or other enzyme disclosed herein, is expressed under the control of an alternate oxygen level-dependent promoter, *e.g.*, DNR (Trunk *et al.*, 2010) or ANR (Ray *et al.*, 1997). In alternate embodiments, the genetically engineered bacteria comprise at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMgf* and/or *brnQ*, is expressed under the control of an alternate oxygen level-dependent promoter, *e.g.*, DNR (Trunk *et al.*, 2010) or ANR (Ray *et al.*, 1997). In these embodiments, catabolism of the branched chain amino acid, *e.g.*, leucine, is particularly activated in a low-oxygen or anaerobic environment, such as in the gut. In some embodiments, gene expression is further optimized by methods known in the art, *e.g.*, by optimizing ribosomal binding sites and/or increasing mRNA stability. In one embodiment, the mammalian gut is a human mammalian gut.

[0282] In some embodiments, the bacterial cell comprises an oxygen-level dependent transcriptional regulator, *e.g.*, FNR, ANR, or DNR, and corresponding promoter from a different bacterial species. The heterologous oxygen-level dependent transcriptional regulator and promoter increase the transcription of genes operably linked to said promoter, *e.g.*, the gene encoding the branched chain amino acid catabolism enzyme, and/or the at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMgf* and/or *brnQ*, and/or the at least one gene encoding a binding protein of a branched chain amino acid in a low-oxygen or anaerobic environment, as compared to the native gene(s) and promoter in the bacteria under the same conditions. In certain embodiments, the non-native oxygen-level dependent transcriptional regulator is an FNR protein from *N. gonorrhoeae* (*see, e.g.*, Isabella *et al.*, 2011). In some embodiments, the corresponding wild-type transcriptional regulator is left intact and retains wild-type activity. In alternate embodiments, the corresponding wild-type transcriptional regulator is deleted or mutated to reduce or eliminate wild-type activity.

[0283] In some embodiments, the genetically engineered bacteria comprise a wild-type oxygen-level dependent transcriptional regulator, *e.g.*, FNR, ANR, or DNR, and corresponding promoter that is mutated relative to the wild-type promoter from bacteria of the same subtype. The mutated promoter enhances binding to the wild-type transcriptional regulator and increases the

transcription of genes operably linked to said promoter, *e.g.*, the gene encoding the branched chain amino acid catabolism enzyme, and/or the at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMGF* and/or *brnQ*, and/or the at least one gene encoding a binding protein of a branched chain amino acid in a low-oxygen or anaerobic environment, as compared to the wild-type promoter under the same conditions. In some embodiments, the genetically engineered bacteria comprise a wild-type oxygen-level dependent promoter, *e.g.*, FNR, ANR, or DNR promoter, and corresponding transcriptional regulator that is mutated relative to the wild-type transcriptional regulator from bacteria of the same subtype. The mutated transcriptional regulator enhances binding to the wild-type promoter and increases the transcription of genes operably linked to said promoter, *e.g.*, the gene encoding the branched chain amino acid catabolism enzyme, and/or the at least one gene encoding a transporter of a branched chain amino acid, *e.g.*, *livKHMGF* and/or *brnQ*, and/or the at least one gene encoding a binding protein of a branched chain amino acid in a low-oxygen or anaerobic environment, as compared to the wild-type transcriptional regulator under the same conditions. In certain embodiments, the mutant oxygen-level dependent transcriptional regulator is an FNR protein comprising amino acid substitutions that enhance dimerization and FNR activity (*see, e.g.*, Moore *et al.*, 2006).

[0284] In some embodiments, the bacterial cells disclosed herein comprise multiple copies of the endogenous gene encoding the oxygen level-sensing transcriptional regulator, *e.g.*, the *FNR* gene. In some embodiments, the gene encoding the oxygen level-sensing transcriptional regulator is present on a plasmid. In some embodiments, the gene encoding the oxygen level-sensing transcriptional regulator and the gene encoding the branched chain amino acid catabolism enzyme are present on different plasmids. In some embodiments, the gene encoding the oxygen level-sensing transcriptional regulator and the gene encoding the branched chain amino acid catabolism enzyme and/or the at least one gene encoding a transporter of a branched chain amino acid and/or the at least one gene encoding a binding protein of a branched chain amino acid are present on different plasmids. In some embodiments, the gene encoding the oxygen level-sensing transcriptional regulator and the gene encoding the branched chain amino acid catabolism enzyme and/or the at least one gene encoding a transporter of a branched chain amino acid and/or the at least one gene encoding a binding protein of a branched chain amino acid are present on the same plasmid.

[0285] In some embodiments, the gene encoding the oxygen level-sensing transcriptional regulator is present on a chromosome. In some embodiments, the gene encoding the oxygen level-sensing transcriptional regulator and the gene encoding the branched chain amino acid catabolism enzyme and/or the at least one gene encoding a transporter of a branched chain amino acid and/or the at least one gene encoding a binding protein of a branched chain amino acid are present on different chromosomes. In some embodiments, the gene encoding the oxygen level-sensing transcriptional regulator and the gene encoding the branched chain amino acid catabolism enzyme and/or the at least one gene encoding a transporter of a branched chain amino acid and/or the

at least one gene encoding a binding protein of a branched chain amino acid are present on the same chromosome. In some instances, it may be advantageous to express the oxygen level-sensing transcriptional regulator under the control of an inducible promoter in order to enhance expression stability. In some embodiments, expression of the transcriptional regulator is controlled by a different promoter than the promoter that controls expression of the gene encoding the branched chain amino acid catabolism enzyme and/or BCAA transporter and/or BCAA binding protein. In some embodiments, expression of the transcriptional regulator is controlled by the same promoter that controls expression of the branched chain amino acid catabolism enzyme and/or BCAA transporter and/or BCAA binding protein. In some embodiments, the transcriptional regulator and the branched chain amino acid catabolism enzyme are divergently transcribed from a promoter region.

RNS-dependent regulation

[0286] In some embodiments, the genetically engineered bacteria comprise a gene encoding a branched chain amino acid catabolism enzyme that is expressed under the control of an inducible promoter. In some embodiments, the genetically engineered bacterium that expresses a branched chain amino acid catabolism enzyme and/or BCAA transporter and/or BCAA binding protein is under the control of a promoter that is activated by inflammatory conditions. In one embodiment, the gene for producing the branched chain amino acid catabolism enzyme and/or BCAA transporter and/or BCAA binding protein is expressed under the control of an inflammatory-dependent promoter that is activated in inflammatory environments, *e.g.*, a reactive nitrogen species or RNS promoter.

[0287] As used herein, “reactive nitrogen species” and “RNS” are used interchangeably to refer to highly active molecules, ions, and/or radicals derived from molecular nitrogen. RNS can cause deleterious cellular effects such as nitrosative stress. RNS includes, but is not limited to, nitric oxide (NO•), peroxynitrite or peroxynitrite anion (ONOO⁻), nitrogen dioxide (•NO₂), dinitrogen trioxide (N₂O₃), peroxynitrous acid (ONOOH), and nitroperoxycarbonate (ONOOCO₂⁻) (unpaired electrons denoted by •). Bacteria have evolved transcription factors that are capable of sensing RNS levels. Different RNS signaling pathways are triggered by different RNS levels and occur with different kinetics.

[0288] As used herein, “RNS-inducible regulatory region” refers to a nucleic acid sequence to which one or more RNS-sensing transcription factors is capable of binding, wherein the binding and/or activation of the corresponding transcription factor activates downstream gene expression; in the presence of RNS, the transcription factor binds to and/or activates the regulatory region. In some embodiments, the RNS-inducible regulatory region comprises a promoter sequence. In some embodiments, the transcription factor senses RNS and subsequently binds to the RNS-inducible regulatory region, thereby activating downstream gene expression. In alternate embodiments, the transcription factor is bound to the RNS-inducible regulatory region in the absence of RNS; in the presence of RNS, the transcription factor undergoes a conformational change, thereby activating

downstream gene expression. The RNS-inducible regulatory region may be operatively linked to a gene or genes, *e.g.*, a branched chain amino acid catabolism enzyme gene sequence(s), *e.g.*, any of the amino acid catabolism enzymes described herein. For example, in the presence of RNS, a transcription factor senses RNS and activates a corresponding RNS-inducible regulatory region, thereby driving expression of an operatively linked gene sequence. Thus, RNS induces expression of the gene or gene sequences.

[0289] As used herein, “RNS-derepressible regulatory region” refers to a nucleic acid sequence to which one or more RNS-sensing transcription factors is capable of binding, wherein the binding of the corresponding transcription factor represses downstream gene expression; in the presence of RNS, the transcription factor does not bind to and does not repress the regulatory region. In some embodiments, the RNS-derepressible regulatory region comprises a promoter sequence. The RNS-derepressible regulatory region may be operatively linked to a gene or genes, *e.g.*, a branched chain amino acid catabolism enzyme gene sequence(s), BCAA transporter sequence(s), BCAA binding protein(s). For example, in the presence of RNS, a transcription factor senses RNS and no longer binds to and/or represses the regulatory region, thereby derepressing an operatively linked gene sequence or gene cassette. Thus, RNS derepresses expression of the gene or genes.

[0290] As used herein, “RNS-repressible regulatory region” refers to a nucleic acid sequence to which one or more RNS-sensing transcription factors is capable of binding, wherein the binding of the corresponding transcription factor represses downstream gene expression; in the presence of RNS, the transcription factor binds to and represses the regulatory region. In some embodiments, the RNS-repressible regulatory region comprises a promoter sequence. In some embodiments, the transcription factor that senses RNS is capable of binding to a regulatory region that overlaps with part of the promoter sequence. In alternate embodiments, the transcription factor that senses RNS is capable of binding to a regulatory region that is upstream or downstream of the promoter sequence. The RNS-repressible regulatory region may be operatively linked to a gene sequence or gene cassette. For example, in the presence of RNS, a transcription factor senses RNS and binds to a corresponding RNS-repressible regulatory region, thereby blocking expression of an operatively linked gene sequence or gene sequences. Thus, RNS represses expression of the gene or gene sequences.

[0291] As used herein, a “RNS-responsive regulatory region” refers to a RNS-inducible regulatory region, a RNS-repressible regulatory region, and/or a RNS-derepressible regulatory region. In some embodiments, the RNS-responsive regulatory region comprises a promoter sequence. Each regulatory region is capable of binding at least one corresponding RNS-sensing transcription factor. Examples of transcription factors that sense RNS and their corresponding RNS-responsive genes, promoters, and/or regulatory regions include, but are not limited to, those shown in **Table 4**.

Table 4. Examples of RNS-sensing transcription factors and RNS-responsive genes

RNS-sensing transcription factor:	Primarily capable of sensing:	Examples of responsive genes, promoters, and/or regulatory regions:
NsrR	NO	<i>norB, aniA, nsrR, hmpA, ytfE, ygbA, hcp, hcr, nrfA, aox</i>
NorR	NO	<i>norVW, norR</i>
DNR	NO	<i>norCB, nir, nor, nos</i>

[0292] In some embodiments, the genetically engineered bacteria of the invention comprise a tunable regulatory region that is directly or indirectly controlled by a transcription factor that is capable of sensing at least one reactive nitrogen species. The tunable regulatory region is operatively linked to a gene or genes capable of directly or indirectly driving the expression of an amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein, thus controlling expression of the branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein relative to RNS levels. For example, the tunable regulatory region is a RNS-inducible regulatory region, and the payload is an amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein, such as any of the amino acid catabolism enzymes, BCAA transporters, and BCAA binding proteins provided herein; when RNS is present, *e.g.*, in an inflamed tissue, a RNS-sensing transcription factor binds to and/or activates the regulatory region and drives expression of the branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein gene or genes. Subsequently, when inflammation is ameliorated, RNS levels are reduced, and production of the branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein is decreased or eliminated.

[0293] In some embodiments, the tunable regulatory region is a RNS-inducible regulatory region; in the presence of RNS, a transcription factor senses RNS and activates the RNS-inducible regulatory region, thereby driving expression of an operatively linked gene or genes. In some embodiments, the transcription factor senses RNS and subsequently binds to the RNS-inducible regulatory region, thereby activating downstream gene expression. In alternate embodiments, the transcription factor is bound to the RNS-inducible regulatory region in the absence of RNS; when the transcription factor senses RNS, it undergoes a conformational change, thereby inducing downstream gene expression.

[0294] In some embodiments, the tunable regulatory region is a RNS-inducible regulatory region, and the transcription factor that senses RNS is NorR. NorR “is an NO-responsive transcriptional activator that regulates expression of the *norVW* genes encoding flavorubredoxin and an associated flavoprotein, which reduce NO to nitrous oxide” (Spiro 2006). The genetically engineered bacteria of the invention may comprise any suitable RNS-responsive regulatory region from a gene that is activated by NorR. Genes that are capable of being activated by NorR are known in the art (see, *e.g.*, Spiro 2006; Vine *et al.*, 2011; Karlinsey *et al.*, 2012). In certain embodiments, the

genetically engineered bacteria of the invention comprise a RNS-inducible regulatory region from norVW that is operatively linked to a gene or genes, *e.g.*, one or more branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein gene sequence(s). In the presence of RNS, a NorR transcription factor senses RNS and activates to the norVW regulatory region, thereby driving expression of the operatively linked gene(s) and producing the amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein.

[0295] In some embodiments, the tunable regulatory region is a RNS-inducible regulatory region, and the transcription factor that senses RNS is DNR. DNR (dissimilatory nitrate respiration regulator) “promotes the expression of the nir, the nor and the nos genes” in the presence of nitric oxide (Castiglione *et al.*, 2009). The genetically engineered bacteria of the invention may comprise any suitable RNS-responsive regulatory region from a gene that is activated by DNR. Genes that are capable of being activated by DNR are known in the art (see, *e.g.*, Castiglione *et al.*, 2009; Giardina *et al.*, 2008; Table 1). In certain embodiments, the genetically engineered bacteria of the invention comprise a RNS-inducible regulatory region from norCB that is operatively linked to a gene or gene cassette, *e.g.*, a butyrogenic gene cassette. In the presence of RNS, a DNR transcription factor senses RNS and activates to the norCB regulatory region, thereby driving expression of the operatively linked gene or genes and producing one or more amino acid catabolism enzymes. In some embodiments, the DNR is *Pseudomonas aeruginosa* DNR.

[0296] In some embodiments, the tunable regulatory region is a RNS-derepressible regulatory region, and binding of a corresponding transcription factor represses downstream gene expression; in the presence of RNS, the transcription factor no longer binds to the regulatory region, thereby derepressing the operatively linked gene or gene cassette.

[0297] In some embodiments, the tunable regulatory region is a RNS-derepressible regulatory region, and the transcription factor that senses RNS is NsrR. NsrR is “an Rrf2-type transcriptional repressor [that] can sense NO and control the expression of genes responsible for NO metabolism” (Isabella *et al.*, 2009). The genetically engineered bacteria of the invention may comprise any suitable RNS-responsive regulatory region from a gene that is repressed by NsrR. In some embodiments, the NsrR is *Neisseria gonorrhoeae* NsrR. Genes that are capable of being repressed by NsrR are known in the art (see, *e.g.*, Isabella *et al.*, 2009; Dunn *et al.*, 2010). In certain embodiments, the genetically engineered bacteria of the invention comprise a RNS-derepressible regulatory region from norB that is operatively linked to a gene or genes, *e.g.*, a branched chain amino acid catabolism enzyme gene or genes. In the presence of RNS, an NsrR transcription factor senses RNS and no longer binds to the norB regulatory region, thereby derepressing the operatively linked branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein gene or genes and producing the encoding an amino acid catabolism enzyme(s).

[0298] In some embodiments, it is advantageous for the genetically engineered bacteria to express a RNS-sensing transcription factor that does not regulate the expression of a significant

number of native genes in the bacteria. In some embodiments, the genetically engineered bacterium of the invention expresses a RNS-sensing transcription factor from a different species, strain, or substrain of bacteria, wherein the transcription factor does not bind to regulatory sequences in the genetically engineered bacterium of the invention. In some embodiments, the genetically engineered bacterium of the invention is *Escherichia coli*, and the RNS-sensing transcription factor is NsrR, *e.g.*, from *Neisseria gonorrhoeae*, wherein the *Escherichia coli* does not comprise binding sites for said NsrR. In some embodiments, the heterologous transcription factor minimizes or eliminates off-target effects on endogenous regulatory regions and genes in the genetically engineered bacteria.

[0299] In some embodiments, the tunable regulatory region is a RNS-repressible regulatory region, and binding of a corresponding transcription factor represses downstream gene expression; in the presence of RNS, the transcription factor senses RNS and binds to the RNS-repressible regulatory region, thereby repressing expression of the operatively linked gene or gene cassette. In some embodiments, the RNS-sensing transcription factor is capable of binding to a regulatory region that overlaps with part of the promoter sequence. In alternate embodiments, the RNS-sensing transcription factor is capable of binding to a regulatory region that is upstream or downstream of the promoter sequence.

[0300] In these embodiments, the genetically engineered bacteria may comprise a two repressor activation regulatory circuit, which is used to express an amino acid catabolism enzyme. The two repressor activation regulatory circuit comprises a first RNS-sensing repressor and a second repressor, which is operatively linked to a gene or gene cassette, *e.g.*, encoding an amino acid catabolism enzyme. In one aspect of these embodiments, the RNS-sensing repressor inhibits transcription of the second repressor, which inhibits the transcription of the gene or gene cassette. Examples of second repressors useful in these embodiments include, but are not limited to, TetR, C1, and LexA. In the absence of binding by the first repressor (which occurs in the absence of RNS), the second repressor is transcribed, which represses expression of the gene or genes. In the presence of binding by the first repressor (which occurs in the presence of RNS), expression of the second repressor is repressed, and the gene or genes, *e.g.*, a branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein gene or genes is expressed.

[0301] A RNS-responsive transcription factor may induce, derepress, or repress gene expression depending upon the regulatory region sequence used in the genetically engineered bacteria. One or more types of RNS-sensing transcription factors and corresponding regulatory region sequences may be present in genetically engineered bacteria. In some embodiments, the genetically engineered bacteria comprise one type of RNS-sensing transcription factor, *e.g.*, NsrR, and one corresponding regulatory region sequence, *e.g.*, from *norB*. In some embodiments, the genetically engineered bacteria comprise one type of RNS-sensing transcription factor, *e.g.*, NsrR, and two or more different corresponding regulatory region sequences, *e.g.*, from *norB* and *aniA*. In some embodiments, the genetically engineered bacteria comprise two or more types of RNS-sensing

transcription factors, *e.g.*, NsrR and NorR, and two or more corresponding regulatory region sequences, *e.g.*, from *norB* and *norR*, respectively. One RNS-responsive regulatory region may be capable of binding more than one transcription factor. In some embodiments, the genetically engineered bacteria comprise two or more types of RNS-sensing transcription factors and one corresponding regulatory region sequence. Nucleic acid sequences of several RNS-regulated regulatory regions are known in the art (see, *e.g.*, Spiro 2006; Isabella *et al.*, 2009; Dunn *et al.*, 2010; Vine *et al.*, 2011; Karlinsey *et al.*, 2012).

[0302] In some embodiments, the genetically engineered bacteria of the invention comprise a gene encoding a RNS-sensing transcription factor, *e.g.*, the *nsrR* gene, that is controlled by its native promoter, an inducible promoter, a promoter that is stronger than the native promoter, *e.g.*, the *GlnRS* promoter or the *P(Bla)* promoter, or a constitutive promoter. In some instances, it may be advantageous to express the RNS-sensing transcription factor under the control of an inducible promoter in order to enhance expression stability. In some embodiments, expression of the RNS-sensing transcription factor is controlled by a different promoter than the promoter that controls expression of the therapeutic molecule. In some embodiments, expression of the RNS-sensing transcription factor is controlled by the same promoter that controls expression of the therapeutic molecule. In some embodiments, the RNS-sensing transcription factor and therapeutic molecule are divergently transcribed from a promoter region.

[0303] In some embodiments, the genetically engineered bacteria of the invention comprise a gene for a RNS-sensing transcription factor from a different species, strain, or substrain of bacteria. In some embodiments, the genetically engineered bacteria comprise a RNS-responsive regulatory region from a different species, strain, or substrain of bacteria. In some embodiments, the genetically engineered bacteria comprise a RNS-sensing transcription factor and corresponding RNS-responsive regulatory region from a different species, strain, or substrain of bacteria. The heterologous RNS-sensing transcription factor and regulatory region may increase the transcription of genes operatively linked to said regulatory region in the presence of RNS, as compared to the native transcription factor and regulatory region from bacteria of the same subtype under the same conditions.

[0304] In some embodiments, the genetically engineered bacteria comprise a RNS-sensing transcription factor, NsrR, and corresponding regulatory region, *nsrR*, from *Neisseria gonorrhoeae*. In some embodiments, the native RNS-sensing transcription factor, *e.g.*, NsrR, is left intact and retains wild-type activity. In alternate embodiments, the native RNS-sensing transcription factor, *e.g.*, NsrR, is deleted or mutated to reduce or eliminate wild-type activity.

[0305] In some embodiments, the genetically engineered bacteria of the invention comprise multiple copies of the endogenous gene encoding the RNS-sensing transcription factor, *e.g.*, the *nsrR* gene. In some embodiments, the gene encoding the RNS-sensing transcription factor is present on a plasmid. In some embodiments, the gene encoding the RNS-sensing transcription factor and the gene or gene cassette for producing the therapeutic molecule are present on different plasmids. In some

embodiments, the gene encoding the RNS-sensing transcription factor and the gene or gene cassette for producing the therapeutic molecule are present on the same plasmid. In some embodiments, the gene encoding the RNS-sensing transcription factor is present on a chromosome. In some embodiments, the gene encoding the RNS-sensing transcription factor and the gene or gene cassette for producing the therapeutic molecule are present on different chromosomes. In some embodiments, the gene encoding the RNS-sensing transcription factor and the gene or gene cassette for producing the therapeutic molecule are present on the same chromosome.

[0306] In some embodiments, the genetically engineered bacteria comprise a wild-type gene encoding a RNS-sensing transcription factor, *e.g.*, the NsrR gene, and a corresponding regulatory region, *e.g.*, a norB regulatory region, that is mutated relative to the wild-type regulatory region from bacteria of the same subtype. The mutated regulatory region increases the expression of the branched chain amino acid catabolism enzyme in the presence of RNS, as compared to the wild-type regulatory region under the same conditions. In some embodiments, the genetically engineered bacteria comprise a wild-type RNS-responsive regulatory region, *e.g.*, the norB regulatory region, and a corresponding transcription factor, *e.g.*, NsrR, that is mutated relative to the wild-type transcription factor from bacteria of the same subtype. The mutant transcription factor increases the expression of the branched chain amino acid catabolism enzyme in the presence of RNS, as compared to the wild-type transcription factor under the same conditions. In some embodiments, both the RNS-sensing transcription factor and corresponding regulatory region are mutated relative to the wild-type sequences from bacteria of the same subtype in order to increase expression of the branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein in the presence of RNS.

[0307] In some embodiments, the gene or gene cassette for producing the anti-inflammation and/or gut barrier function enhancer molecule is present on a plasmid and operably linked to a promoter that is induced by RNS. In some embodiments, expression is further optimized by methods known in the art, *e.g.*, by optimizing ribosomal binding sites, manipulating transcriptional regulators, and/or increasing mRNA stability.

[0308] In some embodiments, any of the gene(s) of the present disclosure may be integrated into the bacterial chromosome at one or more integration sites. For example, one or more copies of a branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein gene(s) may be integrated into the bacterial chromosome. Having multiple copies of the gene or gen(s) integrated into the chromosome allows for greater production of the amino acid catabolism enzyme(s) and also permits fine-tuning of the level of expression. Alternatively, different circuits described herein, such as any of the secretion or exporter circuits, in addition to the therapeutic gene(s) or gene cassette(s) could be integrated into the bacterial chromosome at one or more different integration sites to perform multiple different functions.

ROS-dependent regulation

[0309] In some embodiments, the genetically engineered bacteria comprise a gene for producing an branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein that is expressed under the control of an inducible promoter. In some embodiments, the genetically engineered bacterium that expresses a branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein under the control of a promoter that is activated by conditions of cellular damage. In one embodiment, the gene for producing the branched chain amino acid catabolism enzyme is expressed under the control of a cellular damage-dependent promoter that is activated in environments in which there is cellular or tissue damage, *e.g.*, a reactive oxygen species or ROS promoter.

[0310] As used herein, “reactive oxygen species” and “ROS” are used interchangeably to refer to highly active molecules, ions, and/or radicals derived from molecular oxygen. ROS can be produced as byproducts of aerobic respiration or metal-catalyzed oxidation and may cause deleterious cellular effects such as oxidative damage. ROS includes, but is not limited to, hydrogen peroxide (H₂O₂), organic peroxide (ROOH), hydroxyl ion (OH⁻), hydroxyl radical (•OH), superoxide or superoxide anion (•O₂⁻), singlet oxygen (1O₂), ozone (O₃), carbonate radical, peroxide or peroxy radical (•O₂⁻²), hypochlorous acid (HOCl), hypochlorite ion (OCl⁻), sodium hypochlorite (NaOCl), nitric oxide (NO•), and peroxyxynitrite or peroxyxynitrite anion (ONOO⁻) (unpaired electrons denoted by •). Bacteria have evolved transcription factors that are capable of sensing ROS levels. Different ROS signaling pathways are triggered by different ROS levels and occur with different kinetics (Marinho *et al.*, 2014).

[0311] As used herein, “ROS-inducible regulatory region” refers to a nucleic acid sequence to which one or more ROS-sensing transcription factors is capable of binding, wherein the binding and/or activation of the corresponding transcription factor activates downstream gene expression; in the presence of ROS, the transcription factor binds to and/or activates the regulatory region. In some embodiments, the ROS-inducible regulatory region comprises a promoter sequence. In some embodiments, the transcription factor senses ROS and subsequently binds to the ROS-inducible regulatory region, thereby activating downstream gene expression. In alternate embodiments, the transcription factor is bound to the ROS-inducible regulatory region in the absence of ROS; in the presence of ROS, the transcription factor undergoes a conformational change, thereby activating downstream gene expression. The ROS-inducible regulatory region may be operatively linked to a gene sequence or gene sequence, *e.g.*, a sequence or sequences encoding one or more amino acid catabolism enzyme(s). For example, in the presence of ROS, a transcription factor, *e.g.*, OxyR, senses ROS and activates a corresponding ROS-inducible regulatory region, thereby driving expression of an operatively linked gene sequence or gene sequences. Thus, ROS induces expression of the gene or genes.

[0312] As used herein, “ROS-derepressible regulatory region” refers to a nucleic acid sequence to which one or more ROS-sensing transcription factors is capable of binding, wherein the binding of the corresponding transcription factor represses downstream gene expression; in the presence of ROS, the transcription factor does not bind to and does not repress the regulatory region. In some embodiments, the ROS-derepressible regulatory region comprises a promoter sequence. The ROS-derepressible regulatory region may be operatively linked to a gene or genes, *e.g.*, one or more genes encoding one or more amino acid catabolism enzyme(s). For example, in the presence of ROS, a transcription factor, *e.g.*, OhrR, senses ROS and no longer binds to and/or represses the regulatory region, thereby derepressing an operatively linked gene sequence or gene cassette. Thus, ROS derepresses expression of the gene or gene cassette.

[0313] As used herein, “ROS-repressible regulatory region” refers to a nucleic acid sequence to which one or more ROS-sensing transcription factors is capable of binding, wherein the binding of the corresponding transcription factor represses downstream gene expression; in the presence of ROS, the transcription factor binds to and represses the regulatory region. In some embodiments, the ROS-repressible regulatory region comprises a promoter sequence. In some embodiments, the transcription factor that senses ROS is capable of binding to a regulatory region that overlaps with part of the promoter sequence. In alternate embodiments, the transcription factor that senses ROS is capable of binding to a regulatory region that is upstream or downstream of the promoter sequence. The ROS-repressible regulatory region may be operatively linked to a gene sequence or gene sequences. For example, in the presence of ROS, a transcription factor, *e.g.*, PerR, senses ROS and binds to a corresponding ROS-repressible regulatory region, thereby blocking expression of an operatively linked gene sequence or gene sequences. Thus, ROS represses expression of the gene or genes.

[0314] As used herein, a “ROS-responsive regulatory region” refers to a ROS-inducible regulatory region, a ROS-repressible regulatory region, and/or a ROS-derepressible regulatory region. In some embodiments, the ROS-responsive regulatory region comprises a promoter sequence. Each regulatory region is capable of binding at least one corresponding ROS-sensing transcription factor. Examples of transcription factors that sense ROS and their corresponding ROS-responsive genes, promoters, and/or regulatory regions include, but are not limited to, those shown in **Table 5**.

Table 5. Examples of ROS-sensing transcription factors and ROS-responsive genes

ROS-sensing transcription factor:	Primarily capable of sensing:	Examples of responsive genes, promoters, and/or regulatory regions:
OxyR	H ₂ O ₂	<i>ahpC; ahpF; dps; dsbG; fluF; flu; fur; gor; grxA; hemH; katG; oxyS; sufA; sufB; sufC; sufD; sufE; sufS; trxC; uxuA; yaaA; yaeH; yaiA; ybjM; ydcH; ydeN; ygaQ; yljA; ytfK</i>
PerR	H ₂ O ₂	<i>kata; ahpCF; mrgA; zoaA; fur; hemAXCDBL; srfA</i>
OhrR	Organic peroxides	<i>ohrA</i>

ROS-sensing transcription factor:	Primarily capable of sensing:	Examples of responsive genes, promoters, and/or regulatory regions:
	NaOCl	
SoxR	•O ₂ ⁻ NO• (also capable of sensing H ₂ O ₂)	<i>soxS</i>
RosR	H ₂ O ₂	<i>rbtT; tnp16a; rluC1; tnp5a; mscL; tnp2d; phoD; tnp15b; pstA; tnp5b; xylC; gabD1; rluC2; cgtS9; azlC; narKGHJI; rosR</i>

[0315] In some embodiments, the genetically engineered bacteria comprise a tunable regulatory region that is directly or indirectly controlled by a transcription factor that is capable of sensing at least one reactive oxygen species. The tunable regulatory region is operatively linked to a gene or gene cassette capable of directly or indirectly driving the expression of an amino acid catabolism enzyme, thus controlling expression of the branched chain amino acid catabolism enzyme relative to ROS levels. For example, the tunable regulatory region is a ROS-inducible regulatory region, and the molecule is an amino acid catabolism enzyme; when ROS is present, *e.g.*, in an inflamed tissue, a ROS-sensing transcription factor binds to and/or activates the regulatory region and drives expression of the gene sequence for the amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein thereby producing the amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein. Subsequently, when inflammation is ameliorated, ROS levels are reduced, and production of the branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein is decreased or eliminated.

[0316] In some embodiments, the tunable regulatory region is a ROS-inducible regulatory region; in the presence of ROS, a transcription factor senses ROS and activates the ROS-inducible regulatory region, thereby driving expression of an operatively linked gene or gene cassette. In some embodiments, the transcription factor senses ROS and subsequently binds to the ROS-inducible regulatory region, thereby activating downstream gene expression. In alternate embodiments, the transcription factor is bound to the ROS-inducible regulatory region in the absence of ROS; when the transcription factor senses ROS, it undergoes a conformational change, thereby inducing downstream gene expression.

[0317] In some embodiments, the tunable regulatory region is a ROS-inducible regulatory region, and the transcription factor that senses ROS is OxyR. OxyR “functions primarily as a global regulator of the peroxide stress response” and is capable of regulating dozens of genes, *e.g.*, “genes involved in H₂O₂ detoxification (*katE*, *ahpCF*), heme biosynthesis (*hemH*), reductant supply (*grxA*, *gor*, *trxC*), thiol-disulfide isomerization (*dsbG*), Fe-S center repair (*sufA-E*, *sufS*), iron binding (*yaaA*), repression of iron import systems (*fur*)” and “OxyS, a small regulatory RNA” (Dubbs *et al.*, 2012). The genetically engineered bacteria may comprise any suitable ROS-responsive regulatory

region from a gene that is activated by OxyR. Genes that are capable of being activated by OxyR are known in the art (see, *e.g.*, Zheng *et al.*, 2001; Dubbs *et al.*, 2012; Table 1). In certain embodiments, the genetically engineered bacteria of the invention comprise a ROS-inducible regulatory region from oxyS that is operatively linked to a gene, *e.g.*, a branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein gene. In the presence of ROS, *e.g.*, H₂O₂, an OxyR transcription factor senses ROS and activates to the oxyS regulatory region, thereby driving expression of the operatively linked branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein gene and producing the amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein. In some embodiments, OxyR is encoded by an *E. coli* oxyR gene. In some embodiments, the oxyS regulatory region is an *E. coli* oxyS regulatory region. In some embodiments, the ROS-inducible regulatory region is selected from the regulatory region of katG, dps, and ahpC.

[0318] In alternate embodiments, the tunable regulatory region is a ROS-inducible regulatory region, and the corresponding transcription factor that senses ROS is SoxR. When SoxR is “activated by oxidation of its [2Fe–2S] cluster, it increases the synthesis of SoxS, which then activates its target gene expression” (Koo *et al.*, 2003). “SoxR is known to respond primarily to superoxide and nitric oxide” (Koo *et al.*, 2003), and is also capable of responding to H₂O₂. The genetically engineered bacteria of the invention may comprise any suitable ROS-responsive regulatory region from a gene that is activated by SoxR. Genes that are capable of being activated by SoxR are known in the art (see, *e.g.*, Koo *et al.*, 2003; Table 1). In certain embodiments, the genetically engineered bacteria of the invention comprise a ROS-inducible regulatory region from soxS that is operatively linked to a gene, *e.g.*, an amino acid catabolism enzyme. In the presence of ROS, the SoxR transcription factor senses ROS and activates the soxS regulatory region, thereby driving expression of the operatively linked branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein gene and producing an amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein.

[0319] In some embodiments, the tunable regulatory region is a ROS-derepressible regulatory region, and binding of a corresponding transcription factor represses downstream gene expression; in the presence of ROS, the transcription factor no longer binds to the regulatory region, thereby derepressing the operatively linked gene or gene cassette.

[0320] In some embodiments, the tunable regulatory region is a ROS-derepressible regulatory region, and the transcription factor that senses ROS is OhrR. OhrR “binds to a pair of inverted repeat DNA sequences overlapping the ohrA promoter site and thereby represses the transcription event,” but oxidized OhrR is “unable to bind its DNA target” (Duarte *et al.*, 2010). OhrR is a “transcriptional repressor [that]... senses both organic peroxides and NaOCl” (Dubbs *et al.*, 2012) and is “weakly activated by H₂O₂ but it shows much higher reactivity for organic hydroperoxides” (Duarte *et al.*, 2010). The genetically engineered bacteria of the invention may

comprise any suitable ROS-responsive regulatory region from a gene that is repressed by OhrR. Genes that are capable of being repressed by OhrR are known in the art (see, *e.g.*, Dubbs *et al.*, 2012; Table 1). In certain embodiments, the genetically engineered bacteria of the invention comprise a ROS-derepressible regulatory region from *ohrA* that is operatively linked to a gene or gene cassette, *e.g.*, a branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein gene. In the presence of ROS, *e.g.*, NaOCl, an OhrR transcription factor senses ROS and no longer binds to the *ohrA* regulatory region, thereby derepressing the operatively linked branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein gene and producing an amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein.

[0321] OhrR is a member of the MarR family of ROS-responsive regulators. “Most members of the MarR family are transcriptional repressors and often bind to the -10 or -35 region in the promoter causing a steric inhibition of RNA polymerase binding” (Bussmann *et al.*, 2010). Other members of this family are known in the art and include, but are not limited to, OspR, MgrA, RosR, and SarZ. In some embodiments, the transcription factor that senses ROS is OspR, MgRA, RosR, and/or SarZ, and the genetically engineered bacteria of the invention comprises one or more corresponding regulatory region sequences from a gene that is repressed by OspR, MgRA, RosR, and/or SarZ. Genes that are capable of being repressed by OspR, MgRA, RosR, and/or SarZ are known in the art (see, *e.g.*, Dubbs *et al.*, 2012).

[0322] In some embodiments, the tunable regulatory region is a ROS-derepressible regulatory region, and the corresponding transcription factor that senses ROS is RosR. RosR is “a MarR-type transcriptional regulator” that binds to an “18-bp inverted repeat with the consensus sequence TTGTTGAYRYRTCAACWA (SEQ ID NO: 163)” and is “reversibly inhibited by the oxidant H₂O₂” (Bussmann *et al.*, 2010). RosR is capable of repressing numerous genes and putative genes, including but not limited to “a putative polyisoprenoid-binding protein (cg1322, gene upstream of and divergent from *rosR*), a sensory histidine kinase (cgtS9), a putative transcriptional regulator of the Crp/FNR family (cg3291), a protein of the glutathione S-transferase family (cg1426), two putative FMN reductases (cg1150 and cg1850), and four putative monooxygenases (cg0823, cg1848, cg2329, and cg3084)” (Bussmann *et al.*, 2010). The genetically engineered bacteria of the invention may comprise any suitable ROS-responsive regulatory region from a gene that is repressed by RosR. Genes that are capable of being repressed by RosR are known in the art (see, *e.g.*, Bussmann *et al.*, 2010; Table 1). In certain embodiments, the genetically engineered bacteria of the invention comprise a ROS-derepressible regulatory region from *cgtS9* that is operatively linked to a gene or gene cassette, *e.g.*, an amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein. In the presence of ROS, *e.g.*, H₂O₂, a RosR transcription factor

senses ROS and no longer binds to the *cgtS9* regulatory region, thereby derepressing the operatively linked branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein gene and producing the amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein.

[0323] In some embodiments, it is advantageous for the genetically engineered bacteria to express a ROS-sensing transcription factor that does not regulate the expression of a significant number of native genes in the bacteria. In some embodiments, the genetically engineered bacterium of the invention expresses a ROS-sensing transcription factor from a different species, strain, or substrain of bacteria, wherein the transcription factor does not bind to regulatory sequences in the genetically engineered bacterium of the invention. In some embodiments, the genetically engineered bacterium of the invention is *Escherichia coli*, and the ROS-sensing transcription factor is RosR, *e.g.*, from *Corynebacterium glutamicum*, wherein the *Escherichia coli* does not comprise binding sites for said RosR. In some embodiments, the heterologous transcription factor minimizes or eliminates off-target effects on endogenous regulatory regions and genes in the genetically engineered bacteria.

[0324] In some embodiments, the tunable regulatory region is a ROS-repressible regulatory region, and binding of a corresponding transcription factor represses downstream gene expression; in the presence of ROS, the transcription factor senses ROS and binds to the ROS-repressible regulatory region, thereby repressing expression of the operatively linked gene or gene cassette. In some embodiments, the ROS-sensing transcription factor is capable of binding to a regulatory region that overlaps with part of the promoter sequence. In alternate embodiments, the ROS-sensing transcription factor is capable of binding to a regulatory region that is upstream or downstream of the promoter sequence.

[0325] In some embodiments, the tunable regulatory region is a ROS-repressible regulatory region, and the transcription factor that senses ROS is PerR. In *Bacillus subtilis*, PerR “when bound to DNA, represses the genes coding for proteins involved in the oxidative stress response (*katA*, *ahpC*, and *mrgA*), metal homeostasis (*hemAXCDBL*, *fur*, and *zooA*) and its own synthesis (*perR*)” (Marinho *et al.*, 2014). PerR is a “global regulator that responds primarily to H₂O₂” (Dubbs *et al.*, 2012) and “interacts with DNA at the per box, a specific palindromic consensus sequence (TTATAATNATTATAA (SEQ ID NO: 164)) residing within and near the promoter sequences of PerR-controlled genes” (Marinho *et al.*, 2014). PerR is capable of binding a regulatory region that “overlaps part of the promoter or is immediately downstream from it” (Dubbs *et al.*, 2012). The genetically engineered bacteria of the invention may comprise any suitable ROS-responsive regulatory region from a gene that is repressed by PerR. Genes that are capable of being repressed by PerR are known in the art (see, *e.g.*, Dubbs *et al.*, 2012).

[0326] In these embodiments, the genetically engineered bacteria may comprise a two repressor activation regulatory circuit, which is used to express an amino acid catabolism enzyme. The two repressor activation regulatory circuit comprises a first ROS-sensing repressor, *e.g.*, PerR, and a second repressor, *e.g.*, TetR, which is operatively linked to a gene or gene cassette, *e.g.*, an amino acid catabolism enzyme. In one aspect of these embodiments, the ROS-sensing repressor inhibits transcription of the second repressor, which inhibits the transcription of the gene or gene cassette. Examples of second repressors useful in these embodiments include, but are not limited to, TetR, C1, and LexA. In some embodiments, the ROS-sensing repressor is PerR. In some embodiments, the second repressor is TetR. In this embodiment, a PerR-repressible regulatory region drives expression of TetR, and a TetR-repressible regulatory region drives expression of the gene or gene cassette, *e.g.*, an amino acid catabolism enzyme. In the absence of PerR binding (which occurs in the absence of ROS), tetR is transcribed, and TetR represses expression of the gene or gene cassette, *e.g.*, an amino acid catabolism enzyme. In the presence of PerR binding (which occurs in the presence of ROS), tetR expression is repressed, and the gene or gene cassette, *e.g.*, an amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein is expressed.

[0327] A ROS-responsive transcription factor may induce, derepress, or repress gene expression depending upon the regulatory region sequence used in the genetically engineered bacteria. For example, although “OxyR is primarily thought of as a transcriptional activator under oxidizing conditions... OxyR can function as either a repressor or activator under both oxidizing and reducing conditions” (Dubbs *et al.*, 2012), and OxyR “has been shown to be a repressor of its own expression as well as that of *fhuF* (encoding a ferric ion reductase) and *flu* (encoding the antigen 43 outer membrane protein)” (Zheng *et al.*, 2001). The genetically engineered bacteria of the invention may comprise any suitable ROS-responsive regulatory region from a gene that is repressed by OxyR. In some embodiments, OxyR is used in a two repressor activation regulatory circuit, as described above. Genes that are capable of being repressed by OxyR are known in the art (see, *e.g.*, Zheng *et al.*, 2001; Table 1). Or, for example, although RosR is capable of repressing a number of genes, it is also capable of activating certain genes, *e.g.*, the *narKGHI* operon. In some embodiments, the genetically engineered bacteria comprise any suitable ROS-responsive regulatory region from a gene that is activated by RosR. In addition, “PerR-mediated positive regulation has also been observed...and appears to involve PerR binding to distant upstream sites” (Dubbs *et al.*, 2012). In some embodiments, the genetically engineered bacteria comprise any suitable ROS-responsive regulatory region from a gene that is activated by PerR.

[0328] One or more types of ROS-sensing transcription factors and corresponding regulatory region sequences may be present in genetically engineered bacteria. For example, “OhrR is found in both Gram-positive and Gram-negative bacteria and can coreside with either OxyR or PerR or both” (Dubbs *et al.*, 2012). In some embodiments, the genetically engineered bacteria comprise one type of ROS-sensing transcription factor, *e.g.*, OxyR, and one corresponding regulatory region sequence, *e.g.*,

from *oxyS*. In some embodiments, the genetically engineered bacteria comprise one type of ROS-sensing transcription factor, *e.g.*, OxyR, and two or more different corresponding regulatory region sequences, *e.g.*, from *oxyS* and *katG*. In some embodiments, the genetically engineered bacteria comprise two or more types of ROS-sensing transcription factors, *e.g.*, OxyR and PerR, and two or more corresponding regulatory region sequences, *e.g.*, from *oxyS* and *katA*, respectively. One ROS-responsive regulatory region may be capable of binding more than one transcription factor. In some embodiments, the genetically engineered bacteria comprise two or more types of ROS-sensing transcription factors and one corresponding regulatory region sequence.

[0329] In some embodiments, genetically engineered bacteria comprise a nucleic acid sequence that is at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99% homologous to the DNA sequence of an exemplary OxyR-regulated regulatory region, or a functional fragment thereof. Such sequences include, but are not limited to *aktG*, *dps*, *ahpC*, and *oxyS*.

[0330] In some embodiments, the genetically engineered bacteria of the invention comprise a gene encoding a ROS-sensing transcription factor, *e.g.*, the *oxyR* gene, that is controlled by its native promoter, an inducible promoter, a promoter that is stronger than the native promoter, *e.g.*, the GlnRS promoter or the P(Bla) promoter, or a constitutive promoter. In some instances, it may be advantageous to express the ROS-sensing transcription factor under the control of an inducible promoter in order to enhance expression stability. In some embodiments, expression of the ROS-sensing transcription factor is controlled by a different promoter than the promoter that controls expression of the therapeutic molecule. In some embodiments, expression of the ROS-sensing transcription factor is controlled by the same promoter that controls expression of the therapeutic molecule. In some embodiments, the ROS-sensing transcription factor and therapeutic molecule are divergently transcribed from a promoter region.

[0331] In some embodiments, the genetically engineered bacteria of the invention comprise a gene for a ROS-sensing transcription factor from a different species, strain, or substrain of bacteria. In some embodiments, the genetically engineered bacteria comprise a ROS-responsive regulatory region from a different species, strain, or substrain of bacteria. In some embodiments, the genetically engineered bacteria comprise a ROS-sensing transcription factor and corresponding ROS-responsive regulatory region from a different species, strain, or substrain of bacteria. The heterologous ROS-sensing transcription factor and regulatory region may increase the transcription of genes operatively linked to said regulatory region in the presence of ROS, as compared to the native transcription factor and regulatory region from bacteria of the same subtype under the same conditions.

[0332] In some embodiments, the genetically engineered bacteria comprise a ROS-sensing transcription factor, OxyR, and corresponding regulatory region, *oxyS*, from *Escherichia coli*. In some embodiments, the native ROS-sensing transcription factor, *e.g.*, OxyR, is left intact and retains

wild-type activity. In alternate embodiments, the native ROS-sensing transcription factor, *e.g.*, OxyR, is deleted or mutated to reduce or eliminate wild-type activity.

[0333] In some embodiments, the genetically engineered bacteria of the invention comprise multiple copies of the endogenous gene encoding the ROS-sensing transcription factor, *e.g.*, the oxyR gene. In some embodiments, the gene encoding the ROS-sensing transcription factor is present on a plasmid. In some embodiments, the gene encoding the ROS-sensing transcription factor and the gene or gene cassette for producing the therapeutic molecule are present on different plasmids. In some embodiments, the gene encoding the ROS-sensing transcription factor and the gene or gene cassette for producing the therapeutic molecule are present on the same. In some embodiments, the gene encoding the ROS-sensing transcription factor is present on a chromosome. In some embodiments, the gene encoding the ROS-sensing transcription factor and the gene or gene cassette for producing the therapeutic molecule are present on different chromosomes. In some embodiments, the gene encoding the ROS-sensing transcription factor and the gene or gene cassette for producing the therapeutic molecule are present on the same chromosome.

[0334] In some embodiments, the genetically engineered bacteria comprise a wild-type gene encoding a ROS-sensing transcription factor, *e.g.*, the soxR gene, and a corresponding regulatory region, *e.g.*, a soxS regulatory region, that is mutated relative to the wild-type regulatory region from bacteria of the same subtype. The mutated regulatory region increases the expression of the branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein in the presence of ROS, as compared to the wild-type regulatory region under the same conditions. In some embodiments, the genetically engineered bacteria comprise a wild-type ROS-responsive regulatory region, *e.g.*, the oxyS regulatory region, and a corresponding transcription factor, *e.g.*, OxyR, that is mutated relative to the wild-type transcription factor from bacteria of the same subtype. The mutant transcription factor increases the expression of the branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein in the presence of ROS, as compared to the wild-type transcription factor under the same conditions. In some embodiments, both the ROS-sensing transcription factor and corresponding regulatory region are mutated relative to the wild-type sequences from bacteria of the same subtype in order to increase expression of the branched chain amino acid catabolism enzyme in the presence of ROS.

[0335] In some embodiments, the gene or gene cassette for producing the branched chain amino acid catabolism enzyme is present on a plasmid and operably linked to a promoter that is induced by ROS. In some embodiments, the gene or gene cassette for producing the branched chain amino acid catabolism enzyme is present in the chromosome and operably linked to a promoter that is induced by ROS. In some embodiments, the gene or gene cassette for producing the branched chain amino acid catabolism enzyme is present on a chromosome and operably linked to a promoter that is induced by exposure to tetracycline. In some embodiments, the gene or gene cassette for producing the branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein

is present on a plasmid and operably linked to a promoter that is induced by exposure to tetracycline. In some embodiments, expression is further optimized by methods known in the art, *e.g.*, by optimizing ribosomal binding sites, manipulating transcriptional regulators, and/or increasing mRNA stability.

[0336] In some embodiments, the genetically engineered bacteria may comprise multiple copies of the gene(s) capable of producing an amino acid catabolism enzyme(s), BCAA transporter(s), and/or BCAA binding protein(s). In some embodiments, the gene(s) capable of producing an amino acid catabolism enzyme(s), BCAA transporter(s), and/or BCAA binding protein(s) is present on a plasmid and operatively linked to a ROS-responsive regulatory region. In some embodiments, the gene(s) capable of producing a branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein is present in a chromosome and operatively linked to a ROS-responsive regulatory region.

[0337] Thus, in some embodiments, the genetically engineered bacteria or genetically engineered yeast or virus produce one or more amino acid catabolism enzymes under the control of an oxygen level-dependent promoter, a reactive oxygen species (ROS)-dependent promoter, or a reactive nitrogen species (RNS)-dependent promoter, and a corresponding transcription factor.

[0338] In some embodiments, the genetically engineered bacteria comprise a stably maintained plasmid or chromosome carrying a gene for producing an amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein such that the branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein can be expressed in the host cell, and the host cell is capable of survival and/or growth *in vitro*, *e.g.*, in medium, and/or *in vivo*. In some embodiments, a bacterium may comprise multiple copies of the gene encoding the amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein. In some embodiments, the gene encoding the branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein is expressed on a low-copy plasmid. In some embodiments, the low-copy plasmid may be useful for increasing stability of expression. In some embodiments, the low-copy plasmid may be useful for decreasing leaky expression under non-inducing conditions. In some embodiments, the gene encoding the branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein is expressed on a high-copy plasmid. In some embodiments, the high-copy plasmid may be useful for increasing expression of the amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein. In some embodiments, the gene encoding the branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein is expressed on a chromosome.

[0339] In some embodiments, the bacteria are genetically engineered to include multiple mechanisms of action (MOAs), *e.g.*, circuits producing multiple copies of the same product (*e.g.*, to enhance copy number) or circuits performing multiple different functions. For example, the genetically engineered bacteria may include four copies of the gene encoding a particular branched

chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein inserted at four different insertion sites. Alternatively, the genetically engineered bacteria may include three copies of the gene encoding a particular branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein inserted at three different insertion sites and three copies of the gene encoding a different branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein inserted at three different insertion sites.

[0340] In some embodiments, under conditions where the branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein is expressed, the genetically engineered bacteria of the disclosure produce at least about 1.5-fold, at least about 2-fold, at least about 10-fold, at least about 15-fold, at least about 20-fold, at least about 30-fold, at least about 50-fold, at least about 100-fold, at least about 200-fold, at least about 300-fold, at least about 400-fold, at least about 500-fold, at least about 600-fold, at least about 700-fold, at least about 800-fold, at least about 900-fold, at least about 1,000-fold, or at least about 1,500-fold more of the amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein and/or transcript of the gene(s) in the operon as compared to unmodified bacteria of the same subtype under the same conditions.

[0341] In some embodiments, quantitative PCR (qPCR) is used to amplify, detect, and/or quantify mRNA expression levels of the branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein gene(s). Primers specific for branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein gene(s) may be designed and used to detect mRNA in a sample according to methods known in the art. In some embodiments, a fluorophore is added to a sample reaction mixture that may contain branched chain amino acid catabolism enzyme mRNA, and a thermal cycler is used to illuminate the sample reaction mixture with a specific wavelength of light and detect the subsequent emission by the fluorophore. The reaction mixture is heated and cooled to predetermined temperatures for predetermined time periods. In certain embodiments, the heating and cooling is repeated for a predetermined number of cycles. In some embodiments, the reaction mixture is heated and cooled to 90-100° C, 60-70° C, and 30-50° C for a predetermined number of cycles. In a certain embodiment, the reaction mixture is heated and cooled to 93-97°C, 55-65°C, and 35-45°C for a predetermined number of cycles. In some embodiments, the accumulating amplicon is quantified after each cycle of the qPCR. The number of cycles at which fluorescence exceeds the threshold is the threshold cycle (CT). At least one CT result for each sample is generated, and the CT result(s) may be used to determine mRNA expression levels of the branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein gene(s).

[0342] In some embodiments, quantitative PCR (qPCR) is used to amplify, detect, and/or quantify mRNA expression levels of the branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein gene(s). Primers specific for branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein gene(s) may be designed and

used to detect mRNA in a sample according to methods known in the art. In some embodiments, a fluorophore is added to a sample reaction mixture that may contain branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein mRNA, and a thermal cycler is used to illuminate the sample reaction mixture with a specific wavelength of light and detect the subsequent emission by the fluorophore. The reaction mixture is heated and cooled to predetermined temperatures for predetermined time periods. In certain embodiments, the heating and cooling is repeated for a predetermined number of cycles. In some embodiments, the reaction mixture is heated and cooled to 90-100°C, 60-70°C, and 30-50°C for a predetermined number of cycles. In a certain embodiment, the reaction mixture is heated and cooled to 93-97°C, 55-65°C, and 35-45°C for a predetermined number of cycles. In some embodiments, the accumulating amplicon is quantified after each cycle of the qPCR. The number of cycles at which fluorescence exceeds the threshold is the threshold cycle (CT). At least one CT result for each sample is generated, and the CT result(s) may be used to determine mRNA expression levels of the branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein gene(s).

[0343] In other embodiments, the inducible promoter is a propionate responsive promoter. For example, the *prpR* promoter is a propionate responsive promoter.

Inducible Promoters (nutritional and/or chemical inducer(s) and/or metabolite(s))

[0344] In some embodiments, one or more optimized gene sequence(s), *e.g.*, *kivD*, *leuD^H*, *adh2*, and *brnQ*, is present on a plasmid and operably linked to promoter a directly or indirectly inducible by one or more nutritional and/or chemical inducer(s) and/or metabolite(s). In some embodiments, the bacterial cell comprises a stably maintained plasmid or chromosome carrying the gene encoding the branched chain amino acid catabolism enzyme, which is induced by one or more nutritional and/or chemical inducer(s) and/or metabolite(s), such that the branched chain amino acid catabolism enzyme can be expressed in the host cell, and the host cell is capable of survival and/or growth *in vitro*, *e.g.*, under culture conditions, and/or *in vivo*, *e.g.*, in the gut. In some embodiments, bacterial cell comprises two or more distinct branched chain amino acid catabolism cassette(s), one or more of which are induced by one or more nutritional and/or chemical inducer(s) and/or metabolite(s). In some embodiments, the genetically engineered bacteria comprise multiple copies of the same branched chain amino acid catabolism enzyme gene(s) and/or gene cassette(s) which are induced by one or more nutritional and/or chemical inducer(s) and/or metabolite(s). In some embodiments, the genetically engineered bacteria comprise multiple copies of different branched chain amino acid catabolism enzyme genes or gene cassette(s), one or more of which are induced by one or more nutritional and/or chemical inducer(s) and/or metabolite(s).

[0345] In some embodiments, the gene encoding the branched chain amino acid catabolism enzyme is present on a plasmid and operably linked to a promoter that is induced by one or more nutritional and/or chemical inducer(s) and/or metabolite(s). In some embodiments, the gene encoding the branched chain amino acid catabolism enzyme is present in the chromosome and operably linked

to a promoter that is induced by one or more nutritional and/or chemical inducer(s) and/or metabolite(s).

[0346] In some embodiments, the bacterial cell comprises a stably maintained plasmid or chromosome carrying the one or more gene sequences(s), inducible by one or more nutritional and/or chemical inducer(s) and/or metabolite(s), encoding a transporter of branched chain amino acid(s) and/or one or more metabolites thereof, *e.g.*, *livKHMGF* and/or *brnQ*, such that the transporter can be expressed in the host cell, and the host cell is capable of survival and/or growth *in vitro*, *e.g.*, in medium, and/or *in vivo*, *e.g.*, in the gut. In some embodiments, bacterial cell comprises two or more distinct copies of the one or more gene sequences(s) encoding a branched chain amino acid transporter, which is controlled by a promoter inducible one or more nutritional and/or chemical inducer(s) and/or metabolite(s). In some embodiments, the genetically engineered bacteria comprise multiple copies of the same one or more gene sequences(s) encoding a branched chain amino acid transporter, which is controlled by a promoter inducible one or more nutritional and/or chemical inducer(s) and/or metabolite(s). In some embodiments, the one or more gene sequences(s) encoding a transporter of branched chain amino acid(s), is present on a plasmid and operably linked to a directly or indirectly inducible promoter inducible by one or more nutritional and/or chemical inducer(s) and/or metabolite(s). In some embodiments, the one or more gene sequences(s) encoding a branched chain amino acid transporter, is present on a chromosome and operably linked to a directly or indirectly inducible by one or more nutritional and/or chemical inducer(s) and/or metabolite(s).

[0347] In some embodiments, the gene encoding branched chain amino acid binding protein is present on a plasmid and operably linked to a promoter that is induced by one or more nutritional and/or chemical inducer(s) and/or metabolite(s). In some embodiments, the gene encoding branched chain amino acid binding protein is present in the chromosome and operably linked to a promoter that is induced by one or more nutritional and/or chemical inducer(s) and/or metabolite(s).

[0348] In some embodiments, the promoter that is operably linked to the gene encoding the branched chain amino acid catabolism enzyme and the promoter that is operably linked to the gene encoding the branched chain amino acid transporter and/or BCAA binding protein and/or BCAA exporter, is directly or indirectly induced by one or more nutritional and/or chemical inducer(s) and/or metabolite(s).

[0349] In some embodiments, one or more inducible promoter(s) are useful for or induced during *in vivo* expression of the one or more protein(s) of interest. In some embodiments, the promoters are induced during *in vivo* expression of one or more branched chain amino acid catabolism enzymes and/or branched chain amino acid transporter(s) and/or BCAA binding protein(s) and/or BCAA exporter(s). In some embodiments, expression of one or more branched chain amino acid catabolism enzyme(s) and/or branched chain amino acid transporter(s) and/or BCAA binding protein(s) and/or BCAA exporter(s) is driven directly or indirectly by one or more arabinose inducible promoter(s) *in vivo*. In some embodiments, the promoter is directly or indirectly induced by a

chemical and/or nutritional inducer and/or metabolite which is co-administered with the genetically engineered bacteria of the invention.

[0350] In some embodiments, expression of one or more branched chain amino acid catabolism enzyme gene(s) and/or branched chain amino acid transporter(s) and/or BCAA binding protein(s) and/or BCAA exporter(s), is driven directly or indirectly by one or more promoter(s) induced by a chemical and/or nutritional inducer and/or metabolite during *in vitro* growth, preparation, or manufacturing of the strain prior to *in vivo* administration. In some embodiments, the promoter(s) induced by a chemical and/or nutritional inducer and/or metabolite are induced in culture, *e.g.*, grown in a flask, fermenter or other appropriate culture vessel, *e.g.*, used during cell growth, cell expansion, fermentation, recovery, purification, formulation, and/or manufacture. In some embodiments, the promoter is directly or indirectly induced by a molecule that is added to in the bacterial culture to induce expression and pre-load the bacterium with branched chain amino acid catabolism enzyme(s) and/or branched chain amino acid transporter(s) and/or BCAA binding protein(s) and/or BCAA exporter(s) prior to administration. In some embodiments, the cultures, which are induced by a chemical and/or nutritional inducer and/or metabolite, are grown aerobically. In some embodiments, the cultures, which are induced by a chemical and/or nutritional inducer and/or metabolite, are grown anaerobically.

[0351] In some embodiments, the genetically engineered bacteria encode one or more gene sequence(s) which are inducible through an arabinose inducible system.

[0352] The genes of arabinose metabolism are organized in one operon, AraBAD, which is controlled by the P_{AraBAD} promoter. The P_{AraBAD} (or Para) promoter suitably fulfills the criteria of inducible expression systems. P_{AraBAD} displays tighter control of payload gene expression than many other systems, likely due to the dual regulatory role of AraC, which functions both as an inducer and as a repressor. Additionally, the level of ParaBAD-based expression can be modulated over a wide range of L-arabinose concentrations to fine-tune levels of expression of the payload. However, the cell population exposed to sub-saturating L-arabinose concentrations is divided into two subpopulations of induced and uninduced cells, which is determined by the differences between individual cells in the availability of L-arabinose transporter (Zhang *et al.*, Development and Application of an Arabinose-Inducible Expression System by Facilitating Inducer Uptake in *Corynebacterium glutamicum*; Appl. Environ. Microbiol. August 2012 vol. 78 no. 16 5831-5838). Alternatively, inducible expression from the ParaBad can be controlled or fine-tuned through the optimization of the ribosome binding site (RBS), as described herein.

[0353] In one embodiment, expression of one or more optimized enzyme(s), *e.g.*, *kivD*, *leuDH*, *adh2*, and/or *brnQ*, *e.g.*, as described herein, is driven directly or indirectly by one or more arabinose inducible promoter(s). In one embodiment, the genetically engineered bacteria encode one or more branched chain amino acid transporter(s) *e.g.*, *livKHMGF* and/or *brnQ*, described herein, whose expression is driven directly or indirectly by one or more arabinose inducible promoter(s). In

one embodiment, expression of one or more branched chain amino acid exporter(s), *e.g.*, as described herein, is driven directly or indirectly by one or more arabinose inducible promoter(s). In some embodiments, the arabinose inducible promoter is useful for or induced during *in vivo* expression of the one or more protein(s) of interest. In some embodiments, expression of one or more branched chain amino acid catabolism enzyme(s) and/or branched chain amino acid transporter(s) and/or BCAA binding protein(s) and/or BCAA exporter(s) is driven directly or indirectly by one or more arabinose inducible promoter(s) *in vivo*. In some embodiments, the promoter is directly or indirectly induced by a molecule (*e.g.*, arabinose) that is co-administered with the genetically engineered bacteria of the invention.

[0354] In some embodiments, expression of one or more branched chain amino acid catabolism enzyme(s) and/or branched chain amino acid transporter(s) and/or BCAA binding protein(s) and/or BCAA exporter(s), is driven directly or indirectly by one or more arabinose inducible promoter(s) during *in vitro* growth, preparation, or manufacturing of the strain prior to *in vivo* administration. In some embodiments, the arabinose inducible promoter(s) are induced in culture, *e.g.*, grown in a flask, fermenter or other appropriate culture vessel, *e.g.*, used during cell growth, cell expansion, fermentation, recovery, purification, formulation, and/or manufacture. In some embodiments, the promoter is directly or indirectly induced by a molecule, *e.g.*, arabinose, that is added to in the bacterial culture to induce expression and pre-load the bacterium with branched chain amino acid catabolism enzyme(s) and /or branched chain amino acid transporter(s) and/or BCAA binding protein(s) and/or BCAA exporter(s) prior to administration. In some embodiments, the cultures, which are induced by arabinose, are grown aerobically. In some embodiments, the cultures, which are induced by arabinose, are grown anaerobically.

[0355] . In some embodiments, bacterial cell comprises two or more distinct branched chain amino acid catabolism cassette(s) or transporter(s) or binding protein(s) or exporter(s), one or more of which are induced by arabinose. In some embodiments, the genetically engineered bacteria comprise multiple copies of the same branched chain amino acid catabolism enzyme gene sequence(s) and/or transporter gene sequence(s), *e.g.*, as described herein, which are induced by one or more nutritional and/or chemical inducer(s) and/or metabolite(s). In some embodiments, the genetically engineered bacteria comprise multiple copies of different branched chain amino acid catabolism enzyme genes sequence(s) and/or transporter gene sequence(s) and/or BCAA binding protein gene sequence(s) and/or BCAA exporter gene sequence(s), *e.g.*, as described herein, one or more of which are induced by one or more nutritional and/or chemical inducer(s) and/or metabolite(s).

[0356] In a first example, the arabinose inducible promoter drives the expression of a construct comprising one or more polypeptides of interest described herein jointly with a second promoter, *e.g.*, a second constitutive or inducible promoter. In some embodiments, two promoters are positioned proximally to the construct and drive its expression, wherein the arabinose inducible promoter drives expression under a first set of exogenous conditions, and the second promoter drives

the expression under a second set of exogenous conditions. In second example, the arabinose promoter drives the expression of one or more gene cassette(s) under a first inducing condition and another inducible promoter drives the expression of one or more of the same or different gene cassette(s) expressing one or more polypeptides of interest, under a second inducing condition. In both examples, the first and second conditions can be two sequential inducing culture conditions (*i.e.*, during preparation of the culture in a flask, fermenter or other appropriate culture vessel, *e.g.*, arabinose and IPTG). In another non-limiting example, the first inducing conditions are culture conditions, *e.g.*, the presence of arabinose, and the second inducing conditions are *in vivo* conditions. Such *in vivo* conditions include low-oxygen, microaerobic, or anaerobic conditions, presence of gut metabolites, and/or nutritional and/or chemical inducers and/or metabolites administered in combination with the bacterial strain. In some embodiments, the one or more arabinose promoters drive expression of one or more protein(s) of interest, in combination with the FNR promoter driving the expression of the same gene sequence(s).

[0357] In some embodiments, the gene sequence(s) encoding the branched chain amino acid catabolism enzyme(s) or other polypeptide(s) of interest, are present on a plasmid and operably linked to a promoter that is induced by arabinose. In some embodiments, the gene sequence(s) encoding the branched chain amino acid catabolism enzyme(s) or branched chain amino acid transporter(s) is present in the chromosome and operably linked to a promoter that is induced by arabinose.

[0358] In some embodiments, the arabinose inducible construct further comprises a gene encoding AraC, which is divergently transcribed from the same promoter as the one or more one or more branched chain amino acid catabolism enzyme(s).

[0359] In some embodiments, the genetically engineered bacteria comprise one or more gene sequence(s) which are inducible through a rhamnose inducible system. The genes rhaBAD are organized in one operon which is controlled by the rhaP BAD promoter. The rhaP BAD promoter is regulated by two activators, RhaS and RhaR, and the corresponding genes belong to one transcription unit which divergently transcribed in the opposite direction of rhaBAD. In the presence of L-rhamnose, RhaR binds to the rhaP RS promoter and activates the production of RhaR and RhaS. RhaS together with L-rhamnose then bind to the rhaP BAD and the rhaP T promoter and activate the transcription of the structural genes. In contrast to the arabinose system, in which AraC is provided and divergently transcribed in the gene sequence(s), it is not necessary to express the regulatory proteins in larger quantities in the rhamnose expression system because the amounts expressed from the chromosome are sufficient to activate transcription even on multi-copy plasmids. Therefore, only the rhaP BAD promoter is cloned upstream of the gene that is to be expressed. Full induction of rhaBAD transcription also requires binding of the CRP-cAMP complex, which is a key regulator of catabolite repression. Alternatively, inducible expression from the rhaBAD can be controlled or fine-tuned through the optimization of the ribosome binding site (RBS), as described herein.

[0360] In one embodiment, expression of one or more branched chain amino acid catabolism enzyme(s), *e.g.*, *kivD*, *leuDH*, *adh2*, and/or *brnQ*, *e.g.*, as described herein, is driven directly or indirectly by one or more rhamnose inducible promoter(s). In one embodiment, the genetically engineered bacteria encode one or more branched chain amino acid transporter(s), *e.g.*, *livKHMGF* and/or *brnQ*, described herein, whose expression is driven directly or indirectly by one or more rhamnose inducible promoter(s). In one embodiment, the genetically engineered bacteria encode one or more branched chain amino acid exporter(s), described herein, whose expression is driven directly or indirectly by one or more rhamnose inducible promoter(s).

[0361] In some embodiments, the rhamnose inducible promoter is useful for or induced during *in vivo* expression of the one or more protein(s) of interest. In some embodiments, expression of one or more branched chain amino acid catabolism enzyme(s) and/or branched chain amino acid transporter(s) and/or BCAA binding protein(s) and/or BCAA exporter(s) is driven directly or indirectly by one or more rhamnose inducible promoter(s) *in vivo*. In some embodiments, the promoter is directly or indirectly induced by a molecule (*e.g.*, rhamnose) that is co-administered with the genetically engineered bacteria of the invention.

[0362] In some embodiments, expression of one or more branched chain amino acid catabolism enzyme(s) and/or branched chain amino acid transporter(s) and/or BCAA binding protein(s) and/or BCAA exporter(s), is driven directly or indirectly by one or more rhamnose inducible promoter(s) during *in vitro* growth, preparation, or manufacturing of the strain prior to *in vivo* administration. In some embodiments, the rhamnose inducible promoter(s) are induced in culture, *e.g.*, grown in a flask, fermenter or other appropriate culture vessel, *e.g.*, used during cell growth, cell expansion, fermentation, recovery, purification, formulation, and/or manufacture. In some embodiments, the promoter is directly or indirectly induced by a molecule, *e.g.*, rhamnose, that is added to in the bacterial culture to induce expression and pre-load the bacterium with branched chain amino acid catabolism enzyme(s) and/or BCAA transporter(s) and/or BCAA binding protein(s) and/or BCAA exporter(s) prior to administration. In some embodiments, the cultures, which are induced by rhamnose, are grown aerobically. In some embodiments, the cultures, which are induced by rhamnose, are grown anaerobically.

[0363] In some embodiments, bacterial cell comprises two or more distinct branched chain amino acid catabolism cassette(s) or other polypeptide(s) of interest, one or more of which are induced by rhamnose. In some embodiments, the genetically engineered bacteria comprise multiple copies of the same branched chain amino acid catabolism enzyme gene sequence(s) and/or transporter gene sequence(s) and/or BCAA binding protein gene sequence(s) and/or BCAA exporter gene sequence(s), *e.g.*, as described herein, which are induced by rhamnose. In some embodiments, the genetically engineered bacteria comprise multiple copies of different branched chain amino acid catabolism enzyme genes sequence(s) and/or transporter gene sequence(s) and/or BCAA binding

protein gene sequence(s) and/or BCAA exporter gene sequence(s), *e.g.*, as described herein, one or more of which are induced by rhamnose.

[0364] In a first example, the rhamnose inducible promoter drives the expression of a construct comprising one or more polypeptides of interest described herein jointly with a second promoter, *e.g.*, a second constitutive or inducible promoter. In some embodiments, two promoters are positioned proximally to the construct and drive its expression, wherein the rhamnose inducible promoter drives expression under a first set of exogenous conditions, and the second promoter drives the expression under a second set of exogenous conditions. In second example, the rhamnose promoter drives the expression of one or more gene cassette(s) under a first inducing condition and another inducible promoter drives the expression of one or more of the same or different gene cassette(s) expressing one or more polypeptides of interest, under a second inducing condition. In both examples, the first and second conditions can be two sequential inducing culture conditions (*i.e.*, during preparation of the culture in a flask, fermenter or other appropriate culture vessel, *e.g.*, rhamnose and IPTG). In another non-limiting example, the first inducing conditions are culture conditions, *e.g.*, the presence of rhamnose, and the second inducing conditions are *in vivo* conditions. Such *in vivo* conditions include low-oxygen, microaerobic, or anaerobic conditions, presence of gut metabolites, and/or nutritional and/or chemical inducers and/or metabolites administered in combination with the bacterial strain. In some embodiments, the one or more rhamnose promoters drive expression of one or more protein(s) of interest, in combination with the FNR promoter driving the expression of the same gene sequence(s).

[0365] In some embodiments, the gene sequence(s) encoding the branched chain amino acid catabolism enzyme(s) or other polypeptide(s) of interest, are present on a plasmid and operably linked to a promoter that is induced by rhamnose. In some embodiments, the gene sequence(s) encoding the branched chain amino acid catabolism enzyme(s) or branched chain amino acid transporter(s) and/or BCAA binding protein(s) and/or BCAA exporter(s) is present in the chromosome and operably linked to a promoter that is induced by rhamnose.

[0366] In some embodiments, the genetically engineered bacteria comprise one or more gene sequence(s) which are inducible through an Isopropyl β -D-1-thiogalactopyranoside (IPTG) inducible system or other compound which induced transcription from the Lac Promoter. IPTG is a molecular mimic of allolactose, a lactose metabolite that activates transcription of the lac operon. In contrast to allolactose, the sulfur atom in IPTG creates a non-hydrolyzable chemical bond, which prevents the degradation of IPTG, allowing the concentration to remain constant. IPTG binds to the lac repressor and releases the tetrameric repressor (lacI) from the lac operator in an allosteric manner, thereby allowing the transcription of genes in the lac operon. Since IPTG is not metabolized by *E. coli*, its concentration stays constant and the rate of expression of Lac promoter-controlled is tightly controlled, both *in vivo* and *in vitro*. IPTG intake is independent on the action of lactose permease, since other transport pathways are also involved. Inducible expression from the PLac can be

controlled or fine-tuned through the optimization of the ribosome binding site (RBS), as described herein. Other compounds which inactivate LacI, can be used instead of IPTG in a similar manner.

[0367] In one embodiment, expression of the optimized enzyme(s), *e.g.*, *kivD*, *leuDh*, *adh2*, and *brnQ*, *e.g.*, as described herein, is driven directly or indirectly by one or more IPTG inducible promoter(s). In one embodiment, the genetically engineered bacteria encode one or more branched chain amino acid transporter(s), *e.g.*, *livKHMgf* and/or *brnQ*, described herein, whose expression is driven directly or indirectly by one or more IPTG inducible promoter(s). In one embodiment, the genetically engineered bacteria encode one or more branched chain amino acid exporter(s) described herein, whose expression is driven directly or indirectly by one or more IPTG inducible promoter(s).

[0368] In some embodiments, the IPTG inducible promoter is useful for or induced during *in vivo* expression of the one or more protein(s) of interest. In some embodiments, expression of one or more branched chain amino acid catabolism enzyme(s) and/or branched chain amino acid transporter(s) is driven directly or indirectly by one or more IPTG inducible promoter(s) *in vivo*. In some embodiments, the promoter is directly or indirectly induced by a molecule (*e.g.*, IPTG) that is co-administered with the genetically engineered bacteria of the invention.

[0369] In some embodiments, expression of one or more branched chain amino acid catabolism enzyme(s) and/or branched chain amino acid transporter(s) and/or BCAA binding protein(s) and/or BCAA exporter(s), is driven directly or indirectly by one or more IPTG inducible promoter(s) during *in vitro* growth, preparation, or manufacturing of the strain prior to *in vivo* administration. In some embodiments, the IPTG inducible promoter(s) are induced in culture, *e.g.*, grown in a flask, fermenter or other appropriate culture vessel, *e.g.*, used during cell growth, cell expansion, fermentation, recovery, purification, formulation, and/or manufacture. In some embodiments, the promoter is directly or indirectly induced by a molecule, *e.g.*, IPTG, that is added to in the bacterial culture to induce expression and pre-load the bacterium with branched chain amino acid catabolism enzyme(s) prior to administration. In some embodiments, the cultures, which are induced by IPTG, are grown aerobically. In some embodiments, the cultures, which are induced by IPTG, are grown anaerobically.

[0370] . In some embodiments, bacterial cell comprises two or more distinct branched chain amino acid catabolism cassette(s) or other polypeptide(s) of interest, one or more of which are induced by IPTG. In some embodiments, the genetically engineered bacteria comprise multiple copies of the same branched chain amino acid catabolism enzyme gene sequence(s) and/or transporter gene sequence(s), *e.g.*, as described herein, which are induced IPTG. In some embodiments, the genetically engineered bacteria comprise multiple copies of different branched chain amino acid catabolism enzyme genes sequence(s) and/or transporter gene sequence(s) and/or other gene sequence(s) of interest, as described herein, one or more of which are induced by IPTG.

[0371] In a first example, the IPTG inducible promoter drives the expression of a construct comprising one or more polypeptides of interest described herein jointly with a second promoter, *e.g.*,

a second constitutive or inducible promoter. In some embodiments, two promoters are positioned proximally to the construct and drive its expression, wherein the IPTG inducible promoter drives expression under a first set of exogenous conditions, and the second promoter drives the expression under a second set of exogenous conditions. In second example, the IPTG promoter drives the expression of one or more gene cassette(s) under a first inducing condition and another inducible promoter drives the expression of one or more of the same or different gene cassette(s) expressing one or more polypeptides of interest, under a second inducing condition. In both examples, the first and second conditions can be two sequential inducing culture conditions (*i.e.*, during preparation of the culture in a flask, fermenter or other appropriate culture vessel, *e.g.*, IPTG and IPTG). In another non-limiting example, the first inducing conditions are culture conditions, *e.g.*, the presence of IPTG, and the second inducing conditions are *in vivo* conditions. Such *in vivo* conditions include low-oxygen, microaerobic, or anaerobic conditions, presence of gut metabolites, and/or nutritional and/or chemical inducers and/or metabolites administered in combination with the bacterial strain. In some embodiments, the one or more IPTG promoters drive expression of one or more protein(s) of interest, in combination with the FNR promoter driving the expression of the same gene sequence(s).

[0372] In some embodiments, the gene sequence(s) encoding the branched chain amino acid catabolism enzyme(s) or other polypeptide(s) of interest, are present on a plasmid and operably linked to a promoter that is induced by IPTG. In some embodiments, the gene sequence(s) encoding the branched chain amino acid catabolism enzyme(s) or branched chain amino acid transporter(s) is present in the chromosome and operably linked to a promoter that is induced by IPTG.

[0373] In some embodiments, the IPTG inducible construct further comprises a gene encoding *lacI*, which is divergently transcribed from the same promoter as the one or more one or more branched chain amino acid catabolism enzyme(s) and/or transporters described herein.

[0374] In some embodiments, the genetically engineered bacteria comprise one or more gene sequence(s) which are inducible through a tetracycline inducible system. The initial system Gossen and Bujard (Gossen M & Bujard H. *PNAS*, 1992 Jun 15;89(12):5547-51) developed is known as tetracycline off: in the presence of tetracycline, expression from a tet-inducible promoter is reduced. Tetracycline-controlled transactivator (tTA) was created by fusing *tetR* with the C-terminal domain of VP16 (virion protein 16) from herpes simplex virus. In the absence of tetracycline, the *tetR* portion of tTA will bind *tetO* sequences in the tet promoter, and the activation domain promotes expression. In the presence of tetracycline, tetracycline binds to *tetR*, precluding tTA from binding to the *tetO* sequences. Next, a reverse Tet repressor (rTetR), was developed which created a reliance on the presence of tetracycline for induction, rather than repression. The new transactivator rTA (reverse tetracycline-controlled transactivator) was created by fusing rTetR with VP16. The tetracycline on system is also known as the rTA-dependent system.

[0375] In one embodiment, expression of the optimized enzyme(s), *e.g.*, *kivD*, *leuDH*, *adh2*, and *brnQ*, *e.g.*, as described herein, is driven directly or indirectly by one or more tetracycline

inducible promoter(s). In one embodiment, the genetically engineered bacteria encode one or more branched chain amino acid transporter(s), *e.g.*, *livKHMGF* and/or *brnQ*, described herein, whose expression is driven directly or indirectly by one or more tetracycline inducible promoter(s).

[0376] In one embodiment, the genetically engineered bacteria encode one or more branched chain amino acid exporter(s).

[0377] In some embodiments, the tetracycline inducible promoter is useful for or induced during *in vivo* expression of the one or more protein(s) of interest. In some embodiments, expression of one or more branched chain amino acid catabolism enzyme(s) and/or branched chain amino acid transporter(s) and or other polypeptide(s) of interest is driven directly or indirectly by one or more tetracycline inducible promoter(s) *in vivo*. In some embodiments, the promoter is directly or indirectly induced by a molecule (*e.g.*, tetracycline) that is co-administered with the genetically engineered bacteria of the invention.

[0378] In some embodiments, expression of one or more branched chain amino acid catabolism enzyme(s) and/or branched chain amino acid transporter(s) and or other polypeptide(s) of interest, is driven directly or indirectly by one or more tetracycline inducible promoter(s) during *in vitro* growth, preparation, or manufacturing of the strain prior to *in vivo* administration. In some embodiments, the tetracycline inducible promoter(s) are induced in culture, *e.g.*, grown in a flask, fermenter or other appropriate culture vessel, *e.g.*, used during cell growth, cell expansion, fermentation, recovery, purification, formulation, and/or manufacture. In some embodiments, the promoter is directly or indirectly induced by a molecule, *e.g.*, tetracycline, that is added to in the bacterial culture to induce expression and pre-load the bacterium with branched chain amino acid catabolism enzyme(s) prior to administration. In some embodiments, the cultures, which are induced by tetracycline, are grown aerobically. In some embodiments, the cultures, which are induced by tetracycline, are grown anaerobically.

[0379] In some embodiments, bacterial cell comprises two or more distinct branched chain amino acid catabolism cassette(s) and/or other polypeptide(s) of interest, one or more of which are induced by tetracycline. In some embodiments, the genetically engineered bacteria comprise multiple copies of the same branched chain amino acid catabolism enzyme gene sequence(s) and/or transporter gene sequence(s) and/or gene sequence(s) for the expression of other polypeptide(s) of interest, *e.g.*, as described herein, which are induced by tetracycline. In some embodiments, the genetically engineered bacteria comprise multiple copies of different branched chain amino acid catabolism enzyme genes sequence(s) and/or transporter gene sequence(s) and gene sequence(s) for the expression of other polypeptide(s) of interest, *e.g.*, as described herein, one or more of which are induced by tetracycline.

[0380] In a first example, the tetracycline inducible promoter drives the expression of a construct comprising one or more polypeptides of interest described herein jointly with a second promoter, *e.g.*, a second constitutive or inducible promoter. In some embodiments, two promoters are

positioned proximally to the construct and drive its expression, wherein the tetracycline inducible promoter drives expression under a first set of exogenous conditions, and the second promoter drives the expression under a second set of exogenous conditions. In second example, the tetracycline promoter drives the expression of one or more gene cassette(s) under a first inducing condition and another inducible promoter drives the expression of one or more of the same or different gene cassette(s) expressing one or more polypeptides of interest, under a second inducing condition. In both examples, the first and second conditions can be two sequential inducing culture conditions (*i.e.*, during preparation of the culture in a flask, fermenter or other appropriate culture vessel, *e.g.*, tetracycline and IPTG). In another non-limiting example, the first inducing conditions are culture conditions, *e.g.*, the presence of tetracycline, and the second inducing conditions are *in vivo* conditions. Such *in vivo* conditions include low-oxygen, microaerobic, or anaerobic conditions, presence of gut metabolites, and/or nutritional and/or chemical inducers and/or metabolites administered in combination with the bacterial strain. In some embodiments, the one or more tetracycline promoters drive expression of one or more protein(s) of interest, in combination with the FNR promoter driving the expression of the same gene sequence(s).

[0381] In some embodiments, the gene sequence(s) encoding the branched chain amino acid catabolism enzyme(s) or other polypeptide(s) of interest, are present on a plasmid and operably linked to a promoter that is induced by tetracycline. In some embodiments, the gene sequence(s) encoding the branched chain amino acid catabolism enzyme(s) or branched chain amino acid transporter(s) is present in the chromosome and operably linked to a promoter that is induced by tetracycline.

[0382] In some embodiments, the tetracycline inducible construct further comprises a gene encoding AraC, which is divergently transcribed from the same promoter as the one or more one or more branched chain amino acid catabolism enzyme(s) and/or transporters described herein.

[0383] In some embodiments, the genetically engineered bacteria comprise one or more gene sequence(s) whose expression is controlled by a temperature sensitive mechanism. Thermoregulators are advantageous because of strong transcriptional control without the use of external chemicals or specialized media (see, *e.g.*, Nemani *et al.*, Magnetic nanoparticle hyperthermia induced cytosine deaminase expression in microencapsulated *E. coli* for enzyme-prodrug therapy; J Biotechnol. 2015 Jun 10; 203: 32–40, and references therein). Thermoregulated protein expression using the mutant cI857 repressor and the pL and/or pR phage λ promoters have been used to engineer recombinant bacterial strains. The gene of interest cloned downstream of the λ promoters can then be efficiently regulated by the mutant thermolabile cI857 repressor of bacteriophage λ . At temperatures below 37 °C, cI857 binds to the oL or regions of the pR promoter and blocks transcription by RNA polymerase. At higher temperatures, the functional cI857 dimer is destabilized, binding to the oL or oR DNA sequences is abrogated, and mRNA transcription is initiated. Inducible expression from the thermoregulated promoter can be controlled or further fine-tuned through the optimization of the ribosome binding site (RBS), as described herein.

[0384] In one embodiment, expression of one or more protein(s) of interest is driven directly or indirectly by one or more thermoregulated promoter(s). In one embodiment, expression of the optimized enzyme(s), *e.g.*, *kivD*, *leuD*H, *adh2*, and *brnQ*, *e.g.*, as described herein, is driven directly or indirectly by one or more thermoregulated inducible promoter(s). In one embodiment, the genetically engineered bacteria encode one or more branched chain amino acid transporter(s), *e.g.*, *livKHMGF* and/or *brnQ*, described herein, whose expression is driven directly or indirectly by one or more thermoregulated inducible promoter(s). In one embodiment, the genetically engineered bacteria encode one or more branched chain amino acid exporter(s) described herein, whose expression is driven directly or indirectly by one or more thermoregulated inducible promoter(s).

[0385] In some embodiments, the thermoregulated promoter is useful for or induced during *in vivo* expression of the one or more protein(s) of interest. In some embodiments, expression of one or more branched chain amino acid catabolism enzyme(s) and/or branched chain amino acid transporter(s) and/or other protein(s) of interest is driven directly or indirectly by one or more thermoregulated promoter(s) *in vivo*.

[0386] In some embodiments, expression of one or more protein(s) of interest is driven directly or indirectly by one or more thermoregulated promoter(s) during *in vitro* growth, preparation, or manufacturing of the strain prior to *in vivo* administration. In some embodiments, it may be advantageous to shut off production of the one or more branched chain amino acid catabolism enzyme(s) and/or branched chain amino acid transporter(s). This can be done in a thermoregulated system by growing the strain at lower temperatures, *e.g.*, 30°C. Expression can then be induced by elevating the temperature to 37°C and/or 42°C. In some embodiments, the thermoregulated promoter(s) are induced in culture, *e.g.*, grown in a flask, fermenter or other appropriate culture vessel, *e.g.*, used during cell growth, cell expansion, fermentation, recovery, purification, formulation, and/or manufacture. In some embodiments, the cultures, which are induced by temperatures between 37°C and 42°C, are grown aerobically. In some embodiments, the cultures, which are induced by temperatures between 37°C and 42°C, are grown anaerobically.

[0387] In some embodiments, bacterial cell comprises two or more distinct branched chain amino acid catabolism cassette(s) or branched chain amino acid transporter(s) and/or cassette(s) for the expression of other protein(s) of interest, one or more of which are induced by temperature. In some embodiments, the genetically engineered bacteria comprise multiple copies of the same branched chain amino acid catabolism enzyme gene sequence(s) and/or transporter gene sequence(s) and/or gene sequence(s) for the expression of other proteins of interest, *e.g.*, as described herein, which are induced by temperature. In some embodiments, the genetically engineered bacteria comprise multiple copies of different branched chain amino acid catabolism enzyme genes sequence(s) and/or transporter gene sequence(s) or other gene sequence(s) of interest, *e.g.*, as described herein, one or more of which are induced by temperature.

[0388] In a first example, the temperature inducible promoter drives the expression of a construct comprising one or more polypeptides of interest described herein jointly with a second promoter, *e.g.*, a second constitutive or inducible promoter. In some embodiments, two promoters are positioned proximally to the construct and drive its expression, wherein the temperature inducible promoter drives expression under a first set of exogenous conditions, and the second promoter drives the expression under a second set of exogenous conditions. In second example, the temperature promoter drives the expression of one or more gene cassette(s) under a first inducing condition and another inducible promoter drives the expression of one or more of the same or different gene cassette(s) expressing one or more polypeptides of interest, under a second inducing condition. In both examples, the first and second conditions can be two sequential inducing culture conditions (*i.e.*, during preparation of the culture in a flask, fermenter or other appropriate culture vessel, *e.g.*, temperature regulation and IPTG). In another non-limiting example, the first inducing conditions are culture conditions, *e.g.*, the permissive temperature, and the second inducing conditions are *in vivo* conditions. Such *in vivo* conditions include low-oxygen, microaerobic, or anaerobic conditions, presence of gut metabolites, and/or nutritional and/or chemical inducers and/or metabolites administered in combination with the bacterial strain. In some embodiments, the one or more temperature regulated promoters drive expression of one or more protein(s) of interest, in combination with the FNR promoter driving the expression of the same gene sequence(s).

[0389] In some embodiments, the gene sequence(s) encoding the branched chain amino acid catabolism enzyme(s) or other polypeptide(s) of interest, are present on a plasmid and operably linked to a promoter that is induced by temperature. In some embodiments, the gene sequence(s) encoding the branched chain amino acid catabolism enzyme(s) or branched chain amino acid transporter(s) is present in the chromosome and operably linked to a promoter that is induced by temperature.

[0390] In some embodiments, the thermoregulated construct further comprises a gene encoding mutant cI857 repressor, which is divergently transcribed from the same promoter as the one or more one or more branched chain amino acid catabolism enzyme(s) and/or transporters described herein. In some embodiments, the genetically engineered bacteria comprise one or more gene sequence(s) which are indirectly inducible through a system driven by the PssB promoter. The Pssb promoter is active under aerobic conditions, and shuts off under anaerobic conditions.

[0391] This promoter can be used to express a gene of interest under aerobic conditions. This promoter can also be used to tightly control the expression of a gene product such that it is only expressed under anaerobic conditions. In this case, the oxygen induced PssB promoter induces the expression of a repressor, which represses the expression of a gene of interest. As a result, the gene of interest is only expressed in the absence of the repressor, *i.e.*, under anaerobic conditions. This strategy has the advantage of an additional level of control for improved fine-tuning and tighter control.

[0392] In one embodiment, expression of the optimized enzyme(s), *e.g.*, *kivD*, *leuD*H, *adh2*, and *brnQ*, *e.g.*, as described herein, is driven directly or indirectly by one or more PssB promoter(s). In one embodiment, the genetically engineered bacteria encode one or more branched chain amino acid transporter(s), *e.g.*, *livKHMGF* and/or *brnQ*, described herein, whose expression is driven directly or indirectly by one or more PssB promoter(s). In one embodiment, the genetically engineered bacteria encode one or more branched chain amino acid exporter(s), described herein, whose expression is driven directly or indirectly by one or more PssB promoter(s).

[0393] In some embodiments, the PssB promoter is useful for or induced during *in vivo* expression of the one or more protein(s) of interest. In some embodiments, expression of one or more branched chain amino acid catabolism enzyme(s) and/or branched chain amino acid transporter(s) and/or other protein(s) of interest is driven directly or indirectly by one or more PssB promoter(s) *in vivo*. In some embodiments, the promoter is directly or indirectly induced by a molecule (*e.g.*, arabinose) that is co-administered with the genetically engineered bacteria of the invention.

[0394] In some embodiments, expression of one or more branched chain amino acid catabolism enzyme(s) and/or branched chain amino acid transporter(s) and/or other protein(s) of interest, is driven directly or indirectly by one or more PssB promoter(s) during *in vitro* growth, preparation, or manufacturing of the strain prior to *in vivo* administration. In some embodiments, the PssB promoter(s) are induced in culture, *e.g.*, grown in a flask, fermenter or other appropriate culture vessel, *e.g.*, used during cell growth, cell expansion, fermentation, recovery, purification, formulation, and/or manufacture. In some embodiments, the promoter is directly or indirectly induced by a molecule, *e.g.*, arabinose, that is added to in the bacterial culture to induce expression and pre-load the bacterium with branched chain amino acid catabolism enzyme(s) prior to administration. In some embodiments, the cultures, which are induced by arabinose, are grown aerobically. In some embodiments, the cultures, which are induced by arabinose, are grown anaerobically.

[0395] . In some embodiments, bacterial cell comprises two or more distinct branched chain amino acid catabolism cassette(s) or other polypeptide(s) of interest, one or more of which are induced by arabinose. In some embodiments, the genetically engineered bacteria comprise multiple copies of the same branched chain amino acid catabolism enzyme gene sequence(s) and/or transporter gene sequence(s) and/or other gene sequence(s) of interest, *e.g.*, as described herein, which are induced by one or more nutritional and/or chemical inducer(s) and/or metabolite(s). In some embodiments, the genetically engineered bacteria comprise multiple copies of different branched chain amino acid catabolism enzyme genes sequence(s) and/or transporter gene sequence(s), *e.g.*, as described herein, one or more of which are induced by one or more nutritional and/or chemical inducer(s) and/or metabolite(s).

[0396] In a first example, the PssB promoter drives the expression of a construct comprising one or more polypeptides of interest described herein jointly with a second promoter, *e.g.*, a second constitutive or inducible promoter. In some embodiments, two promoters are positioned proximally to

the construct and drive its expression, wherein the PssB promoter drives expression under a first set of exogenous conditions, and the second promoter drives the expression under a second set of exogenous conditions. In second example, the PssB promoter drives the expression of one or more gene cassette(s) under a first inducing condition and another inducible promoter drives the expression of one or more of the same or different gene cassette(s) expressing one or more polypeptides of interest, under a second inducing condition. In both examples, the first and second conditions can be two sequential inducing culture conditions (*i.e.*, during preparation of the culture in a flask, fermenter or other appropriate culture vessel, *e.g.*, PssB and IPTG). In another non-limiting example, the first inducing conditions are culture conditions, *e.g.*, the presence of arabinose, and the second inducing conditions are *in vivo* conditions. Such *in vivo* conditions include low-oxygen, microaerobic, or anaerobic conditions, presence of gut metabolites, and/or nutritional and/or chemical inducers and/or metabolites administered in combination with the bacterial strain. In some embodiments, the one or more PssB promoters drive expression of one or more protein(s) of interest, in combination with the FNR promoter driving the expression of the same gene sequence(s).

[0397] In some embodiments, the gene sequence(s) encoding the branched chain amino acid catabolism enzyme(s) or other polypeptide(s) of interest, are present on a plasmid and operably linked to a promoter that is induced by arabinose. In some embodiments, the gene sequence(s) encoding the branched chain amino acid catabolism enzyme(s) or branched chain amino acid transporter(s) is present in the chromosome and operably linked to a promoter that is induced by arabinose.

[0398] In another non-limiting example, this strategy can be used to control expression of thyA and/or dapA, *e.g.*, to make a conditional auxotroph. The chromosomal copy of dapA or ThyA is knocked out. Under anaerobic conditions, dapA or thyA -as the case may be- are expressed, and the strain can grow in the absence of dap or thymidine. Under aerobic conditions, dapA or thyA expression is shut off, and the strain cannot grow in the absence of dap or thymidine. Such a strategy can, for example be employed to allow survival of bacteria under anaerobic conditions, *e.g.*, the gut, but prevent survival under aerobic conditions (biosafety switch).

Essential Genes and Auxotrophs

[0399] As used herein, the term “essential gene” refers to a gene which is necessary for cell growth and/or survival. Bacterial essential genes are well known to one of ordinary skill in the art, and can be identified by directed deletion of genes and/or random mutagenesis and screening (see, for example, Zhang and Lin, 2009, DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes, Nucl. Acids Res., 37:D455-D458 and Gerdes *et al.*, Essential genes on metabolic maps, Curr. Opin. Biotechnol., 17(5):448-456, the entire contents of each of which are expressly incorporated herein by reference).

[0400] An “essential gene” may be dependent on the circumstances and environment in which an organism lives. For example, a mutation of, modification of, or excision of an essential gene may result in the recombinant bacteria of the disclosure becoming an auxotroph. An

auxotrophic modification is intended to cause bacteria to die in the absence of an exogenously added nutrient essential for survival or growth because they lack the gene(s) necessary to produce that essential nutrient.

[0401] An auxotrophic modification is intended to cause bacteria to die in the absence of an exogenously added nutrient essential for survival or growth because they lack the gene(s) necessary to produce that essential nutrient. In some embodiments, any of the genetically engineered bacteria described herein also comprise a deletion or mutation in one or more gene(s) required for cell survival and/or growth.

[0402] In some embodiments, the bacterial cell comprises a genetic mutation in one or more endogenous gene(s) encoding a branched chain amino acid biosynthesis gene, wherein the genetic mutation reduces biosynthesis of one or more branched chain amino acids in the bacterial cell. In some embodiments, the endogenous gene encoding a branched chain amino acid biosynthesis gene is a keto acid reductoisomerase gene. Keto acid reductoisomerase gene is required for branched chain amino acid synthesis. Knock-out of this gene creates an auxotroph and requires the cell to import leucine to survive. In some embodiments, the bacterial cell comprises a genetic mutation in *ilvC* gene.

[0403] In one embodiment, the essential gene is an oligonucleotide synthesis gene, for example, *thyA*. In another embodiment, the essential gene is a cell wall synthesis gene, for example, *dapA*. In yet another embodiment, the essential gene is an amino acid gene, for example, *serA* or *MetA*. Any gene required for cell survival and/or growth may be targeted, including but not limited to, *cysE*, *glnA*, *ilvD*, *leuB*, *lysA*, *serA*, *metA*, *glyA*, *hisB*, *ilvA*, *pheA*, *proA*, *thrC*, *trpC*, *tyrA*, *thyA*, *uraA*, *dapA*, *dapB*, *dapD*, *dapE*, *dapF*, *flhD*, *metB*, *metC*, *proAB*, and *thiI*, as long as the corresponding wild-type gene product is not produced in the bacteria.

[0404] Table 6 lists depicts exemplary bacterial genes which may be disrupted or deleted to produce an auxotrophic strain. These include, but are not limited to, genes required for oligonucleotide synthesis, amino acid synthesis, and cell wall synthesis.

Table 6. Non-limiting Examples of Bacterial Genes Useful for Generation of an Auxotroph

Amino Acid	Oligonucleotide	Cell Wall
<i>cysE</i>	<i>thyA</i>	<i>dapA</i>
<i>glnA</i>	<i>uraA</i>	<i>dapB</i>
<i>ilvD</i>		<i>dapD</i>
<i>leuB</i>		<i>dapE</i>
<i>lysA</i>		<i>dapF</i>
<i>serA</i>		
<i>metA</i>		
<i>glyA</i>		
<i>hisB</i>		
<i>ilvA</i>		

pheA		
proA		
thrC		
trpC		
tyrA		

[0405] **Table 7** shows the survival of various amino acid auxotrophs in the mouse gut, as detected 24 hrs and 48 hrs post-gavage. These auxotrophs were generated using BW25113, a non-Nissle strain of *E. coli*.

Table 7. Survival of amino acid auxotrophs in the mouse gut

Gene	AA Auxotroph	Pre-Gavage	24 hours	48 hours
argA	Arginine	Present	Present	Absent
cysE	Cysteine	Present	Present	Absent
glnA	Glutamine	Present	Present	Absent
glyA	Glycine	Present	Present	Absent
hisB	Histidine	Present	Present	Present
ilvA	Isoleucine	Present	Present	Absent
leuB	Leucine	Present	Present	Absent
lysA	Lysine	Present	Present	Absent
metA	Methionine	Present	Present	Present
pheA	Phenylalanine	Present	Present	Present
proA	Proline	Present	Present	Absent
serA	Serine	Present	Present	Present
thrC	Threonine	Present	Present	Present
trpC	Tryptophan	Present	Present	Present
tyrA	Tyrosine	Present	Present	Present
ilvD	Valine/Isoleucine/ Leucine	Present	Present	Absent
thyA	Thiamine	Present	Absent	Absent
uraA	Uracil	Present	Absent	Absent
flhD	FlhD	Present	Present	Present

[0406] For example, thymine is a nucleic acid that is required for bacterial cell growth; in its absence, bacteria undergo cell death. The *thyA* gene encodes thymidylate synthetase, an enzyme that catalyzes the first step in thymine synthesis by converting dUMP to dTMP (Sat *et al.*, 2003). In some embodiments, the bacterial cell of the disclosure is a *thyA* auxotroph in which the *thyA* gene is deleted and/or replaced with an unrelated gene. A *thyA* auxotroph can grow only when sufficient amounts of thymine are present, *e.g.*, by adding thymine to growth media *in vitro*, or in the presence of high thymine levels found naturally in the human gut *in vivo*. In some embodiments, the bacterial cell of the disclosure is auxotrophic in a gene that is complemented when the bacterium is present in the

mammalian gut. Without sufficient amounts of thymine, the *thyA* auxotroph dies. In some embodiments, the auxotrophic modification is used to ensure that the bacterial cell does not survive in the absence of the auxotrophic gene product (*e.g.*, outside of the gut).

[0407] Diaminopimelic acid (DAP) is an amino acid synthesized within the lysine biosynthetic pathway and is required for bacterial cell wall growth (Meadow *et al.*, 1959; Clarkson *et al.*, 1971). In some embodiments, any of the genetically engineered bacteria described herein is a *dapD* auxotroph in which *dapD* is deleted and/or replaced with an unrelated gene. A *dapD* auxotroph can grow only when sufficient amounts of DAP are present, *e.g.*, by adding DAP to growth media *in vitro*. Without sufficient amounts of DAP, the *dapD* auxotroph dies. In some embodiments, the auxotrophic modification is used to ensure that the bacterial cell does not survive in the absence of the auxotrophic gene product (*e.g.*, outside of the gut).

[0408] In other embodiments, the genetically engineered bacterium of the present disclosure is a *uraA* auxotroph in which *uraA* is deleted and/or replaced with an unrelated gene. The *uraA* gene codes for UraA, a membrane-bound transporter that facilitates the uptake and subsequent metabolism of the pyrimidine uracil (Andersen *et al.*, 1995). A *uraA* auxotroph can grow only when sufficient amounts of uracil are present, *e.g.*, by adding uracil to growth media *in vitro*. Without sufficient amounts of uracil, the *uraA* auxotroph dies. In some embodiments, auxotrophic modifications are used to ensure that the bacteria do not survive in the absence of the auxotrophic gene product (*e.g.*, outside of the gut).

[0409] In complex communities, it is possible for bacteria to share DNA. In very rare circumstances, an auxotrophic bacterial strain may receive DNA from a non-auxotrophic strain, which repairs the genomic deletion and permanently rescues the auxotroph. Therefore, engineering a bacterial strain with more than one auxotroph may greatly decrease the probability that DNA transfer will occur enough times to rescue the auxotrophy. In some embodiments, the genetically engineered bacteria comprise a deletion or mutation in two or more genes required for cell survival and/or growth.

[0410] Other examples of essential genes include, but are not limited to *yhbV*, *yagG*, *hemB*, *secD*, *secF*, *ribD*, *ribE*, *thiL*, *dxs*, *ispA*, *dnaX*, *adk*, *hemH*, *lpxH*, *cysS*, *fold*, *rplT*, *infC*, *thrS*, *nadE*, *gapA*, *yeaZ*, *aspS*, *argS*, *pgsA*, *yefM*, *metG*, *folE*, *yefM*, *gyrA*, *nrdA*, *nrdB*, *folC*, *accD*, *fabB*, *gltX*, *ligA*, *zipA*, *dapE*, *dapA*, *der*, *hisS*, *ispG*, *suhB*, *tadA*, *acpS*, *era*, *rnc*, *ftsB*, *eno*, *pyrG*, *chpR*, *lgt*, *fbaA*, *pgk*, *yqgD*, *metK*, *yqgF*, *plsC*, *ygiT*, *pare*, *ribB*, *cca*, *ygjD*, *tdcF*, *yraL*, *yihA*, *ftsN*, *murI*, *murB*, *birA*, *secE*, *nusG*, *rplJ*, *rplL*, *rpoB*, *rpoC*, *ubiA*, *plsB*, *lexA*, *dnaB*, *ssb*, *alsK*, *groS*, *psd*, *orn*, *yjeE*, *rpsR*, *chpS*, *ppa*, *valS*, *yjgP*, *yjgQ*, *dnaC*, *ribF*, *lspA*, *ispH*, *dapB*, *folA*, *imp*, *yabQ*, *ftsL*, *ftsI*, *murE*, *murF*, *mraY*, *murD*, *ftsW*, *murG*, *murC*, *ftsQ*, *ftsA*, *ftsZ*, *lpxC*, *secM*, *secA*, *can*, *folK*, *hemL*, *yadR*, *dapD*, *map*, *rpsB*, *infB*, *nusA*, *ftsH*, *obgE*, *rpmA*, *rplU*, *ispB*, *murA*, *yrbB*, *yrbK*, *yhbN*, *rpsI*, *rplM*, *degS*, *mreD*, *mreC*, *mreB*, *accB*, *accC*, *yrnC*, *def*, *fmt*, *rplQ*, *rpoA*, *rpsD*, *rpsK*, *rpsM*, *entD*, *mrdB*, *mrdA*, *nadD*, *hlepB*, *rpoE*, *pssA*, *yfiO*, *rplS*, *trmD*, *rpsP*, *ffh*, *grpE*, *yfiB*, *csrA*, *ispF*, *ispD*, *rplW*, *rplD*, *rplC*, *rpsJ*, *fusA*, *rpsG*,

rpsL, trpS, yrfF, asd, rpoH, ftsX, ftsE, ftsY, frr, dxr, ispU, rfaK, kdtA, coaD, rpmB, dfp, dut, gmk, spot, gyrB, dnaN, dnaA, rpmH, rnpA, yidC, tnaB, glmS, glmU, wzyE, hemD, hemC, yigP, ubiB, ubiD, hemG, secY, rplO, rpmD, rpsE, rplR, rplF, rpsH, rpsN, rplE, rplX, rplN, rpsQ, rpmC, rplP, rpsC, rplV, rpsS, rplB, cdsA, yaeL, yaeT, lpxD, fabZ, lpxA, lpxB, dnaE, accA, tilS, proS, yafF, tsf, pyrH, ola, rlpB, leuS, lnt, glnS, fldA, cydA, infA, cydC, ftsK, lolA, serS, rpsA, msbA, lpxK, kdsB, mukF, mukE, mukB, asnS, fabA, mviN, rne, yceQ, fabD, fabG, acpP, tmk, holB, lolC, lolD, lolE, purB, ymfK, minE, mind, pth, rsA, ispE, lolB, hemA, prfA, prmC, kdsA, topA, ribA, fabI, racR, dicA, ydfB, tyrS, ribC, ydiL, pheT, pheS, yhhQ, bcsB, glyQ, yibJ, and gpsA. Other essential genes are known to those of ordinary skill in the art.

[0411] In some embodiments, the genetically engineered bacterium of the present disclosure is a synthetic ligand-dependent essential gene (SLiDE) bacterial cell. SLiDE bacterial cells are synthetic auxotrophs with a mutation in one or more essential genes that only grow in the presence of a particular ligand (see Lopez and Anderson “Synthetic Auxotrophs with Ligand-Dependent Essential Genes for a BL21 (DE3 Biosafety Strain, ”ACS Synthetic Biology (2015) DOI: 10.1021/acssynbio.5b00085, the entire contents of which are expressly incorporated herein by reference).

[0412] In some embodiments, the SLiDE bacterial cell comprises a mutation in an essential gene. In some embodiments, the essential gene is selected from the group consisting of *pheS*, *dnaN*, *tyrS*, *metG* and *adk*. In some embodiments, the essential gene is *dnaN* comprising one or more of the following mutations: H191N, R240C, I317S, F319V, L340T, V347I, and S345C. In some embodiments, the essential gene is *dnaN* comprising the mutations H191N, R240C, I317S, F319V, L340T, V347I, and S345C. In some embodiments, the essential gene is *pheS* comprising one or more of the following mutations: F125G, P183T, P184A, R186A, and I188L. In some embodiments, the essential gene is *pheS* comprising the mutations F125G, P183T, P184A, R186A, and I188L. In some embodiments, the essential gene is *tyrS* comprising one or more of the following mutations: L36V, C38A and F40G. In some embodiments, the essential gene is *tyrS* comprising the mutations L36V, C38A and F40G. In some embodiments, the essential gene is *metG* comprising one or more of the following mutations: E45Q, N47R, I49G, and A51C. In some embodiments, the essential gene is *metG* comprising the mutations E45Q, N47R, I49G, and A51C. In some embodiments, the essential gene is *adk* comprising one or more of the following mutations: I4L, L5I and L6G. In some embodiments, the essential gene is *adk* comprising the mutations I4L, L5I and L6G.

[0413] In some embodiments, the genetically engineered bacterium is complemented by a ligand. In some embodiments, the ligand is selected from the group consisting of benzothiazole, indole, 2-aminobenzothiazole, indole-3-butyric acid, indole-3-acetic acid, and L-histidine methyl ester. For example, bacterial cells comprising mutations in *metG* (E45Q, N47R, I49G, and A51C) are complemented by benzothiazole, indole, 2-aminobenzothiazole, indole-3-butyric acid, indole-3-acetic acid or L-histidine methyl ester. Bacterial cells comprising mutations in *dnaN* (H191N, R240C,

I317S, F319V, L340T, V347I, and S345C) are complemented by benzothiazole, indole or 2-aminobenzothiazole. Bacterial cells comprising mutations in *pheS* (F125G, P183T, P184A, R186A, and I188L) are complemented by benzothiazole or 2-aminobenzothiazole. Bacterial cells comprising mutations in *tyrS* (L36V, C38A, and F40G) are complemented by benzothiazole or 2-aminobenzothiazole. Bacterial cells comprising mutations in *adk* (I4L, L5I and L6G) are complemented by benzothiazole or indole.

[0414] In some embodiments, the genetically engineered bacterium comprises more than one mutant essential gene that renders it auxotrophic to a ligand. In some embodiments, the bacterial cell comprises mutations in two essential genes. For example, in some embodiments, the bacterial cell comprises mutations in *tyrS* (L36V, C38A, and F40G) and *metG* (E45Q, N47R, I49G, and A51C). In other embodiments, the bacterial cell comprises mutations in three essential genes. For example, in some embodiments, the bacterial cell comprises mutations in *tyrS* (L36V, C38A, and F40G), *metG* (E45Q, N47R, I49G, and A51C), and *pheS* (F125G, P183T, P184A, R186A, and I188L).

[0415] In some embodiments, the genetically engineered bacterium is a conditional auxotroph whose essential gene(s) is replaced using the arabinose system described herein.

Genetic Regulatory Circuits

[0416] In some embodiments, the genetically engineered bacteria comprise multi-layered genetic regulatory circuits for expressing the constructs described herein (*see, e.g.*, U.S. Provisional Application No. 62/184,811, incorporated herein by reference in its entirety). The genetic regulatory circuits are useful to screen for mutant bacteria that produce a branched chain amino acid catabolism enzyme, BCAA transporter, and/or BCAA binding protein or rescue an auxotroph. In certain embodiments, the invention provides methods for selecting genetically engineered bacteria that produce one or more genes of interest.

[0417] In some embodiments, the invention provides genetically engineered bacteria comprising a gene or gene cassette for producing a payload and a T7 polymerase-regulated genetic regulatory circuit. For example, the genetically engineered bacteria comprise a first gene encoding a T7 polymerase, wherein the first gene is operably linked to a fumarate and nitrate reductase regulator (FNR)-responsive promoter; a second gene or gene cassette for producing a payload, wherein the second gene or gene cassette is operably linked to a T7 promoter that is induced by the T7 polymerase; and a third gene encoding an inhibitory factor, *lysY*, that is capable of inhibiting the T7 polymerase. In the presence of oxygen, FNR does not bind the FNR-responsive promoter, and the payload is not expressed. *LysY* is expressed constitutively (P-lac constitutive) and further inhibits T7 polymerase. In the absence of oxygen, FNR dimerizes and binds to the FNR-responsive promoter, T7 polymerase is expressed at a level sufficient to overcome *lysY* inhibition, and the payload is expressed. In some embodiments, the *lysY* gene is operably linked to an additional FNR binding site. In the absence of oxygen, FNR dimerizes to activate T7 polymerase expression as described above, and also inhibits *lysY* expression.

[0418] In some embodiments, the invention provides genetically engineered bacteria comprising a gene or gene cassette for producing a payload and a protease-regulated genetic regulatory circuit. For example, the genetically engineered bacteria comprise a first gene encoding an mf-lon protease, wherein the first gene is operably linked to a FNR-responsive promoter; a second gene or gene cassette for producing a payload operably linked to a tet regulatory region (tetO); and a third gene encoding an mf-lon degradation signal linked to a tet repressor (tetR), wherein the tetR is capable of binding to the tet regulatory region and repressing expression of the second gene or gene cassette. The mf-lon protease is capable of recognizing the mf-lon degradation signal and degrading the tetR. In the presence of oxygen, FNR does not bind the FNR-responsive promoter, the repressor is not degraded, and the payload is not expressed. In the absence of oxygen, FNR dimerizes and binds the FNR-responsive promoter, thereby inducing expression of mf-lon protease. The mf-lon protease recognizes the mf-lon degradation signal and degrades the tetR, and the payload is expressed.

[0419] In some embodiments, the invention provides genetically engineered bacteria comprising a gene or gene cassette for producing a payload and a repressor-regulated genetic regulatory circuit. For example, the genetically engineered bacteria comprise a first gene encoding a first repressor, wherein the first gene is operably linked to a FNR-responsive promoter; a second gene or gene cassette for producing a payload operably linked to a first regulatory region comprising a constitutive promoter; and a third gene encoding a second repressor, wherein the second repressor is capable of binding to the first regulatory region and repressing expression of the second gene or gene cassette. The third gene is operably linked to a second regulatory region comprising a constitutive promoter, wherein the first repressor is capable of binding to the second regulatory region and inhibiting expression of the second repressor. In the presence of oxygen, FNR does not bind the FNR-responsive promoter, the first repressor is not expressed, the second repressor is expressed, and the payload is not expressed. In the absence of oxygen, FNR dimerizes and binds the FNR-responsive promoter, the first repressor is expressed, the second repressor is not expressed, and the payload is expressed.

[0420] Examples of repressors useful in these embodiments include, but are not limited to, ArgR, TetR, ArsR, AscG, LacI, CscR, DeoR, DgoR, FruR, GalR, GatR, CI, LexA, RafR, QacR, and PtxS (US20030166191).

[0421] In some embodiments, the invention provides genetically engineered bacteria comprising a gene or gene cassette for producing a payload and a regulatory RNA-regulated genetic regulatory circuit. For example, the genetically engineered bacteria comprise a first gene encoding a regulatory RNA, wherein the first gene is operably linked to a FNR-responsive promoter, and a second gene or gene cassette for producing a payload. The second gene or gene cassette is operably linked to a constitutive promoter and further linked to a nucleotide sequence capable of producing an mRNA hairpin that inhibits translation of the payload. The regulatory RNA is capable of eliminating the mRNA hairpin and inducing payload translation via the ribosomal binding site. In the presence of

oxygen, FNR does not bind the FNR-responsive promoter, the regulatory RNA is not expressed, and the mRNA hairpin prevents the payload from being translated. In the absence of oxygen, FNR dimerizes and binds the FNR-responsive promoter, the regulatory RNA is expressed, the mRNA hairpin is eliminated, and the payload is expressed.

[0422] In some embodiments, the invention provides genetically engineered bacteria comprising a gene or gene cassette for producing a payload and a CRISPR-regulated genetic regulatory circuit. For example, the genetically engineered bacteria comprise a Cas9 protein; a first gene encoding a CRISPR guide RNA, wherein the first gene is operably linked to a FNR-responsive promoter; a second gene or gene cassette for producing a payload, wherein the second gene or gene cassette is operably linked to a regulatory region comprising a constitutive promoter; and a third gene encoding a repressor operably linked to a constitutive promoter, wherein the repressor is capable of binding to the regulatory region and repressing expression of the second gene or gene cassette. The third gene is further linked to a CRISPR target sequence that is capable of binding to the CRISPR guide RNA, wherein said binding to the CRISPR guide RNA induces cleavage by the Cas9 protein and inhibits expression of the repressor. In the presence of oxygen, FNR does not bind the FNR-responsive promoter, the guide RNA is not expressed, the repressor is expressed, and the payload is not expressed. In the absence of oxygen, FNR dimerizes and binds the FNR-responsive promoter, the guide RNA is expressed, the repressor is not expressed, and the payload is expressed.

[0423] In some embodiments, the invention provides genetically engineered bacteria comprising a gene or gene cassette for producing a payload and a recombinase-regulated genetic regulatory circuit. For example, the genetically engineered bacteria comprise a first gene encoding a recombinase, wherein the first gene is operably linked to a FNR-responsive promoter, and a second gene or gene cassette for producing a payload operably linked to a constitutive promoter. The second gene or gene cassette is inverted in orientation (3' to 5') and flanked by recombinase binding sites, and the recombinase is capable of binding to the recombinase binding sites to induce expression of the second gene or gene cassette by reverting its orientation (5' to 3'). In the presence of oxygen, FNR does not bind the FNR-responsive promoter, the recombinase is not expressed, the payload remains in the 3' to 5' orientation, and no functional payload is produced. In the absence of oxygen, FNR dimerizes and binds the FNR-responsive promoter, the recombinase is expressed, the payload is reverted to the 5' to 3' orientation, and functional payload is produced.

[0424] In some embodiments, the invention provides genetically engineered bacteria comprising a gene or gene cassette for producing a payload and a polymerase- and recombinase-regulated genetic regulatory circuit. For example, the genetically engineered bacteria comprise a first gene encoding a recombinase, wherein the first gene is operably linked to a FNR-responsive promoter; a second gene or gene cassette for producing a payload operably linked to a T7 promoter; a third gene encoding a T7 polymerase, wherein the T7 polymerase is capable of binding to the T7 promoter and inducing expression of the payload. The third gene encoding the T7 polymerase is

inverted in orientation (3' to 5') and flanked by recombinase binding sites, and the recombinase is capable of binding to the recombinase binding sites to induce expression of the T7 polymerase gene by reverting its orientation (5' to 3'). In the presence of oxygen, FNR does not bind the FNR-responsive promoter, the recombinase is not expressed, the T7 polymerase gene remains in the 3' to 5' orientation, and the payload is not expressed. In the absence of oxygen, FNR dimerizes and binds the FNR-responsive promoter, the recombinase is expressed, the T7 polymerase gene is reverted to the 5' to 3' orientation, and the payload is expressed.

Isolated Plasmids

[0425] In other embodiments, the disclosure provides an isolated plasmid comprising a first nucleic acid encoding a first payload operably linked to a first inducible promoter, and a second nucleic acid encoding a second payload operably linked to a second inducible promoter. In other embodiments, the disclosure provides an isolated plasmid further comprising a third nucleic acid encoding a third payload operably linked to a third inducible promoter. In other embodiments, the disclosure provides a plasmid comprising four, five, six, or more nucleic acids encoding four, five, six, or more payloads operably linked to inducible promoters. In any of the embodiments described here, the first, second, third, fourth, fifth, sixth, etc “payload(s)” can be a branched chain amino acid catabolism enzyme, a transporter of branched chain amino acids, a binding protein of branched chain amino acids, or other sequence described herein. In one embodiment, the nucleic acid encoding the first payload and the nucleic acid encoding the second payload are operably linked to the first inducible promoter. In one embodiment, the nucleic acid encoding the first payload is operably linked to a first inducible promoter and the nucleic acid encoding the second payload is operably linked to a second inducible promoter. In one embodiment, the first inducible promoter and the second inducible promoter are separate copies of the same inducible promoter. In another embodiment, the first inducible promoter and the second inducible promoter are different inducible promoters. In other embodiments comprising a third nucleic acid, the nucleic acid encoding the third payload and the nucleic acid encoding the first and second payloads are all operably linked to the same inducible promoter. In other embodiments, the nucleic acid encoding the first payload is operably linked to a first inducible promoter, the nucleic acid encoding the second payload is operably linked to a second inducible promoter, and the nucleic acid encoding the third payload is operably linked to a third inducible promoter. In some embodiments, the first, second, and third inducible promoters are separate copies of the same inducible promoter. In other embodiments, the first inducible promoter, the second inducible promoter, and the third inducible promoter are different inducible promoters. In some embodiments, the first promoter, the second promoter, and the optional third promoter, or the first promoter and the second promoter and the optional third promoter, are each directly or indirectly induced by low-oxygen or anaerobic conditions. In other embodiments, the first promoter, the second promoter, and the optional third promoter, or the first promoter and the second promoter and the optional third promoter, are each a fumarate and nitrate reduction regulator (FNR) responsive

promoter. In other embodiments, the first promoter, the second promoter, and the optional third promoter, or the first promoter and the second promoter and the optional third promoter are each a ROS-inducible regulatory region. In other embodiments, the first promoter, the second promoter, and the optional third promoter, or the first promoter and the second promoter and the optional third promoter are each a RNS-inducible regulatory region.

[0426] In some embodiments, the heterologous gene encoding a branched chain amino acid catabolism enzyme is operably linked to a constitutive promoter. In one embodiment, the constitutive promoter is a lac promoter. In another embodiment, the constitutive promoter is a tet promoter. In another embodiment, the constitutive promoter is a constitutive *Escherichia coli* σ^{32} promoter. In another embodiment, the constitutive promoter is a constitutive *Escherichia coli* σ^{70} promoter. In another embodiment, the constitutive promoter is a constitutive *Bacillus subtilis* σ^A promoter. In another embodiment, the constitutive promoter is a constitutive *Bacillus subtilis* σ^B promoter. In another embodiment, the constitutive promoter is a *Salmonella* promoter. In other embodiments, the constitutive promoter is a bacteriophage T7 promoter. In other embodiments, the constitutive promoter is and a bacteriophage SP6 promoter. In any of the above-described embodiments, the plasmid further comprises a heterologous gene encoding a transporter of a branched chain amino acid, a BCAA binding protein, and/or a kill switch construct, which may be operably linked to a constitutive promoter or an inducible promoter.

[0427] In some embodiments, the isolated plasmid comprises at least one heterologous branched chain amino acid catabolism enzyme gene operably linked to a first inducible promoter; a heterologous gene encoding a TetR protein operably linked to a P_{araBAD} promoter, a heterologous gene encoding AraC operably linked to a P_{araC} promoter, a heterologous gene encoding an antitoxin operably linked to a constitutive promoter, and a heterologous gene encoding a toxin operably linked to a P_{TetR} promoter. In another embodiment, the isolated plasmid comprises at least one heterologous gene encoding a branched chain amino acid catabolism enzyme operably linked to a first inducible promoter; a heterologous gene encoding a TetR protein and an anti-toxin operably linked to a P_{araBAD} promoter, a heterologous gene encoding AraC operably linked to a P_{araC} promoter, and a heterologous gene encoding a toxin operably linked to a P_{TetR} promoter.

[0428] In one embodiment, the plasmid is a high-copy plasmid. In another embodiment, the plasmid is a low-copy plasmid.

[0429] In another aspect, the disclosure provides a recombinant bacterial cell comprising an isolated plasmid described herein. In another embodiment, the disclosure provides a pharmaceutical composition comprising the recombinant bacterial cell.

Integration

[0430] In some embodiments, any of the gene(s) or gene cassette(s) of the present disclosure may be integrated into the bacterial chromosome at one or more integration sites. One or more copies of the gene (for example, an amino acid catabolism gene, BCAA transporter gene, and/or BCAA

binding protein gene) or gene cassette (for example, a gene cassette comprising an amino acid catabolism gene, an amino acid transporter gene, a BCAA binding protein gene) may be integrated into the bacterial chromosome. Having multiple copies of the gene or gene cassette integrated into the chromosome allows for greater production of the payload, *e.g.*, amino acid catabolism enzyme, BCAA transporter gene, and/or BCAA binding protein gene and other enzymes of a gene cassette, and also permits fine-tuning of the level of expression. Alternatively, different circuits described herein, such as any of the kill-switch circuits, in addition to the therapeutic gene(s) or gene cassette(s) could be integrated into the bacterial chromosome at one or more different integration sites to perform multiple different functions.

Pharmaceutical Compositions and Formulations

[0431] Pharmaceutical compositions comprising the genetically engineered microorganisms of the invention may be used to treat, manage, ameliorate, and/or prevent a disorder associated with branched chain amino acid catabolism or symptom(s) associated with diseases or disorders associated with branched chain amino acid catabolism. Pharmaceutical compositions of the invention comprising one or more genetically engineered bacteria, and/or one or more genetically engineered yeast or virus, alone or in combination with prophylactic agents, therapeutic agents, and/or pharmaceutically acceptable carriers are provided.

[0432] In certain embodiments, the pharmaceutical composition comprises one species, strain, or subtype of bacteria that are engineered to comprise one or more of the genetic modifications described herein, *e.g.*, selected from expression of at least one branched chain amino acid catabolism enzyme, BCAA transporter, BCAA binding protein, auxotrophy, kill-switch, exporter knock-out, etc. In alternate embodiments, the pharmaceutical composition comprises two or more species, strains, and/or subtypes of bacteria that are each engineered to comprise the genetic modifications described herein.

[0433] The pharmaceutical compositions of the invention described herein may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries, which facilitate processing of the active ingredients into compositions for pharmaceutical use. Methods of formulating pharmaceutical compositions are known in the art (*see, e.g.*, "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA). In some embodiments, the pharmaceutical compositions are subjected to tableting, lyophilizing, direct compression, conventional mixing, dissolving, granulating, levigating, emulsifying, encapsulating, entrapping, or spray drying to form tablets, granulates, nanoparticles, nanocapsules, microcapsules, microtablets, pellets, or powders, which may be enterically coated or uncoated. Appropriate formulation depends on the route of administration.

[0434] The genetically engineered microorganisms may be formulated into pharmaceutical compositions in any suitable dosage form (*e.g.*, liquids, capsules, sachet, hard capsules, soft capsules, tablets, enteric coated tablets, suspension powders, granules, or matrix sustained release formations

for oral administration) and for any suitable type of administration (*e.g.*, oral, topical, injectable, intravenous, sub-cutaneous, immediate-release, pulsatile-release, delayed-release, or sustained release). Suitable dosage amounts for the genetically engineered bacteria may range from about 10⁴ to 10¹² bacteria. The composition may be administered once or more daily, weekly, or monthly. The composition may be administered before, during, or following a meal. In one embodiment, the pharmaceutical composition is administered before the subject eats a meal. In one embodiment, the pharmaceutical composition is administered currently with a meal. In one embodiment, the pharmaceutical composition is administered after the subject eats a meal.

[0435] The genetically engineered bacteria or genetically engineered yeast or virus may be formulated into pharmaceutical compositions comprising one or more pharmaceutically acceptable carriers, thickeners, diluents, buffers, buffering agents, surface active agents, neutral or cationic lipids, lipid complexes, liposomes, penetration enhancers, carrier compounds, and other pharmaceutically acceptable carriers or agents. For example, the pharmaceutical composition may include, but is not limited to, the addition of calcium bicarbonate, sodium bicarbonate, calcium phosphate, various sugars and types of starch, cellulose derivatives, gelatin, vegetable oils, polyethylene glycols, and surfactants, including, for example, polysorbate 20. In some embodiments, the genetically engineered bacteria of the invention may be formulated in a solution of sodium bicarbonate, *e.g.*, 1 molar solution of sodium bicarbonate (to buffer an acidic cellular environment, such as the stomach, for example). The genetically engineered bacteria may be administered and formulated as neutral or salt forms. Pharmaceutically acceptable salts include those formed with anions such as those derived from hydrochloric, phosphoric, acetic, oxalic, tartaric acids, etc., and those formed with cations such as those derived from sodium, potassium, ammonium, calcium, ferric hydroxides, isopropylamine, triethylamine, 2-ethylamino ethanol, histidine, procaine, etc.

[0436] The genetically engineered microorganisms may be administered intravenously, *e.g.*, by infusion or injection.

[0437] The genetically engineered microorganisms of the disclosure may be administered intrathecally. In some embodiments, the genetically engineered microorganisms of the invention may be administered orally. The genetically engineered microorganisms disclosed herein may be administered topically and formulated in the form of an ointment, cream, transdermal patch, lotion, gel, shampoo, spray, aerosol, solution, emulsion, or other form well known to one of skill in the art. See, *e.g.*, "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA. In an embodiment, for non-sprayable topical dosage forms, viscous to semi-solid or solid forms comprising a carrier or one or more excipients compatible with topical application and having a dynamic viscosity greater than water are employed. Suitable formulations include, but are not limited to, solutions, suspensions, emulsions, creams, ointments, powders, liniments, salves, etc., which may be sterilized or mixed with auxiliary agents (*e.g.*, preservatives, stabilizers, wetting agents, buffers, or salts) for influencing various properties, *e.g.*, osmotic pressure. Other suitable topical dosage forms include

sprayable aerosol preparations wherein the active ingredient in combination with a solid or liquid inert carrier, is packaged in a mixture with a pressurized volatile (*e.g.*, a gaseous propellant, such as freon) or in a squeeze bottle. Moisturizers or humectants can also be added to pharmaceutical compositions and dosage forms. Examples of such additional ingredients are well known in the art. In one embodiment, the pharmaceutical composition comprising the recombinant bacteria of the invention may be formulated as a hygiene product. For example, the hygiene product may be an antibacterial formulation, or a fermentation product such as a fermentation broth. Hygiene products may be, for example, shampoos, conditioners, creams, pastes, lotions, and lip balms.

[0438] The genetically engineered microorganisms disclosed herein may be administered orally and formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, etc. Pharmacological compositions for oral use can be made using a solid excipient, optionally grinding the resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries if desired, to obtain tablets or dragee cores. Suitable excipients include, but are not limited to, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose compositions such as maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carbomethylcellulose; and/or physiologically acceptable polymers such as polyvinylpyrrolidone (PVP) or polyethylene glycol (PEG). Disintegrating agents may also be added, such as cross-linked polyvinylpyrrolidone, agar, alginic acid or a salt thereof such as sodium alginate.

[0439] Tablets or capsules can be prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (*e.g.*, pregelatinised maize starch, polyvinylpyrrolidone, hydroxypropyl methylcellulose, carboxymethylcellulose, polyethylene glycol, sucrose, glucose, sorbitol, starch, gum, kaolin, and tragacanth); fillers (*e.g.*, lactose, microcrystalline cellulose, or calcium hydrogen phosphate); lubricants (*e.g.*, calcium, aluminum, zinc, stearic acid, polyethylene glycol, sodium lauryl sulfate, starch, sodium benzoate, L-leucine, magnesium stearate, talc, or silica); disintegrants (*e.g.*, starch, potato starch, sodium starch glycolate, sugars, cellulose derivatives, silica powders); or wetting agents (*e.g.*, sodium lauryl sulphate). The tablets may be coated by methods well known in the art. A coating shell may be present, and common membranes include, but are not limited to, polylactide, polyglycolic acid, polyanhydride, other biodegradable polymers, alginate-polylysine-alginate (APA), alginate-polymethylene-co-guanidine-alginate (A-PMCG-A), hydroxymethylacrylate-methyl methacrylate (HEMA-MMA), multilayered HEMA-MMA-MAA, polyacrylonitrilevinylchloride (PAN-PVC), acrylonitrile/sodium methallylsulfonate (AN-69), polyethylene glycol/poly pentamethylcyclotrisiloxane / polydimethylsiloxane (PEG/PD5/PDMS), poly N,N- dimethyl acrylamide (PDMAAm), siliceous encapsulates, cellulose sulphate/sodium alginate/polymethylene-co-guanidine (CS/A/PMCG), cellulose acetate phthalate, calcium alginate, k-carrageenan-locust bean gum gel beads, gellan-xanthan beads, poly(lactide-co-glycolides),

carrageenan, starch poly-anhydrides, starch polymethacrylates, polyamino acids, and enteric coating polymers.

[0440] In some embodiments, the genetically engineered microorganisms are enterically coated for release into the gut or a particular region of the gut, for example, the large intestine. The typical pH profile from the stomach to the colon is about 1-4 (stomach), 5.5-6 (duodenum), 7.3-8.0 (ileum), and 5.5-6.5 (colon). In some diseases, the pH profile may be modified. In some embodiments, the coating is degraded in specific pH environments in order to specify the site of release. In some embodiments, at least two coatings are used. In some embodiments, the outside coating and the inside coating are degraded at different pH levels.

[0441] Liquid preparations for oral administration may take the form of solutions, syrups, suspensions, or a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations may be prepared by conventional means with pharmaceutically acceptable agents such as suspending agents (*e.g.*, sorbitol syrup, cellulose derivatives, or hydrogenated edible fats); emulsifying agents (*e.g.*, lecithin or acacia); non-aqueous vehicles (*e.g.*, almond oil, oily esters, ethyl alcohol, or fractionated vegetable oils); and preservatives (*e.g.*, methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations may also contain buffer salts, flavoring, coloring, and sweetening agents as appropriate. Preparations for oral administration may be suitably formulated for slow release, controlled release, or sustained release of the genetically engineered microorganisms described herein.

[0442] In one embodiment, the genetically engineered microorganisms of the disclosure may be formulated in a composition suitable for administration to pediatric subjects. As is well known in the art, children differ from adults in many aspects, including different rates of gastric emptying, pH, gastrointestinal permeability, etc. (Ivanovska *et al.*, *Pediatrics*, 134(2):361-372, 2014). Moreover, pediatric formulation acceptability and preferences, such as route of administration and taste attributes, are critical for achieving acceptable pediatric compliance. Thus, in one embodiment, the composition suitable for administration to pediatric subjects may include easy-to-swallow or dissolvable dosage forms, or more palatable compositions, such as compositions with added flavors, sweeteners, or taste blockers. In one embodiment, a composition suitable for administration to pediatric subjects may also be suitable for administration to adults.

[0443] In one embodiment, the composition suitable for administration to pediatric subjects may include a solution, syrup, suspension, elixir, powder for reconstitution as suspension or solution, dispersible/effervescent tablet, chewable tablet, gummy candy, lollipop, freezer pop, troche, chewing gum, oral thin strip, orally disintegrating tablet, sachet, soft gelatin capsule, sprinkle oral powder, or granules. In one embodiment, the composition is a gummy candy, which is made from a gelatin base, giving the candy elasticity, desired chewy consistency, and longer shelf-life. In some embodiments, the gummy candy may also comprise sweeteners or flavors.

[0444] In one embodiment, the composition suitable for administration to pediatric subjects may include a flavor. As used herein, "flavor" is a substance (liquid or solid) that provides a distinct taste and aroma to the formulation. Flavors also help to improve the palatability of the formulation. Flavors include, but are not limited to, strawberry, vanilla, lemon, grape, bubble gum, and cherry.

[0445] In certain embodiments, the genetically engineered microorganisms may be orally administered, for example, with an inert diluent or an assimilable edible carrier. The compound may also be enclosed in a hard or soft shell gelatin capsule, compressed into tablets, or incorporated directly into the subject's diet. For oral therapeutic administration, the compounds may be incorporated with excipients and used in the form of ingestible tablets, buccal tablets, troches, capsules, elixirs, suspensions, syrups, wafers, and the like. To administer a compound by other than parenteral administration, it may be necessary to coat the compound with, or co-administer the compound with, a material to prevent its inactivation.

[0446] In another embodiment, the pharmaceutical composition comprising the recombinant bacteria of the invention may be a comestible product, for example, a food product. In one embodiment, the food product is milk, concentrated milk, fermented milk (yogurt, sour milk, frozen yogurt, lactic acid bacteria-fermented beverages), milk powder, ice cream, cream cheeses, dry cheeses, soybean milk, fermented soybean milk, vegetable-fruit juices, fruit juices, sports drinks, confectionery, candies, infant foods (such as infant cakes), nutritional food products, animal feeds, or dietary supplements. In one embodiment, the food product is a fermented food, such as a fermented dairy product. In one embodiment, the fermented dairy product is yogurt. In another embodiment, the fermented dairy product is cheese, milk, cream, ice cream, milk shake, or kefir. In another embodiment, the recombinant bacteria of the invention are combined in a preparation containing other live bacterial cells intended to serve as probiotics. In another embodiment, the food product is a beverage. In one embodiment, the beverage is a fruit juice-based beverage or a beverage containing plant or herbal extracts. In another embodiment, the food product is a jelly or a pudding. Other food products suitable for administration of the recombinant bacteria of the invention are well known in the art. For example, see U.S. 2015/0359894 and US 2015/0238545, the entire contents of each of which are expressly incorporated herein by reference. In yet another embodiment, the pharmaceutical composition of the invention is injected into, sprayed onto, or sprinkled onto a food product, such as bread, yogurt, or cheese.

[0447] In some embodiments, the composition is formulated for intrainestinal administration, intrajejunal administration, intraduodenal administration, intraileal administration, gastric shunt administration, or intracolonic administration, via nanoparticles, nanocapsules, microcapsules, or microtablets, which are enterically coated or uncoated. The pharmaceutical compositions may also be formulated in rectal compositions such as suppositories or retention enemas, using, *e.g.*, conventional suppository bases such as cocoa butter or other glycerides. The

compositions may be suspensions, solutions, or emulsions in oily or aqueous vehicles, and may contain suspending, stabilizing and/or dispersing agents.

[0448] The genetically engineered microorganisms described herein may be administered intranasally, formulated in an aerosol form, spray, mist, or in the form of drops, and conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant (*e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas). Pressurized aerosol dosage units may be determined by providing a valve to deliver a metered amount. Capsules and cartridges (*e.g.*, of gelatin) for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

[0449] The genetically engineered microorganisms may be administered and formulated as depot preparations. Such long acting formulations may be administered by implantation or by injection, including intravenous injection, subcutaneous injection, local injection, direct injection, or infusion. For example, the compositions may be formulated with suitable polymeric or hydrophobic materials (*e.g.*, as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives (*e.g.*, as a sparingly soluble salt).

[0450] In some embodiments, disclosed herein are pharmaceutically acceptable compositions in single dosage forms. Single dosage forms may be in a liquid or a solid form. Single dosage forms may be administered directly to a patient without modification or may be diluted or reconstituted prior to administration. In certain embodiments, a single dosage form may be administered in bolus form, *e.g.*, single injection, single oral dose, including an oral dose that comprises multiple tablets, capsule, pills, etc. In alternate embodiments, a single dosage form may be administered over a period of time, *e.g.*, by infusion.

[0451] Single dosage forms of the pharmaceutical composition may be prepared by portioning the pharmaceutical composition into smaller aliquots, single dose containers, single dose liquid forms, or single dose solid forms, such as tablets, granulates, nanoparticles, nanocapsules, microcapsules, microtablets, pellets, or powders, which may be enterically coated or uncoated. A single dose in a solid form may be reconstituted by adding liquid, typically sterile water or saline solution, prior to administration to a patient.

[0452] In other embodiments, the composition can be delivered in a controlled release or sustained release system. In one embodiment, a pump may be used to achieve controlled or sustained release. In another embodiment, polymeric materials can be used to achieve controlled or sustained release of the therapies of the present disclosure (see *e.g.*, U.S. Patent No. 5,989,463). Examples of polymers used in sustained release formulations include, but are not limited to, poly(2-hydroxy ethyl methacrylate), poly(methyl methacrylate), poly(acrylic acid), poly(ethylene-co-vinyl acetate), poly(methacrylic acid), polyglycolides (PLG), polyanhydrides, poly(N-vinyl pyrrolidone), poly(vinyl alcohol), polyacrylamide, poly(ethylene glycol), polylactides (PLA), poly(lactide-co-glycolides)

(PLGA), and polyorthoesters. The polymer used in a sustained release formulation may be inert, free of leachable impurities, stable on storage, sterile, and biodegradable. In some embodiments, a controlled or sustained release system can be placed in proximity of the prophylactic or therapeutic target, thus requiring only a fraction of the systemic dose. Any suitable technique known to one of skill in the art may be used.

[0453] Dosage regimens may be adjusted to provide a therapeutic response. Dosing can depend on several factors, including severity and responsiveness of the disease, route of administration, time course of treatment (days to months to years), and time to amelioration of the disease. For example, a single bolus may be administered at one time, several divided doses may be administered over a predetermined period of time, or the dose may be reduced or increased as indicated by the therapeutic situation. The specification for the dosage is dictated by the unique characteristics of the active compound and the particular therapeutic effect to be achieved. Dosage values may vary with the type and severity of the condition to be alleviated. For any particular subject, specific dosage regimens may be adjusted over time according to the individual need and the professional judgment of the treating clinician. Toxicity and therapeutic efficacy of compounds provided herein can be determined by standard pharmaceutical procedures in cell culture or animal models. For example, LD₅₀, ED₅₀, EC₅₀, and IC₅₀ may be determined, and the dose ratio between toxic and therapeutic effects (LD₅₀/ED₅₀) may be calculated as the therapeutic index. Compositions that exhibit toxic side effects may be used, with careful modifications to minimize potential damage to reduce side effects. Dosing may be estimated initially from cell culture assays and animal models. The data obtained from *in vitro* and *in vivo* assays and animal studies can be used in formulating a range of dosage for use in humans.

[0454] The ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water-free concentrate in a hermetically sealed container such as an ampoule or sachet indicating the quantity of active agent. If the mode of administration is by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients may be mixed prior to administration.

[0455] The pharmaceutical compositions may be packaged in a hermetically sealed container such as an ampoule or sachet indicating the quantity of the agent. In one embodiment, one or more of the pharmaceutical compositions is supplied as a dry sterilized lyophilized powder or water-free concentrate in a hermetically sealed container and can be reconstituted (*e.g.*, with water or saline) to the appropriate concentration for administration to a subject. In an embodiment, one or more of the prophylactic or therapeutic agents or pharmaceutical compositions is supplied as a dry sterile lyophilized powder in a hermetically sealed container stored between 2° C and 8° C and administered within 1 hour, within 3 hours, within 5 hours, within 6 hours, within 12 hours, within 24 hours, within 48 hours, within 72 hours, or within one week after being reconstituted. Cryoprotectants can be included for a lyophilized dosage form, principally 0-10% sucrose (optimally 0.5-1.0%). Other

suitable cryoprotectants include trehalose and lactose. Other suitable bulking agents include glycine and arginine, either of which can be included at a concentration of 0-0.05%, and polysorbate-80 (optimally included at a concentration of 0.005-0.01%). Additional surfactants include but are not limited to polysorbate 20 and BRIJ surfactants. The pharmaceutical composition may be prepared as an injectable solution and can further comprise an agent useful as an adjuvant, such as those used to increase absorption or dispersion, *e.g.*, hyaluronidase.

[0456] In some embodiments, the genetically engineered viruses are prepared for delivery, taking into consideration the need for efficient delivery and for overcoming the host antiviral immune response. Approaches to evade antiviral response include the administration of different viral serotypes as part of the treatment regimen (serotype switching), formulation, such as polymer coating to mask the virus from antibody recognition and the use of cells as delivery vehicles.

[0457] In another embodiment, the composition can be delivered in a controlled release or sustained release system. In one embodiment, a pump may be used to achieve controlled or sustained release. In another embodiment, polymeric materials can be used to achieve controlled or sustained release of the therapies of the present disclosure (see *e.g.*, U.S. Patent No. 5,989,463). Examples of polymers used in sustained release formulations include, but are not limited to, poly(2-hydroxy ethyl methacrylate), poly(methyl methacrylate), poly(acrylic acid), poly(ethylene-co-vinyl acetate), poly(methacrylic acid), polyglycolides (PLG), polyanhydrides, poly(N-vinyl pyrrolidone), poly(vinyl alcohol), polyacrylamide, poly(ethylene glycol), polylactides (PLA), poly(lactide-co-glycolides) (PLGA), and polyorthoesters. The polymer used in a sustained release formulation may be inert, free of leachable impurities, stable on storage, sterile, and biodegradable. In some embodiments, a controlled or sustained release system can be placed in proximity of the prophylactic or therapeutic target, thus requiring only a fraction of the systemic dose. Any suitable technique known to one of skill in the art may be used.

[0458] The genetically engineered bacteria of the invention may be administered and formulated as neutral or salt forms. Pharmaceutically acceptable salts include those formed with anions such as those derived from hydrochloric, phosphoric, acetic, oxalic, tartaric acids, etc., and those formed with cations such as those derived from sodium, potassium, ammonium, calcium, ferric hydroxides, isopropylamine, triethylamine, 2-ethylamino ethanol, histidine, procaine, etc.

In vivo Methods

[0459] The recombinant bacteria disclosed herein may be evaluated *in vivo*, *e.g.*, in an animal model. Any suitable animal model of a disease or condition associated with catabolism of a branched chain amino acid may be used (see, *e.g.*, Skvorak, *J. Inherit. Metab. Dis.*, 2009, 32:229-246 and Homanics *et al.*, *BMC Med. Genet.*, 2006, 7(33):1-13), including the Dbt^{-/-} model (E2 subunit of BCKDH, which has a 3-fold increase in blood and urine BCAA levels and results in neonatal lethality) (serves as classic MSUD model). This model is partially rescued by two transgenes (LAP-tTA and TRE-E2), allowing 5-6% of normal BCKDH activity and an increase in mice survival to

three or four weeks (serves as an intermediate MSUD model) (as described in Homanics *et al.*, 2006, the contents of which is herein incorporated by reference in its entirety). In addition, intermediate MSUD mice can be used to show development of neuropathology with striking similarity to human MSUD. In this model, branched-chain amino acid accumulation was associated with neurotransmitter deficiency, behavioral changes and limited survival, and providing intermediate MSUD mice with a choice between normal and branched-chain amino acid free diet prevented brain injury and dramatically improved survival (Zinnanti *et al.*, Dual mechanism of brain injury and novel treatment strategy in maple syrup urine disease; *Brain* 2009: 132; 903–918, the contents of which is herein incorporated by reference in its entirety). In some embodiments, the animal model is a mouse model of Maple Syrup Urine Disease. In one embodiment, the mouse model of MSUD is a branched-chain amino transferase knockout mouse (Wu *et al.*, *J. Clin. Invest.*, 113:434-440, 2004 or She *et al.*, *Cell Metabol.*, 6:181-194, 2007). In another embodiment, the mouse model of MSUD is a dihydrolipoamine dehydrogenase (E3) subunit knock-out mouse (Johnson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 94:14512-14517, 1997). In another embodiment, the mouse model of classic MSUD is a deletion of exon 5 and part of exon 4 of the E2 subunit of the branched-chain alpha-keto acid dehydrogenase (Homanics *et al.*, *BMC Med. Genet.*, 7:33, 2006) or the mouse model of intermediate MSUD (Homanics *et al.*, *BMC Med. Genet.*, 7:33, 2006). In another embodiment, the model is a Polled Shorthorn calf model of disease or a Polled Hereford calf model of disease (Harper *et al.*, *Aus. Vet. J.*, 66(2):46-49, 1988). Other relevant animal models include those described in She *et al.*, *Cell Metab.* 2007 September ; 6(3): 181–194; Wu *et al.*, *J. Clin. Invest.* 113:434–440 (2004); Bridi, *et al.*, *J. Neurosci Methods*, 2006 Sep 15;155(2):224-30.

[0460] The recombinant bacterial cells disclosed herein may administered to the animal, *e.g.*, by oral gavage, and treatment efficacy is determined, *e.g.*, by measuring blood leucine levels before and after treatment. The animal may be sacrificed, and tissue samples may be collected and analyzed.

Methods of Screening

Generation of Bacterial Strains with Enhance Ability to Transport Amino Acids

[0461] Due to their ease of culture, short generation times, very high population densities and small genomes, microbes can be evolved to unique phenotypes in abbreviated timescales. Adaptive laboratory evolution (ALE) is the process of passaging microbes under selective pressure to evolve a strain with a preferred phenotype. Most commonly, this is applied to increase utilization of carbon/energy sources or adapting a strain to environmental stresses (*e.g.*, temperature, pH), whereby mutant strains more capable of growth on the carbon substrate or under stress will outcompete the less adapted strains in the population and will eventually come to dominate the population.

[0462] This same process can be extended to any essential metabolite by creating an auxotroph. An auxotroph is a strain incapable of synthesizing an essential metabolite and must therefore have the metabolite provided in the media to grow. In this scenario, by making an auxotroph

and passaging it on decreasing amounts of the metabolite, the resulting dominant strains should be more capable of obtaining and incorporating this essential metabolite.

[0463] For example, if the biosynthetic pathway for producing an amino acid is disrupted a strain capable of high-affinity capture of said amino acid can be evolved via ALE. First, the strain is grown in varying concentrations of the auxotrophic amino acid, until a minimum concentration to support growth is established. The strain is then passaged at that concentration, and diluted into lowering concentrations of the amino acid at regular intervals. Over time, cells that are most competitive for the amino acid – at growth-limiting concentrations – will come to dominate the population. These strains will likely have mutations in their amino acid-transporters resulting in increased ability to import the essential and limiting amino acid.

[0464] Similarly, by using an auxotroph that cannot use an upstream metabolite to form an amino acid, a strain can be evolved that not only can more efficiently import the upstream metabolite, but also convert the metabolite into the essential downstream metabolite. These strains will also evolve mutations to increase import of the upstream metabolite, but may also contain mutations which increase expression or reaction kinetics of downstream enzymes, or that reduce competitive substrate utilization pathways.

[0465] A metabolite innate to the microbe can be made essential via mutational auxotrophy and selection applied with growth-limiting supplementation of the endogenous metabolite. However, phenotypes capable of consuming non-native compounds can be evolved by tying their consumption to the production of an essential compound. For example, if a gene from a different organism is isolated which can produce an essential compound or a precursor to an essential compound this gene can be recombinantly introduced and expressed in the heterologous host. This new host strain will now have the ability to synthesize an essential nutrient from a previously non-metabolizable substrate.

[0466] Hereby, a similar ALE process can be applied by creating an auxotroph incapable of converting an immediately downstream metabolite and selecting in growth-limiting amounts of the non-native compound with concurrent expression of the recombinant enzyme. This will result in mutations in the transport of the non-native substrate, expression and activity of the heterologous enzyme and expression and activity of downstream native enzymes. It should be emphasized that the key requirement in this process is the ability to tether the consumption of the non-native metabolite to the production of a metabolite essential to growth.

[0467] Once the basis of the selection mechanism is established and minimum levels of supplementation have been established, the actual ALE experimentation can proceed. Throughout this process several parameters must be vigilantly monitored. It is important that the cultures are maintained in an exponential growth phase and not allowed to reach saturation/stationary phase. This means that growth rates must be checked during each passaging and subsequent dilutions adjusted accordingly. If growth rate improves to such a degree that dilutions become large, then the concentration of auxotrophic supplementation should be decreased such that growth rate is slowed,

selection pressure is increased and dilutions are not so severe as to heavily bias subpopulations during passaging. In addition, at regular intervals cells should be diluted, grown on solid media and individual clones tested to confirm growth rate phenotypes observed in the ALE cultures.

[0468] Predicting when to halt the stop the ALE experiment also requires vigilance. As the success of directing evolution is tied directly to the number of mutations “screened” throughout the experiment and mutations are generally a function of errors during DNA replication, the cumulative cell divisions (CCD) acts as a proxy for total mutants which have been screened. Previous studies have shown that beneficial phenotypes for growth on different carbon sources can be isolated in about $10^{11.2}$ CCD¹. This rate can be accelerated by the addition of chemical mutagens to the cultures – such as *N*-methyl-*N*-nitro-*N*-nitrosoguanidine (NTG) – which causes increased DNA replication errors. However, when continued passaging leads to marginal or no improvement in growth rate the population has converged to some fitness maximum and the ALE experiment can be halted.

[0469] At the conclusion of the ALE experiment, the cells should be diluted, isolated on solid media and assayed for growth phenotypes matching that of the culture flask. Best performers from those selected are then prepped for genomic DNA and sent for whole genome sequencing. Sequencing with reveal mutations occurring around the genome capable of providing improved phenotypes, but will also contain silent mutations (those which provide no benefit but do not detract from desired phenotype). In cultures evolved in the presence of NTG or other chemical mutagen, there will be significantly more silent, background mutations. If satisfied with the best performing strain in its current state, the user can proceed to application with that strain. Otherwise the contributing mutations can be deconvoluted from the evolved strain by reintroducing the mutations to the parent strain by genome engineering techniques. See Lee, D.-H., Feist, A. M., Barrett, C. L. & Palsson, B. Ø. Cumulative Number of Cell Divisions as a Meaningful Timescale for Adaptive Laboratory Evolution of *Escherichia coli*. *PLoS ONE* 6, e26172 (2011).

[0470] Similar methods can be used to generate *E. coli* Nissle mutants that consume or import branched chain amino acids, *e.g.*, leucine, valine, and/or isoleucine.

Specific Screen to identify strains with improved BCAA degradation enzyme activity

[0471] Screens using genetic selection are conducted to improve BCAA consumption in the genetically engineered bacteria. Toxic BCAA analogs exert their mechanism of action (MOA) by being incorporated into cellular protein, causing cell death. These compounds, *e.g.*, fluoro-leucine and/or aza-leucine, have utility in an untargeted approach to select BCAA enzymes with increased activity. Assuming that these toxic compounds can be metabolized by BCAA enzymes into a non-toxic metabolite, rather than being incorporated into cellular protein, genetically engineered bacteria which have improved BCAA degradation activity can tolerate higher levels of these compounds, and can be screened for and selected on this basis.

Use of Valine and Leucine sensitivity to identify strains with improved BCAA degradation enzyme activity

[0472] Valine and Leucine sensitivity can be used as a genetic screening tool using the *E. coli* K12 strain. There are three AHAS (acetoxybutanoate synthase) isozymes in *E. coli* (AHAS I: *ilvBN*, AHAS II: *ilvGM*, and AHAS III: *ilvIH*). Valine and leucine exert feedback inhibition on AHAS I and AHAS III; AHAS II is resistant to Val and Leu inhibition. *E. coli* K12 has a frameshift mutation in *ilvG* (AHAS II) and is unable to produce BCAA endogenously in the presence of valine and leucine. In contrast, *E. coli* Nissle has a functional *ilvG* and is insensitive to valine and leucine and therefore cannot be used for this screen. A genetically engineered strain derived from *E. coli* K12, which more efficiently degrades leucine, has a greater reduction in sensitivity to leucine (through relieving the feedback inhibition on AHAS I and III). As a result, this pathway can be used as a tool to select and identify a strain with improved resistance to leucine.

Use of Leucine auxotrophy and D-leucine as a method to identify strains with improved BCAA uptake ability.

[0473] Bacterial mutants with increased leucine transport into the bacterial cell may be identified using a leucine auxotroph and providing D-leucine instead of L-leucine in the media, as D-leucine can be imported through the same transporters. The bacteria can grow in the presence of D-leucine, because the bacterial strain has a racemase, which can convert D-leucine to L-leucine. However, the uptake of D-leucine through LivKHMGF is less efficient than the uptake of L-leucine. The leucine auxotroph can still grow if high concentrations of D-Leucine are provided, even though the D-leucine uptake is less efficient than L-leucine uptake. When concentrations of D-leucine in the media are lowered, the cells can no longer grow, unless transport efficiency is increased, ergo, mutants with increased D-leucine uptake can be selected.

Methods of Treatment

[0474] Further disclosed herein are methods of treating a disease or disorder associated with the catabolism of a branched chain amino acid. In some embodiments, disclosed herein are methods for reducing, ameliorating, or eliminating one or more symptom(s) associated with these diseases or disorders. In one embodiment, the disorder involving the catabolism of a branched chain amino acid is a metabolic disorder involving the abnormal catabolism of a branched chain amino acid. Metabolic diseases associated with abnormal catabolism of a branched chain amino acid include maple syrup urine disease (MSUD), isovaleric acidemia, propionic acidemia, methylmalonic acidemia, diabetes ketoacidosis, MCC Deficiency, 3-Methylglutaconyl-CoA hydratase Deficiency, HMG-CoA Lyase Deficiency, Acetyl-CoA Carboxylase Deficiency, Malonyl-CoA Decarboxylase Deficiency, short-branched chain acylCoA dehydrogenase deficiency, 2-methyl-3-hydroxybutyric acidemia, beta-ketothiolase deficiency, isobutyryl-CoA dehydrogenase deficiency, HIBCH deficiency), and 3-Hydroxyisobutyric aciduria.

[0475] In one embodiment, the disease associated with abnormal catabolism of a branched chain amino acid is isovaleric acidemia. In one embodiment, the disease associated with abnormal catabolism of a branched chain amino acid is propionic acidemia. In one embodiment, the disease associated with abnormal catabolism of a branched chain amino acid is methylmalonic acidemia. In another embodiment, the disease associated with abnormal catabolism of a branched chain amino acid is diabetes.

[0476] In one embodiment, the disease is maple syrup urine disease (MSUD). Maple syrup urine disease (MSUD), also known as branched-chain ketoaciduria, is an autosomal recessive metabolic disorder caused by impaired activity of the branched-chain α -keto acid dehydrogenase (BCKDH) complex (Skvorak 2009). The overall incidence for MSUD is 1:185,000, although it is higher in certain populations, such as Mennonites, where the incidence is 1:176. The BCKDH complex is responsible for the oxidative decarboxylation of branched-chain keto acids (BCKAs) derived from branched chain amino acids (BCAAs) (Homanics *et al.* 2006). Patients with MSUD are unable to properly process BCKAs, which can lead to the toxic accumulation of BCAAs and their derivatives in the blood, cerebrospinal fluid and tissues (Skvorak 2009). Specifically, deficiencies of the BCKDH complex in MSUD patients results in accumulation of the BCAAs isoleucine, leucine, and valine, as well as their corresponding branched-chain α -keto acids (BCKAs) α -keto- β -methylvalerate, α -ketoisocaproate, and α -ketoisovalerate) in the tissues in plasma. Clinical manifestations of the disease vary depending on the degree of enzyme deficiency and include poor feeding, vomiting, dehydration, lethargy, hypotonia, seizures, hypoglycemia, ketoacidosis, pancreatitis, coma, and neurological decline (Homanics *et al.* 2009).

[0477] The BCKDH complex is composed of three catalytic components: a dehydrogenase/decarboxylase (E1), which is a heterotetramer composed of two E1 α and two E1 β subunits, a dihydrolipoyl transacylase (E2), and a dihydrolipoamide dehydrogenase (E3) (Skvorak 2009). Additionally, the complex is associated with two regulatory enzymes, a BCKDH kinase and a BCKDH phosphatase, which control its activity through reversible phosphorylation-dephosphorylation of E1 α (Chuang 1998, Homanics *et al.* 2006). To date, MSUD has been associated with mutations in the E1, E2 and E3 subunits of the BCKDH complex (Cheung 1998, Homanics *et al.* 2006).

[0478] MSUD is a very complex, genetically heterogeneous disease. At least 150 mutations in genes encoding BCKDH complex components have been identified that result in MSUD (Skvorak 2009). For example, see **Table 8** below, adapted from Chuang, *J. Pediatrics*, 132(3), Part 2, S17-S23, 1998. As indicated below, E2 mutants are the most prevalent in human disease.

Table 8: MSUD Phenotypes

Molecular Phenotype	Affected Gene	Clinical Phenotype	Number of Mutations Identified
IA	E1 α	Classic, Intermediate MSUD	15
IB	E1 β	Classic MSUD	4
II	E2	Classic, thiamine-responsive MSUD	26
III	E3	E3-deficient	4
IV	Kinase	None reported	None reported
V	Phosphatase	None reported	None reported

[0479] Currently available treatments for MSUD are inadequate for the long term management of the disease and have severe limitations (Svkvorak 2009). A low protein/BCAA-restricted diet, with micronutrient and vitamin supplementation, as necessary, is the widely accepted long-term disease management strategy for MSUD (Homanics *et al.* 2006). However, BCAA-intake restrictions can be particularly problematic since BCAAs can only be acquired through diet and are necessary for several metabolic activities (Skvorak 2009). Even with proper monitoring and patient compliance, BCAA dietary restrictions result in a high incidence of mental retardation and mortality (Skvorak 2009, Homanics *et al.* 2009). Further, a few cases of MSUD have been treated by liver transplantation (Popescu and Dima 2012) or treatment with phenylbutyrate. However, the limited availability of donor organs, the costs associated with the transplantation itself, and the undesirable effects associated with continued immunosuppressant therapy limit the practicality of liver transplantation for treatment of disease (Homanics *et al.* 2012, Popescu and Dima 2012). Therefore, there is significant unmet need for effective, reliable, and/or long-term treatment for MSUD.

[0480] The present disclosure surprisingly demonstrates that pharmaceutical compositions comprising the recombinant bacterial cells disclosed herein may be used to treat metabolic diseases involving the abnormal catabolism of a branched chain amino acid, such as MSUD. In one embodiment, the metabolic disease is selected from the group consisting of classic MSUD, intermediate MSUD, intermittent MSUD, E3-Deficient MSUD, and thiamine-responsive MSUD. In one embodiment, the disease is classic MSUD. In another embodiment, the disease is intermediate MSUD. In another embodiment, the disease is intermittent MSUD. In another embodiment, the disease is E3-deficient MSUD. In another embodiment, the disease is thiamine-responsive MSUD.

[0481] In one embodiment, the subject having MSUD has a mutation in an E1 α gene. In another embodiment, the subject having MSUD has a mutation in the E1 β gene. In another embodiment, the subject having MSUD has a mutation in the E2 gene. In another embodiment, the subject having MSUD has a mutation in the E3 gene.

[0482] In one embodiment, the target degradation rates of branched chain amino acids from food intake in breastfed infants and adults is as indicated in the **Table 9**, below.

Table 9. Target Degradation rates for BCAA.

Age (year)	<1			1-3			4-8			8-12			>12		
Amino acid: Leu (L) Val (V) Ile (I)	L	V	I	L	V	I	L	V	I	L	V	I	L	V	I
MSUD patient daily tolerance (mg/kg)	40	30	20	20	10	5	5	10	5	5	10	5	5	10	5
Recommended Dietary Allowance (RDA) (mg/kg)	93	58	43	63	37	28	49	28	22	46	26	20	46	26	20
Target reduction (mg/kg)	53	28	23	43	27	23	44	18	17	41	16	15	41	16	15
Target reduction (mg); (based on 10, 14, 26, 41 and 70kg weight for the different age groups)	530	280	230	602	378	322	1144	468	442	1681	656	615	2870	1120	1050
Target reduction (mmol)	4.04	2.39	1.75	4.59	3.23	2.45	8.72	3.99	3.37	12.81	5.60	4.69	21.88	9.56	8.00
Target degradation rate ($\mu\text{mol}/10^9$ CFUs/hr); (based on $3 \cdot 10^{11}$ CFUs/day dose)	0.56	0.33	0.24	0.64	0.45	0.34	1.21	0.55	0.47	1.78	0.78	0.65	3.04	1.33	1.11
Combined BCAA target degradation rate ($\mu\text{mol}/10^9$ CFUs/hr)	1.14			1.43			2.23			3.21			5.48		

[0483] In one embodiment, the target degradation rates of branched chain amino acids from food intake in breastfed infants and adults is as indicated in the charts, below.

[0484] The leucine consumption kinetics and dosing are set forth herein. Food intake is based on adult recommended daily allowance of 40 mg/kg/day. MSUD patients are primarily children with restricted protein intake.

[0485] In another embodiment, the disorder involving the catabolism of a branched chain amino acid is a disorder caused by the activation of mTor (mammalian target of rapamycin). mTor is a serine-threonine kinase and has been implicated in a wide range of biological processes including transcription, translation, autophagy, actin organization and ribosome biogenesis, cell growth, cell proliferation, cell motility, and survival. mTOR exists in two complexes, mTORC1 and mTORC2. mTORC1 contains the raptor subunit and mTORC2 contains rictor. These complexes are differentially regulated, and have distinct substrate specificities and rapamycin sensitivity. For example, mTORC1 phosphorylates S6 kinase (S6K) and 4EBP1, promoting increased translation and ribosome biogenesis to facilitate cell growth and cell cycle progression. S6K also acts in a feedback pathway to attenuate PI3K/Akt activation. mTORC2 is generally insensitive to rapamycin and is thought to modulate growth factor signaling by phosphorylating the C-terminal hydrophobic motif of some AGC kinases, such as Akt.

[0486] It is known in the art that mTor activation is caused by branched chain amino acids or alpha keto acids in the subject (see, for example, Harlan *et al.*, Cell Metabolism, 17:599-606, 2013). Specifically, activation of mTorC1 (mTor complex 1) is caused by leucine (see Han *et al.*, Cell, 149:410-424, 2012 and Lynch, J. Nutr., 131(3):861S-865S, 2001). Thus, in one embodiment, the disclosure provides methods of treating disorders involving the catabolism of leucine, caused by the activation of mTor by leucine in the subject. In one embodiment, the leucine levels in the subject are normal, and lowering leucine levels in the subject leads to the decreased activity of mTor and, thus, treatment of the disease. In another embodiment, the leucine levels in the subject are increased, and lowering leucine levels in the subject leads to the decreased activity of mTor and, thus, treatment of the disease. In one embodiment, the activation of mTor is increased as compared to the normal level of activation of mTor in a healthy subject, and lowering leucine levels in the subject leads to the decreased activation of mTor and, thus, treatment of the disease. In one embodiment, the level of activity of mTor is increased as compared to the normal level of activity of mTor in a healthy subject, and lowering leucine levels in the subject leads to the decreased activity of mTor and, thus, treatment of the disease. In another embodiment, the expression of mTor is increased as compared to the normal level of expression of mTor in a healthy subject, and lowering leucine levels in the subject leads to the decreased activity of mTor and, thus, treatment of the disease. In one embodiment, the activation of mTor is an abnormal activation of mTor.

[0487] Diseases caused by the activation of mTor are known in the art. See, for example, Laplante and Sabatini, Cell, 149(2):74-293, 2012. As used herein, the term “disease caused by the activation of mTor” includes cancer, obesity, type 2 diabetes, neurodegeneration, autism, Alzheimer’s disease, Lymphangioleiomyomatosis (LAM), transplant rejection, glycogen storage disease, obesity,

tuberous sclerosis, hypertension, cardiovascular disease, hypothalamic activation, musculoskeletal disease, Parkinson's disease, Huntington's disease, psoriasis, rheumatoid arthritis, lupus, multiple sclerosis, Leigh's syndrome, and Friedrich's ataxia.

[0488] In another aspect, the disclosure provides methods for decreasing the plasma level of at least one branched chain amino acid or branched chain α -keto acid in a subject by administering a pharmaceutical composition comprising a bacterial cell disclosed herein to the subject, thereby decreasing the plasma level of the at least one branched chain amino acid or branched chain α -keto acid or other BCAA metabolite in the subject. In one embodiment, the subject has a disease or disorder involving the catabolism of a branched chain amino acid. In one embodiment, the disorder involving the catabolism of a branched chain amino acid is a metabolic disorder involving the abnormal catabolism of a branched chain amino acid. In another embodiment, the disorder involving the catabolism of a branched chain amino acid is a disorder caused by the activation of mTor. In one embodiment, the disease or disorder is a maple syrup urine disorder (MSUD). In one embodiment, the branched chain amino acid is leucine, isoleucine, or valine. In one embodiment, the branched chain amino acid is leucine. In another embodiment, the branched chain amino acid is isoleucine. In another embodiment, the branched chain amino acid is valine. In another embodiment, the branched chain α -keto acid is α -ketoisocaproic acid (KIC). In another embodiment, the branched chain α -keto acid is α -ketoisovaleric acid (KIV). In another embodiment, the branched chain α -keto acid is α -keto- β -methylvaleric acid (KMV).

[0489] In some embodiments, the disclosure provides methods for reducing, ameliorating, or eliminating one or more symptom(s) associated with these diseases, including but not limited to neurological deficits, mental retardation, brain damage, brain oedema, blindness, branched chain α -keto acid acidosis, myelinization failure, hyperammonaemia, coma, developmental delay, neurological impairment, failure to thrive, ketoacidosis, seizure, ataxia, neurodegeneration, hypotonia, lactic acidosis, recurrent myoglobinuria, and/or liver failure. In some embodiments, the disease is secondary to other conditions, *e.g.*, liver disease.

[0490] In certain embodiments, the bacterial cells disclosed herein are capable of catabolizing branched chain amino acid(s), *e.g.*, leucine, in the diet of the subject in order to treat a disease associated with catabolism of a branched chain amino acid, *e.g.*, MSUD. In these embodiments, the bacterial cells are delivered simultaneously with dietary protein. In another embodiment, the bacterial cells are delivered simultaneously with phenylbutyrate. In another embodiment, the bacterial cells are delivered simultaneously with a thiamine supplement. In some embodiments, the bacterial cells and dietary protein are delivered after a period of fasting or leucine-restricted dieting. In these embodiments, a patient suffering from a disorder involving the catabolism of a branched chain amino acid, *e.g.*, MSUD, may be able to resume a substantially normal diet, or a diet that is less restrictive than a leucine-free diet. In some embodiments, the bacterial cells may be capable of catabolizing leucine from additional sources, *e.g.*, the blood, in order to treat a disease

associated with the catabolism of a branched chain amino acid, *e.g.*, MSUD. In these embodiments, the bacterial cells need not be delivered simultaneously with dietary protein, and a leucine gradient is generated, *e.g.*, from blood to gut, and the recombinant bacteria catabolize the branched chain amino acid, *e.g.*, leucine, and reduce plasma levels of the branched chain amino acid, *e.g.*, leucine.

[0491] The method may comprise preparing a pharmaceutical composition with at least one genetically engineered species, strain, or subtype of bacteria described herein, and administering the pharmaceutical composition to a subject in a therapeutically effective amount. In some embodiments, the bacterial cells disclosed herein are administered orally, *e.g.*, in a liquid suspension. In some embodiments, the bacterial cells disclosed herein are lyophilized in a gel cap and administered orally. In some embodiments, the bacterial cells disclosed herein are administered via a feeding tube or gastric shunt. In some embodiments, the bacterial cells disclosed herein are administered rectally, *e.g.*, by enema. In some embodiments, the genetically engineered bacteria are administered topically, intraintraintestinally, intrajejunally, intraduodenally, intraileally, and/or intracolically.

[0492] In certain embodiments, the administering the pharmaceutical composition described herein reduces branched chain amino acid levels in a subject. In some embodiments, the methods of the present disclosure reduce the branched chain amino acid levels, *e.g.*, leucine levels, in a subject by at least about 10%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, or more. In another embodiment, the methods of the present disclosure reduce the branched chain amino acid levels, *e.g.*, leucine levels, in a subject by at least two-fold, three-fold, four-fold, five-fold, six-fold, seven-fold, eight-fold, nine-fold, or ten-fold. In some embodiments, reduction is measured by comparing the branched chain amino acid level in a subject before and after administration of the pharmaceutical composition. In one embodiment, the branched chain amino acid level is reduced in the gut of the subject. In another embodiment, the branched chain amino acid level is reduced in the blood of the subject. In another embodiment, the branched chain amino acid level is reduced in the plasma of the subject. In another embodiment, the branched chain amino acid level is reduced in the brain of the subject.

[0493] In one embodiment, the pharmaceutical composition described herein is administered to reduce branched chain amino acid levels in a subject to normal levels. In another embodiment, the pharmaceutical composition described herein is administered to reduce branched chain amino acid levels in a subject below normal levels to, for example, decrease the activation of mTor.

[0494] In certain embodiments, the pharmaceutical composition described herein is administered to reduce branched chain α -keto-acid levels in a subject. In some embodiments, the methods of the present disclosure reduce the branched chain α -keto-acid levels, in a subject by at least about 10%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, or more as compared to levels in an untreated or control subject. In another embodiment, the methods of the present disclosure reduce the branched chain α -keto-acid levels, in a subject by at least two-fold, three-fold, four-fold, five-fold, six-fold, seven-fold, eight-fold, nine-fold, or ten-fold. In some

embodiments, reduction is measured by comparing the branched chain α -keto-acid levels in a subject before and after administration of the pharmaceutical composition. In one embodiment, the branched chain α -keto-acid level is reduced in the gut of the subject. In another embodiment, the branched chain α -keto-acid level is reduced in the blood of the subject. In another embodiment, the branched chain α -keto-acid level is reduced in the plasma of the subject. In another embodiment, the branched chain α -keto-acid level is reduced in the brain of the subject.

[0495] In one embodiment, the pharmaceutical composition described herein is administered to reduce the branched chain α -keto-acid level in a subject to a normal level. In another embodiment, the pharmaceutical composition described herein is administered to reduce the branched chain α -keto-acid level in a subject below a normal level to, for example, decrease the activation of mTor.

[0496] In some embodiments, the method of treating the disorder involving the catabolism of a branched chain amino acid, *e.g.*, MSUD, allows one or more symptoms of the condition or disorder to improve by at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or more. In some embodiments, the method of treating the disorder involving the catabolism of a branched chain amino acid, *e.g.*, MSUD, allows one or more symptoms of the condition or disorder to improve by at least about two-fold, three-fold, four-fold, five-fold, six-fold, seven-fold, eight-fold, nine-fold, or ten-fold.

[0497] Before, during, and after the administration of the pharmaceutical composition, branched chain amino acid levels, *e.g.*, leucine levels, in the subject may be measured in a biological sample, such as blood, serum, plasma, urine, peritoneal fluid, cerebrospinal fluid, fecal matter, intestinal mucosal scrapings, a sample collected from a tissue, and/or a sample collected from the contents of one or more of the following: the stomach, duodenum, jejunum, ileum, cecum, colon, rectum, and anal canal. In some embodiments, the methods may include administration of the compositions disclosed herein to reduce levels of the branched chain amino acid, *e.g.*, leucine. In some embodiments, the methods may include administration of the compositions disclosed herein to reduce the branched chain amino acid, *e.g.*, leucine, to undetectable levels in a subject. In some embodiments, the methods may include administration of the compositions disclosed herein to reduce the branched chain amino acid, *e.g.*, leucine, concentrations to undetectable levels, or to less than about 1%, 2%, 5%, 10%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, or 95% of the subject's branched chain amino acid levels prior to treatment.

[0498] In some embodiments, the recombinant bacterial cells disclosed herein produce a branched chain amino acid catabolism enzyme, *e.g.*, KivD, BCKD and/or other BCAA catabolism enzyme, BCAA transporter, BCAA binding protein, etc, under exogenous environmental conditions, such as the low-oxygen environment of the mammalian gut, to reduce levels of branched chain amino acids in the blood or plasma by at least about 1.5-fold, at least about 2-fold, at least about 3-fold, at least about 4-fold, at least about 5-fold, at least about 6-fold, at least about 7-fold, at least about 8-fold, at least about 9-fold, at least about 10-fold, at least about 15-fold, at least about 20-fold, at least

about 30-fold, at least about 40-fold, or at least about 50-fold as compared to unmodified bacteria of the same subtype under the same conditions.

[0499] In one embodiment, the bacteria disclosed herein reduce plasma levels of the branched chain amino acid, *e.g.*, leucine, will be reduced to less than 4 mg/dL. In one embodiment, the bacteria disclosed herein reduce plasma levels of the branched chain amino acid, *e.g.*, leucine, will be reduced to less than 3.9 mg/dL. In one embodiment, the bacteria disclosed herein reduce plasma levels of the branched chain amino acid, *e.g.*, leucine, will be reduced to less than 3.8 mg/dL, 3.7 mg/dL, 3.6 mg/dL, 3.5 mg/dL, 3.4 mg/dL, 3.3 mg/dL, 3.2 mg/dL, 3.1 mg/dL, 3.0 mg/dL, 2.9 mg/dL, 2.8 mg/dL, 2.7 mg/dL, 2.6 mg/dL, 2.5 mg/dL, 2.0 mg/dL, 1.75 mg/dL, 1.5 mg/dL, 1.0 mg/dL, or 0.5 mg/dL.

[0500] In one embodiment, the subject has plasma levels of at least 4 mg/dL prior to administration of the pharmaceutical composition disclosed herein. In another embodiment, the subject has plasma levels of at least 4.1 mg/dL, 4.2 mg/dL, 4.3 mg/dL, 4.4 mg/dL, 4.5 mg/dL, 4.75 mg/dL, 5.0 mg/dL, 5.5 mg/dL, 6 mg/dL, 7 mg/dL, 8 mg/dL, 9 mg/dL, or 10 mg/dL prior to administration of the pharmaceutical composition disclosed herein.

[0501] Certain unmodified bacteria will not have appreciable levels of branched chain amino acid, *e.g.*, leucine, processing. In embodiments using genetically modified forms of these bacteria, processing of branched chain amino acids, *e.g.*, leucine, will be appreciable under exogenous environmental conditions.

[0502] Branched chain amino acid levels, *e.g.*, leucine levels, may be measured by methods known in the art, *e.g.*, blood sampling and mass spectrometry. In some embodiments, branched chain amino acid catabolism enzyme expression is measured by methods known in the art. In another embodiment, branched chain amino acid catabolism enzyme activity is measured by methods known in the art to assess BCAA activity.

[0503] In certain embodiments, the recombinant bacteria are *E. coli* Nissle. The recombinant bacteria may be destroyed, *e.g.*, by defense factors in the gut or blood serum (Sonnenborn *et al.*, 2009) or by activation of a kill switch, several hours or days after administration. Thus, the pharmaceutical composition comprising the recombinant bacteria may be re-administered at a therapeutically effective dose and frequency. In alternate embodiments, the recombinant bacteria are not destroyed within hours or days after administration and may propagate and colonize the gut.

[0504] In one embodiment, the bacterial cells disclosed herein are administered to a subject once daily. In another embodiment, the bacterial cells disclosed herein are administered to a subject twice daily. In another embodiment, the bacterial cells disclosed herein are administered to a subject in combination with a meal. In another embodiment, the bacterial cells disclosed herein are administered to a subject prior to a meal. In another embodiment, the bacterial cells disclosed herein are administered to a subject after a meal. The dosage of the pharmaceutical composition and the frequency of administration may be selected based on the severity of the symptoms and the

progression of the disease. The appropriate therapeutically effective dose and/or frequency of administration can be selected by a treating clinician.

[0505] The methods disclosed herein may comprise administration of a composition disclosed herein alone or in combination with one or more additional therapies, *e.g.*, the phenylbutyrate, thiamine supplementation, and/or a low-branched chain amino acid, *e.g.*, a low-leucine, diet. An important consideration in the selection of the one or more additional therapeutic agents is that the agent(s) should be compatible with the bacteria disclosed herein, *e.g.*, the agent(s) must not interfere with or kill the bacteria. In some embodiments, the genetically engineered bacteria are administered in combination with a low protein diet. In some embodiments, administration of the genetically engineered bacteria provides increased tolerance, so that the patient can consume more protein.

[0506] The methods disclosed herein may further comprise isolating a plasma sample from the subject prior to administration of a composition disclosed herein and determining the level of the branched chain amino acid, *e.g.*, leucine, or branched chain alpha-keto-acid in the sample. In some embodiments, the methods disclosed herein may further comprise isolating a plasma sample from the subject after to administration of a composition disclosed herein and determining the level of the branched chain amino acid, *e.g.*, leucine, or branched chain alpha-keto-acid in the sample.

[0507] In one embodiment, the methods disclosed herein further comprise comparing the level of the branched chain amino acid or branched chain α -keto-acid in the plasma sample from the subject after administration of a composition disclosed herein to the subject to the plasma sample from the subject before administration of a composition disclosed herein to the subject. In one embodiment, a reduced level of the branched chain amino acid or branched chain alpha-keto-acid in the plasma sample from the subject after administration of a composition disclosed herein indicates that the plasma levels of the branched chain amino acid or branched chain alpha-keto-acid are decreased, thereby treating the disorder involving the catabolism of the branched chain amino acid in the subject. In one embodiment, the plasma level of the branched chain amino acid or branched chain α -keto-acid is decreased at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% in the sample after administration of the pharmaceutical composition as compared to the plasma level in the sample before administration of the pharmaceutical composition. In another embodiment, the plasma level of the branched chain amino acid or branched chain α -keto-acid is decreased at least two-fold, three-fold, four-fold, or five-fold in the sample after administration of the pharmaceutical composition as compared to the plasma level in the sample before administration of the pharmaceutical composition.

[0508] In one embodiment, the methods disclosed herein further comprise comparing the level of the branched chain amino acid or branched chain α -keto-acid in the plasma sample from the subject after administration of a composition disclosed herein to a control level of the branched chain amino acid or branched chain alpha-keto-acid. In one embodiment, the control level of the branched

chain amino acid is 4 mg/dL. In one embodiment, the subject is considered treated if the level of branched chain amino acid, *e.g.*, leucine, in the plasma sample from the subject after administration of the pharmaceutical composition disclosed herein, is less than 4 mg/dL. In one embodiment, the subject is considered treated if the level of branched chain amino acid, *e.g.*, leucine, in the plasma sample from the subject after administration of the pharmaceutical composition disclosed herein is less than 3.9, 3.8, 3.7, 3.6, 3.5, 3.4, 3.3, 3.2, 3.1, 3.0, 2.5, 2.0, 1.5, 1.0 or 0.5 mg/dL.

Exemplification

[0509] The present disclosure is further illustrated by the following examples which should not be construed as limiting in any way. The contents of all cited references, including literature references, issued patents, and published patent applications, as cited throughout this application are hereby expressly incorporated herein by reference. It should further be understood that the contents of all the figures and tables attached hereto are also expressly incorporated herein by reference.

Experimental Methods for all Examples

1-1. *Design and construction of LeuDH, KivD, and Adh enzyme libraries*

[0510] *Metagenomic enzyme discovery.* Machine-learning-based bioinformatics tools were used to identify enzyme candidates for each of the three desired activities (leucine dehydrogenase, 1.4.1.9; ketoisovalerate decarboxylase, 4.1.1.1; and alcohol dehydrogenase 1.1.1.1) in public sequence databases (SwissProt and TrEMBL, together known as UniProt). For LeuDH and Adh, sequence diversity was maximized using proprietary algorithms developed at Ginkgo. For KivD, a stratified sampling approach was used.

[0511] *Rational enzyme design.* For LeuDH and Adh, molecular models of the enzyme–transition state complex were built using Rosetta software, and systematic mutations of the active site residues to each of the 20 amino acids were designed.

[0512] *Library synthesis.* DNA sequences for all LeuDH, KivD, and Adh enzymes were codon-optimized for expression in *E. coli* via Ginkgo-developed sequence design tools. Coding sequences were synthesized in an inducible *E. coli* expression vector under the control of the T7 promoter at Twist Bioscience.

1-2. *High-throughput in vitro enzyme assays*

[0513] *Cell growth and enzyme preparation.* Strains harboring library plasmids were transformed into an *E. coli* T7 expression host. 5μL/well of thawed glycerol stocks were stamped into 500μL/well of LB + 100ug/mL Carbenicillin (LB-Carb100) in half-height deepwell plates, which were sealed with AeraSeals. Samples were incubated at 37°C and shaken at 1000 RPM in 80% humidity overnight. 50μL/well of the resulting precultures were stamped into 450μL/well of LB-Carb100 + 1mM IPTG in half-height deepwell plates, which were sealed with AeraSeals. Samples

were incubated at 30°C and shaken at 1000 RPM in 80% humidity overnight. 250µL/well of the resulting production cultures were stamped into deepwell plates containing 500uL of phosphate buffered saline (PBS) and centrifuged for 10 minutes at 4000*G. Supernatant was removed and the resulting cell pellet was resuspended in 200uL of BugBuster Protein Extraction Reagent + 1uL/mL purified Benzonase + 1uL/6mL purified Lysozyme. Samples were incubated for 10 minutes at room temperature to generate the cell lysates used in *in vitro* enzyme assays.

[0514] *LeuDH activity assay*. 10uL of lysate was transferred to a half-area flat-bottom plate containing 90uL/well assay buffer (20mM amino acid [L-Leucine, L-Valine, or L-Isoleucine], 200mM Glycine, 200mM KCl, 0.4mM NAD, pH 10.5). Optical measurements were taken on a plate reader, with absorbance readings taken at 340nm for 10 minutes. The resulting kinetic data was used to resolve the rate of NAD⁺ reduction, a proxy for LeuDH activity.

[0515] *KivD activity assay*. 10 uL of lysate was transferred to a half-area flat-bottom plate containing 90uL/well assay buffer (100 mM PIPES-KOH, 100 mM Potassium glutamate, 1mM Dithiothreitol, 0.4 mM NAD, 1.5 mM Thiamine pyrophosphate, 10 mM Magnesium glutamate, 20 mM ketoisocaproate (KIC), pH 7.5). A coupling enzyme was used to indirectly measure KivD activity on KIC. Optical absorbance measurements were taken over 10 minutes. The resulting kinetic data was used to determine KivD activity.

[0516] *Adh activity assay*. 10 uL of lysate was transferred to a half-area flat-bottom plate containing 90uL/well assay buffer (50 mM MOPS buffer, 0.4mM NADH, and 30mM isovaleraldehyde, pH 7.0). Optical absorbance measurements were taken on a plate reader at 340nm for 10 minutes. The resulting kinetic data was used to resolve the rate of NADH oxidation, a proxy for ADH activity.

[0517] *LeuDH selectivity assay*. To measure LeuDH selectivity (specific deamination of L-Leu in the presence of L-Ile and L-Val), lysate was diluted four-fold in lysis buffer, and 10uL/well of the newly diluted lysate was stamped into 90uL/well of a modified assay buffer from above, featuring 0.5 mM of each amino acid (L-leucine, L-isoleucine, L-valine), 200 mM Glycine, 200 mM Potassium chloride, and 4mM NAD. The reaction was quenched at different timepoints and submitted for LC-MS quantification of leucine, isoleucine, and valine.

1-3. *Design and construction of optimized BCAA pathways*

[0518] Based on their performance in the *in vitro* assay, six (6) LeuDH enzymes, three (3) KivD enzymes, and three (3) Adh enzymes were selected to build candidate optimized BCAA operons. For translational control of each enzyme, we generated 3 ribosome binding sites of differing strengths. These RBS-gene pairs were then assembled into a partial combinatorial library in Synlogic's pathway plasmid. All pathway operons maintained the gene order of Synlogic's prototype pathway (LeuDH-KivD-Adh) and the final BrnQ gene from the parent pathway.

1-4. *High-throughput BCAA Pathway screening*

[0519] *Cell preparation.* BCAA pathway plasmids were transformed into the SYN469 chassis (*AilvC*, *AleuE*, *lacZ::Ptet-livKHMGF*) and plated on LB agar plates supplemented with 100 µg/ml carbenicillin to select for strains harboring operon plasmids. Up to 6 colonies were picked from each transformation, grown in LB media supplemented with 100 µg/ml carbenicillin and cryopreserved in 20% glycerol prior to screening. For screening, strains were inoculated from cryostocks into FM1 medium supplemented with 100 µg/ml carbenicillin and grown overnight at 37°C to generate pre-cultures. Pre-cultures were back-diluted in to FM1 media supplemented with 100 µg/ml carbenicillin and grown at 37°C for approximately 2-2.5 hours to OD₆₀₀ of ~1.5-2. Cultures were then induced at a final concentration of 100 µg/ml aTc and allowed to grow for 2.5 hours at 37°C. Cultures were harvested and cryopreserved in Cryopreservation Buffer in 96-well plates.

[0520] *Pathway screening.* Cells were thawed on ice and the OD₆₀₀ value of each culture plate was measured. The OD₆₀₀ value was averaged across each 96-well plate. The average OD₆₀₀ for each plate was then used to determine the dilution necessary to achieve a final OD₆₀₀ of 2 in 0.8 ml Assay Buffer without leucine. Following this dilution, the samples were centrifuged, and the supernatant was discarded. The resulting cell pellets were resuspended in 0.8 ml Assay Buffer without leucine. An equal volume of 2x Assay Buffer (8 mM leucine) was added to the samples for a final volume of 1.6 ml at an OD₆₀₀ value of 1 to initiate the reaction. Plates were sealed with foil heat seals and incubated at 37°C without shaking for 4 hours. Samples were collected at 0, 2 and 4h intervals. Upon collection, samples were centrifuged to remove cells and supernatant was stored at 4°C prior to analysis by LC-MS, LC-MS-UV and/or GC-MS.

1-5. *Metabolite analysis*

[0521] Leu, Val, and Ile were measured using LC-MS in a multiplexed QQQ method with linearity from 1 µg/ml - 1000 µg/ml. Keto isocaproate (KIC) and isovaleraldehyde was detected by derivatizing assay samples with 2-hydrazinoquinoline. The KIC and isovaleraldehyde derivatives were detected using LC-MS-UV. IPA was measured in assay samples using GC-MS.

1-6. *In vitro assay*

[0522] Materials prepared from flasks or fermenters were thawed on ice, and cell density was measured by light absorption at 600 nm (OD₆₀₀). OD₆₀₀ of 1.0 was assumed to be equal to 10⁹ cells/mL in this method. A volume was calculated to target 1 mL of 2 x 10⁹ cells/mL cell resuspension and the cells were transferred into a 96-deep well plate and washed once with cold PBS. After centrifugation (4000 rpm, 4°C, 10 min), the PBS was discarded, and the cell pellets were then resuspended in 1 mL of 1x M9 + 50 mM MOPS + 0.5% glucose (MMG) buffer. Eight hundred (800) µL of each sample was transferred into a new 96-deep well plate and 800 µL of MMG containing 16 mM leucine was added, mixed well by pipetting. A sample (200 µL) assigned as time zero was collected at this moment. The plate was then covered by a breathable membrane and moved to

anaerobic chamber to incubate at 37 °C. Samples were also collected at 2 h and 4 h during incubation in the anaerobic chamber. The samples were centrifuged for 10 min at 4000 rpm at 4 °C immediately after collection. 100 µL of the supernatant was transferred into a new 96-well plate and stored at -80 °C for future analysis. For different testing purposes, various amounts of branched chain amino acid(s) could be added to the reaction.

1-7. *Flask cell preparation*

[0523] The culture was inoculated into 4 mL of 2xYT medium containing proper antibiotics in a culture tube incubated at 37 °C with shaking (250 RPM) overnight. The next day, 500 µL of the overnight culture was transferred into 50 mL of FM1 medium containing proper antibiotics in a 125 mL baffled flask. The flasks were incubated for 2 hours at 37 °C with shaking at 250 RPM. Inducer was added after this 2-hour incubation for induction of pathway enzymes. The flasks remained shaking for 4 hours after induction. Cultures were then transferred into a 50 mL tube and centrifuged for 16 min at 5000 rpm at 4 °C. The cell pellet was washed once with cold PBS and then resuspended into 800 µL of cold buffer (1xPBS+15% glycerol) and aliquoted into PCR tubes to store at -80 °C.

1-8. *Fermentation cell preparation*

[0524] A frozen master cell bank scrape was used to inoculate 50mL fermentation media in a 500mL ultrayield flask, and grown to stationary phase overnight at 350 RPM, 37°C, resulting in a final density of 30-40 OD₆₀₀. AMBR250 vessels were prepared containing 200mL (unless volume of feed was too high, then reduced to 100 mL) of media, with either 25 g/L glucose or 40 g/L glycerol and appropriate antibiotics, and equilibrated at 37°C, 1500 RPM, 100mL/min air flow, and pH 7 (pH control by ammonium hydroxide). They were inoculated to approximately 0.15 OD₆₀₀ from the overnight seed culture and grown to varying OD₆₀₀ levels before addition of varying levels of inducing agent, tetracycline, and other supplementations (see experimental section below). Cultures were then grown to a final density target of 30 OD₆₀₀ where possible, maintaining dissolved oxygen levels at 60% using agitation up to 3500 RPM and then up to 80% substituted oxygen in the gas (100 mL/min fixed total gas flow) via cascade control. This material was spun down at 2500xg at 4°C in 1 L centrifuge bottles, then resuspended in 100mM K-phosphate pH 7.5 buffer with 15% (vol/vol) glycerol to 1/5th the volume, or approximately 1x10¹¹ cells/mL, and then frozen for storage at -80°C.

1-9. *In vitro gastrointestinal simulation of strain activity*

[0525] To characterize the viability and metabolic activity of engineered bacterial strains, and to predict their function *in vivo*, an *in vitro* gastrointestinal simulation (IVS) model was designed to simulate key aspects of the human gastrointestinal tract, including oxygen concentration, gastric and pancreatic enzymes, and bile. The IVS model is comprised of a series of incubations in 96-well microplate format designed to simulate stomach, small intestine, and colonic conditions. The stomach and small intestinal portions of the IVS model were adapted from Minekus *et al.*, 2014 (3). To study the activity of engineered strains designed for the treatment of MSUD, only the simulated stomach was considered.

[0526] Briefly, frozen aliquots of bacterial cells were first thawed at room temperature and resuspended in 0.077 M sodium bicarbonate buffer at 5.0×10^9 CFU/mL. This solution was then mixed with equal parts of simulated gastric fluid (SGF; Minekus *et al.*, 2014) containing 5 mM leucine and incubated for 2 hours at 37°C with shaking in a Coy microaerobic chamber. The atmosphere within the microaerobic chamber was initially calibrated to 7% oxygen and gradually decreased to 2% oxygen over 2 hours. The cell density in SGF is 2.5×10^9 CFU/mL. For testing the mixture of BCAA, 5 mM of each BCAA was added into the system.

[0527] To determine strain activity over time, SGF aliquots were collected periodically and centrifuged at 4000 rpm for 5 mins using a tabletop centrifuge. Cell free supernatants were collected and stored at -20°C prior to mass spectrometry analysis of leucine and isopentanol concentration.

1-10. Efficacy of engineered E. coli Nissle in iMSUD mice

[0528] The iMSUD mouse model was generated as previously described (3) and breeder pairs sent to Charles River Laboratories' (CRL) Genetically Engineered Models (GEMS) for colony maintenance. Male and female mice aged 21-45 days were shipped to the Synlogic facility and allowed to acclimate for a minimum of 4 days on a mixture of branched chain amino acid free diet (Dyets, 510081) and 18% protein diet (Envigo, TD.2018.15) diet before study initiation. Mice were assigned to study groups according to date of birth and body weight and orally administered vehicle (15% glycerol in PBS), SYN1980 (8.9×10^9 CFU) or SYN5941 (8.9×10^9 CFU) twice a day (BID) for a total of 4 days. Male and female mice were switched to a 70% high-protein diet and provided hydrogel (Envigo, TD.06723) on days 2 and 3, respectively, while receiving vehicle or bacteria BID. On the morning of day 5, mice received 66.7 mg peptone followed by vehicle or bacteria 30 minutes later. Two hours post bacteria dose, submandibular bleedings were performed, and plasma isolated for measurement of branch-chain amino acids (BCAA) leucine, isoleucine and valine levels by mass spectrometry.

1-11. Efficacy of engineered E. coli Nissle in healthy non-human primates

[0529] Twelve monkeys were studied. Each study included 2 conditions (vehicle, SYN1980 or SYN5941). A total of 3 studies were performed and combined so each group would include n = 10-12. Animals were allowed a 1 week break between dosing to allow strains to clear from the intestine between studies.

[0530] On the morning of the study, each monkey was removed from its cage and an oral gavage tube was inserted before 7.8 mL of vehicle (15% glycerol in PBS) or bacteria (1×10^{12} CFUs) was administered to each animal, followed by 5 mL of 0.36 M sodium bicarbonate, 14 mL of 500 g/L peptone and 5 mL water flush. Animals were then returned to their cages and blood was collected at 0, 0.5, 1, 2, 4 and 6 hours post dosing. Plasma was isolated and stored at -70 °C for analysis of BCAAs by mass spectrometry.

1-12. Analysis of branched chain amino acids by LC-MS-MS

[0531] Leucine, Isoleucine, and Valine were quantitated in bacterial supernatant, plasma, and urine by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) using either an Ultimate 3000 UHPLC-TSQ Quantum or a Vanquish UHPLC-TSQ Altis system. Samples were extracted with 9 parts 2:1 acetonitrile: water containing 1 ug/mL leucine-d3 as an internal standard, vortexed, and centrifuged. Supernatants were diluted with 9 parts 0.1% formic acid and analyzed concurrently with standards processed as above from 0.8 to 1000 ug/mL. Samples were separated on a Phenomenex Synergi 4 um Hydro-RP 80A, 75 x 2 mm using a 0.1% formic acid (A), 0.1% formic acid/acetonitrile (B) at 0.3 mL/min and 50°C. After a 2 uL injection and an initial 5% B hold from 0 to 0.5 minutes, analytes were gradient eluted from 5 to 90% B over 0.5 to 1.5 minutes followed by high organic wash and aqueous equilibration steps. Analytes were detected using Selected Reacting Monitoring (SRM) of compound specific collision induced fragments in electrospray positive ion mode (leucine: 132>86, isoleucine: 132>69, Valine: 118>72, leucine-d3: 135>89). SRM chromatograms were integrated and the unknown/internal standard peak area ratios were used to calculate concentrations against the standard curve.

1-13. Analysis of isopentanol by GC-MS

[0532] Isopentanol was quantitated in bacterial supernatant by gas chromatography - mass spectrometry (GC-MS) using an Agilent 7890B-5977B system. Samples were extracted with 2 parts ethyl acetate containing 100 ug/mL isopentanol-d4 as an internal standard, vortexed, and centrifuged. The upper organic layer was removed and analyzed concurrently with standards from 4 to 1000 mg/mL processed as above. Samples or standards, 1 uL injection (split 10:1, inlet at 250 C), were separated on a RESTEK Rxi-35Sil MS, 30 m x 0.25 mmID, 0.25 umdf at 1 mL / min helium. After a 5-minute hold at 35°C, a temperature gradient from 35 to 310°C over 5 minutes was used to elute the analytes. Compound were detected by the Selected Ion Monitoring (SIM) of specific positive fragment ions produced by electron ionization (isopentanol: 55.1, isopentanol-d4: 59.1). SIM chromatograms were integrated, and the unknown/internal standard peak area ratios were used to calculate concentrations against the standard curve.

Example 1. Generation of Various Recombinant Bacterial Strains

[0533] **Table 10** lists all the bacterial strains used in this application. *Escherichia coli* Nissle 1917 (EcN), designated as SYN001 here, was purchased from the German Collection of Microorganisms and Cell Cultures (DSMZ Braunschweig, *E. coli* DSM 6601). The procedures of generating recombinant bacteria (including SYN094, SYN469 and SYN1980) starting from the wild type strain SYN001 were fully disclosed in Examples 1-4 and others of International Application No. PCT/US2016/032565, the entire contents of which are incorporated herein by reference. Existing vectors described below including the operon pSC101-P_{ter}-*leuDH-kivd-adh2-brnQ-ampR* was used to generate SYN1980 and disclosed in Example 19 and others of International Application No.

PCT/US2016/032565. For ease of reference, pSC101-*P_{tet}-leuDH-kivd-adh2-brnQ-ampR* is referred herein as “Original Operon”. Recombinant strains corresponding to ID 5-14 of **Table 10** were newly generated by the present application.

[0534] SYN001 with streptomycin resistance was obtained by streaking $\sim 10^{11}$ cells of SYN001 on LB plate containing 300 ng/mL streptomycin and taking the single colony formed, designated as SYN094. The deletion of the *ilvC* and *leuE* gene was done by P1 transduction from the corresponding BW25113 strains and cured of antibiotic cassette afterwards. The correct colonies were verified by PCR and stored for future usage. For insertion of the *lacZ::Ptet-livKHPMF* cassette, the proper DNA fragment was generated by overlapping PCR and integrated into the chromosome by lambda red recombineering method. Colonies were selected on LB agar containing proper antibiotics (30 μ g/mL chloramphenicol) and correct recombination events were verified by PCR. A verified colony was saved for future operations. Chromosomal insertions of extra *brnQ* cassette into the EcN genome were carried out using the well-characterized lambda red recombineering approach as well. For this insertion, (1) a pKD3- or pKD4-based plasmid containing $\sim 1,000$ bp of 5' and 3' EcN genome homology for recombination was built, followed by (2) insertion of the target fragment into the plasmid by isothermal assembly (HiFi DNA Assembly Master Mix, NEB), (3) amplification of the insertion fragment from the plasmid by PCR (including EcN homology regions and antibiotic resistance cassette, Q5 High-Fidelity Master Mix, NEB), (4) recombineering of the insertion fragment by electroporation via pKD46 and subsequent pKD46 removal, and (5) the removal of antibiotic resistance cassettes via pCP20 and subsequent pCP20 removal.

[0535] A plasmid harboring the BCAA- consuming operon under the control of various promoters was either synthesized or PCR amplified from corresponding vectors containing an ampicillin resistance gene, a low copy number origin of replication (pSC101) and under the regulatory control of the *P_{tet}* promoter. Various promoter systems such as *P_{Cl857}*, *P_{lacI/fnr}*, *P_{BAD}* or *P_{cmT}* promoters were PCR- amplified from existing vectors and subcloned into target plasmids by Gibson assembly method. These plasmids were used to electroporate SYN001, SYN469 or other parental host strains with extra integrated *brnQ* cassette into the *malE/K* or *malP/T* locus. After the electroporation (Eporator, Eppendorf, 1.8-kV pulse, 1-mm gap length electro-cuvettes) the transformed cells were selected as colonies on LB agar (Sigma, L2897) containing proper antibiotics.

Table 10

ID	Strain #	Genotype	Description *	Induction Activity
1	SYN001	Control bacterium	N/A	N/A
2	SYN094	SYN001 with strep resistance	Control bacterium with strep resistance	N/A
3	SYN469	SYN001, $\Delta ilvC$, $\Delta leuE$,	Bacterium with deleted BCAA synthesis enzyme IlvC, leucine	N/A

		<i>lacZ::P_{tet}-livKHPMF</i>	exporter LeuE, and insertion of BCAA importer LivKHPMF under the control of P _{tet} promoter	
4	SYN1980	SYN001, $\Delta ilvC$, $\Delta leuE$, <i>lacZ::P_{tet}-livKHPMF</i> , pSC101-P _{tet} - <i>leuDH-kivd-adh2-brnQ-ampR</i>	Bacterium with plasmid containing original BCAA consumption operon under the control of P _{tet} promoter	aTc
5	SYN5721- SYN5744	SYN001, $\Delta ilvC$, $\Delta leuE$, <i>lacZ::P_{tet}-livKHPMF</i> , pSC101-P _{tet} - <i>optimized operon-brnQ-ampR</i>	Bacterium with plasmid containing optimized BCAA consumption operon under the control of P _{tet} promoter	aTc
6	SYN5941- SYN5943	SYN001, pSC101-P _{tet} - <i>optimized operon-brnQ-ampR</i>	Bacterium with plasmid containing optimized BCAA consumption operon under the control of P _{tet} promoter	aTc
7	SYN5944- SYN5946	SYN001, $\Delta ilvC$, $\Delta leuE$, <i>lacZ::P_{tet}-livKHPMF</i> , <i>malP/T::P_{lacI/fnr}-brnQ</i> , pSC101-P _{tet} - <i>optimized operon-brnQ-ampR</i>	Bacterium with plasmid containing optimized BCAA consumption operon under the control of P _{tet} promoter and extra integrated BCAA importer BrnQ under the control of P _{lacI/fnr} promoter	aTc, anaerobic, IPTG
8	SYN5947- SYN5949	SYN001, $\Delta ilvC$, $\Delta leuE$, <i>lacZ::P_{tet}-livKHPMF</i> , <i>malE/K::P_{cmt}-brnQ</i> , pSC101-P _{tet} - <i>optimized operon-brnQ-ampR</i>	Bacterium with plasmid containing optimized BCAA consumption operon under the control of P _{tet} promoter and extra integrated BCAA importer BrnQ under the control of P _{cmt} promoter	aTc, cumate
9	SYN5950- SYN5952	SYN001, $\Delta ilvC$, $\Delta leuE$, <i>lacZ::P_{tet}-livKHPMF</i> , <i>malE/K::P_{cmt}-brnQ</i> , <i>malP/T::P_{cmt}-brnQ</i> , pSC101-P _{tet} - <i>optimized operon-brnQ-ampR</i>	Bacterium with plasmid containing optimized BCAA consumption operon under the control of P _{tet} promoter and two extra integrated BCAA importer BrnQ under the control of P _{cmt} promoter	aTc, cumate
10	SYN6034	SYN001, pSC101-P _{tet} - <i>leuDH-kivd-adh2-brnQ-ampR</i>	Bacterium with plasmid containing original BCAA consumption operon under the control of P _{tet} promoter	aTc
11	SYN6035	SYN001, pSC101-P _{cmt} - <i>optimized operon-brnQ-ampR</i>	Bacterium with plasmid containing optimized BCAA consumption operon under the control of P _{cmt} promoter	cumate
12	SYN6036	SYN001, pSC101-P _{lacI/fnr} - <i>optimized operon-brnQ-ampR</i>	Bacterium with plasmid containing optimized BCAA consumption operon under the control of P _{lacI/fnr} promoter	IPTG, anaerobic
13	SYN6037	SYN001, pSC101-P _{cl857} - <i>optimized operon-brnQ-ampR</i>	Bacterium with plasmid containing optimized BCAA consumption operon under the control of P _{cl857} promoter	heat
14	SYN6038	SYN001, pSC101-	Bacterium with plasmid containing	arabinose

		P_{BAD} -optimized operon- <i>brnQ-ampR</i>	optimized BCAA consumption operon under the control of P_{BAD} promoter	

*Original BCAA consumption operon refers to pSC101- P_{tet} -*leuDH-kivd-adh2-brnQ-ampR*.
Optimized BCAA consumption operon refers to the operons generated in the present application.

Example 2. Design and Synthesis of LeuDH, KivD, and Adh Enzyme Libraries

[0536] The metagenomic enzyme discovery and rational enzyme design described above were used to design a library for each enzyme family (LeuDH, KivD, and Adh) to identify LeuDH, KivD, and Adh pathway enzymes (i) with increased activity than the parent pathway enzymes from SYN1980: *Bacillus cereus* LeuDH, *Lactococcus lactis* KivD, and *Saccharomyces cerevisiae* ADH2, respectively; and (ii) with increased specificity for leucine (Leu) relative to valine (Val) and isoleucine (Ile). **Table 11** summarizes the enzyme library composition. For each enzyme, a metagenomic library of >1000 enzymes was designed to sample the full metagenomic sequence space available in sequence databases and depicted in **Figure 1**. For LeuDH and Adh, additional designs were constructed using rational design based on available structural data for the *B. cereus* LeuDH and *S. cerevisiae* Adh enzymes. Enzyme sequences for all libraries were optimized for expression in *E. coli*, synthesized in an inducible *E. coli* expression vector, and transformed into *E. coli* for high throughput screening.

Table 11

Library	Bacteria	Fungi	Animal	Plant	Rational	Total Designs
LeuDH	1129	11	23	12	270	1445
KivD	783	508	1	4	0	1296
Adh	654	273	128	122	140	1317

Example 3. Characterization of Pathway Enzyme Libraries

[0537] To screen the 3 x ~1300-member enzyme libraries, high-throughput methods were developed to screen for LeuDH, KivD, and Adh enzyme activities in *E. coli* cell lysates. In brief, strains were cultivated in 96-deepwell plates to induce protein production, with positive and negative control strains included in each plate. Cells were lysed, and enzyme activity was measured in cell lysates using enzyme-specific spectrophotometric assays described above in the method section. Enzyme assays were executed on the fully automated robotic workcell known as ATLAS.

[0538] For each enzyme family, the full library (~1300 members each) was measured in biological duplicate, and hits were selected based on the highest activity in a single replicate for each enzyme. This method of hit selection reduced the impact of false negative activity measurements for a given replicate. For each library, 55-220 enzymes were selected as primary hits. To increase the

confidence in the hit selected from the primary screen and establish enzyme ranking based on activity, the primary hits were re-screened in a secondary screen with additional replication (8 technical replicates).

Leucine dehydrogenase (*leuDH*)

[0539] A total of 1378 LeuDH enzymes were first screened for the ability to deaminate leucine. An initial round of screening identified 220 enzymes with activity similar to or better than the parent LeuDH enzyme from *B. subtilis*. These primary hits were further analyzed in a secondary screen for validation of LeuDH enzyme hits (**Figure 2**). In the secondary screen, LeuDH enzymes with up to a 1.8-fold increase in LeuDH activity on Leu were validated.

[0540] To determine if any of the primary LeuDH hits exhibited increased specificity for Leu over Val and Ile, all 220 primary hits were also screened for activity on Val and Ile. Specificity was measured as the ratio of activity on Leu to the activity on Ile or Val. As shown in **Figure 3**, enzymes exhibited up to ~2.7-fold preference for Leu over Val, and up to a 5-fold preference for Leu over Ile. By comparing specificity for Leu/Ile to Leu/Val, hits were identified with increased specificity for Leu relative to both Leu and Val (**Figure 4**), as compared with the control *B. cereus* LeuDH from the prototype pathway SYN1980.

Ketoisovalerate decarboxylase (*kivD*)

[0541] A total of 1248 KivD enzymes were screened for the decarboxylase activity on ketoisocaproate. An initial round of screening identified 55 enzymes with higher activity than the control KivD enzyme from *S. aureus*, which did not exhibit activity above greater than the background lysate decarboxylase activity in this assay and was equated to the non-zero measurable background activity. These primary KivD hits were further analyzed in a secondary screen (**Figure 5**). In the secondary screen, more than 40 KivD enzymes were identified with at least 6- to 8-fold increase in KivD activity relative to the background lysate activity.

Alcohol dehydrogenase (*adh*)

[0542] A total of 1215 Adh enzymes were screened for the ability to reduce isovaleraldehyde to isopentanol. An initial round of screening identified 55 enzymes with higher activity than the parent ADH2 enzyme from *S. cerevisiae*, which did not exhibit activity greater than the background lysate alcohol dehydrogenase activity in this assay and was equated to the non-zero measurable background activity. Because activity of the ADH2 enzyme for *S. cerevisiae* was indistinguishable from the background activity of the lysate, an *Equus caballus* Adh was used as a positive control for the screen. These primary Adh hits were further analyzed in a secondary screen (**Figure 6**). In the secondary screen, 5 Adh enzymes were identified with at least 20-fold increase in Adh activity relative to the background lysate activity.

Example 4. Selectivity of Top LeuDH Candidate Enzymes for Optimized BCAA Consumption

[0543] In an attempt to better predict the performance of the top LeuDH hits with regard to mixed -substrates pools, the selectivity of LeuDH enzymes for Leu (*i.e.* the preference of LeuDH for Leu when Leu, Val, and Ile are all present in the reaction mixture) were measured. A total of 21 LeuDH enzymes were screened in cell lysate assays similar to the HTP screen, except that the reaction mixture contained Leu, Val, and Ile at 1:1:1 molar ratio. Rate of Leu, Val, and Ile disappearance was monitored in the reaction mixture. **Figure 7** reports consumption of Leu, Ile, and Val within the reaction mixture for each LeuDH enzyme. At least 10 LeuDH enzymes showed improved preference for Leu over Val and Ile when compared to the parent *B. subtilis* LeuDH. For nearly all LeuDH enzymes, least preference was shown for valine.

Example 5. Pathway Enzyme Hit Selection, Operon Assembly and Operon Screening for Optimized BCAA Consumption

MATERIALS AND METHODS

Cell preparation

[0544] Branched-chain amino acid (BCAA) pathway operon plasmids were transformed into *E. coli* Nissle strain 1917, which was purchased from the German Collection of Microorganisms and Cell Cultures (DSMZ Braunschweig, *E. coli* DSM 6601). Transformed cells were thawed on ice and cell density was measured by light absorption at 600 nm (OD600). OD600 of 1.0 was assumed to be equal to 10^9 cells/mL in this method. A volume was calculated to target 1 mL of 2×10^9 cells/mL cell resuspension, and the cells were transferred into a 96-deep well plate and washed once with cold PBS. After centrifugation (4000 rpm, 4°C, 10 min), the PBS was discarded, and the cell pellets were then resuspended in 1 mL of 1x M9 + 50 mM MOPS + 0.5% glucose (MMG) buffer. Eight hundred (800) μ L of each sample was transferred into a new 96-deep well plate and 800 μ L of MMG containing 16 mM leucine was added, mixed well by pipetting. A sample (200 μ L) assigned as time zero was collected at this moment. The plate was then covered by a breathable membrane and moved to an anaerobic chamber to incubate at 37 °C. Samples were also collected at 2 hours and 4 hours during incubation in the anaerobic chamber. The samples were centrifuged for 10 minutes at 4000 rpm at 4 °C immediately after collection. 100 μ L of the supernatant was transferred into a new 96-well plate and stored at -80 °C for future analysis.

Leucine Activity Assay

[0545] Leucine was quantitated in bacterial supernatant by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) using either an Ultimate 3000 UHPLC-TSQ Quantum or a Vanquish UHPLC-TSQ Altis system. Samples were extracted with 9 parts 2:1 acetonitrile:water containing 1 µg/mL leucine-d3 as an internal standard, vortexed, and centrifuged. Supernatants were diluted with 9 parts 0.1% formic acid and analyzed concurrently with standards processed as above from 0.8 to 1000 µg/mL. Samples were separated on a Phenomenex Synergi 4 µm Hydro-RP 80A, 75 x 2 mm using a 0.1% formic acid (A), 0.1% formic acid/acetonitrile (B) at 0.3 mL/min and 50 degrees C. After a 2 µL injection and an initial 5% B hold from 0 to 0.5 minutes, analytes were gradient eluted from 5 to 90% B over 0.5 to 1.5 minutes followed by high organic wash and aqueous equilibration steps. Analytes were detected using Selected Reacting Monitoring (SRM) of compound specific collision induced fragments in electrospray positive ion mode (leucine: 132>86, isoleucine: leucine-d3: 135>89). SRM chromatograms were integrated, and the unknown/internal standard peak area ratios were used to calculate concentrations against the standard curve.

[0546] To improve the overall Leu consumption of the BCAA pathway, multiple enzymes were selected for each step that demonstrated superior performance relative to the *B. cereus* LeuDH enzyme. For LeuDH, 6 hits were selected based on two criteria: (1) enzyme activity on Leu and (2) specificity for Leu relative to Val and Ile. Because LeuDH selectivity analysis was conducted in parallel to operon assembly, the selectivity data set did not factor into LeuDH selection. For KivD and ADH, 3 hits were selected for each enzyme family based on *in vitro* enzyme activity. In total, 12 enzymes were advanced to the final operon design.

[0547] To balance the flux of the Leu-consuming pathway, enzyme identity and enzyme level for each enzyme in the Leu-consuming pathway were varied. Three (3) ribosome binding sites (RBSs) were selected for each pathway enzyme candidate. These RBS-enzyme pairs were then assembled into a partial combinatorial library of 462 pathways, and the gene order was held constant (LeuDH-KivD-Adh), as shown in **Figure 8**. Pathways were synthesized and cloned into a MSUD pathway backbone encoding a branched chain amino acid transporter (BrnQ). The final pathways were composed of four-gene operons (LeuDH-KivD-Adh-BrnQ) in which the LeuDH, KivD, and Adh enzymes and their RBSs were varied while the BrnQ RBS and coding sequence is held constant.

[0548] The operon library was screened in HTP to identify optimal pathway designs for Leu consumption in SYN469 (*AilvC*, *AleuE*, *lacZ::Ptet-livKHMGE*). Of the 462 operons designed, 383 were successfully synthesized and 254 were successfully transformed into SYN469 for screening. All pathway strains were cultured and screened for Leu consumption in 96-well plates. For plate-to-plate

standardization, each plate included control strains which harbored the prototype pathway SYN1980 or versions lacking LeuDH (SYN1980 Δ LeuDH) or BrnQ (SYN1980 Δ BrnQ). As shown in **Figure 9**, multiple pathways were identified that Leu consumption and isopentanol production each occurred at a faster rate than the prototype pathway SYN1980. Strains were ranked based on Leu consumption, and 24 strains were validated in the flask-based model described above in the method section.

Analysis of the enzyme composition of the top 24 pathway operons showed that 4 of 6 LeuDHs and all 3 Adh enzymes were represented in the top 24 operons. However, 23 of the 24 operons contained the same KivD, supporting a correlation between this KivD enzyme and the overall pathway activity.

RESULTS

[0549] The top Leu consuming operons identified through HTP screening were transformed into *E. coli* Nissle 1917 (and labeled as strain 5941, 5942 and 5943) and compared to the prototype strain 1980. Strain 5941 contains the LeuDH enzyme of *Cetobacterium cети*, the KivD enzyme of *Erwinia iniecta*, and the Adh enzyme of *Alcanivorax dieselolei*. Strain 5942 has the LeuDH enzyme of *Cetobacterium cети*, the KivD enzyme of *Erwinia iniecta*, and the Adh enzyme of *Rhizobiales bacterium* NRL2. Strain 5943 has LeuDH enzyme of *Cetobacterium cети*, the KivD enzyme of *Erwinia iniecta*, and the Adh enzyme of *Rhizobiales bacterium* NRL2. The operons further contain BrnQ of *E. coli*. The prototype strain contains *Bacillus cereus* LeuDH, *Lactococcus lactis* KivD, *Saccharomyces cerevisiae* ADH2, as well as *E. coli* BrnQ.

[0550] Samples from the top Leu consuming operons and the prototype strain were analyzed for Leu consumption (FIG. 20). The top Leu consuming operon-containing strains (5941, 5942 and 5943) were found to consume Leu at a significantly faster rate than the prototype strain (1980).

Example 6. Metabolomics Analysis of Top Pathways for Optimized BCAA Consumption

[0551] The top 10 Leu consuming strains identified through HTP screening described above in the method section shown in **Figure 9** were selected for targeted metabolomic analysis of pathway intermediates. In brief, strains harboring optimized or controlled plasmids were grown and induced in a flask-based model in biological triplicate. Cells were harvested, normalized individually by OD₆₀₀, and prepared for the Leu-consumption assay. The assay was initiated by addition of cells to the Leu assay buffer, and reactions were sampled over time and preserved for analysis. Samples were analyzed for Leu, ketoisocaproate (KIC), and isovaleraldehyde (**Figure 10**). Isopentanol measurements were not collected for this experiment. Of the 10 selected pathways, 9 consumed Leu at a faster rate than the prototype strain (SYN1980) and the SYN1980 Δ BrnQ variant (*i.e.* SYN1992). The 9 better performing strains had drastically reduced levels of KIC when compared the control strains, indicating that a KivD bottleneck was relieved through the pathway optimization campaign, providing further evidence that KivD is a bottle neck in the overall pathway. Of the top 9 strains, t214235 exhibited the fastest rate of Leu consumption and a downward IVA trend over time.

Example 7. *In vitro* Activity Validation for BCAA Consuming Strains Generated from Host Strain SYN469

[0552] Plasmids containing optimized operons were transformed into SYN469, resulting in strains SYN5721 to SYN5744. The *in vitro* leucine consumption assay was performed to assess their activity using SYN1980 as a control. As shown in **Figure 11**, all new strains demonstrated improved activity *in vitro* as compared to SYN1980. Top hits such as SYN5721, SYN5722 and SYN5723 were selected for further testing.

[0553] These selected strains were tested using BCAA mixture as substrates *in vitro*. The goal here was to understand the substrate specificity when all three branched chain amino acids were present simultaneously. The leucine: isoleucine: valine ratio was 2:1:1, corresponding to the ratio analyzed previously for *in vivo* GI content. Strains showed some specificity advantage towards leucine versus valine, but not isoleucine (**Table 12**).

Table 12

	SYN1980	SYN5721	SYN5722	SYN5723	SYN5725	SYN5729
Val (2 mM)	0.27	0.49	0.46	0.59	0.64	0.40
Ile (2 mM)	0.31	0.82	0.78	0.59	0.69	0.67
Leu (4 mM)	0.59	1.51	1.43	1.40	1.18	1.25
BCAA	1.17	2.82	2.66	2.58	2.51	2.32

Example 8. Host Strain Screening Study and Promoter Screening Study to Screen for Optimized BCAA Consuming Strains

[0554] In addition to various modifications existed in SYN469 (*ΔilvC*, *ΔleuE*, *lacZ::P_{tet}-livKHMGF*), which was used for the construction of SYN1980 and SYN5721-SYN5744. A panel of new strains was built using the most active operons and different host strains. In these set of host strains, either modifications were removed such as using the wild-type strain SYN001 or extra modifications were introduced such as extra copy(s) of BCAA transporter BrnQ. The *in vitro* leucine consumption activity was measured. SYN5941 demonstrated the highest leucine consumption rate under this test condition (**Figure 12**).

[0555] A panel of selected promoters were chosen to drive the optimized operon and tested for *in vitro* leucine consumption. The operon in SYN1980 was used as a control. The host strain was SYN001 for testing the new promoters. All the clinical candidate strain compatible promoters—FNR_S, P_{cmf}, P_{cl857}, and P_{BAD}—showed reasonable *in vitro* leucine consumption activity, though lower than SYN5941 harvesting the P_{tet} driven operon (**Figure 13** and **Table 13**).

Table 13

Inducer	aTc	aTc	aTc	Cumate	IPTG(fnr)	Heat (37°C)	arabinose
Final Strain #	SYN1980	SYN6034	SYN5941	SYN6035	SYN6036	SYN6037	SYN6038
Activity (Leu)	0.48	0.29	3.08	1.54	1.20	0.91	0.89

($\mu\text{mol}/10^9$ cells/h)							
Host	SYN469	SYN001	SYN001	SYN001	SYN001	SYN001	SYN001
Operon source	1980	1980	5721	5721	5721	5721	5721

Example 9. *In vitro* Activity Study for Optimized BCAA Consuming Strain SYN 5941 in a Fermentation Scale-up Process

[0556] Various parameters were tested such as, but not limited to, carbon source, feeding rate, temperature, time of induction and inducer concentration. Using SYN5941, FM1 media with glucose and 6x anhydrotetracycline (aTc) concentration (600 $\mu\text{g}/\text{mL}$) was chosen as the growth condition and cells were harvested 4 hours post induction from the fermenter for cryopreservation. A greater-than-ten-fold improvement in *in vitro* leucine consumption was observed using the optimized strain SYN5941 with newly developed fermentation process (**Table 14**). As shown in **Table 15**, SYN5941 exhibited a much higher leucine consumption rate than SYN1980.

Table 14

Media	Carbon source	Induction OD	aTc (1x = 100 $\mu\text{g}/\text{mL}$)	Feed	Temp
FM1	25g/L glucose	1.5,10,20	0.1,1,3,6,10X	BCAAs (hi/lo rates)	Shift to 30°C mid-induction
FM5	None	1.5,10	1X	YE+glycerol, YE+glucose, BCAA+glycerol	NA
FM2	40g/L glycerol, 25g/L glucose	1.5,10	1X	NA	Shift to 30°C mid-induction
FM4	25g/L glucose	1.5	1X	BCAAs (hi/lo rates)	NA

FM1: 25g/L YE, K-Phosphate buffer, MgSO_4 , trace metals, ammonium phosphate, iron chloride.

FM2: 24g/L YE, 12g/L soytone, K-Phosphate buffer.

FM5: 8g/L YE, K-Phosphate buffer, MgSO_4 , trace metals, ammonium phosphate, iron chloride.

FM4: K-phosphate buffer, MgSO_4 , trace metals, citric acid, ammonium sulfate, iron citrate.

Table 15

STRAIN#	SYN5941	SYN1980
Final Batch#	SYN5941BR12-13	SYN1980BR16-17
Activity ($\mu\text{mol}/10^9\text{CFU/h}$)	2.69	0.17

Example 10. *In vitro* Activity Study for Optimized BCAA Consuming Strain SYN 5941 in a Gastrointestinal Simulation System

[0557] The *in vitro* activity of the optimized strain, SYN5941, was compared to the that of the original variant, SYN1980, using the *in vitro* gastrointestinal simulation (IVS) model described above in the method section. **Figure 14A** displays representative data concerning leucine consumption (left) and isopentanol production (right) by these two strains during two hours'

incubation in simulated gastric fluid with bicarbonate buffering. The oxygen concentration in this assay was decreased gradually from 7% to 4% oxygen, mimicking the low oxygen conditions of the human upper gastrointestinal tract. SYN5941 displayed increased leucine consumption and isopentanol production in IVS, as compared to SYN1980. Notably, this increased leucine consumption was not attributable to fermentation process changes, as SYN1980 prepared with the optimized process demonstrated comparable activity to the SYN1980 prepared with the base process. The consumption specificity towards leucine only or leucine, isoleucine and valine mixture as the substrate(s) were tested in the IVS system as well. As demonstrated in **Figure 14B**, SYN5941 had the highest activity towards leucine when leucine, isoleucine and valine were present in equal amount of 5 mM. The overall leucine consumption was also similar to the reaction system when only leucine was present.

[0558] Leucine consumption rates and isopentanol production rates were calculated based on the data points collected between time zero and the first 2 hours. As shown in **Table 16**, the rate of leucine consumption was greater for SYN5941 than for SYN1980 (0.85 ± 0.04 versus 0.25 ± 0.12 $\mu\text{mol}/10^9\text{CFU/hr}$). The same trend can be said for isopentanol production (0.45 ± 0.01 versus 0.13 ± 0.02 $\mu\text{mol}/10^9\text{CFU/hr}$).

Table 16

Strain	MSUD Fermentation Process	Leu Consumption Rate ($\mu\text{mol}/10^9$ CFU/h)		Isopentanol Production Rate ($\mu\text{mol}/10^9$ CFU/hr)	
		Mean	StDev	Mean	StDev
SYN5941	Optimized	0.85	0.04	0.45	0.01
SYN1980	Optimized	0.16	0.03	0.04	0.00
SYN1980	Original	0.25	0.12	0.13	0.02
EcN	NA	0.02	0.03	0.00	0.00

Example 11. *In vitro* Characterization of SYN5941

[0559] The *in vitro* activity of the optimized strain, SYN5941, was characterized in various conditions. **Figure 15** depicts leucine consumption rate of SYN5941 in increase leucine concentration (**Figure 15A**), increasing pH (**Figure 15B**), and increasing oxygen concentration (**Figure 15C**). Various substrate concentrations (0, 0.63, 1.25, 5, 10, 15, or 20 mM leucine) were added into the reaction media and the leucine consumption rate was measured for each concentration, respectively. The leucine concentration was measured by HPLC at different time points to obtain the rates. The leucine consumption rates were plotted against the substrate concentration and fitted by a hyperbolic curve. SYN5941 was considered as one enzyme molecule, then the corresponding values of the Ymax and the X value to reach $\frac{1}{2}$ Ymax can be obtained to use as Vmax and Km for SYN5941 during catalysis of leucine degradation. Similar assays were done a fixed substrate concentration (10 mM) but varied the media pH (3, 4, 5, 6, or 7) and oxygen level (0%, 7%, or 21%) to investigate the

pH and oxygen dependence. The optimal strain activity is at pH 7. However, it is relatively insensitive to oxygen level which is consistent with the original reason to choose this redox neutral pathway.

Example 12. *In vivo* Efficacy Study for Optimized BCAA Consuming Strain SYN5941 in iMSUD Mice

[0560] Two independent studies were performed to compare the efficacy of SYN5941 against SYN1980 in a mouse model of MSUD (iMSUD) provided with a high-protein diet. In the first study (**Figure 16**), the leucine levels in the high-protein diet were increased to $1647 \pm 265.8 \mu\text{M}$ in iMSUD males and $1165 \pm 151.9 \mu\text{M}$ in iMSUD females. SYN5941 lowered leucine in males to $843.4 \pm 113.4 \mu\text{M}$, which was more significant than SYN1980 (**Figure 16A**). The high protein diet also increased the levels of isoleucine and valine in males and females but similar to leucine, a more significant effect was observed with SYN5941 in males (**Figure 16B-C**).

[0561] In the second study (**Figure 17**), the leucine levels in the high-protein diet were further increased to $2428 \pm 485.1 \mu\text{M}$ in iMSUD males and $1626 \pm 150.1 \mu\text{M}$ in iMSUD females. In contrast to the first study, neither SYN5941 nor SYN1980 significantly lowered leucine levels in either males or females (**Figure 17A**). The high protein diet also increased the levels of isoleucine and valine in both males and females but similar to leucine, the bacteria had no significant effect on the levels of these amino acids (**Figure 17B-C**).

Example 13. *In vivo* Efficacy Study for Optimized BCAA Consuming Strain SYN5941 in Healthy Non-Human Primates

[0562] To assess the efficacy of SYN5941 and SYN1980 in healthy non-human primates, monkeys received a single dose of vehicle or bacteria concomitantly with the meat digest peptone. Branched chain amino acid levels leucine, isoleucine and valine peaked approximately 1 hour post-peptone administration in vehicle-treated animals to $99.7 \pm 16.5 \mu\text{M}$, $47.7 \pm 9.3 \mu\text{M}$ and $123.2 \pm 21.4 \mu\text{M}$, respectively (**Figure 18A-18C**). SYN5941 lowered the levels of leucine more significantly than SYN1908 (**Figure 18**).

[0563] **Table 17** shows that SYN5941 exhibited significantly higher leucine consumption activities than SYN1980 in all three systems demonstrated above in Examples 9, 10 and 13.

Table 17

Strain	<i>In vitro</i> activity	IVS	NHP
SYN1980*	0.17	0.16	327.3
SYN5941**	2.69	0.85	133.5
UNIT	($\mu\text{mol}/10^9\text{CFU/h}$)	($\mu\text{mol}/10^9\text{CFU/h}$)	(area under curve)
Fold of improvement	15.8	5.3	59.2% decrease

*Original operon in strain SYN1980:

leuDH (*Bacillus cereus*) – kivD (*Lactococcus lactis*) – adh (*Saccharomyces cerevisiae*)

Original chassis for strain SYN1980: SYN469 = SYN001, ΔilvC , ΔleuE , lacZ::Ptet-livKHMGE

**Optimized operon in new strain SYN5941:

leuDH (*Cetobacterium Ceti*) – kivD (*Erwinia iniecta*) – adh (*Alcanivorax dieselolei*)

Optimized chassis for new strain SYN5941: SYN094 = SYN001, strpR

Example 14. Molar Balance Closure of the Isopentanol Pathway

[0564] The performance and molar balance closure of the isopentanol pathway in SYN5941 was assessed in Ambr15 bioreactors. The reactors were filled to 17 mL with M9 media with 0.5% glucose, 10 mM Leu, 10 mM Val, and 5 mM Ile. Conditions were controlled with 0% dissolved oxygen and pH at 7.0. Activated biomass was inoculated to an OD600 of 1, and samples of the supernatant were taken over time to monitor metabolite concentrations.

[0565] The extracellular concentration profiles of pathway intermediates are shown in the Figure below. Over the course of 180 minutes, 4.1 ± 0.3 mM of Leucine was consumed and 4.4 ± 0.5 mM of isopentanol accumulated in the media. The keto-acid (2-oxoisocaproate) and aldehyde (isovaleraldehyde) were not observed in the supernatant. Thus, the flux through the pathway is balanced and accounted for. This is also shown by the conservation of total moles of the pathway intermediates (**Figure 21**).

Methods - Fermentation

[0566] The assay was performed in an AMBR15f, microbioreactor system from Sartorius. The vessels were filled with 17mls of 1x m9 media salts, supplemented with 2.0mM MgSO₄, 0.1mM CaCl₂, 5% glucose, 10mM L-leucine, 5mM L-isoleucine, and 10mM valine. The vessels were filled 18hrs prior to inoculation, to enable both the pH and DO optodes to hydrate. The temperature in the reactors was kept at 37c, the pH was maintained at 7 using 2N NaOH, and the dissolved oxygen was kept at 0 using a 0.14vvm N₂ flow rate. The agitation was set to 500RPM to enable good mixing throughout the experiment. The bioreactors were inoculated to an OD600 of 1, from activated biomass supplied by Synlogic. The bioreactors were sampled at 0, 30, 90, 150, and 180 minutes post inoculation. Samples were immediately centrifuged at 15000xg for 30secs in a microcentrifuge and the supernatant was removed for analysis. Supernatants were stored at -20°C until ready for analysis.

Methods - Analytics

[0567] Analytics were developed for two methods, one method towards liquid chromatography mass spectrometry (LCMS) for the quantification of leucine (Leu), ketoisocaproate acid (Leu acid), and isovaleraldehyde (Leu aldehyde). The method (data not shown), was also validated and used for quantification of valine and isoleucine (and there

respected acid and aldehyde products). While the second method developed for gas chromatography mass spectrometry (GCMS) for the quantification of isopentanol (Leu alcohol). Thus, completing the pathway for mass balance for SYN5941. The GCMS method (data not shown), was also validated and used for quantification of valine and isoleucine alcohol products.

[0568] LCMS analysis was performed on a Thermo Ultimate 3000 UPLC system with a Thermo Q-Exactive quadrupole-orbitrap mass detector and a Thermo Accucore PFP column (2.1x100mm, 2.6um packing) using the following elution solvents: A=0.1% formic acid and 0.1% TFA in water; B=0.1% formic acid in acetonitrile. The gradient was at 0.5mL/min of 1% B in A for 60 seconds, followed by a linear ramp from 1% to 40% B in A over 270 seconds. The column is then flushed with 95% B in A for 60 seconds, and re-equilibrated with 1% B in A for 180 seconds. MS acquisition was from 0.8 to 5.3 minutes.

[0569] Column effluent is introduced into the mass spectrometer via a standard Thermo ESI source with positive mode ionization at +3800V, vaporizer temperature of 400°C, and ion transfer tube temperature of 375°C. Thermo reports gas flow rates in arbitrary units probably approximating L/min at STP. Set points were: sheath gas, 60; aux gas, 30; sweep gas, 1. To increase data acquisition rate, orbitrap resolution was set to 17,500. Quadrupole resolution was 1 m/z.

[0570] This method also derivatizes both aldehydes and keto acids, improving the stability of those analytes. Numerous derivatizing agents were explored, and it was found that 2-(Dimethylamino)ethylhydrazine in methanol resulted in the best sensitivity in positive mode. A buffer of 0.5M acetic acid and 0.5M sodium acetate in methanol was used for the quantification of LEU ACID and LEU ALDEHYDE, while also measuring non-derivatized LEU.

[0571] GC-MS analysis was performed on an Agilent GCMS/MSD with a Gerstel autosampler, using a J&W DB-WAX GC Column (15m) and chloroform as the extraction solvent. Front injector was set at 250°C and a flow rate of 1 mL/min. The oven temperature held at 40°C for 1 minute, followed by a ramp to 130°C (15°C/min), and then ramped up to 200°C (65°C/min). Ms acquisition scan window was at 40-150 m/z, with the MS source and MS quad at 250°C and 200°C respectively.

[0572] Finally to facilitate high throughput and automation, a Gerstel autosampler was used to inject the extracted bottom chloroform layer in a 96 well plate format with the aqueous ambr15 culture matrix on top acting as an overlay to prevent product evaporation.

To account for any other potential alcohol product evaporation, 2-heptanol was added to the chloroform as an internal.

CLAIMS

1. A recombinant bacterium capable of consuming leucine at a rate of at least about 0.5 $\mu\text{mol}/10^9 \text{ CFU/h}$ *in vitro*.
2. The recombinant bacterium of claim 1, wherein the bacterium is capable of consuming leucine at a rate of at least about 0.75, 1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, or 2.75 $\mu\text{mol}/10^9 \text{ CFU/h}$.
3. The recombinant bacterium of claim 1, wherein the bacterium is capable of consuming leucine at a rate of about 2.5 to about 2.75 $\mu\text{mol}/10^9 \text{ CFU/h}$.
4. The recombinant bacterium of any one of the previous claims, wherein the bacterium comprises at least one gene sequence encoding at least one leucine catabolism enzyme operably linked to at least one directly or indirectly inducible promoter that is not associated with a gene encoding the at least one leucine catabolism enzyme in nature.
5. The recombinant bacterium of claim 4, wherein the promoter is selected from the group consisting of an FNRS promoter, a P_{tet} promoter, a P_{cmt} promoter, an FNRS promoter, a P_{cl857} promoter, and a P_{BAD} promoter.
6. The recombinant bacterium of claim 4, wherein the at least one gene sequence encoding the at least one leucine catabolism enzyme is *leuDH*, *kivD*, and/or *adh2*.
7. The recombinant bacterium of claim 6, wherein the bacterium further comprises at least one gene sequence encoding at least one transporter capable of transporting leucine.
8. The recombinant bacterium of claim 7, wherein the at least one gene sequence encoding the at least one transporter is *brnQ*.
9. The recombinant bacterium of claim 8, wherein the bacterium comprises an operon, wherein the operon comprises *leuDH*, *kivD*, *adh2*, and *brnQ*.
10. The recombinant bacterium of claim 9, wherein the operon comprises a sequence having at least 90% identity to a sequence of any one of SEQ ID NOs: 25-48.

11. The recombinant bacterium of any one of claims 8-10,
wherein the *leuDH* gene encodes a LeuDH protein which comprises a sequence having at least 90% identity to a sequence of any one of SEQ ID NOs: 49-72;
wherein the *kivD* gene encodes a KivD protein which comprises a sequence having at least 90% identity to a sequence of any one of SEQ ID NOs: 73-96;
wherein the *adh2* gene encodes an Adh2 protein which comprises a sequence having at least 90% identity to a sequence of any one of SEQ ID NOs: 97-120; and/or
wherein the *brnQ* gene encodes a BrnQ protein which comprises a sequence having at least 90% identity to a sequence of any one of SEQ ID NOs: 121-144.
12. A recombinant bacterium comprising an operon,
wherein the operon comprises an optimized *leuDH* gene which encodes a LeuDH protein, an optimized *kivD* gene which encodes a KivD protein, and an optimized *adh2* gene which encodes an Adh2 protein, and an optimized *brnQ* gene which encodes a BrnQ protein,
wherein the operon comprises a sequence having at least 90% identity to a sequence of any one of SEQ ID NOs: 25-48, and
wherein the operon is operably linked to an inducible promoter that is not associated with a *leuDH* gene, a *kivD* gene, a *adh2* gene, or a *brnQ* gene in nature.
13. The recombinant bacterium of claim 12, wherein the LeuDH protein comprises a sequence having at least 90% identity to a sequence of any one of SEQ ID NOs: 49-72.
14. The recombinant bacterium of claim 12 or claim 13, wherein the KivD protein comprises a sequence having at least 90% identity to a sequence of any one of SEQ ID NOs: 73-96.
15. The recombinant bacterium of any one of claims 12-14, wherein the Adh2 protein comprises a sequence having at least 90% identity to a sequence of any one of SEQ ID NOs: 97-120.
16. The recombinant bacterium of any one of claims 12-15, wherein the BrnQ protein comprises a sequence having at least 90% identity to a sequence of any one of SEQ ID NOs: 121-144.
17. The recombinant bacterium of any one of claims 4-8, wherein the at least one gene sequence encoding the at least one leucine catabolism enzyme and the inducible promoter, are integrated into the chromosome of the bacterium.
18. The recombinant bacterium of any one of claims 9-16, wherein the operon and the inducible promoter are integrated into the chromosome of the bacterium.

19. The recombinant bacterium of any one of claims 8-18, wherein the bacterium further comprises a second *brnQ* gene which encodes a second BrnQ protein, wherein the second *brnQ* gene is operably linked to an inducible promoter that is not associated with a *brnQ* gene in nature, and wherein the second *brnQ* gene is inserted into the *malE/K* or *malP/T* locus on a chromosome of the bacterium.
20. The recombinant bacterium of any one of the previous claims, wherein the bacterium comprises a genetic modification in *leuE* that reduces leucine export from the bacterium.
21. The recombinant bacterium of any one of the previous claims, wherein the bacterium comprises a genetic modification in *ilvC* that reduces endogenous biosynthesis of leucine in the bacterium.
22. The recombinant bacterium of any one of the previous claims, wherein the bacterium further comprises a *livKHMGF* gene sequence encoding at least one transporter capable of transporting leucine.
23. The recombinant bacterium of claim 22, wherein the *livKHMGF* gene sequence is operably linked to at least one promoter that is not associated with a *livKHMGF* gene in nature.
24. The recombinant bacterium of any one of claims 12-23, wherein the bacterium is capable of consuming leucine at a rate of at least about 0.5 $\mu\text{mol}/10^9\text{CFU/h}$ *in vitro*.
25. The recombinant bacterium of any one of the previous claims, wherein the bacterium is capable of producing isopentanol at a rate of at least about 0.2 $\mu\text{mol}/10^9\text{CFU/h}$ *in vitro*.
26. The recombinant bacterium of claim 25, wherein the bacterium is capable of producing isopentanol at a rate of at least about 0.2 to 0.5 $\mu\text{mol}/10^9\text{CFU/h}$ *in vitro*.
27. The recombinant bacterium of any one of the previous claims, wherein the bacterium exhibits preferentially consumes leucine over valine and isoleucine.
28. The recombinant bacterium of claim 27, wherein the bacterium exhibits a leucine/valine activity ratio of at least about 1.1 to at least about 2.75.
29. The recombinant bacterium of claim 27 or claim 28, wherein the bacterium exhibits a leucine/isoleucine activity ratio of at least about 1.1 to at least about 5.

30. The recombinant bacterium of any one of the previous claims, wherein the bacterium is a probiotic bacterium selected from the group consisting of *Bacteroides*, *Bifidobacterium*, *Clostridium*, *Escherichia*, *Lactobacillus*, and *Lactococcus*.
31. The recombinant bacterium of claim 30, wherein the bacterium is *Escherichia coli* strain Nissle.
32. A pharmaceutically acceptable composition comprising the recombinant bacterium of any one of the previous claims and a pharmaceutically acceptable carrier.
33. A method of reducing the level of leucine in a subject, the method comprising a step of administering to the subject the pharmaceutically acceptable composition of claim 32.
34. A method of treating a disease associated with excess leucine and/or a metabolic disorder involving the abnormal catabolism of leucine in a subject, the method comprising a step of administering to the subject the pharmaceutically acceptable composition of claim 32.
35. The method of claim 33 or claim 34, wherein the subject has, or is suspected of having maple syrup urine disease (MSUD).

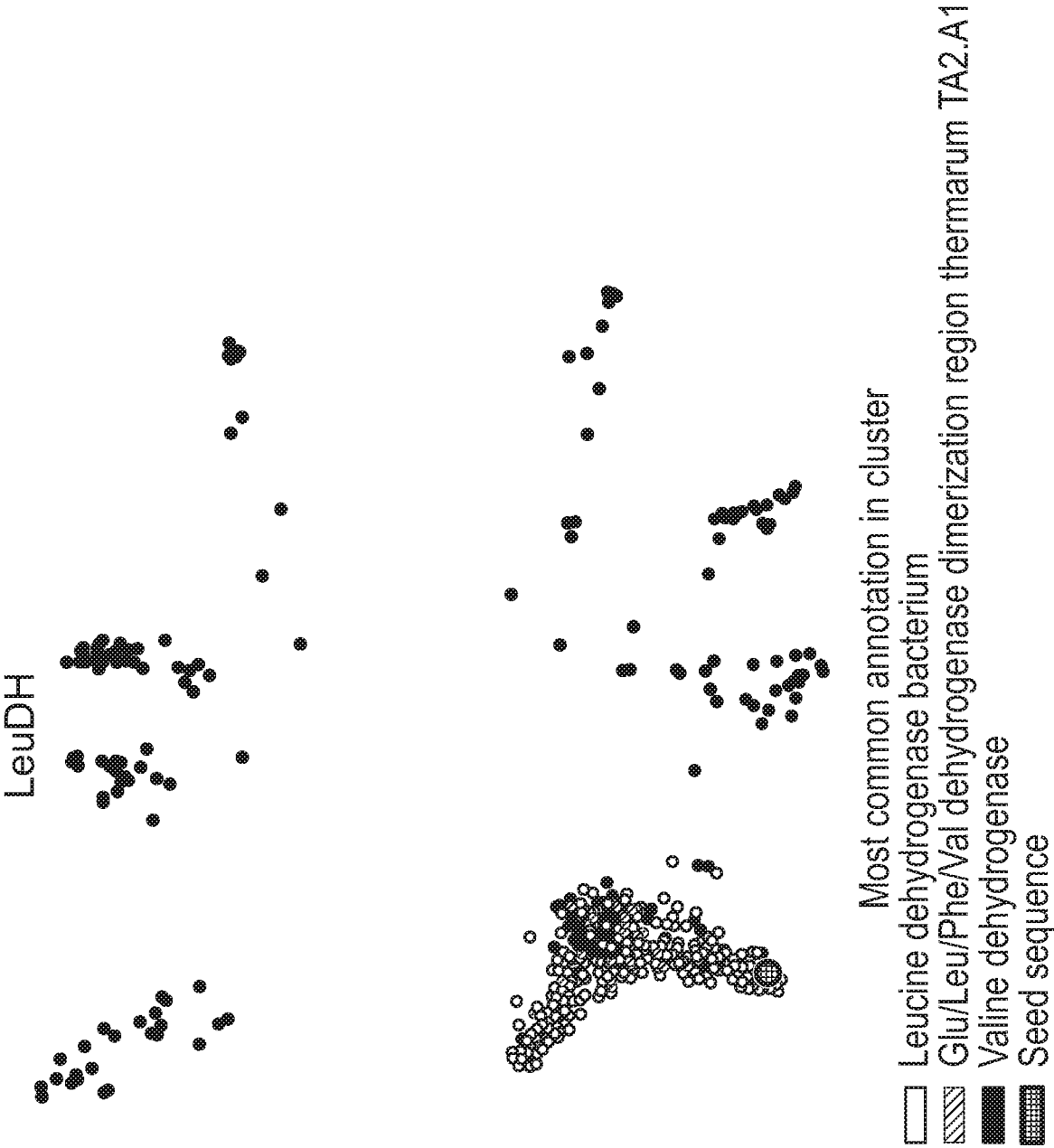


FIG. 1A

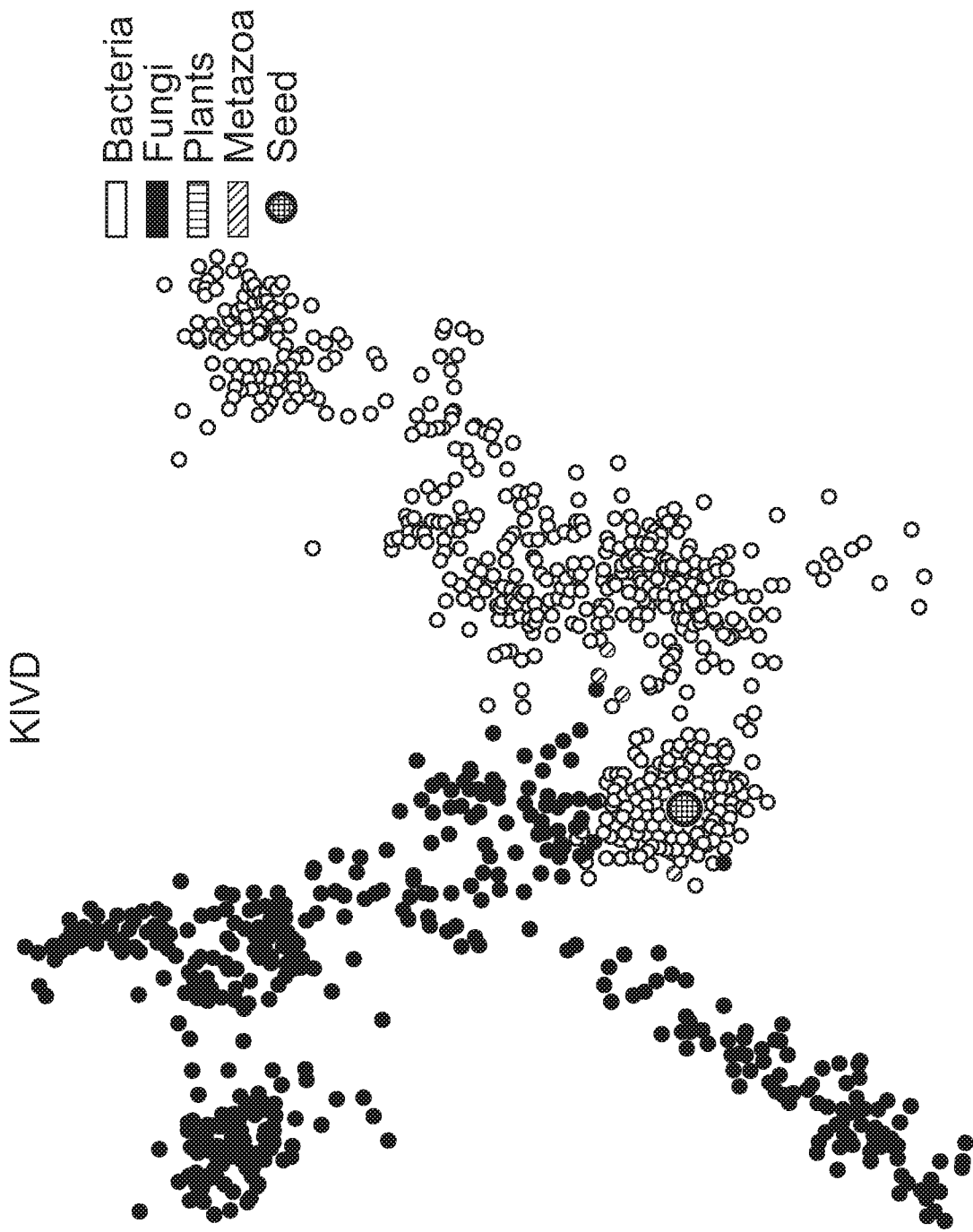
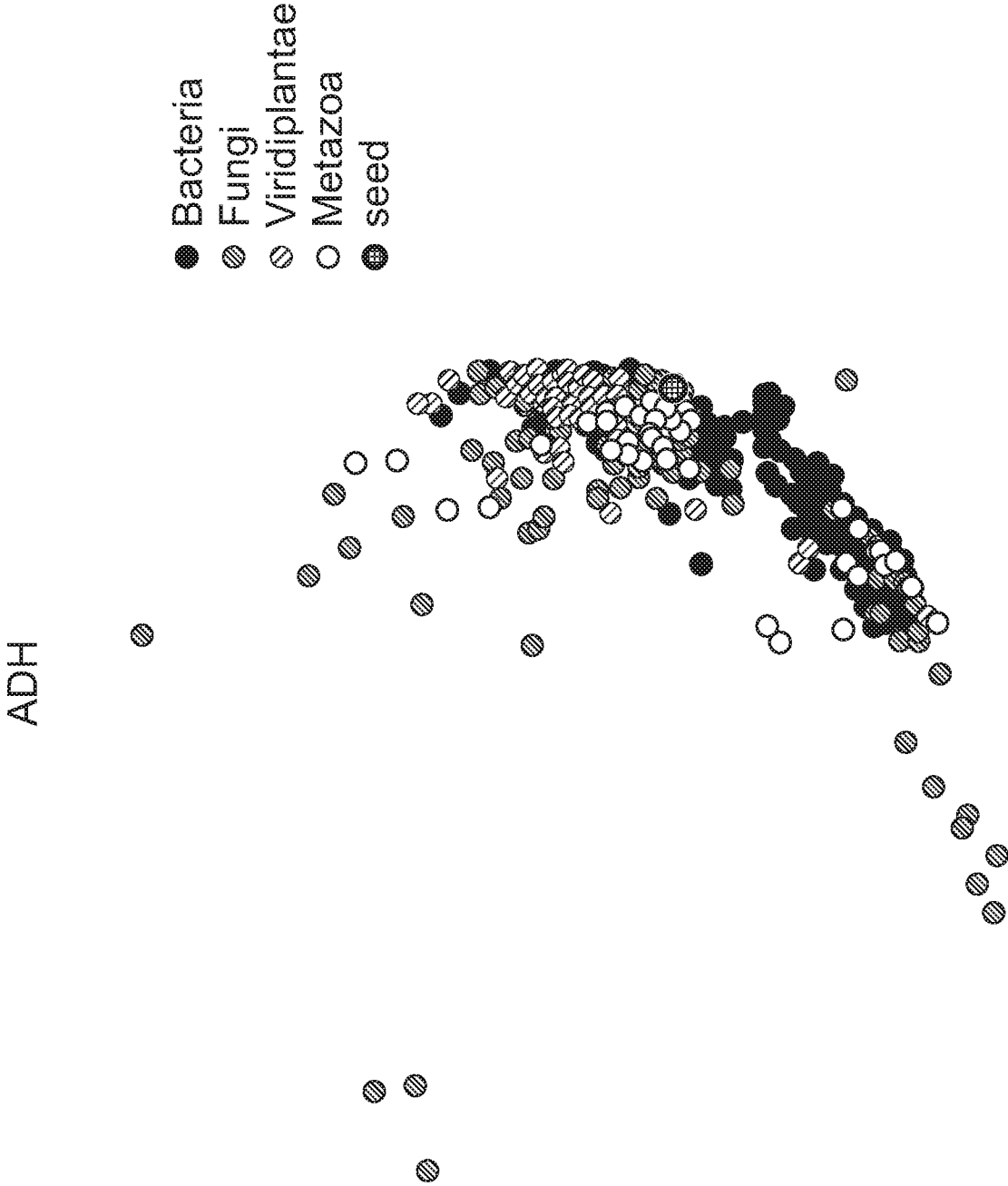


FIG. 1B



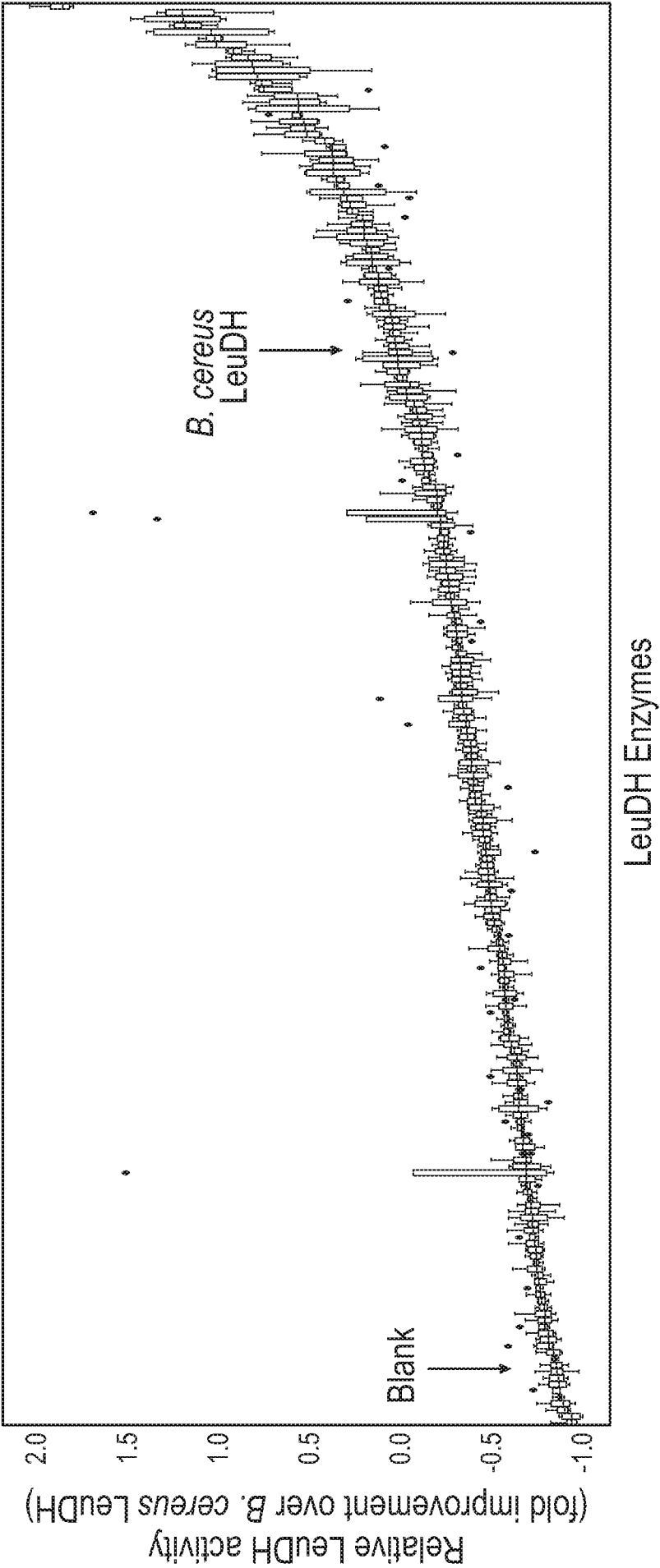


FIG. 2

5/23

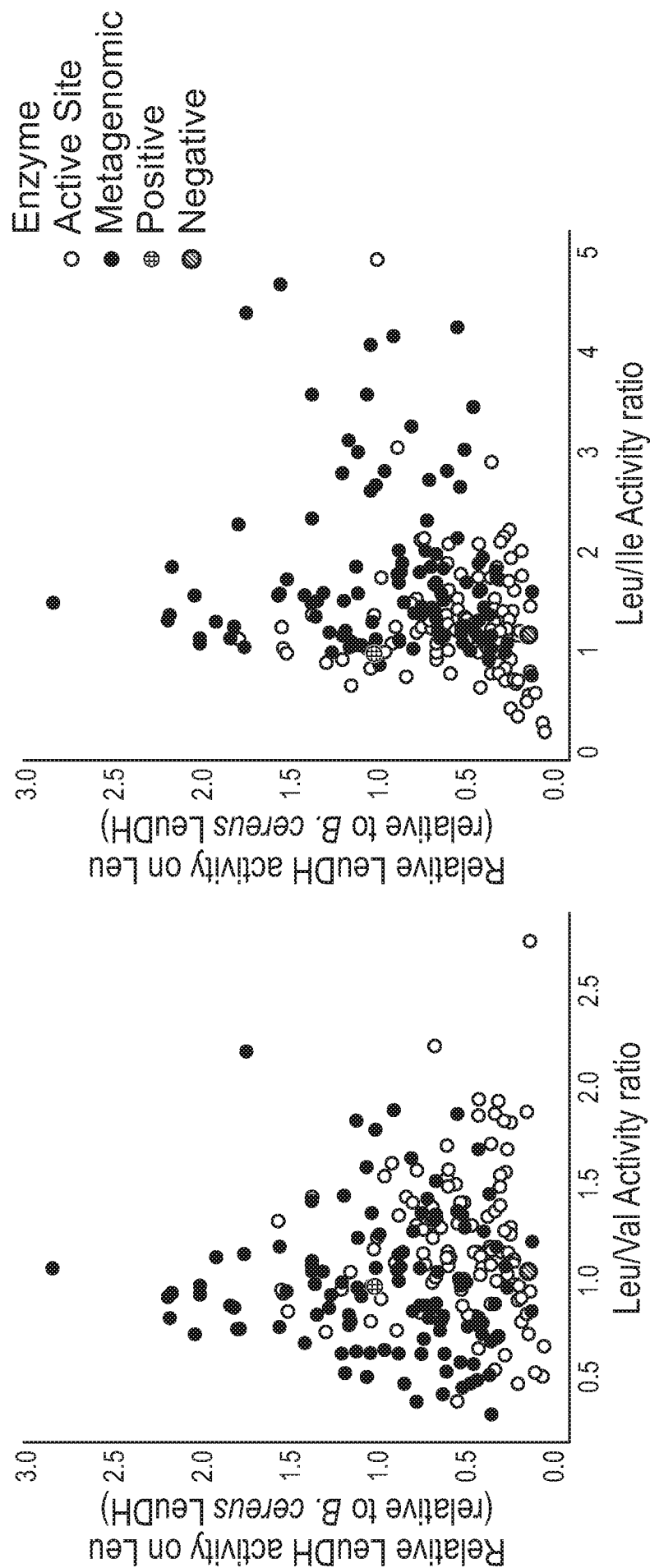


FIG. 3

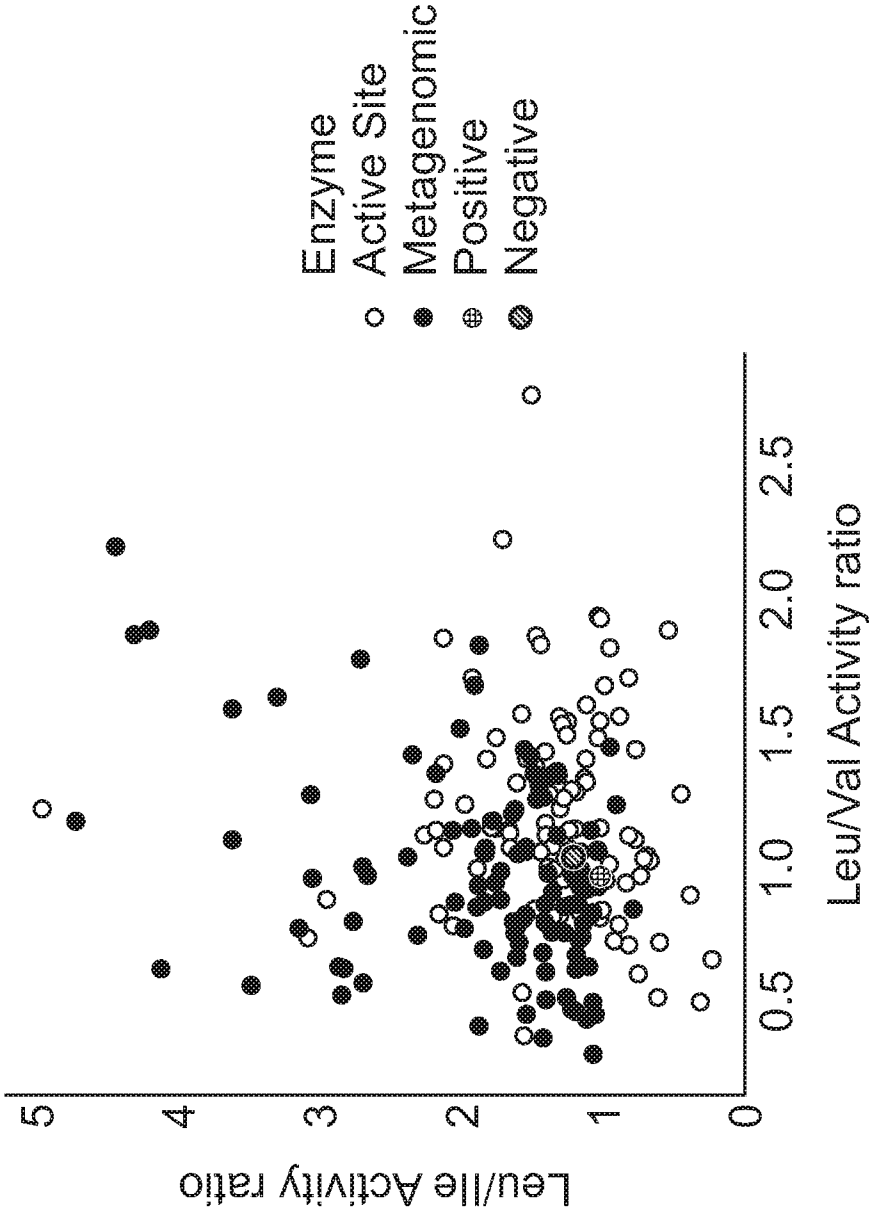
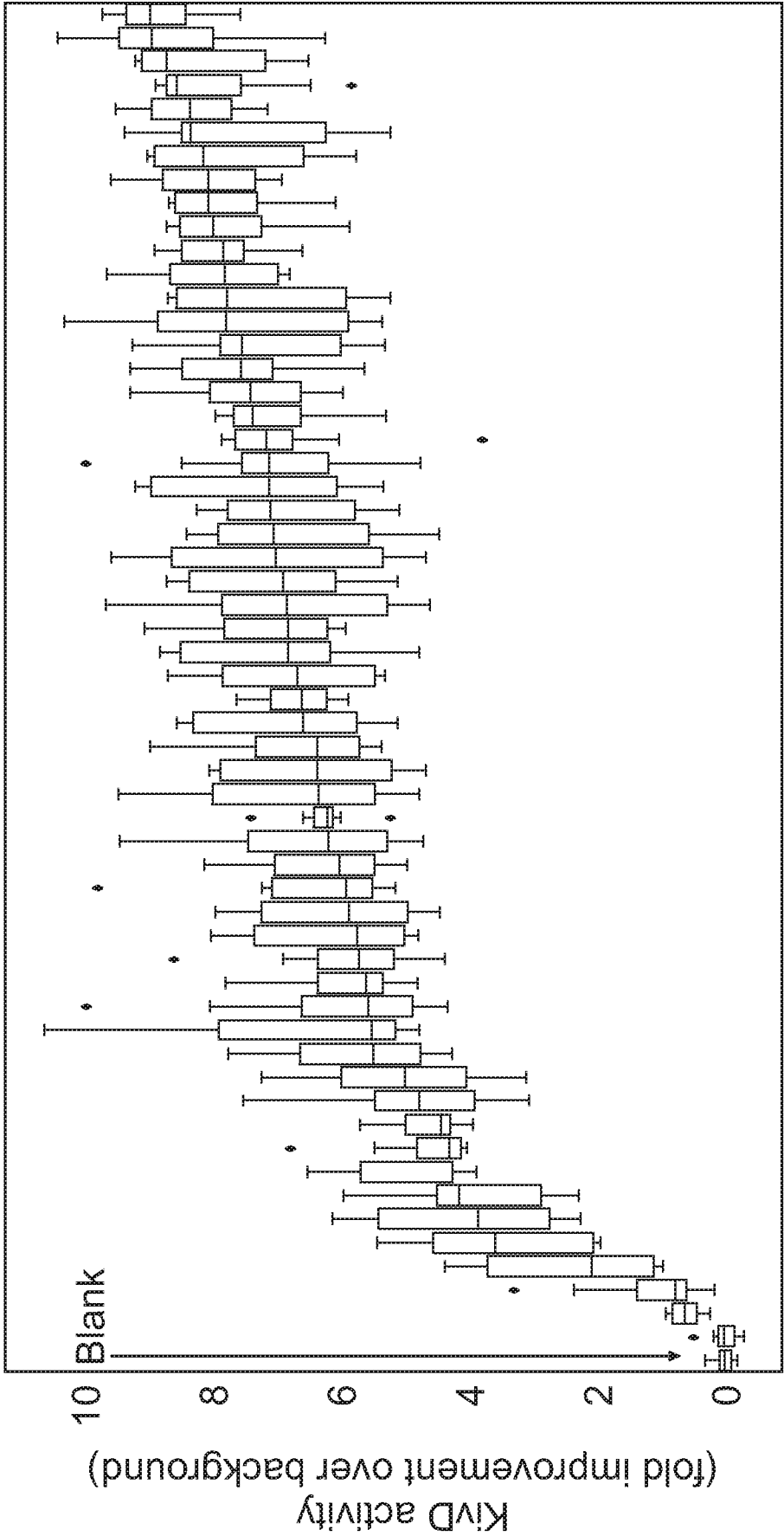


FIG. 4

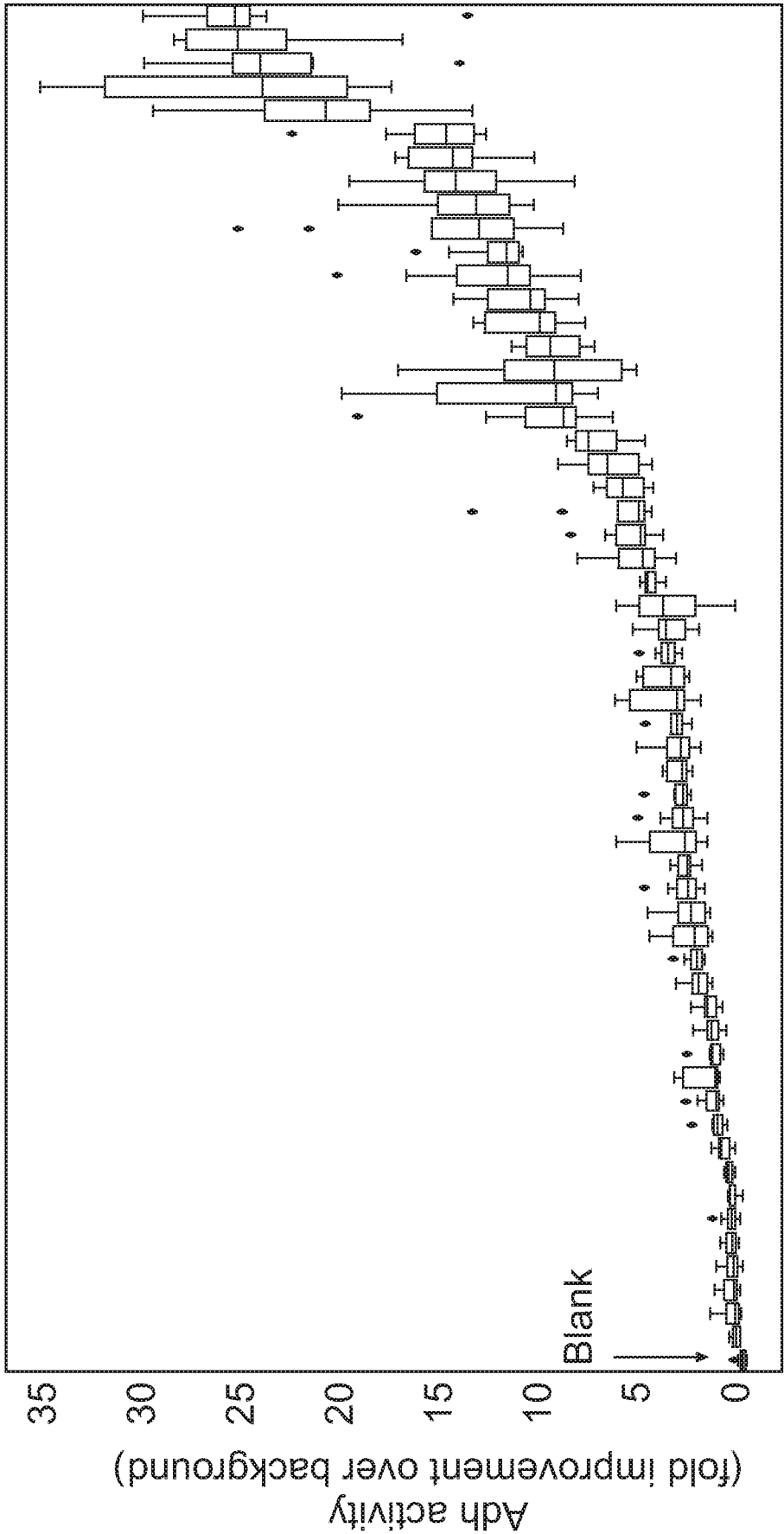
7/23



KivD Enzymes

FIG. 5

8/23



Adh Enzymes

FIG. 6

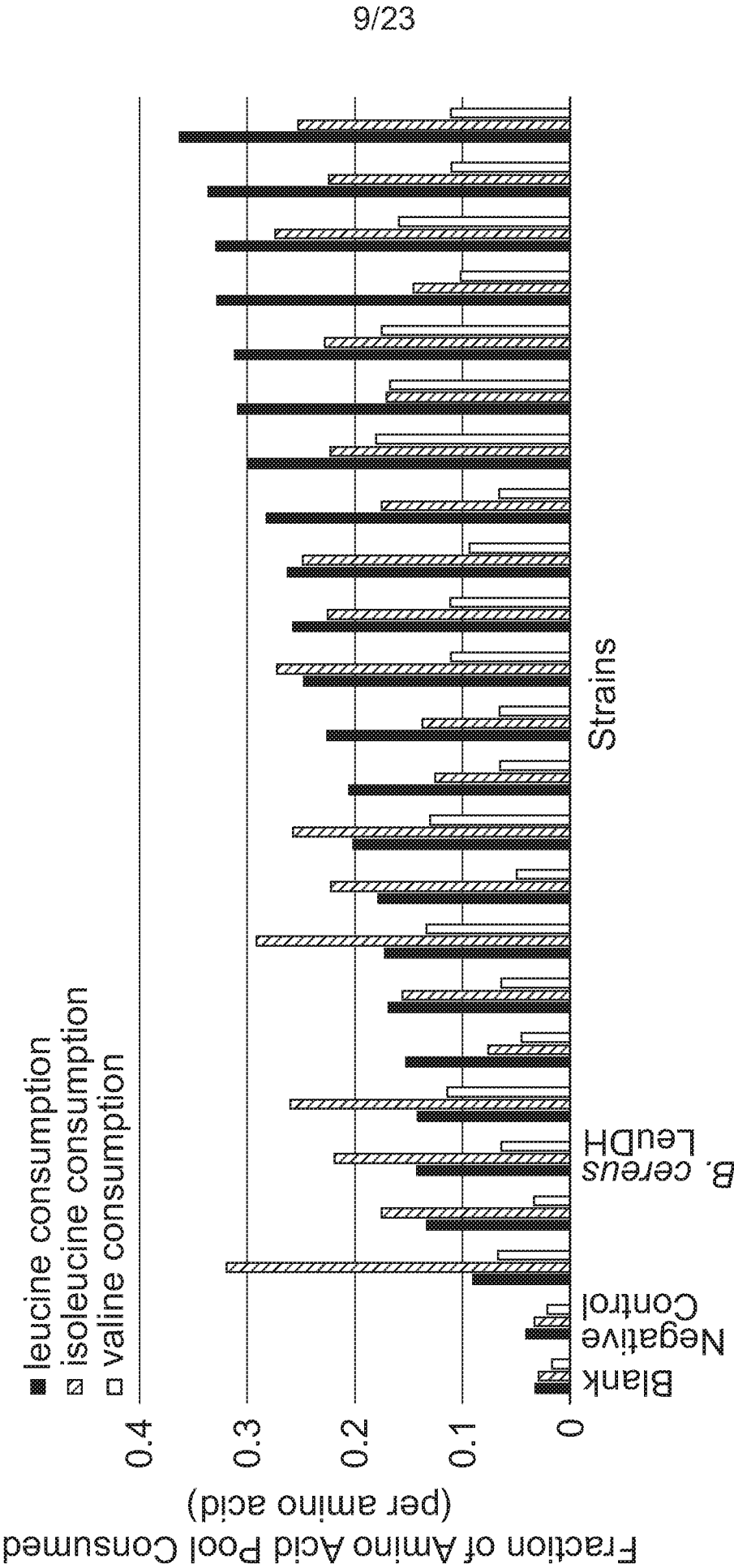


FIG. 7

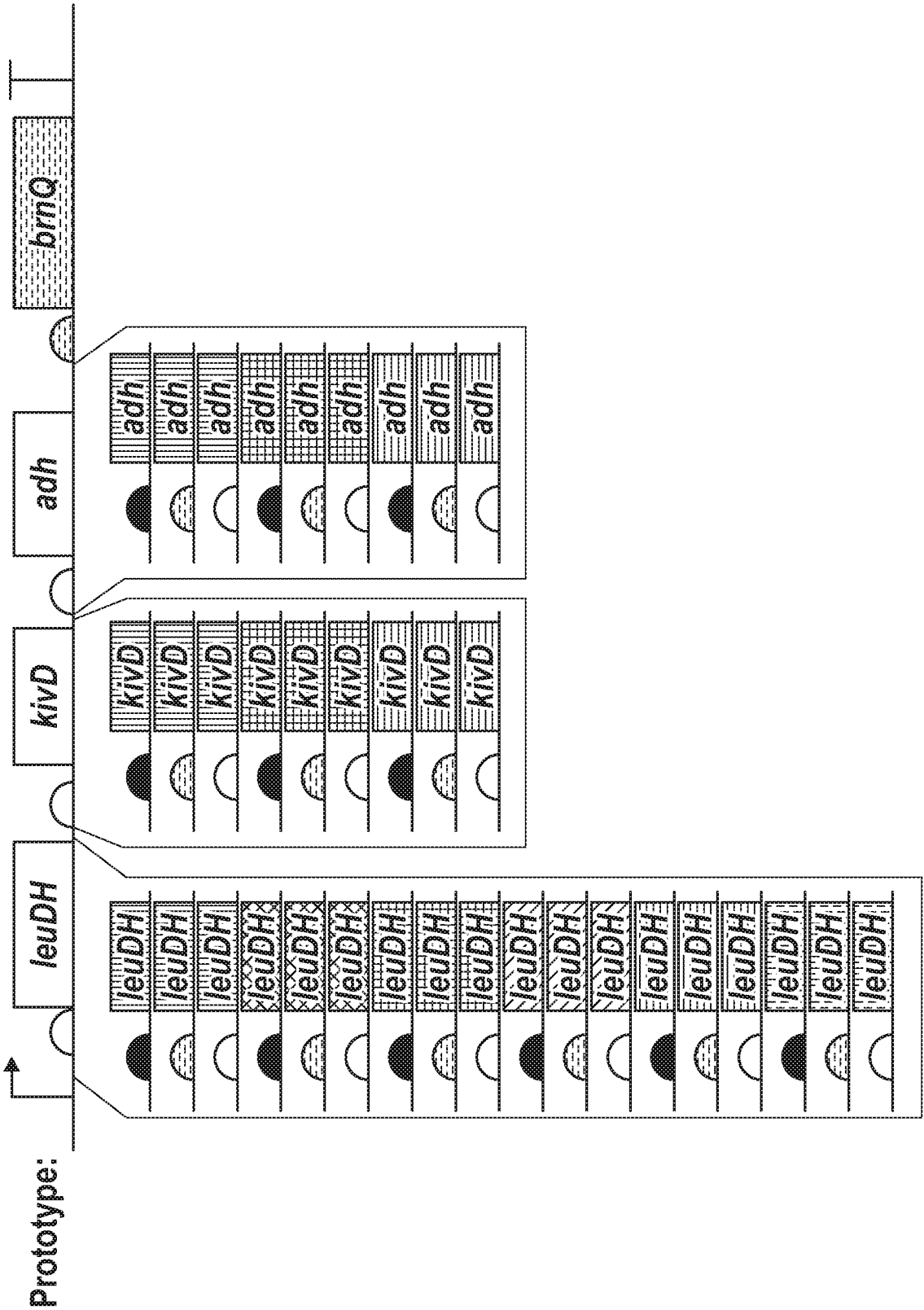
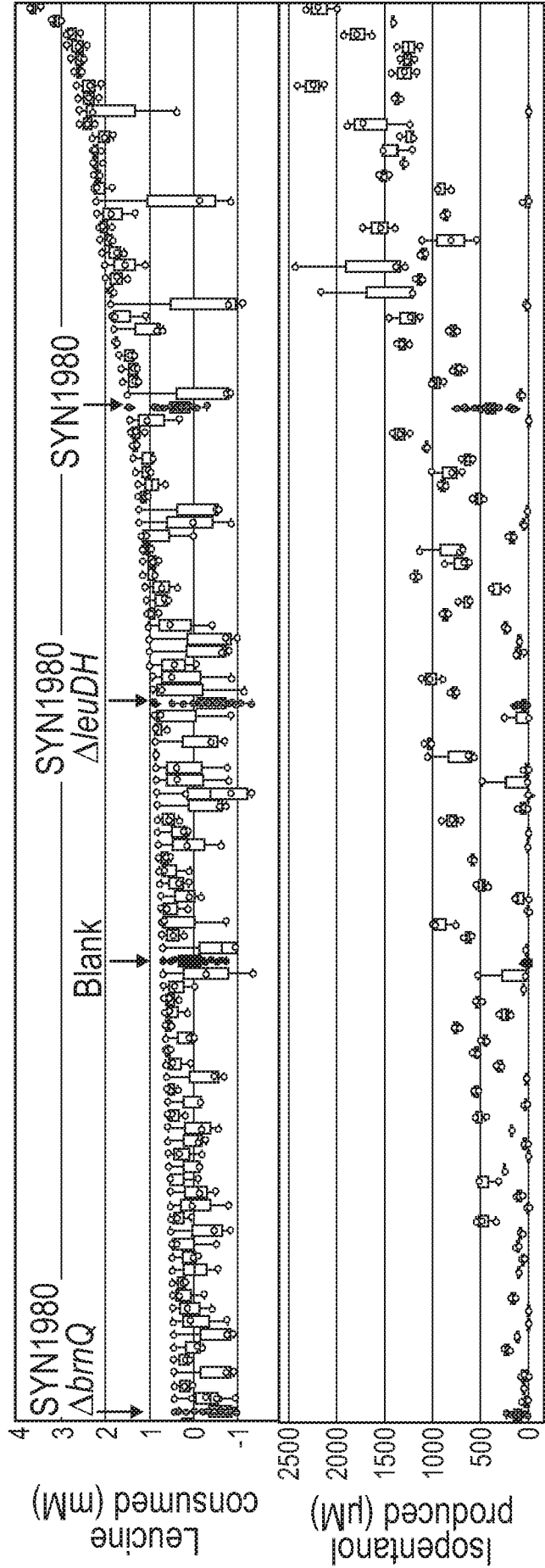


FIG. 8

11/23



Library Strains

FIG. 9

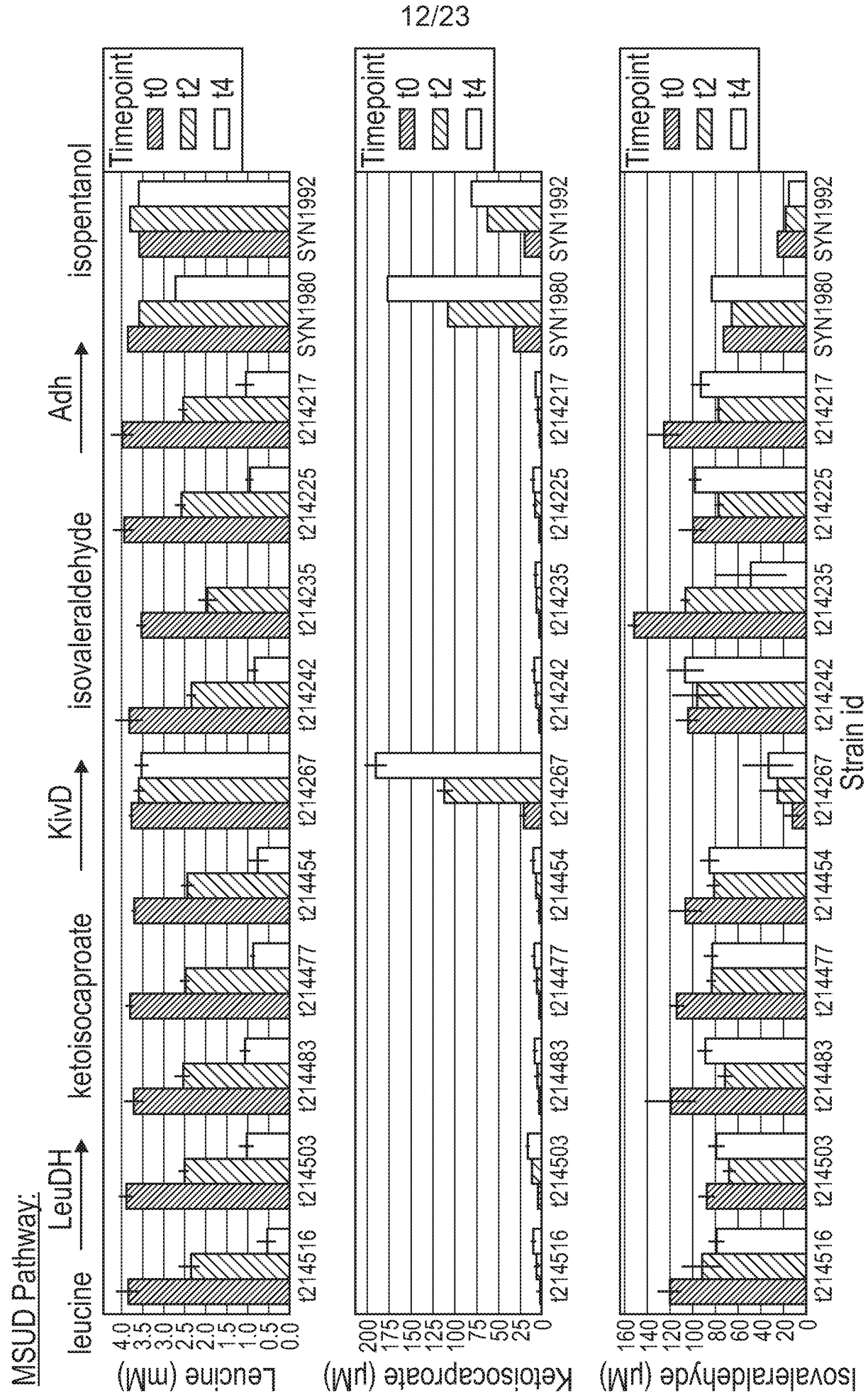


FIG. 10

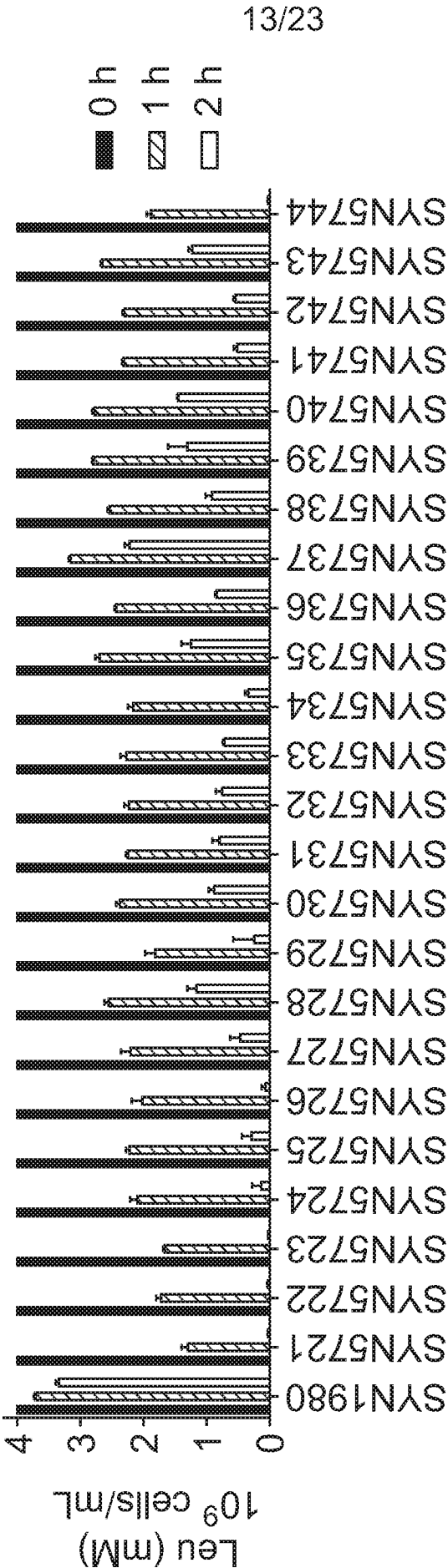


FIG. 11

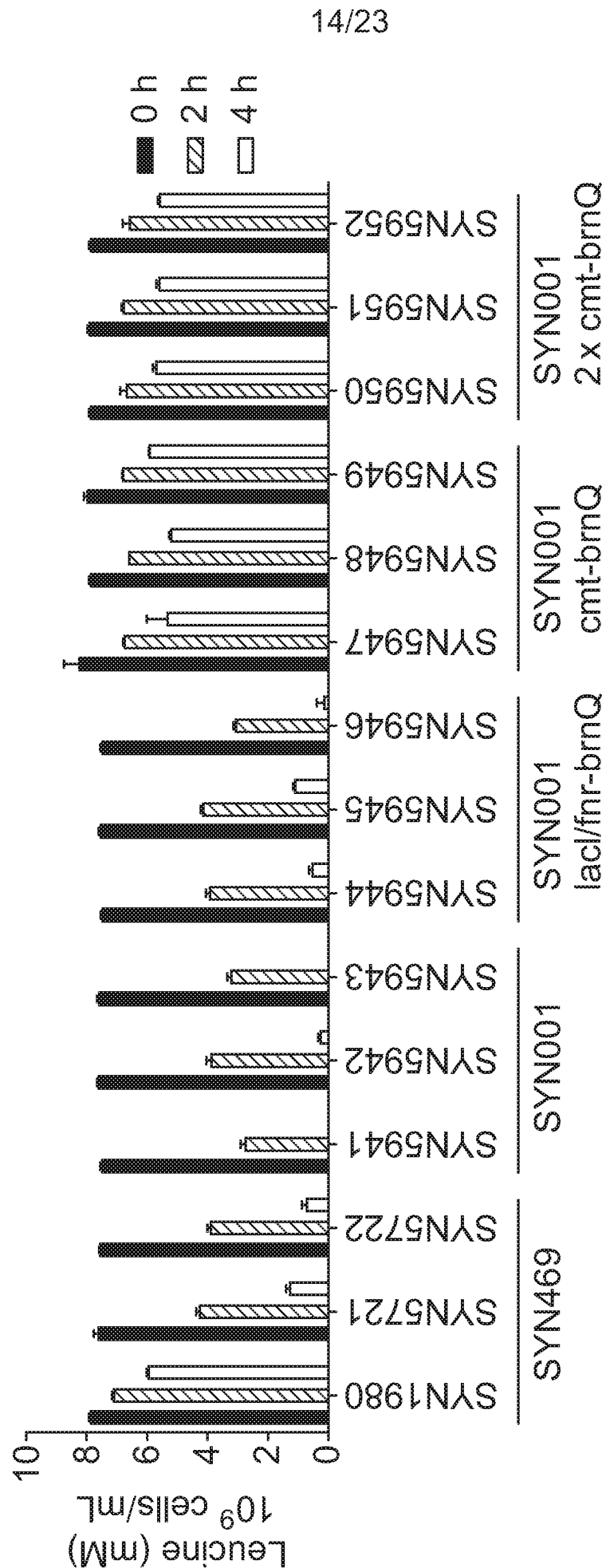


FIG. 12

15/23

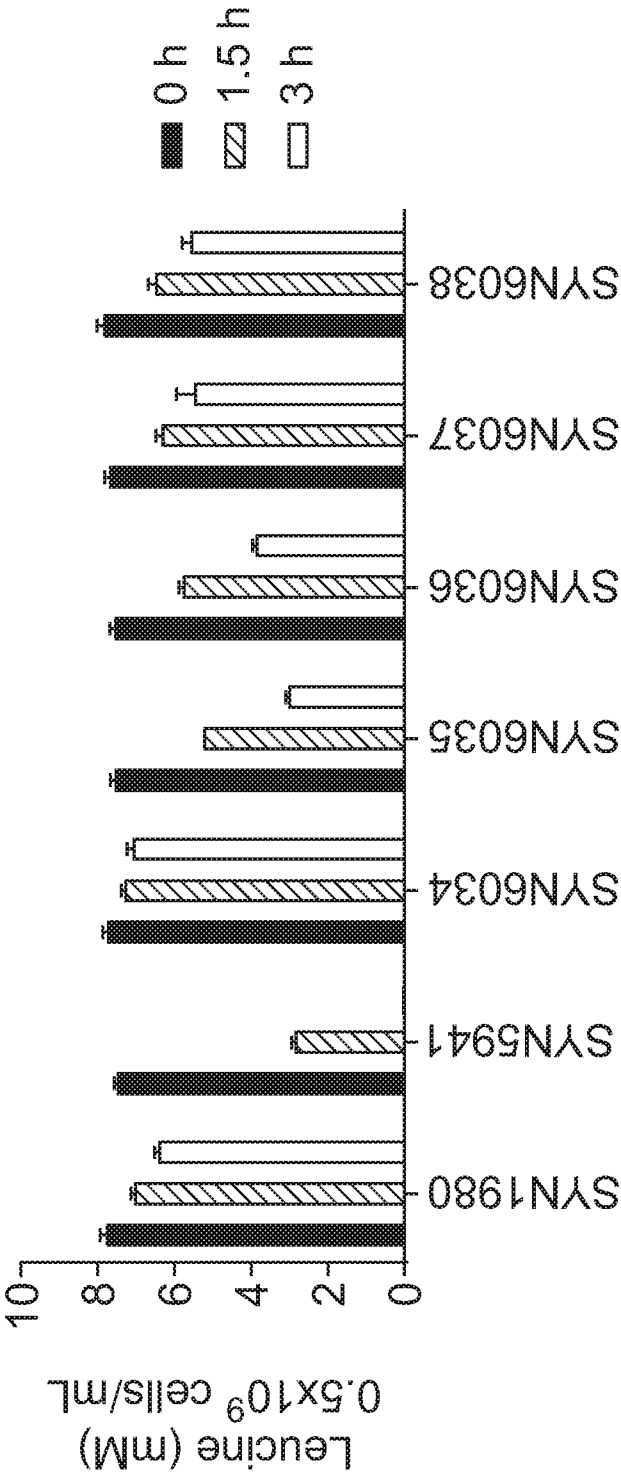


FIG. 13

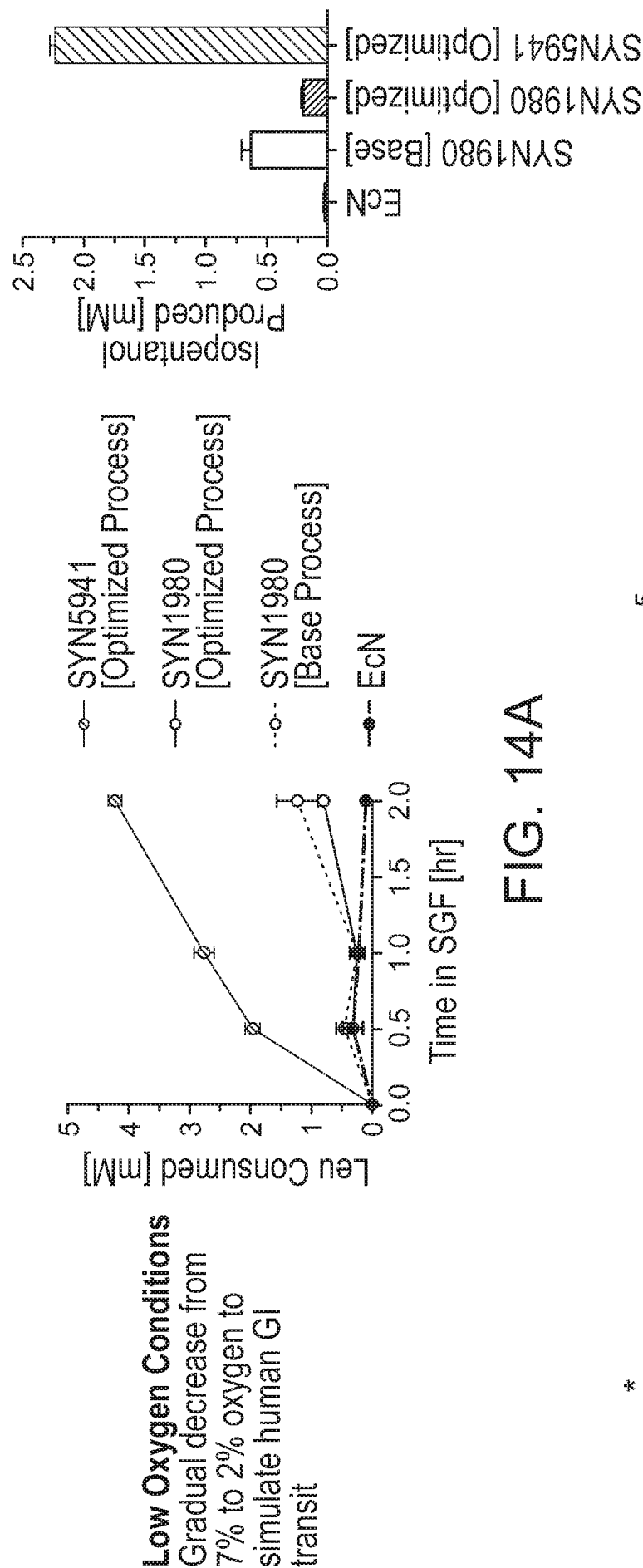


FIG. 14A

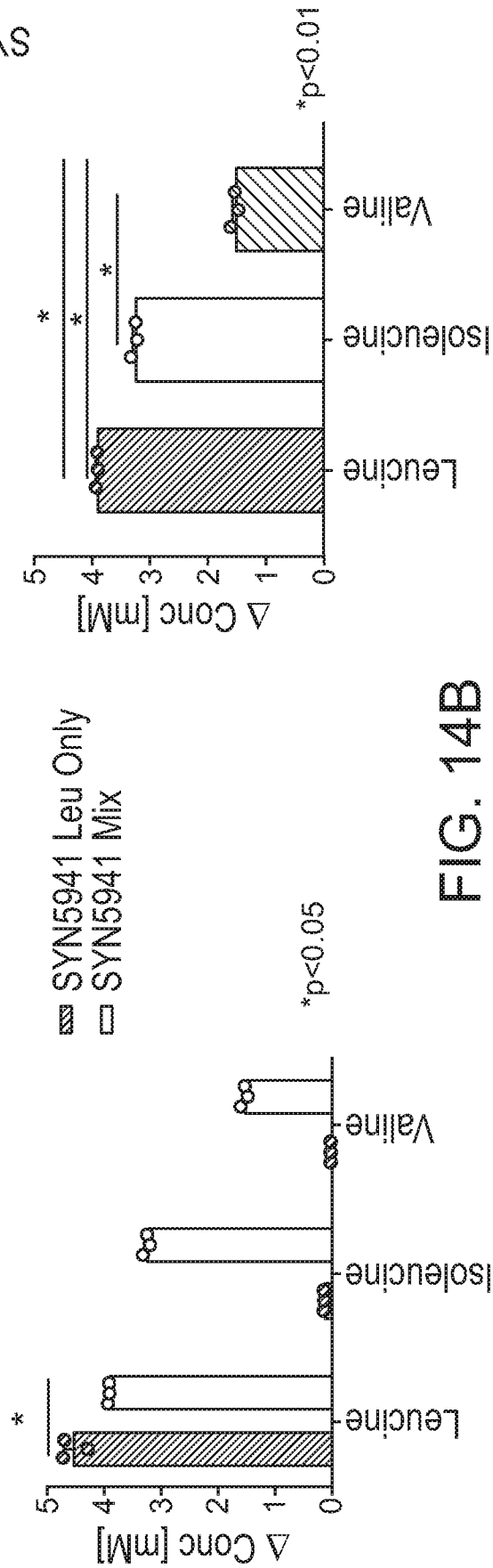


FIG. 14B

FIG. 15A

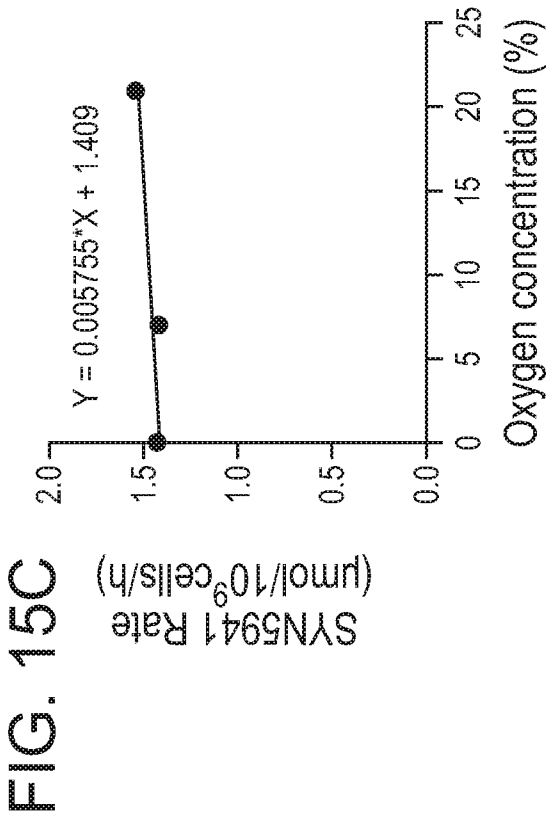
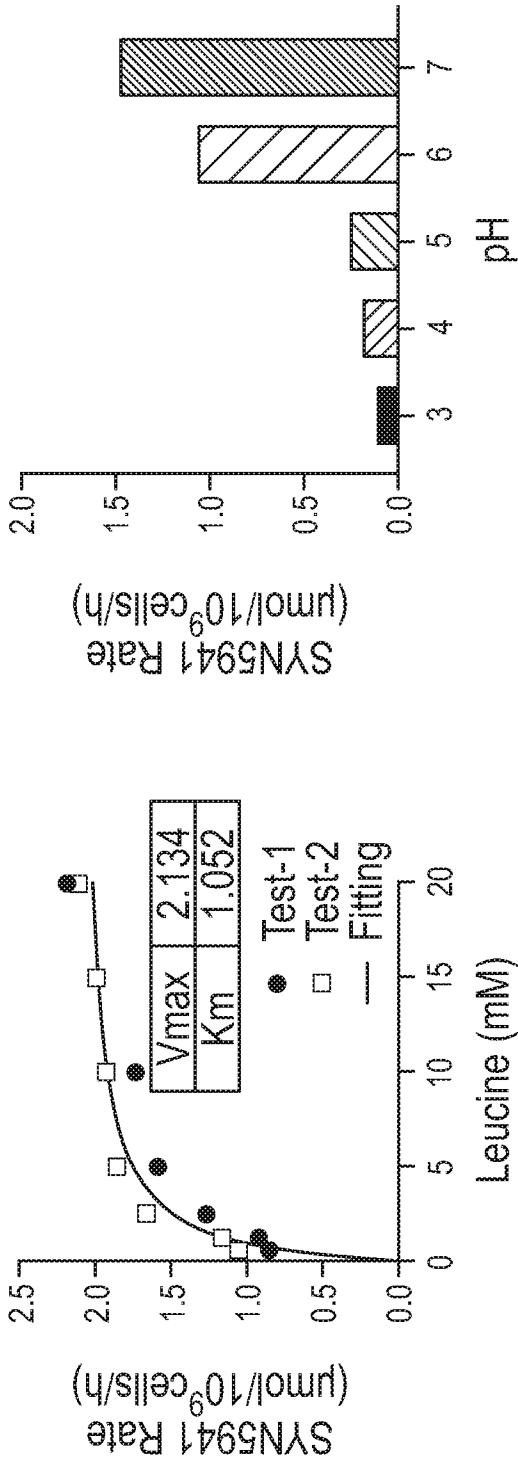


FIG. 15B

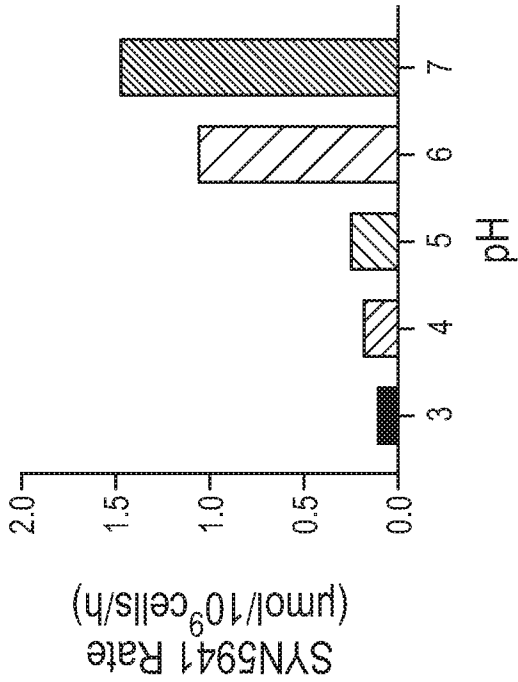


FIG. 15C

FIG. 16A

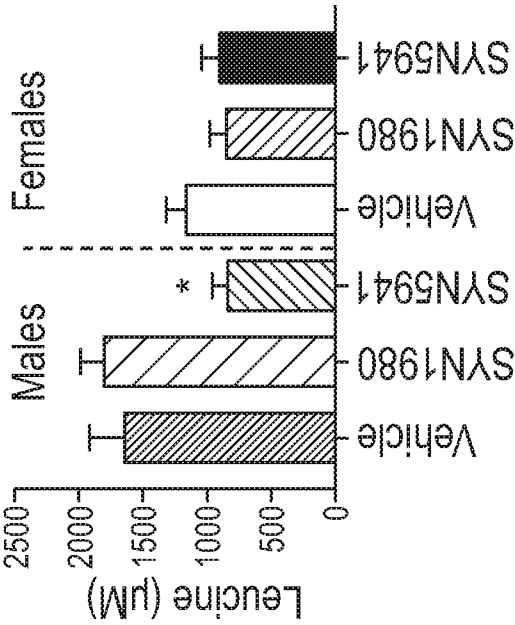


FIG. 16B

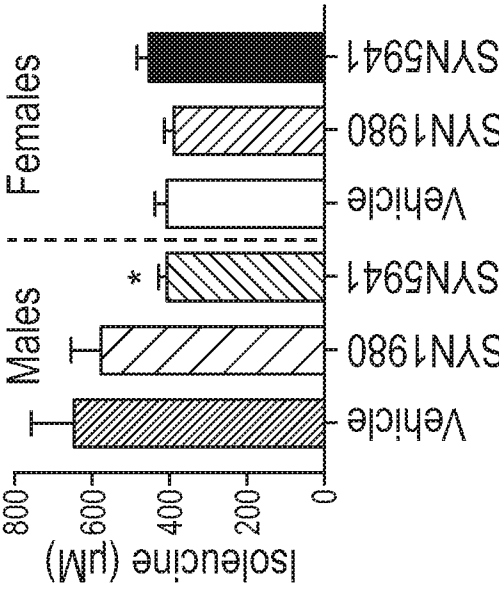
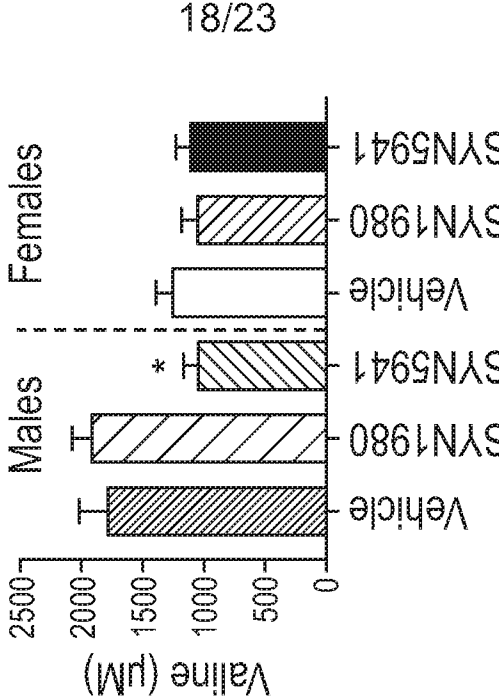


FIG. 16C



18/23

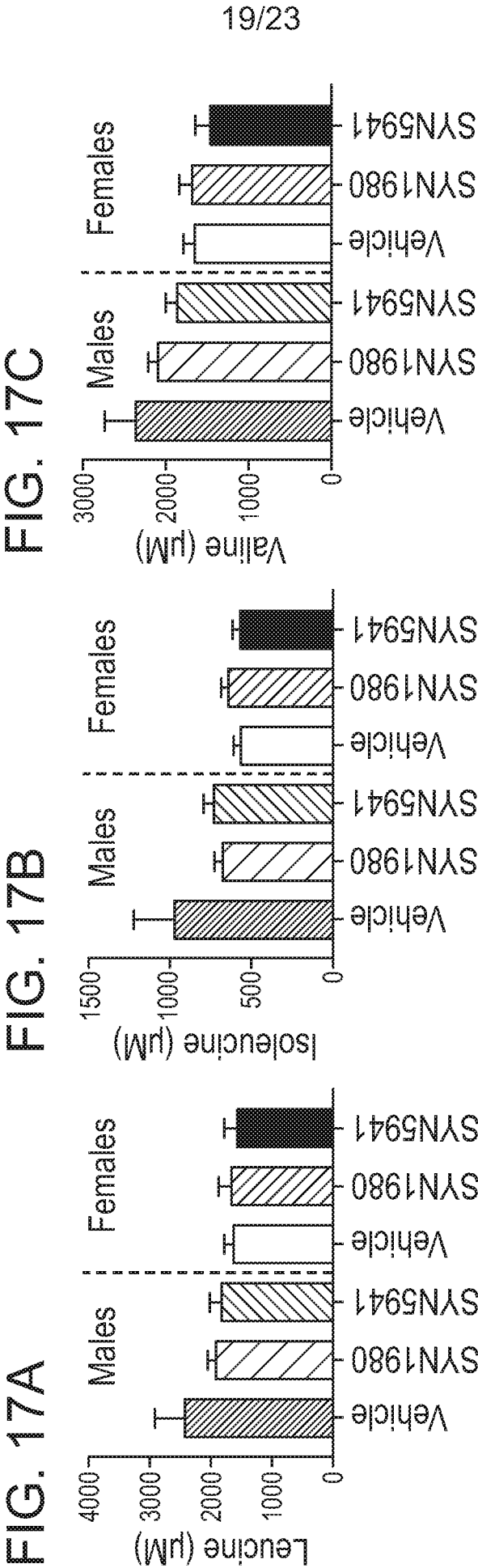


FIG. 18A

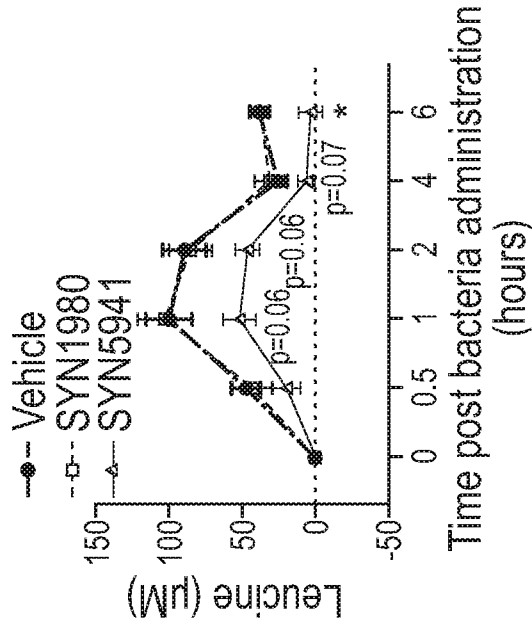


FIG. 18B

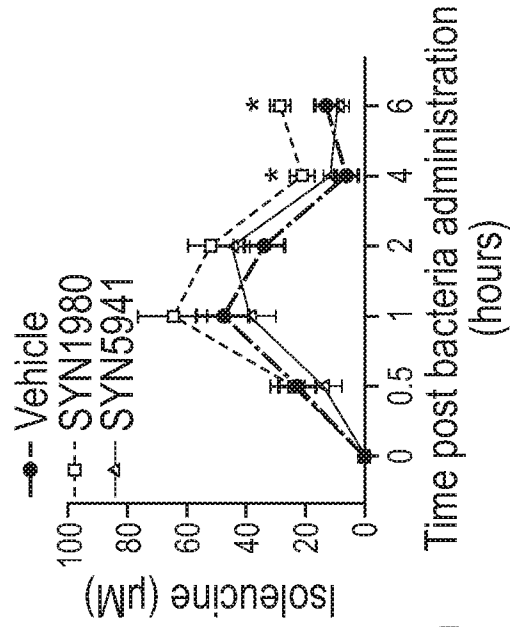


FIG. 18C

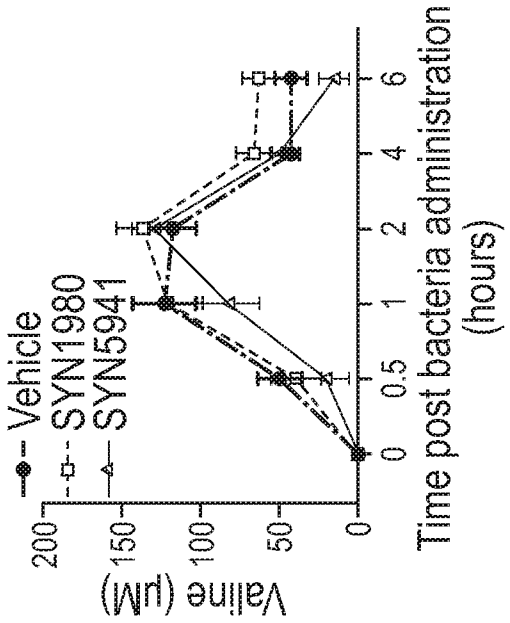


FIG. 18D

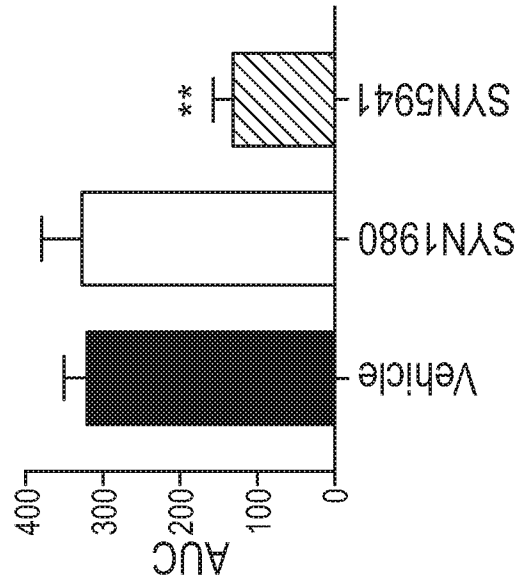


FIG. 18E

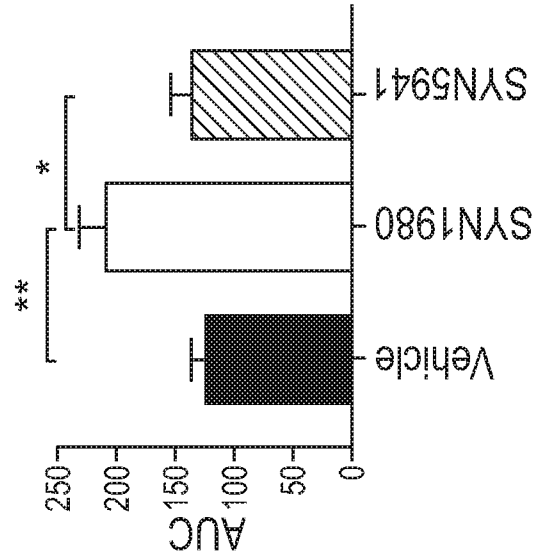
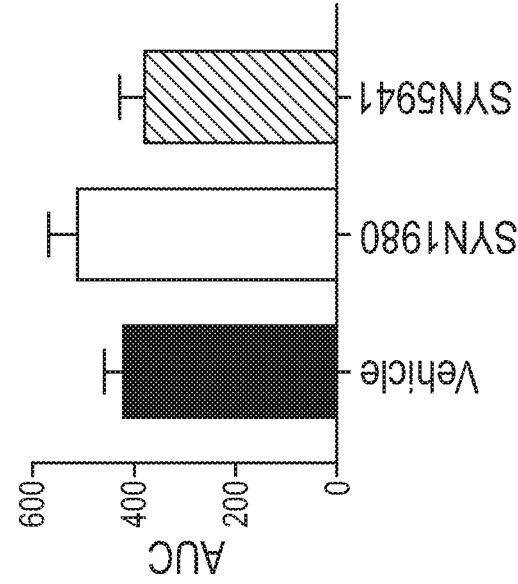


FIG. 18F



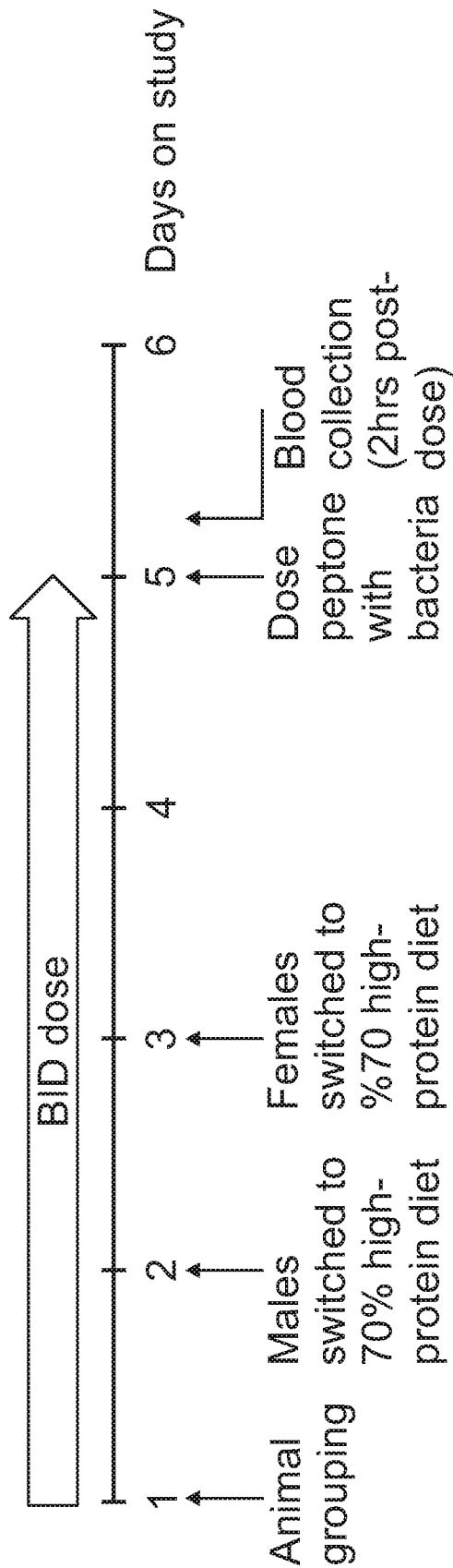


FIG. 19

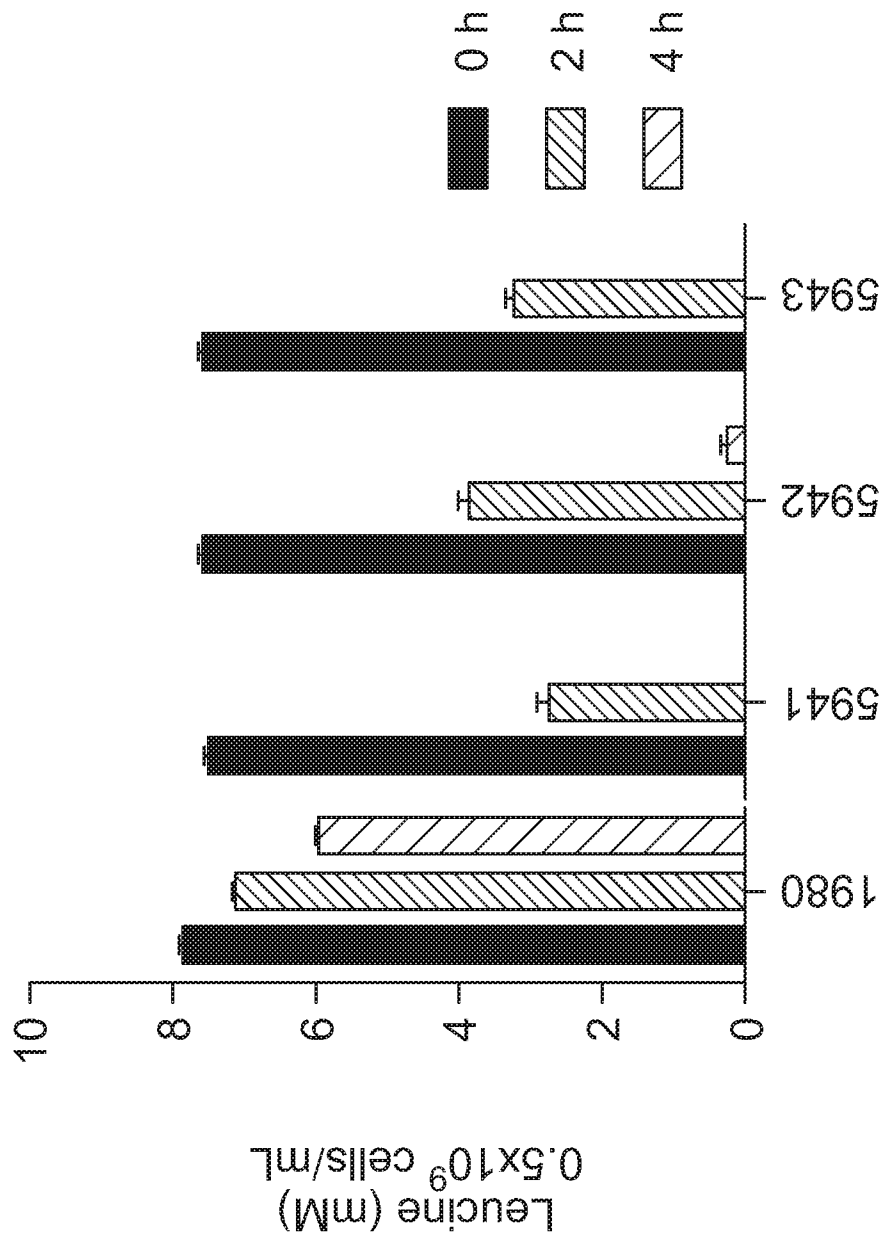


FIG. 20

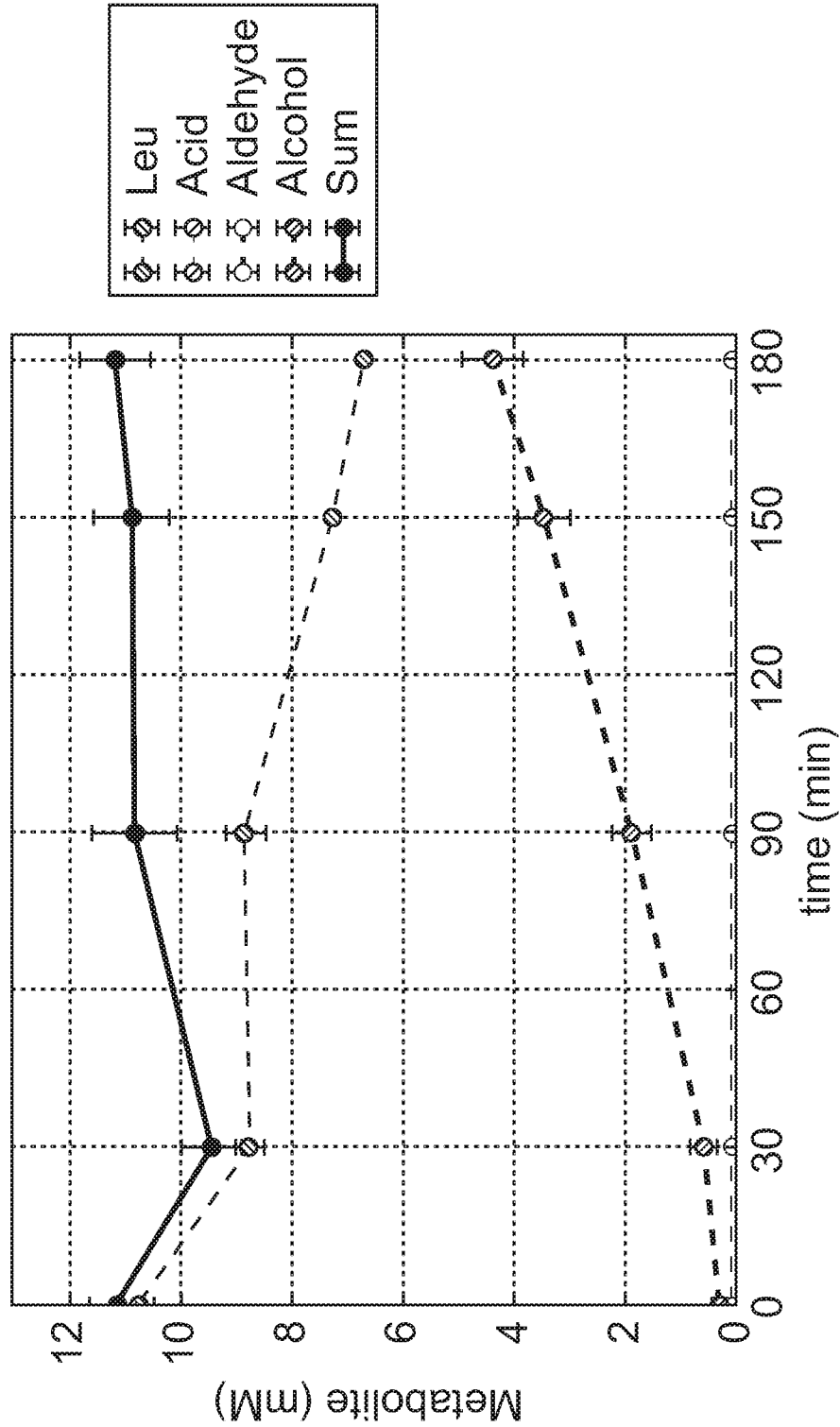


FIG. 21

SEQUENCE LISTING

<110> SYNLOGIC OPERATING COMPANY, INC.
GINKGO BIOWORKS, INC.
CARLIN, DYLAN ALEXANDER
STONE, LAURA K.
TUCKER, ALEX C.
BOYLE, PATRICK M.

<120> OPTIMIZED BACTERIA ENGINEERED TO TREAT DISORDERS INVOLVING THE
CATABOLISM OF LEUCINE, ISOLEUCINE, AND/OR VALINE

<130> 126046-04220

<140>

<141>

<150> 62/865,129

<151> 2019-06-21

<150> 62/864,875

<151> 2019-06-21

<160> 168

<170> PatentIn version 3.5

<210> 1

<211> 10193

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 1

gcgaaaaatc aataaggagg caacaagatg tgcgaaaaac atcttaatca tgctaaggag	60
gttttctaata gaacatcttc aagaaaatgg aggaatttaa ttatgaacaa ctggtctact	120
tctacgacag cgaaacggaa ctcaaaggta ttacctgtat acacaacaca actttagggc	180
cggcattggg cggtacccgc ctttgggaact ataactctga ggaagatgcc gttgaagacg	240
taatccgtct ggctcggggc atgacttaca aagcggcttg cgccggtctg aatctggggc	300
gcggtaaaac cgtgctgata ggtgatgcta aaaagattaa atcagagtcc tacttccgtg	360
gactggggcg ctacgttcag tcgctgaacg gcagatatat caccgcggaa gacgtaaata	420
cttctacgaa ggatatggca tacgttgcta tggaaactga ctatgtggta ggcctgggag	480
gtaaattccg caaccctagt ccagttactg cttacggtgc atttatgggt atcaaagcgg	540

cgctgatgaa	aaaatttgag	gatagctcta	ttgaaggccg	aaccttcgca	gtgcagggtg	600
ctgggcagac	gggttactat	cttatcgatt	acctcctagg	caacaacaag	ttcaaagaaa	660
aggctaaaaa	gatttacttc	accgaaatta	acgagagcta	tatcgagcgt	atgaacaaag	720
aacatccgga	agttgaattt	atttccccgg	acaaaatcta	ctcgctggaa	gtagacgtct	780
tcgtgccctg	cgccctgggc	aaaatcgtta	atgacaaaac	tatcgatgaa	tttaagtgtc	840
cgatcatcgc	aggtactgca	aacaacgtac	tggaaaggga	agcgcacggc	aacatgctta	900
aagaacgtgg	cattctttac	gccccggact	atgtgatcaa	tgctgggtggg	ctgatcaacg	960
tttaccacga	gctgaacggt	tacaataaag	agaacgctat	tctggaagtg	gaattaattt	1020
atgatcgcct	actggaaata	ttcaacatcg	ctgattctct	gaacatcagc	accaatatcg	1080
ctgccaacga	gttcgcggaa	aaacgtatca	agcaaattaa	gtccttgaaa	aacaacttca	1140
ttaaacgcta	agtcaataaa	ggcatataaa	aggagggttaa	taacatgaaa	gttaaagtaa	1200
aacatcttaa	tcatgcacag	gagactttct	aatgtcgacg	acaaccgttg	gtgactactt	1260
gctgtatcgc	ttaaacgaaa	tcggcattga	gcacctcttc	ggagtgccag	gtgattacaa	1320
tctgcaattt	ctggatcatg	taatcgacca	ccctcagctg	acttgggtcg	gctgcactaa	1380
cgaacttaac	gctgcctacg	cagctgatgg	ttatgcgcgt	tgtcgtccgg	ctgcggcact	1440
gctgaccacc	ttcgggggttg	gcgaactgag	cgctattaat	ggcatcgcag	gttcctacgc	1500
ggagtatctg	ccggtaatac	atatcgttgg	tgcaccgagt	ctatcagccc	agcagcaggg	1560
cgacctgatt	caccactctc	ttggcgaagg	tgatTTTTcc	agcttcctga	ggatgtccca	1620
accggtgtct	gttgcgcagg	ctgctctgac	tcctgataac	gcatgcaagg	aaatcgaccg	1680
cgtactggcg	gaagtcctca	ttcagcgtcg	tcccggctac	ctgctgctgt	ctaccgacgt	1740
ggctgctgcg	ccggcgggctc	tgccacaaaag	cactctttct	ttgccgaccg	ccccggatca	1800
tcgcgcagtt	ctggctgctt	tcagcgacgc	tgctgagcag	atgctggctc	aggccaaaag	1860
cgtctctcta	ctggcggact	ttctggctga	tcgtttcggt	gttactcgag	cactggccgc	1920
gtggcttcag	caggttccgc	taccgcacgc	cactctgtta	atgggtaaag	gcgttctgag	1980
tgaacagcaa	ccagggttcg	tgggtaccta	cgctgggtgcg	gcatctatcg	attcgacgcg	2040
tggcgcaatc	gaagaagctg	gggtaattat	cggagtggga	gtagattttt	ccgacactat	2100
cacagcaggc	ttctcgcagc	agatcgacgc	ccgccgtttt	atagacattc	aacccttctt	2160

ctctcgtatt ggcgatcgcc agtttgatca cttgccgatg caggctgccg tcgcagccct	2220
gcatcaactg tgtcttcgtt atcagcagca gtggtctatc accgctccta gcccgcctgc	2280
actgccgccg gctgctggta gcgagctgtc ccagaacgca ttctggcagg cgatgcagaa	2340
cttcatccgc cctggggacc tgttggtggc cgaccaaggt actgcggcgt tcggcgcagc	2400
ggcgtgcgc ttaccgcaga attgccagct gcttgtgcag ccgctgtggg gctcaatcgg	2460
ttacagtctg ccggccacct ttggtgctca gacggcagat acagagcgtc gtgtaatcct	2520
aatcattggc gatggttcag cgcaattaac tattcaggaa ctttccagta tgatgcgtga	2580
cggcttgaia cctatcatct ttctcctgaa caacaacggg tacaccgttg aacgggcgat	2640
tcacggcgcg gagcaacgtt ataacgatat cgctgcttgg aattggaccc aactgccccca	2700
ggcgtgagt gttcattgcc cagcgcagag ctggcgagtc gttgaaacgg tgcagctgac	2760
cgacgtaatg aaagtcatcg ctgcttctcc gcgtctgagc ttggtagaag ttgttctgcc	2820
tgcaatggat gtcccaccgc tgctgcaagc agtgagtgcc gctctgaacc agcgcaactc	2880
ctcttaagac cagaggtaa gaggtaacaa ccatgcgagt gttgaagaaa catcttaatc	2940
atgcagggga gggtttctaa tgcagacggc gttcttgtat aagccaggtc acgaaaactt	3000
agtgcgctcg gagatcccga tacctaaagc tgggcgtggc gaagtcgttc tggaaattaa	3060
agccgctggc atgtgccatt ccgatctgca cgttctcgac ggtggaatcc ccctgccggg	3120
tcaatttgta atgggccatg aaatcgttgg tactattcac gagatcggcc aggacgtgac	3180
cggtttcaaa cagggcgatc tgtacgcagt ccacggccccg aatccgtgtg gtatttgcac	3240
cctgtgcaga gaaggatttg ataacgactg cactacagtg gcgaaaaccg gtcaatggtt	3300
cggactgggt cttgacggcg gctaccagaa gtatatccgt atcccgaacg taaggcttat	3360
cgtaaagtt ccagaagggtg tttcagctga ggcagctgcg agctgtactg atgcagtact	3420
gaccccgtac cgtgcactaa aacaggctgg cgccagcaac tctactcggg tactgattct	3480
gggtctgggt ggcttaggtc tgaatgccct taaactggct aagaccttcg gcagttacgt	3540
ttacgcatct gacctgaaac cttctgcgcg tgaagctgct aaggccgctg gggcggtatga	3600
agtgtggag tccctgccccg aagaccgct ggggtgttgat atcgtgttag acgtcgttgg	3660
cgtgcagagc accttcaacc tcgctcaaaa acacgttggc ccgctggca tcattgtacc	3720
ttaggcctg gcatccccac agctttcgtt taacctaacg gatctggcgc tccgcgaaat	3780

tcgtgttcag ggcacttttt ggggcacgag caatgagctg gctgaatgtc tgcgcctgtg	3840
ccagctgggc ctgatcaacc cgaaatatac tgttggtgcct cttgaagaag cgccgaaata	3900
tatggaagca atggctcatg ggaaagtaga aggtcgtatc gttttccacc cgtaagaagg	3960
agatatacat atgacccatc aattaagatc gcgcgatatc atcgctctgg gctttatgac	4020
atttgcgttg ttcgtcggcg caggtaacat tattttccct ccaatggtcg gcttgcaggc	4080
aggcgaacac gtctggactg cggcattcgg cttcctcatt actgccgttg gcctaccggt	4140
attaacggta gtggcgctgg caaaagttgg cggcggtgtt gacagtctca gcacgccaat	4200
tggtaaagtc gctggcgtag tgctggcaac agtttggttac ctggcggtgg ggccgctttt	4260
tgctacgccg cgtacagcta ccgtttcttt tgaagtgggc attgcgccgc tgacgggtga	4320
ttccgcgctg ccgctgttta ttacagcct ggtctatttc gctatcgta ttctggtttc	4380
gctctatccg ggcaagctgc tggataccgt gggcaacttc cttgcgccgc tgaaaattat	4440
cgcgctggtc atcctgtctg ttgccgcaat tatctggccg gcgggttcta tcagtacggc	4500
gactgaggct tatcaaacg ctgcgttttc taacggcttc gtcaacggct atctgacat	4560
ggatacgctg ggcgcaatgg tgtttggtat cgttattgtt aacgcggcgc gttctcgtgg	4620
cgttaccgaa gcgcgtctgc tgaccggtta taccgtctgg gctggcctga tggcgggtgt	4680
tggctctgact ctgctgtacc tggcgctgtt ccgtctgggt tcagacagcg cgtcgtgggt	4740
cgatcagtct gcaaacgggt cggcgatcct gcatgcttac gttcagcata cttttggcgg	4800
cggcggtagc ttctgtctgg cggcgttaat cttcatcgcc tgcttgggtca cggcggttgg	4860
cctgacctgt gcttgtgcag aattcttcgc ccagtacgta ccgctctctt atcgtagct	4920
ggtgtttatc ctccggcggt tctcgatggt ggtgtctaac ctccggttga gccagctgat	4980
tcagatctct gtaccgggtc tgaccgccat ttatccgccg tgtatcgcac tggttgtatt	5040
aagttttaca cgctcatggt ggcataattc gtcccgcgtg attgctccgc cgatgtttat	5100
cagcctgctt tttggtattc tcgacgggat caaggcatct gcattcagcg atatcttacc	5160
gtcctgggcg cagcgtttac cgctggccga acaaggctct gcgtgggtta tgccaacagt	5220
ggtgatggtg gttctggcca ttatctggga tcgtgcggca ggtcgtcagg tgacctccag	5280
cgctcactaa tacgcatggc atggatgacc gatggtagtg tgggctctcc ccatgcgaga	5340
gtagggaact gccaggcatc aaataaaacg aaaggctcag tcgaaagact gggcctttcg	5400

ttttatctgt tgtttgtcgg tgaacgctct cctgagtagg acaaatccgc cgggagcgga	5460
tttgaacgtt gcgaagcaac ggcccggagg gtggcgggca ggacgcccgc cataaactgc	5520
caggcatcaa attaagcaga aggccatcct gacggatggc ctttttgcgt ggccagtgcc	5580
aagcttgcac gcgtgccagc tgcattaatg aagaaatcat gctggaagaa taacagctca	5640
ctcaaaggcg gtagtacggg ttttgcctgc cgcaaacggg ctgttctggt gttgctagtt	5700
tgttatcaga atcgcagatc cggcttcagc cggtttgccg gctgaaagcg ctatttcttc	5760
cagaattgcc atgatttttt cccacggga ggcgtcactg gctcccgtgt tgtcggcagc	5820
tttgattcga taagcagcat cgcctgtttc aggctgtcta tgtgtgactg ttgagctgta	5880
acaagttgtc tcagggtgttc aatttcatgt tctagttgct ttgttttact ggtttcacct	5940
gttctattag gtgttacatg ctgttcatct gttacattgt cgatctgttc atggtgaaca	6000
gctttgaatg caccaaaaac tcgtaaaagc tctgatgtat ctatcttttt tacaccgttt	6060
tcatctgtgc atatggacag ttttcccttt gatatgtaac ggtgaacagt tgttctactt	6120
ttgtttgtta gtcttgatgc ttcactgata gatacaagag ccataagaac ctcagatcct	6180
tccgtattta gccagtatgt tctctagtgt ggttcgttgt ttttgcgtga gccatgagaa	6240
cgaaccattg agatcatact tactttgcat gtcactcaaa aattttgcct caaaactggt	6300
gagctgaatt tttgcagtta aagcatcgtg tagtgttttt cttagtccgt tatgtaggta	6360
ggaatctgat gtaatggttg ttgggtatttt gtcaccattc atttttatct gggtgttctc	6420
aagttcgggt acgagatcca tttgtctatc tagttcaact tggaaaatca acgtatcagt	6480
cgggcggcct cgcttatcaa ccaccaattt catattgctg taagtgttta aatctttact	6540
tattggtttc aaaaccatt gggttaagcct tttaaactca tggtagttat tttcaagcat	6600
taacatgaac ttaaattcat caaggctaatt ctctatatatt gccttgtagg ttttcttttg	6660
tgtagttct ttttaataacc actcataaat cctcatagag tatttgtttt caaaagactt	6720
aacatgttcc agattatatt ttatgaattt ttttaactgg aaaagataag gcaatatctc	6780
ttcactaaaa actaattcta atttttcgct tgagaacttg gcatagtttg tccactggaa	6840
aatctcaaag cctttaacca aaggattcct gatttccaca gttctcgtca tcagctctct	6900
ggttgcttta gctaatacac cataagcatt ttccctactg atgttcatca tctgagcgta	6960
ttggttataa gtgaacgata ccgtccgttc tttccttgta gggttttcaa tcgtgggggt	7020

gagtagtgcc acacagcata aaattagctt ggtttcatgc tccgttaagt catagcgact	7080
aatcgctagt tcatttgctt tgaaaacaac taattcagac atacatctca attgggtctag	7140
gtgattttaa tcactatacc aattgagatg ggctagtcaa tgataattac tagtcctttt	7200
cctttgagtt gtgggtatct gtaaattctg ctagaccttt gctggaaaac ttgtaaattc	7260
tgctagaccc tctgtaaatt ccgctagacc tttgtgtgtt ttttttgttt atattcaagt	7320
ggttataatt tatagaataa agaaagaata aaaaaagata aaaagaatag atcccagccc	7380
tgtgtataac tcactacttt agtcagttcc gcagtattac aaaaggatgt cgcaaacgct	7440
gtttgctcct ctacaaaaca gaccttaaaa ccctaaaggc ttaagtagca ccctcgcaag	7500
ctcgggcaaa tcgctgaata ttccttttgt ctccgaccat caggcacctg agtcgctgtc	7560
tttttcgtga cattcagttc gctgcgctca cggctctggc agtgaatggg ggtaaattggc	7620
actacaggcg ccttttatgg attcatgcaa ggaaactacc cataatacaa gaaaagcccg	7680
tcacgggctt ctcaggcggt tttatggcgg gtctgctatg tggtgctatc tgactttttg	7740
ctgttcagca gttcctgccc tctgattttc cagtctgacc acttcggatt atcccgtgac	7800
aggtcattca gactggctaa tgcacccagt aaggcagcgg tatcatcaac aggcttacc	7860
gtcttactgt cttttctacg gggctctgacg ctcagtggaa cgaaaactca cgttaaggga	7920
ttttggtcatt gagattatca aaaaggatct tcacctagat ctttttaaat taaaaatgaa	7980
gttttaaatc aatctaaagt atatatgagt aaacttggtc tgacagttac caatgcttaa	8040
tcagtgaggc acctatctca gcgatctgtc tatttcgttc atccatagtt gcctgactcc	8100
ccgtcgtgta gataactacg atacgggagg gcttaccatc tggccccagt gctgcaatga	8160
taccgcgaga accacgtca ccggctccag atttatcagc aataaaccag ccagccggaa	8220
gggccgagcg cagaagtggc cctgcaactt tatccgcctc catccagtct attaatgtt	8280
gccgggaagc tagagtaagt agttcgccag ttaatagttt gcgcaacggt gttgccattg	8340
ctacaggcat cgtgggtgtca cgctcgtcgt ttggtatggc ttcattcagc tccggttccc	8400
aacgatcaag gcgagttaca tgatcccca tgttgtgcaa aaaagcgggt agctccttcg	8460
gtcctccgat cgttgtcaga agtaagttag ccgcagtgtt atcactcatg gttatggcag	8520
cactgcataa ttctcttact gtcatgccat ccgtaagatg cttttctgtg actggtgagt	8580
actcaaccaa gtcattctga gaatagtgtg tgcggcgacc gagttgctct tgccccggcgt	8640

caatacggga taataaccgcg ccacatagca gaacttttaa agtgctcatc attggaaaac	8700
gttcttcggg gcgaaaactc tcaaggatct taccgctgtt gagatccagt tcgatgtaac	8760
ccactcgtgc acccaactga tcttcagcat cttttacttt caccagcggt tctgggtgag	8820
caaaaacagg aaggcaaaat gccgcaaaaa agggaataag ggcgacacgg aaatgttgaa	8880
tactcatact cttccttttt caatattatt gaagcattta tcagggttat tgtctcatga	8940
gcggatacat atttgaatgt atttagaaaa ataaacaaat aggggttccg cgcacatttc	9000
cccgaaaagt gccacctgac gtctaagaaa ccattattat catgacatta acctataaaa	9060
ataggcgtat cacgaggccc tttcgtctcg cgcgtttcgg tgatgacggg gaaaacctct	9120
gacacatgca gctcccggag acggtcacag cttgtctgta agcggatgcc gggagcagac	9180
aagcccgtca gggcgcgtca gcgggtgttg gcgggtgtcg gggctggctt aactatgcgg	9240
catcagagca gattgtactg agagtgcacc atatgcggtg tgaaataaccg cacagatgcg	9300
taaggagaaa ataccgcatc aggcgccatt cgccattcag gctgcgcaac tgttgggaag	9360
ggcgatcggg gcgggcctct tcgctattac gccagctggc gaaaggggga tgtgctgcaa	9420
ggcgattaag ttgggtaacg ccagggtttt ccagtcacg acgttgtaaa acgacggcca	9480
gtgaattcgt taagaccac tttcacattt aagttgtttt tctaattccg atatgatcaa	9540
ttcaaggccg aataagaagg ctggctctgc accttggtga tcaaataatt cgatagcttg	9600
tcgtaataat ggcggcatac tatcagtagt aggtgtttcc ctttcttctt tagcgacttg	9660
atgctcttga tcttccaata cgcaacctaa agtaaaatgc cccacagcgc tgagtgcata	9720
taatgcattc tctagtgaag aaccttggtg gcataaaaag gctaattgat tttcgagagt	9780
ttcatactgt ttttctgtag gccgtgtacc taaatgtact tttgctccat cgcgatgact	9840
tagtaaagca catctaaaac ttttagcggt attacgtaaa aaatcttgcc agctttcccc	9900
ttctaaaggg caaaagtgag tatggtgcct atctaacatc tcaatggcta aggcgtcgag	9960
caaagcccgc ttatTTTTTT catgccaaata caatgtaggc tgctctacac ctagcttctg	10020
ggcgagtTTT cgggttgTTT aaccttcgat tccgacctca ttaagcagct ctaatgcgct	10080
gttaatcact ttacttttat ctaatctaga catcattaat tcctaatttt tgttgacact	10140
ctatcattga tagagttatt ttaccactcc ctatcagtga tagagaaaag tga	10193

<210> 2
<211> 10250
<212> DNA
<213> Artificial Sequence

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

<400> 2
gcgaaaaatc aataaggagg caacaagatg tgcgaaaaac atcttaatca tgctaaggag 60
gttttctaata gaacatcttc aagaaaatgg aggaatttaa ttatgaacaa ctggtctact 120
tctacgacag cgaaacggaa ctcaaaggta ttacctgtat acacaacaca actttagggc 180
cggcattggg cggtacccgc ctttggaact ataactctga ggaagatgcc gttgaagacg 240
taatccgtct ggctcggggc atgacttaca aagcggcttg cgccggctctg aatctgggcg 300
gcggtaaaac cgtgctgata ggtgatgcta aaaagattaa atcagagtcc tacttccgtg 360
gactggggcg ctacgttcag tcgctgaacg gcagatatat caccgcggaa gacgtaaata 420
cttctacgaa ggatatggca tacgttgcta tggaaactga ctatgtggta ggccctgggag 480
gtaaatccgg caaccctagt ccagttactg cttacgggtg atttatgggt atcaaagcgg 540
cgctgatgaa aaaatttgag gatagctcta ttgaaggccg aaccttcgca gtgcagggtg 600
ctgggcagac gggttactat cttatcgatt acctcctagg caacaacaag ttcaaagaaa 660
aggctaaaaa gatttacttc accgaaatta acgagagcta tatcgagcgt atgaacaaag 720
aacatccgga agttgaattt atttccccgg acaaaatcta ctcgctggaa gtagacgtct 780
tcgtgccctg cgccctgggc aaaatcgtaa atgacaaaac tatcgatgaa ttaagtgtc 840
cgatcatcgc aggtactgca aacaacgtac tggaaaggga agcgcacggc aacatgctta 900
aagaacgtgg cattctttac gccccggact atgtgatcaa tgctggtggg ctgatcaacg 960
tttaccacga gctgaacggt tacaataaag agaacgctat tctggaagtg gaattaattt 1020
atgatgcct actggaaata ttcaacatcg ctgattctct gaacatcagc accaatatcg 1080
ctgccaacga gttcgcggaa aaacgtatca agcaaattaa gtccttgaaa aacaacttca 1140
ttaaacgcta agtcaataaa ggcatataaa aggagggttaa taacatgaaa gttaaagtaa 1200
aacatcttaa tcatgcacag gagactttct aatgtcgacg acaaccgttg gtgactactt 1260
gctgtatcgc ttaaacgaaa tcggcattga gcacctcttc ggagtgccag gtgattacaa 1320

tctgcaat	ttt	ctggatcatg	taatcgacca	ccctcagctg	acttgggtcg	gctgcactaa	1380
cgaacttaac		gctgcctacg	cagctgatgg	ttatgcgcgt	tgtcgtccgg	ctgcggcact	1440
gctgaccacc		ttcgggggttg	gcgaactgag	cgctattaat	ggcatcgcag	gttcctacgc	1500
ggagtatctg		ccggtaatac	atatcgttgg	tgcaccgagt	ctatcagccc	agcagcaggg	1560
cgacctgatt		caccactctc	ttggcgaagg	tgatTTTTcc	agcttcctga	ggatgtccca	1620
accggtgtct		gttgcgcagg	ctgctctgac	tcctgataac	gcatgcaagg	aaatcgaccg	1680
cgtactggcg		gaagtcctca	ttcagcgtcg	tcccggctac	ctgctgctgt	ctaccgacgt	1740
ggctgctgcg		ccggcggctc	tgccacaaag	cactctttct	ttgccgaccg	ccccggatca	1800
tcgcgcagtt		ctggctgctt	tcagcgacgc	tgctgagcag	atgctggctc	aggccaaaag	1860
cgtctctcta		ctggcgggact	ttctggctga	tcgtttcggg	gttactcgag	cactggccgc	1920
gtggcttcag		cagggttcgc	taccgcacgc	cactctgtta	atgggtaaag	gcgttctgag	1980
tgaacagcaa		ccagggttcg	tgggtaccta	cgctggtgcg	gcatctatcg	attcgacgcg	2040
tggcgcaatc		gaagaagctg	gggtaattat	cggagtggga	gtagatTTTt	ccgacactat	2100
cacagcaggc		ttctcgcagc	agatcgacgc	ccgccgtttt	atagacattc	aacccttctt	2160
ctctcgtatt		ggcgatcgcc	agtttgatca	cttgccgatg	caggctgccg	tcgcagccct	2220
gcatcaactg		tgtcttcggt	atcagcagca	gtggtctatc	accgctccta	gcccgcctgc	2280
actgccgccg		gctgctggta	gcgagctgtc	ccagaacgca	ttctggcagg	cgatgcagaa	2340
cttcatccgc		cctgggggacc	tgttggtggc	cgaccaaggt	actgcggcgt	tcggcgcagc	2400
ggcgctgcgc		ttaccgcaga	attgccagct	gcttgtgcag	ccgctgtggg	gctcaatcgg	2460
ttacagtctg		ccggccacct	ttggtgctca	gacggcagat	acagagcgtc	gtgtaatcct	2520
aatcattggc		gatggttcag	cgcaattaac	tattcaggaa	ctttccagta	tgatgcgtga	2580
cggcttga		aa	aa	aa	aa	aa	2640
tcacggcgcg		gagcaacgtt	ataacgat	at	at	at	2700
ggcgctgagt		gttcattgcc	cagcgcagag	ctggcgagtc	gttgaaacgg	tgcagctgac	2760
cgacgtaatg		aaagtcatcg	ctgcttctcc	gcgtctgagc	ttggtagaag	ttgttctgcc	2820
tgcaatggat		gtcccaccgc	tgctgcaagc	agtgagtgcc	gctctgaacc	agcgcaactc	2880
ctcttaagac		cagaggtaa	gaggtaa	aa	aa	aa	2940

atgcagggga gggtttctaa tgcgcagcat gcagtttgat gaggtaggtg caccctgaa	3000
agcgttctca tatgaagacc cgaccccgca aggggaaggaa gtagtcgtta ggatcgaagc	3060
ctgtgggtgtg tgccactctg atattcatct tcacgagggc tacttcgaca tgggcgggtg	3120
caataaagct gatgttactc gtgctcgcga actccctttt acattgggtc atgaaatcgt	3180
tggcgaagtg gtagcaactg gaccaggtgt caccggcgct aaaccgggcg acaaacttat	3240
tgtgtacccg tggatcgggt gggcgactg cccgaaatgc aacagtgggtg aggatcagtc	3300
ctgtgcgcgt ccacgtaacc tgggtgttca cgttgacggg ggctattcga cgcacgtaaa	3360
gataccggac gaaaaattcc tgttcgccta cgatgggtatt cctactgagt tagcgggaac	3420
ctatgcttgc agcggcatca ccgcttatgg tgcactgatg aaagcaaagg aagcggctga	3480
aagatctggc tacatcggtc tgattggcgc tgggtggcgtt ggcatggctg gtctgatgct	3540
ggccaaagca gcgatcgggg ctaaaactgt agtctttgat atcgacgacg caaaactgga	3600
agctgcgacc cgtgccgggg cggattacgt gttcaactcc ggtgcaaaag aaacacgcaa	3660
ggaagttagt aaactaacga atggtggcct gtctgggtgct gttgatttcg ttggcagcga	3720
taaaagcgct ctgtttggaa tcaacgcctt gggtcagaac ggcgtgctgg tcataattgg	3780
actgttcggg ggcgctatga ctgttcgggt acccctgttc ccgctgaaag ggatcacctg	3840
acgtggctca tacgtaggtt ccctgcaaga gatgagtgat atgatggagt tagttcgcgc	3900
tgggaaagtt cctccgatgc cggtaaaaac tcggccactg gacgctgcct gggaaaccct	3960
tgaggatcta cgccatggta aaatcgtggg ccgtgttggt ctgaccccat aagaaggaga	4020
tatacatatg acccatcaat taagatcgcg cgatatcatc gctctgggct ttatgacatt	4080
tgcgttggtt gtcggcgcag gtaacattat tttccctcca atggtcggct tgcaggcagg	4140
cgaacacgtc tggactgcgg cattcggctt cctcattact gccgttggcc taccggtatt	4200
aacggtagtg gcgctggcaa aagttggcgg cgggtgttgac agtctcagca cgccaattgg	4260
taaagtcgct ggcgtactgc tggcaacagt ttgttacctg gcggtggggc cgctttttgc	4320
tacgccgcgt acagctaccg tttcttttga agtgggcatt gcgccgctga cgggtgattc	4380
cgcgctgccg ctgtttatct acagcctggg ctatttcgct atcgttattc tggtttcgct	4440
ctatccgggc aagctgctgg ataccgtggg caacttcctt gcgccgctga aaattatcgc	4500
gctggtcatc ctgtctgttg ccgcaattat ctggccggcg gggttctatca gtacggcgac	4560

tgaggcttat caaaacgctg cgtttttctaa cggcttcgtc aacggctatc tgaccatgga	4620
tacgctgggc gcaatggtgt ttggtatcgt tattgttaac gcggcgcgtt ctcgtggcgt	4680
taccgaagcg cgtctgctga cccgttatac cgtctgggct ggccctgatgg cgggtgttgg	4740
tctgactctg ctgtacctgg cgctgttccg tctgggttca gacagcgcgt cgctggtcga	4800
tcagtctgca aacggtgcgg cgatcctgca tgcttacgtt cagcatacct ttggcggcgg	4860
cggtagcttc ctgctggcgg cgttaatctt catcgcctgc ctggtcacgg cggttggcct	4920
gacctgtgct tgtgcagaat tcttcgcca gtacgtaccg ctctcttatc gtacgtgggt	4980
gtttatcctc ggcggcttct cgatggtgggt gtctaacctc ggcttgagcc agctgattca	5040
gatctctgta ccggtgctga ccgccattta tccgccgtgt atcgcaactgg ttgtattaag	5100
ttttacacgc tcatgggtggc ataattcgtc ccgcgtgatt gctccgccga tgtttatcag	5160
cctgcttttt ggtatttctg acgggatcaa ggcatctgca ttcagcgata tcttaccgtc	5220
ctgggcgcag cgtttaccgc tggccgaaca aggtctggcg tggttaatgc caacagtgggt	5280
gatggtgggt ctggccatta tctgggatcg tgcggcaggt cgtcaggtga cctccagcgc	5340
tcactaatac gcatggcatg gatgaccgat ggtagtgtgg gctctcccca tgcgagagta	5400
gggaactgcc aggcatacaa taaaacgaaa ggctcagtcg aaagactggg cctttcgttt	5460
tatctgttgt ttgtcggatga acgctctcct gagtaggaca aatccgccgg gagcggattt	5520
gaacgttgcg aagcaacggc ccggagggtg gcgggcagga cggccgcat aaactgccag	5580
gcatcaaatt aagcagaagg ccatcctgac ggatggcctt ttgctgtggc cagtgccaa	5640
cttgcatgcg tgccagctgc attaatgaag aaatcatgct ggaagaataa cagctcactc	5700
aaaggcggta gtacgggttt tgctgcccgc aaacgggctg ttctgggtgt gctagtttgt	5760
tatcagaatc gcagatccgg cttcagccgg ttgcccggct gaaagcgcta tttcttccag	5820
aattgccatg attttttccc cacgggaggc gtcactggct cccgtgttgt cggcagcttt	5880
gattcgataa gcagcatcgc ctgtttcagg ctgtctatgt gtgactgttg agctgtaaca	5940
agttgtctca ggtgttcaat ttcatgttct agttgctttg ttttactgggt ttcacctgtt	6000
ctattaggtg ttacatgctg ttcatctgtt acattgtcga tctgttcatg gtgaacagct	6060
ttgaatgcac caaaaactcg taaaagctct gatgtatcta tcttttttac accgttttca	6120
tctgtgcata tggacagttt tccctttgat atgtaacgggt gaacagtgtg tctacttttg	6180

tttgtagtc ttgatgcttc actgatagat acaagagcca taagaacctc agatccttcc	6240
gtatttagcc agtatgttct ctagtgtggt tcgttgtttt tgcgtgagcc atgagaacga	6300
accattgaga tcatacttac tttgcatgtc actcaaaaat tttgcctcaa aactgggtgag	6360
ctgaatTTTT gcagttaaag catcgtgtag tgTTTTtctt agtccgttat gtaggtagga	6420
atctgatgta atggttgttg gtattttgtc accattcatt tttatctggt tgttctcaag	6480
ttcggttacg agatccattt gtctatctag ttcaacttgg aaaatcaacg tatcagtcgg	6540
gcggcctcgc ttatcaacca ccaatttcat attgctgtaa gtgttttaaT cttacttat	6600
tggtttcaaa acccattggt taagcctttt aaactcatgg tagttatttt caagcattaa	6660
catgaactta aattcatcaa ggctaattctc tatatttgcc ttgtgagttt tcttttTgtg	6720
tagttctttt aataaccact cataaatcct catagagtat ttgttttcaa aagacttaac	6780
atgttccaga ttatatTTta tgaatTTTT taactggaaa agataaggca atatctcttc	6840
actaaaaact aattctaatt tttcgcttga gaacttggca tagtttTgtcc actggaaaat	6900
ctcaaagcct ttaaccaaag gattcctgat ttccacagtt ctcgtcatca gctctctggt	6960
tgcttttagct aatacaccat aagcattttc cctactgatg tTcatcatct gagcgtattg	7020
gttataagtg aacgataccg tccgttcttt ccttTtaggg ttttcaatcg tggggttgag	7080
tagtgccaca cagcataaaa ttagcttgggt tTcatgctcc gttaagTcat agcgactaat	7140
cgctagtTca tttgctttga aaacaactaa tTcagacata catctcaatt ggtctaggTg	7200
attttaatca ctataccaat tgagatgggc tagtcaatga taattactag tctttttcct	7260
ttgagttgtg ggtatctgta aattctgcta gacctttgct ggaaaacttg taaattctgc	7320
tagaccctct gtaaattccg ctagaccttt gtgtgttttt tttgtttata ttcaagtggT	7380
tataatttat agaataaaga aagaataaaa aaagataaaa agaataTatc ccagccctgt	7440
gtataactca ctactttagt cagttccgca gtattacaaa aggatgtcgc aaacgctgtt	7500
tgctcctcta caaaacagac cttaaaaccc taaaggctta agtagcacc tcgcaagctc	7560
gggcaaTcg ctgaatattc cttttgtctc cgaccatcag gcacctgagt cgctgtcttt	7620
ttcgtgacat tcagttcgct gcgctcacgg ctctggcagt gaatgggggt aaatggcact	7680
acaggcgcct tttatggatt catgcaagga aactacccat aatacaagaa aagcccgtca	7740
cgggccttctc agggcgTTTT atggcgggtc tgctatgtgg tgctatctga ctttttgctg	7800

ttcagcagtt cctgccctct gattttccag tctgaccact tcggattatc ccgtgacagg	7860
tcattcagac tggctaatagc acccagtaag gcagcggtat catcaacagg cttacccgtc	7920
ttactgtctt ttctacgggg tctgacgctc agtggaacga aaactcacgt taagggattt	7980
tggtcatgag attatcaaaa aggatcttca cctagatcct tttaaattaa aaatgaagtt	8040
ttaaatcaat ctaaagtata tatgagtaaa cttgggtctga cagttaccaa tgcttaatca	8100
gtgaggcacc tatctcagcg atctgtctat ttcgttcac catagttgcc tgactccccg	8160
tcgtgtagat aactacgata cgggagggct taccatctgg cccagtgct gcaatgatac	8220
cgcgagaacc acgctcaccg gctccagatt tatcagcaat aaaccagcca gccggaagg	8280
ccgagcgcag aagtggctct gcaactttat ccgcctccat ccagtctatt aattgttgcc	8340
gggaagctag agtaagtagt tcgccagtta atagtttgcg caacgttggt gccattgcta	8400
caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc attcagctcc ggttccaac	8460
gatcaaggcg agttacatga tccccatgt tgtgcaaaaa agcggtagc tccttcggtc	8520
ctccgatcgt tgtcagaagt aagtggccg cagtgttatc actcatggtt atggcagcac	8580
tgcataattc tcttactgtc atgccatccg taagatgctt ttctgtgact ggtgagtact	8640
caaccaagtc attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtcaa	8700
tacgggataa taccgcgcca catagcagaa ctttaaaagt gctcatcatt ggaaaacgtt	8760
cttcggggcg aaaactctca aggatcttac cgctgttgag atccagttcg atgtaacca	8820
ctcgtgcacc caactgatct tcagcatctt ttactttcac cagcgtttct gggtgagcaa	8880
aaacaggaag gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa tgttgaatac	8940
tcatactctt cttttttcaa tattattgaa gcatttatca gggttattgt ctcagagcg	9000
gatacatatt tgaatgtatt tagaaaaata aacaaatagg ggttccgcgc acatttcccc	9060
gaaaagtgcc acctgacgtc taagaaacca ttattatcat gacattaacc tataaaaaata	9120
ggcgtatcac gaggcccttt cgtctcgcgc gtttcgggtga tgacggtgaa aacctctgac	9180
acatgcagct cccggagacg gtcacagctt gtctgtaagc ggatgccggg agcagacaag	9240
cccgtcaggg cgcgtcagcg ggtgttggcg ggtgtcgggg ctggcttaac tatgcggcat	9300
cagagcagat tgtactgaga gtgcaccata tgcggtgtga aataccgcac agatgcgtaa	9360
ggagaaaaata ccgcatcagg cgccattcgc cattcaggct gcgcaactgt tgggaagggc	9420

gatcgggtgcg ggcctcttcg ctattacgcc agctggcgaa aggggggatgt gctgcaaggc	9480
gattaagttg ggtaacgcc gggttttccc agtcacgacg ttgtaaaacg acggccagtg	9540
aattcgttaa gaccacttt cacatttaag ttgtttttct aatccgcata tgatcaattc	9600
aaggccgaat aagaaggctg gctctgcacc ttggtgatca aataattcga tagcttgtcg	9660
taataatggc ggcatactat cagtagtagg tgtttccctt tcttcttttag cgacttgatg	9720
ctcttgatct tccaatacgc aacctaaagt aaaatgcccc acagcgctga gtgcatataa	9780
tgcatctct agtgaaaaac cttgttggca taaaaaggct aattgatttt cgagagtttc	9840
atactgtttt tctgtaggcc gtgtacctaa atgtactttt gctccatcgc gatgacttag	9900
taaagcacat ctaaaacttt tagcgttatt acgtaaaaaa tcttgccagc tttccccttc	9960
taaagggcaa aagtgagtat ggtgcctatc taacatctca atggctaagg cgtcgagcaa	10020
agcccgctta ttttttacct gccaatataa tgtaggctgc tctacaccta gcttctgggc	10080
gagtttacgg gttgttaaac cttcgattcc gacctatta agcagctcta atgcgctgtt	10140
aatcacttta cttttatcta atctagacat cattaattcc taatttttgt tgacactcta	10200
tcattgatag agttatttta ccactcccta tcagtgatag agaaaagtga	10250

<210> 3

<211> 10280

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 3

gcgaaaaatc aataaggagg caacaagatg tgcgaaaaac atcttaatca tgctaaggag	60
gttttctaata gaacatcttc aagaaaatgg aggaatttaa ttatgaacaa ctggtctact	120
tctacgacag cgaaacggaa ctcaaaggta ttacctgtat acacaacaca actttagggc	180
cggcattggg cggtacccgc ctttggaaact ataactctga ggaagatgcc gttgaagacg	240
taatccgtct ggctcggggc atgacttaca aagcggcttg cgccggcttg aatctgggcg	300
gcggtaaaac cgtgctgata ggtgatgcta aaaagattaa atcagagtcc tacttccgtg	360
gactggggcg ctacgttcag tcgctgaacg gcagatatat caccgcggaa gacgtaaata	420
cttctacgaa ggatatggca tacgttgcta tggaaactga ctatgtggta ggcctgggag	480

gtaaatccgg caaccctagt ccagttactg cttacgggtgc atttatgggt atcaaagcgg	540
cgctgatgaa aaaatttgag gatagctcta ttgaaggccg aaccttcgca gtgcagggtg	600
ctgggcagac gggttactat cttatcgatt acctcctagg caacaacaag ttcaaagaaa	660
aggctaaaaa gatttacttc accgaaatta acgagagcta tatcgagcgt atgaacaaag	720
aacatccgga agttgaattt atttccccgg acaaaatcta ctcgctggaa gtagacgtct	780
tcgtgccctg cgccctgggc aaaatcgta atgacaaaac tatcgatgaa tttaaagtgc	840
cgatcatcgc aggtactgca aacaacgtac tggaaaggga agcgcacggc aacatgctta	900
aagaacgtgg cattctttac gccccggact atgtgatcaa tgctgggtggg ctgatcaacg	960
tttaccacga gctgaacggt tacaataaag agaacgctat tctggaagtg gaattaattt	1020
atgatcgcct actggaaata ttcaacatcg ctgatttctt gaacatcagc accaatatcg	1080
ctgccaacga gttcgcggaa aaacgtatca agcaaattaa gtccttgaaa aacaacttca	1140
ttaaacgcta agtcaataaa ggcatataaa aggaggttaa taacatgaaa gttaaagtaa	1200
aacatcttaa tcatgcacag gagactttct aatgtcgacg acaaccgttg gtgactactt	1260
gctgtatcgc ttaaacgaaa tcggcattga gcacctcttc ggagtgccag gtgattacaa	1320
tctgcaattt ctggatcatg taatcgacca cctcagctg acttgggtcg gctgcactaa	1380
cgaacttaac gctgcctacg cagctgatgg ttatgcgcgt tgcgtccgg ctgcggcact	1440
gctgaccacc ttcgggggttg gcgaactgag cgctattaat ggcatcgcag gttcctacgc	1500
ggagtatctg ccggtaatatc atatcgttgg tgcaccgagt ctatcagccc agcagcaggg	1560
cgacctgatt caccactctc ttggcgaagg tgatTTTTCC agcttcctga ggatgtccca	1620
accggtgtct gttgcgcagg ctgctctgac tcctgataac gcatgcaagg aaatcgaccg	1680
cgtactggcg gaagtcctca ttcagcgtcg tcccggtac ctgctgctgt ctaccgacgt	1740
ggctgctgcg ccggcggctc tgccacaaag cactctttct ttgccgaccg ccccggtatca	1800
tcgcgcagtt ctggctgctt tcagcgacgc tgctgagcag atgctggctc aggccaaaag	1860
cgtctctcta ctggcggact ttctggctga tcgtttcggg gttactcgag cactggccgc	1920
gtggcttcag caggttccgc taccgcacgc cactctgtta atgggtaaag gcgttctgag	1980
tgaacagcaa ccagggttcg tgggtaccta cgctgggtcg gcatctatcg attcgacgcg	2040
tggcgcaatc gaagaagctg gggtaattat cggagtggga gttagatttt ccgacactat	2100

cacagcaggc	ttctcgcagc	agatcgacgc	ccgccgtttt	atagacattc	aacccttctt	2160
ctctcgtatt	ggcgatcgcc	agtttgatca	cttgccgatg	caggctgccg	tcgcagccct	2220
gcatcaactg	tgtcttcgtt	atcagcagca	gtggtctatc	accgctccta	gcccgcctgc	2280
actgccgccg	gctgctggta	gcgagctgtc	ccagaacgca	ttctggcagg	cgatgcagaa	2340
cttcatccgc	cctggggacc	tgttggtggc	cgaccaaggt	actgcggcgt	tcggcgcagc	2400
ggcgctgcgc	ttaccgcaga	attgccagct	gcttgtgcag	ccgctgtggg	gctcaatcgg	2460
ttacagtctg	ccggccacct	ttggtgctca	gacggcagat	acagagcgtc	gtgtaatcct	2520
aatcattggc	gatggttcag	cgcaattaac	tattcaggaa	ctttccagta	tgatgcgtga	2580
cggcttgaaa	cctatcatct	ttctcctgaa	caacaacggt	tacaccgttg	aacgggcgat	2640
tcacggcgcg	gagcaacgtt	ataacgatat	cgctgcttgg	aattggaccc	aactgccccca	2700
ggcgctgagt	gttcattgcc	cagcgcagag	ctggcgagtc	gttgaaacgg	tgacagctgac	2760
cgacgtaatg	aaagtcatcg	ctgcttctcc	gcgtctgagc	ttggtagaag	ttgttctgcc	2820
tgcaatggat	gtcccaccgc	tgctgcaagc	agtgagtgcc	gctctgaacc	agcgcaactc	2880
ctcttaagac	cagaggtaag	gaggtaacaa	ccatgcgagt	gttgaagaaa	catcttaatc	2940
atgcagggga	gggttttctaa	tgacagcgga	gcagcaaaat	ggggtatccg	actcacgccg	3000
tttcgaattt	caggaatttg	gtggccctat	cgccccacag	acctatcagc	tccccgcacc	3060
ggctagcgat	gaagttttgt	taaaggtgaa	ctactgcggt	gtctgtcaca	gtgatgttca	3120
tcttcacgac	ggctacttcg	agctgggtgg	cgataaacgt	ctgaacttcg	ctatgccgct	3180
gccgctgacg	ctgggtcacg	aagtaattgg	caccgttgtg	gctgtcggcg	accaggttac	3240
tggtgtaaaa	ccgggggacc	agcgactgat	ctatccgtgg	ataggttgcg	gaaaatgcgg	3300
cgcgtgtcaa	aaaggagaag	aaaacctgtg	cgttactcct	gcacatctgg	gcgtgaacaa	3360
gccgggcggg	tacgctgata	acatcgttgt	acccatttct	cgctaccttc	tggacatttc	3420
gggtctgaac	ccgggtgatg	ccgctaccct	cgcgtgctcc	ggcctgacca	ctttcagcgc	3480
gatcaacaaa	gtgttgccgc	ttgcagatga	ccagtggatt	gttgttatcg	gttgtggtgg	3540
cctcggccag	atggcgctgc	gtatcctgca	agctatggga	attggcaatg	ttatcggtat	3600
tgacctgtct	gaagagaaac	ggaaactggc	tcatgaaagc	ggtgcacgtc	actccttcga	3660
tccaaacact	ccgaagctga	accgcgtggt	cgccgaaacc	tgcccgggta	cggtacaggc	3720

cgcgttagac tttgtgggca atgagcaaac tgctcagctg gcactgtctc tgcttgga	3780
aggtggcaaa tatgttcctg tctgggctgca cggcggcgag ctgcgttacc cattgccgat	3840
catcacgaac aaagctgtaa gtatcatcgg ttcttacgtt ggtaccctga aagaactgga	3900
agacttagtt gctttcgcca aggaaaagaa tctgccgcca attcatattg aacaccgccc	3960
gctggaatcg gcggctcagg ccgtagagga cctggaaaaa ggacaggttg ctgggcgtgt	4020
tatcctggat gcaggtaact aagaaggaga tatacatatg acccatcaat taagatcgcg	4080
cgatatcatc gctctgggct ttatgacatt tgcgttggtc gtcggcgag gtaacattat	4140
tttccctcca atggctcggct tgcaggcagg cgaacacgtc tggactgcgg cattcggctt	4200
cctcattact gccgttggcc taccggtatt aacggtagtg gcgctggcaa aagttggcgg	4260
cgggtgtgac agtctcagca cgccaattgg taaagtcgct ggcgactgc tggcaacagt	4320
ttgttacctg gcgggtggggc cgctttttgc tacgccgct acagctaccg tttcttttga	4380
agtgggcatt gcgccgctga cgggtgattc cgcgctgccg ctgtttatctt acagcctggt	4440
ctatttcgct atcgttattc tggtttcgct ctatccgggc aagctgctgg ataccgtggg	4500
caacttcctt gcgccgctga aaattatcgc gctggtcac cgtctgttg ccgcaattat	4560
ctggccggcg ggttctatca gtacggcgac tgaggcttat caaacgctg cgttttctaa	4620
cggcttcgtc aacggctatc tgaccatgga tacgctgggc gcaatgggtt ttggtatcgt	4680
tattgttaac gcggcgcggt ctcgtggcgt taccgaagcg cgtctgctga cccgttatac	4740
cgtctgggct ggcctgatgg cgggtgttgg tctgactctg ctgtacctgg cgctgttccg	4800
tctgggttca gacagcgct cgctggtcga tcagtctgca aacgggtgcgg cgatcctgca	4860
tgcttacgtt cagcatacct ttggcggcgg cggtagcttc ctgctggcgg cgtaaatctt	4920
catgcctgc ctggtcacgg cggttggcct gacctgtgct tgtgcagaat tcttcgcca	4980
gtacgtaccg ctctcttacc gtacgtgggt gtttatcctc ggcggttct cgatgggtgg	5040
gtctaacctc ggcttgagcc agctgattca gatctctgta ccggtgctga ccgccattta	5100
tccgccgtgt atcgactgg ttgtattaag ttttacacgc tcatgggtggc ataattcgtc	5160
ccgcgtgatt gctccgccga tgtttatcag cctgcttttt ggtattctcg acgggatcaa	5220
ggcatctgca ttcagcgata tcttaccgtc ctgggcgcag cgtttaccgc tggccgaaca	5280
aggtctggcg tggttaatgc caacagtggg gatgggtgggt ctggccatta tctgggatcg	5340

tgcggcaggt cgtcaggtga cctccagcgc tcactaatac gcatggcatg gatgaccgat	5400
ggtagtgtgg gctctcccca tgcgagagta gggaactgcc aggcacaaaa taaaacgaaa	5460
ggctcagtcg aaagactggg cctttcgttt tatctgttgt ttgtcgggtga acgctctcct	5520
gagtaggaca aatccgccgg gagcggattt gaacgttgcg aagcaacggc ccggagggtg	5580
gcgggcagga cgcccgccat aaactgccag gcatcaaatt aagcagaagg ccatcctgac	5640
ggatggcctt tttgcgtggc cagtgccaaag cttgcatgcg tgccagctgc attaatgaag	5700
aaatcatgct ggaagaataa cagctcactc aaaggcggta gtacggggtt tgctgcccgc	5760
aaacgggctg ttctgggtgt gctagtttgt tatcagaatc gcagatccgg cttcagccgg	5820
tttgccggct gaaagcgcta tttcttccag aattgccatg attttttccc cacgggaggc	5880
gtcactggct cccgtgttgt cggcagcttt gattcgataa gcagcatcgc ctgtttcagg	5940
ctgtctatgt gtgactgttg agctgtaaca agttgtctca ggtgttcaat ttcattgtct	6000
agttgctttg ttttactggg ttcacctgtt ctattagggt ttacatgctg ttcattgtt	6060
acattgtcga tctgttcatg gtgaacagct ttgaatgcac caaaaactcg taaaagctct	6120
gatgtatcta tcttttttac accgttttca tctgtgcata tggacagttt tccctttgat	6180
atgtaacggg gaacagttgt tctacttttg tttgttagtc ttgatgcttc actgatagat	6240
acaagagcca taagaacctc agatccttcc gtatttagcc agtatgttct ctagtgtggg	6300
tcgttgtttt tgcgtgagcc atgagaacga accattgaga tcatacttac tttgcatgtc	6360
actcaaaaat tttgcctcaa aactggtgag ctgaattttt gcagttaaag catcgtgtag	6420
tgtttttctt agtccgttat gtaggtagga atctgatgta atggttgttg gtattttgtc	6480
accattcatt tttatctggg tgttctcaag ttcggttacg agatccattt gtctatctag	6540
ttcaacttgg aaaatcaacg tatcagtcgg gcggcctcgc ttatcaacca ccaatttcat	6600
attgctgtaa gtgttttaaat ctttacttat tggtttcaaa acccattggg taagcctttt	6660
aaactcatgg tagttatttt caagcattaa catgaactta aattcatcaa ggctaattctc	6720
tatatattgcc ttgtgagttt tcttttgtgt tagttctttt aataaccact cataaatcct	6780
catagagtat ttgttttcaa aagacttaac atgttccaga ttatatatta tgaatttttt	6840
taactggaag agataaggca atatctcttc actaaaaact aattctaatt tttcgcttga	6900
gaacttggca tagtttgtcc actggaaaat ctcaaagcct ttaaccaaag gattcctgat	6960

ttccacagtt ctcgtcatca gctctctggt tgcttttagct aatacaccat aagcattttc	7020
cctactgatg ttcacatctt gagcgtattg gttataagtg aacgataccg tccgtttcttt	7080
ccttgtaggg ttttcaatcg tgggggttgag tagtgccaca cagcataaaa ttagcttggt	7140
ttcatgctcc gttaagtcac agcgactaat cgctagttca tttgctttga aaacaactaa	7200
ttcagacata catctcaatt ggtctaggtg attttaatca ctataccaat tgagatgggc	7260
tagtcaatga taattactag tccttttctt ttgagttgtg ggtatctgta aattctgcta	7320
gacctttgct ggaaaacttg taaattctgc tagaccctct gtaaattccg ctagaccttt	7380
gtgtgttttt tttgtttata ttcaagtggg tataatttat agaataaaga aagaataaaa	7440
aaagataaaa agaataagtc ccagccctgt gtataactca ctactttagt cagttccgca	7500
gtattacaaa aggatgtcgc aaacgctggt tgctcctcta caaacagac cttaaaaccc	7560
taaaggctta agtagcacc tcgcaagctc gggcaaactg ctgaatattc cttttgtctc	7620
cgaccatcag gcacctgagt cgctgtcttt ttcgtgacat tcagttcgct gcgctcacgg	7680
ctctggcagt gaatgggggt aaatggcact acaggcgcct tttatggatt catgcaagga	7740
aactacccat aatacaagaa aagcccgtca cgggcttctc agggcgtttt atggcgggtc	7800
tgctatgtgg tgctatctga ctttttgctg ttcagcagtt cctgccctct gattttccag	7860
tctgaccact tcggattatc ccgtgacagg tcattcagac tggctaattgc acccagtaag	7920
gcagcgggtat catcaacagg cttaccctgc ttactgtctt ttctacgggg tctgacgctc	7980
agtggaacga aaactcacgt taagggattt tggatcatgag attatcaaaa aggatcttca	8040
cctagatcct tttaaattaa aaatgaagtt ttaaataaat cttaaagtata tatgagtaaa	8100
cttggcttga cagttaccaa tgcttaatca gtgaggcacc tatctcagcg atctgtctat	8160
ttcgttcatc catagttgcc tgactccccg tcgtgtagat aactacgata cgggagggct	8220
taccatctgg cccagtgct gcaatgatac cgcgagaacc acgctcaccg gctccagatt	8280
tatcagcaat aaaccagcca gccggaaggg ccgagcgcag aagtggtcct gcaactttat	8340
ccgcctccat ccagtctatt aattgttgcc gggaagctag agtaagtagt tcgccagtta	8400
atagtttgcg caacgttggt gccattgcta caggcatcgt ggtgtcacgc tcgtcgtttg	8460
gtatggcttc attcagctcc ggttcccaac gatcaaggcg agttacatga tccccatgt	8520
tgtgcaaaaa agcgggttagc tccttcggtc ctccgatcgt tgtcagaagt aagttggccg	8580

cagtgttatc actcatggtt atggcagcac tgcataattc tcttactgtc atgccatccg	8640
taagatgctt ttctgtgact ggtgagtact caaccaagtc attctgagaa tagtgtatgc	8700
ggcgaccgag ttgctcttgc ccggcgtcaa tacgggataa taccgcgcca catagcagaa	8760
ctttaaaagt gctcatcatt ggaaaacgtt cttcggggcg aaaactctca aggatcttac	8820
cgctgttgag atccagttcg atgtaacca ctcgtgcacc caactgatct tcagcatctt	8880
ttactttcac cagcgtttct gggtgagcaa aaacaggaag gcaaaatgcc gcaaaaaagg	8940
gaataagggc gacacggaaa tgttgaatac tcatactctt cttttttcaa tattattgaa	9000
gcatttatca gggttattgt ctcatgagcg gatacatatt tgaatgtatt tagaaaaata	9060
aacaaatagg ggttccgcgc acatttcccc gaaaagtgcc acctgacgtc taagaaacca	9120
ttattatcat gacattaacc tataaaaata ggcgtatcac gaggcccttt cgtctcgcgc	9180
gtttcgggtga tgacggtgaa aacctctgac acatgcagct cccggagacg gtcacagctt	9240
gtctgtaagc ggatgccggg agcagacaag cccgtcaggg cgctcagcg ggtgttggcg	9300
ggtgtcgggg ctggcttaac tatgcggcat cagagcagat tgtactgaga gtgcaccata	9360
tgcgggtgta aataccgcac agatgcgtaa ggagaaaata ccgcatcagg cgccattcgc	9420
cattcaggct gcgcaactgt tgggaagggc gatcgggtcg ggctcttcg ctattacgcc	9480
agctggcgaa aggggggatgt gctgcaaggc gattaagttg ggtaacgcca gggttttccc	9540
agtcacgacg ttgtaaaacg acggccagtg aattcggttaa gaccacttt cacatttaag	9600
ttgtttttct aatccgcata tgatcaattc aaggccgaat aagaaggctg gctctgcacc	9660
ttggtgatca aataattcga tagcttgtcg taataatggc ggcatactat cagtagtagg	9720
tgtttccctt tcttcttttag cgacttgatg ctcttgatct tccaatacgc aacctaaagt	9780
aaaatgcccc acagcgctga gtgcatataa tgcattctct agtgaaaaac cttgttggca	9840
taaaaaggct aattgatttt cgagagtttc atactgtttt tctgtaggcc gtgtacctaa	9900
atgtactttt gctccatcgc gatgacttag taaagcacat ctaaaacttt tagcgttatt	9960
acgtaaaaaa tcttgccagc tttccccttc taaagggcaa aagtgagtat ggtgcctatc	10020
taacatctca atggctaagg cgtcgagcaa agcccgtta ttttttacat gccaatacaa	10080
tgtaggctgc tctacaccta gcttctgggc gagtttacgg gttgttaaac cttcgattcc	10140
gacctatta agcagctcta atgcgctgtt aatcacttta cttttatcta atctagacat	10200

cattaattcc taatTTTTgt tgacactcta tcattgatag agttatttta ccactcccta	10260
tcagtgatag agaaaagtga	10280

<210> 4
 <211> 10202
 <212> DNA
 <213> Artificial Sequence

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

<400> 4	
gcgaaaaatc aataaggagg caacaagatg tgcgaaaaac atcttaatca tgctaaggag	60
gttttctaat ggtagagatc caggctttgc cggaaacttc catttttggg caaatcgcag	120
accaccagca tgaacaggtg gtcttctgcc acgatacga aaccggcctc cgtgcgataa	180
tcggtattca taacacggtt cttggccccg ccttaggtgg aactcgcagtg tggcactatg	240
ctaccgaggc agaagcgctg aatgacgttc tgcgtctgtc tcgcggtatg acctacaagg	300
ctgctatctc gggcctgaac ctgggtggcg gtaaagcagt aatcattggg gatgccaaaa	360
caatcaaaac cgaagcgctg ctgcggaaat tcggcagatt cgtgcagaac ctgaatggta	420
aatacatcac tgctgaagac gttaacatga ctacaaagga tatggagtac attaggatgg	480
aaaccaaaca cgtcgctggc ttacctgaaa gtatgggtgg aagcggtgac ccgtcaccgg	540
taactgcata tggtagctac atgggcatga aagcggcggc caaaaaggcg tttggctctg	600
attccctggc tggcaaactg atcgctgttc aagggtgtggg tcatgttggc acttatctgc	660
ttgagcattt gaccaaagaa ggtgctcaga ttgtgctgac tgactactat aaggaactg	720
ccgaggaagc aggcgcgcgt tttggcgcac aggttggttg cctggacgat atctacgatc	780
aagaggtcga catttactct ccatgtgctc tcggtgctac catcaacgat gacactatcg	840
atgcctgcg ttgcgctggt gtagccggtt gcgcaaaca ccagctgaaa gaagaaaacg	900
tccacggtcc ggcgctggtt gagcgcggga tagtatacgc cccagacttc ctgatcaatg	960
cagggggcct gattaactg tatagcgaag ttacagggtc taccgctcag ggggctttaa	1020
ctcagaccga aaaaatctat gactacacac tccaagttct ggaaaaagcc gcggctgaag	1080
gtctgcaccc gcagcaggct gcgatccgtc aggcggaaca acgcatcgct gcaattggta	1140

aggtgaaaag	cacctactaa	gtcaataaag	gcatataaaa	ggaggttaat	aacatgaaag	1200
ttaaagtaaa	acatcttaat	catgcacagg	agactttcta	atgtcgacga	caaccgttgg	1260
tgactacttg	ctgtatcgct	taaacgaaat	cggcattgag	cacctcttcg	gagtgccagg	1320
tgattacaat	ctgcaatttc	tggatcatgt	aatcgaccac	cctcagctga	cttgggtcgg	1380
ctgcactaac	gaacttaacg	ctgcctacgc	agctgatggg	tatgcgcggt	gtcgtccggc	1440
tgcggcactg	ctgaccacct	tgggggttgg	cgaactgagc	gctattaatg	gcatcgacgg	1500
ttcctacgcg	gagtatctgc	cggtaataca	tatcgttggg	gcaccgagtc	tatcagccca	1560
gcagcagggc	gacctgattc	accactctct	tggcgaagg	gattttttcca	gcttcctgag	1620
gatgtcccaa	ccggtgtctg	ttgcgcaggc	tgctctgact	cctgataacg	catgcaagga	1680
aatcgaccgc	gtactggcgg	aagtcctcat	tcagcgctgt	cccggctacc	tgctgctgtc	1740
taccgacgtg	gctgctgcgc	cggcggctct	gccacaaagc	actctttctt	tgccgaccgc	1800
cccggatcat	cgcgagttc	tggctgcttt	cagcgacgct	gctgagcaga	tgctgggtca	1860
ggccaaaagc	gtctctctac	tggcggactt	tctggctgat	cgtttcgggt	ttactcgagc	1920
actggccgcg	tggcttcagc	aggttccgct	accgcacgcc	actctgttaa	tgggtaaagg	1980
cgttctgagt	gaacagcaac	cagggttcgt	gggtacctac	gctgggtgcgg	catctatcga	2040
ttcgacgcgt	ggcgcaatcg	aagaagctgg	ggtaattatc	ggagtgggag	ttagattttc	2100
cgacactatc	acagcaggct	tctcgcagca	gatcgacgcc	cgccgtttta	tagacattca	2160
acccttcttc	tctcgtattg	gcgatcgcca	gtttgatcac	ttgccgatgc	aggctgccgt	2220
cgcagccctg	catcaactgt	gtcttcgtta	tcagcagcag	tgggtctatca	ccgctcctag	2280
cccgcctgca	ctgccgccgg	ctgctggtag	cgagctgtcc	cagaacgcat	tctggcaggc	2340
gatgcagaac	ttcatccgcc	ctggggacct	gttgggtggc	gaccaaggta	ctgcggcggt	2400
cggcgcagcg	gcgctgcgct	taccgcagaa	ttgccagctg	cttgtgcagc	cgctgtgggg	2460
ctcaatcggt	tacagtctgc	cggccacctt	tgggtgctcag	acggcagata	cagagcgctg	2520
tgtaatccta	atcattggcg	atggttcagc	gcaattaact	attcaggaac	tttccagtat	2580
gatgcgtgac	ggcttgaaac	ctatcatctt	tctcctgaac	aacaacgggt	acaccgttga	2640
acgggcgatt	cacggcgcgg	agcaacgtta	taacgatatc	gctgcttgga	attggacca	2700
actgccccag	gcgctgagtg	ttcattgccc	agcgcagagc	tggcgagtcg	ttgaaacgg	2760

gcagctgacc gacgtaatga aagtcacgc tgcttctccg cgtctgagct tggtagaagt	2820
tgttctgcct gcaatggatg tcccaccgct gctgcaagca gtgagtgccg ctctgaacca	2880
gcgcaactcc tcttaagacc agaggtaagg aggtacaac catgagtg ttgaagaaac	2940
atcttaatca tgcaggggag ggtttctaata gcagacggcg ttcttgata agccaggcca	3000
cgaactta gtgcgctcg agatccgat acctaaagct gggcggtggc aagtcgttct	3060
ggaaattaaa gccgctggca tgtgccattc cgatctgcac gttctcgac gtggaatccc	3120
cctgccgggt caatttgtaa tgggcatga aatcggtgg actattcac agatcgcca	3180
ggagctgacc ggtttcaaac agggcgatct gtacgcagtc cacggcccga atccgtgtgg	3240
tatttgacc ctgtgcagag aaggatttga taacgactgc actacagtgg cgaaccgg	3300
tcaatgggtc ggactgggtc ttgacggcg ctaccagaag tatatccgta tcccgaacgt	3360
aaggctatc gttaaagttc cagaagggtgt ttcagctgag gcagctgcga gctgtactga	3420
tgagctactg acccgtacc gtgcactaaa acaggctggc gccagcaact ctactcgggt	3480
actgattctg ggtctgggtg gcttaggtct gaatgccctt aaactggcta agacctcgg	3540
cagttacgtt tacgcatctg acctgaaacc ttctgcgct gaagctgcta aggccgctgg	3600
ggcgatgaa gtgctggagt ccctgccga agaccgctg ggtgttgata tcgtgttaga	3660
cgtcgttggc gtgcagagca cttcaacct cgctcaaaaa cacgttggc cgctggcat	3720
cattgtacct gtaggcctgg catccccaca gcttctggtt aacctaacgg atctggcgct	3780
ccgcgaaatt cgtgttcagg gcactttttg gggcacgagc aatgagctgg ctgaatgtct	3840
ggcctgtgc cagctgggccc tgatcaaccc gaaatatact gtggtgcctc ttgaagaagc	3900
gccgaaatat atggaagcaa tggctcatgg gaaagtagaa ggtcgatatc tttccacc	3960
gtaagaagga gatatacata tgaccatca attaatgac gcgatatca tcgctctggg	4020
ctttatgaca ttgctgtgt tcgtcggcgc aggtaacatt atttccctc caatggctcg	4080
cttgaggca ggcgaacacg tctggactgc ggcattcggc ttcctatta ctgccgttg	4140
cctaccggtg ttaacggtg tggcgctggc aaaagtggc ggcggtgtg acagtctcag	4200
cacgccaatt ggtaaagtc ctggcgctact gctggcaaca gttgttacc tggcggtggg	4260
gccgctttt gctacgccgc gtacagctac cgttctttt gaagtgggca ttgcgccgct	4320
gacgggtgat tccgcgtgc cgctgtttat ttacagcctg gtctatttcg ctatcgttat	4380

tctggtttcg ctctatccgg gcaagctgct ggataccgtg ggcaacttcc ttgcgccgct	4440
gaaaattatc gcgctgggtca tcctgtctgt tgccgcaatt atctggccgg cgggttctat	4500
cagtacggcg actgaggctt atcaaaacgc tgcgttttct aacggcttcg tcaacggcta	4560
tctgaccatg gatacgtgg gcgcaatggt gtttggatc gttattgtta acgcggcgcg	4620
ttctcgtggc gttaccgaag cgcgtctgct gaccggttat accgtctggg ctggcctgat	4680
ggcgggtgtt ggtctgactc tgctgtacct ggcgctgttc cgtctgggtt cagacagcgc	4740
gtcgctggtc gatcagtctg caaacgggtgc ggcgatcctg catgcttacg ttcagcatac	4800
ctttggcggc ggcggtagct tcctgctggc ggcggttaatc ttcacgcct gcctggtcac	4860
ggcggttggc ctgacctgtg cttgtgcaga attcttcgcc cagtacgtac cgctctctta	4920
tcgtacgtg gtgtttatcc tcggcggctt ctcatgggtg gtgtctaacc tcggcttgag	4980
ccagctgatt cagatctctg taccggtgct gaccgccatt tatccgccgt gtatcgact	5040
ggttgattta agttttacac gctcatgggt gcataattcg tcccgcgtga ttgctccgcc	5100
gatgtttatc agcctgcttt ttggtattct cgacgggatc aaggcatctg cattcagcga	5160
tatcttaccg tcctgggcgc agcgtttacc gctggccgaa caaggctctg cgtgggtta	5220
gccaacagtg gtgatgggtg ttctggccat tatctgggat cgtgcggcag gtcgtcaggt	5280
gacctccagc gtcactaat acgcatggca tggatgaccg atggtagtgt gggctctccc	5340
catgcgagag tagggaactg ccaggcatca aataaaacga aaggctcagt cgaaagactg	5400
ggcctttcgt tttatctgtt gtttgtcggg gaacgctctc ctgagtagga caaatccgcc	5460
gggagcggat ttgaacgttg cgaagcaacg gcccgagggg tggcgggcag gacgcccgcc	5520
ataaactgcc aggcatacaa ttaagcagaa ggccatcctg acggatggcc tttttgcgtg	5580
gccagtgcc agcttgcatg cgtgccagct gcattaatga agaaatcatg ctggaagaat	5640
aacagctcac tcaaaggcgg tagtacgggt tttgctgccc gcaaacgggc tgttctgggtg	5700
ttgctagttt gttatcagaa tcgcagatcc ggcttcagcc ggtttgccgg ctgaaagcgc	5760
tatttcttcc agaattgcc aagatgtttt cccacgggag gcgtcactgg ctcccgtgtt	5820
gtcggcagct ttgattcgat aagcagcatc gcctgtttca ggctgtctat gtgtgactgt	5880
tgagctgtaa caagttgtct caggtgttca atttcatgtt ctagttgctt tgttttactg	5940
gtttcacctg ttctattagg tgttacatgc tgttcatctg ttacattgtc gatctgttca	6000

tggtgaacag ctttgaatgc accaaaaact cgtaaaagct ctgatgtatc tatctttttt	6060
acaccgtttt catctgtgca tatggacagt tttccctttg atatgtaacg gtgaacagtt	6120
gttctacttt tgtttggttag tcttgatgct tctactgatag atacaagagc cataagaacc	6180
tcagatcctt ccgtatttag ccagtatgtt ctctagtgtg gttcgttggt tttgcgtgag	6240
ccatgagaac gaaccattga gatcatactt actttgcatg tctactcaaaa attttgcctc	6300
aaaactggtg agctgaattt ttgcagttaa agcatcgtgt agtgtttttc ttagtccgtt	6360
atgtaggtag gaatctgatg taatggttgt tggatatttg tcaccattca tttttatctg	6420
gttgttctca agttcggtta cgagatccat ttgtctatct agttcaactt ggaaaatcaa	6480
cgtatcagtc gggcggcctc gcttatcaac caccaatttc atattgctgt aagtgtttaa	6540
atctttactt attggtttca aaaccattg gttaagcctt ttaaactcat ggtagttatt	6600
ttcaagcatt aacatgaact taaattcatc aaggctaadc tctatatattg ccttgtaggt	6660
tttcttttgt gttagttctt ttaataacca ctcataaatc ctcatagagt atttgttttc	6720
aaaagactta acatgttcca gattatattt tatgaatttt ttttaactgga aaagataagg	6780
caatatctct tcaactaaaa ctaattctaa tttttcgctt gagaacttgg catagtttgt	6840
ccactggaaa atctcaaagc ctttaaccaa aggattcctg atttccacag ttctcgtcat	6900
cagctctctg gttgctttag ctaatacacc ataagcattt tccctactga tgttcatcat	6960
ctgagcgtat tggttataag tgaacgatac cgtccgttct ttcctttagg ggttttcaat	7020
cgtgggggtg agtagtgcca cacagcataa aattagcttg gtttcatgct ccgttaagtc	7080
atagcgacta atcgctagtt catttgcttt gaaaacaact aattcagaca tacatctcaa	7140
ttggtctagg tgattttaat cactatacca attgagatgg gctagtcaat gataattact	7200
agtccttttc ctttgagttg tgggtatctg taaattctgc tagaccttg ctggaaaact	7260
tgtaaattct gctagaccct ctgtaaattc cgctagacct ttgtgtgttt tttttgttta	7320
tattcaagtg gttataattt atagaataaa gaaagaataa aaaaagataa aaagaataga	7380
tcccagccct gtgtataact cactacttta gtcagttccg cagtattaca aaaggatgtc	7440
gcaaacgctg tttgctcctc tacaaaacag accttaaaac cctaaaggct taagtagcac	7500
cctcgcaagc tcgggcaaatt cgctgaatat tccttttgtc tccgaccatc aggcacctga	7560
gtcgctgtct ttttcgtgac attcagttcg ctgcgctcac ggctctggca gtgaatgggg	7620

gtaaattggca ctacaggcgc cttttatgga ttcattgcaag gaaactaccc ataatacaag	7680
aaaagcccgt cacgggcttc tcagggcgtt ttatggcggg tctgctatgt ggtgctatct	7740
gactttttgc tgttcagcag ttcctgccct ctgattttcc agtctgacca cttcggatta	7800
tcccgtgaca ggtcattcag actggctaata gcacccagta aggagcggg atcatcaaca	7860
ggcttacccg tcttactgtc ttttctacgg ggtctgacgc tcagtgggaa gaaaactcac	7920
gttaagggat tttggctatg agattatcaa aaaggatctt cacctagatc cttttaaat	7980
aaaaatgaag ttttaaatca atctaaagta tatatgagta aacttggtct gacagttacc	8040
aatgcttaat cagtggagca cctatctcag cgatctgtct atttcgttca tccatagttg	8100
cctgactccc cgtcgtgtag ataactacga tacgggaggg cttaccatct ggccccagt	8160
ctgcaatgat accgcgagaa ccacgctcac cggctccaga tttatcagca ataaaccagc	8220
cagccggaag ggccgagcgc agaagtggc ctgcaacttt atccgcctcc atccagtcta	8280
ttaattgttg ccgggaagct agagtaagta gttcgccagt taatagtttg cgcaacgttg	8340
ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt tggatatggc tcattcagct	8400
ccggttccca acgatcaagg cgagttacat gatccccat gttgtgcaaa aaagcggtta	8460
gctccttcgg tcctccgatc gttgtcagaa gtaagtggc cgcagtgtta tcactcatgg	8520
ttatggcagc actgcataat tctcttactg tcatgccatc cgtaagatgc ttttctgtga	8580
ctggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg agttgctctt	8640
gcccggcgtc aatacgggat aataccgcgc cacatagcag aactttaaaa gtgctcatca	8700
ttggaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg agatccagtt	8760
cgatgtaacc cactcgtgca cccaactgat cttcagcatc ttttactttc accagcgttt	8820
ctgggtgagc aaaaacagga aggcaaaatg ccgcaaaaaa gggaataagg gcgacacgga	8880
aatgttgaat actcactc ttcctttttc aatattattg aagcatttat cagggttatt	8940
gtctcatgag cggatacata tttgaatgta tttagaaaaa taaacaaata ggggttccgc	9000
gcacatttcc ccgaaaagt ccacctgacg tctaagaaac cattattatc atgacattaa	9060
cctataaaaa taggcgtatc acgaggccct ttcgtctcgc gcgtttcggg gatgacgggtg	9120
aaaacctctg acacatgcag ctcccggaga cggtcacagc ttgtctgtaa gcggatgccg	9180
ggagcagaca agcccgtcag ggcgcgtcag cgggtgttgg cgggtgtcgg ggctggctta	9240

actatgcggc atcagagcag attgtactga gagtgcacca tatgcggtgt gaaataaccgc	9300
acagatgcgt aaggagaaaa taccgcatca ggcgccattc gccattcagg ctgcgcaact	9360
gttgggaagg gcgatcgggtg cgggcctctt cgctattacg ccagctggcg aaagggggat	9420
gtgctgcaag gcgattaagt tgggtaacgc cagggttttc ccagtcacga cgttgtaaaa	9480
cgacggccag tgaattcgtt aagaccact ttcacattta agttgttttt ctaatccgca	9540
tatgatcaat tcaaggccga ataagaaggc tggctctgca ccttggtgat caaataattc	9600
gatagcttgt cgtaataatg gcggcatact atcagtagta ggtgtttccc tttcttcttt	9660
agcgacttga tgctcttgat cttccaatac gcaacctaaa gtaaaatgcc ccacagcgct	9720
gagtgcatat aatgcattct ctagtgaaaa accttgttgg cataaaaagg ctaattgatt	9780
ttcgagagtt tcatactggt tttctgtagg ccgtgtacct aaatgtactt ttgctccatc	9840
gcgatgactt agtaaagcac atctaaaact tttagcgta ttacgtaaaa aatcttgcca	9900
gctttcccct tctaaagggc aaaagtgagt atggtgccta tctaacatct caatggctaa	9960
ggcgtcgagc aaagcccgt tattttttac atgccaatac aatgtaggct gctctacacc	10020
tagcttctgg gcgagtttac gggttgttaa accttcgatt ccgacctcat taagcagctc	10080
taatgcgctg ttaatcactt tacttttatac taatctagac atcattaatt cctaattttt	10140
gttgacactc tatcattgat agagttatth taccactccc tatcagtgat agagaaaagt	10200
ga	10202

<210> 5

<211> 10259

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 5

gcgaaaaatc aataaggagg caacaagatg tgcgaaaaac atcttaatca tgctaaggag	60
gttttctaata ggtagagatc caggctttgc cggaacttc catttttggg caaatcgag	120
accaccagca tgaacagggt gtcttctgcc acgatcacga aaccggcctc cgtgcgataa	180
tcggtattca taacacggtt cttggccccg ccttaggtgg aactcgcatg tggcactatg	240
ctaccgaggc agaagcgctg aatgacgttc tgcgtctgtc tcgcggtatg acctacaagg	300

ctgctatctc	gggcctgaac	ctgggtggcg	gtaaagcagt	aatcattggg	gatgccaaaa	360
caatcaaaac	cgaagcgctg	ctgcggaaat	tcggcagatt	cgtgcagaac	ctgaatggta	420
aatacatcac	tgctgaagac	gttaacatga	ctacaaagga	tatggagtac	attaggatgg	480
aaaccaaaca	cgtcgctggc	ttacctgaaa	gtatgggtgg	aagcggtgac	ccgtcaccgg	540
taactgcata	tggtacgtac	atgggcatga	aagcggcggc	caaaaaggcg	tttggctctg	600
attccctggc	tggcaaacgt	atcgctgttc	aaggtgtggg	tcatgttggc	acttatctgc	660
ttgagcattt	gaccaaagaa	ggtgctcaga	ttgtgctgac	tgactactat	aaggaacgtg	720
ccgaggaagc	aggcgcgcgt	tttggcgcac	aggttgttgg	cctggacgat	atctacgatc	780
aagaggtcga	catttactct	ccatgtgttc	tcgggtgtac	catcaacgat	gacactatcg	840
atgcctgcg	ttgcgctggt	gtagccggtt	gcgcaaaca	ccagctgaaa	gaagaaaacg	900
tccacggtcc	ggcgctgggt	gagcgcggga	tagtatacgc	cccagacttc	ctgatcaatg	960
cagggggcct	gattaacgtg	tatagcgaag	ttacagggtc	tacccgtcag	ggggctttaa	1020
ctcagaccga	aaaaatctat	gactacacac	tccaagttct	ggaaaaagcc	gcggctgaag	1080
gtctgcaccc	gcagcaggct	gcgatccgtc	aggcgggaaca	acgcatcgct	gcaattggta	1140
aggtgaaaag	cacctactaa	gtcaataaag	gcatataaaa	ggagggttaat	aacatgaaag	1200
ttaaagtaaa	acatcttaat	catgcacagg	agactttcta	atgtcgacga	caaccgttgg	1260
tgactacttg	ctgtatcgct	taaacgaaat	cggcattgag	cacctcttcg	gagtgccagg	1320
tgattacaat	ctgcaatttc	tggatcatgt	aatcgaccac	cctcagctga	cttgggtcgg	1380
ctgcactaac	gaacttaacg	ctgcctacgc	agctgatggg	tatgcgcgtt	gtcgtccggc	1440
tgcggcactg	ctgaccacct	tcgggggttg	cgaactgagc	gctattaatg	gcatcgcagg	1500
ttcctacgcg	gagtatctgc	cggtaataca	tatcgttggg	gcaccgagtc	tatcagccca	1560
gcagcagggc	gacctgattc	accactctct	tggcgaagg	gattttttcca	gcttcctgag	1620
gatgtcccaa	ccggtgtctg	ttgcgcaggc	tgctctgact	cctgataacg	catgcaagga	1680
aatcgaccgc	gtactggcgg	aagtcctcat	tcagcgtcgt	cccggctacc	tgctgctgtc	1740
taccgacgtg	gctgctgcgc	cggcggctct	gccacaaagc	actctttctt	tgccgaccgc	1800
cccggatcat	cgcgagttc	tggctgcttt	cagcgacgct	gctgagcaga	tgctggctca	1860
ggccaaaagc	gtctctctac	tggcggactt	tctggctgat	cgtttcgggtg	ttactcgagc	1920

actggccgcg tggcttcagc aggttccgct accgcacgcc actctgttaa tgggtaaagg	1980
cgttctgagt gaacagcaac cagggttcgt gggtagctac gctgggtgcgg catctatcga	2040
ttcgacgcgt ggcgcaatcg aagaagctgg ggtaattatc ggagtgggag ttagattttc	2100
cgacactatc acagcaggct tctcgcagca gatcgacgcc cgccgtttta tagacattca	2160
acccttcttc tctcgtattg gcgatcgcca gtttgatcac ttgccgatgc aggctgccgt	2220
cgcagccctg catcaactgt gtcttcgtta tcagcagcag tggcttatca ccgctcctag	2280
cccgctgca ctgccgccgg ctgctggtag cgagctgtcc cagaacgcat tctggcaggc	2340
gatgcagaac ttcattccgc ctggggacct gttgggtggcc gaccaaggta ctgcggcggt	2400
cggcgcagcg gcgctgcgt taccgcagaa ttgccagctg cttgtgcagc cgctgtgggg	2460
ctcaatcggg tacagtctgc cgccacctt tgggtgctcag acggcagata cagagcgtcg	2520
tgtaatccta atcattggcg atggttcagc gcaattaact attcaggaac tttccagtat	2580
gatgcgtgac ggcttgaaac ctatcatctt tctcctgaac aacaacgggt acaccgttga	2640
acgggcgatt cacggcgcgg agcaacgtta taacgatatc gctgcttgga attggacca	2700
actgccccag gcgctgagtg ttcatcgccc agcgcagagc tggcgagtcg ttgaaacggg	2760
gcagctgacc gacgtaatga aagtcatcgc tgcttctccg cgtctgagct tggtagaagt	2820
tgttctgcct gcaatggatg tcccaccgct gctgcaagca gtgagtgccg ctctgaacca	2880
gcgcaactcc tcttaagacc agaggtaagg aggtacaac catgagtgatg ttgaagaaac	2940
atcttaatca tgcaggggag ggtttctaata gcgcagcatg cagtttgatg agtacgggtgc	3000
accctgaaa gcgttctcat atgaagacc gaccccgcaa gggaagggaag tagtcgttag	3060
gatcgaagcc tgtggtgtgt gccactctga tattcatctt cacgagggt acttcgacat	3120
gggcgggtggc aataaagctg atgttactcg tgctcgcgaa ctccctttta cattgggtca	3180
tgaaatcggt ggcaagtgg tagcaactgg accagggtgtc accggcgcta aaccgggcga	3240
caaacgtatt gtgtaccgt ggatcgggtg cggcgactgc ccgaaatgca acagtgggtga	3300
ggatcagtcc tgtgcgcgtc cacgtaacct ggggtgttcac gttgacgggt gctattcgac	3360
gcacgtaaag ataccggacg aaaaattcct gttcgcctac gatgggtattc ctactgagtt	3420
agcgggaacc tatgcttgca gcggcatcac cgcttatggg gactgatga aagcaaagga	3480
agcggctgaa agatctggct acatcgggtc gattggcgct ggtggcggtg gcatggctgg	3540

tctgatgctg gccaaagcag cgatcggggc taaaactgta gtctttgata tcgacgacgc	3600
aaaactggaa gctgcgaccc gtgccggggc ggattacgtg ttcaactccg gtgcaaaaga	3660
aacacgcaag gaagttatga aactaacgaa tgggtggcctg tctgggtgctg ttgatttcgt	3720
tggcagcgat aaaagcgctc tgtttggaat caacgccttg ggtcagaacg gcgtgctggc	3780
cataattgga ctgttcgggtg gcgctatgac tgttccggta cccctgttcc cgctgaaagg	3840
gatcaccgta cgtggctcat acgtaggttc cctgcaagag atgagtata tgatggagtt	3900
agttcgcgct gggaaagtcc ctccgatgcc ggtaaaaact cggccactgg acgtgcctg	3960
ggaaaccctt gaggatctac gccatggtaa aatcgtgggc cgtgttgctc tgacccata	4020
agaaggagat atacatatga cccatcaatt aagatcgcg gatatacatg ctctgggctt	4080
tatgacattt gcgttgctcg tcggcgcagg taacattatt ttccctcaa tggcggctt	4140
gcaggcaggc gaacacgtct ggactgcggc attcggcttc ctctactg ccgttggcct	4200
accggtatta acggtagtgg cgctggcaaa agttggcggc ggtgttgaca gtctcagcac	4260
gccaattggg aaagtcgctg gcgtactgct ggcaacagtt tgttacctgg cgggtggggc	4320
gctttttgct acgccgcgta cagctaccgt ttcttttgaa gtgggcattg cgccgctgac	4380
gggtgattcc gcgctgccgc tgtttattta cagcctggc tatttcgcta tcgttattct	4440
ggtttcgctc tatccgggca agctgctgga taccgtgggc aacttccttg cgccgctgaa	4500
aattatcgcg ctggatcatc tgtctgttgc cgcaattatc tggccggcgg gttctatcag	4560
tacggcgact gaggcttatc aaaacgctgc gttttctaac ggcttcgtca acggctatct	4620
gaccatggat acgctgggcg caatggtgtt tggatatggt attgttaacg cggcgcgttc	4680
tcgtggcggt accgaagcgc gtctgctgac ccgttatacc gtctgggctg gcctgatggc	4740
gggtgttggt ctgactctgc tgtacctggc gctgttccgt ctgggttcag acagcgcgtc	4800
gctggtcgat cagtctgcaa acggtgcggc gatcctgcat gcttacgttc agcatacctt	4860
tggcggcggc ggtagcttcc tgctggcggc gttaatcttc atcgctgcc tggtcacggc	4920
ggttggcctg acctgtgctt gtgcagaatt cttcgcccag tacgtaccgc tctcttatcg	4980
tacgtgggtg tttatcctcg gcggcttctc gatgggtgggt tctaacctcg gcttgagcca	5040
gctgattcag atctctgtac cgggtgctgac cgccatttat ccgccgtgta tcgactggc	5100
tgtattaagt ttacacgct catggtggca taattcgtcc gcgtgattg ctccgccgat	5160

gtttatcagc ctgctttttg gtattctcga cgggatcaag gcatctgcat tcagcgatat	5220
cttaccgtcc tggg'gcagc gtttaccgct ggccgaacaa ggtctggcgt ggttaatgcc	5280
aacagtgggtg atgggtggttc tggccattat ctgggatcgt gcggcaggtc gtcaggtgac	5340
ctccagcgct cactaatacg catggcatgg atgaccgatg gtagtgtggg ctctccccat	5400
gcgagagtag ggaactgcca ggcatcaaataaaaacgaaag gctcagtcga aagactgggc	5460
ctttcgtttt atctgttggt tgtcggtgaa cgctctcctg agtaggacaa atccgccggg	5520
agcggatttg aacgttgcca agcaacggcc cggagggtgg cgggcaggac gcccgccata	5580
aactgccagg catcaaatta agcagaaggc catcctgacg gatggccttt ttgcgtggcc	5640
agtccaagc ttgcatgcgt gccagctgca ttaatgaaga aatcatgctg gaagaataac	5700
agctcactca aaggcggtag tacgggtttt gctgcccgc aacgggctgt tctggtgttg	5760
ctagtttggt atcagaatcg cagatccggc ttccagccggt ttgccggctg aaagcgctat	5820
ttcttccaga attgccatga ttttttcccc acgggaggcg tcaactggctc ccgtgttgct	5880
ggcagctttg attcgataag cagcatcgcc tgtttcaggc tgtctatgtg tgactgttga	5940
gctgtaacaa gttgtctcag gtgttcaatt tcatgttcta gttgctttgt tttactgggt	6000
tcacctgttc tattaggtgt tacatgctgt tcatctgtta cattgtcgat ctgttcatgg	6060
tgaacagctt tgaatgcacc aaaaactcgt aaaagctctg atgtatctat cttttttaca	6120
ccgttttcat ctgtgcatat ggacagtttt ccctttgata tgtaacgggtg aacagttggt	6180
ctacttttgt ttgttagtct tgatgcttca ctgatagata caagagccat aagaacctca	6240
gatccttccg tatttagcca gtatgttctc tagtgtgggt cgttgttttt gcgtgagcca	6300
tgagaacgaa ccattgagat catacttact ttgcatgtca ctcaaaaatt ttgcctcaaa	6360
actggtgagc tgaatttttg cagttaaagc atcgtgtagt gtttttctta gtccgttatg	6420
taggtaggaa tctgatgtaa tggttggttg tattttgtca ccattcattt ttatctgggt	6480
gttctcaagt tcggttacga gatccatttg tctatctagt tcaacttgga aaatcaacgt	6540
atcagtcggg cggcctcgct tatcaaccac caatttcata ttgctgtaag tgtttaaadc	6600
tttacttatt ggtttcaaaa cccattgggt aagcctttta aactcatggt agttattttc	6660
aagcattaac atgaacttaa attcatcaag gctaactctt atatttgcct tgtgagtttt	6720
cttttggtt agttctttta ataaccactc ataaatcctc atagagtatt tgttttcaaa	6780

agacttaaca	tgttccagat	tatatTTTTat	gaatTTTTTT	aactggaaaa	gataaggcaa	6840
tatctcttca	ctaaaaacta	attctaattt	ttcgcttgag	aacttggcat	agtttgtcca	6900
ctggaaaatc	tcaaagcctt	taaccaaagg	attcctgatt	tccacagttc	tcgtcatcag	6960
ctctctggtt	gcttttagcta	atacaccata	agcattttcc	ctactgatgt	tcatcatctg	7020
agcgtattgg	ttataagtga	acgataccgt	ccgttctttc	cttgtagggg	tttcaatcgt	7080
ggggttgagt	agtgccacac	agcataaaat	tagcttgggt	tcatgctccg	ttaagtcata	7140
gcgactaatc	gctagttcat	ttgctttgaa	aacaactaat	tcagacatac	atctcaattg	7200
gtctaggtga	ttttaatcac	tataccaatt	gagatgggct	agtcaatgat	aattactagt	7260
ccttttcctt	tgagttgtgg	gtatctgtaa	attctgctag	acctttgctg	gaaaacttgt	7320
aaattctgct	agaccctctg	taaattccgc	tagacctttg	tgtgtTTTTt	ttgtttatat	7380
tcaagtgggt	ataatttata	gaataaagaa	agaataaaaa	aagataaaaa	gaatagatcc	7440
cagccctgtg	tataactcac	tacttttagtc	agttccgcag	tattacaaaa	ggatgtcgca	7500
aacgctgttt	gctcctctac	aaaacagacc	ttaaaaccct	aaaggcttaa	gtagcaccct	7560
cgcaagctcg	ggcaaatcgc	tgaatattcc	ttttgtctcc	gaccatcagg	cacctgagtc	7620
gctgtctttt	tcgtgacatt	cagttcgctg	cgctcacggc	tctggcagtg	aatgggggta	7680
aatggcacta	caggcgcctt	ttatggattc	atgcaaggaa	actaccata	atacaagaaa	7740
agcccgtcac	gggcttctca	gggcgtttta	tggcgggtct	gctatgtggg	gctatctgac	7800
tttttgctgt	tcagcagttc	ctgccctctg	attttccagt	ctgaccactt	cggattatcc	7860
cgtgacaggt	cattcagact	ggctaattgca	cccagtaagg	cagcggtatc	atcaacaggc	7920
ttaccctgtc	tactgtcttt	tctacggggg	ctgacgtca	gtggaacgaa	aactcacgtt	7980
aagggatttt	ggtcattgaga	ttatcaaaaa	ggatcttcac	ctagatcctt	ttaaattaaa	8040
aatgaagttt	taaatcaatc	taaagtatat	atgagtaaac	ttggtctgac	agttaccaat	8100
gcttaatcag	tgaggcacct	atctcagcga	tctgtctatt	tcgttcatcc	atagttgcct	8160
gactccccgt	cgtgtagata	actacgatac	gggaggggct	accatctggc	cccagtgtgt	8220
caatgatacc	gcgagaacca	cgctcaccgg	ctccagattt	atcagcaata	aaccagccag	8280
ccggaagggc	cgagcgcaga	agtggctctg	caactttatc	cgctccatc	cagtctatta	8340
attgttgccg	ggaagctaga	gtaagtagtt	cgccagttaa	tagtttgcgc	aacgttggtg	8400

ccattgctac aggcacgtg gtgtcacgct cgtcgtttgg tatggcttca ttcagctccg	8460
gttcccaacg atcaaggcga gttacatgat ccccatgtt gtgcaaaaaa gcggtttagct	8520
ccttcggtcc tccgatcgtt gtcagaagta agttggccgc agtgttatca ctcatggtta	8580
tggcagcact gcataattct cttactgtca tgccatccgt aagatgcttt tctgtgactg	8640
gtgagtactc aaccaagtca ttctgagaat agtgtatgcg gcgaccgagt tgctcttgcc	8700
cggcgtcaat acgggataat accgcgccac atagcagaac tttaaaagtg ctcatcattg	8760
gaaaacgttc ttcggggcga aaactctcaa ggatcttacc gctgttgaga tccagttcga	8820
tgtaaccac tcgtgcaccc aactgatctt cagcatcttt tactttcacc agcgtttctg	8880
ggtgagcaaa aacaggaagg caaatgccg caaaaaaggg aataaggcg acacggaaat	8940
gttgaatact catactcttc ctttttcaat attattgaag catttatcag ggttattgtc	9000
tcatgagcgg atacatattt gaatgtattt agaaaaataa acaaataggg gttccgcgca	9060
catttccccg aaaagtgcc cctgacgtct aagaaacat tattatcatg acattaacct	9120
ataaaaaatag gcgtatcacg aggccctttc gtctcgcgcg tttcgggat gacggtgaaa	9180
acctctgaca catgcagctc ccggagacgg tcacagcttg tctgtaagcg gatgccggga	9240
gcagacaagc ccgtcagggc gcgtcagcgg gtgttgccgg gtgtcggggc tggcttaact	9300
atgcggcatc agagcagatt gtactgagag tgcaccatat gcggtgtgaa ataccgcaca	9360
gatgcgtaag gagaaaatac cgcacagggc gccattcgcc attcaggctg cgcaactgtt	9420
gggaaggcg atcgggtcgg gcctcttcgc tattacgcca gctggcgaaa gggggatgtg	9480
ctgcaaggcg attaagttgg gtaacgccag ggttttcca gtcacgacgt tgtaaaacga	9540
cggccagtga attcgtaag acccactttc acatttaagt tgtttttcta atccgcatat	9600
gatcaattca aggccgaata agaaggctgg ctctgcacct tggatgatcaa ataattcgat	9660
agcttgtcgt aataatggcg gcatactatc agtagtaggt gtttcccttt cttcttagc	9720
gacttgatgc tcttgatctt ccaatacgca acctaaagta aaatgcccc aagcgtgag	9780
tgcatataat gcattctcta gtgaaaaacc ttgttgcat aaaaaggcta attgattttc	9840
gagagtttca tactgttttt ctgtaggccg tgtacctaaa tgtacttttg ctccatcgcg	9900
atgacttagt aaagcacatc taaaactttt agcgttatta cgtaaaaaat cttgccagct	9960
ttcccccttct aaagggcaaa agtgagtatg gtgcctatct aacatctcaa tggctaaggc	10020

gtcgagcaaa gcccgcttat tttttacatg ccaataacaat gtaggctgct ctacacctag	10080
cttctgggcg agtttacggg ttgttaaacc ttcgattccg acctcattaa gcagctctaa	10140
tgcgctgtta atcactttac ttttatctaa tctagacatc attaattcct aatttttgtt	10200
gacactctat cattgataga gttattttac cactccctat cagtgataga gaaaagtga	10259

<210> 6

<211> 10289

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 6

gcgaaaaatc aataaggagg caacaagatg tgcgaaaaac atcttaatca tgctaaggag	60
gttttctaata ggtagagatc caggctttgc cggaaacttc catttttggg caaatcgcag	120
accaccagca tgaacagggtg gtcttctgcc acgatcacga aaccggcctc cgtgcgataa	180
tcggtattca taacacgggtt cttggccccg ccttaggtgg aactcgcatg tggcactatg	240
ctaccgaggc agaagcgctg aatgacgttc tgcgtctgtc tcgcggtatg acctacaagg	300
ctgctatctc gggcctgaac ctgggtggcg gtaaagcagt aatcattggg gatgccaaaa	360
caatcaaaac cgaagcgctg ctgcggaaat tcggcagatt cgtgcagaac ctgaatggta	420
aatacatcac tgctgaagac gttaacatga ctacaaagga tatggagtac attaggatgg	480
aaaccaaaca cgtcgctggc ttacctgaaa gtatgggtgg aagcgggtgac ccgtcaccgg	540
taactgcata tggtagctac atgggcatga aagcggcggc caaaaaggcg tttggctctg	600
attccctggc tggcaaactg atcgctgttc aagggtgtgg tcatgttggc acttatctgc	660
ttgagcattt gaccaaagaa ggtgctcaga ttgtgctgac tgactactat aaggaaactg	720
ccgaggaagc aggcgcgcgt tttggcgcac aggttggttg cctggacgat atctacgatc	780
aagaggtcga catttactct ccatgtgctc tcggtgctac catcaacgat gacactatcg	840
atgcctgcg ttgcgctgtt gtagccggtt gcgcaaaca ccagctgaaa gaagaaaacg	900
tccacggtcc ggcgctgggt gagcgcggga tagtatacgc ccagacttc ctgatcaatg	960
cagggggcct gattaacgtg tatagcgaag ttacagggtc taccggtcag ggggctttta	1020

ctcagaccga	aaaaatctat	gactacacac	tccaagttct	ggaaaaagcc	gcggctgaag	1080
gtctgcaccc	gcagcaggct	gcgatccgtc	aggcgggaaca	acgcatcgct	gcaattggta	1140
aggtgaaaag	cacctactaa	gtcaataaag	gcatataaaa	ggagggttaat	aacatgaaag	1200
ttaaagtaaa	acatcttaat	catgcacagg	agacttttcta	atgtcgacga	caaccgttgg	1260
tgactacttg	ctgtatcgct	taaacgaaat	cggcattgag	cacctcttcg	gagtgccagg	1320
tgattacaat	ctgcaatttc	tggatcatgt	aatcgaccac	cctcagctga	cttgggtcgg	1380
ctgcactaac	gaacttaacg	ctgcctacgc	agctgatggg	tatgcgcggt	gtcgtccggc	1440
tgcggcactg	ctgaccacct	tcgggggttg	cgaactgagc	gctattaatg	gcatcgcagg	1500
ttcctacgcg	gagtatctgc	cggtaataca	tatcgttggg	gcaccgagtc	tatcagccca	1560
gcagcagggc	gacctgattc	accactctct	tggcgaagg	gatttttcca	gcttcctgag	1620
gatgtcccaa	ccgggtgtctg	ttgcgcaggc	tgctctgact	cctgataacg	catgcaagga	1680
aatcgaccgc	gtactggcgg	aagtcctcat	tcagcgtcgt	cccggctacc	tgctgctgtc	1740
taccgacgtg	gctgctgcgc	cggcggctct	gccacaaagc	actctttctt	tgccgaccgc	1800
cccggatcat	cgcgagttc	tggctgcttt	cagcgacgct	gctgagcaga	tgctggctca	1860
ggccaaaagc	gtctctctac	tggcggactt	tctggctgat	cgtttcggtg	ttactcgagc	1920
actggccgcg	tggcttcagc	aggttccgct	accgcacgcc	actctgttaa	tgggtaaagg	1980
cgttctgagt	gaacagcaac	cagggttcgt	gggtacctac	gctgggtgcgg	catctatcga	2040
ttcgacgcgt	ggcgcaatcg	aagaagctgg	ggtaattatc	ggagtgggag	ttagattttc	2100
cgacactatc	acagcaggct	tctcgcagca	gatcgacgcc	cgccgtttta	tagacattca	2160
acccttcttc	tctcgtattg	gcgatcgcca	gtttgatcac	ttgccgatgc	aggctgccgt	2220
cgcagccctg	catcaactgt	gtcttcgtta	tcagcagcag	tggtctatca	ccgctcctag	2280
cccgctgca	ctgccgccgg	ctgctggtag	cgagctgtcc	cagaacgcat	tctggcaggc	2340
gatgcagaac	ttcatccgcc	ctggggacct	gttggtggcc	gaccaaggta	ctgcggcggt	2400
cggcgcagcg	gcgctgcgct	taccgcagaa	ttgccagctg	cttggtgcagc	cgctgtgggg	2460
ctcaatcggg	tacagtctgc	cggccacctt	tggtgctcag	acggcagata	cagagcgtcg	2520
tgtaatccta	atcattggcg	atggttcagc	gcaattaact	attcaggaac	tttccagtat	2580
gatgcgtgac	ggcttgaaac	ctatcatctt	tctcctgaac	aacaacgggt	acaccgttga	2640

acgggcgatt	cacggcgcgg	agcaacgtta	taacgatatc	gctgcttgga	attggacca	2700
actgccccag	gcgctgagtg	ttcattgccc	agcgcagagc	tggcgagtcg	ttgaaacggg	2760
gcagctgacc	gacgtaatga	aagtcacgcg	tgcttctccg	cgtctgagct	tggtagaagt	2820
tgttctgcct	gcaatggatg	tcccaccgct	gctgcaagca	gtgagtgccg	ctctgaacca	2880
gcgcaactcc	tcttaagacc	agaggtaagg	aggtaacaac	catgcgagtg	ttgaagaaac	2940
atcttaatca	tgcaggggag	ggtttcta	gacagcggag	cagcaaaatg	gggtatccga	3000
ctcacgccgt	ttcgaatttc	aggaatttgg	tggccctatc	gccccacaga	cctatcagct	3060
ccccgcaccg	gctagcgatg	aagttttgtt	aaaggtgaac	tactgcggtg	tctgtcacag	3120
tgatgttcat	cttcacgacg	gctacttcga	gctgggtggc	gataaacgtc	tgaacttcgc	3180
tatgccgctg	ccgctgacgc	tgggtcacga	agtaattggc	accgttgtgg	ctgtcggcga	3240
ccaggttact	ggtgtaaaac	cgggggacca	gcgactgac	tatccgtgga	taggttgcgg	3300
aaaatgcggc	gcgtgtcaaa	aaggagaaga	aaacctgtgc	gttactcctg	cacatctggg	3360
cgtgaacaag	ccgggcgggt	acgctgatca	catcgttgta	ccccattctc	gctaccttct	3420
ggacatttcg	ggtctgaacc	cgggtgatgc	cgctaccctc	gcgtgctccg	gcctgaccac	3480
tttcagcgcg	atcaacaaag	tgttgccgct	tgcagatgac	cagtggattg	ttgttatcgg	3540
ttgtgggtgg	ctcggccaga	tggcgctgcg	tatcctgcaa	gctatgggaa	ttggcaatgt	3600
tatcgggtatt	gacctgtctg	aagagaaacg	gaaactggct	catgaaagcg	gtgcacgtca	3660
ctccttcgat	caaacactc	cgaagctgaa	ccgcgtggtc	gccgaaacct	gcccgggtac	3720
ggtacaggcc	gcgttagact	ttgtgggcaa	tgagcaaact	gctcagctgg	cactgtctct	3780
gcttggaana	ggtggcaaat	atgttcctgt	cgggctgcac	ggcggcgagc	tgcgttacc	3840
attgccgatc	atcacgaaca	aagctgtaag	tatcatcggg	tcttacgttg	gtaccctgaa	3900
agaactggaa	gacttagttg	ctttcgccaa	ggaaaagaat	ctgccgcaa	ttcatattga	3960
acaccgccc	ctggaatcgg	cggctcaggc	cgtagaggac	ctggaaaaag	gacaggttgc	4020
tgggcgtgtt	atcctggatg	caggtaacta	agaaggagat	atacatatga	cccatcaatt	4080
aagatcgcgc	gatatcatcg	ctctgggctt	tatgacattt	gcgttggtcg	tcggcgcagg	4140
taacattatt	ttccctcaa	tggctggctt	gcaggcaggc	gaacacgtct	ggactgcggc	4200
attcggcttc	ctcattactg	ccgttggcct	accggtatta	acggtagtgg	cgctggcaaa	4260

agttggcggc ggtgttgaca gtctcagcac gccaatgggt aaagtcgctg gcgtactgct	4320
ggcaacagtt tgttacctgg cgggtggggcc gctttttgct acgccgcgta cagctaccgt	4380
ttcttttgaa gtgggcattg cgccgctgac ggggtgattcc gcgctgccgc tgtttattta	4440
cagcctggtc tatttcgcta tcgttattct ggtttcgctc tatccgggca agctgctgga	4500
taccgtgggc aacttccttg cgccgctgaa aattatcgcg ctggtcattcc tgtctgttgc	4560
cgcaattatc tggccggcgg gttctatcag tacggcgact gaggcttatc aaaacgctgc	4620
gttttctaac ggcttcgtca acggctatct gaccatggat acgctgggcg caatggtggt	4680
tggtatcggt attgttaacg cggcgcgttc tcgtggcgtt accgaagcgc gtctgctgac	4740
ccgttatacc gtctgggctg gcctgatggc ggggtgttgggt ctgactctgc tgtacctggc	4800
gctgttccgt ctgggttcag acagcgcgtc gctggtcgat cagtctgcaa acggtgcggc	4860
gacctgcat gcttacgttc agcatacctt tggcggcggc ggtagcttcc tgctggcggc	4920
gttaatcttc atgcctgcc tggtcacggc ggttggcctg acctgtgctt gtgcagaatt	4980
cttcgccag tacgtaccgc tctcttatcg tacgctgggtg tttatcctcg gcggcttctc	5040
gatggtggtg tctaacctcg gcttgagcca gctgattcag atctctgtac cggtgctgac	5100
cgccatttat ccgccgtgta tcgcactgggt tgtattaagt tttacacgct catgggtggca	5160
taattcgtcc cgcgtgattg ctccgccgat gtttatcagc ctgctttttg gtattctcga	5220
cgggatcaag gcatctgcat tcagcgatat cttaccgtcc tgggcgcagc gtttaccgct	5280
ggccgaacaa ggtctggcgt ggttaatgcc aacagtgggtg atggtggttc tggccattat	5340
ctgggatcgt gcggcaggtc gtcaggtgac ctccagcgct cactaatacg catggcatgg	5400
atgaccgatg gtagtgtggg ctctcccat gcgagagtag ggaactgcca ggcatcaa	5460
aaaacgaaag gctcagtcga aagactgggc ctttcgtttt atctgttggt tgtcggtgaa	5520
cgctctcctg agtaggacaa atccgccggg agcggatttg aacgttgcca agcaacggcc	5580
cggagggtgg cgggcaggac gcccgccata aactgccagg catcaaatta agcagaaggc	5640
catcctgacg gatggccttt ttgcgtggcc agtgccaagc ttgcatgcgt gccagctgca	5700
ttaatgaaga aatcatgctg gaagaataac agctcactca aaggcggtag tacgggtttt	5760
gctgcccga aacgggctgt tctggtgttg ctagtttggt atcagaatcg cagatccggc	5820
ttcagccggt ttgccggctg aaagcgctat ttcttccaga attgccatga tttttcccc	5880

acgggaggcg	tcactggctc	ccgtgttgtc	ggcagctttg	attcgataag	cagcatcgcc	5940
tgtttcaggc	tgtctatgtg	tgactgttga	gctgtaacaa	gttgtctcag	gtgttcaatt	6000
tcatgttcta	gttgctttgt	tttactggtt	tcacctgttc	tattaggtgt	tacatgctgt	6060
tcatctgtta	cattgtcgat	ctgttcatgg	tgaacagctt	tgaatgcacc	aaaaactcgt	6120
aaaagctctg	atgtatctat	cttttttaca	ccgttttcat	ctgtgcatat	ggacagtttt	6180
ccctttgata	tgtaacgggtg	aacagttgtt	ctacttttgt	ttgttagtct	tgatgcttca	6240
ctgatagata	caagagccat	aagaacctca	gacaccttcg	tatttagcca	gtatgttctc	6300
tagtgtgggt	cgttgttttt	gcgtgagcca	tgagaacgaa	ccattgagat	catacttact	6360
ttgcatgtca	ctcaaaaatt	ttgcctcaaa	actggtgagc	tgaatttttg	cagttaaagc	6420
atcgtgtagt	gtttttctta	gtccgttatg	taggtaggaa	tctgatgtaa	tggttggttg	6480
tattttgtca	ccattcattt	ttatctgggt	gttctcaagt	tcggttacga	gatccatttg	6540
tctatctagt	tcaacttgga	aaatcaacgt	atcagtcggg	cggcctcgct	tatcaaccac	6600
caatttcata	ttgctgtaag	tgtttaaatc	tttacttatt	ggtttcaaaa	cccattgggt	6660
aagcctttta	aactcatggg	agttattttc	aagcattaac	atgaacttaa	attcatcaag	6720
gctaactctc	atatttgcct	tgtgagtttt	cttttgtgtt	agttctttta	ataaccactc	6780
ataaatcctc	atagagtatt	tgttttcaaa	agacttaaca	tgttccagat	tatattttat	6840
gaattttttt	aactggaaaa	gataaggcaa	tatctcttca	ctaaaaacta	attctaattt	6900
ttcgcttgag	aacttggcat	agtttgtcca	ctggaaaatc	tcaaagcctt	taaccaaagg	6960
attcctgatt	tccacagttc	tcgtcatcag	ctctctgggt	gcttttagcta	atacaccata	7020
agcattttcc	ctactgatgt	tcatcatctg	agcgtattgg	ttataagtga	acgataccgt	7080
ccgttctttc	cttgtagggg	tttcaatcgt	ggggttgagt	agtgccacac	agcataaaat	7140
tagcttgggt	tcatgctccg	ttaagtcata	gcgactaatc	gctagttcat	ttgctttgaa	7200
aacaactaat	tcagacatac	atctcaattg	gtctaggtga	ttttaatcac	tataccaatt	7260
gagatgggct	agtcaatgat	aattactagt	ccttttcctt	tgagttgtgg	gtatctgtaa	7320
attctgctag	acctttgctg	gaaaacttgt	aaattctgct	agaccctctg	taaattccgc	7380
tagacctttg	tgtgtttttt	ttgtttatat	tcaagtgggt	ataatttata	gaataaagaa	7440
agaataaaaa	aagataaaaa	gaatagatcc	cagccctgtg	tataactcac	tacttttagtc	7500

agttccgcag tattacaaaa ggatgtcgca aacgctgttt gtcctcttac aaaacagacc	7560
ttaaaaccct aaaggcttaa gtagcaccct cgcaagctcg ggcaaatcgc tgaatattcc	7620
ttttgtctcc gaccatcagg cacctgagtc gctgtctttt tcgtgacatt cagttcgctg	7680
cgctcacggc tctggcagtg aatgggggta aatggcacta caggcgcctt ttatggattc	7740
atgcaaggaa actaccata atacaagaaa agcccgtcac gggcttctca gggcgtttta	7800
tggcgggtct gctatgtggg gctatctgac tttttgctgt tcagcagttc ctgccctctg	7860
attttccagt ctgaccactt cggattatcc cgtgacaggc cattcagact ggctaatagca	7920
cccagtaagg cagcgggtatc atcaacaggc ttaccctgtc tactgtcttt tctacggggg	7980
ctgacgctca gtggaacgaa aactcacgtt aagggaatttt ggcatgaga ttatcaaaaa	8040
ggatcttcac ctagatcctt ttaaattaaa aatgaagttt taaatcaatc taaagtatat	8100
atgagtaaac ttgggtctgac agttaccaat gcttaatcag tgaggcacct atctcagcga	8160
tctgtctatt tcgttcatcc atagttgcct gactccccgt cgtgtagata actacgatac	8220
gggaggggctt accatctggc cccagtgtcg caatgatacc gcgagaacca cgctcaccgg	8280
ctccagattt atcagcaata aaccagccag ccggaagggc cgagcgcaga agtggctctg	8340
caactttatc cgcctccatc cagtctatta attgttgccg ggaagctaga gtaagtagtt	8400
cgccagttaa tagtttgccg aacgttggtt ccattgctac aggcacgtg gtgtcacgct	8460
cgtcgtttgg tatggcttca ttcagctccg gttcccaacg atcaaggcga gttacatgat	8520
cccccatgtt gtgcaaaaaa gcggtagct ccttcgggtc tccgatcgtt gtcagaagta	8580
agttggccgc agtggttatca ctcatggtta tggcagcact gcataattct cttactgtca	8640
tgccatccgt aagatgcttt tctgtgactg gtgagtactc aaccaagtca ttctgagaat	8700
agtgatatgcg gcgaccgagt tgctcttgcc cggcgtcaat acgggataat accgcgccac	8760
atagcagaac tttaaaagtg ctcatcattg gaaaacgttc ttcggggcga aaactctcaa	8820
ggatcttacc gctgttgaga tccagttcga tgtaaccac tcgtgcaccc aactgatctt	8880
cagcatcttt tactttcacc agcgtttctg ggtgagcaaa aacaggaagg caaaatgccg	8940
caaaaaaggg aataagggcg acacggaaat gttgaatact catactcttc ctttttcaat	9000
attattgaag catttatcag gggtattgtc tcatgagcgg atacatattt gaatgtattt	9060
agaaaaataa acaaataggg gttccgcgca catttccccg aaaagtgcca cctgacgtct	9120

aagaaacat tattatcatg acattaacct ataaaaatag gcgtatcacg aggccctttc	9180
gtctcgcgcg tttcgggtgat gacgggtgaaa acctctgaca catgcagctc ccggagacgg	9240
tcacagcttg tctgtaagcg gatgccggga gcagacaagc ccgtcagggc gcgtcagcgg	9300
gtgttggcgg gtgtcggggc tggcttaact atgcggcatc agagcagatt gtactgagag	9360
tgcaccatat gcggtgtgaa ataccgcaca gatgcgtaag gagaaaatac cgcacacaggc	9420
gccattcgcc attcaggctg cgcaactgtt gggaagggcg atcgggtgcgg gcctcttcgc	9480
tattacgcca gctggcgaaa gggggatgtg ctgcaaggcg attaagttgg gtaacgccag	9540
ggttttccca gtcacgacgt tgtaaaacga cggccagtga attcgttaag acccactttc	9600
acatttaagt tgtttttcta atccgcatat gatcaattca aggccgaata agaaggctgg	9660
ctctgcacct tggatgatcaa ataattcgat agcttgtcgt aataatggcg gcatactatc	9720
agtagtaggt gtttcccttt cttcttttagc gacttgatgc tcttgatctt ccaatacgca	9780
acctaaagta aaatgcccc aagcgtgag tgcatataat gcattctcta gtgaaaaacc	9840
ttgttggcat aaaaaggcta attgattttc gagagtttca tactgttttt ctgtaggccg	9900
tgtacctaaa tgtacttttg ctccatcgcg atgacttagt aaagcacatc taaaactttt	9960
agcgttatta cgtaaaaaat cttgccagct ttccccttct aaagggcaaa agtgagtatg	10020
gtgcctatct aacatctcaa tggctaaggc gtcgagcaaa gcccgttat tttttacatg	10080
ccaatacaat gtaggctgct ctacacctag cttctgggcg agtttacggg ttgttaaacc	10140
ttcgattccg acctcattaa gcagctctaa tgcgctgtta atcactttac ttttatctaa	10200
tctagacatc attaatcct aatttttggt gacactctat cattgataga gttattttac	10260
cactccctat cagtgataga gaaaagtga	10289

<210> 7

<211> 10268

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 7

gcgaaaaatc aataaggagg caacaagatg tgcgaaaaac atcttaatca tgctaaggag	60

gttttctaata gatcttcgag acaatttcta cgtcgaatca cgaagaagtt gtgtattgcc	120

ataacaagga cgccggcttg aaagcaatca tcgcgattca caacactgta ctcgggtccgg	180
ctctgggtgg cactcgcacg tggccctacg ctagcgaaga ggaagcactg aaagatgtcc	240
ttcgtttatc ccgtgggatg acctacaaag ctgcggtttc aggtctaaac ctgggcggcg	300
gtaaagctgt gatctggggg gatccgaata aagacaagtc tgaagcgctg tttagagcct	360
tcggacgggt tgtaaacagc ctgggcggac gctacattac cgcggaggac gttggcattg	420
atgttaacga catggaatat gtgctgcgtg aaactgatta cgtcaccggg gtacatcagg	480
ttcacgggtg gagtggtgat ctttctccat tcaccgcata tggcactctg caaggcctga	540
tggccgctct gcaagtgaag ttcggtaacg aagacgtagg caattacagc tacgctgttc	600
aggggtgtggg tcacgttggc atggaatttg ttaaactgct gcgtgagcgc ggtgcaaagg	660
ttttcgtcac tgacatcaac aaagatgcgg tccagcgtgc tgtggacgaa tttggttgtg	720
aggcagtagc cctggatgaa atctatgacg ttgattgcga cgtgtactcc ccgaccgctc	780
tgggcggcac cgtgaacgat aaaactttac cgcgtctgaa atgtaaggta atctgcgggtg	840
cggcaaacia ccagttagct aatgatgaga taggcgtgga actggaaaag aaaggcatcc	900
tctatgctcc ggactacgcg gtcaacgcgg gtgggctgat gaacgttagc ctggaaatcg	960
atggatacaa ccgcgaacgt gcgatgcgta tgatgcgtac catttattac aatttgggtc	1020
gcattttcga aatctctaag cgcgacggca tccctacatt ccgagccgcc gatcgtatgg	1080
ctgaagaacg cataacggcc atcggtaaac tgcgtttacc gcatttgggc gctgcggcac	1140
cgcgcttcca gggccgacgt ggcaactaag tcaataaagg catataaaag gaggttaata	1200
acatgaaagt taaagtaaaa catcttaatc atgcacagga gactttctaa tgtcgacgac	1260
aaccgttggt gactacttgc tgtatcgctt aaacgaaatc ggcatagagc acctcttcgg	1320
agtgccagggt gattacaatc tgcaatttct ggatcatgta atcgaccacc ctcagctgac	1380
ttgggtcggc tgcactaacg aacttaacgc tgcctacgca gctgatgggt atgcgcgttg	1440
tcgtccggct gcggcactgc tgaccacctt cggggttggc gaactgagcg ctattaatgg	1500
catcgcagggt tcctacgcgg agtatctgcc ggtaatacat atcgttggtg caccgagtct	1560
atcagcccag cagcagggcg acctgattca ccactctctt ggcaagggtg atttttccag	1620
cttcctgagg atgtcccaac cgggtgtctgt tgcgcaggct gctctgactc ctgataacgc	1680
atgcaaggaa atcgaccgcg tactggcgga agtcctcatt cagcgtcgtc ccggctacct	1740

gctgctgtct accgacgtgg ctgctgcgcc ggcggctctg ccacaaagca ctctttcttt	1800
gccgaccgcc ccggatcatc gcgcagttct ggctgctttc agcgacgctg ctgagcagat	1860
gctggctcag gccaaaagcg tctctctact ggcggacttt ctggctgatc gtttcggtgt	1920
tactcgagca ctggccgcgt ggcttcagca ggttccgcta ccgcacgcca ctctgttaat	1980
gggtaaaggc gttctgagtg aacagcaacc agggttcgtg ggtacctacg ctggtgcggc	2040
atctatcgat tcgacgcgtg gcgcaatcga agaagctggg gtaattatcg gagtgggagt	2100
tagattttcc gacactatca cagcaggctt ctgcagcag atcgacgcc gccgttttat	2160
agacattcaa cccttcttct ctctgtattgg cgatcgccag tttgatcact tgccgatgca	2220
ggctgccgtc gcagccctgc atcaactgtg tcttcgttat cagcagcagt ggtctatcac	2280
cgctcctagc ccgcctgcac tgccgccggc tgctggtagc gagctgtccc agaacgcatt	2340
ctggcaggcg atgcagaact tcatccgcc tggggacctg ttggtggccg accaaggtac	2400
tgccggcgtt gccgcagcgg cgctgcgctt accgcagaat tgccagctgc ttgtgcagcc	2460
gctgtggggc tcaatcggtt acagtctgcc ggccacctt ggtgctcaga cggcagatac	2520
agagcgtcgt gtaatcctaa tcattggcga tggttcagcg caattaacta ttcaggaact	2580
ttccagtatg atgcgtgacg gcttgaaacc tatcatcttt ctctgaaca acaacggtta	2640
caccgttgaa cgggcgattc acggcgcgga gcaacgttat aacgatatcg ctgcttgga	2700
ttggacccaa ctgccccagg cgctgagtggt tcattgcccc gcgcagagct ggcgagtcgt	2760
tgaaacggtg cagctgaccg acgtaatgaa agtcatcgct gcttctccgc gtctgagctt	2820
ggtagaagtt gttctgcctg caatggatgt cccaccgctg ctgcaagcag tgagtgccgc	2880
tctgaaccag cgcaactcct cttagacca gaggtaagga ggtaacaacc atgcgagtgt	2940
tgaagaaaca tcttaatcat gcaggggagg gtttctaatt cgagcatgc agtttgatga	3000
gtacggtgca cccctgaaag cgttctcata tgaagacctg accccgcaag ggaaggaagt	3060
agtcgttagg atcgaagcct gtggtgtgtg ccactctgat attcatcttc acgagggcta	3120
cttcgacatg ggcggtggca ataaagctga tgttactcgt gctcgcgaac tcccttttac	3180
attgggtcat gaaatcgttg gcgaagtggg agcaactgga ccagggtgtca ccggcgctaa	3240
accgggcgac aaacgtattg tgtacctgtg gatcgggtgc ggcgactgcc cgaaatgcaa	3300
cagtggtagg gatcagtcct gtgcgcgtcc acgtaacctg ggtgttcacg ttgacggtgg	3360

ctattcgacg cacgtaaaga taccggacga aaaattcctg ttcgcctacg atggtattcc	3420
tactgagtta gcgggaacct atgcttgcag cggcatcacc gcttatgggtg cactgatgaa	3480
agcaaaggaa gcggctgaaa gatctggcta catcgggtctg attggcgctg gtggcggttg	3540
catggctgggt ctgatgctgg ccaaagcagc gatcggggct aaaactgtag tctttgatata	3600
cgacgacgca aaactggaag ctgcgacccg tgccggggcg gattacgtgt tcaactccgg	3660
tgcaaaagaa acacgcaagg aagttatgaa actaacgaat ggtggcctgt ctggtgctgt	3720
tgatttcgtt ggcagcgata aaagcgctct gtttggaaatc aacgccttgg gtcagaacgg	3780
cgtgctgggtc ataattggac tgttcgggtg cgctatgact gttccgggtac ccctgttccc	3840
gctgaaaggg atcaccgtac gtggctcata cgtaggttcc ctgcaagaga tgagtatat	3900
gatggagtta gttcgcgctg ggaaagtcc tccgatgccg gtaaaaactc ggccactgga	3960
cgctgcctgg gaaacccttg aggatctacg ccatggtaaa atcgtgggcc gtgtgttct	4020
gaccccataa gaaggagata tacatatgac ccatcaatta agatcgcgcg atatcatcgc	4080
tctgggcttt atgacatttg cgttgttcgt cggcgcaggt aacattattt tccctccaat	4140
ggtcggcttg caggcaggcg aacacgtctg gactgcggca ttcggcttcc tcattactgc	4200
cgttggccta ccggtattaa cggtagtggc gctggcaaaa gttggcggcg gtgttgacag	4260
tctcagcacg ccaattggta aagtcgctgg cgtactgctg gcaacagttt gttacctggc	4320
ggtggggccg ctttttgcta cgccgcgtac agctaccgtt tcttttgaag tgggcattgc	4380
gccgctgacg ggtgattccg cgctgccgct gtttatttac agcctgggtct atttcgctat	4440
cgttattctg gtttcgctct atccgggcaa gctgctggat accgtgggca acttccttgc	4500
gccgctgaaa attatcgcgc tggtcatcct gtctgttgcc gcaattatct ggccggcggg	4560
ttctatcagt acggcgactg aggcttatca aaacgctgcg ttttctaacg gcttcgtcaa	4620
cggctatctg accatggata cgctgggcgc aatgggtgtt ggtatcgtaa ttgttaacgc	4680
ggcgcgttct cgtggcgtaa ccgaagcgcg tctgctgacc cgttataaccg tctgggctgg	4740
cctgatggcg ggtgttggtc tgactctgct gtacctggcg ctgttccgtc tgggttcaga	4800
cagcgcgtcg ctggtcgac agtctgcaaa cggcgcggcg atcctgcatg cttacgttca	4860
gcataccttt ggccggcgcg gtagcttcct gctggcggcg ttaatcttca tcgcctgcct	4920
ggtcacggcg gttggcctga cctgtgcttg tgcagaattc ttcgccagat acgtaccgct	4980

ctcttatcgt acgctgggtgt ttatcctcgg cggcttctcg atggtggtgt ctaacctcgg	5040
cttgagccag ctgattcaga tctctgtacc ggtgctgacc gccatttatc cgccgtgtat	5100
cgcaactgggtt gtattaagtt ttacacgctc atgggtggcat aattcgtccc gcgtgattgc	5160
tccgccgatg tttatcagcc tgcttttttg tattctcgac gggatcaagg catctgcatt	5220
cagcgatatc ttaccgtcct gggcgcagcg tttaccgctg gccgaacaag gtctggcgtg	5280
gttaatgcca acagtgggtga tgggtggttct ggccattatc tgggatcgtg cggcaggctg	5340
tcaggtgacc tccagcgctc actaatacgc atggcatgga tgaccgatgg tagtgtgggc	5400
tctcccatg cgagagtagg gaactgccag gcatcaaata aaacgaaagg ctcaagtcgaa	5460
agactggggc tttcgtttta tctgtttgtt gtcggtgaac gctctcctga gtaggacaaa	5520
tccgccggga gcggatttga acgttgcgaa gcaacggccc ggagggtggc gggcaggacg	5580
cccgccataa actgccaggc atcaaattaa gcagaaggcc atcctgacgg atggcctttt	5640
tgctgggcca gtgccaagct tgcatgcgtg ccagctgcat taatgaagaa atcatgctgg	5700
aagaataaca gctcactcaa aggcggtagt acgggttttg ctgcccgcaa acgggctgtt	5760
ctggtgttgc tagtttgtaa tcagaatcgc agatccggct tcagccgggtt tgccggctga	5820
aagcgctatt tcttcagaa ttgcatgat tttttccca cgggaggcgt cactggctcc	5880
cgtgttgtcg gcagctttga ttcgataagc agcatcgctt gtttcaggct gtctatgtgt	5940
gactgttgag ctgtaacaag ttgtctcagg tgttcaattt catgttctag ttgctttgtt	6000
ttactggttt cacctgttct attaggtgtt acatgctgtt catctgttac attgtcgatc	6060
tgttcatggg gaacagcttt gaatgcacca aaaactcgta aaagctctga tgtatctatc	6120
ttttttacac cgttttcac tgtgcatatg gacagttttc cttttgatat gtaacgggtga	6180
acagttgttc tacttttgtt tgtagtctt gatgcttcac tgatagatac aagagccata	6240
agaacctcag atccttccgt atttagccag tatgttctct agtgtgggtt gttgtttttg	6300
cgtgagccat gagaacgaac cattgagatc atacttactt tgcatgtcac tcaaaaattt	6360
tgcctcaaaa ctgggtgagct gaatttttgc agttaaaagca tcgtgtagtg tttttcttag	6420
tccgttatgt aggtaggaat ctgatgtaat ggttggttgg attttgtcac cattcatttt	6480
tatctgggtg ttctcaagtt cggttacgag atccatttgt ctatctagtt caacttgga	6540
aatcaacgta tcagtcgggc ggcctcgctt atcaaccacc aatttcatat tgctgtaagt	6600

gtttaaatct ttacttattg gtttcaaaac ccattgggta agccttttaa actcatggta	6660
gttattttca agcattaaca tgaacttaaa ttcacaaagg ctaatctcta tatttgcctt	6720
gtgagttttc ttttgtgtta gttcttttaa taaccactca taaatcctca tagagtattt	6780
gttttcaaaa gacttaacat gttccagatt atattttatg aattttttta actggaaaag	6840
ataaggcaat atctcttcac taaaaactaa ttctaatttt tcgcttgaga acttggcata	6900
gtttgtccac tggaaaatct caaagccttt aaccaaagga ttcctgattt ccacagttct	6960
cgtcatcagc tctctgggtg ctttagctaa tacaccataa gcattttccc tactgatgtt	7020
catcatctga gcgtattggt tataagtga cgateaccgtc cgttctttcc ttgtaggggt	7080
ttcaatcgtg gggttgagta gtgccacaca gcataaaatt agcttgggtt catgctccgt	7140
taagtcatag cgactaatcg ctagttcatt tgctttgaaa acaactaatt cagacataca	7200
tctcaattgg tctaggtgat tttaatcact ataccaattg agatgggcta gtcaatgata	7260
attactagtc cttttccttt gagttgtggg tatctgtaaa ttctgctaga cttttgctgg	7320
aaaacttgta aattctgcta gaccctctgt aaattccgct agacctttgt gtgttttttt	7380
tgtttatatt caagtgggta taatttatag aataaagaaa gaataaaaaa agataaaaaa	7440
aatagatccc agccctgtgt ataactcact acttttagtca gttccgcagt attacaaaag	7500
gatgtcgcaa acgctgtttg ctctctaca aaacagacct taaaacccta aaggcttaag	7560
tagcaccctc gcaagctcgg gcaaatcgct gaatattcct tttgtctccg accatcaggc	7620
acctgagtcg ctgtcttttt cgtgacattc agttcgctgc gctcacggct ctggcagtga	7680
atgggggtaa atggcactac aggcgccttt tatggattca tgcaaggaaa ctaccataa	7740
tacaagaaaa gcccgtcacg ggcttctcag ggcgttttat ggcgggtctg ctatgtgggtg	7800
ctatctgact ttttgctgtt cagcagttcc tgccctctga ttttccagtc tgaccacttc	7860
ggattatccc gtgacaggtc attcagactg gctaatgcac ccagtaaggc agcggtatca	7920
tcaacaggct taccgtctt actgtctttt ctacgggggtc tgacgctcag tggaacgaaa	7980
actcacgtta agggattttg gtcattgagat tatcaaaaag gatcttcacc tagatccttt	8040
taaattaata atgaagtttt aaatcaatct aaagtatata tgagtaaact tggcttgaca	8100
gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcatcca	8160
tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggtta ccatctggcc	8220

ccagtgcctgc aatgataccg cgagaaccac gctcaccggc tccagattta tcagcaataa	8280
accagccagc cggaagggcc gagcgcagaa gtggctctgc aactttatcc gcctccatcc	8340
agtctattaa ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca	8400
acgttgctgc cattgctaca ggcatcgtgg tgtcacgctc gtcgtttggg atggcttcat	8460
tcagctccgg ttcccaacga tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag	8520
cggttagctc cttcggtcct ccgatcgttg tcagaagtaa gttggccgca gtgttatcac	8580
tcatgggttat ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt	8640
ctgtgactgg tgagtactca accaagtcac tctgagaata gtgtatgcgg cgaccgagtt	8700
gctcttgccc gggttcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc	8760
tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat	8820
ccagttcgat gtaaccact cgtgcacca actgatcttc agcatctttt actttcacca	8880
gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataaggcgca	8940
cacggaaatg ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg	9000
gttattgtct catgagcgga tacatatctg aatgtattta gaaaaataaa caaatagggg	9060
ttccgcgcac atttccccga aaagtgccac ctgacgtcta agaaaccatt attatcatga	9120
cattaaccta taaaaatagg cgtatcacga ggccctttcg tctcgcgcgt ttcggtgatg	9180
acggtgaaaa cctctgacac atgcagctcc cggagacggg cacagcttgt ctgtaagcgg	9240
atgccgggag cagacaagcc cgtcagggcg cgtcagcggg tgttggcggg tgtcggggct	9300
ggcttaacta tgcggcatca gagcagattg tactgagagt gcaccatatg cgggtgtgaaa	9360
taccgcacag atgcgtaagg agaaaatacc gcatcaggcg ccattcgcca ttcaggctgc	9420
gcaactgttg ggaagggcga tcggtgcggg cctcttcgct attacgccag ctggcgaaag	9480
ggggatgtgc tgcaaggcga ttaagttggg taacgccagg gttttcccag tcacgacgtt	9540
gtaaaacgac ggccagtga ttcgttaaga cccactttca catttaagtt gtttttctaa	9600
tccgcatatg atcaattcaa ggccgaataa gaaggctggc tctgcacctt ggtgatcaaa	9660
taattcgata gcttgtcgta ataattggcg catactatca gtagtaggtg tttccctttc	9720
ttcttttagcg acttgatgct cttgatcttc caatacgcaa cctaaagtaa aatgccccac	9780
agcgctgagt gcatataatg cattctctag tgaaaaacct tgttggcata aaaaggctaa	9840

ttgatttttcg agagtttcat actgtttttc tgtaggccgt gtacctaaat gtacttttgc	9900
tccatcgcgga tgacttagta aagcacatct aaaactttta gcgttattac gtaaaaaatc	9960
ttgccagctt tccccttcta aagggcacaa gtgagtatgg tgcctatcta acatctcaat	10020
ggctaaggcg tcgagcaaag cccgcttatt ttttacatgc caatacaatg taggctgctc	10080
tacacctagc ttctgggcga gtttacgggt tggttaaact tcgattccga cctcattaag	10140
cagctctaata gcgctgttaa tcactttact tttatctaata ctagacatca ttaattccta	10200
atttttgttg acactctatc attgatagag ttattttacc actccctatc agtgatagag	10260
aaaagtga	10268

<210> 8

<211> 10256

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 8

gcgaaaaatc aataaggagg caacaagatg tgcgaaaaac atcttaatca tgctaaggag	60
gttttctaata gcagatcttc gacactttgc aatcaatggg ccatgagcag gttgtcctat	120
gtagcgataa gaccacgggt ctgcgcgcca ttatcgctat acacgataca tccttagggc	180
cggcgcttgg tggtaaccgt atgtggcagt atgcaactga cgacgatgct attactgacg	240
cactccgtct gtctcggggc atgacctaca aagctgcggt ttctggcgta aatctgggcg	300
gtggtaaagc cgttatcatc ggaaaccctc acagtataa aagcgaagcg ctgtttcgcg	360
cttacggcag aatggtggaa tcccagcgtg ggcgttacat caccgccgaa gacgttggtg	420
ctagcgtacg tgatatggag tggattcgca tggaaaccaa atatgtaacg ggcgtgggtg	480
gcaacggagg ctctggtgac ccctctccag ttaccgctct ggggtgtttac tcgggcatga	540
aggcatgcmc taaatcagtc tatggtactg atgcgctgag cggtaaaagg atcgtgggtc	600
agggcgcggg taacgttgca tcccatctgg ttcacagtct ggtaaaagaa ggcgctgtgg	660
ttttcgtcac tgacatctac gaagaaaagg ccaaagcatt agcggctgaa acgggcgcta	720
ccgtgattcg caccgacgag gtttttacta cacaatgcga tatcttctct ccgaacgctc	780

tgggggccgt cctgaacgat gaaactattc cgcagctcac atgcgctatc gtagctggtg	840
gtgcaaacia tcagcttaaa atcgaacaac gtcacgccac ggctctgcaa gagaaaggca	900
ttctgtatgc gccggattac gtaatcaacg ccgggggcct catgaatgtg gcgagcgaag	960
ttgacggcta caaccgtgaa aaggttatgc gccaggctga aggtatttac gatattacta	1020
tgaacatcct aaataccgcg cgtgagcgta acatcctgac catcgaagca tccaacgcga	1080
ttgctgaaga gcggatcaac aaagttcgcc atgttcacgg gaacttcatc gggtccccgt	1140
ctattcgcgg agtataagtc aataaaggca tataaaagga ggtaataaac atgaaagtta	1200
aagtaaaaca tcttaatcat gcacaggaga ctttctaata tcgacgacaa ccgttggtga	1260
ctacttgctg tatcgcttaa acgaaatcgg cattgagcac ctcttcggag tgccagggtga	1320
ttacaatctg caatttctgg atcatgtaat cgaccaccct cagctgactt gggtcggctg	1380
cactaacgaa cttaacgctg cctacgcagc tgatgggttat gcgcgttgtc gtccggctgc	1440
ggcactgctg accaccttcg gggttggcga actgagcgct attaattggca tcgcaggttc	1500
ctacgcggag tatctgccgg taatacatat cgttgggtgca ccgagtctat cagcccagca	1560
gcagggcgac ctgattcacc actctcttgg cgaagggtgat ttttccagct tcctgaggat	1620
gtcccaaccg gtgtctgttg cgcaggctgc tctgactcct gataacgcat gcaaggaaat	1680
cgaccgcgta ctggcggaag tcctcattca gcgtcgtccc ggctacctgc tgctgtctac	1740
cgacgtggct gctgcgccgg cggctctgcc acaaagcact ctttctttgc cgaccgcccc	1800
ggatcatcgc gcagttctgg ctgctttcag cgacgctgct gagcagatgc tggctcaggc	1860
caaaagcgtc tctctactgg cggactttct ggctgatcgt ttcgggtgta ctcgagcact	1920
ggccgcgtgg cttcagcagg ttccgctacc gcacgccact ctgttaatat gtaaaggcgt	1980
tctgagtga cagcaaccag ggttcgtggg tacctacgct ggtgcggcat ctatcgattc	2040
gacgcgtggc gcaatcgaag aagctggggg aattatcgga gtgggagtta gattttccga	2100
cactatcaca gcaggcttct cgcagcagat cgacgcccgc cgttttatag acattcaacc	2160
cttcttctct cgtattggcg atcgccagtt tgatcacttg ccgatgcagg ctgccgtcgc	2220
agccctgcat caactgtgtc ttcgttatca gcagcagtgg tctatcaccg ctctagccc	2280
gcctgcactg ccgccggctg ctggtagcga gctgtcccag aacgcattct ggaggcgat	2340
gcagaacttc atccgccctg gggacctgtt ggtggccgac caagggtact cggcgttcgg	2400

cgcagcggcg	ctgcgcttac	cgcagaattg	ccagctgctt	gtgcagccgc	tgtggggctc	2460
aatcggttac	agtctgccgg	ccacctttgg	tgctcagacg	gcagatacag	agcgtcgtgt	2520
aatcctaadc	attggcgatg	gttcagcgca	attaactatt	caggaacttt	ccagtatgat	2580
gcgtgacggc	ttgaaaccta	tcattctttct	cctgaacaac	aacggttaca	ccgttgaacg	2640
ggcgattcac	ggcgcggagc	aacgtttataa	cgatatcgct	gcttggaatt	ggacccaact	2700
gccccaggcg	ctgagtgttc	attgcccagc	gcagagctgg	cgagtcgttg	aaacggtgca	2760
gctgaccgac	gtaatgaaag	tcatcgctgc	ttctccgcgt	ctgagcttgg	tagaagttgt	2820
tctgcctgca	atggatgtcc	caccgctgct	gcaagcagtg	agtgccgctc	tgaaccagcg	2880
caactcctct	taagaccaga	ggtaaggagg	taacaacat	gcgagtgttg	aagaaacatc	2940
ttaatcatgc	aggggagggt	ttctaatacg	cagcatgcag	tttgatgagt	acggtgcacc	3000
cctgaaagcg	tttcatatg	aagacccgac	cccgaaggg	aaggaagtag	tcgttaggat	3060
cgaagcctgt	ggtgtgtgcc	actctgatat	tcattctcac	gagggtact	tcgacatggg	3120
cgggtggcaat	aaagctgatg	ttactcgtgc	tcgcgaactc	ccttttacat	tgggtcatga	3180
aatcgttggc	gaagtggtag	caactggacc	aggtgtcacc	ggcgctaaac	cgggcgacaa	3240
acgtattgtg	tacccgtgga	tcgggtgcgg	cgactgcccg	aaatgcaaca	gtggtgagga	3300
tcagtcctgt	gcgcgtccac	gtaacctggg	tgttcacgtt	gacgggtggct	attcgacgca	3360
cgtaaagata	ccggacgaaa	aattcctgtt	cgcctacgat	ggtattccta	ctgagttagc	3420
gggaacctat	gcttgacgag	gcatcaccgc	ttatgggtgca	ctgatgaaag	caaaggaagc	3480
ggctgaaaga	tctggctaca	tcggtctgat	tggcgctggg	ggcggttgca	tggctggtct	3540
gatgctggcc	aaagcagcga	tcgggggctaa	aactgtagtc	tttgatatcg	acgacgcaaa	3600
actggaagct	gcgacccgtg	ccggggcgga	ttacgtgttc	aactccggtg	caaaagaaac	3660
acgcaaggaa	gttatgaaac	taacgaatgg	tggcctgtct	ggtgctgttg	atttcgttgg	3720
cagcgataaa	agcgctctgt	ttggaatcaa	cgccttgggt	cagaacggcg	tgctggtcat	3780
aattggactg	ttcgggtggcg	ctatgactgt	tccggtaccc	ctgttcccgc	tgaaggggat	3840
caccgtacgt	ggctcatacg	taggttcctt	gcaagagatg	agtgatatga	tggagttagt	3900
tcgcgtggg	aaagttcctc	cgatgccggt	aaaaactcgg	ccactggacg	ctgcctggga	3960
aacccttgag	gatctacgcc	atggtaaaat	cgtgggccgt	gttggttctga	ccccataaga	4020

aggagatata catatgaccc atcaattaag atcgcgcgat atcatcgctc tgggctttat	4080
gacatttgcg ttgttcgtcg gcgaggttaa cattattttc cctccaatgg tcggcttgca	4140
ggcaggcgaa cacgtctgga ctgcggcatt cggcttcctc attactgccg ttggcctacc	4200
ggtattaacg gtagtggcgc tggcaaaagt tggcggcggg gttgacagtc tcagcacgcc	4260
aattggtaaa gtcgctggcg tactgctggc aacagtttgt tacctggcgg tggggccgct	4320
ttttgctacg ccggtacag ctaccgtttc ttttgaagtg ggcatcgcg cgctgacggg	4380
tgattccgcg ctgccgtgt ttatttacag cctgggtctat ttcgctatcg ttattctggt	4440
ttcgctctat ccgggcaagc tgctggatac cgtgggcaac ttccttgcg cgctgaaaat	4500
tatcgcgctg gtcacctgt ctgttgccgc aattatctgg ccggcgggtt ctatcagtac	4560
ggcgactgag gcttatcaaa acgctgcgtt ttctaacggc ttcgtcaacg gctatctgac	4620
catggatacg ctgggcgcaa tgggtgttgg tatcgttatt gttaacgcgg cgcgttctcg	4680
tggcgttacc gaagcgcgtc tgctgaccg ttataccgtc tgggctggcc tgatggcggg	4740
tgttggtctg actctgctgt acctggcgt gttccgtctg gggtcagaca gcgcgtcgct	4800
ggtcgatcag tctgcaaacg gtgcggcgat cctgcatgct tacgttcagc ataccttgg	4860
cggcggcggg agcttcctgc tggcggcgtt aatcttcac gcctgcctgg tcacggcggg	4920
tggcctgacc tgtgcttggt cagaattctt cgcccagtac gtaccgtct cttatcgta	4980
gctgggtgtt atcctcggcg gcttctcgat ggtggtgtct aacctcggct tgagccagct	5040
gattcagatc tctgtaccgg tgctgaccgc catttatccg ccgtgtatcg cactgggtgt	5100
attaagtttt acacgctcat ggtggcataa ttcgtccgc gtgattgctc cgccgatgtt	5160
tatcagcctg ctttttggtt ttctcgacgg gatcaaggca tctgcattca gcgatatctt	5220
accgtcctgg gcgcagcgtt taccgtggc cgaacaagg ctggcgtgg taatgccaac	5280
agtgggtgat gtggttctgg ccattatctg ggatcgtgc gcaggctcgtc aggtgacctc	5340
cagcgtcac taatacgc atggatggat accgatggta gtgtgggctc tccccatgcg	5400
agagtaggga actgccaggc atcaaataaa acgaaaggct cagtcgaaag actgggcctt	5460
tcgttttatc tgttgtttgt cgggtgaacgc tctcctgagt aggacaaatc cgccgggagc	5520
ggatttgaac gttgcgaagc aacggcccgg aggggtggcg gcaggacgcc cgccataaac	5580
tgccaggcat caaattaagc agaaggccat cctgacggat ggccctttttg cgtggccagt	5640

gccaagcttg catgcgtgcc agctgcatta atgaagaaat catgctggaa gaataacagc	5700
tcactcaaag gcggtagtagc gggttttgct gcccgcacaaac gggctgttct ggtgttgcta	5760
gtttgttatc agaatcgtagc atccggcttc agccgggttg ccggctgaaa gcgctatttc	5820
ttccagaatt gccatgattt tttccccacg ggaggcgtca ctggctcccg tgttgtcggc	5880
agctttgatt cgataagcag catcgcctgt ttcaggctgt ctatgtgtga ctgttgagct	5940
gtaacaagtt gtctcagggtg ttcaatttca tgttctagtt gctttgtttt actggtttca	6000
cctgttctat taggtgttac atgctgttca tctgtttacat tgtcgatctg ttcattgggtga	6060
acagctttga atgcacacaaa aactcgtaaa agctctgatg tatctatctt ttttacaccg	6120
ttttcatctg tgcataatgga cagttttccc tttgatattg aacgggtgaac agttgttcta	6180
cttttgtttg ttagtcttga tgcttcactg atagatacaa gagccataag aacctcagat	6240
ccttccgtat ttagccagta tgttctctag tgtgggttcgt tgtttttgcg tgagccatga	6300
gaacgaacca ttgagatcat acttactttg catgtcactc aaaaattttg cctcaaaact	6360
ggtgagctga atttttgcag ttaaagcatc gtgtagtggt tttcttagtc cgttatgtag	6420
gtaggaatct gatgtaatgg ttgttggtat tttgtcacca ttcattttta tctggttgtt	6480
ctcaagttcg gttacgagat ccatttgtct atctagttca acttggaata tcaacgtatc	6540
agtcgggcgg cctcgcttat caaccaccaa tttcatattg ctgtaagtgt ttaaattctt	6600
acttattggt ttcaaaaccc attgggttaag ctttttaaac tcatggtagt tattttcaag	6660
cattaacatg aacttaaatt catcaaggct aatctctata tttgccttgt gagttttctt	6720
ttgtgttagt tcttttaata accactcata aatcctcata gagtatttgt tttcaaaaga	6780
cttaacatgt tccagattat attttatgaa tttttttaac tggaaaagat aaggcaatat	6840
ctcttcacta aaaactaatt ctaatttttc gcttgagaac ttggcatagt ttgtccactg	6900
gaaaatctca aagcctttaa ccaaaggatt cctgatttcc acagttctcg tcatcagctc	6960
tctggttgct ttagctaata caccataagc attttcccta ctgatgttca tcatctgagc	7020
gtattgggtta taagtgaacg ataccgtccg ttctttcctt gtagggtttt caatcgtggg	7080
gttgagtagt gccacacagc ataaaattag cttgggttca tgctccgtta agtcatagcg	7140
actaatcgct agttcatttg ctttgaaaac aactaattca gacatacatc tcaattggtc	7200
taggtgattt taatcactat accaattgag atgggctagt caatgataat tactagtcct	7260

tttcctttga gttgtgggta tctgtaaatt ctgctagacc tttgctggaa aacttgtaaa	7320
ttctgctaga ccctctgtaa attccgctag acctttgtgt gttttttttg tttatatattca	7380
agtgggtata atttatagaa taaagaaaga ataaaaaaag ataaaaagaa tagatcccag	7440
ccctgtgtat aactcactac tttagtcagt tccgcagtat tacaaaagga tgtcgcaaac	7500
gctgtttgct cctctacaaa acagacctta aaaccctaaa ggcttaagta gcaccctcgc	7560
aagctcgggc aaatcgctga atattccttt tgtctccgac catcaggcac ctgagtcgct	7620
gtctttttcg tgacattcag ttcgctgcgc tcacggctct ggccagtgaat gggggtaaat	7680
ggcactacag gcgcctttta tggattcatg caaggaaact acccataata caagaaaagc	7740
ccgtcacggg cttctcaggg cgttttatgg cgggtctgct atgtggtgct atctgacttt	7800
ttgctgttca gcagttcctg ccctctgatt ttccagtctg accacttcgg attatcccgt	7860
gacaggcatc tcagactggc taatgcaccc agtaaggcag cggatatcatc aacaggctta	7920
cccgtcttac tgtcttttct acgggggtctg acgctcagtg gaacgaaaac tcacgttaag	7980
ggatttttgt catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat	8040
gaagttttta atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct	8100
taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcattccata gttgcctgac	8160
tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa	8220
tgataccgcg agaaccacgc tcaccggctc cagatttatc agcaataaac cagccagccg	8280
gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt	8340
gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca	8400
ttgctacagg catcgtggtg tcacgctcgt cgtttgggtat ggcttcattc agctccggtt	8460
cccaacgatac aaggcgagtt acatgatccc ccatgtttgtg caaaaaagcg gttagctcct	8520
tcggtcctcc gatcgttgctc agaagtaagt tggccgcagt gttatcactc atggttatgg	8580
cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg	8640
agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg	8700
cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgtc atcattggaa	8760
aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt	8820
aaccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt	8880

gagcaaaaac aggaaggcaa aatgccgcaa aaaaggggaat aagggcgaca cggaaatgtt	8940
gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca	9000
tgagcggata catatttgaa tgtatttaga aaaataaaca aataggggtt ccgcgcacat	9060
ttccccgaaa agtgccacct gacgtctaag aaaccattat tatcatgaca ttaacctata	9120
aaaataggcg tatcacgagg ccttttcgtc tcgcgcgttt cggatgatgac ggtgaaaacc	9180
tctgacacat gcagctcccg gagacgggtca cagcttgtct gtaagcggat gccgggagca	9240
gacaagccccg tcagggcgcg tcagcgggtg ttggcgggtg tcggggctgg cttaaactatg	9300
cggcatcaga gcagattgta ctgagagtgc accatatgcg gtgtgaaata ccgcacagat	9360
gcgtaaggag aaaataccgc atcaggcgcc attcgccatt caggctgcgc aactgttggg	9420
aagggcgatc ggtgcgggccc tcttcgctat tacgccagct ggcgaaaggg ggatgtgctg	9480
caaggcgatt aagttgggta acgccagggt tttcccagtc acgacgttgt aaaacgacgg	9540
ccagtgaatt cgtaaagacc cactttcaca ttttaagttgt ttttctaatac cgcatatgat	9600
caattcaagg ccgaataaga aggctggctc tgcaccttgg tgatcaaata attcgatagc	9660
ttgtcgtaat aatggcggca tactatcagt agtaggtgtt tccctttctt ctttagcgac	9720
ttgatgctct tgatcttcca atacgaacc taaagtaaaa tgccccacag cgctgagtgc	9780
atataatgca ttctctagtg aaaaaccttg ttggcataaa aaggctaatt gattttcgag	9840
agtttcatac tgtttttctg taggccgtgt acctaaatgt acttttgctc catcgcgatg	9900
acttagtaaa gcacatctaa aacttttagc gttattacgt aaaaaatctt gccagctttc	9960
cccttctaaa gggcaaaagt gagtatggtg cctatctaac atctcaatgg ctaaggcgtc	10020
gagcaaagcc cgcttatttt ttacatgcca atacaatgta ggctgctcta cacctagctt	10080
ctgggcgagt ttacgggttg ttaaaccctc gattccgacc tcattaagca gctctaatagc	10140
gctgttaatc actttacttt tatctaatac agacatcatt aattcctaata tttgttgac	10200
actctatcat tgatagagtt attttaccac tccctatcag tgatagagaa aagtga	10256

<210> 9

<211> 10110

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 9

gataaacgaa tcacaaaggc gaggcaagat gaacatcttc aagaaaatgg aggaatttaa	60
ttatgaacaa ctgggtctact tctacgacag cgaaacggaa ctcaaaggta ttacctgtat	120
acacaacaca acttttagggc cggcattggg cggtacccgc ctttggaact ataactctga	180
ggaagatgcc gttgaagacg taatccgtct ggctcggggc atgacttaca aagcggcttg	240
cgccggtctg aatctggggc gcggtaaaac cgtgctgata ggtgatgcta aaaagattaa	300
atcagagtcc tacttccgtg gactggggcg ctacgttcag tcgctgaacg gcagatatat	360
caccgcggaa gacgtaaata cttctacgaa ggatatggca tacgttgcta tggaaactga	420
ctatgtggta ggcctgggag gtaaattcgg caaccctagt ccagttactg cttacgggtgc	480
atttatgggt atcaaagcgg cgctgatgaa aaaatttgag gatagctcta ttgaaggccg	540
aaccttcgca gtgcagggtg ctgggcagac gggttactat cttatcgatt acctcctagg	600
caacaacaag ttcaaagaaa aggctaaaaa gatttacttc accgaaatta acgagagcta	660
tatcgagcgt atgaacaaag aacatccgga agttgaattt atttccccgg acaaaatcta	720
ctcgctggaa gtagacgtct tcgtgccctg cgccctgggc aaaatcgta atgacaaaac	780
tatcgatgaa tttaagtgtc cgatcatcgc aggtactgca aacaacgtac tggaaaggga	840
agcgcacggc aacatgctta aagaacgtgg cattctttac gccccggact atgtgatcaa	900
tgctgggtggg ctgatcaacg tttaccacga gctgaacggg tacaataaag agaacgctat	960
tctggaagtg gaattaattt atgatcgcct actggaaata ttcaacatcg ctgattctct	1020
gaacatcagc accaatatcg ctgccaacga gttcgcggaa aaacgtatca agcaaattaa	1080
gtccttgaaa aacaacttca ttaaacgcta agaccattaa taagggaacat tttatgtcga	1140
cgacaaccgt tggtgactac ttgctgtatc gcttaaacga aatcggcatt gagcacctct	1200
tcggagtgcc aggtgattac aatctgcaat ttctggatca tgtaatcgac caccctcagc	1260
tgacttgggt cggctgcact aacgaactta acgctgccta cgcagctgat ggttatgcgc	1320
gttgtcgtcc ggctgcggca ctgctgacca cttcgggggt tggcgaactg agcgcattaa	1380
atggcatcgc aggttcctac gcggagtatc tgccggtaat acatatcggt ggtgcaccga	1440
gtctatcagc ccagcagcag ggcgacctga ttcaccactc tcttggcgaa ggtgattttt	1500
ccagcttcct gaggatgtcc caaccggtgt ctgttgcgca ggctgctctg actcctgata	1560

acgcatgcaa ggaaatcgac cgcgtactgg cggaagtcct cattcagcgt cgtccccggct	1620
acctgctgct gtctaccgac gtggctgctg cgccggcggc tctgccacaa agcactcttt	1680
ctttgccgac cgccccggat catcgcgcag ttctggctgc tttcagcgac gctgctgagc	1740
agatgctggc tcaggccaaa agcgtctctc tactggcggg ctttctggct gatcgtttcg	1800
gtgttactcg agcactggcc gcgtggcttc agcaggttcc gctaccgcac gccactctgt	1860
taatgggtaa aggcgttctg agtgaacagc aaccagggtt cgtgggtacc tacgctgggtg	1920
cggcatctat cgattcgacg cgtggcgcaa tcgaagaagc tggggtaatt atcggagtg	1980
gagttagatt ttccgacact atcacagcag gcttctcgca gcagatcgac gcccgccgtt	2040
ttatagacat tcaacccttc ttctctcgta ttggcgatcg ccagtttgat cacttgccga	2100
tgcaggctgc cgtcgcagcc ctgcatcaac tgtgtcttcg ttatcagcag cagtgggtcta	2160
tcaccgctcc tagcccgctt gcactgccgc cggctgctgg tagcgagctg tcccagaacg	2220
cattctggca ggcgatgcag aacttcatcc gccctgggga cctgttggtg gccgaccaag	2280
gtactgcggc gttcggcgca gcggcgctgc gcttaccgca gaattgccag ctgcttgctg	2340
agccgctgtg gggctcaatc gggttacagtc tgccggccac ctttggtgct cagacggcag	2400
atacagagcg tcgtgtaatc ctaatcattg gcgatggttc agcgcaatta actattcagg	2460
aactttccag tatgatgcgt gacggcttga aacctatcat ctttctcctg aacaacaacg	2520
gttacaccgt tgaacgggcg attcacggcg cggagcaacg ttataacgat atcgctgctt	2580
ggaattggac ccaactgccc caggcgctga gtgttcattg cccagcgag agctggcgag	2640
tcgttgaaac ggtgcagctg accgacgtaa tgaaagtcac cgctgcttct ccgctctga	2700
gcttggtaga agttgttctg cctgcaatgg atgtcccacc gctgctgcaa gcagtgagtg	2760
ccgctctgaa ccagcgcaac tcctcttaat ccctccctac gggttaagga gcccctgcta	2820
tgcgcagcat gcagtttgat gagtacggtg caccctgaa agcgttctca tatgaagacc	2880
cgaccccgca agggaaggaa gtagtcgtta ggatcgaagc ctgtgggtgtg tgccactctg	2940
atattcatct tcacgagggc tacttcgaca tgggcgggtg caataaagct gatgttactc	3000
gtgctcgcga actccctttt acattgggtc atgaaatcgt tggcgaagtg gtagcaactg	3060
gaccaggtgt caccggcgct aaaccgggcg acaaacgtat tgtgtaccg tggatcgggt	3120
gcggcgactg cccgaaatgc aacagtgggtg aggatcagtc ctgtgcgcgt ccacgtaacc	3180

tgggtgttca cgttgacggt ggctattcga cgcacgtaaa gataccggac gaaaaattcc	3240
tgttcgccta cgatgggtatt cctactgagt tagcgggaac ctatgcttgc agcggcatca	3300
ccgcttatgg tgcactgatg aaagcaaagg aagcggctga aagatctggc tacatcggtc	3360
tgattggcgc tgggtggcggt ggcatggctg gtctgatgct ggccaaagca gcgatcgggg	3420
ctaaaactgt agtctttgat atcgacgacg caaaactgga agctgcgacc cgtgccgggg	3480
cggattacgt gttcaactcc ggtgcaaaag aaacacgcaa ggaagttatg aaactaacga	3540
atggtggcct gtctgggtgct gttgatttcg ttggcagcga taaaagcgct ctgtttggaa	3600
tcaacgcctt gggtcagaac ggcgtgctgg tcataattgg actgttcggt ggcgctatga	3660
ctgttcgggt acccctgttc ccgctgaaag ggatcacctg acgtggctca tacgtagggt	3720
ccctgcaaga gatgagtgat atgatggagt tagttcgcgc tgggaaagtt cctccgatgc	3780
cggtaaaaac tcggccactg gacgctgcct gggaaaccct tgaggatcta cgccatggta	3840
aaatcgtggg ccgtgttggt ctgaccccat aagaaggaga tatacatatg acccatcaat	3900
taagatcgcg cgatatcatc gctctgggct ttatgacatt tgcgttggtc gtcggcgcag	3960
gtaacattat tttccctcca atggtcggct tgcaggcagg cgaacacgct tggactgcgg	4020
cattcggctt cctcattact gccgttggcc taccgggtatt aacggtagtg gcgctggcaa	4080
aagttggcgg cgggtgtgac agtctcagca cgccaattgg taaagtcgct ggcgtactgc	4140
tggcaacagt ttgttacctg gcggtggggc cgctttttgc tacgccgctg acagctaccg	4200
tttcttttga agtgggcatt gcgccgctga cgggtgattc cgcgctgccg ctgtttattt	4260
acagcctggg ctatttcgct atcgttattc tggtttcgct ctatccgggc aagctgctgg	4320
ataccgtggg caacttcctt gcgccgctga aaattatcg cgtggtcatc ctgtctgttg	4380
ccgcaattat ctggccggcg ggttctatca gtacggcgac tgaggcttat caaaacgctg	4440
cgttttctaa cggcttcgtc aacggctatc tgaccatgga tacgctgggc gcaatggtgt	4500
ttggtatcgt tattgttaac gcggcgcgtt ctcgtggcgt taccgaagcg cgtctgctga	4560
cccgttatac cgtctgggct ggcctgatgg cgggtgttgg tctgactctg ctgtacctgg	4620
cgctgttccg tctgggttca gacagcgcgt cgctggctga tcagtctgca aacggtgcgg	4680
cgatcctgca tgcttacgtt cagcatacct ttggcggcgg cggtagcttc ctgctggcgg	4740
cgttaatctt catcgcctgc ctggtcacgg cggttggcct gacctgtgct tgtgcagaat	4800

tcttcgcccc	gtacgtaccg	ctctcttata	gtacgctggg	gtttatcctc	ggcggcttct	4860
cgatggtggg	gtctaacctc	ggcttgagcc	agctgattca	gatctctgta	ccggtgctga	4920
ccgccattta	tccgccgtgt	atcgcaactg	ttgtattaag	ttttacacgc	tcatggtggc	4980
ataattcgtc	ccgcgtgatt	gctccgccga	tgtttatcag	cctgcttttt	ggtattctcg	5040
acgggatcaa	ggcatctgca	ttcagcgata	tcttaccgtc	ctgggcgcag	cgtttaccgc	5100
tggccgaaca	aggtctggcg	tggttaatgc	caacagtggg	gatggtgggt	ctggccatta	5160
tctgggatcg	tgccgcaggt	cgtcaggtga	cctccagcgc	tcactaatac	gcatggcatg	5220
gatgaccgat	ggtagtgtgg	gctctcccca	tgcgagagta	gggaactgcc	aggcatcaaa	5280
taaaacgaaa	ggctcagtcg	aaagactggg	cctttcgttt	tatctgttgt	ttgtcgggtg	5340
acgctctcct	gagtaggaca	aatccgccgg	gagcggatgt	gaacgttgcg	aagcaacggc	5400
ccggagggtg	gcgggcagga	cgcccgccat	aaactgccag	gcatcaaatt	aagcagaagg	5460
ccatcctgac	ggatggcctt	tttgctgggc	cagtgccaaag	cttgcattgc	tgccagctgc	5520
attaatgaag	aatcatgctt	ggaagaataa	cagctcactc	aaaggcggta	gtacgggttt	5580
tgctgcccgc	aaacgggctg	ttctgggtgt	gctagtttgt	tatcagaatc	gcagatccgg	5640
cttcagccgg	tttgccggct	gaaagcgcta	tttcttcag	aattgccatg	atttttccc	5700
cacgggaggc	gtcactggct	cccgtgttgt	cggcagcttt	gattcgataa	gcagcatcgc	5760
ctgtttcagg	ctgtctatgt	gtgactgttg	agctgtaaca	agttgtctca	ggtgttcaat	5820
ttcatgttct	agttgctttg	ttttactggg	ttcacctgtt	ctattagggt	ttacatgctg	5880
ttcatctgtt	acattgtcga	tctgttcatg	gtgaacagct	ttgaatgcac	caaaaactcg	5940
taaaagctct	gatgtatcta	tcttttttac	accgttttca	tctgtgcata	tggacagttt	6000
tccctttgat	atgtaacggg	gaacagtgtg	tctacttttg	tttgtttagt	ttgatgcttc	6060
actgatagat	acaagagcca	taagaacctc	agatccttcc	gtatttagcc	agtatgttct	6120
ctagtgtggg	tcgttggttt	tgctgagcc	atgagaacga	accattgaga	tcatacttac	6180
tttgcattgc	actcaaaaat	tttgccctca	aactgggtgag	ctgaattttt	gcagttaaag	6240
catcgtgtag	tggttttctt	agtccgttat	gtaggtagga	atctgatgta	atggttggtg	6300
gtattttgtc	accattcatt	tttatctggg	tgttctcaag	ttcggttacg	agatccattt	6360
gtctatctag	ttcaacttgg	aaaatcaacg	tatcagtcgg	gcggcctcgc	ttatcaacca	6420

ccaatttcat attgctgtaa gtgttttaaat ctttacttat tggtttcaaa acccattggt	6480
taagcctttt aaactcatgg tagttatttt caagcattaa catgaactta aattcatcaa	6540
ggctaattctc tatatttgcc ttgtgagttt tcttttgtgt tagttctttt aataaccact	6600
cataaatcct catagagtat ttgttttcaa aagacttaac atgttccaga ttatatttta	6660
tgaatttttt taactggaaa agataaggca atatctcttc actaaaaact aattctaatt	6720
tttcgcttga gaacttggca tagtttgtcc actggaaaat ctcaaagcct ttaaccaaag	6780
gattcctgat ttccacagtt ctcgatcatca gctctctggt tgcttttagct aatacaccat	6840
aagcattttc cctactgatg ttcattcatct gagcgtattg gttataagtg aacgataccg	6900
tccgttcttt cttttagagg ttttcaatcg tggggttgag tagtgccaca cagcataaaa	6960
ttagcttggt ttcattgctcc gttaagtcatt agcgactaat cgctagttca ttgctttga	7020
aaacaactaa ttcagacata catctcaatt ggtctagggtg attttaatca ctataccaat	7080
tgagatgggc tagtcaatga taattactag tccttttcct ttgagttgtg ggtatctgta	7140
aattctgcta gacctttgct ggaaaacttg taaattctgc tagaccctct gtaaattccg	7200
ctagaccttt gtgtgttttt tttgtttata ttcaagtgggt tataatttat agaataaaga	7260
aagaataaaa aaagataaaa agaataagatc ccagccctgt gtataactca ctactttagt	7320
cagttccgca gtattacaaa aggatgtcgc aaacgctggt tgctcctcta caaacagac	7380
cttaaaacc taaaggctta agtagcacc tcgcaagctc gggcaaactcg ctgaatattc	7440
cttttgtctc cgaccatcag gcacctgagt cgctgtcttt ttcgtgacat tcagttcgct	7500
gcgctcacgg ctctggcagt gaatgggggt aaatggcact acaggcgcct tttatggatt	7560
catgcaagga aactacccat aatacaagaa aagcccgatc cgggcttctc agggcgtttt	7620
atggcggggtc tgctatgtgg tgctatctga ctttttgctg ttcagcagtt cctgccctct	7680
gattttccag tctgaccact tcggattatc ccgtgacagg tcattcagac tggctaattgc	7740
accagtaag gcagcgggtat catcaacagg cttaccgctc ttactgtctt ttctacgggg	7800
tctgacgctc agtggaaacga aaactcacgt taagggattt tggatcatgag attatcaaaa	7860
aggatcttca cctagatcct tttaaattaa aaatgaagtt ttaaataaat ctaaagtata	7920
tatgagtaaa cttggtctga cagttaccaa tgcttaatca gtgaggcacc tatctcagcg	7980
atctgtctat ttcgttcatc catagttgcc tgactccccg tcgtgtagat aactacgata	8040

cgggagggct	taccatctgg	ccccagtgct	gcaatgatac	cgcgagaacc	acgctcaccg	8100
gctccagatt	tatcagcaat	aaaccagcca	gccggaaggg	ccgagcgcag	aagtggtcct	8160
gcaactttat	ccgcctccat	ccagtctatt	aattgttgcc	gggaagctag	agtaagtagt	8220
tcgccagtta	atagtttgcg	caacgttggt	gccattgcta	caggcatcgt	ggtgtcacgc	8280
tcgtcgtttg	gtatggcttc	attcagctcc	ggttcccaac	gatcaaggcg	agttacatga	8340
tccccatgt	tgtgcaaaaa	agcggttagc	tccttcggtc	ctccgatcgt	tgtcagaagt	8400
aagttggccg	cagtgttatc	actcatgggt	atggcagcac	tgcataattc	tcttactgtc	8460
atgccatccg	taagatgctt	ttctgtgact	ggtgagtact	caaccaagtc	attctgagaa	8520
tagtgtatgc	ggcgaccgag	ttgtctttgc	ccggcgtcaa	tacgggataa	taccgcgcca	8580
catagcagaa	ctttaaaagt	gctcatcatt	ggaaaacgtt	cttcggggcg	aaaacttca	8640
aggatcttac	cgctgttgag	atccagttcg	atgtaacca	ctcgtgcacc	caactgatct	8700
tcagcatctt	ttactttcac	cagcgtttct	gggtgagcaa	aaacaggaag	gcaaaatgcc	8760
gcaaaaaagg	gaataagggc	gacacggaaa	tgttgaatac	tcatactctt	cctttttcaa	8820
tattattgaa	gcatttatca	gggttattgt	ctcatgagcg	gatacatatt	tgaatgtatt	8880
tagaaaaata	aacaaatagg	ggttccgcgc	acatttcccc	gaaaagtgcc	acctgacgtc	8940
taagaaacca	ttattatcat	gacattaacc	tataaaaata	ggcgtatcac	gaggcccttt	9000
cgtctcgcgc	gtttcgggtga	tgacggtgaa	aacctctgac	acatgcagct	cccggagacg	9060
gtcacagctt	gtctgtaagc	ggatgccggg	agcagacaag	cccgtcaggg	cgcgtcagcg	9120
ggtgttggcg	ggtgtcgggg	ctggcttaac	tatgcggcat	cagagcagat	tgtactgaga	9180
gtgcaccata	tgcggtgtga	aataccgcac	agatgcgtaa	ggagaaaaata	ccgcatcagg	9240
cgccattcgc	cattcaggct	gcgcaactgt	tgggaagggc	gatcggtgcg	ggcctcttcg	9300
ctattacgcc	agctggcgaa	agggggatgt	gctgcaaggc	gattaagttg	ggtaacgcca	9360
gggtttttccc	agtcacgacg	ttgtaaaacg	acggccagtg	aattcgtaa	gaccactttt	9420
cacatttaag	ttgtttttct	aatccgcata	tgatcaattc	aaggccgaat	aagaaggctg	9480
gctctgcacc	ttgggtgatca	aataattcga	tagcttgtcg	taataatggc	ggcatactat	9540
cagtagtagg	tgtttccctt	tcttcttttag	cgacttgatg	ctcttgatct	tccaatacgc	9600
aacctaaagt	aaaatgcccc	acagcgtga	gtgcatataa	tgcattctct	agtgaaaaac	9660

cttgttggca taaaaaggct aattgatttt cgagagtttc atactgtttt tctgtaggcc	9720
gtgtacctaa atgtactttt gtcctatcgc gatgacttag taaagcacat ctaaaacttt	9780
tagcgttatt acgtaaaaaa tcttgccagc tttccccttc taaagggcaa aagtgagtat	9840
ggtgcctatc taacatctca atggctaagg cgtcgagcaa agcccgctta ttttttacct	9900
gccaatacaa tgtaggctgc tctacaccta gcttctgggc gagttttacgg gttgttaaac	9960
cttcgattcc gacctatta agcagctcta atgcgctgtt aatcacttta cttttatcta	10020
atctagacat cattaattcc taatttttgt tgacactcta tcattgatag agttatttta	10080
ccactcccta tcagtgatag agaaaagtga	10110

<210> 10

<211> 10108

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 10

gagctagata ataaggacgg cacaagatga acatcttcaa gaaaatggag gaatttaatt	60
atgaacaact ggtctacttc tacgacagcg aaacggaact caaagggtatt acctgtatac	120
acaacacaac tttagggccg gcattgggcg gtacccgcct ttggaactat aactctgagg	180
aagatgccgt tgaagacgta atccgtctgg ctcggggcat gacttacaaa gcggccttgcg	240
ccggtctgaa tctgggcggc ggtaaaaccg tgctgatcgg tgatgctaaa aagattaaat	300
cagagtccta cttccgtgga ctggggcgct acgttcagtc gctgaacggc agatatatca	360
ccgcggaaga cgtaaatact tctacgaagg atatggcata cgttgctatg gaaactgact	420
atgtggtagg cctgggaggt aaatccggca accctagtcc agttactgct tacggtgcat	480
ttatgggtat caaagcggcg ctgatgaaaa aatttgagga tagctctatt gaaggccgaa	540
ccttcgcagt gcagggtgct gggcagacgg gttactatct tatcgattac ctccaggca	600
acaacaagtt caaagaaaag gctaaaaaga ttactttcac cgaaattaac gagagctata	660
tcgagcgtat gaacaaagaa catccggaag ttgaatttat ttccccggac aaaatctact	720
cgctggaagt agacgtcttc gtgccctgcg ccctgggcaa aatcgттаат gacaaaacta	780

tcgatgaatt taagtgtccg atcatcgcag gtactgcaaa caacgtactg gaaaggggaag	840
cgcacggcaa catgcttaaa gaacgtggca ttcttttacgc cccggactat gtgatcaatg	900
ctgggtgggct gatcaacggt taccacgagc tgaacggtta caataaagag aacgctattc	960
tggaaagtga attaatttat gatcgcctac tggaaatatt caacatcgct gattctctga	1020
acatcagcac caatatcgct gccaacgagt tcgcggaaaa acgtatcaag caaattaagt	1080
ccttgaaaaa caacttcatt aaacgctaag accattaata aggaacattt tatgtcgacg	1140
acaaccgttg gtgactactt gctgtatcgc ttaaacgaaa tcggcattga gcacctcttc	1200
ggagtgccag gtgattacaa tctgcaattt ctggatcatg taatcgacca ccctcagctg	1260
acttgggtcg gctgcactaa cgaacttaac gctgcctacg cagctgatgg ttatgcgcgt	1320
tgtcgtccgg ctgcggcact gctgaccacc ttcggggttg gcgaactgag cgctattaat	1380
ggcatcgcag gttcctacgc ggagtatctg ccggtaatac atatcgttgg tgcaccgagt	1440
ctatcagccc agcagcaggg cgacctgatt caccactctc ttggcgaagg tgatttttcc	1500
agcttcctga ggatgtccca accggtgtct gttgcgcagg ctgctctgac tcctgataac	1560
gcatgcaagg aaatcgaccg cgtactggcg gaagtccctca ttcagcgtcg tcccggctac	1620
ctgctgctgt ctaccgacgt ggctgctgcg ccggcggctc tgccacaaag cactctttct	1680
ttgccgaccg ccccggatca tcgcgcagtt ctggctgctt tcagcgacgc tgctgagcag	1740
atgctggctc aggccaaaag cgtctctcta ctggcggact ttctggctga tcgtttcggg	1800
gttactcgag cactggccgc gtggcttcag caggttccgc taccgcacgc cactctgtta	1860
atgggtaaag gcgttctgag tgaacagcaa ccagggttcg tgggtaccta cgctgggtgcg	1920
gcatctatcg attcgacgcg tggcgcaatc gaagaagctg gggtaattat cggagtggga	1980
gtagatttt ccgacactat cacagcaggc ttctcgcagc agatcgacgc ccgccgtttt	2040
atagacattc aacccttctt ctctcgtatt ggcgatcgcc agtttgatca cttgccgatg	2100
caggctgccg tcgcagccct gcatcaactg tgtcttcgtt atcagcagca gtggtctatc	2160
accgctccta gcccgcctgc actgccgccg gctgctggta gcgagctgtc ccagaacgca	2220
ttctggcagg cgatgcagaa cttcatccgc cctggggacc tgttggtggc cgaccaaggt	2280
actgcggcgt tcggcgcagc ggcgctgcgc ttaccgcaga attgccagct gcttgtgcag	2340
ccgctgtggg gctcaatcgg ttacagtctg ccggccacct ttggtgctca gacggcagat	2400

acagagcgtc	gtgtaatcct	aatcattggc	gatggttcag	cgcaattaac	tattcaggaa	2460
ctttccagta	tgatgcgtga	cggcttgaaa	cctatcatct	ttctcctgaa	caacaacggg	2520
tacaccgttg	aacggggcgt	tcacggcgcg	gagcaacggt	ataacgatat	cgctgcttgg	2580
aattggaccc	aactgcccc	ggcgctgagt	gttcattgcc	cagcgagag	ctggcgagtc	2640
gttgaaacgg	tgacgctgac	cgacgtaatg	aaagtcacg	ctgcttctcc	gcgtctgagc	2700
ttggtagaag	ttgttctgcc	tgcaatggat	gtcccaccgc	tgctgcaagc	agtgagtgcc	2760
gctctgaacc	agcgcaactc	ctcttaatcc	ctccctacgg	gttaaggagc	ccctgctatg	2820
cgcagcatgc	agtttgatga	gtacggtgca	cccctgaaag	cgttctcata	tgaagacccg	2880
accccgcaag	ggaaggaagt	agtcgttagg	atcgaagcct	gtgggtgtgt	ccactctgat	2940
attcatcttc	acgagggcta	cttcgacatg	ggcgggtggc	ataaagctga	tgttactcgt	3000
gctcgcgaac	tcccttttac	attgggtcat	gaaatcgttg	gcgaagtggg	agcaactgga	3060
ccagggtgtc	ccggcgctaa	accgggcgac	aaacgtattg	tgtacccgtg	gatcgggtgc	3120
ggcgactgcc	cgaaatgcaa	cagtgggtgag	gatcagtcct	gtgcgcgtcc	acgtaacctg	3180
gggtgttcacg	ttgacgggtg	ctattcgacg	cacgtaaaga	taccggacga	aaaattcctg	3240
ttgcctacg	atgggtattcc	tactgagtta	gcgggaacct	atgcttgacg	cggcatcacc	3300
gcttatgggtg	cactgatgaa	agcaaaggaa	gcggctgaaa	gatctggcta	catcgggtctg	3360
attggcgctg	gtggcggttg	catggctggg	ctgatgctgg	ccaaagcagc	gatcggggct	3420
aaaactgtag	tctttgatat	cgacgacgca	aaactggaag	ctgcgacccg	tgccggggcg	3480
gattacgtgt	tcaactccgg	tgcaaaagaa	acacgcaagg	aagttatgaa	actaacgaat	3540
gggtggcctgt	ctgggtgctgt	tgatttcgtt	ggcagcgata	aaagcgctct	gtttggaatc	3600
aacgccttgg	gtcagaacgg	cgtgctggtc	ataattggac	tgttcggtgg	cgctatgact	3660
gttccgggtac	ccctgttccc	gctgaaaggg	atcaccgtac	gtggctcata	cgtaggttcc	3720
ctgcaagaga	tgagtgatat	gatggagtta	gttcgcgctg	ggaaagtcc	tccgatgccg	3780
gtaaaaactc	ggccactgga	cgctgcctgg	gaaacccttg	aggatctacg	ccatggtaaa	3840
atcgtgggcc	gtgttgttct	gaccccataa	gaaggagata	tacatatgac	ccatcaatta	3900
agatcgcgcg	atatcatcgc	tctgggcttt	atgacatttg	cgttgttcgt	cggcgaggt	3960
aacattattt	tccctccaat	ggtcggcttg	caggcaggcg	aacacgtctg	gactgcggca	4020

ttcggcttcc tcattactgc cgttggccta ccggtattaa cggtagtggc gctggcaaaa	4080
gttggcggcg gtgttgacag tctcagcacg ccaattggta aagtcgctgg cgtactgctg	4140
gcaacagttt gttacctggc ggtggggccg ctttttgcta cgccgcgtac agctaccgtt	4200
tcttttgaag tgggcattgc gccgctgacg ggtgattccg cgctgccgct gtttatttac	4260
agcctggctt atttcgctat cgttattctg gtttcgctct atccgggcaa gctgctggat	4320
accgtgggca acttccttgc gccgctgaaa attatcgcg tggtcatcct gtctgttgcc	4380
gcaattatct ggccggcggg ttctatcagt acggcgactg aggcttatca aaacgctgcg	4440
ttttctaacg gcttcgtcaa cggctatctg accatggata cgctgggcgc aatgggtgtt	4500
ggtatcgtta ttgttaacgc ggcgcgttct cgtggcgtta ccgaagcgcg tctgctgacc	4560
cgttataccg tctgggctgg cctgatggcg ggtgttggtc tgactctgct gtacctggcg	4620
ctgttccgtc tgggttcaga cagcgcgctg ctggctgata agtctgcaaa cggtgccgcg	4680
atcctgcatg cttacgttca gcatacctt ggccggcggcg gtagcttcct gctggcggcg	4740
ttaatcttca tcgcctgcct ggtcacggcg gttggcctga cctgtgcttg tgcagaattc	4800
ttcgcccagt acgtaccgct ctcttatcgt acgctgggtg ttatcctcgg cggcttctcg	4860
atgggtgggtg ctaacctcgg cttgagccag ctgattcaga tctctgtacc ggtgctgacc	4920
gccatttatc cgccgtgtat cgcactgggt gtattaagtt ttacacgctc atgggtggcat	4980
aattcgtccc gcgtgattgc tccgccgatg tttatcagcc tgcttttttg tattctcgac	5040
gggatcaagg catctgcatt cagcgatata ttaccgtcct gggcgcagcg ttaccgctg	5100
gccgaacaag gtctggcgtg gttaatgcca acagtgggtg tgggtggttct ggccattatc	5160
tgggatcgtg cggcaggtcg tcaggtgacc tccagcgctc actaatacgc atggcatgga	5220
tgaccgatgg tagtgtgggc tctcccatg cgagagtagg gaactgccag gcatcaaata	5280
aaacgaaagg ctcagtcgaa agactgggcc tttcgtttta tctgttggtt gtcggtgaac	5340
gctctcctga gtaggacaaa tccgccggga gcggatttga acgttgcgaa gcaacggccc	5400
ggagggtggc gggcaggacg cccgccataa actgccaggc atcaaattaa gcagaaggcc	5460
atcctgacgg atggcctttt tgcgtggcca gtgccaagct tgcatgcgtg ccagctgcat	5520
taatgaagaa atcatgctgg aagaataaca gctcactcaa aggcggtagt acgggttttg	5580
ctgcccgcaa acgggctgtt ctggtgttgc tagtttgta tcagaatcgc agatccggct	5640

tcagccgggtt tgccggctga aagcgctatt tcttccagaa ttgccatgat tttttcccca	5700
cgggaggcgt cactggctcc cgtgttgctg gcagctttga ttcgataagc agcatcgcct	5760
gtttcaggct gtctatgtgt gactgttgag ctgtaacaag ttgtctcagg tgttcaattt	5820
catgttctag ttgctttgtt ttactggttt cacctgttct attaggtgtt acatgctgtt	5880
catctgttac attgtcgatc tgttcatggg gaacagcttt gaatgcacca aaaactcgta	5940
aaagctctga tgtatctatc ttttttacac cgttttcatc tgtgcatatg gacagttttc	6000
cctttgatat gtaacgggta acagttgttc tacttttgtt tgttagtctt gatgcttcac	6060
tgatagatac aagagccata agaacctcag atccttccgt atttagccag tatgttctct	6120
agtgtgggtc gttgtttttg cgtgagccat gagaacgaac cattgagatc atacttactt	6180
tgcatgtcac tcaaaaattt tgcctcaaaa ctggtgagct gaatttttgc agttaaaagca	6240
tcgtgtagtg tttttcttag tccgttatgt aggttaggaat ctgatgtaat ggttgttggg	6300
attttgtcac cattcatttt tatctgggtg ttctcaagtt cggttacgag atccatttgt	6360
ctatctagtt caacttgga aatcaacgta tcagtcgggc ggcttcgctt atcaaccacc	6420
aatttcatat tgctgtaagt gtttaaactt ttacttattg gtttcaaaac ccattgggta	6480
agccttttaa actcatggta gttattttca agcattaaca tgaacttaaa ttcatcaagg	6540
ctaactctta tatttgcctt gtgagttttc ttttgtgtta gttcttttaa taaccactca	6600
taaatcctca tagagtattt gttttcaaaa gacttaacat gttccagatt atattttatg	6660
aattttttta actggaaaag ataaggcaat atctcttcac taaaaactaa ttctaatttt	6720
tcgcttgaga acttggcata gtttgtccac tggaaaatct caaagccttt aaccaaagga	6780
ttcctgattt ccacagttct cgtcatcagc tctctgggtg ctttagctaa tacaccataa	6840
gcattttccc tactgatgtt catcatctga gcgtattggg tataagtga cgataccgtc	6900
cgttctttcc ttgtaggggt ttcaatcgtg gggttgagta gtgccacaca gcataaaatt	6960
agcttggttt catgctccgt taagtcatag cgactaatcg ctagttcatt tgctttgaaa	7020
acaactaatt cagacataca tctcaattgg tctaggtgat tttaatcact ataccaattg	7080
agatgggcta gtcaatgata attactagtc cttttccttt gagttgtggg tatctgtaaa	7140
ttctgctaga ctttgctgg aaaacttgta aattctgcta gaccctctgt aaattccgct	7200
agacctttgt gtgttttttt tgtttatatt caagtgggta taatttatag aataaagaaa	7260

gaataaaaaa agataaaaag aatagatccc agccctgtgt ataactcact acttttagtca	7320
gttccgcagt attacaaaag gatgtcgcaa acgctgtttg ctcctctaca aaacagacct	7380
taaaacccta aaggcttaag tagcacctc gcaagctcgg gcaaatacgct gaatattcct	7440
tttgtctccg accatcaggc acctgagtcg ctgtcttttt cgtgacattc agttcgctgc	7500
gctcacggct ctggcagtga atgggggtaa atggcactac aggcgccttt tatggattca	7560
tgcaaggaaa ctaccataa tacaagaaaa gcccgtcacg ggcttctcag ggcgttttat	7620
ggcgggtctg ctatgtggtg ctatctgact ttttgctggt cagcagttcc tgccctctga	7680
ttttccagtc tgaccacttc ggattatccc gtgacaggtc attcagactg gctaatagcac	7740
ccagtaaggc agcggatatca tcaacaggct taccgctctt actgtctttt ctacggggtc	7800
tgacgctcag tggaacgaaa actcacgtta agggaatttg gtcatagagat tatcaaaaag	7860
gatcttcacc tagatccttt taaattaaaa atgaagtttt aaatcaatct aaagtatata	7920
tgagtaaaact tggctctgaca gttaccaatg cttaatcagt gaggcaccta tctcagcgat	7980
ctgtctattt cgttcatcca tagttgcctg actccccgtc gtgtagataa ctacgatacg	8040
ggagggccta ccatctggcc ccagtgcctg aatgataccg cgagaaccac gtcaccggc	8100
tccagattta tcagcaataa accagccagc cggaagggcc gagcgcagaa gtggtcctgc	8160
aactttatcc gcctccatcc agtctattaa ttgttgccgg gaagctagag taagtagttc	8220
gccagttaat agtttgcgca acgttggtgc cattgctaca ggcatcgtgg tgtcacgctc	8280
gtcgtttggt atggcttcat tcagctccgg ttcccaacga tcaaggcgag ttacatgatc	8340
ccccatgttg tgcaaaaaag cggttagctc cttcggtcct ccgatcgttg tcagaagtaa	8400
gttggccgca gtgttatcac tcatggttat ggcagcactg cataattctc ttactgtcat	8460
gccatccgta agatgctttt ctgtgactgg tgagtactca accaagtcac tctgagaata	8520
gtgtatgcgg cgaccgagtt gctcttgccc ggcgtcaata cgggataata ccgcgccaca	8580
tagcagaact ttaaaagtgc tcatcattgg aaaacgttct tcggggcgaa aactctcaag	8640
gatcttaccg ctgttgagat ccagttcgat gtaaccact cgtgcacca actgatcttc	8700
agcatctttt actttcacca gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc	8760
aaaaaaggga ataaggcgga cacggaaatg ttgaatactc atactcttcc tttttcaata	8820
ttattgaagc atttatcagg gttattgtct catgagcgga tacatatattg aatgtattta	8880

gaaaaataaa caaatagggg ttccgcgcac atttccccga aaagtgccac ctgacgtcta	8940
agaaaccatt attatcatga cattaaccta taaaaatagg cgtatcacga ggccctttcg	9000
tctcgcgcgt ttcggtgatg acggtgaaaa cctctgacac atgcagctcc cggagacggt	9060
cacagcttgt ctgtaagcgg atgccgggag cagacaagcc cgtcagggcg cgtcagcggg	9120
tgttggcggg tgtcggggct ggcttaacta tgcggcatca gagcagattg tactgagagt	9180
gcaccatatg cgggtgtgaaa taccgcacag atgcgtaagg agaaaatacc gcatcaggcg	9240
ccattcgcca ttcaggctgc gcaactgttg ggaagggcga tcggtgcggg cctcttcgct	9300
attacgccag ctggcgaaag ggggatgtgc tgcaaggcga ttaagttggg taacgccagg	9360
gttttcccag tcacgacgtt gtaaaacgac ggccagtga ttcgttaaga cccactttca	9420
catttaagtt gtttttctaa tccgcatatg atcaattcaa ggccgaataa gaaggctggc	9480
tctgcacctt ggtgatcaaa taattcgata gcttgtcgta ataatggcgg catactatca	9540
gtagtaggtg tttccctttc ttcttttagcg acttgatgct cttgatcttc caatacgcaa	9600
cctaaagtaa aatgccccac agcgtgagt gcatataatg cattctctag tgaaaaacct	9660
tgttggcata aaaaggctaa ttgattttcg agagtttcat actgtttttc tgtaggccgt	9720
gtacctaaat gtacttttgc tccatcgcg tgacttagta aagcacatct aaaactttta	9780
gcgttattac gtaaaaaatc ttgccagctt tccccttcta aagggcaaaa gtgagtatgg	9840
tgcctatcta acatctcaat ggctaaggcg tcgagcaaag cccgcttatt tttacatgc	9900
caatacaatg taggctgctc tacacctagc ttctgggcga gtttacgggt tgttaaacct	9960
tcgattccga cctcattaag cagctcta at gcgctgttaa tcactttact tttatcta at	10020
ctagacatca ttaattccta atttttgttg acactctatc attgatagag ttattttacc	10080
actccctatc agtgatagag aaaagtga	10108

<210> 11

<211> 10157

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 11

actacagtat ataaggaggg taatatgatc ttcgagacaa tttctacgtc gaatcacgaa	60

gaagttgtgt attgccataa caaggacgcc ggcttgaaag caatcatcgc gattcacaac	120
actgtactcg gtccggctct gggtggcact cgcattgtggc cctacgctag cgaagaggaa	180
gcactgaaag atgtccttcg tttatcccgt gggatgacct acaaagctgc ggtttcaggt	240
ctaaacctgg gcggcggtaa agctgtgatc tggggtgatc cgaataaaga caagtctgaa	300
gcgctgttta gagccttcgg acggtttgta aacagcctgg gcggacgcta cattaccgcg	360
gaggacgttg gcattgatgt taacgacatg gaatatgtgc tgcgtgaaac tgattacgtc	420
accggtgtac atcaggttca cgggtgggagt ggtgatcctt ctccattcac cgcatatggc	480
actctgcaag gcctgatggc cgctctgcaa gtgaaattcg gtaacgaaga cgtaggcaat	540
tacagctacg ctgttcaggg tgtgggtcac gttggcatgg aatttgtaa actgctgcgt	600
gagcgcggtg caaaggtttt cgtcactgac atcaacaaag atgcgggtcca gcgtgctgtg	660
gacgaatttg gttgtgaggc agtagccctg gatgaaatct atgacgttga ttgcgacgtg	720
tactccccga ccgctctggg cggcaccgtg aacgataaaa ctttaccgcg tctgaaatgt	780
aaggtaatct gcggtgcggc aaacaaccag ttagctaata atgagatagg cgtggaactg	840
gaaaagaaag gcatacctcta tgctccggac tacgcggtca acgcgggtgg gctgatgaac	900
gtagccttg aaatcgatgg atacaaccgc gaacgtgcga tgcgtatgat gcgtaccatt	960
tattacaatt tgggtcgcatt tttcgaaatc tctaagcgcg acggcatccc tacattccga	1020
gccgccgatc gtatggctga agaacgcata acggccatcg gtaaaactgcg tttaccgcat	1080
ttgggcgctg cggcaccgcg cttccagggc cgacgtggca actaattacg acattttaag	1140
gaggttcgca tgtcgacgac aaccgttggt gactacttgc tgtatcgctt aaacgaaatc	1200
ggcattgagc acctcttcgg agtgccaggt gattacaatc tgcaatttct ggatcatgta	1260
atcgaccacc ctacgctgac ttgggtcggc tgcactaacg aacttaacgc tgcctacgca	1320
gctgatgggt atgcgcgttg tcgtccggct gcggcactgc tgaccacctt cggggttggc	1380
gaactgagcg ctattaatgg catcgcaggt tcctacgcgg agtatctgcc ggtaatacat	1440
atcgttgggt caccgagtct atcagcccag cagcagggcg acctgattca ccactctctt	1500
ggcgaagggt atttttccag cttcctgagg atgtcccaac cgggtgtctgt tgcgcaggct	1560
gctctgactc ctgataacgc atgcaaggaa atcgaccgcg tactggcgga agtcctcatt	1620
cagcgtcgtc ccggctacct gctgctgtct accgacgtgg ctgctgcgcc ggcggtctctg	1680

ccacaaagca	ctctttcttt	gccgaccgcc	ccggatcatc	gcgcagttct	ggctgctttc	1740
agcgacgctg	ctgagcagat	gctggctcag	gccaaaagcg	tctctctact	ggcggacttt	1800
ctggctgata	gtttcgggtg	tactcgagca	ctggccgcgt	ggcttcagca	ggttccgcta	1860
ccgcacgcca	ctctgttaat	gggtaaaggc	gttctgagtg	aacagcaacc	agggttcgtg	1920
ggtacctacg	ctggtgcggc	atctatcgat	tcgacgcgtg	gcgcaatcga	agaagctggg	1980
gtaattatcg	gagtgggagt	tagattttcc	gacactatca	cagcaggctt	ctcgagcag	2040
atcgacgccc	gccgttttat	agacattcaa	cccttcttct	ctcgtattgg	cgatcgccag	2100
tttgatcact	tgccgatgca	ggctgccgtc	gcagccctgc	atcaactgtg	tcttcgttat	2160
cagcagcagt	ggtctatcac	cgctcctagc	ccgcctgcac	tgccgccggc	tgctggtagc	2220
gagctgtccc	agaacgcatt	ctggcaggcg	atgcagaact	tcatccgccc	tggggacctg	2280
ttggtggccg	accaaggtac	tgccggcgttc	ggcgcagcgg	cgctgcgctt	accgcagaat	2340
tgccagctgc	ttgtgcagcc	gctgtggggc	tcaatcgggt	acagtctgcc	ggccaccttt	2400
ggtgctcaga	cggcagatac	agagcgtcgt	gtaatcctaa	tcattggcga	tggttcagcg	2460
caattaacta	ttcaggaact	ttccagtatg	atgcgtgacg	gcttgaaacc	tatcatcttt	2520
ctcctgaaca	acaacggtta	caccgttgaa	cgggcgattc	acggcgcgga	gcaacgttat	2580
aacgatatac	ctgcttgga	ttggacccaa	ctgccccagg	cgctgagtg	tcattgccc	2640
gcgagagct	ggcgagtcgt	tgaaacgggt	cagctgaccg	acgtaatgaa	agtcatcgct	2700
gcttctccgc	gtctgagctt	ggtagaagtt	gttctgcctg	caatggatgt	cccaccgctg	2760
ctgcaagcag	tgagtccgc	tctgaaccag	cgcaactcct	cttaaatttt	tcaacggcct	2820
atcataagga	ggaataatga	cagcggagca	gcaaatggg	gtatccgact	cacgccgttt	2880
cgaatttcag	gaatttggtg	gccctatcgc	cccacagacc	tatcagctcc	ccgcaccggc	2940
tagcgatgaa	gttttggtta	aggtgaacta	ctgcggtgtc	tgacacagtg	atgttcatct	3000
tcacgacggc	tacttcgagc	tgggtggcga	taaacgtctg	aacttcgcta	tgccgctgcc	3060
gctgacgctg	ggtcacgaag	taattggcac	cgttgtggct	gtcggcgacc	aggttactgg	3120
tgtaaaaccg	ggggaccagc	gactgatcta	tccgtggata	ggttgcggaa	aatgcggcgc	3180
gtgtcaaaaa	ggagaagaaa	acctgtgcgt	tactcctgca	catctgggcg	tgaacaagcc	3240
gggcgggttac	gctgatcaca	tcgttgtacc	ccattctcgc	taccttctgg	acatttcggg	3300

tctgaacccg ggtgatgccg ctaccctcgc gtgctccggc ctgaccactt tcagcgcgat	3360
caacaaagtg ttgccgcttg cagatgacca gtggattggt gttatcgggt gtggtggcct	3420
cggccagatg gcgctgcgta tcctgcaagc tatgggaatt ggcaatgtta tcggtattga	3480
cctgtctgaa gagaaacgga aactggctca tgaaagcggg gcacgtcact ctttcgatcc	3540
aaacactccg aagctgaacc gcgtggctgc cgaaacctgc ccgggtacgg tacaggccgc	3600
gtagactttt gtgggcaatg agcaaactgc tcagctggca ctgtctctgc ttggaaaagg	3660
tggcaaatat gttcctgtcg ggctgcacgg cggcgagctg cgttacccat tgccgatcat	3720
cacgaacaaa gctgtaagta tcatcggttc ttacgttggt accctgaaag aactggaaga	3780
cttagttgct ttcgccaagg aaaagaatct gccgccaatt catattgaac accgcccgt	3840
ggaatcggcg gctcaggccg tagaggacct ggaaaaagga caggttgctg ggcgtgttat	3900
cctggatgca ggtaactaag aaggagatat acatatgacc catcaattaa gatcgcgcga	3960
tatcatcgct ctgggcttta tgacatttgc gttgttcgtc ggcgaggta acattatttt	4020
ccctccaatg gtcggcttgc aggcaggcga acacgtctgg actgcggcat tcggcttcct	4080
cattactgcc gttggcctac cggtattaac ggtagtggcg ctggcaaaag ttggcggcgg	4140
tgttgacagt ctcagcacgc caattggtaa agtcgctggc gtactgctgg caacagtttg	4200
ttacctggcg gtggggccgc tttttgctac gccgcgtaca gctaccgttt cttttgaagt	4260
gggcattgcg ccgctgacgg gtgattccgc gctgccgctg tttatttaca gcctgggtcta	4320
tttcgctatc gttattctgg tttcgtctta tccgggcaag ctgctggata ccgtgggcaa	4380
cttccttgcg ccgctgaaaa ttatcgcgct ggtcatcctg tctgttgccg caattatctg	4440
gccggcgggt tctatcagta cggcgactga ggcttatcaa aacgctgcgt tttctaacgg	4500
cttcgtcaac ggctatctga ccatggatac gctgggcgca atggtgtttg gtatcgttat	4560
tgtaaacgcg gcgcgttctc gtggcgttac cgaagcgcgt ctgctgaccc gttataccgt	4620
ctgggctggc ctgatggcgg gtgttggtct gactctgctg tacctggcgc tgttccgtct	4680
gggttcagac agcgcgtcgc tggtcgatca gtctgcaaac ggtgcggcga tcctgcatgc	4740
ttacgttcag catacctttg gcggcggcgg tagcttcctg ctggcggcgt taatcttcat	4800
cgcctgcctg gtcacggcgg ttggcctgac ctgtgcttgt gcagaattct tcgccagta	4860
cgtaccgctc tcttatcgta cgctggtgtt tatcctcggc ggcttctcga tggtggtgtc	4920

taacctcggc ttgagccagc tgattcagat ctctgtaccg gtgctgaccg ccatttatcc	4980
gccgtgtatc gcaactggttg tattaagttt tacacgctca tgggtggcata attcgtccccg	5040
cgtgattgct ccgccgatgt ttatcagcct gcttttttggg attctcgcgc ggatcaaggc	5100
atctgcattc agcgatatct taccgtcctg ggcgagcgt ttaccgctgg ccgaacaagg	5160
tctggcgtgg ttaatgccaa cagtgggtgat ggtggttctg gccattatct gggatcgtgc	5220
ggcaggtcgt caggtgacct ccagcgctca ctaatacgca tggcatggat gaccgatggt	5280
agtgtgggct ctccccatgc gagagtaggg aactgccagg catcaaataa aacgaaaggc	5340
tcagtcgaaa gactgggcct ttcgtttttat ctgttgtttg tcggtgaacg ctctcctgag	5400
taggacaaat ccgccgggag cggatttgaa cgttgcgaag caacggccccg gaggggtggcg	5460
ggcaggacgc ccgccataaa ctgccaggca tcaaattaag cagaaggcca tcctgacgga	5520
tggccttttt gcgtggccag tgccaagctt gcatgcgtgc cagctgcatt aatgaagaaa	5580
tcatgctgga agaataacag ctcaactcaa ggcggttagta cgggttttgc tgcccgcaaa	5640
cgggctgttc tgggtgttgct agtttgttat cagaatcgca gatccggctt cagccggttt	5700
gccggctgaa agcgctatct cttccagaat tgccatgatt ttttccccac gggaggcgtc	5760
actggctccc gtgttgctcg cagctttgat tcgataagca gcatgcctg tttcaggctg	5820
tctatgtgtg actgttgagc tgtaacaagt tgtctcaggt gttcaatttc atgttctagt	5880
tgctttgttt tactggtttc acctgttcta ttaggtgtta catgctgttc atctgttaca	5940
ttgtcgatct gttcatggtg aacagctttg aatgcaccaa aaactcgtaa aagctctgat	6000
gtatctatct tttttacacc gttttcatct gtgcatatgg acagttttcc ctttgatatg	6060
taacggtgaa cagttgttct acttttgttt gttagtcttg atgcttcact gatagataca	6120
agagccataa gaacctcaga tccttccgta ttagccagt atgttctcta gtgtggttcg	6180
ttgtttttgc gtgagccatg agaacgaacc attgagatca tacttacttt gcatgtcact	6240
caaaaatttt gcctcaaaac tgggtgagctg aatttttgca gttaaagcat cgtgtagtgt	6300
ttttcttagt ccgttatgta ggtaggaatc tgatgtaatg gttgttggtta tttgtcacc	6360
attcattttt atctggttgt tctcaagttc ggttacgaga tccatttgtc tatctagttc	6420
aacttgaaa atcaacgtat cagtcgggcg gcctcgctta tcaaccacca atttcatatt	6480
gctgtaagtg tttaaactct tacttattgg tttcaaaacc cattgggtaa gccttttaaa	6540

ctcatggtag ttatitttcaa gcattaacat gaacttaaat tcatcaaggc taatctctat	6600
atttgccttg tgagttttct tttgtgttag ttcttttaaat aaccactcat aaatcctcat	6660
agagtatttg ttttcaaaag acttaacatg ttccagatta tatttttatga atttttttaa	6720
ctggaaaaga taaggcaata tctcttcact aaaaactaat tctaattttt cgcttgagaa	6780
cttggcatag tttgtccact ggaaaatctc aaagccttta accaaaggat tcctgatttc	6840
cacagtcttc gtcacagct ctctggttgc tttagctaata acaccataag cattttccct	6900
actgatgttc atcatctgag cgtattgggtt ataagtgaac gataccgtcc gttctttcct	6960
tgtagggttt tcaatcgtgg ggttgagtag tgccacacag cataaaatta gcttggtttc	7020
atgctccgtt aagtcatagc gactaatcgc tagttcattt gctttgaaaa caactaattc	7080
agacatacat ctcaattgggt ctaggtgatt ttaatcacta taccaattga gatgggctag	7140
tcaatgataa ttactagtcc ttttcctttg agttgtgggt atctgtaaat tctgctagac	7200
ctttgctgga aaacttgtaa attctgctag accctctgta aattccgcta gacctttgtg	7260
tgtttttttt gtttatattc aagtggttat aatttataga ataaagaaag aataaaaaaa	7320
gataaaaaga atagatccca gccctgtgta taactcacta ctttagtcag ttccgcagta	7380
ttacaaaagg atgtcgcaaa cgctgtttgc tcctctacaa aacagacctt aaaaccctaa	7440
aggcttaagt agcaccctcg caagctcggg caaatcgctg aatattcctt ttgtctccga	7500
ccatcaggca cctgagtcgc tgtctttttc gtgacattca gttcgctgcg ctcacggctc	7560
tggcagtga tgggggtaaa tggcactaca ggcgcctttt atggattcat gcaaggaaac	7620
taccataat acaagaaaag cccgtcacgg gcttctcagg gcgttttatg gcgggtctgc	7680
tatgtggtgc tatctgactt tttgctgttc agcagttcct gccctctgat tttccagtct	7740
gaccacttcg gattatcccc tgacagggtca ttcagactgg ctaatgcacc cagtaaggca	7800
gcggtatcat caacaggctt acccgtctta ctgtcttttc tacgggggtct gacgctcagt	7860
ggaacgaaaa ctcacgttaa gggatttttg tcatgagatt atcaaaaagg atcttcacct	7920
agatcctttt aaattaaaaa tgaagtttta aatcaatcta aagtatatat gagtaaactt	7980
ggtctgacag ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc tgtctatttc	8040
gttcatccat agttgcctga ctcccgtcg tgtagataac tacgatacgg gagggcttac	8100
catctggccc cagtgtctga atgataccgc gagaaccacg ctcaccggct ccagatttat	8160

cagcaataaa ccagccagcc ggaagggccg agcgcagaag tggtcctgca actttatccg	8220
cctccatcca gtctattaat tgttgccggg aagctagagt aagtagttcg ccagttaata	8280
gtttgcgcaa cgttgttgcc attgctacag gcatcgtggg gtcacgctcg tcgtttggta	8340
tggcttcatt cagctccggg tcccaacgat caaggcgagt tacatgatcc cccatgttgt	8400
gcaaaaaagc ggtagctcc ttcggctctc cgatcgttgt cagaagtaag ttggccgcag	8460
tgttatcact catggttatg gcagcactgc ataattctct tactgtcatg ccatccgtaa	8520
gatgcttttc tgtgactggg gagtactcaa ccaagtcatt ctgagaatag tgtatgcggc	8580
gaccgagttg ctcttgcccg gcgtcaatac gggataatac cgcgccacat agcagaactt	8640
taaaagtgct catcattgga aaacgttctt cggggcgaaa actctcaagg atcttaccgc	8700
tgttgagatc cagttcgatg taaccactc gtgcacccaa ctgatcttca gcatctttta	8760
ctttcaccag cgtttctggg tgagcaaaaa caggaaggca aaatgccgca aaaaagggaa	8820
taagggcgac acggaaatgt tgaatactca tactcttcct ttttcaatat tattgaagca	8880
tttatcaggg ttattgtctc atgagcggat acatatttga atgtatttag aaaaataaac	8940
aaataggggt tccgcgcaca tttccccgaa aagtgccacc tgacgtctaa gaaaccatta	9000
ttatcatgac attaacctat aaaaataggc gtatcacgag gccctttcgt ctcgcgcggt	9060
tcggtgatga cggtgaaaac ctctgacaca tgcagctccc ggagacggtc acagcttgct	9120
tgtaagcggg tgccgggagc agacaagccc gtcagggcgc gtcagcgggt gttggcgggt	9180
gtcggggctg gcttaactat gcggcatcag agcagattgt actgagagtg caccatatgc	9240
ggtgtgaaat accgcacaga tgcgtaagga gaaaataccg catcaggcgc cattcgccat	9300
tcaggctgcg caactgttgg gaagggcgat cgggtgcgggc ctcttcgcta ttacgccagc	9360
tggcgaaagg gggatgtgct gcaaggcgat taagttgggt aacgccaggg ttttcccagt	9420
cacgacgttg taaaacgacg gccagtgaat tcgttaagac ccactttcac atttaagttg	9480
tttttctaata cgcatatga tcaattcaag gccgaataag aaggctggct ctgcaccttg	9540
gtgatcaaat aattcgatag cttgtcgtaa taatggcggc atactatcag tagtaggtgt	9600
ttccctttct tctttagcga cttgatgctc ttgatcttcc aatacgcaac ctaaagtaaa	9660
atgccccaca gcgctgagtg catataatgc attctctagt gaaaaacctt gttggcataa	9720
aaaggctaata tgattttcga gagtttcata ctgtttttct gtaggccgtg tacctaaatg	9780

tacttttgct ccacgcgat gacttagtaa agcacatcta aaacttttag cgttattacg	9840
taaaaaatct tgccagcttt ccccttctaa agggcaaaag tgagtatggt gcctatctaa	9900
catctcaatg gctaaggcgt cgagcaaagc ccgcttattt ttacatgcc aatacaatgt	9960
aggctgctct acacctagct tctgggcgag ttacgggtt gttaaaccctt cgattccgac	10020
ctcattaagc agctctaagc cgctgttaat cactttactt ttatctaatac tagacatcat	10080
taattcctaa tttttgttga cactctatca ttgatagagt tattttacca ctccctatca	10140
gtgatagaga aaagtga	10157

<210> 12

<211> 10156

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 12

agtttacccg cgggattgga ggtacacatg cagatcttcg acactttgca atcaatgggc	60
catgagcagg ttgtcctatg tagcgataag accacgggtc tgcgcgccat tatcgctata	120
cacgatacat ccttagggcc ggcgcttggt ggtaccgta tgtggcagta tgcaactgac	180
gacgatgcta ttactgacgc actccgtctg tctcggggca tgacctacaa agctgcgggt	240
tctggcgtaa atctgggcgg tggtaaagcc gttatcatcg gaaaccctca cagtgataaa	300
agcgaagcgc tgtttcgcgc ttacggcaga atgggtggaat ccagcgtgg gcgttacatc	360
accgccgaag acgttggtac tagcgtacgt gatatggagt ggattcgcac ggaaaccaa	420
tatgtaacgg gcgtgggtgg caacggaggc tctggtgacc cctctccagt taccgctctg	480
ggtgtttact cgggcatgaa ggcatgcgct aaatcagtct atggtactga tgcgctgagc	540
ggtaaaagga tcgtggttca gggcgcggtt aacgttgcat cccatctggt tcacagtctg	600
gtaaaagaag gcgctgtggt tttcgtcact gacatctacg aagaaaaggc caaagcatta	660
gctggctgaaa cgggcgctac cgtgattcgc accgacgagg tttttactac acaatgcgat	720
atcttctctc cgaacgtctt gggggccgct ctgaacgatg aaactattcc gcagctcaca	780
tgcgctatcg tagctggtgg tgcaacaat cagcttaaaa tcgaacaacg tcacgccacg	840

gctctgcaag agaaaggcat tctgtatgcg ccggattacg taatcaacgc cgggggcctc	900
atgaatgtgg cgagcgaagt tgacggctac aaccgtgaaa aggttatgcg ccaggctgaa	960
ggtatttacg atattactat gaacatccta aataccgcgc gtgagcgtaa catcctgacc	1020
atcgaagcat ccaacgcgat tgctgaagag cggatcaaca aagttcgcca tgttcacggg	1080
aacttcacg gttccccgtc tattcgcgga gtataataaa tcaaaacacc aaattaagga	1140
gaaattttat gtcgacgaca accgttggtg actacttgct gtatcgctta aacgaaatcg	1200
gcattgagca cctcttcgga gtgccaggtg attacaatct gcaatttctg gatcatgtaa	1260
tcgaccacc tcagctgact tgggtcggct gcactaacga acttaacgct gcctacgcag	1320
ctgatggtta tgcgcgttgt cgtccggctg cggcactgct gaccaccttc ggggttggcg	1380
aactgagcgc tattaatggc atcgcaggtt cctacgcgga gtatctgccg gtaatacata	1440
tcgttggtgc accgagtcta tcagcccagc agcagggcga cctgattcac cactctcttg	1500
gcgaaggtga tttttccagc ttcctgagga tgtcccaacc ggtgtctgtt gcgcaggctg	1560
ctctgactcc tgataacgca tgcaaggaaa tcgaccgcgt actggcggaa gtcctcattc	1620
agcgtcgtcc cggctacctg ctgctgtcta ccgacgtggc tgctgcgccg gcggctctgc	1680
cacaaagcac tctttctttg ccgaccgcc cggatcatcg cgcagttctg gctgctttca	1740
gcgacgctgc tgagcagatg ctggctcagg ccaaaagcgt ctctctactg gcggactttc	1800
tggctgatcg tttcgggtgt actcgagcac tggccgcgtg gcttcagcag gttccgctac	1860
cgcacgccac tctgttaatg ggtaaaggcg ttctgagtga acagcaacca gggttcgtgg	1920
gtacctacgc tggcgcggca tctatcgatt cgacgcgtgg cgcaatcgaa gaagctgggg	1980
taattatcgg agtgggagtt agattttccg aactatcac agcaggcttc tcgcagcaga	2040
tcgacgcccg ccgttttata gacattcaac ctttcttctc tcgtattggc gatcgccagt	2100
ttgatcactt gccgatgcag gctgccgtcg cagccctgca tcaactgtgt cttcgttatc	2160
agcagcagtg gtctatcacc gctcctagcc cgcctgcact gccgccggct gctggtagcg	2220
agctgtccca gaacgcattc tggcaggcga tgcagaactt catccgccct ggggacctgt	2280
tggcggccga ccaaggctact gcggcgttcg gcgcagcggc gctgcgctta ccgcagaatt	2340
gccagctgct tgtgcagccg ctgtggggct caatcggtta cagtctgccg gccacctttg	2400
gtgctcagac ggcagatata gagcgtcgtg taatcctaatt cattggcgat ggttcagcgc	2460

aattaactat tcaggaactt tccagtatga tgcgtgacgg cttgaaacct atcatctttc	2520
tcctgaacaa caacggttac accgttgaac gggcgattca cggcgcgag caacggtata	2580
acgatatcgc tgcttggaaat tggaccaaac tgccccaggc gctgagtgtt cattgcccag	2640
cgcagagctg gcgagtcgtt gaaacggtgc agctgaccga cgtaatgaaa gtcacgctg	2700
cttctccgcg tctgagcttg gtagaagttg ttctgcctgc aatggatgtc ccaccgctgc	2760
tgcaagcagt gagtgccgct ctgaaccagc gcaactcctc ttaaattttt caacggccta	2820
tcataaggag gaataatgac agcggagcag caaaatgggg tatccgactc acgccgtttc	2880
gaatttcagg aatttggtgg ccctatcgcc ccacagacct atcagctccc cgcaccggct	2940
agcgatgaag ttttgttaaa ggtgaactac tgcggtgtct gtcacagtga tgttcatctt	3000
cacgacggct acttcgagct ggggtggcgat aaacgtctga acttcgctat gccgctgccg	3060
ctgacgctgg gtcacgaagt aattggcacc gttgtggctg tcggcgacca ggttactggt	3120
gtaaaaccgg gggaccagcg actgatctat ccgtggatag gttgcggaaa atgcggcgcg	3180
tgtcaaaaag gagaagaaaa cctgtgcgtt actcctgcac atctgggcgt gaacaagccg	3240
ggcggttacg ctgatcacat cgttgtagcc cattctcgct accttctgga catttcgggt	3300
ctgaaccgg gtgatgccgc taccctcgcg tgctccggcc tgaccacttt cagcgcgatc	3360
aacaaagtgt tgccgcttgc agatgaccag tggattgttg ttatcggttg tgggtggcctc	3420
ggccagatgg cgctgcgtat cctgcaagct atgggaattg gcaatgttat cggatttgac	3480
ctgtctgaag agaaacggaa actggctcat gaaagcgggt cacgtcactc cttcgatcca	3540
aacactccga agctgaaccg cgtggctgcc gaaacctgcc cgggtacgggt acaggccgcg	3600
ttagactttg tgggcaatga gcaaactgct cagctggcac tgtctctgct tggaaaaggt	3660
ggcaaatatg ttctgtcgg gctgcacggc ggcgagctgc gttaccatt gccgatcatc	3720
acgaacaaag ctgtaagtat catcggttct tacgttggta ccctgaaaga actggaagac	3780
ttagttgctt tcgccaagga aaagaatctg ccgccaattc atattgaaca ccgcccgtg	3840
gaatcggcgg ctcaggccgt agaggacctg gaaaaaggac aggttgctgg gctgtttatc	3900
ctggatgcag gtaactaaga aggagatata catatgacct atcaattaag atcgcgcgat	3960
atcatcgctc tgggctttat gacatttgcg ttgttcgtcg gcgcaggtaa cattattttc	4020
cctccaatgg tcggcttgca ggcaggcgaa cacgtctgga ctgcggcatt cggcttcctc	4080

attactgccg ttggcctacc ggtattaacg gtagtggcgc tggcaaaagt tggcggcggg	4140
gttgacagtc tcagcacgcc aattggtaaa gtcgctggcg tactgctggc aacagtttgt	4200
tacctggcgg tggggccgct ttttgctacg ccgctacag ctaccgtttc ttttgaagtg	4260
ggcattgcgc cgctgacggg tgattccgcg ctgccgtgt ttatttacag cctgggtctat	4320
ttcgctatcg ttattctggg ttcgctctat ccgggcaagc tgctggatac cgtgggcaac	4380
ttccttgccg cgctgaaaat tatcgcgctg gtcacacctgt ctgttgccgc aattatctgg	4440
ccggcggggt ctatcagtac ggcgactgag gcttatcaaa acgctgcgtt ttctaacggc	4500
ttcgtcaacg gctatctgac catggatacg ctgggcgcaa tgggtgtttgg tatcgttatt	4560
gttaacgcgg cgcgttctcg tggcgttacc gaagcgcgtc tgctgacccg ttataccgtc	4620
tgggctggcc tgatggcggg tgttggtctg actctgctgt acctggcgct gttccgtctg	4680
ggttcagaca gcgcgtcgct ggtcgatcag tctgcaaacg gtgcggcgat cctgcatgct	4740
tacgttcagc atacctttgg cggcggcggg agcttcctgc tggcggcgtt aatcttcac	4800
gcctgcctgg tcacggcggg tggcctgacc tgtgcttggt cagaattctt cgcccagtac	4860
gtaccgctct cttatcgta gctgggtgtt atcctcggcg gcttctcgat ggtggtgtct	4920
aacctcggct tgagccagct gattcagatc tctgtaccgg tgctgaccgc catttatccg	4980
ccgtgtatcg cactggttgt attaaagttt acacgctcat ggtggcataa ttcgtcccgc	5040
gtgattgctc cgccgatgtt tatcagcctg ctttttggtt ttctcgacgg gatcaaggca	5100
tctgcattca gcgatatctt accgtcctgg gcgcagcggt taccgctggc cgaacaagg	5160
ctggcgtggg taatgccaac agtggtgatg gtggttctgg ccattatctg ggatcgtgcg	5220
gcaggtcgtc aggtgacctc cagcgtcac taatacgcat ggcatggatg accgatggta	5280
gtgtgggctc tccccatgcg agagtaggga actgccaggc atcaaataaa acgaaaggct	5340
cagtcgaaag actgggcctt tcgttttatc tgttgtttgt cggatgaacgc tctcctgagt	5400
aggacaaatc cgccgggagc ggatttgaac gttgcgaagc aacggcccgg aggggtggcgg	5460
gcaggacgcc cgccataaac tgccaggcat caaattaagc agaaggccat cctgacggat	5520
ggcctttttg cgtggccagt gccaagcttg catgcgtgcc agctgcatta atgaagaaat	5580
catgctggaa gaataacagc tcactcaaag gcggtagtac gggttttgct gcccgcaaac	5640
gggctgttct ggtgttgcta gtttgttatc agaatcgag atccggcttc agccggtttg	5700

ccggctgaaa gcgctatttc ttccagaatt gccatgattt tttccccacg ggaggcgtca	5760
ctggctcccc tggtgtcggc agctttgatt cgataagcag catcgcctgt ttcaggctgt	5820
ctatgtgtga ctgttgagct gtaacaagtt gtctcagggtg ttcaatttca tgttctagtt	5880
gctttgtttt actggtttca cctgttctat taggtgttac atgctgttca tctgttacat	5940
tgtcgatctg ttcattggtga acagctttga atgcacaaaa aactcgtaaa agctctgatg	6000
tatctatctt ttttacaccg ttttcatctg tgcataatgga cagttttccc tttgatatgt	6060
aacgggtgaac agttgttcta cttttgtttg ttagtcttga tgcttcactg atagatacaa	6120
gagccataag aacctcagat ccttccgtat ttagccagta tgttctctag tgtggttcgt	6180
tgtttttgcg tgagccatga gaacgaacca ttgagatcat acttactttg catgtcactc	6240
aaaaattttg cctcaaaact ggtgagctga atttttgcag ttaaagcatc gtgtagtggt	6300
tttcttagtc cgttatgtag gtaggaatct gatgtaatgg ttgttggtat tttgtcacca	6360
ttcattttta tctggttggt ctcaagttcg gttacgagat ccatttgtct atctagttca	6420
acttgaaaa tcaacgtatc agtcgggcgg cctcgcttat caaccaccaa tttcatattg	6480
ctgtaagtgt ttaaactctt acttattggt ttcaaaaccc attgggtaag ctttttaaac	6540
tcatggtagt tattttcaag cattaacatg aacttaaatt catcaaggct aatctctata	6600
tttgccttgt gagttttctt ttgtgttagt tcttttaata accactcata aatcctcata	6660
gagtatttgt tttcaaaaga cttaacatgt tccagattat attttatgaa tttttttaac	6720
tggaaaagat aaggcaatat ctcttcacta aaaactaatt ctaatttttc gcttgagaac	6780
ttggcatagt ttgtccactg gaaaatctca aagcctttaa ccaaaggatt cctgatttcc	6840
acagttctcg tcatcagctc tctggttgct ttagctaata caccataagc attttcccta	6900
ctgatgttca tcatctgagc gtattgggtta taagtgaacg ataccgtccg ttctttcctt	6960
gtagggtttt caatcgtggg gttgagtagt gccacacagc ataaaattag cttggtttca	7020
tgctccgtta agtcatagcg actaatcgct agttcatttg ctttgaaaac aactaattca	7080
gacatacatc tcaattgggtc taggtgattt taatcactat accaattgag atgggctagt	7140
caatgataat tactagtcct tttcctttga gttgtgggta tctgtaaaatt ctgctagacc	7200
tttgctggaa aacttgtaaa ttctgctaga ccctctgtaa attccgctag acctttgtgt	7260
gttttttttg tttatattca agtggttata atttatagaa taaagaaaga ataaaaaaag	7320

ataaaaagaa tagatcccag ccctgtgtat aactcactac tttagtcagt tccgcagtat	7380
tacaaaagga tgtcgcaaac gctgttttgct cctctacaaa acagacctta aaaccctaaa	7440
ggcttaagta gcaccctcgc aagctcgggc aaatcgctga atattccttt tgtctccgac	7500
catcaggcac ctgagtcgct gtcttttttcg tgacattcag ttcgctgcgc tcacggctct	7560
ggcagtgaat gggggtaaataa ggcactacag gcgcctttta tggattcatg caaggaaact	7620
accataata caagaaaagc ccgtcacggg cttctcaggg cgttttatgg cgggtctgct	7680
atgtggtgct atctgacttt ttgctgttca gcagttcctg ccctctgatt ttccagtctg	7740
accacttcgg attatcccgt gacagggtcat tcagactggc taatgcaccc agtaaggcag	7800
cggtatcatc aacaggctta cccgtcttac tgtcttttct acggggctctg acgctcagtg	7860
gaacgaaaac tcacgttaag ggatttttgg catgagatta tcaaaaagga tcttcaccta	7920
gatcctttta aattaataat gaagttttaa atcaatctaa agtatatatg agtaaaactg	7980
gtctgacagt taccaatgct taatcagtga ggcacctatc tcagcgatct gtctatttcg	8040
ttcatccata gttgcctgac tccccgtcgt gtagataact acgatacggg agggcttacc	8100
atctggcccc agtgctgcaa tgataccgcg agaaccacgc tcaccggctc cagatttatc	8160
agcaataaac cagccagccg gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc	8220
ctccatccag tctattaatt gttgccggga agctagagta agtagttcgc cagttaatag	8280
tttgcgcaac gttgttgcca ttgctacagg catcgtgggtg tcacgctcgt cgtttggtat	8340
ggcttcattc agctccggtt cccaacgatc aaggcgagtt acatgatccc ccatgtttgtg	8400
caaaaaagcg gttagctcct tcggtcctcc gatcgttgctc agaagtaagt tggccgcagt	8460
gttatcactc atgggttatgg cagcactgca taattctctt actgtcatgc catccgtaag	8520
atgcttttct gtgactgggtg agtactcaac caagtcattc tgagaatagt gtatgcggcg	8580
accgagttgc tcttgcccgg cgtcaatacg ggataatacc gcgccacata gcagaacttt	8640
aaaagtgtc atcattggaa aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct	8700
gttgagatcc agttcgatgt aaccactcg tgcacccaac tgatcttcag catcttttac	8760
tttcaccagc gtttctgggt gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat	8820
aaggcgaca cggaatgtt gaatactcat actcttcctt tttcaatatt attgaagcat	8880
ttatcagggt tattgtctca tgagcggata catatttgaa tgtatttaga aaaataaaca	8940

aataggggtt ccgcgcacat ttccccgaaa agtgccacct gacgtctaag aaaccattat	9000
tatcatgaca ttaacctata aaaataggcg tatcacgagg ccctttcgtc tcgcgcgttt	9060
cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca cagcttgtct	9120
gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg ttggcgggtg	9180
tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc accatatgcg	9240
gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcaggcgcc attcgccatt	9300
caggctgctgc aactgttggg aagggcgatc ggtgcgggcc tcttcgctat tacgccagct	9360
ggcgaaaggg ggatgtgctg caaggcgatt aagttgggta acgccagggt tttcccagtc	9420
acgacgttgt aaaacgacgg ccagtgaatt cgtaaagacc cactttcaca ttttaagttgt	9480
ttttctaata cgcatatgat caattcaagg ccgaataaga aggctggctc tgcaccttgg	9540
tgatcaaata attcgatagc ttgtcgtaat aatggcggca tactatcagt agtaggtgtt	9600
tccctttctt ctttagcgac ttgatgctct tgatcttcca atacgcaacc taaagtaaaa	9660
tgccccacag cgctgagtgc atataatgca ttctctagtg aaaaaccttg ttggcataaa	9720
aaggctaatt gattttcgag agtttcatac tgtttttctg taggccgtgt acctaaatgt	9780
acttttgctc catcgcgatg acttagtaaa gcacatctaa aacttttagc gttattacgt	9840
aaaaaatctt gccagctttc cctttctaaa gggcaaaagt gagtatgggtg cctatctaac	9900
atctcaatgg ctaaggcgct gagcaaagcc cgcttatttt ttacatgcca atacaatgta	9960
ggctgctcta cacctagctt ctgggcgagt ttacgggttg ttaaaccctc gattccgacc	10020
tcattaagca gctctaatac gctgttaatc actttacttt tatctaatac agacatcatt	10080
aattcctaata ttttggtgac actctatcat tgatagagtt attttaccac tccctatcag	10140
tgatagagaa aagtga	10156

<210> 13

<211> 10122

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 13

aatattcacg taaggaggta atagatgcag atcttcgaca ctttgcaatc aatgggccat	60

gagcaggttg tcctatgtag cgataagacc acgggtctgc gcgccattat cgctatacac	120
gatacatcct tagggccggc gcttggtggt acccgatatgt ggagtagatgc aactgacgac	180
gatgctatta ctgacgcact ccgtctgtct cggggcatga cctacaaagc tgcggtttct	240
ggcgtaaatc tgggcggtgg taaagccgtt atcatcggaa accctcacag tgataaaagc	300
gaagcgctgt ttcgcgctta cggcagaatg gtggaatccc agcgtgggcg ttacatcacc	360
gccgaagacg ttggtactag cgtacgtgat atggagtgga ttcgcatgga aaccaaatat	420
gtaacgggcg tgggtggcaa cggaggctct ggtgaccctt ctccagttac cgctctgggt	480
gtttactcgg gcatgaaggc atgcgctaaa tcagtctatg gtactgatgc gctgagcgg	540
aaaaggatcg tggttcaggg cgcgggtaac gttgcatccc atctggttca cagtctggta	600
aaagaaggcg ctgtggtttt cgtcactgac atctacgaag aaaaggccaa agcattagcg	660
gctgaaacgg gcgctaccgt gattcgcacc gacgaggttt ttactacaca atgcgatatc	720
ttctctccga acgctctggg ggccgtcctg aacgatgaaa ctattccgca gctcacatgc	780
gctatcgtag ctggtggtgc aaacaatcag cttaaaatcg aacaacgtca cgccacggct	840
ctgcaagaga aaggcattct gtatgcgccg gattacgtaa tcaacgccgg gggcctcatg	900
aatgtggcga gcgaagttga cggctacaac cgtgaaaagg ttatgcgcca ggctgaaggt	960
atttacgata ttactatgaa catcctaaat accgcgcgtg agcgtaacat cctgaccatc	1020
gaagcatcca acgcgattgc tgaagagcgg atcaacaaag ttcgccatgt tcacgggaac	1080
ttcatcgggt ccccgcttat tcgcggagta taataaatca aaacaccaa ttaaggagaa	1140
attttatgtc gacgacaacc gttggtgact acttgctgta tcgcttaaac gaaatcggca	1200
ttgagcacct cttcggagtgc ccaggtgatt acaatctgca atttctggat catgtaatcg	1260
accaccctca gctgacttgg gtcggctgca ctaacgaact taacgctgcc tacgcagctg	1320
atggttatgc gcgttgctgt ccggctgcgg cactgctgac caccttcggg gttggcgaac	1380
tgagcgctat taatggcatc gcaggttcct acgcggagta tctgccgta atacatatcg	1440
ttggtgcacc gagtctatca gccagcagc agggcgacct gattcaccac tctcttggcg	1500
aaggtgattt ttccagcttc ctgaggatgt cccaaccggg gtctgttgcg caggctgctc	1560
tgactcctga taacgcatgc aaggaaatcg accgcgtact ggcggaagtc ctattcagc	1620
gtcgtcccgg ctacctgctg ctgtctaccg acgtggctgc tgcgccggcg gctctgccac	1680

aaagcactct	ttctttgccg	accgccccgg	atcatcgcg	agttctggct	gctttcagcg	1740
acgctgctga	gcagatgctg	gctcaggcca	aaagcgtctc	tctactggcg	gacttttctgg	1800
ctgatcgttt	cgggtgttact	cgagcactgg	ccgcgtggct	tcagcagggt	ccgctaccgc	1860
acgccactct	gttaatgggt	aaaggcgttc	tgagtgaaca	gcaaccaggg	ttcgtgggta	1920
cctacgctgg	tgcggcatct	atcgattcga	cgcgtggcg	aatcgaagaa	gctggggtaa	1980
ttatcgaggt	gggagttaga	ttttccgaca	ctatcacagc	aggcttctcg	cagcagatcg	2040
acgcccgcg	ttttatagac	attcaaccct	tcttctctcg	tattggcgat	cgccagtttg	2100
atcacttgcc	gatgcaggct	gccgtcgcag	ccctgcatca	actgtgtctt	cgttatcagc	2160
agcagtggtc	tatcaccgct	cctagcccg	ctgcaactgc	gccggctgct	ggtagcgagc	2220
tgtcccagaa	cgcatcttgg	caggcgatgc	agaacttcat	ccgccctggg	gacctgttgg	2280
tggccgacca	aggtactg	gcgttcggcg	cagcggcgct	gcgcttaccg	cagaattgcc	2340
agctgcttgt	gcagccgctg	tggggctcaa	tcggttacag	tctgccggcc	acctttggtg	2400
ctcagacggc	agatacagag	cgctgtgtaa	tcctaactcat	tggcgatggt	tcagcgcaat	2460
taactattca	ggaactttcc	agtatgatgc	gtgacggctt	gaaacctatc	atctttctcc	2520
tgaacaacaa	cggttacacc	gttgaacggg	cgattcacgg	cgcggagcaa	cgttataacg	2580
atatcgctgc	ttggaattgg	acccaactgc	cccaggcgct	gagtgttcat	tgcccagcgc	2640
agagctggcg	agtcgttgaa	acggtgcagc	tgaccgacgt	aatgaaagtc	atcgctgctt	2700
ctccgcgtct	gagcttggtg	gaagtgttgc	tgccgtgcaat	ggatgtccca	ccgctgctgc	2760
aagcagttag	tgccgctctg	aaccagcgca	actcctctta	atccctccct	acgggttaag	2820
gagcccctgc	tatgcgcagc	atgcagtttg	atgagtacgg	tgcacccctg	aaagcgttct	2880
catatgaaga	cccgaccccc	caagggaagg	aagtagtcgt	taggatcgaa	gcctgtggtg	2940
tgtgccactc	tgatattcat	cttcacgagg	gctacttcga	catgggcggg	ggcaataaag	3000
ctgatgttac	tcgtgctcgc	gaactccctt	ttacattggg	tcatgaaatc	gttggcgaag	3060
tggtagcaac	tggaccagggt	gtcaccggcg	ctaaaccggg	cgacaaacgt	attgtgtacc	3120
cgtggatcgg	gtgcggcgac	tgcccgaaat	gcaacagtgg	tgaggatcag	tcctgtgcgc	3180
gtccacgtaa	cctgggtgtt	cacgttgacg	gtggctattc	gacgcacgta	aagataccgg	3240
acgaaaaatt	cctgttcgcc	tacgatggta	ttcctactga	gttagcggga	acctatgctt	3300

gcagcggcat caccgcttat ggtgcactga tgaaagcaaa ggaagcggct gaaagatctg	3360
gctacatcgg tctgattggc gctgggtggcg ttggcatggc tggctctgatg ctggccaaag	3420
cagcgatcgg ggctaaaact gtagtctttg atatcgacga cgcaaaactg gaagctgcga	3480
cccgtgccgg ggcggattac gtgttcaact ccggtgcaaa agaaacacgc aaggaagtta	3540
tgaaactaac gaatggtggc ctgtctggtg ctgttgattt cgttggcagc gataaaagcg	3600
ctctgttttg aatcaacgcc ttgggtcaga acggcgtgct ggtcataatt ggactgttcg	3660
gtggcgctat gactgttccg gtaccctgt tcccgctgaa agggatcacc gtacgtggct	3720
catacgtagg ttccctgcaa gagatgagtg atatgatgga gttagtctgc gctgggaaag	3780
ttcctccgat gccggtaaaa actcggccac tggacgctgc ctgggaaacc cttgaggatc	3840
tacgccatgg taaaatcgtg ggccgtgttg ttctgacccc ataagaagga gatatacata	3900
tgacccatca attaagatcg cgcgatatca tcgctctggg ctttatgaca tttgcgttgt	3960
tcgtcggcgc aggtaacatt attttccctc caatggtcgg cttgcaggca ggcgaacacg	4020
tctggactgc ggcattcggc ttcctcatta ctgccgttgg cctaccggta ttaacggtag	4080
tggcgctggc aaaagttggc ggcggtgttg acagtctcag cacgccaatt ggtaaagtcg	4140
ctggcgctact gctggcaaca gtttgttacc tggcggtggg gccgcttttt gctacgccgc	4200
gtacagctac cgtttctttt gaagtgggca ttgcgccgct gacgggtgat tccgcgctgc	4260
cgctgtttat ttacagcctg gtctatttcg ctatcgttat tctggtttcg ctctatccgg	4320
gcaagctgct ggataccgtg ggcaacttcc ttgcgccgct gaaaattatc gcgctggtca	4380
tcctgtctgt tgccgcaatt atctggccgg cgggttctat cagtacggcg actgaggctt	4440
atcaaaacgc tgcgttttct aacggcttcg tcaacggcta tctgaccatg gatacgctgg	4500
gcgcaatggg gtttggtatc gttattgtta acgcggcgcg ttctcgtggc gttaccgaag	4560
cgcgtctgct gaccgttat accgtctggg ctggcctgat ggcgggtggt ggtctgactc	4620
tgctgtacct ggcgctgttc cgtctggggt cagacagcgc gtcgctggtc gatcagtctg	4680
caaacggcgc ggcgatcctg catgcttacg ttcagcatac ctttggcggc ggcggtagct	4740
tcctgctggc ggcgttaatc ttcatcgcct gcctggtcac ggcggttggc ctgacctgtg	4800
cttgtgcaga attcttcgcc cagtacgtac cgctctctta tcgtacgctg gtgtttatcc	4860
tcggcggctt ctcgatggtg gtgtctaacc tcggcttgag ccagctgatt cagatctctg	4920

taccggtgct gaccgccatt tatccgccgt gtatcgcaact ggttgtatta agttttacac	4980
gctcatgggtg gcataattcg tcccgcgtga ttgctccgcc gatgtttatc agcctgcttt	5040
ttgggtattct cgacgggatc aaggcatctg cattcagcga tatcttaccg tcctggggcgc	5100
agcgtttacc gctggccgaa caaggctctgg cgtgggtaat gccaacagtg gtgatgggtg	5160
ttctggccat tatctgggat cgtgcggcag gtcgtcaggt gacctccagc gtcactaat	5220
acgcatggca tggatgaccg atggtagtgt gggctctccc catgcgagag tagggaactg	5280
ccaggcatca aataaaacga aaggctcagt cgaaagactg ggcctttcgt tttatctgtt	5340
gtttgtcgggt gaacgctctc ctgagtagga caaatccgcc gggagcggat ttgaacgttg	5400
cgaagcaacg gcccggaggg tggcgggcag gacggccgcc ataaactgcc aggcataaaa	5460
ttaagcagaa ggccatcctg acggatggcc tttttgcgtg gccagtgcc aagcttgcag	5520
cgtgccagct gcattaatga agaaatcatg ctggaagaat aacagctcac tcaaaggcgg	5580
tagtacgggt tttgctgccc gcaaacgggc tgttctgggtg ttgctagttt gttatcagaa	5640
tcgcagatcc ggcttcagcc ggtttgccgg ctgaaagcgc tatttcttcc agaattgcc	5700
tgattttttc cccacgggag gcgtcactgg ctcccgtgtt gtcggcagct ttgattcgat	5760
aagcagcatc gcctgtttca ggctgtctat gtgtgactgt tgagctgtaa caagttgtct	5820
caggtgttca atttcatgtt ctagtgtctt tgttttactg gtttcacctg ttctattagg	5880
tgttacatgc tgttcatctg ttacattgtc gatctgttca tgggtgaacag ctttgaatgc	5940
acaaaaact cgtaaaagct ctgatgtatc tatctttttt acaccgtttt catctgtgca	6000
tatggacagt tttccctttg atatgtaacg gtgaacagtt gttctacttt tgtttgtag	6060
tcttgatgct tcatgatag atacaagagc cataagaacc tcagatcctt ccgtatttag	6120
ccagtatgtt ctctagtgtg gttcgttggt tttgcgtgag ccatgagaac gaaccattga	6180
gatcatactt actttgcatg tcatcaaaa attttgccctc aaaactgggtg agctgaattt	6240
ttgcagttaa agcatcgtgt agtgtttttc ttagtccgtt atgtaggtag gaatctgatg	6300
taatggttgt tggatatttg tcaccattca tttttatctg gttgttctca agttcggtta	6360
cgagatccat ttgtctatct agttcaactt ggaaaatcaa cgtatcagtc gggcggcctc	6420
gcttatcaac caccaatttc atattgctgt aagtgtttta atctttactt attggtttca	6480
aaaccattg gttaagcctt ttaaactcat ggtagttatt ttcaagcatt aacatgaact	6540

taaattcatc aaggctaadc tctatatattg ccttgtgagt tttcttttgt gttagttctt	6600
ttaataacca ctcataaadc ctcataagagt atttgttttc aaaagactta acatgtttcca	6660
gattatattt tatgaatttt ttttaactgga aaagataagg caatatctct tcactaaaaa	6720
ctaattctaa tttttcgctt gagaacttgg catagtttgt ccactggaaa atctcaaagc	6780
ctttaaccaa aggattcctg atttccacag ttctcgtcat cagctctctg gttgctttag	6840
ctaatacacc ataagcattt tccctactga tgttcatcat ctgagcgtat tggttataag	6900
tgaacgatac cgtccgttct ttccttgtag ggttttcaat cgtgggggtg agtagtgcca	6960
cacagcataa aattagcttg gtttcatgct ccgttaagtc atagcgacta atcgctagtt	7020
catttgcttt gaaaacaact aattcagaca tacatctcaa ttggtctagg tgattttaat	7080
cactatacca attgagatgg gctagtcaat gataattact agtccttttc ctttgagttg	7140
tgggtatctg taaattctgc tagacccttg ctggaaaact tgtaaattct gctagaccct	7200
ctgtaaattc cgctagacct ttgtgtgttt tttttgttta tattcaagtg gttataattt	7260
atagaataaa gaaagaataa aaaaagataa aaagaataga tcccagccct gtgtataact	7320
cactacttta gtcagttccg cagtattaca aaaggatgtc gcaaacgctg tttgctcctc	7380
tacaaaacag accttaaaac cctaaaggct taagtagcac cctcgcaagc tcgggcaaatt	7440
cgctgaatat tccttttgtc tccgaccatc aggcacctga gtcgctgtct ttttcgtgac	7500
attcagttcg ctgcgctcac ggctctggca gtgaatgggg gtaaattggca ctacaggcgc	7560
cttttatgga ttcatgcaag gaaactaccc ataatacaag aaaagcccgt cacgggcttc	7620
tcagggcgtt ttatggcggg tctgctatgt ggtgctatct gactttttgc tgttcagcag	7680
ttcctgccct ctgattttcc agtctgacca cttcggatta tcccgtgaca ggtcattcag	7740
actggctaatt gcaccagta aggacagcgt atcatcaaca ggcttaccgc tcttactgtc	7800
ttttctacgg ggtctgacgc tcagtggaac gaaaactcac gttaagggat tttggtcatg	7860
agattatcaa aaaggatctt cacctagatc cttttaaat aaaaatgaag ttttaaatca	7920
atctaaagta tatatgagta aacttggctt gacagttacc aatgcttaat cagtgaggca	7980
cctatctcag cgatctgtct atttcgttca tccatagttg cctgactccc cgtcgtgtag	8040
ataactacga tacgggaggg cttaccatct ggccccagtg ctgcaatgat accgcgagaa	8100
ccacgctcac cggctccaga tttatcagca ataaaccagc cagccggaag ggccgagcgc	8160

agaagtggtc	ctgcaacttt	atccgcctcc	atccagtcta	ttaattgttg	ccgggaagct	8220
agagtaagta	gttcgccagt	taatagtttg	cgcaacgttg	ttgccattgc	tacaggcatc	8280
gtgggtgtcac	gctcgtcgtt	tggtatggct	tcattcagct	ccggttccca	acgatcaagg	8340
cgagttacat	gatccccc	gttgtgcaaa	aaagcggtta	gctccttcgg	tcctccgatc	8400
gttgtcagaa	gtaagttggc	cgcagtgtta	tcactcatgg	ttatggcagc	actgcataat	8460
tctcttactg	tcatgccatc	cgtaagatgc	ttttctgtga	ctgggtgagta	ctcaaccaag	8520
tcattctgag	aatagtgtat	gcggcgaccg	agttgctctt	gcccggcgtc	aatacgggat	8580
aataccgcgc	cacatagcag	aacttttaaaa	gtgctcatca	ttggaaaacg	ttcttcgggg	8640
cgaaaactct	caaggatctt	accgctgttg	agatccagtt	cgatgtaacc	cactcgtgca	8700
cccaactgat	cttcagcatc	ttttactttc	accagcgttt	ctgggtgagc	aaaaacagga	8760
aggcaaaatg	ccgcaaaaaa	gggaataagg	gcgacacgga	aatgttgaat	actcatactc	8820
ttcctttttc	aatattattg	aagcatttat	cagggttatt	gtctcatgag	cggatacata	8880
tttgaatgta	tttagaaaaa	taaacaaata	ggggttccgc	gcacatttcc	ccgaaaagtg	8940
ccacctgacg	tctaagaaac	cattattatc	atgacattaa	cctataaaaa	taggcgtatc	9000
acgaggccct	ttcgtctcgc	gcgtttcggt	gatgacgggtg	aaaacctctg	acacatgcag	9060
ctcccggaga	cggtcacagc	ttgtctgtaa	gcggatgccg	ggagcagaca	agcccgtcag	9120
ggcgcgtcag	cgggtgttgg	cgggtgtcgg	ggctggctta	actatgcggc	atcagagcag	9180
attgtactga	gagtgcacca	tatgcggtgt	gaaataccgc	acagatgcgt	aaggagaaaa	9240
taccgcatca	ggcgccattc	gccattcagg	ctgcgcaact	gttgggaagg	gcgatcgggtg	9300
cgggcctctt	cgctattacg	ccagctggcg	aaagggggat	gtgctgcaag	gcgattaagt	9360
tgggtaacgc	cagggttttc	ccagtcacga	cgttgtaaaa	cgacggccag	tgaattcggtt	9420
aagaccact	ttcacattta	agttgttttt	ctaataccgca	tatgatcaat	tcaaggccga	9480
ataagaaggc	tggctctgca	ccttggtgat	caaataattc	gatagcttgt	cgtaataatg	9540
gcggcatact	atcagtagta	ggtgtttccc	tttcttcttt	agcgacttga	tgctcttgat	9600
cttccaatac	gcaacctaaa	gtaaaatgcc	ccacagcgct	gagtgcatat	aatgcattct	9660
ctagtgaaaa	accttgttgg	cataaaaagg	ctaattgatt	ttcgagagtt	tcatactggtt	9720
tttctgtagg	ccgtgtacct	aaatgtactt	ttgctccatc	gcgatgactt	agtaaagcac	9780

atctaaaact tttagcgtaa ttacgtaaaa aatcttgcca gctttcccct tctaaagggc	9840
aaaagtgagt atgggtgccta tctaacatct caatggctaa ggcgtcgagc aaagcccgct	9900
tatTTTTTtac atgccaatac aatgtaggct gctctacacc tagcttctgg gcgagtttac	9960
gggttggttaa accttcgatt ccgacctcat taagcagctc taatgcgctg ttaatcactt	10020
tacttttatc taatctagac atcattaatt cctaattttt gttgacactc tatcattgat	10080
agagttatTTT taccactccc tatcagtgat agagaaaagt ga	10122

<210> 14

<211> 10153

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 14

aatattcacg taaggaggta atagatgcag atcttcgaca ctttgcaatc aatgggccat	60
gagcagggtg tcctatgtag cgataagacc acgggtctgc gcgccattat cgctatacac	120
gatacatcct tagggccggc gcttggtggt acccgatatgt ggcagtatgc aactgacgac	180
gatgctatta ctgacgcact ccgtctgtct cggggcatga cctacaaagc tgcggtttct	240
ggcgtaaadc tgggcggtgg taaagccgtt atcatcgga accctcacag tgataaaagc	300
gaagcgctgt ttcgcgctta cggcagaatg gtggaatccc agcgtgggcg ttacatcacc	360
gccgaagacg ttggtactag cgtacgtgat atggagtgga ttcgcatgga aaccaaatat	420
gtaacgggcg tgggtggcaa cggaggctct ggtgaccct ctccagttac cgctctgggt	480
gtttactcgg gcatgaaggc atgcgctaaa tcagtctatg gtactgatgc gctgagcggc	540
aaaaggatcg tggttcaggg cgcgggtaac gttgcatccc atctgggtca cagtctggta	600
aaagaaggcg ctgtggTTTT cgtcactgac atctacgaag aaaaggccaa agcattagcg	660
gctgaaacgg gcgctaccgt gattcgcacc gacgaggttt ttactacaca atgcgatatc	720
ttctctccga acgctctggg ggccgtcctg aacgatgaaa ctattccgca gtcacatgc	780
gctatcgtag ctgggtgtgc aaacaatcag cttaaaatcg aacaacgtca cgccacggct	840
ctgcaagaga aaggcattct gtatgcgccg gattacgtaa tcaacgccgg gggcctcatg	900

aatgtggcga gcgaagttga cggctacaac cgtgaaaagg ttatgcgcca ggctgaaggt	960
atttacgata ttactatgaa catcctaaat accgcgcgtg agcgtaacat cctgaccatc	1020
gaagcatcca acgcgattgc tgaagagcgg atcaacaaag ttcgccatgt tcacgggaac	1080
ttcatcgggt ccccgctctat tcgcggagta taataaatca aaacaccaaa ttaaggagaa	1140
atTTTTatgtc gacgacaacc gttggtgact acttgctgta tcgcttaaAC gaaatcggca	1200
ttgagcacct cttcggagtG ccaggtgatt acaatctgca atttctggat catgtaatcg	1260
accaccctca gctgacttgg gtcggctgca ctaacgaact taacgctgcc tacgcagctg	1320
atggttatgc gcgttgtcgt ccggctgcgg cactgctgac caccttcggg gttggcgaac	1380
tgagcgctat taatggcatc gcaggttcct acgcggagta tctgccggtA atacatatcg	1440
ttggtgcacc gagtctatca gcccagcagc agggcgacct gattcaccac tctcttggcg	1500
aaggtgattt ttccagcttc ctgaggatgt cccaaccggt gtctgttgcg caggctgctc	1560
tgactcctga taacgcatgc aaggaaatcg accgcgtact ggcggaagtc ctcatcagc	1620
gtcgtcccgg ctacctgctg ctgtctaccg acgtggctgc tgcgccggcg gctctgccac	1680
aaagcactct ttctttgccg accgccccgg atcatcgcgC agttctggct gctttcagcg	1740
acgctgctga gcagatgctg gctcaggcca aaagcgtctc tctactggcg gactttctgg	1800
ctgatcgttt cggtgttact cgagcactgg ccgcgtggct tcagcaggtt ccgctaccgc	1860
acgccactct gttaatgggt aaaggcgttc tgagtgaaca gcaaccaggg ttcgtgggta	1920
cctacgctgg tgcggcatct atcgattcga cgcgtggcgc aatcgaagaa gctggggtaa	1980
ttatcggagt gggagttaga ttttccgaca ctatcacagc aggcttctcg cagcagatcg	2040
acgcccgccg ttttatagac attcaaccct tcttctctcg tattggcgat cgccagtttg	2100
atcacttgcc gatgcaggct gccgtcgcag ccctgcatca actgtgtctt cgttatcagc	2160
agcagtggtc tatcaccgct cctagcccgc ctgcactgcc gccggctgct ggtagcgagc	2220
tgtcccagaa cgcattctgg caggcgatgc agaacttcat ccgccctggg gacctgttgg	2280
tggccgacca aggtactgcg gcgttcggcg cagcggcgct gcgcttaccg cagaattgcc	2340
agctgcttgt gcagccgctg tggggctcaa tcggttacag tctgccggcc acctttggtg	2400
ctcagacggc agatacagag cgtcgtgtaa tcctaatacat tggcgatggt tcagcgcaat	2460
taactattca ggaactttcc agtatgatgc gtgacggctt gaaacctatc atctttctcc	2520

tgaacaacaa cggttacacc gttgaacggg cgattcacgg cgcggagcaa cgttataacg	2580
atatcgctgc ttggaattgg acccaactgc cccaggcgct gagtggtcat tgcccagcgc	2640
agagctggcg agtcgttgaa acggtgcagc tgaccgacgt aatgaaagtc atcgctgctt	2700
ctccgcgtct gagcttggtg gaagttgttc tgcctgcaat ggatgtcca ccgctgctgc	2760
aagcagttag tgccgctctg aaccagcgcg actcctctta aatTTTTTcaa cggcctatca	2820
taaggaggaa taatgacagc ggagcagcaa aatgggggtat ccgactcacg ccgtttcgaa	2880
tttcaggaat ttggtggccc tatcgcccca cagacctatc agctccccgc accggctagc	2940
gatgaagttt tgtaaaggt gaactactgc ggtgtctgtc acagtgatgt tcatcttcac	3000
gacggctact tcgagctggg tggcgataaa cgtctgaact tcgctatgcc gctgccgctg	3060
acgctgggtc acgaagtaat tggcaccgtt gtggctgtcg gcgaccaggt tactgggtgta	3120
aaaccggggg accagcgact gatctatccg tggatagggt gcggaaaatg cggcgcgtgt	3180
caaaaaggag aagaaaacct gtgcgttact cctgcacatc tgggcgtgaa caagccgggc	3240
ggttacgctg atcacatcgt tgtaccccat tctcgctacc ttctggacat ttcgggtctg	3300
aaccgggtg atgccgctac cctcgcgtgc tccggcctga ccactttcag cgcgatcaac	3360
aaagtgttgc cgcttgacga tgaccagtgg attgttggtta tcggttggtg tggcctcggc	3420
cagatggcgc tgcgtatcct gcaagctatg ggaattggca atgttatcgg tattgacctg	3480
tctgaagaga aacggaaact ggctcatgaa agcggcgcac gtcactcctt cgatccaaac	3540
actccgaagc tgaaccgcgt ggtcgccgaa acctgcccgg gtacggtaca ggccgcgtta	3600
gactttgtgg gcaatgagca aactgctcag ctggcactgt ctctgcttgg aaaaggtggc	3660
aaatatgttc ctgtcgggct gcacggcggc gagctgcgtt acccattgcc gatcatcacg	3720
aacaaagctg taagtatcat cggttcttac gttggtaccc tgaaagaact ggaagactta	3780
gttgctttcg ccaaggaaaa gaatctgccg ccaattcata ttgaacaccg cccgctggaa	3840
tcggcggctc aggccgtaga ggacctggaa aaaggacagg ttgctgggcg tgttatcctg	3900
gatgcaggta actaagaagg agatatacat atgacctatc aattaagatc gcgcgatatc	3960
atcgctctgg gctttatgac atttgcgttg ttcgctggcg caggtaacat tattttccct	4020
ccaatggctg gcttgaggc aggcgaacac gtctggactg cggcattcgg cttcctcatt	4080
actgccgttg gcctaccggt attaacggta gtggcgtg caaaagtgtg cggcgggtgtt	4140

gacagtctca gcacgccaat tggtaaagtc gctggcgtag tgctggcaac agtttgttac	4200
ctggcggtgg ggccgctttt tgctacgccg cgtacagcta ccgtttcttt tgaagtgggc	4260
attgcgccgc tgacgggtga ttccgcgctg ccgctgttta ttacagcct ggtctatttc	4320
gctatcgta ttctggtttc gctctatccg ggcaagctgc tggataccgt gggcaacttc	4380
cttgcgccgc tgaaaattat cgcgctggtc atcctgtctg ttgccgcaat tatctggccg	4440
gcgggttcta tcagtacggc gactgaggct tatcaaaacg ctgcgttttc taacggcttc	4500
gtcaacggct atctgaccat ggatacgctg ggcgcaatgg tgtttggtat cgttattgtt	4560
aacgcggcgc gttctcgtgg cgttaccgaa gcgcgtctgc tgacccgtta taccgtctgg	4620
gctggcctga tggcggtgt tggtctgact ctgctgtacc tggcgctgtt ccgtctgggt	4680
tcagacagcg cgtcgctggt cgatcagtct gcaaacggtg cggcgatcct gcatgcttac	4740
gttcagcata ctttggcgg cggcggtagc ttctgctgg cggcgtaat cttcatcgcc	4800
tgcttggtca cggcggttgg cctgacctgt gcttggtcag aattcttcgc ccagtacgta	4860
ccgtctctt atcgtacgct ggtgtttatc ctggcggtt tctcgatggt ggtgtctaac	4920
ctcggcttga gccagctgat tcagatctct gtaccggtgc tgaccgcat ttatccgccg	4980
tgtatcgcac tggttgtatt aagttttaca cgctcatggt ggcataattc gtccgcgtg	5040
attgctccgc cgatgtttat cagcctgctt tttggtattc tcgacgggat caaggcatct	5100
gcattcagcg atatcttacc gtcctgggcg cagcgtttac cgctggccga acaaggctctg	5160
gcgtgggtta tgccaacagt ggtgatggtg gttctggcca ttatctggga tcgtgcggca	5220
ggtcgtcagg tgacctcag cgctcactaa tacgcatggc atggatgacc gatggtagt	5280
tgggctctcc ccatgcgaga gtagggaact gccaggcatc aaataaaacg aaaggctcag	5340
tcgaaagact gggcctttcg ttttatctgt tgtttgtcgg tgaacgctct cctgagtagg	5400
acaaatccgc cgggagcgga tttgaacgtt gcgaagcaac ggcccggagg gtggcgggca	5460
ggacgccgc cataaactgc caggcatcaa attaagcaga aggccatcct gacggatggc	5520
ctttttgcgt ggccagtgcc aagcttgcac gcgtgccagc tgcattaatg aagaaatcat	5580
gctggaagaa taacagctca ctcaaaggcg gtagtacggg ttttgctgcc cgcaaacggg	5640
ctgttctggt gttgctagtt tgttatcaga atcgcagatc cggcttcagc cggtttgccg	5700
gctgaaagcg ctatttcttc cagaattgcc atgatttttt cccacggga ggcgtcactg	5760

gctcccgtgt tgtcggcagc tttgattcga taagcagcat cgcctgtttc aggctgtcta	5820
tgtgtgactg ttgagctgta acaagttgtc tcagggtgttc aatttcatgt tctagttgct	5880
ttgttttact ggtttcacct gttctattag gtgttacatg ctgttcatct gttacattgt	5940
cgatctgttc atgggtgaaca gctttgaatg caccaaaaac tcgtaaaagc tctgatgtat	6000
ctatcttttt tacaccgttt tcatctgtgc atatggacag ttttcccttt gatatgtaac	6060
ggatgaacagt tgttctactt ttgtttgtta gtcttgatgc ttcactgata gatacaagag	6120
ccataagaac ctcagatcct tccgtattta gccagtatgt tctctagtgt ggttcgttgt	6180
ttttgcbtga gccatgagaa cgaaccattg agatcatact tactttgcat gtcactcaaa	6240
aattttgcct caaaactggg gagctgaatt tttgcagtta aagcatcgtg tagtgttttt	6300
cttagtccgt tatgtaggta ggaatctgat gtaatggttg ttggatattt gtcaccattc	6360
atttttatct gggtgttctc aagttcgggt acgagatcca tttgtctatc tagttcaact	6420
tggaaaatca acgtatcagt cgggcggcct cgcttatcaa ccaccaattt catattgctg	6480
taagtgttta aatctttact tattggtttc aaaaccatt ggtaagcct tttaaactca	6540
tggtagtatt tttcaagcat taacatgaac ttaaattcat caaggctaatt ctctatattt	6600
gccttgtagg ttttcttttg tgtagttct ttttaataacc actcataaat cctcatagag	6660
tatttgtttt caaaagactt aacatgttcc agattatatt ttatgaattt ttttaactgg	6720
aaaagataag gcaatatctc ttcactaaaa actaattcta atttttcgct tgagaacttg	6780
gcatagtttg tccactggaa aatctcaaag cctttaacca aaggattcct gatttccaca	6840
gttctcgtca tcagctctct gggtgcttta gctaatacac cataagcatt ttccctactg	6900
atgttcatca tctgagcgta ttgggtataa gtgaacgata ccgtccgttc tttccttgta	6960
gggttttcaa tcgtgggggt gagtagtgcc acacagcata aaattagctt ggtttcatgc	7020
tccgttaagt catagcgact aatcgctagt tcatttgctt tgaaaacaac taattcagac	7080
atacatctca attggcttag gtgattttta tcactataacc aattgagatg ggctagtcaa	7140
tgataattac tagtcctttt cctttgagtt gtgggtatct gtaaattctg ctagaccttt	7200
gctggaaaac ttgtaaattc tgctagaccc tctgtaaatt ccgctagacc tttgtgtgtt	7260
ttttttgttt atattcaagt gggtataatt tatagaataa agaaagaata aaaaaagata	7320
aaaagaatag atcccagccc tgtgtataac tcactacttt agtcagttcc gcagtattac	7380

aaaaggatgt cgcaaacgct gtttgctcct ctacaaaaca gaccttaaaa ccctaaaggc	7440
ttaagtagca ccctcgcaag ctcgggcaaa tcgctgaata ttccttttgt ctccgaccat	7500
caggcacctg agtcgctgtc tttttcgtga cattcagttc gctgcgctca cggctctggc	7560
agtgaatggg ggtaaattggc actacaggcg ccttttatgg attcatgcaa ggaaactacc	7620
cataatacaa gaaaagcccg tcacgggctt ctcagggcgt tttatggcgg gtctgctatg	7680
tggtgctatc tgactttttg ctgttcagca gttcctgccc tctgattttc cagtctgacc	7740
acttcggatt atcccgtgac aggtcattca gactggctaa tgcacccagt aaggcagcgg	7800
tatcatcaac aggcttacct gtcttactgt cttttctacg gggctctgacg ctgagtggaa	7860
cgaaaactca cgttaaggga ttttggtcat gagattatca aaaaggatct tcacctagat	7920
ccttttaaat taaaaatgaa gttttaaatc aatctaaagt atatatgagt aaacttggtc	7980
tgacagttac caatgcttaa tcagtgaggc acctatctca gcgatctgtc tatttcgttc	8040
atccatagtt gcctgactcc ccgtcgtgta gataactacg atacgggagg gcttaccatc	8100
tggccccagt gctgcaatga taccgcgaga accacgtca ccggctccag atttatcagc	8160
aataaaccag ccagccggaa gggccgagcg cagaagtggc cctgcaactt tatccgcctc	8220
catccagtct attaattggt gccgggaagc tagagtaagt agttcgccag ttaatagttt	8280
gcgcaacggt gttgccattg ctacaggcat cgtgggtgtca cgctcgtcgt ttggtatggc	8340
ttcattcagc tccggttccc aacgatcaag gcgagttaca tgatcccca tgttgtgcaa	8400
aaaagcgggt agctccttcg gtcctccgat cgttgtcaga agtaagttgg ccgcagtgtt	8460
atcactcatg gttatggcag cactgcataa ttctcttact gtcatgccat ccgtaagatg	8520
cttttctgtg actggtgagt actcaaccaa gtcattctga gaatagtgtg tgcggcgacc	8580
gagttgctct tgcccggcgt caatacggga taataccgcg ccacatagca gaactttaaa	8640
agtgtcatc attggaaaac gttcttcggg gcgaaaactc tcaaggatct taccgctgtt	8700
gagatccagt tcgatgtaac ccactcgtgc acccaactga tcttcagcat cttttacttt	8760
caccagcgtt tctgggtgag caaaaacagg aaggcaaaat gccgcaaaaa agggaataag	8820
ggcgacacgg aaatgttgaa tactcatact cttccttttt caatattatt gaagcattta	8880
tcagggttat tgtctcatga gcggatacat atttgaatgt atttagaaaa ataaacaaat	8940
aggggttccg cgcacatttc cccgaaaagt gccacctgac gtctaagaaa ccattattat	9000

catgacatta acctataaaa ataggcgtat cacgaggccc tttcgtctcg cgcgtttcgg	9060
tgatgacggg gaaaacctct gacacatgca gctcccggag acggtcacag cttgtctgta	9120
agcggatgcc gggagcagac aagcccgtca gggcgcgta gcgggtgttg gcgggtgtcg	9180
gggctggctt aactatgcgg catcagagca gattgtactg agagtgcacc atatgcggtg	9240
tgaaataaccg cacagatgcg taaggagaaa ataccgcatc aggcgccatt cgccattcag	9300
gctgcgcaac tgttgggaag ggcgatcggg gcgggcctct tcgctattac gccagctggc	9360
gaaaggggga tgtgctgcaa ggcgattaag ttgggtaacg ccagggtttt cccagtcacg	9420
acgttgtaaa acgacggcca gtgaattcgt taagaccac tttcacattt aagttgtttt	9480
tctaattccg atatgatcaa ttcaaggccg aataagaagg ctggctctgc accttggtga	9540
tcaaataatt cgatagcttg tcgtaataat ggcggcatac tatcagtagt aggtgtttcc	9600
ctttcttctt tagcgacttg atgctcttga tcttccaata cgcaacctaa agtaaaatgc	9660
cccacagcgc tgagtgcata taatgcattc tctagtga aaaccttggtg gcataaaaag	9720
gctaattgat tttcgagagt ttcatactgt ttttctgtag gccgtgtacc taaatgtact	9780
tttgctccat cgcgatgact tagtaaagca catctaaaac ttttagcggtt attacgtaaa	9840
aaatcttgcc agctttcccc ttctaaaggg caaaagtga tatggtgcct atctaacatc	9900
tcaatggcta aggcgtcgag caaagccgc ttatTTTTTA catgccaata caatgtaggc	9960
tgctctacac ctagcttctg ggcgagttta cgggttggtta aaccttcgat tccgacctca	10020
ttaagcagct ctaatgcgct gttaatcact ttacttttat ctaatctaga catcattaat	10080
tcctaatttt tgttgacact ctatcattga tagagttatt ttaccactcc ctatcagtga	10140
tagagaaaag tga	10153

<210> 15

<211> 10058

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 15

aatattcacg taaggaggta atagatgcag atcttcgaca ctttgcaatc aatgggccat	60

gagcagggtg tcctatgtag cgataagacc acgggtctgc gcgccattat cgctatacac	120

gatacatcct tagggccggc gcttggtggt acccgatatgt ggtagtatgc aactgacgac	180
gatgctatta ctgacgcact ccgtctgtct cggggcatga cctacaaagc tgcggtttct	240
ggcgtaaatc tgggcggtgg taaagccgtt atcatcgga accctcacag tgataaaagc	300
gaagcgctgt ttcgcgctta cggcagaatg gtggaatccc agcgtgggcg ttacatcacc	360
gccgaagacg ttggtactag cgtacgtgat atggagtgga ttcgcatgga aaccaaatat	420
gtaacgggcg tgggtggcaa cggaggctct ggtgacctt ctccagttac cgctctgggt	480
gtttactcgg gcatgaaggc atgcgctaaa tcagtctatg gtactgatgc gctgagcgg	540
aaaaggatcg tggttcaggc cgcgggtaac gttgcatccc atctggttca cagtctggta	600
aaagaaggcg ctgtggtttt cgtcactgac atctacgaag aaaaggccaa agcattagcg	660
gctgaaacgg gcgctaccgt gattcgcacc gacgaggttt ttactacaca atgcgatatc	720
ttctctccga acgctctggg ggccgtcctg aacgatgaaa ctattccgca gtcacatgc	780
gctatcgtag ctggtggtgc aaacaatcag cttaaaatcg aacaacgtca cgccacggct	840
ctgcaagaga aaggcattct gtatgcgccg gattacgtaa tcaacgccgg gggcctcatg	900
aatgtggcga gcgaagttga cggctacaac cgtgaaaagg ttatgcgcca ggctgaagg	960
atttacgata ttactatgaa catcctaaat accgcgcgtg agcgtaacat cctgaccatc	1020
gaagcatcca acgcgattgc tgaagagcgg atcaacaaag ttcgccatgt tcacgggaac	1080
ttcatcggtt ccccgcttat tcgcggagta taataaatca aaacaccaaa ttaaggagaa	1140
attttatgtc gacgacaacc gttggtgact acttgctgta tcgcttaaac gaaatcggca	1200
ttgagcacct cttcggagtg ccagggtgatt acaatctgca atttctggat catgtaatcg	1260
accaccctca gctgacttgg gtcggctgca ctaacgaact taacgctgcc tacgcagctg	1320
atggttatgc gcgttgctgt cgggctgcgg cactgctgac caccttcggg gttggcgaac	1380
tgagcgctat taatggcatc gcaggttcct acgcggagta tctgccggtata acatatac	1440
ttggtgcacc gactctatca gccagcagc agggcgacct gattcaccac tctcttggcg	1500
aagggtgattt ttccagcttc ctgaggatgt cccaaccggt gtctgttgcg caggctgctc	1560
tgactcctga taacgcatgc aaggaaatcg accgcgtact ggcggaagtc ctattcagc	1620
gtcgtcccgg ctacctgctg ctgtctaccg acgtggctgc tgcgccggcg gctctgccac	1680
aaagcactct ttctttgccg accgccccgg atcatcgcg agttctggct gctttcagcg	1740

acgctgctga	gcagatgctg	gctcaggcca	aaagcgtctc	tctactggcg	gacttttctgg	1800
ctgatcgttt	cgggtgttact	cgagcactgg	ccgcgtggct	tcagcagggt	ccgctaccgc	1860
acgccactct	gttaatgggt	aaaggcgttc	tgagtgaaca	gcaaccaggg	ttcgtgggta	1920
cctacgctgg	tgcggcatct	atcgattcga	cgcggtggcg	aatcgaagaa	gctggggtaa	1980
ttatcggagt	gggagttaga	ttttccgaca	ctatcacagc	aggcttctcg	cagcagatcg	2040
acgcccgccg	ttttatagac	attcaaccct	tcttctctcg	tattggcgat	cgccagtttg	2100
atcacttgcc	gatgcaggct	gccgtcgcag	ccctgcatca	actgtgtctt	cgttatcagc	2160
agcagtggtc	tatcaccgct	cctagcccgc	ctgcactgcc	gccggctgct	ggtagcgagc	2220
tgtcccagaa	cgcatctctg	caggcgatgc	agaacttcat	ccgccctggg	gacctgttgg	2280
tggccgacca	aggtactgcg	gcgttcggcg	cagcggcgct	gcgcttaccg	cagaattgcc	2340
agctgcttgt	gcagccgctg	tggggctcaa	tcggttacag	tctgccggcc	acctttgggtg	2400
ctcagacggc	agatacagag	cgtcgtgtaa	tcctaatacat	tggcgatggg	tcagcgcaat	2460
taactattca	ggaactttcc	agtatgatgc	gtgacggctt	gaaacctatc	atctttctcc	2520
tgaacaacaa	cggttacacc	gttgaacggg	cgattcacgg	cgcgagcaa	cgttataacg	2580
atatcgctgc	ttggaattgg	acccaactgc	cccaggcgct	gagtgttcat	tgcccagcgc	2640
agagctggcg	agtcgttgaa	acggtgcagc	tgaccgacgt	aatgaaagtc	atcgctgctt	2700
ctccgcgtct	gagcttggtg	gaagtgtgtc	tgacctgcaat	ggatgtccca	ccgctgctgc	2760
aagcagttag	tgccgctctg	aaccagcgca	actcctctta	attcttaata	ataaggagct	2820
aattatgcag	acggcgttct	tgtataagcc	aggtcacgaa	aacttagtgc	gctcggagat	2880
cccgatacct	aaagctgggc	gtggcgaagt	cgttctggaa	attaaagccg	ctggcatgtg	2940
ccattccgat	ctgcacgttc	tcgacgggtg	aatccccctg	ccgggtcaat	ttgtaatggg	3000
ccatgaaatc	gttggtacta	ttcacgagat	cggccaggac	gtgaccggtt	tcaaacaggg	3060
cgatctgtac	gcagtccacg	gcccgaatcc	gtgtgggtatt	tgcaccctgt	gcagagaagg	3120
atttgataac	gactgcacta	cagtggcgaa	aaccgggtcaa	tgggttcggac	tgggtcttga	3180
cggcggctac	cagaagtata	tccgtatccc	gaacgtaagg	tctatcgtaa	aagttccaga	3240
aggtgtttca	gctgaggcag	ctgcgagctg	tactgatgca	gtactgacct	cgtaccgtgc	3300
actaaaacag	gctggcgcca	gcaactctac	tcgggtactg	attctgggtc	tgggtggctt	3360

aggtctgaat gcccttaaac tggctaagac cttcggcagt tacgtttacg catctgacct	3420
gaaaccttct gcgcgtgaag ctgctaaggc cgctggggcg gatgaagtgc tggagtcctt	3480
gcccgaagac ccgctgggtg ttgatatcgt gttagacgtc gttggcgtgc agagcacctt	3540
caacctcgct caaaaacacg ttggcccgcg tggcatcatt gtacctgtag gcctggcatc	3600
cccacagctt tcgtttaacc taacggatct ggcgctccgc gaaattcgtg ttcagggcac	3660
tttttggggc acgagcaatg agctggctga atgtctgcgc ctgtgccagc tgggcctgat	3720
caaccgaaa tatactgtgg tgcctcttga agaagcgccg aaatatatgg aagcaatggc	3780
tcatgggaaa gtagaaggct gtatcgtttt ccacccgtaa gaaggagata tacatatgac	3840
ccatcaatta agatcgcgcg atatcatcgc tctgggcttt atgacatttg cgttgttcgt	3900
cggcgcagggt aacattattt tccctccaat ggtcggcttg caggcaggcg aacacgtctg	3960
gactgcggca ttcggcttcc tcattactgc cgttggccta ccggtattaa cggtagtggc	4020
gctggcaaaa gttggcggcg gtgttgacag tctcagcacg ccaattggta aagtcgctgg	4080
cgtactgctg gcaacagttt gttacctggc ggtggggccg ctttttgcta cgccgcgtac	4140
agctaccgtt tcttttgaag tgggcattgc gccgctgacg ggtgattccg cgctgccgct	4200
gtttatttac agcctggtct atttcgctat cgttattctg gtttcgctct atccgggcaa	4260
gctgctggat accgtgggca acttccttgc gccgctgaaa attatcgcgc tggtcacctt	4320
gtctgttgcc gcaattatct ggccggcggg ttctatcagt acggcgactg aggcttatca	4380
aaacgctgcg ttttctaacg gcttcgtcaa cggctatctg accatggata cgctgggcgc	4440
aatggtgttt ggtatcgtta ttgttaacgc ggcgcgttct cgtggcgtta ccgaagcgcg	4500
tctgctgacc cgttataccg tctgggctgg cctgatggcg ggtgttggtc tgactctgct	4560
gtacctggcg ctgttccgtc tgggttcaga cagcgcgctc ctggtcgac agtctgcaaa	4620
cggtgccggc atcctgcatg cttacgttca gcataccttt ggccggcggcg gtagcttcct	4680
gctggcggcg ttaatcttca tcgcctgcct ggtcacggcg gttggcctga cctgtgcttg	4740
tgcagaattc ttcgcccagt acgtaccgct ctcttatcgt acgtgggtgt ttatcctcgg	4800
cggcttctcg atgggtgtgt ctaacctcgg cttgagccag ctgattcaga tctctgtacc	4860
ggtgctgacc gccatttatc cgccgtgtat cgcactgggt gtattaagtt ttacacgctc	4920
atggtggcat aattcgtccc gcgtgattgc tccgccgatg tttatcagcc tgctttttgg	4980

tattctcgac gggatcaagg catctgcatt cagcgatatc ttaccgtcct gggcgcagcg	5040
tttaccgctg gccgaacaag gtctggcgtg gttaatgcca acagtgggtga tgggtggttct	5100
ggccattatc tgggatcgtg cggcaggtcg tcaggtgacc tccagcgctc actaatacgc	5160
atggcatgga tgaccgatgg tagtgtgggc tctccccatg cgagagtagg gaactgccag	5220
gcatcaaata aaacgaaagg ctgagtcgaa agactggggc tttcgtttta tctgttgttt	5280
gtcggatgaac gctctcctga gtaggacaaa tccgccggga gcggatttga acgttgcgaa	5340
gcaacggccc ggagggtggc gggcaggacg cccgccataa actgccaggc atcaaattaa	5400
gcagaaggcc atcctgacgg atggcctttt tgcgtggcca gtgccaagct tgcattgcgtg	5460
ccagctgcat taatgaagaa atcatgctgg aagaataaca gctcactcaa aggcggtagt	5520
acggggtttt ctgcccgaac acgggctgtt ctgggtgttg tagtttgta tcagaatcgc	5580
agatccggct tcagccggtt tgccggctga aagcgctatt tcttccagaa ttgccatgat	5640
tttttcccca cgggaggcgt cactggctcc cgtgttgctg gcagctttga ttcgataagc	5700
agcatcgcct gtttcaggct gtctatgtgt gactgttgag ctgtaacaag ttgtctcagg	5760
tgttcaattt catgttctag ttgctttgtt ttactggttt cacctgttct attaggtgtt	5820
acatgctgtt catctgttac attgtcgatc tgttcatggg gaacagcttt gaatgcacca	5880
aaaactcgta aaagctctga tgtatctatc ttttttacac cgttttcatc tgtgcatatg	5940
gacagttttc cttttgatat gtaacgggtga acagttgttc tacttttggt tgtagtctt	6000
gatgcttcac tgatagatac aagagccata agaacctcag atccttccgt atttagccag	6060
tatgttctct agtgtggttc gttgtttttg cgtgagccat gagaacgaac cattgagatc	6120
atacttactt tgcattgtcac tcaaaaattt tgcctcaaaa ctggtgagct gaatttttgc	6180
agttaaagca tcgtgtagtg tttttcttag tccgttatgt aggtaggaat ctgatgtaat	6240
ggttggttgg attttgtcac cattcatttt tatctggttg ttctcaagtt cggttacgag	6300
atccatttgt ctatctagtt caacttggaa aatcaacgta tcagtcgggc ggcctcgctt	6360
atcaaccacc aatttcatat tgctgtaagt gtttaaatct ttacttattg gtttcaaaac	6420
ccattgggta agccttttaa actcatggta gttattttca agcattaaca tgaacttaaa	6480
ttcatcaagg ctaatctcta tatttgcctt gtgagttttc ttttgtgtta gttcttttaa	6540
taaccactca taaatcctca tagagtattt gttttcaaaa gacttaacat gttccagatt	6600

atattttatg aattttttta actggaaaag ataaggcaat atctcttcac taaaaactaa	6660
ttctaatttt tcgcttgaga acttggcata gtttgtccac tggaatatct caaagccttt	6720
aaccaaagga ttcctgattt ccacagttct cgtcatcagc tctctggttg ctttagctaa	6780
tacaccataa gcattttccc tactgatgtt catcatctga gcgtattggt tataagtga	6840
cgataccgtc cgttctttcc ttgtagggtt ttcaatcgtg gggttgagta gtgccacaca	6900
gcataaaatt agcttgggtt catgctccgt taagtcatag cgactaatcg ctagttcatt	6960
tgctttgaaa acaactaatt cagacataca tctcaattgg tctaggatgat tttaatcact	7020
ataccaattg agatgggcta gtcaatgata attactagtc cttttccttt gagttgtggg	7080
tatctgtaaa ttctgctaga cctttgctgg aaaacttgta aattctgcta gaccctctgt	7140
aaattccgct agacctttgt gtgttttttt tgtttatatt caagtgggta taatttatag	7200
aataaagaaa gaataaaaaa agataaaaag aatagatccc agccctgtgt ataactcact	7260
actttagtca gttccgcagt attacaaaag gatgtcgcaa acgctgtttg ctcctctaca	7320
aaacagacct taaaacccta aaggcttaag tagcaccctc gcaagctcgg gcaaactgct	7380
gaatattcct tttgtctccg accatcaggc acctgagtcg ctgtcttttt cgtgacattc	7440
agttcgctgc gtcacaggct ctggcagtga atgggggtaa atggcactac aggcgccttt	7500
tatggattca tgcaaggaaa ctaccataa tacaagaaaa gcccgtcacg ggcttctcag	7560
ggcgttttat ggcggtctg ctatgtggtg ctatctgact ttttgctgtt cagcagttcc	7620
tgccctctga ttttccagtc tgaccacttc ggattatccc gtgacaggtc attcagactg	7680
gctaatacac ccagtaaggc agcggatatca tcaacaggct taccgtctt actgtctttt	7740
ctacggggtc tgacgctcag tggaacgaaa actcacgtta agggattttg gtcattgagat	7800
tatcaaaaag gatcttcacc tagatccttt taaattaaaa atgaagtttt aatcaatct	7860
aaagtatata tgagtaaact tggcttgaca gttaccaatg cttaatcagt gaggcaccta	7920
tctcagcgat ctgtctattt cgttcatcca tagttgcctg actccccgtc gtgtagataa	7980
ctacgatacg ggagggtta ccatctggcc ccagtgtgc aatgataccg cgagaaccac	8040
gctcaccggc tccagattta tcagcaataa accagccagc cggaagggcc gagcgagaa	8100
gtggtcctgc aactttatcc gcctccatcc agtctattaa ttgttgccgg gaagctagag	8160
taagtagttc gccagttaat agtttgcgca acgttggtgc cattgctaca ggcatcgtgg	8220

tgtcacgctc gtcgttttgg atggcttcat tcagctccgg ttcccaacga tcaaggcgag	8280
ttacatgatc ccccatgttg tgcaaaaaag cggtagctc cttcggtcct ccgatcgttg	8340
tcagaagtaa gttggccgca gtgttatcac tcatggttat ggcagcactg cataattctc	8400
ttactgtcat gccatccgta agatgctttt ctgtgactgg tgagtactca accaagtcac	8460
tctgagaata gtgtatgcgg cgaccgagtt gctcttgccc ggcgtaata cgggataata	8520
ccgcgccaca tagcagaact ttaaaagtgc tcatcattgg aaaacgttct tcggggcgaa	8580
aactctcaag gatcttaccg ctgttgagat ccagttcgat gtaaccact cgtgcacca	8640
actgatcttc agcatctttt actttcacca gcgtttctgg gtgagcaaaa acaggaaggc	8700
aaaatgccgc aaaaaaggga ataaggcgca cacggaaatg ttgaatactc atactcttcc	8760
tttttcaata ttattgaagc atttatcagg gttattgtct catgagcgga tacatatttg	8820
aatgtattta gaaaaataaa caaatagggg ttccgcgcac atttccccga aaagtgccac	8880
ctgacgtcta agaaaccatt attatcatga cattaaccta taaaaatagg cgtatcacga	8940
ggccctttcg tctcgcgcgt ttcggtgatg acggtgaaaa cctctgacac atgcagctcc	9000
cggagacggg cacagcttgt ctgtaagcgg atgccgggag cagacaagcc cgtcagggcg	9060
cgtcagcggg tgttggcggg tgtcggggct ggcttaacta tgcggcatca gagcagattg	9120
tactgagagt gcaccatatg cgggtgtgaaa taccgcacag atgcgtaagg agaaaatacc	9180
gcatcaggcg ccattcgcca ttcaggctgc gcaactgttg ggaaggcgga tcggtgcggg	9240
cctcttcgct attacgccag ctggcgaaag ggggatgtgc tgcaaggcga ttaagttggg	9300
taacgccagg gttttccag tcacgacgtt gtaaaacgac ggccagtga ttcgttaaga	9360
cccactttca catttaagtt gtttttctaa tccgcatatg atcaattcaa ggccgaataa	9420
gaaggctggc tctgcacctt ggtgatcaaa taattcgata gcttgtcgta ataatggcgg	9480
catactatca gtagtaggtg tttccctttc ttcttttagcg acttgatgct cttgatcttc	9540
caatacgcaa cctaaagtaa aatgccccac agcgctgagt gcatataatg cattctctag	9600
tgaaaaacct tgttggcata aaaaggctaa ttgattttcg agagtttcat actgtttttc	9660
tgtaggccgt gtacctaaat gtacttttgc tccatcgcgga tgacttagta aagcacatct	9720
aaaactttta gcgttattac gtaaaaaatc ttgccagctt tccccttcta aagggcaaaa	9780
gtgagtatgg tgcctatcta acatctcaat ggctaaggcg tcgagcaaag cccgcttatt	9840

ttttacatgc caatacaatg taggctgctc tacacctagc ttctgggcga gtttacgggt	9900
tgttaaacct tcgattccga cctcattaag cagctctaata gcgctgttaa tcactttact	9960
tttatctaata ctagacatca ttaattccta atttttgttg acactctatc attgatagag	10020
ttattttacc actccctatc agtgatagag aaaagtga	10058

<210> 16

<211> 10136

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 16

gataaacgaa tcacaaaggc gaggcaagat gaacatcttc aagaaaatgg aggaatttaa	60
ttatgaacaa ctggtctact tctacgacag cgaaacggaa ctcaaaggta ttacctgtat	120
acacaacaca actttagggc cggcattggg cggtagccgc ctttggaact ataactctga	180
ggaagatgcc gttgaagacg taatccgtct ggctcggggc atgacttaca aagcggcttg	240
cgccggtctg aatctggggc gcggtaaaac cgtgctgata ggtgatgcta aaaagattaa	300
atcagagtcc tacttccgtg gactggggcg ctacgttcag tcgctgaacg gcagatatat	360
caccgcggaa gacgtaaata cttctacgaa ggatatggca tacgttgcta tggaaactga	420
ctatgtggta ggcctgggag gtaaattcgg caaccctagt ccagttactg cttacgggtgc	480
atttatgggt atcaaagcgg cgctgatgaa aaaatttgag gatagctcta ttgaaggccg	540
aaccttcgca gtgcagggtg ctgggcagac gggttactat cttatcgatt acctcctagg	600
caacaacaag ttcaaagaaa aggctaaaaa gatttacttc accgaaatta acgagagcta	660
tatcgagcgt atgaacaaag aacatccgga agttgaattt atttccccgg acaaaatcta	720
ctcgctggaa gtagacgtct tcgtgccctg cgccctgggc aaaatcgta atgacaaaac	780
tatcgatgaa ttttaagtgtc cgatcatcgc aggtactgca aacaacgtac tggaaaggga	840
agcgcacggc aacatgctta aagaacgtgg cattctttac gccccggact atgtgatcaa	900
tgctgggtggg ctgatcaacg tttaccacga gctgaacggg tacaataaag agaacgctat	960
tctggaagtg gaattaattt atgatgcct actggaaata ttcaacatcg ctgattctct	1020

gaacatcagc	accaatatcg	ctgccaacga	gttcgcggaa	aaacgtatca	agcaaattaa	1080
gtccttgaaa	aacaacttca	ttaaagcta	agaccattaa	taaggaacat	tttatgtcga	1140
cgacaaccgt	tggtgactac	ttgctgtatc	gcttaaacga	aatcggcatt	gagcacctct	1200
tcggagtgcc	aggtgattac	aatctgcaat	ttctggatca	tgtaatcgac	caccctcagc	1260
tgacttgggt	cggctgcact	aacgaactta	acgctgccta	cgcagctgat	ggttatgcgc	1320
gttgtcgtcc	ggctgcggca	ctgctgacca	ccttcgggggt	tggcgaactg	agcgctatta	1380
atggcatcgc	aggttcctac	gcggagtatc	tgccggtaat	acatatcggt	ggtgcaccga	1440
gtctatcagc	ccagcagcag	ggcgacctga	ttcaccactc	tcttggcgaa	ggtgattttt	1500
ccagcttcct	gaggatgtcc	caaccggtgt	ctgttgcgca	ggctgctctg	actcctgata	1560
acgcatgcaa	ggaaatcgac	cgcgtactgg	cggaagtcc	cattcagcgt	cgtcccggct	1620
acctgctgct	gtctaccgac	gtggctgctg	cgccggcggc	tctgccacaa	agcactcttt	1680
ctttgccgac	cgccccggat	catcgcgacg	ttctggctgc	tttcagcgac	gctgctgagc	1740
agatgctggc	tcaggccaaa	agcgtctctc	tactggcgga	ctttctggct	gatcgtttcg	1800
gtgttactcg	agcactggcc	gcgtggcttc	agcaggttcc	gctaccgcac	gccactctgt	1860
taatgggtaa	aggcgttctg	agtgaacagc	aaccagggtt	cgtgggtacc	tacgctgggtg	1920
cggcatctat	cgattcgacg	cgtggcgcaa	tcgaagaagc	tggggtaatt	atcggagtgg	1980
gagttagatt	ttccgacact	atcacagcag	gcttctcgca	gcagatcgac	gcccggcggt	2040
ttatagacat	tcaacccttc	ttctctcgta	ttggcgatcg	ccagtttgat	cacttgccga	2100
tgcaggctgc	cgtcgcagcc	ctgcatcaac	tgtgtcttcg	ttatcagcag	cagtggctcta	2160
tcaccgctcc	tagcccgctt	gcactgccgc	cggctgctgg	tagcgagctg	tcccagaacg	2220
cattctggca	ggcgatgcag	aacttcattc	gccctgggga	cctgttggtg	gccgaccaag	2280
gtactgcggc	gttcggcgca	gcggcgctgc	gcttaccgca	gaattgccag	ctgcttgctgc	2340
agccgctgtg	gggctcaatc	ggttacagtc	tgccggccac	ctttggtgct	cagacggcag	2400
atacagagcg	tcgtgtaatc	ctaattcattg	gcgatggttc	agcgcaatta	actattcagg	2460
aactttccag	tatgatgcgt	gacggcttga	aacctatcat	ctttctcctg	aacaacaacg	2520
gttacaccgt	tgaacgggcg	attcacggcg	cggagcaacg	ttataacgat	atcgctgctt	2580
ggaattggac	ccaactgccc	caggcgctga	gtgttcattg	cccagcgagc	agctggcgag	2640

tcgttgaaac ggtgcagctg accgacgtaa tgaaagtcac cgctgcttct ccgctgctga	2700
gcttggtaga agttgttctg cctgcaatgg atgtcccacc gctgctgcaa gcagtgagtg	2760
ccgctctgaa ccagcgcaac tcctcttaat tggataaaga ataaggaggt attttatgac	2820
agcggagcag caaaatgggg tatccgactc acgccgtttc gaatttcagg aatttggtgg	2880
ccctatcgcc ccacagacct atcagctccc cgcaccggct agcgatgaag ttttgtaaaa	2940
ggtgaactac tgcggtgtct gtcacagtga tgttcatctt cacgacggct acttcgagct	3000
gggtggcgat aaacgtctga acttcgctat gccgctgccg ctgacgctgg gtcacgaagt	3060
aattggcacc gttgtggctg tcggcgacca gggtactggg gtaaaaccgg gggaccagcg	3120
actgatctat ccgtggatag gttgcggaaa atgcggcgcg tgtcaaaaag gagaagaaaa	3180
cctgtgcgtt actcctgcac atctgggcgt gaacaagccg ggcggttacg ctgatcacat	3240
cgttgtaccc cattctcgct accttctgga catttcgggt ctgaaccggg gtgatgccgc	3300
taccctcgcg tgctccggcc tgaccacttt cagcgcgatc aacaaagtgt tgccgcttgc	3360
agatgaccag tggattgttg ttatcggttg tggtagcctc ggccagatgg cgctgcgtat	3420
cctgcaagct atgggaattg gcaatgttat cggtattgac ctgtctgaag agaaacggaa	3480
actggctcat gaaagcgggt cacgtcactc cttcgatcca aacactccga agctgaaccg	3540
cgtggctgcc gaaacctgcc cgggtacggg acaggccgcg ttagactttg tgggcaatga	3600
gcaaactgct cagctggcac tgtctctgct tggaaaagggt ggcaaatatg ttcctgtcgg	3660
gctgcacggc ggcgagctgc gttaccatt gccgatcatc acgaacaaag ctgtaagtat	3720
catcggttct tacgttggtg ccctgaaaga actggaagac ttagttgctt tcgccaagga	3780
aaagaatctg ccgccaattc atattgaaca ccgccgctg gaatcggcgg ctcaggccgt	3840
agaggacctg gaaaaaggac aggttgctgg gcgtgttatc ctggatgcag gtaactaaga	3900
aggagatata catatgacct atcaattaag atcgcgcgat atcatcgctc tgggctttat	3960
gacatttgcg ttgttcgtcg gcgcaggtaa cattattttc cctccaatgg tcggcttgca	4020
ggcaggcgaa cacgtctgga ctgcggcatt cggcttcctc attactgccg ttggcctacc	4080
ggtattaacg gtagtggcgc tggcaaaagt tggcggcggg gttgacagtc tcagcacgcc	4140
aattggtaaa gtcgtggcg tactgctggc aacagtttgt tacctggcgg tggggccgct	4200
ttttgctacg ccgcgtacag ctaccgtttc ttttgaagtg ggcatcgcg cgctgacggg	4260

tgattccgcg ctgccgctgt ttattttacag cctgggtctat ttcgctatcg ttatttctggt	4320
ttcgctctat ccgggcaagc tgctggatac cgtgggcaac ttccttgcg cgtgaaaat	4380
tatcgcgctg gtcacacctgt ctgttgccgc aattatctgg ccggcgggtt ctatcagtac	4440
ggcgactgag gcttatcaaa acgctgcgtt ttctaacggc ttcgtcaacg gctatctgac	4500
catggatac cgtggcgcaa tgggtgtttgg tatcgttatt gttaacgcgg cgcgttctcg	4560
tggcgttacc gaagcgcgtc tgctgaccgc ttataccgtc tgggctggcc tgatggcggg	4620
tgttggtctg actctgctgt acctggcgct gttccgtctg gggttcagaca gcgcgtcgct	4680
ggtcgatcag tctgcaaacg gtgcggcgat cctgcatgct tacgttcagc atacctttgg	4740
cggcggcggg agcttcctgc tggcggcggt aatcttcatc gcctgcctgg tcacggcggg	4800
tggcctgacc tgtgcttggt cagaattctt cgcccagtac gtaccgctct cttatcgtac	4860
gctgggtgtt atcctcggcg gcttctcgat ggtgggtgtt aacctcggct tgagccagct	4920
gattcagatc tctgtaccgg tgctgaccgc catttatccg ccgtgtatcg cactggttgt	4980
attaagtttt acacgctcat ggtggcataa ttcgtccgc gtgattgctc cgccgatgtt	5040
tatcagcctg ctttttggtt ttctcgacgg gatcaaggca tctgcattca gcgatatctt	5100
accgtcctgg gcgcagcgtt taccgctggc cgaacaaggc ctggcgtggt taatgccaac	5160
agtgggtgat gtggttctgg ccattatctg ggatcgtgcg gcaggctcgtc aggtgacctc	5220
cagcgctcac taatacgcac ggcatggatg accgatggta gtgtgggctc tccccatgcg	5280
agagtaggga actgccaggc atcaaataaa acgaaaggct cagtcgaaag actgggcctt	5340
tcgttttatc tgttggttgt cgggtgaacgc tctcctgagt aggacaaatc cgccgggagc	5400
ggatttgaac gttgcgaagc aacggcccgg aggggtggcg gcaggacgcc cgccataaac	5460
tgccaggcat caaattaagc agaaggccat cctgacggat ggcctttttg cgtggccagt	5520
gccaagcttg catgcgtgcc agctgcatta atgaagaaat catgctggaa gaataacagc	5580
tcactcaaag gcggtagtac gggttttgct gcccgcaaac gggctgttct ggtgttgcta	5640
gtttgttatc agaatcgag atccggcttc agccggtttg ccggctgaaa gcgctatttc	5700
ttccagaatt gccatgattt tttccccacg ggaggcgtca ctggctcccg tgttgtcggc	5760
agctttgatt cgataagcag catcgcctgt ttcaggctgt ctatgtgtga ctgttgagct	5820
gtaacaagtt gtctcaggtg ttcaatttca tgttctagtt gctttgtttt actggtttca	5880

cctgttctat taggtgttac atgctgttca tctgtttacat tgtcgatctg ttcatgggtga	5940
acagctttga atgcaccaaaa aactcgtaaa agctctgatg tatctatctt ttttacaccg	6000
ttttcatctg tgcatatgga cagttttccc tttgatatgt aacgggtgaac agttgtttcta	6060
cttttgtttg ttagtcttga tgcttcactg atagatacaa gagccataag aacctcagat	6120
ccttccgtat ttagccagta tgttctctag tgtggttcgt tgtttttgcg tgagccatga	6180
gaacgaacca ttgagatcat acttactttg catgtcactc aaaaattttg cctcaaaact	6240
ggtagagctga atttttgcag ttaaagcatc gtgtagtggt tttcttagtc cgttatgtag	6300
gtaggaatct gatgtaatgg ttgttggtat tttgtcacca ttcattttta tctggttggt	6360
ctcaagttcg gttacgagat ccatttgtct atctagttca acttggaaaa tcaacgtatc	6420
agtcgggcgg cctcgcttat caaccaccaa tttcatattg ctgtaagtgt ttaaattctt	6480
acttattggt ttcaaaacc atttggttaag ctttttaaac tcatggtagt tattttcaag	6540
cattaacatg aacttaaatt catcaaggct aatctctata tttgccttgt gagttttctt	6600
ttgtgttagt tcttttaata accactcata aatcctcata gagtatttgt tttcaaaaga	6660
cttaacatgt tccagattat attttatgaa ttttttaaac tggaaaagat aaggcaatat	6720
ctcttcacta aaaactaatt ctaatttttc gcttgagaac ttggcatagt ttgtccactg	6780
gaaaatctca aagcctttaa ccaaaggatt cctgatttcc acagttctcg tcatcagctc	6840
tctggttgct ttagctaata caccataagc attttcccta ctgatgttca tcatctgagc	6900
gtattgggta taagtgaacg ataccgtccg ttctttcctt gtagggtttt caatcgtggg	6960
gttgagtagt gccacacagc ataaaattag cttggtttca tgctccgtta agtcatagcg	7020
actaatcgct agttcatttg ctttgaaaac aactaattca gacatacatc tcaattggtc	7080
taggtgattt taatcactat accaattgag atgggctagt caatgataat tactagtcct	7140
tttcctttga gttgtgggta tctgtaaatt ctgctagacc tttgctggaa aacttgtaaa	7200
ttctgctaga ccctctgtaa attccgctag acctttgtgt gttttttttg tttatattca	7260
agtggttata atttatagaa taaagaaaga ataaaaaag ataaaaagaa tagatcccag	7320
ccctgtgtat aactcactac tttagtcagt tccgcagtat tacaaaagga tgtcgcaaac	7380
gctgtttgct cctctacaaa acagacctta aaaccctaaa ggcttaagta gcaccctcgc	7440
aagctcgggc aaatcgctga atattccttt tgtctccgac catcaggcac ctgagtcgct	7500

gtctttttcg tgacattcag ttcgctgcgc tcacggctct ggcagtgaat gggggtaa	7560
ggcactacag ggccttttta tggattcatg caaggaaact acccataata caagaaaagc	7620
ccgtcacggg cttctcaggg cgttttatgg cgggtctgct atgtggtgct atctgacttt	7680
ttgctgttca gcagttcctg ccctctgatt ttccagtctg accacttcgg attatcccgt	7740
gacaggtcat tcagactggc taatgcaccc agtaaggcag cggatatcatc aacaggctta	7800
cccgtcttac tgtcttttct acggggctctg acgctcagtg gaacgaaaac tcacgttaag	7860
ggattttggg catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaat	7920
gaagtttta atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct	7980
taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcattccata gttgcctgac	8040
tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa	8100
tgataccgcg agaaccacgc tcaccggctc cagatttatc agcaataaac cagccagccg	8160
gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt	8220
gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca	8280
ttgctacagg catcgtgggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt	8340
cccaacgatc aaggcgagtt acatgatccc ccatgttggtg caaaaaagcg gtagctcct	8400
tcggtcctcc gatcgttgctc agaagtaagt tggccgcagt gttatcactc atggttatgg	8460
cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactgggtg	8520
agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg	8580
cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgtc atcattggaa	8640
aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt	8700
aaccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt	8760
gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aaggcgaca cggaaatgtt	8820
gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca	8880
tgagcggata catatttgaa tgtatttaga aaaataaaca aataggggtt ccgcgcacat	8940
ttccccgaaa agtgccacct gacgtctaag aaaccattat tatcatgaca ttaacctata	9000
aaaataggcg tatcacgagg ccctttcgtc tcgcgcgttt cggatgatgac ggtgaaaacc	9060
tctgacacat gcagctcccg gagacgggtca cagcttgtct gtaagcggat gccgggagca	9120

gacaagcccg tcagggcgcg tcagcgggtg ttggcgggtg tcggggctgg cttactatg	9180
cggcatcaga gcagattgta ctgagagtgc accatatgcg gtgtgaaata ccgcacagat	9240
gcgtaaggag aaaataccgc atcaggcgcc attcgccatt caggctgcgc aactgttggg	9300
aagggcgatc ggtgcgggcc tcttcgctat tacgccagct ggcgaaaggg ggatgtgctg	9360
caaggcgatt aagttgggta acgccagggt tttcccagtc acgacgttgt aaaacgacgg	9420
ccagtgaatt cgtaaagacc cactttcaca ttttaagttgt ttttctaatac cgcatatgat	9480
caattcaagg ccgaataaga aggctggctc tgcaccttgg tgatcaaata attcgatagc	9540
ttgtcgtaat aatggcggca tactatcagt agtaggtgtt tccctttctt ctttagcgac	9600
ttgatgctct tgatcttcca atacgaacc taaagtaaaa tgccccacag cgctgagtgc	9660
atataatgca ttctctagtg aaaaaccttg ttggcataaa aaggctaatt gattttcgag	9720
agtttcatac tgtttttctg taggccgtgt acctaaatgt acttttgctc catcgcgatg	9780
acttagtaaa gcacatctaa aacttttagc gttattacgt aaaaaatctt gccagctttc	9840
cccttctaaa gggcaaaagt gagtatggtg cctatctaac atctcaatgg ctaaggcgtc	9900
gagcaaagcc cgcttatttt ttacatgcca atacaatgta ggctgctcta cacctagctt	9960
ctgggcgagt ttacgggttg ttaaaccctt gattccgacc tcattaagca gctctaagc	10020
gctgttaatc actttacttt tatctaattc agacatcatt aattcctaatt tttgttgac	10080
actctatcat tgatagagtt attttaccac tccctatcag tgatagagaa aagtga	10136

<210> 17

<211> 10118

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 17

gagctagata ataaggacgg cacaagatga acatcttcaa gaaaatggag gaatttaatt	60
atgaacaact ggtctacttc tacgacagcg aaacggaact caaaggattt acctgtatac	120
acaacacaac tttagggccg gcattgggcg gtacccgcct ttggaactat aactctgagg	180
aagatgccgt tgaagacgta atccgtctgg ctgggggcat gacttacaaa gcggcttgcg	240
ccggtctgaa tctgggcggc ggtaaaaccg tgctgatcgg tgatgctaaa aagattaaat	300

cagagtccta cttccgtgga ctggggcgct acgttcagtc gctgaacggc agatatatca	360
ccgcggaaga cgtaaatact tctacgaagg atatggcata cgttgctatg gaaactgact	420
atgtggtagg cctgggaggt aaatccggca accctagtcc agttactgct tacgggtgcat	480
ttatgggtat caaagcggcg ctgatgaaaa aatttgagga tagctctatt gaaggccgaa	540
ccttcgcagt gcagggtgct gggcagacgg gttactatct tatcgattac ctccctaggca	600
acaacaagtt caaagaaaag gctaaaaaga tttacttcac cgaaattaac gagagctata	660
tcgagcgtat gaacaaagaa catccggaag ttgaatttat ttccccggac aaaatctact	720
cgctggaagt agacgtcttc gtgccctgcg ccctgggcaa aatcgттаат gacaaaacta	780
tcgatgaatt taagtgtccg atcatcgcag gtactgcaaa caacgtactg gaaagggag	840
cgcacggcaa catgcttaaa gaacgtggca ttctttacgc cccggactat gtgatcaatg	900
ctggtgggct gatcaacgtt taccacgagc tgaacggtta caataaagag aacgctattc	960
tggaagtgga attaatttat gatcgcctac tggaaatatt caacatcgct gattctctga	1020
acatcagcac caatatcgct gccaacgagt tcgcggaaaa acgtatcaag caaattaagt	1080
ccttgaaaaa caacttcatt aaacgctaag accattaata aggaacattt tatgtcgacg	1140
acaaccgttg gtgactactt gctgtatcgc ttaaacgaaa tcggcattga gcacctcttc	1200
ggagtgccag gtgattacaa tctgcaattt ctggatcatg taatcgacca ccctcagctg	1260
acttgggtcg gctgcactaa cgaacttaac gctgcctacg cagctgatgg ttatgcgcgt	1320
tgtcgtccgg ctgcggcact gctgaccacc ttcggggttg gcgaactgag cgctattaat	1380
ggcatcgcag gttcctacgc ggagtatctg ccggtataac atatcgttgg tgcaccgagt	1440
ctatcagccc agcagcaggg cgacctgatt caccactctc ttggcgaagg tgatTTTTcc	1500
agcttcctga ggatgtccca accggtgtct gttgcgcagg ctgctctgac tcctgataac	1560
gcatgcaagg aaatcgaccg cgtactggcg gaagtcctca ttcagcgtcg tccgggtac	1620
ctgctgctgt ctaccgacgt ggctgctgcg ccggcggctc tgccacaaag cactctttct	1680
ttgccgaccg ccccgatca tcgcgcagtt ctggctgctt tcagcgacgc tgctgagcag	1740
atgctggctc aggccaaaag cgtctctcta ctggcggact ttctggctga tcgtttcggt	1800
gttactcgag cactggccgc gtggcttcag caggttccgc taccgcacgc cactctgtta	1860
atgggtaaag gcgttctgag tgaacagcaa ccagggttcg tgggtaccta cgctggtgcg	1920

gcatctatcg attcgacgcg tggcgcaatc gaagaagctg gggtaattat cggagtggga	1980
gtagatTTTT cCGacactat cacagcaggc ttctcgCagc agatcgacgc ccgCCgtttt	2040
atagacattc aacccttctt ctctcgtatt ggcgatcgcc agtttgatca cttgccgatg	2100
caggctgccg tcgcagccct gcatcaactg tgtcttcgtt atcagcagca gtggtctatc	2160
accgctccta gcccgcctgc actgccgccg gctgctggta gcgagctgtc ccagaacgca	2220
ttctggcagg cgatgcagaa cttcatccgc cctggggacc tgttggTggc cgaccaaggt	2280
actgcggcgt tcggcgCagc ggcgctgcgc ttaccgcaga attgccagct gcttgTgcag	2340
ccgctgtggg gctcaatcgg ttacagtctg ccggccacct ttggtgctca gacggcagat	2400
acagagcgtc gtgtaatcct aatcattggc gatggTtcag cgcaattaac tattcaggaa	2460
ctttccagta tgatgcgtga cggcttgaaa cctatcatct ttctcctgaa caacaacggT	2520
tacaccgttg aacgggcgat tcacggcgCg gagcaacgtt ataacgatat cgctgcttg	2580
aattggaccc aactgcccc a ggcgctgagt gttcattgcc cagcgCagag ctggcgagtc	2640
gttgaaacgg tgcagctgac cgacgtaatg aaagtcatcg ctgcttctcc gcgtctgagc	2700
ttggtagaag ttgttctgcc tgcaatggat gtcccaccgc tgctgcaagc agtgagtgcc	2760
gctctgaacc agcgcaactc ctcttaattt ttaaagtaac acacgagcta agaactaagg	2820
aggatttatg cgcagcatgc agtttgatga gtacggTgca cccctgaaag cgTtctcata	2880
tgaagacccg accccgcaag ggaaggaagt agtcgttagg atcgaagcct gtggtgtgtg	2940
ccactctgat attcatcttc acgagggcta cttcgacatg ggcggTggca ataaagctga	3000
tgTtactcgt gctcgCgaac tcccttttac attgggtcat gaaatcgTtg gcgaagtggT	3060
agcaactgga ccaggTgtca ccggcgctaa accgggcgac aaacgtattg tgtacccgtg	3120
gatcgggtgc ggcgactgcc cgaaatgcaa cagtggTgag gatcagTcct gtgcgcgtcc	3180
acgtaacctg ggtgttcacg ttgacggTgg ctattcgacg cacgtaaaga taccggacga	3240
aaaattcctg ttcgcctacg atggtattcc tactgagtta gcgggaacct atgcttgCag	3300
cggcatcacc gcttatggTg cactgatgaa agcaaaggaa gcggctgaaa gatctggcta	3360
catcggTctg attggcgctg gtggcgTtg catggctggT ctgatgctgg ccaaagcagc	3420
gatcggggct aaaactgtag tctttgatat cgacgacgca aaactggaag ctgcgacccg	3480
tgccggggcg gattacgtgt tcaactccgg tgcaaaagaa acacgcaagg aagttatgaa	3540

actaacgaat ggtggcctgt ctggtgctgt tgatttcgtt ggcagcgata aaagcgctct	3600
gtttggaatc aacgccttgg gtcagaacgg cgtgctggtc ataattggac tgttcggtgg	3660
cgctatgact gttccggtac ccctgttccc gctgaaaggg atcaccgtac gtggctcata	3720
cgtaggttcc ctgcaagaga tgagtgatat gatggagtta gttcgcgctg ggaaagtcc	3780
tccgatgccg gtaaaaactc ggccactgga cgctgcctgg gaaacccttg aggatctacg	3840
ccatggtaaa atcgtggggc gtgttgttct gaccccataa gaaggagata tacatatgac	3900
ccatcaatta agatcgcgcg atatcatcgc tctgggcttt atgacatttg cgttgttcgt	3960
cggcgcaggt aacattatth tccctccaat ggtcggcttg caggcaggcg aacacgtctg	4020
gactgcggca ttcggcttcc tcattactgc cgttggccta ccggtattaa cggtagtggc	4080
gctggcaaaa gttggcggcg gtgttgacag tctcagcacg ccaattggta aagtcgctgg	4140
cgtactgctg gcaacagttt gttacctggc ggtggggccg ctttttgcta cgccgcgtac	4200
agctaccgtt tcttttgaag tgggcattgc gccgctgacg ggtgattccg cgctgccgct	4260
gtttatttac agcctggtct atttcgctat cgttattctg gtttcgctct atccgggcaa	4320
gctgctggat accgtgggca acttccttgc gccgctgaaa attatcgcgc tggtcattcct	4380
gtctgttgcc gcaattatct ggccggcggg ttctatcagt acggcgactg aggcttatca	4440
aaacgctgcg ttttctaacg gcttcgtcaa cggctatctg accatggata cgctgggcgc	4500
aatggtgttt ggtatcgtha ttgttaacgc ggcgcgttct cgtggcgtha ccgaagcgcg	4560
tctgctgacc cgttataccg tctgggctgg cctgatggcg ggtgttggtc tgactctgct	4620
gtacctggcg ctgttccgtc tgggttcaga cagcgcgtcg ctggtcgatc agtctgcaaa	4680
cggcgcggcg atcctgcatg cttacgttca gcataccttt ggccggcgcg gtagcttcct	4740
gctggcggcg ttaatcttca tcgcctgcct ggtcacggcg gttggcctga cctgtgcttg	4800
tgagaattc ttcgcccagt acgtaccgct ctcttatcgt acgctgggtg ttatcctcgg	4860
cggcttctcg atggtggtgt ctaacctcgg cttgagccag ctgattcaga tctctgtacc	4920
ggtgctgacc gccatttatc cgccgtgtat cgactgggtt gtattaagtt ttacacgctc	4980
atggtggcat aattcgtccc gcgtgattgc tccgccgatg tttatcagcc tgctttttgg	5040
tattctcgac gggatcaagg catctgcatt cagcgatatc ttaccgtcct gggcgcagcg	5100
tttaccgctg gccgaacaag gtctggcgtg gttaatgcca acagtgggtga tgggtggttct	5160

ggccattatc tgggatcgtg cggcaggtcg tcaggtgacc tccagcgctc actaatacgc	5220
atggcatgga tgaccgatgg tagtgtgggc tctcccatg cgagagtagg gaactgccag	5280
gcatcaaata aaacgaaagg ctgagtcgaa agactggggc tttcgtttta tctgttgttt	5340
gtcggtgaaac gctctcctga gtaggacaaa tccgccggga gcggatttga acgttgcgaa	5400
gcaacggccc ggagggtggc gggcaggacg cccgccataa actgccaggc atcaaattaa	5460
gcagaaggcc atcctgacgg atggcctttt tgcgtggcca gtgccaagct tgcattgcgtg	5520
ccagctgcat taatgaagaa atcatgctgg aagaataaca gctcactcaa aggcggtagt	5580
acgggttttg ctgcccgaac acgggctggt ctgggtgttg tagtttgta tcagaatcgc	5640
agatccggct tcagccggtt tgccggctga aagcgctatt tcttcagaa ttgccatgat	5700
tttttcccca cgggaggcgt cactggctcc cgtgttgctg gcagcttga ttcgataagc	5760
agcatcgcct gtttcaggct gtctatgtgt gactgttgag ctgtaacaag ttgtctcagg	5820
tgttcaattt catgttctag ttgctttgtt ttactgggtt cacctgttct attaggtgtt	5880
acatgctgtt catctgttac attgtcgatc tgttcatggt gaacagcttt gaatgcacca	5940
aaaactcgta aaagctctga tgtatctatc ttttttacac cgttttcatc tgtgcatatg	6000
gacagttttc ctttgatat gtaacggtga acagttgttc tacttttggt ttgtagtctt	6060
gatgcttcac tgatagatac aagagccata agaacctcag atccttccgt atttagccag	6120
tatgttctct agtgtggttc gttgtttttg cgtgagccat gagaacgaac cattgagatc	6180
atacttactt tgcattgtcac tcaaaaattt tgcctcaaaa ctggtgagct gaatttttgc	6240
agttaaagca tcgtgtagtg tttttcttag tccgttatgt aggtaggaat ctgatgtaat	6300
ggttgttggt attttgtcac cattcatttt tatctgggtt ttctcaagtt cggttacgag	6360
atccatttgt ctatctagtt caacttgga aatcaacgta tcagtcgggc ggcctcgctt	6420
atcaaccacc aatttcatat tgctgtaagt gtttaaactt ttacttattg gtttcaaac	6480
ccattgggta agccttttaa actcatggta gttattttca agcattaaca tgaacttaaa	6540
ttcatcaagg ctaatctcta tatttgcctt gtgagttttc ttttgtgtta gttcttttaa	6600
taaccactca taaatcctca tagagtattt gttttcaaaa gacttaacat gttccagatt	6660
atattttatg aattttttta actggaaaag ataaggcaat atctcttcac taaaaactaa	6720
ttctaatttt tcgcttgaga acttggcata gtttgtccac tggaaaatct caaagccttt	6780

aaccaaagga	ttcctgattt	ccacagttct	cgtcatcagc	tctctggttg	ctttagctaa	6840
tacaccataa	gcattttccc	tactgatgtt	catcatctga	gcgtattggt	tataagtga	6900
cgataccgtc	cgttctttcc	ttgtaggggt	ttcaatcgtg	gggttgagta	gtgccacaca	6960
gcataaaatt	agcttggttt	catgctccgt	taagtcatag	cgactaatcg	ctagttcatt	7020
tgctttgaaa	acaactaatt	cagacataca	tctcaattgg	tctaggtgat	tttaatcact	7080
ataccaattg	agatgggcta	gtcaatgata	attactagtc	cttttccttt	gagttgtggg	7140
tatctgtaaa	ttctgctaga	cctttgctgg	aaaacttgta	aattctgcta	gaccctctgt	7200
aaattccgct	agacctttgt	gtgttttttt	tgtttatatt	caagtgggta	taatttatag	7260
aataaagaaa	gaataaaaaa	agataaaaag	aatagatccc	agccctgtgt	ataactcact	7320
actttagtca	gttccgcagt	attacaaaag	gatgtcgcaa	acgctgtttg	ctcctctaca	7380
aaacagacct	taaaacccta	aaggcttaag	tagcaccctc	gcaagctcgg	gcaaatacgct	7440
gaatattcct	tttgtctccg	accatcaggc	acctgagtcg	ctgtcttttt	cgtgacattc	7500
agttcgctgc	gctcacggct	ctggcagtga	atgggggtaa	atggcactac	aggcgccttt	7560
tatggattca	tgcaaggaaa	ctaccataa	tacaagaaaa	gcccgtcacg	ggcttctcag	7620
ggcgttttat	ggcgggtctg	ctatgtggtg	ctatctgact	ttttgctggt	cagcagttcc	7680
tgccctctga	ttttccagtc	tgaccacttc	ggattatccc	gtgacaggtc	attcagactg	7740
gctaatagcac	ccagtaaggc	agcggtatca	tcaacaggct	taccgtctt	actgtctttt	7800
ctacggggtc	tgacgctcag	tggaacgaaa	actcacgtta	agggattttg	gtcatgagat	7860
tatcaaaaag	gatcttcacc	tagatccttt	taaattaaaa	atgaagtttt	aatcaatct	7920
aaagtatata	tgagtaaact	tggtctgaca	gttaccaatg	cttaatcagt	gaggcaccta	7980
tctcagcgat	ctgtctattt	cgttcatcca	tagttgcctg	actccccgtc	gtgtagataa	8040
ctacgatacg	ggagggttta	ccatctggcc	ccagtgtctg	aatgataaccg	cgagaaccac	8100
gctcaccggc	tccagattta	tcagcaataa	accagccagc	cggaagggcc	gagcgcagaa	8160
gtggtcctgc	aactttatcc	gcctccatcc	agtctattaa	ttgttgccgg	gaagctagag	8220
taagtagttc	gccagttaat	agtttgcgca	acgttgttgc	cattgctaca	ggcatcgtgg	8280
gtgcacgctc	gtcgtttggt	atggcttcat	tcagctccgg	ttccaacga	tcaaggcgag	8340
ttacatgatc	ccccatgttg	tgcaaaaaag	cggttagctc	cttcggtcct	ccgatcgttg	8400

tcagaagtaa gttggccgca gtgttatcac tcatggttat ggcagcactg cataattctc	8460
ttactgtcat gccatccgta agatgctttt ctgtgactgg tgagtactca accaagtcac	8520
tctgagaata gtgtatgcgg cgaccgagtt gctcttgccc ggcgtcaata cgggataata	8580
ccgcgccaca tagcagaact ttaaaagtgc tcatcattgg aaaacgttct tcggggcgaa	8640
aactctcaag gatcttaccg ctgttgagat ccagttcgat gtaaccact cgtgcacca	8700
actgatcttc agcatctttt actttcacca gcgtttctgg gtgagcaaaa acaggaaggc	8760
aaaatgccgc aaaaaaggga ataaggcgca cacggaaatg ttgaatactc atactcttcc	8820
tttttcaata ttattgaagc atttatcagg gttattgtct catgagcgga tacatatttg	8880
aatgtattta gaaaaataaa caaatagggg ttccgcgcac atttccccga aaagtgccac	8940
ctgacgtcta agaaaccatt attatcatga cattaaccta taaaaatagg cgtatcacga	9000
ggccctttcg tctcgcgcgt ttcggtgatg acggtgaaaa cctctgacac atgcagctcc	9060
cggagacggc cacagcttgt ctgtaagcgg atgccgggag cagacaagcc cgtcagggcg	9120
cgtcagcggg tgttggcggg tgtcggggct ggcttaacta tgcggcatca gagcagattg	9180
tactgagagt gcaccatatg cgggtgtgaaa taccgcacag atgcgtaagg agaaaatacc	9240
gcatcaggcg ccattcgcca ttcaggctgc gcaactgttg ggaagggcga tcggtgcggg	9300
cctcttcgct attacgccag ctggcgaaaag ggggatgtgc tgcaaggcga ttaagttggg	9360
taacgccagg gttttccag tcacgacgtt gtaaaacgac ggccagtga ttcgttaaga	9420
cccactttca catttaagtt gtttttctaa tccgcatatg atcaattcaa ggccgaataa	9480
gaaggctggc tctgcacctt ggtgatcaaa taattcgata gcttgtcgta ataatggcgg	9540
catactatca gtagtaggtg tttccctttc ttcttttagcg acttgatgct cttgatcttc	9600
caatacgcaa cctaaagtaa aatgccccac agcgtgagt gcatataatg cattctctag	9660
tgaaaaacct tgttggcata aaaaggctaa ttgattttcg agagtttcat actgtttttc	9720
tgtaggccgt gtacctaaat gtacttttgc tccatcgcga tgacttagta aagcacatct	9780
aaaactttta gcgttattac gtaaaaaatc ttgccagctt tccccttcta aagggcaaaa	9840
gtgagtatgg tgcctatcta acatctcaat ggctaaggcg tcgagcaaag cccgcttatt	9900
ttttacatgc caatacaatg taggctgctc tacacctagc ttctgggcga gtttacgggt	9960
tgttaaacct tcgattccga cctcattaag cagctctaata gcgctgttaa tcactttact	10020

```
tttatctaat ctagacatca ttaattccta atttttgttg acactctatc attgatagag    10080
ttattttacc actccctatc agtgatagag aaaagtga                               10118
```

```
<210> 18
<211> 10134
<212> DNA
<213> Artificial Sequence
```

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

<400> 18						
gagctagata	ataaggacgg	cacaagatga	acatcttcaa	gaaaatggag	gaatttaatt	60
atgaacaact	ggtctacttc	tacgacagcg	aaacggaact	caaagggtatt	acctgtatac	120
acaacacaac	tttagggccg	gcattggggcg	gtacccgcct	ttggaactat	aactctgagg	180
aagatgccgt	tgaagacgta	atccgtctgg	ctcggggcat	gacttacaaa	gcggcttgcg	240
ccggctctgaa	tctgggcggc	ggtaaaaccg	tgctgatcgg	tgatgctaaa	aagattaaat	300
cagagtccta	cttcctgga	ctggggcgct	acgttcagtc	gctgaacggc	agatatatca	360
ccgcggaaga	cgtaaatact	tctacgaagg	atatggcata	cgttgctatg	gaaactgact	420
atgtggtagg	cctgggaggt	aaatccggca	accctagtcc	agttactgct	tacggtgcat	480
ttatgggtat	caaagcggcg	ctgatgaaaa	aatttgagga	tagctctatt	gaaggccgaa	540
ccttcgcagt	gcagggtgct	gggcagacgg	gttactatct	tatcgattac	ctcctaggca	600
acaacaagtt	caaagaaaag	gctaaaaaga	tttacttcac	cgaaattaac	gagagctata	660
tcgagcgtat	gaacaaagaa	catccggaag	ttgaatttat	ttccccggac	aaaatctact	720
cgctggaagt	agacgtcttc	gtgccctgcg	ccctgggcaa	aatcgttaat	gacaaaacta	780
tcgatgaatt	taagtgtccg	atcatcgcag	gtactgcaaa	caacgtactg	gaaagggag	840
cgcacggcaa	catgcttaaa	gaacgtggca	ttctttacgc	cccggactat	gtgatcaatg	900
ctgggtgggct	gatcaacgtt	taccacgagc	tgaacggtta	caataaagag	aacgctattc	960
tggaagtgga	attaatttat	gatcgcctac	tggaatatatt	caacatcgct	gattctctga	1020
acatcagcac	caatatcgct	gccaacgagt	tcgcggaaaa	acgtatcaag	caaattaagt	1080
ccttgaaaaa	caacttcatt	aaacgctaag	accattaata	aggaacattt	tatgtcgacg	1140

acaaccgttg gtgactactt gctgtatcgc ttaaacgaaa tcggcattga gcacctcttc	1200
ggagtgccag gtgattacaa tctgcaatth ctggatcatg taatcgacca ccctcagctg	1260
acttgggtcg gctgcactaa cgaacttaac gctgcctacg cagctgatgg ttatgcgcgt	1320
tgtcgtccgg ctgcggcact gctgaccacc ttcggggttg gcgaactgag cgctattaat	1380
ggcatcgag gttcctacgc ggagtatctg ccggtaatac atatcgttgg tgcaccgagt	1440
ctatcagccc agcagcaggg cgacctgatt caccactctc ttggcgaagg tgatthttcc	1500
agcttcctga ggatgtccca accggtgtct gttgcgcagg ctgctctgac tcctgataac	1560
gcatgcaagg aaatcgaccg cgtactggcg gaagtcctca ttcagcgtcg tcccggtac	1620
ctgctgctgt ctaccgacgt ggctgctgcg ccggcggctc tgccacaaag cactctttct	1680
ttgccgaccg ccccggtatca tcgcgcagtt ctggctgctt tcagcgacgc tgctgagcag	1740
atgctggctc aggccaaaag cgtctctcta ctggcggact ttctggctga tcgtttcggt	1800
gttactcgag cactggccgc gtggcttcag caggttccgc taccgcacgc cactctgtta	1860
atgggtaaag gcgttctgag tgaacagcaa ccagggttcg tgggtacctg cgctggtgcg	1920
gcatctatcg attcgacgcg tggcgcaatc gaagaagctg gggtaattat cggagtggga	1980
gtagatthtt ccgacactat cacagcaggc ttctcgcagc agatcgacgc ccgccgthtt	2040
atagacattc aacccttctt ctctcgtatt ggcgatcgcc agthttgatca cttgccgatg	2100
caggctgccg tcgcagccct gcatcaactg tgtcttcgth atcagcagca gtggtctatc	2160
accgctccta gcccgcctgc actgccgccg gctgctggta gcgagctgtc ccagaacgca	2220
ttctggcagg cgatgcagaa cttcatccgc cctggggacc tgttgggtggc cgaccaaggt	2280
actgcggcgt tcggcgacgc ggcgctgcgc ttaccgcaga attgccagct gcttgtgcag	2340
ccgctgtggg gctcaatcgg ttacagtctg ccggccacct ttggtgctca gacggcagat	2400
acagagcgtc gtgtaatcct aatcattggc gatggthtcag cgcaattaac tattcaggaa	2460
ctttccagta tgatgcgtga cggcttgaaa cctatcatct ttctcctgaa caacaacggt	2520
tacaccgttg aacgggcgat tcacggcgcg gagcaacgth ataacgatat cgctgcttgg	2580
aattggaccc aactgccccg ggcgctgagt gthcattgcc cagcgacagc ctggcgagtc	2640
gttgaaacgg tgcagctgac cgacgtaatg aaagtcacg ctgcttctcc gcgtctgagc	2700
ttggtagaag ttgttctgcc tgcaatggat gtcccaccgc tgctgcaagc agtgagtgcc	2760

gctctgaacc agcgcaactc ctcttaattg gataaagaat aaggaggat tttatgacag	2820
cggagcagca aaatggggta tccgactcac gccgtttcga atttcaggaa tttggtggcc	2880
ctatcgcccc acagacctat cagctccccg caccggctag cgatgaagtt ttgttaaagg	2940
tgaactactg cgggtgtctgt cacagtgatg ttcattctca cgacggctac ttcgagctgg	3000
gtggcgataa acgtctgaac ttcgctatgc cgctgccgct gacgctgggt cacgaagtaa	3060
ttggcaccgt tgtggctgtc ggcgaccagg ttactgggtg aaaaccgggg gaccagcgac	3120
tgatctatcc gtggataggt tgcggaaaat gcggcgctg tcaaaaagga gaagaaaacc	3180
tgtgcgttac tcctgcacat ctgggcgtga acaagccggg cggttacgct gatcacatcg	3240
ttgtaccca ttctcgctac cttctggaca tttcgggtct gaaccgggt gatgccgcta	3300
ccctcgctg ctccggcctg accactttca gcgcgatcaa caaagtgtt ccgcttgacg	3360
atgaccagtg gattgttggt atcggttggt gtggcctcgg ccagatggcg ctgcgtatcc	3420
tgcaagctat gggaattggc aatgttatcg gtattgacct gtctgaagag aaacggaaac	3480
tggctcatga aagcggtgca cgtcactcct tcgatccaaa cactccgaag ctgaaccgcg	3540
tggtcgccga aacctgccc ggtacggtac aggccgcgtt agactttgtg ggcaatgagc	3600
aaactgctca gctggcactg tctctgcttg gaaaagggtg caaatatgtt cctgtcgggc	3660
tgcacggcgg cgagctgcgt taccattgc cgatcatcac gaacaaagct gtaagtatca	3720
tcggttctta cgttggtacc ctgaaagaac tggaagactt agttgctttc gccaaggaaa	3780
agaatctgcc gccaatcat attgaacacc gcccgctgga atcggcggct caggccgtag	3840
aggacctgga aaaaggacag gttgctgggc gtgttatcct ggatgcaggt aactaagaag	3900
gagatataca tatgacctat caattaagat cgcgcgatat catcgctctg ggctttatga	3960
catttgcggt gttcgtcggc gcaggttaaca ttattttccc tccaatggtc ggcttgacgg	4020
caggcgaaca cgtctggact gcggcattcg gcttcctcat tactgccgtt ggctaccgg	4080
tattaacggg agtggcgctg gcaaaagttg gcggcggtgt tgacagtctc agcacgcaa	4140
ttggtaaagt cgctggcgta ctgctggcaa cagtttggtta cctggcggtg gggccgcttt	4200
ttgctacgcc gcgtacagct accgtttctt ttgaagtggg cattgcgccg ctgacgggtg	4260
attccgcgct gccgctgttt atttacagcc tgggtctatct cgctatcggt attctggttt	4320
cgctctatcc gggcaagctg ctggataccg tgggcaactt ccttgcgccg ctgaaaatta	4380

tcgcgctggt catcctgtct gttgccgcaa ttatctggcc ggcgggttct atcagtacgg	4440
cgactgaggc ttatcaaaac gctgcgtttt ctaacggctt cgtcaacggc tatctgacca	4500
tggatacgct gggcgcaatg gtgttttgta tcgttattgt taacgcggcg cgttctcgtg	4560
gcgttaccga agcgcgtctg ctgaccggtt ataccgtctg ggctggcctg atggcgggtg	4620
ttggtctgac tctgctgtac ctggcgctgt tccgtctggg ttcagacagc gcgtcgctgg	4680
tcgatcagtc tgcaaacggg gggcgcatcc tgcattgctta cgttcagcat acctttggcg	4740
gcggcggttag cttcctgctg gggcggttaa tcttcatcgc ctgcctgggtc acggcggttg	4800
gcctgacctg tgcttgtgca gaattcttcg cccagtacgt accgctctct tatcgtacgc	4860
tggtgtttat cctcggcggc ttctcgatgg tgggtgtctaa cctcggcttg agccagctga	4920
ttcagatctc tgtaccggtg ctgaccgcca tttatccgcc gtgtatcgca ctggttgat	4980
taagttttac acgctcatgg tggcataatt cgtcccgcgt gattgctccg ccgatgttta	5040
tcagcctgct ttttggtatt ctgcacggga tcaaggcatc tgcattcagc gatattctac	5100
cgtcctgggc gcagcgttta ccgctggccg aacaaggctt ggcgtggtta atgccaacag	5160
tggtgatggt ggttctggcc attatctggg atcgtgcggc aggtcgtcag gtgacctcca	5220
gcgctcacta atacgcatgg catggatgac cgatggtagt gtgggctctc cccatgcgag	5280
agtagggaac tgccaggcat caaataaaac gaaaggctca gtcgaaagac tgggcctttc	5340
gttttatctg ttgtttgtcg gtgaacgctc tcctgagtag gacaaatccg ccgggagcgg	5400
atttgaacgt tgcgaagcaa cggcccggag ggtggcgggc aggacgcccg ccataaactg	5460
ccaggcatca aattaagcag aaggccatcc tgacggatgg cctttttgcg tggccagtgc	5520
caagcttgca tgcgtgccag ctgcattaat gaagaaatca tgctggaaga ataacagctc	5580
actcaaaggc ggtagtacgg gttttgctgc ccgcaaacgg gctgttctgg tgttgctagt	5640
ttgttatcag aatcgcagat ccggcttcag ccggtttgcc ggctgaaagc gctatttctt	5700
ccagaattgc catgatTTTT tccccacggg aggcgtcact ggctcccgtg ttgtcggcag	5760
ctttgattcg ataagcagca tcgcctgttt caggctgtct atgtgtgact gttgagctgt	5820
aacaagttgt ctcagggtgt caatttcatt ttctagttgc tttgttttac tggtttcacc	5880
tgttctatta ggtgttacat gctgttcatt tgttacattg tcgatctggt catggtgaac	5940
agctttgaat gcacaaaaa ctcgtaaaag ctctgatgta tctatctttt ttacaccgtt	6000

ttcatctgtg catatggaca gttttccctt tgatatgtaa cggatgaacag ttgttctact	6060
tttgtttgtt agtcttgatg cttcactgat agatacaaga gccataagaa cctcagatcc	6120
ttccgtattt agccagtatg ttctctagtg tggttcgttg tttttgcgtg agccatgaga	6180
acgaaccatt gagatcatac ttactttgca tgtcactcaa aaattttgcc tcaaaactgg	6240
tgagctgaat ttttgcagtt aaagcatcgt gtagtgtttt tcttagtccg ttatgtaggt	6300
aggaatctga tgtaatggtt gttgggtatth tgtcaccatt catttttatc tggttgttct	6360
caagttcggg tacgagatcc atttgtctat ctagttcaac ttggaaaatc aacgtatcag	6420
tcgggcgggc tcgcttatca accaccaatt tcatattgct gtaagtgttt aaatctttac	6480
ttattggttt caaaacccat tggttaagcc ttttaaacctc atggtagtta ttttcaagca	6540
ttaacatgaa cttaaattca tcaaggctaa tctctatatt tgccttgtga gttttctttt	6600
gtgtagttc ttttaataac cactcataaa tcctcataga gtatttgttt tcaaaagact	6660
taacatgttc cagattatat tttatgaatt tttttaactg gaaaagataa ggcaatatct	6720
cttcactaaa aactaattct aatttttcgc ttgagaactt ggcatagttt gtccactgga	6780
aaatctcaaa gcctttaacc aaaggattcc tgatttccac agttctcgtc atcagctctc	6840
tggttgcttt agctaataca ccataagcat tttccctact gatgttcac atctgagcgt	6900
attggttata agtgaacgat accgtccgtt ctttccttgt agggttttca atcgtggggg	6960
tgagtagtgc cacacagcat aaaattagct tggtttcatg ctccgttaag tcatagcgac	7020
taatcgctag ttcatthgtt ttgaaaacaa ctaattcaga catacatctc aattgggtcta	7080
ggtagtttta atcactatac caattgagat gggctagtca atgataatta ctagtccttt	7140
tcctttgagt tgtgggtatc tgtaaatctt gctagacctt tgctggaaaa cttgtaaat	7200
ctgctagacc ctctgtaaat tccgctagac ctttgttgtt tttttttgtt tatattcaag	7260
tggttataat ttatagaata aagaaagaat aaaaaagat aaaaagaata gatcccagcc	7320
ctgtgtataa ctactactt tagtcagttc cgcagtatta caaaaggatg tcgcaaacgc	7380
tgtttgctcc tctacaaaac agaccttaaa accctaaagg cttaaagtagc accctcgcaa	7440
gctcgggcaa atcgctgaat attccttttg tctccgacca tcaggcacct gagtcgctgt	7500
ctttttcgtg acattcagtt cgctgcgtc acggctctgg cagtgaatgg gggtaaatgg	7560
cactacaggc gccttttatg gattcatgca aggaaactac ccataatata agaaaagccc	7620

gtcacgggct tctcagggcg ttttatggcg ggtctgctat gtgggtgctat ctgacttttt	7680
gctgttcagc agttcctgcc ctctgatttt ccagtctgac cacttcggat tatcccgatga	7740
caggctattc agactggcta atgcaccag taaggcagcg gtatcatcaa caggcttacc	7800
cgtcttactg tcttttctac ggggtctgac gctcagtggga acgaaaactc acgttaaggg	7860
attttgggtca tgagattatc aaaaaggatc ttcacctaga tcctttttaa ttaaaaatga	7920
agttttaaat caatctaaag tatatatgag taaacttggg ctgacagtta ccaatgctta	7980
atcagtgagg cacctatctc agcgatctgt ctatttcgtt catccatagt tgcctgactc	8040
cccgctcgtg agataactac gatacgggag ggcttaccat ctggccccag tgctgcaatg	8100
ataccgagc aaccacgctc accggctcca gatttatcag caataaacca gccagccgga	8160
agggccgagc gcagaagtgg tcctgcaact ttatccgcct ccatccagtc tattaattgt	8220
tgccgggaag ctagagtaag tagttcgcca gttaatagtt tgcgcaacgt tgttgccatt	8280
gctacaggca tcgtggtgtc acgctcgtcg tttggtatgg cttcattcag ctccggttcc	8340
caacgatcaa ggcgagttac atgatcccc atgttgtgca aaaaagcggg tagctccttc	8400
ggtcctccga tcgttgtcag aagtaagttg gccgcagtgt tatcactcat gggtatggca	8460
gcactgcata attctcttac tgtcatgcca tccgtaagat gcttttctgt gactggtgag	8520
tactcaacca agtcattctg agaatagtgt atgcggcgac cgagttgctc ttgcccggcg	8580
tcaatacggg ataataccgc gccacatagc agaactttaa aagtgtcat cattggaaaa	8640
cgttcttcgg ggcgaaaact ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa	8700
cccactcgtg caccgaactg atcttcagca tcttttactt tcaccagcgt ttctgggtga	8760
gcaaaaacag gaaggcaaaa tgccgcaaaa aagggaataa gggcgacacg gaaatgttga	8820
atactcatac tcttcctttt tcaatattat tgaagcattt atcagggtta ttgtctcatg	8880
agcggataca tatttgaatg tatttagaaa aataaacaaa taggggttcc gcgcacattt	8940
ccccgaaaag tgccacctga cgtctaagaa accattatta tcatgacatt aacctataaa	9000
aataggcgta tcacgaggcc ctttcgtctc gcgcgtttcg gtgatgacgg tgaaaacctc	9060
tgacacatgc agctcccgga gacggtcaca gcttgtctgt aagcggatgc cgggagcaga	9120
caagcccgtc agggcgcgctc agcgggtgtt ggcgggtgtc ggggctggct taactatgcg	9180
gcatcagagc agattgtact gagagtgcac catatgcggg gtgaaatacc gcacagatgc	9240

gtaaggagaa aataccgcat caggcgccat tcgccattca ggctgcgcaa ctgttgggaa	9300
gggcgatcgg tgcgggcctc ttcgctatta cgccagctgg cgaaaggggg atgtgctgca	9360
aggcgattaa gttgggtaac gccagggttt tcccagtcac gacgttgtaa aacgacggcc	9420
agtgaattcg ttaagacca ctttcacatt taagttgttt ttctaattccg catatgatca	9480
attcaaggcc gaataagaag gctggctctg caccttggtg atcaaataat tcgatagctt	9540
gtcgtaataa tggcggcata ctatcagtag taggtgtttc cttttcttct ttagcgactt	9600
gatgctcttg atcttccaat acgcaaccta aagtaaaatg ccccacagcg ctgagtgcac	9660
ataatgcatt ctctagtga aaaccttggt ggcatataaa ggctaattga ttttcgagag	9720
tttcatactg tttttctgta ggccgtgtac ctaaattgtac ttttgctcca tcgcatgac	9780
ttagtaaagc acatctaaaa ctttttagcgt tattacgtaa aaaatcttgc cagctttccc	9840
cttctaaagg gcaaaagtga gtatggtgcc tatctaact ctcaatggct aaggcgtcga	9900
gcaaagcccg cttatTTTTT acatgccaat acaatgtagg ctgctctaca cctagcttct	9960
gggcgagttt acgggttggt aaaccttcga ttccgacctc attaagcagc tctaattgcg	10020
tgtaatacac ttacttttta tctaattctag acatcattaa ttcctaattt ttgttgacac	10080
tctatcattg atagagttat ttaccactc cctatcagtg atagagaaaa gtga	10134

<210> 19

<211> 10142

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 19

aaacaggcca aacgtaggct ccggaattta aggaggtatt attatggtag agatccaggc	60
tttgccgga acttccattt ttgggcaa atcgagaccac cagcatgaac aggtggtctt	120
ctgccacgat cacgaaaccg gcctccgtgc gataatcggg attcataaca cggttcttgg	180
ccccgcctta ggtggaactc gcatgtggca ctatgctacc gaggcagaag cgctgaatga	240
cggtctgcgt ctgtctcgcg gtatgacctc caaggctgct atctcgggcc tgaacctggg	300
tggcggtaaa gcagtaatca ttggggatgc caaaacaatc aaaaccgaag cgctgctgcg	360
gaaattcggc agattcgtgc agaacctgaa tggtaaatac atcactgctg aagacgttaa	420

catgactaca aaggatatgg agtacattag gatggaaacc aaacacgtcg ctggcttacc	480
tgaaagtatg ggtggaagcg gtgacccgtc accggtaact gcatatggta cgtacatggg	540
catgaaagcg gcggccaaaa aggcgtttgg ctctgattcc ctggctggca aacgtatcgc	600
tgttcaaggt gtgggtcatg ttggcactta tctgcttgag catttgacca aagaaggtgc	660
tcagatttg ctgactgact actataagga acgtgccgag gaagcaggcg cgcgttttgg	720
cgcacagggtt gttggcctgg acgatatcta cgatcaagag gtcgacattt actctccatg	780
tgctctcggg gctaccatca acgatgacac tatcgatcgc ctgcgttgcg ctgtttagc	840
cggttgcgca aacaaccagc tgaaagaaga aaacgtccac ggtccggcgc tggttgagcg	900
cgggatatga tacgccccag acttctgat caatgcaggg ggctgatta acgtgtatag	960
cgaagttaca ggggtctacc gtcagggggc tttactcag accgaaaaa tctatgacta	1020
cacactcaa gttctggaag aagccgcggc tgaaggctctg caccgcagc aggctgcgat	1080
ccgtcaggcg gaacaacgca tcgctgcaat tggtaagggtg aaaagcacct actaatgaga	1140
aataaggagg atttcatgtc gacgacaacc gttggtgact acttgctgta tcgcttaaac	1200
gaaatcggca ttgagcacct cttcggagtgc ccagggtgatt acaatctgca atttctggat	1260
catgtaatcg accaccctca gctgacttgg gtcggctgca ctaacgaact taacgctgcc	1320
tacgcagctg atggttatgc gcgttgtcgt ccggctgcgg cactgctgac caccttcggg	1380
gttggcgaac tgagcgctat taatggcatc gcaggttcct acgcggagta tctgccggta	1440
atacatatcg ttggtgcacc gagtctatca gccagcagc agggcgacct gattcaccac	1500
tctcttggcg aaggtgattt ttccagcttc ctgaggatgt cccaaccggg gtctgttgcg	1560
caggctgctc tgactcctga taacgcatgc aaggaaatcg accgcgtact ggcggaagtc	1620
ctcattcagc gtcgtcccgg ctacctgtg ctgtctaccg acgtggctgc tgcgccggcg	1680
gctctgccac aaagcactct ttctttgccg accgccccgg atcatcgcg agttctggct	1740
gctttcagcg acgctgctga gcagatgctg gctcaggcca aaagcgtctc tctactggcg	1800
gactttctgg ctgatcgttt cgggtgttact cgagcactgg ccgcgtggct tcagcagggt	1860
ccgctaccgc acgccactct gttaatgggt aaaggcgttc tgagtgaaca gcaaccaggg	1920
ttcgtgggta cctacgtgg tgcggcatct atcgattcga cgcgtggcgc aatcgaagaa	1980
gctggggtaa ttatcggagt gggagttaga ttttccgaca ctatcacagc aggcttctcg	2040

cagcagatcg acgcccgccg ttttatagac attcaaccct tcttctctcg tattggcgat	2100
cgccagtttg atcacttgcc gatgcaggct gccgtcgag ccctgcatca actgtgtctt	2160
cgttatcagc agcagtggtc tatcaccgct cctagcccg cctgactgcc gccggctgct	2220
ggtagcgagc tgtcccagaa cgcattctgg caggcgatgc agaacttcat ccgccctggg	2280
gacctgttgg tggccgacca aggtactgcg gcgttcggcg cagcggcgct gcgcttaccg	2340
cagaattgcc agctgcttgt gcagccgctg tggggctcaa tcggttacag tctgccggcc	2400
acctttggtg ctgagacggc agatacagag cgctcgtgtaa tcctaatacat tggcgatggt	2460
tcagcgcaat taactattca ggaactttcc agtatgatgc gtgacggctt gaaacctatc	2520
atctttctcc tgaacaacaa cggttacacc gttgaacggg cgattcacgg cgcggagcaa	2580
cgttataacg atatcgctgc ttggaattgg acccaactgc cccaggcgct gagtgttcat	2640
tgcccagcgc agagctggcg agtcgttgaa acgggtgcagc tgaccgacgt aatgaaagtc	2700
atcgctgctt ctccgcgtct gagcttggtg gaagttgttc tgcctgcaat ggatgtccca	2760
ccgctgctgc aagcagtgag tgccgctctg aaccagcgca actcctctta atttttaaag	2820
taacacacga gctaagaact aaggaggatt tatgcgcagc atgcagtttg atgagtacgg	2880
tgcaccctg aaagcgttct catatgaaga cccgaccccg caagggaagg aagtagtcgt	2940
taggatcgaa gcctgtggtg tgtgccactc tgatattcat cttcacgagg gctacttcga	3000
catgggcggg ggcaataaag ctgatgttac tcgtgctcgc gaactccctt ttacattggg	3060
tcatgaaatc gttggcgaag tggtagcaac tggaccagggt gtcaccggcg ctaaaccggg	3120
cgacaaacgt attgtgtacc cgtggatcgg gtgcggcgac tgcccgaat gcaacagtgg	3180
tgaggatcag tcctgtgcgc gtccacgtaa cctgggtggt cacgttgacg gtggctattc	3240
gacgcacgta aagataccgg acgaaaaatt cctgttcgcc tacgatggta ttcctactga	3300
gtagcggga acctatgctt gcagcggcat caccgcttat ggtgcactga tgaaagcaaa	3360
ggaagcggct gaaagatctg gctacatcgg tctgattggc gctgggtggcg ttggcatggc	3420
tggctctgat ctggccaaag cagcgatcgg ggctaaaact gtagtctttg atatcgacga	3480
cgcaaaactg gaagctgcga cccgtgccgg ggcggattac gtgttcaact ccggtgcaaa	3540
agaaacacgc aaggaagtta tgaaactaac gaatggtggc ctgtctggtg ctgttgattt	3600
cgttggcagc gataaaagcg ctctgttttg aatcaacgcc ttgggtcaga acggcgtgct	3660

ggtcataatt ggactgttcg gtggcgctat gactgttccg gtaccctgtg tcccgtgaa	3720
agggatcacc gtacgtggct catacgtagg ttccctgcaa gagatgagtg atatgatgga	3780
gtagttcgc gctgggaaag ttctctcgat gccggtaaaa actcggccac tggacgctgc	3840
ctgggaaacc cttgaggatc tacgccatgg taaaatcgtg ggccgtgttg ttctgacccc	3900
ataagaagga gatatacata tgacccatca attaagatcg cgcgatatca tcgctctggg	3960
ctttatgaca tttgcgttgt tcgtcggcgc aggtaacatt attttccctc caatggtcgg	4020
cttgaggca ggcgaacacg tctggactgc ggcatcggc ttctcatta ctgccgttg	4080
cctaccggtg ttaacggtag tggcgctggc aaaagtggc ggccgtgttg acagtctcag	4140
cacgccaatt ggtaaagtcg ctggcgctact gctggcaaca gtttggtacc tggcggtggg	4200
gccgcttttt gctacgccgc gtacagctac cgtttctttt gaagtgggca ttgcgccgct	4260
gacgggtgat tccgcgctgc cgctgtttat ttacagcctg gtctatttcg ctatcgttat	4320
tctggtttcg ctctatccgg gcaagctgct ggataccgtg ggcaacttcc ttgcgccgct	4380
gaaaattatc gcgctggtca tcctgtctgt tgccgcaatt atctggccgg cgggttctat	4440
cagtacggcg actgaggctt atcaaaacgc tgcgttttct aacggcttcg tcaacggcta	4500
tctgaccatg gatacgtgg gcgcaatggt gtttggtatc gttattgtta acgcggcgcg	4560
ttctcgtggc gttaccgaag cgcgtctgct gaccggttat accgtctggg ctggcctgat	4620
ggcgggtgtt ggtctgactc tgctgtacct ggcgtgttc cgtctgggtt cagacagcgc	4680
gtcgctggtc gatcagtctg caaacgggtgc ggcgatcctg catgcttacg ttcagcatac	4740
ctttggcggc ggcggtagct tcctgctggc ggcgttaatc ttcacgcct gcctggtcac	4800
ggcggttggc ctgacctgtg cttgtgcaga attcttcgcc cagtacgtac cgctctctta	4860
tcgtacgctg gtgtttatcc tcggcggctt ctcgatgggtg gtgtctaacc tcggcttgag	4920
ccagctgatt cagatctctg taccggtgct gaccgccatt tatccgccgt gtatcgact	4980
ggttgtatta agttttacac gtcctggtg gcataattcg tcccgcgtga ttgctccgcc	5040
gatgtttatc agcctgcttt ttggtattct cgacgggatc aaggcatctg cattcagcga	5100
tatcttaccg tcctgggcgc agcgtttacc gctggccgaa caaggctctg cgtggttaat	5160
gccaacagtg gtgatggtgg ttctggccat tatctgggat cgtgcggcag gtcgtcaggt	5220
gacctccagc gctcactaat acgcatggca tggatgaccg atggtagtgt gggctctccc	5280

catgcgagag tagggaactg ccaggcatca aataaaacga aaggctcagt cgaaagactg	5340
ggccttttcgt tttatctgtt gtttgtcggg gaacgctctc ctgagtagga caaatccgcc	5400
gggagcggat ttgaacgttg cgaagcaacg gcccgagggg tggcgggcag gacgcccgcc	5460
ataaactgcc aggcatcaaa ttaagcagaa ggccatcctg acggatggcc tttttgcgtg	5520
gccagtgcc aagcttgcag cgtgccagct gcattaatga agaaatcatg ctggaagaat	5580
aacagctcac tcaaaggcgg tagtacgggt tttgctgccc gcaaacgggc tgttctggtg	5640
ttgctagttt gttatcagaa tcgcagatcc ggcttcagcc ggtttgccgg ctgaaagcgc	5700
tatttcttcc agaattgcc a tgattttttc cccacgggag gcgtcactgg ctcccgtgtt	5760
gtcggcagct ttgattcgat aagcagcatc gcctgtttca ggctgtctat gtgtgactgt	5820
tgagctgtaa caagttgtct caggtgttca atttcatgtt ctagttgctt tgttttactg	5880
gtttcacctg ttctattagg tgttacatgc tgttcacatg ttacattgtc gatctgttca	5940
tggatgaacag ctttgaatgc accaaaaact cgtaaaagct ctgatgtatc tatctttttt	6000
acaccgtttt catctgtgca tatggacagt tttccctttg atatgtaacg gtgaacagtt	6060
gttctacttt tgtttgttag tcttgatgct tctactgatag atacaagagc cataagaacc	6120
tcagatcctt ccgtatttag ccagtatgtt ctctagtgtg gttcgttggt tttgcgtgag	6180
ccatgagaac gaaccattga gatcatactt actttgcatg tctactcaaaa attttgcctc	6240
aaaactgggtg agctgaattt ttgcagttaa agcatcgtgt agtgtttttc ttagtccgtt	6300
atgtaggtag gaatctgatg taatggttgt tggatatttg tcaccattca tttttatctg	6360
gttgttctca agttcggtta cgagatccat ttgtctatct agttcaactt ggaaaatcaa	6420
cgtatcagtc gggcggcctc gcttatcaac caccaatttc atattgctgt aagtgtttaa	6480
atctttactt attggtttca aaaccattg gttaagcctt ttaaactcat ggtagttatt	6540
ttcaagcatt aacatgaact taaattcatc aaggctaata tctatatattg ccttgtgagt	6600
tttcttttgt gttagttctt ttaataacca ctcataaatc ctcatagagt atttgttttc	6660
aaaagactta acatgttcca gattatatatt tatgaatttt ttttaactgga aaagataagg	6720
caatatctct tactaaaaa ctaattctaa tttttcgctt gagaacttgg catagtttgt	6780
ccactggaaa atctcaaagc ctttaaccaa aggattcctg atttccacag ttctcgtcat	6840
cagctctctg gttgcttttag ctaatacacc ataagcattt tccctactga tgttcacat	6900

ctgagcgtat tggttataag tgaacgatac cgtccgttct ttcctttag ggTTTTcaat	6960
cgtgggggtt agtagtgcca cacagcataa aattagcttg gtttcatgct ccgttaagtc	7020
atagcgacta atcgctagtt catttgcttt gaaaacaact aattcagaca tacatctcaa	7080
ttggtctagg tgattttaat cactatacca attgagatgg gctagtcaat gataattact	7140
agtccttttc ctttgagttg tgggtatctg taaattctgc tagaccttg ctggaaaact	7200
tgtaaattct gctagaccct ctgtaaattc cgctagacct ttgtgtgttt tttttgttta	7260
tattcaagtg gttataatTT atagaataaa gaaagaataa aaaaagataa aaagaataga	7320
tcccagccct gtgtataact cactacttta gtcagttccg cagtattaca aaaggatgtc	7380
gcaaacgctg tttgctcctc tacaaaacag accttaaaac cctaaaggct taagtagcac	7440
cctcgcaagc tcgggcaaatt cgctgaatat tccttttgtc tccgaccatc aggcacctga	7500
gtcgctgtct ttttcgtgac attcagttcg ctgcgctcac ggctctggca gtgaatgggg	7560
gtaaatggca ctacaggcgc cttttatgga ttcattgcaag gaaactaccc ataatacaag	7620
aaaagcccgt cacgggcttc tcagggcgtt ttatggcggg tctgctatgt ggtgctatct	7680
gactttttgc tgttcagcag ttcctgccct ctgattttcc agtctgacca cttcggatta	7740
tcccgtgaca ggtcattcag actggctaatt gcaccagta aggcagcggg atcatcaaca	7800
ggcttaccgg tcttactgtc ttttctacgg ggtctgacgc tcagtggaac gaaaactcac	7860
gttaagggat tttggtcatt agattatcaa aaaggatctt cacctagatc cttttaaat	7920
aaaaatgaag ttttaaatca atctaaagta tatatgagta aacttggtct gacagttacc	7980
aatgcttaat cagtgaggca cctatctcag cgatctgtct atttcgttca tccatagttg	8040
cctgactccc cgtcgtgtag ataactacga tacgggaggg cttaccatct ggccccagt	8100
ctgcaatgat accgcgagaa ccacgctcac cggctccaga tttatcagca ataaaccagc	8160
cagccggaag ggccgagcgc agaagtggc ctgcaacttt atccgcctcc atccagtcta	8220
ttaattgttg ccgggaagct agagtaagta gttcgccagt taatagtttg cgcaacgttg	8280
ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt tggatatggct tcattcagct	8340
ccggttcca acgatcaagg cgagttacat gatccccat gttgtgcaaa aaagcggtta	8400
gctccttcgg tcctccgacg gttgtcagaa gtaagttggc cgcagtgtta tcactcatgg	8460
ttatggcagc actgcataat tctcttactg tcatgccatc cgtaagatgc ttttctgtga	8520

ctggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg agttgctctt	8580
gccccggcgtc aatacgggat aataccgcgc cacatagcag aactttaaaa gtgctcatca	8640
ttggaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg agatccagtt	8700
cgatgtaacc cactcgtgca cccaactgat cttcagcatc ttttactttc accagcgttt	8760
ctgggtgagc aaaaacagga aggcaaaatg ccgcaaaaaa gggaataagg gcgacacgga	8820
aatgttgaat actcatactc ttcctttttc aatattattg aagcatttat cagggttatt	8880
gtctcatgag cggatacata tttgaatgta tttagaaaaa taaacaaata ggggttccgc	8940
gcacatttcc ccgaaaagtg ccacctgacg tctaagaaac cattattatc atgacattaa	9000
cctataaaaa taggcgtatc acgaggccct ttcgtctcgc gcgtttcggg gatgacggtg	9060
aaaacctctg acacatgcag ctcccggaga cggtcacagc ttgtctgtaa gcggatgccg	9120
ggagcagaca agcccgtcag ggcgcgtcag cgggtgttgg cgggtgtcgg ggctggctta	9180
actatgcggc atcagagcag attgtactga gagtgcacca tatgcggtgt gaaataccgc	9240
acagatgcgt aaggagaaaa taccgcatca ggccgcatc gccattcagg ctgcgcaact	9300
gttggaagg gcgatcgggt cgggcctctt cgctattacg ccagctggcg aaagggggat	9360
gtgctgcaag gcgattaagt tgggtaacgc cagggttttc ccagtcacga cgttgtaaaa	9420
cgacggccag tgaattcgtt aagaccact ttcacattta agttgttttt ctaatccgca	9480
tatgatcaat tcaaggccga ataagaaggc tggctctgca ccttggtgat caaataattc	9540
gatagcttgt cgtaataatg gcggcatact atcagtagta ggtgtttccc tttcttcttt	9600
agcgacttga tgctcttgat cttccaatac gcaacctaaa gtaaaatgcc ccacagcgct	9660
gagtgcatat aatgcattct ctagtgaaaa acctgttggt cataaaaagg ctaattgatt	9720
ttcgagagtt tcatactgtt tttctgtagg ccgtgtacct aaatgtactt ttgctccatc	9780
gcgatgactt agtaaagcac atctaaaact tttagcgtaa ttacgtaaaa aatcttgcca	9840
gctttcccct tctaaagggc aaaagtgagt atgggtgccta tctaacatct caatggctaa	9900
ggcgtcgagc aaagcccgt tatttttttac atgccaatac aatgtaggct gctctacacc	9960
tagcttctgg gcgagtttac gggttgttaa accttcgatt ccgacctcat taagcagctc	10020
taatgcgctg ttaatcactt tacttttatc taatctagac atcattaatt cctaattttt	10080
gttgacactc tatcattgat agagttatct taccactccc tatcagtgat agagaaaagt	10140

<210> 20

<211> 10158

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 20

aaacaggcca aacgtaggct ccggaattta aggaggtatt attatggtag agatccaggc	60
tttgccggaa acttccattt ttgggcaaat cgcagaccac cagcatgaac aggtgggtctt	120
ctgccacgat cacgaaaccg gcctccgtgc gataatcggg attcataaca cggttcttgg	180
ccccgcctta ggtggaactc gcatgtggca ctatgctacc gaggcagaag cgctgaatga	240
cggttctgcgt ctgtctcgcg gtatgacctt caaggctgct atctcgggcc tgaacctggg	300
tggcggtaaa gcagtaatca ttgggggatgc caaaacaatc aaaaccgaag cgctgctgcg	360
gaaattcggc agattcgtgc agaacctgaa tggtaaatac atcactgctg aagacgttaa	420
catgactaca aaggatatgg agtacattag gatggaaacc aaacacgtcg ctggcttacc	480
tgaaagtatg ggtggaagcg gtgacctgtc accggtaact gcatatggta cgtacatggg	540
catgaaagcg gcggccaaaa aggcgttttg ctctgattcc ctggctggca aacgtatcgc	600
tgttcaagggt gtgggtcatg ttggcactta tctgcttgag catttgacca aagaagggtc	660
tcagattgtg ctgactgact actataagga acgtgccgag gaagcaggcg cgcgttttgg	720
cgcacagggt gttggcctgg acgatattct cgatcaagag gtcgacattt actctccatg	780
tgctctcggg gctaccatca acgatgacac tatcgatcgc ctgcgttgcg ctgttgtagc	840
cggttgcgca aacaaccagc tgaaagaaga aaacgtccac ggtccggcgc tggttgagcg	900
cgggatagta tacgccccag acttcctgat caatgcaggg ggcctgatta acgtgtatag	960
cgaagttaca gggctctacc gtcagggggc tttaactcag accgaaaaaa tctatgacta	1020
cacactccaa gttctggaaa aagccgcggc tgaagggtct caccgcagc aggctgcgat	1080
ccgtcaggcg gaacaacgca tcgctgcaat tggttaagggtg aaaagcacct actaatgaga	1140
aataaggagg atttcatgtc gacgacaacc gttggtgact acttgctgta tcgcttaaac	1200

gaaatcggca ttgagcacct cttcggagtg ccaggtgatt acaatctgca atttctggat	1260
catgtaatcg accaccctca gctgacttgg gtcggctgca ctaacgaact taacgctgcc	1320
tacgcagctg atgggttatgc gcgttgtcgt ccggctgcgg cactgctgac caccttcggg	1380
gttggcgaac tgagcgctat taatggcatc gcaggttcct acgcggagta tctgccggt	1440
atacatatcg ttggtgcacc gagtctatca gcccagcagc agggcgacct gattcaccac	1500
tctcttggcg aaggtgattt ttccagcttc ctgaggatgt cccaaccggt gtctgttgcg	1560
caggctgctc tgactcctga taacgcatgc aaggaaatcg accgcgtact ggcggaagtc	1620
ctcattcagc gtcgtcccgg ctacctgtg ctgtctaccg acgtggctgc tgcgccggcg	1680
gctctgccac aaagcactct ttctttgccg accgccccgg atcatcgcg agttctggct	1740
gctttcagcg acgctgctga gcagatgctg gctcaggcca aaagcgtctc tctactggcg	1800
gactttctgg ctgatcgttt cgggtgttact cgagcactgg ccgcgtggct tcagcaggtt	1860
ccgctaccgc acgccactct gttaatgggt aaaggcgctc tgagtgaaca gcaaccaggg	1920
ttcgtgggta cctacgctgg tgcggcatct atcgattcga cgcgtaggcg aatcgaagaa	1980
gctggggtaa ttatcggagt gggagttaga ttttccgaca ctatcacagc aggcttctcg	2040
cagcagatcg acgcccgcg ttttatagac attcaaccct tcttctctcg tattggcgat	2100
cgccagtttg atcacttgcc gatgcaggct gccgtcgcag ccctgcatca actgtgtctt	2160
cgttatcagc agcagtggct tatcaccgct cctagccgc ctgactgcc gccggctgct	2220
ggtagcgagc tgtcccagaa cgcatcttg caggcgatgc agaacttcat ccgccctggg	2280
gacctgttgg tggccgacca aggtactgcg gcgttcggcg cagcggcgct gcgcttaccg	2340
cagaattgcc agctgcttgt gcagccgctg tggggctcaa tcggttacag tctgccggcc	2400
acctttggtg ctcagacggc agatacagag cgtcgtgtaa tcctaatacat tggcgatggt	2460
tcagcgcaat taactattca ggaactttcc agtatgatgc gtgacggctt gaaacctatc	2520
atctttctcc tgaacaacaa cggttacacc gttgaacggg cgattcacgg gcgaggacaa	2580
cgttataacg atatcgctgc ttggaattgg acccaactgc ccaggcgct gagtgttcat	2640
tgcccagcgc agagctggcg agtcgttgaa acgggtgcagc tgaccgacgt aatgaaagtc	2700
atcgctgctt ctccgcgtct gagcttggtga gaagttgttc tgcctgcaat ggatgtccca	2760
ccgctgctgc aagcagttag tgccgctctg aaccagcgca actcctctta attggataaa	2820

gaataaggag	gtatTTTTatg	acagcggagc	agcaaaatgg	ggtatccgac	tcacgccgtt	2880
tcgaatttca	ggaatttggT	ggccctatcg	ccccacagac	ctatcagctc	cccgcaccgg	2940
ctagcgatga	agttttgtta	aaggtgaact	actgcggtgt	ctgtcacagt	gatgttcatc	3000
ttcacgacgg	ctacttcgag	ctgggtggcg	ataaacgtct	gaacttcgct	atgccgctgc	3060
cgctgacgct	gggtcacgaa	gtaattggca	ccgttgtggc	tgtcggcgac	caggttactg	3120
gtgtaaaacc	gggggaccag	cgactgatct	atccgtggat	aggttgcgga	aaatgcggcg	3180
cgtgtcaaaa	aggagaagaa	aacctgtgcg	ttactcctgc	acatctgggc	gtgaacaagc	3240
cgggcggtta	cgctgatcac	atcgttgtac	cccattctcg	ctaccttctg	gacatttcgg	3300
gtctgaaccc	gggtgatgcc	gctaccctcg	cgtgctccgg	cctgaccact	ttcagcgcga	3360
tcaacaaagt	gttgccgctt	gcagatgacc	agtggattgt	tgttatcggt	tgtgggtggcc	3420
tcggccagat	ggcgctgcgt	atcctgcaag	ctatgggaat	tggcaatgtt	atcggatttg	3480
acctgtctga	agagaaacgg	aaactggctc	atgaaagcgg	tgcacgtcac	tccttcgatc	3540
caaacactcc	gaagctgaac	cgcgTggTcg	ccgaaacctg	cccgggtacg	gtacaggccg	3600
cgtagactt	tgtgggcaat	gagcaaactg	ctcagctggc	actgtctctg	cttggaaaag	3660
gtggcaaata	tgTtcctgtc	gggctgcacg	gcggcgagct	gcgttaccca	ttgccgatca	3720
tcacgaacaa	agctgtaagt	atcatcggtt	cttacgttgg	taccctgaaa	gaactggaag	3780
acttagttgc	tttcgccaaG	gaaaagaatc	tgccgccaat	tcataattgaa	caccgcccgc	3840
tggaatcggc	ggctcaggcc	gtagaggacc	tggaaaaagg	acaggttgct	gggcgtgtta	3900
tcctggatgc	aggtaactaa	gaaggagata	tacatatgac	ccatcaatta	agatcgcgcg	3960
atatcatcgc	tctgggcttt	atgacatttg	cgttgTtcgt	cggcgCaggt	aacattatTT	4020
tccctccaat	ggtcggcttg	caggcaggcg	aacacgtctg	gactgcggca	ttcggcttcc	4080
tcattactgc	cgttggccta	ccggtattaa	cggtagtggc	gctggcaaaa	gttggcggcg	4140
gtgttgacag	tctcagcacg	ccaattggta	aagtcgctgg	cgtactgctg	gcaacagttt	4200
gttacctggc	ggtggggccg	ctttttgcta	cgccgcgtac	agctaccgtt	tcttttgaag	4260
tgggcattgc	gccgctgacg	ggtgattccg	cgtgcccgtc	gtttattttac	agcctggTct	4320
atttcgctat	cgttattctg	gtttcgctct	atccgggcaa	gctgctggat	accgtgggca	4380
acttccttgc	gccgctgaaa	attatcgcgC	tggTcatcct	gtctgttgcc	gcaattatct	4440

ggccggcggg	ttctatcagt	acggcgactg	aggcttatca	aaacgctgcg	ttttctaacg	4500
gcttcgtcaa	cggctatctg	accatggata	cgctgggcgc	aatgggtgtt	ggtatcgtta	4560
ttgttaacgc	ggcgcgttct	cgtggcggtta	ccgaagcgcg	tctgctgacc	cgttataaccg	4620
tctgggctgg	cctgatggcg	ggtgttggtc	tgactctgct	gtacctggcg	ctgttccgtc	4680
tgggttcaga	cagcgcgtcg	ctggtcgata	agtctgcaaa	cggtgcggcg	atcctgcatg	4740
cttacgttca	gcataccttt	ggcggcggcg	gtagcttcct	gctggcggcg	ttaatcttca	4800
tcgcctgcct	ggtcacggcg	gttggcctga	cctgtgcttg	tgcagaattc	ttcgcccagt	4860
acgtaccgct	ctcttatcgt	acgctgggtg	ttatcctcgg	cggcttctcg	atgggtgggtg	4920
ctaacctcgg	cttgagccag	ctgattcaga	tctctgtacc	ggtgctgacc	gccatttatc	4980
cgccgtgtat	cgcactgggt	gtattaagtt	ttacacgctc	atggtggtcat	aattcgtccc	5040
gcgtgattgc	tccgccgatg	tttatcagcc	tgcttttttg	tattctcgac	gggatcaagg	5100
catctgcatt	cagcgatatc	ttaccgtcct	gggcgcagcg	tttaccgctg	gccgaacaag	5160
gtctggcgtg	gttaatgcca	acagtgggtga	tggtggttct	ggccattatc	tgggatcgtg	5220
cggcaggctg	tcaggtgacc	tccagcgctc	actaatacgc	atggcatgga	tgaccgatgg	5280
tagtgtgggc	tctcccatg	cgagagtagg	gaactgccag	gcatcaaata	aaacgaaagg	5340
ctcagtcgaa	agactgggcc	tttcgtttta	tctgttggtt	gtcgggtgaac	gctctcctga	5400
gtaggacaaa	tccgccggga	gcggatttga	acgttgcgaa	gcaacggccc	ggagggtggc	5460
gggcaggacg	cccgccataa	actgccaggc	atcaaattaa	gcagaaggcc	atcctgacgg	5520
atggcccttt	tgcgtggcca	gtgccaagct	tgcatgcgtg	ccagctgcat	taatgaagaa	5580
atcatgctgg	aagaataaca	gctcactcaa	aggcggtagt	acgggttttg	ctgcccgcaa	5640
acgggctgtt	ctgggtgttg	tagtttggtta	tcagaatcgc	agatccggct	tcagccggtt	5700
tgccggctga	aagcgctatt	tcttcagaa	ttgcatgat	tttttcccca	cgggaggcgt	5760
cactggctcc	cgtgttgctg	gcagctttga	ttcgataagc	agcatcgcct	gtttcaggct	5820
gtctatgtgt	gactgttgag	ctgtaacaag	ttgtctcagg	tgttcaattt	catgttctag	5880
ttgctttgtt	ttactggttt	cacctgttct	attaggtgtt	acatgctgtt	catctgttac	5940
attgtcgatc	tgttcatggt	gaacagcttt	gaatgcacca	aaaactcgta	aaagctctga	6000
tgtatctatc	ttttttacac	cgttttcatc	tgtgcatatg	gacagttttc	cctttgatat	6060

gtaacggtga acagttgttc tacttttgtt tgttagtctt gatgcttcac tgatagatac	6120
aagagccata agaacctcag atccttccgt atttagccag tatgttctct agtgtggttc	6180
gttgtttttg cgtgagccat gagaacgaac cattgagatc atacttactt tgcattgtcac	6240
tcaaaaattt tgcctcaaaa ctggtgagct gaatttttgc agttaagca tcgtgtagtg	6300
tttttcttag tccgttatgt aggttaggaat ctgatgtaat gggtgttggt attttgtcac	6360
cattcatttt tatctggttg ttctcaagtt cggttacgag atccatttgt ctatctagtt	6420
caacttgga aatcaacgta tcagtcgggc ggcctcgctt atcaaccacc aatttcatat	6480
tgctgtaagt gtttaaatct ttacttattg gtttcaaac ccattgggta agccttttaa	6540
actcatggta gttattttca agcattaaca tgaacttaaa ttcattcaagg ctaatctcta	6600
tatttgcctt gtgagttttc ttttgtgtta gttcttttaa taaccactca taaatcctca	6660
tagagtattt gttttcaaaa gacttaacat gttccagatt atattttatg aattttttta	6720
actggaaaag ataaggcaat atctcttcac taaaaactaa ttctaatttt tcgcttgaga	6780
acttggcata gtttgtccac tggaaaatct caaagccttt aaccaaagga ttcctgattt	6840
ccacagttct cgtcatcagc tctctggttg ctttagctaa tacaccataa gcattttccc	6900
tactgatgtt catcatctga gcgtattgggt tataagtga cgataccgtc cgttctttcc	6960
ttgtaggggt ttcaatcgtg gggttgagta gtgccacaca gcataaaaatt agcttggttt	7020
catgctccgt taagtcatag cgactaatcg ctagttcatt tgctttgaaa acaactaatt	7080
cagacataca tctcaattgg tctaggtgat tttaatcact ataccaattg agatgggcta	7140
gtcaatgata attactagtc cttttccttt gagttgtggg tatctgtaaa ttctgctaga	7200
cctttgctgg aaaacttgta aattctgcta gaccctctgt aaattccgct agacctttgt	7260
gtgttttttt tgtttatatt caagtgggta taatttatag aataaagaaa gaataaaaaa	7320
agataaaaag aatagatccc agccctgtgt ataactcact acttttagtca gttccgcagt	7380
attacaaaag gatgtcgcaa acgctgtttg ctctctaca aaacagacct taaaacccta	7440
aaggcttaag tagcaccctc gcaagctcgg gcaaatcgct gaatattcct tttgtctccg	7500
accatcaggc acctgagtcg ctgtcttttt cgtgacattc agttcgctgc gtcacggct	7560
ctggcagtga atgggggtaa atggcactac aggcgccttt tatggattca tgcaaggaaa	7620
ctaccataa tacaagaaaa gcccgtcacg ggcttctcag ggcgttttat ggcgggtctg	7680

ctatgtggtg ctatctgact ttttgctgtt cagcagttcc tgccctctga ttttccagtc	7740
tgaccacttc ggattatccc gtgacaggtc attcagactg gctaatgcac ccagtaaggc	7800
agcggtatca tcaacaggct taccctgtctt actgtctttt ctacggggtc tgacgctcag	7860
tggaacgaaa actcacgtta agggattttg gtcattgagat tatcaaaaag gatcttcacc	7920
tagatccttt taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact	7980
tggtctgaca gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt	8040
cgttcatcca tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggccta	8100
ccatctggcc ccagtgtgc aatgataccg cgagaaccac gctcaccggc tccagattta	8160
tcagcaataa accagccagc cggaagggcc gagcgcagaa gtggtcctgc aactttatcc	8220
gcctccatcc agtctattaa ttgttgccgg gaagctagag taagtagttc gccagttaat	8280
agtttgcgca acgttggtgc cattgctaca ggcatcgtgg tgtcacgctc gtcgtttggt	8340
atggcttcat tcagctccgg ttcccaacga tcaaggcgag ttacatgatc ccccatgttg	8400
tgcaaaaaag cggtagctc cttcggtcct ccgatcgttg tcagaagtaa gttggccgca	8460
gtgttatcac tcatggttat ggcagcactg cataattctc ttactgtcat gccatccgta	8520
agatgctttt ctgtgactgg tgagtactca accaagtcatt tctgagaata gtgtatgcgg	8580
cgaccgagtt gctcttgccc ggcgtcaata cgggataata ccgcgccaca tagcagaact	8640
ttaaaagtgc tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg	8700
ctgttgagat ccagttcgat gtaaccact cgtgcaccca actgatcttc agcatctttt	8760
actttcacca gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga	8820
ataagggcga cacggaaatg ttgaatactc atactcttcc tttttcaata ttattgaagc	8880
atttatcagg gttattgtct catgagcgga tacatatttg aatgtattta gaaaaataaa	8940
caaatagggg ttccgcgcac atttcccga aaagtgccac ctgacgtcta agaaaccatt	9000
attatcatga cattaaccta taaaaatagg cgtatcacga ggccctttcg tctcgcgcgt	9060
ttcgggtgatg acggtgaaaa cctctgacac atgcagctcc cggagacggg cacagcttgt	9120
ctgtaagcgg atgccgggag cagacaagcc cgtcagggcg cgtcagcggg tgttggcggg	9180
tgtcggggct ggcttaacta tgcggcatca gagcagattg tactgagagt gcaccatatg	9240
cgggtgtgaaa taccgcacag atgcgtaagg agaaaatacc gcatcaggcg ccattcgcca	9300

ttcaggctgc gcaactgttg ggaagggcga tcggtgcggg cctcttcgct attacgccag	9360
ctggcgaaag ggggatgtgc tgcaaggcga ttaagttggg taacgccagg gttttcccag	9420
tcacgacgtt gtaaaacgac ggccagtga ttcgttaaga cccactttca catttaagtt	9480
gtttttctaa tccgcatatg atcaattcaa ggccgaataa gaaggctggc tctgcacctt	9540
ggtgatcaaa taattcgata gcttgtcgtg ataatggcgg catactatca gtagtaggtg	9600
tttccctttc ttcttttagcg acttgatgct cttgatcttc caatacgcaa cctaaagtaa	9660
aatgccccac agcgctgagt gcatataatg cattctctag tgaaaaacct tgttggcata	9720
aaaaggctaa ttgattttcg agagtttcat actgtttttc tgtaggccgt gtacctaata	9780
gtacttttgc tccatcgcg tgacttagta aagcacatct aaaactttta gcgttattac	9840
gtaaaaaatc ttgccagctt tccccttcta aagggcacaaa gtgagtatgg tgcctatcta	9900
acatctcaat ggctaaggcg tcgagcaaag cccgcttatt ttttacctgc caatacaatg	9960
taggctgctc tacacctagc ttctgggcga gtttacgggt tggttaaact tcgattccga	10020
cctcattaag cagctctaata gcgctgttaa tcactttact tttatctaata ctagacatca	10080
ttaattccta atttttgttg acactctatc attgatagag ttattttacc actccctatc	10140
agtgatagag aaaagtga	10158

<210> 21

<211> 10101

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 21

actacagtat ataaggaggg taatatgatac ttcgagacaa tttctacgtc gaatcacgaa	60
gaagttgtgt attgccataa caaggacgcc ggcttgaaag caatcatcgc gattcacaac	120
actgtactcg gtccggctct ggggtggcact cgcattgtggc cctacgctag cgaagaggaa	180
gcactgaaag atgtccttcg tttatcccgt gggatgacct acaaagctgc ggtttcaggt	240
ctaaacctgg gcggcggtaa agctgtgatac tgggggtgata cgaataaaga caagtctgaa	300
gcgctgttta gagccttcgg acggtttgta aacagcctgg gcggacgcta cattaccgcg	360
gaggacgttg gcattgatgt taacgacatg gaatatgtgc tgcgtgaaac tgattacgtc	420

accggtgtac atcaggttca cgggtgggagt ggtgatcctt ctccattcac cgcatatggc	480
actctgcaag gcctgatggc cgctctgcaa gtgaaattcg gtaacgaaga cgtaggcaat	540
tacagctacg ctgttcaggg tgtgggtcac gttggcatgg aatttggtta actgctgcgt	600
gagcgcggtg caaagggtttt cgtcactgac atcaacaaag atgcggtcca gcgtgctgtg	660
gacgaatttg gttgtgaggc agtagccctg gatgaaatct atgacgttga ttgcgacgtg	720
tactccccga ccgctctggg cggcaccgtg aacgataaaa ctttaccgcg tctgaaatgt	780
aaggtaatct gcggtgcggc aaacaaccag ttagctaattg atgagatagg cgtggaactg	840
gaaaagaaag gcatacctcta tgctccggac tacgcggtca acgcgggtgg gctgatgaac	900
gtagcctgg aaatcgatgg atacaaccgc gaacgtgcga tgcgtatgat gcgtaccatt	960
tattacaatt tgggtcgcatt tttcgaatc tctaagcgcg acggcatccc tacattccga	1020
gccgccgatc gtatggctga agaacgcata acggccatcg gtaaaactgcg tttaccgcat	1080
ttgggcgctg cggcaccgcg cttccagggc cgacgtggca actaacagaa atagtaaagg	1140
aggtttttta tgtcggagat cacattgggt agataccttt tcgaacgctt aaaccaactg	1200
caagtgcaga ctatTTTTTgg gctgcccggc gacttcaatc tgtccctgct ggataagatc	1260
tatgaagttg atggcatgcg ttgggcaggt aacgctaacg aactcaacgc cgcttacgcg	1320
gctgacggtt atagccgtgt caaaggcctc gcatgtctgg ttaccacttt tgggtgtaggc	1380
gagctaagtg cgctgaatgg tgtgggtggc gcttacgcag aacacgttgg gctgctgcat	1440
gtagtgggcg tcccatcaat ctctagccag gcgaaacagc tgctgctgca ccataccctg	1500
ggtaacggag atttcacggt tttccaccgc atgtccaaca acatttctca gaccacggct	1560
tttatcagcg acattaattc tgctcctggt gaaatcgata ggtgcatccg tgaggcctgg	1620
gtacatcagc gtccggttta cgtcggcctg ccggcgaacc tagttgacct gactgtgccg	1680
gcgtctctgt tagacactcc gatcgatctg tccttgaaaa agaacgaccc ggatgccag	1740
gaagaagtta ttgaaaccgt ccttgatctg gtagacaagt ctaaaaaccc tataatctta	1800
gttgacgcat gcgctagccg tcaactcatgc cgcgatgaag tacgccggtt ggtggactcc	1860
accagcttcc cggttttcgt tactccaatg ggtaaattctg ctgtaaatga gagtcacccg	1920
cgttttggcg gtgtttacgt gggcagcctc agcgagccaa acgtaaaaga agccgttgaa	1980
aacgctgacc tgggtgctgtc cataggcgcc ctgttgagcg acttcaacac tggatcgttc	2040

tcttattcct	acaaaactaa	gaacattggt	gaatttcact	ctgattatac	caaaatccgt	2100
caagcaacgt	tcccgggtgt	tcagatgaaa	gaagcactga	atgtcctggt	ggaaaaaatc	2160
ccgagccatg	tcgctaacta	caaacctctg	ccggttccgc	agcgtcgcgt	tattccgagc	2220
ccaggggata	aggctgcgat	ctctcaggag	tggctgtggt	cgcgtctgtc	tagctggttc	2280
cgcgagggcg	acatcgtcat	tacagaaacc	ggtaccagtg	cgtttggaat	tgtacagtcc	2340
tatttcccag	ataactgcat	cggcatcagt	caggtgctgt	ggggttcgat	cggcttcacc	2400
gtaggtgcaa	cgctgggcgc	ggtgatggct	gcacaagaaa	tcgatccgaa	aaaacgtgtg	2460
attttatttg	tcggtgacgg	ttctctgcaa	cttactgtac	aggaaatttc	taccatgggt	2520
aagtgggaaa	ccactcccta	cctgtttgtg	ctgaacaacg	atgggtacac	tatcgaacgc	2580
cttatccatg	gcgagactgc	tacgtataac	gatattcagc	cgtgggataa	tctgggtctg	2640
ttgccgctgt	tcaaagctcg	tgactacgaa	accaaccgag	ttgcgactgt	aggcgaaatt	2700
gaagcgctat	tcaacaattc	agctttcaat	gagaatacaa	agatccgtat	ggtggaggtc	2760
atgctgccgc	ggatggatgc	accacagaac	ctggttaaac	aggctgaatt	ttcctccaag	2820
accaacagcg	aaaactaaac	aaggcacgga	aaaattggag	gagaaagatg	cagacggcgt	2880
tcttgtataa	gccaggtcac	gaaaacttag	tgcgctcgga	gatcccgata	cctaaagctg	2940
ggcgtggcga	agtcgttctg	gaaattaaag	ccgctggcat	gtgccattcc	gatctgcacg	3000
ttctcgacgg	tggaatcccc	ctgccgggtc	aatttgtaat	gggccatgaa	atcgttggtg	3060
ctattcacga	gatcggccag	gacgtgaccg	gtttcaaaca	gggcgatctg	tacgcagtcc	3120
acggcccgaa	tccgtgtggt	atttgcaccc	tgtgcagaga	aggatttgat	aacgactgca	3180
ctacagtggc	gaaaaccggt	caatggttcg	gactgggtct	tgacggcggc	taccagaagt	3240
atatccgtat	cccgaacgta	aggtctatcg	ttaaagttcc	agaaggtggt	tcagctgagg	3300
cagctgcgag	ctgtactgat	gcagtactga	ccccgtaccg	tgcactaaaa	caggctggcg	3360
ccagcaactc	tactcgggta	ctgattctgg	gtctgggtgg	cttaggtctg	aatgccctta	3420
aactgggctaa	gaccttcggc	agttacgttt	acgcatctga	cctgaaacct	tctgcgcgtg	3480
aagctgctaa	ggccgctggg	gcggatgaag	tgctggagtc	cctgcccgaa	gacccgctgg	3540
gtgttgatat	cgtgttagac	gtcgttggcg	tgcagagcac	cttcaacctc	gctcaaaaac	3600
acgttggccc	gcgtggcatc	attgtacctg	taggcctggc	atccccacag	ctttcgttta	3660

acctaacgga tctggcgctc cgcgaaattc gtgttcaggg cacttttttg ggcacgagca	3720
atgagctggc tgaatgtctg cgcctgtgcc agctgggcct gatcaacccg aaatatactg	3780
tggcgcctct tgaagaagcg ccgaaatata tggaagcaat ggctcatggg aaagtagaag	3840
gtcgtatcgt tttccacccg taagaaggag atatacatat gacccatcaa ttaagatcgc	3900
gcgatatcat cgctctgggc tttatgacat ttgcgttggt cgtcggcgca ggtaacatta	3960
ttttccctcc aatggtcggc ttgcaggcag gcgaacacgt ctggactgcg gcattcggct	4020
tcctcattac tgccgttggc ctaccggtat taacggtagt ggcgctggca aaagtggcg	4080
gcggtgttga cagtctcagc acgccaattg gttaaagtcgc tggcgtactg ctggcaacag	4140
tttgttacct ggcggtgggg ccgctttttg ctacgccgcg tacagctacc gtttcttttg	4200
aagtgggcat tgcgccgctg acgggtgatt ccgcgctgcc gctgtttatt tacagcctgg	4260
tctatttcgc tatcgttatt ctggtttcgc tctatccggg caagctgctg gataccgtgg	4320
gcaacttcct tgcgccgctg aaaattatcg cgctgggtcat cctgtctgtt gccgcaatta	4380
tctggccggc gggttctatc agtacggcga ctgaggctta tcaaaacgct gcgttttcta	4440
acggcttcgt caacggctat ctgaccatgg atacgctggg cgcaatgggtg tttggtatcg	4500
ttattgttaa cgcggcgcgt tctcgtggcg ttaccgaagc gcgtctgctg acccgttata	4560
ccgtctgggc tggcctgatg gcgggtgttg gtctgactct gctgtacctg gcgctgttcc	4620
gtctgggttc agacagcgcg tcgctggctg atcagtctgc aaacggtgcg gcgatcctgc	4680
atgcttacgt tcagcatacc tttggcggcg gcggtagctt cctgctggcg gcgttaatct	4740
tcatgcctg cctggtcacg gcggttgcc tgacctgtgc ttgtgcagaa ttcttcgccc	4800
agtacgtacc gctctcttat cgtacgctgg tgtttatcct cggcggcttc tcgatgggtg	4860
tgtctaacct cggcttgagc cagctgattc agatctctgt accggtgctg accgccattt	4920
atccgccgtg tatcgactg gttgtattaa gttttacacg ctcatgggtg cataattcgt	4980
cccgctgat tgctccgccg atgtttatca gcctgctttt tggatttctc gacgggatca	5040
aggcatctgc attcagcgat atcttaccgt cctgggcgca gcgtttaccg ctggccgaac	5100
aaggctcggc gtggttaatg ccaacagtgg tgatgggtgg tctggccatt atctgggatc	5160
gtcgggcagg tcgtcaggtg acctccagcg ctactaata cgcatggcat ggatgaccga	5220
tggtagtgtg ggctctcccc atgcgagagt agggaaactgc caggcatcaa ataaaacgaa	5280

aggctcagtc gaaagactgg gcctttcgtt ttatctgttg tttgtcggtg aacgctctcc	5340
tgagtaggac aaatccgccg ggagcggatt tgaacgttgc gaagcaacgg cccggagggt	5400
ggcgggcagg acgcccgcc aaaactgcc ggcacaaat taagcagaag gccatcctga	5460
cggatggcct ttttgcgtgg ccagtgccaa gcttgcacgc gtgccagctg cattaatgaa	5520
gaaatcatgc tggaagaata acagctcact caaaggcggg agtacgggtt ttgctgccc	5580
caaacgggct gttctgggtg tgctagtttg ttatcagaat cgcagatccg gttcagccg	5640
gtttgccggc tgaaagcgct atttcttcca gaattgccat gattttttcc ccacgggagg	5700
cgtcactggc tcccgtgttg tcggcagctt tgattcgata agcagcatcg cctgtttcag	5760
gctgtctatg tgtgactgtt gagctgtaac aagtgtgttc aggtgttcaa tttcatgttc	5820
tagttgcttt gttttactgg tttcacctgt tctattaggt gttacatgct gttcatctgt	5880
tacattgtcg atctgttcat ggtgaacagc tttgaatgca ccaaaaactc gtaaaagctc	5940
tgatgtatct atctttttta caccgttttc atctgtgcat atggacagtt ttccctttga	6000
tatgtaacgg tgaacagttg ttctactttt gtttgtagt cttgatgctt cactgataga	6060
tacaagagcc ataagaacct cagatccttc cgtatttagc cagtatgttc tctagtgtgg	6120
ttcgttgctt ttgcgtgagc catgagaacg aaccattgag atcatactta ctttgcacgt	6180
cactcaaaaa ttttgcctca aaactgggtga gctgaatttt tgcagttaaa gcatcgtgta	6240
gtgtttttct tagtccgtta tgtaggtagg aatctgatgt aatgggtgtt ggtattttgt	6300
caccattcat ttttatctgg ttgtttctca gttcgggttac gagatccatt tgtctatcta	6360
gttcaacttg gaaaatcaac gtatcagtcg ggccgacctg cttatcaacc accaatttca	6420
tattgctgta agtgtttaaa tctttactta ttgggtttcaa aaccattgg ttaagccttt	6480
taaactcatg gtagttatct tcaagcatta acatgaactt aaattcatca aggctaatt	6540
ctatatttgc cttgtgagtt ttcttttgtg ttagttcttt taataaccac tcataaatcc	6600
tcataagagta tttgttttca aaagacttaa catgttccag attatatatt atgaattttt	6660
ttaactggaa aagataaggc aatatctctt cactaaaaac taattctaatt ttttcgcttg	6720
agaacttggc atagtttgtc cactggaaaa tctcaaagcc tttaaccaa ggattcctga	6780
ttccacagc tctcgtcatc agctctctgg ttgcttttagc taatacacca taagcatttt	6840
ccctactgat gttcatcatc tgagcgtatt gggtataagt gaacgatacc gtccgttctt	6900

tcctttagg	gttttcaatc	gtggggttga	gtagtgccac	acagcataaa	attagcttgg	6960
tttcatgctc	cgtaaagtca	tagcgactaa	tcgctagttc	atttgctttg	aaaacaacta	7020
attcagacat	acatctcaat	tgggtctaggt	gattttaatc	actataccaa	ttgagatggg	7080
ctagtcaatg	ataattacta	gtccttttcc	tttgagttgt	gggtatctgt	aaattctgct	7140
agacctttgc	tggaaaactt	gtaaattctg	ctagaccctc	tgtaaattcc	gctagacctt	7200
tgtgtgtttt	ttttgtttat	attcaagtgg	ttataattta	tagaataaag	aaagaataaa	7260
aaaagataaa	aagaatagat	cccagccctg	tgtataactc	actactttag	tcagttccgc	7320
agtattacaa	aaggatgtcg	caaacgctgt	ttgctcctct	acaaaacaga	ccttaaaacc	7380
ctaaaggctt	aagtagcacc	ctcgcaagct	cgggcaaatc	gctgaatatt	ccttttgtct	7440
ccgaccatca	ggcacctgag	tcgctgtctt	tttcgtgaca	ttcagttcgc	tgcgctcacg	7500
gctctggcag	tgaatggggg	taaatggcac	tacaggcgcc	ttttatggat	tcatgcaagg	7560
aaactaccca	taatacaaga	aaagcccgtc	acgggcttct	cagggcgttt	tatggcgggt	7620
ctgctatgtg	gtgctatctg	actttttgct	gttcagcagt	tcctgccctc	tgattttcca	7680
gtctgaccac	ttcggattat	cccgtgacag	gtcattcaga	ctggctaatt	caccagtaa	7740
ggcagcggta	tcatcaacag	gcttaccctg	cttactgtct	tttctacggg	gtctgacgct	7800
cagtggaacg	aaaactcacg	ttaagggatt	ttggctatga	gattatcaaa	aaggatcttc	7860
acctagatcc	ttttaaat	aaaatgaagt	tttaaatcaa	tctaaagtat	atatgagtaa	7920
acttggtctg	acagttacca	atgcttaatc	agtgaggcac	ctatctcagc	gatctgtcta	7980
tttcgttcat	ccatagttgc	ctgactcccc	gtcgtgtaga	taactacgat	acgggagggc	8040
ttaccatctg	gccccagtgc	tgcaatgata	ccgcgagAAC	cacgctcacc	ggctccagat	8100
ttatcagcaa	taaaccagcc	agccggaagg	gccgagcgca	gaagtgggtcc	tgcaacttta	8160
tccgcctcca	tccagtctat	taattgttgc	cgggaagcta	gagtaagtag	ttcgccagtt	8220
aatagtttgc	gcaacgttgt	tgccattgct	acaggcatcg	tgggtgtcacg	ctcgtcgttt	8280
ggtatggctt	cattcagctc	cggttcccaa	cgatcaaggc	gagttacatg	atcccccatg	8340
ttgtgcaaaa	aagcggttag	ctccttcggt	cctccgatcg	ttgtcagaag	taagttggcc	8400
gcagtgttat	cactcatggt	tatggcagca	ctgcataatt	ctcttactgt	catgccatcc	8460
gtaagatgct	tttctgtgac	tgggtgagtac	tcaaccaagt	cattctgaga	atagtgtatg	8520

cggcgaccga gttgctcttg cccggcgtca atacgggata ataccgcgcc acatagcaga	8580
actttaaaag tgctcatcat tggaaaacgt tcttcggggc gaaaactctc aaggatctta	8640
ccgctgttga gatccagttc gatgtaaccc actcgtgcac ccaactgac ttcagcatct	8700
tttactttca ccagcgtttc tgggtgagca aaaacaggaa ggcaaaatgc cgcaaaaaag	8760
ggaataaggg cgacacggaa atgttgaata ctcatactct tcctttttca atattattga	8820
agcatttatc agggttattg tctcatgagc ggatacatat ttgaatgtat ttagaaaaat	8880
aaacaaatag gggttccgcg cacatttccc cgaaaagtgc cacctgacgt ctaagaaacc	8940
attattatca tgacattaac ctataaaaat aggcgtatca cgaggccctt tcgtctcgcg	9000
cgtttcggtg atgacggtga aaacctctga cacatgcagc tcccggagac ggtcacagct	9060
tgtctgtaag cggatgccgg gagcagacaa gcccgtcagg gcgcgtcagc ggggtgttggc	9120
gggtgtcggg gctggcttaa ctatgcggca tcagagcaga ttgtactgag agtgcaccat	9180
atgcggtgtg aaataccgca cagatgcgta aggagaaaat accgcatcag gcgccattcg	9240
ccattcaggc tgcgcaactg ttgggaaggg cgatcgggtg gggcctcttc gctattacgc	9300
cagctggcga aagggggatg tgctgcaagg cgattaagtt gggtaacgcc agggttttcc	9360
cagtcacgac gttgtaaaac gacggccagt gaattcgta agaccactt tcacatttaa	9420
gttgtttttc taatccgcat atgatcaatt caaggccgaa taagaaggct ggctctgcac	9480
cttggtgatc aaataattcg atagcttgtc gtaataatgg cggcatacta tcagtagtag	9540
gtgtttccct ttcttcttta gcgacttgat gctcttgatc ttccaatacg caacctaaag	9600
taaaatgccc cacagcgctg agtgcataata atgcattctc tagtgaaaaa ccttgttggc	9660
ataaaaaggc taattgattt tcgagagttt catactgttt ttctgtaggc cgtgtaccta	9720
aatgtacttt tgctccatcg cgatgactta gtaaagcaca tctaaaactt ttagcgttat	9780
tacgtaaaaa atcttgccag ctttcccctt ctaaagggca aaagtgagta tggcgcctat	9840
ctaakatctc aatggctaag gcgtcgagca aagcccgtt attttttaca tgccaataca	9900
atgtaggctg ctctacacct agcttctggg cgagttttacg gggtgttaaa ccttcgattc	9960
cgacctcatt aagcagctct aatgcgctgt taatcacttt acttttatct aatctagaca	10020
tcattaattc ctaatttttg ttgacactct atcattgata gagttatttt accactccct	10080
atcagtgata gagaaaagtg a	10101

<210> 22
<211> 10141
<212> DNA
<213> Artificial Sequence

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

<400> 22
actacagtat ataaggaggg taatatgatac ttcgagacaa tttctacgtc gaatcacgaa 60
gaagttgtgt attgccataa caaggacgcc ggcttgaaag caatcatcgc gattcacaac 120
actgtactcg gtccggctct gggtggcact cgcattgtggc cctacgctag cgaagaggaa 180
gcactgaaag atgtccttcg tttatcccgt gggatgacct acaaagctgc ggtttcaggt 240
ctaaacctgg gcggcggtta agctgtgatac tggggtgatac cgaataaaga caagtctgaa 300
gcgctgttta gagccttcgg acggtttgta aacagcctgg gcggacgcta cattaccgcg 360
gaggacgttg gcattgatgt taacgacatg gaatatgtgc tgcgtgaaac tgattacgtc 420
accggtgtac atcaggttca cgggtgggagt ggtgatacct ctccattcac cgcataatggc 480
actctgcaag gcctgatggc cgctctgcaa gtgaaattcg gtaacgaaga cgtaggcaat 540
tacagctacg ctgttcaggg tgtgggtcac gttggcatgg aatttgtaa actgctgcgt 600
gagcgcggtg caaaggtttt cgtcactgac atcaacaaag atgcggtcca gcgtgctgtg 660
gacgaatttg gttgtgaggc agtagccctg gatgaaatct atgacgttga ttgcgacgtg 720
tactccccga ccgctctggg cggcaccgtg aacgataaaa ctttaccgcg tctgaaatgt 780
aaggtaatct gcggtgcggc aaacaaccag ttagctaata atgagatagg cgtggaactg 840
gaaaagaaag gcatacctta tgctccggac tacgcggtca acgcgggtgg gctgatgaac 900
gtagcctgg aaatcgatgg atacaaccgc gaacgtgcga tgcgtatgat gcgtaccatt 960
tattacaatt tgggtcgcata tttcgaaatc tctaagcgcg acggcatccc tacattccga 1020
gccgccgata gtatggctga agaacgcata acggccatcg gtaaaactgcg tttaccgcat 1080
ttgggcgctg cggcaccgcg cttccagggc cgacgtggca actaaaattc actttaacaa 1140
ctaaggggagc aagaatgtcg acgacaaccg ttggtgacta cttgctgtat cgcttaaacg 1200
aaatcggcat tgagcacctc ttcggagtgc cagggtgatta caatctgcaa tttctggatc 1260

atgtaatcga ccaccctcag ctgacttggg tcggctgcac taacgaactt aacgctgcct	1320
acgcagctga tggttatgcg cgttgctgctc cggctgcggc actgctgacc accttcgggg	1380
ttggcgaact gagcgctatt aatggcatcg caggttccta cgcggagtat ctgccggtaa	1440
tacatatcgt tgggtgcaccg agtctatcag cccagcagca gggcgacctg attcaccact	1500
ctcttggcga aggtgatttt tccagcttcc tgaggatgtc ccaaccgggtg tctgttgctc	1560
aggctgctct gactcctgat aacgcatgca aggaaatcga ccgcgtactg gcggaagtcc	1620
tcattcagcg tcgtcccggc tacctgctgc tgtctaccga cgtggctgct gcgccggcgg	1680
ctctgccaca aagcactctt tctttgccga ccgccccgga tcatcgcgca gttctggctg	1740
ctttcagcga cgctgctgag cagatgctgg ctcaggccaa aagcgtctct ctactggcgg	1800
actttctggc tgatcgtttc ggtgttactc gagcactggc cgctgggtt cagcaggttc	1860
cgctaccgca cgccactctg ttaatgggta aaggcgttct gagtgaacag caaccagggt	1920
tcgtgggtac ctacgtgggt gcggcatcta tcgattcgac gcgtggcgca atcgaagaag	1980
ctggggtaat tatcggagtg ggagttagat tttccgacac tatcacagca ggcttctcgc	2040
agcagatcga cgcccgccgt tttatagaca ttcaaccctt cttctctcgt attggcgatc	2100
gccagtttga tcacttgccg atgcaggctg ccgtcgcagc cctgcatcaa ctgtgtcttc	2160
gttatcagca gcagtggctt atcaccgctc ctagcccgcc tgcaactgccg ccggctgctg	2220
gtagcgagct gtcccagaac gcattctggc aggcgatgca gaacttcacg cgccctgggg	2280
acctgttgggt ggccgaccaa ggtactgcgg cgttcggcgc agcggcgctg cgcttaccgc	2340
agaattgccg gctgcttggt cagccgctgt ggggctcaat cggttacagt ctgccggcca	2400
cctttgggtg tcagacggca gatacagagc gtcgtgtaat cctaattcatt ggcgatggtt	2460
cagcgcaatt aactattcag gaactttcca gtatgatgcg tgacggcttg aaacctatca	2520
tctttctcct gaacaacaac ggttacaccg ttgaacgggc gattcacggc gcggagcaac	2580
gttataacga tatcgctgct tggaattgga cccaactgcc ccaggcgctg agtgttcatt	2640
gcccagcgca gagctggcga gtcgttgaaa cgggtgcagct gaccgacgta atgaaagtca	2700
tcgctgcttc tccgctctg agcttggtag aagttgttct gcctgcaatg gatgtcccac	2760
cgctgctgca agcagtgagt gccgctctga accagcgcaa ctcctcttaa tttttaaggt	2820
aacacacgag ctaagaacta aggaggattt atgcgcagca tgcagtttga tgagtacggt	2880

gcacccctga aagcgttctc atatgaagac ccgaccccg	aagggaagga agtagtcgtt	2940
aggatcgaag cctgtggtgt gtgccactct gatattcatc	ttcacgaggg ctacttcgac	3000
atgggcggtg gcaataaagc tgatgttact cgtgctcgcg	aactcccttt tacattgggt	3060
catgaaatcg ttggcgaagt ggtagcaact ggaccaggtg	tcaccggcg	3120
gacaaacgta ttgtgtaccc gtggatcggg tgcggcgact	gcccgaatg caacagtgg	3180
gaggatcagt cctgtgcgcg tccacgtaac ctgggtgttc	acgttgacgg tggctattcg	3240
acgcacgtaa agataccgga cgaaaaattc ctgttcgcct	acgatgggtat tcctactgag	3300
ttagcgggaa cctatgcttg cagcggcatc accgcttatg	gtgcactgat gaaagcaaag	3360
gaagcggctg aaagatctgg ctacatcgg	ctgattggcg ctggtggcgt	3420
ggtctgatgc tggccaaagc agcgatcggg gctaaaactg	tagtctttga tatcgacgac	3480
gcaaaactgg aagctgcgac ccgtgccggg gcggattacg	tgttcaactc cggtgcaaaa	3540
gaaacacgca aggaagttat gaaactaacg aatggtggcc	tgtctggtgc tgttgatttc	3600
gttggcagcg ataaaagcgc tctgtttgga atcaacgcct	tgggtcagaa cggcgtgctg	3660
gtcataattg gactgttcgg tggcgctatg actgttccgg	taccctgtt cccgctgaaa	3720
gggatcaccg tacgtggctc atacgtaggt tccctgcaag	agatgagtga tatgatggag	3780
ttagttcgcg ctgggaaagt tcctccgatg ccggtaaaaa	ctcggccact ggacgctgcc	3840
tgggaaacc ttgaggatct acgcatgggt aaaatcgtgg	gccgtgttgt tctgaccca	3900
taagaaggag atatacatat gaccatcaa ttaagatcg	gcgatatcat cgctctgggc	3960
tttatgacat ttgcgttggt cgtcggcgca ggtaacatta	ttttccctcc aatggtcggc	4020
ttgcaggcag gcgaacacgt ctggactgcg gcattcggct	tcctcattac tgccgttggc	4080
ctaccggtat taacggtagt ggcgctggca aaagttggcg	gcggtgttga cagtctcagc	4140
acgccaattg gtaaagtcgc tggcgtagt ctggcaacag	tttgttacct ggcggtgggg	4200
ccgctttttg ctacgcccg tacagctacc gtttcttttg	aagtgggcat tgcgccgtg	4260
acgggtgatt ccgcgctgcc gctgtttatt tacagcctgg	tctatttcgc tatcgttatt	4320
ctggtttcgc tctatccggg caagctgctg gataccgtgg	gcaacttcct tgcgccgtg	4380
aaaattatcg cgctgggtcat cctgtctgtt gccgcaatta	tctggccggc ggttctatc	4440
agtacggcga ctgaggctta tcaaacgct gcgttttcta	acggcttcgt caacggctat	4500

ctgaccatgg atacgctggg cgcaatggtg tttggtatcg ttattgttaa cgcggcgcgt	4560
tctcgtggcg ttaccgaagc gcgtctgctg acccgttata ccgtctgggc tggcctgatg	4620
gcgggtgttg gtctgactct gctgtacctg gcgctgttcc gtctgggttc agacagcgcg	4680
tcgctggtcg atcagtctgc aaacggtgcg gcgatcctgc atgcttacgt tcagcatacc	4740
tttggcggcg gcggtagctt cctgctggcg gcgttaatct tcatcgctg cctggtcacg	4800
gcggttggcc tgacctgtgc ttgtgcagaa ttcttcgccc agtacgtacc gctctcttat	4860
cgtacgctgg tgtttatcct cggcggcttc tcgatgggtg tgtctaacct cggcttgagc	4920
cagctgattc agatctctgt accggtgctg accgccatct atccgccgtg tatcgcactg	4980
gttgatttaa gttttacacg ctcatggtgg cataattcgt cccgcgtgat tgctccgccg	5040
atgtttatca gcctgctttt tgggtattctc gacgggatca aggcattctgc attcagcgat	5100
atcttaccgt cctggggcgca gcgtttaccg ctggccgaac aaggctctggc gtgggttaatg	5160
ccaacagtgg tgatggtggt tctggccatt atctgggatc gtgcggcagg tcgtcaggtg	5220
acctccagcg ctactaata cgcatggcat ggatgaccga tggtagtgtg ggctctcccc	5280
atgcgagagt agggaaactgc caggcatcaa ataaaacgaa aggcctcagtc gaaagactgg	5340
gcctttcgtt ttatctgttg tttgtcggtg aacgctctcc tgagtaggac aaatccgccg	5400
ggagcggatt tgaacgttgc gaagcaacgg cccggagggt ggcgggcagg acgcccgcca	5460
taaactgcca ggcataaat taagcagaag gccatcctga cggatggcct ttttgcgtgg	5520
ccagtgccaa gcttgcatgc gtgccagctg cattaatgaa gaaatcatgc tggaagaata	5580
acagctcact caaaggcggc agtacgggtt ttgctgcccg caaacgggct gttctggtgt	5640
tgctagtttg ttatcagaat cgcagatccg gcttcagccg gtttgccggc tgaaagcgct	5700
atttcttcca gaattgccat gattttttcc ccacgggagg cgtcactggc tcccggtgtg	5760
tcggcagctt tgattcgata agcagcatcg cctgtttcag gctgtctatg tgtgactgtt	5820
gagctgtaac aagttgtctc aggtgttcaa tttcatgttc tagttgcttt gttttactgg	5880
ttcacctgt tctattaggt gttacatgct gttcatctgt tacattgtcg atctgttcat	5940
ggtgaacagc tttgaatgca ccaaaaactc gtaaaagctc tgatgtatct atctttttta	6000
caccgttttc atctgtgcat atggacagtt ttccctttga tatgtaacgg tgaacagttg	6060
ttctactttt gtttgttagt cttgatgctt cactgataga tacaagagcc ataagaacct	6120

cagatccttc cgtatcttagc cagtatgttc tctagtgtgg ttcgttggtt ttgcgtgagc	6180
catgagaacg aaccattgag atcatactta ctttgcattg cactcaaaaa ttttgcctca	6240
aaactgggtga gctgaatctt tgcagttaaa gcatcgtgta gtgtttttct tagtccgtta	6300
tgtaggtagg aatctgatgt aatgggtgtt ggtatcttgt caccattcat ttttatctgg	6360
ttgttctcaa gtctgggtac gagatccatt tgtctatcta gttcaacttg gaaaatcaac	6420
gtatcagtcg ggccgctcgt cttatcaacc accaatttca tattgctgta agtgttttaa	6480
tctttactta ttgggttcaa aaccattggg ttaagccttt taaactcatg gtagttatct	6540
tcaagcatta acatgaactt aaattcatca aggctaattc ctatatttgc cttgtgagtt	6600
ttcttttgtg ttagttcttt taataaccac tcataaatcc tcatagagta tttgttttca	6660
aaagacttaa catgttccag atttatcttt atgaattttt ttaactggaa aagataaggc	6720
aatatctctt cactaaaaac taattctaatt ttttcgcttg agaacttggc atagtgtgtc	6780
cactggaaaa tctcaaagcc ttttaaccaa ggattcctga tttccacagt tctcgtcatc	6840
agctctctgg ttgcttttagc taatacacca taagcatttt ccctactgat gttcatcatc	6900
tgagcgtatt gggtataagt gaacgatacc gtccgttctt tcctttagg gttttcaatc	6960
gtgggggtga gtagtgccac acagcataaa attagcttgg tttcatgctc cgtaaagtca	7020
tagcgactaa tcgctagtct atttgctttg aaaacaacta attcagacat acatctcaat	7080
tggcttaggt gattttaatc actataccaa ttgagatggg ctagtcaatg ataattacta	7140
gtccttttcc tttgagttgt gggtatctgt aaattctgct agaccttgc tggaaaactt	7200
gtaaattctg ctagaccctc tgtaaatcc gctagacctt tgtgtgtttt tttgtttat	7260
attcaagtgg ttataattta tagaataaag aaagaataaa aaaagataaa aagaatagat	7320
cccagccctg tgtataactc actactttag tcagttccgc agtattacaa aaggatgtcg	7380
caaacgctgt ttgctcctct acaaaacaga cttaaaacc ctaaaggctt aagtagcacc	7440
ctcgcaagct cgggcaaact gctgaatatt cttttgtct ccgaccatca ggcacctgag	7500
tcgctgtctt tttcgtgaca ttcagttcgc tgcgctcacg gctctggcag tgaatggggg	7560
taaatggcac tacaggcgcc ttttatggat tcatgcaagg aaactaccca taatacaaga	7620
aaagcccgtc acgggcttct cagggcgttt tatggcgggt ctgctatgtg gtgctatctg	7680
actttttgct gttcagcagt tcctgccctc tgattttcca gtctgaccac ttcggattat	7740

cccgtagacag gtcattcaga ctggctaata caccagtaa ggcagcggta tcatcaacag	7800
gcttaccctgt cttactgtct tttctacggg gtctgacgct cagtggaaacg aaaactcacg	7860
ttaagggatt ttgggtcatga gattatcaaa aaggatcttc acctagatcc ttttaaatta	7920
aaaatgaagt tttaaatcaa tctaaagtat atatgagtaa acttgggtctg acagttacca	7980
atgcttaatc agtgaggcac ctatctcagc gatctgtcta tttcggtcat ccatagttgc	8040
ctgactcccc gtcgtgtaga taactacgat acgggagggc ttaccatctg gccccagtgc	8100
tgcaatgata ccgcgagaac cacgctcacc ggctccagat ttatcagcaa taaaccagcc	8160
agccggaagg gccgagcgca gaagtgggtcc tgcaacttta tccgcctcca tccagtctat	8220
taattgttgc cgggaagcta gagtaagtag ttcgccagtt aatagtttgc gcaacgttgt	8280
tgccattgct acaggcatcg tgggtgtcacg ctctctggtt ggtagggctt cattcagctc	8340
cggttcccaa cgatcaaggc gagttacatg atccccatg ttgtgcaaaa aagcggtag	8400
ctccttcggg cctccgatcg ttgtcagaag taagttggcc gcagtgttat cactcatggt	8460
tatggcagca ctgcataatt ctcttactgt catgccatcc gtaagatgct tttctgtgac	8520
tggtagtac tcaaccaagt cattctgaga atagtgtatg cggcgaccga gttgctcttg	8580
cccggcgtca atacgggata ataccgcgcc acatagcaga actttaaaag tgctcatcat	8640
tggaacacgt tcttcggggc gaaaactctc aaggatctta ccgctgttga gatccagttc	8700
gatgtaacc actcgtgcac ccaactgatc ttcagcatct tttactttca ccagcgtttc	8760
tgggtgagca aaaacaggaa ggcaaaatgc cgcaaaaaag ggaataaggg cgacacggaa	8820
atgttgaata ctcatactct tcctttttca atattattga agcatttatc agggttattg	8880
tctcatgagc ggatacatat ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg	8940
cacatttccc cgaaaagtgc cacctgacgt ctaagaaacc attattatca tgacattaac	9000
ctataaaaaat aggcgtatca cgaggccctt tcgtctcgcg cgtttcgggtg atgacgggtga	9060
aaacctctga cacatgcagc tcccggagac gggtcacagct tgtctgtaag cggatgccgg	9120
gagcagacaa gcccgtagg gcgcgtcagc ggggtgttggc ggggtgtcggg gctggcttaa	9180
ctatgaggca tcagagcaga ttgtactgag agtgcacat atgcgggtgtg aaataaccgca	9240
cagatgcgta aggagaaaat accgcatcag gcgccattcg ccattcaggc tgcgcaactg	9300
ttgggaaggg cgatcgggtgc gggcctcttc gctattacgc cagctggcga aagggggatg	9360

tgctgcaagg cgattaagtt gggtaacgcc agggttttcc cagtcacgac gttgtaaaac	9420
gacggccagt gaattcgtta agaccactt tcacatttaa gttgtttttc taatccgcat	9480
atgatcaatt caaggccgaa taagaaggct ggctctgcac cttggtgatc aaataattcg	9540
atagcttgtc gtaataatgg cggcatacta tcagtagtag gtgtttccct ttcttcttta	9600
gcgacttgat gctcttgatc ttccaatacg caacctaaag taaaatgccc cacagcgctg	9660
agtgcataata atgcattctc tagtgaaaaa cttgttggc ataaaaaggc taattgattt	9720
tcgagagttt catactgttt ttctgtaggc cgtgtacctt aatgtacttt tgctccatcg	9780
cgatgactta gtaaagcaca tctaaaactt ttagcgttat tacgtaaaaa atcttgccag	9840
ctttccctt ctaaagggca aaagtgagta tggcgcctat ctaacatctc aatggctaag	9900
gcgtcgagca aagcccgctt attttttaca tgccaataca atgtaggctg ctctacacct	9960
agcttctggg cgagtttacg ggttgttaaa ccttcgattc cgacctcatt aagcagctct	10020
aatgcgctgt taatcacttt acttttatct aatctagaca tcattaattc ctaatttttg	10080
ttgacactct atcattgata gagttatttt accactccct atcagtgata gagaaaagtg	10140
a	10141

<210> 23

<211> 10132

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 23

aatattcacg taaggaggta atagatgcag atcttcgaca ctttgcaatc aatgggccat	60
gagcaggttg tcctatgtag cgataagacc acgggtctgc gcgccattat cgctatacac	120
gatacatcct tagggccggc gcttggtggt acccgatatgt ggcagtatgc aactgacgac	180
gatgctatta ctgacgcact ccgtctgtct cggggcatga cctacaaagc tgcggtttct	240
ggcgtaaatc tgggcgggtgg taaagccgtt atcatcgga accctcacag tgataaaagc	300
gaagcgctgt ttcgcgctta cggcagaatg gtggaatccc agcgtgggcg ttacatcacc	360
gccgaagacg ttggtactag cgtacgtgat atggagtgga ttcgcatgga aaccaaatat	420
gtaacgggcg tgggtggcaa cggaggctct ggtgaccct ctccagttac cgctctgggt	480

gtttactcgg gcatgaaggc atgcgctaaa tcagtctatg gtactgatgc gctgagcgg	540
aaaaggatcg tggttcaggg cgcgggtaac gttgcatccc atctgggtca cagtctggta	600
aaagaaggcg ctgtgggttt cgtcactgac atctacgaag aaaaggccaa agcattagcg	660
gctgaaacgg gcgctaccgt gattcgcacc gacgaggttt ttactacaca atgcgatatc	720
ttctctccga acgctctggg ggccgtcctg aacgatgaaa ctattccgca gctcacatgc	780
gctatcgtag ctgggtgggtc aaacaatcag cttaaaatcg aacaacgtca cgccacggct	840
ctgcaagaga aaggcattct gtatgcgccg gattacgtaa tcaacgccgg gggcctcatg	900
aatgtggcga gcgaagttga cggctacaac cgtgaaaagg ttatgcgcca ggctgaaggt	960
atttacgata ttactatgaa catcctaaat accgcgcgtg agcgtaacat cctgaccatc	1020
gaagcatcca acgcgattgc tgaagagcgg atcaacaaag ttcgccatgt tcacgggaac	1080
ttcatcgggt ccccgctctat tcgcggagta taataaatca aaacaccaaa ttaaggagaa	1140
atthttatgtc gacgacaacc gttgggtgact acttgctgta tcgcttaaac gaaatcggca	1200
ttgagcacct cttcggagtgc ccaggtgatt acaatctgca atthtttgat catgtaatcg	1260
accaccctca gctgacttgg gtcgggtgca ctaacgaact taacgctgcc tacgcagctg	1320
atggttatgc gcgttgctgt ccggctgcgg cactgctgac caccttcggg gttggcgaac	1380
tgagcgctat taatggcatc gcaggttcct acgcggagta tctgccggta atacatatcg	1440
ttgggtgcacc gagtctatca gcccagcagc agggcgacct gattcaccac tctcttggcg	1500
aaggtgattt ttccagcttc ctgaggatgt cccaaccggg gtctgttgcg caggctgctc	1560
tgactcctga taacgcatgc aaggaaatcg accgcgtact ggcggaagtc ctcatcagc	1620
gtcgtcccgg ctacctgctg ctgtctaccg acgtggctgc tgcgccggcg gctctgccac	1680
aaagcactct ttctttgccg accgccccgg atcatcgcg agttctggct gctttcagcg	1740
acgctgctga gcagatgctg gctcaggcca aaagcgtctc tctactggcg gactttctgg	1800
ctgatcgttt cgggtgttact cgagcactgg ccgcgtggct tcagcagggt ccgctaccgc	1860
acgccactct gttaatgggt aaaggcggtt tgagtgaaca gcaaccaggg ttcgtgggta	1920
cctacgctgg tgcggcatct atcgattcga cgcgtggcgc aatcgaagaa gctggggtaa	1980
ttatcggagt gggagttaga ttttccgaca ctatcacagc aggttctctg cagcagatcg	2040
acgcccgcg ttttatagac attcaaccct tcttctctcg tattggcgat cgccagtttg	2100

atcacttgcc gatgcaggct gccgtcgcag ccctgcatca actgtgtctt cgttatcagc	2160
agcagtggtc tatcaccgct cctagcccgc ctgcaactgcc gccggctgct ggtagcgagc	2220
tgtcccagaa cgcattctgg caggcgatgc agaacttcat ccgccctggg gacctgttgg	2280
tggccgacca aggtactgcg gcgttcggcg cagcggcgct gcgcttaccg cagaattgcc	2340
agctgcttgt gcagccgctg tggggctcaa tcggttacag tctgccggcc acctttggtg	2400
ctcagacggc agatacagag cgtcgtgtaa tcctaatacat tggcgatggt tcagcgcaat	2460
taactattca ggaactttcc agtatgatgc gtgacggctt gaaacctatc atctttctcc	2520
tgaacaacaa cggttacacc gttgaacggg cgattcacgg cgcggagcaa cgttataacg	2580
atatcgctgc ttggaattgg acccaactgc cccaggcgct gagggttcat tgcccagcgc	2640
agagctggcg agtcgttgaa acggtgcagc tgaccgacgt aatgaaagtc atcgctgctt	2700
ctccgctct gagcttggtg gaagttgttc tgcctgcaat ggatgtcca ccgctgctgc	2760
aagcagttag tgccgctctg aaccagcgca actcctctta atttttaaaag taacacacga	2820
gctaagaact aaggaggatt tatgcgcagc atgcagtttg atgagtacgg tgcacccctg	2880
aaagcgttct catatgaaga cccgaccccg caagggaagg aagtagtcgt taggatcgaa	2940
gcctgtggtg tgtgccactc tgatattcat cttcacgagg gctacttcga catgggcggt	3000
ggcaataaag ctgatgttac tcgtgctcgc gaactccctt ttacattggg tcatgaaatc	3060
gttggcgaag tggtagcaac tggaccaggt gtcaccggcg ctaaaccggg cgacaaacgt	3120
attgtgtacc cgtggatcgg gtgcggcgac tgcccgaat gcaacagtgg tgaggatcag	3180
tcctgtgcgc gtccacgtaa cctgggtggt cacgttgacg gtggctattc gacgcacgta	3240
aagataccgg acgaaaaatt cctgttcgcc tacgatggta ttcctactga gttagcggga	3300
acctatgctt gcagcggcat caccgcttat ggtgcactga tgaaagcaaa ggaagcggct	3360
gaaagatctg gctacatcgg tctgattggc gctggtggcg ttggcatggc tggctctgatg	3420
ctggccaaag cagcgatcgg ggctaaaact gtagtctttg atatcgacga cgcaaaactg	3480
gaagctgcga cccgtgccgg ggcggattac gtgttcaact ccggtgcaaa agaaacacgc	3540
aaggaagtta tgaaactaac gaatggtggc ctgtctggtg ctgttgattt cgttggcagc	3600
gataaaagcg ctctgttttg aatcaacgcc ttgggtcaga acggcgtgct ggtcataatt	3660
ggactgttcg gtggcgctat gactgttccg gtaccctgt tcccgtgaa agggatcacc	3720

gtacgtggct catacgtagg ttccctgcaa gagatgagtg atatgatgga gttagttcgc	3780
gctgggaaag ttcctccgat gccggtaaaa actcggccac tggacgctgc ctgggaaacc	3840
cttgaggatc tacgccatgg taaaatcgtg ggccgtgttg ttctgacccc ataagaagga	3900
gatatacata tgacccatca attaagatcg cgcgatatca tcgctctggg ctttatgaca	3960
tttgcgttgt tcgtcggcgc aggtaacatt attttcctc caatggtcgg cttgcaggca	4020
ggcgaacacg tctggactgc ggcattcggc ttcctcatta ctgccgttgg cctaccggta	4080
ttaacggtag tggcgctggc aaaagttggc ggcggtgttg acagtctcag cacgccaatt	4140
ggtaaagtcg ctggcgctact gctggcaaca gtttgttacc tggcggtggg gccgcttttt	4200
gctacgccgc gtacagctac cgtttctttt gaagtgggca ttgcgccgct gacgggtgat	4260
tccgcgctgc cgctgtttat ttacagcctg gtctatttcg ctatcgttat tctggtttcg	4320
ctctatccgg gcaagctgct ggataccgtg ggcaacttcc ttgcgccgct gaaaattatc	4380
gcgctgggtca tcctgtctgt tgccgcaatt atctggccgg cgggttctat cagtacggcg	4440
actgaggctt atcaaaacgc tgcgttttct aacggcttcg tcaacggcta tctgaccatg	4500
gatacgctgg gcgcaatggg gtttggtatc gttattgtta acgcggcgcg ttctcgtggc	4560
gttaccgaag cgcgtctgct gaccggttat accgtctggg ctggcctgat ggcgggtgtt	4620
ggtctgactc tgctgtacct ggcgctgttc cgtctgggtt cagacagcgc gtcgctggtc	4680
gatcagtctg caaacggcgc ggcgatcctg catgcttacg ttcagcatac ctttggcggc	4740
ggcggtagct tcctgctggc ggcgttaatc ttcatcgcct gcctggtcac ggcggttggc	4800
ctgacctgtg cttgtgcaga attcttcgcc cagtacgtac cgctctctta tcgtacgtg	4860
gtgtttatcc tcggcggcct ctcgatggtg gtgtctaacc tcggcttgag ccagctgatt	4920
cagatctctg taccggtgct gaccgccatt tatccgccgt gtatcgcaact ggttgattta	4980
agttttacac gtcctgggtg gcataattcg tcccgcgtga ttgctccgcc gatgtttatc	5040
agcctgcttt ttggtattct cgacgggatc aaggcatctg cattcagcga tatcttaccg	5100
tcctgggcgc agcgtttacc gctggccgaa caaggctctg cgtgggttaat gccaacagtg	5160
gtgatgggtg ttctggccat tatctgggat cgtgcggcag gtcgtcaggt gacctccagc	5220
gctcactaat acgcatggca tggatgaccg atggtagtgt gggctctccc catgcgagag	5280
tagggaactg ccaggcatca aataaaacga aaggctcagt cgaaagactg ggcctttcgt	5340

tttatctgtt gtttgtcggg gaacgctctc ctgagtagga caaatccgcc gggagcggat	5400
ttgaacgttg cgaagcaacg gcccggaggg tggcgggcag gacgcccgcc ataaactgcc	5460
aggcatcaaa ttaagcagaa ggccatcctg acggatggcc tttttgcgtg gccagtgcc	5520
agcttgcatg cgtgccagct gcattaatga agaaatcatg ctggaagaat aacagctcac	5580
tcaaaggcgg tagtacgggt tttgctgccc gcaaacgggc tgttctgggtg ttgctagttt	5640
gttatcagaa tcgcagatcc ggcttcagcc ggtttgccgg ctgaaagcgc tatttcttcc	5700
agaattgcc	5760
tgattttttc cccacgggag gcgtcactgg ctcccgtgtt gtcggcagct	5760
ttgattcgat aagcagcatc gcctgtttca ggctgtctat gtgtgactgt tgagctgtaa	5820
caagttgtct caggtgttca atttcatgtt ctagttgctt tgttttactg gtttcacctg	5880
ttctattagg tgttacatgc tgttcatctg ttacattgtc gatctgttca tggatgaacag	5940
ctttgaatgc accaaaaact cgtaaaagct ctgatgtatc tatctttttt acaccgtttt	6000
catctgtgca tatggacagt tttccctttg atatgtaacg gtgaacagtt gttctacttt	6060
tgtttgttag tcttgatgct tcactgatag atacaagagc cataagaacc tcagatcctt	6120
ccgtatttag ccagtatgtt ctctagtgtg gttcgttggt tttgcgtgag ccatgagaac	6180
gaaccattga gatcatactt actttgcatg tcactcaaaa attttgcctc aaaactgggtg	6240
agctgaattt ttgcagttaa agcatcgtgt agtggtttttc ttagtccgtt atgtaggtag	6300
gaatctgatg taatggttgt tggatattttg tcaccattca tttttatctg gttgttctca	6360
agttcgggta cgagatccat ttgtctatct agttcaactt ggaaaatcaa cgtatcagtc	6420
gggcggcctc gcttatcaac caccaatttc atattgctgt aagtgtttta atctttactt	6480
attggtttca aaaccattg gtttaagcctt ttaaactcat ggtagttatt ttcaagcatt	6540
aacatgaact taaattcatc aaggctaata tctatatattg ccttgtaggt tttcttttgt	6600
gtagttctt ttaataacca ctcataaatc ctcatagagt atttgttttc aaaagactta	6660
acatgttcca gattatattt tatgaatttt tttaactgga aaagataagg caatatctct	6720
tcactaaaaa ctaattctaa tttttcgtt gagaacttgg catagtttgt ccactggaaa	6780
atctcaaagc ctttaaccaa aggatcctg atttccacag ttctcgtcat cagctctctg	6840
gttgctttag ctaatacacc ataagcattt tccctactga tgttcatcat ctgagcgtat	6900
tggttataag tgaacgatac cgtccgttct ttcctttag ggttttcaat cgtgggggtt	6960

agtagtgcca	cacagcataa	aattagcttg	gtttcatgct	ccgttaagtc	atagcgacta	7020
atcgctagtt	catttgcttt	gaaaacaact	aattcagaca	tacatctcaa	ttgggtctagg	7080
tgattttaat	cactatacca	attgagatgg	gctagtcaat	gataattact	agtccttttc	7140
ctttgagttg	tgggtatctg	taaattctgc	tagacctttg	ctggaaaact	tgtaaattct	7200
gctagaccct	ctgtaaattc	cgctagacct	ttgtgtgttt	tttttgttta	tattcaagtg	7260
gttataatth	atagaataaa	gaaagaataa	aaaaagataa	aaagaataga	tcccagccct	7320
gtgtataact	cactacttta	gtcagttccg	cagtattaca	aaaggatgtc	gcaaacgctg	7380
tttgctcctc	tacaaaacag	accttaaaac	cctaaaggct	taagtagcac	cctcgcaagc	7440
tcggggcaa	cgctgaatat	tccttttgtc	tccgaccatc	aggcacctga	gtcgtgtct	7500
ttttcgtgac	attcagttcg	ctgcgtcac	ggctctggca	gtgaatgggg	gtaaatggca	7560
ctacaggcgc	cttttatgga	ttcatgcaag	gaaactaccc	ataatacaag	aaaagcccgt	7620
cacgggcttc	tcagggcgtt	ttatggcggg	tctgctatgt	ggtgctatct	gactttttgc	7680
tgttcagcag	ttcctgccct	ctgattttcc	agtctgacca	cttcggatta	tcccgtgaca	7740
ggtcattcag	actggcta	gcaccagta	aggcagcgg	atcatcaaca	ggcttacccg	7800
tcttactgtc	ttttctacgg	ggcttgacgc	tcagtggaac	gaaaactcac	gttaagggat	7860
tttggctatg	agattatcaa	aaaggatctt	cacctagatc	cttttaaatt	aaaaatgaag	7920
ttttaaatca	atctaaagta	tatatgagta	aacttggtct	gacagttacc	aatgcttaat	7980
cagtgaggca	cctatctcag	cgatctgtct	atttcgttca	tccatagttg	cctgactccc	8040
cgtcgtgtag	ataactacga	tacgggaggg	cttaccatct	ggccccagtg	ctgcaatgat	8100
accgcgagaa	ccacgctcac	cggctccaga	tttatcagca	ataaaccagc	cagccggaag	8160
ggccgagcgc	agaagtgggtc	ctgcaacttt	atccgcctcc	atccagtcta	ttaattgttg	8220
ccgggaagct	agagtaagta	gttcgccagt	taatagtttg	cgcaacgttg	ttgccattgc	8280
tacaggcatc	gtggtgtcac	gctcgtcgtt	tggatatggct	tcattcagct	ccggttccca	8340
acgatcaagg	cgagttacat	gatcccccat	gttgtgcaaa	aaagcgggta	gctccttcgg	8400
tcctccgatc	gttgtcagaa	gtaagttggc	cgcagtgtta	tcactcatgg	ttatggcagc	8460
actgcataat	tctcttactg	tcatgccatc	cgtaagatgc	ttttctgtga	ctggtgagta	8520
ctcaaccaag	tcattctgag	aatagtgtat	gcggcgaccg	agttgctctt	gccccgcgtc	8580

aatacgggat	aataccgcgc	cacatagcag	aactttaaaa	gtgctcatca	ttggaaaacg	8640
ttcttcgggg	cgaaaactct	caaggatctt	accgctgttg	agatccagtt	cgatgtaacc	8700
cactcgtgca	cccaactgat	cttcagcatc	ttttactttc	accagcgttt	ctgggtgagc	8760
aaaaacagga	aggcaaatg	ccgcaaaaaa	gggaataagg	gcgacacgga	aatgttgaat	8820
actcatactc	ttcctttttc	aatattattg	aagcatttat	cagggttatt	gtctcatgag	8880
cggatacata	tttgaatgta	tttagaaaaa	taaacaaata	ggggttccgc	gcacatttcc	8940
ccgaaaagtg	ccacctgacg	tctaagaaac	cattattatc	atgacattaa	cctataaaaa	9000
taggcgtatc	acgaggccct	ttcgtctcgc	gcgtttcggg	gatgacgggtg	aaaacctctg	9060
acacatgcag	ctcccggaga	cggtcacagc	ttgtctgtaa	gcggatgccg	ggagcagaca	9120
agcccgtcag	ggcgcgtcag	cgggtgttgg	cgggtgtcgg	ggctggctta	actatgcggc	9180
atcagagcag	attgtactga	gagtgcacca	tatgcgggtg	gaaataccgc	acagatgcgt	9240
aaggagaaaa	taccgcatca	ggcgccattc	gccattcagg	ctgcgcaact	gttggaagg	9300
gcgatcgggtg	cgggcctctt	cgctattacg	ccagctggcg	aaagggggat	gtgctgcaag	9360
gcgattaagt	tgggtaacgc	cagggttttc	ccagtcacga	cgttgtaaaa	cgacggccag	9420
tgaattcggt	aagaccact	ttcacattta	agttgttttt	ctaataccga	tatgatcaat	9480
tcaaggccga	ataagaaggc	tggctctgca	ccttggtgat	caaataattc	gatagcttgt	9540
cgtaataatg	gcggcatact	atcagtagta	gggtgtttccc	tttcttcttt	agcgacttga	9600
tgctcttgat	cttccaatac	gcaacctaaa	gtaaaatgcc	ccacagcgct	gagtgcatat	9660
aatgcattct	ctagtgaaaa	accttgttgg	cataaaaagg	ctaattgatt	ttcgagagtt	9720
tcatactgtt	tttctgtagg	ccgtgtacct	aaatgtactt	ttgctccatc	gcgatgactt	9780
agtaaagcac	atctaaaact	tttagcggtta	ttacgtaaaa	aatcttgcca	gctttcccct	9840
tctaaagggc	aaaagtgagt	atggtgccta	tctaacatct	caatggctaa	ggcgtcgagc	9900
aaagcccgct	tatttttttac	atgccaatac	aatgtaggct	gctctacacc	tagcttctgg	9960
gcgagtttac	gggttggttaa	accttcgatt	ccgacctcat	taagcagctc	taatgcgctg	10020
ttaatcactt	tactttttatc	taatctagac	atcattaatt	cctaattttt	gttgacactc	10080
tatcattgat	agagttattt	taccactccc	tatcagtgat	agagaaaagt	ga	10132

<210> 24
<211> 10148
<212> DNA
<213> Artificial Sequence

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

<400> 24
aatattcacg taaggaggta atagatgcag atcttcgaca ctttgcaatc aatgggccat 60
gagcaggttg tcctatgtag cgataagacc acgggtctgc gcgccattat cgctatacac 120
gatacatcct tagggccggc gcttggtggt acccgatatgt ggcagtatgc aactgacgac 180
gatgctatta ctgacgcact ccgtctgtct cggggcatga cctacaaagc tgcggtttct 240
ggcgtaaatc tgggcgggtgg taaagccgtt atcatcgga accctcacag tgataaaagc 300
gaagcgctgt ttcgcgctta cggcagaatg gtggaatccc agcgtgggcg ttacatcacc 360
gccgaagacg ttggtactag cgtacgtgat atggagtgga ttcgcatgga aaccaaatat 420
gtaacgggcg tgggtggcaa cggaggctct ggtgaccct ctccagttac cgctctgggt 480
gtttactcgg gcatgaaggc atgcgctaaa tcagtctatg gtactgatgc gctgagcgggt 540
aaaaggatcg tggttcaggg cgcgggtaac gttgcatccc atctggttca cagtctggta 600
aaagaaggcg ctgtggtttt cgtcactgac atctacgaag aaaaggccaa agcattagcg 660
gctgaaacgg gcgctaccgt gattcgcacc gacgaggttt ttactacaca atgcgatatc 720
ttctctccga acgctctggg ggccgtcctg aacgatgaaa ctattccgca gtcacatgc 780
gctatcgtag ctggtggtgc aaacaatcag cttaaaatcg aacaacgtca cgccacggct 840
ctgcaagaga aaggcattct gtatgcgccg gattacgtaa tcaacgccgg gggcctcatg 900
aatgtggcga gcgaagttga cggctacaac cgtgaaaagg ttatgcgcca ggctgaaggt 960
atttacgata ttactatgaa catcctaaat accgcgcgtg agcgtaacat cctgaccatc 1020
gaagcatcca acgcgattgc tgaagagcgg atcaacaaag ttcgccatgt tcacgggaac 1080
ttcatcggtt ccccgcttat tcgcggagta taataaatca aaacaccaaa ttaaggagaa 1140
attttatgtc gacgacaacc gttggtgact acttgctgta tcgcttaaac gaaatcggca 1200
ttgagcacct cttcggagtg ccaggtgatt acaatctgca atttctggat catgtaatcg 1260
accaccctca gctgacttgg gtcggctgca ctaacgaact taacgctgcc tacgcagctg 1320

atggttatgc gcgttgtcgt ccggctgcgg cactgctgac caccttcggg gttggcgaac	1380
tgagcgctat taatggcatc gcaggttcct acgcggagta tctgccggta atacatatcg	1440
ttggtgcacc gagtctatca gcccagcagc agggcgacct gattcaccac tctcttggcg	1500
aaggtgattt ttccagcttc ctgaggatgt cccaaccggt gtctgttgcg caggctgctc	1560
tgactcctga taacgcatgc aaggaaatcg accgcgtact ggcggaagtc ctcatcagc	1620
gtcgtcccgg ctacctgctg ctgtctaccg acgtggctgc tgcgccggcg gctctgccac	1680
aaagcactct ttctttgccg accgccccgg atcatcgcg agttctggct gctttcagcg	1740
acgctgctga gcagatgctg gctcaggcca aaagcgtctc tctactggcg gactttctgg	1800
ctgatcgttt cgggtgttact cgagcactgg ccgcgtggct tcagcaggtt ccgctaccgc	1860
acgccactct gttaatgggt aaaggcggtt tgagtgaaca gcaaccaggg ttcgtgggta	1920
cctacgctgg tgcggcatct atcgattcga cgcgtggcgc aatcgaagaa gctggggtaa	1980
ttatcggagt gggagttaga ttttccgaca ctatcacagc aggcttctcg cagcagatcg	2040
acgcccgcg ttttatagac attcaaccct tcttctctcg tattggcgat cgccagtttg	2100
atcacttgcc gatgcaggct gccgtcgcag ccctgcatca actgtgtctt cgttatcagc	2160
agcagtggtc tatcaccgct cctagcccgc ctgcactgcc gccggctgct ggtagcgagc	2220
tgtcccagaa cgcattctgg caggcgatgc agaacttcat ccgccctggg gacctgttgg	2280
tggccgacca aggtactgcg gcgttcggcg cagcggcgct gcgcttaccg cagaattgcc	2340
agctgcttgt gcagccgctg tggggctcaa tcggttacag tctgccggcc acctttggtg	2400
ctcagacggc agatacagag cgtcgtgtaa tcctaatacat tggcgatggg tcagcgcaat	2460
taactattca ggaactttcc agtatgatgc gtgacggctt gaaacctatc atctttctcc	2520
tgaacaacaa cggttacacc gttgaacggg cgattcacgg cgcgagcaa cgttataacg	2580
atatcgctgc ttggaattgg acccaactgc ccaggcgct gagtgttcat tgcccagcgc	2640
agagctggcg agtcgttgaa acggtgcagc tgaccgacgt aatgaaagtc atcgctgctt	2700
ctccgctct gagcttggtg gaagttgttc tgcctgcaat ggatgtccca ccgctgctgc	2760
aagcagtgag tgccgctctg aaccagcgca actcctctta attggataaa gaataaggag	2820
gtattttatg acagcggagc agcaaatgg ggtatccgac tcacgccgtt tcgaatttca	2880
ggaatttggt ggccctatcg cccacagac ctatcagctc cccgcaccgg ctagcgatga	2940

agttttgtta aaggtgaact actgcggtgt ctgtcacagt gatgttcac	ttcacgacgg	3000			
ctacttcgag ctgggtggcg ataaacgtct gaacttcgct atgccgctgc	cgctgacgct	3060			
gggtcacgaa gtaattggca ccgttgtggc tgtcggcgac caggttactg	gtgtaaaacc	3120			
gggggaccag cgactgatct atccgtggat aggttgcgga aaatgcggcg	cgtgtcaaaa	3180			
aggagaagaa aacctgtgcg ttactcctgc acatctgggc gtgaacaagc	cgggcggtta	3240			
cgctgatcac atcgttgtac cccattctcg ctaccttctg gacatttcgg	gtctgaaccc	3300			
gggtgatgcc gctaccctcg cgtgctccgg cctgaccact ttcagcgcg	tcaacaaagt	3360			
gttgccgctt gcagatgacc agtggattgt tgttatcggc	tgtgggtggc	tcggccagat	3420		
ggcgctgcgt atcctgcaag ctatgggaat tggcaatgtt atcggtat	acctgtctga	3480			
agagaaacgg aaactggctc atgaaagcgg tgcacgtcac tccttcgatc	caaacactcc	3540			
gaagctgaac cgcgtggctg ccgaaacctg cccgggtacg gtacaggccg	cgtagactt	3600			
tgtgggcaat gagcaaactg ctcagctggc actgtctctg cttggaaaag	gtggcaaata	3660			
tgttcctgtc gggctgcacg gcggcgagct gcgttaccca ttgccgatca	tcacgaacaa	3720			
agctgtaagt atcatcggtt cttacgttgg taccctgaaa gaactggaag	acttagttgc	3780			
tttcgccaag gaaaagaatc tgccgccaat tcatattgaa caccgcccgc	tggaatcggc	3840			
ggctcaggcc gtagaggacc tggaaaaagg acaggttgct gggcgtgtta	tcctggatgc	3900			
aggtaactaa gaaggagata tacatatgac ccatcaatta agatcgcgcg	atatcatcgc	3960			
tctgggcttt atgacatttg cgttgttcgt cggcgcaggt aacattattt	tccctccaat	4020			
ggtcggcttg caggcaggcg aacacgtctg gactgcggca ttcggcttcc	tcattactgc	4080			
cgttggccta ccggtattaa cggtagtggc gctggcaaaa gttggcggcg	gtgttgacag	4140			
tctcagcacg ccaattggta aagtcgctgg cgtactgctg gcaacagttt	gttacctggc	4200			
ggtaggggccc ctttttgcta cgccgcgtac agctaccgtt tcttttgaag	tgggcattgc	4260			
gccgctgacg ggtgattccg cgctgccgct gtttatttac agcctggctt	atttcgctat	4320			
cgttattctg gtttcgctct atccgggcaa gctgctggat accgtgggca	acttccttgc	4380			
gccgctgaaa attatcgcg	tggtcacct	gtctgttgcc	gcaattatct	ggccggcggg	4440
ttctatcagt acggcgactg	aggcttatca	aaacgctgcg	ttttctaacg	gcttcgtcaa	4500
cggctatctg accatggata	cgctgggcgc	aatggtgttt	ggtatcgta	ttgttaacgc	4560

ggcgcgttct cgtggcggtta ccgaagcgcg tctgctgacc cgttataaccg tctgggctgg	4620
cctgatggcg ggtgttggtc tgactctgct gtacctggcg ctgttccgtc tgggttcaga	4680
cagcgcgtcg ctggtcgatac agtctgcaaa cgggtcggcg atcctgcatg cttacgttca	4740
gcataccttt ggcgggcgcg gtagcttcct gctggcgggcg ttaatcttca tcgcctgcct	4800
ggtcacggcg gttggcctga cctgtgcttg tgcagaattc ttcgcccagt acgtaccgct	4860
ctcttatcgt acgctggtgt ttatcctcgg cggcttctcg atggtggtgt ctaacctcgg	4920
cttgagccag ctgattcaga tctctgtacc ggtgctgacc gccatttatc cgccgtgtat	4980
cgcactgggtt gtattaagtt ttacacgctc atggtggcat aattcgtccc gcgtgattgc	5040
tccgccgatg tttatcagcc tgcttttttg tattctcgac gggatcaagg catctgcatt	5100
cagcgatatc ttaccgtcct gggcgcagcg tttaccgctg gccgaacaag gtctggcgtg	5160
gttaatgcca acagtgggtga tgggtggttct ggccattatc tgggatcgtg cggcaggctg	5220
tcaggtgacc tccagcgctc actaatacgc atggcatgga tgaccgatgg tagtgtgggc	5280
tctccccatg cgagagtagg gaactgccag gcatcaaata aaacgaaagg ctcagtcgaa	5340
agactgggcc tttcgtttta tctgttggtt gtcggtgaac gctctcctga gtaggacaaa	5400
tccgccggga gcggatttga acgttgcgaa gcaacggccc ggagggtggc gggcaggacg	5460
cccgccataa actgccaggc atcaaattaa gcagaaggcc atcctgacgg atggcctttt	5520
tgcgtggcca gtgccaagct tgcatacgtg ccagctgcat taatgaagaa atcatgctgg	5580
aagaataaca gctcactcaa aggcggtagt acgggttttg ctgccgcaa acgggctgtt	5640
ctggtgttgc tagtttggtt tcagaatcgc agatccggct tcagccgggt tgccggctga	5700
aagcgctatt tcttccagaa ttgccatgat tttttccca cgggaggcgt cactggctcc	5760
cgtgttgctg gcagctttga ttcgataagc agcatcgcct gtttcaggct gtctatgtgt	5820
gactgttgag ctgtaacaag ttgtctcagg tgttcaattt catgttctag ttgctttgtt	5880
ttactggttt cacctgttct attaggtgtt acatgctgtt catctgttac attgtcgatc	5940
tgttcatggg gaacagcttt gaatgcacca aaaactcgta aaagctctga tgtatctatc	6000
ttttttacac cgttttcatc tgtgcatatg gacagttttc cttttgatat gtaacgggtga	6060
acagttgttc tacttttggt tgtagtctt gatgcttcac tgatagatac aagagccata	6120
agaacctcag atccttccgt atttagccag tatgttctct agtgtgggtt gttgtttttg	6180

cgtgagccat gagaacgaac cattgagatc atacttactt tgcattgtcac tcaaaaattt	6240
tgcctcaaaa ctgggtgagct gaatttttgc agttaaagca tcgtgtagtg tttttcttag	6300
tccgttatgt aggttaggaat ctgatgtaat ggttggttggt attttgtcac cattcatttt	6360
tatctgggtg ttctcaagtt cggttacgag atccatttgt ctatctagtt caacttgga	6420
aatcaacgta tcagtcgggc ggcctcgctt atcaaccacc aatttcatat tgctgtaagt	6480
gtttaaatct ttacttattg gtttcaaaac ccattgggta agccttttaa actcatggta	6540
gttattttca agcattaaca tgaacttaaa ttcatcaagg ctaatctcta tatttgcctt	6600
gtgagttttc ttttgtgta gttcttttaa taaccactca taaatcctca tagagtattt	6660
gttttcaaaa gacttaacat gttccagatt atattttatg aattttttta actggaaaag	6720
ataaggcaat atctcttcac taaaaactaa ttctaatttt tcgcttgaga acttggcata	6780
gtttgtccac tggaaaatct caaagccttt aaccaaagga ttcctgattt ccacagttct	6840
cgtcatcagc tctctgggtg ctttagctaa tacaccataa gcattttccc tactgatgtt	6900
catcatctga gcgtattggt tataagtga cgateaccgtc cgttctttcc ttgtaggggt	6960
ttcaatcgtg gggttgagta gtgccacaca gcataaaatt agcttggttt catgctccgt	7020
taagtcatag cgactaatcg ctagttcatt tgctttgaaa acaactaatt cagacataca	7080
tctcaattgg tctaggtgat tttaatcact ataccaattg agatgggcta gtcaatgata	7140
attactagtc cttttccttt gagttgtggg tatctgtaaa ttctgctaga ctttgctgg	7200
aaaacttgta aattctgcta gaccctctgt aaattccgct agaccttgt gtgttttttt	7260
tgtttatatt caagtggta taatttatag aataaagaaa gaataaaaaa agataaaaaag	7320
aatagatccc agccctgtgt ataactcact acttttagtca gttccgcagt attacaaaag	7380
gatgtcgcaa acgctgtttg ctcctctaca aaacagacct taaaacccta aaggcttaag	7440
tagcaccctc gcaagctcgg gcaaactcgt gaatattcct tttgtctccg accatcaggc	7500
acctgagtcg ctgtcttttt cgtgacattc agttcgctgc gctcacggct ctggcagtga	7560
atgggggtaa atggcactac aggcgccttt tatggattca tgcaaggaaa ctaccataa	7620
tacaagaaaa gcccgtcacg ggcttctcag ggcgttttat ggcgggtctg ctatgtggtg	7680
ctatctgact ttttgctgtt cagcagttcc tgccctctga ttttccagtc tgaccattc	7740
ggattatccc gtgacaggtc attcagactg gctaatagcac ccagtaaggc agcggtatca	7800

tcaacaggct taccgctctt actgtctttt ctacggggtc tgacgctcag tggaacgaaa	7860
actcacgtta agggattttg gtcattgagat tatcaaaaag gatcttcacc tagatccttt	7920
taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagttaaact tggctctgaca	7980
gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcattcca	8040
tagttgcctg actccccgtc gtgtagataa ctacgatacg ggaggggctta ccatctggcc	8100
ccagtgcctgc aatgataccg cgagaaccac gctcaccggc tccagattta tcagcaataa	8160
accagccagc cggaaggggc gagcgcagaa gtggctcctgc aacttttatcc gcctccatcc	8220
agtctattaa ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca	8280
acgttgctgc cattgctaca ggcatcgtgg tgtcacgctc gtcgtttggg atggcttcatt	8340
tcagctccgg ttcccaacga tcaaggcgag ttacatgata ccccatgttg tgcaaaaaag	8400
cggtagctc cttcggctct ccgatacgtt tcagaagtaa gttggccgca gtgttatcac	8460
tcattggttat ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt	8520
ctgtgactgg tgagtactca accaagtcatt tctgagaata gtgtatgcgg cgaccgagtt	8580
gctcttgccc ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc	8640
tcattcattgg aaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat	8700
ccagttcgat gtaaccact cgtgcaccca actgatcttc agcatctttt actttcacca	8760
gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataaggcgca	8820
cacggaaatg ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg	8880
gttattgtct catgagcgga tacatatttg aatgtattta gaaaaataaa caaatagggg	8940
ttccgcgcac atttccccga aaagtgccac ctgacgtcta agaaaccatt attatcatga	9000
cattaacctt taaaaatagg cgtatcacga ggccctttcg tctcgcgcgt ttcggtgatg	9060
acggtgaaaa cctctgacac atgcagctcc cggagacggg cacagcttgt ctgtaagcgg	9120
atgccgggag cagacaagcc cgtcagggcg cgtcagcggg tgttgggcgg tgtcggggct	9180
ggcttaacta tgcggcatca gagcagattg tactgagagt gcacatatg cgggtgtgaaa	9240
taccgcacag atgcgtaagg agaaaatacc gcatcaggcg ccattcgcca ttcaggctgc	9300
gcaactgttg ggaaggcgga tcggtgcggg cctcttcgct attacgccag ctggcgaaag	9360
ggggatgtgc tgcaaggcga ttaagttggg taacgccagg gttttcccag tcacgacgtt	9420

gtaaaacgac ggccagtga ttcgttaaga cccactttca catttaagtt gtttttctaa	9480
tccgcatatg atcaattcaa ggccgaataa gaaggctggc tctgcacctt ggtgatcaaa	9540
taattcgata gcttgtcgta ataatggcgg catactatca gtagtaggtg tttccctttc	9600
ttcttttagcg acttgatgct cttgatcttc caatacgcaa cctaaagtaa aatgccccac	9660
agcgctgagt gcatataatg cattctctag tgaaaaacct tgttggcata aaaaggctaa	9720
ttgattttcg agagtttcat actgtttttc tgtaggccgt gtacctaaat gtacttttgc	9780
tccatcgca tgacttagta aagcacatct aaaactttta gcgttattac gtaaaaaatc	9840
ttgccagctt tcccccttcta aagggcaaaa gtgagtatgg tgcctatcta acatctcaat	9900
ggctaaggcg tcgagcaaag cccgcttatt ttttacatgc caatacaatg taggctgctc	9960
tacacctagc ttctgggcga gtttacgggt tgtaaacct tcgattccga cctcatthaag	10020
cagctctaata gcgctgttaa tcactttact tttatctaata ctagacatca ttaattccta	10080
atttttgttg acactctatc attgatagag ttattttacc actccctatc agtgatagag	10140
aaaagtga	10148

<210> 25

<211> 5421

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 25

gttgacactc tatcattgat agagttatth taccactccc tatcagtgat agagaaaagt	60
gagcgaaaaa tcaataagga ggcaacaaga tgtgcgaaaa acatcttaata catgctaagg	120
aggttttcta atgaacatct tcaagaaaat ggaggaatth aattatgaac aactgggtcta	180
cttctacgac agcgaaacgg aactcaaagg tattacctgt atacacaaca caactttagg	240
gccggcattg ggcggtaccc gccttttgaa ctataactct gaggaagatg ccgttgaaga	300
cgtaatccgt ctggctcggg gcatgactta caaagcggct tgcgccggtc tgaatctggg	360
cggcggtaaa accgtgctga tcggtgatgc taaaaagatt aaatcagagt cctacttccg	420
tggactgggg cgctacgttc agtcgctgaa cggcagatat atcaccgcgg aagacgtaaa	480
tacttctacg aaggatatgg catacgttgc tatggaaact gactatgtgg taggcctggg	540

aggtaaatcc	ggcaacccta	gtccagttac	tgcttacggt	gcatttatgg	gtatcaaagc	600
ggcgctgatg	aaaaaatttg	aggatagctc	tattgaaggc	cgaaccttcg	cagtgcaggg	660
tgctgggcag	acgggttact	atcttatcga	ttacctccta	ggcaacaaca	agttcaaaga	720
aaaggctaaa	aagatttact	tcaccgaaat	taacgagagc	tatatcgagc	gtatgaacaa	780
agaacatccg	gaagttgaat	ttattttccc	ggacaaaatc	tactcgctgg	aagtagacgt	840
cttcgtgccc	tgcgccctgg	gcaaaatcgt	taatgacaaa	actatcgatg	aatttaagt	900
tccgatcatc	gcaggtactg	caaacaacgt	actggaaagg	gaagcgcacg	gcaacatgct	960
taaagaacgt	ggcattcttt	acgccccgga	ctatgtgatc	aatgctgggtg	ggctgatcaa	1020
cgtttaccac	gagctgaacg	gttacaataa	agagaacgct	attctggaag	tggaattaat	1080
ttatgatcgc	ctactggaaa	tattcaacat	cgctgattct	ctgaacatca	gcaccaatat	1140
cgctgccaac	gagttcgcgg	aaaaacgtat	caagcaaatt	aagtccttga	aaaacaactt	1200
cattaaacgc	taagtcaata	aaggcatata	aaaggagggt	aataacatga	aagttaaagt	1260
aaaacatctt	aatcatgcac	aggagacttt	ctaattgtcg	cgacaaccgt	tggtgactac	1320
ttgctgtatc	gcttaaacga	aatcggcatt	gagcacctct	tcggagtgcc	aggtgattac	1380
aatctgcaat	ttctggatca	tgtaatcgac	caccctcagc	tgacttgggt	cggctgcact	1440
aacgaactta	acgctgccta	cgcagctgat	ggttatgcgc	gttgctgtcc	ggctgcggca	1500
ctgctgacca	ccttcgggggt	tggcgaactg	agcgctatta	atggcatcgc	aggttcctac	1560
gcggagtatc	tgccggtaat	acatatcggt	ggtgcaccga	gtctatcagc	ccagcagcag	1620
ggcgacctga	ttcaccactc	tcttggcgaa	ggtgattttt	ccagcttcct	gaggatgtcc	1680
caaccggtgt	ctgttgcgca	ggctgctctg	actcctgata	acgcatgcaa	ggaaatcgac	1740
cgcgtactgg	cggaagtcct	cattcagcgt	cgtcccggct	acctgctgct	gtctaccgac	1800
gtggctgctg	cgccggcggc	tctgccacaa	agcactcttt	ctttgccgac	cgtcccggat	1860
catcgcgcag	ttctggctgc	tttcagcgac	gctgctgagc	agatgctggc	tcaggccaaa	1920
agcgtctctc	tactggcgga	ctttctggct	gatcgtttcg	gtgttactcg	agcactggcc	1980
gcgtggcttc	agcaggttcc	gctaccgcac	gccactctgt	taatgggtaa	aggcgttctg	2040
agtgaacagc	aaccagggtt	cgtgggtacc	tacgctgggtg	cggcatctat	cgattcgacg	2100
cgtggcgcaa	tcgaagaagc	tggggtaatt	atcggagtgg	gagttagatt	ttccgacact	2160

atcacagcag gcttctcgca gcagatcgac gcccgccgtt ttatagacat tcaacccttc	2220
ttctctcgta ttggcgatcg ccagtttgat cacttgccga tgcaggctgc cgtcgcagcc	2280
ctgcatcaac tgtgtcttcg ttatcagcag cagtgggtcta tcaccgctcc tagcccgctt	2340
gcactgccgc cggctgctgg tagcgagctg tcccagaacg cattctggca ggcatgcag	2400
aacttcatcc gccctgggga cctgttggtg gccgaccaag gtactgcggc gttcggcgca	2460
gcggcgctgc gcttaccgca gaattgccag ctgcttgctg agccgctgtg gggctcaatc	2520
ggttacagtc tgccggccac ctttggtgct cagacggcag atacagagcg tcgtgtaatc	2580
ctaatcattg gcgatggttc agcgcaatta actattcagg aactttccag tatgatgcgt	2640
gacggcttga aacatatcat ctttctcctg aacaacaacg gttacaccgt tgaacgggcg	2700
attcacggcg cggagcaacg ttataacgat atcgctgctt ggaattggac ccaactgccc	2760
caggcgctga gtgttcattg cccagcgag agctggcgag tcgttgaaac ggtgcagctg	2820
accgacgtaa tgaaagtcac cgctgcttct ccgcgtctga gcttggtaga agttgttctg	2880
cctgcaatgg atgtcccacc gctgctgcaa gcagtgagtg ccgctctgaa ccagcgcaac	2940
tcctcttaag accagaggta aggaggtaac aaccatgcga gtgttgaaga aacatcttaa	3000
tcatgcaggg gagggtttct aatgcagacg gcgttcttgt ataagccagg tcacgaaaac	3060
ttagtgcgct cggagatccc gatacctaaa gctgggcgtg gcgaagtcgt tctggaaatt	3120
aaagccgctg gcatgtgcca ttccgatctg cacgttctcg acggtggaat ccccctgccg	3180
ggtcaatttg taatgggcca tgaaatcggt ggtactattc acgagatcgg ccaggacgtg	3240
accggtttca aacagggcga tctgtacgca gtccacggcc cgaatccgtg tggatattgc	3300
accctgtgca gagaaggatt tgataacgac tgcactacag tggcgaaaac cggatcaatgg	3360
ttcggactgg gtcttgacgg cggctaccag aagtatatcc gtatcccgaa cgtaaggctt	3420
atcgtaaag ttccagaagg tgtttcagct gaggcagctg cgagctgtac tgatgcagta	3480
ctgaccccgt accgtgcact aaaacaggct ggccgagca actctactcg ggtactgatt	3540
ctgggtctgg gtggcttagg tctgaatgcc cttaaactgg ctaagacctt cggcagttac	3600
gtttacgcat ctgacctgaa accttctgcg cgtgaagctg ctaaggccgc tggggcggat	3660
gaagtgcctg agtccctgcc cgaagaccg ctgggtgttg atatcgtgtt agacgtcgtt	3720
ggcgtgcaga gcaccttcaa cctcgtcaa aaacacgttg gccgcgtgg catcattgta	3780

cctgtaggcc	tggcatcccc	acagcttttcg	tttaacctaa	cggatctggc	gctccgcgaa	3840
attcgtgttc	agggcacttt	ttggggcacg	agcaatgagc	tggctgaatg	tctgcgctg	3900
tgccagctgg	gcctgatcaa	cccgaaatat	actgtggtgc	ctcttgaaga	agcgccgaaa	3960
tatatggaag	caatggctca	tgggaaagta	gaaggtcgta	tcgttttcca	cccgtaagaa	4020
ggagatatac	atatgaccca	tcaattaaga	tcgcgcgata	tcatcgctct	gggctttatg	4080
acatttgcgt	tgttcgtcgg	cgcaggtaac	attattttcc	ctccaatgg	cggcttgacg	4140
gcaggcgaac	acgtctggac	tgcggcattc	ggcttcctca	ttactgccgt	tggcctaccg	4200
gtattaacgg	tagtggcgct	ggcaaaagtt	ggcggcggtg	ttgacagtct	cagcacgcca	4260
attggtaaag	tcgctggcgt	actgctggca	acagtttggt	acctggcggt	ggggccgctt	4320
tttgctacgc	cgcgtacagc	taccgtttct	tttgaagtgg	gcattgcgcc	gctgacgggt	4380
gattccgcgc	tgccgctggt	tatttacagc	ctggctctatt	tcgctatcgt	tattctggtt	4440
tcgctctatc	cgggcaagct	gctggatacc	gtgggcaact	tccttgcgcc	gctgaaaatt	4500
atcgcgctgg	tcatcctgtc	tgttgccgca	attatctggc	cggcgggttc	tatcagtacg	4560
gcgactgagg	cttatcaaaa	cgtgcgtttt	tctaacggct	tcgtcaacgg	ctatctgacc	4620
atggatacgc	tgggcgcaat	ggtgtttggt	atcgttattg	ttaacgcggc	gcgttctcgt	4680
ggcgttaccg	aagcgcgtct	gctgaccctg	tataccgtct	gggctggcct	gatggcgggt	4740
gttggctctga	ctctgctgta	cctggcgctg	ttccgtctgg	gttcagacag	cgcgtcgctg	4800
gtcgatcagt	ctgcaaacgg	tgcggcgatc	ctgcatgctt	acgttcagca	tacctttggc	4860
ggcggcggtg	gcttcctgct	ggcggcggtt	atcttcatcg	cctgcctggt	cacggcggtt	4920
ggcctgacct	gtgcttgtgc	agaattcttc	gcccagtagc	taccgctctc	ttatcgtagc	4980
ctggtgttta	tcctcggcgg	cttctcgatg	gtggtgtcta	acctcggctt	gagccagctg	5040
attcagatct	ctgtaccggt	gctgaccgcc	atztatccgc	cgtgtatcgc	actggttgta	5100
ttaagtttta	cacgctcatg	gtggcataat	tcgtcccgcg	tgattgctcc	gccgatgttt	5160
atcagcctgc	tttttggtat	tctcgacggg	atcaaggcat	ctgcattcag	cgatatctta	5220
ccgtcctggg	cgcagcgttt	accgctggcc	gaacaaggct	tggcgtgggt	aatgccaaca	5280
gtggtgatgg	tggttctggc	cattatctgg	gatcgtgcgg	caggctgtca	ggtgacctcc	5340
agcgctcact	aatacgcatg	gcatggatga	ccgatggtag	tgtgggctct	ccccatgcga	5400

gagtagggaa ctgccaggca t

5421

<210> 26

<211> 5478

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 26

gttgacactc tatcattgat agagttat	60
ttt taccactccc tatcagtgat agagaaaagt	
gagcgaaaaa tcaataagga ggcaacaaga	120
tgtgcgaaaa acatcttaat catgctaagg	
aggttttcta atgaacatct tcaagaaaat	180
ggaggaat	
ttt aattatgaac aactgggtcta	
cttctacgac agcgaaacgg aactcaaagg	240
tattacctgt atacacaaca caactttagg	
gccggcattg ggcggtaccc gcctttggaa	300
ctataactct gaggaagatg ccgttgaaga	
cgtaatccgt ctggctcggg gcatgactta	360
caaagcggct tgcgccggtc tgaatctggg	
cggcggtaaa accgtgctga tcggtgatgc	420
taaaaagatt aaatcagagt cctacttccg	
tggaactgggg cgctacgttc agtcgctgaa	480
cggcagatat atcaccgcgg aagacgtaaa	
tacttctacg aaggatatgg catacgttgc	540
tatggaaact gactatgtgg taggcctggg	
aggtaaatcc ggcaacccta gtccagttac	600
tgcttacggg gcatttatgg gtatcaaagc	
ggcgctgatg aaaaaatttg aggatagctc	660
tattgaaggc cgaaccttcg cagtgcaggg	
tgctgggcag acgggttact atcttatcga	720
ttacctccta ggcaacaaca agttcaaaga	
aaaggctaaa aagatttact tcaccgaaat	780
taacgagagc tatatcgagc gtatgaacaa	
agaacatccg gaagttgaat ttatttcccc	840
ggacaaaatc tactcgctgg aagtagacgt	
cttcgtgccc tgcgccctgg gcaaaatcgt	900
taatgacaaa actatcgatg aatttaagt	
tccgatcatc gcaggtactg caaacaacgt	960
actggaaagg gaagcgcacg gcaacatgct	
taaagaacgt ggcattcttt acgccccgga	1020
ctatgtgatc aatgctgggt ggctgatcaa	
cgtttaccac gagctgaacg gttacaataa	1080
agagaacgct attctggaag tggaattaat	
ttatgatcgc ctactggaaa tattcaacat	1140
cgctgattct ctgaacatca gcaccaatat	
cgctgccaac gagttcgcgg aaaaacgtat	1200
caagcaaatt aagtccttga aaaacaactt	

cattaaacgc taagtcaata aaggcatata aaaggagggtt aataacatga aagttaaagt	1260
aaaacatctt aatcatgcac aggagacttt ctaatgtcga cgacaaccgt tgggtgactac	1320
ttgctgtatc gcttaaacga aatcggcatt gagcacctct tcggagtgcc aggtgattac	1380
aatctgcaat ttctggatca tgtaatcgac caccctcagc tgacttgggt cggctgcact	1440
aacgaactta acgctgccta cgcagctgat ggttatgctc gttgtcgtcc ggctgcggca	1500
ctgctgacca ctttcgggggt tggcgaactg agcgtatta atggcatcgc aggttcctac	1560
gcggagtatc tgccggtaat acatatcggt ggtgcaccga gtctatcagc ccagcagcag	1620
ggcgacctga ttcaccactc tcttggcgaa ggtgattttt ccagcttcct gaggatgtcc	1680
caaccggtgt ctgttgcgca ggctgctctg actcctgata acgcatgcaa ggaaatcgac	1740
cgcgtactgg cggaagtcct cattcagcgt cgtcccggt acctgctgct gtctaccgac	1800
gtggctgctg cgccggcggc tctgccacaa agcactcttt ctttgccgac cgccccggat	1860
catcgcgag ttctggctgc tttcagcgac gctgctgagc agatgctggc tcaggccaaa	1920
agcgtctctc tactggcgga ctttctggct gatcgtttcg gtgttactcg agcactggcc	1980
gcgtggcttc agcaggttcc gctaccgcac gccactctgt taatgggtaa aggcgttctg	2040
agtgaacagc aaccagggtt cgtgggtacc tacgctgggt cggcattctat cgattcgacg	2100
cgtggcgcaa tcgaagaagc tggggtaatt atcggagtgg gagttagatt ttccgacact	2160
atcacagcag gcttctcgca gcagatcgac gcccgccgtt ttatagacat tcaacccttc	2220
ttctctcgta ttggcgatcg ccagtttgat cacttgccga tgcaggctgc cgtcgcagcc	2280
ctgcatcaac tgtgtcttcg ttatcagcag cagtgggtcta tcaccgctcc tagcccgctt	2340
gcactgccgc cggctgctgg tagcgagctg tcccagaacg cattctggca ggcatgacg	2400
aacttcatcc gccctgggga cctgttggtg gccgaccaag gtactgcggc gttcggcgca	2460
gcggcgctgc gcttaccgca gaattgccag ctgcttgctg agccgctgtg gggctcaatc	2520
ggttacagtc tgccggccac ctttggtgct cagacggcag atacagagcg tcgtgtaatc	2580
ctaattcattg gcgatggttc agcgcaatta actattcagg aactttccag tatgatgcgt	2640
gacggcttga aacctatcat ctttctcctg aacaacaacg gttacaccgt tgaacgggcg	2700
attcacggcg cggagcaacg ttataacgat atcgtgctt ggaattggac ccaactgccc	2760
caggcgctga gtgttcattg cccagcgag agctggcgag tcgttgaaac ggtgcagctg	2820

accgacgtaa tgaaagtcac cgctgcttct ccgcgtctga gcttggtaga agttgttctg	2880
cctgcaatgg atgtcccacc gctgctgcaa gcagtgaagt ccgctctgaa ccagcgcaac	2940
tcctcttaag accagaggta aggaggtaac aaccatgcga gtgttgaaga aacatcttaa	3000
tcatgcaggg gagggtttct aatgcgcagc atgcagtttg atgagtacgg tgcaccctg	3060
aaagcgttct catatgaaga cccgaccccg caagggaagg aagtagtcgt taggatcgaa	3120
gcctgtggtg tgtgccactc tgatattcat cttcacgagg gctacttcga catgggcggt	3180
ggcaataaag ctgatgttac tcgtgctcgc gaactccctt ttacattggg tcatgaaatc	3240
gttggcgaag tggtagcaac tggaccaggt gtcaccggcg ctaaaccggg cgacaaacgt	3300
atttgttacc cgtggatcgg gtgcggcgac tgcccgaat gcaacagtgg tgaggatcag	3360
tcctgtgctc gtccacgtaa cctgggtgtt cacgttgacg gtggctattc gacgcacgta	3420
aagataccgg acgaaaaatt cctgttcgcc tacgatggta ttcctactga gttagcggga	3480
acctatgctt gcagcggcat caccgcttat ggtgcactga tgaaagcaaa ggaagcggct	3540
gaaagatctg gctacatcgg tctgattggc gctggtggtg ttggcatggc tggctctgatg	3600
ctggccaaag cagcgatcgg ggctaaaact gtagtctttg atatcgacga cgcaaaactg	3660
gaagctgcga cccgtgccgg ggcggtattac gtgttcaact ccggtgcaaa agaaacacgc	3720
aaggaagtta tgaaactaac gaatggtggc ctgtctggtg ctgttgattt cgttggcagc	3780
gataaaagcg ctctgtttgg aatcaacgcc ttgggtcaga acggcgtgct ggtcataatt	3840
ggactgttcg gtggcgctat gactgttccg gtaccctgt tcccgtgaa agggatcacc	3900
gtacgtggct catacgtagg ttccctgcaa gagatgagtg atatgatgga gttagtctgc	3960
gctgggaaag ttctccgat gccggtaaaa actcggccac tggacgctgc ctgggaaacc	4020
cttgaggatc tacgccatgg taaaatcgtg ggccgtgttg ttctgacccc ataagaagga	4080
gatatacata tgacccatca attaagatcg cgcgatatca tcgctctggg ctttatgaca	4140
tttgcgttgt tcgtcggcgc aggtaacatt atttccctc caatggtcgg cttgcaggca	4200
ggcgaacacg tctggactgc ggcatcggc ttctcatta ctgccgttg cctaccggta	4260
ttaacggtag tggcgctggc aaaagtggc ggcggtgttg acagtctcag cacgccaatt	4320
ggtaaagtcg ctggcgtact gctggcaaca gtttgttacc tggcgggtgg gccgctttt	4380
gctacgccgc gtacagctac cgtttctttt gaagtgggca ttgcgccgct gacgggtgat	4440

tccgcgctgc cgctgtttat ttacagcctg gtctatttcg ctatcgttat tctggtttcg	4500
ctctatccgg gcaagctgct ggataccgtg ggcaacttcc ttgcgccgct gaaaattatc	4560
gcgctgggtca tcctgtctgt tgccgcaatt atctggccgg cgggttctat cagtacggcg	4620
actgaggctt atcaaaacgc tgcgttttct aacggcttcg tcaacggcta tctgaccatg	4680
gatacgctgg gcgcaatggt gtttggtatc gttattgtta acgcggcgcg ttctcgtggc	4740
gttaccgaag cgctgtctgct gacccgttat accgtctggg ctggcctgat ggcggtgtt	4800
ggtctgactc tgctgtacct ggcgctgttc cgtctgggtt cagacagcgc gtcgctggtc	4860
gatcagtctg caaacggtgc ggcgatcctg catgcttacg ttcagcatac ctttggcggc	4920
ggcggtagct tcctgctggc ggcgttaatc ttcatcgct gcctggtcac ggcggttggc	4980
ctgacctgtg cttgtgcaga attcttcgcc cagtacgtac cgctctctta tcgtacgtg	5040
gtgtttatcc tcggcggctt ctcgatggtg gtgtctaacc tcggcttgag ccagctgatt	5100
cagatctctg taccggtgct gaccgccatt tatccgccgt gtatcgact ggttgatt	5160
agttttacac gtcattggtg gcataattcg tcccgctga ttgctccgcc gatgtttatc	5220
agcctgcttt ttggtattct cgacgggatc aaggcatctg cattcagcga tatcttaccg	5280
tcctgggcgc agcgtttacc gctggccgaa caaggtctgg cgtggttaat gccaacagt	5340
gtgatggtgg ttctggccat tatctgggat cgtgcggcag gtcgtcaggt gacctccagc	5400
gctcactaat acgcatggca tggatgaccg atggtagtgt gggctctccc catgcgagag	5460
tagggaactg ccaggcat	5478

<210> 27

<211> 5508

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 27

gttgacactc tatcattgat agagttatth taccactccc tatcagtgat agagaaaagt	60
gagcgaaaaa tcaataagga ggcaacaaga tgtgcgaaaa acatcttaat catgctaagg	120
aggttttcta atgaacatct tcaagaaaat ggaggaatth aattatgaac aactggtcta	180
cttctacgac agcgaaacgg aactcaaagg tattacctgt atacacaaca caactttagg	240

gccggcattg ggcggtaccc gccttttgaa ctataactct gaggaagatg ccgttgaaga	300
cgtaatccgt ctggctcggg gcatgactta caaagcggct tgcgccggtc tgaatctggg	360
cggcggtaaa accgtgctga tcggtgatgc taaaaagatt aaatcagagt cctacttccg	420
tggactgggg cgctacgttc agtcgctgaa cggcagatat atcaccgcgg aagacgtaaa	480
tactttctacg aaggatatgg catacgttgc tatggaaact gactatgtgg taggcctggg	540
aggtaaattcc ggcaacccta gtccagttac tgcttacggg gcatttatgg gtatcaaagc	600
ggcgctgatg aaaaaatttg aggatagctc tattgaaggc cgaaccttcg cagtgcaggg	660
tgctgggcag acgggttact atcttatcga ttacctccta ggcaacaaca agttcaaaga	720
aaaggctaaa aagatttact tcaccgaaat taacgagagc tatatcgagc gtatgaacaa	780
agaacatccg gaagttgaat ttatttcccc ggacaaaatc tactcgctgg aagtagacgt	840
cttcgtgccc tgcgccctgg gcaaaatcgt taatgacaaa actatcgatg aatttaagt	900
tccgatcatc gcaggtactg caaacaacgt actggaaagg gaagcgcacg gcaacatgct	960
taaagaacgt ggcattcttt acgccccgga ctatgtgatc aatgctgggt ggctgatcaa	1020
cgtttaccac gagctgaacg gttacaataa agagaacgct attctggaag tggaattaat	1080
ttatgatcgc ctactggaaa tattcaacat cgctgattct ctgaacatca gcaccaatat	1140
cgctgccaac gagttcgcgg aaaaacgtat caagcaaatt aagtccttga aaaacaactt	1200
cattaaacgc taagtcaata aaggcatata aaaggagggt aataacatga aagttaaagt	1260
aaaacatctt aatcatgcac aggagacttt ctaatgtcga cgacaaccgt tggtgactac	1320
ttgctgtatc gcttaaacga aatcggcatt gagcacctct tcggagtgcc aggtgattac	1380
aatctgcaat ttctggatca tgtaatcgac caccctcagc tgacttgggt cggctgcact	1440
aacgaactta acgctgccta cgcagctgat ggttatgcgc gttgtcgtcc ggctgcggca	1500
ctgctgacca ctttcggggg tggcgaactg agcgctatta atggcatcgc aggttcctac	1560
gcggagtatc tgccggtaat acatatcggt ggtgcaccga gtctatcagc ccagcagcag	1620
ggcgacctga ttcaccactc tcttggcgaa ggtgatTTTT ccagcttcct gaggatgtcc	1680
caaccgggtg ctgttgcgca ggctgctctg actcctgata acgcatgcaa ggaaatcgac	1740
cgcgtactgg cggaagtcct cattcagcgt cgtcccggt acctgctgct gtctaccgac	1800
gtggctgctg cgccggcggc tctgccacaa agcactcttt ctttgccgac cgccccggat	1860

catcgcgcag ttctggctgc tttcagcgac gctgctgagc agatgctggc tcaggccaaa	1920
agcgtctctc tactggcgga ctttctggct gatcgtttcg gtgttactcg agcactggcc	1980
gcgtggcctc agcagggttc gctaccgcac gccactctgt taatgggtaa aggcgttctg	2040
agtgaacagc aaccagggtt cgtgggtacc tacgctgggtg cggcatctat cgattcgacg	2100
cgtggcgcaa tcgaagaagc tggggtaatt atcggagtg gagttagatt ttccgacact	2160
atcacagcag gcttctcgca gcagatcgac gcccgccgtt ttatagacat tcaacccttc	2220
ttctctcgta ttggcgatcg ccagtttgat cacttgccga tgcaggctgc cgtcgcagcc	2280
ctgcatcaac tgtgtcttcg ttatcagcag cagtgggtcta tcaccgctcc tagcccgct	2340
gcactgccgc cggctgctgg tagcgagctg tcccagaacg cattctggca ggcgatgcag	2400
aacttcatcc gccctgggga cctgttggtg gccgaccaag gtactgcggc gttcggcgca	2460
gcggcgctgc gcttaccgca gaattgccag ctgcttggtc agccgctgtg gggctcaatc	2520
ggttacagtc tgccggccac ctttggtgct cagacggcag atacagagcg tcgtgtaatc	2580
ctaattcattg gcgatggttc agcgcaatta actattcagg aactttccag tatgatgcgt	2640
gacggcttga aacctatcat ctttctcctg aacaacaacg gttacaccgt tgaacgggcg	2700
attcacggcg cggagcaacg ttataacgat atcgtgctt ggaattggac ccaactgccc	2760
caggcgctga gtgttcattg cccagcgcag agctggcgag tcgttgaaac ggtgcagctg	2820
accgacgtaa tgaaagtcac cgctgcttct ccgcgtctga gcttggtaga agttgttctg	2880
cctgcaatgg atgtcccacc gctgctgcaa gcagtgagtg ccgctctgaa ccagcgcaac	2940
tcctcttaag accagaggta aggaggtaac aaccatgcga gtgttgaaga aacatcttaa	3000
tcatgcaggg gagggtttct aatgacagcg gagcagcaaa atgggggtatc cgactcacgc	3060
cgtttcgaat ttcaggaatt tgggtggcct atcgccccac agacctatca gctccccgca	3120
ccggctagcg atgaagtttt gttaaagggtg aactactgcg gtgtctgtca cagtgatgtt	3180
catcttcacg acggctactt cgagctgggt ggcgataaac gtctgaactt cgctatgccg	3240
ctgccgctga cgctgggtca cgaagtaatt ggcaccgttg tggctgtcgg cgaccagggtt	3300
actggtgtaa aaccggggga ccagcgactg atctatccgt ggatagggtg cggaaaatgc	3360
ggcgcgtgtc aaaaaggaga agaaaacctg tgcgttactc ctgcacatct gggcgtgaac	3420
aagccgggcg gttacgtga tcacatcggt gtacccatt ctcgctacct tctggacatt	3480

tcgggtctga	acccgggtga	tgccgctacc	ctcgcgtgct	ccggcctgac	cactttcagc	3540
gcgatcaaca	aagtgttgcc	gcttgcagat	gaccagtgga	ttgttggttat	cggttggtggt	3600
ggcctcggcc	agatggcgct	gcgtatcctg	caagctatgg	gaattggcaa	tgttatcgggt	3660
attgacctgt	ctgaagagaa	acggaaactg	gctcatgaaa	gcgggtgcacg	tcactccttc	3720
gatccaaaca	ctccgaagct	gaaccgcgtg	gtcgccgaaa	cctgcccggg	tacggtacag	3780
gccgcgttag	actttgtggg	caatgagcaa	actgctcagc	tggcactgtc	tctgcttgga	3840
aaaggtggca	aatatgttcc	tgtcgggctg	cacggcggcg	agctgcgtta	cccattgccg	3900
atcatcacga	acaaagctgt	aagtatcatc	ggttcttacg	ttggtaccct	gaaagaactg	3960
gaagacttag	ttgctttcgc	caaggaaaag	aatctgccgc	caattcatat	tgaacaccgc	4020
ccgctggaat	cggcggctca	ggccgtagag	gacctggaaa	aaggacaggt	tgctgggcgt	4080
gttatcctgg	atgcaggtaa	ctaagaagga	gatatacata	tgacccatca	attaagatcg	4140
cgcgatatca	tcgctctggg	ctttatgaca	tttgcgttgt	tcgtcggcgc	aggtaacatt	4200
attttccctc	caatggctgg	cttgcaggca	ggcgaacacg	tctggactgc	ggcattcggc	4260
ttcctcatta	ctgccgttgg	cctaccggtg	ttaacggtag	tggcgctggc	aaaagttggc	4320
ggcgggtgtg	acagtctcag	cacgccaatt	ggtaaagtcg	ctggcgctact	gctggcaaca	4380
gtttgttacc	tggcgggtggg	gccgcttttt	gctacgccgc	gtacagctac	cgtttctttt	4440
gaagtgggca	ttgcgccgct	gacgggtgat	tccgcgctgc	cgctgtttat	ttacagcctg	4500
gtctatttcg	ctatcgttat	tctggtttcg	ctctatccgg	gcaagctgct	ggataccgtg	4560
ggcaacttcc	ttgcgccgct	gaaaattatc	gcgctgggtca	tcctgtctgt	tgccgcaatt	4620
atctggccgg	cgggttctat	cagtacggcg	actgaggctt	atcaaaacgc	tgcgttttct	4680
aacggcttcg	tcaacggcta	tctgaccatg	gatacgctgg	gcgcaatggt	gtttggtatc	4740
gttatgttta	acgcggcgcg	ttctcgtggc	gttaccgaag	cgcgtctgct	gacccgttat	4800
accgtctggg	ctggcctgat	ggcgggtgtt	ggctctgactc	tgctgtacct	ggcgctgttc	4860
cgtctggggt	cagacagcgc	gtcgctggtc	gatcagtctg	caaacgggtgc	ggcgatcctg	4920
catgcttacg	ttcagcatac	ctttggcggc	ggcggtagct	tcctgctggc	ggcgtaaatc	4980
ttcatcgcct	gcctggtcac	ggcggttggc	ctgacctgtg	cttgtgcaga	attcttcgcc	5040
cagtacgtac	cgctctctta	tcgtacgctg	gtgttttatcc	tcggcggctt	ctcgtatggtg	5100

gtgtctaacc tcggcttgag ccagctgatt cagatctctg taccggtgct gaccgccatt	5160
tatccgccgt gtatcgact gggtgtatta agttttacac gtcattgggtg gcataattcg	5220
tcccgcgtga ttgctccgcc gatgtttatc agcctgcttt ttggtattct cgacgggatc	5280
aaggcatctg cattcagcga tatcttaccg tcctgggcgc agcgtttacc gctggccgaa	5340
caaggctctgg cgtgggtaat gccaacagtg gtgatgggtg ttctggccat tatctgggat	5400
cgtgcggcag gtcgtcaggt gacctccagc gtcactaat acgcatggca tggatgaccg	5460
atggtagtgt gggctctccc catgcgagag tagggaactg ccaggcat	5508

<210> 28

<211> 5430

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 28

gttgacactc tatcattgat agagttatth taccactccc tatcagtgat agagaaaagt	60
gagcgaaaaa tcaataagga ggcaacaaga tgtgcgaaaa acatcttaat catgctaagg	120
aggttttcta atggtagaga tccaggcttt gccggaaact tccatttttg ggcaaatcgc	180
agaccaccag catgaacagg tggctcttctg ccacgatcac gaaaccggcc tccgtgcgat	240
aatcggtatt cataacacgg ttcttgcccc cgccttaggt ggaactcgca tgtggcacta	300
tgctaccgag gcagaagcgc tgaatgacgt tctgcgtctg tctcgcggta tgacctaaa	360
ggctgctatc tcgggcctga acctgggttg cggtaaagca gtaatcattg gggatgccaa	420
aacaatcaaa accgaagcgc tgctgcggaa attcggcaga ttcgtgcaga acctgaatgg	480
taaatacatc actgctgaag acgttaacat gactacaaag gatattggagt acattaggat	540
ggaaacaaa cacgtcgctg gcttacctga aagtatgggt ggaagcgggtg acccgtcacc	600
ggtaactgca tatggtacgt acatgggcat gaaagcggcg gccaaaaagg cgtttggtc	660
tgattccctg gctggcaaac gtatcgctgt tcaagggtgtg ggtcatgttg gcacttatct	720
gcttgagcat ttgaccaaag aaggtgctca gattgtgctg actgactact ataaggaacg	780
tgccgaggaa gcaggcgcgc gttttggcgc acaggttggtt ggcctggacg atatctacga	840

tcaagaggtc gacatttact ctccatgtgc tctcgggtgct accatcaacg atgacactat	900
cgatcgcctg cgttgcgctg ttgtagccgg ttgcgcaaac aaccagctga aagaagaaaa	960
cgtccacggt ccggcgctgg ttgagcgcgg gatagtatac gccccagact tcctgatcaa	1020
tgcagggggc ctgattaacg tgtatagcga agttacaggg tctacccgctc agggggcctt	1080
aactcagacc gaaaaaatct atgactacac actccaagtt ctggaaaaag ccgcggtga	1140
aggtctgcac ccgcagcagg ctgcgatccg tcaggcggaa caacgcatcg ctgcaattgg	1200
taaggatgaaa agcacctact aagtcaataa aggcatataa aaggagggtta ataacatgaa	1260
agttaaagta aaacatctta atcatgcaca ggagactttc taatgtcgac gacaaccgtt	1320
ggtgactact tgctgtatcg cttaaacgaa atcggcattg agcacctctt cggagtgcga	1380
ggtgattaca atctgcaatt tctggatcat gtaatcgacc accctcagct gacttgggtc	1440
ggctgcacta acgaacttaa cgctgcctac gcagctgatg gttatgcgcg ttgtcgtccg	1500
gctgcggcac tgctgaccac cttcgggggtt ggcgaactga gcgctattaa tggcatcgca	1560
ggttcctacg cggagtatct gccggtaata catatcgttg gtgcaccgag tctatcagcc	1620
cagcagcagg gcgacctgat tcaccactct cttggcgaag gtgatttttc cagcttcctg	1680
aggatgtccc aaccggtgtc tgttgcgag gctgctctga ctctgataa cgcatgcaag	1740
gaaatcgacc gcgtactggc ggaagtcctc attcagcgtc gtcccggcta cctgctgctg	1800
tctaccgacg tggctgctgc gccggcggct ctgccacaaa gcactctttc tttgccgacc	1860
gccccggatc atcgcgcagt tctggctgct ttcagcgacg ctgctgagca gatgctggct	1920
caggccaaaa gcgtctctct actggcggac tttctggctg atcgtttcgg tgttactcga	1980
gcactggccg cgtggcttca gcaggttccg ctaccgcacg ccactctgtt aatgggtaaa	2040
ggcgttctga gtgaacagca accagggttc gtgggtacct acgctgggtgc ggcatctatc	2100
gattcgacgc gtggcgcaat cgaagaagct ggggtaatta tcggagtggg agttagattt	2160
tccgacacta tcacagcagg cttctcgcag cagatcgacg cccgccgttt tatagacatt	2220
caacccttct tctctcgtat tggcgatcgc cagtttgatc acttgccgat gcaggctgcc	2280
gtcgcagccc tgcatcaact gtgtcttcgt tatcagcagc agtgggtctat caccgctcct	2340
agcccgctg cactgccgcc ggctgctggg agcgagctgt cccagaacgc attctggcag	2400
gcgatgcaga acttcatccg ccctggggac ctgttggtgg ccgaccaagg tactgcggcg	2460

ttcggcgag	cggcgctgcg	cttaccgcag	aattgccagc	tgcttgca	gccgctgtgg	2520
ggctcaatcg	gttacagtct	gccggccacc	tttgggtgctc	agacggcaga	tacagagcgt	2580
cgtgtaatcc	taatcattgg	cgatgggttca	gcgcaattaa	ctattcagga	actttccagt	2640
atgatgcgtg	acggcttgaa	acctatcatc	tttctcctga	acaacaacgg	ttacaccgtt	2700
gaacgggcga	ttcacggcgc	ggagcaacgt	tataacgata	tcgctgcttg	gaattggacc	2760
caactgcccc	aggcgctgag	tgttcattgc	ccagcgcaga	gctggcgagt	cgttgaaacg	2820
gtgcagctga	ccgacgtaat	gaaagtcatc	gctgcttctc	cgctctgag	cttggtagaa	2880
gttgcttctgc	ctgcaatgga	tgtcccaccg	ctgctgcaag	cagtgagtgc	cgctctgaac	2940
cagcgcaact	cctcttaaga	ccagaggtaa	ggaggtaaca	accatgcgag	tggtgaagaa	3000
acatcttaat	catgcagggg	agggtttcta	atgcagacgg	cgctcttgta	taagccaggt	3060
cacgaaaact	tagtgcgctc	ggagatcccg	atacctaaag	ctgggcgtgg	cgaagtcgtt	3120
ctggaaatta	aagccgctgg	catgtgccat	tccgatctgc	acgttctcga	cgggtggaatc	3180
cccctgccgg	gtcaatttgt	aatgggccat	gaaatcgttg	gtactattca	cgagatcggc	3240
caggacgtga	ccggtttcaa	acagggcgat	ctgtacgcag	tccacggccc	gaatccgtgt	3300
ggatatttgca	ccctgtgcag	agaaggattt	gataacgact	gcactacagt	ggcgaaaacc	3360
ggatcaatgg	tcggactggg	tcttgacggc	ggctaccaga	agtatatccg	tatcccgaac	3420
gtaaggctcta	tcgttaaagt	tccagaaggt	gtttcagctg	aggcagctgc	gagctgtact	3480
gatgcagtac	tgaccccgta	ccgtgcacta	aaacaggctg	gcgccagcaa	ctctactcgg	3540
gtactgattc	tgggtctggg	tggcttaggt	ctgaatgccc	ttaaactggc	taagaccttc	3600
ggcagttacg	tttacgcac	tgacctgaaa	ccttctgcgc	gtgaagctgc	taaggccgct	3660
ggggcggatg	aagtgcctga	gtccctgccc	gaagaccgcg	tgggtgttga	tatcgtgtta	3720
gacgtcgttg	gcgtgcagag	caccttcaac	ctcgtcaaaa	aacacgttgg	cccgcgtggc	3780
atcattgtac	ctgtaggcct	ggcatcccca	cagctttcgt	ttaacctaac	ggatctggcg	3840
ctccgcgaaa	ttcgtgttca	gggcactttt	tggggcacga	gcaatgagct	ggctgaatgt	3900
ctgcgcctgt	gccagctggg	cctgatcaac	ccgaaatata	ctgtgggtgcc	tcttgaagaa	3960
gcgccgaaat	atatggaagc	aatggctcat	gggaaagtag	aaggctgtat	cgttttccac	4020
ccgtaagaag	gagatataca	tatgacctat	caattaagat	cgcgcgatat	catcgctctg	4080

ggctttatga catttgcgtt gttcgtcggc gcaggtaaca ttattttccc tccaatggtc	4140
ggcttgcagg caggcgaaca cgtctggact gcggcattcg gcttcctcat tactgccgtt	4200
ggcctaccgg tattaacggt agtggcgctg gcaaaagttg gcggcgggtg tgacagtctc	4260
agcacgcaa ttggtaaagt cgctggcgta ctgctggcaa cagtttgta cctggcggtg	4320
gggccgcttt ttgctacgcc gcgtacagct accgtttctt ttgaagtggg cattgcgccg	4380
ctgacgggtg attccgcgct gccgctgttt atttacagcc tggctctatt cgctatcggt	4440
attctgggtt cgctctatcc gggcaagctg ctggataccg tgggcaactt ccttgcgccg	4500
ctgaaaatta tcgcgctggt catcctgtct gttgccgcaa ttatctggcc ggcggttct	4560
atcagtacgg cgactgaggc ttatcaaac gctgcgtttt ctaacggctt cgtcaacggc	4620
tatctgacca tggatacgct gggcgcaatg gtgtttggta tcgttattgt taacgcggcg	4680
cgttctcgtg gcgttaccga agcgcgtctg ctgaccggtt ataccgctctg ggctggcctg	4740
atggcgggtg ttggtctgac tctgctgtac ctggcgctgt tccgtctggg ttcagacagc	4800
gcgtcgctgg tcgatcagtc tgcaaacggt gcggcgatcc tgcatgctta cgttcagcat	4860
acctttggcg gcggcggtag cttcctgctg gcggcggtta tcttcacgc ctgcctggtc	4920
acggcggttg gcctgacctg tgcttgtgca gaattcttcg cccagtacgt accgctctct	4980
tatcgtagcg tgggtgttat cctcggcggc ttctcgatgg tgggtgtctaa cctcggcttg	5040
agccagctga ttcagatctc tgtaccggtg ctgaccgcca tttatccgcc gtgtatcgca	5100
ctggttgat taagttttac acgctcatgg tggcataatt cgtcccgcgt gattgctccg	5160
ccgatgttta tcagcctgct ttttggtatt ctcgacggga tcaaggcatc tgcatcagc	5220
gatatcttac cgtcctgggc gcagcgttta ccgctggccg aacaaggctt ggctgggtta	5280
atgccaacag tggatgatgt ggttctggcc attatctggg atcgtgcggc aggtcgtcag	5340
gtgacctcca gcgctcacta atacgcatgg catggatgac cgatggtagt gtgggctctc	5400
cccatgcgag agtagggaac tgccaggcat	5430

<210> 29

<211> 5487

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 29

gttgacactc tatcattgat agagttat	ttt taccactccc tatcagtgat agagaaaagt	60
gagcgaaaaa tcaataagga ggcaacaaga	tgtgcgaaaa acatcttaat catgctaagg	120
aggttttcta atggttagaga tccaggcttt	gccggaaact tccatttttg ggcaaatcgc	180
agaccaccag catgaacagg tggctcttctg	ccacgatcac gaaaccggcc tccgtgcat	240
aatcggtatt cataacacgg ttcttggccc	cgccttaggt ggaactcgca tgtggcacta	300
tgctaccgag gcagaagcgc tgaatgacgt	tctgcgtctg tctcgcggta tgacctaaa	360
ggctgctatc tcgggcctga acctgggtgg	cggtaaagca gtaatcattg gggatgccaa	420
aacaatcaaa accgaagcgc tgctgcggaa	attcggcaga ttcgtgcaga acctgaatgg	480
taaatacatc actgctgaag acgttaacat	gactacaaag gatattggagt acattaggat	540
ggaaaccaa cagctcgctg gcttacctga	aagtattgggt ggaagcgggtg acccgtcacc	600
ggtaactgca tatggtacgt acatgggcat	gaaagcggcg gccaaaaagg cgtttggctc	660
tgattccctg gctggcaaac gtatcgctgt	tcaagggtgtg ggtcatgttg gcacttatct	720
gcttgagcat ttgaccaaag aagggtgctca	gattgtgctg actgactact ataaggaacg	780
tgccgaggaa gcaggcgcgc gttttggcgc	acaggttgtt ggcttgacg atatctacga	840
tcaagaggtc gacatttact ctccatgtgc	tctcgggtgct accatcaacg atgacactat	900
cgatcgctg cgttgcgctg ttgtagccgg	ttgcgcaaac aaccagctga aagaagaaaa	960
cgtccacggg cgggcgctgg ttgagcgcgg	gatagtatac gcccagact tcctgatcaa	1020
tgcagggggc ctgattaacg tgtatagcga	agttacaggg tctaccgctc agggggcttt	1080
aactcagacc gaaaaaatct atgactacac	actccaagtt ctggaaaaag ccgcggctga	1140
aggtctgcac ccgcagcagg ctgcgatccg	tcaggcggaa caacgcatcg ctgcaattgg	1200
taagggtgaaa agcacctact aagtcaataa	aggcatataa aaggagggtta ataacatgaa	1260
agttaaagta aaacatctta atcatgcaca	ggagactttc taatgtcgac gacaaccgtt	1320
ggtgactact tgctgtatcg cttaaacgaa	atcggcattg agcacctctt cggagtgcc	1380
ggtgattaca atctgcaatt tctggatcat	gtaatcgacc accctcagct gacttgggtc	1440
ggctgcacta acgaacttaa cgctgcctac	gcagctgatg gttatgcgcg ttgtcgtccg	1500
gctgcggcac tgctgaccac cttcgggggt	ggcgaactga gcgctattaa tggcatcgca	1560

ggttcctacg cggagtatct gccggtaata catatcgttg gtgcaccgag tctatcagcc	1620
cagcagcagg gcgacctgat tcaccactct cttggcgaag gtgatttttc cagcttcctg	1680
aggatgtccc aaccggtgtc tgttgcgag gctgctctga ctcttgataa cgcattgcaag	1740
gaaatcgacc gcgtactggc ggaagtcctc attcagcgtc gtcccggtc cctgctgctg	1800
tctaccgacg tggctgctgc gccggcggct ctgccacaaa gcactctttc tttgccgacc	1860
gccccggatc atcgcgagc tctggctgct ttcagcgacg ctgctgagca gatgctggct	1920
caggccaaaa gcgtctctct actggcggac tttctggctg atcgtttcgg tgttactcga	1980
gcactggccg cgtggcttca gcaggttccg ctaccgcacg ccactctgtt aatgggtaaa	2040
ggcgttctga gtgaacagca accagggttc gtgggtacct acgctgggtg gccatctatc	2100
gattcgacgc gtggcgcaat cgaagaagct ggggtaatta tcggagtggg agttagattt	2160
tccgacacta tcacagcagg cttctcgag cagatcgacg cccgccgtt tatagacatt	2220
caacccttct tctctctgtat tggcgatcgc cagtttgatc acttgccgat gcaggctgcc	2280
gtcgagccc tgcattcaact gtgtcttctg tatcagcagc agtgggtctat caccgctcct	2340
agcccgctg cactgccgc ggctgctggg agcgagctgt cccagaacgc attctggcag	2400
gcgatgcaga acttcatccg ccctggggac ctgttggtgg ccgaccaagg tactgcggcg	2460
ttcggcgag cggcgctgcg cttaccgcag aattgccagc tgcttggtga gccgctgtgg	2520
ggctcaatcg gttacagtct gccggccacc tttgggtgctc agacggcaga tacagagcgt	2580
cgtgtaatcc taatcattgg cgatggttca gcgcaattaa ctattcagga actttccagt	2640
atgatgcgtg acggcttgaa acctatcatc tttctctga acaacaacgg ttacaccgtt	2700
gaacgggcca ttcacggcgc ggagcaacgt tataacgata tcgctgcttg gaattggacc	2760
caactgcccc aggcgctgag tgttcattgc ccagcgagc gctggcgagt cgttgaaacg	2820
gtgcagctga ccgacgtaat gaaagtcacg gctgcttctc cgcgtctgag cttggtagaa	2880
gttggtctgc ctgcaatgga tgtcccaccg ctgctgcaag cagtgagtgc cgctctgaac	2940
cagcgcaact cctcttaaga ccagaggtaa ggaggtaaca accatgcgag tgttgaagaa	3000
acatcttaat catgcagggg agggtttcta atgcgcagca tgcagtttga tgagtacggt	3060
gcaccctga aagcgttctc atatgaagac ccgacccgc aagggaagga agtagtcgtt	3120
aggatcgaag cctgtggtgt gtgccactct gatattcatc ttcacgaggg ctacttcgac	3180

atgggcggtg gcaataaagc tgatgttact cgtgctcgcg aactcccttt tacattgggt	3240
catgaaatcg ttggcgaagt ggtagcaact ggaccagggtg tcaccggcgc taaaccgggc	3300
gacaaacgta ttgtgtaccc gtggatcggg tgcggcgact gcccgaaatg caacagtgggt	3360
gaggatcagt cctgtgcgcg tccacgtaac ctgggtgttc acgttgacgg tggctattcg	3420
acgcacgtaa agataccgga cgaaaaattc ctgttcgcct acgatgggat tcctactgag	3480
ttagcgggaa cctatgcttg cagcggcatc accgcttatg gtgcactgat gaaagcaaag	3540
gaagcggctg aaagatctgg ctacatcgggt ctgattggcg ctggtggcgt tggcatggct	3600
ggtctgatgc tggccaaagc agcgatcggg gctaaaactg tagtctttga tatcgacgac	3660
gcaaaactgg aagctgcgac ccgtgccggg gcggattacg tgttcaactc cggtgcaaaa	3720
gaaacacgca aggaagttat gaaactaacg aatggtggcc tgtctggtgc tgttgatttc	3780
gttggcagcg ataaaagcgc tctgtttgga atcaacgcct tgggtcagaa cggcgtgctg	3840
gtcataattg gactgttcgg tggcgctatg actgttccgg taccctgtt cccgctgaaa	3900
gggatcaccg tacgtggctc atacgtaggt tccctgcaag agatgagtga tatgatggag	3960
ttagttcgcg ctgggaaagt tcctccgatg ccggtaaaaa ctcggccact ggacgctgcc	4020
tgggaaacct ttgaggatct acgccatggt aaaatcgtgg gccgtgttgt tctgaccca	4080
taagaaggag atatacatat gacccatcaa ttaagatcgc gcgatatcat cgctctgggc	4140
tttatgacat ttgcgttggt cgtcggcgca ggtaacatta tttccctcc aatggtcggc	4200
ttgcaggcag gcgaacacgt ctggactgcg gcattcggct tcctcattac tgccgttggc	4260
ctaccggtat taacggtagt ggcgctggca aaagttggcg gcggtgttga cagtctcagc	4320
acgccaattg gtaaagtcgc tggcgctactg ctggcaacag tttgttacct ggcggtgggg	4380
ccgctttttg ctacgccgcg tacagctacc gtttcttttg aagtgggcat tgcgccgctg	4440
acgggtgatt ccgcgctgcc gctgtttatt tacagcctgg tctatttcgc tatcgttatt	4500
ctggtttcgc tctatccggg caagctgctg gataccgtgg gcaacttcct tgcgccgctg	4560
aaaattatcg cgctgggtcat cctgtctgtt gccgcaatta tctggccggc gggttctatc	4620
agtacggcga ctgaggctta tcaaaacgct gcgttttcta acggcttcgt caacggctat	4680
ctgaccatgg atacgtggg cgcaatggtg tttggtatcg ttattgttaa cgcggcgcgt	4740
tctcgtggcg ttaccgaagc gcgtctgctg acccgttata ccgtctgggc tggcctgatg	4800

gcgggtgttg gtctgactct gctgtacctg gcgctgttcc gtctgggttc agacagcgcg	4860
tcgctggtcg atcagtctgc aaacggtgcg gcgacacctg atgcttacgt tcagcatacc	4920
tttggcggcg gcggtagctt cctgctggcg gcgttaatct tcatcgacctg cctggtcacg	4980
gcggttggcc tgacctgtgc ttgtgcagaa ttcttcgccc agtacgtacc gctctcttat	5040
cgtacgctgg tgtttatcct cggcggcttc tcgatgggtg tgtctaacct cggcttgagc	5100
cagctgattc agatctctgt accggtgctg accgccattt atccgccgtg tatcgcaactg	5160
gttgatttaa gttttacacg ctcatgggtg cataattcgt cccgcgtgat tgctccgccg	5220
atgtttatca gcctgctttt tgggtattctc gacgggatca aggcattctgc attcagcgat	5280
atcttaccgt cctgggcgca gcgtttaccg ctggccgaac aaggctctggc gtgggttaatg	5340
ccaacagtgg tgatgggtgg tctggccatt atctgggatc gtgcggcagg tcgtcaggtg	5400
acctccagcg ctactaata cgcatggcat ggatgaccga tggtagtggt ggctctcccc	5460
atgcgagagt agggaactgc caggcat	5487

<210> 30

<211> 5517

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 30

gttgacactc tatcattgat agagttatct taccactccc tatcagtgat agagaaaagt	60
gagcgaaaaa tcaataagga ggcaacaaga tgtgcgaaaa acatcttaat catgctaagg	120
aggttttcta atggtagaga tccaggcttt gccggaaact tccatttttg ggcaaatcgc	180
agaccaccag catgaacagg tggctcttctg ccacgatcac gaaaccggcc tccgtgcgat	240
aatcggtatt cataacacgg ttcttggtccc cgccttaggt ggaactcgca tgtggcacta	300
tgctaccgag gcagaagcgc tgaatgacgt tctgcgtctg tctcgcggtg tgacctacaa	360
ggctgctatc tcgggcctga acctgggtgg cggtaaagca gtaatcattg gggatgccaa	420
aacaatcaaa accgaagcgc tgctgcggaa attcggcaga ttcgtgcaga acctgaatgg	480
taaatacatc actgctgaag acgttaacat gactacaaag gatatggagt acattaggat	540

ggaaacaaa	cacgtcgctg	gcttacctga	aagtatgggt	ggaagcgggtg	acccgtcacc	600
ggtaactgca	tatggtacgt	acatgggcat	gaaagcggcg	gccaaaaagg	cgtttggtc	660
tgattccctg	gctggcaaac	gtatcgctgt	tcaagggtgtg	ggtcattgtt	gcacttatct	720
gcttgagcat	ttgaccaaag	aagggtgctca	gattgtgctg	actgactact	ataaggaacg	780
tgccgaggaa	gcaggcgcg	gttttggcg	acaggttgtt	ggcctggacg	atatctacga	840
tcaagaggtc	gacatttact	ctccatgtgc	tctcgggtgct	accatcaacg	atgacactat	900
cgatcgctg	cgttgcgctg	ttgtagccgg	ttgcgcaaac	aaccagctga	aagaagaaaa	960
cgccacgggt	ccggcgctgg	ttgagcgcg	gatagtatac	gccccagact	tcctgatcaa	1020
tgcagggggc	ctgattaacg	tgtatagcga	agttacagg	tctacccgtc	agggggcctt	1080
aactcagacc	gaaaaaatct	atgactacac	actccaagtt	ctggaaaaag	ccgcggctga	1140
aggtctgcac	ccgcagcagg	ctgcgatccg	tcaggcggaa	caacgcatcg	ctgcaattgg	1200
taagggtgaa	agcacctact	aagtcaataa	aggcatataa	aaggagggtta	ataacatgaa	1260
agttaaagta	aaacatctta	atcatgcaca	ggagactttc	taatgtcgac	gacaaccgtt	1320
ggtgactact	tgctgtatcg	cttaaacgaa	atcggcattg	agcacctctt	cggagtgcc	1380
ggtgattaca	atctgcaatt	tctggatcat	gtaatcgacc	accctcagct	gacttgggtc	1440
ggctgcacta	acgaacttaa	cgctgcctac	gcagctgatg	gttatgcgcg	ttgtcgtccg	1500
gctgcggcac	tgctgaccac	cttcgggggt	ggcgaactga	gcgctattaa	tggcatcgca	1560
ggttcctacg	cggagtatct	gccggtaata	catatcgttg	gtgcaccgag	tctatcagcc	1620
cagcagcagg	gcgacctgat	tcaccactct	cttggcgaag	gtgatttttc	cagcttcctg	1680
aggatgtccc	aaccgggtgtc	tgttgcgcag	gctgctctga	ctcctgataa	cgcatgcaag	1740
gaaatcgacc	gcgtactggc	ggaagtcctc	attcagcgtc	gtcccggcta	cctgctgctg	1800
tctaccgacg	tggctgctgc	gccggcggct	ctgccacaaa	gcactctttc	tttgccgacc	1860
gccccggatc	atcgcgcagt	tctggctgct	ttcagcgacg	ctgctgagca	gatgctggct	1920
caggccaaaa	gcgtctctct	actggcggac	tttctggctg	atcgtttcgg	tgttactcga	1980
gcactggccg	cgtaggcttca	gcaggttccg	ctaccgcacg	ccactctgtt	aatgggtaaa	2040
ggcgttctga	gtgaacagca	accagggttc	gtgggtacct	acgctgggtg	ggcatctatc	2100
gattcgacgc	gtggcgcaat	cgaagaagct	ggggtaatta	tcggagtggg	agttagattt	2160

tccgacacta tcacagcagg cttctcgcag cagatcgacg cccgccgttt tatagacatt	2220
caacccttct tctctcgtat tggcgatcgc cagtttgatc acttgccgat gcaggctgcc	2280
gtcgcagccc tgcatacaact gtgtcttcgt tatcagcagc agtgggtctat caccgctcct	2340
agcccgccctg cactgccgcc ggctgctggt agcgagctgt cccagaacgc attctggcag	2400
gcgatgcaga acttcatccg ccctggggac ctgttggtgg ccgaccaagg tactgcggcg	2460
ttcggcgcag cggcgctgcg cttaccgcag aattgccagc tgcttggtgca gccgctgtgg	2520
ggctcaatcg gttacagtct gccggccacc tttggtgctc agacggcaga tacagagcgt	2580
cgtgtaatcc taatcattgg cgatggttca gcgcaattaa ctattcagga actttccagt	2640
atgatgcgtg acggcttgaa acctatcatc tttctcctga acaacaacgg ttacaccgtt	2700
gaacggggcga ttcacggcgc ggagcaacgt tataacgata tcgctgcttg gaattggacc	2760
caactgcccc aggcgctgag tgttcattgc ccagcgcaga gctggcgagt cgttgaaacg	2820
gtgcagctga ccgacgtaat gaaagtcac gctgcttctc cgcgtctgag cttggtagaa	2880
gttgttctgc ctgcaatgga tgtcccaccg ctgctgcaag cagtgagtgc cgctctgaac	2940
cagcgcgaact cctcttaaga ccagaggtaa ggaggtaaca accatgcgag tgttgaagaa	3000
acatcttaat catgcagggg agggtttcta atgacagcgg agcagcaaaa tggggatatcc	3060
gactcacgcc gtttcgaatt tcaggaattt ggtggcccta tcgccccaca gacctatcag	3120
ctccccgcac cggctagcga tgaagttttg ttaaagggtga actactgcgg tgtctgtcac	3180
agtgatgttc atcttcacga cggctacttc gagctgggtg gcgataaacg tctgaacttc	3240
gctatgccgc tgccgctgac gctgggtcac gaagtaattg gcaccgttgt ggctgtcggc	3300
gaccaggtta ctggtgtaaa accgggggac cagcgactga tctatccgtg gataggttgc	3360
ggaaaatgcg gcgcgtgtca aaaaggagaa gaaaacctgt gcgttactcc tgcacatctg	3420
ggcgtgaaca agccgggcgg ttacgtgat cacatcgttg taccatttc tcgctacctt	3480
ctggacattt cgggtctgaa cccgggtgat gccgctaccc tcgcgtgctc cggcctgacc	3540
actttcagcg cgatcaacaa agtgttgccg cttgcagatg accagtggat tgttgttatc	3600
ggttggtgtg gcctcggcca gatggcgctg cgtatcctgc aagctatggg aattggcaat	3660
gttatcggtg ttgacctgtc tgaagagaaa cggaaactgg ctcatgaaag cgggtgcacgt	3720
cactccttcg atccaaacac tccgaagctg aaccgcgtgg tcgccgaaac ctgcccgggt	3780

acggtacagg ccgcgttaga ctttgtgggc aatgagcaaa ctgctcagct ggcaactgtct	3840
ctgcttgga aaggtggcaa atatgttcct gtcgggctgc acggcggcga gctgcgttac	3900
ccattgccga tcatcacgaa caaagctgta agtatcatcg gttcttacgt tggtagcctg	3960
aaagaactgg aagacttagt tgctttcgcc aaggaaaaga atctgccgcc aattcatatt	4020
gaacaccgcc cgctggaatc ggcggctcag gccgtagagg acctggaaaa aggacaggtt	4080
gctgggcgtg ttatcctgga tgcaggtaac taagaaggag atatacatat gacccatcaa	4140
ttaagatcgc gcgatcatcg cgctctgggc tttatgacat ttgcgttggt cgtcggcgca	4200
ggtaacatta ttttcctcc aatggctggc ttgcaggcag gcgaacacgt ctggactgcg	4260
gcattcggct tcctcattac tgccgttggc ctaccggtat taacggtagt ggcgctggca	4320
aaagtggcg gcggtgtga cagtctcagc acgccaattg gtaaagtcgc tggcgtagctg	4380
ctggcaacag tttgttacct ggcgggtggg ccgctttttg ctacgccgcg tacagctacc	4440
gtttcttttg aagtgggcat tgcgccgctg acgggtgatt ccgcgctgcc gctgtttatt	4500
tacagcctgg tctatctgc tatcgttatt ctggtttcgc tctatccggg caagctgctg	4560
gataccgtgg gcaacttcct tgcgccgctg aaaattatcg cgctgggtcat cctgtctgtt	4620
gccgcaatta tctggccggc gggttctatc agtacggcga ctgaggctta tcaaacgct	4680
gcgttttcta acggcttcgt caacggctat ctgacatgg atacgctggg cgcaatgggtg	4740
tttggtagctg ttattgttaa cgcggcgctg tctcgtggcg ttaccgaagc gcgtctgctg	4800
acccgttata ccgtctgggc tggcctgatg gcgggtgttg gtctgactct gctgtacctg	4860
gcgctgttcc gtctgggttc agacagcgcg tcgctggctg atcagtctgc aaacgggtgcg	4920
gcgatcctgc atgcttacgt tcagcatacc tttggcggcg gcggtagctt cctgctggcg	4980
gcgttaatct tcatgcctg cctggtcacg gcggttggcc tgacctgtgc ttgtgcagaa	5040
ttcttcgccc agtacgtacc gctctcttat cgtacgctgg tgtttatcct cggcggcttc	5100
tcgatgggtg tgtctaacct cggcttgagc cagctgattc agatctctgt accgggtgctg	5160
accgccattt atccgccgtg tatcgactg gttgtattaa gttttacacg ctcatgggtg	5220
cataattcgt cccgcgtgat tgctccgcc atgtttatca gcctgctttt tggtagtctc	5280
gacgggatca aggcatctgc attcagcgat atcttaccgt cctgggcgca gcgtttaccg	5340
ctggccgaac aaggtctggc gtggttaatg ccaacagtgg tgatgggtgt tctggccatt	5400

atctgggatc gtgcggcagg tcgtcaggtg acctccagcg ctactaata cgcatggcat	5460
ggatgaccga tggtagtgtg ggctctcccc atgcgagagt agggaaactgc caggcat	5517

<210> 31
 <211> 5496
 <212> DNA
 <213> Artificial Sequence

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

<400> 31	
gttgacactc tatcattgat agagttatth taccactccc tatcagtgat agagaaaagt	60
gagcgaaaaa tcaataagga ggcaacaaga tgtgcgaaaa acatcttaat catgctaagg	120
aggthtttcta atgatcttcg agacaatttc tacgtcgaat cacgaagaag ttgtgtattg	180
ccataacaag gacgccggct tgaaagcaat catcgcgatt cacaacactg tactcggctc	240
ggctctgggt ggcactcgca tgtggcccta cgctagcgaa gaggaagcac tgaaagatgt	300
ccttcgttta tcccgtggga tgacctaca agctgcggtt tcaggtctaa acctgggcgg	360
cggtaaagct gtgatctggg gtgatccgaa taaagacaag tctgaagcgc tgtttagagc	420
cttcggacgg tttgtaaaca gcctgggcgg acgctacatt accgcggagg acgttgcat	480
tgatgttaac gacatggaat atgtgctgcg tgaaactgat tacgtcaccg gtgtacatca	540
ggttcacggg gggagtgggtg atccttctcc attcaccgca tatggcactc tgcaaggcct	600
gatggccgct ctgcaagtga aattcggtaa cgaagacgta ggcaattaca gctacgtgt	660
tcagggtgtg ggtcacgttg gcatggaatt tgttaaactg ctgcgtgagc gcggtgcaaa	720
ggttttcgtc actgacatca acaaagatgc ggtccagcgt gctgtggacg aatttggttg	780
tgaggcagta gccctggatg aaatctatga cgttgattgc gacgtgtact ccccgaccgc	840
tctgggcggc accgtgaacg ataaaacttt accgcgtctg aaatgtaagg taatctgcgg	900
tgcggcaaac aaccagttag ctaatgatga gataggcgtg gaactggaaa agaaaggcat	960
cctctatgct ccggactacg cgggtcaacgc ggggtgggctg atgaacgtta gcctggaaat	1020
cgatggatac aaccgcgaac gtgcgatgcg tatgatgcgt accattttatt acaatttggg	1080
tcgcattttc gaaatctcta agcgcgacgg catccctaca ttccgagccg ccgatcgtat	1140
ggctgaagaa cgcataacgg ccatcggtaa actgcgttta ccgcatttgg gcgctgcggc	1200

accgcgcttc	cagggccgac	gtggcaacta	agtcaataaa	ggcatataaa	aggaggttaa	1260
taacatgaaa	gttaaagtaa	aacatcttaa	tcatgcacag	gagactttct	aatgtcgacg	1320
acaaccgttg	gtgactactt	gctgtatcgc	ttaaacgaaa	tcggcattga	gcacctcttc	1380
ggagtgccag	gtgattacaa	tctgcaatth	ctggatcatg	taatcgacca	ccctcagctg	1440
acttgggtcg	gctgcactaa	cgaacttaac	gctgcctacg	cagctgatgg	ttatgcgcgt	1500
tgtcgtccgg	ctgcggcact	gctgaccacc	ttcgggggtt	gcgaactgag	cgctattaat	1560
ggcatcgcag	gttcctacgc	ggagtatctg	ccggtaatac	atatcgttgg	tgcaccgagt	1620
ctatcagccc	agcagcaggg	cgacctgatt	caccactctc	ttggcgaagg	tgatthttcc	1680
agcttcctga	ggatgtccca	accgggtgtct	gttgccgcagg	ctgctctgac	tcctgataac	1740
gcatgcaagg	aaatcgaccg	cgtactggcg	gaagtctca	ttcagcgtcg	tcccggctac	1800
ctgctgctgt	ctaccgacgt	ggctgctgcg	ccggcggctc	tgccacaaag	cactctttct	1860
ttgccgaccg	ccccggatca	tcgcgcagtt	ctggctgctt	tcagcgacgc	tgctgagcag	1920
atgctggctc	aggccaaaag	cgtctctcta	ctggcggact	ttctggctga	tcgtttcggt	1980
gttactcgag	cactggccgc	gtggcttcag	caggttccgc	taccgcacgc	cactctgtta	2040
atgggtaaag	gcgttctgag	tgaacagcaa	ccagggttcg	tgggtaccta	cgctgggtgcg	2100
gcatctatcg	attcgacgcg	tggcgcaatc	gaagaagctg	gggtaattat	cggagtggga	2160
gtagatthtt	ccgacactat	cacagcaggc	ttctcgcagc	agatcgacgc	ccgccgthtt	2220
atagacattc	aacccttctt	ctctcgtatt	ggcgatcgcc	agthttgatca	cttgccgatg	2280
caggctgccg	tcgcagccct	gcatcaactg	tgtcttcgth	atcagcagca	gtggtctatc	2340
accgctccta	gcccgcctgc	actgccgccg	gctgctggta	gcgagctgtc	ccagaacgca	2400
ttctggcagg	cgatgcagaa	cttcatccgc	cctgggggacc	tgthgggtggc	cgaccaaggth	2460
actgcggcgt	tcggcgcagc	ggcgctgcgc	ttaccgcaga	attgccagct	gcttgctgag	2520
ccgctgtggg	gctcaatcgg	ttacagtctg	ccggccacct	ttggtgctca	gacggcagat	2580
acagagcgtc	gtgtaatcct	aatcattggc	gatggthtcag	cgcaattaac	tattcaggaa	2640
ctthccagta	tgatgcgtga	cggcttgaaa	cctatcatct	ttctcctgaa	caacaacggth	2700
tacaccgttg	aacgggcgat	tcacggcgcg	gagcaacgth	ataacgatat	cgctgcttg	2760
aattggaccc	aactgcccc	ggcgctgagth	gttcattgcc	cagcgagag	ctggcgagth	2820

gttgaacg	gtcagctgac	cgacgtaatg	aaagtcac	ctgcttctcc	gcgtctgagc	2880
ttggtagaag	ttgttctgcc	tgcaatggat	gtcccaccgc	tgctgcaagc	agtgagtgcc	2940
gctctgaacc	agcgcaactc	ctcttaagac	cagaggtaag	gaggtaacaa	ccatgcgagt	3000
gttgaagaaa	catcttaatc	atgcagggga	gggtttctaa	tgcgcagcat	gcagtttgat	3060
gagtacggtg	caccctgaa	agcgttctca	tatgaagacc	cgaccccgca	agggaaggaa	3120
gtagtcgtta	ggatcgaagc	ctgtgggtgtg	tgccactctg	atattcatct	tcacgagggc	3180
tacttcgaca	tgggcggtgg	caataaagct	gatgttactc	gtgctcgcga	actccctttt	3240
acattgggtc	atgaaatcgt	tggcgaagtg	gtagcaactg	gaccaggtgt	caccggcgct	3300
aaaccggg	caaacgtat	tgtgtacccg	tggatcgggt	gcggcgactg	cccgaaatgc	3360
aacagtgggtg	aggatcagtc	ctgtgcgcgt	ccacgtaacc	tgggtgttca	cgttgacggt	3420
ggctattcga	cgcacgtaaa	gataccggac	gaaaaattcc	tgttcgcc	cgatggtatt	3480
cctactgagt	tagcgggaac	ctatgcttgc	agcggcatca	ccgcttatgg	tgcactgatg	3540
aaagcaaagg	aagcggctga	aagatctggc	tacatcggtc	tgattggcgc	tgggtggcgtt	3600
ggcatggctg	gtctgatgct	ggccaaagca	gcgatcgggg	ctaaaactgt	agtctttgat	3660
atcgacgacg	caaaactgga	agctgcgacc	cgtgccgggg	cggattacgt	gttcaactcc	3720
ggtgcaaaag	aaacacgcaa	ggaagttatg	aaactaacga	atggtggcct	gtctggtgct	3780
gttgatttcg	ttggcagcga	taaaagcgct	ctgtttggaa	tcaacgcctt	gggtcagaac	3840
ggcgtgctgg	tcataattgg	actgttcggt	ggcgctatga	ctgttccggt	accctgttc	3900
ccgctgaaag	ggatcacctg	acgtggctca	tacgtagggt	ccctgcaaga	gatgagtgat	3960
atgatggagt	tagttcgcgc	tgggaaagtt	cctccgatgc	cggtaaaaac	tcggccactg	4020
gacgctgcct	gggaaaccct	tgaggatcta	cgccatggta	aaatcgtggg	ccgtgttggt	4080
ctgaccccat	aagaaggaga	tatacatatg	acccatcaat	taagatcgcg	cgatatcatc	4140
gctctgggct	ttatgacatt	tgcgttggtc	gtcggcgcag	gtaacattat	tttccctcca	4200
atggtcggct	tgcaaggcagg	cgaacacgtc	tggactgcgg	cattcggcctt	cctcattact	4260
gccgttggcc	taccggtatt	aacggtagtg	gcgctggcaa	aagttggcgg	cgggtgtgac	4320
agtctcagca	cgccaattgg	taaagtcgct	ggcgtactgc	tggcaacagt	ttgttacctg	4380
gcggtggggc	cgctttttgc	tacgccgcgt	acagctaccg	tttcttttga	agtgggcatt	4440

gcgccgctga cgggtgattc cgcgctgccg ctgtttatatt acagcctggc ctatttcgct	4500
atcgttattc tggtttcgct ctatccgggc aagctgctgg ataccgtggg caacttcctt	4560
gcgccgctga aaattatcgc gctgggtcatc ctgtctgttg ccgcaattat ctggccggcg	4620
ggttctatca gtacggcgac tgaggcttat caaacgctg cgttttctaa cggcttcgtc	4680
aacggctatc tgaccatgga tacgctgggc gcaatgggtg ttggtatcgt tattgttaac	4740
gcggcgcggt ctcgtggcgt taccgaagcg cgtctgctga cccgttatac cgtctgggct	4800
ggcctgatgg cgggtgttgg tctgactctg ctgtacctgg cgctgttccg tctgggttca	4860
gacagcgctg cgctgggtcga tcagtctgca aacgggtgcgg cgatcctgca tgcttacgtt	4920
cagcatacct ttggcggcgg cggtagcttc ctgctggcgg cgttaatctt catcgctgc	4980
ctggtcacgg cggttggcct gacctgtgct tgtgcagaat tcttcgcca gtacgtaccg	5040
ctctcttatt gtacgctggg gtttatcctc ggcggttctt cgatgggtgg gtctaacctc	5100
ggcttgagcc agctgattca gatctctgta ccgggtgctga ccgccattta tccgccgtgt	5160
atcgactagg ttgtattaag ttttacacgc tcatgggtggc ataattcgtc ccgcgtgatt	5220
gctccgccga tgtttatcag cctgcttttt ggtattctcg acgggatcaa ggcatctgca	5280
ttcagcgata tcttaccgtc ctgggcgcag cgtttaccgc tggccgaaca aggtctggcg	5340
tggttaatgc caacagtggg gatgggtggt ctggccatta tctgggatcg tgcggcaggt	5400
cgtcaggtga cctccagcgc tcactaatac gcatggcatg gatgaccgat ggtagtgtgg	5460
gctctcccca tgcgagagta gggaactgcc aggcatt	5496

<210> 32

<211> 5484

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 32

gttgacactc tatcattgat agagttatatt taccactccc tatcagtgat agagaaaagt	60
gagcgaaaaa tcaataagga ggcaacaaga tgtgcgaaaa acatcttaat catgctaagg	120
aggttttcta atgcagatct tcgacacttt gcaatcaatg ggccatgagc aggttgtcct	180

atgtagcgat aagaccacgg gtctgcgcgc cattatcgct atacacgata catccttagg	240
gccggcgctt ggtggtaccc gtatgtggca gtatgcaact gacgacgatg ctattactga	300
cgcactccgt ctgtctcggg gcatgaccta caaagctgcg gtttctggcg taaatctggg	360
cgggtggtaaa gccgttatca tcggaaaccc tcacagtgat aaaagcgaag cgctgtttcg	420
cgcttacggc agaatggtgg aatcccagcg tgggcgttac atcaccgccg aagacgttgg	480
tactagcgta cgtgatatgg agtggattcg catggaaacc aaatatgtaa cgggcgtggg	540
tggcaacgga ggctctggtg acccctctcc agttaccgct ctgggtgttt actcgggcat	600
gaaggcatgc gctaaatcag tctatggtac tgatgcgctg agcggtaaaa ggatcgtggt	660
tcagggcgcg ggtaacgttg catcccatct ggttcacagt ctggtaaaag aaggcgctgt	720
ggttttcgtc actgacatct acgaagaaaa ggccaaagca ttagcggctg aaacgggcgc	780
taccgtgatt cgcaccgacg aggtttttac tacacaatgc gatatcttct ctccgaacgc	840
tctgggggcc gtcctgaacg atgaaactat tccgcagctc acatgcgcta tcgtagctgg	900
tggtgcaaac aatcagctta aaatcgaaca acgtcacgcc acggctctgc aagagaaagg	960
cattctgtat gcgccggatt acgtaatcaa cgccgggggc ctcataaatg tggcgagcga	1020
agttgacggc tacaaccgtg aaaaggttat gcgccaggct gaaggatattt acgatattac	1080
tatgaacatc ctaaataaccg cgcgtgagcg taacatcctg accatcgaag catccaacgc	1140
gattgctgaa gagcggatca acaaagttcg ccatgttcac gggaacttca tcggttcccc	1200
gtctattcgc ggagtataag tcaataaagg catataaaag gaggttaata acatgaaagt	1260
taaagtaaaa catcttaatc atgcacagga gactttctaa tgtcgacgac aaccgttggt	1320
gactacttgc tgtatcgctt aaacgaaatc ggcatcgagc acctcttcgg agtgccaggt	1380
gattacaatc tgcaatttct ggatcatgta atcgaccacc ctcagctgac ttgggtcggc	1440
tgactaacg aacttaacgc tgcctacgca gctgatgggt atgcgcgttg tcgtccggct	1500
gcggcactgc tgaccacctt cggggttggc gaactgagcg ctattaatgg catcgcaggt	1560
tcctacgcgg agtatctgcc ggtaatacat atcgttggtg caccgagtct atcagcccag	1620
cagcagggcg acctgattca ccactctctt ggcgaagggt atttttccag cttcctgagg	1680
atgtcccaac cgggtgtctgt tgcgcaggct gctctgactc ctgataacgc atgcaaggaa	1740
atcgaccgcg tactggcgga agtcctcatt cagcgtcgtc ccggctacct gctgctgtct	1800

accgacgtgg	ctgctgcgcc	ggcggctctg	ccacaaagca	ctcttttcttt	gccgaccgcc	1860
ccggatcatc	gcgcagttct	ggctgctttc	agcgacgctg	ctgagcagat	gctggctcag	1920
gccaaaagcg	tctctctact	ggcggacttt	ctggctgata	gtttcggtgt	tactcgagca	1980
ctggccgcgt	ggcttcagca	ggttccgcta	ccgcacgcca	ctctgttaat	gggtaaaggc	2040
gttctgagt	aacagcaacc	agggttcgtg	ggtacctacg	ctggcgcggc	atctatcgat	2100
tcgacgcgtg	gcgcaatcga	agaagctggg	gtaattatcg	gagtgggagt	tagattttcc	2160
gacactatca	cagcaggctt	ctcgcagcag	atcgacgccc	gccgttttat	agacattcaa	2220
cccttcttct	ctcgtattgg	cgatcgccag	tttgatcact	tgccgatgca	ggctgccgtc	2280
gcagccctgc	atcaactgtg	tcttcgttat	cagcagcagt	ggtctatcac	cgctcctagc	2340
ccgcctgcac	tgccgccggc	tgctggtagc	gagctgtccc	agaacgcatt	ctggcaggcg	2400
atgcagaact	tcattccgcc	tggggacctg	ttggcggccg	accaagggtac	tgccggcggtc	2460
ggcgcagcgg	cgctgcgctt	accgcagaat	tgccagctgc	ttgtgcagcc	gctgtggggc	2520
tcaatcggtt	acagtctgcc	ggccaccttt	ggtgctcaga	cggcagatac	agagcgtcgt	2580
gtaatcctaa	tcattggcga	tggttcagcg	caattaacta	ttcagggaact	ttccagtatg	2640
atgcgtgacg	gcttgaaacc	tatcatcttt	ctcctgaaca	acaacgggta	caccgttgaa	2700
cgggcgattc	acggcgcgga	gcaacgttat	aacgatatac	ctgcttgga	ttggacccaa	2760
ctgccccagg	cgctgagtgt	tcattgcccc	gcgcagagct	ggcagagtcgt	tgaacgggtg	2820
cagctgaccg	acgtaatgaa	agtcacgct	gcttctccgc	gtctgagctt	ggtagaagtt	2880
gttctgcctg	caatggatgt	cccaccgctg	ctgcaagcag	tgagtgccgc	tctgaaccag	2940
cgcaactcct	cttaagacca	gaggtaagga	ggtaacaacc	atgcgagtgt	tgaagaaaca	3000
tcttaatcat	gcaggggagg	gtttctaatg	cgcagcatgc	agtttgatga	gtacggtgca	3060
cccctgaaag	cgttctcata	tgaagaccgc	accccgcaag	ggaagggaagt	agtcgttagg	3120
atcgaagcct	gtggtgtgtg	ccactctgat	attcatcttc	acgagggcta	cttcgacatg	3180
ggcgggtggca	ataaagctga	tgttactcgt	gctcgcgaac	tcccttttac	attgggtcat	3240
gaaatcgttg	gcgaagtggg	agcaactgga	ccagggtgtca	ccggcgctaa	accgggcgac	3300
aaacgtattg	tgtaccgctg	gatcgggtgc	ggcgactgcc	cgaaatgcaa	cagtgggtgag	3360
gatcagtcct	gtgcgcgtcc	acgtaacctg	ggtgttcacg	ttgacgggtg	ctattcgacg	3420

cacgtaaaga taccggacga aaaattcctg ttcgcctacg atggatttcc tactgagtta	3480
gcgggaacct atgcttgacg cggcatcacc gcttatgggtg cactgatgaa agcaaaggaa	3540
gcggctgaaa gatctggcta catcgggtctg attggcgctg gtggcggttg catggctggt	3600
ctgatgctgg ccaaagcagc gatcggggct aaaactgtag tctttgatata cgacgacgca	3660
aaactggaag ctgcgacccg tgccggggcg gattacgtgt tcaactccgg tgcaaaagaa	3720
acacgcaagg aagttatgaa actaacgaat ggtggcctgt ctggtgctgt tgatttcgtt	3780
ggcagcgata aaagcgctct gtttggaaac aacgccttgg gtcagaacgg cgtgctggtc	3840
ataattggac tgttcgggtg cgctatgact gttccggtac ccctgttccc gctgaaaggg	3900
atcaccttac gtggctcata cgtagggttcc ctgcaagaga tgagtatat gatggagtta	3960
gttcgcgctg ggaaagtacc tccgatgccg gtaaaaactc ggccactgga cgctgcctgg	4020
gaaacccttg aggatctacg ccatggtaaa atcgtgggcc gtgttggttct gacccataa	4080
gaaggagata tacatatgac ccatcaatta agatcgcgcg atatcatcgc tctgggcttt	4140
atgacatttg cgttggtcgt cggcgcgagt aacattatct tccctccaat ggtcggcttg	4200
caggcaggcg aacacgtctg gactgcggca ttcggcttcc tcattactgc cgttggccta	4260
ccggtattaa cggtagtggc gctggcaaaa gttggcggcg gtgttgacag tctcagcacg	4320
ccaattggta aagtcgctgg cgtactgctg gcaacagttt gttacctggc ggtggggccg	4380
ctttttgcta cgccgcgtac agctaccgtt tcttttgaag tgggcattgc gccgctgacg	4440
ggtgattccg cgctgccgtt gtttattttac agcctggtct atttcgctat cgttattctg	4500
gtttcgctct atccgggcaa gctgctggat accgtgggca acttccttgc gccgctgaaa	4560
attatcgcgc tggatcatct gtctgttgcc gcaattatct ggccggcggg ttctatcagt	4620
acggcgactg aggcttatca aaacgtgcg ttttctaacg gcttcgtcaa cggctatctg	4680
accatggata cgctgggcgc aatgggtgtt ggtatcggtt ttgttaacgc ggcgcgttct	4740
cgtggcggtt ccgaagcgcg tctgctgacc cgttataccg tctgggctgg cctgatggcg	4800
ggtgttggtc tgactctgct gtacctggcg ctgttccgtc tgggttcaga cagcgcgctc	4860
ctggtcgatc agtctgcaaa cgggtcggcg atcctgcatg cttacgttca gcataccttt	4920
ggcggcggcg gtagcttctt gctggcggcg ttaatcttca tcgcctgcct ggtcacggcg	4980
gttggcctga cctgtgcttg tgcagaattc ttcgccagc acgtaccgct ctcttatcgt	5040

acgctggtgt ttatcctcgg cggcttctcg atggtggtgt ctaacctcgg cttgagccag	5100
ctgattcaga tctctgtacc ggtgctgacc gccatttata cgcctgttat cgcactgggtt	5160
gtattaagtt ttacacgctc atggtggcat aattcgtccc gcgtgattgc tccgccgatg	5220
tttatcagcc tgcttttttg tattctcgac gggatcaagg catctgcatt cagcgatata	5280
ttaccgtcct gggcgcagcg ttaccgctg gccgaacaag gtctggcgtg gttaatgcca	5340
acagtgggtga tgggtggttct ggccattata tgggatcgtg cggcaggctc tcaggtgacc	5400
tccagcgctc actaatacgc atggcatgga tgaccgatgg tagtgtgggc tctccccatg	5460
cgagagtagg gaactgccag gcat	5484

<210> 33

<211> 5338

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 33

gttgacactc tatcattgat agagttatth taccactccc tatcagtgat agagaaaagt	60
gagataaacg aatcacaaag gcgaggcaag atgaacatct tcaagaaaat ggaggaattt	120
aattatgaac aactggtcta cttctacgac agcgaaacgg aactcaaagg tattacctgt	180
atacacaaca caactttagg gccggcattg ggcggtaccc gcctttggaa ctataactct	240
gaggaagatg ccgttgaaga cgtaatccgt ctggctcggg gcatgactta caaagcggct	300
tgcgccggtc tgaatctggg cggcggtaaa accgtgctga tcggtgatgc taaaaagatt	360
aaatcagagt cctacttccg tggactgggg cgctacgttc agtcgctgaa cggcagatat	420
atcaccgcgg aagacgtaaa tactttctac aaggatatgg catacgttgc tatggaaact	480
gactatgtgg taggcctggg aggtaaatcc ggcaacccta gtccagttac tgcttacggt	540
gcatttatgg gtatcaaagc ggcgctgatg aaaaaatttg aggatagctc tattgaaggc	600
cgaaccttcg cagtgcaggg tgctgggcag acgggttact atcttatcga ttacctccta	660
ggcaacaaca agttcaaaga aaaggctaaa aagatttact tcaccgaaat taacgagagc	720
tatatcgagc gtatgaacaa agaacatccg gaagttgaat ttatttcccc ggacaaaatc	780
tactcgctgg aagtagacgt cttcgtgccc tgcgccctgg gcaaaatcgt taatgacaaa	840

actatcgatg aattttaagtg tccgatcatc gcaggtactg caaacaacgt actggaaagg	900
gaagcgcacg gcaacatgct taaagaacgt ggcatctctt acgccccgga ctatgtgatc	960
aatgctgggtg ggctgatcaa cgtttaccac gagctgaacg gttacaataa agagaacgct	1020
attctggaag tggaattaat ttatgatcgc ctactggaaa tattcaacat cgctgattct	1080
ctgaacatca gcaccaatat cgctgccaac gagttcgcgg aaaaacgtat caagcaaatt	1140
aagtccttga aaaacaactt cattaaacgc taagaccatt aataaggaac attttatgtc	1200
gacgacaacc gttggtgact acttgctgta tcgcttaaac gaaatcggca ttgagcacct	1260
cttcggagtg ccaggtgatt acaatctgca atttctggat catgtaatcg accaccctca	1320
gctgacttgg gtcggctgca ctaacgaact taacgctgcc tacgcagctg atggttatgc	1380
gcgttgtcgt ccggctgcgg cactgctgac caccttcggg gttggcgaac tgagcgctat	1440
taatggcatc gcaggttcct acgcggagta tctgccggta atacatatcg ttggtgcacc	1500
gagtctatca gcccagcagc agggcgacct gattcaccac tctcttggcg aaggtgattt	1560
ttccagcttc ctgaggatgt cccaaccggt gtctgttgcg caggctgctc tgactcctga	1620
taacgcatgc aaggaaatcg accgcgtact ggcggaagtc ctcatcagc gtcgtcccgg	1680
ctacctgctg ctgtctaccg acgtggctgc tgcgccggcg gctctgccac aaagcactct	1740
ttctttgccg accgccccgg atcatcgcgc agttctggct gctttcagcg acgctgctga	1800
gcagatgctg gctcaggcca aaagcgtctc tctactggcg gactttctgg ctgatcgttt	1860
cgggtgttact cgagcactgg ccgcgtggct tcagcagggt cgcctaccgc acgccactct	1920
gttaatgggt aaaggcgttc tgagtgaaca gcaaccaggg ttcgtgggta cctacgctgg	1980
tgcggcatct atcgattcga cgcgtggcgc aatcgaagaa gctggggtaa ttatcggagt	2040
gggagttaga ttttccgaca ctatcacagc aggcttctcg cagcagatcg acgcccgcg	2100
ttttatagac attcaaccct tcttctctcg tattggcgat cgccagtttg atcacttgcc	2160
gatgcaggct gccgtcgcag ccctgcatca actgtgtctt cgttatcagc agcagtggtc	2220
tatcaccgct cctagcccg cctgactgcc gccggctgct ggtagcgagc tgtcccagaa	2280
cgcatcttgg caggcgatgc agaacttcat ccgccctggg gacctgttgg tggccgacca	2340
aggtactgcg gcgttcggcg cagcggcgct gcgcttaccg cagaattgcc agctgcttgt	2400
gcagccgctg tggggctcaa tcggttacag tctgccggcc acctttggtg ctcagacggc	2460

agatacagag cgtcgtgtaa tcctaatacat tggcgatggg tcagcgcaat taactattca	2520
ggaactttcc agtatgatgc gtgacggctt gaaacctatc atctttctcc tgaacaacaa	2580
cggttacacc gttgaacggg cgattcacgg cgcggagcaa cgttataacg atatcgctgc	2640
ttggaattgg acccaactgc cccaggcgct gagtgttcat tgcccagcgc agagctggcg	2700
agtcgttgaa acgggtgcagc tgaccgacgt aatgaaagtc atcgctgctt ctccgcgtct	2760
gagcttggtg gaagttgttc tgcctgcaat ggatgtccca ccgctgctgc aagcagtga	2820
tgccgctctg aaccagcgca actcctctta atccctccct acgggttaag gagcccctgc	2880
tatgcgcagc atgcagtttg atgagtacgg tgcacccctg aaagcgttct catatgaaga	2940
cccgaccccg caagggaagg aagtagtcgt taggatcgaa gcctgtgggtg tgtgccactc	3000
tgatattcat cttcacgagg gctacttcga catgggcggg ggcaataaag ctgatgttac	3060
tcgtgctcgc gaactccctt ttacattggg tcatgaaatc gttggcgaag tggtagcaac	3120
tggaccaggt gtcaccggcg ctaaaccggg cgacaaacgt attgtgtacc cgtggatcgg	3180
gtcgggcgac tgcccgaat gcaacagtgg tgaggatcag tcctgtgcgc gtccacgtaa	3240
cctgggtgtt cacgttgacg gtggctattc gacgcacgta aagataccgg acgaaaaatt	3300
cctgttcgcc tacgatggta ttctactga gttagcggga acctatgctt gcagcggcat	3360
caccgcttat ggtgcactga tgaaagcaaa ggaagcggct gaaagatctg gctacatcgg	3420
tctgattggc gctgggtggcg ttggcatggc tggctgatg ctggccaaag cagcgatcgg	3480
ggctaaaact gtagtctttg atatcgacga cgcaaaactg gaagctgcga cccgtgccgg	3540
ggcggattac gtgttcaact ccggtgcaaa agaaacacgc aaggaagtta tgaaactaac	3600
gaatggtggc ctgtctgggtg ctgttgattt cgttggcagc gataaaaagcg ctctgtttgg	3660
aatcaacgcc ttgggtcaga acggcgtgct ggtcataatt ggactgttcg gtggcgctat	3720
gactgttccg gtacccctgt tcccgtgaa agggatcacc gtacgtggct catacgtagg	3780
ttccctgcaa gagatgagtg atatgatgga gttagtctgc gctgggaaag ttccctccgat	3840
gccggtaaaa actcggccac tggacgctgc ctgggaaacc cttgaggatc tacgccatgg	3900
taaaatcgtg ggccgtgttg ttctgacccc ataagaagga gatatacata tgacccatca	3960
attaagatcg cgcgatatca tcgctctggg ctttatgaca tttgcgttgt tcgtcggcgc	4020
aggtaacatt attttccctc caatggtcgg cttgcaggca ggcgaaacag tctggactgc	4080

ggcattcggc ttcctcatta ctgccgttgg cctaccggta ttaacggtag tggcgctggc	4140
aaaagttagc ggcggtgttg acagtctcag caccgaatt ggtaaagtcg ctggcgctact	4200
gctggcaaca gtttgttacc tggcggtggg gccgcttttt gctacgccgc gtacagctac	4260
cgtttctttt gaagtgggca ttgcgccgct gacgggtgat tccgcgctgc cgctgtttat	4320
ttacagcctg gtctatttcg ctatcgttat tctggtttcg ctctatccgg gcaagctgct	4380
ggataccgtg ggcaacttcc ttgcgccgct gaaaattatc gcgctgggtca tcctgtctgt	4440
tgccgcaatt atctggccgg cgggttctat cagtacggcg actgaggctt atcaaaacgc	4500
tgcgttttct aacggcttcg tcaacggcta tctgaccatg gatacgctgg gcgcaatgg	4560
gtttggtatc gttattgtta acgcggcgcg ttctcgtggc gttaccgaag cgcgtctgct	4620
gacccgttat accgtctggg ctggcctgat ggcgggtgtt ggtctgactc tgctgtacct	4680
ggcgctgttc cgtctgggtt cagacagcgc gtcgctggtc gatcagtctg caaacgggtgc	4740
ggcgatcctg catgcttacg ttcagcatac ctttggcggc ggcggtagct tcctgctggc	4800
ggcgtaatac ttcacgcct gcctggtcac ggcggttggc ctgacctgtg cttgtgcaga	4860
attcttcgcc cagtacgtac cgctctctta tcgtacgctg gtgtttatcc tcggcggctt	4920
ctcgatggtg gtgtctaacc tcggcttgag ccagctgatt cagatctctg taccggtgct	4980
gaccgccatt tatccgccgt gtatcgcaact ggttgtatta agttttacac gctcatggtg	5040
gcataattcg tcccgctga ttgctccgcc gatgtttatc agcctgcttt ttggtattct	5100
cgacgggatc aaggcatctg cattcagcga tatcttaccg tcctgggcgc agcgtttacc	5160
gctggccgaa caaggtctgg cgtggttaat gccaacagtg gtgatggtgg ttctggccat	5220
tatctgggat cgtgcggcag gtcgtcaggt gacctccagc gctcactaat acgcatggca	5280
tggatgaccg atggtagtgt gggctctccc catgcgagag tagggaactg ccaggcat	5338

<210> 34

<211> 5336

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 34

gttgacactc tatcattgat agagttat	ttt taccactccc tatcagtgat agagaaaagt	60
gagagctaga taataaggac ggcacaagat	gaacatcttc aagaaaatgg aggaatttaa	120
ttatgaacaa ctggtctact tctacgacag	cgaaacggaa ctcaaaggta ttacctgtat	180
acacaacaca acttttagggc cggcattggg	cggtagccgc ctttggaact ataactctga	240
ggaagatgcc gttgaagacg taatccgtct	ggctcggggc atgacttaca aagcggcttg	300
cgccggtctg aatctgggcg gcggtaaaac	cgtgctgata ggtgatgcta aaaagattaa	360
atcagagtcc tacttccgtg gactggggcg	ctacgttcag tcgctgaacg gcagatatat	420
caccgcggaa gacgtaaata cttctacgaa	ggatatggca tacgttgcta tggaaactga	480
ctatgtggta ggcctgggag gtaaattcgg	caaccctagt ccagttactg cttacgggtgc	540
atttatgggt atcaaagcgg cgctgatgaa	aaaatttgag gatagctcta ttgaaggccg	600
aaccttcgca gtgcagggtg ctgggcagac	gggttactat cttatcgatt acctcctagg	660
caacaacaag ttcaaagaaa aggctaaaaa	gatttacttc accgaaatta acgagagcta	720
tatcgagcgt atgaacaaag aacatccgga	agttgaattt atttccccgg acaaaatcta	780
ctcgctggaa gtagacgtct tcgtgccctg	cgccctgggc aaaatcgta atgacaaaac	840
tatcgatgaa tttaagtgtc cgatcatcgc	aggtactgca aacaacgtac tggaaaggga	900
agcgcacggc aacatgctta aagaacgtgg	cattctttac gccccggact atgtgatcaa	960
tgctgggtggg ctgatcaacg tttaccacga	gctgaacggg tacaataaag agaacgctat	1020
tctggaagtg gaattaattt atgatcgcct	actggaaata ttcaacatcg ctgattctct	1080
gaacatcagc accaatatcg ctgccaacga	gttcgcggaa aaacgtatca agcaaatata	1140
gtccttgaaa aacaacttca ttaaacgcta	agaccattaa taaggaacat tttatgtcga	1200
cgacaaccgt tgggtgactac ttgctgtatc	gcttaaacga aatcggcatt gagcacctct	1260
tcggagtgcc aggtgattac aatctgcaat	ttctggatca tgtaatcgac caccctcagc	1320
tgacttgggt cggctgcact aacgaactta	acgctgccta cgcagctgat ggttatgcgc	1380
gttgtcgtcc ggctgcggca ctgctgacca	ccttcggggg tggcgaactg agcgcattta	1440
atggcatcgc aggttcctac gcggagtatc	tgccggtaat acatatcggt ggtgcaccga	1500
gtctatcagc ccagcagcag ggcgacctga	ttcaccactc tcttggcgaa ggtgattttt	1560
ccagcttcct gaggatgtcc caaccggtgt	ctgttgcgca ggctgctctg actcctgata	1620

acgcatgcaa ggaaatcgac cgcgtactgg cggaagtcct cattcagcgt cgtccccggct	1680
acctgctgct gtctaccgac gtggctgctg cgccggcggc tctgccacaa agcactcttt	1740
ctttgccgac cgccccggat catcgcgcag ttctggctgc tttcagcgac gctgctgagc	1800
agatgctggc tcaggccaaa agcgtctctc tactggcggc ctttctggct gatcgtttcg	1860
gtgttactcg agcactggcc gcgtggcttc agcaggttcc gctaccgcac gccactctgt	1920
taatgggtaa aggcgttctg agtgaacagc aaccagggtt cgtgggtacc tacgctggtg	1980
cggcatctat cgattcgacg cgtggcgcaa tcgaagaagc tggggtaatt atcggagtgg	2040
gagttagatt ttccgacact atcacagcag gcttctcgca gcagatcgac gcccgccgtt	2100
ttatagacat tcaacccttc ttctctcgta ttggcgatcg ccagtttgat cacttgccga	2160
tgcaggctgc cgtcgcagcc ctgcatcaac tgtgtcttcg ttatcagcag cagtgggtcta	2220
tcaccgctcc tagcccgctt gcactgccgc cggctgctgg tagcgagctg tcccagaacg	2280
cattctggca ggcgatgcag aacttcatcc gccctgggga cctgttggtg gccgaccaag	2340
gtactgcggc gttcggcgca gcggcgctgc gcttaccgca gaattgccag ctgcttgctc	2400
agccgctgtg gggctcaatc ggttacagtc tgccggccac ctttggtgct cagacggcag	2460
atacagagcg tcgtgtaatc ctaatcattg gcgatggttc agcgcaatta actattcagg	2520
aactttccag tatgatgcgt gacggcttga aacctatcat ctttctcctg aacaacaacg	2580
gttacaccgt tgaacgggcg attcacggcg cggagcaacg ttataacgat atcgctgctt	2640
ggaattggac ccaactgccc caggcgctga gtgttcattg cccagcgagc agctggcgag	2700
tcgttgaaac ggtgcagctg accgacgtaa tgaaagtcac cgctgcttct ccggtctga	2760
gcttggtaga agttgttctg cctgcaatgg atgtcccacc gctgctgcaa gcagtgagtg	2820
ccgctctgaa ccagcgcaac tcctcttaat ccctccctac gggttaagga gccctgcta	2880
tgcgcagcat gcagtttgat gagtacggtg caccctgaa agcgttctca tatgaagacc	2940
cgaccccgca agggaaggaa gtagtcgtta ggatcgaagc ctgtgggtgtg tgccactctg	3000
atattcatct tcacgagggc tacttcgaca tgggcgggtg caataaagct gatgttactc	3060
gtgctcgcga actccctttt acattgggtc atgaaatcgt tggcgaagtg gtagcaactg	3120
gaccaggtgt caccggcgct aaaccgggcg acaaacgtat tgtgtaccg tggatcgggt	3180
gcggcgactg ccgaaatgc aacagtgggtg aggatcagtc ctgtgcgcgt ccacgtaacc	3240

tgggtgttca cgttgacggt ggctattcga cgcacgtaaa gataccggac gaaaaattcc	3300
tgttcgccta cgatgggtatt cctactgagt tagcgggaac ctatgcttgc agcggcatca	3360
ccgcttatgg tgcactgatg aaagcaaagg aagcggctga aagatctggc tacatcggtc	3420
tgattggcgc tgggtggcggt ggcatggctg gtctgatgct ggccaaagca gcgatcgggg	3480
ctaaaactgt agtctttgat atcgacgacg caaaactgga agctgcgacc cgtgccgggg	3540
cggattacgt gttcaactcc ggtgcaaaag aaacacgcaa ggaagttatg aaactaacga	3600
atgggtggcct gtctgggtgct gttgatttcg ttggcagcga taaaagcgct ctgtttggaa	3660
tcaacgcctt gggtcagaac ggcgtgctgg tcataattgg actgttcggt ggcgctatga	3720
ctgttccggt acccctgttc ccgctgaaag ggatcacctg acgtggctca tacgtagggt	3780
ccctgcaaga gatgagtgat atgatggagt tagttcgcgc tgggaaagtt cctccgatgc	3840
cggtaaaaac tcggccactg gacgctgcct gggaaaccct tgaggatcta cgccatggta	3900
aaatcgtggg ccgtgttggt ctgaccccat aagaaggaga tatacatatg acccatcaat	3960
taagatcgcg cgatatcatc gctctgggct ttatgacatt tgcgttggtc gtcggcgcag	4020
gtaacattat tttccctcca atggtcggct tgcaggcagg cgaacacgtc tggactgcgg	4080
cattcggctt cctcattact gccgttggcc taccgggtatt aacggtagtg gcgctggcaa	4140
aagttggcgg cgggtgtgac agtctcagca cgccaattgg taaagtcgct ggcgtactgc	4200
tggcaacagt ttgttacctg gcggtggggc cgctttttgc tacgccgcgt acagctaccg	4260
tttcttttga agtgggcatt gcgccgctga cgggtgattc cgcgctgccg ctgtttattt	4320
acagcctggt ctatttcgct atcgttattc tggtttcgct ctatccgggc aagctgctgg	4380
ataccgtggg caacttcctt gcgccgctga aaattatcgc gctggtcatc ctgtctgttg	4440
ccgcaattat ctggccggcg ggttctatca gtacggcgac tgaggcttat caaaacgctg	4500
cgttttctaa cggcttcgtc aacggctatc tgaccatgga tacgctgggc gcaatgggtg	4560
ttggtatcgt tattgttaac gcggcgcgtt ctcgtggcgt taccgaagcg cgtctgctga	4620
cccgttatac cgtctgggct ggcctgatgg cgggtgttgg tctgactctg ctgtacctgg	4680
cgctgttccg tctgggttca gacagcgcgt cgctggctga tcagtctgca aacggtgcgg	4740
cgatcctgca tgcttacgtt cagcatacct ttggcggcgg cggtagcttc ctgctggcgg	4800
cgtaaatctt catcgcctgc ctggtcacgg cggttggcct gacctgtgct tgtgcagaat	4860

tcttcgcca gtacgtaccg ctctcttatac gtacgctggg gtttatcctc ggcggttct	4920
cgatgggtggg gtctaacctc ggcttgagcc agctgattca gatctctgta ccggtgctga	4980
ccgccattta tccgccgtgt atcgcaactgg ttgtattaag ttttacacgc tcatgggtggc	5040
ataattcgtc ccgcgtgatt gctccgccga tgtttatcag cctgcttttt ggtattctcg	5100
acgggatcaa ggcatctgca ttcagcgata tcttaccgtc ctgggcgcag cgtttaccgc	5160
tggccgaaca aggtctggcg tggttaatgc caacagtggg gatgggtgggt ctggccatta	5220
tctgggatcg tgcggcaggt cgtcaggtga cctccagcgc tcactaatac gcatggcatg	5280
gatgaccgat ggtagtgtgg gctctcccca tgcgagagta gggaactgcc aggcatt	5336

<210> 35

<211> 5385

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 35

gttgacactc tatcattgat agagttatct taccactccc tatcagtgat agagaaaagt	60
gaactacagt atataaggag ggtaatatga tcttcgagac aatttctacg tcgaatcacg	120
aagaagttgt gtattgccat aacaaggacg ccggcttgaa agcaatcatc gcgattcaca	180
acactgtact cgggtccggc ctgggtggca ctgcgatgtg gccctacgct agcgaagagg	240
aagcactgaa agatgtcctt cgtttatccc gtgggatgac ctacaaagct gcggtttcag	300
gtctaaacct gggcggcggg aaagctgtga tctgggggtga tccgaataaa gacaagtctg	360
aagcgtgtt tagagccttc ggacggtttg taaacagcct gggcggacgc tacattaccg	420
cggaggacgt tggcattgat gttaacgaca tggaatatgt gctgcgtgaa actgattacg	480
tcaccggtgt acatcagggt cacgggtggga gtgggtgatcc ttctccattc accgcatatg	540
gcactctgca aggcctgatg gccgctctgc aagtgaaatt cggtaacgaa gacgtaggca	600
attacagcta cgctgttcag ggtgtgggtc acgttggcat ggaatttggt aaactgctgc	660
gtgagcgcgg tgcaaagggt ttcgtcactg acatcaacaa agatgcggtc cagcgtgctg	720
tggacgaatt tggttgtgag gcagtagccc tggatgaaat ctatgacgtt gattgacgac	780
tgtactcccc gaccgctctg ggcggcaccg tgaacgataa aactttaccg cgtctgaaat	840

gtaaghtaata	ctgcggtgca	gcaaaacaacc	agttagctaa	tgatgagata	ggcgtggaac	900
tggaataagaa	aggcatcctc	tatgctccgg	actacgcggt	caacgcgggt	gggctgatga	960
acgttagcct	ggaaatcgat	ggatacaacc	gcgaacgtgc	gatgcgtatg	atgcgtacca	1020
tttattacaa	tttgggtcgc	attttcgaaa	tctctaagcg	cgacggcatc	cctacattcc	1080
gagccgccga	tcgtatggct	gaagaacgca	taacggccat	cggtaaaactg	cgtttaccgc	1140
at ttggggcgc	tgcggcaccg	cgcttccagg	gccgacgtgg	caactaatta	cgacatttta	1200
aggaggttcg	catgtcgacg	acaaccgttg	gtgactactt	gctgtatcgc	ttaaacgaaa	1260
tcggcattga	gcacctcttc	ggagtgccag	gtgattacaa	tctgcaattt	ctggatcatg	1320
taatcgacca	ccctcagctg	acttgggtcg	gctgcactaa	cgaacttaac	gctgcctacg	1380
cagctgatgg	ttatgcgcgt	tgctgtccgg	ctgcggcact	gctgaccacc	ttcgggggtg	1440
gcgaactgag	cgctattaat	ggcatcgcag	gttcctacgc	ggagtatctg	ccggtaatac	1500
atatcgttgg	tgcaaccgagt	ctatcagccc	agcagcaggg	cgacctgatt	caccactctc	1560
ttggcgaagg	tgat ttttcc	agcttcctga	ggatgtccca	accggtgtct	gttgcgcagg	1620
ctgctctgac	tcctgataac	gcatgcaagg	aaatcgaccg	cgtactggcg	gaagtcctca	1680
ttcagcgtcg	tcccggctac	ctgctgctgt	ctaccgacgt	ggctgctgcg	ccggcggtc	1740
tgccacaaag	cactctttct	ttgccgaccg	ccccggatca	tcgcgagtt	ctggctgctt	1800
tcagcgacgc	tgctgagcag	atgctggctc	aggccaaaag	cgtctctcta	ctggcggact	1860
ttctggctga	tcgtttcggt	gttactcgag	cactggccgc	gtggcttcag	caggttccgc	1920
taccgcacgc	cactctgtta	atgggtaaag	gcgttctgag	tgaacagcaa	ccagggttcg	1980
tgggtaccta	cgctgggtgcg	gcatctatcg	attcgacgcg	tggcgcaatc	gaagaagctg	2040
gggtaattat	cggagtggga	gttagat ttt	ccgacactat	cacagcaggc	ttctcgcagc	2100
agatcgacgc	ccgccgtttt	atagacattc	aacccttctt	ctctcgtatt	ggcgatcgcc	2160
agtttgatca	cttgccgatg	caggctgccg	tcgcagccct	gcatcaactg	tgtcttcggt	2220
atcagcagca	gtggctctatc	accgctccta	gcccgcctgc	actgccgccg	gctgctggta	2280
gcgagctgtc	ccagaacgca	ttctggcagg	cgatgcagaa	cttcatccgc	cctggggacc	2340
tgttgggtggc	cgaccaaggt	actgcggcgt	tcggcgagc	ggcgctgctc	ttaccgcaga	2400
attgccagct	gcttgtgcag	ccgctgtggg	gctcaatcgg	ttacagtctg	ccggccacct	2460

ttggtgctca gacggcagat acagagcgtc gtgtaatcct aatcattggc gatggttcag	2520
cgcaattaac tattcaggaa ctttccagta tgatgcgtga cggcttgaaa cctatcatct	2580
ttctcctgaa caacaacggt tacaccgttg aacgggcgat tcacggcgcg gagcaacggt	2640
ataacgatat cgctgcttgg aattggaccc aactgcccc a ggcgctgagt gttcattgcc	2700
cagcgcagag ctggcgagtc gttgaaacgg tgcagctgac cgacgtaatg aaagtcatcg	2760
ctgcttctcc gcgtctgagc ttggtagaag ttgttctgcc tgcaatggat gtcccaccgc	2820
tgctgcaagc agtgagtgcc gctctgaacc agcgcaactc ctcttaaatt tttcaacggc	2880
ctatcataag gaggaataat gacagcggag cagcaaaatg gggatatccga ctcacgccgt	2940
ttcgaatttc aggaatttgg tggccctatc gccccacaga cctatcagct ccccgccaccg	3000
gctagcgatg aagttttgtt aaaggtgaac tactgcggtg tctgtcacag tgatgttcat	3060
cttcacgacg gctacttcga gctgggtggc gataaacgtc tgaacttcgc tatgccgctg	3120
ccgctgacgc tgggtcacga agtaattggc accgttgtgg ctgtcggcga ccaggttact	3180
ggtgtaaaac cgggggacca gcgactgac tatccgtgga taggttgagg aaaatgcggc	3240
gcgtgtcaaa aaggagaaga aaacctgtgc gttactcctg cacatctggg cgtgaacaag	3300
ccgggcggtt acgctgatca catcgttgta cccattctc gctaccttct ggacatttcg	3360
ggtctgaacc cgggtgatgc cgctaccctc gcgtgctccg gcctgaccac tttcagcgcg	3420
atcaacaaag tgttgccgct tgcagatgac cagtggattg ttgttatcgg ttgtgggtggc	3480
ctcggccaga tggcgctgcg tatcctgcaa gctatgggaa ttggcaatgt tatcggtatt	3540
gacctgtctg aagagaaacg gaaactggct catgaaagcg gtgcacgtca ctccttcgat	3600
ccaaacactc cgaagctgaa ccgcgtggtc gccgaaacct gcccgggtac ggtacaggcc	3660
gcgttagact ttgtgggcaa tgagcaaact gctcagctgg cactgtctct gcttggaaaa	3720
ggtggcaa atgttcctgt cgggctgcac ggcggcgagc tgcgttacct attgccgatc	3780
atcacgaaca aagctgtaag tatcatcggg tcttacgttg gtaccctgaa agaactggaa	3840
gacttagttg ctttcgcaa ggaaaagaat ctgccgcaa ttcataattga acaccgccc	3900
ctggaatcgg cggctcaggc cgtagaggac ctggaaaaag gacaggttgc tgggcgtgtt	3960
atcctggatg caggtaacta agaaggagat atacatatga cccatcaatt aagatcgcgc	4020
gatatcatcg ctctgggctt tatgacattt gcgttggttcg tcggcgagg taacattatt	4080

ttccctccaa tggtcggcctt gcaggcaggc gaacacgtct ggactgcggc attcggcttc	4140
ctcattactg ccgttggcct accggtatta acggtagtgg cgctggcaaa agttggcggc	4200
ggtgttgaca gtctcagcac gccaattgggt aaagtcgctg gcgtactgct ggcaacagtt	4260
tgttacctgg cggtggggcc gctttttgct acgccgcgta cagctaccgt ttcttttgaa	4320
gtgggcattg cgccgctgac ggggtgattcc gcgctgccgc tgtttattta cagcctggtc	4380
tatttcgcta tcgttattct ggtttcgtc tatccgggca agctgctgga taccgtgggc	4440
aatttccttg cgccgctgaa aattatcgcg ctgggtcatcc tgtctgttgc cgcaattatc	4500
tggccggcgg gttctatcag tacggcgact gaggcttatc aaaacgctgc gttttctaac	4560
ggcttcgtca acggctatct gaccatggat acgctgggcg caatgggtgtt tggatatcgtt	4620
attgttaacg cggcgcgttc tcgtggcggt accgaagcgc gtctgctgac ccgttatacc	4680
gtctgggctg gcctgatggc ggggtgttgggt ctgactctgc tgtacctggc gctgttccgt	4740
ctgggttcag acagcgcgtc gctggtcgat cagtctgcaa acggtgccgc gatcctgcat	4800
gcttacgttc agcatacctt tggcggcggc ggtagcttcc tgctggcggc gttaatcttc	4860
atgcctgcc tggtcacggc ggttggcctg acctgtgctt gtgcagaatt cttcgcccag	4920
tacgtaccgc tctcttatcg tacgctgggtg tttatcctcg gcggcttctc gatgggtggtg	4980
tctaacctcg gcttgagcca gctgattcag atctctgtac cggtgctgac cgccatttat	5040
ccgccgtgta tcgcactggt tgtattaagt ttacacgct catgggtggca taattcgtcc	5100
cgcgtgattg ctccgccgat gtttatcagc ctgctttttg gtattctcga cgggatcaag	5160
gcatctgcat tcagcgatat cttaccgtcc tgggcgcagc gtttaccgct ggccgaacaa	5220
ggtctggcgt ggttaatgcc aacagtgggtg atgggtggttc tggccattat ctgggatcgt	5280
gcggcaggtc gtcaggtgac ctccagcgct cactaatacg catggcatgg atgaccgatg	5340
gtagtgtggg ctctcccat gcgagagtag ggaactgccca ggcat	5385

<210> 36

<211> 5384

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 36

gttgacactc tatcattgat agagttat	taccactccc tatcagtgat agagaaaagt	60
gaagtttacc cgcgggattg gaggtacaca	tgcagatctt cgacactttg caatcaatgg	120
gccatgagca ggttgtccta tgtagcgata	agaccacggg tctgcgcgcc attatcgcta	180
tacacgatac atccttaggg ccggcgcttg	gtggtacccg tatgtggcag tatgcaactg	240
acgacgatgc tattactgac gcactccgtc	tgtctcgggg catgacctac aaagctgcgg	300
tttctggcgt aaatctgggc ggtggtaaag	ccgttatcat cggaaccct cacagtgata	360
aaagcgaagc gctgtttcgc gcttacggca	gaatgggtgga atcccagcgt gggcgttaca	420
tcaccgccga agacgttggt actagcgtag	gtgatatgga gtggattcgc atggaaacca	480
aatatgtaac gggcggtgggt ggcaacggag	gctctggtga cccctctcca gttaccgctc	540
tgggtgttta ctcgggcatg aaggcatgcg	ctaaatcagt ctatggtact gatgcgctga	600
gcggtaaaag gatcgtgggt cagggcgcg	gtaacgttgc atcccatctg gttcacagtc	660
tggtaaaaga aggcgctgtg gttttcgtca	ctgacatcta cgaagaaaag gccaaagcat	720
tagcggctga aacgggcgct accgtgattc	gcaccgacga ggtttttact acacaatgcg	780
atatcttctc tccgaacgct ctgggggccc	tcctgaacga tgaaactatt ccgcagctca	840
catgcgctat cgtagctggt ggtgcaaaca	atcagcttaa aatcgaacaa cgtcacgcca	900
cggctctgca agagaaaggc attctgtatg	cgccggatta cgtaatcaac gccggggggcc	960
tcatgaatgt ggcgagcgaa gttgacggct	acaaccgtga aaaggttatg cgccaggctg	1020
aaggatttta cgatattact atgaacatcc	taaataccgc gcgtgagcgt aacatcctga	1080
ccatcgaagc atccaacgcg attgctgaag	agcggatcaa caaagttcgc catgttcacg	1140
ggaacttcat cggttccccg tctattcgcg	gagtataata aatcaaaaaca ccaaattaag	1200
gagaaatfff atgtcgacga caaccgttgg	tgactacttg ctgtatcgct taaacgaaat	1260
cggcattgag cacctcttcg gagtgccagg	tgattacaat ctgcaatttc tggatcatgt	1320
aatcgaccac cctcagctga cttgggtcgg	ctgcactaac gaacttaacg ctgcctacgc	1380
agctgatggt tatgcgcgtt gtcgtccggc	tgcggcactg ctgaccacct tcgggggttg	1440
cgaactgagc gctattaatg gcatcgcagg	ttcctacgcg gagtatctgc cggtaataca	1500
tatcgttggt gcaccgagtc tatcagccca	gcagcagggc gacctgattc accactctct	1560

tggcgaaggt gattttttcca gcttcctgag gatgtcccaa ccggtgtctg ttgcgaggc	1620
tgctctgact cctgataacg catgcaagga aatcgaccgc gtactggcgg aagtcctcat	1680
tcagcgtcgt cccggctacc tgctgctgtc taccgacgtg gctgctgcgc cggcggctct	1740
gccacaaagc actcttttctt tgccgaccgc cccggatcat cgcgcagttc tggctgcttt	1800
cagcgacgct gctgagcaga tgctggctca ggccaaaagc gtctctctac tggcggactt	1860
tctggctgat cgtttcggtg ttactcgagc actggccgcg tggcttcagc aggttccgct	1920
accgcacgcc actctgttaa tgggtaaagg cgttctgagt gaacagcaac cagggttcgt	1980
gggtacctac gctggtgcgg catctatcga ttcgacgcgt ggcgcaatcg aagaagctgg	2040
ggtaattatc ggagtgggag ttagattttc cgacactatc acagcaggct tctcgcagca	2100
gatcgacgcc cgccgtttta tagacattca acccttcttc tctcgtattg gcgatcgcca	2160
gtttgatcac ttgccgatgc aggctgccgt cgcagccctg catcaactgt gtcttcgtta	2220
tcagcagcag tggcttatca ccgctcctag cccgcctgca ctgccgccgg ctgctggtag	2280
cgagctgtcc cagaacgcat tctggcaggc gatgcagaac ttcattccgcc ctggggacct	2340
gttggtggcc gaccaaggta ctgcggcggt cggcgcagcg gcgctgcgct taccgcagaa	2400
ttgccagctg cttgtgcagc cgctgtgggg ctcaatcggg tacagtctgc cggccacctt	2460
tgggtgctcag acggcagata cagagcgctg tgtaatccta atcattggcg atggttcagc	2520
gcaattaact attcaggaac tttccagtat gatgcgtgac ggcttgaaac ctatcatctt	2580
tctcctgaac aacaacggtt acaccgttga acgggcgatt cacggcgcgg agcaacgtta	2640
taacgatatc gctgcttgga attggacca actgccccag gcgctgagtg ttcattgccc	2700
agcgagagc tggcgagtcg ttgaaacggt gcagctgacc gacgtaatga aagtcatcgc	2760
tgcttctccg cgtctgagct tggtagaagt tgttctgcct gcaatggatg tcccaccgct	2820
gctgcaagca gtgagtgccg ctctgaacca gcgcaactcc tcttaaattt ttcaacggcc	2880
tatcataagg aggaataatg acagcggagc agcaaaatgg ggtatccgac tcacgccgtt	2940
tcgaatttca ggaatttggt ggccctatcg cccacagac ctatcagctc cccgcaccgg	3000
ctagcgatga agttttgtta aaggatgaact actgcggtgt ctgtcacagt gatgttcac	3060
ttcacgacgg ctacttcgag ctgggtggcg ataaacgtct gaacttcgct atgccgctgc	3120
cgctgacgct gggtcacgaa gtaattggca ccgttggtggc tgtcggcgac caggttactg	3180

gtgtaaaacc	gggggaccag	cgactgatct	atccgtggat	aggttgcgga	aaatgcggcg	3240
cgtgtcaaaa	aggagaagaa	aacctgtgcg	ttactcctgc	acatctgggc	gtgaacaagc	3300
cgggcgggta	cgctgatcac	atcgttgtac	cccatttctg	ctaccttctg	gacatttcgg	3360
gtctgaaccc	gggtgatgcc	gctaccctcg	cgtgctccgg	cctgaccact	ttcagcgcga	3420
tcaacaaagt	gttgccgctt	gcagatgacc	agtggattgt	tgttatcggg	tgtgggtggcc	3480
tcggccagat	ggcgtgctg	atcctgcaag	ctatgggaat	tggcaatgtt	atcgggtattg	3540
acctgtctga	agagaaacgg	aaactggctc	atgaaagcgg	tgcacgtcac	tccttcgatc	3600
caaacactcc	gaagctgaac	cgcggtggtc	ccgaaacctg	cccgggtacg	gtacaggccg	3660
cgtagacttt	tgtgggcaat	gagcaaactg	ctcagctggc	actgtctctg	cttggaaaag	3720
gtggcaaata	tgttcctgtc	gggctgcacg	gcggcgagct	gcgttaccca	ttgccgatca	3780
tcacgaacaa	agctgtaagt	atcatcggtt	cttacgttgg	taccctgaaa	gaactggaag	3840
acttagttgc	tttcgccaa	gaaaagaatc	tgccgccaat	tcatattgaa	caccgcccgc	3900
tggaatcggc	ggctcaggcc	gtagaggacc	tggaaaaagg	acaggttgct	gggcgtgtta	3960
tcctggatgc	aggtaactaa	gaaggagata	tacatatgac	ccatcaatta	agatcgcgcg	4020
atatcatcgc	tctgggcttt	atgacatttg	cgttgttcgt	cggcgcaggt	aacattattt	4080
tccctccaat	ggtcggcttg	caggcaggcg	aacacgtctg	gactgcggca	ttcggcttcc	4140
tcattactgc	cgttggccta	ccggtattaa	cggtagtggc	gctggcaaaa	gttggcggcg	4200
gtgttgacag	tctcagcacg	ccaattggta	aagtcgctgg	cgtactgctg	gcaacagttt	4260
gttacctggc	ggtggggccg	ctttttgcta	cgccgcgtac	agctaccggt	tcttttgaag	4320
tgggcattgc	gccgctgacg	ggtgattccg	cgctgccgct	gtttattttac	agcctgggtct	4380
atctcgctat	cgttatttctg	gtttcgctct	atccgggcaa	gctgctggat	accgtgggca	4440
acttccttgc	gccgctgaaa	attatcgcg	tggtcatcct	gtctgttgcc	gcaattatct	4500
ggccggcggg	ttctatcagt	acggcgactg	aggcttatca	aaacgctgcg	ttttctaacg	4560
gcttcgtcaa	cggctatctg	accatggata	cgctgggcgc	aatgggtgtt	ggtatcggtta	4620
ttgttaacgc	ggcgcgttct	cgtggcggtta	ccgaagcgcg	tctgctgacc	cggtataaccg	4680
tctgggctgg	cctgatggcg	ggtgttggtc	tgactctgct	gtacctggcg	ctgttccgtc	4740
tgggttcaga	cagcgcgtcg	ctggtcgatc	agtctgcaaa	cgggtgcggcg	atcctgcatg	4800

cttacgttca gcataccttt ggcggcggcg gtagcttcct gctggcggcg ttaatcttca	4860
tcgcctgcct gggtcacggcg gttggcctga cctgtgcttg tgcagaattc ttcgcccagt	4920
acgtaccgct ctcttatcgt acgctgggtgt ttatcctcgg cggcttctcg atgggtgggtgt	4980
ctaacctcgg cttgagccag ctgattcaga tctctgtacc ggtgctgacc gccatttatc	5040
cgccgtgtat cgcactgggtt gtattaagtt ttacacgctc atgggtggcat aattcgtccc	5100
gcgtgattgc tccgccgatg tttatcagcc tgcttttttg tattctcgac gggatcaagg	5160
catctgcatt cagcgatata ttaccgtcct gggcgcagcg tttaccgctg gccgaacaag	5220
gtctggcgtg gttaatgcca acagtgggtga tgggtggttct ggccattatc tgggatcgtg	5280
cggcaggtcg tcaggtgacc tccagcgctc actaatacgc atggcatgga tgaccgatgg	5340
tagtgtgggc tctcccatg cgagagtagg gaactgccag gcat	5384

<210> 37

<211> 5350

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 37

gttgacactc tatcattgat agagttatth taccactccc tatcagtgat agagaaaagt	60
gaaatattca cgtaaggagg taatagatgc agatcttcga cactttgcaa tcaatgggcc	120
atgagcaggt tgtcctatgt agcgataaga ccacgggtct gcgcgccatt atcgctatac	180
acgatacatc cttagggccg gcgcttggtg gtacccgtat gtggcagtat gcaactgacg	240
acgatgctat tactgacgca ctccgtctgt ctcggggcat gacctacaaa gctgcggttt	300
ctggcgtaaa tctgggcggg ggtaaagccg ttatcatcgg aaaccctcac agtgataaaa	360
gcgaagcgct gtttcgcgct tacggcagaa tgggtggaatc ccagcgtggg cgttacatca	420
ccgccgaaga cgttgggtact agcgctacgtg atatggagtg gattcgcagt gaaaccaa	480
atgtaacggg cgtgggtggc aacggaggct ctggtgacct ctctccagtt accgctctgg	540
gtgtttactc gggcatgaag gcatgcgcta aatcagtcta tgggtactgat gcgctgagcg	600
gtaaaaggat cgtgggttcag ggcgcgggta acgttgcatc ccatctgggt cacagtctgg	660
taaaagaagg cgctgtggtt ttcgtcactg acatctacga agaaaaggcc aaagcattag	720

cggctgaaac gggcgctacc gtgattcgca cgcacgaggt ttttactaca caatgcgata	780
tcttctctcc gaacgctctg ggggccgtcc tgaacgatga aactattccg cagctcacat	840
gcgctatcgt agctgggtggg gcaaacaatc agcttaaaat cgaacaacgt cacgccacgg	900
ctctgcaaga gaaaggcatt ctgtatgcgc cggattacgt aatcaacgcc gggggcctca	960
tgaatgtggc gagcgaagtt gacggctaca accgtgaaaa ggttatgcgc caggctgaag	1020
gtatttacga tattactatg aacatcctaa ataccgcgcg tgagcgtaac atcctgacca	1080
tcgaagcatc caacgcgatt gctgaagagc ggatcaacaa agttcgccat gttcacggga	1140
acttcatcgg ttccccgtct attcgcggag tataataaat caaaacacca aattaaggag	1200
aaattttatg tcgacgacaa ccgttggtga ctacttgctg tatcgcttaa acgaaatcgg	1260
cattgagcac ctcttcggag tgccagggtga ttacaatctg caatttctgg atcatgtaat	1320
cgaccaccct cagctgactt gggtcggctg cactaacgaa cttaacgctg cctacgcagc	1380
tgatggttat gcgcgttgct gtccggctgc ggacttgctg accaccttcg gggttggcga	1440
actgagcgct attaatggca tcgcaggttc ctacgcggag tatctgccgg taatacatat	1500
cgttggtgca ccgagtctat cagcccagca gcagggcgac ctgattcacc actctcttgg	1560
cgaaggtgat ttttccagct tcctgaggat gtcccaaccg gtgtctgttg cgcaggctgc	1620
tctgactcct gataacgcat gcaaggaaat cgaccgcgta ctggcggaag tcctcattca	1680
gcgtcgtccc ggctacctgc tgctgtctac cgacgtggct gctgcgccgg cggctctgcc	1740
acaaagcact ctttctttgc cgaccgcccc ggatcatcgc gcagttcttg ctgctttcag	1800
cgacgctgct gagcagatgc tggctcaggc caaaagcgtc tctctactgg cggactttct	1860
ggctgatcgt ttcggtgtta ctcgagcact ggccgcgtgg cttcagcagg ttccgctacc	1920
gcacgccact ctgttaatgg gtaaaggcgt tctgagtga cagcaaccag ggttcgtggg	1980
tacctacgct ggtgcggcat ctatcgattc gacgcgtggc gcaatcgaag aagctggggg	2040
aattatcgga gtgggagtta gatcttccga cactatcaca gcaggcttct cgcagcagat	2100
cgacgcccgc cgttttatag acattcaacc cttcttctct cgtattggcg atcgccagtt	2160
tgatcacttg ccgatgcagg ctgccgtcgc agccctgcat caactgtgtc ttcgttatca	2220
gcagcagtgg tctatcaccg ctcctagccc gcctgcactg ccgccggctg ctggtagcga	2280
gctgtcccag aacgcattct ggcaggcgat gcagaacttc atccgccctg gggacctgtt	2340

ggtggccgac caaggtactg cggcgttcgg cgcagcggcg ctgcgcttac cgcagaattg	2400
ccagctgctt gtgcagccgc tgtggggctc aatcggttac agtctgccgg ccacctttgg	2460
tgctcagacg gcagatacag agcgtcgtgt aatcctaadc attggcgatg gttcagcgca	2520
attaactatt caggaacttt ccagtatgat gcgtgacggc ttgaaaccta tcatctttct	2580
cctgaacaac aacggttaca ccgttgaacg ggcgattcac ggcgcgaggc aacgttataa	2640
cgatatcgct gcttggaatt ggacccaact gccccaggcg ctgagtgttc attgcccagc	2700
gcagagctgg cgagtcgttg aaacggtgca gctgaccgac gtaatgaaag tcatcgctgc	2760
ttctccgcgt ctgagcttgg tagaagttgt tctgcctgca atggatgtcc caccgctgct	2820
gcaagcagtg agtgccgctc tgaaccagcg caactcctct taatccctcc ctacgggtta	2880
aggagcccct gctatgcgca gcatgcagtt tgatgagtac ggtgcacccc tgaaagcggt	2940
ctcatatgaa gacccgaccc cgcaagggaa ggaagtagtc gttaggatcg aagcctgtgg	3000
tgtgtgccac tctgatattc atcttcacga gggctacttc gacatgggcg gtggcaataa	3060
agctgatgtt actcgtgctc gcgaactccc ttttacattg ggtcatgaaa tcgttggcga	3120
agtggttagca actggaccag gtgtcaccgg cgctaaaccg ggcgacaaac gtattgtgta	3180
cccgtggatc ggggtgcggcg actgcccga atgcaacagt ggtgaggatc agtcctgtgc	3240
gcgtccacgt aacctgggtg ttcacgttga cgggtggctat tcgacgcacg taaagatacc	3300
ggacgaaaaa ttcctgttcg cctacgatgg tattcctact gagttagcgg gaacctatgc	3360
ttgcagcggc atcaccgctt atggtgcact gatgaaagca aaggaagcgg ctgaaagatc	3420
tggctacatc ggtctgattg gcgctggtgg cgttggcatg gctgggtctga tgctggccaa	3480
agcagcgatc ggggctaana ctgtagtctt tgatatcgac gacgcaaaac tggaagctgc	3540
gacccgtgcc ggggaggatt acgtgttcaa ctccggtgca aaagaaacac gcaaggaagt	3600
tatgaaacta acgaatggtg gcctgtctgg tgctgttgat ttcgttggca gcgataaaag	3660
cgctctgttt ggaatcaacg ccttgggtca gaacggcgtg ctggtcataa ttggactgtt	3720
cgggtggcgct atgactgttc cgggtaccct gttcccgctg aaagggatca ccgtacgtgg	3780
ctcatacgta ggttccctgc aagagatgag tgatatgatg gagttagtgc gcgctgggaa	3840
agttcctccg atgccggtaa aaactcggcc actggacgct gcctgggaaa cccttgagga	3900
tctacgccat ggtaaaatcg tgggccgtgt tgttctgacc ccataagaag gagatataca	3960

tatgacccat caattaagat cgcgcgatat catcgctctg ggctttatga catttgcgtt	4020
gttcgtcggc gcaggtaaca ttattttccc tccaatggtc ggcttgcagg caggcgaaca	4080
cgtctggact gcggcattcg gcttcctcat tactgccgtt ggccctaccgg tattaacggt	4140
agtggcgctg gcaaaagttg gcggcgggtg tgacagtctc agcacgcaa ttggtaaagt	4200
cgctggcgta ctgctggcaa cagtttggtta cctggcggtg gggccgcttt ttgctacgcc	4260
gcgtacagct accgtttctt ttgaagtggg cattgcgccg ctgacgggtg attccgcgct	4320
gccgctgttt atttacagcc tgggtctatct cgctatcgtt attctgggtt cgctctatcc	4380
gggcaagctg ctggataccg tgggcaactt ccttgcgccg ctgaaaatta tcgcgctggt	4440
catcctgtct gttgccgcaa ttatctggcc ggcggttct atcagtacgg cgactgaggc	4500
ttatcaaaac gctgcgtttt ctaacggctt cgtcaacggc tatctgacca tggatacgct	4560
gggcgcaatg gtgtttggta tcgttattgt taacgcggcg cgttctcgtg gcgttaccga	4620
agcgcgtctg ctgaccggtt ataccgtctg ggctggcctg atggcgggtg ttggtctgac	4680
tctgctgtac ctggcgctgt tccgtctggg ttcagacagc gcgtcgctgg tcgatcagtc	4740
tgcaaacggt gcggcgatcc tgcatgctta cgttcagcat acctttggcg gcggcggtag	4800
cttcctgctg gcggcggtta tcttcatcgc ctgcctggtc acggcggttg gcctgacctg	4860
tgcttggtgca gaattcttcg cccagtacgt accgctctct tatcgtacgc tgggtgttat	4920
cctcggcggc ttctcgatgg tgggtgtctaa cctcggcttg agccagctga ttcagatctc	4980
tgtaccggtg ctgaccgcca tttatccgcc gtgtatcgca ctggttgat taagttttac	5040
acgctcatgg tggcataatt cgtcccgcgt gattgctccg ccgatgttta tcagcctgct	5100
ttttggtatt ctcgacggga tcaaggcatc tgcattcagc gatatcttac cgtcctgggc	5160
gcagcgttta ccgctggccg aacaaggctt ggcgtgggtta atgccaacag tggatgatgt	5220
ggttctggcc attatctggg atcgtgcggc aggtcgtcag gtgacctcca gcgctcacta	5280
atacgcatgg catggatgac cgatggtagt gtgggctctc cccatgcgag agtagggaac	5340
tgccaggcat	5350

<210> 38

<211> 5381

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 38

gttgacactc tatcattgat agagttatTTt taccactccc tatcagtgat agagaaaagt	60
gaaatattca cgtaaggagg taatagatgc agatcttcga cactttgcaa tcaatgggcc	120
atgagcaggt tgtcctatgt agcgataaga ccacgggtct gcgcgccatt atcgctatac	180
acgatacatc ctaggggccg gcgcttggtg gtaccggtat gtggcagtat gcaactgacg	240
acgatgctat tactgacgca ctccgtctgt ctcggggcat gacctacaaa gctgcggttt	300
ctggcgtaaa tctgggcggt ggtaaagccg ttatcatcgg aaaccctcac agtgataaaa	360
gcgaagcgct gtttcgcgct tacggcagaa tgggtggaatc ccagcgtggg cgttacatca	420
ccgccgaaga cgttggtact agcgtacgtg atatggagtg gattcgcatg gaaaccaaT	480
atgtaacggg cgtgggtggc aacggaggct ctggtgacct ctctccagtt accgctctgg	540
gtgtttactc gggcatgaag gcatgcgcta aatcagtcta tggTactgat gcgctgagcg	600
gtaaaaggat cgtggttcag ggcgcgggta acgttgcatc ccatctggtt cacagtctgg	660
taaaagaagg cgctgtggtt ttcgtcactg acatctacga agaaaaggcc aaagcattag	720
cggctgaaac gggcgctacc gtgattcgca ccgacgaggt ttttactaca caatgcgata	780
tcttctctcc gaacgctctg ggggccgtcc tgaacgatga aactattccg cagctcacat	840
gcgctatcgt agctgggtggT gcaaacaatc agcttaaaat cgaacaacgt cacgccacgg	900
ctctgcaaga gaaaggcatt ctgtatgcgc cggattacgt aatcaacgcc gggggcctca	960
tgaatgtggc gagcgaagtt gacggctaca accgtgaaaa ggTtatgcgc caggctgaag	1020
gtatttacga tattactatg aacatcctaa ataccgcgcg tgagcgtaac atcctgacca	1080
tcgaagcatc caacgcgatt gctgaagagc ggatcaacaa agttcgccat gttcacggga	1140
acttcatcgg ttccccgtct attcgcggag tataataaat caaaacacca aattaaggag	1200
aaatTTtatg tcgacgacaa ccgttggtga ctacttgctg tatcgcttaa acgaaatcgg	1260
cattgagcac ctcttcggag tgccagggtga ttacaatctg caatttctgg atcatgtaat	1320
cgaccaccct cagctgactt gggtcggctg cactaacgaa cttaacgctg cctacgcagc	1380
tgatggttat gcgcgttgTc gtccggctgc ggcactgctg accaccttcg gggttggcga	1440

actgagcgcct	attaatggca	tcgcaggttc	ctacgcggag	tatctgccgg	taatacatat	1500
cgttggtgca	ccgagtctat	cagcccagca	gcagggcgac	ctgattcacc	actctcttgg	1560
cgaaggtgat	ttttccagct	tcctgaggat	gtcccaaccg	gtgtctgttg	cgcaggctgc	1620
tctgactcct	gataacgcat	gcaaggaaat	cgaccgcgta	ctggcggaag	tcctcattca	1680
gcgtcgtccc	ggctacctgc	tgctgtctac	cgacgtggct	gctgcgccgg	cggtcttgcc	1740
acaaagcact	ctttctttgc	cgaccgcccc	ggatcatcgc	gcagttcttg	ctgctttcag	1800
cgacgctgct	gagcagatgc	tggctcaggc	caaaagcgtc	tctctactgg	cggactttct	1860
ggctgatcgt	ttcggtgтта	ctcgagcact	ggccgcgtgg	cttcagcagg	ttccgctacc	1920
gcacgccact	ctgttaatgg	gtaaaggcgt	tctgagtga	cagcaaccag	ggttcgtggg	1980
tacctacgct	ggtgcggcat	ctatcgattc	gacgcgtggc	gcaatcgaag	aagctgggggt	2040
aattatcgga	gtgggagtta	gattttccga	cactatcaca	gcaggcttct	cgagcagat	2100
cgacgcccgc	cgttttatag	acattcaacc	cttcttctct	cgtattggcg	atcgccagtt	2160
tgatcacttg	ccgatgcagg	ctgccgtcgc	agccctgcat	caactgtgtc	ttcgttatca	2220
gcagcagtgg	tctatcaccg	ctcctagccc	gcctgcactg	ccgccggctg	ctggtagcga	2280
gctgtcccag	aacgcattct	ggcaggcgat	gcagaacttc	atccgccctg	gggacctgtt	2340
ggtggccgac	caaggtactg	cggcgttcgg	cgagcgggcg	ctgcgcttac	cgcagaattg	2400
ccagctgctt	gtgcagccgc	tgtggggctc	aatcggttac	agtctgccgg	ccacctttgg	2460
tgctcagacg	gcagatacag	agcgtcgtgt	aatcctaatc	attggcgatg	gttcagcgca	2520
attaactatt	caggaacttt	ccagtatgat	gcgtgacggc	ttgaaaccta	tcatttttct	2580
cctgaacaac	aacggttaca	ccgttgaacg	ggcgattcac	ggcgcgagc	aacgttataa	2640
cgatatcgct	gcttggaatt	ggacccaact	gccccaggcg	ctgagtgttc	attgcccagc	2700
gcagagctgg	cgagtcgttg	aaacggtgca	gctgaccgac	gtaatgaaag	tcatcgctgc	2760
ttctccgcgt	ctgagcttgg	tagaagttgt	tctgcctgca	atggatgtcc	caccgctgct	2820
gcaagcagtg	agtgccgctc	tgaaccagcg	caactcctct	taaatttttc	aacggcctat	2880
cataaggagg	aataatgaca	gcggagcagc	aaaatgggggt	atccgactca	cgccgtttcg	2940
aatttcagga	atttggtggc	cctatcgccc	cacagaccta	tcagctcccc	gcaccggcta	3000
gcgatgaagt	tttgttaaag	gtgaactact	gcggtgtctg	tcacagtgat	gttcattctt	3060

acgacggcta	cttcgagctg	ggtggcgata	aacgtctgaa	cttcgctatg	ccgctgccgc	3120
tgacgctggg	tcacgaagta	attggcaccg	ttgtggctgt	cggcgaccag	gttactgggtg	3180
taaaaccggg	ggaccagcga	ctgatctatc	cgtggatagg	ttgcggaaaa	tgcggcgcgt	3240
gtcaaaaagg	agaagaaaac	ctgtgcgtta	ctcctgcaca	tctgggcgtg	aacaagccgg	3300
gcggttacgc	tgatcacatc	gttgtacccc	attctcgcta	ccttctggac	atttcgggtc	3360
tgaacccggg	tgatgccgct	accctcgcgt	gctccggcct	gaccactttc	agcgcgatca	3420
acaaagtgtt	gccgcttgca	gatgaccagt	ggattgttgt	tatcggttgt	ggtggcctcg	3480
gccagatggc	gctgcgtatc	ctgcaagcta	tgggaattgg	caatgttatc	ggtattgacc	3540
tgtctgaaga	gaaacggaaa	ctggctcatg	aaagcgggtg	acgtcactcc	ttcgatccaa	3600
acactccgaa	gctgaaccgc	gtggtcgccg	aaacctgccc	gggtacggta	caggccgcgt	3660
tagactttgt	gggcaatgag	caaactgctc	agctggcact	gtctctgctt	ggaaaagggtg	3720
gcaaatatgt	tcctgtcggg	ctgcacggcg	gcgagctgcg	ttaccattg	ccgatcatca	3780
cgaacaaagc	tgtaagtatc	atcggttctt	acgttggtac	cctgaaagaa	ctggaagact	3840
tagttgcttt	cgccaaggaa	aagaatctgc	cgccaattca	tattgaacac	cgcccgctgg	3900
aatcggcggc	tcaggccgta	gaggacctgg	aaaaaggaca	ggttgctggg	cgtgttatcc	3960
tggaatgcagg	taactaagaa	ggagatatatac	atatgaccca	tcaattaaga	tcgcgcgata	4020
tcatacgtct	gggctttatg	acatttgcgt	tgttcgtcgg	cgcaggtaac	attattttcc	4080
ctccaatggg	cggcttgtag	gcaggcgaac	acgtctggac	tgcggcattc	ggcttcctca	4140
ttactgccgt	tggcctaccg	gtattaacgg	tagtggcgct	ggcaaaaagt	ggcggcgggtg	4200
ttgacagtct	cagcacgcca	attggtaaag	tcgctggcgt	actgctggca	acagtttggt	4260
acctggcggg	ggggccgctt	tttgctacgc	cgcgtacagc	taccgtttct	tttgaagtgg	4320
gcattgcgcc	gctgacgggt	gattccgcgc	tgccgctgtt	tatttacagc	ctggtctatt	4380
tcgctatcgt	tattctgggt	tcgctctatc	cgggcaagct	gctggatacc	gtgggcaact	4440
tccttgcgcc	gctgaaaatt	atcgcgctgg	tcatacctgtc	tgttgccgca	attatctggc	4500
cggcggggtt	tatcagtacg	gcgactgagg	cttatcaaaa	cgctgcgttt	tctaacggct	4560
tcgtcaacgg	ctatctgacc	atggatacgc	tgggcgcaat	ggtgtttggt	atcgttattg	4620
ttaacgcggc	gcgttctcgt	ggcgttaccg	aagcgcgtct	gctgaccctg	tataccgtct	4680

gggctggcct gatggcgggt gttggtctga ctctgctgta cctggcgctg ttccgtctgg	4740
gttcagacag cgcgtcgctg gtcgatcagt ctgcaaacgg tgcggcgatc ctgcatgctt	4800
acgttcagca tacctttggc ggcggcggta gcttcctgct ggcggcgtta atcttcatcg	4860
cctgcctgggt cacggcggtt ggccctgacct gtgcttggtc agaattcttc gcccagtagc	4920
taccgctctc ttatcgtacg ctggtgttta tcctcggcgg cttctcgatg gtggtgtcta	4980
acctcggctt gagccagctg attcagatct ctgtaccggt gctgaccgcc atttatccgc	5040
cgtgtatcgc actggttgta ttaagtttta cacgctcatg gtggcataat tcgtcccgcg	5100
tgattgctcc gccgatgttt atcagcctgc tttttggtat tctcgacggg atcaaggcat	5160
ctgcattcag cgatatctta ccgtcctggg cgcagcgttt accgctggcc gaacaaggctc	5220
tggcgtgggt aatgccaaca gtggtgatgg tggttctggc cattatctgg gatcgtgcgg	5280
caggtcgtca ggtgacctcc agcgctcact aatacgcatg gcatggatga ccgatggtag	5340
tgtgggctct ccccatgcga gagtagggaa ctgccaggca t	5381

<210> 39

<211> 5286

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 39

gttgacactc tatcattgat agagttatth taccactccc tatcagtgat agagaaaagt	60
gaaatattca cgtaaggagg taatagatgc agatcttcga cactttgcaa tcaatgggcc	120
atgagcaggt tgtcctatgt agcgataaga ccacgggtct gcgcgccatt atcgctatac	180
acgatacatc cttagggccg gcgcttggtg gtacccgtat gtggcagtat gcaactgacg	240
acgatgctat tactgacgca ctccgtctgt ctcggggcat gacctacaaa gctgcggttt	300
ctggcgtaaa tctgggcgggt ggtaaagccg ttatcatcgg aaaccctcac agtgataaaa	360
gcgaagcgct gtttcgcgct tacggcagaa tgggtggaatc ccagcgtggg cgttacatca	420
ccgccgaaga cgttggtact agcgtagctg atatggagtg gattcgcatg gaaaccaa	480
atgtaacggg cgtgggtggc aacggaggct ctggtgacct ctctccagtt accgctctgg	540
gtgtttactc gggcatgaag gcatgcgcta aatcagtcta tggtagtgat gcgctgagcg	600

gtaaaaggat	cgtaggttcag	ggcgcgggta	acgttgcata	ccatctggtt	cacagtctgg	660
taaaagaagg	cgctgtgggt	ttcgtcactg	acatctacga	agaaaaggcc	aaagcattag	720
cggctgaaac	gggcgctacc	gtgattcgca	ccgacgaggt	ttttactaca	caatgcgata	780
tcttctctcc	gaacgctctg	ggggccgtcc	tgaacgatga	aactattccg	cagctcacat	840
gcgctatcgt	agctgggtgt	gcaaacaatc	agcttaaaat	cgaacaacgt	cacgccacgg	900
ctctgcaaga	gaaaggcatt	ctgtatgcgc	cggattacgt	aatcaacgcc	gggggcctca	960
tgaatgtggc	gagcgaagtt	gacggctaca	accgtgaaaa	ggttatgcgc	caggctgaag	1020
gtatttacga	tattactatg	aacatcctaa	ataccgcgcg	tgagcgtaac	atcctgacca	1080
tcgaagcatc	caacgcgatt	gctgaagagc	ggatcaacaa	agttcgccat	gttcacggga	1140
acttcatcgg	ttccccgtct	attcgcggag	tataataaat	caaaacacca	aattaaggag	1200
aaattttatg	tcgacgacaa	ccgttggtga	ctacttgctg	tatcgcttaa	acgaaatcgg	1260
cattgagcac	ctcttcggag	tgccagggtga	ttacaatctg	caatttctgg	atcatgtaat	1320
cgaccaccct	cagctgactt	gggtcggctg	cactaacgaa	cttaacgctg	cctacgcagc	1380
tgatggttat	gcgcgttgct	gtccggctgc	ggcactgctg	accaccttcg	gggttggcga	1440
actgagcgct	attaatggca	tcgcaggttc	ctacgcggag	tatctgccgg	taatacatat	1500
cgttggtgca	ccgagtctat	cagcccagca	gcagggcgac	ctgattcacc	actctcttgg	1560
cgaaggatgat	ttttccagct	tcctgaggat	gtcccaaccg	gtgtctgttg	cgcaggctgc	1620
tctgactcct	gataacgcat	gcaaggaaat	cgaccgcgta	ctggcggaag	tcctcattca	1680
gcgtcgtccc	ggctacctgc	tgctgtctac	cgacgtggct	gctgcgccgg	cggctctgcc	1740
acaaagcact	ctttctttgc	cgaccgcccc	ggatcatcgc	gcagttcttg	ctgctttcag	1800
cgacgctgct	gagcagatgc	tggtcaggc	caaaagcgct	tctctactgg	cggactttct	1860
ggctgatcgt	ttcgggtgta	ctcgagcact	ggccgcgtgg	cttcagcagg	ttccgctacc	1920
gcacgccact	ctgttaatgg	gtaaaggcgt	tctgagtga	cagcaaccag	ggttcgtggg	1980
tacctacgct	ggtgcggcat	ctatcgattc	gacgcgtggc	gcaatcgaag	aagctgggggt	2040
aattatcgga	gtgggagtta	gattttccga	cactatcaca	gcaggcttct	cgcagcagat	2100
cgacgcccgc	cgttttatag	acattcaacc	cttcttctct	cgtattggcg	atcgccagtt	2160
tgatcacttg	ccgatgcagg	ctgccgtcgc	agccctgcat	caactgtgtc	ttcgttatca	2220

gcagcagtgg tctatcacccg ctctagccc gcctgcactg ccgccggctg ctggtagcga	2280
gctgtcccag aacgcattct ggcaggcgat gcagaacttc atccgccctg gggacctgtt	2340
ggtggccgac caaggtactg cggcgttcgg cgcagcggcg ctgcgcttac cgcagaattg	2400
ccagctgctt gtgcagccgc tgtggggctc aatcggttac agtctgccgg ccacctttgg	2460
tgctcagacg gcagatacag agcgtcgtgt aatcctaatac attggcgatg gttcagcgca	2520
attaactatt caggaacttt ccagtatgat gcgtgacggc ttgaaacctt tcatctttct	2580
cctgaacaac aacggttaca ccgttgaacg ggcgattcac ggcgcggagc aacgttataa	2640
cgatatcgct gcttgggaatt ggacccaact gccccaggcg ctgagtgttc attgcccagc	2700
gcagagctgg cgagtcgttg aaacggtgca gctgaccgac gtaatgaaag tcatcgctgc	2760
ttctccgcgt ctgagcttgg tagaagttgt tctgcctgca atggatgtcc caccgctgct	2820
gcaagcagtg agtgccgctc tgaaccagcg caactcctct taattcttaa taataaggag	2880
ctaattatgc agacggcggtt cttgtataag ccaggtcacg aaaacttagt gcgctcggag	2940
atcccgatac ctaaagctgg gcgtggcgaa gtcgttcttg aaattaaagc cgctggcatg	3000
tgccattccg atctgcacgt tctcgacggt ggaatcccc tgccgggtca atttgtaatg	3060
ggccatgaaa tcgttggtag tattcacgag atcggccagg acgtgaccgg tttcaaacag	3120
ggcgatctgt acgcagtcca cggcccgaat ccgtgtggta tttgcaccct gtgcagagaa	3180
ggatttgata acgactgcac tacagtggcg aaaaccggtc aatgggttcgg actgggtctt	3240
gacggcggct accagaagta tatccgtatc ccgaacgtaa ggtctatcgt taaagttcca	3300
gaagggtgtt cagctgaggc agctgcgagc tgtactgatg cagtactgac cccgtaccgt	3360
gcactaaaac aggctggcgc cagcaactct actcgggtac tgattctggg tctgggtggc	3420
ttaggtctga atgcccttaa actggctaag accttcggca gttacgttta cgcacttgac	3480
ctgaaacctt ctgcgcgtga agctgctaag gccgctgggg cggaatgaagt gctggagtcc	3540
ctgcccgaag acccgctggg tgttgatatc gtgttagacg tcgttggcgt gcagagcacc	3600
ttcaacctcg ctcaaaaaca cgttggcccg cgtggcatca ttgtacctgt aggcctggca	3660
tccccacagc tttcgtttaa cctaacggat ctggcgctcc gcgaaattcg tgttcagggc	3720
actttttggg gcacgagcaa tgagctggct gaatgtctgc gcctgtgcca gctgggcctg	3780
atcaaccgca aatatactgt ggtgcctctt gaagaagcgc cgaaatatat ggaagcaatg	3840

gctcatggga aagtagaagg tcgtatcggt ttccacccgt aagaaggaga tatacatatg	3900
acccatcaat taagatcgcg cgatatcatc gctctgggct ttatgacatt tgcgttggtc	3960
gtcggcgcag gtaacattat tttccctcca atggtcggct tgcaggcagg cgaacacgtc	4020
tggactgcgg cattcggctt cctcattact gccgttggcc taccgggtatt aacggtagtg	4080
gcgctggcaa aagttggcgg cgggtgtgac agtctcagca cgccaattgg taaagtcgct	4140
ggcgtactgc tggcaacagt ttgttacctg gcggtggggc cgctttttgc tacgccgcgt	4200
acagctaccg tttcttttga agtgggcatt gcgccgctga cgggtgattc cgcgctgccg	4260
ctgtttatct acagcctggc ctatttcgct atcgttattc tggtttcgct ctatccgggc	4320
aagctgctgg ataccgtggg caacttcctt gcgccgctga aaattatcgc gctgggtcatc	4380
ctgtctgttg ccgcaattat ctggccggcg ggttctatca gtacggcgac tgaggcttat	4440
caaaacgctg cgttttctaa cggcttcgtc aacggctatc tgacccatgga tacgctgggc	4500
gcaatggtgt ttggtatcgt tattgttaac gcggcgcggt ctcgtggcgt taccgaagcg	4560
cgtctgctga cccgttatac cgtctgggct ggcctgatgg cgggtgttgg tctgactctg	4620
ctgtacctgg cgctgttccg tctgggttca gacagcgctg cgctggtcga tcagtctgca	4680
aacggtgccg cgatcctgca tgcttacgtt cagcatacct ttggcggcgg cggtagcttc	4740
ctgctggcgg cgttaatctt catcgcctgc ctggtcacgg cggttggcct gacctgtgct	4800
tgtgcagaat tcttcgcccc gtacgtaccg ctctcttatac gtacgtgggt gtttatcctc	4860
ggcgggttct cgatgggtgg gtctaacctc ggcttgagcc agctgattca gatctctgta	4920
ccggtgctga ccgccattta tccgccgtgt atcgactgg ttgtattaag ttttacacgc	4980
tcatggtggc ataattcgtc ccgctgatt gctccgccga tgtttatcag cctgcttttt	5040
ggtattctcg acgggatcaa ggcatctgca ttcagcgata tcttaccgtc ctgggcgcag	5100
cgtttaccgc tggccgaaca aggtctggcg tggttaatgc caacagtggg gatgggtggt	5160
ctggccatta tctgggatcg tgcggcaggt cgtcagggtga cctccagcgc tcactaatac	5220
gcatggcatg gatgaccgat ggtagtgtgg gctctcccca tgcgagagta gggaactgcc	5280
aggcat	5286

<210> 40

<211> 5364

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 40

gttgacactc tatcattgat agagttatTTt taccactccc tatcagtgat agagaaaagt	60
gagataaacg aatcacaaag gcgaggcaag atgaacatct tcaagaaaat ggaggaattt	120
aattatgaac aactgggtcta cttctacgac agcgaaacgg aactcaaagg tattacctgt	180
atacacaaca caacttttagg gccggcattg ggcggtaccc gcctttggaa ctataactct	240
gaggaagatg ccgttgaaga cgtaatccgt ctggctcggg gcatgactta caaagcggct	300
tgcgccggtc tgaatctggg cggcggtaaa accgtgctga tcggtgatgc taaaaagatt	360
aaatcagagt cctacttccg tggactgggg cgctacgttc agtcgctgaa cggcagatat	420
atcaccgcgg aagacgtaaa tacttctacg aaggatatgg catacgttgc tatggaaact	480
gactatgtgg taggcctggg aggtaaatcc ggcaacccta gtccagttac tgcttacggt	540
gcatttatgg gtatcaaagc ggcgctgatg aaaaaatttg aggatagctc tattgaaggc	600
cgaaccttcg cagtgcaggg tgctgggcag acgggttact atcttatcga ttacctccta	660
ggcaacaaca agttcaaaga aaaggctaaa aagatttact tcaccgaaat taacgagagc	720
tatatcgagc gtatgaacaa agaacatccg gaagttgaat ttatttcccc ggacaaaatc	780
tactcgctgg aagtagacgt cttcgtgccc tgcgccctgg gcaaaatcgt taatgacaaa	840
actatcgatg aatttaagtg tccgatcatc gcaggctactg caaacaacgt actggaaagg	900
gaagcgcacg gcaacatgct taaagaacgt ggcattcttt acgccccgga ctatgtgatc	960
aatgctggtg ggctgatcaa cgtttaccac gagctgaacg gttacaataa agagaacgct	1020
attctggaag tggaattaat ttatgatcgc ctactggaaa tattcaacat cgctgattct	1080
ctgaacatca gcaccaatat cgctgccaac gagttcgcgg aaaaacgtat caagcaaatt	1140
aagtccttga aaaacaactt cattaaacgc taagaccatt aataagggaac attttatgtc	1200
gacgacaacc gttggtgact acttgctgta tcgcttaaac gaaatcggca ttgagcacct	1260
cttcggagtg ccaggtgatt acaatctgca atttctggat catgtaatcg accaccctca	1320
gctgacttgg gtcggctgca ctaacgaact taacgctgcc tacgcagctg atggttatgc	1380

gcgttgtcgt ccggctgcgg cactgctgac caccttcggg gttggcgaac tgagcgctat	1440
taatggcatc gcaggttcct acgcgagta tctgccggta atacatatcg ttggtgcacc	1500
gagtctatca gcccagcagc agggcgacct gattcaccac tctcttggcg aaggtgattt	1560
ttccagcttc ctgaggatgt cccaaccggt gtctgttgcg caggctgctc tgactcctga	1620
taacgcatgc aaggaaatcg accgcgtact ggcggaagtc ctcatcagc gtcgtcccgg	1680
ctacctgctg ctgtctaccg acgtggctgc tgcgccggcg gctctgccac aaagcactct	1740
ttctttgccg accgccccgg atcatcgcgc agttctggct gctttcagcg acgtgctga	1800
gcagatgctg gctcaggcca aaagcgtctc tctactggcg gactttctgg ctgatcgttt	1860
cgggtgttact cgagcactgg ccgcgtggct tcagcagggt ccgctaccgc acgccactct	1920
gttaatgggt aaaggcgttc tgagtgaaca gcaaccagggt ttcgtgggta cctacgctgg	1980
tgcggcatct atcgattcga cgcgtggcgc aatcgaagaa gctggggtaa ttatcggagt	2040
gggagttaga ttttccgaca ctatcacagc aggcttctcg cagcagatcg acgcccgcg	2100
ttttatagac attcaaccct tcttctctcg tattggcgat cgccagtttg atcacttgcc	2160
gatgcaggct gccgtcgcag ccctgcatca actgtgtctt cgttatcagc agcagtggtc	2220
tatcaccgct cctagccgc ctgcactgcc gccggctgct ggtagcgagc tgtcccagaa	2280
cgcattctgg caggcgatgc agaacttcat ccgccctggg gacctgttgg tggccgacca	2340
aggtactgcg gcgttcggcg cagcggcgct gcgcttaccg cagaattgcc agctgcttgt	2400
gcagccgctg tggggctcaa tcggttacag tctgccggcc acctttggtg ctcagacggc	2460
agatacagag cgtcgtgtaa tcctaatacat tggcgatggg tcagcgcaat taactattca	2520
ggaactttcc agtatgatgc gtgacggctt gaaacctatc atctttctcc tgaacaacaa	2580
cggttacacc gttgaacggg cgattcacgg cgcggagcaa cgttataacg atatcgctgc	2640
ttggaattgg acccaactgc ccaggcgct gagtgttcat tgcccagcgc agagctggcg	2700
agtcgttgaa acggtgcagc tgaccgacgt aatgaaagtc atcgctgctt ctccgcgtct	2760
gagcttggtg gaagttgttc tgcctgcaat ggatgtccca ccgtgctgc aagcagtga	2820
tgccgctctg aaccagcgca actcctctta attggataaa gaataaggag gtattttatg	2880
acagcggagc agcaaaatgg ggtatccgac tcacgccgtt tcgaatttca ggaatttgg	2940
ggccctatcg cccacagac ctatcagctc ccgcaccgg ctagcgatga agttttgtta	3000

aaggtgaact actgcggtgt ctgtcacagt gatgttcac ttcacgacgg ctacttcgag	3060
ctgggtggcg ataaacgtct gaacttcgct atgccgctgc cgctgacgct gggtcacgaa	3120
gtaattggca ccgttgtggc tgtcggcgac caggttactg gtgtaaaacc gggggaccag	3180
cgactgatct atccgtggat aggttgcgga aaatgcggcg cgtgtcaaaa aggagaagaa	3240
aacctgtgcg ttactcctgc acatctgggc gtgaacaagc cgggcggtta cgctgatcac	3300
atcgttgtag cccattctcg ctaccttctg gacatttcgg gtctgaaccc gggtgatgcc	3360
gctaccctcg cgtgctccgg cctgaccact ttcagcgcg tcaacaaagt gttgccgctt	3420
gcagatgacc agtggattgt tgttatcggt tgtgggtggcc tcggccagat ggcgctgcgt	3480
atcctgcaag ctatgggaat tggcaatgtt atcgggtattg acctgtctga agagaaacgg	3540
aaactggctc atgaaagcgg tgcacgtcac tccttcgac caaacactcc gaagctgaac	3600
cgcgtggctc ccgaaacctg cccgggtacg gtacaggccg cgtagactt tgtgggcaat	3660
gagcaaacctg ctcagctggc actgtctctg cttggaaaag gtggcaaata tgttcctgtc	3720
gggctgcacg gcggcgagct gcgttaccca ttgccgatca tcacgaacaa agctgtaagt	3780
atcatcgggt cttacgttgg taccctgaaa gaactggaag acttagttgc tttcgccaag	3840
gaaaagaatc tgccgccaat tcatattgaa caccgcccgc tggaatcggc ggctcaggcc	3900
gtagaggacc tggaaaaagg acaggttgct gggcgtgtta tcctggatgc aggtaaactaa	3960
gaaggagata tacatatgac ccatcaatta agatcgcgcg atatcatcgc tctgggcttt	4020
atgacatttg cgttgttcgt cggcgcaggt aacattatth tccctccaat ggtcggcttg	4080
caggcaggcg aacacgtctg gactgcggca ttcggcttcc tcattactgc cgttggccta	4140
ccggtattaa cggtagtggc gctggcaaaa gttggcggcg gtgttgacag tctcagcacg	4200
ccaattggta aagtcgctgg cgtactgctg gcaacagttt gttacctggc ggtggggccg	4260
ctttttgcta cgccgcgtac agctaccgtt tcttttgaag tgggcattgc gccgctgacg	4320
ggtgattccg cgctgccgct gtttattttac agcctggtct atttcgctat cgttattctg	4380
gtttcgctct atccgggcaa gctgctggat accgtgggca acttccttgc gccgctgaaa	4440
attatcgcgc tggatcatcct gtctgttgcc gcaattatct ggccggcggg ttctatcagt	4500
acggcgactg aggcttatca aaacgtgcg ttttctaacg gcttcgtcaa cggctatctg	4560
accatggata cgctgggcgc aatggtgttt ggtatcgtaa ttgttaacgc ggcgcttct	4620

cgtggcggtta ccgaagcgcg tctgctgacc cgttataccg tctgggctgg cctgatggcg	4680
ggtgttggtc tgactctgct gtacctggcg ctgttccgtc tgggttcaga cagcgcgctc	4740
ctggctgatc agtctgcaaa cgggtcggcg atcctgcatg cttacgttca gcataccttt	4800
ggcggcggcg gtagcttcct gctggcggcg ttaatcttca tcgcctgcct ggtcacggcg	4860
gttggcctga cctgtgcttg tgcagaattc ttgcccagt acgtaccgct ctcttatcgt	4920
acgctggtgt ttatcctcgg cggcttctcg atgggtggtg ctaacctcgg cttgagccag	4980
ctgattcaga tctctgtacc ggtgctgacc gccatttata cgccgtgtat cgcactgggt	5040
gtattaagtt ttacacgctc atgggtggcat aattcgtccc gcgtgattgc tccgccgatg	5100
tttatcagcc tgcttttttg tattctcgac gggatcaagg catctgcatt cagcgatatc	5160
ttaccgtcct gggcgagcg tttaccgctg gccgaacaag gtctggcgtg gttaatgcca	5220
acagtgggtga tgggtggttct ggccattatc tgggatcgtg cggcaggctc tcaggtgacc	5280
tccagcgctc actaatacgc atggcatgga tgaccgatgg tagtgtgggc tctcccatg	5340
cgagagtagg gaactgccag gcat	5364

<210> 41

<211> 5346

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 41

gttgacactc tatcattgat agagttatth taccactccc tatcagtgat agagaaaagt	60
gagagctaga taataaggac ggcacaagat gaacatcttc aagaaaatgg aggaatttaa	120
ttatgaacaa ctggtctact tctacgacag cgaaacggaa ctcaaaggta ttacctgtat	180
acacaacaca actttagggc cggcattggg cggtacccgc ctttggaact ataactctga	240
ggaagatgcc gttgaagacg taatccgtct ggctcggggc atgacttaca aagcggcttg	300
cgccgggtctg aatctgggcg gcggtaaaac cgtgctgatc ggtgatgcta aaaagattaa	360
atcagagtcc tacttccgtg gactggggcg ctacgttcag tcgctgaacg gcagatatat	420
caccgcggaa gacgtaaata cttctacgaa ggatatggca tacgttgcta tggaaactga	480
ctatgtggta ggcctgggag gtaaataccg caaccctagt ccagttactg cttacgggtg	540

atztatgggt atcaaagcgg cgctgatgaa aaaatttgag gatagctcta ttgaaggccg	600
aaccttcgca gtgcagggtg ctgggcagac gggttactat cttatcgatt acctcctagg	660
caacaacaag ttcaaagaaa aggctaaaaa gattttacttc accgaaatta acgagagcta	720
tatcgagcgt atgaacaaag aacatccgga agttgaattt atttccccgg acaaaatcta	780
ctcgctggaa gtagacgtct tcgtgccctg cgccctgggc aaaatcgta atgacaaaac	840
tatcgatgaa tttaaagtgtc cgatcatcgc aggtactgca aacaacgtac tggaaaggga	900
agcgcacggc aacatgctta aagaacgtgg cattctttac gccccggact atgtgatcaa	960
tgctgggtggg ctgatcaacg tttaccacga gctgaacggg tacaataaag agaacgctat	1020
tctggaagtg gaattaattt atgatcgcct actggaaata ttcaacatcg ctgattctct	1080
gaacatcagc accaatatcg ctgccaacga gttcgcggaa aaacgtatca agcaaattaa	1140
gtccttgaaa aacaacttca ttaaacgcta agaccattaa taaggaacat tttatgtcga	1200
cgacaaccgt tgggtgactac ttgctgtatc gcttaaacga aatcggcatt gagcacctct	1260
tcggagtgcc aggtgattac aatctgcaat ttctggatca tgtaatcgac caccctcagc	1320
tgacttgggt cggctgcact aacgaactta acgctgccta cgcagctgat ggttatgcgc	1380
gttgtcgtcc ggctgcggca ctgctgacca cttcgggggt tggcgaactg agcgctatta	1440
atggcatcgc aggttcctac gcggagtatc tgccggtaat acatatcgtt ggtgcaccga	1500
gtctatcagc ccagcagcag ggcgacctga ttcaccactc tcttggcgaa ggtgattttt	1560
ccagcttcct gaggatgtcc caaccggtgt ctgttgcgca ggctgctctg actcctgata	1620
acgcatgcaa ggaaatcgac cgcgtactgg cggaagtcct cattcagcgt cgtccccggt	1680
acctgctgct gtctaccgac gtggctgctg cgccggcggc tctgccacaa agcactcttt	1740
ctttgccgac cgccccggat catcgcgcag ttctggctgc tttcagcgac gctgctgagc	1800
agatgctggc tcaggccaaa agcgtctctc tactggcgga ctttctggct gatcgtttcg	1860
gtgttactcg agcactggcc gcgtggcttc agcaggttcc gctaccgcac gccactctgt	1920
taatgggtaa aggcgttctg agtgaacagc aaccagggtt cgtgggtacc tacgctgggtg	1980
cggcatctat cgattcgacg cgtggcgcaa tcgaagaagc tggggtaatt atcggagtgg	2040
gagttagatt ttccgacact atcacagcag gcttctcgca gcagatcgac gcccgccgtt	2100
ttatagacat tcaacccttc ttctctcgta ttggcgatcg ccagtttgat cacttgccga	2160

tgcaggctgc	cgtcgcagcc	ctgcatcaac	tgtgtcttcg	ttatcagcag	cagtggctcta	2220
tcaccgctcc	tagcccgctt	gcactgccgc	cggctgctgg	tagcgagctg	tcccagaacg	2280
cattctggca	ggcgatgcag	aacttcatcc	gccctgggga	cctgttggtg	gccgaccaag	2340
gtactgcggc	gttcggcgca	gcggcgctgc	gcttaccgca	gaattgccag	ctgcttgtgc	2400
agccgctgtg	gggctcaatc	ggttacagtc	tgccggccac	ctttgggtgct	cagacggcag	2460
atacagagcg	tcgtgtaatc	ctaattcattg	gcgatgggtc	agcgcaatta	actattcagg	2520
aactttccag	tatgatgcgt	gacggcttga	aacctatcat	ctttctcctg	aacaacaacg	2580
gttacaccgt	tgaacgggcg	attcacggcg	cggagcaacg	ttataacgat	atcgctgctt	2640
ggaattggac	ccaactgccc	caggcgctga	gtgttcattg	cccagcgag	agctggcgag	2700
tcgttgaaac	ggtgcagctg	accgacgtaa	tgaaagtcac	cgctgcttct	ccgctctga	2760
gcttggtaga	agttgttctg	cctgcaatgg	atgtcccacc	gctgctgcaa	gcagtgagtg	2820
ccgctctgaa	ccagcgcaac	tcctcttaac	ttttaagta	acacacgagc	taagaactaa	2880
ggaggattta	tgcgcagcat	gcagtttgat	gagtacgggtg	cacccctgaa	agcgttctca	2940
tatgaagacc	cgaccccgca	agggaaggaa	gtagtcgtta	ggatcgaagc	ctgtgggtgtg	3000
tgccactctg	atattcatct	tcacgagggc	tacttcgaca	tgggcgggtg	caataaagct	3060
gatgttactc	gtgctcgcga	actccctttt	acattgggtc	atgaaatcgt	tggcgaagtg	3120
gtagcaactg	gaccaggtgt	caccggcgct	aaaccgggcg	acaaacgtat	tgtgtacccg	3180
tggatcgggt	gcggcgactg	cccgaaatgc	aacagtgggtg	aggatcagtc	ctgtgcgcgt	3240
ccacgtaacc	tgggtgttca	cgttgacggt	ggctattcga	cgcacgtaaa	gataccggac	3300
gaaaaattcc	tgttcgccta	cgatggtatt	cctactgagt	tagcgggaac	ctatgcttgc	3360
agcggcatca	ccgcttatgg	tgcactgatg	aaagcaaagg	aagcggctga	aagatctggc	3420
tacatcggtc	tgattggcgc	tggtggcggt	ggcatggctg	gtctgatgct	ggccaaagca	3480
gcgatcgggg	ctaaaactgt	agtctttgat	atcgacgacg	caaaactgga	agctgcgacc	3540
cgtgccgggg	cggattacgt	gttcaactcc	ggtgcaaaag	aaacacgcaa	ggaagttatg	3600
aaactaacga	atgggtggcct	gtctggtgct	gttgatttcg	ttggcagcga	taaaagcgct	3660
ctgtttggaa	tcaacgcctt	gggtcagaac	ggcgtgctgg	tcataattgg	actgttcggt	3720
ggcgctatga	ctgttccggt	accctgttc	ccgctgaaag	ggatcaccgt	acgtggctca	3780

tacgtaggtt ccctgcaaga gatgagtgat atgatggagt tagttcgcgc tgggaaagtt	3840
cctccgatgc cggtaaaaac tcggccactg gacgctgcct gggaaaccct tgaggatcta	3900
cgccatggta aaatcgtggg ccgtgttggt ctgaccccat aagaaggaga tatacatatg	3960
acccatcaat taagatcgcg cgatatcatc gctctgggct ttatgacatt tgcgttggtc	4020
gtcggcgcag gtaacattat tttccctcca atggctggct tgcaggcagg cgaacacgtc	4080
tggactgcgg cattcggctt cctcattact gccgttggcc taccggtatt aacggtagtg	4140
gcgctggcaa aagttggcgg cgggtgtgac agtctcagca cgccaattgg taaagtcgct	4200
ggcgtactgc tggcaacagt ttgttacctg gcggtggggc cgctttttgc tacgccgcgt	4260
acagctaccg tttcttttga agtgggcatt gcgccgctga cgggtgattc cgcgctgccg	4320
ctgtttatth acagcctggt ctatttcgct atcgttattc tggtttcgct ctatccgggc	4380
aagctgctgg ataccgtggg caacttcctt gcgccgctga aaattatcgc gctggtcac	4440
ctgtctgttg ccgcaattat ctggccggcg ggttctatca gtacggcgac tgaggcttat	4500
caaaacgctg cgttttctaa cggcttcgtc aacggctatc tgaccatgga tacgctgggc	4560
gcaatggtgt ttggtatcgt tattgttaac gcggcgcgtt ctcgtggcgt taccgaagcg	4620
cgtctgctga cccgttatac cgtctgggct ggcctgatgg cgggtgttgg tctgactctg	4680
ctgtacctgg cgctgttccg tctgggttca gacagcgcgt cgctggtcga tcagtctgca	4740
aacggtgccg cgatcctgca tgcttacgtt cagcatacct ttggcggcgg cggtagcttc	4800
ctgctggcgg cgtaaatctt catcgcctgc ctggtcacgg cggttggcct gacctgtgct	4860
tgtgcagaat tcttcgcca gtacgtaccg ctctcttatac gtacgctggt gtttatctc	4920
ggcggcttct cgatggtggt gtctaacctc ggcttgagcc agctgattca gatctctgta	4980
ccggtgctga ccgccattta tccgccgtgt atcgcaactgg ttgtattaag ttttacacgc	5040
tcatggtggc ataattcgtc ccgcgtgatt gctccgccga tgtttatcag cctgcttttt	5100
ggtattctcg acgggatcaa ggcatctgca ttcagcgata tcttaccgtc ctgggcgcag	5160
cgtttaccgc tggccgaaca aggtctggcg tggttaatgc caacagtggg gatggtggtt	5220
ctggccatta tctgggatcg tgcggcaggt cgtcaggtga cctccagcgc tactaatac	5280
gcatggcatg gatgaccgat ggtagtgtgg gctctcccca tgcgagagta gggaactgcc	5340
aggcat	5346

<210> 42
<211> 5362
<212> DNA
<213> Artificial Sequence

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

<400> 42
gttgacactc tatcattgat agagttatth taccactccc tatcagtgat agagaaaagt 60
gagagctaga taataaggac ggcacaagat gaacatcttc aagaaaatgg aggaatttaa 120
ttatgaacaa ctggtctact tctacgacag cgaaacggaa ctcaaaggta ttacctgtat 180
acacaacaca actttagggc cggcattggg cggtaaccgc ctttggaact ataactctga 240
ggaagatgcc gttgaagacg taatccgtct ggctcggggc atgacttaca aagcggcttg 300
cgccgggtctg aatctgggag gcggtaaaac cgtgctgatc ggtgatgcta aaaagattaa 360
atcagagtcc tacttccgtg gactggggcg ctacgttcag tcgctgaacg gcagatatat 420
caccgcggaa gacgtaaata cttctacgaa ggatatggca tacgttgcta tggaaactga 480
ctatgtggta ggcctgggag gtaaattcgg caaccctagt ccagttactg cttacgggtgc 540
atthtatgggt atcaaagcgg cgctgatgaa aaaatttgag gatagctcta ttgaaggccg 600
aaccttcgca gtgcagggtg ctgggcagac gggttactat cttatcgatt acctcctagg 660
caacaacaag ttcaaagaaa aggctaaaaa gatttacttc accgaaatta acgagagcta 720
tatcgagcgt atgaacaaag aacatccgga agttgaatth atttccccgg acaaaatcta 780
ctcgctggaa gtagacgtct tcgtgccctg cgccctgggc aaaatcgtha atgacaaaac 840
tatcgatgaa tthtaagtgtc cgatcatcgc aggtactgca aacaacgtac tggaaaggga 900
agcgcacggc aacatgctta aagaacgtgg cattctthac gccccggact atgtgatcaa 960
tgctgggtggg ctgatcaacg thtaccacga gctgaacggg tacaataaag agaacgctat 1020
tctggaagtg gaatthattt atgatcgcct actggaaata thcaacatcg ctgattctct 1080
gaacatcagc accaatatcg ctgccaacga gttcgcggaa aaacgtatca agcaaattaa 1140
gtccttgaaa aacaacttca thaaacgcta agaccattaa taagggaacat thtatgtcga 1200
cgacaaccgt tggtgactac ttgctgtatc gctthaaacga aatcggcatt gagcacctct 1260

tcggagtgcc aggtgattac aatctgcaat ttctggatca tgtaatcgac caccctcagc	1320
tgacttgggt cggctgcact aacgaactta acgctgccta cgagctgat ggttatgcgc	1380
gttgtcgtcc ggctgcggca ctgctgacca ccttcgggggt tggcgaactg agcgctatta	1440
atggcatcgc aggttcctac gcggagtatc tgccggtaat acatatcggtt ggtgcaccga	1500
gtctatcagc ccagcagcag ggcgacctga ttcaccactc tcttggcgaa ggtgattttt	1560
ccagcttcct gaggatgtcc caaccgggtgt ctgttgcgca ggctgctctg actcctgata	1620
acgcatgcaa ggaaatcgac cgcgtactgg cggaagtcct cattcagcgt cgtcccggct	1680
acctgctgct gtctaccgac gtggctgctg cgccggcggc tctgccacaa agcactcttt	1740
ctttgccgac cgccccggat catcgcgcag ttctggctgc tttcagcgac gctgctgagc	1800
agatgctggc tcaggccaaa agcgtctctc tactggcgga ctttctggct gatcgtttcg	1860
gtgttactcg agcactggcc gcgtggcttc agcaggttcc gctaccgcac gccactctgt	1920
taatgggtaa aggcgttctg agtgaacagc aaccagggtt cgtgggtacc tacgctggtg	1980
cggcatctat cgattcgacg cgtggcgcaa tcgaagaagc tggggtaatt atcggagtgg	2040
gagttagatt ttccgacact atcacagcag gcttctcgca gcagatcgac gcccgccgtt	2100
ttatagacat tcaacccttc ttctctcgta ttggcgatcg ccagtttgat cacttgccga	2160
tgcaggctgc cgtcgcagcc ctgcatcaac tgtgtcttcg ttatcagcag cagtggctcta	2220
tcaccgctcc tagcccgctt gcaactgccg cggtgctgg tagcgagctg tcccagaacg	2280
cattctggca ggcgatgcag aacttcatcc gccctgggga cctgttggtg gccgaccaag	2340
gtactgcggc gttcggcgca gcggcgctgc gcttaccgca gaattgccag ctgcttgctc	2400
agccgctgtg gggctcaatc ggttacagtc tgccggccac ctttggtgct cagacggcag	2460
atacagagcg tcgtgtaatc ctaatcattg gcgatggttc agcgcaatta actattcagg	2520
aactttccag tatgatgcgt gacggcttga aacctatcat ctttctcctg aacaacaacg	2580
gttacaccgt tgaacgggcg attcacggcg cggagcaacg ttataacgat atcgctgctt	2640
ggaattggac ccaactgccc caggcgctga gtgttcattg cccagcgag agctggcgag	2700
tcgttgaaac ggtgcagctg accgacgtaa tgaaagtcac cgctgcttct ccgctctga	2760
gcttggtaga agttgttctg cctgcaatgg atgtcccacc gctgctgcaa gcagtgagtg	2820
ccgctctgaa ccagcgcaac tcctcttaat tggataaaga ataaggaggt attttatgac	2880

agcggagcag caaaatgggg tatccgactc acgccgtttc gaatttcagg aatttggtgg	2940
ccctatcgcc ccacagacct atcagctccc cgcaccggct agcgatgaag ttttgttaaa	3000
ggtgaactac tgcggtgtct gtcacagtga tgttcatctt cacgacggct acttcgagct	3060
gggtggcgat aaacgtctga acttcgctat gccgctgccg ctgacgctgg gtcacgaagt	3120
aattggcacc gttgtggctg tcggcgacca ggttactggg gtaaaaccgg gggaccagcg	3180
actgatctat ccgtggatag gttgcggaaa atgcggcgcg tgtcaaaaag gagaagaaaa	3240
cctgtgcgtt actcctgcac atctgggcgt gaacaagccg ggcggttacg ctgatcacat	3300
cgttgtagcc cattctcgct accttctgga catttcgggt ctgaaccggg gtgatgccgc	3360
taccctcgcg tgctccggcc tgaccacttt cagcgcgatc aacaaagtgt tgccgcttgc	3420
agatgaccag tggattgttg ttatcggttg tggtagcctc ggccagatgg cgctgcgtat	3480
cctgcaagct atgggaattg gcaatgttat cggtattgac ctgtctgaag agaaacggaa	3540
actggctcat gaaagcgggt cacgtcactc cttcgatcca aacactccga agctgaaccg	3600
cgtggtcgcc gaaacctgcc cgggtacggg acaggccgcg ttagactttg tgggcaatga	3660
gcaaactgct cagctggcac tgtctctgct tggaaaagggt ggcaaatatg ttcctgtcgg	3720
gctgcacggc ggcgagctgc gttaccatt gccgatcatc acgaacaaag ctgtaagtat	3780
catcggttct tacgttggtta ccctgaaaga actggaagac ttagttgctt tcgccaagga	3840
aaagaatctg ccgccaattc atattgaaca ccgcccgctg gaatcggcgg ctgaggccgt	3900
agaggacctg gaaaaaggac aggttgctgg gcgtgttatc ctggatgcag gtaactaaga	3960
aggagatata catatgacct atcaattaag atcgcgcgat atcatcgctc tgggctttat	4020
gacatttgcg ttgttcgtcg gcgcaggtaa cattattttc cctccaatgg tcggcttgca	4080
ggcaggcgaa cacgtctgga ctgcggcatt cggcttcctc attactgccg ttggcctacc	4140
ggtattaacg gtagtggcgc tggcaaaagt tggcggcggg gttgacagtc tcagcacgcc	4200
aattggtaaa gtcgctggcg tactgctggc aacagtttgt tacctggcgg tggggccgct	4260
ttttgctacg ccgcgtacag ctaccgtttc ttttgaagtg ggcatcgcg cgctgacggg	4320
tgattccgcg ctgccgctgt ttatttacag cctgggtctat ttcgctatcg ttattctggt	4380
ttcgctctat ccgggcaagc tgctggatac cgtgggcaac ttccttgcg cgctgaaaat	4440
tatcgcgctg gtcacacctgt ctgttgccgc aattatctgg ccggcgggtt ctatcagtac	4500

ggcgactgag gcttatcaaa acgctgcggt ttctaacggc ttcgtcaacg gctatctgac	4560
catggatacg ctgggcgcaa tgggtgtttgg tatcggttatt gttaacgcgg cgcgttctcg	4620
tggcggttacc gaagcgcgtc tgctgacccg ttataccgctc tgggctggcc tgatggcggg	4680
tgttggtctg actctgctgt acctggcgct gttccgtctg ggttcagaca gcgcgtcgct	4740
ggtcgatcag tctgcaaacg gtgcggcgat cctgcatgct tacgttcagc atacctttgg	4800
cggcggcggg agcttctgc tggcggcggt aatcttcac gcctgcctgg tcacggcggg	4860
tggcctgacc tgtgcttgtg cagaattctt cgcccagtac gtaccgctct cttatcgta	4920
gctgggtgtt atcctcggcg gcttctcgat ggtgggtgtt aacctcggct tgagccagct	4980
gattcagatc tctgtaccgg tgctgaccgc ctttatccg ccgtgtatcg cactgggtgt	5040
attaagtttt acacgctcat ggtggcataa ttcgtccgc gtgattgctc cgccgatgtt	5100
tatcagcctg ctttttggtt ttctcgacgg gatcaaggca tctgcattca gcgatatctt	5160
accgtcctgg gcgcagcgtt taccgctggc cgaacaaggc ctggcgtggg taatgccaac	5220
agtgggtgat gtggttctgg ccattatctg ggatcgtgcg gcaggctcgc aggtgacctc	5280
cagcgtcac taatacgcat ggcatggatg accgatggta gtgtgggctc tccccatgcg	5340
agagtaggga actgccaggc at	5362

<210> 43

<211> 5370

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 43

gttgacactc tatcattgat agagttatct taccactccc tatcagtgat agagaaaagt	60
gaaaacaggc caaacgtagg ctccggaatt taaggaggta ttattatggg agagatccag	120
gctttgccgg aaacttccat ttttgggcaa atcgcagacc accagcatga acaggtggc	180
ttctgccacg atcacgaaac cggcctccgt gcgataatcg gtattcataa cacggttctt	240
ggccccgcct taggtggaac tcgcatgtgg cactatgcta ccgaggcaga agcgtgaat	300
gacgttctgc gtctgtctcg cggtatgacc tacaaggctg ctatctcggg cctgaacctg	360
ggtggcggta aagcagtaat cattggggat gccaaaacaa tcaaaaccga agcgtgctg	420

cggaaattcg gcagattcgt gcagaacctg aatggtaaata acatcactgc tgaagacgtt	480
aacatgacta caaaggatat ggagttacatt aggatggaaa ccaaacacgt cgctggctta	540
cctgaaagta tgggtggaag cgggtgacccg tcaccggtaa ctgcatatgg tacgtacatg	600
ggcatgaaag cggcggccaa aaaggcgttt ggctctgatt ccctggctgg caaacgtatc	660
gctgttcaag gtgtgggtca tgttggcact tatctgcttg agcatttgac caaagaaggt	720
gctcagattg tgctgactga ctactataag gaacgtgccg aggaagcagg cgcgcgtttt	780
ggcgcacagg ttgttggcct ggacgatatc tacgatcaag aggtcgacat ttactctcca	840
tgtgctctcg gtgctaccat caacgatgac actatcgatc gcctgcgttg cgctgttgta	900
gccggttgcg caaacaacca gctgaaagaa gaaaacgtcc acggtccggc gctggttgag	960
cgcgggatag tatacgcccc agacttcctg atcaatgcag ggggcctgat taacgtgtat	1020
agcgaagtta cagggtctac ccgtcagggg gctttaactc agaccgaaaa aatctatgac	1080
tacacactcc aagttctgga aaaagccgcg gctgaaggtc tgcacccgca gcaggctgcg	1140
atccgtcagg cggaacaacg catcgtgca attggtaagg tgaagacac ctactaatga	1200
gaaataagga ggatttcattg tcgacgacaa ccgttggtga ctacttgctg tatcgcttaa	1260
acgaaatcgg cattgagcac ctcttcggag tgccagggtga ttacaatctg caatttctgg	1320
atcatgtaat cgaccacct cagctgactt gggtcggctg cactaacgaa cttaacgctg	1380
cctacgcagc tgatggttat gcgcgttgtc gtccggctgc ggactgctg accaccttcg	1440
gggttggcga actgagcgct attaatggca tcgcaggttc ctacgcggag tatctgccgg	1500
taatacatat cgttggtgca ccgagtctat cagcccagca gcagggcgac ctgattcacc	1560
actctcttgg cgaagggtgat ttttccagct tcctgaggat gtcccaaccg gtgtctgttg	1620
cgcaggctgc tctgactcct gataacgcat gcaaggaaat cgaccgcgta ctggcgggaag	1680
tcctcattca gcgtcgtccc ggctacctgc tgctgtctac cgacgtggct gctgcgccgg	1740
cggctctgcc acaaagcact ctttctttgc cgaccgcccc ggatcatcgc gcagttctgg	1800
ctgctttcag cgacgctgct gagcagatgc tggctcaggc caaaagcgtc tctctactgg	1860
cggactttct ggctgatcgt ttcgggtgta ctcgagcact ggccgcgtgg cttcagcagg	1920
ttccgctacc gcacgccact ctgttaatgg gtaaaggcgt tctgagtga cagcaaccag	1980
ggttcgtggg tacctacgct ggtgcggcat ctatcgattc gacgcgtggc gcaatcgaag	2040

aagctgggggt aattatcgga gtgggagtta gattttccga cactatcaca gcaggcttct	2100
cgcagcagat cgacgcccgc cgttttatag acattcaacc cttcttctct cgtattggcg	2160
atcgccagtt tgatcacttg ccgatgcagg ctgccgtcgc agccctgcat caactgtgtc	2220
ttcgttatca gcagcagtgg tctatcaccg ctccctagccc gcctgcactg ccgccggctg	2280
ctggtagcga gctgtcccag aacgcattct ggcaggcgat gcagaacttc atccgccctg	2340
gggacctgtt ggtggccgac caaggtactg cggcgttcgg cgcagcggcg ctgcgcttac	2400
cgcagaattg ccagctgctt gtgcagccgc tgtggggctc aatcggttac agtctgccgg	2460
ccacctttgg tgctcagacg gcagatacag agcgtcgtgt aatcctaatac attggcgatg	2520
gttcagcgca attaactatt caggaacttt ccagtatgat gcgtgacggc ttgaaaccta	2580
tcatctttct cctgaacaac aacggttaca ccgttgaacg ggcgattcac ggcgcggagc	2640
aacgttataa cgatatcgct gcttgggaatt ggacccaact gccccaggcg ctgagtgttc	2700
attgcccagc gcagagctgg cgagtcgttg aaacggtgca gctgaccgac gtaatgaaag	2760
tcatcgctgc ttctccgctg ctgagcttgg tagaagttgt tctgcctgca atggatgtcc	2820
caccgctgct gcaagcagtg agtgccgctc tgaaccagcg caactcctct taatttttaa	2880
agtaacacac gagctaagaa ctaaggagga tttatgcgca gcatgcagtt tgatgagtac	2940
ggtgcacccc tgaaagcggt ctcatatgaa gacccgacct cgcaagggaa ggaagtagtc	3000
gtaggatcg aagcctgtgg tgtgtgccac tctgatattc atcttcacga gggctacttc	3060
gacatgggcg gtggcaataa agctgatgtt actcgtgctc gcgaactccc ttttacattg	3120
ggtcatgaaa tcgttggcga agtggttagca actggaccag gtgtcaccgg cgctaaaccg	3180
ggcgacaaac gtattgtgta cccgtggatc ggggtcggcg actgcccga atgcaacagt	3240
ggtgaggatc agtcctgtgc gcgtccacgt aacctgggtg ttcacgttga cgggtggctat	3300
tcgacgcacg taaagatacc ggacgaaaaa ttcctgttcg cctacgatgg tattcctact	3360
gagttagcgg gaacctatgc ttgcagcggc atcaccgctt atggtgcact gatgaaagca	3420
aaggaagcgg ctgaaagatc tggctacatc ggtctgattg gcgctgggtg cgttggcatg	3480
gctggcttga tgctggccaa agcagcgatc ggggctaata ctgtagtctt tgatatcgac	3540
gacgcaaaac tggaagctgc gacccgtgcc ggggcggatt acgtgttcaa ctccggtgca	3600
aaagaaacac gcaaggaagt tatgaaacta acgaatgggtg gcctgtctgg tgctgttgat	3660

ttcgttggca gcgataaaag cgctctgttt ggaatcaacg ccttgggtca gaacggcgtg	3720
ctggtcataa ttggactggt cgggtggcgct atgactgttc cggtagccct gttcccgtg	3780
aaagggatca ccgtacgtgg ctcatagta gggtccctgc aagagatgag tgatatgatg	3840
gagttagttc gcgctgggaa agttcctccg atgccggtaa aaactcggcc actggacgct	3900
gcctgggaaa cccttgagga tctacgccat ggtaaaatcg tgggccgtgt tgttctgacc	3960
ccataagaag gagatataca tatgacccat caattaagat cgcgcgatat catcgctctg	4020
ggctttatga catttgcggt gttcgtcggc gcaggtaaca ttattttccc tccaatggtc	4080
ggcttgagg caggcgaaca cgtctggact gcggcattcg gcttcctcat tactgccgtt	4140
ggcctaccgg tattaacggt agtggcgctg gcaaaagttg gcggcggtgt tgacagtctc	4200
agcacgcaa ttggtaaagt cgctggcgta ctgctggcaa cagtttgta cctggcggtg	4260
gggccgcttt ttgctacgcc gcgtacagct accgtttctt ttgaagtggg cattgcgccg	4320
ctgacgggtg attccgcgct gccgctgttt atttacagcc tgggtctattt cgctatcggt	4380
attctgggtt cgctctatcc gggcaagctg ctggataccg tgggcaactt ccttgcgccg	4440
ctgaaaatta tcgcgctggt catcctgtct gttgccgcaa ttatctggcc ggcgggttct	4500
atcagtacgg cgactgaggc ttatcaaac gctgcgtttt ctaacggctt cgtcaacggc	4560
tatctgacca tggatacgct gggcgcaatg gtgtttggta tcgttattgt taacgcggcg	4620
cgttctctg gcgttaccga agcgcgtctg ctgaccggtt ataccgtctg ggctggcctg	4680
atggcgggtg ttggtctgac tctgctgtac ctggcgctgt tccgtctggg ttcagacagc	4740
gcgtcgctgg tcgatcagtc tgcaaacggt gcggcgatcc tgcatgctta cgttcagcat	4800
acctttggcg gcggcggtag cttcctgctg gcggcggtta tcttcatcgc ctgcctggtc	4860
acggcgggtg gcctgacctg tgcttgtgca gaattcttcg cccagtacgt accgctctct	4920
tatcgtacgc tgggtgttat cctcggcggc ttctcgatgg tgggtgtctaa cctcggcttg	4980
agccagctga ttcagatctc tgtaccggtg ctgaccgcca tttatccgcc gtgtatcgca	5040
ctggttgat taagttttac acgctcatgg tggcataatt cgtcccgcgt gattgctccg	5100
ccgatgttta tcagcctgct ttttgggtatt ctgcacggga tcaaggcatc tgcatcagc	5160
gatatcttac cgtcctgggc gcagcgttta ccgctggccg aacaaggctt ggctgggtta	5220
atgccaacag tggatgatgt ggttctggcc attatctggg atcgtgcggc aggtcgtcag	5280

gtgacctcca gcgctcacta atacgcatgg catggatgac cgatggtagt gtgggctctc	5340
cccatgcgag agtagggaac tgccaggcat	5370

<210> 44
 <211> 5386
 <212> DNA
 <213> Artificial Sequence

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

<400> 44	
gttgacactc tatcattgat agagttatTTt taccactccc tatcagtgat agagaaaagt	60
gaaaacaggc caaacgtagg ctccggaatt taaggaggta ttattatggT agagatccag	120
gctttgccgg aaacttccat ttttgggcaa atcgCagacc accagcatga acaggTggTc	180
ttctgccacg atcacgaaac cggcctccgt gcgataatcg gtattcataa cacggTtctt	240
ggccccgcct taggtggaac tcgcatgtgg cactatgcta ccgaggcaga agcgctgaat	300
gacgttctgc gtctgtctcg cggtatgacc tacaaggctg ctatctcggg cctgaacctg	360
ggTggcggtA aagcagtaat cattggggat gccaaaacaa tcaaaaccga agcgctgctg	420
cggaaattcg gcagattcgt gcagaacctg aatggtaaT acatcactgc tgaagacgtt	480
aacatgacta caaaggatat ggagtacatt aggatggaaa ccaaacacgt cgctggctta	540
cctgaaagta tgggtggaag cggtgacctg tcaccggtaa ctgcatatgg tacgtacatg	600
ggcatgaaag cggcggccaa aaaggcgTtt ggctctgatt ccctggctgg caaacgtatc	660
gctgttcaag gtgtgggtca tgttggcact tatctgcttg agcatttgac caaagaaggT	720
gctcagattg tgctgactga ctactataag gaacgtgccg aggaagcagg cgcgcgTTTT	780
ggcgcacagg ttgttgcct ggacgatatc tacgatcaag aggtcgacat ttactctcca	840
tgtgctctcg gtgctaccat caacgatgac actatcgatc gcctgcgttg cgctgttgta	900
gccggttgcg caaacaacca gctgaaagaa gaaaacgtcc acggtccggc gctggttgag	960
cgcgggatag tatacgcccc agacttcttg atcaatgcag ggggcctgat taacgtgtat	1020
agcgaagtta cagggtctac ccgtcagggg gctttaactc agaccgaaaa aatctatgac	1080
tacacactcc aagttctgga aaaagccgcg gctgaaggTc tgcacccgca gcaggctgcg	1140

atccgtcagg cggaacaacg catcgtgca attggtaagg tgaaaagcac ctactaatga	1200
gaaataagga ggatttcatg tcgacgacaa ccgttggtga ctacttgctg tatcgcttaa	1260
acgaaatcgg cattgagcac ctcttcggag tgccagggtga ttacaatctg caatttctgg	1320
atcatgtaat cgaccaccct cagctgactt gggtcggctg cactaacgaa cttaacgctg	1380
cctacgcagc tgatggttat gcgcgttgtc gtccggctgc ggcaactgctg accaccttcg	1440
gggttggcga actgagcgct attaatggca tcgcaggttc ctacgcggag tatctgccgg	1500
taatacatat cgttgggtgca ccgagtctat cagcccagca gcagggcgac ctgattcacc	1560
actctcttgg cgaagggtgat ttttccagct tcctgaggat gtcccaaccg gtgtctgttg	1620
cgcaggctgc tctgactcct gataacgcag gcaaggaaat cgaccgcgta ctggcggaag	1680
tcctcattca gcgtcgtccc ggctacctgc tgctgtctac cgacgtggct gctgcgccgg	1740
cggctctgcc acaaagcact ctttctttgc cgaccgcccc ggatcatcgc gcagttctgg	1800
ctgctttcag cgacgtgct gagcagatgc tggctcaggc caaaagcgtc tctctactgg	1860
cggactttct ggctgatcgt ttcgggtgta ctcgagcact ggccgcgtgg cttcagcagg	1920
ttccgctacc gcacgccact ctgttaatgg gtaaaggcgt tctgagtga cagcaaccag	1980
ggttcgtggg tacctacgct ggtgcggcat ctatcgattc gacgcgtggc gcaatcgaag	2040
aagctggggg aattatcgga gtgggagtta gattttccga cactatcaca gcaggcttct	2100
cgcagcagat cgacgcccgc cgttttatag acattcaacc cttcttctct cgtattggcg	2160
atcgccagtt tgatcacttg ccgatgcagg ctgccgtcgc agccctgcat caactgtgtc	2220
ttcgttatca gcagcagtgg tctatcaccg ctctagccc gcctgcactg ccgccggctg	2280
ctggtagcga gctgtcccag aacgcattct ggcaggcgat gcagaacttc atccgccctg	2340
gggacctgtt ggtggccgac caagggtactg cggcgttcgg cgacgcggcg ctgcgcttac	2400
cgcagaattg ccagctgctt gtgcagccgc tgtggggctc aatcggttac agtctgccgg	2460
ccacctttgg tgctcagacg gcagatacag agcgtcgtgt aatcctaatac attggcgatg	2520
gttcagcgca attaaactatt caggaacttt ccagtatgat gcgtgacggc ttgaaaccta	2580
tcatttttct cctgaacaac aacggttaca ccgttgaacg ggcgattcac ggcgcgaggc	2640
aacgttataa cgatatcgct gcttgggaatt ggacccaact gccccaggcg ctgagtgttc	2700
attgcccagc gcagagctgg cgagtcgttg aaacgggtgca gctgaccgac gtaatgaaag	2760

tcatcgctgc ttctccgctg ctgagcttgg tagaagttgt tctgcctgca atggatgtcc	2820
caccgctgct gcaagcagtg agtgccgctc tgaaccagcg caactcctct taattggata	2880
aagaataagg aggtatttta tgacagcgga gcagcaaaat ggggtatccg actcacgccg	2940
tttcgaattt caggaatttg gtggccctat cgccccacag acctatcagc tccccgcacc	3000
ggctagcgat gaagttttgt taaagggtgaa ctactgcggg gtctgtcaca gtgatgttca	3060
tcttcacgac ggctacttcg agctgggttg cgataaacgt ctgaacttcg ctatgccgct	3120
gccgctgacg ctgggtcacg aagtaattgg caccgttggt gctgtcggcg accaggttac	3180
tggtgtaaaa ccgggggacc agcgactgat ctatccgtgg ataggttgcg gaaaatgcgg	3240
cgcgtgtcaa aaaggagaag aaaacctgtg cgttactcct gcacatctgg gcgtgaacaa	3300
gccgggagggt tacgctgac acatcgttgt accccattct cgctaccttc tggacatttc	3360
gggtctgaac ccgggtgatg ccgctaccct cgcgtgctcc ggctgacca ctttcagcgc	3420
gatcaacaaa gtgttgccgc ttgcagatga ccagtggatt gttgttatcg gttgtgggtg	3480
cctcggccag atggcgctgc gtatcctgca agctatggga attggcaatg ttatcggtat	3540
tgacctgtct gaagagaaac ggaaactggc tcatgaaagc ggtgcacgtc actccttcga	3600
tccaaacact ccgaagctga accgcgtggg cgccgaaacc tgcccgggta cggtagaggc	3660
cgcgttagac tttgtgggca atgagcaaac tgctcagctg gcaactgtct tgcttgga	3720
aggtggcaaa tatgttcctg tcgggctgca cggcggcgag ctgcgttacc cattgccgat	3780
catcacgaac aaagctgtaa gtatcatcgg ttcttacgtt ggtaccctga aagaactgga	3840
agacttagtt gctttcgcca aggaaaagaa tctgccgcca attcatattg aacaccgccc	3900
gctggaatcg gcggctcagg ccgtagagga cctggaaaaa ggacagggtg ctgggcgtgt	3960
tatcctggat gcaggtaact aagaaggaga tatacatatg acccatcaat taagatcgcg	4020
cgatatcatc gctctgggct ttatgacatt tgcgttggtt gtcggcgag gtaacattat	4080
tttccctcca atggctggct tgcaggcagg cgaacacgtc tggactgcgg cattcggtt	4140
cctcattact gccgttggcc taccggtatt aacggtagtg gcgctggcaa aagttggcgg	4200
cgggtgtgac agtctcagca cgccaattgg taaagtcgct ggcgtagtgc tggcaacagt	4260
ttgttacctg gcgggtggggc cgctttttgc tacgccgctg acagctaccg tttcttttga	4320
agtgggcatt gcgccgtga cgggtgattc cgcgctgccg ctgtttattt acagcctggg	4380

ctatttcgct atcgtttattc tggtttcgct ctatccgggc aagctgctgg ataccgtggg	4440
caacttcctt gcgccgctga aaattatcgc gctgggtcatc ctgtctgttg ccgcaattat	4500
ctggccggcg ggttctatca gtacggcgac tgaggcttat caaaacgctg cgttttctaa	4560
cggcttcgtc aacggctatc tgaccatgga tacgctgggc gcaatggtgt ttggtatcgt	4620
tattgttaac gcggcgcggt ctcgtggcgt taccgaagcg cgtctgctga cccgttatac	4680
cgtctgggct ggcctgatgg cgggtgttgg tctgactctg ctgtacctgg cgctgttccg	4740
tctgggttca gacagcgcggt cgctggtcga tcagtctgca aacggtgcgg cgatcctgca	4800
tgcttacgtt cagcatacct ttggcggcgg cggtagcttc ctgctggcgg cgtaaatctt	4860
catcgcctgc ctgggtcacgg cggttggcct gacctgtgct tgtgcagaat tcttcgcca	4920
gtacgtaccg ctctcttata gtacgtgggt gtttatcctc ggcggttctt cgatgggtgg	4980
gtctaacctc ggcttgagcc agctgattca gatctctgta ccggtgctga ccgccattta	5040
tccgccgtgt atcgactgg ttgtattaag ttttacacgc tcatgggtggc ataattcgtc	5100
ccgcgtgatt gtcgccga tgtttatcag cctgcttttt ggtattctcg acgggatcaa	5160
ggcatctgca ttcagcgata tcttaccgtc ctgggcgcag cgtttaccgc tggccgaaca	5220
aggtctggcg tggttaatgc caacagtgggt gatggtgggt ctggccatta tctgggatcg	5280
tgcggcaggt cgtcaggtga cctccagcgc tcactaatac gcatggcatg gatgaccgat	5340
ggtagtgtgg gctctcccca tgcgagagta gggaactgcc aggcatt	5386

<210> 45

<211> 5329

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 45

gttgacactc tatcattgat agagttatth taccactccc tatcagtgat agagaaaagt	60
gaactacagt atataaggag ggtaatatga tcttcgagac aatttctacg tcgaatcacg	120
aagaagttgt gtattgccat aacaaggacg ccggcttgaa agcaatcatc gcgattcaca	180
acactgtact cgggtccggct ctgggtggca ctgcacgtg gccctacgct agcgaagagg	240
aagcactgaa agatgtcctt cgtttatccc gtgggatgac ctacaaagct gcggtttcag	300

gtctaaacct gggcggcggg aaagctgtga tctgggggtga tccgaataaa gacaagtctg	360
aagcgcgtgtt tagagccttc ggacgggtttg taaacagcct gggcggacgc tacattaccg	420
cggaggacgt tggcattgat gttaacgaca tggaatatgt gctgcgtgaa actgattacg	480
tcaccggtgt acatcagggt cacgggtggga gtggtgatcc ttctccattc accgcatatg	540
gcactctgca aggcctgatg gccgctctgc aagtgaatc cggtaacgaa gacgtaggca	600
attacagcta cgctgttcag ggtgtgggtc acgttggcat ggaatttggt aaactgctgc	660
gtgagcgcgg tgcaaagggt ttcgtcactg acatcaacaa agatgcgggtc cagcgtgctg	720
tggacgaatt tggttgtgag gcagtagccc tggatgaaat ctatgacgtt gattgcgacg	780
tgtactcccc gaccgctctg ggcggcaccg tgaacgataa aactttaccg cgtctgaaat	840
gtaaggtaat ctgcggtgcg gcaaacaacc agttagctaa tgatgagata ggcgtggaac	900
tggaaaagaa aggcatcctc tatgctccgg actacgcggg caacgcgggt gggctgatga	960
acgttagcct ggaaatcgat ggatacaacc gcgaacgtgc gatgcgtatg atgcgtacca	1020
tttattacaa tttgggtcgc attttcgaaa tctctaagcg cgacggcatc cctacattcc	1080
gagccgccga tcgtatggct gaagaacgca taacggccat cggtaaactg cgtttaccgc	1140
at ttggggcgc tgcggcaccg cgcttcagg gccgacgtgg caactaacag aaatagtaaa	1200
ggagggttttt aatgtcggag atcacattgg gtagatacct tttcgaacgc ttaaaccaac	1260
tgcaagtgca gactatTTTT gggctgcccg gcgacttcaa tctgtccctg ctggataaga	1320
tctatgaagt tgatggcatg cgttgggcag gtaacgctaa cgaactcaac gccgcttacg	1380
cggctgacgg ttatagccgt gtcaaaggcc tcgcatgtct ggttaccact tttggtgtag	1440
gcgagctaag tgcgctgaat ggtgtgggtg gcgcttacgc agaacacgtt gggctgctgc	1500
atgtagtggg cgtcccatca atctctagcc aggcgaaaca gctgctgctg caccataccc	1560
tggtgaacgg agatttcacg gttttccacc gcatgtccaa caacatttct cagaccacgg	1620
cttttatcag cgacattaat tctgctcctg gtgaaatcga taggtgcatc cgtgaggcct	1680
gggtacatca gcgtccggtt tacgtcggcc tgccggcgaa cctagttgac ctgactgtgc	1740
cggcgtctct gttagacact ccgatcgatc tgtccttgaa aaagaacgac ccggatgccc	1800
aggaagaagt tattgaaacc gtccttgatc tggtagacaa gtctaaaaac cctataatct	1860
tagttgacgc atgcgctagc cgtcactcat gccgcgatga agtacgccgg ttggtggact	1920

ccaccagctt	cccggttttc	gttactccaa	tgggtaaatc	tgctgtaaat	gagagtcacc	1980
cgcgtttttg	cggtgtttac	gtgggcagcc	tcagcgagcc	aaacgtaaaa	gaagccgttg	2040
aaaacgctga	cctggtgctg	tccataggcg	ccctgttgag	cgacttcaac	actggatcgt	2100
tctcttattc	ctacaaaact	aagaacattg	ttgaatttca	ctctgattat	acaaaaatcc	2160
gtcaagcaac	gttcccgggt	gttcagatga	aagaagcact	gaatgtcctg	ttggaaaaaa	2220
tcccagacca	tgtcgctaac	tacaaacctc	tgccggttcc	gcagcgtcgc	gttattccga	2280
gcccagggga	taaggctgcg	atctctcagg	agtggctgtg	gtcgcgtctg	tctagctggt	2340
tccgcgaggg	cgacatcgtc	attacagaaa	ccggtaccag	tgcgtttgga	attgtacagt	2400
cctattttcc	agataactgc	atcggcatca	gtcagggtgt	gtggggttcg	atcggcttca	2460
ccgtaggtgc	aacgctgggc	gcggtgatgg	ctgcacaaga	aatcgatccg	aaaaaacgtg	2520
tgattttatt	tgtcggtgac	ggttctctgc	aacttactgt	acaggaaatt	tctaccatgg	2580
ttaagtggga	aaccactccc	tacctgtttg	tgctgaacaa	cgatgggtac	actatcgaac	2640
gccttatcca	tggcgagact	gctacgtata	acgatattca	gccgtgggat	aatctgggtc	2700
tgttgccgct	gttcaaagct	cgtgactacg	aaaccaaccg	agttgcgact	gtaggcgaaa	2760
ttgaagcgct	attcaacaat	tcagctttca	atgagaatac	aaagatccgt	atggtggagg	2820
tcatgctgcc	gcggatggat	gcaccacaga	acctggttaa	acaggctgaa	ttttcctcca	2880
agaccaacag	cgaaaactaa	acaaggcacg	gaaaaattgg	aggagaaaga	tgacagcggc	2940
gttcttgtat	aagccaggtc	acgaaaactt	agtgcgctcg	gagatcccga	tacctaaagc	3000
tgggcgtggc	gaagtcgttc	tggaaattaa	agccgctggc	atgtgccatt	ccgatctgca	3060
cgttctcgac	ggtggaatcc	ccctgccggg	tcaatttgta	atgggccatg	aaatcgttgg	3120
tactattcac	gagatcggcc	aggacgtgac	cggtttcaaa	cagggcgatc	tgtacgcagt	3180
ccacggcccc	aatccgtgtg	gtatttgcac	cctgtgcaga	gaaggatttg	ataacgactg	3240
cactacagtg	gcgaaaaccg	gtcaatggtt	cggactgggt	cttgacggcg	gctaccagaa	3300
gtatatccgt	atcccgaacg	taaggctctat	cgttaaagtt	ccagaagggtg	tttcagctga	3360
ggcagctgcg	agctgtactg	atgcagtact	gaccccgtag	cgtgcactaa	aacaggctgg	3420
cgccagcaac	tctactcggg	tactgattct	gggtctgggt	ggcttaggtc	tgaatgccct	3480
taaactggct	aagaccttcg	gcagttacgt	ttacgcatct	gacctgaaac	cttctgcgcg	3540

tgaagctgct aaggccgctg gggcggatga agtgctggag tccctgcccg aagacccgct	3600
gggtgttgat atcgtgttag acgtcgttgg cgtgcagagc accttcaacc tcgctcaaaa	3660
acacgttggc ccgcgtggca tcattgtacc tgtaggcctg gcatccccac agctttcgtt	3720
taacctaacg gatctggcgc tccgcgaaat tcgtgttcag ggcacttttt ggggcacgag	3780
caatgagctg gctgaatgtc tgcgcctgtg ccagctgggc ctgatcaacc cgaaatatac	3840
tgtggtgcct cttgaagaag cgccgaaata tatggaagca atggctcatg ggaaagtaga	3900
aggtcgtatc gttttccacc cgtaagaagg agatatacat atgacccatc aattaagatc	3960
gcgcgatatc atcgtctctg gctttatgac atttgcgttg ttcgtcggcg caggtaacat	4020
tattttccct ccaatggtcg gcttgcaggc aggccaacac gtctggactg cggcattcgg	4080
cttcctcatt actgccgttg gcctaccggt attaacggta gtggcgctgg caaaagttgg	4140
cggcgggtgtt gacagtctca gcacgccaat tggtaaagtc gctggcgtag tgctggcaac	4200
agtttgttac ctggcggtag ggccgctttt tgctacgccg cgtacagcta ccgtttcttt	4260
tgaagtgggc attgcgccgc tgacgggtga ttccgcgctg ccgctgttta ttacagcct	4320
ggtctatttc gctatcgtaa ttctggtttc gctctatccg ggcaagctgc tggataccgt	4380
gggcaacttc cttgcgccgc tgaaaattat cgcgctggtc atcctgtctg ttgccgcaat	4440
tatctggccg gcgggttcta tcagtacggc gactgaggct tatcaaaacg ctgcgttttc	4500
taacggcttc gtcaacggct atctgaccat ggatacgctg ggcgcaatgg tgtttggtat	4560
cgttattgtt aacgcggcgc gttctcgtgg cgttaccgaa gcgcgtctgc tgacccgtaa	4620
taccgtctgg gctggcctga tggcgggtgt tggctgact ctgctgtacc tggcgctgtt	4680
ccgtctgggt tcagacagcg cgtcgtgggt cgatcagtct gcaaacggtg cggcgatcct	4740
gcatgcttac gttcagcata cttttggcgg cggcggtagc ttcctgctgg cggcgtaa	4800
cttcacgcc tgcctggtca cggcgggttg cctgacctgt gcttgtgcag aattcttcgc	4860
ccagtacgta ccgctctctt atcgtacgct ggtgtttatc ctcggcggct tctcgatgg	4920
ggtgtctaac ctcggcttga gccagctgat tcagatctct gtaccgggtgc tgaccgccat	4980
ttatccgccg tgtatgcac tggttgtatt aagttttaca cgctcatggt ggcataattc	5040
gtcccgcgtg attgctccgc cgatgtttat cagcctgctt tttggtattc tcgacgggat	5100
caaggcatct gcattcagcg atatcttacc gtcctgggcg cagcgtttac cgctggccga	5160

acaaggtctg gcgtgggttaa tgccaacagt ggtgatgggtg gttctggcca ttatctggga	5220
tcgtgctggca ggtcgtcagg tgacctccag cgctcactaa tacgcatggc atggatgacc	5280
gatggtagtg tgggctctcc ccatgcgaga gtagggaact gccaggcat	5329

<210> 46

<211> 5369

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 46

gttgacactc tatcattgat agagttatth taccactccc tatcagtgat agagaaaagt	60
gaactacagt atataaggag ggtaatatga tcttcgagac aatttctacg tcgaatcacg	120
aagaagttgt gtattgccat aacaaggacg ccggcttgaa agcaatcatc gcgattcaca	180
acactgtact cgggtccggct ctgggtggca ctgcatgtg gccctacgct agcgaagagg	240
aagcactgaa agatgtcctt cgthttatccc gtgggatgac ctacaaagct gcggtttcag	300
gtctaaacct gggcggcgggt aaagctgtga tctgggggtga tccgaataaa gacaagtctg	360
aagcgctgtt tagagccttc ggacggthtg taaacagcct gggcggacgc tacattaccg	420
cggaggacgt tggcattgat gttaacgaca tggaatatgt gctgcgtgaa actgattacg	480
tcaccgggtgt acatcaggth cacgggtggga gtgggtgatcc thctccattc accgcatatg	540
gcactctgca aggctgatg gccgctctgc aagtgaatt cggtaacgaa gacgtaggca	600
attacagcta cgctgttcag ggtgtgggtc acgttggcat ggaatttgth aaactgctgc	660
gtgagcgagg tgcaaaggth ttcgtcactg acatcaacaa agatgcggtc cagcgtgctg	720
tggacgaatt tggthgtgag gcagtagccc tggatgaaat ctatgacgth gattgcgacg	780
tgtactcccc gaccgctctg ggcggcaccc tgaacgataa aactthaccg cgtctgaaat	840
gtaaggtaat ctgcgggtgc gcaaacaacc agttagctaa tgatgagata ggcgtggaac	900
tggaaaagaa aggcatcctc tatgctccgg actacgggt caacgcgggt gggctgatga	960
acgttagcct ggaaatcgat ggatacaacc gcgaacgtgc gatgcgtatg atgcgtacca	1020
thtattacaa thtgggtcgc atthttcgaat tctctaagcg cgacggcatc cctacattcc	1080

gagccgccga tcgtatggct gaagaacgca taacggccat cggtaaactg cgtttaccgc	1140
at ttggggcg tgccggcaccg cgcttccagg gccgacgtgg caactaaaat tcactttaac	1200
aactaagggg cgaagaatgt cgacgacaac cg ttgggtgac tacttgctgt atcgcttaaa	1260
cgaaatcggc attgagcacc tcttcggagt gccagggtgat tacaatctgc aatttctgga	1320
tcatgtaatc gaccaccctc agctgacttg ggtcggctgc actaacgaac ttaacgctgc	1380
ctacgcagct gatggttatg cgcgttgctg tccggctgcg gcaactgctga ccaccttcgg	1440
gg ttggcgaa ctgagcgcta ttaatggcat cgcaggttcc tacgcggagt atctgccggt	1500
aatacatatc gttgggtgcac cgagtcctatc agcccagcag cagggcgacc tgattcacca	1560
ctctcttggc gaaggtgatt tttccagctt cctgaggatg tccaaccgg tgtctgttgc	1620
gcaggctgct ctgactcctg ataacgcatg caaggaaatc gaccgcgtac tggcggaagt	1680
cctcattcag cgtcgtccc gctacctgct gctgtctacc gacgtggctg ctgcgccggc	1740
ggctctgcca caaagcactc tttctttgcc gaccgccccg gatcatcgcg cagttctggc	1800
tgctttcagc gacgctgctg agcagatgct ggctcaggcc aaaagcgtct ctctactggc	1860
ggactttctg gctgatcggt tcggtgttac tcgagcactg gccgcgtggc ttcagcaggt	1920
tccgctaccg cacgccactc tg ttaatggg taaaggcggt ctgagtgaac agcaaccagg	1980
gttcgtgggt acctacgctg gtgcggcatc tatcgattcg acgcgtggcg caatcgaaga	2040
agctggggta attatcggag tgggagttag attttccgac actatcacag caggcttctc	2100
gcagcagatc gacgcccgc gttttataga cattcaacctc ttcttctctc gtattggcga	2160
tcgccagttt gatcattgc cgatgcaggc tgccgtcgca gccctgcatc aactgtgtct	2220
tcgttatcag cagcagtggc ctatcaccgc tcctagcccc cctgcactgc cgccggctgc	2280
tggtagcgag ctgtcccaga acgcattctg gcaggcgatg cagaacttca tccgccctgg	2340
ggacctgttg gtggccgacc aaggtactgc ggcgttcggc gcagcggcg tgcgcttacc	2400
gcagaattgc cagctgcttg tgcagccgct gtggggctca atcggttaca gtctgccggc	2460
cacctttggg gctcagacgg cagatacaga gcgtcgtgta atcctaataca ttggcgatgg	2520
ttcagcgcaa ttaactattc aggaactttc cagtatgatg cgtgacggct tgaaacctat	2580
catctttctc ctgaacaaca acggttacac cg ttgaacgg gcgattcacg gcgcggagca	2640
acgttataac gatatcgctg cttggaattg gacccaactg ccccgaggcg tgagtgttca	2700

ttgcccagcg cagagctggc gagtcgttga aacggtgcag ctgaccgacg taatgaaagt	2760
catcgctgct tctccgcgtc tgagcttgggt agaagttggt ctgcctgcaa tggatgtccc	2820
accgctgctg caagcagtga gtgccgctct gaaccagcgc aactcctctt aatttttaaa	2880
gtaacacacg agctaagaac taaggaggat ttatgcgcag catgcagttt gatgagtacg	2940
gtgcacccct gaaagcgttc tcatatgaag acccgacccc gcaaggggaag gaagtagtcg	3000
ttaggatcga agcctgtgggt gtgtgccact ctgatattca tcttcacgag ggctacttcg	3060
acatgggagg tggcaataaa gctgatgtta ctcgtgctcg cgaactccct ttacattgg	3120
gtcatgaaat cgttggcgaa gtggtagcaa ctggaccagg tgtcaccggc gctaaaccgg	3180
gcgacaaacg tatttgtgtac ccgtggatcg ggtgcggcga ctgcccgaag tgcaacagtg	3240
gtgaggatca gtcctgtgcg cgtccacgta acctgggtgt tcacgttgac ggtggctatt	3300
cgacgcacgt aaagataccg gacgaaaaat tcctgttcgc ctacgatgggt attcctactg	3360
agttagcggg aacctatgct tgcagcggca tcaccgctta tgggtgactg atgaaagcaa	3420
aggaagcggc tgaaagatct ggctacatcg gtctgattgg cgctgggtggc gttggcatgg	3480
ctgggtctgat gctggccaaa gcagcgatcg gggctaaaac tgtagtcttt gatatcgacg	3540
acgcaaaact ggaagctgcg acccgtgccg gggcggatta cgtgttcaac tccgggtgcaa	3600
aagaaacacg caaggaagtt atgaaactaa cgaatggtgg cctgtctgggt gctgttgatt	3660
tcgttggcag cgataaaaagc gctctgtttg gaatcaacgc cttgggtcag aacggcgtgc	3720
tggtcataat tggactgttc ggtggcgcta tgactgttcc ggtacccctg ttcccgtga	3780
aagggatcac cgtacgtggc tcatacgtag gttccctgca agagatgagt gatatgatgg	3840
agttagtctg cgctgggaaa gttcctccga tgccggtaaa aactcggcca ctggacgctg	3900
cctgggaaac cttgaggat ctacgccatg gtaaaatcgt gggccgtgtt gttctgacct	3960
cataagaagg agatatacat atgaccatc aattaagatc gcgcgatatc atcgctctgg	4020
gctttatgac atttgcgttg ttcgtcggcg caggtaacat tattttccct ccaatggctg	4080
gcttgaggc aggcgaaac gtctggactg cggcattcgg cttcctcatt actgccgttg	4140
gcctaccggt attaacggta gtggcgctgg caaaagttgg cggcgggtgtt gacagtctca	4200
gcacgccaat tggtaaagtc gctggcgtag tgctggcaac agtttggttac ctggcgggtg	4260
ggccgctttt tgctacgccg cgtacagcta ccgtttcttt tgaagtgggc attgcgccgc	4320

tgacgggtga ttccgcgctg ccgctgttta ttacagcct ggtctatttc gctatcgta	4380
ttctggtttc gctctatccg ggcaagctgc tggataccgt gggcaacttc cttgcgccgc	4440
tgaattat cgcgctggtc atcctgtctg ttgccgaat tatctggccg gcgggttcta	4500
tcagtacggc gactgaggct tatcaaacg ctgcgttttc taacggcttc gtcaacggct	4560
atctgaccat ggatacgctg ggcgcaatgg tgtttggat cgttattgtt aacgcggcgc	4620
gttctcgtgg cgttaccgaa gcgcgtctgc tgaccgta taccgtctgg gctggcctga	4680
tggcgggtgt tggctgact ctgctgtacc tggcgctgtt ccgtctgggt tcagacagcg	4740
cgtcgctggc cgatcagtct gcaaacggcg cggcgatcct gcatgcttac gttcagcata	4800
cctttggcgg cggcggtagc ttctgtctgg cggcgtaaat cttcatcgcc tgcctggcga	4860
cggcggttgg cctgacctgt gcttgtgcag aattcttcgc ccagtacgta ccgctctctt	4920
atcgtacgct ggtgtttatc ctccggcgct tctcgatggg ggtgtctaac ctccgcttga	4980
gccagctgat tcagatctct gtaccggctg tgaccgcat ttatccgccg tgtatcgac	5040
tggttgtatt aagttttaca cgctcatggg ggcataattc gtcccgctg attgctccgc	5100
cgatgtttat cagcctgctt ttgtgtatc tcgacgggat caaggcatct gcattcagcg	5160
atatcttacc gtcctgggcg cagcgtttac cgctggccga acaaggctctg gcgtgggtta	5220
tgccaacagt ggtgatggcg gttctggcca ttatctggga tcgtgcggca ggtcgtcagg	5280
tgacctcag cgctcactaa tacgcatggc atggatgacc gatggtagtg tgggctctcc	5340
ccatgcgaga gtagggaact gccaggcat	5360

<210> 47

<211> 5360

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 47

gttgacactc tatcattgat agagttattt taccactccc tatcagtgat agagaaaagt	60
gaaatattca cgtaaggagg taatagatgc agatcttcga cactttgcaa tcaatgggcc	120
atgagcaggc tgtcctatgt agcgataaga ccacgggtct gcgcgccatt atcgctatac	180
acgatacatc cttagggccg gcgcttggtg gtaccggtat gtggcagtat gcaactgacg	240

acgatgctat tactgacgca ctccgtctgt ctcggggcat gacctacaaa gctgcggttt	300
ctggcgtaaa tctggggcgtt ggtaaagccg ttatcatcgg aaaccctcac agtgataaaa	360
gcgaagcgct gtttcgcgct tacggcagaa tgggtggaatc ccagcgtggg cgttacatca	420
ccgccgaaga cgttggtact agcgtacgtg atatggagtg gattcgcgcatg gaaaccaa	480
atgtaacggg cgtgggtggc aacggaggct ctggtgaccc ctctccagtt accgctctgg	540
gtgtttactc gggcatgaag gcatgcgcta aatcagtcta tgggtactgat gcgctgagcg	600
gtaaaaggat cgtgggttcag ggcgcgggta acgttgcatc ccatctgggtt cacagtctgg	660
taaaagaagg cgctgtgggtt ttcgtcactg acatctacga agaaaaggcc aaagcattag	720
cggctgaaac gggcgctacc gtgattcgca ccgacgaggt ttttactaca caatgcgata	780
tcttctctcc gaacgctctg ggggccgtcc tgaacgatga aactattccg cagctcacat	840
gcgctatcgt agctgggtgg gcaaacaatc agcttaaaat cgaacaacgt cacgccacgg	900
ctctgcaaga gaaaggcatt ctgtatgcgc cggattacgt aatcaacgcc gggggcctca	960
tgaatgtggc gagcgaagtt gacggctaca accgtgaaaa ggttatgcgc caggctgaag	1020
gtatttacga tattactatg aacatcctaa ataccgcgcg tgagcgtaac atcctgacca	1080
tcgaagcatc caacgcgatt gctgaagagc ggatcaacaa agttcgccat gttcacggga	1140
acttcatcgg ttccccgtct attcgcggag tataataaat caaaacacca aattaaggag	1200
aaattttatg tcgacgacaa ccgttggtga ctacttgctg tatcgcttaa acgaaatcgg	1260
cattgagcac ctcttcggag tgccagggtga ttacaatctg caatttctgg atcatgtaat	1320
cgaccaccct cagctgactt gggtcggctg cactaacgaa cttaacgctg cctacgcagc	1380
tgatggttat gcgcgttgct gtccggctgc ggcactgctg accaccttcg gggttggcga	1440
actgagcgct attaatggca tcgcagggtt ctacgcggag tatctgccgg taatacatat	1500
cgttggtgca ccgagtctat cagcccagca gcaggcgac ctgattcacc actctcttgg	1560
cgaaggtgat ttttccagct tcctgaggat gtcccaaccg gtgtctgttg cgcaggctgc	1620
tctgactcct gataacgcat gcaaggaaat cgaccgcgta ctggcggaag tcctcattca	1680
gcgtcgtccc ggctacctgc tgctgtctac cgacgtggct gctgcgccgg cggctctgcc	1740
acaaagcact ctttctttgc cgaccgcccc ggatcatcgc gcagttctgg ctgctttcag	1800
cgacgctgct gagcagatgc tggctcaggc caaaagcgtc tctctactgg cggactttct	1860

ggctgatcgt ttcggtgtta ctcgagcact ggccgcgtgg cttcagcagg ttccgctacc	1920
gcacgccact ctgttaatgg gtaaaggcgt tctgagtga cagcaaccag ggttcgtggg	1980
tacctacgct ggtgcggcat ctatcgattc gacgcgtggc gcaatcgaag aagctgggggt	2040
aattatcgga gtgggagtta gattttccga cactatcaca gcaggcttct cgagcagat	2100
cgacgccgc cgttttatag acattcaacc cttcttctct cgtattggcg atcgccagtt	2160
tgatcacttg ccgatgcagg ctgccgtcgc agccctgcat caactgtgtc ttcgttatca	2220
gcagcagtgg tctatcaccg ctctagccc gcctgcactg ccgccggctg ctggtagcga	2280
gctgtcccag aacgcattct ggcaggcgat gcagaacttc atccgccctg gggacctgtt	2340
ggtggccgac caaggtactg cggcgttcgg cgcagcggcg ctgcgcttac cgcagaattg	2400
ccagctgctt gtgcagccgc tgtggggctc aatcggttac agtctgccgg ccacctttgg	2460
tgctcagacg gcagatacag agcgtcgtgt aatcctaadc attggcgatg gttcagcgca	2520
attaactatt caggaacttt ccagtatgat gcgtgacggc ttgaaaccta tcatctttct	2580
cctgaacaac aacggttaca ccgttgaacg ggcgattcac ggccgaggagc aacgttataa	2640
cgatatcgct gcttgaatt ggacccaact gcccaggcg ctgagtgttc attgcccagc	2700
gcagagctgg cgagtcgttg aaacggtgca gctgaccgac gtaatgaaag tcatcgctgc	2760
ttctccgcgt ctgagcttgg tagaagttgt tctgcctgca atggatgtcc caccgctgct	2820
gcaagcagtg agtgccgctc tgaaccagcg caactcctct taatttttaa agtaacacac	2880
gagctaagaa ctaaggagga tttatgcgca gcatgcagtt tgatgagtac ggtgcacccc	2940
tgaaagcgtt tcatatgaa gaccgaccc cgcaaggga ggaagtagtc gttaggatcg	3000
aagcctgtgg tgtgtgccac tctgatattc atcttcacga gggctacttc gacatgggcg	3060
gtggcaataa agctgatgtt actcgtgctc gcgaactccc ttttacattg ggtcatgaaa	3120
tcgttggcga agtggttagca actggaccag gtgtcaccgg cgctaaaccg ggcgacaaac	3180
gtattgtgta cccgtggatc ggggtcggcg actgcccga atgcaacagt ggtgaggatc	3240
agtcctgtgc gcgtccacgt aacctgggtg ttcacgttga cgggtggctat tcgacgcacg	3300
taaagatacc ggacgaaaaa ttcctgttcg cctacgatgg tattcctact gagttagcgg	3360
gaacctatgc ttgcagcggc atcaccgctt atggtgcact gatgaaagca aaggaagcgg	3420
ctgaaagatc tggctacatc ggtctgattg gcgctgggtg cgttggcatg gctggtctga	3480

tgctggccaa agcagcgatc ggggctaaaa ctgtagtctt tgatatcgac gacgcaaaac	3540
tggaagctgc gacccgtgcc ggggcggatt acgtgttcaa ctccggtgca aaagaaacac	3600
gcaaggaagt tatgaaacta acgaatggtg gcctgtctgg tgctgttgat ttcgttggca	3660
gcgataaaaag cgctctgttt ggaatcaacg ccttgggtca gaacggcgtg ctggtcataa	3720
ttggactgtt cgggtggcgct atgactgttc cggtagccct gttcccgtg aaagggatca	3780
ccgtacgtgg ctcatagta ggttcctgc aagagatgag tgatatgatg gagttagttc	3840
gcgctgggaa agttcctccg atgccggtaa aaactcggcc actggacgct gcctgggaaa	3900
cccttgagga tctacgcat ggtaaaatcg tgggccgtgt tgttctgacc ccataagaag	3960
gagatataca tatgacccat caattaagat cgcgcgatat catcgctctg ggctttatga	4020
catttgcggtt gttcgtcggc gcaggtaca ttattttccc tccaatggtc ggcttgcagg	4080
caggcgaaca cgtctggact gcggcattcg gcttcctcat tactgccgtt ggcttaccgg	4140
tattaacggt agtggcgctg gcaaaagttg gcggcgggtg tgacagtctc agcacgcaa	4200
ttggtaaagt cgctggcgta ctgctggcaa cagtttgta cctggcggtg gggccgcttt	4260
ttgctacgcc gcgtacagct accgtttctt ttgaagtggg cattgcgccg ctgacgggtg	4320
attccgcgct gccgctgttt atttacagcc tggctctattt cgctatcggtt attctgggtt	4380
cgctctatcc gggcaagctg ctggataccg tgggcaactt ccttgcgccg ctgaaaatta	4440
tcgcgctggg catcctgtct gttgccgcaa ttatctggcc ggcggttct atcagtacgg	4500
cgactgaggc ttatcaaaac gctgcgtttt ctaacggctt cgtaacggc tatctgacca	4560
tggatacgct gggcgcaatg gtgtttgta tcgttattgt taacgcggcg cgttctcgtg	4620
gcgttaccga agcgcgtctg ctgacccgtt ataccgtctg ggctggcctg atggcgggtg	4680
ttggtctgac tctgctgtac ctggcgctgt tccgtctggg ttcagacagc gcgtcgctgg	4740
tcgatcagtc tgcaaacggt gcggcgatcc tgcattgcta cgttcagcat acctttggcg	4800
gcggcggtag cttcctgctg gcggcggtta tcttcatcgc ctgcctggtc acggcgggtg	4860
gcctgacctg tgcttgtgca gaattcttcg ccagtagct accgctctct tatcgtacgc	4920
tgggtgttat cctcggcggc ttctcgatgg tgggtgtctaa cctcggcttg agccagctga	4980
ttcagatctc tgtaccggtg ctgaccgcca tttatccgcc gtgtatcgca ctggttgat	5040
taagttttac acgctcatgg tggcataatt cgtcccgct gattgctccg ccgatgttta	5100

tcagcctgct ttttgggtatt ctcgacggga tcaaggcatc tgcattcagc gatattcttac	5160
cgtcctgggc gcagcgttta ccgctggccg aacaaggctt ggctgtggtta atgccaacag	5220
tggatgatggg ggttctggcc attatctggg atcgtgcggc aggtcgtcag gtgacctcca	5280
gcgctcacta atacgcatgg catggatgac cgatggtagt gtgggctctc cccatgagag	5340
agtagggaac tgccaggcat	5360

<210> 48

<211> 5376

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 48

gttgacactc tatcattgat agagttatct taccactccc tatcagtgat agagaaaagt	60
gaaatattca cgtaaggagg taatagatgc agatcttcga cactttgcaa tcaatgggcc	120
atgagcaggt tgtcctatgt agcgataaga ccacgggtct gcgcgccatt atcgctatac	180
acgatacatc cttagggccg gcgcttggtg gtacccgtat gtggcagtat gcaactgacg	240
acgatgctat tactgacgca ctccgtctgt ctcggggcat gacctacaaa gctgcggttt	300
ctggcgtaaa tctgggcggg ggtaaagccg ttatcatcgg aaaccctcac agtgataaaa	360
gcgaagcgct gtttcgcgt tacggcagaa tgggtggaatc ccagcgtggg cgttacatca	420
ccgccgaaga cgttggtact agcgtacgtg atatggagtg gattcgcatt gaaaccaa	480
atgtaacggg cgtgggtggc aacggaggct ctggtgacct ctctccagtt accgctctgg	540
gtgtttactc gggcatgaag gcatgcgcta aatcagtcta tgggtactgat gcgctgagcg	600
gtaaaaggat cgtgggttcag ggcgcggtta acgttgcatc ccatctgggt cacagtctgg	660
taaaagaagg cgctgtggtt ttcgtcactg acatctacga agaaaaggcc aaagcattag	720
cggctgaaac gggcgctacc gtgattcgca ccgacgaggt ttttactaca caatgcgata	780
tcttctctcc gaacgctctg ggggcccgtc tgaacgatga aactattccg cagctcacat	840
gcgctatcgt agctgggtgg gcaaacaatc agcttaaaat cgaacaacgt cacgccacgg	900
ctctgcaaga gaaaggcatt ctgtatgcgc cggattacgt aatcaacgcc gggggcctca	960

tgaatgtggc gagcgaagtt gacggctaca accgtgaaaa ggttatgcg	caggctgaag	1020
gtattttacga tattactatg aacatcctaa ataccgcgcg tgagcgtaac	atcctgacca	1080
tcgaagcatc caacgcgatt gctgaagagc ggatcaacaa agttcgccat	gttcacggga	1140
acttcacggtt ttccccgtct attcgcggag tataataaat caaaacacca	aattaaggag	1200
aaattttatg tcgacgacaa ccgttggtga ctacttgctg tatcgcttaa	acgaaatcgg	1260
cattgagcac ctcttcggag tgccagggtga ttacaatctg caatttctgg	atcatgtaat	1320
cgaccaccct cagctgactt gggtcggctg cactaacgaa cttaacgctg	cctacgcagc	1380
tgatggttat gcgcgttgct gtccggctgc ggcaactgctg accaccttcg	gggttggcga	1440
actgagcgct attaatggca tcgcagggtt ctacgcggag tatctgccgg	taatacatat	1500
cgttggtgca ccgagtctat cagcccagca gcagggcgac ctgattcacc	actctcttgg	1560
cgaaggatgat ttttccagct tcctgaggat gtcccaaccg gtgtctgttg	cgcaggctgc	1620
tctgactcct gataacgcat gcaaggaaat cgaccgcgta ctggcgggaag	tcctcattca	1680
gcgtcgtccc ggctacctgc tgctgtctac cgacgtggct gctgcgccgg	cggctctgcc	1740
acaaagcact ctttctttgc cgaccgcccc ggatcatcgc gcagttcttg	ctgctttcag	1800
cgacgctgct gagcagatgc tggctcaggc caaaagcgtc tctctactgg	cggactttct	1860
ggctgatcgt ttcggtgtta ctcgagcact ggccgcgtgg cttcagcagg	ttccgctacc	1920
gcacgccact ctgttaatgg gtaaaggcgt tctgagtga cagcaaccag	ggttcgtggg	1980
tacctacgct ggtgcggcat ctatcgattc gacgcgtggc gcaatcgaag	aagctgggggt	2040
aattatcgga gtgggagtta gattttccga cactatcaca gcaggcttct	cgcagcagat	2100
cgacgcccgc cgttttatag acattcaacc cttcttctct cgtattggcg	atcgccagtt	2160
tgatcacttg ccgatgcagg ctgccgtcgc agccctgcat caactgtgtc	ttcgttatca	2220
gcagcagtgg tctatcaccg ctctagccc gcctgcactg ccgccggctg	ctggtagcga	2280
gctgtcccag aacgcattct ggcaggcgat gcagaacttc atccgccctg	gggacctgtt	2340
ggtggccgac caaggtactg cggcgttcgg cgcagcggcg ctgcgcttac	cgcagaattg	2400
ccagctgctt gtgcagccgc tgtggggctc aatcggttac agtctgccgg	ccacctttgg	2460
tgctcagacg gcagatacag agcgtcgtgt aatcctaadc attggcgatg	gttcagcgca	2520
attaactatt caggaacttt ccagtatgat gcgtgacggc ttgaaaccta	tcattcttct	2580

cctgaacaac aacggttaca ccgttgaacg ggcgattcac ggcgcggagc aacgttataa	2640
cgatatcgct gcttgggaatt ggacccaact gccccaggcg ctgagtgttc attgcccagc	2700
gcagagctgg cgagtcgttg aaacggtgca gctgaccgac gtaatgaaag tcatcgctgc	2760
ttctccgcgt ctgagcttgg tagaagttgt tctgcctgca atggatgtcc caccgctgct	2820
gcaagcagtg agtgccgctc tgaaccagcg caactcctct taattggata aagaataagg	2880
aggtatttta tgacagcgga gcagcaaaat ggggtatccg actcacgccg tttcgaattt	2940
caggaatttg gtggccctat cgccccacag acctatcagc tccccgcacc ggctagcgat	3000
gaagttttgt taaaggtgaa ctactgcggt gtctgtcaca gtgatgttca tcttcacgac	3060
ggctacttcg agctgggtgg cgataaacgt ctgaacttcg ctatgccgct gccgctgacg	3120
ctgggtcacg aagtaattgg caccgtttgt gctgtcggcg accaggttac tgggtgtaaaa	3180
ccggggggacc agcgactgat ctatccgtgg ataggttgcg gaaaatgcgg cgcgtgtcaa	3240
aaaggagaag aaaacctgtg cgttactcct gcacatctgg gcgtgaacaa gccgggcggt	3300
tacgctgata acatcgttgt accccattct cgctaccttc tggacatttc gggctctgaac	3360
ccgggtgatg ccgctaccct cgcgtgctcc ggcctgacca ctttcagcgc gatcaacaaa	3420
gtgttgccgc ttgcagatga ccagtggatt gttgttatcg gttgtggtgg cctcggccag	3480
atggcgctgc gtatcctgca agctatggga attggcaatg ttatcggtat tgacctgtct	3540
gaagagaaac ggaaactggc tcatgaaagc ggtgcacgtc actccttcga tccaaacact	3600
ccgaagctga accgcgtggt cgccgaaacc tgcccgggta cggtacaggc cgcgttagac	3660
tttgtgggca atgagcaaac tgctcagctg gcaactgtct tgcttggaag aggtggcaaa	3720
tatgttcctg tcgggctgca cggcggcgag ctgcgttacc cattgccgat catcacgaac	3780
aaagctgtaa gtatcatcgg ttctttacgtt ggtaccctga aagaactgga agacttagtt	3840
gctttcgcca aggaaaagaa tctgccgcca attcatattg aacaccgccc gctggaatcg	3900
gcggctcagg ccgtagagga cctggaaaaa ggacaggttg ctgggcgtgt tatcctggat	3960
gcaggtaact aagaaggaga tatacatatg acccatcaat taagatcgcg cgatatcatc	4020
gctctgggct ttatgacatt tgcgttggtt gtcggcgcag gtaacattat tttccctcca	4080
atggtcggct tgcaggcagg cgaacacgtc tggactgcgg cattcggctt cctcattact	4140
gccgttggcc taccggtatt aacggtagtg gcgctggcaa aagttggcgg cgggtgttgac	4200

agtctcagca cgccaattgg taaagtcgct ggcgtactgc tggcaacagt ttgttacctg	4260
gcgggtggggc cgctttttgc tacgccgcgt acagctaccg tttcttttga agtgggcatt	4320
gcgccgctga cgggtgattc cgcgctgccg ctgtttattht acagcctggg ctatttcgct	4380
atcgttattc tggtttcgct ctatccgggc aagctgctgg ataccgtggg caacttcctt	4440
gcgccgctga aaattatcgc gctggtcatt ctgtctgttg ccgcaattat ctggccggcg	4500
ggttctatca gtacggcgac tgaggcttat caaaacgctg cgttttctaa cggcttcgct	4560
aacggctatc tgaccatgga tacgctgggc gcaatgggtg ttggtatcgt tattgttaac	4620
gcggcgcgtht ctcgtagcggt taccgaagcg cgtctgctga cccgttatac cgtctgggct	4680
ggcctgatgg cgggtgttgg tctgactctg ctgtacctgg cgctgttccg tctgggttca	4740
gacagcgcggt cgctggctga tcagtctgca aacgggtcgg cgatcctgca tgcttacgtht	4800
cagcatacct ttggcggcggt cggtagcttht ctgctggcggt cgthaatctt catcgctgct	4860
ctggtcacgg cggttggcct gacctgtgct tgtgcagaat tcttcgcca gtacgtaccg	4920
ctctcttatt gtacgctgggt gtttatctct ggccgcttct cgatgggtgg gtctaacctc	4980
ggcttgagcc agctgattca gatctctgta ccgggtgctga ccgccattta tccgccgtgt	5040
atcgcactgg ttgtattaag ttttacacgc tcatgggtggc ataattcgtc ccgcgtgatt	5100
gctccgccga tgthttatcag cctgctthttt ggtattctcg acgggatcaa ggcatctgca	5160
ttcagcgata tcttaccgtc ctgggcgcag cgthttaccgc tggccgaaca aggtctggcg	5220
tggthaatgc caacagtgggt gatgggtggt ctggccatta tctgggatcg tgcggcaggt	5280
cgtcaggtga cctccagcgc tcactaatac gcatggcatg gatgaccgat ggtagtgtgg	5340
gctctcccca tgcgagagta gggaactgcc aggcatt	5376

<210> 49

<211> 360

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 49

Met	Asn	Ile	Phe	Lys	Lys	Met	Glu	Glu	Phe	Asn	Tyr	Glu	Gln	Leu	Val
1				5					10					15	

Tyr Phe Tyr Asp Ser Glu Thr Glu Leu Lys Gly Ile Thr Cys Ile His
20 25 30

Asn Thr Thr Leu Gly Pro Ala Leu Gly Gly Thr Arg Leu Trp Asn Tyr
35 40 45

Asn Ser Glu Glu Asp Ala Val Glu Asp Val Ile Arg Leu Ala Arg Gly
50 55 60

Met Thr Tyr Lys Ala Ala Cys Ala Gly Leu Asn Leu Gly Gly Gly Lys
65 70 75 80

Thr Val Leu Ile Gly Asp Ala Lys Lys Ile Lys Ser Glu Ser Tyr Phe
85 90 95

Arg Gly Leu Gly Arg Tyr Val Gln Ser Leu Asn Gly Arg Tyr Ile Thr
100 105 110

Ala Glu Asp Val Asn Thr Ser Thr Lys Asp Met Ala Tyr Val Ala Met
115 120 125

Glu Thr Asp Tyr Val Val Gly Leu Gly Gly Lys Ser Gly Asn Pro Ser
130 135 140

Pro Val Thr Ala Tyr Gly Ala Phe Met Gly Ile Lys Ala Ala Leu Met
145 150 155 160

Lys Lys Phe Glu Asp Ser Ser Ile Glu Gly Arg Thr Phe Ala Val Gln
165 170 175

Gly Ala Gly Gln Thr Gly Tyr Tyr Leu Ile Asp Tyr Leu Leu Gly Asn
180 185 190

Asn Lys Phe Lys Glu Lys Ala Lys Lys Ile Tyr Phe Thr Glu Ile Asn
195 200 205

Glu Ser Tyr Ile Glu Arg Met Asn Lys Glu His Pro Glu Val Glu Phe
210 215 220

Ile Ser Pro Asp Lys Ile Tyr Ser Leu Glu Val Asp Val Phe Val Pro

225	230										235			240		
Cys	Ala	Leu	Gly	Lys	Ile	Val	Asn	Asp	Lys	Thr	Ile	Asp	Glu	Phe	Lys	
				245					250					255		
Cys	Pro	Ile	Ile	Ala	Gly	Thr	Ala	Asn	Asn	Val	Leu	Glu	Arg	Glu	Ala	
			260					265					270			
His	Gly	Asn	Met	Leu	Lys	Glu	Arg	Gly	Ile	Leu	Tyr	Ala	Pro	Asp	Tyr	
		275					280					285				
Val	Ile	Asn	Ala	Gly	Gly	Leu	Ile	Asn	Val	Tyr	His	Glu	Leu	Asn	Gly	
	290					295					300					
Tyr	Asn	Lys	Glu	Asn	Ala	Ile	Leu	Glu	Val	Glu	Leu	Ile	Tyr	Asp	Arg	
305					310					315					320	
Leu	Leu	Glu	Ile	Phe	Asn	Ile	Ala	Asp	Ser	Leu	Asn	Ile	Ser	Thr	Asn	
				325					330					335		
Ile	Ala	Ala	Asn	Glu	Phe	Ala	Glu	Lys	Arg	Ile	Lys	Gln	Ile	Lys	Ser	
			340					345					350			
Leu	Lys	Asn	Asn	Phe	Ile	Lys	Arg									
		355					360									
<210> 50																
<211> 360																
<212> PRT																
<213> Artificial Sequence																
<220>																
<223> Description of Artificial Sequence: Synthetic polypeptide																
<400> 50																
Met	Asn	Ile	Phe	Lys	Lys	Met	Glu	Glu	Phe	Asn	Tyr	Glu	Gln	Leu	Val	
1				5					10					15		
Tyr	Phe	Tyr	Asp	Ser	Glu	Thr	Glu	Leu	Lys	Gly	Ile	Thr	Cys	Ile	His	
			20					25					30			
Asn	Thr	Thr	Leu	Gly	Pro	Ala	Leu	Gly	Gly	Thr	Arg	Leu	Trp	Asn	Tyr	

35

40

45

Asn Ser Glu Glu Asp Ala Val Glu Asp Val Ile Arg Leu Ala Arg Gly
 50 55 60

Met Thr Tyr Lys Ala Ala Cys Ala Gly Leu Asn Leu Gly Gly Gly Lys
 65 70 75 80

Thr Val Leu Ile Gly Asp Ala Lys Lys Ile Lys Ser Glu Ser Tyr Phe
 85 90 95

Arg Gly Leu Gly Arg Tyr Val Gln Ser Leu Asn Gly Arg Tyr Ile Thr
 100 105 110

Ala Glu Asp Val Asn Thr Ser Thr Lys Asp Met Ala Tyr Val Ala Met
 115 120 125

Glu Thr Asp Tyr Val Val Gly Leu Gly Gly Lys Ser Gly Asn Pro Ser
 130 135 140

Pro Val Thr Ala Tyr Gly Ala Phe Met Gly Ile Lys Ala Ala Leu Met
 145 150 155 160

Lys Lys Phe Glu Asp Ser Ser Ile Glu Gly Arg Thr Phe Ala Val Gln
 165 170 175

Gly Ala Gly Gln Thr Gly Tyr Tyr Leu Ile Asp Tyr Leu Leu Gly Asn
 180 185 190

Asn Lys Phe Lys Glu Lys Ala Lys Lys Ile Tyr Phe Thr Glu Ile Asn
 195 200 205

Glu Ser Tyr Ile Glu Arg Met Asn Lys Glu His Pro Glu Val Glu Phe
 210 215 220

Ile Ser Pro Asp Lys Ile Tyr Ser Leu Glu Val Asp Val Phe Val Pro
 225 230 235 240

Cys Ala Leu Gly Lys Ile Val Asn Asp Lys Thr Ile Asp Glu Phe Lys
 245 250 255

Cys Pro Ile Ile Ala Gly Thr Ala Asn Asn Val Leu Glu Arg Glu Ala
260 265 270

His Gly Asn Met Leu Lys Glu Arg Gly Ile Leu Tyr Ala Pro Asp Tyr
275 280 285

Val Ile Asn Ala Gly Gly Leu Ile Asn Val Tyr His Glu Leu Asn Gly
290 295 300

Tyr Asn Lys Glu Asn Ala Ile Leu Glu Val Glu Leu Ile Tyr Asp Arg
305 310 315 320

Leu Leu Glu Ile Phe Asn Ile Ala Asp Ser Leu Asn Ile Ser Thr Asn
325 330 335

Ile Ala Ala Asn Glu Phe Ala Glu Lys Arg Ile Lys Gln Ile Lys Ser
340 345 350

Leu Lys Asn Asn Phe Ile Lys Arg
355 360

<210> 51

<211> 360

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 51

Met Asn Ile Phe Lys Lys Met Glu Glu Phe Asn Tyr Glu Gln Leu Val
1 5 10 15

Tyr Phe Tyr Asp Ser Glu Thr Glu Leu Lys Gly Ile Thr Cys Ile His
20 25 30

Asn Thr Thr Leu Gly Pro Ala Leu Gly Gly Thr Arg Leu Trp Asn Tyr
35 40 45

Asn Ser Glu Glu Asp Ala Val Glu Asp Val Ile Arg Leu Ala Arg Gly
50 55 60

Met Thr Tyr Lys Ala Ala Cys Ala Gly Leu Asn Leu Gly Gly Gly Lys
65 70 75 80

Thr Val Leu Ile Gly Asp Ala Lys Lys Ile Lys Ser Glu Ser Tyr Phe
85 90 95

Arg Gly Leu Gly Arg Tyr Val Gln Ser Leu Asn Gly Arg Tyr Ile Thr
100 105 110

Ala Glu Asp Val Asn Thr Ser Thr Lys Asp Met Ala Tyr Val Ala Met
115 120 125

Glu Thr Asp Tyr Val Val Gly Leu Gly Gly Lys Ser Gly Asn Pro Ser
130 135 140

Pro Val Thr Ala Tyr Gly Ala Phe Met Gly Ile Lys Ala Ala Leu Met
145 150 155 160

Lys Lys Phe Glu Asp Ser Ser Ile Glu Gly Arg Thr Phe Ala Val Gln
165 170 175

Gly Ala Gly Gln Thr Gly Tyr Tyr Leu Ile Asp Tyr Leu Leu Gly Asn
180 185 190

Asn Lys Phe Lys Glu Lys Ala Lys Lys Ile Tyr Phe Thr Glu Ile Asn
195 200 205

Glu Ser Tyr Ile Glu Arg Met Asn Lys Glu His Pro Glu Val Glu Phe
210 215 220

Ile Ser Pro Asp Lys Ile Tyr Ser Leu Glu Val Asp Val Phe Val Pro
225 230 235 240

Cys Ala Leu Gly Lys Ile Val Asn Asp Lys Thr Ile Asp Glu Phe Lys
245 250 255

Cys Pro Ile Ile Ala Gly Thr Ala Asn Asn Val Leu Glu Arg Glu Ala
260 265 270

His Gly Asn Met Leu Lys Glu Arg Gly Ile Leu Tyr Ala Pro Asp Tyr

275

280

285

Val Ile Asn Ala Gly Gly Leu Ile Asn Val Tyr His Glu Leu Asn Gly
 290 295 300

Tyr Asn Lys Glu Asn Ala Ile Leu Glu Val Glu Leu Ile Tyr Asp Arg
 305 310 315 320

Leu Leu Glu Ile Phe Asn Ile Ala Asp Ser Leu Asn Ile Ser Thr Asn
 325 330 335

Ile Ala Ala Asn Glu Phe Ala Glu Lys Arg Ile Lys Gln Ile Lys Ser
 340 345 350

Leu Lys Asn Asn Phe Ile Lys Arg
 355 360

<210> 52

<211> 363

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
 polypeptide

<400> 52

Met Val Glu Ile Gln Ala Leu Pro Glu Thr Ser Ile Phe Gly Gln Ile
 1 5 10 15

Ala Asp His Gln His Glu Gln Val Val Phe Cys His Asp His Glu Thr
 20 25 30

Gly Leu Arg Ala Ile Ile Gly Ile His Asn Thr Val Leu Gly Pro Ala
 35 40 45

Leu Gly Gly Thr Arg Met Trp His Tyr Ala Thr Glu Ala Glu Ala Leu
 50 55 60

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

Ser Gly Leu Asn Leu Gly Gly Gly Lys Ala Val Ile Ile Gly Asp Ala

85

90

95

Lys Thr Ile Lys Thr Glu Ala Leu Leu Arg Lys Phe Gly Arg Phe Val
 100 105 110

Gln Asn Leu Asn Gly Lys Tyr Ile Thr Ala Glu Asp Val Asn Met Thr
 115 120 125

Thr Lys Asp Met Glu Tyr Ile Arg Met Glu Thr Lys His Val Ala Gly
 130 135 140

Leu Pro Glu Ser Met Gly Gly Ser Gly Asp Pro Ser Pro Val Thr Ala
 145 150 155 160

Tyr Gly Thr Tyr Met Gly Met Lys Ala Ala Ala Lys Lys Ala Phe Gly
 165 170 175

Ser Asp Ser Leu Ala Gly Lys Arg Ile Ala Val Gln Gly Val Gly His
 180 185 190

Val Gly Thr Tyr Leu Leu Glu His Leu Thr Lys Glu Gly Ala Gln Ile
 195 200 205

Val Leu Thr Asp Tyr Tyr Lys Glu Arg Ala Glu Glu Ala Gly Ala Arg
 210 215 220

Phe Gly Ala Gln Val Val Gly Leu Asp Asp Ile Tyr Asp Gln Glu Val
 225 230 235 240

Asp Ile Tyr Ser Pro Cys Ala Leu Gly Ala Thr Ile Asn Asp Asp Thr
 245 250 255

Ile Asp Arg Leu Arg Cys Ala Val Val Ala Gly Cys Ala Asn Asn Gln
 260 265 270

Leu Lys Glu Glu Asn Val His Gly Pro Ala Leu Val Glu Arg Gly Ile
 275 280 285

Val Tyr Ala Pro Asp Phe Leu Ile Asn Ala Gly Gly Leu Ile Asn Val
 290 295 300

Tyr Ser Glu Val Thr Gly Ser Thr Arg Gln Gly Ala Leu Thr Gln Thr
305 310 315 320

Glu Lys Ile Tyr Asp Tyr Thr Leu Gln Val Leu Glu Lys Ala Ala Ala
325 330 335

Glu Gly Leu His Pro Gln Gln Ala Ala Ile Arg Gln Ala Glu Gln Arg
340 345 350

Ile Ala Ala Ile Gly Lys Val Lys Ser Thr Tyr
355 360

<210> 53

<211> 363

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 53

Met Val Glu Ile Gln Ala Leu Pro Glu Thr Ser Ile Phe Gly Gln Ile
1 5 10 15

Ala Asp His Gln His Glu Gln Val Val Phe Cys His Asp His Glu Thr
20 25 30

Gly Leu Arg Ala Ile Ile Gly Ile His Asn Thr Val Leu Gly Pro Ala
35 40 45

Leu Gly Gly Thr Arg Met Trp His Tyr Ala Thr Glu Ala Glu Ala Leu
50 55 60

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

Ser Gly Leu Asn Leu Gly Gly Gly Lys Ala Val Ile Ile Gly Asp Ala
85 90 95

Lys Thr Ile Lys Thr Glu Ala Leu Leu Arg Lys Phe Gly Arg Phe Val
100 105 110

Gln Asn Leu Asn Gly Lys Tyr Ile Thr Ala Glu Asp Val Asn Met Thr
115 120 125

Thr Lys Asp Met Glu Tyr Ile Arg Met Glu Thr Lys His Val Ala Gly
130 135 140

Leu Pro Glu Ser Met Gly Gly Ser Gly Asp Pro Ser Pro Val Thr Ala
145 150 155 160

Tyr Gly Thr Tyr Met Gly Met Lys Ala Ala Ala Lys Lys Ala Phe Gly
165 170 175

Ser Asp Ser Leu Ala Gly Lys Arg Ile Ala Val Gln Gly Val Gly His
180 185 190

Val Gly Thr Tyr Leu Leu Glu His Leu Thr Lys Glu Gly Ala Gln Ile
195 200 205

Val Leu Thr Asp Tyr Tyr Lys Glu Arg Ala Glu Glu Ala Gly Ala Arg
210 215 220

Phe Gly Ala Gln Val Val Gly Leu Asp Asp Ile Tyr Asp Gln Glu Val
225 230 235 240

Asp Ile Tyr Ser Pro Cys Ala Leu Gly Ala Thr Ile Asn Asp Asp Thr
245 250 255

Ile Asp Arg Leu Arg Cys Ala Val Val Ala Gly Cys Ala Asn Asn Gln
260 265 270

Leu Lys Glu Glu Asn Val His Gly Pro Ala Leu Val Glu Arg Gly Ile
275 280 285

Val Tyr Ala Pro Asp Phe Leu Ile Asn Ala Gly Gly Leu Ile Asn Val
290 295 300

Tyr Ser Glu Val Thr Gly Ser Thr Arg Gln Gly Ala Leu Thr Gln Thr
305 310 315 320

Glu Lys Ile Tyr Asp Tyr Thr Leu Gln Val Leu Glu Lys Ala Ala Ala

325

330

335

Glu Gly Leu His Pro Gln Gln Ala Ala Ile Arg Gln Ala Glu Gln Arg
 340 345 350

Ile Ala Ala Ile Gly Lys Val Lys Ser Thr Tyr
 355 360

<210> 54

<211> 363

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
 polypeptide

<400> 54

Met Val Glu Ile Gln Ala Leu Pro Glu Thr Ser Ile Phe Gly Gln Ile
 1 5 10 15

Ala Asp His Gln His Glu Gln Val Val Phe Cys His Asp His Glu Thr
 20 25 30

Gly Leu Arg Ala Ile Ile Gly Ile His Asn Thr Val Leu Gly Pro Ala
 35 40 45

Leu Gly Gly Thr Arg Met Trp His Tyr Ala Thr Glu Ala Glu Ala Leu
 50 55 60

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

Ser Gly Leu Asn Leu Gly Gly Gly Lys Ala Val Ile Ile Gly Asp Ala
 85 90 95

Lys Thr Ile Lys Thr Glu Ala Leu Leu Arg Lys Phe Gly Arg Phe Val
 100 105 110

Gln Asn Leu Asn Gly Lys Tyr Ile Thr Ala Glu Asp Val Asn Met Thr
 115 120 125

Thr Lys Asp Met Glu Tyr Ile Arg Met Glu Thr Lys His Val Ala Gly

130

135

140

Leu Pro Glu Ser Met Gly Gly Ser Gly Asp Pro Ser Pro Val Thr Ala
 145 150 155 160

Tyr Gly Thr Tyr Met Gly Met Lys Ala Ala Ala Lys Lys Ala Phe Gly
 165 170 175

Ser Asp Ser Leu Ala Gly Lys Arg Ile Ala Val Gln Gly Val Gly His
 180 185 190

Val Gly Thr Tyr Leu Leu Glu His Leu Thr Lys Glu Gly Ala Gln Ile
 195 200 205

Val Leu Thr Asp Tyr Tyr Lys Glu Arg Ala Glu Glu Ala Gly Ala Arg
 210 215 220

Phe Gly Ala Gln Val Val Gly Leu Asp Asp Ile Tyr Asp Gln Glu Val
 225 230 235 240

Asp Ile Tyr Ser Pro Cys Ala Leu Gly Ala Thr Ile Asn Asp Asp Thr
 245 250 255

Ile Asp Arg Leu Arg Cys Ala Val Val Ala Gly Cys Ala Asn Asn Gln
 260 265 270

Leu Lys Glu Glu Asn Val His Gly Pro Ala Leu Val Glu Arg Gly Ile
 275 280 285

Val Tyr Ala Pro Asp Phe Leu Ile Asn Ala Gly Gly Leu Ile Asn Val
 290 295 300

Tyr Ser Glu Val Thr Gly Ser Thr Arg Gln Gly Ala Leu Thr Gln Thr
 305 310 315 320

Glu Lys Ile Tyr Asp Tyr Thr Leu Gln Val Leu Glu Lys Ala Ala Ala
 325 330 335

Glu Gly Leu His Pro Gln Gln Ala Ala Ile Arg Gln Ala Glu Gln Arg
 340 345 350

Ile Ala Ala Ile Gly Lys Val Lys Ser Thr Tyr
355 360

<210> 55

<211> 366

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 55

Met Ile Phe Glu Thr Ile Ser Thr Ser Asn His Glu Glu Val Val Tyr
1 5 10 15

Cys His Asn Lys Asp Ala Gly Leu Lys Ala Ile Ile Ala Ile His Asn
20 25 30

Thr Val Leu Gly Pro Ala Leu Gly Gly Thr Arg Met Trp Pro Tyr Ala
35 40 45

Ser Glu Glu Glu Ala Leu Lys Asp Val Leu Arg Leu Ser Arg Gly Met
50 55 60

Thr Tyr Lys Ala Ala Val Ser Gly Leu Asn Leu Gly Gly Gly Lys Ala
65 70 75 80

Val Ile Trp Gly Asp Pro Asn Lys Asp Lys Ser Glu Ala Leu Phe Arg
85 90 95

Ala Phe Gly Arg Phe Val Asn Ser Leu Gly Gly Arg Tyr Ile Thr Ala
100 105 110

Glu Asp Val Gly Ile Asp Val Asn Asp Met Glu Tyr Val Leu Arg Glu
115 120 125

Thr Asp Tyr Val Thr Gly Val His Gln Val His Gly Gly Ser Gly Asp
130 135 140

Pro Ser Pro Phe Thr Ala Tyr Gly Thr Leu Gln Gly Leu Met Ala Ala
145 150 155 160

Leu Gln Val Lys Phe Gly Asn Glu Asp Val Gly Asn Tyr Ser Tyr Ala
165 170 175

Val Gln Gly Val Gly His Val Gly Met Glu Phe Val Lys Leu Leu Arg
180 185 190

Glu Arg Gly Ala Lys Val Phe Val Thr Asp Ile Asn Lys Asp Ala Val
195 200 205

Gln Arg Ala Val Asp Glu Phe Gly Cys Glu Ala Val Ala Leu Asp Glu
210 215 220

Ile Tyr Asp Val Asp Cys Asp Val Tyr Ser Pro Thr Ala Leu Gly Gly
225 230 235 240

Thr Val Asn Asp Lys Thr Leu Pro Arg Leu Lys Cys Lys Val Ile Cys
245 250 255

Gly Ala Ala Asn Asn Gln Leu Ala Asn Asp Glu Ile Gly Val Glu Leu
260 265 270

Glu Lys Lys Gly Ile Leu Tyr Ala Pro Asp Tyr Ala Val Asn Ala Gly
275 280 285

Gly Leu Met Asn Val Ser Leu Glu Ile Asp Gly Tyr Asn Arg Glu Arg
290 295 300

Ala Met Arg Met Met Arg Thr Ile Tyr Tyr Asn Leu Gly Arg Ile Phe
305 310 315 320

Glu Ile Ser Lys Arg Asp Gly Ile Pro Thr Phe Arg Ala Ala Asp Arg
325 330 335

Met Ala Glu Glu Arg Ile Thr Ala Ile Gly Lys Leu Arg Leu Pro His
340 345 350

Leu Gly Ala Ala Ala Pro Arg Phe Gln Gly Arg Arg Gly Asn
355 360 365

<211> 362
<212> PRT
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 56

Met	Gln	Ile	Phe	Asp	Thr	Leu	Gln	Ser	Met	Gly	His	Glu	Gln	Val	Val
1				5					10					15	

Leu	Cys	Ser	Asp	Lys	Thr	Thr	Gly	Leu	Arg	Ala	Ile	Ile	Ala	Ile	His
			20					25					30		

Asp	Thr	Ser	Leu	Gly	Pro	Ala	Leu	Gly	Gly	Thr	Arg	Met	Trp	Gln	Tyr
		35					40					45			

Ala	Thr	Asp	Asp	Asp	Ala	Ile	Thr	Asp	Ala	Leu	Arg	Leu	Ser	Arg	Gly
	50					55					60				

Met	Thr	Tyr	Lys	Ala	Ala	Val	Ser	Gly	Val	Asn	Leu	Gly	Gly	Gly	Lys
65					70					75					80

Ala	Val	Ile	Ile	Gly	Asn	Pro	His	Ser	Asp	Lys	Ser	Glu	Ala	Leu	Phe
				85					90					95	

Arg	Ala	Tyr	Gly	Arg	Met	Val	Glu	Ser	Gln	Arg	Gly	Arg	Tyr	Ile	Thr
			100					105					110		

Ala	Glu	Asp	Val	Gly	Thr	Ser	Val	Arg	Asp	Met	Glu	Trp	Ile	Arg	Met
		115					120					125			

Glu	Thr	Lys	Tyr	Val	Thr	Gly	Val	Gly	Gly	Asn	Gly	Gly	Ser	Gly	Asp
	130					135					140				

Pro	Ser	Pro	Val	Thr	Ala	Leu	Gly	Val	Tyr	Ser	Gly	Met	Lys	Ala	Cys
145					150					155					160

Ala	Lys	Ser	Val	Tyr	Gly	Thr	Asp	Ala	Leu	Ser	Gly	Lys	Arg	Ile	Val
				165					170					175	

Val	Gln	Gly	Ala	Gly	Asn	Val	Ala	Ser	His	Leu	Val	His	Ser	Leu	Val

180

185

190

Lys Glu Gly Ala Val Val Phe Val Thr Asp Ile Tyr Glu Glu Lys Ala
 195 200 205

Lys Ala Leu Ala Ala Glu Thr Gly Ala Thr Val Ile Arg Thr Asp Glu
 210 215 220

Val Phe Thr Thr Gln Cys Asp Ile Phe Ser Pro Asn Ala Leu Gly Ala
 225 230 235 240

Val Leu Asn Asp Glu Thr Ile Pro Gln Leu Thr Cys Ala Ile Val Ala
 245 250 255

Gly Gly Ala Asn Asn Gln Leu Lys Ile Glu Gln Arg His Ala Thr Ala
 260 265 270

Leu Gln Glu Lys Gly Ile Leu Tyr Ala Pro Asp Tyr Val Ile Asn Ala
 275 280 285

Gly Gly Leu Met Asn Val Ala Ser Glu Val Asp Gly Tyr Asn Arg Glu
 290 295 300

Lys Val Met Arg Gln Ala Glu Gly Ile Tyr Asp Ile Thr Met Asn Ile
 305 310 315 320

Leu Asn Thr Ala Arg Glu Arg Asn Ile Leu Thr Ile Glu Ala Ser Asn
 325 330 335

Ala Ile Ala Glu Glu Arg Ile Asn Lys Val Arg His Val His Gly Asn
 340 345 350

Phe Ile Gly Ser Pro Ser Ile Arg Gly Val
 355 360

<210> 57

<211> 360

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 57

Met Asn Ile Phe Lys Lys Met Glu Glu Phe Asn Tyr Glu Gln Leu Val
1 5 10 15

Tyr Phe Tyr Asp Ser Glu Thr Glu Leu Lys Gly Ile Thr Cys Ile His
20 25 30

Asn Thr Thr Leu Gly Pro Ala Leu Gly Gly Thr Arg Leu Trp Asn Tyr
35 40 45

Asn Ser Glu Glu Asp Ala Val Glu Asp Val Ile Arg Leu Ala Arg Gly
50 55 60

Met Thr Tyr Lys Ala Ala Cys Ala Gly Leu Asn Leu Gly Gly Gly Lys
65 70 75 80

Thr Val Leu Ile Gly Asp Ala Lys Lys Ile Lys Ser Glu Ser Tyr Phe
85 90 95

Arg Gly Leu Gly Arg Tyr Val Gln Ser Leu Asn Gly Arg Tyr Ile Thr
100 105 110

Ala Glu Asp Val Asn Thr Ser Thr Lys Asp Met Ala Tyr Val Ala Met
115 120 125

Glu Thr Asp Tyr Val Val Gly Leu Gly Gly Lys Ser Gly Asn Pro Ser
130 135 140

Pro Val Thr Ala Tyr Gly Ala Phe Met Gly Ile Lys Ala Ala Leu Met
145 150 155 160

Lys Lys Phe Glu Asp Ser Ser Ile Glu Gly Arg Thr Phe Ala Val Gln
165 170 175

Gly Ala Gly Gln Thr Gly Tyr Tyr Leu Ile Asp Tyr Leu Leu Gly Asn
180 185 190

Asn Lys Phe Lys Glu Lys Ala Lys Lys Ile Tyr Phe Thr Glu Ile Asn
195 200 205

Glu Ser Tyr Ile Glu Arg Met Asn Lys Glu His Pro Glu Val Glu Phe
210 215 220

Ile Ser Pro Asp Lys Ile Tyr Ser Leu Glu Val Asp Val Phe Val Pro
225 230 235 240

Cys Ala Leu Gly Lys Ile Val Asn Asp Lys Thr Ile Asp Glu Phe Lys
245 250 255

Cys Pro Ile Ile Ala Gly Thr Ala Asn Asn Val Leu Glu Arg Glu Ala
260 265 270

His Gly Asn Met Leu Lys Glu Arg Gly Ile Leu Tyr Ala Pro Asp Tyr
275 280 285

Val Ile Asn Ala Gly Gly Leu Ile Asn Val Tyr His Glu Leu Asn Gly
290 295 300

Tyr Asn Lys Glu Asn Ala Ile Leu Glu Val Glu Leu Ile Tyr Asp Arg
305 310 315 320

Leu Leu Glu Ile Phe Asn Ile Ala Asp Ser Leu Asn Ile Ser Thr Asn
325 330 335

Ile Ala Ala Asn Glu Phe Ala Glu Lys Arg Ile Lys Gln Ile Lys Ser
340 345 350

Leu Lys Asn Asn Phe Ile Lys Arg
355 360

<210> 58

<211> 360

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 58

Met Asn Ile Phe Lys Lys Met Glu Glu Phe Asn Tyr Glu Gln Leu Val
1 5 10 15

Tyr Phe Tyr Asp Ser Glu Thr Glu Leu Lys Gly Ile Thr Cys Ile His
20 25 30

Asn Thr Thr Leu Gly Pro Ala Leu Gly Gly Thr Arg Leu Trp Asn Tyr
35 40 45

Asn Ser Glu Glu Asp Ala Val Glu Asp Val Ile Arg Leu Ala Arg Gly
50 55 60

Met Thr Tyr Lys Ala Ala Cys Ala Gly Leu Asn Leu Gly Gly Gly Lys
65 70 75 80

Thr Val Leu Ile Gly Asp Ala Lys Lys Ile Lys Ser Glu Ser Tyr Phe
85 90 95

Arg Gly Leu Gly Arg Tyr Val Gln Ser Leu Asn Gly Arg Tyr Ile Thr
100 105 110

Ala Glu Asp Val Asn Thr Ser Thr Lys Asp Met Ala Tyr Val Ala Met
115 120 125

Glu Thr Asp Tyr Val Val Gly Leu Gly Gly Lys Ser Gly Asn Pro Ser
130 135 140

Pro Val Thr Ala Tyr Gly Ala Phe Met Gly Ile Lys Ala Ala Leu Met
145 150 155 160

Lys Lys Phe Glu Asp Ser Ser Ile Glu Gly Arg Thr Phe Ala Val Gln
165 170 175

Gly Ala Gly Gln Thr Gly Tyr Tyr Leu Ile Asp Tyr Leu Leu Gly Asn
180 185 190

Asn Lys Phe Lys Glu Lys Ala Lys Lys Ile Tyr Phe Thr Glu Ile Asn
195 200 205

Glu Ser Tyr Ile Glu Arg Met Asn Lys Glu His Pro Glu Val Glu Phe
210 215 220

Ile Ser Pro Asp Lys Ile Tyr Ser Leu Glu Val Asp Val Phe Val Pro

225		230		235		240									
Cys	Ala	Leu	Gly	Lys	Ile	Val	Asn	Asp	Lys	Thr	Ile	Asp	Glu	Phe	Lys
				245					250					255	
Cys	Pro	Ile	Ile	Ala	Gly	Thr	Ala	Asn	Asn	Val	Leu	Glu	Arg	Glu	Ala
			260					265					270		
His	Gly	Asn	Met	Leu	Lys	Glu	Arg	Gly	Ile	Leu	Tyr	Ala	Pro	Asp	Tyr
		275					280					285			
Val	Ile	Asn	Ala	Gly	Gly	Leu	Ile	Asn	Val	Tyr	His	Glu	Leu	Asn	Gly
	290					295					300				
Tyr	Asn	Lys	Glu	Asn	Ala	Ile	Leu	Glu	Val	Glu	Leu	Ile	Tyr	Asp	Arg
305					310					315					320
Leu	Leu	Glu	Ile	Phe	Asn	Ile	Ala	Asp	Ser	Leu	Asn	Ile	Ser	Thr	Asn
				325					330					335	
Ile	Ala	Ala	Asn	Glu	Phe	Ala	Glu	Lys	Arg	Ile	Lys	Gln	Ile	Lys	Ser
			340					345					350		
Leu	Lys	Asn	Asn	Phe	Ile	Lys	Arg								
		355					360								

<210> 59
 <211> 366
 <212> PRT
 <213> Artificial Sequence

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

<400>	59														
Met	Ile	Phe	Glu	Thr	Ile	Ser	Thr	Ser	Asn	His	Glu	Glu	Val	Val	Tyr
1				5					10					15	

Cys	His	Asn	Lys	Asp	Ala	Gly	Leu	Lys	Ala	Ile	Ile	Ala	Ile	His	Asn
			20					25					30		

Thr	Val	Leu	Gly	Pro	Ala	Leu	Gly	Gly	Thr	Arg	Met	Trp	Pro	Tyr	Ala

35

40

45

Ser Glu Glu Glu Ala Leu Lys Asp Val Leu Arg Leu Ser Arg Gly Met
 50 55 60

Thr Tyr Lys Ala Ala Val Ser Gly Leu Asn Leu Gly Gly Gly Lys Ala
 65 70 75 80

Val Ile Trp Gly Asp Pro Asn Lys Asp Lys Ser Glu Ala Leu Phe Arg
 85 90 95

Ala Phe Gly Arg Phe Val Asn Ser Leu Gly Gly Arg Tyr Ile Thr Ala
 100 105 110

Glu Asp Val Gly Ile Asp Val Asn Asp Met Glu Tyr Val Leu Arg Glu
 115 120 125

Thr Asp Tyr Val Thr Gly Val His Gln Val His Gly Gly Ser Gly Asp
 130 135 140

Pro Ser Pro Phe Thr Ala Tyr Gly Thr Leu Gln Gly Leu Met Ala Ala
 145 150 155 160

Leu Gln Val Lys Phe Gly Asn Glu Asp Val Gly Asn Tyr Ser Tyr Ala
 165 170 175

Val Gln Gly Val Gly His Val Gly Met Glu Phe Val Lys Leu Leu Arg
 180 185 190

Glu Arg Gly Ala Lys Val Phe Val Thr Asp Ile Asn Lys Asp Ala Val
 195 200 205

Gln Arg Ala Val Asp Glu Phe Gly Cys Glu Ala Val Ala Leu Asp Glu
 210 215 220

Ile Tyr Asp Val Asp Cys Asp Val Tyr Ser Pro Thr Ala Leu Gly Gly
 225 230 235 240

Thr Val Asn Asp Lys Thr Leu Pro Arg Leu Lys Cys Lys Val Ile Cys
 245 250 255

Gly Ala Ala Asn Asn Gln Leu Ala Asn Asp Glu Ile Gly Val Glu Leu
260 265 270

Glu Lys Lys Gly Ile Leu Tyr Ala Pro Asp Tyr Ala Val Asn Ala Gly
275 280 285

Gly Leu Met Asn Val Ser Leu Glu Ile Asp Gly Tyr Asn Arg Glu Arg
290 295 300

Ala Met Arg Met Met Arg Thr Ile Tyr Tyr Asn Leu Gly Arg Ile Phe
305 310 315 320

Glu Ile Ser Lys Arg Asp Gly Ile Pro Thr Phe Arg Ala Ala Asp Arg
325 330 335

Met Ala Glu Glu Arg Ile Thr Ala Ile Gly Lys Leu Arg Leu Pro His
340 345 350

Leu Gly Ala Ala Ala Pro Arg Phe Gln Gly Arg Arg Gly Asn
355 360 365

<210> 60

<211> 362

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 60

Met Gln Ile Phe Asp Thr Leu Gln Ser Met Gly His Glu Gln Val Val
1 5 10 15

Leu Cys Ser Asp Lys Thr Thr Gly Leu Arg Ala Ile Ile Ala Ile His
20 25 30

Asp Thr Ser Leu Gly Pro Ala Leu Gly Gly Thr Arg Met Trp Gln Tyr
35 40 45

Ala Thr Asp Asp Asp Ala Ile Thr Asp Ala Leu Arg Leu Ser Arg Gly
50 55 60

Met Thr Tyr Lys Ala Ala Val Ser Gly Val Asn Leu Gly Gly Gly Lys
65 70 75 80

Ala Val Ile Ile Gly Asn Pro His Ser Asp Lys Ser Glu Ala Leu Phe
85 90 95

Arg Ala Tyr Gly Arg Met Val Glu Ser Gln Arg Gly Arg Tyr Ile Thr
100 105 110

Ala Glu Asp Val Gly Thr Ser Val Arg Asp Met Glu Trp Ile Arg Met
115 120 125

Glu Thr Lys Tyr Val Thr Gly Val Gly Gly Asn Gly Gly Ser Gly Asp
130 135 140

Pro Ser Pro Val Thr Ala Leu Gly Val Tyr Ser Gly Met Lys Ala Cys
145 150 155 160

Ala Lys Ser Val Tyr Gly Thr Asp Ala Leu Ser Gly Lys Arg Ile Val
165 170 175

Val Gln Gly Ala Gly Asn Val Ala Ser His Leu Val His Ser Leu Val
180 185 190

Lys Glu Gly Ala Val Val Phe Val Thr Asp Ile Tyr Glu Glu Lys Ala
195 200 205

Lys Ala Leu Ala Ala Glu Thr Gly Ala Thr Val Ile Arg Thr Asp Glu
210 215 220

Val Phe Thr Thr Gln Cys Asp Ile Phe Ser Pro Asn Ala Leu Gly Ala
225 230 235 240

Val Leu Asn Asp Glu Thr Ile Pro Gln Leu Thr Cys Ala Ile Val Ala
245 250 255

Gly Gly Ala Asn Asn Gln Leu Lys Ile Glu Gln Arg His Ala Thr Ala
260 265 270

Leu Gln Glu Lys Gly Ile Leu Tyr Ala Pro Asp Tyr Val Ile Asn Ala

275

280

285

Gly Gly Leu Met Asn Val Ala Ser Glu Val Asp Gly Tyr Asn Arg Glu
 290 295 300

Lys Val Met Arg Gln Ala Glu Gly Ile Tyr Asp Ile Thr Met Asn Ile
 305 310 315 320

Leu Asn Thr Ala Arg Glu Arg Asn Ile Leu Thr Ile Glu Ala Ser Asn
 325 330 335

Ala Ile Ala Glu Glu Arg Ile Asn Lys Val Arg His Val His Gly Asn
 340 345 350

Phe Ile Gly Ser Pro Ser Ile Arg Gly Val
 355 360

<210> 61

<211> 362

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
 polypeptide

<400> 61

Met Gln Ile Phe Asp Thr Leu Gln Ser Met Gly His Glu Gln Val Val
 1 5 10 15

Leu Cys Ser Asp Lys Thr Thr Gly Leu Arg Ala Ile Ile Ala Ile His
 20 25 30

Asp Thr Ser Leu Gly Pro Ala Leu Gly Gly Thr Arg Met Trp Gln Tyr
 35 40 45

Ala Thr Asp Asp Asp Ala Ile Thr Asp Ala Leu Arg Leu Ser Arg Gly
 50 55 60

Met Thr Tyr Lys Ala Ala Val Ser Gly Val Asn Leu Gly Gly Gly Lys
 65 70 75 80

Ala Val Ile Ile Gly Asn Pro His Ser Asp Lys Ser Glu Ala Leu Phe

85

90

95

Arg Ala Tyr Gly Arg Met Val Glu Ser Gln Arg Gly Arg Tyr Ile Thr
 100 105 110

Ala Glu Asp Val Gly Thr Ser Val Arg Asp Met Glu Trp Ile Arg Met
 115 120 125

Glu Thr Lys Tyr Val Thr Gly Val Gly Gly Asn Gly Gly Ser Gly Asp
 130 135 140

Pro Ser Pro Val Thr Ala Leu Gly Val Tyr Ser Gly Met Lys Ala Cys
 145 150 155 160

Ala Lys Ser Val Tyr Gly Thr Asp Ala Leu Ser Gly Lys Arg Ile Val
 165 170 175

Val Gln Gly Ala Gly Asn Val Ala Ser His Leu Val His Ser Leu Val
 180 185 190

Lys Glu Gly Ala Val Val Phe Val Thr Asp Ile Tyr Glu Glu Lys Ala
 195 200 205

Lys Ala Leu Ala Ala Glu Thr Gly Ala Thr Val Ile Arg Thr Asp Glu
 210 215 220

Val Phe Thr Thr Gln Cys Asp Ile Phe Ser Pro Asn Ala Leu Gly Ala
 225 230 235 240

Val Leu Asn Asp Glu Thr Ile Pro Gln Leu Thr Cys Ala Ile Val Ala
 245 250 255

Gly Gly Ala Asn Asn Gln Leu Lys Ile Glu Gln Arg His Ala Thr Ala
 260 265 270

Leu Gln Glu Lys Gly Ile Leu Tyr Ala Pro Asp Tyr Val Ile Asn Ala
 275 280 285

Gly Gly Leu Met Asn Val Ala Ser Glu Val Asp Gly Tyr Asn Arg Glu
 290 295 300

Lys Val Met Arg Gln Ala Glu Gly Ile Tyr Asp Ile Thr Met Asn Ile
305 310 315 320

Leu Asn Thr Ala Arg Glu Arg Asn Ile Leu Thr Ile Glu Ala Ser Asn
325 330 335

Ala Ile Ala Glu Glu Arg Ile Asn Lys Val Arg His Val His Gly Asn
340 345 350

Phe Ile Gly Ser Pro Ser Ile Arg Gly Val
355 360

<210> 62

<211> 362

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 62

Met Gln Ile Phe Asp Thr Leu Gln Ser Met Gly His Glu Gln Val Val
1 5 10 15

Leu Cys Ser Asp Lys Thr Thr Gly Leu Arg Ala Ile Ile Ala Ile His
20 25 30

Asp Thr Ser Leu Gly Pro Ala Leu Gly Gly Thr Arg Met Trp Gln Tyr
35 40 45

Ala Thr Asp Asp Asp Ala Ile Thr Asp Ala Leu Arg Leu Ser Arg Gly
50 55 60

Met Thr Tyr Lys Ala Ala Val Ser Gly Val Asn Leu Gly Gly Gly Lys
65 70 75 80

Ala Val Ile Ile Gly Asn Pro His Ser Asp Lys Ser Glu Ala Leu Phe
85 90 95

Arg Ala Tyr Gly Arg Met Val Glu Ser Gln Arg Gly Arg Tyr Ile Thr
100 105 110

Ala Glu Asp Val Gly Thr Ser Val Arg Asp Met Glu Trp Ile Arg Met
115 120 125

Glu Thr Lys Tyr Val Thr Gly Val Gly Gly Asn Gly Gly Ser Gly Asp
130 135 140

Pro Ser Pro Val Thr Ala Leu Gly Val Tyr Ser Gly Met Lys Ala Cys
145 150 155 160

Ala Lys Ser Val Tyr Gly Thr Asp Ala Leu Ser Gly Lys Arg Ile Val
165 170 175

Val Gln Gly Ala Gly Asn Val Ala Ser His Leu Val His Ser Leu Val
180 185 190

Lys Glu Gly Ala Val Val Phe Val Thr Asp Ile Tyr Glu Glu Lys Ala
195 200 205

Lys Ala Leu Ala Ala Glu Thr Gly Ala Thr Val Ile Arg Thr Asp Glu
210 215 220

Val Phe Thr Thr Gln Cys Asp Ile Phe Ser Pro Asn Ala Leu Gly Ala
225 230 235 240

Val Leu Asn Asp Glu Thr Ile Pro Gln Leu Thr Cys Ala Ile Val Ala
245 250 255

Gly Gly Ala Asn Asn Gln Leu Lys Ile Glu Gln Arg His Ala Thr Ala
260 265 270

Leu Gln Glu Lys Gly Ile Leu Tyr Ala Pro Asp Tyr Val Ile Asn Ala
275 280 285

Gly Gly Leu Met Asn Val Ala Ser Glu Val Asp Gly Tyr Asn Arg Glu
290 295 300

Lys Val Met Arg Gln Ala Glu Gly Ile Tyr Asp Ile Thr Met Asn Ile
305 310 315 320

Leu Asn Thr Ala Arg Glu Arg Asn Ile Leu Thr Ile Glu Ala Ser Asn

325

330

335

Ala Ile Ala Glu Glu Arg Ile Asn Lys Val Arg His Val His Gly Asn
 340 345 350

Phe Ile Gly Ser Pro Ser Ile Arg Gly Val
 355 360

<210> 63

<211> 362

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
 polypeptide

<400> 63

Met Gln Ile Phe Asp Thr Leu Gln Ser Met Gly His Glu Gln Val Val
 1 5 10 15

Leu Cys Ser Asp Lys Thr Thr Gly Leu Arg Ala Ile Ile Ala Ile His
 20 25 30

Asp Thr Ser Leu Gly Pro Ala Leu Gly Gly Thr Arg Met Trp Gln Tyr
 35 40 45

Ala Thr Asp Asp Asp Ala Ile Thr Asp Ala Leu Arg Leu Ser Arg Gly
 50 55 60

Met Thr Tyr Lys Ala Ala Val Ser Gly Val Asn Leu Gly Gly Gly Lys
 65 70 75 80

Ala Val Ile Ile Gly Asn Pro His Ser Asp Lys Ser Glu Ala Leu Phe
 85 90 95

Arg Ala Tyr Gly Arg Met Val Glu Ser Gln Arg Gly Arg Tyr Ile Thr
 100 105 110

Ala Glu Asp Val Gly Thr Ser Val Arg Asp Met Glu Trp Ile Arg Met
 115 120 125

Glu Thr Lys Tyr Val Thr Gly Val Gly Gly Asn Gly Gly Ser Gly Asp

130

135

140

Pro Ser Pro Val Thr Ala Leu Gly Val Tyr Ser Gly Met Lys Ala Cys
 145 150 155 160

Ala Lys Ser Val Tyr Gly Thr Asp Ala Leu Ser Gly Lys Arg Ile Val
 165 170 175

Val Gln Gly Ala Gly Asn Val Ala Ser His Leu Val His Ser Leu Val
 180 185 190

Lys Glu Gly Ala Val Val Phe Val Thr Asp Ile Tyr Glu Glu Lys Ala
 195 200 205

Lys Ala Leu Ala Ala Glu Thr Gly Ala Thr Val Ile Arg Thr Asp Glu
 210 215 220

Val Phe Thr Thr Gln Cys Asp Ile Phe Ser Pro Asn Ala Leu Gly Ala
 225 230 235 240

Val Leu Asn Asp Glu Thr Ile Pro Gln Leu Thr Cys Ala Ile Val Ala
 245 250 255

Gly Gly Ala Asn Asn Gln Leu Lys Ile Glu Gln Arg His Ala Thr Ala
 260 265 270

Leu Gln Glu Lys Gly Ile Leu Tyr Ala Pro Asp Tyr Val Ile Asn Ala
 275 280 285

Gly Gly Leu Met Asn Val Ala Ser Glu Val Asp Gly Tyr Asn Arg Glu
 290 295 300

Lys Val Met Arg Gln Ala Glu Gly Ile Tyr Asp Ile Thr Met Asn Ile
 305 310 315 320

Leu Asn Thr Ala Arg Glu Arg Asn Ile Leu Thr Ile Glu Ala Ser Asn
 325 330 335

Ala Ile Ala Glu Glu Arg Ile Asn Lys Val Arg His Val His Gly Asn
 340 345 350

Phe Ile Gly Ser Pro Ser Ile Arg Gly Val
355 360

<210> 64

<211> 360

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 64

Met Asn Ile Phe Lys Lys Met Glu Glu Phe Asn Tyr Glu Gln Leu Val
1 5 10 15

Tyr Phe Tyr Asp Ser Glu Thr Glu Leu Lys Gly Ile Thr Cys Ile His
20 25 30

Asn Thr Thr Leu Gly Pro Ala Leu Gly Gly Thr Arg Leu Trp Asn Tyr
35 40 45

Asn Ser Glu Glu Asp Ala Val Glu Asp Val Ile Arg Leu Ala Arg Gly
50 55 60

Met Thr Tyr Lys Ala Ala Cys Ala Gly Leu Asn Leu Gly Gly Gly Lys
65 70 75 80

Thr Val Leu Ile Gly Asp Ala Lys Lys Ile Lys Ser Glu Ser Tyr Phe
85 90 95

Arg Gly Leu Gly Arg Tyr Val Gln Ser Leu Asn Gly Arg Tyr Ile Thr
100 105 110

Ala Glu Asp Val Asn Thr Ser Thr Lys Asp Met Ala Tyr Val Ala Met
115 120 125

Glu Thr Asp Tyr Val Val Gly Leu Gly Gly Lys Ser Gly Asn Pro Ser
130 135 140

Pro Val Thr Ala Tyr Gly Ala Phe Met Gly Ile Lys Ala Ala Leu Met
145 150 155 160

Lys Lys Phe Glu Asp Ser Ser Ile Glu Gly Arg Thr Phe Ala Val Gln
165 170 175

Gly Ala Gly Gln Thr Gly Tyr Tyr Leu Ile Asp Tyr Leu Leu Gly Asn
180 185 190

Asn Lys Phe Lys Glu Lys Ala Lys Lys Ile Tyr Phe Thr Glu Ile Asn
195 200 205

Glu Ser Tyr Ile Glu Arg Met Asn Lys Glu His Pro Glu Val Glu Phe
210 215 220

Ile Ser Pro Asp Lys Ile Tyr Ser Leu Glu Val Asp Val Phe Val Pro
225 230 235 240

Cys Ala Leu Gly Lys Ile Val Asn Asp Lys Thr Ile Asp Glu Phe Lys
245 250 255

Cys Pro Ile Ile Ala Gly Thr Ala Asn Asn Val Leu Glu Arg Glu Ala
260 265 270

His Gly Asn Met Leu Lys Glu Arg Gly Ile Leu Tyr Ala Pro Asp Tyr
275 280 285

Val Ile Asn Ala Gly Gly Leu Ile Asn Val Tyr His Glu Leu Asn Gly
290 295 300

Tyr Asn Lys Glu Asn Ala Ile Leu Glu Val Glu Leu Ile Tyr Asp Arg
305 310 315 320

Leu Leu Glu Ile Phe Asn Ile Ala Asp Ser Leu Asn Ile Ser Thr Asn
325 330 335

Ile Ala Ala Asn Glu Phe Ala Glu Lys Arg Ile Lys Gln Ile Lys Ser
340 345 350

Leu Lys Asn Asn Phe Ile Lys Arg
355 360

<210> 65

<211> 360
<212> PRT
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 65

Met	Asn	Ile	Phe	Lys	Lys	Met	Glu	Glu	Phe	Asn	Tyr	Glu	Gln	Leu	Val
1				5					10					15	

Tyr	Phe	Tyr	Asp	Ser	Glu	Thr	Glu	Leu	Lys	Gly	Ile	Thr	Cys	Ile	His
			20					25					30		

Asn	Thr	Thr	Leu	Gly	Pro	Ala	Leu	Gly	Gly	Thr	Arg	Leu	Trp	Asn	Tyr
		35					40					45			

Asn	Ser	Glu	Glu	Asp	Ala	Val	Glu	Asp	Val	Ile	Arg	Leu	Ala	Arg	Gly
	50					55					60				

Met	Thr	Tyr	Lys	Ala	Ala	Cys	Ala	Gly	Leu	Asn	Leu	Gly	Gly	Gly	Lys
65					70					75					80

Thr	Val	Leu	Ile	Gly	Asp	Ala	Lys	Lys	Ile	Lys	Ser	Glu	Ser	Tyr	Phe
				85					90					95	

Arg	Gly	Leu	Gly	Arg	Tyr	Val	Gln	Ser	Leu	Asn	Gly	Arg	Tyr	Ile	Thr
		100						105					110		

Ala	Glu	Asp	Val	Asn	Thr	Ser	Thr	Lys	Asp	Met	Ala	Tyr	Val	Ala	Met
		115					120					125			

Glu	Thr	Asp	Tyr	Val	Val	Gly	Leu	Gly	Gly	Lys	Ser	Gly	Asn	Pro	Ser
	130					135					140				

Pro	Val	Thr	Ala	Tyr	Gly	Ala	Phe	Met	Gly	Ile	Lys	Ala	Ala	Leu	Met
145					150					155					160

Lys	Lys	Phe	Glu	Asp	Ser	Ser	Ile	Glu	Gly	Arg	Thr	Phe	Ala	Val	Gln
				165					170					175	

Gly	Ala	Gly	Gln	Thr	Gly	Tyr	Tyr	Leu	Ile	Asp	Tyr	Leu	Leu	Gly	Asn

180

185

190

Asn Lys Phe Lys Glu Lys Ala Lys Lys Ile Tyr Phe Thr Glu Ile Asn
195 200 205

Glu Ser Tyr Ile Glu Arg Met Asn Lys Glu His Pro Glu Val Glu Phe
210 215 220

Ile Ser Pro Asp Lys Ile Tyr Ser Leu Glu Val Asp Val Phe Val Pro
225 230 235 240

Cys Ala Leu Gly Lys Ile Val Asn Asp Lys Thr Ile Asp Glu Phe Lys
245 250 255

Cys Pro Ile Ile Ala Gly Thr Ala Asn Asn Val Leu Glu Arg Glu Ala
260 265 270

His Gly Asn Met Leu Lys Glu Arg Gly Ile Leu Tyr Ala Pro Asp Tyr
275 280 285

Val Ile Asn Ala Gly Gly Leu Ile Asn Val Tyr His Glu Leu Asn Gly
290 295 300

Tyr Asn Lys Glu Asn Ala Ile Leu Glu Val Glu Leu Ile Tyr Asp Arg
305 310 315 320

Leu Leu Glu Ile Phe Asn Ile Ala Asp Ser Leu Asn Ile Ser Thr Asn
325 330 335

Ile Ala Ala Asn Glu Phe Ala Glu Lys Arg Ile Lys Gln Ile Lys Ser
340 345 350

Leu Lys Asn Asn Phe Ile Lys Arg
355 360

<210> 66

<211> 360

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 66

Met Asn Ile Phe Lys Lys Met Glu Glu Phe Asn Tyr Glu Gln Leu Val
1 5 10 15

Tyr Phe Tyr Asp Ser Glu Thr Glu Leu Lys Gly Ile Thr Cys Ile His
20 25 30

Asn Thr Thr Leu Gly Pro Ala Leu Gly Gly Thr Arg Leu Trp Asn Tyr
35 40 45

Asn Ser Glu Glu Asp Ala Val Glu Asp Val Ile Arg Leu Ala Arg Gly
50 55 60

Met Thr Tyr Lys Ala Ala Cys Ala Gly Leu Asn Leu Gly Gly Gly Lys
65 70 75 80

Thr Val Leu Ile Gly Asp Ala Lys Lys Ile Lys Ser Glu Ser Tyr Phe
85 90 95

Arg Gly Leu Gly Arg Tyr Val Gln Ser Leu Asn Gly Arg Tyr Ile Thr
100 105 110

Ala Glu Asp Val Asn Thr Ser Thr Lys Asp Met Ala Tyr Val Ala Met
115 120 125

Glu Thr Asp Tyr Val Val Gly Leu Gly Gly Lys Ser Gly Asn Pro Ser
130 135 140

Pro Val Thr Ala Tyr Gly Ala Phe Met Gly Ile Lys Ala Ala Leu Met
145 150 155 160

Lys Lys Phe Glu Asp Ser Ser Ile Glu Gly Arg Thr Phe Ala Val Gln
165 170 175

Gly Ala Gly Gln Thr Gly Tyr Tyr Leu Ile Asp Tyr Leu Leu Gly Asn
180 185 190

Asn Lys Phe Lys Glu Lys Ala Lys Lys Ile Tyr Phe Thr Glu Ile Asn
195 200 205

Glu Ser Tyr Ile Glu Arg Met Asn Lys Glu His Pro Glu Val Glu Phe
210 215 220

Ile Ser Pro Asp Lys Ile Tyr Ser Leu Glu Val Asp Val Phe Val Pro
225 230 235 240

Cys Ala Leu Gly Lys Ile Val Asn Asp Lys Thr Ile Asp Glu Phe Lys
245 250 255

Cys Pro Ile Ile Ala Gly Thr Ala Asn Asn Val Leu Glu Arg Glu Ala
260 265 270

His Gly Asn Met Leu Lys Glu Arg Gly Ile Leu Tyr Ala Pro Asp Tyr
275 280 285

Val Ile Asn Ala Gly Gly Leu Ile Asn Val Tyr His Glu Leu Asn Gly
290 295 300

Tyr Asn Lys Glu Asn Ala Ile Leu Glu Val Glu Leu Ile Tyr Asp Arg
305 310 315 320

Leu Leu Glu Ile Phe Asn Ile Ala Asp Ser Leu Asn Ile Ser Thr Asn
325 330 335

Ile Ala Ala Asn Glu Phe Ala Glu Lys Arg Ile Lys Gln Ile Lys Ser
340 345 350

Leu Lys Asn Asn Phe Ile Lys Arg
355 360

<210> 67

<211> 363

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 67

Met Val Glu Ile Gln Ala Leu Pro Glu Thr Ser Ile Phe Gly Gln Ile
1 5 10 15

Ala Asp His Gln His Glu Gln Val Val Phe Cys His Asp His Glu Thr
20 25 30

Gly Leu Arg Ala Ile Ile Gly Ile His Asn Thr Val Leu Gly Pro Ala
35 40 45

Leu Gly Gly Thr Arg Met Trp His Tyr Ala Thr Glu Ala Glu Ala Leu
50 55 60

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

Ser Gly Leu Asn Leu Gly Gly Gly Lys Ala Val Ile Ile Gly Asp Ala
85 90 95

Lys Thr Ile Lys Thr Glu Ala Leu Leu Arg Lys Phe Gly Arg Phe Val
100 105 110

Gln Asn Leu Asn Gly Lys Tyr Ile Thr Ala Glu Asp Val Asn Met Thr
115 120 125

Thr Lys Asp Met Glu Tyr Ile Arg Met Glu Thr Lys His Val Ala Gly
130 135 140

Leu Pro Glu Ser Met Gly Gly Ser Gly Asp Pro Ser Pro Val Thr Ala
145 150 155 160

Tyr Gly Thr Tyr Met Gly Met Lys Ala Ala Ala Lys Lys Ala Phe Gly
165 170 175

Ser Asp Ser Leu Ala Gly Lys Arg Ile Ala Val Gln Gly Val Gly His
180 185 190

Val Gly Thr Tyr Leu Leu Glu His Leu Thr Lys Glu Gly Ala Gln Ile
195 200 205

Val Leu Thr Asp Tyr Tyr Lys Glu Arg Ala Glu Glu Ala Gly Ala Arg
210 215 220

Phe Gly Ala Gln Val Val Gly Leu Asp Asp Ile Tyr Asp Gln Glu Val

225		230		235		240									
Asp	Ile	Tyr	Ser	Pro	Cys	Ala	Leu	Gly	Ala	Thr	Ile	Asn	Asp	Asp	Thr
				245					250					255	
Ile	Asp	Arg	Leu	Arg	Cys	Ala	Val	Val	Ala	Gly	Cys	Ala	Asn	Asn	Gln
			260						265				270		
Leu	Lys	Glu	Glu	Asn	Val	His	Gly	Pro	Ala	Leu	Val	Glu	Arg	Gly	Ile
		275					280					285			
Val	Tyr	Ala	Pro	Asp	Phe	Leu	Ile	Asn	Ala	Gly	Gly	Leu	Ile	Asn	Val
	290					295					300				
Tyr	Ser	Glu	Val	Thr	Gly	Ser	Thr	Arg	Gln	Gly	Ala	Leu	Thr	Gln	Thr
305					310					315					320
Glu	Lys	Ile	Tyr	Asp	Tyr	Thr	Leu	Gln	Val	Leu	Glu	Lys	Ala	Ala	Ala
				325					330					335	
Glu	Gly	Leu	His	Pro	Gln	Gln	Ala	Ala	Ile	Arg	Gln	Ala	Glu	Gln	Arg
			340					345					350		
Ile	Ala	Ala	Ile	Gly	Lys	Val	Lys	Ser	Thr	Tyr					
		355					360								

<210> 68
 <211> 363
 <212> PRT
 <213> Artificial Sequence

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

<400> 68
 Met Val Glu Ile Gln Ala Leu Pro Glu Thr Ser Ile Phe Gly Gln Ile
 1 5 10 15

Ala Asp His Gln His Glu Gln Val Val Phe Cys His Asp His Glu Thr
 20 25 30

Gly Leu Arg Ala Ile Ile Gly Ile His Asn Thr Val Leu Gly Pro Ala

35

40

45

Leu Gly Gly Thr Arg Met Trp His Tyr Ala Thr Glu Ala Glu Ala Leu
 50 55 60

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

Ser Gly Leu Asn Leu Gly Gly Gly Lys Ala Val Ile Ile Gly Asp Ala
 85 90 95

Lys Thr Ile Lys Thr Glu Ala Leu Leu Arg Lys Phe Gly Arg Phe Val
 100 105 110

Gln Asn Leu Asn Gly Lys Tyr Ile Thr Ala Glu Asp Val Asn Met Thr
 115 120 125

Thr Lys Asp Met Glu Tyr Ile Arg Met Glu Thr Lys His Val Ala Gly
 130 135 140

Leu Pro Glu Ser Met Gly Gly Ser Gly Asp Pro Ser Pro Val Thr Ala
 145 150 155 160

Tyr Gly Thr Tyr Met Gly Met Lys Ala Ala Ala Lys Lys Ala Phe Gly
 165 170 175

Ser Asp Ser Leu Ala Gly Lys Arg Ile Ala Val Gln Gly Val Gly His
 180 185 190

Val Gly Thr Tyr Leu Leu Glu His Leu Thr Lys Glu Gly Ala Gln Ile
 195 200 205

Val Leu Thr Asp Tyr Tyr Lys Glu Arg Ala Glu Glu Ala Gly Ala Arg
 210 215 220

Phe Gly Ala Gln Val Val Gly Leu Asp Asp Ile Tyr Asp Gln Glu Val
 225 230 235 240

Asp Ile Tyr Ser Pro Cys Ala Leu Gly Ala Thr Ile Asn Asp Asp Thr
 245 250 255

Ile Asp Arg Leu Arg Cys Ala Val Val Ala Gly Cys Ala Asn Asn Gln
260 265 270

Leu Lys Glu Glu Asn Val His Gly Pro Ala Leu Val Glu Arg Gly Ile
275 280 285

Val Tyr Ala Pro Asp Phe Leu Ile Asn Ala Gly Gly Leu Ile Asn Val
290 295 300

Tyr Ser Glu Val Thr Gly Ser Thr Arg Gln Gly Ala Leu Thr Gln Thr
305 310 315 320

Glu Lys Ile Tyr Asp Tyr Thr Leu Gln Val Leu Glu Lys Ala Ala Ala
325 330 335

Glu Gly Leu His Pro Gln Gln Ala Ala Ile Arg Gln Ala Glu Gln Arg
340 345 350

Ile Ala Ala Ile Gly Lys Val Lys Ser Thr Tyr
355 360

<210> 69

<211> 366

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 69

Met Ile Phe Glu Thr Ile Ser Thr Ser Asn His Glu Glu Val Val Tyr
1 5 10 15

Cys His Asn Lys Asp Ala Gly Leu Lys Ala Ile Ile Ala Ile His Asn
20 25 30

Thr Val Leu Gly Pro Ala Leu Gly Gly Thr Arg Met Trp Pro Tyr Ala
35 40 45

Ser Glu Glu Glu Ala Leu Lys Asp Val Leu Arg Leu Ser Arg Gly Met
50 55 60

Thr Tyr Lys Ala Ala Val Ser Gly Leu Asn Leu Gly Gly Gly Lys Ala
65 70 75 80

Val Ile Trp Gly Asp Pro Asn Lys Asp Lys Ser Glu Ala Leu Phe Arg
85 90 95

Ala Phe Gly Arg Phe Val Asn Ser Leu Gly Gly Arg Tyr Ile Thr Ala
100 105 110

Glu Asp Val Gly Ile Asp Val Asn Asp Met Glu Tyr Val Leu Arg Glu
115 120 125

Thr Asp Tyr Val Thr Gly Val His Gln Val His Gly Gly Ser Gly Asp
130