



US 20070092943A1

(19) **United States**

(12) **Patent Application Publication**

Laible et al.

(10) **Pub. No.: US 2007/0092943 A1**

(43) **Pub. Date: Apr. 26, 2007**

(54) **VERSATILE VECTORS FOR EXPRESSION OF FOREIGN PROTEINS IN PHOTOSYNTHETIC BACTERIA**

Related U.S. Application Data

(60) Provisional application No. 60/721,423, filed on Sep. 28, 2005.

(75) Inventors: **Philip D. Laible**, Villa Park, IL (US);
Deborah K. Hanson, Villa Park, IL (US)

Publication Classification

(51) **Int. Cl.**
C12P 21/06 (2006.01)
C07H 21/04 (2006.01)
C12N 15/74 (2006.01)
C12N 1/21 (2006.01)
C07K 14/195 (2006.01)
(52) **U.S. Cl.** **435/69.1**; 435/252.3; 435/471; 530/350; 536/23.7

Correspondence Address:
BARNES & THORNBURG LLP
P.O. BOX 2786
CHICAGO, IL 60690-2786 (US)

(73) Assignee: **THE UNIVERSITY OF CHICAGO**,
Chicago, IL (US)

(57) **ABSTRACT**

Methods for expressing and purifying foreign (heterologous) proteins in photosynthetic organisms employ expression of both heterologous membrane proteins and a means for compartmentalizing or sequestering of the protein.

(21) Appl. No.: **11/536,331**

(22) Filed: **Sep. 28, 2006**

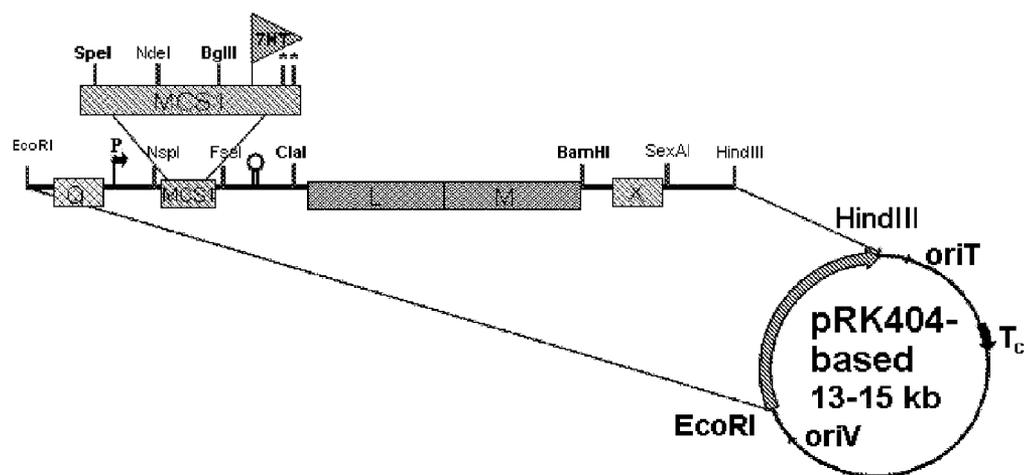


FIG. 1

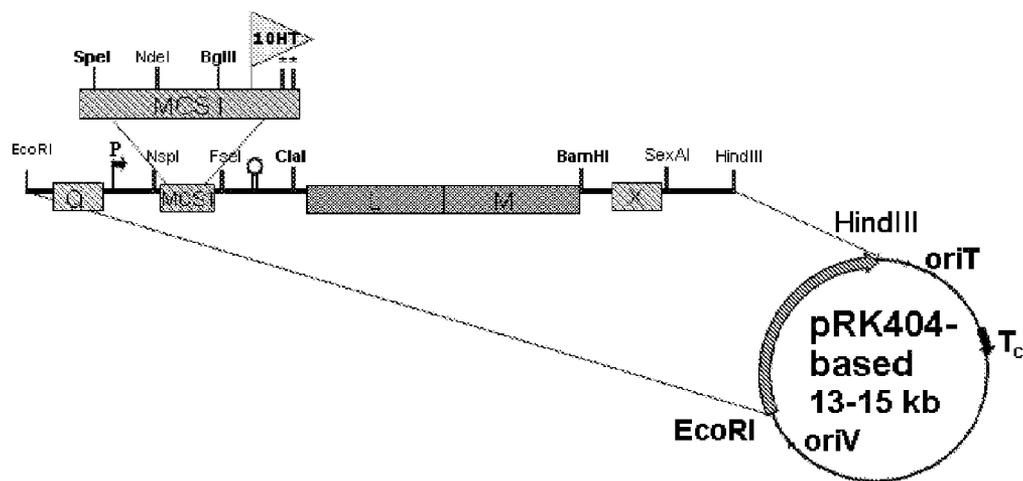


FIG. 2

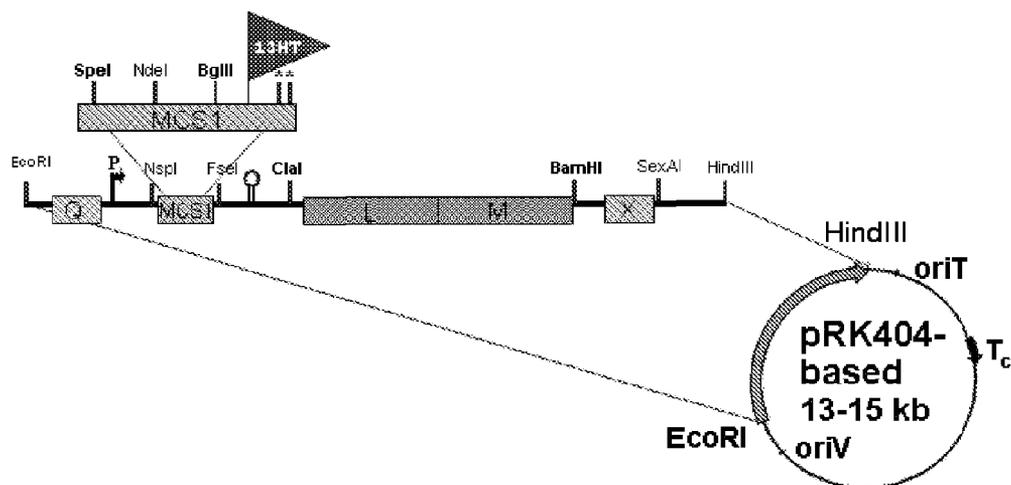


FIG. 3

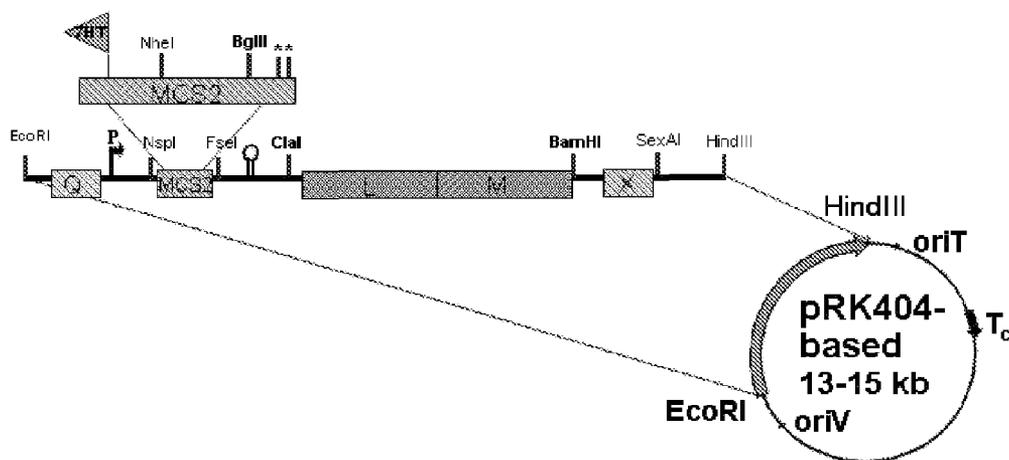


FIG. 4

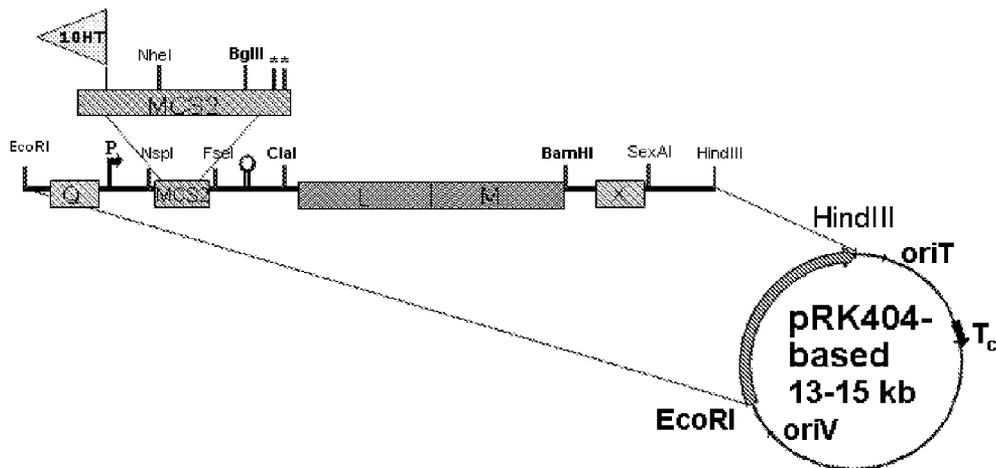


FIG. 5

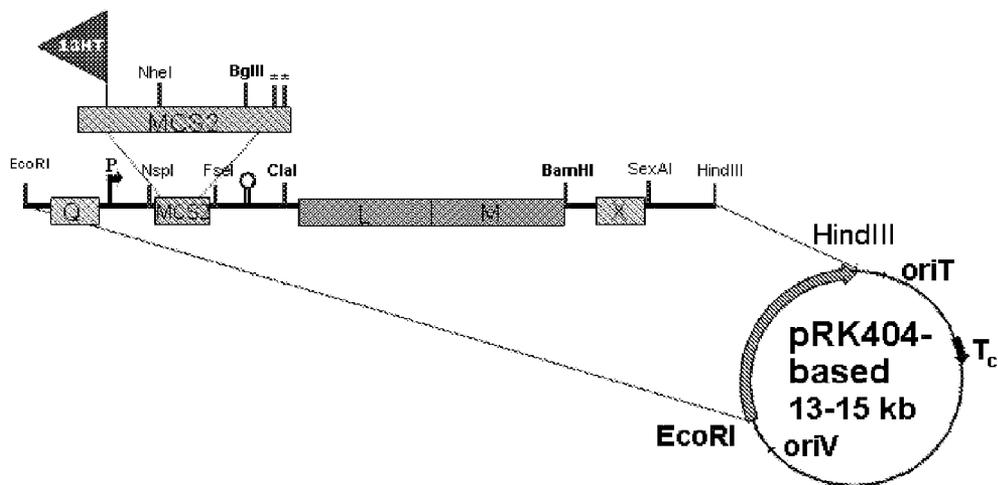


FIG. 6

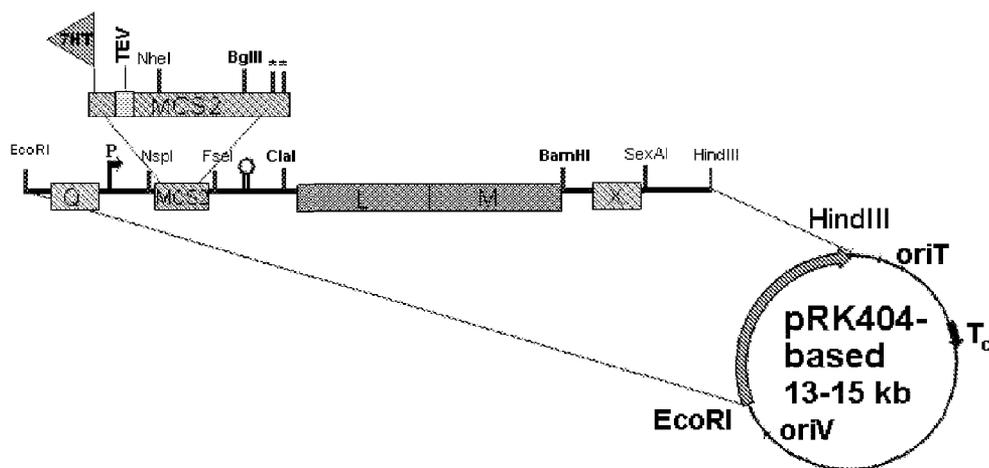


FIG. 7

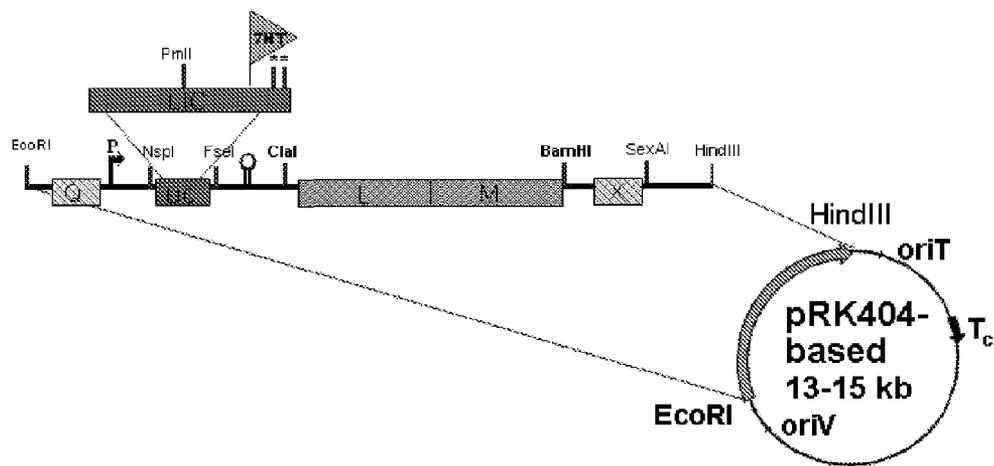


FIG. 8

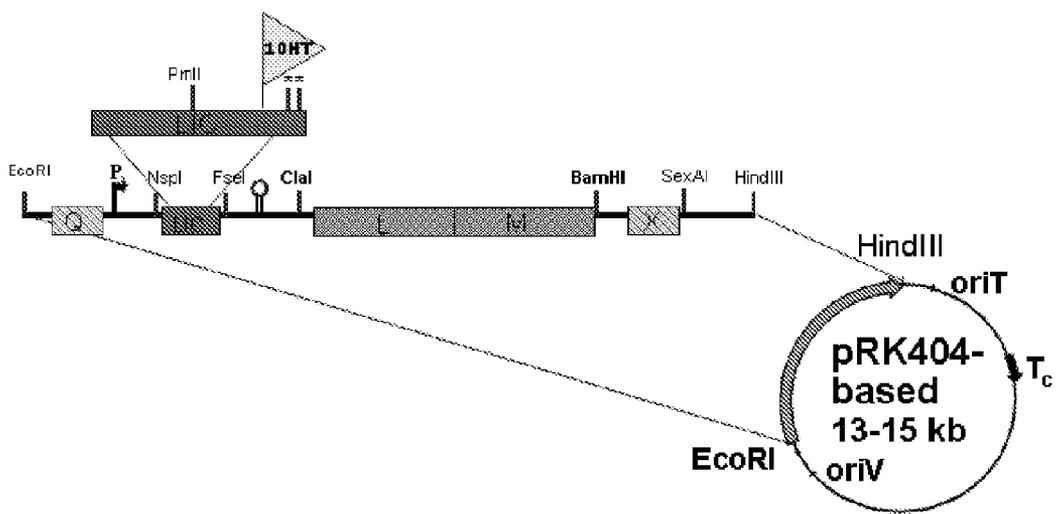


FIG. 9

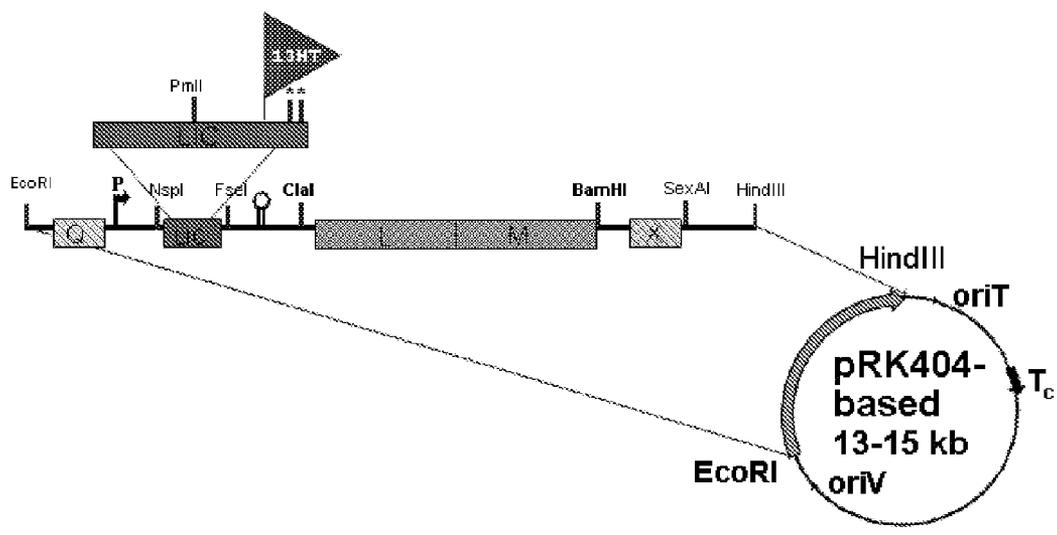


FIG. 10

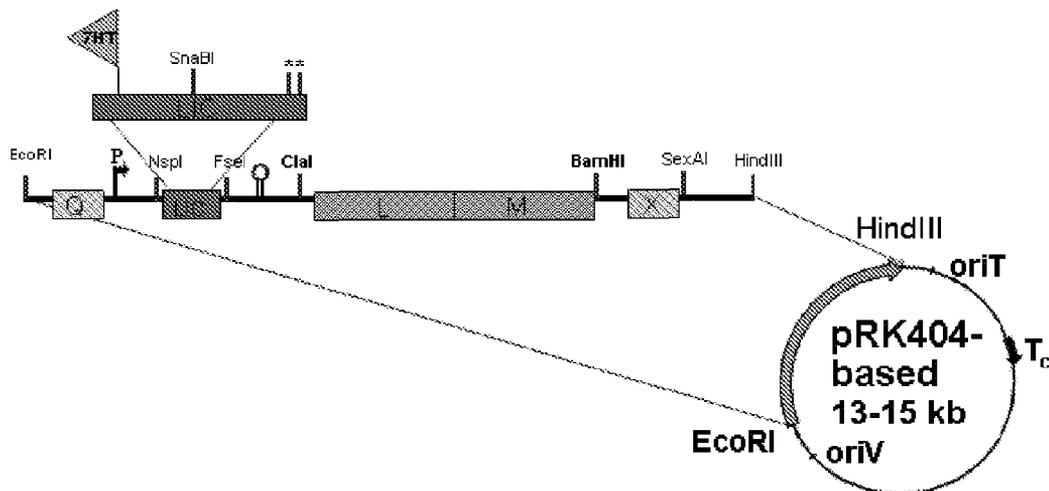


FIG. 11

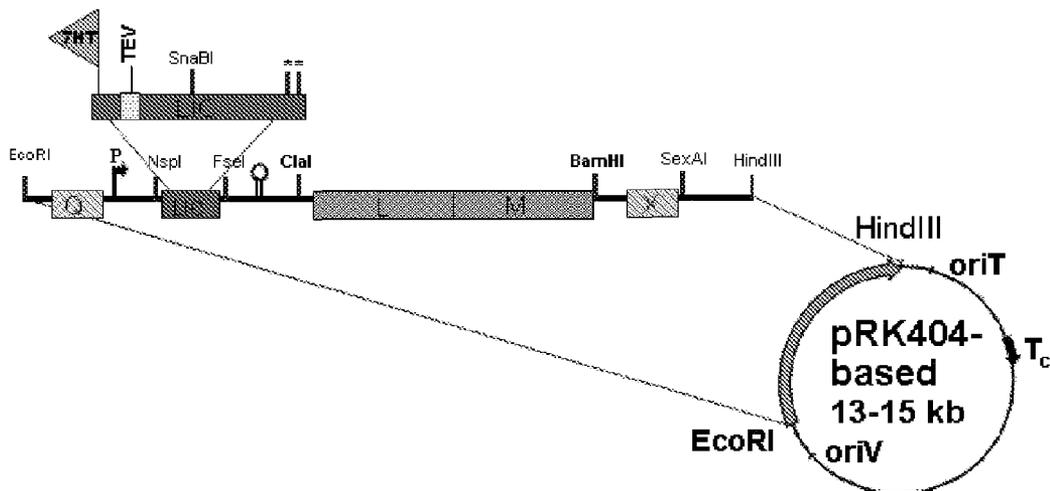


FIG. 12

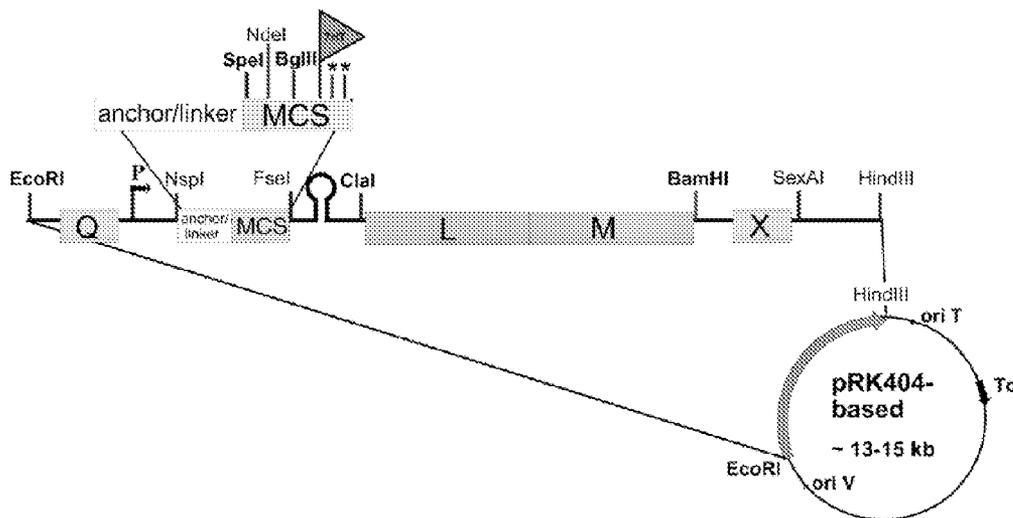


FIG. 13

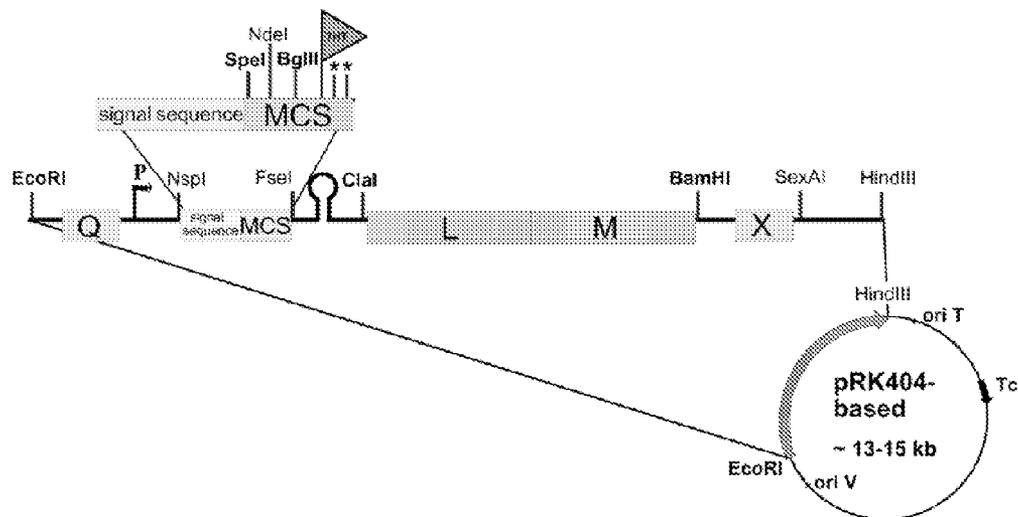


FIG. 14

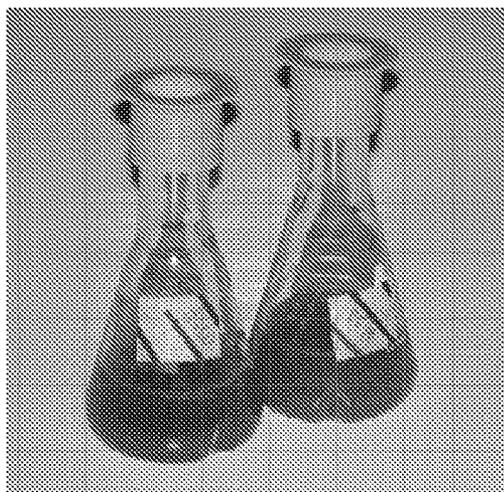


FIG. 15

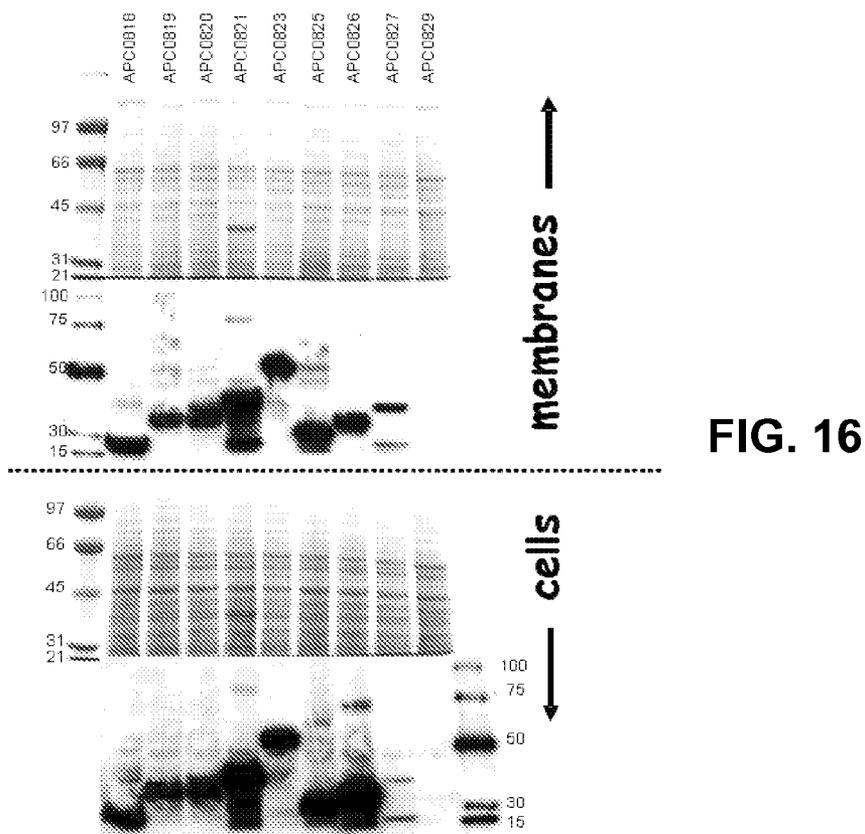


FIG. 16

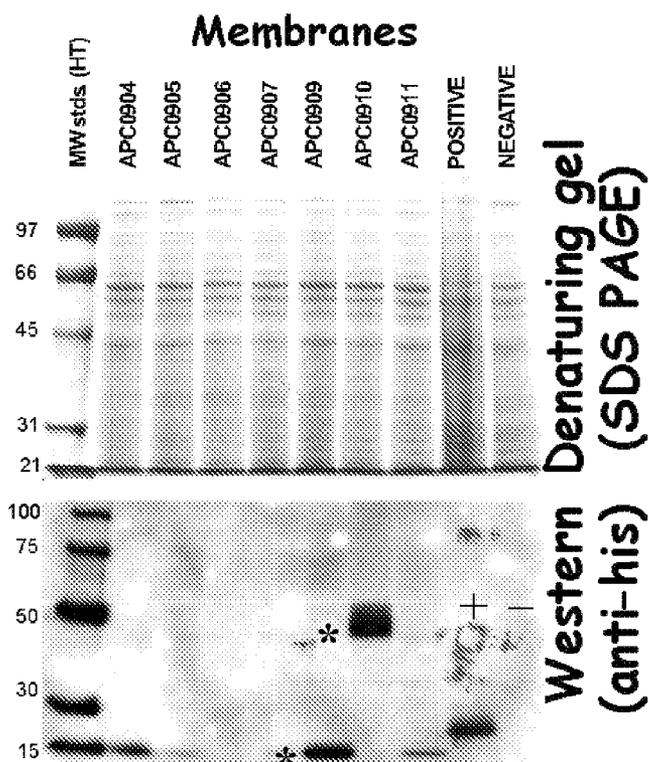


FIG. 17

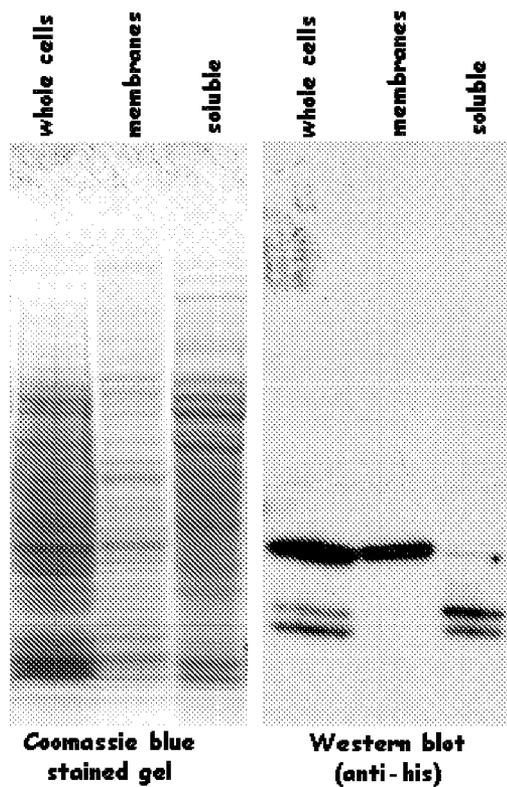


FIG. 18

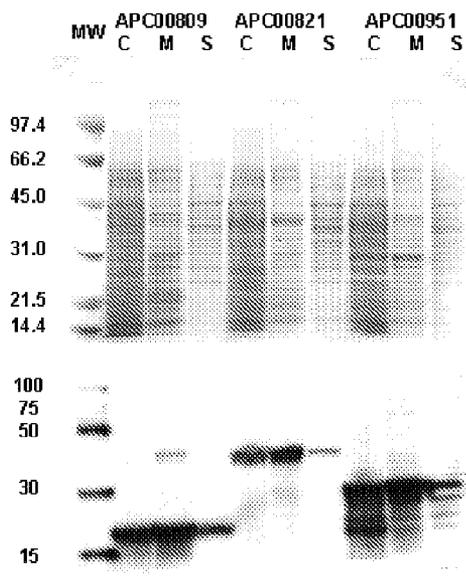


FIG. 19

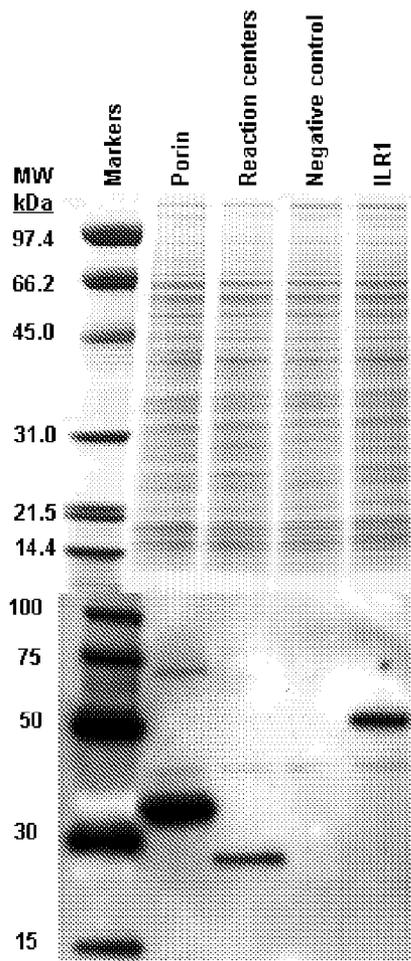


FIG. 20

Typical primers for ligation-dependent cloning

APC00809.top ($T_m = 59.7^\circ \text{C}$)
 5' - CTGTACTAGTGGAGGATGGTATGAAGCGCAAAGTATTGTTAATTCCG - 3'
 APC00809.bot ($T_m = 66.4^\circ \text{C}$)
 5' - TGEAAGATCCTTGTGCGGCCTCCTTACTGTATTTCTCCC - 3'

FIG. 21

Typical primers for ligation-independent cloning

APC00809.top ($T_m = 59.7^\circ \text{C}$)
 5' - AGGCAGGAGGAACACTATGAAGCGCAAAGTATTGTTAATTCCG - 3'
 APC00809.bot ($T_m = 66.4^\circ \text{C}$)
 5' - GACCACCGGACGACACTGTTGTGCGGCCTCCTTACTGTATTTCTCCC - 3'

FIG. 22

PCR fragment

AGGCAGGAGGAACACT	ATG • Target • CAA	GTGTCGTCGGTGGTC
TCCGTCCTCCTTGTGA	TAC • protein • GTT	CACAGCAGGCCACCAG

↓ T4 DNA polymerase + dATP

AGGCAGGAGGAACACT	ATG • Target • CAA	
A	TAC • protein • GTT	CACAGCAGGCCACCAG

FIG. 23

VERSATILE VECTORS FOR EXPRESSION OF FOREIGN PROTEINS IN PHOTOSYNTHETIC BACTERIA

[0001] This applications claims priority to U.S. Ser. No. 60/721,423 filed Sep. 28, 2005.

[0002] The United States Government has rights in this invention under Contract No. W-31-109-ENG-38 between the United States Department of Energy and the University of Chicago representing Argonne National Laboratory.

BACKGROUND

[0003] Methods and compositions are used for expressing foreign (heterologous) genes in photosynthetic organisms and sequestering and isolating resulting heterologous proteins. Versatile vectors facilitate the cloning and expression of a wide variety of natively-folded, functionally intact target proteins in photosynthetic organisms.

[0004] "Proteins" include soluble proteins and membrane proteins. Although many systems exist for the overexpression of soluble proteins for their input into structural and functional studies and applications, some target soluble proteins prove problematic when expressed heterologously. If soluble proteins are produced too quickly or are exposed to an environment that disfavors their folding in such systems, they are often found to form large-order aggregates and precipitate, or are rapidly degraded by inherent host machinery designed to maintain order within the cell. With limited ability to change the membrane content of host cells employed routinely as expression vehicles, these soluble proteins, in most cases, are abandoned because no satisfactory system or condition can be found whereby they can be produced in useful quantity or quality. Many of these 'problematic' soluble proteins could benefit from an increased volume of the cytoplasmic membrane with which they can associate.

[0005] Membrane proteins are extremely important for normal cell function. They provide the means by which cells communicate, transduce signals and transport metabolites between internal compartments, and build gradients of ions which are used to fuel all ingrained activities. Membrane proteins are one of the early defenses against invading foreign organisms.

[0006] Although roughly 35% of the proteins known or expected to be found in most organisms are membrane-associated, little structural or functional information exists on these proteins relative to soluble proteins. New information on membrane protein structures would aid biologists, physicists and chemists in their understanding of important structural relationships necessary for essential protein functions in lipid bilayer environments and could provide strategies to develop drugs that need to interact with membrane functions. Quantities of native membrane-associated proteins are difficult to purify in quantities sufficient for analysis. Inasmuch as the functional properties and stability of membrane proteins are dependent upon the lipid bilayer surrounding them, these proteins often denature or otherwise deviate from their native states when removed from their natural environs. Additionally, most membrane proteins are often expressed at very low levels, in amounts insufficient for purification and crystallization. To date, the three dimensional structures of only about 60 unique membrane proteins

are known, in comparison to the structures of representatives of more than 4000 families of soluble proteins.

[0007] Knowledge of the structures, and a determination of the functions, of membrane proteins would contribute greatly to understanding of biological processes and facilitate applications for clinical use. For example, structure-based rational drug design has produced powerful competitive inhibitors of cofactor binding in enzyme catalysis.

[0008] Because of their importance in cellular functions that can contribute to various disease states, membrane proteins are targets for drug discovery that impacts disease control and prevention.

[0009] Purification of membrane proteins from their host cells has been attempted by removing the proteins from hydrophobic surroundings and placing them in small detergent micelles which attempt to mimic the lipid environment. Following this solubilization process, routine chromatography or precipitation techniques (which have been perfected for soluble proteins) are utilized to purify and crystallize the solubilized membrane proteins. However, such adaptations rarely yield large amounts of the membrane protein in functional form.

[0010] Efforts have been made to create a process whereby membrane-associated proteins are over-expressed and subsequently purified from host cells of another organism (i.e., heterologous expression). To some degree, these efforts have all utilized a combination of a desired coding sequence with a foreign promoter known to induce high levels of protein synthesis. Fusion proteins, comprising a coding sequence of a desired (target) protein and the coding gene sequence of an affinity peptide, wherein the affinity peptide is attached and used to purify the desired protein product, are reported. This process provides an additional means of purifying the desired protein through chemical or enzymatic cleavage at a strategic cleavage site. No provision for maintaining the intact, tertiary and quaternary structure of the desired hydrophobic protein is reported. Purification is accomplished using for example, metal chelate affinity chromatography in nitrilotriacetic acid resins. However, no provision exists for circumventing the unique and inherent difficulties associated with purifying intact hydrophobic proteins.

[0011] Heterologous overexpression of hydrophobic proteins has been reported when coding regions of desired membrane proteins are juxtaposed with the bacterio-opsin (bop) regulatory sequences in the cell membrane of *Halo-bacterium salinarum*. However, the process does not provide for simultaneous production and sequestration or compartmentalization of the desired protein.

[0012] A heterologous overexpression system based on *Hansenula polymorpha* suggests the utilization of peroxisomes in which produced proteins may accumulate. However, as with the *H. salinarum* system, no provision exists for the simultaneous production and compartmentalization of the targeted components, inasmuch as the promoters utilized therein are for the most part constitutive.

Photosynthetic Organisms May be Hosts for Heterologous Expression of Proteins.

[0013] Members of the *Rhodobacter* genus are extremely robust and among the most versatile organisms known to

biology. These bacteria are characterized by a metabolic diversity that allows them to adapt readily to a wide variety of environmental conditions.

[0014] They thrive in dark or well-lit environments, in the presence or absence of oxygen. They can biochemically exploit an assortment of substrates for cell growth and division, or can harvest energy from the sun for that same purpose. As an example, single members of the genus *Rhodobacter* are known to reduce nitrogen compounds, fix carbon dioxide, utilize carbon sources in an aerobic environment, or grow photosynthetically under anaerobic conditions—depending on resources available in their immediate vicinity. The mechanisms by which environmental cues are sensed and are used to turn on or off the biochemical machinery necessary to survive in a particular setting are complex, as is the composition of the membranes in this organism.

[0015] A heterologous overexpression system based on *Rhodospirillum rubrum* has been reported whereby proteins can be expressed under control of the regulatable promoters of the *puf* and *puc* operons. The photosynthetic apparatus in this organism is less evolved than *Rhodobacter* and lacks the *puc* operon encoding the structural genes of the peripheral (and highly abundant in low light regimes) light-harvesting antenna. The latter operon in *Rhodobacter* species is controlled by changes in both oxygen tension and light intensity, and the transmembrane proteins encoded by it are widely utilized for survival in marginal photoautotrophic conditions.

[0016] Proteins associated with the inner membranes of *Rhodobacter* cells (those proteins that adhere to, span, or are tethered to the membrane) are quite dynamic and are a key feature of the multifaceted nature of the organism. The robust nature of photosynthetic organisms such as *Rhodobacter* and their complex and dynamic membrane systems are potential cellular factories for the production of foreign proteins.

SUMMARY

[0017] Methods and compositions for simultaneous production and sequestration of a wide variety of heterologous (foreign) proteins in photosynthetic organisms employ the design and construction of: 1) vectors carrying extended affinity tags for improved efficiency in protein purification, wherein “extended” means more than 7 residues, 2) vectors that vary in the placement of the affinity tags within the coding sequence to maintain structural and functional integrity, 3) vectors that incorporate cleavable affinity tags to yield a protein following purification that is as native as possible for structural and functional analyses; and 4) vectors that enable ligation-independent cloning (LIC) of target sequences to enable adaptation of the methods to high-throughput screening scenarios.

[0018] A method for expressing heterologous proteins in photosynthetic organisms such as the *Rhodobacter* species, includes the steps of producing and sequestering the protein within an inducible intracytoplasmic membrane system, wherein the protein and membrane are produced simultaneously. The protein may contain a plurality of affinity tags. The protein may be a complex of mutually co-dependent proteins. The expression of the heterologous proteins and the inducible membrane system may depend upon the same environmental stimuli.

[0019] The environmental stimuli activate a *puf* promoter or a *puc* promoter from the *Rhodobacter* genus. The coding sequence for the heterologous protein is inserted within the *puf* operon or the *puc* operon of the *Rhodobacter* genus. The inducible membrane system is controlled by the same environmental stimuli which induce expression of genes controlled by the *puf* promoter or the *puc* promoter of the *Rhodobacter* genus, including the target gene of interest. Examples of environmental stimuli include oxygen tension and light.

[0020] A method for producing and sequestering a functional protein within the *Rhodobacter* intracytoplasmic membrane wherein the expression of the membrane protein is under control of a *Rhodobacter* inducible promoter and wherein the functional protein is synthesized at the same time the sequestering membrane is synthesized includes the steps of:

[0021] a) supplying a DNA sequence containing the code for the target functional protein under control of a *Rhodobacter* inducible promoter and a host strain that produces inducible intracytoplasmic membranes in response to the same environmental cues; and

[0022] b) subjecting the resulting plasmid-bearing *Rhodobacter* strain to the environmental cue.

[0023] In the expression systems disclosed herein, auto-induction of both protein synthesis and intracytoplasmic membrane synthesis occurs as, for example, the oxygen tension of the culture decreases as the cell density increases.

[0024] The *Rhodobacter* Expression System offers several additional advantages over *E. coli*-based alternatives. The experimenter has more control over rates of protein expression in *Rhodobacter* than in *E. coli*. Slower, but nonetheless complete, induction in *Rhodobacter* is possible since this process is automatic and is controlled by oxygen tension. The resulting kinetics of induction and semi-aerobic growth rate are correspondingly slower in *Rhodobacter* and may shift the equilibrium towards the production of the folded, functional state of the target protein. Furthermore, induction in the *Rhodobacter* system is accompanied by the concomitant synthesis of new intracellular membranes which are available to newly synthesized proteins. If the target protein requires membrane association for its integrity, *Rhodobacter* possesses a greatly increased membrane surface area with which it can interact.

[0025] A method for purifying transmembrane proteins appends an affinity tag to the protein. An advantage is that the tag facilitates simple, rapid, and less disruptive extraction of the formed protein from its native membrane environment so that the protein retains its structural and functional integrity.

[0026] DNA sequences that transcribe mRNA, include a *puf*-promoted or a *puc*-promoted gene that results in a stable transcript and the translation of biologically active polypeptides linked to an affinity peptide that will also result in the simultaneous isolation/purification of the polypeptides in their native state.

[0027] Generally, light induced growth is facilitated in photosynthetic bacteria through the absorption of photons by specialized light-harvesting (LH) complexes, known as

antennae. These antennae transfer excited states to reaction centers (RC) where primary charge separation occurs.

BRIEF DESCRIPTION OF THE DRAWINGS

[0028] FIG. 1: Diagram of the broad-host-range expression vector used for ligation-dependent cloning with 7-member, C-terminal histidine tag. The *puf* operon of the *Rhodobacter* species of photosynthetic bacteria encodes six transmembrane proteins of the photosynthetic apparatus, that has been cloned into a broad-host-range vector (based upon pRK404). To facilitate the expression of foreign genes, the highest expressed native genes (*pufB* and *pufA*) have been replaced by a multiple cloning site (MCS). A region of stable hairpin structures is located between the MCS and the *pufL* gene; the major oxygen-regulated *puf* promoter is indicated (P). Foreign genes are mobilized into this vector by standard restriction endonuclease and ligation strategies after proper amplification. Foreign genes are fused in frame to a vector-based, seven-membered, C-terminal histidine tag (7HT) followed by appropriate translation terminators (stop codons; *). Restriction sites in bold are unique in this expression vector.

[0029] FIG. 2: Diagram of the broad-host-range expression vector used for ligation-dependent cloning with 10-member, C-terminal histidine tag.

[0030] FIG. 3: Diagram of the broad-host-range expression vector used for ligation-dependent cloning with 13-member, C-terminal histidine tag.

[0031] FIG. 4: Diagram of the broad-host-range expression vector used for ligation-dependent cloning with 7-member, N-terminal histidine tag.

[0032] FIG. 5: Diagram of the broad-host-range expression vector used for ligation-dependent cloning with 10-member, N-terminal histidine tag.

[0033] FIG. 6: Diagram of the broad-host-range expression vector used for ligation-dependent cloning with 13-member, N-terminal histidine tag.

[0034] FIG. 7: Diagram of the broad-host-range expression vector used for ligation-dependent cloning with 7-member, N-terminal histidine tag followed by a protease site for cleavage by such from Tobacco Etch virus (TEV).

[0035] FIG. 8: Diagram of the broad-host-range expression vector used for ligation-independent cloning (LIC) with 7-member, C-terminal histidine tag.

[0036] FIG. 9: Diagram of the broad-host-range expression vector used for ligation-independent cloning with 10-member, C-terminal histidine tag.

[0037] FIG. 10: Diagram of the broad-host-range expression vector used for ligation-independent cloning with 13-member, C-terminal histidine tag.

[0038] FIG. 11: Diagram of the broad-host-range expression vector used for ligation-independent cloning with 7-member, N-terminal histidine tag.

[0039] FIG. 12: Diagram of the broad-host-range expression vector used for ligation-independent cloning with 7-member, N-terminal histidine tag followed by a protease site for cleavage by such from Tobacco Etch virus.

[0040] FIG. 13: Diagram of the broad-host-range expression vector used for ligation-dependent cloning with an N-terminal membrane anchor/linker domain and a 7-member, C-terminal histidine tag.

[0041] FIG. 14: Diagram of the broad-host-range expression vector used for ligation-dependent cloning with an N-terminal, cleavable signal sequence and a t-member, C-terminal histidine tag.

[0042] FIG. 15: Small volume (80 mL) cultures of expression strains of *Rhodobacter* are grown semi-aerobically. Coordinated synthesis of target protein and membrane is autoinduced as oxygen tension lowers when the cell density increases.

[0043] FIG. 16: Screening for successful *Rhodobacter* expression and ICM insertion using 80-mL cell cultures grown semi-aerobically. Whole cell lysates (bottom) and membrane fractions (top) are extracted and analyzed using Western blotting techniques (anti-His; Novagen). Overexpressed bands are not always clearly visible in Coomassie-stained gels. Westerns are conclusive, general, and help identify membrane proteins with anomalous mobility on gels.

[0044] FIG. 17: Quantitation of heterologous expression of membrane in *Rhodobacter*. Western blots (anti-His antibody; Novagen) with well-characterized controls are employed to probe the level of expression of membrane proteins from *E. coli* (APC#s) in *Rhodobacter* ICMs. Experimental membranes from expression strains are compared with membranes carrying his-tagged reaction centers (expressed at 1 mg/L culture; +control) and with membranes from a recombinant strain lacking a cloned gene (-control). Any target protein expressed at or above the +control level is considered a "hit" (the two * in the above gel and blot). The expression levels of some targets rival those of native ICM proteins that can be purified to yields of >10 mg/L culture.

[0045] FIG. 18: Determination of the cellular localization in *Rhodobacter* of heterologously expressed membrane proteins is simplified by the presence of the polyhistidine tag. In this analysis, data are presented on an equal volume basis rather than on an equal protein basis. Here, target protein (APC 951) is found almost exclusively in the membranes. Target protein that is found in the soluble fraction results from small membrane fragments that do not pellet during ultracentrifugation; co-purifying host proteins reside quantitatively in the soluble fraction.

[0046] FIG. 19: Determination of the cellular localization in *Rhodobacter* of membrane proteins heterologously expressed from pRKLICHT1Dpuf. No differences in expression levels are apparent when comparing results from the same gene expressed from pRKPLHT1Dpuf or pRKLICHT1Dpuf.

[0047] FIG. 20: Determination of the level of *Rhodobacter* production of a soluble protein, ILR1, derived from *Arabidopsis thaliana*. Analysis from Western blots of whole cell lysates, where signals from the target protein can be compared to signals from well-characterized controls (porin, a β -barrel membrane protein expressed at >10 mg/L cell culture, and reaction centers produced by an engineered strain where expression has been downregulated to 1 mg/L

cell culture), suggests that ILR1 is produced in *Rhodobacter* cells at a level that equals or exceeds 2 mg/L of cell culture.

[0048] FIG. 21: A typical set of oligonucleotides used to amplify and subsequently clone a target membrane protein gene (APCO0809) into pRKLICHT1Dpuf. Boxes depict the LIC overhangs generated by T4 DNA polymerase digestion; underlined bases denote the ribosome binding site; long dashed lines represent the regions that are complementary to the template (target gene); and in circle are non-complementary, obligate bases that are necessary for generation of the LIC overhang.

[0049] FIG. 22: A typical set of oligonucleotides used to amplify and subsequently clone a target membrane protein gene (APCO0809) into pRKPLHT1Dpuf. Boxes depict "dummy" bases that enable efficient digestion of the amplicon by the restriction enzymes;

[0050] circle denotes the SpeI site, black and underlined is the consensus *Rhodobacter* ribosome binding site (RBS); arrow is the six base spacer between RBS and initiation codon; long dashed line is the region complementary to the template (target gene); and dotted line denotes the BglII site.

[0051] FIG. 23: A typical set of oligonucleotides used to amplify and subsequently clone a target membrane protein gene (APCO0809) into pRKLICHT1Dpuf. Box depicts the LIC overhang that is generated by T4 DNA polymerase digestion; underlined bases denote the ribosome binding site; long dashed line represents the region that is complementary to the template (target gene); and in circle are non-complementary, obligate bases that are necessary for generation of the LIC overhang.

[0052] FIG. 24: LIC handles for amplified target genes compatible for insertion into pRKLICHT1Dpuf are generated by 3'-5' exonuclease activity of T4 DNA polymerase in the presence of excess dATP. The resulting T_m s of the overhangs are sufficient to allow the transformation of competent *E. coli* to tetracycline resistance after a brief annealing process at room temperature.

[0053] FIG. 25: Strategy employed to clone target membrane protein genes into pRKPLHT1Dpuf using restriction enzymes SpeI and BglII. This vector is designed to fuse a C-terminal, seven-membered histidine tag. The protein sequence of the tag and "linker" amino acids are shown in single letter code.

[0054] FIG. 27: LIC strategy employed to clone target membrane protein genes into pRKLICHT1Dpuf using semi-automated methodologies. This vector is designed to fuse a C-terminal, seven-membered histidine tag (such is partially displayed). "Linker" residues between target gene and tag are shown.

DETAILED DESCRIPTION

[0055] Expression of heterologous (foreign) proteins is achieved the introduction of gene sequences encoding the proteins into photosynthetic organism e.g. bacteria of the genus *Rhodobacter*.

[0056] A suite of expression vectors makes the *Rhodobacter* membrane protein expression system a versatile tool for functional and structural studies and possibly large-scale structural and functional genomics efforts. These vectors limit interference of affinity tags in the native folding of the

target protein, thus helping it retain its native structure and function and increasing expression yields. If the tag's placement and composition cannot achieve this goal, then other vectors include sites enabling removal of affinity tags following purification of the protein of interest. Other vectors are engineered to facilitate cloning of the target gene in a manner that is not dependent upon restriction endonuclease digestion, enabling the cloning of genes that would otherwise be excluded because they contain sites for the cloning enzymes within their coding sequences.

[0057] The broad-host-range expression vectors for *Rhodobacter* disclosed herein, include vectors with extended tags engineered to be positioned at the C-terminal or at the N-terminal end of a protein of interest (a target protein). Some of these extended tags also include a cleavable peptide moiety that is recognized by a peptide cleavage enzyme to separate the tag from the protein of interest during purification. Some of these vectors have cloning sites that enable ligation independent cloning (LIC) of a nucleic acid sequence encoding a protein of interest into the vector. Others append native-ICM-protein-derived membrane anchors and signal sequences that help to target the membrane to a particular cellular compartment to increase levels of expression of target proteins in functional form. These broad-host-range expression vectors disclosed herein include one of the features listed below (illustrated in FIGS. 1-14 and sequences disclosed herein):

[0058] (i) N-terminal 7×His tag for ligation independent cloning

[0059] (ii) N-terminal 7×His tag followed by cleavage site for Tobacco Etch Virus protease for ligation-independent cloning

[0060] (iii) N-terminal 7×His tag

[0061] (iv) N-terminal 7×His tag followed by cleavage site for Tobacco Etch Virus protease

[0062] (v) C-terminal 10×His tag for ligation-independent cloning

[0063] (vi) C-terminal 13×His tag for ligation-independent cloning (14386 bp)

[0064] (vii) C-terminal 7×His tag for ligation-independent cloning

[0065] (viii) C-terminal 10×His tag

[0066] (ix) C-terminal 13×His tag.

[0067] (x) Broad-host-range expression vector used for ligation-dependent cloning with an N-terminal membrane anchor/linker domain and a 7-member, C-terminal histidine tag.

[0068] (xi) Broad-host-range expression vector used for ligation-dependent cloning with an N-terminal, cleavable signal sequence and a 7-member, C-terminal histidine tag.

[0069] The N-terminal or C-terminal tag can include any affinity tag that is of suitable length to promote better access to a purification system, such as for example, an immobilized metal ion affinity chromatography (IMAC). The N-terminal or C-terminal tag can also include a spacer or a linker that provides extended length for an affinity tag. For example, an affinity tag can include a spacer or a linker and a stretch of 6 or 7 histidine residues (spacer/linker plus

his-tag). The spacer or a linker generally includes a stretch of random or non-random amino acids. The spacer or linker can range from about 1 to about 50 amino acids; 1 to about 20; 1 to about 10; or 1 to about 5 amino acids in length. The spacer or linker may or may not exhibit affinity for a purification system. The spacer or linker in combination with an affinity tag can range in length from about 5 amino acids to about 50 amino acids; from about 10 amino acids to about 20 amino acids; or from about 15 amino acids to about 30 amino acids in length. An affinity tag may also have a longer stretch of affinity residues.

[0070] Broad-host range vectors are capable of replicating in more than one host species. For example, vectors disclosed herein are capable of replicating in *E. coli*, *Rhodobacter* and other host species.

[0071] Host strain *Rhodobacter sphaeroides* $\Delta\Delta 11$ (Delta-Delta11) was deposited in the ATCC and the accession number is designated as PTA-5921. This host strain has an increased capacity for incorporating heterologously expressed membrane proteins into its intracytoplasmic membranes. This engineered host lacks three native transmembrane complexes of the photosynthetic apparatus that normally populate the intracytoplasmic membrane in the wild-type organism.

[0072] A protocol was developed to facilitate the parallel induction of foreign proteins and host membranes. A heterologous protein is created and encapsulated in its natural state. The protein can be a membrane protein, a membrane-anchored protein, a soluble protein, a protein targeted to a specific cellular compartment, one protein, separate proteins, or a complex of mutually co-dependent proteins, such as a multi-subunit membrane-associated protein complex.

[0073] Different promoters are suitable, which respond to the same environmental stimuli, by actuating target promoters, to simultaneously induce foreign protein formation and sequestration. The co-expressed intracytoplasmic membrane (ICM) serves as a means to simultaneously compartmentalize, and therefore segregate, the developing heterologous membrane-bound proteins from the majority of other cellular components. The system has also produced functional, soluble proteins from genes derived from an unrelated organism.

[0074] Broad-host-range plasmids/vectors have been engineered to facilitate the cloning, expression and purification process. Generally, the fragment of host-chromosomal DNA containing the operon for producing the LH/RC machinery is transferred to a vector. The gene for the desired protein is then inserted to replace one or more genes of the operon. When this expression plasmid is transferred back to the photosynthetic host organism, the target protein is generated when the culture is subjected to the environmental cues that are specific for activating the promoter of that particular operon.

[0075] In *Rhodobacter* species, cells become pigmented as the ICM develops. This new membrane takes the form of vesicles. ICM is contiguous with a cell membrane. The interior of these vesicles contains periplasmic components. For example, a region of the ICM houses the reaction center (RC), which in photosynthetic organisms comprises a central complex of pigments and proteins. The RC is comprised of three separate components, or subunits, called H (heavy),

M (medium) and L (light) based on the way these units migrate in an electric field. RC complexes house the cofactors of the photosynthesis complex, which include bacteriochlorophylls, bacteriopheophytins, quinones and a non-heme iron.

[0076] Upon cell disruption, the vesicles break apart from the cell membrane, thus becoming sealed "inside-out" particles, termed chromatophores. These vesicles (basically ICM) are easily isolated by virtue of their size. Chromatophores are much smaller than cellular debris and thus remain soluble during low-speed centrifugation. Then, during brief ultracentrifugation, they are readily separable from cellular components in forming a pellet. This pellet is rich in ICM. Therefore, proteins residing in the ICM are already significantly purified following these two simple fractionation steps with a total duration of typically less than two hours, and often less than one hour. Target proteins which are truly soluble will be found in either the cytoplasm or the periplasm. The supernatant from this brief ultracentrifugation contains both of these cellular compartments and would be used as starting material for the purification of the majority of the 'problematic' soluble proteins expressed in this system.

[0077] *Rhodobacter* produces large quantities of membrane that is filled with proteins of the photosynthetic apparatus. Using methods and compositions described herein, the photosynthetic proteins are replaced with foreign proteins. The *Rhodobacter* genus of photosynthetic bacteria can produce large quantities of intracytoplasmic membrane;

[0078] placing the expression of heterologous proteins under control of a promoter that controls synthesis of intracytoplasmic membrane components induces expression of the heterologous protein as well. Among the bacteria in the *Rhodobacter* genus, *R. sphaeroides* and *R. capsulatus* are suitable for use in the protein production and isolation method disclosed herein.

[0079] The *Rhodobacter* genus of photosynthetic bacteria can be grown in a variety of conditions, such as anaerobic, semi-aerobic, aerobic, light or dark. This is because the cytoplasmic membrane in *Rhodobacter* contains components of the respiratory chain, transport systems, and other energy-transducing complexes. The physiology of this genus under each of these conditions is different.

[0080] For example, when *Rhodobacter* cultures are switched from aerobic chemotrophic conditions to phototrophic growth conditions, large quantities of a new intracytoplasmic membrane (ICM) that houses the newly synthesized photosynthetic machinery are induced. This ICM is formed as invaginations of the cytoplasmic membrane and in its nascency, is contiguous with the cytoplasmic membrane. Since it houses the newly synthesized photosynthetic machinery of the cell, the lipid, chemical, and protein composition, and its kinetics of biogenesis differ from the cytoplasmic membrane.

[0081] *Rhodobacter* can also be induced to synthesize ICM in dark-grown cultures which are limited for oxygen, since this stimulus also directs the organism to prepare for a switch from oxidative phosphorylation to anaerobic phototrophic growth.

[0082] *Rhodobacter* is induced to synthesize ICM and ICM-protein, either native or foreign. During cell disruption,

the ICMs break away from the cytoplasmic membrane to become discrete entities with physical properties that are different from other cellular components. Inasmuch as the cells become pigmented as these ICMs form, this phenomenon was exploited to indicate the presence of heterologous proteins formed concomitantly with the ICMs. Therefore, the heterologous proteins residing in the ICMs are easily isolated from other protein-containing cellular fractions.

[0083] To facilitate heterologous protein purification (through isolation of the heterologous protein from other ICM components), an affinity tag is engineered into the protein-coding sequence. The affinity tag is used to readily sequester the heterologous proteins in native form by chromatography with the correspondingly compatible resin. This results in a 4-5 hour purification protocol, versus the more than three day isolation procedure provided by the state-of-the-art for the purification of unengineered proteins from native hosts.

Intracytoplasmic Membrane and *Rhodobacter* Operon Details

[0084] The intracytoplasmic membrane (ICM) is formed when photosynthetic bacteria are switched from chemotrophic conditions to phototrophic growth conditions or when grown in the absence of light and limited oxygen.

[0085] The ICM forms from invaginations of the cell membrane and is thus contiguous with the cell membrane, while also having different characteristics vis-a-vis the cell membrane. The ICM differs from the cell membrane in its kinetics of biogenesis. Specifically, the ICM forms when ICM-protein is being actively expressed and folded, an event which occurs separate from the formation of the cell membrane.

[0086] The majority of natural ICM protein belongs to three transmembrane protein complexes of the photosynthetic apparatus: the reaction center (RC) and the two different light harvesting complexes (LH1 and LH2). The *puf* operon, and specifically the *puf* promoter, coordinates expression of Light Harvesting Complex 1 (LH1) and RC complexes. The *puc* operon, coordinates expression of the Light Harvesting Complex 2 (LH2), via its *puc* promoter. The *puf* operon will be discussed first.

[0087] The *puf* operon encodes six transmembrane proteins, specifically the two subunits of the LH1 complex, (the genes for the subunits represented as A and B in the drawing, respectively), the L and M subunits of the RC complex, and two regulatory proteins, PufQ and PufX, which are present in small amounts in the membrane. A region of stable hairpin structures is located between the *pufA* and *pufL* genes. While the *puc* promoter for the LH2 complex is controlled by both light and oxygen, the *puf* promoter, located upstream of *pufQ*, directs synthesis of RC and LH1 complex and is controlled solely by oxygen tension. At high oxygen tensions, the *puf* operon is repressed. When the oxygen tension is lowered, transcription of the *puf* operon is induced, and the transmembrane proteins that it encodes are produced in relative stoichiometries, determined in part by mRNA stability. The hairpin structure located between *pufA* and *pufL* confers this stability to varying degrees by protecting the transcript from exonuclease digestion, according to the positions of *puf* genes relative to its own location. The hairpin structures serve as a means for blocking exonuclease

action beyond the location of the hairpin. The result of this blocking mechanism is an increase in mRNA stability leading ultimately to production of a larger quantity of the protein of interest. The LH1-B and LH1-A proteins are present in 15-20 fold excess over the RC-L and RC-M subunits because the stable hairpin structure prevents degradation of the mRNA of the former.

[0088] All of the *puf* operon proteins are inserted into the developing ICM, whose synthesis is induced coordinately.

[0089] Transcription of the operon and synthesis of the ICM is induced in the lab by growing cells under semi-aerobic, chemoheterotrophic conditions in the dark per the protocols provided herein. Under these conditions, complexes of the photosynthetic apparatus are synthesized and assembled and the ICM is produced even though the cell is not using these components to grow.

[0090] The *R. sphaeroides* operon is cloned into a modified version of broad-host-range vector pRK404, an 11.2 kb derivative of pRK292 which carries the polylinker from pUC9 and tetracycline resistance. It is transferred to *Rhodobacter* via conjugation with *E. coli* donor strain S17-1; its copy number in *Rhodobacter* strains is 4-6/cell. Plasmid pRK404 was subsequently engineered to remove a second EcoRI site, and the HindIII site in the polylinker has also been removed to leave a single HindIII site within the *puf* operon. This modified vector is designated pRK442(H). These modifications facilitated the shuttling of singly- or multiply-mutated L and M genes in and out of the plasmid. For expression of mutant or wild-type RCs, plasmid pRKHT (or a derivative of it) is used to complement, in trans, a strain of *R. sphaeroides* ($\Delta\Delta 11$) that carries an engineered deletion of the chromosomal copy of this operon. The genes for the LH2 complex are also deleted in strain $\Delta\Delta 11$, thus the phenotype of this strain is LH1⁺ LH2⁻ RC⁺.

[0091] Site-specific mutagenesis is used to append a seven-histidine tail to the C-terminus of the M subunit of RCs of *R. capsulatus*. This tail is on the periplasmic surface of the pigment-protein complex and associates with Ni- or Co-NTA (nitrilotriacetic acid) resin for rapid IMAC. Starting from a cell suspension, extremely pure RCs are isolated using a 4-5 hour protocol. The previous purification methodology took 3 days and produced complexes that were less pure.

[0092] Inasmuch as the His-tag improved the *R. capsulatus* RC purification so dramatically, it was added to a vector for production of *R. sphaeroides* RCs. This modification is useful for *R. sphaeroides* RCs because, unlike *R. capsulatus* RCs, the former have a greater propensity to form diffraction-quality crystals. To facilitate the addition of an analogous His-tag to this RC, an expression vector carrying a his-tagged *R. sphaeroides* M gene was obtained. In a multi-step cloning strategy, the His-tagged M gene was added to the *R. sphaeroides* system for site-directed mutagenesis.

[0093] By coupling this expression system with the IMAC purification protocol discussed herein, large quantities of exceptionally pure RCs from both mutant and wild-type strains of *R. sphaeroides* are obtained.

[0094] Based on the success of the system for expressing native and mutant RCs, the expression vectors were modified to facilitate the heterologous expression of any target gene in *Rhodobacter*. These engineered vectors are designed

to place expression of a foreign gene under control of the oxygen-regulated *puf* operon promoter (P). The position of the gene relative to the region of stable hairpin structure in the operon dictates the relative level of expression. A multiple cloning site (MCS) replaces genes of the LH1 complex for high-level expression of the foreign protein. A multiple cloning site that allows for insertion of the foreign gene in place of reaction center genes (L and M) to obtain a moderate expression level. Dual expression of two genes is possible by combining these strategies.

[0095] Other broad-host-range vectors, host-specific vectors, or vectors utilizing ligation-independent cloning (LIC) strategies are also appropriate vehicles to facilitate protein expression in trans. LIC protocols utilize the proof-reading capabilities inherent in some DNA polymerases to generate lengthy complementary cohesive ends between the insert and vector which when annealed in the absence of ligating enzymes yield molecules that transform organisms with high efficiency. Vectors containing an N-terminal membrane anchor/linker domain help to target a fused heterologously-expressed protein to the ICM.

[0096] In a similar manner, vectors have been designed to fuse an N-terminal, cleavable signal sequence to the coding sequence of the target protein in order to direct a soluble protein or the N-terminus of a membrane protein to the periplasmic space. Conjugation is utilized to shuttle LIC plasmids into *Rhodobacter*.

[0097] For extended tags, because many target proteins have low affinity for Ni-NTA resin with just a simple heptaoligomeric histidine tag fused to the end of its normal amino acid chain, in the absence of a linker, more residues were successful in improving adherence of target proteins to resin, allowing more quantitative removal of impurities that bind either non-specifically or with lower affinity to these columns; N-terminal tags solved the problem that some targets have a buried C-terminus that is inaccessible to chromatographic resin; extended cleavable tags addressed the problem that most structural biologists prefer to work with native protein in crystallization trials over 'inferior' products with tags still attached for crystallization trials; and LIC strategies eliminate concerns about sites for cloning enzymes within the sequence of the gene of interest and increase the speed by which expression constructs are generated.

[0098] Cloning with vectors featuring extended tags, N-terminal tags, extended cleavable tags, and designs enabling LIC strategies was achieved. Vectors employing N-terminal tag strategies work well and are a 'salvage' method employed if C-terminal tags are not satisfactory (e.g., expression yields are low, protein cannot be purified because tag is inaccessible, or tag prevents proper folding of the protein and thus disrupts its function);

[0099] cleavable tags are useful although the protease employed is not compatible with many commonly used surfactants; and the LIC strategy is such that for example, 96 expression clones are created in one experiment—a paradigm shift for previous results using *Rhodobacter* at a pace of one gene at a time.

[0100] For purification processes, associated with isolating the generated proteins, a suitable moiety with an affinity for a predetermined structure is appended to the generated

protein for subsequent separation. The His-tag improves the ability to purify and manipulate RCs for functional studies. A polyhistidine tail (HT) is inserted in frame at the C-terminus of the MCS before stop codons (*) which terminate protein translation. This HT expedites purification of the expressed protein. The histidine tag also can be attached to the N-terminus. Other tags also are appropriate, including, but not limited to intein, maltose binding protein, and small peptide tags with high-affinity antibody-based recovery systems. A myriad of suitable peptide tags is commercially available, including, but not limited to, E-tag™ of GE Healthcare, Inc., Piscataway, N.J., and the S-tag™ of Novagen, Inc., Madison, Wis. Any of the attached tags can be designed to be cleaved with a compatible protease.

[0101] The His-tag facilitates the use of different surfactants with a wider range of properties to remove the complex from its native membrane environment. For example, when IMAC protocols are used in combination with a mild charged detergent (which is incompatible with traditional ion exchange chromatography), the cofactors of the resulting product remain in their native states as evidenced by spectral properties—dimeric bacteriochlorophyll in *R. capsulatus* RCs absorbs at its native 870 nanometer position versus a shift to 850 nm when other detergents are utilized. Small crystals of His-tagged RCs of *R. capsulatus* were obtained.

[0102] IMAC was also used to isolate LH1RC superassemblies in large quantity for crystallization trials. The non-covalent association between the RC and LH1 is strong enough to allow purification of the entire superassembly utilizing the single poly-histidine tail on the RC. Crystals of the superassembly were obtained.

[0103] The His tag also enables the changing of surfactants after removal of the complex from the native lipid bilayer. The functional or structural integrity of the complex is maintained during the purification process. In fact, four different types of spectroscopic experiments that measure electron transfer, proton transfer, or energy transfer reactions in the RC have indicated that the poly-histidine tag does not interfere with the normal functions of the complex.

Methods for Purifying the Multi-subunit RC and the LH1/RC

[0104] Superassembly complexes with a single his-tag were adapted to exploit the *Rhodobacter* heterologous expression system to co-purify proteins which are members of larger membrane complexes. This adaptation requires and enables the simultaneous expression of interacting proteins. Genes for many proteins that associate into functional complexes are organized into conserved DNA segments. The ability to express clusters of mutually dependent proteins enables methods in which systematic co-expression of two or more membrane-associated proteins results in successful production of proteins and/or complexes heretofore recalcitrant to efforts of mono-molecular expression.

[0105] Coordinated expression of multiple genes is accomplished by shuttling a gene cluster, containing one gene that is affinity tagged (such as with histidine), into one of the above *Rhodobacter* expression plasmids. If the members of the cluster physically interact, the single protein which is affinity tagged will facilitate purification of the entire complex, thereby allowing for the identity of proteins

which associate to form a functional multi-subunit macromolecular membrane-associated machine. A vector that was designed for the tandem expression of two genes whose protein products associate in a stoichiometry other than 1:1.

Construction of Versatile Vectors

[0106] An expression vector the *puf* operon (FIG. 1) was cloned into pRK442, a modified version of the broad-host-range vector pRK404, an 11.2 kb derivative of pRK292 which carries the polylinker from pUC9 and tetracycline resistance. Later, a more generalized 'platform' version was engineered that allowed for introduction of foreign genes in place of structural genes of the photosynthetic apparatus. The best yield of heterologous expression was through extensive testing with several foreign genes obtained with a vector that placed a multiple-cloning-site (harboring recognition sequences for *SpeI*, *NdeI*, and *BglII*) in place of the *pufB* and *pufA* genes. Synthesis of the foreign protein is directed by the oxygen-/light sensitive *puf* promoter. Routinely, the foreign genes are amplified such that a *SpeI* site is inserted at the N-terminus and a *BglII* site is appended to the C-terminus. Cloning of the amplicon using these (or compatible) sites inserts the gene into the vector such that it is fused in frame to a C-terminal 7×His tag followed by two stop codons.

[0107] The platform vectors are based upon a large (11.2 kb) broad-host-range vector, pRK404, whose sequence was largely unknown. For ease in designing future constructs, the sequence of the pRK404 derivative being used in the project was determined [2], with assistance from MWG Biotech (Highpoint, N.C.). Knowledge of the vector sequence has been of extreme utility in design and construction of the later generations of expression vectors described herein. Because this vector is large, smaller, broad-host-range vectors (derivatives of pBBR1 were evaluated; [3-6]) that carry a variety of antibiotic resistance genes and extensive multiple cloning sites. Although higher copy number was expected, surprisingly, expression from these vectors was lower than those for genes borne on pRK404-based plasmids.

[0108] Genes for some target proteins may fail to encode compartmentalization signals that are recognized by the *Rhodobacter* host. Thus, a platform vector was constructed that encodes an N-terminal membrane anchor/linker domain derived from cytochrome *c₂* of *R. capsulatus* (13, FIG. 13). In addition, a vector containing a cleavable, N-terminal signal sequence derived from cytochrome *c₂* of *R. sphaeroides* (15, FIG. 14) was also constructed to enable targeting a soluble foreign protein or the N-terminus of a foreign membrane protein to the periplasmic space of the *Rhodobacter* host cell.

[0109] Platform vectors include affinity tags of altered composition and position. In order to accommodate a target protein whose C-terminus is completely or partially buried, platform vectors with 7-membered histidine tags fused in frame to the N-terminus were constructed. Vectors were constructed in which a site for Tobacco Etch virus (TEV) protease was inserted between the His tag and the start of the foreign gene. Cleavage of the tag results in the addition of three amino acids (SAS) to the N-terminus of the foreign gene. Vectors containing longer C-terminal tags with 10 or 13 consecutive histidines were also constructed and did assist with affinity purification of target proteins, because the

longer tags bind more tightly to immobilized metal resin and allow more quantitative removal of impurities that bind either non-specifically or with lower affinity to these columns.

[0110] Cloning of genes into the above platform vectors utilizes PCR and ligation methodologies. A vector that enables ligation-independent cloning of genes encoding foreign membrane proteins for expression in *R. sphaeroides* was designed, constructed, tested and employed in a plate-based automated manner wherein clones were generated in 96-well format utilizing a method based on microtiter plates. The LIC versions of vectors featuring extended tags, N-terminal tags, and cleavable tags have also been designed and/or constructed. FIGS. 2-14 are schematic representations of vectors.

Problematic Soluble Proteins

[0111] The *Rhodobacter* Expression System has been applied more generally to the expression of soluble proteins or multisubunit complexes thereof whose expression has proven to be especially problematic for *E. coli*-based expression systems. This new soluble protein strategy functions in the absence of the aforementioned membrane protein tether. The *Rhodobacter* system, in addition to serving as a tool for heterologous expression of membrane proteins, also offers utility for soluble protein expression.

[0112] This application of the *Rhodobacter* Expression System is especially important because large percentage of proteins in current structural genomics efforts (up to and possibly exceeding 50%) are "triated" when they prove to be expressed at low levels or primarily in insoluble forms in *E. coli*. Detection of expressed proteins with the anti-polyhistidine antibody has never indicated that expressed proteins form inclusion bodies in *Rhodobacter*. This is in sharp contrast to T7 polymerase-based *E. coli* expression systems, where high-level overexpression often results in aggregation and precipitation of incompletely folded polypeptides as inclusion bodies.

[0113] Soluble protein expression is accomplished with the same vectors and strategies that have already been used successfully or designed for use with membrane proteins. In an analysis of a wide range of target proteins, the *Rhodobacter* expression system handles adequately some problematic soluble proteins even IN THE ABSENCE of the membrane protein tether. The only small adaptation of the method is to purify proteins from the cytoplasm and/or periplasm (cell fractions combined as the supernatant from an ultracentrifuge spin at >100,000 g after cell breakage), thereby eliminating the need for solubilization steps or the inclusion of detergents in any chromatographic buffers.

[0114] Problematic, supposedly "soluble" proteins that failed in *E. coli* expression systems, have been produced successfully in soluble form—albeit with reduced yield—in the *Rhodobacter* system. These target proteins are currently being produced in large scale for crystallization trials.

[0115] Because it is capable of successfully expressing soluble proteins which are otherwise lost to inclusion bodies, this suite of expression vectors and hosts forms the basis for a likely "salvage strategy" that will improve the efficiency of existing structural genomics programs. It, thus, expands the versatility of the *Rhodobacter* Expression Sys-

tem as a vehicle that enables functional and structural studies (and possibly large-scale genomics efforts) for problematic target proteins.

EXAMPLES

[0116] The following examples are illustrative and do not limit the scope of the various methods and compositions disclosed herein.

Example 1

Small-Scale Screening for Expression and Localization of Target Protein in *Rhodobacter*.

[0117] For initial expression screening, the cells are grown in small culture volume, and the expression levels and cellular localization of the target protein are determined by Western blotting following SDS-PAGE. Coordinated synthesis of nascent membrane and target membrane protein is autoinduced by decreasing oxygen tension as the cell density increases during semi-aerobic culture. Those conditions are achieved as described below.

Small-Scale Growth and Preparation of Samples for SDS-PAGE.

Growth and harvest of expression strains

[0118] Cells are grown in 80 mL of YCC/tet_r medium in a 125 mL baffled flask (see FIG. 15). This flask is stoppered with a silicone sponge closure (Bellco Glass, Inc., Cat. No. 2004-00003).

[0119] Incubate at 32-34° C., shaking at 125 rpm, for 72-96 hours. Remove 5 mL of media to determine turbidity using a Klett-Summerson calorimeter. The equivalent OD₆₀₀ may also be used. After measurement is complete, refrigerate this sample in a 15 mL Falcon tube for later use in SDS-PAGE. When the Klett value reaches 210-260 or the OD₆₀₀ is ~2, pellet the remaining 75 mL of cells for 10 minutes at 12,500×g. Discard supernatant and wash cells with 25 mL Buffer 1. Pellet cells as above. Resuspend cells in 25 mL Buffer 1.

Cell Lysis

[0120] Add 300 units of DNase (Sigma D-5025) in 20 μL Buffer 1. Sonicate on ice in a small beaker to disrupt cell aggregates.

[0121] Lyse cells in a French press or a microfluidizer at approximately 18,000 psi. Collect into a beaker on ice.

[0122] Pellet cell debris for 15 minutes at 22,000×g.

[0123] The supernatant is transferred to an ultracentrifuge tube and membranes are pelleted for 45 minutes at 245,000×g. The supernatant is discarded.

[0124] Preparing whole cell samples for SDS-PAGE

[0125] Pellet the cells from the 5 mL of cells removed previously (for turbidity measurement) for 10 minutes at 12,500×g. Supernatant is decanted and discarded.

[0126] Wash the cells once by resuspending and vortexing in 1 mL of Buffer 1. Pellet the cells and discard the supernatant.

[0127] Resuspend cells in 150 μL of 0.1 M Tris, pH 8.5, then add 150 μL of Sample Quench. Vortex for 30 seconds.

[0128] To shear DNA, sonicate each tube with a microtip probe until foam appears (2-3 seconds).

[0129] Place tubes in a 90° C. bath for 10 minutes.

[0130] Add 618 μL of TE to each tube.

[0131] Vortex each tube for 30 seconds, then heat again at 90° C. for 5 minutes.

[0132] Short term storage of tubes is at 4° C. Longer term storage requires freezing at -80° C.

Preparing Membrane Samples for SDS-PAGE

[0133] To each ultracentrifuge tube, add 1 mL of 0.1 M Tris, pH 8.5.

[0134] Resuspend the membranes by vortexing, scraping if necessary. A paint brush works very well here. Add 1 mL Sample Quench and mix.

[0135] Transfer 1 mL of the resuspended membrane pellet to a microfuge tube and heat for 10 minutes at 90° C.

[0136] Vortex this sample for 30 seconds and transfer 100 μL to another microfuge tube.

[0137] Save the remaining 900 μL at -20° C. To the 100 μL aliquot, add 206 μL of TE.

[0138] Heat each tube for another 5 minutes at 90° C., then vortex for 30 seconds.

[0139] Short term storage of tubes is at 4° C. Longer term storage requires freezing at -80° C.

SDS-PAGE Followed by Electroblothing of Proteins to PVDF Membrane

[0140] Replica gels are run in parallel. One gel is stained with Coomassie Brilliant Blue. If heterologously-expressed target proteins are not well-visualized by this method, then proteins of the replica gel are electroblotted to PVDF membrane and the target protein is detected on a Western blot with an anti-polyhistidine antibody.

[0141] For SDS-PAGE, assemble gels on apparatus (e.g., Mini-Protean III system from Bio-Rad) with the running buffer required by the gel manufacturer. See gel product manual for the appropriate buffer recipes. Load samples and run gels according to gel manufacturer's specifications.

[0142] For gels that will be stained directly, follow these steps: Stain and destain according to instructions from the gel manufacturer and stain/destain manufacturer.

[0143] For gels that are to be electroblotted to PVDF membranes for Western blots, follow these steps: Prepare adequate quantities of the blotting buffer "TGMS". This is a 1×solution that is prepared by dilution of 10×blotting buffer. Approximately 1 L of TGMS is required per electrotransfer tank.

[0144] While the gel is running, prepare the PVDF membrane for transfer by first wetting in a minimal amount of methanol, then placing it in 50 mL TGMS for further wetting (with rocking).

[0145] When SDS-PAGE is complete, disassemble the plates and remove the gel. Soak it in 50 mL TGMS for only five minutes, with rocking. This short soak ensures that some SDS remains to prevent membrane proteins from precipitating in the gel.

[0146] Assemble the blotting sandwich according to directions provided by the manufacturer.

[0147] Be sure that all air bubbles are removed, especially between the gel and the PVDF membrane. Everything should be thoroughly wetted in TGMS at the time of assembly.

[0148] Transfer the blotting sandwich to the blotting tank filled with TGMS. Place a small stir bar in the bottom of the tank and use an ice reservoir to keep the initial transfer cold.

[0149] Transfer at 300 mA (1 hour), then overnight at 100 mA with slow stirring.

[0150] Disassembly and waste disposal: Separate the sandwich layer by layer, taking care to note the orientation of the PVDF membrane, and place it in container with protein side up. Either proceed immediately to development of the Western blot or allow the PVDF membrane to air dry for later processing. If the PVDF is allowed to dry, it must be wetted again in methanol prior to transfer to any aqueous solution for further processing.

[0151] Stain the electroblotted gel to determine transfer efficiency. Dispose of the tank blotting buffer in a hazardous waste container. Western Blot Development Using an Antipolyhistidine Antibody. (Protocol adapted from those found at www.novagen.com and www.piercenet.com).

[0152] Resuspension of His-Tag Monoclonal Antibody: The His-Tag Monoclonal Antibody (Novagen 70796-3) is provided as a lyophilized powder and must be resuspended prior to use in the following protocols. Dissolve the lyophilized antibody in 500 sterile water per 100 μ g vial or 15 μ L sterile water per 3 μ g vial (final concentration 0.2 mg/mL).

[0153] Chemiluminescent detection: Alkali-soluble Casein (Novagen 70955-3; stored at 4°C) is the recommended blocking reagent for chemiluminescent detection on nitrocellulose membranes because it results in the lowest background and can be used as a blocking reagent throughout the protocol. The following conditions work well for the hydrophobic PVDF blotting membranes recommended. Note that different membranes may require different blocking conditions (e.g. longer blocking incubations, higher concentration of blocking reagent).

[0154] Reagent Preparation: Prepare 30 mL of blocking solution (1% Alkali-soluble casein in 1 \times TBS) per blot by mixing 6 mL of 5% Alkali-soluble Casein with 24 mL of deionized water. Fresh blocking solution should be prepared each time. Reserve the blocking solution throughout the procedure because it will also be used for the primary and secondary antibody dilution.

[0155] Prepare 1 L each of 1 \times TBS and 1 \times TBSTT. They may be prepared by diluting 10 \times stocks. Filter sterilize the 1 \times TBSTT.

[0156] The resuspended His-Tag Monoclonal 1E Antibody will be used at a dilution of 1:1000 in blocking solution (7.5 μ L in 7.5 mL total).

[0157] The Goat Anti-Mouse IgG HRP conjugate 2E antibody (Novagen 71045-3) will be used at a dilution of 1:50,000 in blocking solution. Total volume is 20 mL for this step.

[0158] Development of the Western blot: The following steps should be performed at room temperature, with gentle agitation or rocking during incubations. For the standard 5.5 cm \times 8.5 cm pieces of PVDF that fit purchased mini-gels, use clear plastic 6.5 cm \times 9 cm trays for all incubations. Place the membrane in the tray with the protein-side up, as determined by marking it or using colored molecular weight standards. The solution volumes used in this protocol are based on a 5.5 cm \times 8.5 cm blot. Larger or smaller membranes will require adjustment of the volumes.

[0159] If starting with a dried PVDF membrane, first re-wet it by soaking in methanol. Transfer the membrane to 15 mL 1 \times TBS and perform two washes, each of 10 minutes.

[0160] Discard the washes and incubate the membrane in 15 mL blocking solution for at least 1 hour. Remove 7.5 mL of blocking solution from the tray and save it for later. To the remaining 7.5 mL of blocking solution in the tray, add 7.5 μ L of the His-Tag Monoclonal 1E Antibody (thus diluted 1:1000) and incubate for 1 hour with rocking.

[0161] Wash twice, for 10 minutes each time, with 20 mL 1 \times TBSTT to remove unbound 1E antibody.

[0162] Wash for 10 minutes with 15 mL 1 \times TBS. Incubate for 1 hour with 20 mL Goat anti-Mouse IgG HRP Conjugate 2E antibody diluted 1:50,000 in blocking solution (see Reagent Preparation, step 4, above).

[0163] Wash at least five times, for 10 minutes each wash, using 20 mL 1 \times TBSTT per wash. It is important to thoroughly wash the membrane at this point to achieve maximum signal:noise ratios.

[0164] After the final washing step is complete, drain as much TBSTT from the membrane as possible.

[0165] Addition of the substrate: For a typical 5.5 cm \times 8.5 cm membrane, use 3 mL each of Pierce Pico peroxide solution (#1856135) and Pierce Pico enhancer (#1856136), and 0.25 mL each of Pierce Dura peroxide solution (#1856158) and Pierce Dura enhancer (#1856157) for a total volume of 6.5 mL. Incubate the membrane in the substrate at room temperature for 5 minutes with rocking.

[0166] Remove the membrane from the substrate. Drain any excess substrate from the membrane by touching the edge to a paper towel. Place the membrane in a clear plastic development folder and fold the plastic over the membrane. Remove any bubbles between the plastic and the membrane. Gently remove any liquid from the exterior of the plastic.

[0167] Use the membrane to expose x-ray film for identification of expressed target proteins. Typical results from a screening experiment of this type using the pRKPLHT1Dpuf (Table 1) expression vector are shown in FIG. 16. Expression yield can be crudely estimated as outlined in FIG. 17 (again shown here for expression vector pRKPLHT1Dpuf).

[0168] Differential centrifugation may be used to determine the cellular localization of the expressed target membrane protein in *Rhodobacter* cells (e.g., FIG. 16). Expression in whole cells is compared (on an equal volume basis

using Western analysis with an anti-polyhistidine antibody) with the supernatant (soluble fraction) and pellet (membrane fraction) obtained from ultracentrifuge separation (245,000×g) of lysates that are devoid of cellular debris. Most of the target membrane proteins that have been studied are expressed predominantly in the *Rhodobacter* ICM. Very few target membrane proteins show any significant presence in the soluble fraction. The sum of the signals from the soluble and membrane fractions should equal the total expression level observed in the cells. If this is not the case, one should investigate the debris pellet obtained from the lysate to test for the presence of target protein that may have aggregated as inclusion bodies—a phenomenon not yet observed with the expression of membrane proteins in *Rhodobacter*.

Example 2

Summary of Results with Platform Vector pRKPLICHT1Dpuf.

[0169] Expression Analysis with recently designed and constructed vector pRKLICHT1Dpuf.

[0170] The ligation-independent-cloning vector pRKLICHT1Dpuf was initially tested with target genes that were characterized by good expression using pRKPLHT1Dpuf. These prokaryotic membrane protein genes were numbered APC00809, APC00821, and APC00951. Expression analysis in whole cells, crude membrane preparations, and the soluble fraction are shown in FIG. 17. No differences in expression levels are apparent when comparing results from the same gene expressed from pRKPLHT1Dpuf or pRKLICHT1Dpuf.

[0171] Membrane proteins from *E. coli* that have no known homolog in the PDB are selected for expression. If a *Rhodobacter* homolog of the *E. coli* target exists, then it is also selected. Information obtained from a single structure by focusing on large protein families is maximized. Targets exhibiting a wide range of MW, pIs, and hydrophathy plot signatures are selected initially.

[0172] A typical set of oligonucleotides used to amplify and subsequently clone a target membrane protein gene (APC00809) into pRKPLHT1Dpuf is shown in FIG. 22. This success spawned the use of pRKLICHT1Dpuf for semi-automated cloning of 288 membrane protein genes (from *E. coli* and *B. subtilis*).

[0173] A set of oligonucleotides (FIG. 23) was used to amplify and subsequently clone a target membrane protein gene (APC00809) into pRKLICHT1Dpuf. FIGS. 24-25 show strategy for cloning target membrane protein genes in to versatile vectors.

Example 3

Adaptability of Versatile Vectors to Various Photosynthetic Bacteria

[0174] Several vectors disclosed herein, for example in Tables 1 and 2, can be adapted for use in other bacterial species using methodology known to a skilled artisan. In some of the embodiments disclosed herein, the pufB and pufA genes (B and A subunits) of the light-harvesting I complex (LHI) of *Rhodobacter sphaeroides* were replaced with a gene of interest. The gene of interest is flanked at the C-terminal or N-terminal end by an affinity tag and may be

followed by a protease digestion site. Similarly, a skilled artisan can clone an appropriate, functionally similar operon from another bacterial species to replace the puf or puc operon backbone present in some of the vectors disclosed herein. Host genes that are not essential in membrane formation, membrane integrity or survival of host bacteria, may be replaced with a gene of interest under an appropriate promoter to obtain a suitable level of expression. The N-terminal or C-terminal affinity tags as part of the vector backbone can be used in the design of vectors capable of multiplication in a traditional host such as *E. coli* and are also capable of expressing a desired gene in a photosynthetic bacteria such as, for example, *Rhodospseudomonas*, *Rhodocyclus*, and *Chlorobium*.

[0175] Coordinately, it is desirable to delete chromosomal copies of the non-essential host gene whose plasmid-borne copy is being replaced by the foreign gene, thus engineering added capacity in the host membranes for accommodating over-expressed heterologous foreign membrane proteins.

[0176] Some commercially applicable target membrane proteins that can be expressed using the vectors disclosed herein include receptors including G-protein coupled receptors, ion channels, transporters, membrane-bound enzymes, cytoskeletal membrane proteins, and membrane proteins specific to prokaryotic pathogens.

Example 4

Production of a Soluble Protein in *Rhodobacter* that Proved Problematic when Expressed in *E. coli*.

[0177] When expression of ILR1, a soluble protein derived from *Arabidopsis thaliana*, was attempted in *E. coli* using an expression system based upon T7 polymerase, the expressed protein aggregated and precipitated in a non-functional state in the form of inclusion bodies. To test whether this problem could be circumvented by features of the *Rhodobacter* expression system, this gene was cloned into pRKPLHT1Dpuf for expression in *R. sphaeroides*. Results from small-scale screening suggest that the protein associates with the ICM and that the elaboration of additional membranes in *Rhodobacter* allows for successful expression of this protein in an unaggregated state (FIG. 20). *Rhodobacter* produced this protein at an approximate level of 2 mg protein per liter of cell culture (sufficient for subsequent larger-scale purification efforts).

TABLE 1

Versatile Vectors designed for use in <i>Rhodobacter</i> Expression System				
Vector Properties				
Vector Name	Cloning Strategy	Tag Position	Tag Length	Cleavable Components [†]
pRKMALICHT1Dpuf	LIC	C	7	none
pRKMALICHT10Dpuf	LIC	C	10	none
pRKMALICHT13Dpuf	LIC	C	13	none
pRKHTMALIC1Dpuf	LIC	N	7	none
pRK10HTMAPLIC1Dpuf	LIC	N	10	none
pRK13HTMALIC1Dpuf	LIC	N	13	none
pRKMATEVLICHT1Dpuf	LIC	C	7	membrane anchor
pRKMATEVLICHT10Dpuf	LIC	C	10	membrane anchor
pRKMATEVLICHT13Dpuf	LIC	C	13	membrane anchor
pRKMALICTEVHT1Dpuf	LIC	C	7	tag

TABLE 1-continued

Versatile Vectors designed for use in Rhodobacter Expression System Vector Properties				
Vector Name	Clon- ing Strat- egy	Tag Posi- tion	Tag Length	Cleavable Components [‡]
pRKMALICTEVHT10Dpuf	LIC	C	10	tag
pRKMALICTEVHT13Dpuf	LIC	C	13	tag
pRKHTTEVMALIC1Dpuf	LIC	N	7	tag
pRK10HTTEVMALIC1Dpuf	LIC	N	10	tag
pRK13HTTEVMALIC1Dpuf	LIC	N	13	tag
pRKHTMATEVLIC1Dpuf	LIC	N	7	tag and anchor
pRK10HTMATEVLIC1Dpuf	LIC	N	10	tag and anchor
pRK13HTMATEVLIC1Dpuf	LIC	N	13	tag and anchor
pRKSSLICHT1Dpuf	LIC	C	7	none
pRKSSLICHT10Dpuf	LIC	C	10	none
pRKSSLICHT13Dpuf	LIC	C	13	none
pRKSSLICTEVHT1Dpuf	LIC	C	7	tag
pRKSSLICTEVHT10Dpuf	LIC	C	10	tag
pRKSSLICTEVHT13Dpuf	LIC	C	13	tag
pRKSSHTLIC1Dpuf	LIC	N	7	none
pRKSS10HTLIC1Dpuf	LIC	N	10	none
pRKSS13HTLIC1Dpuf	LIC	N	13	none

TABLE 1-continued

Versatile Vectors designed for use in Rhodobacter Expression System Vector Properties				
Vector Name	Clon- ing Strat- egy	Tag Posi- tion	Tag Length	Cleavable Components [‡]
pRKSSHTTEVLIC1Dpuf	LIC	N	7	tag
pRKSS10HTTEVLIC1Dpuf	LIC	N	10	tag
pRKSS13HTTEVLIC1Dpuf	LIC	N	13	tag

[‡]The signal sequence, by definition, is cleaved and not denoted as such here.
Definitions:
LDC = ligation dependent cloning
LIC = ligation independent cloning
HT = polyhistidine tag
TEV = tobacco etch virus
MA = membrane anchor
SS = signal sequence

[0178]

TABLE 2

Description of cloning sites and affinity tags in the expression vectors		
Vector name	DNA sequence position	Site Description
pRKHTLIC1Dpuf	2563–2565	ATG start site (underlined)
(N-terminal 7 x His tag	2566–2586	7x Histidine tag (gray)
for ligation-independent	2601–2606	SnaBI restriction site (bold)
cloning)		
pRKHTPL1Dpuf	2563–2565	ATG start site (underlined)
(N-terminal 7 x His	2566–2586	7x Histidine tag (gray)
tag)	2587–2592	NheI restriction site (bold)
	2599–2604	BglIII restriction site (bold)
pRKHTTEVLIC1Dpuf	2563–2565	ATG start site (underlined)
(N-terminal 7 x His tag	2566–2586	7x Histidine tag (gray)
followed by cleavage site for	2587–2607	TEV protease recognition site
Tobacco Etch Virus protease	2610–2615	(double underlined)
for ligation-independent		SnaBI restriction site (bold)
cloning)		
pRKHTTEVPL1Dpuf	2563–2565	ATG start site (underlined)
(N-terminal 7 x His tag	2566–2586	7x Histidine tag (gray)
followed by cleavage site for	2587–2607	TEV protease recognition site
Tobacco Etch Virus protease)	2608–2613	(double underlined)
	2620–2625	NheI restriction site (bold)
		BglIII restriction site (bold)
pRKLICHT10Dpuf	2562–2567	PmlI restriction site (bold)
(C-terminal 10 x His	2580–2609	10x Histidine tag (gray)
tag for ligation-independent	2610–2615	Stop codons (italicized)
cloning)		
pRKLICHT13Dpuf	2562–2567	PmlI restriction site (bold)
(C-terminal 13 x His	2580–2618	13x Histidine tag (gray)
tag for ligation-independent	2619–2623	Stop codons (italicized)
cloning)		
pRKLICHT1Dpuf	2562–2567	PmlI restriction site (bold)
(C-terminal 7 x His tag	2583–2600	7x Histidine tag (gray)
for ligation-independent	2601–2605	Stop codons (italicized)
cloning)		
pRKPLHT10Dpuf	2564–2569	SpeI restriction site (bold)
(C-terminal 10 x His	2572–2577	BglIII restriction site (bold)
tag)	2578–2607	10x Histidine tag (gray)
	2608–2613	Stop codons (italicized)
pRKPLHT13Dpuf	2564–2569	SpeI restriction site (bold)

duces an in-frame TGA stop codon; this may be desirable if the preference is to express the target protein without the C-terminal polyhistidine tag.

[0197] Typical 5'- and 3'-oligonucleotides ("top" and "bottom", respectively) are shown in FIG. 19. Four to six "dummy" bases are included at the 5'-end of each oligonucleotide to enable efficient digestion of the amplicon by the restriction enzyme. This sequence is followed in the top primer by the restriction site sequence and a ribosome binding site (*Rhodobacter* RBS=GGAGG) placed 4-12 bases before the start codon; typically, the RBS is placed six bases before the start codon. The bottom primer incorporates the sequence for the second restriction enzyme site followed by the gene sequence. A polyhistidine tag and stop codons are encoded by the platform vectors, thus the native stop codon of the target gene should not be included in the amplicon. Oligonucleotides should be designed such that they have good GC-clamps at the 3' ends; at least three contiguous Gs or Cs are recommended.

[0198] Using any standard software, examine the oligonucleotide sequences to determine the melting temperature of the complementary region for use in determining annealing temperature for PCR reactions. T_m s of the complementary regions of the primer sets should match within 5° C.

Ligation-Independent Cloning Using pRKLICHT1Dpuf

[0199] Typical 5'- and 3'-oligonucleotides for use in ligation-independent cloning of a target gene are shown in FIG. 20. The 5'-end of the top primer begins with the sequence that provides a LIC overhang which is complementary to that of the platform vector (FIG. 21), followed by the RBS placed 4-12 bases before the start codon; typically, the RBS is placed six bases before the start codon. The 5'-end of the bottom primer begins with the other complementary LIC overhang, followed by the gene sequence. A polyhistidine tag and stop codons are encoded by the platform vectors, thus the native stop codon of the target gene should not be included in the amplicon. Oligonucleotides should be designed such that they have good GC-clamps at the 3' ends; at least three contiguous Gs or Cs are recommended.

[0200] Using any standard software, examine the oligonucleotide sequences to determine the melting temperature of the complementary region and to check for regions of stable secondary structure.

Platform Vector Preparation

[0201] In order to prepare the platform vectors for ligation-dependent (pRKPLHT1Dpuf) or ligation-independent (pRKLICHT1Dpuf) cloning, steps are provided herein by which the relatively large vectors are linearized and compatible, cohesive ends are generated. The protocols for ligation-dependent cloning that use restriction enzymes are outlined separately from the protocols for ligation-independent cloning that use the proof-reading exonuclease activity of T4 DNA polymerase. Examples include one for ligation-dependent cloning and the other for ligation-independent cloning, both using vectors with C-terminal, non-cleavable polyhistidine tags—that prepare platform vectors for insertion of foreign genes for expression in the *Rhodobacter* system. Similar steps are used in the preparation of other vectors (Table 1) described herein.

Large-Scale Vector Preparation Protocol for Ligation-Dependent Cloning using pRKPLHT1Dpuf

[0202] The platform vector, pRKPLHT1Dpuf, used for ligation-dependent cloning (FIG. 22) has a simple multiple cloning site with three unique restriction sites (SpeI, NdeI, and BglII) for target gene insertion. Routine cloning has been achieved using SpeI and BglII, and protocols below are designed and written based on the assumption that these restriction endonucleases will be utilized.

[0203] The conditions that work well for small reactions do not scale well to large volumes, thus multiple small reactions are preferred to one larger reaction to keep background levels of uncut plasmid low. Set up multiple tubes using this protocol to generate a large supply of digested vector. Typically, 3-4 reactions are good since this yields 150+ μ L of cloning vector, which is adequate for the cloning of approximately 150 target genes. Allow the reaction to incubate at 37° C. for at least 2 hours to ensure complete digestion.

[0204] Preparatory Digestion:

[0205] 25 μ L pRKPLHT1Dpuf

[0206] 2.5 μ L SpeI

[0207] 2.5 μ L BglII

[0208] 8 μ L Promega buffer B

[0209] 42 μ L sterile ddH₂O

[0210] For a total volume of 80 μ L

[0211] When using plasmid DNA prepared with basic alkaline lysis miniprep protocols, RNase should be included in the reaction. Most modern miniprep kits employ RNase during cell lysis and, hence, RNase can be excluded from the typical restriction endonuclease reaction, as presented above. The preparatory digest above assumes that the concentration of the plasmid DNA stock is between 0.3 and 2 μ g/ μ L.

[0212] When preparing to gel purify the DNA fragments, pour an 0.8% agarose gel and use a preparative comb.

[0213] Run the gel for at least 1.5 hours at 60 volts to help determine if the digestion was complete and to be able to separate linear from circular uncut plasmid DNA and then excise the band.

Purification and Evaluation of Digested Vector:

[0214] Extract the DNA from the excised agarose slice using a commercially available gel extraction kit (e.g., MoBio UltraClean GelSpin kit). Use a maximum of 0.2 g minced agarose per spin filter.

[0215] The QiaEx II does have a greater recovery rate by about 40-50%, however it is very time consuming (over an hour). Recovery from the MoBio kit is much quicker (~7 minutes) and the yield is lower. MoBio kit is satisfactory.

[0216] Before using the digested vector in an experimental reaction, run a control ligation (no insert) to determine background of colonies resulting from contamination of it by uncut or singly-cut vector. Store the digested vector at 4° C.

Large-Scale Vector Preparation Protocol for Ligation-Independent Cloning using pRKLICHT1Dpuf.

[0217] To generate the LIC overhangs, platform vector pRKLICHT1Dpuf is first linearized by digestion with PmlI and then treated with T4 DNA polymerase in the presence of dTTP. The exonuclease activity of the polymerase yields the overhangs that are shown in red in FIG. 23.

[0218] Since conditions that work well for small reactions often do not scale well to large volumes, the best results are achieved when multiple small reactions are performed and then combined following enzymatic digestion. The following steps indicate amounts of DNA used in typical preparations of vector carrying the LIC overhangs.

[0219] Vector linearization with PmlI: Digest 10 μ L pRKLICHT1Dpuf (1/5 of the yield of plasmid DNA from a standard miniprep protocol) with PmlI in 70 μ L reaction volume for one hour at 37° C. PmlI is an unstable enzyme and best results are achieved by adding a second aliquot half-way through the incubation. Clean up the reaction with any standard purification kit that is suitable for plasmids larger than 10 kb.

[0220] Generation of LIC overhangs: One half of the PmlI-digested DNA should be used in generating the sticky ends with LIC-qualified T4 DNA polymerase.

[0221] pRKLICHT 1 Dpuf/PmlI

[0222] 1 μ L 100 mM dTTP

[0223] 2 μ L 100 mM DTT

[0224] 4 μ L 10 \times T4 polymerase reaction buffer

[0225] 1 unit T4 DNA polymerase

[0226] Total volume of 40 μ L

[0227] Incubate at room temperature for 30 minutes, then inactivate the polymerase at 75° C. for 20 minutes. This inactivated mixture can be used directly in annealing reactions or it can be cleaned up using a standard purification kit that is suitable for plasmids larger than 10 kb.

[0228] Before using the digested vector in an experimental reaction, determine the background of colonies resulting from contamination of it by undigested vector. Store the digested vector at 4° C.

[0229] Vector Sequences

Broad-host-range expression vector with N-terminal 7 \pm His tag for Ligation

Independent Cloning

pRKHTLIC1Dpuf.seq Length: 14386

```

1  CCACCCAGGC  CGCCGCCCTC  ACTGCCCGGC  ACCTGGTCGC  TGAATGTGCA
51  TGCCAGCACC  TGCGGCACGT  CAATGCTTCC  GGGCGTCGGC  CTCGGGCTGA
101  TCGCCCATCC  CGTTACTGCC  CCGATCCCGG  CAATGGCAAG  GACTGCCAGC
151  GCTGCCATTT  TTGGGGTGAG  GCCGTTCCGG  GCCGAGGGGC  GCAGCCCCTG
201  GGGGGATGGG  AGGCCCGCGT  TAGCGGGCCG  GGAGGGTTCG  AGAAGGGGGG
251  GCACCCCCCT  TCGCGCTGCG  CGGTCACGGC  CACAGGGCCG  AGCCCTGGTT
301  AAAAACAAGG  TTTATAAATA  TTGGTTTAAA  AGCAGGTTAA  AAGACAGGTT
351  AGCGGTGGCC  GAAAAACGGG  CGGAAACCCF  TGCAAATGCT  GGATTTTCTG
401  CCTGTGGACA  GCCCTCAAAA  TGTCATAGG  TGCGCCCTC  ATCTGTACGC
451  ACTCTGCCCC  TCAAGTGTCA  AGGATCGCGC  CCCTCATCTG  TCAGTAGTCC
501  CGCCCTCAA  GTGTCAATAC  CGCAGGGCAC  TTATCCCCAG  GCTTGTCCAC
551  ATCATCTGTG  GAAACTCGC  GTAAAATCAG  GCGTTTTCGC  CGATTTGCCA
601  GGCTGGCCAG  CTCCACGTCG  CCGGCCGAAA  TCGAGCCTGC  CCCTCATCTG
651  TCAACGCCGC  GCCGGGTGAG  TCGGCCCTC  AAGTGTCAAC  GTCGCCCTC
701  CATCTGTGAG  TGAGGGCCAA  GTTTCCCGC  AGGTATCCAC  AACGCCGGCG
751  GCCCGGTGT  CTGCGACAGC  GCTTCGACGG  CGTTTCTGGC  GCGTTTGAGC
801  GGCCATAGAC  GGCCGCCAGC  CCAGCGGCGA  GGGCAACCAG  CCGGTGAGC
851  GTCGAAAGG  CGCTCTCCG  CTFCTCCGCT  CACTGACTCG  CTGGCTCGG
901  TCGTTCGGCT  GCGCGAGCG  GTATCAGCTC  ACTCAAAGGC  GGTAATACGG
951  TTATCCACAG  AATCAGGGGA  TAACGCAGGA  AAGAATATGT  GAGCAAAAGG
1001  CCAGCAAAAG  GCCAGGAACC  GTAAAAGGC  CGCGTTGCTG  GCGTTTTCCT
1051  ATAGGCTCCG  CCCCCGTGAC  GAGCATCACA  AAAATCGACG  CTCAGTCCAG
1101  AGGTGGCGAA  ACCCGACAGG  ACTATAAAGA  TACCAGGCGT  TTCCCCCTGG
1151  AAGTCCCTC  GTGCGCTCTC  CTGTTCCGAC  CCTGCCGCTT  ACCGGATACC
1201  TGTCCGCTT  TCTCCCTTCG  GGAAGCGTGG  CGCCATTCGC  CATTCAGGCT
1251  GCGCAACTGT  TGGGAAGGGC  GATCGGTGCG  GGCTTCTTCG  CTATTACGCC
1301  AGCTGGCGAA  AGGGGATGT  GCTGCAAGGC  GATTAAGTTG  GGTAAACGCCA
1351  GGGTTTTCCC  AGTCACGACG  TTGTAACAGC  ACGGGCAGTG  AATTCGGCCG
1401  CCGGCTGGCC  GAGGTGCTGG  GCAAGCCCTA  CCTCCAGGCC  CCCATCGGGG
1451  TCGAGAGCAC  GACCGCTTTC  CTGCGCCGCC  TGGCGAGAT  TCTGGGCGCT
1501  GATCCGGAGC  CCTTCATCGA  GCGCGAGAAG  CACTCGACGC  TGAAGCCCCT
1551  GTGGGATCTG  TGGCGGAGTG  TCACGCAGGA  CTTCTCCGGG  ACGGCCAATT
1601  TCGGAATCGT  GCGGACCGAA  ACTTATGCAA  GAGGCATCCG  AAACATATCTC
1651  GAAGGCGATC  TCGGGCTGCC  CTGCGCCTTC  GCCgTGCCCC  GCAAGAGGGG

```

-continued

1701 CTCGAAGACC GACAACGAAG CGGTGCGCGG ACTGATCCGC CAGCACCGTC
1751 CGCTCGTGCT CATGGGGTCC ATCAACGAGA AGATTACCT TCGGGAAC TG
1801 AAAGCCGGTC ACGGCCCGCA ACCCTCTTTC ATCGCTGCCT CTTTCCCGGG
1851 TCGCGCGGAT CCGCGCGCTA CCGGAACGCC CGTTATGGGA TATGCAGGTG
1901 CTACGTGGTT ACTGCAGGAA GTTTGCAACG CCCTGTTCGA CGCCCTGTTC
1951 CACATTCGTC CCCTCGGGAC GGAGATGGAC AGCGCCCGCC CCACACCGAC
2001 GACACTGCGC CCGGACTTCC CGTGGGATGC CGATGCACAA gcGGCCCTGG
2051 ACCGCATCGT AGAGGAGCAT CCGTCTTCA CCGGATCAG CGCCGCGGCT
2101 GCCTTGCGCGC ACGCCGCCGA GAAGGCTGCC CTCGATGCCG GTCGCCGAGG
2151 GGTCTGTGAGA GAGACTGTCC AAGCCCTGCC TGGGCCGGGC TTCGGCGAGA
2201 GGAAGGGAGA GAACCAATGA GCATCATGC CGTCAACACG CCGGTCCATG
2251 CCGCCAGGGC CCACGGGCAC CGAGCACCCAC GTGCCGAGTT CTACGTCTAC
2301 TTGCGCCGTA TTCTGTGGG CGCCTTCCCG GTGGCCTTCG TGAGTGGAT
2351 CGTCTCGACG ATCCGCCACC GCAGGCTTCC CAAGCGCGC CCCTTCGCGT
2401 CCGCCTGGTT CGATGCCAAG GCATCACGC CGCTGATTTT CCGCGCCTGA
2451 CCGCAGGTCA GGTTCGCACA CGCCATTCGT CGTCTCCCA AGGGCGGGC
2501 GATTAATCGG GAGGGCATGG TGCCTTACCG TAACCCACGC CACCAGCATG
2551 TGGAGGATCG CCATGCGGAGG GCGGCGGCTT GCGGCGGCT ATTCCAATCC
2601 TACGTAGAAG GGAAGATCTT GAGGGCGGGC CTCCGTCGC GGGCGGCACC
2651 CACGCCCGCA tCGATTCCAA GGTTACGCCA TTGAGACGGC TCCGCTTCGC
2701 GCGCAAGCGC GGGTTGGGCC GACTGCAAGC GGAGAGGAA GCATGGCACT
2751 GCTCAGCTTC GAGCGAAAAT ATCGCGTGCC GGGGGGCACG CTGGTCGGCG
2801 GAAACCTGTT CGACTTCTGG GTCCGCCCTT TCTATGTCGG CTCTCTCGGG
2851 GTTCGCGACG TTTTCTTCGC GGCCCTGGGT ATCATCTCTG ATGCTGGAG
2901 TGCCGTACTC CAGGGTACCT GGAACCCCA ACTCATCTCT GTCTACCCGC
2951 CGGCCCTTGA ATATGGCCTG GGAGGTGCAC CCCTCGCAA AGGGCGGCTG
3001 TGGCAGATCA TCACGATCTG CGCCACTGGT GCCTTCGTCA GCTGGGCGCT
3051 GCGCGAAGTC GAAATCTGCC GTAAGCTGGG CATCGGGTAC CACATCCCGT
3101 TCGCCTTCGC GTTCGCCATC CTGGCCATACC TGACGCTGGT GCTGTTCGGC
3151 CCGGTGATGA TGGCGCCTG GGGCTATGCC TTCCCTACG GGATCTGGAC
3201 GCACCTGCAC TGGGTGTCGA ACACGGGCTA CACCTACGGC AACTTCCACT
3251 ACAACCTGTC CCACATGATC GCCATCTCGT TCTTCTTAC GAACGCGCTG
3301 GCTCTGGCGC TGCACGGCGC CTTGTGTCTC TCCCGGCCA ACCCCGAGAA
3351 GCGCAAGGAA ATGCGGACCG CGGATCACGA GGATACGTTT TTCGCGGATC

3401 TGGTCGGGTA CTCGATCGGG ACGCTCGGCA TCCACGCCT CGGCCTGCTG
3451 CTCTCGCTGA GCGCCGTCTT CTTCAGCGCC CTCTGCATGA TCATFACCGG
3501 CACCATCTGG TTCGATCAGT GGTCGACTG GTGGCAATGG TGGGTGAAGC
3551 TGCCGTGGTG GCGGAACATC CCGGGAGGCA TCAATGGCTG AGTATCAGAA
3601 CATCTTCTCC CAGGTCCAGG TCCGCGGACC GGCCGACCTG GGGATGACCG
3651 AAGACGTCAA CCTGGCCAAC CGTTCCGGCG TCGGTCCCTT CTCGACCCTG
3701 CTCGGCTGGT TCGGCAACGC CAGCTCGGC CCGATCTATC TCGGCTCGCT
3751 CGGCCTCCTG TCCCTCTTCT CGGCCCTGAT GTGGTCTTTC ACCATCGGGA
3801 TCTGGTCTG GTATCAGGCG GGCTGGAACC CGGCCGTCTT CTTGCGCGAC
3851 CTGTCTTCTT TCTCGCTCGA GCCCGCCGCA CCGAATACG GTCTGTCTT
3901 CGCGGCTCCG CTGAAGGAAG GCGGGCTGTG GCTGATCGCG TCGTCTTCA
3951 TGTTCTGTCG GGTCTGGTCC TGGTGGGGCC GCACCTATCT CCGCGCTCAG
4001 GCGCTGGGCA TGGCAAGCA CACCCCTGG CGGTTCCTCT CGGCCATCTG
4051 GCTGTGGATG GTGCTGGCT TCATCCGTCG GATCCTCATG GGGTCTGGT
4101 CGGAAGCGGT TCCTACGGC ATCTTCTCGC ACCTCGACTG GACGAACAAC
4151 TTCTCGCTCG TCCACGGCAA CCTGTTCTAC AACCCCTTCC ACGGTCTCTC
4201 GATCGCCTTC CTCTACGGGT CGGCCCTGCT CTTCGCGATG CACGGTCCGA
4251 CCATCCCTGC GGTCTCCCGC TTCCGGCGGG AGCGCGAGCT GGAGCAGATC
4301 GCGCACCAGC GGACGGCAGC GGAGCGGGCC GCCCTCTTCT GCGCTGGAC
4351 CATGGGTTTC AACGCCACGA TGAAGGCAT CCACCCGCTG GCCATCTGGA
4401 TGGCGGTCTT CGTGACCCTC ACCGGCGGCA TCGGgATCCT GCTCTCGGGC
4451 ACGGTCTGGT ACAACTGGTA CGTCTGGGGC CAGAACCACG GCATGGCGCC
4501 GCTGAACCTGA GGAGCGATCA CAATGGCTGA CAAGACCATC TTCAACGATC
4551 ACCTCAACAC CAATCCGAAG ACCAACCTTC GCCTCTGGGT CGCTTCCAG
4601 ATGATGAAGG GTGCGGGCTG GGCTGGCGGC GTGTTCTTCG GGACGCTCCT
4651 TCTCATCGGG TTCTTCCGGG TGGTCCGGCG GATGCTTCCG ATCCAGGAGA
4701 ACCAGGCTCC GGGCCCGAAC ATCACCGGCG CTCTGGAGAC CCGGATCGAG
4751 CTGATCAAGC ATCTCGTCTG AGACAAGTCT CCGGGCAGGG CCGCGCAGG
4801 CCGCCCGCTC CTCCAAGTCC GGGCCATATC GCCGGCCCGG GTCCGGGGCG
4851 ACACCACAGC CCGGTTCCTT TCCTGTGGC GACAGGGACC TGGTCCCTG
4901 TGAAGACCG CACGGCACCC TTTTGACATT CACGGGAGGC TCTGATGACC
4951 AATCCACCCG CCGGACCCGA AACCCCGCTT TTGGATCGCG TCTGTGCCCC
5001 GCGCGACATG AAGGCGCTGA GTGACCCGA ACTGGAGCGG CTCGGCCAGC
5051 AAGTGCCTTC CGAGGTCAGT GATAGGGGTA GTTCTTATT TTAGGCAGTT

-continued

5101 TATATGAAAT TAAGACATGC AGATGTCACA GTGGATATTG AACTGGTCTC
5151 GAAAGCTCAA TATCCCCAA AGCACAAAGCA CAAACTTCGA CATCATGCAG
5201 AAGCGTTTCC CGAAaccggt ctctgacgtg ggcacgcgag agcagcatgc
5251 cgtgaccttc gggcgcgcc tcgCcgGGGc cggGatgaa cctttctgag
5301 cgtctctatc ctggttctct caacggggtt acgaccagat cgcccatgac
5351 gtggcgctgc agaaccttcc cgtccgcttc gtgatcgacc gggcggggct
5401 cgtggggggc gatggcgcgga cccatgcggg ggcTTCGAC GTTGGCTTCA
5451 TCACTTCGCT GCCCAACATG ACCGTGATGG CCGCGGCCGA CGAGGCCGAG
5501 CTCATCCACA TGATCgcCAC CGCCGTGGCC TTCGACGAGG GCCCATCCGC
5551 CTTcCGCTTC CCGCGGGGCG AGGGGGTGGG CGTCGAGATG CCCGAGCCCG
5601 GGACGGTGCT GGAGCCCGcC CGGGCCCGCG TGGTGCGCGA AGGGACggat
5651 gtcgcgatcc tctccttCGG CGCGcATCTG CAcGAGgccT TgcAGGcggc
5701 GAAACTTCTC GAGGccGAGG GGGTGAGCGT GACCGTGGCC GACgcccgCT
5751 TctCgCgCCC GCTCgAcACG GGGCTCATCG ACCAGCTCGT gcGCCATCAC
5801 GCGGCGCTGG TAACGGTGGG GCATGGGGCC ATGGGCGGCT TCGGCGCCCA
5851 TGTCATGCAC TATCTCGCCA ATTCGGCGCG CTTCGACGGG GGCCTCGGCG
5901 TCCGGTTCAT GACGCTGCCC GACCGCTTCA TCGAGCAGG GAGCCCGGAG
5951 gACATGTATG CCGATGCGGG GCTGCGGGCC GAGGATATCA AGCTTGGCGT
6001 AATCATGGTC ATAGCTGTTT CCTGTGTGAA ATGTATTATC GCTACAATT
6051 CCACACAACA TACGAGCCGG AAGCATAAAG TGTAAAGCCT GGGGTGCCTA
6101 ATGAGTGAGC TAATCACCAT TAATTGCGTT GCGCTCACTG CCCGCTTTC
6151 AGTCGGGAAA CCTGTCTGTC CAGCTGCATT AATGAATCGG CCAACGGCG
6201 GGGAGAGCGG GTTTCGCTAT TGGGCGCTCG GTCTTGCCCT GCTCGTCGGT
6251 GATGTACTTC ACCAGCTCCG CGAAGTCGCT CTCTTGTATG GAGCGCATGG
6301 GGACGTGCTT GGCATACAGC CGCACCCCCC GGCCGTTTTA GCGGTAAAA
6351 AAGTCATGGC TCTGCCCTCG GCGGACCAC GCCATCATG ACCTTGCCAA
6401 GCTCGTCTCG CTCTCTCTCG ATCTTCGCCA GCAGGGCGAG GATCGTGGCA
6451 TCACCGAACC GCGCCGTGCG CGGGTCGTCG GTGAGCCAGA GTTTCAGCAG
6501 GCCGCCCAGG CGGCCAGGT CGCCATTGAT GCGGGCCAGC TCGCGGACGT
6551 GCTCATAGTC CACGACGCCG GTGATTTTGT AGCCCTGGCC GACGGCCAGC
6601 AGGTAGGCCG ACAGGCTCAT GCCGCGCCCG CCGCCTTTTT CCTCAATCGC
6651 TCTTCGTTTC TCTGGAAGGC AGTACACCTT GATAGTGGG CTGCCCTTCC
6701 TGGTTGGCTT GGTTCATCA GCCATCCGCT TGCCCTCATC TGTTACGCCG
6751 GCGGTAGCCG GCCAGCCTCG CAGAGCAGGA TTCCCGTTGA GCACCCCGAG

6801 GTGCGAATAA GGGACAGTGA AGAAGGAACA CCCGCTCGCG GGTGGGCCTA
6851 CTTACCTAT CCTGCCCGCG TGACGCCGTT GGATACACCA AGGAAAGTCT
6901 ACACGAACCC TTTGGCAAAA TCCTGTATAT CGTGCGAAAA AGGATGGATA
6951 TACCGAAAAA ATCGCTATAA TGACCCCGAA GCAGGTTTAT GCAGCGGAAA
7001 AGCGCCACGC TTCCCGAAGG GAGAAAGGCG GACAGGTATC CGGTAAGCGG
7051 CAGGGTCGGA ACAGGAGAGC GCACGAGGGA GCTTCCAGGG GGAACGCCT
7101 GGTATCTTTA TAGTCTGTGC GGGTTTCGCC ACCTCTGACT TGAGCGTCGA
7151 TTTTGTGTAT GCTCGTCAGG GGGCGGAGC CTATGAAAA ACGCCAGCAA
7201 CCGGCCCTTT TTACGGTTCG TGGCCTTTTG CTGGCCTTTT GCTCACATGT
7251 TCTTTCCTGC GTTATCCCTT GATTCGTGTG ATAACCGTAT TACCCCTTT
7301 GAGTGAGCTG ATACCGCTCG CCGCAGCCGA ACGACCGAGC GCAGCGAGTC
7351 AGTGAGCGAG GAAGCGGAG AGCGCCAGAA GGCCCGCAGA GAGGCCGAGC
7401 CCGCCGTGA GGCTTGACG CTAGGGCAGG GCATGAAAA GCCCGTAGCG
7451 GGCTGCTACG GCGCTCTGAC GCGGTGAAA GGGGAGGGG ATGTTGTCTA
7501 CATGGCTCTG CTGTAGTGAG TGGTTGCGC TCCGGCAGCG GTCTGTATCA
7551 ATCGTACCC TTTCTCGGTC CTTCAACGTT CCTGACAACG AGCCCTCTTT
7601 TCGCAATCC ATCGACAATC ACCGCGAGTC CCTGCTCGAA CGCTCGCTCC
7651 GGACCCGCTT CGTGAAAGGC GTCTATCGCG GCCCGCAACA CGGGCGAGAG
7701 CGGACCTGTT TCAACGGTGC CGCCGCGCTC GCCCGCATCG CTGTCCCGGG
7751 CCTGCTCCTC AAGCACGGCC CCAACAGTGA AGTAGCTGAT TGTATCAGC
7801 GCATTGACGG CGTCCCGGCG CGAAAAACCC GCCTCGCAGA GGAAGCGAAG
7851 CTGCGCGTCG GCCGTTTCCA TCTGCGGTGC GCCCGGTGCG GTGCCGGCAT
7901 GGATGCGCGC GCCATCGCGG TAGGCGAGCA GCGCCTGCCT GAAGCTGCGG
7951 GCATTCCCGA TCAGAAATGA GCGCCAGTCG TCGTCGGCTC TCGGCACCGA
8001 ATGCGTATGA TTCTCCGCCA GCATGGCTTC GGCCAGTGCG TCGAGCAGCG
8051 CCCGCTTGTT CCTGAAGTGC CAGTAAAGCG CCGGCTGCTG AACCCCAAC
8101 CGTTCCGCCA GTTTCGCTGT CGTCAGACCG TCTACGCCGA CCTCGTCAA
8151 CAGGTCCAGG GCGGCACGGA TCACTGTATT CGGCTGCAAC TTTGTCTATG
8201 TTGACACTTT ATCACTGATA AACATAATAT GTCCACCAAC TTATCAGTGA
8251 TAAAGAAATCC GCGGTTCAA TCGGACCAGC GGAGGCTGGT CCGGAGGCCA
8301 GACaTGAAC CCAACATACC CCTGATCGTA ATTCTGAGCA CTGTCCGGCT
8351 CGACCGTGTG GGCATCGGCC TGATTATGCC GGTGCTGCCG GGCCCTCTGC
8401 CCGATCTGGT TCACTCGAAC GACGTCACCG CCCACTATGG CATTCTGCTG
8451 GCGCTGTATG CGTTGGTGCA ATTTGCTCTG GCACCTGTGC TGGGCGCGCT

-continued

8501 GTCGGATCGT TTCGGGCGGC GGCCAATCCTT GTCGCTCTCG CTGGCCGGCG
8551 CCACTGTGCGA CTAGCCCATC ATGGCGACAG CGCCTTCTCT FTGGGTCTC
8601 TATATCGGGC GGATCGTGGC CGGCATCACC GGGCGACTG GGGCGGTAGC
8651 CGGCGCTTAT ATTGCCGATA TCACTGATGG CGATGAGCGC GCGCGGCACT
8701 TCGGCTTCAT GAGCGCCTGT TTCGGGTTCC GGATGGTCCG GGGACCTGTG
8751 CTCGGTGGCG TGATGGGCGG TTCTCCTCCC CACGCTCCGT TCTTCGCGCG
8801 GGCAGCCTTG AACGGCCTCA ATTTCTGAC GGGCTGTTTC CTTTTGCGCG
8851 AGTCGCACAA AGGCGAAGCG CGGCCGTTAC GCGGGGAGGC TCTCAACCCG
8901 CTCGCTTCGT TCCGGTGGGC CCGGGGCATG ACCGTCTGCG CCGCCCTGAT
8951 GCGGCTCTTC TTCATCATGC AACTTGTCCG ACAGGTGCCG GCCCGCTTT
9001 GGGTCATTTT CCGCGAGGAT CGCTTCTACT GGGACGCGAC CACGATCGGC
9051 ATTTCCGCTTG CCGCATTGG CATTCTGCAT TCACTCGCCC AGGCAATGAT
9101 CACCGGCCCT GTAGCCGCCG GGCTCGGCGA AAGGGGGGCA CTCATGCTCG
9151 GAATGATTGC CGACGGCACA GGCTACATCC TGCTTGCTT GCACACGG
9201 GGATGGATGG CGTTCCCGAT CATGGTCCCTG CTGCTTCGG GTGGCATCGG
9251 AATGCCGCGC CTGCAAGCAA TGTGTCCAG GCAGGTGGAT GAGGAACGTC
9301 AGGGGAGCT GCAAGGCTCA CTGGCGGCGC TCACCAGCCT GACCTCGATC
9351 GTCGGAGCCC TCCTCTTCAC GCGGATCTAT GCGGCTTCTA TAACAACGTC
9401 GAACGGGTGG GCATGGATTG CAGGCGCTGC CCTCTACTTG CTCTGCTGC
9451 CGGCGCTGCG TCGCGGCTT TGAGCGGCGC CAGGGCAAG AGCCGATCGC
9501 TGATCGTGG AAGCATAGGC CTATGCCATG CCGGTCAAG CGACTTCCGG
9551 CAAGCTATAC GCGCCCTAGG AGTGGGTTG GAACGTTGGC CCAGCCAGAT
9601 ACTCCCAGTC ACAGGAGGCA CGCCGATGAT TTGAAGCGCA CTCAGCGTCT
9651 GATCCAGAAA CAACCACTCT AGCAACACCG CGGTCCCCGG GCTGAGAAAG
9701 CCCAGTAAG AAACAACCTGT AGTTCGAGT CCGGAGATCC CCGGAACCA
9751 AAGGAAGTAG GTTAAACCCG CTCGATCAG GCCGAGCCAC GCCAGGCCGA
9801 GAACATTGGT TCCTGTAGGC ATCGGGATTG CCGGATCAAA CACTAAAGCT
9851 ACTGGAACGA GCAGAAGTCC TCCGGCCGCC AGTTGCCAGG CGGTAAGGT
9901 GAGCAGAGGC ACGGGAGGTT GCCACTTGGC GGTGAGCAG GTTCCGAAAG
9951 CCAATGAAAC CGCCCCGCC AGGCCCGCTG CGACGCGGAC AGGATCTAGC
10001 GCTGCGTTTG GTGTCAACAC CAACAGCGCC ACGCCCGCAG TTCCGCAAT
10051 AGCCCCCAG ACCGCCATCA ATCGTATCGG GCTACCTAGC AGAGCGGCAG
10101 AGATGAACAC GACCATCAGC GGCTGCACAG CCCTACCCTG CGCCGCGACC
10151 CCGCCCGGCA GCGGCTAGAC CGAAATAAAC AACAAGCTCC AGAATAGCGA

10201 AATATTAAGT GCGCCGAGGA TGAAGATGCG CATCCACCAG ATTCCCGTGT
10251 GAATCTGTG GAGGATCATC ACAGCAATA AACCCGCCGG CAACGCCCGC
10301 AGCAGCATAC CCGCGACCCC TCGGCTCTCG TGTTCGGGCT CCACGAAAAC
10351 GCGGACAGA TGCGCTTGT GAGCGTCTTT GGGGCCGTCC TCCTGTTTGA
10401 AGACCGACAG CCCAATGATC TCGCCGTCGA TGTAGCGGCC GAATGCCAGC
10451 GCATCTCGCA ACCGTTGAGC GAACGCTCTC ATGGGCTTTT TCTCCTCGTG
10501 CTCGTAACG GACCCGAAAC TCCTGGAGC TTCTTTCAGG GCCGACAATC
10551 GGATCTCGCG GAAATCCTGC ACCTCGGCGC CTCCAAGCCG TCGAATCTGA
10601 GCCTTAATCA CAATTGTCAA TTTTAATCCT CTGTTTATCG GCAGTTCGTA
10651 GAGCGCGCCG TGCGTCCCGA GCGATACTGA GCGAAGCAAG TGCGTCGAGC
10701 AGTGCCCGCT TGTTCCTGAA ATGCCAGTAA AGCGCTGGCT GCTGAACCCC
10751 CAGCCGGAAC TGACCCACA AGGCCCTAGC GTTTGCAATG CACCAGGTCA
10801 TCATTGACCC AGGCGTGTTC CACCAGGCGC CTGCCTCGCA ACTCTTCGCA
10851 GCGTTCGCGC ACCTGTCTGC GCCACTTCTT CACGCGGGTG GAATCCGATC
10901 CGCACATGAG GCGGAAGGTT TCAGCTTGA GCGGTACGG CTCCCGGTGC
10951 GAGCTGAAAT AGTCAACAT CCGTCGGGCC GTCGGCGACA GCTTGCGGTA
11001 CTTCTCCCAT ATGAATTTCC TGTAGTGGTC GCCAGCAAAC AGCAGCAGCA
11051 TTTCTCTGTC GATCAGGACC TGGCAACGGG ACGTTTCTT GCCACGGTCC
11101 AGGACCGGGA AGCGGTGCGC CAGCGACACC GATTCAGGT GCCAACCGG
11151 GTCGGACGTG AAGCCCATCG CCGTCGCCTG TAGGCGCGAC AGGCATTCTG
11201 CCGCCTTCGT GTAATACCGG CCATTGATCG ACCAGCCAG GTCTGGCAA
11251 AGCTCGTAGA ACGTGAAGGT GATCGGCTCG CCGATAGGGG TCGCTTCGC
11301 GTACTCCAAC ACCTGTCTGC ACACCAGTTC GTCATCGTGC GCCCGCAGCT
11351 CGACCGCGGT GTAGGTGATC TTCACGCTCT TGTGACGTG GAAAATGACC
11401 TTGTTTTCGA GCGCCTCGCG CCGGATTTTC TTGTTGCGCG TGGTGAACAG
11451 GGCAGAGCGG GCCGTGTCTG TTGGCATCGC TCGCATCGTG TCCGGCCAGC
11501 GCGCAATATC GAACAAGGAA AGCTGCATTT CCTTGATCTG CTGCTTCGTG
11551 TGTTTCAGCA ACGCGGCTCG CTTGGCTCTG CTGACCTGTT TTGCCAGGTC
11601 CTCGCCGCGG GTTTTCGCTC TCTTGGTCTG CATAGTTCCT CGCGTTCGCA
11651 TGGTCATCGA CTTGCCCCAA CCTGCGCCTC CCTGTTTCGAG ACGACGCGAA
11701 CGTCCACGCG CGGCCGATGG CCGGGGCGAG GCAGGGGAG CCAGTTGCAC
11751 GCTGTCCGCT TCGATCTTGG CCGTAGCTTG CTGGACCATC GAGCCGACGG
11801 ACTGGAAGGT TTCGCGGGGC GCACGATGA CCGTGCGGCT TCGCATGGTT
11851 TCGGCATCCT CCGCGGAAAA CCCCAGCTCG ATCAGTTCTT GCCTGTATGC

-continued

```

11901 CTTCCGGTCA AACGTCCGAT TCATTACCC TCCTTGCGGG ATFGCCCCGA
11951 CTCACGCGG GGAATGTGC CCTTATTCCT GATTTGACCC GCCTGGTGCC
12001 TTGGTGTCCA GATAATCCAC CTTATCGGCA ATGAAGTCGG TCCCGTAGAC
12051 CGTCTGGCCG TCCTTCTCGT ACTTGGTATT CCGAATCTTG CCCTGCACGA
12101 ATACCAGCGA CCCCTTGCCC AAATACTTGC CGTGGGCTC GGCCTGAGAG
12151 CCAAACACT TGATGCGGAA GAAGTCGGTG CGCTCCTGCT TGTCCCGGT
12201 CGTGGCCGCG CCAACCTTTG CGATCCGCAA GCGCGCGGTC GCCATCTTCA
12251 CGCTGGAACA GTACGTCGAG GCGGCATCA TGACCCGCGA GCAATACGAG
12301 GTCATTAATA GCGCCGTGAT TGATGATATA GCGGCCCGC TGCTCCTGGT
12351 TCTCGCGCAC CGAAATGGGT GACTTCACCC CGCGCTCTTT GATCGTGGCA
12401 CCGATTTCCG CGATGCTCTC CGGGGAAAAG CCGGGGTTGT CGGCCGTCCG
12451 CCGGTGATGC GGATCTTCGT CGATCAGGTC CAGGTCCAGC TCGATAGGGC
12501 CGGAACCGCC CTGAGACGCC GCAGGAGCGT CCAGGAGGCT CGACAGGTCG
12551 CCGATGCTAT CCAACCCAG GCGGACGGC TGCGCCGCGC CTGCGGCTTC
12601 CTGAGCGGCC GCAGCGGTGT TTTTCTTGGT GGTCTTGGCT TGAGCCGCG
12651 TCATGGGGAA ATCTCCATCT TCGTGAACAC GTAATCAGCC AGGGCGCGAA
12701 CCTCTTTTCA TGCCTTGCGC GCGGCCGTTT TCTTGATCTT CCAGACCGGC
12751 ACACCGGATG CAGAGGGATC GCGGATGCTG CTGCGCAGGC CAACGGTGGC
12801 CCGAATCATC ATCTTGGGGT ACGCGGCCAG CAGCTCGGCT TGGTGGCGCG
12851 CGTGGCCGCG ATTCGCGCCA TCGACCTTGC TGGGCCACAT GCCAAGGAAT
12901 TGCAGCTTGG CGTTCTTCTG GCGCACGTTT GCAATGGTCC TGACCATCTT
12951 CTTGATGCCC TGGATGCTGT ACGCCTCAAG CTCGATGGGG GACAGCACAT
13001 AGTCGGCCGC GAAGAGGGCG GCGCCAGGC CGACGCCAAG GGTCCGGGCC
13051 GTGTCGATCA GGCACACGTC GAAGCCTTGG TTCCGCCAGG CTTGATGTT
13101 CGCCCCGAAC AGCTCGCGGG GTCTGTCAG CGACAGCCGT TCGCGTTCG
13151 CCAGTACCGG GTTGGACTCG ATGAGGCGA GCGCGCGGC CTGGCCGTCG
13201 CCGGCTGCGG GTGCGGTTTC GGTCCAGCCG CCGCAGGGA CAGCGCCGAA
13251 CAGCTTGCTT GCATGCAGGC CGGTAGCAAA GTCCTTGAGC GTGTAGGACG
13301 CATTGCCCTG GGGTCCAGG TCGATCACGG CAACCCGCAA CCGCGCTCG
13351 AAAAAGTCGA AGGCAAGATG CACAAGGGTC GAAGTCTTGC CGACCGCCG
13401 TTTCTGTTG GCGGTGACCA AAGTTTCAT CGTTTGGTTT CCTGTTTTTT
13451 CTTGGCGTCC GCTTCCACT TCCGACGAT GTACGCCTGA TGTCCGGCA
13501 GAACCGCGT TACCCGCGG TACCCCTCGG GCAAGTCTT GTCTCGAAC
13551 GCGGCCACA CGCATGCAC CGCTTGCAC ACTGCGCCCC TGGTCAGTCC

13601 CAGCGAGTT GCGAACGTCG CCTGTGGCTT CCCATCGACT AAGACGCCCC
13651 GCGTATCTC GATGGTCTGC TGCCCACCTT CCAGCCCTG GATCGCCTCC
13701 TGGAACTGGC TTTCGTAAG CCGTTTCTT ATGATAACA CCCATAAFTT
13751 GCTCCGCGCC TTGGTTGAAC ATAGCGGTGA CAGCCGCCAG CACATGAGAG
13801 AAGTTAGTCT AAACATTTC TCGCACGTCAA CACCTTTAGC TTCGTCTCTT
13851 CGTCCTTGGC GTAACAAAAC AAAAGCCCCG AAADDGGGCT TTCGTCTCTT
13901 GCGCTTATG GCTCTGACCC CGGCTCCATC ACCAACAGGT CGCGCACGCG
13951 CTTCACTCGG TTGCGGATCG AACTGCGAG CCAACAAAAG CCGGTTGCCG
14001 CCGCCGCCAG GATCGCGCC ATGATGCGCG CCACACGGC CATCGCCAC
14051 CAGTCCGCG CCTTCCGGTT CCATTCTGCT TGGTACTGCT TCGCAATGCT
14101 GGAACCTCGC TCACCATAGG CTGACCGCTC GATGGCGTAT GCCGCTTCTC
14151 CCCTTGGCGT AAAACCCAGC GCGCAGGGC GCATTGCCAT GCTGCCCGCC
14201 GCTTTCCCGA CCACGACGCG CGCACCAGGC TTGCGGTCCA GACCTTCGGC
14251 CACGGCAGC TGGCAAGGA CATAATCAGC CGCCGACTTG GCTCCACGCG
14301 CCTCGATCAG CTCCTGCACT CGCGGAAAT CTTGGCCTC CACGGCCGCC
14351 ATGAATCGCG CACGCGCGA AGGCTCCGCA GGGCCG

```

-continued

Broad-host-range expression vector with N-terminal 7 ± His tag

pRKHTPL1Dpuf.seq Length: 14370

```

1   CCACCCAGGC CGCCGCCCTC ACTGCCCGGC ACCTGGTCGC TGAATGTGCA
51  TGCCAGCACC TGCGGCACGT CAATGCTTCC GGGCGTCGCG CTCGGGCTGA
101 TCGCCATFCC CGTTACTGCC CCGATCCCGG CAATGGCAAG GACTGCCAGC
151 GCTGCCATTT TTGGGGTGAG GCCGTTCCGG GCCGAGGGGC GCAGCCCCTG
201 GGGGGATGGG AGGCCCGCGT TAGCGGGCCG GGAGGGTTCC AGAAGGGGGG
251 GCACCCCCCTC TCGCGCTGCC CGGTCACGCG CACAGGGCGC AGCCCTGGTT
301 AAAAACAAAGG TTTATAAATA TTGGTTTAAA AGCAGGTAA AAGACAGGTT
351 AGCGGTGGCC GAAAAACGGG CGGAAACCC TGC AATGTG GATTTTCTG
401 CCTGTGGACA GCCCCTCAA TGTCAATAGG TCGCCCTC ATCTGTACGC
451 ACTCTGCCCC TCAAGTGTCA AGGATCGCGC CCCTCATCTG TCAGTAGTCC
501 CGCCCTCAA GTGTCAATAC CGCAGGGCAC TTATCCCCAG GCTTGTCCAC
551 ATCATCTGTG GGAACCTCGC GTAAAAATCAG GCGTTTTCGC CGATTTGCGA
601 GGCTGGCCAG CTCCACCTCG CCGGCCGAAA AAGTGTCAAC FTCCGCCCTC
651 TCAACGGCGC GCCGGGTGAG TCGGCCCTC AAGTGTCAAC FTCCGCCCTC
701 CATCTGTACG TGAGGGCCAA GTTTTCCCGG AGGTATCCAC AACGCCGGCG
751 GCCCGGTGT CTGCACACG GCTTCGACGG CTTTTCTGGC GCGTTTGCAG
801 GGGCATAGAG GGGCGCCAGC CCAGCGGCGA GGGCAACCAG CCCGGTAGC
851 GTCGGAAAGG CGCTCTTCCG CTTCCTCGCT CACTGACTCG CTGCGCTCGG
901 TCGTTCGGCT GCGCGAGCG GTATCAGCTC ACTCAAAGGC GGTAATACGG
951 TTATCCACAG AATCAGGGGA TAACGCAGGA AAGAATATG GAGCAAAAGG
1001 CCAGCAAAGG GCCAGGAACC GTAAAAAGGC CCGTTCGCTG GCGTTTTTCC
1051 ATAGGCTCCG CCCCCGTGAC GAGCATCACA AAAATCGACG CTC AAGTCCAG
1101 AGGTGGCGAA ACCCGACAGG ACTATAAAGA TACCAGGCGT TTCCCCCTGG
1151 AAGTCCCTC GTGGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC
1201 TGTCGCCCTT TCTCCTTCG GGAAGCGTGG CGCCATTTCG CATT CAGGCT
1251 GCGCAACTGT TGGGAAGGGC GATCGGTGCG GCCTCTTCG CTATTACGCC
1301 AGCTGGCGAA AGGGGATGT GCTGCAAGGC GATTAAGTTG GGTAACGCCA
1351 GGGTTTTTCC AGTACACGAG TTGTAACAGC ACGGCCAGTG AATTCCGGCCG
1401 CGGGCTGGCC GAGGTGCTGG GCAAGCCCTA CCTCCAGGCC CCCATCGGGG
1451 TCGAGAGCAC GACCGCCTTC CTGCGCCGCC TGGGCGAGAT TCTGGGCCTC
1501 GATCCGGAGC CCTTCATCGA GCGCGAGAAG CACTCGACGC TGAAGCCCGT
1551 GTGGATCTG TGGCGGAGTG TCACGCAGGA CTTCCTCGGG ACGGCCAATT
1601 TCGGAATCGT GGGGACCAGG ACTTATGCAA GAGGCATCCG AAATATCTC
1651 GAAGGCATC TCGGCTGCC CTGCGCCTTC GCCGTGGCCC GCAAGAGGGG

1701 CTCGAAGACC GACAACGAAG CGGTGCGCG ACTGATCCGC CAGCACCGTC
1751 CGCTCGTGCT CATGGGGTCC ATCAACGAGA AGATTTACCT TGCGGAACATG
1801 AAAGCCGGTC ACGGCCGCA ACCCTCTTTC ATCGCTGCCT CTTTCCCGGG
1851 TGCGGCGATC CCGCGCGCTA CCGGAACGCC CTTATGGGA TATGCAGGTG
1901 CTACGTGGTT ACTGCAGGAA GTTTGCAACG CCCTGTTCGA CGCCCTGTTC
1951 CACATTTCTG CCTTCGGGAC GGAGATGGAC AGCGCCGCGC CCACACCAGC
2001 GACACTCGGC CCGGACTTCC CGTGGGATGC CGATGCGCAA GCGGCCCTGG
2051 ACCGCATCGT AGAGGAGCAT CCGGTTCTCA CCGGATCAG GCGCGCGGCT
2101 GCCTTGGCGG ACGCCGCCGA GAAGGCTGCC CTGATGCCG GTGCCGAGAG
2151 GGTGCTGAGA GAGACTGTCC AAGCCCTGCC TGGCCGGGC TTCGGCGAGA
2201 GGAAGGGAGA GAACCAATGA GCAATCATGC CGTCAACACG CCGGTCCATG
2251 CCGCCAGGGC CCACGGGCAC CGAGCACACC GTGCCGAGT CTACGCTTAC
2301 TTGCGCGTCA TTCTGCTGGG CGCCTTCCCG GTGGCCTTCG TGAGCTGGAT
2351 CGTCTCGACG ATCCGCCACC GCAGGCTTCC CAAGCGGGC CCCTTCGGGT
2401 CCGCCTGGTT CGATGCCAAG GCATCAGCG CGCTGATTTT CCGCGCTGA
2451 CCGCAGGTCA GGTGCGACA CGCCATTCTG CTCTCCCA AGGGCGGGC
2501 GATTAATCGG GAGGGCATGG TGCCTTACCG TAACCCACGC CACCAGCATG
2551 TGGAGGATCG CCAATGCTGCG GCGGCTGCGA GCGGCTGCGA GCGCGGGTTG
2601 ATCTTGAGCG CCGCCCTCCG TCGCGGGCG CACCCACGCC CGcATCGATT
2651 CCAAGGTTCA GCCATTGAGA CGGCTCGCT TCGCGGCAA GCGCGGGTTG
2701 GGCCGACTGC AAGCGGAGG GGAAGCATGG CACTGCTCAG CTTCGAGCGA
2751 AAATATCGCG TGCCGGGGGG CACGCTGGTC GCGGAAACC TGTTGCACTT
2801 CTGGGTGCGC CCTTCTATG TCGGCTTCTT CCGGGTTGCG ACGTTTTTCT
2851 TCGCGGCCCT GGTATCATT CTGATTGCC TGGAGTCCCT ACTCCAGGGT
2901 ACCTGGAACC CCAACTCAT CTCTGTCTAC CCGCGGCC TGAATATGG
2951 CCTGGGAGGT GCACCCCTCG CAAAAGCGG GCTGTGGCAG ATCATCAGGA
3001 TCTGCGCCAC TGGTGCCTTC GTCAGCTGGC CGCTGCGCGA AGTCGAAATC
3051 TGCCGTAAGC TGGGCATCGG GTACCATC CCGTTCCGCT TCGCGTTCCG
3101 CATCCTGGCC TACCTGACGC TGGTGTCTGT CCGCCCGGTG ATGATGGGCG
3151 CCTGGGGCTA TGCCCTTCCC TACGGGATCT GGACGCACCT CGACTGGGTG
3201 TCGAACAGGC GCTACACCTA CGGCACTTC CACTACAACC CTGCCACAT
3251 GATCGCCATC TCGTCTTCT TCACGAACGC GCTGGCTCTG GCGCTGCAGC
3301 GCGCCCTTGT GCTCTCCGCG GCCAACCCCG AGAAGGGCAA GGAAATGCGG
3351 ACGCCGGATC ACGAGGATAC GTTCTTCCCG GATCTGGTGC GCTACTCGAT

```

-continued

3401 CGGGACGCTC GGCATCCACC GCCTCGGCCT GGTGCTCTCG CTGAGCGCCG
3451 TCTTCTTCAG CGCCCTCTGC ATGATCATTAC CCGGCACCAT CTGGTTCGAT
3501 CAGTGGGTCC ACTGGTGGCA ATGGTGGGTG AAGCTGCCGT GGTGGGGCAA
3551 CATCCCCGGA GGCATCAATG GCTGAGTATC AGAACATCTT CTCCCAGGTC
3601 CAGGTCCGCG GACC CGCCGA CCTGGGGATG ACCGAAGACG TCAACCTGGC
3651 CAACCGTTCG GCGCTCGGTC CCTTCTCGAC CTGCTCGGC TGGTTCGGCA
3701 ACGCCAGCT CGGCCGATC TATCTCGGCT CGCTCGGCGT CCTGTCCCTC
3751 TTCTCGGGCC TGATGTGGTT CTTCACCATC GGGATCTGGT TCTGGTATCA
3801 GCGGGGCTGG AACCCGGCCG TCTTCTCTCG CGACCTGTTC TTCTTCTCGC
3851 TCGAGCCGCC GGCACCCGAA TACGGTCTGT CCTTCGCGGC TCCGCTGAAG
3901 GAAGCGGGGC TGTGGCTGAT CGCGTCTGTC TTCATGTTTC TCGCGGCTCG
3951 GTCCCTGGTGG GCGCCACCT ATCTCCGCGC TCAGGCGCTG GGCATGGGCA
4001 AGCACACCCG CTGGGCGTTC CTCTCGGCCA TCTGGCTGTG GATGGTGTG
4051 GGCTTCAATCC GTCCGATCCT CATGGGTGCC TGGTGGAAAG CGGTTCCTTA
4101 CGGCATCTTC TCGCACCTCG ACTGGACGAA CAACTTCTCG CTCGTCCACG
4151 GCAACCTGTT CTACAACCCC TTCCACGGTC TCTCGATCGC CTTCCTCTAC
4201 GGGTCCGGCC TGCTCTTCGC GATGACCGGT GCGACCATCC TCGCGGTCTC
4251 CCGCTTCGGC GCGGAGCGCG AGCTGGAGCA GATCGCCGAC CGCGGGACGG
4301 CAGCGGAGCG GCGCCCTCTC TTCTGGCGCT GGACCATGGG TTCAACGCC
4351 ACGATGGAA GCAATCCACC CTGGGCATC TGGATGGCGG TCCTCGTGAC
4401 CCTCACCGGC GGCATCGGgA TCCTGCTCTC GGGCACGGTC GTGGACAAC
4451 GGTACGTCGT GGGCCAGAAC CACGGCATGG CGCCGCTGAA CTGAGGAGCG
4501 ATCACAATGG CTGACAAGAC CATCTTCAAC GATCACCTCA ACACCAATCC
4551 GAAGACCAAC CTTCGCCTCT GGGTCTGTTT CCAGATGATG AAGGGTGGCG
4601 GCTGGGCTGG GCGGCTGTTT TCCGGACGC TCCTTCTCAT CGGGTCTTTC
4651 CGGGTGGTCC GCGGATGCT TCCGATCCAG GAGAACCAGG CGCCGGCGCC
4701 GAACATCACC GCGGCTCTGG AGACCGGGAT CGAGCTGATC AAGCATCTCG
4751 TCTGAGACAA GTCTCGGGGC AGGGCGCGCC GAGGCGCGCC GCTCTCCAA
4801 GTCCGGGCCA TATCGCCGCG CCGGGTCCGG GCGGACACCA CAGCCCGGTT
4851 CCCTTCTGTG TGGCGACAGG GACCTGGTGC CGTGTGGAAG ACCGCACGGC
4901 ACCCTTTTGA CATTCACGGG AGGCTCTGAT GACCAATCCC ACCCCGCGAC
4951 CCGAAACCCC GCTTTTGGAT CGCGTCTGCT GCCCGGCCGA CATGAAGGCG
5001 CTGAGTGAGC CCGAACCTGA GCGGCTGGCC GACGAAGTGC GTTCCGAGGT
5051 CAGTGATAGG GGTAGTTTCT TATTTTAGGC AGTTTTATATG AAATTAAGAC

5101 ATGCAGATGT CACAGTGGAT ATTGAACTGG TCTCGAAAGC TCAATATCCC
5151 CCAAAGCACA AGCACAACCT TCGACATCAT GCAGAAGCGT TTCCCGAAcc
5201 gcytctctega cgtgggcatc gccgagcagc atgcccgtgac cttccggycc
5251 ggccctcgcg GGGccggGat gaagcccttc tgcgcatctc attcctcgtt
5301 cctgcaacggy ggttacgacc agatcgccca tgacgtggcg ctgacagaacc
5351 ttcccgctccg cttcgtgacg gaccgggcyg ggtcgtggg ggcgcatggy
5401 gcyaccatg cgggggctTT CGACGTGGC TTCATCACTT CGCTGCCAA
5451 CATGACCGCT ATGGCCCGCG CCGACGAGGC CGAGCTCATC CACATGATCy
5501 cCACCCCGT GGCCTTCGAC GAGGGCCCCA TCGCCTTeCG CTTCGCCGG
5551 GCGGAGGGG TGGCGTOGA GATGCCCGAG CGCGGACCGG TGCTGGAGCC
5601 CGGCCGGGG CCGTGGTGC GCGAAGGGAC ggatgtcgcg atcctctcct
5651 tCGCGCGcA TCTGCACGAG gccTTGcAGG CggcGAAACT TCTCGAGGcc
5701 GAGGGGTGA CCGTGACCGT GGCCGACgce cgtTtctCgc gCCCCTCgA
5751 cACGGGGCTC ATCGACCAGC TCGTgcGCCA TCACGGGGCG CTGGTAACGG
5801 TGGAGCAGGG GGCcATGGGC GGCTTCGGCG CCCATGTCTC GCACATATCTC
5851 GCCAATTCCG GCGGCTTCGA CCGGGCCCTC CGCCTCCGGG TCATGACGCT
5901 GCCCGACCGC TTCATCGAGC AGGCGAGCCC CGAGgACATG TATGCCGATG
5951 CCGGGCTGCG GCGCGAGGAT ATCAAGCTTG GCGTAATCAT GGTCTAGCT
6001 GTTTCCTGTG TGAATTTGTT ATCCGCTCAC AATTCCACAC AACATACGAG
6051 CCGGAAGCAT AAAGTGTAAA GCCTGGGGTG CCTAATGAGT GAGCTAATCTC
6101 ACATTAATTG CGTTGCGCTC ACTGCCCGCT TTCCAGTCCG GAAACCTGTC
6151 GTGCCAGCTG CATTAATGAA TCGGCCAACG CGCGGGGAGA GCGGTTTTCG
6201 GTATTGGGCG CTCGGTCTTG CCTTGCTCGT CCGTGTGATG TTTACCCAGC
6251 TCCGGAAGT CGCTTCTTCT GATGGAGCGC ATGGGGACGT GCTTGGCAAT
6301 CACGCGCACC CCCC GGCCGT TTTAGCGGCT AAAAAAGTCA TGGCTTGCC
6351 CTCGGGCGGA CCACGCCCAT CATGACCTTG CCAAGCTCGT CCTGCTTCTC
6401 TTCGATCTTC GCCAGCAGGG CGAGGATCTG GGCATCACCG AACCGCGCCG
6451 TCGCGGGGTC GTCGGTGGAG CAGAGTTTCA GCAGGCGGCC CAGGCGGCCC
6501 AGTTCGCCAT TGATCGGGG CAGCTCGCGG ACGTGTCTAT AGTCCACGAC
6551 GCCCGTGATT TTGTAGCCCT GGCCGACGGC CAGCAGGTAG GCCGACGGC
6601 TCATGCCGGC CGCCCGCCCT TTTTCTCAA TCGCTCTTCG TTCGTCTGGA
6651 AGGCAGTACA CCTTGATAGG TGGGCTGCC TTCTGGTTG GCTTGGTTTC
6701 ATCAGCCATC CGCTTGCCCT CATCTGTAC GCCGGCGGTA CCGCGCCAGC
6751 CTCGAGAGC AGGATTTCCC TTGAGCACCG CCAGGTGCGA ATAAGGGACA

-continued

6801 GTGAAGAAGG AACACCCGCT CGCGGGTGGG CCTACTTCAC CTATCCTGCC
6851 CGGCTGACGC CGTTGGATAC ACCAAGGAAA GTCTACACGA ACCCTTTGGC
6901 AAAATCCTGT ATATCGTGCC AAAAAGGATG GATATACCGA AAAAATCGCT
6951 ATAATGACCC CGAAGCAGGG TTATGCAGCG GAAAAGCGCC ACGCCTCCCG
7001 AAGGGAGAAA GCGGGACAGG TATCCGGTAA GCGGCAGGGT CGGAACAGGA
7051 GAGCGCACGA GGGAGCTTCC AGGGGAAAAC GCCTGGTATC TTATAGTCC
7101 TGTCGGGTTT CGCCACCTCT GACTTGAGCG TCGATTTTTG TGATGCTCGT
7151 CAGGGGGGCG GAGCCTATGG AAAAACGCCA GCAACGCGGC CTTTTTACGG
7201 TTCCTGGCCT TTTGCTGGCC TTTTGCTCAC ATGTTCTTTC CTGCGTTATC
7251 CCCTGATTCT GTGGATAACC GTATTACCCG CTTTGAGTGA GCTGATACCG
7301 CTCGCCGCG CCGAACGACC GAGCGCAGCG AGTCAGTGAG CGAGGAAGCG
7351 GAAGAGCGCC AGAAGGCCCG CAGAGAGGCC GAGCGCGGCC GTGAGGCTTG
7401 GACGCTAGGG CAGGGCATGA AAAAGCCCGT AGCGGGCTGC TACGGCGCTC
7451 TGACCGGGTG GAAAGGGGGA GGGGATGTTG TCTACATGGC TCTGCTGTAG
7501 TGAGTGGGTT CGCTCCGGC AGCGGTCCTG ATCAATCGTC ACCCTTCTC
7551 GGTCTTCAA CGTTCCTGAC AACGAGCCCT CTTTTCGCCA ATCCATCGAC
7601 AATCACCGCG AGTCCCTGCT CGAACGCTGC GTCCGGACCG GCTTCGTCGA
7651 AGGCGTCTAT CGCGGCCCGC AACAGCGGGC AGAGCGGAGC CTGTTCAACG
7701 GTGCCGCCGC GCTCGCCGGC ATCGCTGTCG CCGGCCTGCT CCTCAAGCAC
7751 GGCCCAACA GTGAAGTAGC TGATTGTATC CAGCGCATFG ACGGCGTCCC
7801 CGGCCGAAAA ACCCGCCTCG CAGAGGAAGC GAAGCTGCGC GTCGGCCGTT
7851 TCCATCTGCG GTGCGCCCGG TCGCGTGCCG GCATGGATGC GCGCGCCATC
7901 CGGTAGGCG AGCAGCCCTT GCCTGAAGCT GCGGGCATTG CCGATCAGAA
7951 ATGAGCGCCA GTCGTCTGCG GC'TCTCGGCA CCGAATGCGT ATGATTCTCC
8001 GCCAGCATGG CTTGCGCCAG TCGTCTGAGC AGCGCCCGCT TGTTCCTGAA
8051 GTGCCAGTAA AGCGCCGGCT GCTGAAACCC CAACCCTTCC GCCAGTTTGC
8101 GTGTCGTCAG ACCGCTTACG CCGACCTCGT TCAACAGGTC CAGGGCGGCA
8151 CGGATCACTG TATTCGCGTG CAACTTGTGC ATGCTTGACA CTTTATCACT
8201 GATAAACATA ATATGTCCAC CAACTTATCA GTGATAAAGA ATCCCGCGGT
8251 TCAATCGGAC CAGCGGAGGC TGGTCCGGAG GCCAGACaTG AAACCCAACA
8301 TACCCTGAT CGTAAATTCTG AGCACTGTCC GCCTGACGC TGTGCGCATC
8351 GGCCTGATTA TGCCGGTGCT GCCGGGCCTC CTGCGCGATC TGGTTCACCT
8401 GAACGACGTC ACCGCCCACT ATGGCATTCT GTGCGCGCTG TATGCGTTGG
8451 TGCAATTTGC CTGCGCACCT GTGCTGGGCG CGCTGTCGGA TCGTTTCGGG

8501 CGGCGGCCAA TCTTGCTCGT CTGCTGGCC GCGGCCACTG TCGACTACGC
8551 CATCATGGCG ACAGCGCCTT TCCTTTGGGT TCTCTATATC GGGCGGATCG
8601 TGGCCGGCAT CACCGGGGCG ACTGGGGCGG TAGCCGGCGC TTATATTGCC
8651 GATATCACTG ATGGCGATGA GCGCGCGCGG CACTTCGGGT TCATGAGCGC
8701 CTGTTTCGGG TTCGGGATGG TCGCGGGACC TGTGCTCGGT GGGCTGATGG
8751 CGGTTTCTC CCCCACGCT CCGTTCCTCG CCGCGGACG CTTGAACGGC
8801 CTC AATTTCC TGACGGGCTG TTTCTTTTG CCGGAGTCCG ACAAAGGGGA
8851 ACGCCGGCCG TTACGCCGGG AGGCTCTCAA CCCGCTCGCT TCGTTCGGGT
8901 GGGCCCGGGG CATGACCCTG GTGCGCCGCC TGATGGCGGT CTTCTTCATC
8951 ATGCAACTTG TCGGACAGGT GCCGCGCGCG CTTTGGGTCA TTTTCGGCGA
9001 GGATCGCTTT CACTGGGACG CGACCACGAT CGGCATTTCC CTTGCGGCAT
9051 TTGGCATTCT GCATTCACTC GCCCAGGCAA TGATCACCAG CCCTGTAGCC
9101 GCCCGGCTCG GCGAAAGGCG GGCACCTCAT CTCGGAATGA TTGCGCAGG
9151 CACAGGCTAC ATCCTGCTTG CCTTCGCGAC ACGGGGATGG ATGGCGTTCC
9201 CGATCATGGT CCTGCTTGCT TCGGGTGGCA TCGGAATGCC GGCCTGCAA
9251 GCAATGTTGT CCAGGCAGGT GGATGAGGAA CGTCAGGGGC AGCTGCAAGG
9301 CTCACTGGCG GCGCTCACCA GCCTGACCTC GATCGTCGGA CCCCTCCTCT
9351 TCACGGCGAT CTATGCGGCT TCTATAACAA CGTGGAAACG GTGGGCATGG
9401 ATTGCAGGCG CTGCCCTCTA CTTGCTCTGC CTGCCGGCGC TGCGTCCGGG
9451 GCTTTGGAGC GCGCGAGGGC AACGAGCCGA TCGCTGATCG TGGAAACGAT
9501 AGGCCTATGC CATGCGGGTC AAGGCGACTT CCGGCAAGCT ATACGCGCCC
9551 TAGGAGTGCG GTTGAACGCT TGGCCAGCC AGATACTCCC GATCACGAGC
9601 AGGACGCCGA TGATTTGAAG CGCACTCAGC GTCTGATCCA AGAACAAACA
9651 TCCTAGCAAC ACGGCGGTCC CCGGGCTGAG AAAGCCAGT AAGGAAACAA
9701 CTGTAGGTTT GAGTCGCGAG ATCCCCGGA ACCAAAGGAA GTAGTATAAA
9751 CCCGCTCGA TCAGGCCGAG CCACGCCAGG CCGAGAACAT TGGTTCCTGT
9801 AGGCATCGGG ATTGGCGGAT CAAACTATAA AGCTACTGGA ACGAGCAGAA
9851 GTCCCTCCGG CGCCAGTTGC CAGGCGGTAA AGGTGAGCAG AGGCACGGGA
9901 GGTTCGCACT TGCGGGTFCAG CACGGTTCGG AACGCCATGG AAACCGCCCC
9951 CGCCAGGCCG GCTGCGACGC CGACAGGATC TAGCGCTGCG TTTGGTGTCA
10001 ACACCAACAG CGCCACGCCG GCAGTTCGGC AAATAGCCCC CAGGACCGCC
10051 ATCAATCGTA TCGGGCTACC TAGCAGAGCG GCAGAGATGA ACACGACCAT
10101 CAGCGGCTGC ACAGCGCCTA CCGTCGCCCC GACCCCGCCC GGCAGGCGGT
10151 AGACCGAAAT AAACAACAAG CTCCAGAATA GCGAAATATT AAGTGCGCCG

-continued

10201 AGGATGAAGA TGGCATCCA CCAGATFCCC GTTGGAACTCT GTCGGACGAT
10251 CATCACGAGC AATAAACCCG CCGGCAACGC CCGCAGCAGC ATACCGGGCA
10301 CCCCCTCGGC TCCTGTTCG GGCTCCACGA AAACGCCGGA CAGATGCGCC
10351 TTGTGAGCGT CCTTGGGGCC GTCCCTCTGT TTGAAGACCG ACAGCCCAAT
10401 GATCTCGCCG TCGATGTAGG CGCCGAATGC CACGGCATCT CGCAACCGTT
10451 CAGCGAAGCG CTCCATGGGC TTTTTCFCTT CGTGCFCGTA AACGGACCCG
10501 AACATCTCTG GAGCTTCTT CAGGGCCGAC AATCGGATCT CGCGAAATC
10551 CTGCACGTCG CCGCTCCAA GCCGTGCAAT CTGAGCCTTA ATCACAATF
10601 TCAATTTTAA TCCTCTGTTT ATCGGCAGTT CGTAGAGCGC GCCGTGCGTC
10651 CCGAGCGATA CTGAGCGAAG CAAGTGCCTC GAGCAGTGCC CGCTTGTTC
10701 TGAATGCGCA GTAAGCGCT GGCTGCTGAA CCCCCAGCCG GAACTGACCC
10751 CACAAGGCC TAGCGTTTGC AATGCACCAG GTCATATTG ACCCAGCGT
10801 GTTCCACCAG GCCGCTGCTT CGCAACTCTT CGCAGGCTTC GCCGACCTGC
10851 TCGCGCCACT TCTTACGCG GGTTGGAATCC GATCCGCACA TGAGGCGGAA
10901 GGTTCACGCT TTAGCGGGT ACGGCTCCCG GTGCGAGCTG AAATAGTCGA
10951 ACATCCGTCG GGCCTGCGC GACAGCTTGC GGTACTTCTC CCATATGAAT
11001 TTCGTGTAGT GGTGCGCAGC AAACAGCAGC ACGATTTCCT CGTCGATCAG
11051 GACCTGGCAA CGGGACGTTT TCTTGCCACG GTCCAGGACG CGGAAGCGGT
11101 CGACGAGCGA CACCGATTC AGGTGCCCAA CGCGTCCGGA CGTGAAGCC
11151 ATCGCCGTCG CCTGTAGCG CGACAGGCAT TCCTCGGCTC TCGTGAATA
11201 CCGGCCATTT ATCGACCAGC CCAGGTCCTG GCAAAGCTCG TAGAACGTGA
11251 AGGTGATCGG CTCGCCGATA GGGGTGCGCT TCGCGTACTC CAACACCTGC
11301 TGCCACACCA GTTCTGTCATC GTCGGCCCGC AGCTCGACGC CGGTGTAGGT
11351 GATCTTACAG TCCTTGTGTA CGTGGAAAAT GACCTTGTTC TGACGCGCT
11401 CGCGCGGAT TTTCTTGTG CGCGTGTGTA ACAGGGCAGA GCGGCGGTC
11451 TCGTTTGGCA TCCTCGCAT CGTGTCCGCG CACGGCGCAA TATCGAACAA
11501 GGAAGCTGCT ATTTCTTGA TCTGCTGCTT CGTGTGTTTC AGCAACCGCG
11551 CCTGCTGCGC CTCGCTGACC TGTTTTGCCA GGTCCTCGCC GCGGTTTTTT
11601 CGCTTCTTGG TCGTCAATAG TCCTCGCGTG TCGATGGTCA TCGACTTCGC
11651 CAAACCTGCC GCCTCTGTT CGAGACGACG CGAACCTCC ACGGCGGCG
11701 ATGGCGCGGG CAGGGCAGGG GGAGCCAGTT GCACGCTGTC GCGCTCGATC
11751 TTGGCCGTAG CTTGCTGGAC CATCGAGCCG ACGGACTGGA AGGTTTCGCG
11801 GGGCGCACGC ATGACGGTGC GGCTTCCGAT GGTTCGGCA TCCTCGGCG
11851 AAAACCCCGC GTCGATCAGT TCTTGCTGT ATGCCCTCCG GTCAAACGTC

11901 CGATTCAATC ACCCTCCTTG CGGGATGACC CCGACTCAGC CCGGGCAAT
11951 GTGCCCTTAT TCCTGATTTG ACCCGCTGCG TGCCCTGGTG TCCAGATAAT
12001 CCACCTTATC GGAATGAAG TCGGTCCCGT AGACCGTCTG GCCGTCTCTC
12051 TCGTACTTGG TATTCCGAAT CTTGCCCTGC ACGAATACCA GCGACCCCTT
12101 GCCCAAATAC TTGCCGTGGG CCTCGGCTCG AGAGCCAAAA CACTTGATGC
12151 GGAAGAAGTC GGTGCGCTCC TGCTTGTCCG CGGTCTGTCG CGCGCAACCC
12201 TTTGCGATCC GCAAGCGCGC GGTCCGCATC TTCACGCTGG AACAGTACGT
12251 CGAGGGCGGG ATCATGACCC GCAGCAATA CGAGGTCATT AAAAGCGCG
12301 TGATTGATGA TATAGCGGCC CGGCTGCTCC TGGTCTCTCG GCACCGAAT
12351 GGGTACTTTC ACCCCGCGCT CTTTGATCGT GGCACCGATT TCCGCGATGC
12401 TCTCCGGGGA AAAGCCGGGG TTGTCGGCGG TCCCGGGCTG ATGCGGATCT
12451 TCGTCCGATC GGTCCAGGTC CAGCTCGATA GGGCCGGAAC CGCCCTGAGA
12501 CGCGCGAGGA GCGTCCAGGA GGCTCGACAG GTCCGCGATG CTATCCAACC
12551 CCAGGCGGGA CCGCTGCGCC GCGCCTGCGG CTTCCCTGAGC GGCGGAGCG
12601 GTGTTTTTCT TGGTGGTCTT GGCTTGAGCC GCAGTCAATG GGAATCTCC
12651 ATCTTCGTTA ACACGTAATC AGCCAGGGCG CGAACCTCTT TCGATGCCTT
12701 GCGCGCGGCC GTTTTCTTGA TCCTCCAGAC CGGCACACCG GATGCGAGGG
12751 CATCGCGGAT GCTGCTGCGC AGGCCAACCG TGCCCGGAAT CATCATCTTG
12801 GGGTACCGCG CACAGAGCTC GGCTTGGTGG CGCGCTGGGC GCGGATTCGG
12851 CGCATCGACC TTGCTGGGCA CCATGCCAAG GAATTGACGC TTGGCGTTCT
12901 TCTGGCGCAC GTTCGCAATG GTCTGACCA TCTTCTGAT GCCCTGGATG
12951 CTGTACGCCCT CAAGCTCGAT GGGGACAGC ACATAGTCCG CCGCGAAGAG
13001 GCGCGCGGCC AGGCCGACGC CAAGGGTCCG GGCCTGTCTG ATCAGGCACA
13051 CGTCGAAGCG TTGGTTGCGC AGGGCCTTGA TGTTCGCCCC GAACAGCTCG
13101 CCGGCGTCTG CACGCGACAG CGGTTCGGCG TTCGCGAGTA CCGGTTGGA
13151 CTCGATGAGG GCGAGGCGCG CGGCTGGGCC GTCCGCGGCT GCGGTTGCGG
13201 TTTCCGTTCA GCCCGCGCA GGGACAGCG CGAACAGCTT GCTTGCATGC
13251 AGGCCGCTAG CAAAGTCTTT GAGCGTGTAG GACGCATTGC CCTGGGGGTC
13301 CAGGTCGATC ACGCAACCC GCAAGCCGCG CTCGAAAAAG TCGAAGGCAA
13351 GATGCACAAG GGTGCAAGTC TTGCCGACCG CGCCTTCTG GTTGGCCGTG
13401 ACCAAAGTTT TCATCGTTTG GTTTCCTGTT TTTTCTTGGC FTCCGCTTCC
13451 CACTTCCGGA CGATGTACGC CTGATGTTC GGCAGAACCG CCGTTACCCG
13501 CGCGTACCC TCGGGCAAGT TCTTGTCTTC GAACGCGGCC CACACGCGAT
13551 GCACCGCTTG CGACACTGCG CCCCCTGTC A GTCAGCGCA CGTTGCGAAC

13601 GTCGCTGTG GCTTCCATC GACTAAGAC CCCCAGCTA TCTCGATGTT
13651 CTGCTGCCCT ACTTCCAGCC CCTGGATCGC CTCCCTGGAAC TGGCTTTCGG
13701 TAAGCCGTTT CTTTATGAT AACACCCATA ATTTGCTCCG CGCCTTGGTT
13751 GAACATAGCG GTGACAGCCG CCAGCACATG AGAGAAGTTT AGCTAAACAT
13801 TTCTCGCACG TCAACACCTT TAGCCGCTAA AACTCGTCTT TGGCTAACA
13851 AAACAAAAGC CCGGAAACCG GGCTTTCGTC TCTTGCSSCT TATGGCTCTG
13901 CACCCGCGTC CATCACCAAC AGGTGCGGCA CGCGCTTAC TCGGTTGCGG
13951 ATCGACACTG CAGCCCAAC AAAGCCGTTT GCCCGCGCG CAGGATCGC
14001 CCGCATGATG CCGGCCACAC CGGCCATCGC CCACAGGTC GCCGCTTCC
14051 GGTTCATTTT CTGCTGGTAC TGCTTCGCAA TGCTGGACCT GCGCTACCA
14101 TAGGCTGACC GCTCGATGGC GTATGCGCTT TCTCCCTTG GGTAAAACC
14151 CAGCGCCGCA GCGGCGATTG CCATGCTGCC CGCCGCTTC CCGACCACGA
14201 CCGCGCACGC AGGCTTGGCG TCCAGACCTT CCGCCACGGC GAGCTGCGCA
14251 AGGACATAAT CAGCCCGCA CTTGGCTCCA CGCCCTCGA TCAGCTCTTG
14301 CACTCGCGCG AAATCCTTGG CCTCCACGGC CGCCATGAAT CCGCGACCGG
14351 GCGAAGGCTC GCGAGGCGC

-continued

3401 TCTTCCGCGA TCTGGTCGGC TACTCGATCG GGACGCTCGG CATCCACCGC
 3451 CTCGGCCTGC TGCTCTCGCT GAGCGCCGTC TTCTTCAGCG CCCTCTGCAT
 3501 GATCATTACC GGCACCATCT GGTTTCGATCA GTGGGTCGAC TGGTGGCAAT
 3551 GGTGGGTGAA GCTGCCGTGG TGGGCGAACA TCCCGGGAGG CATCAATGGC
 3601 TGAGTATCAG AACATCTTCT CCCAGGTCCA GGTCCGCGGA CCGGCCGACC
 3651 TGGGGATGAC CGAAGACGTC AACCTGGCCA ACCGTTCGGG CGTCCGGTCC
 3701 TTCTCGACCC TGCTCGGCTG GTTCGGCAAC GCCAGCTCG GCCCGATCTA
 3751 TCTCGGCTCG CTCGGCTGCC TGCCCTCTTT CTCGGGCTGT ATGTGGTTCT
 3801 TCACCATCGG GATCTGGTTC TGATATCAGG CGGGCTGGAA CCCGGCCGTC
 3851 TTCCTGCGCG ACCTGTTCTT CTCTCTCGTC GAGCCGCCGG CACCCGAATA
 3901 CGGTCTGTCC TTCGCGGCTC CGCTGAAGGA AGCGGGCTGT TGGCTGATCG
 3951 CGTCGTTCTT CATGTTCTGT GCGGTCTGGT CCTGGTGGGG CCGCACCTAT
 4001 CTCGCGCTCG AGGCGCTGGG CATGGGCAAG CACACCCTCT GGGCGTCTCT
 4051 CTCGGCCATC TGGCTGTGGA TGGTGTGGG CTTTCATCCGT CCGATCCTCA
 4101 TGGGGTCTCG GTCGGAAGCG GTTCCCTACG GCATCTTCTC GCACCTCGAC
 4151 TGGACGAACA ACTTCTCGCT CGTCCACGGC AACCTGTTCT ACAACCCCTT
 4201 CCACGGTCTC TCGATCGCCT TCCTCTACGG GTCGGCCCTG CTCTTCGCGA
 4251 TGCACGGTGC GACCATCCTC GCGGTCTCCC GCTTCGGCGG CGAGCGCGAG
 4301 CTGGAGCAGA TCGCCGACCG CGGGACGGCA GCGGAGCGGG CCGCCTCTTT
 4351 CTGGCGTGG ACCATGGGTT TCAACGCCAC GATGGAAGGC ATCCACCGCT
 4401 GGGCCATCTG GATGGCGGTC CTCGTGACCC TCACCGCGGG CATCGGgATC
 4451 CTGCTCTCGG GCACGGTCTG GGACAACCTG TACGTCTGGG GCCAGAACCA
 4501 CCGCATGGCG CCGCTGAAGT GAGGAGCGAT CACAATGGGT GACAAGACCA
 4551 TCTTCAACGA TCACTCAAC ACCAATCCGA AGACCAACCT TCGCCTCTGG
 4601 GTCGCTTTCC AGATGATGAA GGGTGCGGGG TGGGCTGGCG GCGTGTCTTT
 4651 CCGGACGCTC CTTCTCATCG GGTCTTCCG GGTGGTCCGG CGGATGCTTC
 4701 CGATCCAGGA GAACCAGGCT CCGGCGCCGA ACATCACCAG CGCTCTGGAG
 4751 ACCGGGATCG AGCTGATCAA GCATCTCGTC TGAGACAAAT CTCGGGGCAG
 4801 GCGCGCGCGA GCGCGCCCGC TCCTCCAAGT CCGGGCCATA TCGCCGGCCC
 4851 GGGTCCGGGG CGACACCACA GCCCGGTTCG CTTCTGTTG GCGACAGGGA
 4901 CCTGGTCCGG TGTGGAAGAC CGCACGGCAC CCTTTTGACA TTCACGGGAG
 4951 GCTCTGATGA CCAATCCAC CCCGCGACCC GAAACCCCGC TTTTGGATCG
 5001 CGCTCTGTCC CCGGCCGACA TGAAGGCGCT GAGTGACSCC GAAC TGAGGC
 5051 GGCTGGCCGA CGAAGTCCGT TCCGAGGTCA GTGATAGGGG TAGTTTCTTA

 5101 TTTTAGGCAG TTTATATGAA ATTAAGACAT GCAGATGTCA CAGTGGATAT
 5151 TGAAGTGGTC TCGAAGCTC AATATCCCC AAAGCACAAG CACAACTTC
 5201 GACATCATGC AGAAGCGTTT CCCGAaccgc gtctttagacg tgggcatcgc
 5251 cgagcagcatg gccgtgacct tcgcgcccg cctcgCcgGG GccggGatga
 5301 agccctctctg cgcgatctat tcctcgttcc tgcaacgggg ttacgaccag
 5351 atgcccctatg acgtggcgct gcagaacctt cccgtccgct tcgtgatcga
 5401 ccggggcgggg cctcgtggggg ccgatggcgc gaccatcgc ggggctTTCG
 5451 ACTTTGGCTT CATCATTTCG CTGCCAACA TGACCGTGAT GGCCGCGGCC
 5501 GACGAGGCCG AGCTCATCCA CATGATCgcc ACCGCCGTGG CCTTCGACGA
 5551 GGGCCCCATC GCCTTccGCT TCCCGCGGGG CGAGGGGGTG GGCGTCGAGA
 5601 TGCCCGAGCG CCGGACGGTG CTGGAGCCCC gCCGGGGCCG CGTGGTGGCG
 5651 GAAGGGAGcg atgtcgcgat cctctccttc GCGCGcATC TGAcGAGgc
 5701 cTTGcAGGCG gcGAACTTC TCGAGGccGA GGGGTGAGC GTGACCGTGG
 5751 CCGACgcccg CTTctCgCgC CCCTCgAcA CGGGGTCAT CGACCAGCTC
 5801 GTgcGCCATC ACGCGGCGCT GGTAAAGGTT GAGCAGGGGG CeATGGGCGG
 5851 CTTCCGGCGCC CATGTCTATGC ACTATCTCGC CAATTCCGGC GGCTTCGACG
 5901 GGGCCCTCGC GCTCCGGGTC ATGACGCTGC CCGACCGCTT CATCGAGCAG
 5951 GCGAGCCCCG AggACATGTA TGCCGATGCG GGGCTGCGGG CCGAGGATAT
 6001 CAAGCTTGGC GTAATCATGG TCATAGCTGT TTCTGTGTG AAATGTTTAT
 6051 CCGCTCACAA TTCCACACAA CATAACGACC GGAAGCATAA AGTGTAAAGC
 6101 CTGGGGTGGC TAATGAGTGA GCTAACTCAC ATTAATTGCG TTGCGCTCAC
 6151 TGCCCGCTTT CCACTCGGGA AACCTGTCGT GCCAGCTGCA TTAATGAATC
 6201 GGCCAACGCG CCGGGAGAGG CGGTTTGCCT ATTTGGCGCT CGGTCTTGCC
 6251 TTGCTCGTGC GTGATGTA CTACCAGCTC CGGAAGTGC CTCTTCTTGA
 6301 TGGAGCGCAT GGGACGTGC TTGGCAATCA CGCGACCCC CCGCCGCTTT
 6351 TAGCGGCTAA AAAAGTCATG GCTCTGCCCT CCGCGGACC ACGCCCATCA
 6401 TGACCTTGGC AAGCTCGTCC TGCTTCTTCT CGATCTTCGC CAGCAGGGCG
 6451 AGGATCTGTG CATCACCGAA CCGCGCCGTC CGCGGGTCTG CGGTGAGCCA
 6501 GAGTTTCAGC AGGCCGCCCA GCGCGCCAG GTCGCCATTG ATGCGGGCCA
 6551 GCTCGCGGAC GTGCTCATAG TCCACGACGC CCGTGATTTT GTAGCCCTGG
 6601 CCGACGGCCA GCAGGTAGGC CGACAGGCTC ATGCCGGCCG CCGCCGCTTT
 6651 TTCCTCAATC GCTCTCGTCT CGTCTGGAAG GCAGTACACC TTGATAGGTG
 6701 GGCTGCCCTT CCTGGTTGGC TTGTTTTCAT CAGCCATCCG CTTGCCCTCA
 6751 TCTGTTACGC CCGCGGTAGC CGGCCAGCCT CGCAGAGCAG GATTCCTGTT

-continued

6801 GAGCACCGCC AGGTGCGAAT AAGGGACAGT GAAGAAGGAA CACCCGCTCG
6851 CGGGTGGGCC TACTTCACCT ATCCTGCCCC GCTGACGCCG TTGGATACAC
6901 CAAGGAAAGT CTACACGAAC CCTTTGGCAA AATCCTGTAT ATCGTGGCAA
6951 AAAGGATGGA TATACCGAAA AAATCGCTAT AATGACCCCG AAGCAGGGTT
7001 ATGCAGCGGA AAAGCGCCAC GC'TTCCGAA GGGAGAAAGG CGGACAGGTA
7051 TCCGGTAAGC GGCAGGGTCG GAACAGGAGA GCGCACGAGG GAGCTTCCAG
7101 GGGGAAACGC CTGGTATCTT TATAGTCTCG TCGGGTTTCG CCACCTCTGA
7151 CTTGAGCGTC GATTTTTGTG ATGCTCGTCA GGGGGCGGA GCCTATGGAA
7201 AAACGCCAGC AACGCGGCC TTTTACGGTT CCTGGCCTTT TGCTGGCCTT
7251 TTGCTCACAT GTTCTTTCC TCGTTATCCC CTGATCTGT GGATAACCGT
7301 ATTACCCCTT TTGAGTGAGC TGATACCGCT CGCCGCGACC GAACGACCGA
7351 GCGCAGCGAG TCAGTGAGCG AGGAAGCGGA AGAGCCCGAG AAGGCCCGCA
7401 GAGAGCCCGA GCGCGGCCGT GAGGCTTGGA CGTAGGGGA GGGCATGAAA
7451 AAGCCCGTAG CGGGCTGCTA CGGGCTCTG ACGCGGTGGA AAGGGGAGG
7501 GGATGTTGTC TACATGGCTC TGCTGTAGTG AGTGGTTGC GCTCCGGCAG
7551 CGGTCTGAT CAATCGTCAC CCTTCTCGG TCCTTCAACG TCCTTGACAA
7601 CGAGCCTCCT TTTCCGCAAT CCATCGACAA TCACCCGAG TCCTTGCTCG
7651 AACGCTCGGT CCGGACCCGC TTCGTGGAAG GCGTCTATCG CGGCCCGCAA
7701 CAGCGCGGAG AGCGGAGCCT GTTCAACGGT GCCCGCGCG TCGCCGGCAT
7751 CGTGTGCGCC GGCCTGTCC TCAGCACGG CCCAACAGT GAAGTAGCTG
7801 ATGTGATCA GCGCATGAC GCGTCCCGG GCCGAAAAAC CCGCTCGCA
7851 GAGGAAGCGA AGTGGCGGT CGGCCGTTT CATCTGCGGT GCGCCCGGTC
7901 GCGTGCCGGC ATGATGCGC GCGCCATCG GGTAGCGAG CAGCCCTGTC
7951 CTGAAGCTGC GGCATTTCC GATCAGAAAT GAGCGCCAGT CGTCTCGGC
8001 TCTCGGCACC GAATGCGTAT GATTCGCCG CAGCATGGCT TCGGCCAGTG
8051 CGTCGAGCAG CGCCCGTTG TTCTGAAGT GCCAGTAAAG CGCCGGCTGC
8101 TGAACCCCA ACCGTTCCGC CAGTTTGGCT GTCGTCAGAC CGTCTACGCC
8151 GACCTCGTTC AACAGGTCCA GGGCGGCAG GATCACTGTA TTCGGCTGCA
8201 ACTTTGTCAT GCTTGACACT TTATCACTGA TAAACATAAT ATGTCCACCA
8251 ACTTATCAGT GATAAAGAA CCGCCGCTTC AATCGGACCA GCGGAGGCTG
8301 TCCCGGAGGC CAGACaTGAA ACCCAACATA CCCCTGATCG TAATTTGAG
8351 CACTGTGCGC CTCGACCTG TCGGCATCG CCTGATTATG CCGGTGCTGC
8401 CGGGCCTCCT GCGCGATCTG GTTCACTCGA ACGACGTCAC CGCCCACTAT
8451 GGCATTTCTG TGGCGCTGTA TGCCTTGGTG CAATTTGCGT GCGCACCTGT

8501 GCTGGGCGCG CTGTCGGATC GTTTCGGGCG GCGGCCAATC TTGCTCGTCT
8551 CGCTGGCCGG CGCCACTGTC GACTACGCCA TCATGGCGAC AGCGCCTTTC
8601 CTTTGGGTTT TCTATATCGG GCGGATCGTG GCCGGCATCA CCGGGGGGAC
8651 TGGGGCGGTA GCCGGCGCTT ATATTGCCGA TATCACTGAT GCGGATGAGC
8701 GCGCGCGGCA CTTCCGGTTC ATGAGCGCCT GTTTCGGGTT CCGGATGGTC
8751 GCGGGACCTG TGCTCGGTGG GCTGATGGGC GGTTTCTCCC CCCACGCTCC
8801 GTTCTTCGCG GCGCGAGCCT TGAACGGCCT CAATTTCTCG ACGGGCTGTT
8851 TCCTTTTGCC GGAGTCCGAC AAAGCGGAAC GCCGGCCGTT ACGCCGGGAG
8901 GCTCTCAACC CGCTCGCTTC GTTCCGGTGG GCCCGGGGCA TGACCGTCTG
8951 CGCCGCCCTG ATGGCGGTCT TCTTCATCAT GCAACTTGTG GGACAGGTGC
9001 CGGCCGCGCT TTGGGTCAAT TTCGGCGAGG ATCGCTTTCA CTGGGACGCG
9051 ACCACGATCG GCATTTGCGT TGCCGCAATTT GCATTTCTGC ATTCACTCGC
9101 CCAGGCAATG ATCACCGGCC CTGTAGCCGC CCGGCTCGGC GAAAGCCGGG
9151 CACTCATGCT CGGAATGATT GCCGACGGCA CAGGCTACAT CCTGCTTGCC
9201 TTCGCGACAC GGGGATGGAT GCGTTCCTCC ATCATGGTCC TGCTTGCTTC
9251 GGGTGGCATC GGAATGCCGG CGCTGCAAGC AATGTTGTCC AGGCAGGTGG
9301 ATGAGGAACG TCAGGGGCGC CTGCAAGGCT CACTGGCGGC GCTCACCAGC
9351 CTGACCTCGA TCGTCCGACC CCTCCTCTTC ACGCGGATCT ATGCGGCTTC
9401 TATAACAACG TGGAACGGGT GGGCATGGAT TGCAGGCGCT GCCCTCTACT
9451 TGCTCTGCCT GCCGGCGCTG CGTCCGCGGC TTTGGAGCGG CGCAGGGCAA
9501 CGAGCCGATC GCTGATCGTG GAAACGATAG GCCTATGCCA TGCCGGTCAA
9551 GGGGACTTCC GGAAGCTAT ACGCCCCCTA GGAGTCCGGT TGAACGTTG
9601 GCCCAGCCAG ATACTCCCGA TCACGAGCAG GACGCCGATG ATTTGAAGCG
9651 CACTCAGCGT CTGATCCAAG AACAAACATC CTAGCAACAC GCGGTTCCCC
9701 GGGCTGAGAA AGCCAGTAA GAAACAACCT GTAGGTTCTGA CTCGCGAGAT
9751 CCCCAGGAAC CAAAGGAAGT AGGTTAAACC CGCTCCGATC AGGCCGAGCC
9801 ACGCCAGGCC GAGAACATTG GTTCTGTAG GCATCGGGAT TGGCGGATCA
9851 AACACTAAAG CTACTGGAAC GAGCAGAAGT CCTCCGGCCG CCAGTTGCCA
9901 GCGGGTAAAG GTGAGCAGAG GCACGGGAGG TTGCCACTTG CCGGTGAGCA
9951 CGTTCCGAAA CGCCATGGAA ACCGCCCCCG CCAGGCCCGC TCGACGCGC
10001 ACAGGATCTA CGCTGCGTT TGGTGTCAAC ACCAACAGCG CCACGCCCGC
10051 AGTTCCGCAA ATAGCCCCCA GGACCGCCAT CAATCGTATC GGGCTACCTA
10101 GCAGAGCGGC AGAGATGAAC ACGACCATCA GCGGCTGCAC AGCGCTACC
10151 GTCGCCCGCA CCCCGCCCGG CAGGCGGTAG ACCGAAATAA ACAACAAGCT

-continued

10201	CCAGAAATAGC	GAATATATTA	GTGCGCCGAG	GATGAAGATG	CGCATCCACC
10251	AGATTCCCGT	TGGAATCTGT	CGGACGATCA	TCACGAGCAA	TAAACCCGCC
10301	GGCAACGCC	GCAGCAGCAT	ACCGCGGACC	CCTCGGCCTC	GCTGTTCCGG
10351	CTCCACGAAA	ACGCCGGACA	GATGCGCCTT	GTGAGCGTCC	TTGGGGCCGT
10401	CCTCCTGTTT	GAAGACCGAC	AGCCCAATGA	TCTCGCCGTC	GATGTAGGGC
10451	CCGAATGCCA	CGGCATCTCG	CAACCGTTCA	GCGAACGCCT	CCATGGGCTT
10501	TTTCTCCTCG	TGCTCGTAAA	CGGACCCGAA	CATCTCTGGA	GCTTTCTTCA
10551	GGGCCGACAA	TCCGATCTCG	CGGAAATCCT	GCACGTCGGC	CGTCCAAGC
10601	CGTCCAACT	GAGCCTTAAT	CACAATFTGC	AATTTTAATC	CTCTGTTTAT
10651	CGGCAGTTCG	TAGAGCGCGC	CGTGCCTCCC	GAGCGATACT	GAGCGAAGCA
10701	AGTGCCTCGA	GCAGTGCCTCG	CCTGTTCCCTG	AAATGCCAGT	AAAGCGCTGG
10751	CTGTGTAACC	CCCAGCCGGA	ACTGACCCCA	CAAGGCCCTA	GCCTTTGCAA
10801	TGCACCAAGT	CATCATTGAC	CCAGGCGTGT	TCCACCAGGC	CGCTGCCTCG
10851	CAACTCTTCG	CAGGCTTCGC	CGACCTGCTC	GCGCCACTTC	FTCACGGCGG
10901	TGGAATCCGA	TCCGCACATG	AGCCGGAAGG	TTTCCAGCTT	GAGCGGGTAC
10951	GGTCCCGGTT	GCGAGCTGAA	ATAGTCGAAC	ATCCGTCCGG	CCGTCCGCGA
11001	CAGTTGCGGG	TACTTCTCCC	ATATGAATTT	CGTGTAGTGG	TCGCCAGCAA
11051	ACAGCACGAC	GATTTCCCTCG	TGCATCAGGA	CCTGGCAACG	GGACGTTTTTC
11101	TTGCCACGGT	CCAGGACCGC	GAAGCGGTGC	AGCAGCGACA	CCGATTCCAG
11151	GTGCCCAACG	CGTTCGGACG	TGAAGCCCAT	CGCCGTCGCC	TGTAGGCGCG
11201	ACAGGCATT	CTCGGCCTTC	GTGTAATACC	GGCCATTGAT	CGACCAGCCC
11251	AGGTCCCTGG	AAAGCTCGTA	GAACGTGAAG	GTGATCGGGT	CGCCGATAGG
11301	GGTGCCTTTC	GCGTACTCCA	ACACCTGCTG	CCACACCAGT	TCGTCACTGT
11351	CGGCCCGCAG	CTCGACCGCC	GTGTAGGTGA	TCTTCACGTC	CTTGTGACG
11401	TGGAATAATGA	CCTGTGTTTTG	CAGCCCTCG	CGCGGGATT	TCTTGTGCG
11451	CGTGGTGAAC	AGGGCAGAGC	GGCCGTGTCT	GTTTGGCATC	GCTCGCATCG
11501	TGTCGGCCCA	CGGCCAATA	TCGAACAAGG	AAAGTGCAT	TTCTTTGATC
11551	TGCTGCTTCG	TGTGTTTCAG	CAACGCGGCC	TGCTTGGCCT	CGCTGACCTG
11601	TTTTTGCCAGG	TCTTCGCGGG	CGTTTTTTCG	CTTCTTGGTG	GTCATAGTTC
11651	CTCGCGTGT	GATGGTCATC	GACTTCGCCA	AACCTGCCGC	CTCCTGTTTC
11701	AGACGACGCG	AACGCTCCAC	GGCGCGCAT	GGCGGGGCA	GGGACGGGGG
11751	AGCCAGTTGC	ACGCTGTFCG	GCTCGATCTT	GGCCGTAGCT	TGCTGGACCA
11801	TCGAGCCGAC	GGACTGGAAG	GTTCGCGGG	GCGCACGCAT	GACGGTGC
11851	CTTGCGATGG	TTTTCGGCATC	CTCGGCGGAA	AACCCCGCGT	CGATCAGTTC
11901	TTGCCTGTAT	GCCTTCCGGT	CAAACGTCGG	ATTCATTCAC	CCTCCTTGCG
11951	GGATTGCCCC	GACTCACGCC	GGGCAATGT	GCCCTTATTC	CTGATTTGAC
12001	CCGCCTGGTG	CCTTGGTGTG	CAGATAATCC	ACCTTATCGG	CAATGAAATC
12051	GGTCCCGTAG	ACCGTCTGGC	CGTCCCTCTC	GTACTTGGTA	TTCCGAATCT
12101	TGCCCTGCAC	GAATACCAGC	GACCCCTTGC	CCAAATACTT	GCCGTGGGGC
12151	TCGGCCTGAG	AGCCAAAACA	CTTGATGCGG	AAGAAGTCCG	TGCGCTCCTG
12201	CTTGTCCCGG	GTCGTGGCCG	CGCCAACCTT	TGCGATCCCG	AAGCGCGCGG
12251	TCGCCATCTT	CACGCTGGAA	CAGTACGTCG	AGGCGGGCAT	CATGACCCGC
12301	GAGCAATACG	AGGTCAATTA	AAGCGCCGTG	ATTGATGATA	TAGCGGCCCG
12351	GCTGCTCCTG	GTTCTCGCGC	ACCGAAATGG	GTGACTTAC	CCCGCGCTCT
12401	TTGATCTGGG	CACCATTTC	CGCATGCTC	TCCGGGAAA	AGCCGGGGTT
12451	GTCGGCCGTC	CGCGGCTGAT	CGGGATCTTC	GTGATCAGG	TCCAGGTCCA
12501	GCTCGATAGG	GCCGGAACCG	CCCTGAGACG	CCGCAGGAGC	GTCCAGGAGG
12551	CTCGACAGGT	CGCCGATGCT	ATCCAACCCC	AGGCCGGACG	GCTGCGCCGC
12601	GCCTGCGGCT	TCTTGCAGCG	CCGACGCGGT	GTTTTTCTTG	GTGGTCTTGG
12651	CTTGAGCCGC	AGTCAATGGG	AAATCTCCAT	CTTCGTGAAC	ACGTAATCAG
12701	CCAGGGCGCG	AACCTCTTTC	GATGCCCTTC	GCGCGCCGT	TTCTTTGATC
12751	TTCCAGACCG	GCACACCGGA	TGCGAGGGCA	TCGGCGATGC	TGCTGCGCAG
12801	GCCAACCGTG	GCCGGAATCA	TCATCTTGGG	GTACGCGGCC	AGCAGCTCGG
12851	CTTGGTGGCG	CGCGTGGCGC	GGATTCGCGG	CATCGACCTT	GCTGGGCACC
12901	ATGCCAAGGA	ATTGCAGCTT	GGCGTTCTTC	TGGCGCACGT	TCGCAATGGT
12951	CGTGACCATC	TTCTTGATGC	CTTGATGCT	GTACGCCTCA	AGCTCGATGG
13001	GGGACAGCAC	ATAGTCCGGC	GCGAAGAGGG	CGGCCGCCAG	GCCGACGCCA
13051	AGGGTCCGGG	CCGTGTCGAT	CAGGCACACG	TCGAAGCCTT	GGTTCGCCAG
13101	GGCCTTGATG	TTCCGCCCGA	ACAGCTCGCG	GGCGTCTCC	AGCGACAGCC
13151	GTTCCGGGTT	CGCCAGTACC	GGGTGGACT	CGATGAGGGC	GAGGCGCGCG
13201	GCCTGGCCGT	CGCCGGCTGC	GGGTGCGGTT	TCGGTCCAGC	CGCCGGCAGG
13251	GACAGCGCCG	AACAGCTTGC	TTGCATGCAG	GCCGGTAGCA	AAGTCTTGA
13301	GCGTGTAGGA	CGCATTTGCC	TGGGGGTCCA	GTCGATCAC	GGCAACCCGC
13351	AAGCCGCGCT	CGAAAAAGTC	GAAGCAAGA	TGCACAAGGG	TCGAAGTCTT
13401	GCCGACGCCG	CCTTTCTGGT	TGGCCGTGAC	CAAAGTTTTT	ATCGTTTGGT
13451	TTCTGTGTTT	TTCTTGGCGT	CCGCTTCCCA	CTTCCGGACG	ATGTACGCCCT
13501	GATGTCCCGG	CAGAACCGCC	GTTACCCGCG	CGTACCCTTC	GGGCAAGTTC
13551	TTGTCTCGA	ACGCGGCCCA	CACGCGATGC	ACCGCTTGGC	ACACTGCGCC
13601	CCTGGTCAAGT	CCCAGCGACG	TTGCGAAGCT	CGCCTGTGGC	TTCCCATCGA
13651	CTAAGACGCC	CCGCGCTATC	TCGATGGTCT	GCTGCCCCAC	TTCCAGCCCC
13701	TGGATCCGCT	CCTGGAACCTG	GCTTTCGGTA	AGCCGTTTCT	TCATGGATAA
13751	CACCCATAAT	TTGCTCCGCG	CCTTGGTTGA	ACATAGCCGT	GACAGCCGCC
13801	AGCACATGAG	AGAAGTTTAG	CTAAACATTT	CTCGCACGTC	AACACCTTTA
13851	GCCGCTAAAA	CTGCTCCTTG	GCATAACAAA	ACAAAAGCCC	GGAAACCCGG
13901	CTTTCGCTCT	TTGCCGCTTA	TGGCTCTGCA	CCCGCTCCA	TCACCAACAG
13951	GTCGCGCACG	CGCTTCACTC	GGTTGCGGAT	CGACACTGCC	AGCCCAACAA
14001	AGCCGGTTTC	CGCCGCCGCC	AGGATCGCGC	CGATGATGCC	GGCCACCCG
14051	GCCATCGCCC	ACCAGGTCGC	CGCCTTCCCG	TTCCATTCCCT	GCTGGTACTG
14101	CTTCGCAATG	CTGACCTCG	GCTCACCATA	GGCTGACCCG	TCGATGGCGT
14151	ATGCCGCTTC	TCCCTTGGC	GTAACACCCA	GCGCCGACG	GGCATTGCC
14201	ATGCTGCCCG	CCGCTTTCCC	GACCACGACG	CGCGACCAG	GCTTGGCGTC
14251	CAGACCTTCG	GCCACGGCGA	GCTGCGCAAG	GACATAATCA	GCCGCGGACT
14301	TGGCTCCACG	CGCCTCGATC	AGCTCTTGCA	CTCGCGCGAA	ATCCTTGGCC
14351	TCCACGGCCG	CCATGAATCG	CGCACGCGGC	GAAGGTCCTC	CAGGGCCG

-continued

3401 CGATCTGGTC GGCTACTCGA TCGGGACGCT CGGCATCCAC CGCCTCGGCC
3451 TGCTGCTCTC GCTGAGCGCC GTCTTCTTCA GCGCCCTCTG CATGATCATT
3501 ACCGGCACCA TCTGGTTCGA TCAGTGGGTC GACTGGTGGC AATGGTGGGT
3551 GAAGCTCGCC TGGTGGGCGA ACATCCCGGG AGGCATCAAT GGCTGAGTAT
3601 CAGAACATCT TCTCCCAGGT CCAGGTCCGC GGACCGGCCG ACCTGGGGAT
3651 GACCGAAGAC GTCAACCTGG CCAACCGTTC GGGCGTCGGT CCCTTCTCGA
3701 CCCTGCTCGG CTGGTTCGGC AACGCCCAGC TCGGCCCGAT CTATCTCGGC
3751 TCGTTCGGCG TCCTGTCCCT CTCTCGGGC CTGATGTGGT TCTTACCAT
3801 CGGGATCTGG TTCTGGTATC AGGCGGGCTG GAACCCGGCC GTCTTCTGTC
3851 GCGACCTGTT CTCTTCTCG CTCGAGCCGC CGGCACCCGA ATACGGTCTG
3901 TCCTTCGCGG CTCGCTGAA GGAAGCGGG CTGTGGCTGA TCGCTCGTT
3951 CTTCATGTTC GTGCGGTCT GTTCCTGGTG GGGCCGCACC TATCTCGCG
4001 CTCAGGCGCT GGGCATGGC AAGCACACCG CCTGGGCGTT CCTCTCGGCC
4051 ATCTGGCTGT GGATGGTGT GGGCTTTCATC CGTCCGATCC TCATGGGGTC
4101 CTGGTCCGAA GCGGTTCCTT ACGGCATCTT CTCGCACCTC GACTGGACGA
4151 ACAACTTCTC GCTCGTCCAC GGCAACTGT TCTACAACCC CTTCACGGT
4201 CTCTCGATCG CCTTCTCTTA CCGGTTCGGC CTGCTCTTCG CGATGCACGG
4251 TCGGACCATC CTCGCGGTCT CCGCTTCGG CCGCGAGCGC GAGCTGGAGC
4301 AGATCGCCGA CCGCGGGACG GCAGCGGAGC GGGCCGCCCT CTCTGGCCG
4351 TGGACCATGG GTTTC AACCG CACGATGGAA GGCATCCACC GCTGGGCCAT
4401 CTGGATGGCG GTCCTCGTGA CCCTCACCGG CGGCATCGGG ATCTTGCTCT
4451 CGGGCACGGT CGTGGAACAAC TGGTACGTCT GGGGCCAGAA CCACGGCATG
4501 GCGCCGCTGA ACTGAGGAGC GATCACAATG GGTGACAAGA CCATCTTCAA
4551 CGATCACCTC AACACCAATC CGAAGACCAA CCTTCGCCCT TGGGTCCGCT
4601 TCCAGATGAT GAAGGGTGGC GGCTGGGCTG GCGGCGTGT CTTCGGGACG
4651 CTCCTTCTCA TCGGGTTCCT CCGGTGGTC GGGCGGATGC TTCCGATCCA
4701 GGAGAACCCG GCTCCGGCGC CGAACATCAC CCGCGCTCTG GAGACCCGGA
4751 TCGAGCTGAT CAAGCATCTC GTCTGAGACA AGTCTCGGG CAGGGCGCGC
4801 CGAGGCGGCC CGTCTCTCCA AGTCCGGGCC ATATCGCCCG CCGGGTCCG
4851 GGGCGACACC ACAGCCCGGT TCCTTTCCTG TTGGCGACAG GGACCTGGTG
4901 CCGTGTGGAA GACCGCACGG CACCTTTTGG ACATTACAGG GAGGCTCTGA
4951 TGACCAATCC CACCCCGCA CCCGAAACCC CGCTTTTGG TCGCTCTGC
5001 TGCCCGGCCG ACATGAAGGC GCTGAGTGAC CGCGAACTGG AGCGGCTGGC
5051 GCACGAAGTG CGTTCGGAG TCAGTGATAG GGGTAGTTTC TTATTTTAGG

5101 CAGTTTATAT GAAATTAAGA CATGCAGATG TCACAGTGGG TATTGAACTG
5151 GTCTCGAAGG CTCAATATCC CCCAAAGCAC AAGCACAAC TTCGACATCA
5201 TGCAGAAGCG TTTCCCGAAG cgcgttcttc acgtgggeat cgcgagcag
5251 catgcgctga ccttcggggc cgcctcgc cgggcccggg tgaagccctt
5301 ctgcgctgat tctctcctct tctcgaacg gggttacgac cagatcgcgc
5351 atgacgtggc gctgcagaac cttcccgtcc gcttcgtgat cgaccgggag
5401 gggctcgtgg gggccgatgg cgcgaccat gggggggcct TCGACGTTGG
5451 CTTCATCACT TCGCTGCCCA ACATGACCGT GATGGCCGCG GCCACGAGG
5501 CCGAGCTCAT CCACATGATC gcCACCGCCG TGGCCTTCGA CGAGGGCCCC
5551 ATCGCCTTcC GCTTCCCGCG GGGCGAGGGG GTGGGCGTCG AGATCCCCGA
5601 GCGCGGGACG GTGCTGGAGC CCGCGCGGGG CCGCGTGGTG CGCGAAGGGA
5651 Cggatgtgct gatcctctcc ttCGGCGCGc ATCTGCAcGA GgctTGGcAG
5701 GcggcGAAAC TTCTCGAGGc cGAGGGGGTG AGCGTGACCG TGGCCGACgc
5751 ccgCTTctCg CgCCCGCTCg AcACGGGGCT CATCGACCAG CTCGTgcGCC
5801 ATCACCGCGC GCTGGTAACG GTGGAGCAGG GGGccATGGG CGGCTTCGGC
5851 GCCCATGTCA TGCACTATCT CGCCAATTC GCGGCTTCG ACGGGGGCCT
5901 CCGCTCCCG GTCATGACCG TGCCCGACCG CTTCATCGAG CAGGGGAGCC
5951 CCGAGgACAT GTATGCCGAT GCGGGGCTGC GGGCCGAGGA TATCAAGCTT
6001 GCGTAATCA TGGTCATAGC TGTTCTCTGT GTGAAATGT TATCCGCTCA
6051 CAATCCACA CAACATACGA GCCGGAAGCA TAAAGTGTAA AGCCTGGGGT
6101 GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC
6151 TTTCCAGTCG GGAACCTGT CGTGCCAGCT GCATTAATGA ATCGGCCAAC
6201 GCGCGGGGAG AGGCGGTTTG CGTATTGGGC GCTCGTCTT GCCTTGCTCG
6251 TCGGTGATGT ACTTACCAG CTCCGCGAAG TCGCTTCTT TGATGGAGCG
6301 CATGGGACG TGCTTGGCAA TCACGCGCAC CCCCCGGCCG TTTTAGCGGC
6351 TAAAAAAGTC ATGGCTCTGC CCTCGGGCGG ACCACGCCA TCATGACCTT
6401 GCCAAGCTCG TCCTGTCTCT CTTCGATCTT CGCCAGCAGG GCGAGGATCG
6451 TGGCATCACC GAACCCGCGC GTGCGCGGGT CGTCGGTGAG CCAGAGTTTC
6501 AGCAGGCGCC CAGGCGGCC CAGGTCGCCA TTGATGCGGG CCAGCTCGCG
6551 GACGTGCTCA TAGTCCACGA CGCCCGTGAT TTGTAGCCC TGGCCGACGG
6601 CCAGCAGGTA GGCAGCAGG CTATGCCCCG CCGCCGCCG CTTTTCTCTA
6651 ATCGCTCTTC GTTCGTCTGG AAGGCAGTAC ACCTTGATAG GTGGGCTGCC
6701 CTCTCTGGTT GGCTTGGTTT CATCAGCCAT CCGCTTGCCT TCATCTGTTA
6751 CCGCGGCGGT AGCCGGCCAG CCTCGCAGAG CAGGATTTCC GTTGAGCACC

-continued

6801 GCCAGGTGCG AATAAGGGAC AGTGAAGAAG GAACACCCGC TCGCGGGTGG
6851 GCCTACTTCA CCTATCCTGC CCGGCTGACG CCGTTGGATA CACCAAGGAA
6901 AGTCTACACG AACCTTTGG CAAAATCCTG TATATCGTGC GAAAAAGGAT
6951 GGATATACCG AAAAAATCGC TATAATGACC CCGAAGCAGG GTTATGCAGC
7001 GGAAGAGCGC CACGCTTCCC GAAGGGAGAA AGGCGGACAG GTATCCGGTA
7051 AGCGGCAGGG TCGGAACAGG AGAGCGCAGC AGGGAGCTTC CAGGGGGAAA
7101 CGCCTGGTAT CTTTATAGTC CTGTGCGGTT TCGCCACCTC TGACTTGAGC
7151 GTCGATTTTT GTGATGCTCG TCAGGGGGGC GGAGCCTATG GAAAAACGCC
7201 AGCAACGCGG CCTTTTTACG GTTCCCTGGCC TTTTGCTGGC CTTTGGCTCA
7251 CATGTTCTTT CCTGCGTTAT CCCCTGATTC TGTGGATAAC CGTATTACCG
7301 CCTTTGAGTG AGCTGATACC GCTCGCCGCA GCCGAACGAC CGAGCGCAGC
7351 GAGTCAGTGA GCGAGGAAGC GGAAGAGCGC CAGAAGGCCG CCAGAGAGGC
7401 CGAGCGCGGC CGTGAGGCTT GGACGCTAGG GCAGGGCATG AAAAAAGCCG
7451 TAGCGGGGTG CTACGGGCGT CTGACGCGGT GGAAGGGGGG AGGGGATGTT
7501 GTCTACATGG CTCGCTGTA GTGAGTGGGT TCGCTCCGG CAGCGTCTCT
7551 GATCAATCGT CACCCTTCT CGTCCCTCA ACCTTCCCTGA CAACGAGCCT
7601 CCTTTTCGCC AATCCATCGA CAATCACCCG GAGTCCCTGC TCGAACCGTG
7651 CGTCCGGACC GGTTCGTCG AAGGCGTCTA TCGCGGCCCG CAACAGCGGC
7701 GAGAGCGGAC CCTGTTCAAC GGTGCCGCG CGCTCGCCGG CATCGCTGTC
7751 GCCGCCCTGC TCCTCAAGCA CGGCCCAAC AGTGAAGTAG CTGATGTGTA
7801 TCAGCGCATT GACGGCGTCC CCGGCCGAAA AACCCGCCTC GCAGAGGAAG
7851 CGAAGCTGCG CGTGGGCGGT TTCCATCTGC GGTGCGCCCG GTCGCGTGCC
7901 GGCATGGATG CGCGCGCCAT CGCGGTAGGC GAGCAGCGCC TGCCGTAAGC
7951 TCGGGGCATT CCGGATCAGA AATGAGCGCC AGTCTGCTGC GGCTCTCGGC
8001 ACCGAATGCG TATGATTCTC CGCCAGCATG GCTCGGCCA GTGCTCGAG
8051 CAGCGCCGCG TTGTTCTGA AGTGCCAGTA AAGCGCCGCG TGCTGAACCC
8101 CCAACCGTTC CGCCAGTTG CGTGTCTGTA GACCGTCTAC GCCGACCTCG
8151 TTCAACAGGT CCAGGGCGGC ACGGATCACT GTATTGCGGT GCAACTTTGT
8201 CATGCTTGAC ACTTTATCAC TGATAAACAT AATATGTCCA CCAACTTATC
8251 AGTGATAAAG AATCCGCGCG TTCAATCGGA CCAGCGGAGG CTGGTCCGGA
8301 GGCCAGACaT GAAACCCAAC ATACCCCTGA TCGTAATTCT GAGCACTGTC
8351 GCGCTCGACG CTGTCGCGAT CGGCTGATT ATGCCGGTGC TGCCGGGCTC
8401 CCTGCGCGAT CTGTTCACT CGAACGACGT CACCCGCCAC TATGGCATTC
8451 TGCTGGCCGT GTATGCGTTG GTGCAATTTG CCTGCGCACC TGTGCTGGGC

8501 GCGTGTGCG ATCGTTTCGG GCGGCGGCCA ATCTTGCTCG TCTCGTGGC
8551 CGGCGCCACT GTCGACTACG CCATCATGGC GACAGCGCCT TTCTTTGGG
8601 TTCTCTATAT CCGGCGGATC GTGGCCGGCA TCACCGGGGC GACTGGGGCG
8651 GTAGCCGGCG CTTATATTCG CGATATCACT GATGGCGATG AGCGCGGGCG
8701 GCACCTCGCG TTCATGAGCG CCTGTTTCGG GTTCGGGATG GTCGCGGGAC
8751 CTGTGCTCGG TGGGCTGATG GCGGTTTCT CCCCCACGC TCCGTTCTTC
8801 GCGCGGCGAG CCTTGAACGG CCTCAATTTT CTGACGGGCT GTTTCCTTTT
8851 GCCGGAGTGC CACAAGGGCG AACGCGGGCC GTTACGCCGG GAGGCTCTCA
8901 ACCGCTCGCG TTCGTTCCGG TGGGCGCGGG GCATGACCGT CGTCGCGCC
8951 CTGATGGCGG TCTTCTTCAT CATGCAACTT GTCGGACAGG TGCCGGCCGC
9001 GCTTTGGGTC ATTTTCGGCG AGGATCGCTT TCACTGGGAC GCGACCACGA
9051 TCGGCATTTT GCTTGCAGCA TTTGGCATTC TGCAATCACT CGCCAGGCA
9101 ATGATCACCG GCCCTGTAGC CGCCCGGCTC GCGAAAGGC GGGCACTCAT
9151 GCTCGGAATG ATTGCGGACG GCACAGGCTA CATCTGCTT GCCTTCGCGA
9201 CACGGGGATG GATGGCGTTC CCGATCATGG TCCTGCTTGC TTCGGGTGGC
9251 ATCGGAATGC CCGCGCTGCA AGCAATGTTG TCCAGGCAGG TGGATGAGGA
9301 ACCTCAGGGG CAGCTGCAAG GCTCACTGGC GCGCTCACC AGCCTGACCT
9351 CGATCGTCCG ACCCTCTCTC TTCACGGCGA TCTATGCGGC TTCTATAACA
9401 ACCTGGAACG GGTGGGCATG GATTGACGGC GCTGCCCTCT ACTTGCTCTG
9451 CCTGCCGGCG CTGCTGCGCG GGCTTTGGAG CCGCGCAGGG CAACGAGCCG
9501 ATCGCTGATC GTGGAACGA TAGGCCTATG CCATGCGGGT CAAGCGACT
9551 TCCGGCAAGC TATACCGGCC CTAGGAGTGC GGTGGAAAGC TTGGCCAGC
9601 CAGATACTCC CGATCACGAG CAGGACGCCG ATGATTTGAA GCGCACTCAG
9651 CGTCTGATCC AAGAACAACC ATCCTAGCAA CACGGCGGTC CCCGGGCTGA
9701 GAAAGCCAG TAAGGAAACA ACTGTAGGTT CGAGTCGCGA GATCCCCGG
9751 AACCAAGGA AGTAGGTTAA ACCCGCTCCG ATCAGGCCGA GCCACGCCAG
9801 GCGGAGAACA TTGGTTCCTG TAGGCATCGG GATTGGCGGA TCAAACACTA
9851 AAGCTACTGG AACGAGCAGA AGTCTCCCG CCGCCAGTTG CCAGGCGGTA
9901 AAGGTGAGCA GAGGCACGGG AGGTTGCCAC TTGCGGGTCA GCACGGTTC
9951 GAACGCCATG GAAACCGCCC CCGCCAGGCC CGCTGCGACG CCGACAGGAT
10001 CTAGCGTTCG GTTGTGTGTC AACACCAACA GCGCCACGCC CGCAGTTCGG
10051 CAAATAGCCC CAGGACCGC CATCAATCGT ATCGGGCTAC CTAGCAGAGC
10101 GGCAGAGATG AACACGACCA TCAGCGGCTG CACAGCGCCT ACCGTCGCG
10151 CGACCCCGCC CGCGAGGCGG TAGACCGAAA TAAACAACAA GCTCCAGAAT

-continued

10201 AGCGAAATAT TAAGTGCGCC GAGGATGAAG ATGCGCATCC ACCAGATTCC
10251 CGTTGGAAATC TGTCGGACGA TCATCACGAG CAATAAACCC GCCGGCAACG
10301 CCCCGCAGCAG CATAACCGCC ACCCTCGGC CTCCGTGTTT GGGCTCCACG
10351 AAAACGCCGG ACAGATGCGC CTTGTGAGCG TCCTTGGGGC CGTCTCCTG
10401 TTTGAAAGACC GACAGCCCAA TGATCTCGCC GTCCGATGTAG GCGCCGAATG
10451 CCACGGCATT TCGCAACCGT TCAGCGAACG CCTCCATGGG GGGGGCTCC
10501 TCGTGCCTGT AAACGGACCC GAACATCTCT GGAGCTTCTC TCAGGGCCGA
10551 CAATFCGGATC TCGCGAAAT CCTGCACGTC GGCCGCTCCA AGCCGTGAA
10601 TCTGAGCCTT AATCACAATT GTCAATTTTA ATCCTCTGTT TATCGGCAGT
10651 TCGTAGAGCG CGCCGTGCGT CCCGAGCGAT ACTGAGCGAA GCAAGTGCCT
10701 CGAGCAGTGC CCGCTTGTTT CTGAAATGCC AGTAAAGCCG TGGCTGCTGA
10751 ACCCCAGCCG GAAACTGACC CCACAAGGCC CTAGCGTTTG CAATGCACCA
10801 GGTCAATCATT GACCCAGGCG TGTCCACCA GGCCGCTGCC TCGCAACTCT
10851 TCGCAGGCTT CGCCGACCTG CTGCGCCAC TTCTTACAGC GGGTGAATC
10901 CGATCCGCAC ATGAGGCGGA AGGTTTCCAG CTTGAGCGGG TACGGCTCCC
10951 GGTGCGAGCT GAAATAGTCG AACATCCGTC GGGCCGTCGG CGACAGCTTG
11001 CGGTACTTCT CCCATATGAA TTTCGTGTAG TGGTCGCCAG CAAACAGCAC
11051 GACGATTTCC TCGTCGATCA GGACCTGGCA ACGGGACSTT TTCTTGCCAC
11101 GGTCCAGGAC GCGGAAGCGG TCGAGCAGCG ACACCGATTG CAGGTGCCCA
11151 ACGCGGTCGG ACGTGAAGCC CATCGCCGTC GCCTGTAGG CCGACAGGCA
11201 TTCCTCGGCC TTCTGTAAAT ACCGGCCATT GATCGACCAG CCCAGGTCCT
11251 GGCAAAGCTC GTAGAACGTG AAGGTGATCG GCTCGCCGAT AGGGGTGCCG
11301 TTGCGGTACT CCAACACCTG CTGCCACACC AGTTCGTCTC CGTCGGCCCG
11351 CAGCTCGACG CCGGTGTAGG TGATCTTACG GTCCCTGTTG ACGTGGAAAA
11401 TGACCTTGTG TTGCGAGGCC TCGCGCGGGA TTTCTTGTG GCGCGTGGTG
11451 AACAGGCGAG AGCGGGCCGT GTCGTTTGGC ATCGCTCGCA TCGTGTCCGG
11501 CCACGGCGCA ATATCGAACA AGGAAAGCTG CATTTCCCTG ATCTGCTGCT
11551 TCGTGTGTTT CAGCAACGCG GCCTGCTTGG CCTCGCTGAC CTGTTTGGCC
11601 AGGTCCCTCG CCGCGGTTTT TCGCTTCTTG GTCCGTATAG TTCCTCGCGT
11651 TCGATGTTTC ATCGACTTCG CCAAACCTGC CGCCTCTGT TCGAGACGAC
11701 GCGAACGCTC CACGGCGGCC GATGGCGCGG GCAGGGCAGG GGGAGCCAGT
11751 TGCACGCTGT CGCGCTCGAT CTTGGCCGTA GCTTGCTGGA CCATCGAGCC
11801 GACGGACTGG AAGGTTTCGC GGGGCGCAG CATGACGGTG CGGCTTGGCA
11851 TGGTTTCGGC ATCTTCGGCG GAAAACCCCG CGTCGATCAG TTCTTGCCCTG

11901 TATGCCCTTC GGTCAAACGT CCGATTCAAT CACCCTCCTT GCGGGATTGC
11951 CCCGACTCAC GCCGGGGCAA TGTGCCCTTA TTCTTGATTG GACCAGCCTG
12001 GTGCTTGTGT GTCCAGATAA TCACCTTAT CGGCAATGAA GTCCGTCCCG
12051 TAGACCGTCT GGCCGTCTCT CTGTACTTGT GTATTCCGAA TCTTGCCCTG
12101 CACGAATACC AGCGACCCCT TGCCAAATA CTTGCCGTGG GCCTCGCCCT
12151 GAGAGCCAAA ACACTTGATG CGGAAAGAAT CGGTGCGCTC CTGCTTGTGG
12201 CCGTCTGTGG CCGCGCCAAC CTTTGCATC CGCAAGCGCG CGGTCCCAT
12251 CTTACCGCTG GAACAGTACG TCGAGGCGGG CATCATGACC CGCGAGCAAT
12301 ACGAGGTCAT TAAAAGCGCC GTGATTGATG ATATAGCGGC CCGGCTGCTC
12351 CTGGTTCTCG CGCACCGAAA TGGGTGACTT CACCCCGCGC TCTTTGATCG
12401 TGGCACCGAT TTCCCGGATG CTCTCCGGGG AAAAGCCGGG GTTGTCCGGC
12451 GTCCCGCGGT GATGCGGATC TTCTGCTGAT AGGTCCAGGT CCAGCTCGAT
12501 AGGGCCGGAA CCGCCCTGAG ACGCCGAGG AGCGTCCAGG AGGCTCGACA
12551 GGTGCGCGAT GCTATCCAAC CCCAGGCGCG ACGGCTGCGC GCGGCTGCG
12601 GCTTCCGTAG CCGCGCGAGC GGTGTTTTTC TTGGTGGTCT TGGCTGAGC
12651 CGCAGTCAAT GGGAAATCTC CATCTTCTGT AACACGTAAT CAGCCAGGGC
12701 GCGAACCTCT TTCGATGCCT TCGCGCGCGC CGTTTTCTTG ATCTTCCAGA
12751 CCGGCACACC GGATGCGAGG GCATCGCGGA TGCTGTCTGG CAGGCCAACG
12801 GTGGCCGGAA TCATCATCTT GGGGTACCGG GCCAGCAGCT CGGCTTGGTG
12851 GCGCGCGTGG CCGCGGATCC GCGCATCGAC CTGCTGCGG ACCATGCCAA
12901 GGAATTGCGAG CTTGGCGTTC TTCTGGCGCA CGTTCGCAAT GGTCTGACC
12951 ATCTTCTTGA TGCCCTGGAT GCTGTACGCC TCAAGCTCGA TGGGGACAG
13001 CACATAGTCG GCCCGAAGA GGGCGCGCCG CAGGCCGAGC CCAAGGGTGG
13051 GGGCCGTGTC GATCAGGCAC ACCTCGAAGC CTTGGTTCGC CAGGGCCCTG
13101 ATGTTCCGCC CGAACAGCTC GCGGGCGTGG TCCAGCGACA GCGTTCGGC
13151 GTTCGCCAGT ACCGGGTGGG ACTCGATGAG GCGAGGGCGC GCGGCTGGC
13201 CGTCCCGCGG TCGGGGTGGG GTTTCGGTCC AGCCCGCGC AGGGACAGCG
13251 CCGAACAGCT TGCTTGCATG CAGGCCGGTA GCAAAGTCTT TGAGCGTGTA
13301 GGACGCATTT CCCTGGGGGT CCAGGTCGAT CACGGCAACC CGCAAGCCGC
13351 GCTCGAAAAA GTCGAAGGCA AGATGCACAA GGGTCGAAGT CTTGCCGAGC
13401 CCGCCTTCTT GGTGCGCGT GACCAAAGTT TTCAATCGTTT GGTTCCTGT
13451 TTTTCTTGGG CGTCCGCTTC CCACTTCCGG ACGATGTACG CCTGATGTTT
13501 CGCAGAACCC CGCGTTACCC GCGCGTACCC CTCGGCAAG TTCTTCTCCT
13551 CGAACCGCGC CCACACGCGA TGCACCGCTT GCGACACTGC GCCCTGGTTC

13601 AGTCCAGCG ACCTTGCAGG CGTCCGCTGT GGCTTCCCAT CACTAAGAC
13651 GCCCCGCGCT ATCTCGATGG TCTGCTGCC CACTTCCAGC CCCTGGATCG
13701 CCTCCTGGAA CTGGCTTTCG GTAAGCCGTT TCTTTCATGA TAACCCCAT
13751 AATTTGCTCC GCGCCTTGGT TGAACATAGC GGTGACAGCC GCCAGCAGT
13801 GAGAGAAGTT TAGCTAAACA TTCTTCGCAC GTCAACACCT TTAGCCGCTA
13851 AAATCTGCTC TTGGCGTAA AAAACAAAAG CCCGGAACCC GGGCTTCTGT
13901 CTCTTGGCCG TTATGGCTCT GCACCCGGCT CCATCACCAA CAGGTCCGGC
13951 ACGCGCTTCA CTCGTTGCGG GATCGACACT GCCAGCCCAA CAAAGCCGGT
14001 TGCCCGCGCC GCCAGGATCG CGCCGATGAT GCCGGCCACA CCGGCCATCG
14051 CCCACAGGTT CGCGCCCTTC CGGTTCCTAT CCTGCTGGTA CTGCTTCGCA
14101 ATGCTGGACC TCGGCTCACC ATAGGCTGAC CGCTCGATGG CGTATGCCG
14151 TTCTCCCTTT GCGTAAAAC CCAGCGCCGC AGGCGGCAAT GCCATGCTG
14201 CCGCGCCTTT CCCGACCACG ACGCGCGCAC CAGGCTTGGC GTCCAGACCT
14251 TCGGCCACGG CGAGCTGCGC AAGGACATAA TCAGCCCGCC ACTTGGCTCC
14301 ACGCGCCTCG ATCAGCTCTT GCACCTCGCG GAAATCCTTG GCCTCCACGG
14351 CCGCATGAA TCGCGCACGG GCGAAGGCT CCGCAGGGCC G

-continued

Broad-host-range expression vector for ligation-independent cloning

featuring C-terminal 10 ± His tag

pRKLIC2HT10Dpuf.seq Length: 14377

```

1 CCACCCAGGC CGCCGCCCTC ACTGCCCGGC ACCTGGTTCG TGAATGTGCA
51 TGCCAGCACC TGCGGCACGT CAATGCTTCC GGGCGTCCGC CTCGGGCTGA
101 TCGCCATFCC CGTTACTGCC CGGATCCCGG CAATGGCAAG GACTGCCAGC
151 GCTGCCATTT TTGGGGTGAG GCCGTTCCGG GCCGAGGGGC GCAGCCCCTG
201 GGGGGATGGG AGGCCCGCGT TAGCGGGCCG GGAGGGTTCC AGAAGGGGGG
251 GCACCCCCCTC TCGCGCTGCC CGGTCACGCG CACAGGGCGC AGCCCTGGTT
301 AAAAACAAGG TTTATAAATA TTGGTTTAAA AGCAGGTTAA AAGACAGGTT
351 AGCGGTGGCC GAAAACCGGG CGGAAACCCCT TGC AATGTT GGATTTTCCTG
401 CCTGTGGACA GCCCCTCAA TGTCAATAGG TGCGCCCTC ATCTGTCAGC
451 ACTCTGCCCC TCAAGTGTC AAGGATCGCG CCCTCATGTG TCAGTAGTCCG
501 CGCCCTCAA GTGTCAATAC CGCAGGGCAC TTATCCCCAG GCTTGTCCAC
551 ATCATCTGTG GGA AACTCGC GTAAAATCAG GCGTTTTCGC CGATTTGCCA
601 GGCTGGCCAG CTCACCTCG CGGCCGAAA TCGAGCCTGC CCCTCATCTG
651 TCAACGCCGC GCCGGGTGAG TCGGCCCTC AAGTGTCAAC FTCCGCCCTC
701 CATCTGTCTG TGAGGGCCAA GTTTTCCCGG AGGTATCCAC AACGCCGGCG
751 GCCCGGTGT CTGCGACAGC GCTTCGACGG CTTTTCTGGC GCGTTTGCAG
801 GGCCATAGAC GGCCGCCAGC CCAGCGCGCA GGGCAACCAG CCGCGTAGC
851 GTCGAAAGG CGCTTCTCCG CTTCCTCGCT CACTGACTCG CTGCGCTCGG
901 TCGTTCGGCT GCGCGAGCG GTATCAGCTC ACTCAAAGGC GGTAATACGG
951 TTATCCACAG AATCAGGGGA TAACGCAGGA AAGAACATGT GAGCAAAGG
1001 CCAGCAAAG GCCAGGAACC GTAAAAGGC CGCGTTGCTG GCGTTTTTCC
1051 ATAGGCTCCG CCCCCTGAC GAGCATCACA AAAATCGACG CTC AAGTCCAG
1101 AGGTGGCGAA ACCCGACAGG ACTATAAAGA TACGAGCGCT TTCCCCCTGG
1151 AAGTTCCTTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC
1201 TGTCGCCCTT TCTCCTTCG GGAAGCGTGG CGCCATTTCG CATT CAGGCT
1251 GCGCACTGT TGGGAAGGGC GATCGGTGCG GCCTCTTCG CTATTACGCC
1301 AGCTGGCGAA AGGGGATGT GCTGCAAGGC GATTAAGTTG GGTACGCCA
1351 GGGTTTTTCC AGTCACGAGC TTGTA A AAGC ACGGCCAGTG AATTCGGCCG
1401 CCGGCTGGCC GAGGTGCTGG GCAAGCCCTA CCTCCAGGCC CCCATCGGGG
1451 TCGAGAGCAC GACCGCCTTC CTGCGCCGCC TGGGCAGAT TCTGGCCCTC
1501 GATCCGGAGC CCTTCATCGA GCGCGAGAAG CACTCGACGC TGAAGCCCGT
1551 GTGGATCTG TGGCGGAGT TCACGCAGGA CTTCCTCGG ACGGCCAATF
1601 TCGGAATCGT GGC GACCGAA ACTTATGCAA GAGGCATCCG AAAC TATCTC
1651 GAAGGCATC TCGGCTGCC CTGCGCCTC GCCTGGCCCG CAAGAGGGGC

1701 TCGAAGACCG ACAACGAAGC GGTGCGCGGA CTGATCCGCC AGCACCCTCC
1751 GCTCGTGTCT ATGGGTTCGA TCAACGAGAA GATTTACCTT GCGGAAGTGA
1801 AAGCCGGTCA CGGCCCGGCA ACCCTCTTTC ATCGTGCCTT CTTTCCCGGG
1851 TGCGGCGATC CGGCGCGCTA CCGGAACGCC CGTTATGGGA TATGCGAGTG
1901 CTACGTGGTT ACTCGAGGAA GTTTGCAACG CCCTGTTTCA GCGCCTGTTC
1951 CACATTTCTG CCTTCGGGAC GGAGATGGAC AGCGCCGCGC CCACACCGAC
2001 GACACTGCGC CGCGACTTCC CGTGGGATGC CGATCGCAA CGGGCCCTGG
2051 ACCGATCGT AGAGGAGCAT CCGTTCCTCA CCGGATCAG CGCCGCGCTG
2101 GCCTTGGCGG ACGCCGCGCA GAAGGCTTGC CCTCGATGCC GGTGCCGAGA
2151 GGGTCGTGAG AGAGACTGTC GAAGCCCTGA CGTGGGCCGG GCTTCGGCGA
2201 GAGGAAGGGA GAGAACCAAT GAGCGATCAT GCCGTCAACA CGCCGTCCA
2251 TGCCGCCAGG GCCACGCGGC ACCGAGCACC TCGTGGCAG TTCTACGTTT
2301 ACTTCGCGCT CATCTGCTG GCGCCTTCC CGGTGGCCTT CGTGAGCTGG
2351 ATCGTCTCGA CGATCCGCCA CCGCAGGCTT CCAAGCGCGG GCCCTTCGC
2401 GTCCGCCCTG TTCGATGCCA AGGCGATCAC GCCGTGATT TTCCGCGCCT
2451 GACCGCAGGT CAGGTTGCGA CACGCCATTC GTCGTCTCCC CAAGGGGCGG
2501 CGGATTAATC GGGAGGGCAT GGTGCCTTAC CGTAACCCAC GCCACAGTA
2551 GCGAGGAGGA ACACGTGTCG TCCGTTGGIT XXXXXXXXXXXXXXXXXXXX
2601 XXXXXXXXXXXX AATAGGCCGG CCTCCGTCG CGGGCGCAC CCACGCCGC
2651 atCGATTCCA AGGTT CAGC ATTGAGACGG CTCCGCTTCG CGCGAAGCG
2701 CGGGTTGGGC CGACTGCAAG CCGAGAGGGA AGCATGGCAC TGCTCAGCTT
2751 CGAGCGAAAA TATCGCTGTC CGGGGGGCAC GCTGGTCCGC GGAACCTGT
2801 TCGACTTCTG GGTCCGGCCCT TTCTATGTCG GCTTCTTCGG GGTGCGACG
2851 TTTTCTTCFG CGGCCCTGGG TATCATCTGT ATTGCCTGGA GTGCCGTACT
2901 CCAGGGTACC TGGAAACCCC AACTCATCTC TGTTCTACCG CCGGCCCTFG
2951 AATATGGCCT GGGAGGTGCA CCCCTCGCAA AAGGCGGGCT GTGGCAGATC
3001 ATCACGATCT GCGCCACTGG TGCTTCTGTC AGCTGGGGCG TGCGCGAAGT
3051 CGAAATCTGC CGTAAGCTGG GCATCGGGTA CCACATCCCG TTCGCTTCG
3101 CGTTCGCCAT CCTGGCCTAC CTGACGCTGG TGCTGTTCCG CCCGTTGATG
3151 ATGGGCGCCT GGGGCTATGC CTTCCTTAC GGGATCTGGA GCACCTCGA
3201 CTGGGTGTCG AACACGGGCT ACACCTACGG CAACTTCCAC TACAACCTG
3251 CCCACATGAT CGCCATCTCG TTCTTCTTCA CGAACCGCTG GGCTCTGGCG
3301 CTGACGGCGC CCCTTGTGCT CTCCGCGGCC AACCCGAGA AGGGCAAGGA
3351 AATGCGGACG CCGGATCAGG AGGATACGTT CTTCGCGCAT CTGGTCCGGT

```

-continued

3401 ACTCGATCGG GACGCTCGGC ATCCACCGCC TCGGCCTGCT GCTCTCGCTG
3451 AGCGCCGTCT TCTTCAGCGC CCTCTGCATG ATCATTACCG GCACCATCTG
3501 GTTCGATCAG TGGGTGCGACT GGTGGCAATG GTGGGTGAAG CTGCCGTGGT
3551 GGGCGAACAT CCCGGGAGGC ATCAATGGCT GAGTATCAGA ACATCTTCTC
3601 CCAGGTCCAG GTCCGCGGAC CGCCCGACCT GGGGATGACC GAAGACGTCA
3651 ACCTGGCCAA CCGTTGCGGC GTCCGTCCCT TCTCGACCCT GCTCGGCTGG
3701 TTCGGCAACG CCCAGCTCGG CCCGATCTAT CTCGGCTCGC TCGGCGTCTC
3751 GTCCTCCTTC TCGGGCTCGA TGTGGTCTCT CACCATCGGG ATCTGGTCTC
3801 GGTATCAGCG GGGCTGGAAC CCGGCCGTCT TCCTGCGCGA CCTGTTCTTC
3851 TTCTCGCTCG AGCCGCCGGC ACCGAATAC GGTCTGTCTT TCGCGGCTCC
3901 GCTGAAGGAA GGGGGCTGT GGTGATCGC GTCTGTCTTC ATGTTCTGCTG
3951 CCGTCTGGTC CTGGTGGGGC CGCACCTATC TCCGCGCTCA GGCGCTGGGC
4001 ATGGCAAGC ACACGCCTG GCGTTCCTC TCGGCCATCT GGCTGTGGAT
4051 GGTCTGGGC TTCATCCGTC CGATCCTCAT GGGTCTCTGG TCGGAAGCGG
4101 TTCCCTACGG CATCTTCTCG CACCTCGACT GGACGAACAA CTCTCGCTC
4151 GTCCACGGCA ACCTGTCTTA CAACCCCTTC CACGGTCTCT CGATCGCCTT
4201 CCTCTACGGG TCGGCCCTGC TCTTCGCGAT GCACGGTGGC ACCATCCTCG
4251 CCGTCTCCCG CTTGCGGGGC GAGCGCGAGC TGGAGCAGAT CGCCGACCGC
4301 GGGACGGCAG CGGAGCGGGC CGCCCTCTTC TGGCGCTGGA CCATGGTFTT
4351 CAACGCCACG ATGGAAGGCA TCCACCGCTG GGCCATCTGG ATGGCGGTCC
4401 TCGTGACCCCT CACCGCGGGC ATCGGgATCC TGCTCTCGGG CACGGTCTGG
4451 GACAACTGGT ACGTCTGGGG CCAGAACCAC GGCATGGCGC CGCTGAACTG
4501 AGGAGCGATC ACAATGGCTG ACAAGACCAT CTTC AACGAT CACCTCAACA
4551 CCAATCCGAA GACCAACCTT CGCCTCTGGG TCGCTTTCCA GATGATGAAG
4601 GGTGCGGGCT GGGCTGGCGG CGTGTCTTTC GGGACGCTCC TTCTCATCGG
4651 GTTCTTCCGG GTGCTCGGGC GGATGCTTCC GATCCAGGAG AACAGGCTC
4701 CCGCGCCGAA CATCACCGGC GCTCTGGAGA CCGGGATCGA GCTGATCAAG
4751 CATCTCGTCT GAGACAAGTC TCGGGCAGG GCGCGCGGAG GCCGCCCGCT
4801 CCTCCAAGTC CGGGCCATAT CGCCGGCCCG GGTCCGGGGC GACACCACAG
4851 CCCGGTTCCT TTCTGTGTTG CGACAGGGAC CTGGTGCCGT GTGGAAGACC
4901 GCACGGCAC CTTTGTGACAT TCACGGGAGG CTCTGATGAC CAATCCCACC
4951 CCGCGACCCG AAACCCCGCT TTTGGATCGC GTCTGTCTGC CGGCCGACAT
5001 GAAGCGCGTG AGTGACCGCG AACTGGAGCG GCTGGCCGAG GAAGTGCCTT
5051 CCGAGGTCAG TGATAGGGGT AGTTTCTTAT TTTAGGCAGT TTATATGAAA

5101 TTAAGACATG CAGATGTCAC AGTGGATATT GAACTGGTCT CGAAAGCTCA
5151 ATATCCCCCA AAGCACAAAG ACAAACTTCG ACATCATGCA GAAGCGTTC
5201 CCGAAccgcy tcttcgacgt gggcatcgcc gagcagcatg ccgtgacctt
5251 cgggcecgcc ctcgCcgGGG cgggGatgaa gcccttctgc gcgatctatt
5301 cctcgctcct gcaaccgggt tacgaccaga tcgcccatga cgtggcgctg
5351 cagaaccttc ccgtccgctt cgtgatcgac cgggccccgc tcgtgggggc
5401 cgatggcgcy acccatcgcy gggccTFCGA CGTTGGCTTC ATCACTTCGC
5451 TGCCCAACAT GACCCTGATG GCCCGGCGCG ACGAGGCCGA GCTCATCCAC
5501 ATGATCgcCA CCGCCGTGGC CTTCGACGAG GGCCCATCG CCTTccGCTT
5551 CCCCGGGGGC GAGGGGGTGG GCCTCGAGAT GCCCGAGCG GGGACGGTGC
5601 TGGAGCCCCG CCGGGGCCGC GTGGTGC CGG AAGGGACgga tgtcgcgac
5651 ctctccttcG CCGCGcATCT GCACGAGgcc TTGcAGGCGg cGAAACTTCT
5701 CGAGGccGAG GGGGTGAGCG TGACCGTGGC CGACgccccG TtctCgCgCC
5751 CGCTCgAcAC GGGGCTCATC GACCAGCTCG TgcGCATCA CGCGCGCTG
5801 GTAACGGTGG AGCAGGGGGC cATGGGGCGC TTCGGCGCCC ATGTCTATGCA
5851 CTATCTCGCC AATTCGGCG GCTTCGACGG GGGCCTCGCG CTCCGGGTCA
5901 TGACGCTGCC CGACCGCTTC ATCGAGCAGG CGAGCCCGA GgACATGTAT
5951 GCCGATGCGG GGCTGCGGGC CGAGGATATC AAGCTTGGCG TAATCATGGT
6001 CATAGCTGTT TCCTGTGTGA AATTGTATAT CGCTCACAA TCCACACAAC
6051 ATACGAGCCC GAAGCATAAA GTGTAAGCC TGGGGTGCCT AATGACTGAG
6101 CTAACTCACA TTAATTGCGT TCGCTCACT GCCCGCTTTC CAGTCGGGAA
6151 ACCTGTCTGT CCACTGTCAT TAATGAATCG GCCAACCGCG GGGGAGAGGC
6201 GGTTTGTGTA TTGGGCGCTC GGTCTTGCCT TGCTCGTCTG TGATGTACTT
6251 CACCAGCTCC GCGAAGTCCG TCTTCTGTAT GGAGCGCATG GGGACGTGCT
6301 TGGCAATCAC GCGCACCCCC CGGCCGTTTT AGCGGCTAAA AAAGTCATGG
6351 CTCTGCCCTC GGGCGGACCA CGCCCATCAT GACCTTGCCA AGCTCGTCTC
6401 GCTTCTCTTC GATCTTCGCC AGCAGGGCGA GGATCGTGGC ATCACCGAAC
6451 CGCGCCGTGC CCGGGTCTGC GGTGAGCCAG AGTTTCAGCA GGCCGCCAG
6501 CGCGCCAGG TCGCCATGTA TCGGGCCAG CTCGCGGACG TGCTCATAGT
6551 CCACGACGCC CGTGATTTG TAGCCCTGGC CGACGGCCAG CAGGTAGGCC
6601 GACAGGCTCA TGCCGGCCGA CGCCGCTTTT TCCTCAATCG CTCTTCGFTC
6651 GTCTGGAAGG CAGTACACCT TGATAGGTGG GCTGCCCTTC CTGGTTGGCT
6701 TGGTTTCATC AGCCATCCGC TTGCCCTCAT CTGTTACGCC GGCGGTAGCC
6751 GGCCAGCCTC GCAGAGCAGG ATTCCCGTTG AGCACGCCA GGTGCGAATA

-continued

6801 AGGGACAGTG AAGAAGGAAC ACCCGCTCGC GGGTGGGCCT ACTTCACCTA
6851 TCCTGCCCGG CTGACGCCGT TGGATACACC AAGGAAAGTC TACACGAACC
6901 CTTTGGCAA AATCCTGTATA TCGTGGGAAA AAGGATGGAT ATACCGAAAA
6951 AATCGCTATA ATGACCCCGA AGCAGGGTTA TGCAGCGGAA AAGCGCCACG
7001 CTTCCCGAAG GGAGAAAGGC GGACAGGTAT CCGGTAAGCG GCAGGGTCGG
7051 AACAGGAGAG CGCACGAGGG AGCTTCCAGG GGGAAACGCC TGGTATCTTT
7101 ATAGTCCGTG CGGGTGTCGC CACCTCTGAC TTGAGCGTCG ATTTTTGTGA
7151 TGCTCGTCAG GGGGGCGGAG CCTATGGAAA AACGCCAGCA ACGCGCCCTT
7201 TTTACGGTTC CTGGCCTTTT GCTGGCCTTT TGCTCACATG TTCTTTCCFG
7251 CGTTATCCCC TGATTCTGTG GATAACCGTA TTACCGCCTT TGAGTGAGCT
7301 GATACCGCTC GCCCGAGCCG AACGACCGAG CGCAGCGAGT CAGTGAGCGA
7351 GGAAGCGGAA GAGCGCCAGA AGGCCGCCAG AGAGGCCGAG CGCGGCCGTG
7401 AGGCTTGGAG GCTAGGGCAG GGCATGAAAA AGCCCCAGC GGGCTGTCTAC
7451 GGGCGTCTGA CGCGGTGGAA AGGGGGAGGG GATGTTGTCT ACATGGCTCT
7501 GCTGTAGTGA GTGGGTTGGC CTCCGGCAGC GGTCTGTATC AATCGTCACC
7551 CTTTCTCGGT CCTTCAACGT TCCTGACAAC GAGCCTCCTT TTCGCCAATC
7601 CATCGACAAT CACCGCGAGT CCCTGCTCGA ACGCTGCGTC CGGACCGGCT
7651 TCGTCAAGG CGTCTATCGC GGCCCGCAAC AGCGGCGAGA GCGGAGCCGT
7701 TTCAACGGTG CCGCCGCGCT CGCCGGCATC GCTGTCCCGC GCCTGCTCCT
7751 CAAGCAGCGC CCAACACAGT AAGTAGCTGA TTGTCATCAG GCATTTGACG
7801 GCGTCCCGCG CCGAAAAACC CGCCTCGCAG AGGAAGCGAA GCTGCGCGTC
7851 GGGCGTTTCC ATCTGCGGTG CGCCCGTCCG CGTGCCGGCA TGGATGCGCG
7901 CGCCATCGCG GTAGGCGAGC AGCGCCTGCC TGAAGCTGCG GGCATTTCCG
7951 ATCAGAAAT AGCGCCAGTC GTCGTCCGGT CTCGGCACCG AATGCGTATG
8001 ATTCCTCGCC AGCATGGCTT CGGCCAGTGC GTCGAGCAGC GCCCGCTTGT
8051 TCCTGAAGTG CCAATAAAGC CGCGGCTGCT GAACCCCAA CCGTTCCGCG
8101 AGTTTGGGTG TCGTCAAGC GTCTACGCCG ACCTCGTTCA ACAGGTCCAG
8151 GGGCGCACGG ATCACTGTAT TCGGCTGCAA CTTTGTCTAT CTGTACACTT
8201 TATCACTGAT AACATAATA TGTCACCAA CTTATCAGT ATAAAGAAATC
8251 CGCGGTTTCA ATCGGACAGC CGGAGGCTGG TCCGGAGGCC AGACATGAAA
8301 CCCAACATAC CCTGATCGT AATTCTGAGC ACTGTCCGCG TCGACGCTGT
8351 CGGCATCGGC CTGATTATGC CGGTGCTGCC GGGCCTCCTG CGCGATCTGG
8401 TTCACTCGAA CGACGTCACC GCCCACTATG GCATTCTGCT GGCGCTGTAT
8451 GCGTTGGTGT AATTTGCCCTG CGCACCTGTG CTGGGCGCGC TGTCCGATCG

8501 TTTCCGGCGG CGGCCAATCT TGCTCGTCTC GCTGGCCGGC GCCACTGTGC
8551 ACTACGCCAT CATGGCGACA GCGCCTTTCC TTTGGGTTC TATATCGGG
8601 CGGATCGTGG CCGGCATCAC CGGGGCGACT GGGGCGGTAG CCGGCGCTTA
8651 TATTGCGGAT ATCACTGATG GCGATGAGCG CGCGCGCAC TTCGGCTTCA
8701 TGAGCGCCTG TTTCCGGTTC GGGATGGTCG CCGGACCTGT GCTCGGTGGG
8751 CTGATGGCGG GTTCTCTCCC CCACGCTCCG TTCTTCGCGC GCGGAGCCTT
8801 GAACGGCCTC AATTCTCTGA CCGGCTGTTT CCTTTGCGCG GAGTGCACA
8851 AAGCGAACC CCGGCCGTTA CGCCGGGAGG CTCTCAACC GCTCGCTTCG
8901 TTCGGGTGGG CCGGGGCGAT GACCGTCTG GCGCCCTGA TGGCGGTCTT
8951 CTTTATCATG CAACTTGTGC GACAGGTGCC GCGCGCTTT TGGGTCAATT
9001 TCGGCGAGGA TCGCTTTCAC TGGGACGCGA CCACGATCG CATTTCGCTT
9051 GCGGCATTTG GCATTCGCA TTCACTCGCC CAGGCAATGA TCACCGGCC
9101 TGTAGCCGCC CGGCTCGGCG AAAGCGGGC ACTCATGCTC GGAATGATG
9151 CCGACGGCAC AGGCTACATC CTGCTTGCT TCGGACACG GGGATGGATG
9201 GCGTTCGCCA TCATGGTCTT GCTTGTCTCG GGTGGCATCG GAATCGCCG
9251 GCTGCAAGCA ATGTTGTCCA GGCAGGTGGA TGAGGAACGT CAGGGCAGC
9301 TGCAAGGCTC ACTGGCGGCG CTCACCAGCC TGACCTCGAT CGTCCGACCC
9351 CTCCTCTTCA CCGGATCTA TGCGGCTTCT ATAACAACGT GGAACGGGTG
9401 GGCATGGATT GCAGCGCTG CCTCTFACTT GCTCTGCTG CCGGCGCTGC
9451 GTCGCGGGCT TTGGAGCGGC GCAGGGCAAC GAGCCGATCG CTGATCGTGG
9501 AAACGATAGG CCTATGCCAT GCGGGTCAAG GCGACTTCCG GCAAGCTATA
9551 CGCGCCCTAG GAGTGCGGTT GGAACGTTGG CCCAGCCAGA TACTCCCGAT
9601 CACGAGCAGG ACGCCGATGA TTTGAAGCGC ACTCAGCGTC TGATCCAAGA
9651 ACAACATFCC TAGCAACACG GCGGTCCCGG GGCTGAGAAA GCCAGTAA
9701 GAAACAACCT TAGGTTGAG TCGCGAGATC CCCCAGAAC AAAGGAAGTA
9751 GGTTAAACCC GCTCCGATCA GGCCGAGCCA CGCCAGGCCG AGAACATTGG
9801 TTCTGTAGG CATCGGGATT GCGGATCAA AACTAAAGC TACTGGAACG
9851 AGCAGAAGTC CTCCGGCCG CAGTTGCCAG GCGGTAAGG TGAGCAGAGG
9901 CACGGGAGGT TGCCACTTGC GGTTCAGCAC GGTTCGCAAC GCCATGGAAA
9951 CGGCCCGCGC CAGGCCGCT GCGACGCCGA CAGGATCTAG CGTTCGCTT
10001 GGTGTCAACA CCAACAGCGC CACGCCCGCA GTTCCGCAA TAGCCCCAG
10051 GACCGCATC AATCGTATCG GGCTACTTAG CAGAGCGGCA GAGATGAACA
10101 CGACCATCAG CGGCTGCACA GCGCTTACCG TCGCCGCGAC CCGCCCGCGC
10151 AGGCGGTAGA CCGAAATAA CAACAAGCTC CAGAATAGCG AAATATTAAG

-continued

10201	TGAGCCGAGG	ATGAAGATGC	GCATCCACCA	GATTCCCGTT	GGAATCTGTC
10251	GGACGATCAT	CACGAGCAAT	AAACCCGCGG	GCAACGCCCG	CAGCAGCATA
10301	CCGGCGACCC	CTCGGCCTCG	CTGTTGCGGC	TCCACGAAAA	CGCCGGACAG
10351	ATGCGCCTTG	TGAGCGTCC	TGGGGCCGTC	CTCCTGTTTG	AAGACCCGACA
10401	GCCCAATGAT	CTCGCCGTCG	ATGTAGGCGC	CGAATGCCAC	GGCATCTCGC
10451	AACCGTTCAG	CGAACGCCTC	CATGGGCTTT	TTCTCCTCGT	GCTCGTAAAC
10501	GGACCCGAAC	ATCTCTGGAG	CTTCTTCAG	GGCCGACAAT	CGGATCTCGC
10551	GAAATCCTG	CACGTGCGCC	GCTCCAAGCC	GTGCAATCTG	AGCCTTAATC
10601	ACAATTGTC	ATTTAATCC	TCGTTFATC	GGCAGTTCGT	AGAGCGCGCC
10651	GTGCGTCCCG	AGCGATACTG	AGCGAAGCAA	GTGCGTCCG	CAGTGCCCGC
10701	TTGTTCCTGA	AATGCCAGTA	AAGCGCTGGC	TGCTGAACCC	CCAGCCGGAA
10751	CTGACCCAC	AAGGCCCTAG	CGTTTGAAT	GCACAGGTC	ATCATTGACC
10801	CAGCGTGTT	CCACCAGGCC	GCCTGCTCGC	AACTCTTCGC	AGGCTTCGCC
10851	GACCTGCTCG	CGCCACTTCT	TCACGCGGGT	GGAATCCGAT	CCGCACATGA
10901	GGCGGAAGT	TTCCAGCTTG	AGCGGTACG	GCTCCCGTGG	CGAGCTGAAA
10951	TAGTCCGAAT	TCCGTGCGGC	CGTCGGCGAC	AGCTTCCGGT	ACTTCTCCCA
11001	TATGAATTT	GTGTAGTGGT	CGCCAGCAAA	CAGCACGAGC	ATTTCTCTGT
11051	CGATCAGGAC	CTGGCAACGG	GACGTTTCT	TGCCACGGTC	CAGGACCGGG
11101	AAGCGGTGCA	CGACGACAC	CGATTCCAGG	TGCCAACGCG	GGTCGGACGT
11151	GAAGCCCATC	GCCGTGCGCT	GTAGGCGCGA	CAGGCATTCC	TCCGGCTTCG
11201	TGTAATACCG	GCCATTGATC	GACCAGCCCA	GGTCTGGCA	AAGCTCGTAG
11251	AACGTGAAGG	TGATCGGCTC	GCCGATAGGG	GTGCGCTTCG	CGTACTCCAA
11301	CACCTGCTGC	CACACCAGTT	CGTCATCGTC	GGCCCGCAGC	TCGACGCCGG
11351	TGTAGGTGAT	CTTCACGTCC	TTGTTGACGT	GGAAAATGAC	CTTGTTTTGC
11401	AGCGCCTCGC	GCGGGATTTT	CTTGTTCGCG	GTGGTGAAAC	GGGACAGAGC
11451	GGCCGTGTGC	TTTGGCATCG	CTCGCATCGT	GTCCGGCCAC	GGCGCAATAT
11501	CGAACAAAGG	AAGCTGCATT	TCCTTGATCT	GCTGCTTCGT	GTGTTTCAGC
11551	AACGCGCCCT	GCTTGGCCTC	GCTGACCTGT	TTTGCCAGGT	CCTCGCCGGC
11601	GGTTTTTCGC	TTCTTGGTCG	TCATAGTTCC	TCGCGTGTGC	ATGGTCATCG
11651	ACTTCCGCAA	ACCTGCCGCC	TCCTGTTCGA	GACGACGCGA	ACGCTCCACG
11701	GCGGCCGATG	GCGCGGGCAG	GGCAGGGGGA	GCCAGTTGCA	CGCTGTCCGC
11751	CTCGATCTTG	GCCGTAGCTT	GCTGGACCAT	CGAGCCGACG	GACTGGAAGG
11801	TTTCGCGGGG	CGCACGCATG	ACGGTGCGGC	TTGCGATGGT	TTCCGGCATCC
11851	TCGGCGGAAA	ACCCCGCGTC	GATCAGTTCT	TGCTGTATG	CCTTCCGGTC
11901	AAACGTCCGA	TTCATTACCC	CTCCTTGGCG	GATTGCCCGG	ACTCACGCCG
11951	GGGCAATGTG	CCCTTATTC	TGATTTGACC	CGCCTGGTGC	CTTGGTGTCC
12001	AGATAATCCA	CCTTATCGGC	AAATGAAGTCG	GTCCCGTAGA	CCGTCTGGCC
12051	GCTCTTCTCG	TACTTGGTAT	TCCGAACTTT	GCCCTGCACG	AATACCAGCG
12101	ACCCCTTGCC	CAAATACTTG	CCGTGGGCTT	CGGCCTGAGA	GCCAAAACAC
12151	TTGATGCGGA	AGAAGTCCGT	GCCTCCTGCG	TTGTCGCCGG	TCGTGGCCGC
12201	GCCAACCTTT	GCGATCCGCA	AGCGCGCGGT	CGCCATCTTC	ACGCTGGAAC
12251	AGTACGTGCA	GGCGGGCAGC	ATGACCCCGG	AGCAATACGA	GGTCATTAAC
12301	AGCGCCGTGA	TTGATGATAT	AGCGGCCCGG	CTGCTCCTGG	TCTTCGCGCA
12351	CCGAAATGGG	TGACTTCACC	CCGCGCTCTT	TGATCGTGGC	ACCGATTTC
12401	GCGATGCTCT	CCGGGGAAAA	GCCGGGGTTG	TCGGCCGTC	GCGGCTGATG
12451	CGGATCTTCG	TCGATCAGGT	CCAGGTCCAG	CTCGATAGGG	CCGGAACCGC
12501	CCTGAGACGC	CGCAGGAGCG	TCCAGGAGGC	TCGACAGGTC	GCCGATGCTA
12551	TCCAACCCCA	GGCCGGACCG	CTGCGCCGCG	CCTGCGGGCT	CCTGAGCCGG
12601	CGCAGCGGTG	TTTTCTTGG	TGGTCTTGGC	TTGAGCCGCA	GTTCATTTGGA
12651	AATCTCCATC	TTCTGTAACA	CGTAATCAGC	CAGGGCGCGA	ACCTCTTTCG
12701	ATGCTTGGC	CGCGCCGTT	TTCTTGATCT	TCCAGACCGG	CACACCGGAT
12751	GCGAGGGCAT	CGCCGATGCT	GCTGCGCAGG	CCAACGGTGG	CCGGAATCAT
12801	CATCTTGGGG	TACGCGGCCA	GCAGCTCGGC	TTGGTGGCGC	CGCTGGCGCG
12851	GATTCCCGCG	ATCGACCTTG	CTGGGCACCA	TGCCAAGGAA	TTGCAGCTTG
12901	GCTTCTTCT	GGCGCACGTT	CGCAATGGTC	GTGACCATCT	TCTTGATGCC
12951	CTGGATGCTG	TACGCCCTCAA	GCTCGATGGG	GGACAGCACA	TAGTCGGCCG
13001	CGAAGAGGGG	GGCCGCCAGG	CCGACGCCAA	GGGTCGGGGC	CGTGTGATC
13051	AGGCACACGT	CGAAGCCTTG	GTTCGCCAGG	GCCTTGATGT	TCGCCCCGAA
13101	CAGCTCGCGG	GCGTCGTCCA	GCGACAGCCG	TTCCGGCTTC	GCCAGTACCG
13151	GGTTGGACTC	GATGAGGGCG	AGCGCGCGCG	CCTGGCCGTC	GCCGGCTGCG
13201	GGTGGGTTT	CGGTCCAGCC	GCCGGCAGGG	ACAGCGCCGA	ACAGCTTGCT
13251	TGCATGCGAG	CCGTAGCAA	AGTCTTGAG	CGTGTAGGAC	GCATTGCCCT
13301	GGGGTCCAG	GTGATCACG	GCAACCCGCA	AGCCCGGCTC	GAAAAAGTCG
13351	AAGGCAAGAT	GCACAAGGGT	CGAAGTCTTG	CCGACGCCGC	CTTTCTGGTT
13401	GGCCGTGACC	AAAGTTTCA	TCGTTTGGTT	TCCTGTTTTT	TCTTGGCGTC
13451	CGCTTCCAC	TTCCGGACGA	TGTACGCCCTG	ATGTTCCGGC	AGAACCCCGC
13501	TTACCCGCGC	GTACCCCTCG	GGCAAGTTCT	TGTCTCGAA	GCGGCCAC
13551	ACGCGATGCA	CCGCTTGCGA	CACTGCGCCC	CTGGTACGTC	CCAGCGACGT
13601	TGCGAACGTC	GCCTGTGGCT	TCCCATCGAC	TAAGACGCC	CGCGTATCT
13651	CGATGGTCTG	CTGCCCCACT	TCCAGCCCTT	GGATCGCCTC	CTGGAATCTG
13701	CTTTCCGTTAA	GCCGTTTCTT	CATGGATAAC	ACCCATAAAT	TGCTCCGCGC
13751	CTTGGTTGAA	CATAGCGGGT	ACAGCCGCCA	GCACATGAGA	GAAGTTTAGC
13801	TAAACATTT	TCCGACGTCA	ACACCTTAG	CCGCTAAAAC	TCGCTCTTGG
13851	CGTAACAAAA	CAAAAGCCCG	GAAACCGGGC	TTTCTGCTCT	TGCCGCTTAT
13901	GGCTCTGCAC	CCGGCTCCAT	CACCAACAGG	TCGCGCACGC	GCTTCACTCG
13951	GTGCGGATC	GACACTGCCA	GCCCCAACAA	GCCGGTTGCC	GCCCGCCGCA
14001	GGATCGCGCC	GATGATCGCC	GCCACACCGG	CCATCGCCCA	CCAGTCCGCC
14051	GCCTTCCGGT	TCCATTCTCT	CTGGTACTGC	TTCCGCAATG	TGGACCTCGG
14101	CTCACCATAG	GCTGACCGCT	CGATGGCGTA	TGCCGCTTCT	CCCTTTGGCG
14151	TAAACCCAG	CGCCGACGGC	GGCATTTGCCA	TGCTGCCCGC	CGCTTTCCCG
14201	ACCACGACGC	GCGCACCAGG	CTTGGCGTCC	AGACCTTCGG	CCACGGCGAG
14251	CTGCGCAAGG	ACATAATCAG	CCGCCGACTT	GGCTCCACGC	GCCTCGATCA
14301	GCTCTTGAC	TCGCGCGAAA	TCCTTGGCCT	CCACGGCCGC	CATGAATCGC
14351	GCACCGCGCG	AAGGCTCCCG	AGGGCCG		

-continued

3401 TGGTCGGCTA CTCGATCGGG ACGCTCGGCA TCCACCGCCT CGGCCTGCTG
3451 CTCTCGCTGA GCGCCGTCTT CTTACGCGCC CTCTGCATGA TCATTACCGG
3501 CACCATCTGG TTCGATCAGT GGTTCGACTG GTGGCAATGG TGGGTGAAGC
3551 TGCCGTGGTG GGGCAACATC CCGGGAGGCA TCAATGGCTG AGTATCAGAA
3601 CATCTTCTCC CAGGTCCAGG TCCGCGGACC GGCCGACCTG GGGATGACCG
3651 AAGACGTCAA CCTGGCCAAAC CGTTCGGGGC TCGGTCCCTT CTCGACCCCTG
3701 CTCTGGCTGGT TCGGCAACGC CCAGCTCGGC CCGATCTATC TCGGCTCGCT
3751 CGGGCTCCTG TCCCTCTTCT CGGGCCTGAT GTGGTTCTTC ACCATCGGGA
3801 TCTGGTTCTG GTATCAGGCG GGCTGGAACC CGGCCGTCTT CCTGCGCGAC
3851 CTGTTCTTCT TCTCGCTCGA GCCGCCGGCA CCCGAATACG GTCTGTCCCT
3901 CGCGGCTCCG CTGAAGGAAG CGGGGCTGTG GCTGATCGCG TCGTTCTTCA
3951 TGTTCTGTCG GGTCTGGTCC TGGTGGGGCC GCACCTATCT CCGCGCTCAG
4001 CGCTGGGCA TGGCAAGCA CACCCCTGG GCGTTCCTCT CGGCCATCTG
4051 GCTGTGGATG GTGCTGGGCT TCATCCGTCC GATCCTCATG GGTCTCTGGT
4101 CGGAAGCGGT TCCCTACGGC ATCTTCTCGC ACCTCGACTG GACGAACAAC
4151 TTCTCGCTCG TCCACGGCAA CCTGTCTAC AACCCCTTCC ACGGTCTCTC
4201 GATCGCCTTC CTCTACGGGT CGGCCCTGCT CTTTCGCGATG CACGGTGCGA
4251 CCATCCTCGC GGTCTCCCGC TTCGGCGGGC AGCGCGAGCT GGAGCAGATC
4301 GCCGACCGCG GGACGGCAGC GGAGCGGGCC GCCCTCTTCT GGCGCTGGAC
4351 CATGGGTTTC AACGCCACGA TGAAGGCAT CCACCCGCTG GCCATCTGGA
4401 TGGCGGTCTT CGTGACCCTC ACCGGCGGCA TCGGGATCCT GCTCTCGGGC
4451 ACCGTCTGTG ACAACTGGTA CGTCTGGGGC CAGAACCACG GCATGGCCGC
4501 GCTGAACCTG GAGCGATCA CAATGGCTGA CAAGACCATC TTCAACGATC
4551 ACCCTAACAC CAATCCGAAG ACCAACCTTC GCCTCTGGGT GCCTTTCCAG
4601 ATGATGAAGG GTGCGGGCTG GGCTGGCGGC GTGTTCTTCG GGACGCTCCT
4651 TCTCATCGGG TTCTTCCGGG TGCTCGGGCC GATGCTTCCG ATCCAGGAGA
4701 ACCAGGCTTC GCGCGCGAAC ATCACCGGCG CTTCTGGAGC CGGGATCGAG
4751 CTGATCAAGC ATCTCGTCTG AGACAAGTCT CCGGGCAGGG CGGCGCGAGG
4801 CGCCCGGCTC CTCCAAGTCC GGGCCATATC GCCGGCCCGG GTCCGGGGCG
4851 ACACCACAGC CCGGTTCCCT TCCTGTGTGG GACAGGGACC TGGTCCCGTG
4901 TGAAGACCGC CACGGCACCC TTTTGACATT CACGGGAGGC TCTGATGACC
4951 AATCCACCC CGCGACCCGA AACCCCGCTT TTGGATCGCG TCTGTGCGCC
5001 GGCCGACATG AAGCGCTGA GTGACGCCGA ACTGGAGCGG CTGGCCGACG
5051 AAGTGCCTTC CGAGGTCAGT GATAGGGGTA GTTCTTATT TTAGCAGTT

5101 TATATGAAAT TAAGACATGC AGATGTCACA GTGGATATTG AACTGGTCTC
5151 GAAAGCTCAA TATCCCCAA AGCACAGCA CAAACTTCGA CATCATGCAG
5201 AAGCGTTTTC CGAAccgctg cttcgacgtg ggcacgccc agcagcatgc
5251 cgtgaccttc gcgccggccc tcgCcgGGGc cggGatgaa gccctctgcg
5301 cgatctattc ctctgttcctg caacgggggtt acgaccagat cgcccatgac
5351 gtggcgctgc agaaccttcc cgtccgcttc gtgatcgacc gggcggggct
5401 cgtggggggc gatggcgcca cccatcgggg ggcctTCGAC GTTGGCTTCA
5451 TCACTTCCCT GCCCAACATG ACCGTGATGG CCGCGGCCGA CGAGGCCGAG
5501 CTCATCCACA TGATCgcCAC CGCCGTGGCC TTCGACGAGG GCCCATCGC
5551 CTTCGCTTC CCGCGGGGCG AGGGGTGGG CGTCGAGATG CCGGAGCGCG
5601 GGACGGTGT GAGCCCGGc CGGGCCGCG TGGTGC CGA AGGGACggat
5651 gtcgagatcc tctccttCG CGCGcATCTG CAcGAGgccc TGcAGGCGgc
5701 GAAACTTCTC GAGGccGAGG GGGTGA GCGT GACCGTGGCC GACgcccgCT
5751 TctCgCgCCC GCTCgAcACG GGGCTCATCG ACCAGTCTGT gcGCCATCAC
5801 GCGCGCTGG TAACGGTGGG GCATGGGGc ATGGGCGGCT TCGGCGCCCA
5851 TGTCATGCAC TATCTCGCCA ATTCCGGCGG CTTTCGACGG GGCCTCGCGC
5901 TCCGGGTCAT GACGCTGCC GACCGCTTCA TCGAGCAGG GAGCCCCGAG
5951 gACATGTATG CCGATGCGGG GCTGCGGGCC GAGGATATCA AGCTTGGCGT
6001 AATCATGGTC ATAGCTGTTT CCTGTGTGAA ATGTATTATC GCTCACAAAT
6051 CCACACAACA TACGAGCCGG AACATAAAG TGTAAAGCCT GGGGTGCCA
6101 ATGAGTGAGC TAACCTACAT TAATTGCGTT GCGCTCACTG CCGCTTTCC
6151 AGTCGGGAAA CCTTCTGTC CAGCTGCATT AATGAATCGG CCAACCGCG
6201 GGGAGAGGCG GTTTGCGTAT TGGGCGCTCG GTCTTGCCCT GCTCGTGGT
6251 GATGACTTCT ACCAGCTCCG CGAAGTCGCT CTCTTGTATG GAGCGCATGG
6301 GGACGTGCTT GGCAATCACG CGCACCCCC GGCCGTTTTA GCGGCTAAAA
6351 AAGTCATGGC TCTGCCCTCG GCGGACCAC GCCCATCATG ACCTTGCCAA
6401 GCTCGTCTCG CTCTCTTTCG ATCTTCGCCA GCAGGGCGAG GATCGTGGCA
6451 TCACCGAACC GCGCCGTGCG CGGGTCGTCG GTGAGCCAGA GTTTCAGCAG
6501 GCCCGCCAGG CGGCCAGGT CGCCATTGAT GCGGGCCAGC TCGCGGACGT
6551 GCTCATAGTC CACGACGCC GTGATTTGT AGCCCTGGCC GACGGCCAGC
6601 AGGTAGGCCG ACAGGCTCAT GCCGGCGGCC GCCGCTTTT CCTCAATCGC
6651 TCTTCGTTTC TCTGGAAGGC AGTACACCTT GATAGTGGG CTGCCCTTCC
6701 TGGTTGGCTT GGTTCATCA GCCATCCGCT TGCCCTCATC TGTTCAGCCG
6751 CGGGTAGCCG GCCAGCCTCG CAGAGCAGGA TTCCCGTTGA GCACCGCCAG

-continued

6801 GTGCGAATAA GGGACAGTGA AGAAGGAACA CCCGCTCGCG GGTGGGCCATA
 6851 CTTACCTAT CCTGCCCGGC TGACGCCGTT GGATACACCA AGGAAAGTCT
 6901 ACACGAACCC TTTGGCAAAA TCCTGTATAT CGTGCAGAAA AGGATGGATA
 6951 TACCGAAAAA ATCGCTATAA TGACCCCGAA GCAGGGTTAT GCAGCGGAAA
 7001 AGCGCCACGC TTCCCGAAGG GAGAAAGCGG GACAGGTATC CGGTAAAGCGG
 7051 CAGGGTCGGA ACAGGAGAGC GCACGAGGGA GCTTCCAGGG GGAACGCCT
 7101 GGTATCTTTA TAGTCTGTGC GGGTTTCGCC ACCTCTGACT TGAGCGTCGA
 7151 TTTTGTGTAT GCTCGTCAGG GGGCGGAGC CTATGGAAAA ACGCCAGCAA
 7201 CGCGGCCTTT TTACGGTTCG TGCCCTTTTG CTGGCCTTTT GCTCACATGT
 7251 TCTTCTCTGC GTTATCCCTT GATTCGTGGG ATAACCGTAT TACCGCCTTT
 7301 GAGTGAGCTG ATACCGCTCG CCGCAGCCGA ACGACCGAGC GCAGCGAGTC
 7351 AGTGAGCGAG GAAGCGGAAG AGCGCCAGAA GGCCCGCAGA GAGCGCGAGC
 7401 CGGCGCTGTA GGCCTGGAGC CTAGGGCAGG GCATGAAAA GCGCGTAGCG
 7451 GGCTGCTACG GCGCTCTGAC CCGGTGGAAA GGGGAGGGG ATGTTGTCTA
 7501 CATGGCTCTG CTGTAGTGAG TGGGTTGCCG TCCGGCAGCG GTCCTGATCA
 7551 ATCGTCACCC TTTCTCGGTC CTTCAACGTT CDTGACAACG AGCCTCCTTT
 7601 CTGCCAATCC ATCGACAATC ACCCGAGTCC CDTGCTCGAA CGCTGCGTCC
 7651 GGACCGGCTT CGTCAAGGCG GTCTATCGCG GCCCGCAACA GCGCGGAGAG
 7701 CGGAGCCTGT TCAACGGTGC CGCCCGCTC CCGGCATCG CTGTCCCGGG
 7751 CCTGCTCCTC AAGCACGGCC CCAACAGTGA AGTAGCTGAT TGTCTCAGC
 7801 GCATTGACGG CGTCCCGGCG CGAAAAACCC GCCTCGCAGA GGAAGCGAAG
 7851 CTGGCGCTCG GCCGTTTCCA TCTGCGGTGC GCCCGTCCG GTGCGCGCAT
 7901 GGATGCGCGC GCCATCGCGG TAGGCGAGCA GCGCCTGCCT GAAGTGCAGG
 7951 GCATTCCCTT TCAGAAATGA CGCCAGTCC TCGTGGCTC TCGGCACCGA
 8001 ATGCGTATGA TTTCCGCCA GCATGGCTTC GGCCAGTGGC TCGAGCAGCG
 8051 CCCGCTTGT TCTGAAGTGC CAGTAAAGCG CCGGCTGCTG AACCCCAAC
 8101 CGTTCCGCCA GTTTGCGTGT CGTCAGACCG TCTACGCCGA CCTCGTTCAA
 8151 CAGGTCCAGG GCGGCACCGA TCACTGTATT CCGCTGCAAC TTTGTCAATG
 8201 TTGACACTTT ATCACTGATA AACATAATAT GTCCACCAAC TTATCAGTGA
 8251 TAAAGATACC GCGCGTTCAA TCGGACACGC GGAGGCTGTT CCGGAGCCCA
 8301 GACATGA AACCAACATACC CCTGATCGTA ATTCTGAGCA CTGTCCGCT
 8351 CGACGCTGTC GGCATCGGCC TGATTATGCC GGTGCTGCCG GGCCTCCTGC
 8401 CGCATCTGTT TCACTCGAAC GACGTCACCG CCCACTATGG CATTTCTGCT
 8451 GCGCTGTATG CGTTGGTGCA ATTTGCTGTC GCACCTGTGC TGGGCGCGCT

 8501 GTCGGATCGT TTCGGGCGGC GGCCAATCTT GCTCGTCTCG CTGGCCGGCG
 8551 CCACGTGTCGA CTACGCCATC ATGGCGACAG CGCCTTCTCT TTGGGTTCTC
 8601 TATATCGGGC GGATCGTGCC CGGCATCACC GGGCGACTG GGGCGGTAGC
 8651 CGCGCTTTAT ATTGCCGATA TCACTGATGG CGATGAGCGC GCGCGGCACT
 8701 TCGGCTTCAT GAGCGCTGTT TTCGGGTTCC GGATGGTCCG GGGACCTGTG
 8751 CTCGGTGGGC TGATGGGCGG TTTCTCCCCC CACGCTCCCT TCTTCGCGCG
 8801 GGCAGCCTTG AACGGCCTCA ATTTCTGAC GGGCTGTTT CTTTTGCCGG
 8851 AGTCGCACA AGGCGAACC CGGCCGTTAC GCCGGGAGC TCTCAACCCG
 8901 CTCGCTCGT TCCGGTGGGC CCGGGCATG ACCGTCTGTC CCGCCCTGAT
 8951 GCGGCTTCT TTCATCATGC AACTTGTCCG ACAGGTGCCG GCCCGCTTT
 9001 GGGTCAATTT CGGCGAGGAT CGCTTTCCT GGGACGCGAC CACGATCGGC
 9051 ATTTTCGCTTG CCGCATTTGG CATTTCTGAT TCACTCGCCC AGGCAATGAT
 9101 CACCGGCCCT GTAGCCGCC GGCTCGGCGA AAGGCGGCA CTCATGCTCG
 9151 GAATGATTGC CGACGGCACA GGCTACATCC TGCTTGCTT CCGGACACGG
 9201 GGATGGATGG CGTTCGCGAT CATGGTCCCT CTGCTTCCG GTGGCATCGG
 9251 AATGCCGGCG CTGCAAGCAA TGTGTCCAG GCAGGTGGAT GAGGAACGTC
 9301 AGGGGACGCT GCAAGGCTCA CTGGCGGCGC TCACCAGCCT GACCTCGATC
 9351 GTCGGACCCC TCTCTTTCAC GCGGATCTAT GCGGCTTCTA TAACAACGTG
 9401 GAACGGTGG GCATGGATTG CAGGCGCTGC CCTCTACTTG CTCTGCTGC
 9451 CGGCGCTGCG TCGCGGGCTT TGGAGCGGCG CAGGGCAACG AGCCGATCGC
 9501 TGATCGTGG AACGATAGGC CTATGCCATG CCGGTCAAGG CGACTTCCGG
 9551 CAAGCTATAC GCGCCCTAGG AGTGCGGTTG GAACGTTGGC CCAGCCAGAT
 9601 ACTCCCGATC ACGAGCAGGA CGCCGATGAT TTGAAGCGCA CTCAGCGTCT
 9651 GATCCAAGAA CAACCATCCT AGCAACACGG CGGTCCCGG GCTGAGAAAG
 9701 CCCAGTAAG AAACAACGTG AGTTCGAGT CGCGAGATCC CCCGAACCA
 9751 AAGGAAGTAG GTTAAACCCG CTCCGATCAG GCCGAGCCAC GCCAGCCGA
 9801 GAACATTGTT TCCTGTAGGC ATCGGGATTG GCGGATCAAA CACTAAAGCT
 9851 ACTGGAACGA GCAGAAGTCC TCCGGCCGCC AGTTGCCAGG CCGTAAAGGT
 9901 GAGCAGAGGC ACGGGAGGTT GCCACTTGGC GGTGAGCAGC GTTCCGAACG
 9951 CCATGGAAC CGCCCCGCC AGGCCGCTG CGACGCGGAC AGGATCTAGC
 10001 GCTGCGTTTG GTGTCAACAC CAACAGCGCC ACGCCCGCAG TTCCGCAAT
 10051 AGCCCCCAGG ACCGCCATCA ATCGTATCGG GCTACCTAGC AGAGCGGCAG
 10101 AGATGAACAC GACCATCAGC GGCTGCACAG CGCCTACCGT CGCCGCGACC
 10151 CCGCCCCGCA GCGGGTAGAC CGAAATAAAC AACAGCTCC AGAATAGCGA

-continued

10201 AATATTAAGT GCGCCGAGGA TGAAGATGCG CATCCACCAG ATTCCCGTGT
10251 GAACTCTGTC GACGATCATC ACGAGCAATA AACCCGCCGG CAACGCCCGC
10301 AGCAGCATAC CGGCGACCCC TCGCCCTCGC TGTTCGGGCT CCACGAAAAC
10351 GCCGGACAGA TGCCCTTGTG GAGCGTCCTT GGGGCCGTCC TCCGTGTTGA
10401 AGACCGACAG CCCAATGATC TCGCCGTCGA TGTAGCGGCC GAATGCCACG
10451 GCATCTCGCA ACCGTTGAGC GAACGCCTCC ATGGGCTTTT TCTCCTCGTG
10501 CTCGTAACG GACCCGAACA TCTCTGGAGC TTCTTCAGG GCCGACAATC
10551 GGATCTCGCG GAAATCCTGC ACGTCCGGCC CTCCAAGCCG TCGAATCTGA
10601 GCCTTAATCA CAATGTGCAA TTTTAATCCT CTGTTTATCG GCAGTTCGTA
10651 GAGCGCGCCG TCGTCCCGA GCGATACTGA GCGAAGCAAG TCGTTCGAGC
10701 AGTGCCCGCT TGTTCCTGAA ATGCCAGTAA AGCGCTGGCT GCTGAACCCC
10751 CAGCCGGAAC TGACCCACA AGGCCCTAGC GTTTGCAATG CACCAGGTCA
10801 TCATTGACCC AGGCGTGTTC CACCAGGCCG CTGCCTCGCA ACTCTTCGCA
10851 GGCTTCGCCG ACCTGCTCGC GCCACTTCTT CACGGGGGTG GAATCCGATC
10901 CGCACATGAG GCGGAAGGTT TCCAGCTTGA GCGGGTACGG CTCCCGGTGC
10951 GAGCTGAAAT AGTCGAACAT CCGTCGGGCC GTCGGCGACA GCTTCCGTTA
11001 CTTCTCCCAT ATGAATTTCC TGTAGTGGTC GCCAGCAAAC AGCAGCAGCA
11051 TTTCTCCTGC GATCAGGACC TGGCAACGGG ACGTTTCTT GCCACGGTCC
11101 AGGACGGCGA AGCGGTGCGC CAGCGACACC GATTCAGGTT GCCAACGGC
11151 GTCGGACGTG AAGCCCATCG CCGTCGCCCTG TAGGCGCGAC AGGCATTCTT
11201 CGGCCCTCGT GTAATACCGG CCATTGATCG ACCAGCCAG GTCCTGGCAA
11251 AGCTCGTAGA ACGTGAAGTT GATCGGCTCG CCGATAGGGG TCGCTTCGCG
11301 GTACTCCAAC ACCTGCTGCC ACACCAAGTT GTCATCGTCG GCCCGAGCT
11351 CGACGGCGGT GTAGGTGATC TTCACGTCCT TGTGACGTG GAAAATGACC
11401 TGTGTTGCA GCGCCTCGCG CGGGATTTT TGTGCGCGG TGGTGAACAG
11451 GCGAGAGCGG GCGCTGTCGT TTGGCATCGC TCGCATCGTG TCCGGCCAG
11501 GCGCAATATC GAACAAGGAA AGCTGCATTT CCTTGATCTG CTGCTTCGTG
11551 TGTTCAGCA ACGCGGCTG CTTGGCCTCG CTGACCTGTT TFGCCAGGTC
11601 CTCGCGCGCG GTTTTCGCT TCTTGGTCTG CATAGTTCCT CGCGTGTGCA
11651 TGGTCATCGA CTTCCGCAAA CTTGCCGCTT CCGTTCGAG ACGACGCGAA
11701 CGTCCACGGG CGGCCGATGG CGCGGGCAGG GCAGGGGAG CCAGTTGCAC
11751 GCTGTCGCGC TCGATCTTGG CCGTAGCTTG CTGGACCATC GAGCCGACGG
11801 ACTGGAAGGT TTCGCGGGGC GCACGCATGA CCGTGGCGGT TCGCATGGTT
11851 TCGGCATCCT CGGCGGAAAA CCCCCTGTCG ATCAGTTCTT GCCTGTATGC

11901 CTTCCGGTCA AACGTCGGAT TCATTACCC TCCTTGCGGG ATFGCCCGCA
11951 CTCACGCCGG GGCAATGTGC CCTTATTCCT GATTTGACCC GCCTGGTGCC
12001 TTGGTGTCCA GATAATCCAC CTTATCGGCA ATGAAGTCGG TCCCGTAGAC
12051 CGTCTGGCCG TCCTTCTCGT ACTTGGTATT CCGAATCTTG CCTGCACGA
12101 ATACCAGCGA CCCCTTGCCC AAATACTTGC CGTGGGCTTC GGCCGTAGAG
12151 CCAAACACT TGATGCGGAA GAAGTCGGTG CGCTCCTGCT TGTCCCGGT
12201 CGTGGCCGCG CCAACCTTTG CGATCCGCAA GCGCGCGGTC GCCATCTTCA
12251 CGCTGGAACA GTACGTCGAG GCGGGCATCA TGACCCGCGA GCAATACGAG
12301 GTCAATTA AAA GCGCCGTGAT TGATGATATA GCGGCCCGGC TGTCTCTGGT
12351 TCTCGCCGAC CGAATGGGT GACTTCACCC CGCGCTCTTT GATCGTGGCA
12401 CCGATTTCCG CGATGCTCTC CGGGGAAAAG CCGGGTGTGT CGGCCGTCCG
12451 CGGTGATGCG GGATCTTCGT CGATCAGGTC CAGGTCCAGC TCGATAGGGC
12501 CGGAACCGCC CTGAGACGCC GCAGGAGCGT CCAGGAGGCT CGACAGGTCG
12551 CCGATGCTAT CCAACCCAG GCGGACCGGC TCGCCCGCGC CTGCGCGTTC
12601 CTGAGCGCCG CGACCGGTGT TTTCTTGGT GGTCTTGGCT TGAGCCGCGA
12651 TCATGGGGAA ATCTCCATCT TCGTGAACAC GTAATCAGCC AGGGCGCGAA
12701 CCTCTTTCGA TGCCCTGCGC GCGGCCGTTT TCTTGATCTT CCAGACCGGC
12751 ACACCGGATG CGAGGGCATC GCGGATGCTG CTGCGCAGGC CAACGGTGGC
12801 CCGAATCATC ATCTTGGGGT ACGCGGCCAG CAGCTCGGCT TGGTGGCGCG
12851 CGTGGCGCGG ATTCGCGGCA TCGACCTTGC TGGGACCAT GCGAAGGAAT
12901 TGCAGCTTGG CGTCTTCTG GCGCACGTTT GCAATGGTGC TGACCATCTT
12951 CTTGATGCCC TGGATGCTGT ACGCCTCAAG CTGATGGGG GACAGCAGAT
13001 AGTCGGCCCG GAAGAGGGCG GCGCCAGGC CGACGCCAAG GGTCCGGGGC
13051 GTGTCGATCA GGCACACGTC GAAGCCTTGG TTCGCCAGGG CCTTGATGTT
13101 CGCCCCGAAC AGCTCGCGGG CGTCTCCAG CGACAGCCGT TCGGCGTTCG
13151 CCAGTACCGG GTTGGACTCG ATGAGGGCGA GCGCGCGCGC CTGGCCGTGCG
13201 CCGGCTCGCG GTGCGGTTTC GGTCCAGCCG CCGGCAGGGA CAGCGCCGAA
13251 CAGCTTGTCT GCATGCAGGC CGGTAGCAAA GTCTTGTAGC GTGTAGGACG
13301 CATTGCCCTG GGGGTCCAGG TCGATCAGGC CAACCCGCAA GCGCGCTCG
13351 AAAAAGTCGA AGGCAAGATG CACAAGGGTC GAAGTCTTGC CGACGCCGCG
13401 TTTCTGTTG GCGGTGACCA AAGTTTTTCAT CGTTTTGTTT CCTGTTTTTT
13451 CTTGGCGTCC GCTTCCCACT TCCGACGAT GTACGCTTGA TGTCCCGGCA
13501 GAACCGCCGT TACCCGCGCG TACCCCTCGG GCAAGTTCTT GTCTCGAAC
13551 GCGGCCACA CGCGATGCAC CGCTTGCAG ACTGCGCCCC TGGTCAGTCC

13601 CAGCGACGTT GCGAACGTCG CCTGTGGCTT CCCATCGACT AAGACGCCCC
13651 GCGCTATCTC GATGGTCTGC TGCCCCACTT CCAGCCCTG GATCGCTCC
13701 TGGAACTGGC TTTCCGGTAA GCGTTTCTTC ATGGATAACA CCCATAAATT
13751 GCTCCGCGCC TTGGTTGAAC ATAGCCGGTGA CAGCCGCCAG CACATGAGAG
13801 AAGTTTGTAGT AAACATTCT CGCACGTCAA CACCTTGTAG TFCGCTCTT
13851 CGTCTTGGC GTAACAAAAC AAAAGCCCGG AAADDGGGCT TTCGTCTCTT
13901 GCCGCTTATG GCTCTGCACC CGGCTCCATC ACCAACAGGT CGCGCACGGC
13951 CTTCACTCGG TTGCGGATCG ACTGTCAGG CCAACAAAAG CCGGTTGCGG
14001 CCGCCCGCAG GATCGCGCGG ATGATGCCGG CCACACCGGC CATCGCCAC
14051 CAGGTCCCGC CTTTCCGGTT CCATTCTGTC TGGTACTGCT TCGCAATGCT
14101 GGACCTCGCG TCACCATAGG CTGACCGCTC GATGGCGTAT GCCGCTTCTC
14151 CCCTTGGCGT AAAACCAGC GCGCAGGGCG GCATTGCCAT GCTGCCCGCC
14201 GCTTCCCGCA CCACGACGCG CGCACCAGGC TTGCGGTCCA GACCTTCGGC
14251 CACGGCGAGC TGCGCAAGGA CATAATCAGC CCGCGACTTG GCTCCACGGC
14301 CCTCGATCAG CTCTTGCAC TCGCGGAAAT CCTTGGCCTC CACGGCCGCG
14351 ATGAATCGCG CACGCGCGCA AGGCTCCGCA GGGCCG

-continued

Broad-host-range expression vector for ligation-independent cloning

featuring C-terminal 7 ± His tag

pRKLIC2HT1Dpuf.seq Length: 14368

```

1  CCACCCAGGC  CGCCGCCCTC  ACTGCCCGGC  ACCTGGTTCG  TGAATGTCGA
51  TGCCAGCAC  TGGGGCACGT  CAATGCTTCC  GGGCGTCGCG  CTCGGGCTGA
101  TCGCCATPC  CGTTACTGCC  CCGATCCCGG  CAATGGCAAG  GACTGCCAGC
151  GCTGCCATTT  TTGGGGTGAG  GCCGTTCCGG  GCCGAGGGGG  GCAGCCCCTG
201  GGGGGATGG  AGGCCCGCGT  TAGCGGGCCG  GGAGGGTTTC  AGAAGGGGGG
251  GCACCCCCCT  TCGCGCTGCG  CGGTACACCG  CACAGGGGCG  AGCCCTGGTT
301  AAAAACAAAG  TTTATAAATA  TTGGTTTAAA  AGCAGGTTAA  AAGACAGGTT
351  AGCGGTGGCC  GAAAAACGGG  CGGAAACCC  TGCAAATGCT  GGATTTTCTG
401  CCTGTGGACA  GCCCCTCAA  TGTCAATAG  TCGCCCTC  ATCTGTACG
451  ACTCTGCCCC  TCAAGTGTCA  AGGATCGCG  CCTCATGTG  TCAGTAGTCC
501  CGCCCTCAA  GTGTCAATAC  CGCAGGGCAC  TTATCCCCAG  GCTTGTCCAC
551  ATCATCTGTG  GAAACTCGC  GTAAAAACAG  GCGTTTTCG  CGATTTGCGA
601  GGCTGGCCAG  CTCACCTCG  CCGGCCGAAA  TCGAGCCTG  CCCTCATCTG
651  TCAACGCCGC  GCCGGGTGAG  TCGCCCTC  AAGTGTCAAC  GTCCGCCCTC
701  CATCTGTGAC  TGAGGGCCAA  GTTTTCCGG  AGGTATCCAC  AACGCCGGCG
751  GCCCGGTGT  CTCGCACACG  GCTTCGACGG  CGTTTCTGG  GCCTTTGCAG
801  GGCCATAGAC  GGGCCGCCAG  CCAGCGGCGA  GGGCAACCAG  CCGCGTGAGC
851  GTCGGAAAG  CGCTCTTCCG  CTTCCTCGCT  CACTGACTCG  CTGCGCTCGG
901  TCGTTCGGCT  GCGCGAGCG  GTATCAGCTC  ACTCAAAGGC  GGTAATACGG
951  TTATCCACAG  AATCAGGGGA  TAACGCAGGA  AAGAACATGT  GAGCAAAGG
1001  CCAGCAAAG  GCCAGGAACC  GTAAAAAGGC  CGCGTTGCTG  GCGTTTTTCC
1051  ATAGGCTCCG  CCCCCCTGAC  GAGCATCACA  AAAATCGACG  CTCAAGTCAG
1101  AGGTGGCGAA  ACCCGACAGG  ACTATAAAGA  TACCAGGCGT  TTCCCCCTGG
1151  AAGTCTCCCT  GTGGCTCTC  CTGTTCGGAC  CCTGCCGCTT  ACCGGATACC
1201  TGTCCGCTT  TCTCCCTTCG  GGAAGCGTGG  CGCCATTTCG  CATTACGGCT
1251  GCGCAACTGT  TGGGAAGGG  GATCGGTGCG  GGCCTCTTCG  CTATTACGCC
1301  AGCTGGCGAA  AGGGGGATGT  GCTGCAAGGC  GATTAAGTTG  GGTAACGCCA
1351  GGGTTTTTCC  AGTCAACGAC  TTGTAACAG  ACGGCCAGTG  AATTCCGGCG
1401  CCGGCTGGCC  GAGGTGCTGG  GCAAGCCCTA  CCTCCAGGCC  CCCATCGGGG
1451  TCGAGAGCAC  GACCGCCTTC  CTGCGCGGCC  TGGCGGAGAT  TCTGGGCCTC
1501  GATCCGGAGC  CCTTCATCGA  GCGCGAGAAG  CACTCGACGC  TGAAGCCCGT
1551  GTGGGATCTG  TGGCGGAGTG  TCACGCAGGA  CTTCTTCGGG  ACGGCCAATT
1601  TCGGAATCGT  GGGGACCGAA  ACTTATGCAA  GAGGCATCCG  AAACATATCT
1651  GAAGGCGATC  TCGGCTGCC  CTGCGCCTTC  GCCTGGCCCG  CAAGAGGGGC

1701  TCGAAGACCG  ACAACGAAGC  GGTGCGCGGA  CTGATCCGCC  AGCACCGTCC
1751  GCTCGTGTCT  ATGGGGTCGA  TCAACGAGAA  GATTTACCTT  CGGGAACGTA
1801  AAGCCGGTCA  CGGCCCGGCA  ACCCTCTTTC  ATCGTGCCTT  CTTTCCCGGG
1851  TCGCGGATC  CGGCGCGCTA  CCGGAACGCC  CGTTATGGGA  TATGCGAGTG
1901  CTACGTGGTT  ACTGCAGGAA  GTTTGCAACG  CCCTGTTCGA  CGCCCTGTTC
1951  CACATTTCTG  CCTTCGGGAC  GGAGATGGAC  AGCGCCGCGC  CCACACCGAC
2001  GACACTGCGC  CGCGACTTCC  CGTGGGATGC  CGATGCGCAA  CGGGCCCTGG
2051  ACCGCATCGT  AGAGGAGCAT  CCGGTTCTCA  CCGGATCAG  CGCCGCGCGT
2101  GCCTTGCGCG  ACGCCGCGCA  GAAGGCTTGC  CCTCGATGCC  GGTGCCGAGA
2151  GGGTCGTGAG  AGAGACTGTG  GAAGCCCTGA  CGTGGGCCGG  GCTTCGGCGA
2201  GAGGAAGGGA  GAGAACCAAT  GAGCGATCAT  GCCGTCAACA  CGCCGGTCCA
2251  TGCCGCCAGG  GCCACGGGG  ACCGAGCACC  TCGTCCGAG  TTCTACGTCT
2301  ACTTCGCGCT  CATTCTGCTG  GCGCCTTCC  CGGTGGCCTT  CGTGAGCTGG
2351  ATCGTCTCGA  CGATCCGCCA  CCGCAGGCTT  CCAAGCGCG  GCCCTTCGC
2401  GTCCGCCTGG  TTCGATGCCA  AGCGATCAC  GCCGTGATT  TTCCGCGCCT
2451  GACCGAGGT  CAGGTTGCGA  CACGCCATTC  GTCGTCTCCC  CAAGGGGCGG
2501  CGGATTAATC  GGGAGGGCAT  GGTGCCTTAC  CGTAACCCAC  GCCACCAGTA
2551  GCGAGGAGGA  ACACGTGTCG  TCCGTTGGTC  ACGAGGAGGAG
2601  TAATAGGCCG  GCCCTCCGTC  GCGGGCGGCA  CCCACGCCCG  CAtCGATTCC
2651  AAGGTTACAG  CATTGAGACG  GCTCCGCTTC  GCGCGCAAGC  GCGGGTTGGG
2701  CCGACTGCAA  GCGGAGAGGG  AAGCATGGCA  CTGCTCAGCT  TCGAGCGAAA
2751  ATATCGCGTG  CCGGGGGGCA  CGCTGGTCCG  CGGAAACCTG  TTCGACTTCT
2801  GGGTCGGCCC  TTTCTATGTC  GGCTTCTTCG  GGGTTGCGAC  GTTTTCTTTC
2851  GCGGCCCTGG  GTATCATCTT  GATTGCTCTG  AGTGCCTGAC  TCCAGGGTAC
2901  CTGGAACCCC  CAACTCATCT  CTGTCTACCC  GCCGGCCCTT  GAATATGGCC
2951  TGGGAGGTGC  ACCCCTCGCA  AAAGGCGGGC  TGTGGCAGAT  CATCACGATC
3001  TGCGCCACTG  GTGCCCTTCT  CAGCTGGGGC  CTCGCCGAAG  TCGAAATCTG
3051  CCGTAAGCTG  GGCATCGGGT  ACCACATCCC  GTTCGCCTTC  GCGTTCCGCA
3101  TCCTGGCCTA  CCTGACGCTG  GTGCTGTTC  GCCCGGTGAT  GATGGGGCCG
3151  TGGGGCTATG  CCTTCCCTTA  CGGGATCTGG  ACGCACCTCG  ACTGGGTGTC
3201  GAACACGGGC  TACACCTACC  GCAACTTCCA  CTACAACCC  GCCCACATGA
3251  TCGCCATCTC  GTTCTTCTTC  ACGAACCGGC  TGGCTCTGGC  GCTGCACGGC
3301  GCCCTTGTG  TCTCCGCGGC  CAACCCCGAG  AAGGGCAAGG  AAATCGGGAC
3351  GCCGGATCAC  GAGGATACGT  TCTTCCGCGA  TCTGGTCCGG  TACTCGATCG

```

-continued

3401	GGACGCTCGG	CATCCACCGC	CTCGGCCTGC	TGCTCTCGCT	GAGCGCCGTC
3451	TTCTTCAGCG	CCCTCTGCAT	GATCATFACC	GGCACCATCT	GGTTCGATCA
3501	GTGGGTGCGA	TGGTGGCAAT	GGTGGGTGAA	GCTGCCGTGG	TGGGCGAACA
3551	TCCCGGGAGG	CATCAATGGC	TGAGTATCAG	AACATCTTCT	CCCAGGTCCA
3601	GGTCCCGCGA	CCGCGCGACC	TGGGGATGAC	CGAAGACGTC	AACCTGGCCA
3651	ACCGTTCGGG	CGTCGGTCCC	TTCTCGACCC	TGCTCGGCTG	GTTCGGCAAC
3701	GCCGAGCTCG	GCCCGATCTA	TCFCGGCTCG	CTCGGCGTCC	TGTCCCTCTT
3751	CTCGGGCTCG	ATGTGGTCTT	TCACCATFCG	GATCTGGTTC	TGGTATCAGG
3801	CGGGCTGGAA	CCGCGCCGTC	TTCTTGCGCG	ACCTGTTCTT	CTTCTCGCTC
3851	GAGCCGCGCG	CACCCGAATA	CGGTCTGTCC	TTCCGCGGCTC	CGCTGAAGGA
3901	AGGCGGGCTG	TGGCTGATCG	CGTCGTCTTT	CATGTTCCGT	GCGGTCTGGT
3951	CCTGGTGGGG	CCGCACCTAT	CTCCGCGCTC	AGGCGCTGGG	CATGGGCAAG
4001	CACACCCGCT	GGGCGTTCCT	CTCGGCCATC	TGGCTGTGGA	TGGTGTCTGG
4051	CTTCATCCGT	CCGATCCTCA	TGGGTCTCTG	GTCCGAAGCG	GTTCCTACG
4101	GCATCTTCTC	GCACCTCGAC	TGGACGAACA	ACTTCTCGCT	CGTCCACGGC
4151	AACCTGTTCT	ACAACCCCTT	CCACGGTCTC	TCGATCGCCT	TCCTCTACGG
4201	TTCCGGCCCTG	CTCTTCGCGA	TGCACGGTGC	GACCATCCTC	GCGGTCTCCC
4251	GCTTCGGCGG	CGAGCGCGAG	CTGGAGCAGA	TCGCCGACCG	CGGGACGGCA
4301	CCGGAGCGGG	CCGCCCTCTT	CTGGCGCTGG	ACCATGGGTT	TCAACGCCAC
4351	GATGGAAGGC	ATCCACCCTG	GGCCATCTGT	GATGGCGGTC	CTCGTGACCC
4401	TCACCGCGCG	CATCGGgATC	CTGCTCTCGG	GCACGGTCTG	GGACAACCTGG
4451	TACGTCTGGG	GCCAGAACCA	CGGCATGGCG	CCGCTGAAC	GAGGAGCGAT
4501	CACAATGGCT	GACAAGACCA	TCCTCAACGA	TCACCTCAAC	ACCAATCCGA
4551	AGACCAACCT	TCGCCTCTGG	GTGCTTTTCC	AGATGATGAA	GGGTGCGGGC
4601	TGGGCTGGCG	GCGTGTCTTT	CGGGACGCTC	CTTCTCATCG	GGTCTTCCG
4651	GGTGGTCCGG	CGGATGCTTC	CGATCCAGGA	GAACCAGGCT	CCGCGCCCGA
4701	ACATCACCGG	CGCTCTGGAG	ACCGGGATCG	AGCTGATCAA	GCATCTCGTC
4751	TGAGACAAGT	CTCGGGGCGAG	GGCGGCGCGA	GGCCGCCCGC	TCCTCAAGT
4801	CCGGGCCATA	TCGCCGGCCC	GGTCCGGGG	CGACACCACA	GCCCGGTTCC
4851	CTTCTCTGTTG	GCGACAGGGA	CTTGGTGCCG	TGTGGAAGAC	GCCACGGCAC
4901	CTTTTGGACA	TTACGGGGAG	GCTCTGATGA	CCAATCCAC	CCCGGACCC
4951	GAACCCCGC	TTTTGGATCG	CGTCTGTGTC	CCGGCCGACA	TGAAGGCGCT
5001	GAGTGACCGC	GAACTGGAGC	GGCTGGCCGA	CGAAGTGCCT	TCCGAGGTCA
5051	GTGATAGGAG	TAGTTTCTTA	TTTTAGGCAG	TTTATATGAA	ATTAAGACAT
5101	GCAGATGTCA	CAGTGGATAT	TGAACGTGTC	TCGAAAGCTC	AATATCCCCC
5151	AAAGCACAA	CACAACTTC	GACATCATGC	AGAAGCGTTT	CCCCAAcgc
5201	gtcttcgacg	tgggcatcgc	cgagcagcat	gccgtgacct	tcgcgccggy
5251	cctcgcgGG	GccggGatga	agccctctcg	cgcgatctat	tcctcgttcc
5301	tgaacgggg	ttacgaccag	atcgcccatg	acgtggcgct	gcagaacctt
5351	ccggtccgct	tcgtgatcga	ccggcggggg	ctcgtggggg	ccgatggcgc
5401	gaccatcgcg	ggggcctTCG	ACGTTGGCTT	CATCACTTCG	CTGCCAACA
5451	TGACCGTGAT	GGCCCGGGCC	GACGAGGCCG	AGCTCATCCA	CATGATcgcc
5501	ACCGCCGTGG	CCTTCGACGA	GGCCCCATC	GCCTTCGCT	TCCCGCGGGG
5551	CGAGGGGGTG	GGCTTCGAGA	TGCCCGAGCG	CGGGACGGTG	CTGGAGCCCG
5601	gCCGGGGCCG	CGTGGTGGCG	GAAGGGACgg	atgtcgcgat	ccctcccttc
5651	GGCGGCATC	TGCACGAGgc	cTTGcAGGCG	gcGAACTTC	TCGAGGccGA
5701	GGGGGTGAGC	GTGACCGTGG	CCGACgcccg	CTTctCgCgC	CCGCTCgAcA
5751	CGGGGCTCAT	CGACCAGCTC	GTgcGCCATC	ACGCGGCGCT	GGTAACGGTG
5801	GAGCAGGGGG	CcATGGGCGG	CTTCGGCGCC	CATGTATGTC	ACTATCTCGC
5851	CAATTCCGGC	GGCTTCGACG	GGGGCTTCGC	GCTCCGGGTC	ATGACGCTGC
5901	CCGACCCGTT	CATCGAGCAG	GCGAGCCCGG	AGgACATGTA	TGCCGATGCG
5951	GGGCTGGGG	CCGAGGATAT	CAAGCTTGGC	GTAATCATGG	TCATAGCTGT
6001	TTCTGTGTG	AAATGTTTAT	CCGCTCACAA	TTCCACACAA	CATACGAGCC
6051	GGAAAGCATA	AGTGTAAAGC	CTGGGGTGCC	TAATGAGTGA	GCTAACTCAC
6101	ATTAATTGCG	TTGCGCTCAC	TGCCCGCTTT	CCAGTCGGGA	AACCTGTCTG
6151	GCCAGCTGCA	TTAATGAATC	GGCCAACGCG	CGGGGAGAGG	CGGTTTGGCT
6201	ATTGGGGCCT	CGGTCTTGCC	TTGCTCGTCC	GTGATGTACT	TCACCAGCTC
6251	CGCGAAGTCG	CTCTTCTTGA	TGGAGCGCAT	GGGGACGTGC	TTGGCAATCA
6301	CGCGCACCC	CCGCGCTTT	TAGCGGCTAA	AAAAGTCATG	GCTCTGCCCT
6351	CGGGCGGACC	ACGCCCATCA	TGACCTTGCC	AAGCTCGTCC	TGCTTCTCTT
6401	CGATCTTCGC	CAGCAGGGCG	AGGATCGTGG	CATCACCGAA	CCGCGCCGTG
6451	CGCGGGTCTG	CGGTGAGCCA	GAGTTTCAGC	AGGCCGCCCA	GGCGGCCAG
6501	GTCGCCATTG	ATCGGGGCCA	GCTCGCGGAC	GTGCTCATAG	TCCACGACGC
6551	CCGTGATTTT	GTAGCCCTGG	CCGACGGCCA	GCAGGTAGGC	CGACAGGCTC
6601	ATGCGGGCCG	CCGCGCCCTT	TTCTCAATC	GCTCTCGTGT	CGTCTGGAAG
6651	GCAGTACAC	TTGATAGGTG	GGCTGCCCTT	CCTGGTTGGC	TTGGTTTCAT
6701	CAGCCATCCG	CTTGCCCTCA	TCGTTCACGC	CGGCGTAGC	CGGCCACCT
6751	CGCAGAGCAG	GATTCGCCGT	GAGCACCGCC	AGGTGCGAAT	AAGGGACAGT

-continued

6801 GAAGAAGGAA CACCCGCTCG CGGGTGGGCC TACTTCACCT ATCCTGCCCC
6851 GCTGACGCCG TTGGATACAC CAAGGAAAGT CTACACGAAC CCTTTGGCAA
6901 AATCCTGTAT ATCGTGCGAA AAAGGATGGA TATACGAAA AAATCGCTAT
6951 AATGACCCCG AAGCAGGGTT ATGCAGCGGA AAAGCGCCAC GCTTCCCGAA
7001 GGGAGAAAAG CGGACAGGTA TCCGGTAAGC GGCAGGGTCG GAACAGGAGA
7051 GCGCACGAGG GAGCTTCCAG GGGGAAACGC CTGGTATCTT TATAGTCCTG
7101 TCGGGTTTCG CCACCTCTGA CTTGAGCGTC GATTTTTGTG ATGCTCGTCA
7151 GGGGGGGCGA GCCTATGGAA AAACGCCAGC AACCGGCCCT TTTTACGGTT
7201 CCTGGCCTTT TGCTGGCCTT TTGCTCACAT GTTCTTTCCT GCGTTATCCC
7251 CTGATTCTGT GGATAACCGT ATTACCGCCT TTGAGTGAGC TGATACCGCT
7301 CGCCGACGCC GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAAGCGGA
7351 AGAGCGCCAG AAGGCCGCCA GAGAGGCCGA GCGCGGCCGT GAGGCTTGGA
7401 CGCTAGGGCA GGGCATGAAA AAGCCCGTAG CGGGCTGCTA CGGGCGCTG
7451 ACGCGGTGGA AAGGGGGAGG GGATGTGTGTC TACATGGCTC TGCTGTAGTG
7501 AGTGGGTTGC GCTCCGGCAG CGTCTGAT CAATCGTAC CCTTCTCGG
7551 TCCTTCAACG TTCTTGACAA CGAGCCTCCT TTTCCCAAT CCATCGACAA
7601 TCACCCGGAG TCCTTGCTCG AACGCTGCGT CCGGACCCGC TTCGTGGAAG
7651 GCGTCTATCG CGGCCCGCAA CAGCGCCGAG AGCGGAGCCT GTTCAACGGT
7701 GCCGCCCGCG TCGCCGGCAT CGCTGTCGCC GCCTGCTCC TCAAGCACGG
7751 CCCCAACAGT GAAGTAGCTG ATTGTATCA GCGCATTGAC GCGTCCCGG
7801 GCCGAAAAAC CCGCCTCGCA GAGGAAGCGA AGCTGCGCGT CGGCCCTTTC
7851 CATCTGCGGT GCGCCCGGTC GCGTCCGGC ATGGATGCGC GCGCCATCGC
7901 GGTAGGCGAG CAGCGCCTGC CTGAACTGCG GGGCATTCCC GATCAGAAT
7951 GAGCGCCAGT CGTCTGCGGC TCFCGGCACC GAATGCGTAT GATTCTCCGC
8001 CAGCATGGCT TCGGCCAGTG CGTFCGAGCAG CCGCCGCTTG TFCCTGAAGT
8051 GCCAGTAAAG CGCCGGCTGC TGAACCCCA ACCGTCCCGC CAGTTGCGT
8101 GTCGTACAGC CGTCTACGCC GACCTCGTTC AACAGTCCA GGGCGGCACG
8151 GATCACTGTA TTCCGGTGCA ACTTTGTCAT GCTTGACACT TTATCACTGA
8201 TAAACATAAT ATGTCCACCA ACTTATCAGT GATAAAGAAT CCGCCGTTTC
8251 AATCGGACCA GCGGAGGCTG GTCCGGAGGC CAGACATGAA ACCCAACATA
8301 CCCCTGATCG TAATCTGAG CACTGTCGCG CTCGACGCTG TCGGCATCGG
8351 CCTGATTATG CCGGTGCTGC CGGCCCTCCT GCGGATCTG GTTCACTCGA
8401 ACGACGTCC CCGCCACTAT GGCATTCTGC TGGCGTGTG TGCGTTGGTG
8451 CAATTTGCGC GCGCACCTGT GCTGGGCGCG CTGTGCGGATC CTTTCGGGCG

8501 GCGGCCAATC TTGCTCGTCT CGCTGGCCGG CGCCACTGTC GACTACGCCA
8551 TCATGGGAC AGGCCTTTC CTTTGGGTTT TCTATATCGG GCGGATCGTG
8601 GCCCGCATCA CCGGGCGGAC TGGGGCGGTA GCCGGCGCTT ATATTGCCGA
8651 TATCACTGAT GCGGATGAGC GCGCGCGGCA CTTCCGGCTT ATGAGCGCCT
8701 GTTTCCGGTT CCGGATGGTC GCGGGACCTG TGCTCGGTGG GCTGATGGGC
8751 GTTTTCTCCC CCCACGCTCC GTTCTTCGCC GCGGCAGCCT TGAACGGCCT
8801 CAATTTCTGT ACGGGCTGTT TCCTTTTGCC GGAGTGCAC AAAGCGAAC
8851 GCCGGCCGTT ACGCCGGGAG GCTCTCAACC CGCTCGCTTC GTTCCGGTGG
8901 CCCCAGGCA TGACCGTCTG CGCCGCCCTG ATGGCGTCT TCTTCATCAT
8951 GCAACTTGTC GGACAGGTGC CGGCCGCGCT TTGGGTCATT TTCGGCGAGG
9001 ATCGCTTTCA CTGGGACGCG ACCACGATCG GCATTTTCGCT TGCCGCATTT
9051 GGCATTCGTC ATTCACTGCG CCAGGCAATG ATCACCAGGC CTGTAGCCGC
9101 CCGGCTCGGC GAAAGCGGG CACTCATGCT CGGAATGATT GCCGACGGCA
9151 CAGGCTACAT CCTGCTTGCC TTCGCGACAC GGGGATGAT GCGTTCCTCG
9201 ATCATGGTCC TGCTTGCTTC GGTGGGCATC GGAATGCCGG CGTGAACG
9251 AATGTTGTCC AGGCAGGTGG ATGAGGAACG TCAGGGGCGAG CTGAAGGCT
9301 CACTGGCGGC GCTCACCAGC CTGACCTCGA TCGTCCGACC CCTCTCTTC
9351 ACGGCGATCT ATGCGGCTTC TATAACAACG TGGAACGGGT GGGCATGGAT
9401 TGCAGGCGCT GCCCTTACT TGCTCTGCTT GCGCGCGCTG CGTCCGGGCG
9451 TTTGGAGCGG CGCAGGGCAA CGAGCCGATC GCTGATCGTG GAAACGATAG
9501 GCCTATGCCA TGCGGGTCAA GCGGACTTCC GGCAAGCTAT ACGCGCCCTA
9551 GGAGTGGCGT TGAACGTTG GCCAGCCAG ATACTCCCGA TCACGAGCAG
9601 GACGCCGATG ATTTGAAGCG CACTCAGCGT CTGATCCAAG AACAAACATC
9651 CTAGCAACAC GCGGCTCCCC GGGCTGAGAA AGCCAGTAA GGAACAACCT
9701 GTAGGTTCTGA GTCGCGAGAT CCCCCGGAAC CAAAGGAAAGT AGTTAAAC
9751 CGCTCCGATC AGGCCGAGCC ACGCCAGGCC GAGAACATTG GTTCTGTAG
9801 GCATCGGGAT TGGCGGATCA AACACTAAAG CTAAGTGAAC GAGCAGAAGT
9851 CCTCCGGCCG CCAGTTGCCA GCGGTAAAG GTGAGCAGAG GCACGGGAGG
9901 TTGGCACTTG CCGGTACGCA CGGTTCGGAA CGCCATGGAA ACCGCCCCCG
9951 CCAGGCCCGC TCGGACGCCG ACAGGATCTA GCGCTGCGTT TGGTGTCAAC
10001 ACCAACAGCG CCAGCCCGCG AGTTCCGCAA ATAGCCCCCA GGACCCCAT
10051 CAATCGTATC GGGCTACCTA GCAGAGCGGC AGAGATGAAC ACGACCATCA
10101 GCGGCTGCAC AGCGCTTACC GTCGCCGCGA CCCCCCGCGG CAGGCGGTAG
10151 ACCGAAATAA ACAACAAGCT CCAGAATAGC GAAATATTA GTCGCCGCGG

-continued

10201 GATGAAGATG CGCATCCACC AGATTCCCGT TGGAACTCTG CGGACGATCA
 10251 TCACGAGCAA TAAACCCGCC GGCAACGCC GCAGCAGCAT ACCGGCGACC
 10301 CCTCGGCCCTC GCTGTTCCGG CTCCACGAAA ACGCCGGACA GATGCGCCTT
 10351 GTGAGCGTCC TTGGGGCCGT CCTCCTGTTT GAAGACCGAC AGCCCAATGA
 10401 TCTCGCCGTC GATGTAGGCG CCGAATGCCA CGGCATCTCG CAACCGTTCA
 10451 GCGAACGCCCT CCATGGGCTT TTTCTCCTCG TGCTCGTAAA CGGACCCGAA
 10501 CATCTCTGGA GCTTCTTCA GGGCCGACAA TCGGATCTCG CGGAAATCCT
 10551 GCACGTCCGGC CGCTCCAAGC CGTCAATCT GAGCCTAAT CACAATTGTC
 10601 AATTTAATC CTCFTGTTAT CGGCAGTTCG TAGAGCGCCG CGTGCCTCC
 10651 GAGCGATACT GAGCGAAGCA AGTGCCTCGA GCAGTGCCCG CTGTTCCCTG
 10701 AAATGCCAGT AAAGCGCTGG CTGCTGAACC CCCAGCCGGA ACTGACCCCA
 10751 CAAGGCCCTA CGCTTTGCAA TGACCAGGT CATCATTTGAC CCAGGCGTGT
 10801 TCCACCAGGC CGCTGCCTCG CAACTCTTCG CAGGCTTCG CGACCTGCTC
 10851 GGGCCACTTC TTCACGGGG TGGAATCCGA TCCGCACATG AGGCGGAAGG
 10901 TTTCCAGCTT GAGCGGTAC GGTCCCGGT CCGAGCTGAA ATAGTCAAC
 10951 ATCCGTCGGG CCGTCGGCGA CAGCTTCCG TACTTCTCCC ATATGAATTT
 11001 CGTGTAGTGG TCGCCAGCAA ACAGCAGCAC GATTTCTCTG TCGATCAGGA
 11051 CTGGCAACG GAGCGTTTTC TTGCCACGGT CCAGGACGCG GAAGCGGTGC
 11101 AGCAGCGACA CCGATTCCAG GTGCCAACG CGTCCGGACG TGAAGCCCAT
 11151 CGCGGTCCGC TGTAGGCGCG ACAGGCATTC CTCGGCCTTC GTGTAATACC
 11201 GGCCATTGAT CGACCAGCCC AGTCCCTGGC AAAGCTCGTA GAACGTGAAG
 11251 GTGATCGGCT CGCCGATAGG GGTGCGCTTC GCGTACTCCA ACACCTGCTG
 11301 CCACACCAGT TCGTCATCGT CGGCCCGCAG CTCGACGCGG GTGTAGGTGA
 11351 TCTTCACGTC CTTGTGACG TGGAAAATGA CCTTGTTTTG CAGCGCCTCG
 11401 CGCGGGATTT TCTGTGTCG CGTGGTGAAC AGGGCAGAGC GGGCGGTGTC
 11451 GTTTGGCATC GCTCGCATCG TGTCCGGCCA CGCGCAATA TCGAACAGG
 11501 AAAGCTGCAT TTCTTTGATC TGCTGCTTCG TGTGTTTCAG CAACCGGGCC
 11551 TGCTTGGCCT CGCTGACCTG TTTTGCCAGG TCCTCGCCCG CGGTTTTTCG
 11601 CTCTTGGTTC GTCATAGTTC CTCGCGTTCG GATGGTTCAT GACTTCGCCA
 11651 AACCTGCCGC CTCCTGTTCG AGACGACGCG AACCGTCCAC GCGCGCCGAT
 11701 GGGCGGGGCA GGGCAGGGGG AGCCAGTTGC ACGCTGTCCG GCTCGATCTT
 11751 GGGCGTAGCT TGCTGGACCA TCGAGCCGAC GGACTGGAAG GTTTCGCGGG
 11801 GCGCACGCAT GACGGTCCGG CTTGCCGATG TTTCCGCATC CTCGGCGGAA
 11851 AACCCCGCGT CGATCAGTTC TTGCCTGTAT GCCTTCCGGT CAAACGTCCG

 11901 ATTCAATCAC CCTCCTTGGC GGATTGCCCC GACTCACGCC GGGGCAATGT
 11951 GCCCTTATTC CTGATTTGAC CCGCCTGGTG CCTTGGTGTG CAGATAATCC
 12001 ACCTTATCGG CAATGAAGTC GGTCCCGTAG ACCGTCTGGC CGTCTCTC
 12051 GTACTTGGTA TTCCGAATCT TGCCCTGCAC GAATACCAGC GACCCTTGC
 12101 CCAATACTTT GCCGTGGGCC TCGCCCTGAG AGCCAAAACA CTGTATGCGG
 12151 AAGAAGTCGG TGGCCTCCTG CTTGTGCGCG GTCGTGGCCG CGCCAACCTT
 12201 TCGCATCCGC AAGCGCGCGG TCGCCATCTT CACGCTGGA CAGTACGTCG
 12251 AGGCGGGCAT CATGACCCCG GAGCAATACG AGGTCAATTA AAGCGCCGTG
 12301 ATTGATGATA TAGCGGCCCG GCTGCTCCTG GTTCTCGCGC ACCGAAATGG
 12351 GTGACTTCAC CCGCGCTCTT TTGATCGTGG CACCGATTTT CGCGATGCTC
 12401 TCCGGGGAAA AGCCGGGGTT GTCCGCCGTC CGCGCTGAT GCGGATCTTC
 12451 GTCGATCAGG TCCAGGTCCA GCTCGATAGG GCCGGAAACCG CCTTGAGAGC
 12501 CCGCAGGAGC GTCCAGGAGG CTCGACAGGT CGCCGATGCT ATCCAACCC
 12551 AGGCGCGACG GCTGCGCCCG GCCTGCGGCT TCCTGAGCGG CCGCAGCGGT
 12601 GTTTTTCTTG GTGGCTCTGG CTTGAGCCCG AGTCATTTGG AATCTCCAT
 12651 CTTCTGTAAC ACCTAATCAG CCAGGCGCGG AACCTCTTTC GATGCTTGC
 12701 GCGCGCCGCT TTTCTTGATC TTCCAGACCG GCACACCGGA TCGGAGGGCA
 12751 TCGGCGATGC TGCTGCGCAG GCCAACGGTG GCCGGAAATCA TCATCTTGGG
 12801 GTACGCGGCC AGCAGCTCCG CTTGGTGGCG CGCGTGGCGG GGATTCGCGC
 12851 CATCGACCTT GCTGGGCACC ATGCCAAGGA ATTGCAGCTT GCGTCTCTTC
 12901 TGGCGCAGCT TCGCAATGGT CGTGACCATC TTTCTGATGC CTGGATGCT
 12951 GTACGCCTCA AGCTCGATGG GGGACAGCAC ATAGTCCGGC GCGAAGAGGG
 13001 CGGCGCCAGC GCGGACGCCA AGGGTCGGGG CCGTGTCCAT CAGGCACAGC
 13051 TCGAAGCCTT GGTTCGCCAG GGCCTTGATG TTCGCCCCGA ACAGCTCCG
 13101 GGCCTCGTCC AGCGACAGCC GTTCCGGCCT CGCCAGTACC GGGTTGGACT
 13151 CGATGAGGGC GAGGCGCGCG GCCTGGCCGT CGCCGGCTGC GGGTCCGGTT
 13201 TCGGTCCAGC CGCCGGCAGG GACAGCGCCG AACAGCTTGC TTGCATGCAG
 13251 GCCGGTAGCA AAGTCCCTGA GCCTGTAGGA CGCATTGCC TGGGGTCCA
 13301 GGTGATCAC GGCACCCCGC AAGCCCGCT CGAAAAGTC GAAGGCAAGA
 13351 TGCACAAGGG TCGAAGTCTT GCCGACGCGC CCTTTCTGTT TGGCCGTGAC
 13401 CAAAGTTTTC ATCGTTGGT TTTCTGTTTT TTTCTGGCGT TGGCCGTGAC
 13451 CTTCCGACG ATGTACGCC TGTGTTCCCG CAGAACCCTG GTTACCCCGC
 13501 CTACCCCTC GGGCAAGTTC TTGTCTCGA ACGCGGCCA CACGCGATGC
 13551 ACCGCTTGGC ACATGCGCC CCTGGTCACT CCCAGCGACG TTGCGAACGT

 13601 CGCCTGTGGC TTCCATCGA CTAAGACGCC CCGCGTATC TCGATGGTCT
 13651 GCTGCCCCAC TTCCAGCCCC TGGATCGCCT CCTGGAATG GCTTTCGGTA
 13701 AGCCGTTTCT TCATGGATAA CACCCATAAT TTGCTCCGCG CCTTGGTTGA
 13751 ACATAGCCGT GACAGCCGCC AGCACATGAG AGAAGTTTAG CTAACATTT
 13801 CTCGCAGGTC AACACCTTTA GCCGTAATAA CTCGTCTTTC GCGTAACAAA
 13851 ACAAAGCCG GGAACCCGGG CTTTCGTCTC TTGCCGCTTA TGGCTCTGCA
 13901 CCCGGCTCCA TCACCAACAG GTCGCGCACG CGCTTCACT GGTTCGGGAT
 13951 CGACACTGCC AGCCCCAACAA AGCCGGTTGC CGCCCGCCCG AGGATCGCGC
 14001 CGATGATGCC GGGCACACCG GCCATCGCCC ACCAGTCCG CGCCTTCCGG
 14051 TTCCATTCCT GCTGGTACTG CTTCCGCAATG CTGGACCTCG GCTCACCATA
 14101 GGCTGACCGC TCGATGGCGT ATGCCGCTTC TCCCTTGGC GTAAAACCCA
 14151 GCGCCGACG CGGCATTGCC ATGCTGCCCG CCGCTTTCCT GACCACGACG
 14201 CGCGCACCG GCTTGGCGTC CAGACCTTCC GCCACGGCGA GCTGCGCAAG
 14251 GACATAATCA GCGCCGACT TGGCTCCACG CGCCTCGATC AGCTCTTGCA
 14301 CTCGCGCGAA ATCCTTGGCC TCCACGGCCC CCATGAATCG CGCACGCGGC
 14351 GAAGGCTCCG CAGGGCCG

-continued

Broad-host-range expression vector featuring C-terminal 10 ± His tag

pRKPLHT10Dpuf Length: 14375

```

1 CCACCCAGGC CGCGCCCTC ACTGCCCGGC ACCTGGTTCG TGAATGTCGA
51 TGCCAGCACG TGGGGCACGT CAATGCTTCC GGGCGTCGCG CTCGGGCTGA
101 TCGCCCATCC CGTTACTGCC CCGATCCCGG CAATGGCAAG GACTGCCAGC
151 GCTGCCATTT TTGGGGTGAG GCGTTCGCGG GCGGAGGGGC GCAGCCCCTG
201 GGGGGATGGG AGGCCCGCGT TAGCGGGCCG GGAGGGTTTC AGAAGGGGGG
251 GCACCCCCCT TCGGCGTGCG CCGTACCGCG CACAGGGGCG AGCCCTGGTT
301 AAAAACAAAG TTTATAAATA TTGGTTTAAA AGCAGGTTAA AAGACAGGTT
351 AGCGGTGGCC GAAAACCGGG CGGAAACCCCT TGCAAATGCT GGATTTTCTG
401 CCTGTGGACA GCCCCTCAA TGTCAATAGG TCGCCCTC ATCTGTACGC
451 ACTCTGCCCC TCAAGTGTC AAGGATCGCG CCGTATGTG TCAGTAGTTCG
501 CGCCCCTCAA GTGTCAATAC CGCAGGGCAC TTATCCCCAG GCTTGTCCAC
551 ATCATCTGTG GAAACTCGC GTAAAAACAG CGGTTTTCGC CGATTTGCGA
601 GGCTGGCCAG CTCACGTCG CCGGCCGAAA TCGAGCCTGC CCCTCATCTG
651 TCAACGCCGC GCGGGTGAG TCGCCCTC AAGTGTCAAC FTCCGCCCTC
701 CATCTGTGCG TGAGGGCCAA GTTTTCCGCG AGGTATCCAC AACGCCGGCG
751 GCGCGGTGTG CTCGCACAGC GCTTCGACGG CGTTTCTGGC GCCTTTGCAG
801 GGCCATAGAG GCGCCGACG CCAGCGGCGA GGGCAACCAG CCGGTGAGC
851 GTCCGAAAGG CGCTCTCCG CTTCCTCGCT CACTGACTCG CTGCGCTCGG
901 TCGTTCGGCT GCGCGAGCG GTATCAGCTC ACTCAAAGGC GGAATACGG
951 TTATCCACAG AATCAGGGGA TAACGCAGGA AAGAACATGT GAGCAAAGG
1001 CCAGCAAAGG GCCAGGAACC GTAAAAAGGC CGCGTTGCTG GCGTTTTTCC
1051 ATAGGCTCCG CCCCCCTGAC GAGCATCACA AAAATCGAGC CTCGAAGTCAG
1101 AAGTGGCGAA ACCCGACAGG ACTATAAAGA TACCAGGCGT TTCCCCCTGG
1151 AAGTCTCCCT GTGCGCTCTC CTGTTCCGAC CCGTCCGCTT ACCGGATACC
1201 TGTCGGCTTT TCTCCCTTCG GGAAGCGTGG CGCCATTTCG CATTAGGCTT
1251 GCGCAACTGT TGGGAAGGGC GATCGGTGCG GGCCTTTCG CTATTACGCC
1301 AGCTGGCGAA AGGGGGATGT GCTGCAAGGC GATTAAGTTG GGTAAACGCCA
1351 GGGTTTTTCC AGTCAACGAG TTGTAACAGC ACGGCCAGTG AATTCCGGCCG
1401 CCGGCTGGCC GAGGTGCTGG GCAAGCCCTA CCTCCAGGCC CCCATCGGGG
1451 TCGAGAGCAC GACCGCTTC CTGCGCGGCC TGGCGAGAT TCTGGGCCCTC
1501 GATCCGGAGC CCTTCATCGA GCGCGAGAAG CACTCGACGC TGAAGCCCGT
1551 GTGGGATCTG TGGCGGAGTG TCACGCAGGA CTTCTCGGG ACGGCCAATT
1601 TCGGAATCGT GCGGACCGAA ACTTATGCAA GAGGCATCCG AAACATATCTC
1651 GAAGGCGATC TCGGGCTGCC CTGCGCTTC GCCGTGGCCC GCAAGAGGGG

1701 CTCGAAGACC GACAACGAAG CGGTGCGCG ACTGATCCGC CAGCACCGTC
1751 CGCTCGTGCT CATGGGTGCG ATCAACGAGA AGATTACCTT TCGGGAACCTG
1801 AAAGCCGGTC ACGGCCCGCA ACCCTCTTTC ATCGCTGCCT CTTTCCCGGG
1851 TCGGGCGATC CGGCGCGCTA CCGGAACGCC CGTTATGGGA TATGCGAGTG
1901 CTACGTGGTT ACTCGAGGAA GTTTGCAACG CCTGTTCGA CGCCCTGTTC
1951 GACATTTCTG CCTTCGGGAC GGAGATGGAC AGCGCCGCGC CCACACCAGC
2001 GACACTGCGC CGGACTTCC CGTGGGATGC CGATGCGCAA GCGGCCCTGG
2051 ACCGCATCGT AGAGGAGCAT CCGTCTTCA CCCGGATCAG CGCCGCGCGT
2101 GCCTTGCGCG ACGCCGCGA GAAGGCTGCC CTCGATGCCG GTGCCGAGAG
2151 GGTCTGAGAG GAGACTGTCC AAGCCCTGCC TGGGCCGGGC TTCGGCGAGA
2201 GGAAGGGAGA GAACCAATGA GCGATCATGC CGTCAACACG CCGGTCCATG
2251 CCGCAGGGC CCACGGGAC CGAGCACACC GTGCCGAGTT CTACGTCTAC
2301 TTCGCCGTC TCTGCTGGG CGCCTTCCCG GTGGCCTTCG TGAGCTGGAT
2351 CGTCTGACG ATCCGCCACC GCAGGCTTCC CAAGCGGGC CCCTTCGCGT
2401 CCGCCTGGTT CGATGCCAAG GCATCACGC CGCTGATTTT CCGCGCCTGA
2451 CCGCAGGTCA GGTGCGACA CGCCATTCGT CGTCTCCCA AGGGCGGGCG
2501 GATTAATCGG GAGGGCATGG TGCCTTACCG TAACCCACGC CACAGCATG
2551 TGGCGACTAG TTCCATATGA TAGATCTTAA
2601 TAA TAGGCCGGCC CTCCGTCGCG GCGGCACCC ACGCCCGC at
2651 CGATTCCAAG GTTCAGCCAT TGAGACGGCT CCGCTTCGCG CGCAAGCGCG
2701 GGTGGGGCCG ACTGCAAGCG GAGAGGGAAG CATGGCACTG CTCAGCTTCG
2751 AGCGAAAATA TCGCGTCCCG GGGGGCACCG TGGTCCGGCG AAACCTGTTC
2801 GACTTCTGGG TCGGCCCTTT CTATGTCGGC TTCTTCGGGG FTGCGACGTT
2851 TTTCTTCGCG GCCCTGGGTA TCATTCATG TGCTGGAGT GCCGTACTCC
2901 AAGGTACCTG GAACCCCAA CTCACTCTG TCTACCCGCC GGCCCTTGAA
2951 TATGGCCTGG GAGGTGCACC CCTCGCAAAA GCGGGCTGT GGCAGATCAT
3001 CACGATCTGC GCCACTGGTG CCTTCGTCAG CTGGGCGCTG CGCGAAGTCG
3051 AAATCTGCCG TAAGCTGGGC ATCGGGTACC ACATCCCGTT GCCTTCGCG
3101 TFCGCCATCC TGGCCTACCT GACGCTGGTG CTGTTCCGCC CGGTGATGAT
3151 GGGCGCCTGG GGCTATGCT TCCCTACGG GATCTGGACG CACTCGACT
3201 GGTGTTCGAA CACGGGCTAC ACCTACGGCA ACTTCCACTA CAACCTGCC
3251 CACATGATGC CCATCTCGTT CTCTTCACG AACCGCTGG CTCTGGCGCT
3301 GCACGGCCCG CTTGTGCTCT CCGGCCCAA CCCCAGAAAG GCAAGGAAA
3351 TCGGAGCGCC GGATCACGAG GATACGTTCT TCCGGGATCT GGTCCGCTAC

```

-continued

3401 TCGATCGGGA CGCTCGGCAT CCACCGCCTC GGCCTGCTGC TCTCGCTGAG
3451 CGCCGTCTTC TTCAGCGCCC TCTGCATGAT CATTACCGGC ACCATCTGGT
3501 TCGATCAGTG GGTGACTGG TGGAATGGT GGGTGAAGCT GCCGTGGTGG
3551 GCGAACATCC CGGGAGGCAT CAATGGCTGA GTATCAGAAC ATCTTCTCCC
3601 AGGTCCAGGT CCGCGGACCG GCCGACCTGG GGATGACCGA AGACGTCAAC
3651 CTGGCCAACC GTTCGGGCGT CGGTCCCTTC TCGACCTGC TCGGTGGTT
3701 CGGCAACGCC CAGCTCGGCC CGATCTATCT CGGCTCGCTC GGCGTCTGT
3751 CCCTCTTCTC GGGCTGATG TGTTCTTCA CCATCGGGAT CTGGTCTTGG
3801 TATCAGGCGG GCTGGAACCC GGCCTCTTC CTGCGCGACC TGTCTTCTT
3851 CTCGCTCGAG CCGCCGGCAC CCGAATACGG TCTGTCTTC GCGGCTCCGC
3901 TGAAGGAAG CGGCTGTGG CTGATCGCGT CGTCTTCAT GTTCGTCGG
3951 GTCTGGTCTT GGTGGGGCCG CACCTATCTC CGCGCTCAGG CGCTGGGCAT
4001 GGGCAAGCAC ACCGCTGGG CGTTCCTCTC GGCCATCTGG CTGTGGATGG
4051 TGCTGGGCTT CATCCGTCCG ATCCTCATGG GGTCTGGTC GGAAGCGGTT
4101 CCCTACGGCA TCTTCTCGCA CCTCGACTGG ACGAACAACT TCTCGCTCGT
4151 CCACGGCAAC CTGTTCTACA ACCCTTCCA CGGTCTCTCG ATCGCCTTCC
4201 TCTACGGTTC GGCCTTGCTC TTCGCGATGC ACGGTGCGAC CATCCTCGCG
4251 GTCTCCCGCT TCGCGGCGGA GCGCAGCTG GAGCAGATCG CCGACCCGGG
4301 GACGGCAGCG GAGCGGGCCG CCCTCTTCTG GCGCTGGACC ATGGGTTTCA
4351 ACGCCACGAT GGAAGGCATC CACCGCTGGG CCATCTGGAT GCGGTCTC
4401 GTGACCTCA CCGGCGGCAT CGGgATCCTG CTCTCGGGCA CGGTCTGGA
4451 CAACTGTGTAC GTCTGGGGCC AGAACACCG CATGGCGCCG CTGAACTGAG
4501 GAGCGATCAC AATGGCTGAC AAGACCATCT TCAACGATCA CCTCAACACC
4551 AATCCGAAGA CCAACCTTCG CCTCTGGGTC GCTTTCCAGA TGATGAAGGG
4601 TCGGGCTGG GCTGGCGCGG TGTTCTTCGG GACGCTCCTT CTCATCGGGT
4651 TCTTCCGGGT GGTGCGGCGG ATGCTTCCGA TCCAGGAGAA CCAGGCTCCG
4701 GCGCCGAACA TCACCGGCGC TCTGGAGACC GGGATCGAGC TGATCAAGCA
4751 TCTCGTCTGA GACAAGTCTC GGGGCGAGGC GCGCGGAGC CGCCCGCTCC
4801 TCCAAGTCCG GGCATATCG CCGCCCGGGG TCCGGGGCGA CACCACAGCC
4851 CGGTTCCTTT CCTGTGGCGG ACAGGGACCT GGTGCCGTGT GGAAGACCGC
4901 ACGGCACCTT TTTGACATTC ACGGGAGGCT CTGATGACCA ATCCACCCCC
4951 GCGACCCGAA ACCCCGCTTT TGATGCGGT CTGCTGCCCC GCCGACATGA
5001 AGGCGCTGAG TGAGCCCGAA CTGGAGCGGC TGGCCGACGA AGTGCCTTCC
5051 GAGTCTAGTG ATAGGGGTAG TTTCTTATTT TAGGCAGTTT ATATGAAATT

5101 AAGACATGCA GATGTCACAG TGGATATTGA ACTGGTCTCG AAAGCTCAAT
5151 ATCCCCAAA GCACAAGCAC AAACCTCGAC ATCATGCAGA AGCGTTTCCC
5201 GAACcgcgctc ttcgcagctgg gcatcgccga gcagcatgcc gtgaccttcg
5251 cggcccgccct cgCcgGGGcc ggGatgaagc ccttctgccc gatctattcc
5301 tcgcttcctgc aacgggggta cgaccagatc gccccatgac tgggcgtgca
5351 gaaccttccc gtcgcttccg tgatcgaccg ggcggggctc gtggggcccg
5401 atggcgcgac ccatgcgggg gccTTCGAGG TTGGCTTCAT CACTTCGCTG
5451 CCCAACATGA CCGTGATGGC CGCGGCCGAC GAGGCCGAGC TCATCCACAT
5501 GATCgcCACC GCCGTGGCCT TCGACGAGGG CCCCATCGCC TTeCGCTTCC
5551 CGCGGGCGGA GGGGGTGGGC GTCGAGATGC CCGAGCGCGG GACCGTGTG
5601 GAGCCCCgCC GGGGCCGCGT GGTGCGCGAA GGGACggatg tgcgcatcct
5651 ctccctCGGC GCGcATCTGC AcGAGgcccT GcAGGcggcG AAACCTTCTC
5701 AGGcccGAGGG GGTGAGCGTG ACCGTGGCCG AcgcccgCTT ctCgCgCCG
5751 CTCgAcACGG GGCATCATGA CCAGCTCGTg cGCCATCAGC CGGCCTGGT
5801 AACGGTGGAG CAGGGGGcA TGGCGGCTT CGGCGCCAT GTCATGCACT
5851 ATCTCGCCAA TTCCGGCGGC TTCGACGGGG GCCTCGCGCT CCGGGTCATG
5901 ACGTGGCCG ACCGCTTCAT CGAGCAGGCG AGCCCCGAGg ACATGTATGC
5951 CGATGCGGGG CTGCGGGCCG AGGATATCAA GCTTGGCGTA ATCATGGTCA
6001 TAGTGTFTTC CTGTGTAAA TTGTATCCG CTCACATTC CACACAACAT
6051 ACGAGCCGGA AGCATAAAGT GTAAAGCCTG GGGTGCCTAA TGAGTGAGCT
6101 AACTCACATT AATTGCGTTG CGCTCACTGC CCGCTTTCCA GTCGGGAAAC
6151 CTGTCGTGCC AGTGCATTA ATGAATCGGC CAACGCGCGG GGAGAGGCGG
6201 TTTGCGTATT GGGCGCTCGG TCTTGCTTGT CTCGTGCGTG ATGTACTTCA
6251 CCAGCTCCGC GAAGTCGCTC TTCTGTATGG AGCGCATGGG GACGTGCTTG
6301 GCAATCACGC GCACCCCGC GCCGTTTTAG CCGCTAAAAA AGTCATGGCT
6351 CTGCCCTCGG GCGGACCACG CCCATCATGA CCTTGCCAAG CTCGTCTTGC
6401 TTCTCTTCGA TCTTCGCCAG CAGGGCGAGG ATCGTGGCAT CACCGAACCG
6451 CGCCGTGCGC GGGTCGTCGG TGAGCCAGAG TTTCAGCAGG CCGCCAGGC
6501 GGCCAGGTC GCCATTGATG GGGCCAGCT CCGGACGTG CTCATAGTCC
6551 ACAGCCCGCG TGATTTGTA GCCCTGCCCC ACGGCCAGCA GGTAGCCGA
6601 CAGGCTCATG CCGGCCCGC CCGCCTTTTC CTCAATCGCT CTTCGTCTGT
6651 CTGGAAGGCA GTACACCTTG ATAGGTGGGC TGCCCTTCTT GGTGGCTTGG
6701 GTTTCATCAG CCATCCGCTT GCCCTCATCT GTTACGCGCG CGGTAGCCGG
6751 CCAGCCTCGC AGAGCAGGAT TCCCCTGAG CACCGCCAGG TGCGAATAAG

-continued

6801	GGACAGTGAA	GAAGGAACAC	CCGCTCGCGG	GTGGGCCTAC	TTCACCTATC
6851	CTGCCCGGCT	GACGCCGTTG	GATACACCAA	GGAAAGTCTA	CACGAACCCT
6901	TTGGCAAAT	CCTGTATATC	GTGCGAAAA	GGATGGATAT	ACCGAAAAAA
6951	TCGCTATAAT	SACCCGAAG	CAGGGTTATG	CAGCGAAAA	GCGCCACGCT
7001	TCCCGAAGGG	AGAAAGGCGG	ACAGGTATCC	GGTAAGCGGC	AGGGTCCGGAA
7051	CAGGAGAGCG	CACGAGGGAG	CTTCCAGGGG	GAAACGCCTG	GTATCTTTAT
7101	AGTCTGTGCG	GGTTTCGCCA	CCTCTGACTT	GAGCGTCGAT	TTTTGTGATG
7151	CTCGTCAGGG	GGGCGGAGCC	TATGGAAAAA	CGCCAGCAAC	GCGGCCTTTT
7201	TACGGTTCCT	GGCCTTTTGC	TGGCCTTTTG	CTCACATGTT	CTTTCCTGCG
7251	TTATCCCCTG	ATTCTGTGGA	TAACCGTATT	ACCGCCTTTG	AGTGAGCTGA
7301	TACCGCTCGC	CGCAGCCGAA	CGACCGAGCG	CAGCGAGTCA	GTGAGCGAGG
7351	AAGCGGAAGA	GCGCCAGAAG	GCCGCCAGAG	AGGCCGAGCG	GCGCCGTGAG
7401	GCTTGGACGC	TAGGGCAGGG	CATGAAAAAG	CCCGTAGCGG	GCTGTACGGG
7451	GCGTCTGACG	CGGTGGAAG	GGGGAGGGGA	TGTTGTCTAC	ATGGCTCTGC
7501	TGTAAGTAGT	GGTTGCGCT	CCGGCAGCGG	TCCTGATCAA	TCGTACCCTT
7551	TTCTCGGTCC	TTCAACGTTT	CTGACAACGA	GCCTCCTTTT	CGCCAATCCA
7601	TCGACAATCA	CCGCGAGTCC	CTGCTCGAAC	GCTGCGTCCG	GACCCGCTTC
7651	GTGCAAGGCG	TCTATCGCGG	CCCGCAACAG	CGCGGAGAGC	GGAGCCTGTT
7701	CAACGGTGCC	GCCGCGCTCG	CCGGCATCGC	TGTCGCCGGC	CTGCTCCTCA
7751	AGCACGGGCC	CAACAGTGAA	GTAGCTGATT	GTCTATCAGG	CATTGACGGC
7801	GTCCCGGCGC	GAAAAAACCC	CCTCGCAGAG	GAAGCGAAGC	TGCGCCTCGG
7851	CCGTTTCCAT	CTGCGGTGCG	CCCGGTGCGG	TGCCGCGATG	CATGCGCGCG
7901	CCATCGCGGT	AGGCGAGCAG	CGCCTGCCTG	AAGCTGCGGG	CATTCCCGAT
7951	CAGAAATGAG	CGCCAGTCTG	CGTCCGCTCT	CGGCACCGAA	TGCGTATGAT
8001	TCTCCGCCAG	CATGGCTTCG	GCCAGTGCCT	CGAGCAGCGC	CCGCTGTGTC
8051	CTGAAGTGCC	AGTAAAGCGC	CGGCTGCTGA	ACCCCAACCC	GTTCGCGCAG
8101	TTTGCCTGTC	GTCAGACCGT	CTACGCCGAC	CTCGTTCAAC	AGGTCAGGGG
8151	CGGCACGGAT	CACGTGATTC	GGCTGCAACT	TTGTCATGCT	TGACACTTTA
8201	TCACTGATAA	ACATAATATG	TCCACCAACT	TATCAGTGAT	AAAGAATCCG
8251	CGCGTTCAAT	CGGACCAGCG	GAGGCTGGTC	CGGAGGCCAG	AcATGAAACC
8301	CAACATACCC	CTGATCGTAA	TTCTGAGCAC	TGTCGCGCTC	GACGCTGTGCG
8351	GCATCGGCCT	GATTATGCCG	GTGCTGCCGG	GCCTCCTGCG	CGATCTGGTT
8401	CACTCGAACG	ACGTCACCCG	CCACTATGGC	ATTCTGCTGG	CGCTGTATGC
8451	GTTGGTGCAA	TTTGCTGCGC	CACCTGTGCT	GGGCGCGCTG	TCGGATCGTT
8501	TCGGCGCGCG	GCCAATCTTG	CTCGTCTCGC	TGGCCGGCGC	CACTGTGACG
8551	TACGCCATCA	TGGCGCAGCG	GCCTTTCCCT	TGGGTCTCTT	ATATCGGGCG
8601	GATCGTGGCC	GGCATCACCG	GGCGACTGGG	GGCGGTAGCC	GCGCTTATA
8651	TTGCCGATAT	CACGTGATGG	GATGAGCGCG	CGCGGCACCT	GCGCTTCATG
8701	AGCGCCTGTT	TCGGGTTCCG	GATGGTCCGG	GGACCTGTGC	TCGGTGGGCT
8751	GATGGGCGGT	TTCTCCCCC	ACGCTCCGTT	CTTCGCCGCG	GCAGCCTTGA
8801	ACGGCCTCAA	TTTCTGACG	GGCTGTTTCC	TTTTCGCCGA	GTGCGACAAA
8851	GGCGAACGCC	GGCCGTTACG	CGGGGAGGCT	CTCAACCCCG	TCGCTTCGTT
8901	CCGTTGGGCC	CGGGGCATGA	CCGTCGTCGC	CGCCCTGATG	GCGGTCTTCT
8951	TCATCATGCA	ACTTGTCCGA	CAGGTGCCGG	CCGCGCTTTG	GGTCAATTTT
9001	GGCGAGGATC	GCTTTCACCTG	GGACGCGACC	ACGATCGGCA	TTTCGCTTGC
9051	CGCATTTGGC	ATTCTGCATT	CACCTCGCCA	GGCAATGATC	ACCGGCCCTG
9101	TAGCCGCCCG	GCTCGGCGAA	AGGCGGGCAC	TCATGCTCGG	AATGATTGCC
9151	GACGGCACAG	GCTACATCCT	GCTTGCTTTC	GCGACACGGG	GATGGATGGC
9201	GTTCCCGATC	ATGGTCTGTC	TTGCTTCGGG	TGGCATCGGA	ATGCCGGCGC
9251	TGCAAGCAAT	GTTGTCCAGG	CAGGTGGATG	AGGAACGTCA	GGGGCAGCTG
9301	CAAGGCTCAC	TGGCGCGCCT	CACCAGCCTG	ACCTCGATCG	TCGGACCCCT
9351	CCTCTTACCG	GCGATCTATG	CGGCTTCTAT	AACAACGTGG	AACGGGTGGG
9401	CATGGATTGC	AGGCGCTGCC	CTCTACTTGC	TCTGCCTGCC	GGCGTGGCT
9451	CGCGGCTTTT	GGAGCGGCGC	AGGCAACGA	GCCGATCGCT	GATCGTGGAA
9501	ACGATAGGCC	TATGCCATGC	GGTCAAGGC	GACTTCCGGC	AAGCTATACG
9551	CGCCCTAGGA	GTGCGGTTGG	AACGTTGGCC	CAGCCAGATA	CTCCCGATCA
9601	CGAGCAGGAC	GCCGATGATT	TGAAGCGCAC	TCAGCGTCTG	ATCCAAGAAC
9651	AACCATCCTA	GCAACACGGC	GGTCCCGGGG	CTGAGAAAGC	CCAGTAAGGA
9701	AACAACCTGA	GGTTCGAGTC	GCGAGATCCC	CCGGAACCAA	AGGAAGTAGG
9751	TTAAACCCCG	TCCGATCAGG	CCGAGCCAGC	CCAGGCCGAG	AACATTGGTT
9801	CCTGTAGGCA	TCGGGATTTG	CGGATCAAAC	ACTAAAGCTA	CTGGAACGAG
9851	CAGAAGTCCT	CCGCCCCCCA	GTTGCCAGGC	GGTAAAGGTG	AGCAGAGGCA
9901	CGGGAGGTTG	CCACTTGC GG	GTCAGCACGG	TTCCGAACGC	CATGGAACCC
9951	GCCCCCGCCA	GGCCCCGCTG	GACGCCGACA	GGATCTAGCG	CTGCGTTTGG
10001	TGTCAACACC	AACAGCGCCA	CGCCCGCAGT	TCCGCAATA	GCCCCAGGA
10051	CCGCCATCAA	TCGTATCGGG	CTACCTAGCA	GAGCGGCAGA	GATGAACACG
10101	ACCATCAGCG	GCTGCACAGC	GCCTACCGTC	GCCCGGACCC	GCCCCGCGAG
10151	GCGGTAGACC	GAAATAAACA	ACAAGCTCCA	GAATAGCGAA	ATATTAAGTG

-continued

10201	CGCCGAGGAT	GAAGATGCGC	ATCCACCAGA	TTCCCGTTGG	AATCTGTCGG
10251	ACGATCATCA	CGAGCAATAA	ACCCGCCGGC	AACGCCCGCA	GCAGCATACC
10301	GGCGACCCCT	CGGCCTCGCT	GTTCCGGGCTC	CACGAAAACG	CCGGACAGAT
10351	GCGCCTTG TG	AGCGTCCTTG	GGGCCGTCTC	CCTGTTTGAA	GACCGACAGC
10401	CCAATGATCT	CGCCGTCGAT	GTAGGCGCCG	AATGCCACGG	CATCTCGCAA
10451	CCGTTCAAGC	AACGCCCTCA	TGGGCTTTTT	CTCCTCGTGC	TCGTAACGG
10501	ACCCGAACAT	CTCTGGAGCT	TTCTTCAGGG	CCGACAATCG	GATCTCGCGG
10551	AAATCCTGCA	CGTCGGCCGC	TCCAAGCCGT	CGAATCTGAG	CCTTAATCAC
10601	AATFGTCAAT	TTTAATCCTC	TGTTTATCGG	CAGTTCGTAG	AGCGCCCGGT
10651	GCGTCCCAG	CGATACTGAG	CGAAGCAAGT	GCGTCGAGCA	GTGCCCGCTT
10701	GTTCTCGAAA	TGCCAGTAAA	GCGCTGGCTG	CTGAACCCCC	AGCCGGAACT
10751	GACCCACAAA	GGCCCTAGCG	TTTGCAATGC	ACCAGTTCAT	CATTGACCCA
10801	GGCGTGTTC	ACCAGGCCGC	TGCCCTCGCAA	CTCTTCGCAG	GCTTCGCCGA
10851	CCTGCTCGCG	CCACTTCTTC	ACGCGGGTGG	AATCCGATCC	GCACATGAGG
10901	CGGAAGGTTT	CCAGCTTGAG	CGGGTACGGC	TCCCGGTGCG	AGCTGAAATA
10951	GTCGAAACATC	CGTCGGGCCG	TCGGCGACAG	CTTGCGGTAC	TTCTCCATA
11001	TGAATTTTCGT	GTAGTGGTCCG	CCAGCAAACA	GCACGACGAT	TTCTCTCGTC
11051	ATCAGGACCT	GGCAACGGGA	CGTTTTCTTG	CCACGGTCCA	GGACCGGAA
11101	GCGGTGCAGC	AGGCACACCG	ATTCCAGGTG	CCCAACGCGG	TCGGACGTGA
11151	AGCCCATCGC	CGTCGCCTGT	AGGCGCGACA	GGCATTCCTC	GGCCTTCTGT
11201	TAATACCGGC	CATTGATCGA	CCAGCCAGG	TCCTGGCAA	GCTCGTAGAA
11251	CGTGAAGGTG	ATCGGCTCGC	CGATAGGGGT	GCGTTCGCG	TACTCCAACA
11301	CCTGCTGCCA	CACCAGTTCG	TCATCGTCGG	CCCGCAGCTC	GACGCGGTG
11351	TAGTGATCT	TCACGTCCTT	GTTGACGTGG	AAAATGACCT	TGTTTTGCAG
11401	CGCCTCGCGC	GGGATTTTCT	TGTTGCGCGT	GGTGAACAGG	GCAGAGCGGG
11451	CCGTGCTGCT	TGGCATCGCT	CGCATCGTGT	CCGGCCACGG	CGCAATATCG
11501	AAACAAGGAAA	GCTGCATTTT	CTTGATCTGC	TGCTTCGTGT	GTTTCAGCAA
11551	CGCGGCTGTC	TTGGCCTCGC	TGACCTGTTT	TGCCAGGTTC	TCGCCGCGG
11601	TTTTTTCGCTT	CTTGGTCTGC	ATAGTTCCCTC	GCGTGTGCAT	GGTCATCGAC
11651	TTTCGCAAAC	CTGCCGCCTC	CTGTTGAGAG	CGACGCGAAC	GCTCCACGGC
11701	GGCCGATGGC	GCGGGCAGGG	CAGGGGGAGC	CAGTTGCACG	CTGTCCGCGT
11751	CGATCTGGC	CGTAGCTTGC	TGGACCATCG	AGCCGACGGA	CTGGAAGGTT
11801	TCGCGGGGCG	CACGCATGAC	GGTGGCGCTT	GCGATGGTTT	CGGCATCCTC
11851	GGCGGAAAAC	CCCGCCTCGA	TCAGTTCTTG	CCTGTATGCC	TTCCGGTCAA
11901	ACGTCCGATT	CATTCACCCCT	CCTTGCGGGA	TTGCCCCGAC	TCACGCCGGG
11951	GCAATGTGCC	CTTATTCCTG	ATTTGACCCG	CCTGGTGCCT	TGGTGTCCAG
12001	ATAATCCACC	TTATCGGCAA	TGAAGTCGGT	CCCGTAGACC	GTCTGGCCGT
12051	CCTTCTCGTA	CTTGGTATTC	CGAATCTTGC	CCTGCACGAA	TACCAGCGAC
12101	CCCTTGCCCA	AACTACTTGC	GTGGCCTCG	GCCTGAGAGC	CAAAACACTT
12151	GATGCGGAAG	AAGTCGGTGC	GCTCCTGCTT	GTCGCCGGTC	GTGGCCGCGC
12201	CAACCTTTGC	GATCCGCAAG	CGCGCGTTCG	CCATCTTCAC	GCTGGAACAG
12251	TACGTCGAGG	CGGCATCAT	GACCCGCGAG	CAATACGAGG	TCATTAAG
12301	CGCCGTGATT	GATGATATAG	CGGCCCGGCT	GCTCCTGGTT	CTCGCGCAC
12351	GAAATGGGTG	ACTTACACCC	GCGCTCTTTG	ATCGTGGCAC	CGATTTCCGC
12401	GATGCTCTCC	GGGAAAAGC	CGGFTTGTTC	GGCCGTCCCG	GGTGTGTGCG
12451	GATCTTCGTC	GATCAGGTCC	AGGTCCAGCT	CGATAGGGCC	GGAAACGCC
12501	TGAGACCGCC	CAGGAGCGTC	CAGGAGGCTC	GACAGTTCGC	CGATGCTATC
12551	CAACCCAGG	CCGGACGGCT	GCGCCGCGCC	TGCCGCTTCC	TGAGCGCCCG
12601	CAGCGGTGTT	TTTTCTGGTG	GTCTTGGCTT	GAGCCGCAGT	CATTGGGAAA
12651	TCTCCATCTT	CGTGAACACG	TAATCAGCCA	GGCGCGAAC	CTCTTTCGAT
12701	GCCTTGCCTG	CGGCCGTTTT	CTTGATCTTC	CAGACCGCA	CACCGGATGC
12751	GAGGGCATCG	GCGATGCTGC	TGCGCAGGCC	AACGGTGGCC	GGAAATCATCA
12801	TCTTGGGGTA	CGCGCCAGC	AGCTCGGCTT	GGTGGCGCGC	GTGGCGCGGA
12851	TTCCGCGCAT	CGACCTTGCT	GGGCACCATG	CCAAGGAATT	GCAGCTTGGC
12901	GTTCTTCTGG	CGCACGTTCC	CAATGGTCTG	GACCATCTTC	TTGATGCCCT
12951	GGATGCTGTA	CGCTCAAGC	TCGATGGGGG	ACAGCACATA	GTCCGCGCGG
13001	AAGAGGGCGG	CCGCCAGGCC	GACGCCAAGG	GTCCGGCCCG	TGTCGATCAG
13051	GCACACGTCG	AAGCCTTGGT	TCGCCAGGGC	CTTGATGTTT	GCCCGGAACA
13101	GCTCGCGGGC	GTCGTCAGC	GACAGCCGTT	CGCGCTTCGC	CAGTACCGGG
13151	TTGGACTCGA	TGAGGGCGAG	GCGCGCGGCC	TGGCCGTTCG	CGGCTGCGGG
13201	TGCGGTTTCG	GTCACGCGC	CGGCAGGGAC	AGCGCCGAAC	AGCTTGTCTG
13251	CATGCAGGCC	GTTAGCAAAG	TCCTTGAGCG	TGTAGGACCG	ATTGCCCTGG
13301	GGGTCCAGGT	CGATCACGGC	AACCCGCAAG	CCGCGCTCGA	AAAAGTCGAA
13351	GGCAAGATGC	ACAAGGGTCG	AAGTCTTGCC	GACGCCGCTT	TTCTGGTTGG
13401	CCGTGACCAA	AGTTTTTCATC	GTTTTGTTTC	CTGTTTTTTT	TTGGCGTCCG
13451	CTTCCCACCT	CCGGACGATG	TACGCCCTGAT	GTTCGGGCAG	AACCCGCGTT
13501	ACCCGCGCGT	ACCCCTCGGG	CAAGTTCTTG	TCCTCGAACG	CGGCCACAC
13551	GCGATGCACC	GCTTGCAGCA	CTGCCGCCCT	GGTCAGTCCC	AGCGACGTTG
13601	GAAAGCTCGC	CTGTGGCTTC	CCATCGACTA	AGACGCCCGG	CGTATCTCG
13651	ATGGTCTGCT	GCCCACTTC	CAGCCCTCGG	ATCGCCTCCT	GGAACTGGCT
13701	TTCCGTAAGC	CGTTTCTTCA	TGATAACAC	CCATAATTTG	CTCCGCGCCT
13751	TGGTTGAACA	TAGCGGTGAC	AGCCGCGCAG	ACATGAGAGA	AGTTTAGCTA
13801	AACATTTCTC	GCACGTCAAC	ACCTTTAGCC	GCTAAAACCT	GTCTTGGCG
13851	TAACAAAACA	AAAGCCCGGA	AACCGGGCTT	TCGTCCTTTG	CCGCTTATGG
13901	CTCTGCACCC	GGTCCATCA	CCAACAGGTC	GCGCACGCGC	TTCACTCGGT
13951	TGCGGATCGA	CACTGCCAGC	CCAACAAAGC	CGGTTGCCCG	CGCCGCCAGG
14001	ATCGCGCCGA	TGATGCCCGC	CACACCGGCC	ATCGCCACCC	AGGTGCGCGC
14051	CTTCCGGTTC	CATTCCTGCT	GGTACTGCTT	CGCAATGCTG	GACCTCGGTA
14101	CACCATAGGC	TGACCGCTCG	ATGGCGTATG	CCGCTTCTCC	CCTTGGCGTA
14151	AAACCCAGCG	CCGACGGCGG	CATTGCCATG	CTGCCCGCCG	CTTTCCCGAC
14201	CACGACGCGC	GCACAGGCTT	TGCGGCTCCG	ACCTTCCGCC	ACGGCGAGCT
14251	GCGCAAGGAC	ATAATCAGCC	GCCGACTTGG	CTCCACGCGC	CTCGATCAGC
14301	TCTTGCACCT	GCGGAAAATC	CTTGGCCCTC	ACGGCCGCCA	TGAATCGCGC
14351	ACGGGGCGAA	GGCTCCCGAG	GGCG		

-continued

Broad-host-range expression vector featuring C-terminal 13 ± His tag

pRKPLHT13Dpuf.seq Length: 14384

```

1  CCACCCAGGC  CGCCGCCCTC  ACTGCCCGGC  ACCTGGTCGC  TGAATGTCGA
51  TGCCAGCACC  TGCGGCACGT  CAATGCTTCC  GGGCGTCGCG  CTCGGGCTGA
101  TCGCCATFCC  CGTFACTGCC  CCGATCCCGG  CAATGGCAAG  GACTGCCAGC
151  GCTGCCATTT  TTGGGGTGAG  GCCGTTCCGG  GCCGAGGGGC  GCAGCCCCTG
201  GGGGGATGGG  AGGCCCGCGT  TAGCGGGCCG  GGAGGGTTTC  AGAAGGGGGG
251  GCACCCCCCT  TCGCGCTGCG  CGGTCACGCG  CACAGGGGCG  AGCCCTGGTT
301  AAAAACAAAG  TTTATAAATA  TTGGTTTAAA  AGCAGGTAA  AAGACAGGTT
351  AGCGGTGGCC  GAAAAACGGG  CGGAAACCC  TGCAAATGTT  GGATTTTCGT
401  CCTGTGGACA  GCCCCTCAA  TGTCAAATAG  TCGCCCCCTC  ATCTGTCAGC
451  ACTCTGCCCC  TCAAGTGTC  AGGATCGCG  CCCTCATGTG  TCAGTAGTCG
501  CGCCCTCAA  GTGTCAATAC  CGCAGGGCAC  TTATCCCCAG  GCTTGTCCAC
551  ATCATCTGTG  GGAACCTCG  GTAAAAATCA  GCGTTTTCGC  CGATTTGCCA
601  GGCTGGCCAG  CTCACCTCG  CCGGCCGAAA  TCGAGCCTGC  CCCTCATCTG
651  TCAACGCCGC  GCCGGTGAG  TCGCCCCCTC  AAGTGTC AAC  FTCCGCCCTC
701  CATCTGTCTG  TGAGGGCCAA  GTTTTCCGCG  AGGTATCCAC  AACGCCGGCG
751  GCCCGGTGT  CTCGCACAG  GCTTCGACGG  CGTTTCTGGC  GCGTTTGCAG
801  GGCCATAGAG  GGCAGCCAG  CCAGCGCCGA  GGCACACCAG  CCGCGTGAGC
851  TCGCGAAAGG  CGCTTCTCC  CTTCCTCGCT  CACTGACTCG  CTGCGCTCGG
901  TCGTTCGGCT  CGCGCAGCG  GTATCAGCTC  ACTCAAAGGC  GGTAATACGG
951  TTATCCACAG  AATCAGGGGA  TAACGCAGGA  AAGAATCATG  GAGCAAAAGG
1001  CCAGCAAAAG  GCCAGGAACC  GTAAAAAGGC  CGCGTTGCTG  GCGTTTTTCC
1051  ATAGGCTCCG  CCCCCGTGAC  GAGCATCACA  AAAATCGACG  CTCAAAGTCAG
1101  AGGTGGCGAA  ACCCGACAGG  ACTATAAAGA  TACCAGGCGT  TTCCCCCTGG
1151  AAGTCTCCCT  GTGCGCTCTC  CTGTTCCGAC  CCTGCCGCTT  ACCGGATACC
1201  TGTCGGCCTT  TCTCCCTTC  GGAAGCGTGG  CGCCATTTCG  CATTCAAGCT
1251  GCGCAACTGT  TGGGAAGGG  GATCGGTGCG  GGCCCTCTTC  CTATTACGCC
1301  AGCTGGCGAA  AGGGGATGT  GCTGCAAGGC  GATTAAGTTG  GGTAACGCCA
1351  GGGTTTTTCC  AGTCAACAG  TTGTAACAG  ACGGCCAGTG  AATTCCGGCC
1401  CGGGCTGGCC  GAGGTGCTGG  GCAAGCCCTA  CCTCCAGGCC  CCCATCGGGG
1451  TCGAGAGCAC  GACCGCCTTC  CTGCGCCGCC  TGGGCAGAT  TCTGGCCCTC
1501  GATCCGGAG  CCTTCATCG  GCAGGAGAG  CACTCGACGC  TGAAGCCCGT
1551  GTGGATCTG  TGGCGGAGT  TCACGCAGGA  CTTCCTCGGG  ACGGCCAATT
1601  TCGGAATCGT  GGGGACCGAA  ACTTATGCAA  GAGGCATCCG  AAATATCTC
1651  GAAGGCATC  TCGGCTGCC  CTGCGCCTTC  GCGGTGCGCG  ACTGATCCGC  CAGCACCGTC
1701  CTGAAAGACC  GACAACGAAG  CGGTGCGCG  ACTGATCCGC  CAGCACCGTC
1751  CGCTCGTCT  CATGGGTGCG  ATCAACGAGA  AGATTTACCT  TGCGGAACCTG
1801  AAAGCCGGTC  ACGGCCCGCA  ACCCTCTTTC  ATCGTGCCTT  CTTTCCCGGG
1851  TCGGGCAGTC  CGGCGCGCTA  CCGAAACGCC  CGTTATGGGA  TATGCGAGTG
1901  CTACGTGGTT  ACTCGAGGAA  GTTTGCAACG  CCTGTTCGA  CGCCCTGTTT
1951  CACATTTCTG  CCTTCGGGAC  GGAGATGGAC  AGCGCCGCGC  CCACACCGAC
2001  GACACTGCGC  CGCGACTTCC  CGTGGGATGC  CGATCGCAA  GCGGCCCTGG
2051  ACCGATCTGT  AGAGGAGCAT  CCGTTTCTCA  CCGGATCAG  CGCCGCGCTG
2101  GCCTTGGCG  ACGCCGCGA  GAAGGCTGCC  CTCGATGCCG  GTGCCGAGAG
2151  GGTCTGTAGA  GAGACTGTGC  AAGCCCTGCC  TGGGCCGGGC  TTCGGCGAGA
2201  GGAAGGGAGA  GAACCAATGA  GCGATCATGC  CGTCAACACG  CCGGTCCATG
2251  CCGCCAGGGC  CCACGGGCAC  CGAGCACACC  GTGCCGAGTT  CTACGTCTAC
2301  TTCGCCGTC  TTCTGTGGG  CGCCTTCCCG  GTGGCCTTCG  TGAGCTGGAT
2351  CGTCTCGAC  ATCCGCCACC  GCAGGCTTCC  CAAGCGCGGC  CCCTTCGCGT
2401  CCGCCTGGT  CGATGCCAAG  GCATCACGC  CGCTGATTTT  CCGCGCCTGA
2451  CCGCAGGTCA  GGTGCGACA  CGCCATTCGT  CGTCTCCCA  AGGGCGGGCG
2501  GATTAATCG  GAGGGCATGG  TGCTTACC  TAACCCACGC  CACAGCATG
2551  TGGCGACTAG  TTCCATATGA  TAGATCTCAG  TAGATCTCAG  TAGATCTCAG
2601  TAGATCTCAG  TAGATCTCAG  TAGATCTCAG  TAGATCTCAG  TAGATCTCAG
2651  CGCCCGC  GATFCCAAG  TTCAGCCATT  GAGACGGCTC  CGCTTCGCCG
2701  GCAAGCGCG  GTTGGGCGA  CTGCAAGCGG  AGAGGGGAGC  ATGGCACTGC
2751  TCAGCTTCGA  GCGAAAATAT  CGCGTGCCGG  GGGGCACGCT  GGTCCGGCGA
2801  AACCTGTTTC  ACTTCTGGGT  CGGCCCTTTC  TATGTGCGGT  TCTTCGGGGT
2851  TCGCAGGTTT  TTCTTCGGG  CCCTGGGTAT  CATCTGAT  GCCTGGAGTG
2901  CCGTACTCCA  GGGTACCTGG  AACCCCAAC  TCATCTCTGT  CTACCCGCGC
2951  GCCCTTGAAT  ATGGCCTGGG  AGGTGCACCC  CTCGCAAAAG  GCGGGCTGTG
3001  GCAGATCAT  ACGATCTGG  CCACTGGTGC  CTTCGTCAGC  TGGGGCTGTC
3051  GCGAAGTCGA  AATCTGCCGT  AAGCTGGGCA  TCGGGTACCA  CATCCCGTTC
3101  GCCTTCGGGT  TCGCCATCCT  GGCCTACCTG  ACGCTGGTGC  TGTTCCGCCC
3151  GGTGATGATG  GCGCCTGGG  GCTATGCCTT  CCCTACGGG  ATCTGGACGC
3201  ACCTCGACTG  GGTTCGAAC  ACGGGTACA  CCTACGGCAA  CTTCCTACTC
3251  AACCTGCC  ACATGATCG  CATCTCGTTC  TTCTTACGA  ACGCGCTGGC
3301  TCTGGCGCTG  CACGGCGCC  TTGTGCTCTC  CGCGGCCAAC  CCGGAGAAGG
3351  GCAAGGAAAT  GCGGACGCC  GATCAGGAG  ATACGTTCTT  CCGCGATCTG

```

-continued

3401 GTCGGCTACT CGATCGGGAC GCTCGGCATC CACCGCCTCG GCCTGCTGCT
3451 CTCGCTGAGC GCCGTCTTCT TCAGCGCCCT CTGCATGATC ATTACCGGCA
3501 CCACTCTGGTT CGATCAGTGG GTCGACTGGT GGCAATGGTG GGTGAAGCTG
3551 CCGTGGTGGG CGAACATCCC GGGAGGCATC AATGGCTGAG TATCAGAACA
3601 TCTTCTCCCA GGTCCAGGTC CGCGGACCGG CCGACCTGGG GATGACCGAA
3651 GACGTCAACC TGGCCAACCG TTCCGGCGTC GTTCCCTTCT CGACCTGTCT
3701 CCGCTGGTTC GGCAACGCC AGCTCGGCC GATCTATCTC GGCTCGCTCG
3751 GCGTCTGTCT CCTTCTTCTG GGCTGATGT GGTCTTCTAC CATCGGGATC
3801 TGGTCTGTCT ATCAGGCGGG CTGGAACCCG GCCGTCTTCC TCGCGGACCT
3851 GTTCTTCTTC TCGCTCGAGC CGCCGGCACC CGAATACGGT CTGTCTTCTG
3901 CCGCTCCGCT GAAGGAAGGC GGGCTGTGGC TGATCGCGTC GTTCTTCTATG
3951 TTCGTCCGGG TCTGGTCTCT GTGGGGCCCG ACCTATCTCC GCGCTCAGGC
4001 GCTGGGCATG GGCAAGCACA CCGCCGTGGC GTTCTTCTCG GCCATCTGGC
4051 TGTGGATGGT GCTGGGCTTC ATCCGTCCGA TCCTCATGGG GTTCTGGTCTG
4101 GAAGCGGTTT CCTACGGCAT CTTCTCGCAC CTCGACTGGA CGAACAACTT
4151 CTCGCTCGTC CACGGCAACC TGTTCTACAA CCCCTTCCAC GGTCTCTCGA
4201 TCGCCTTCCCT CTACGGGTCTG GCCCTGCTCT TCGCGATGCA CCGTCCGACC
4251 ATCTCTCGCG TCTCCCGCTT CGGCGGCGAG CGCGAGCTGG AGCAGATCGC
4301 CGACCCGCGG ACGGCAGCGG AGCGGGCCCG CCTTCTTCTG CGCTGGACCA
4351 TGGGTTTCAA CGCCACGATG GAAGGCATCC ACCGCTGGGC CATCTGGATG
4401 GCGGTCTCTG TGACCTTCCG CGGCGGCATC GGgATCCTGC TCTCGGGCAC
4451 GGTCTGTGGC AACTGGTACG TCTGGGGCCA GAACCAACGGC ATGGCGCCGC
4501 TGAAGTGAAG AGCGATCACA ATGGCTGACA AGACCATCTT CAACGATCAC
4551 CTCAACACCA ATCCGAAGAC CAACCTTCCG CTCTGGGTCTG CTTTCCAGAT
4601 GATGAAGGGT GCGGGCTGGG CTGGCGCGGT GTTCTTCTGG ACCTCTCTTC
4651 TCATCGGGTT CTTCGGGTG CTGCGGCGGA TGCTTCCGAT CCAGGAGAAC
4701 CAGGCTCCGG CGCCGAACAT CACCGGCGCT CTGGAGACCG GGATCGAGCT
4751 GATCAAGCAT CTCTGTCTGAG ACAAGTCTCG GGGCAGGGCG GCGCGAGGCC
4801 GCCCGCTCCT CCAAGTCCGG GCCATATCCG CGGCCCGGGT CCGGGGCGAC
4851 ACCACAGCCC GGTTCCTTCT CTGTGGCGGA CAGGGACCTG GTGCCGTGTG
4901 GAAGACCGCA CGGCACCCTT TTGACATTCA CCGGAGGCTC TGATGACCAA
4951 TCCCACCCCG CGACCCGAAA CCCCGCTTTT GGATCGCGTC TGCTGCCCGG
5001 CCGACATGAA GGGCGTGTGAG GACGCCGAAC TGGAGCGGGT GGCCGACGAA
5051 GTGCGTTCGG AGGTCACTGA TAGGGGTAGT TTCTTATTTT AGGCAGTTTA

5101 TATGAAATTA AGACATGCAG ATGTCACAGT GGATATTGAA CTGGTCTCGA
5151 AAGCTCAATA TCCCCCAAG CACAAGCACA AACTTCCGACA TCATCGAGAA
5201 GCGTTTCCCG Aaccgegtct tcgacgtggg catcgccgag cagcatgccc
5251 tgaccttgcg ggcgggectc gCcgGGGccg gGatgaagcc cttctgcccg
5301 atctattcct cgttctctgca acggggttac gaccagatcg cccatgacgt
5351 gggcgtgcag aaccttcccg tccgcttctg gatcgaccg gggggctcgt
5401 tgggggcccga tggcgcgacc catgcggggg cctTCGACGT TGGCTTCTATC
5451 ACTTCTCTGA CCAACATGAC CGTGTATGGCC GCGGCCGACG AGGCCGAGCT
5501 CATCCACATG ATCgcCACCG CCGTGGCCTT CGACGAGGGC CCCATCGCCT
5551 TcCGCTTCCC GCGGGGCGAG GGGGTGGGCG TCGAGATGCC CGAGCGCGGG
5601 ACGGTGTCTG AGCCCCgCCG GGGCCGCGTG GTGCGCGAAG GGACggatgt
5651 cgggatcctc tccttCGGCG CGcATCTGCA cGAGgctTTG cAGCGgcGA
5701 AACTTCTCGA GgcccGAGGG GTGAGCGTGA CCGTGGCCGA GgcccgCTTc
5751 tCgCgCCCC TCgAcACGGG GCTCATCGAC CAGCTCGTgc GCCATCACGC
5801 GGCCTGGTGA ACGGTGGAGC AGGGGGcAT GGGCGCTTC GGCGCCCATG
5851 TCATGACTA TCTCGCCAAT TCCGGCGGCT TCGACGGGG CCTCGCGCTC
5901 CGGGTATGTA CGCTGCCGA CCGCTTCTATC GAGCAGGGCA GCCCCGAGgA
5951 CATGTATGCC GATGCGGGG TGCGGGCCGA GGATATCAAG CTTGGCGTAA
6001 TCATGGTCTAT AGCTGTTTCC TGTGTGAAAT TGTATCCCG TCACAATTCC
6051 ACACAACATA CGAGCCGGAA GCATAAAGTG TAAAGCCTGG GGTGCCTAAT
6101 GAGTGAGCTA ACTCACATTA ATTGCGTTGC GCTCACTGCC CGCTTTCAG
6151 TCGGAAAC TGTCGTGCCA GCTGCATTA TGAATCGGCC AACCGCGGG
6201 GAGAGGCGGT TTGCGTATTG GCGCTCGGT CTTGCCTTGC TCGTGGGTGA
6251 TGTACTTCC CAGCTCCGCG AAGTCTGCTT TCTGTATGGA GCGCATGGGG
6301 ACGTGTCTGG CAATCACGCG CACCCCCCG CCGTTTTAGC GGCTAAAAA
6351 GTCATGGCTC TGCCCTCGGG CGGACCACCG CCATCATGAC CTTGCCAAGC
6401 TCGTCTGTCT TCTTCTCGAT CTTCGCCAGC AGGGCGAGGA TCGTGGCATC
6451 ACCGAACCG GCCGTGCGCG GGTCTCGGT GAGCCAGAGT TTCAGCAGGC
6501 CGCCAGGCG GCCCAGGTCC CCATTGATGC GGGCCAGCTC GCGGACGTGC
6551 TCATAGTCCA CGACGCCGT GATTTGTAG CCCTGGCCGA GGGCCAGCAG
6601 GTAGCCGAC AGGCTCATGC CGGCCGCGC GCCTTTTCC TCAATCGCTC
6651 TTCGTTCTGT TGGAAAGCAG TACACCTTGA TAGGTGGGCT GCCCTTCTCTG
6701 GTTGGCTTGG TTTTATCAGC CATCCGCTTG CCTCATCTG TTACGCCGGC
6751 GGTAGCCGGC CAGCCTCGCA GAGCAGGATT CCCGTTGAGC ACCGCCAGGT

-continued

6801	GCGAATAAGG	GACAGTGAAG	AAGGAACACC	CGCTCGCGGG	TGGGCCTACT
6851	TCACCTATCC	TGCCCGGCTG	ACGCCGTTGG	ATACACCAAG	GAAAGTCTAC
6901	ACGAACCCTT	TGGCAAATC	CTGTATATCG	TGCGAAAAAG	GATGGATATA
6951	CCGAAAAAAT	CGCTATAATG	ACCCCGAAGC	AGGGTTATGC	AGCGGAAAAG
7001	CGCCACGCTT	CCCGAAGGGA	GAAAGGCGGA	CAGGTATCCG	GTAAGCGGCA
7051	GGGTCGGAAC	AGGAGAGCGC	ACGAGGGAGC	TTCCAGGGGG	AAACGCCCTGG
7101	TATCTTTATA	GTCCTGTCCG	GTTTCGCCAC	CTCTGACTTG	AGCGTCGATT
7151	TTTGTGATGC	TCGTCAGGGG	GGCGGAGCCT	ATGGA AAAAC	GCCAGCAACG
7201	CGGCCTTTTT	ACGGTTCCTG	GCCTTTTGCT	GGCCTTTTGC	TCACATGFTC
7251	TTTCTGCGT	TATCCCTGA	TTCTGTGGAT	AACCGTATTA	CCGCCTTTGA
7301	GTGAGCTGAT	ACCGCTCGCC	GCAGCCGAAC	GACCGAGCGC	AGCGAGTCAG
7351	TGAGCGAGGA	AGCGGAAGAG	CGCCAGAAGG	CCGCCAGAGA	GGCCGAGCGC
7401	GGCCGTGAGG	CTTGAGCGCT	AGGGCAGGGC	ATGAAAAAGC	CCGTAGCGGG
7451	CTGCTACGGG	CGTCTGACGC	GGTGGAAAGG	GGGAGGGGAT	GTTGTCTACA
7501	TGGCTCTGCT	GTAGTGAGTG	GGTTGCGCTC	CGGCAGCGGT	CCTGATCAAT
7551	CGTCACCCTT	TCTCGGTCTT	TCAACGTTCC	TGACAACGAG	CCTCCTTTTC
7601	GCCAATCCAC	CGACAATCAC	CGCGAGTCCC	TGCTCGAACG	CTGCGTCCGG
7651	ACCGGCTTCG	TGGAAGCGCT	CTATCGCGCC	CCGCAACAGC	GGCGAGAGCG
7701	GAGCCTGTTC	AACGGTGCCG	CCGCGCTCGC	CGGCATCGCT	GTCGCCGGCC
7751	TGCTCCTCAA	GCACGGCCCC	AACAGTGAAG	TAGCTGATTG	TCATCAGCGC
7801	ATTGACGGCC	TCCCGGCGG	AAAAACCCGC	CTCGCAGAGG	AAGCGAAGCT
7851	GCGCGTCGGC	CGTTTCCATC	TGCGGTGCGC	CCGGTCCGCT	GCCGCGATGG
7901	ATGCGCGCGC	CATCGCGGTA	GGCGAGCAGC	GCCTGCCTGA	AGCTGCGGGC
7951	ATTCCCGATC	AGAAATGAGC	GCCAGTCTGC	GTCGGCTCTC	GGCACCGAAT
8001	CGGTATGATT	CTCCGCCAGC	ATGGCTTCGG	CCAGTGCCTC	GAGCAGCGCC
8051	CGCTTGTTCC	TGAAGTGCCA	GTAAGCGGCC	GGCTGTGTA	CCCCAACCCG
8101	TTCCGCCAGT	TTGCGGTGCG	TCAGACCCTC	TACGCCGACC	TCGTTCAACA
8151	GGTCACGGCC	GGCACGGATC	ACTGTATTCG	GCTGCAACTT	TGTCATGCTT
8201	GACACTTTAT	CACCTGATAAA	CATAATATGT	CCACCAACTT	ATCAGTGATA
8251	AAGAATCCCG	GCGTTCAATC	GGACCAGCGG	AGGCTGGTCC	GGAGGCCAGA
8301	CaTGA AACCC	AACATACCCC	TGATCGTAAT	TCTGAGCACT	GTCGCGCTCG
8351	ACGCTGTCGG	CATCGGCCCTG	ATTATGCCGG	TGCTGCCGGG	CCTCCTGCGC
8401	GATCTGGTTC	ACTCGAACGA	CGTCACCGCC	CACATGCGCA	TTCTGCTGGC
8451	GCTGTATGCG	TTGGTGCAAT	TTGCCGTCGC	ACCTGTGCTG	GGCGCGCTGT
8501	CGGATCGTTT	CGGGCGGCGG	CCAATCTTGC	TCGTCTCGCT	GGCCGGCGCC
8551	ACTGTGCACT	ACGCCATCAT	GGCGACAGCG	CCTTTCTCTT	GGTTCTCTTA
8601	TATCGGGCGG	ATCGTGGCCG	GCATCACCCG	GGCGACTGGG	GCGGTAGCCG
8651	GCGCTTATAT	TGCCGATATC	ACTGATGGCG	ATGAGCGCGC	GCGGCACTTC
8701	GGCTTCATGA	GCGCCTGTTT	CGGGTTCGGG	ATGGTTCGGG	GACCTGTGCT
8751	CGGTGGGCTG	ATGGGCGGTT	TCTCCCCCCA	CGCTCCGTTT	TTCCGCCGGG
8801	CAGCCTTGAA	CGGCCTCAAT	TTCTGACAGG	GCTGTTTCTT	TTTGCCGGAG
8851	TCGCACAAAG	GCGAACGCCG	GCCGTTACCG	CGGGAGGCTC	TCAACCCGCT
8901	CGCTTCGTTT	CGGTGGGCCC	GGGCATGAC	CGTCGTCGCC	GCCCTGATGG
8951	CGGTCTTCTT	CATCATGCAA	CTTGTCCGGC	AGGTGCCGGC	CGCGCTTTGG
9001	GTCATTTTTC	GCGAGGATCG	CTTTCACTGG	GACCGGACCA	CGATCGGCAT
9051	TTGCTTGGCC	GCAATTTGGCA	TTCTGCATTC	ACTCGCCAG	GCAATGATCA
9101	CCGGCCCTGT	AGCCGCCCGG	CTCGGCGAAA	GGCGGGCACT	CATGCTCGGA
9151	ATGATTGCGG	ACGGCACAGG	CTACATCCTG	CTTGCCTTCG	CGACACGGGG
9201	ATGGATGGCG	TTCCCGATCA	TGTTCTGCTG	TGCTTCGGGT	GGCATCGGAA
9251	TGCCCGGCGT	GCAAGCAATG	TTGTCCAGGC	AGGTGGATGA	GGAACGTCAG
9301	GGCGAGCTGC	AAGGCTCACT	GGCGGCGCTC	ACCAGCCTGA	CCTCGATCGT
9351	CGGACCCCTC	CTCTTACCGG	CGATCTATGC	GGCTTCTATA	ACAACGTTGA
9401	ACGGGTGGGC	ATGGATTGCA	GGCGTGCCTC	TCTACTTGCT	CTGCCTGCCG
9451	GCGCTGCGTC	GCGGCTTTTG	GAGCGGCGCA	GGCAACGAG	CCGATCGCTG
9501	ATCGTGGAAA	CGATAGGCCT	ATGCCATGCG	GGTCAAGGCG	ACTTCCGGCA
9551	AGCTATACGC	GCCCTAGGAG	TGCGGTGGA	ACGTTGGCCC	AGCCAGATAC
9601	TCCCGATCAC	GAGCAGGACG	CCGATGATTT	GAAGCGCACT	CAGCGTCTGA
9651	TCCAAGAAC	ACCATCCTAG	CAACACGGCG	GTCCCGGGC	TGAGAAAGCC
9701	CAGTAAGGAA	ACAACCTGTAG	GTTGAGTCTG	CGAGATCCCC	CGGAACCAAA
9751	GGAAGTAGGT	TAAACCCGCT	CCGATCAGGC	CGAGCCACGC	CAGGCCGAGA
9801	ACATTGGTTC	CTGTAGGCAT	CGGGATTGGC	GGATCAACA	CTAAAGCTAC
9851	TGGAACGAGC	AGAAGTCCCT	CGGCCGCCAG	TTGCCAGGCG	GTAAGGTGA
9901	CGAGAGGCAC	GGGAGGTTGC	CACCTGCGGG	TCAGCACGGT	TCCGAACGCC
9951	ATGGAACCCG	CCCCGCCAG	GCCCGCTGCG	ACGCCGACAG	GATCTAGCGC
10001	TGCCTTTGGT	GTC AACACCA	ACAGCGCCAC	GCCCGCAGTT	CCGCAATATG
10051	CCCCCAGGAC	CGCCATCAAT	CGTATCGGGC	TACCTAGCAG	AGCGGCAGAG
10101	ATGAACACGA	CCATCAGCGG	CTGCACAGCG	CCTACCGTGC	CGCGCACCCC
10151	GCCCGGCAGG	CGGTAGACCG	AAATAAACAA	CAAGCTCCAG	AATAGCGAAA

-continued

10201	TATTAAGTGC	GCCGAGGATG	AAGATGCGCA	TCCACCAGAT	TCCCCTTGG
10251	ATCTGTGCGA	CGATCATCAC	GAGCAATAAA	CCCGCCGGCA	ACGCCCGCAG
10301	CAGCATAACC	GGACCCCTC	GGCCTCGCTG	TTCCGGCTCC	ACGAAAACGC
10351	CGGACAGATG	CGCCTTGTTG	GCCTCCTTGG	GGCCGTCTTC	CTGTTTGAAG
10401	ACCGACAGCC	CAATGATCTC	GCCGTCGATG	TAGGGCCCGA	ATGCCACGGC
10451	ATCTCGCAAC	CGTTCAGCGA	ACGCCCTCCAT	GGGCTTTTTC	TCTCTGTGCT
10501	CGTAAACGGA	CCCGAACATC	TCTGGAGCTT	TCTTCAGGGC	CGACAATCGG
10551	ATCTCGCGGA	AATCCTGCAC	GTCCGGCCGT	CCAAGCCGTC	GAATCTGAGC
10601	CTTAATCACA	ATTGTCAATT	TTAATCCTCT	GTTTATCGGC	AGTTCGTAGA
10651	GCGCGCCGTG	CGTCCCAGC	GATACTGAGC	GAAGCAAGTG	CGTCGAGCAG
10701	TGCCCGCTTG	TTCTGAAAT	GCCAGTAAAG	CGCTGGCTGC	TGAACCCCA
10751	GCCGGAAGT	ACCCACAAG	GCCCTAGCGT	TTGCAATGCA	CCAGGTATC
10801	ATTGACCCAG	GCGTGTCCA	CCAGGCCGCT	GCCTCGCAAC	TCTTCGCAGG
10851	CTTCGCCGAC	CTGCTCGCGC	CACTTCTTCA	CGCGGGTGA	ATCCGATCCG
10901	CACATGAGGC	GGAAGGTTTC	CAGCTTGAGC	GGGTACGGCT	CCCCTGCGA
10951	GCTGAAATAT	TCGAACATCC	GTCCGGCCGT	CGCGACAGC	TTCCGGTACT
11001	TCTCCCATAT	GAATTTCTGT	TAGTGGTCGC	CAGCAAAACAG	CACGACGATT
11051	TCCTCGTGA	TCAGGACCTG	GCAACGGGAC	GTTTTCTTGC	CACGGTCCAG
11101	GACGCGGAAG	CGGTGCAGCA	GCGACACCGA	TTCCAGGTGC	CCAACCGCGT
11151	CGGACGTGAA	GCCCATCGCC	GTCCCTGTGA	GCGCGACAG	GCATTCCTCG
11201	GCCTTCGTGT	AATACCGGCC	ATTGATCGAC	CAGCCAGGTT	CCTGGCAAAG
11251	CTCGTAGAAC	GTGAAGGTGA	TCGGCTCGCC	GATAGGGGTT	CGCTTCGGCT
11301	ACTCCAACAC	CTGCTGCCAC	ACCAGTTCGT	CATCGTCGGC	CCGACGCTCG
11351	ACGCGCGTGT	AGGTGATCTT	CACGTCTCTG	TTGACGTGGA	AAATGACCTT
11401	GTTTTGACAG	GCCCTCGCGC	GGATTTCTT	GTTCGCGGTG	GTGAACAGGG
11451	CAGAGCGGGC	CGTCTCGTTT	GGCATCGCTC	GCATCGTGTG	CGGCCACGGC
11501	GCAATATCGA	ACAAGGAAAG	CTGCATTTCC	TTGATCTGCT	GCTTCGTGTG
11551	TTTCAGCAAC	GCGCCCTGCT	TGCCTCGCT	GACCTGTTTT	GCCAGTCCCT
11601	CGCCGGCGGT	TTTTTCGGTTC	TTGGTCGTCA	TAGTTCCTCG	CGTGTGATG
11651	GTCATCGACT	TCGCCAAACC	TGCCGCCTCC	TGTTTCGAGAC	GACGCGAAGC
11701	CTCCACGGCG	GCCGATGGCG	CGGGCAGGGC	AGGGGAGGCC	AGTTGCACGC
11751	TGTCGGGCTC	GATCTTGGCC	GTAGCTTGGT	GGACCATCGA	GCCGACGGAC
11801	TGGAAGGTTT	CGCGGGGCGC	ACGCATGAGC	GTCCGGCTTG	CGATGGTTTC
11851	GGCATCCTCG	GCGGAAAACC	CCGCTCGAT	CAGTTCCTTG	CTGTATGCCT
11901	TCCGGTCAA	CGTCCGATTC	ATTACCCCTC	CTTGCGGGAT	TGCCCCGACT
11951	CACGCCGGGG	CAATGTGCC	TTATTCCTGA	TTTGACCCGC	CTGGTGCCTT
12001	GGTGTCCAGA	TAATCCACCT	TATCGGCAAT	GAAGTCGGTC	CCGTAGACCG
12051	TCTGGCCGTC	CTTCTCGTAC	TTGGTATTCC	GAATCTTGCC	CTGCACGAAT
12101	ACCAGCGACC	CCTTGCCCAA	ATACTTGCCG	TGGCCCTCGG	CCTGAGAGCC
12151	AAAACACTTG	ATGCGGAAGA	AGTCCGGTGG	CTCCTGCTTG	TCGCCGGTCC
12201	TGGCCCGGCC	AACCTTTGCG	ATCCGCAAGC	GCGCGGTCCG	CATCTTCACG
12251	CTGGAAACAGT	ACGTCGAGGC	GGGCATCATG	ACCCCGGAGC	AATACGAGGT
12301	CATTAAGAGC	GCCGTGATTG	ATGATATAGC	GGCCCGGCTG	CTCCTGGTTC
12351	TCGCGCACCC	AAATGGGTGA	CTTACCCCGC	CGCTCTTTGA	TCGTGGCACC
12401	GATTTCCGCG	ATGCTCTCCG	GGGAAAAGCC	GGGGTGTGCG	GCCGTCCGGG
12451	GCTGATGCGG	ATCTTCTGTC	ATCAGGTCCA	GGTCCAGCTC	GATAGGGCCG
12501	GAACCCCGCT	GAGACGCCGC	AGGAGCGTCC	AGGAGGCTCG	ACAGGTCCGC
12551	GATGCTATCC	AACCCAGGCC	CGGACGGCTG	CGCCCGCCCT	CGGGCTTCCT
12601	GAGCGGCCGC	AGCGGTGTTT	TTCTTGGTGG	TCTTGGCTTG	AGCCCGAGTC
12651	ATTGGGAAAT	CTCCATCTTC	GTGAACACGT	AATCAGCCAG	GGCGCGAACC
12701	TCTTTCGATG	CCTTGCAGCG	GGCCGTFTTC	TTGATCTTCC	AGACCCGAC
12751	ACCGGATGCG	AGGGCATCGG	CGATGCTGCT	GCGCAGGCCA	ACGGTGGCCG
12801	GAATCATCAT	CTTGGGGTAC	GCGGCCAGCA	GCTCGGCTTG	GTGGCGCGCG
12851	TGGCGCGGAT	TCCGCGCATC	GACCTTGGCTG	GGCACCATGC	CAAGGAAATG
12901	CAGCTTGGCG	TTCTTCTGGC	GCACGTTCCG	AATGGTCCGTG	ACCATCTTCT
12951	TGATGCCCTG	GATGCTGTAC	GCCTCAAGCT	CGATGGGGGA	CAGCACATAG
13001	TCGGCCCGCA	AGAGGGCGGC	CGCCAGGCCG	ACGCCAAGGG	TCGGGGCCGT
13051	GTCGATCAGG	CACACGTCGA	AGCCTTGGTT	CGCCAGGGCC	TTGATGTTTCG
13101	CCCCGAACAG	CTCCGCGGCG	TCGTCCAGCG	ACAGCCGTTT	GGCGTTCGCC
13151	AGTACCGGGT	TGGACTCGAT	GAGGGCGAGG	CGCGCCGCTT	GGCCGTCCGC
13201	GGCTGCGGGT	CGGCTTTCGG	TCCAGCCGCC	GGCAGGGACA	GCGCCGAACA
13251	GCTTGTCTGC	ATGCAGGCCG	GTAGCAAAGT	CCTTGAGCGT	GTAGGACGCA
13301	TTGCCCTGGG	GGTCCAGGTC	GATCACGGCA	ACCCGCAAGC	CGCGCTCGAA
13351	AAAGTCCGAG	GCAAGATGCA	CAAGGGTCCGA	AGTCTTGGCG	ACGCCCGCTT
13401	TCTGGTTGGC	CGTGACCAAA	GTTTTTCATCG	TTTGGTTTCC	TGTTTTTTCT
13451	TGGCGTCCCG	TTCCCATCTC	CGGACGATGT	ACGCCGTATG	TTCCGGCAGA
13501	ACCGCCGTTA	CCCGCGCGTA	CCCTCGGGC	AAGTCTTGT	CCTCGAACGC
13551	GGCCCAACAG	CGATGCACCG	CTTGCAGAC	TGCGCCCTTG	GTGAGTCCCA
13601	GCGACGTTGC	GAACGTCGCC	TGTGGCTTCC	CATCGACTAA	GACGCCCCGC
13651	GCTATCTCGA	TGGTCTGCTG	CCCCACTTCC	AGCCCTGGA	TCGCCTCCTG
13701	GAACTGGCTT	TCGGTAAAGCC	GTTTTCTCAT	GGATAACACC	CATAAATTTG
13751	TCCGCGCCTT	GGTGAACAT	AGCGGTGACA	GCCCGCAGCA	CATGAGAGAA
13801	GTTTAGCTAA	ACATTTCTCG	CACGTCAACA	CCTTTAGCCG	CTAAAATCTG
13851	TCCTTGGCGT	AACAAAACAA	AAGCCCGGGA	ACCGGCTTTT	CGTCTCTTGC
13901	CGCTTATGGC	TCTGCACCCG	GTCCATCAC	CAACAGGTCG	CGCACGGCTT
13951	TCACTCGGTT	GCGGATCGAC	ACTGCCAGCC	CAACAAAGCC	GGTTGCCGCC
14001	GCCGCCAGGA	TCGCGCCGAT	GATGCCGGCC	ACACCCGGCA	TCGCCCAACA
14051	GGTCGCGGCC	TTCCGGTTCC	ATTCCTGCTG	GTACTGCTTC	GCAATGCTGG
14101	ACCTCGGCTC	ACCATAGGCT	GACCGCTCGA	TGGCGTATGC	CGCTTCTCC
14151	CTTGGCGTAA	AACCCAGCGC	CGCAGGCGGC	ATTGCCATGC	TGCCCGCGCC
14201	TTTCCCGACC	ACGACGCGCG	CACCAGGCTT	GCGGTCCAGA	CCTTCGGCCA
14251	GCGGAGCTG	CGCAAGGACA	TAATCAGCCG	CCGACTTGGC	TCCACGCGCC
14301	TCGATCAGCT	CTTGCACCTG	GCGGAAATCC	TTGGCTCCCA	GCGCCGCCAT
14351	GAATCGCGCA	GCGGCGGAA	GCTCCGCGG	GCCG	

-continued

pRKMAHT1Dpuf.seq

```

1 CCACCCAGGC CGCCGCCCTC ACTGCCCGGC ACCTGGTTCG TGAAATGTCGA
51 TGCCAGCACCG TGCGGCACGT CAATGCTTCC GGGCGTCGCG CTCGGGCTGA
101 TCGCCATFCC CGTACTGACC CGGATCCCGG CAATGGCAAG GACTGCCAGC
151 GCTGCCATTT TTGGGGTGAG GCCGTTCCGG GCCGAGGGGC GCAGCCCTTG
201 GGGGGATGGG AGGCCCGCGT TAGCGGGCCG GGAGGGTTCG AGAAGGGGGG
251 GCACCCCCCTC TCGCGCTGCG CGGTCACGCG CACAGGGCGC AGCCCTGGTT
301 AAAAACAAAGG TTTATAAATA TTGGTTTAAA AGCAGGTAA AAGACAGGTT
351 AGCGGTGGCC GAAAACCGGG CGGAAACCCCT TGCAAATGCT GGATTTTCGT
401 CCTGTGGACA GCCCCTCAA GCCCCTCAA TGTCAAATAGG TGCGCCCTC ATCTGTACG
451 ACTCTGCCCC TCAAGTGTCA AGGATCGCGC CCCTCATGTG TCAGTAGTCC
501 CGCCCTCAA GTGTCAATAC CGCAGGGCAC TTATCCCCAG GCTTGTCCAC
551 ATCATCTGTG GGAAACTCGC GTAAAATCAG GCGTTTTCG CGATTTGCGA
601 GGCTGGCCAG CTCACGCTCG CGGGCCGAAA TCGAGCCTGC CCCTCATCTG
651 TCAACGCGCG GCCGGGTGAG TCGGCCCTC AAGTGTCAAC FTCCGCCCTC
701 CATCTGTACG TGAGGGCCAA GTTTTCCGG AGGTATCCAC AACGCCGGCG
751 GCCCGGGTGT CTCGCACAGC GCTTCGACGG CGTTTCTGGC GCGTTTGCAG
801 GGCCATAGAG GGCCGCCAGC CCAGCGCGCA GGGCAACCAG CCCGGTGAGC
851 GTCGGAAAGG CGCTCTTCCG CTTCCTCGCT CACTGACTCG CTGCCTCGG
901 TCGTTCGGCT GCGCGAGCG GTATCAGCTC ACTCAAAGGC GGTAATACGG
951 TTATCCAGAG AATCAGGGGA TAACGCAGGA AAGAATATGT GAGCAAAAGG
1001 CCAGCAAAGG GCCAGGAACC GTAAAAGGC CGCGTTGCTG GCGTTTTC
1051 ATAGGCTCCG CCCCCCTGAC GAGCATCACA AAAATCGAGC CTCAAGTCAG
1101 AGGTGGCGAA ACCCGACAGG ACTATAAAGA TACCAGGCGT TTCCCCCTGG
1151 AAGTCCCTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC
1201 TGTCGGCTT TCTCCCTTCG GGAAGCGTGG CGCCATTCG CATTACGGCT
1251 GCGCAACTGT TGGGAAGGGC GATCGGTGCG GCCTCTTCG CTATTACGCC
1301 AGCTGGCGAA AGGGGATGT GCTGCAAGGC GATTAAGTTG GGTACGCCA
1351 GGGTTTTC CAGTACGAGG TTGTAACAGC ACGGCCAGTG AATTCCGGCG
1401 CCGGCTGGCC GAGGTGCTGG GCAAGCCCTA CCTCCAGGCC CCCATCGGGG
1451 TCGAGAGCAC GACCGCCTTC CTGCGCCGCC TGGGCGAGAT TCTGGGCCTC
1501 GATCCGGAGC CCTTCATCGA GCGCGAGAAG CACTCGACGC TGAAGCCCGT
1551 GTGGATCTG TGGCGGAGT TCACGCAGGA CTTCCTCGG ACGGCCAATT
1601 TCGGAATCTG GGCACCGAA ACTTATGCAA GAGGCATCCG AAATAATCTC
1651 GAAGGCGATC TCGGCTGCC CTGCGCTTC GCCGTGGCC GCAAGAGGGG

1701 CTCGAAGACC GACAACGAAG CGGTGCGCG ACTGATCCGC CAGCACCGTC
1751 CGCTCGTGCT CATGGGGTGC ATCAACGAGA AGATTTACCT TGCGGAACATG
1801 AAAGCCGGTC ACGGCCGCA ACCCTCTTTC ATCGCTGCCT CTTTCCCGGG
1851 TGCGGCGATC CGGCGCGCTA CCGGAACGCC CGTTATGGGA TATGCAGGTG
1901 CTACGTGGTT ACTGCAGGAA GTTTGCAACG CCCTGTTCGA CGCCCTGTTC
1951 CACATTTCTG CCTTCGGGAC GGAGATGGAC AGCGCCGCGC CCACCCGAC
2001 GACACTGCGC CGGCACTTCC CGTGGGATGC CGATGCGCAA GCGGCCCTGG
2051 ACCGCATCTG AGAGGAGCAT CCGTTTCTCA CCGGATCAG GCGCGCGCTG
2101 GCCTTGGCGG ACGCCGCCGA GAAGGCTGCC CTGATGCGG GTGCCGAGAG
2151 GGTGCTGAGA GAGACTGTG AAGCCCTGCG TGGCCCGGGC TTCGGCGAGA
2201 GGAAGGGAGA GAACCAATGA GCGATCATGC CGTCAACACG CCGGTCCATG
2251 CCGCCAGGGC CCACGGGCAC CGAGCACACC GTGCCGAGTT CTACGCTTAC
2301 TTCGCCGTFCA TTCTGCTGGG CGCCTTCCCG GTGGCCTTCG TGAGCTGGAT
2351 CGTCTGACG ATCCGCCACC GCAGGCTTCC CAAGCGGGC CCCTTCGGCT
2401 CCGCCTGGTT CGATGCCAAG GCGATCACGC CGCTGATTTT CCGCGCTGA
2451 CCGCAGGTCA GGTGCGACA CGCCATTCTG CTCTCCCCA AGGGCGGGG
2501 GATTAATCGG GAGGGCATGG TGCTTACCG TAACCCACGC CACAGCATG
2551 TGCGGACTAG GGGAGGATAG TAATGCTCGT CAAGACGCAC ATCAGGAAGA
2601 TTGGGGTAC GCTGTTCGCC GTGGCTCTGT TCTACGGGT CATCTACATG
2651 CTTTCCAACT CGCTCTTTC GACGCGGCG GCCACCGCG TTGCGGTTCG
2701 TGCGGACGGG AAAGCCCTGA CTAGTTCAT ATGATAGATC TCGCGGGCGG
2751 TAAATAGGC CGGCCCTCCG TCGCGGGCGG CACCCACGCC
2801 CGCATCGATT CCAAGGTTC GGCATTGAGA CCGCTCCGCT TCGCGCGCAA
2851 GCGCGGGTTG GGCCGACTGC AAGCGGAGAG GGAAGCATGG CACTGCTCAG
2901 CTTGAGCGA AAATATCGCG TGCCGGGGG CACGCTGGTC GGCGGAAACC
2951 TGTTGACTT CTGGTCCGG CCTTCTATG TCGGCTTCTT CCGGGTTCG
3001 ACGTTTTTCT TCGCGGCCCT GGTATCATT CTGATGCTT GGAGTCCGCT
3051 ACTCCAGGGT ACCTGGAACC CCAACTCAT CTCTGTCTAC CCGCCGCCCT
3101 TTGAATATGG CCTGGGAGGT GCACCCCTCG CAAAAGGCGG CGTGTGGCAG
3151 ATCATCACGA TCTGCGCCAC TGGTGCCTTC GTCAGTGGGGCTGCGCGA
3201 AGTCGAAATC TGCCGTAAGC TGGGCATCGG GTACCACATC CCGTTCGGCT
3251 TCGGTTTCGC CATCTGGCC TACTGACGC TGGTGTGTT CCGCCGGTGG
3301 ATGATGGGCG CCTGGGGCTA TGCTTCCCC TACGGGATCT GGACGCACCT
3351 CGACTGGGTG TCGAACACGG GTTACACCTA CGGCAACTTC CACTACAACT

```

-continued

3401 CTGCCACAT GATCGCCATC TCCTTCTTCT TCACGAACGC GCTGGCTCTG
3451 GCGGTGCACG GCGCCCTTGT GCTCTCCGCG GCCAACCCCG AGAAGGGCAA
3501 GGAATATGCG ACGCCGATC ACGAGGATAC GTTCTTCCCG GATCTGGTCTG
3551 GCTACTCGAT CCGGACGCTC GGCATCCACC GCCTCGGCCT GCTGTCTCTG
3601 CTGAGCGCCG TCTTCTTTCAG CGCCCTCTGC ATGATCATTG CCGGCACCAT
3651 CTGGTTCGAT CAGTGGGTCC ACTGGTGGCA ATGGTGGGTG AAGCTGCCGT
3701 GGTGGGCGAA CATCCCGGGA GGCATCAATG GCTGAGTATC AGAACATCTT
3751 CTCCAGGTC CAGTCCGCG GACC GGCCGA CCTGGGGATG ACCGAAGACG
3801 TCAACCTGGC CAACCGTTCG GCGTCCGTC CCTTCTCGAC CCTGCTCGGC
3851 TGGTTCGGCA ACGCCAGCT CGCCCGATC TATCTCGGCT CGCTCGGCGT
3901 CCTGTCCCTC TTCTCGGGCC TGATGTGGTT CTTCACCATC GGGATCTGGT
3951 TCTGGTATCA GCGGGGCTGG AACCCGGCCG TCTTCCTGCG CGACTGTTC
4001 TTCTTCTGCG TCGAGCCGCC GGCACCCGAA TACGGTCTGT CCTTCGCGGC
4051 TCCGCTGAAG GAAGCGGGC TGTGGCTGAT CGCGTCTTC TFCATGTTTCG
4101 TCGCGGTCTG GTCCTGGTGG GCGCCACCT ATCTCCGCGC TCAGGCGCTG
4151 GGCATGGGCA AGCACACCCG CTGGGCGTTC CTCTCGGCCA TCTGGCTGTG
4201 GATGGTGTG GGTTCATCC GTCCGATCCT CATGGGTTC TGGTCGGAAG
4251 CCGTTCCTTA CCGCATCTTC TCGCACTCCG ACTGGACGAA CAACTTCTCG
4301 CTCTGTCACG GCAACCTGTT CTACAACCCC TTCCACGGTC TCTCGATCG
4351 CTCTCTTAC GGTTCGGCCC TGCTCTTCCG GATGCACGGT GCGACCATCC
4401 TCGCGGTCTC CCGCTTCGGC GCGGAGCGCG AGCTGGAGCA GATCGCCGAC
4451 CCGGGGACCG CAGCGGAGCG GCGCCCGCTC TTCTGGCGCT GGACCATGGG
4501 TTTCAACGCC ACGATGGAAG GCATCCACCG CTGGGCCATC TGGATGGCGG
4551 TCCTCGTGGC CCTCACCGGC GGCATCGGGA TCCTGTCTTC GGGCACGGTC
4601 GTGGACAAC GTTACGCTCG GGGCCAGAAC CACGGCATGG GCGCGTGA
4651 CTGAGGAGCG ATCACAATGG CTGACAAGAC CATCTTCAAC GATCACCTCA
4701 ACACCAATCC GAAGACCAAC CTTCGCCTCT GGGTCGCTTT CCAGATGATG
4751 AAGGGTCCGG GCTGGGCTGG CCGCGTGTTC TTCGGGACGC TCCTTCTCAT
4801 CCGGTTCCTC CCGGTGGTCC GCGGATGCT TCCGATCCAG GAGAACCAGG
4851 CTCCGGCGCC GAACATCACC GCGCTCTGCG AGACCGGGAT CGAGTGTATC
4901 AAGCATCTCG TCTGAGACAA GTCTCGGGCG AGGGCGGCGC GAGGCGGCC
4951 GCTCCTCAA GTCCGGGCA TATCGCCGCG CCGGGTCCGG GCGACACCA
5001 CAGCCCGGTT CCCTTCTGTT TGGCGACAGG GACCTGGTGC CGTGTGGAAG
5051 ACCGCACGCG ACCCTTTTGA CATTACGCGG AGGCTCTGAT GACCAATCCC

5101 ACCCGCGCAC CCGAAACCCC GCTTTTGGAT CGCGTCTGCT GCCCGCCGA
5151 CATGAAGCGC CTGAGTGACG CCGAACTGGA GCGGCTGGCC GACGAAGTGC
5201 GTTCCGAGGT CAGTGATAGG GGTAGTTTCT TATTTTAGGC AGTTTATATG
5251 AAATTAAGAC ATGCAGATGT CACAGTGGAT ATTGAAGTGG TCTCGAAAGC
5301 TCAATATCCC CCAAGACACA AGCACAACT TCGACATCAT GCAGAAGCGT
5351 TTCCCGAACC GCGTCTTCGA CGTGGGCATC GCGGAGCAGC ATGCGGTGAC
5401 CTTCGGCGCC GGCCTCGCCG GGGCCGGGAT GAAGCCCTTC TGCGCGATCT
5451 ATTCTCGTTC CCTGCAACCG GGTACGACC AGATCGCCCA TGACGTGGCG
5501 CTGCAGAACC TTCCCGTCCG CTTCGTGATC GACCGGGCGG GGCTCGTGGG
5551 GGCCGATGGC GCGACCCATG CCGGGGCTTC CGACGTTGGC TTCATCACTT
5601 CGCTGCCCCA CATGACCGTG ATGGCCGCGG CCGACGAGGC CGAGCTCATC
5651 CACATGATCG CCACCGCGGT GGCCTTCGAC GAGGGCCCCA TCGCCTCCG
5701 CTTCGGCGCG GCGGAGGGGG TGGCGTTCGA GATGCCCGAG CGCGGGACGG
5751 TGCTGGAGCC CCGCGGGGCG CCGTGTGTGC GCGAAGGGAC GGATGTCGCG
5801 ATCTCTCTCT TCGCGCGCA TCTGCACGAG GCCTTGACAG CGCGGAAACT
5851 TCTCGAGGCC GAGGGGGTGA GCGTGACCGT GGCCGACGCC CGCTTCTCGC
5901 GCGCGCTCGA CACGGGGCTC ATCGACAGC TCGTGGCCCA TCACGCGGCG
5951 CTGGTAACGG TGGAGCAGGG GGCCATGGGC GGCTTCGGCG CCCATGTATC
6001 GCACTATCTC GCCAATTCGG GCGGCTTCGA CCGGGGCTTC GCGCTCCGGG
6051 TCATGACGCT GCCCGACCGC TTCATCGAGC AGGCGAGCCC CGAGGACATG
6101 TATGCCGATG CCGGGCTGCG GGCCGAGGAT ATCAAGCTTG GCGTAATCAT
6151 GGTATAGACT GTTCTCTGTG TGAATTTGTT ATCCGCTCAC AATTCCACAC
6201 AACATACGAG CCGGAAGCAT AAAGTGTAAG GCCTGGGGTG CCTAATGAGT
6251 GAGCTAAGTC ACATTAATTG CGTTGCGCTC ACTGCCCGCT TTCCAGTCCG
6301 GAAACCTGTC GTGCCAGCTG CATTAAATGAA TCGGCCAACG CCGGGGAGA
6351 GCGGTTTTCG GTATTGGGCG CTCGGTCTTG CCTTGCTCGT CCGTGTGTA
6401 CTTCACGAGC TCCGCAAGT CGCTCTTCTT GATGGAGCGC ATGGGGAGCT
6451 GCTTGGCAAT CACGCGCAC CCGCGCCGT TTTAGCGGCT AAAAAAGTCA
6501 TGGTCTGCCC CTCGGGCGGA CCACGCCCAT CATGACCTTG CCAAGCTCGT
6551 CCTGCTTCTC TTCGATCTTC GCCAGCAGGG CGAGGATCGT GGCATCACCG
6601 AACCGCGCGC TCGCGGGTCC GTCGGTGGAG CAGAGTTTCA GCAGCGCCCG
6651 CAGGCGGCC AGGTCGCCAT TGATCGGGCG CAGCTCGCGG ACGTGTCTAT
6701 AGTCCACGAC GCCCGTGAAT TTGTAGCCCT GGCCGACGGC CAGCAGGTAG
6751 GCCGACAGGC TCATGCGGCG CCGCGCGCGC TTTTCTCTCA TCCTCTCTCG

-continued

6801 TTCGTCTGGA AGGCAGTACA CCTTGATAGG TGGGCTGCCC TTCCTGGTGTG
6851 GCTTGGTTTC ATCAGCCATC CGCTTGCCCT CATCTGTTAC GCCGGCGGTA
6901 GCCGGCCAGC CTCGCAGAGC AGGATTCGCC TTGAGCACCG CCAGGTGCCA
6951 ATAAGGGACA GTGAAGAAGG AACACCCGCT CGCGGGTGGG CCTACTTCAC
7001 CTATCCTGCC CGGCTGACGC CGTTGGATAC ACCAAGGAAA GTCTACACGA
7051 ACCCTTTGGC AAAATCCTGT ATATCGTGCC AAAAAGGATG GATATAACCGA
7101 AAAAATCGCT ATAATGACCC CGAAGCAGGG TTATGCAGCG GAAAAGCGCC
7151 ACGCTTCCCG AAGGGAGAAA GCGGACAGG TATCCGGTAA GCGGCAGGTT
7201 CGGATACCGA GAGCGCACGA GGGAGCTTCC AGGGGAAAC GCCTGGTATC
7251 TTTATAGTCC TGTCGGGTTT CGCCACCTCT GACTTGAGCG TCGATTTTTG
7301 TGATGCTCGT CAGGGGGCGG GAGCCTATGG AAAAAAGCCA GCAACGCGGC
7351 CTTTTTACGG TTCCTGGCCT TTTGCTGGCC TTTTGCTCAC ATGTTCTTTC
7401 CTGCGTTATC CCCTGATCTT GTGGATAACC GTATTACCGC CTTTGAGTGA
7451 GCTGATACCG CTCGCCGAG CCGAACGACC GAGCGCAGCG AGTCAGTGAG
7501 CGAGGAAGCG GAAGAGCGCC AGAAGCGCCG CAGAGAGGCC GAGCGCGGCC
7551 GTGAGGCTTG GAGGCTAGGG CAGGGCATGA AAAAGCCCGT AGCGGGCTGC
7601 TACGGGCGTC TGACGCGGTG GAAAGGGGGA GGGGATGTTG TCTACATGGC
7651 TCTGCTGTAG TGAGTGGGTT GCGCTCCGGC AGCGGTCCCT ATCAATCGTC
7701 ACCCTTCTCT GGTCTTCAA CGTTCTGAC AACGAGCCTC CTTTTCGCCA
7751 ATCCATCGAC AATCACCAGC AGTCCCTGCT CGAACGCTGC GTCCGGACCG
7801 GCTTCGTCGA AGGCGTCTAT CGCGGCCCGC AACAGCGCGG AGAGCGGAGC
7851 CTGTTCAACG GTGCCGCGCG GCTCGCCGGC ATCGCTGTGCG CCGGCTGCT
7901 CCTCAAGCAC GGGCCCAACA GTGAAGTAGC TGATTGTCAT CAGCGCATTG
7951 ACGGCGTCCC CGGCCGAAAA ACCCGCCTCG CAGAGGAAGC GAAGCTGCGC
8001 GTCGGCCGTT TCCATCTGCG GTGCGCCCGG TCGCGTGCCG GCATGGATGC
8051 GCGGCCATC GCGGTAGGCG AGCAGCGCCT GCCTGAAGCT CCGGCATTC
8101 CCGATCAGAA ATGAGCGCCA GTCGTCGTCG GCTCTCGGCA CCGAATGCGT
8151 ATGATTCTCC GCCAGCATGG CTTCGGCCAG TCGCTCGAGC AGCGCCCGCT
8201 TGTTCCGTAA GTGCCAGTAA AGCGCCGGCT GCTGAACCCC CAACCGTTCC
8251 GCCAGTTTGC GTGTCGTCAG ACCGCTACG CCGACCTCGT TCAACAGGTC
8301 CAGGCGCGCA CGGATCACTG TATTGCGCTG CAACTTTGTC ATGCTTGACA
8351 CTTTATCACT GATAAACATA ATATGTCCAC CAACTTATCA GTGATAAAGA
8401 ATCCGCGCGT TCAATCGGAC CAGCGGAGGC TGGTCGGAG GCCAGACATG
8451 AAACCCAAACA TACCCTGAT ACCCTGATG CHTAATTCG AGCACTGTGCG CGCTCGACCG

8501 TGTCGGCATC GGCCTGATTA TGCCGGTGCT GCCGGGCCTC CTGCGCGATC
8551 TGGTTCACCT GAACGACGTC ACCGCCACT ATGGCATTCT GCTGGCGCTG
8601 TATGCGTTGG TGCAATTTGC CTGCGCACCT GTGCTGGGCG CGCTGTCCGA
8651 TCGTTTCGGG CGGCGGCCAA TCTTGCTCGT CTCGCTGGCC GCGGCCACTG
8701 TCGACTACGC CATCATGGCG ACAGCGCCTT TCCTTTGGGT TCTCTATATC
8751 GGGCGGATCG TGGCCGGCAT CACCGGGCGG ACTGGGGCGG TAGCCGGCGC
8801 TTATATTGCC GATATCACTG ATGGCGATGA GCGCGCGCGG CACTTCGGCT
8851 TCATGAGCGC CTGTTTCGGG TTCGGGATGG TCGCGGGACC TGTGCTCGGT
8901 GGGCTGATGG GCGGTTTCTC CCCCACGCT CCGTTCCTCG CCGCGGCAGC
8951 CTTGAACGGC CTCAAATTTCC TGACGGGCTG TTTCTTTTG CCGGAGTCCG
9001 ACAAGGCGA ACGCCGGCCG TTACGCGGGG AGGCTCTCAA CCGCTCGCT
9051 TCGTTCGGGT GGGCCCGGGG CATGACCGTC GTCCGCGCCC TGATGGCGGT
9101 CTTCTTCACT ATGCAACTTG TCGGACAGGT GCCGGCCGCG CTTTGGGTCA
9151 TTTTCGGCGA GGATCGCTTT CACTGGGACG CGACCAGCAT CGGCATTTCC
9201 CTTGCCGAT TTGGCAATCT GCATTCACTC GCCCAGCAA TGATCACCGG
9251 CCCTGTAGCC GCCCGGCTCG GCGAAAGCGG GGCACTCATG CTCGGAATGA
9301 TTGCCGACGG CACAGGCTAC ATCCTGCTTG CCTTCGCGAC ACGGGGATGG
9351 ATGGCGTTCC CGATCATGGT CCTGCTTGCT TCGGGTGGCA TCGGAATGCC
9401 GCGGCTGCAA GCAATGTTGT CCAGGCAGGT GGATGAGGAA CGTCAGGGCG
9451 AGCTGCAAGG CTCACTGGCG GCGCTACCA GCCTGACCTC GATCGTCCGA
9501 CCCCTCCTCT TCACGGCGAT CTATGCGGCT TCTATAACAA CGTGAACCGG
9551 GTGGGCATGG ATTGCAGGCG CTGCCCTCTA CTTGCTCTGC CTGCCGGCGC
9601 TCGCTCGCGG GCTTTGGAGC GCGCAGGGC AACGAGCCGA TCGCTGATCG
9651 TGAAACGAT AGGCCTATGC CATGCGGGTC AAGGCACACT CCGGCAAGCT
9701 ATACGCGCCC TAGGAGTGGG GTTGAACGT TGGCCAGCC AGATACTCCC
9751 GATCAGAGC AGGACGCCGA TGATTTGAAG CGCACTCAGC GTCGATCCA
9801 AGAACAAACA TCCTAGCAAC ACGGCGGTCC CCGGGCTGAG AAAGCCCAGT
9851 AAGGAAACAA CTGTAGGTTT GAGTCGCGAG ATCCCCCGGA ACCAAAGGAA
9901 GTAGGTTAAA CCCGCTCCGA TCAGGCGGAG CCACGCCAGG CCGAGAACAT
9951 TGGTTCCTGT AGGCATCGGG ATTGGCGGAT CAAACACTAA AGCTACTGGA
10001 ACGAGCAGAA GTCTCCGGC CGCCAGTTGC CAGGCGGTAA AGGTGAGCAG
10051 AGGCAGGGA GGTGCCCCT TCGGGTFCAG CACGGTCCG AACCCATGG
10101 AAACCGCCCC CGCCAGGCC GCTGCGACCG CGACAGGATC TAGCGCTGCG
10151 TTTGGTGTCA ACACCAACAG CGCCACGCC GCAGTTCCCG AAATAGCCCC

-continued

10201	CAGGACCGCC	ATCAATCGTA	TCGGGCTACC	TAGCAGAGCG	GCAGAGATGA
10251	ACACGACCAT	CAGCGGCTGC	ACAGCGCCTA	CCGTGCGCCG	GACCCCGCCC
10301	GGCAGCGCGT	AGACCGAAAT	AAACAACAAG	CTCCAGAATA	GCGAAATATT
10351	AAGTGCGCCG	AGGATGAAGA	TGCGCATCCA	CCAGATTCCC	GTTGGAATCT
10401	GTCGGACGAT	CATCAGGAGC	AATAAACCCG	CCGGCAACGC	CCGAGCAGC
10451	ATACCGCGCA	CCCTCGGCC	TCGCTGTTCG	GGCTCCACGA	AAACCGCCGA
10501	CAGATGCGCC	TTGTGAGCGT	CCTTGGGGCC	GTCTCTCTGT	TTGAAGACCG
10551	ACAGCCCAAT	GATCTCGCCG	TCGATGTAGG	CGCCGAATGC	CACGGCATCT
10601	CGCAACCGTT	CAGCGAACGC	CTCCATGGGC	TTTTTCTCCT	CGTGCCTGTA
10651	AACGGACCCG	AACATCTCTG	GAGCTTCTTT	CAGGGCCGAC	AATCGGATCT
10701	CGCGAAATC	CTGACAGTCG	GCCGCTCCAA	GCCGTCGAAT	CTGAGCCTTA
10751	ATCACAAATG	TCAATTTTAA	TCCTCTGTTT	ATCGGCAGTT	CGTAGAGCGC
10801	CGCGTGCCTC	CCGAGCGATA	CTGAGCGAAG	CAAGTGCGTC	GAGCAGTGCC
10851	CGCTTGTFTC	TGAAATGCCA	GTAAGCGCCT	GGCTGTCTGA	CCCCAGCCG
10901	GAACGACCC	CACAAGGCC	TAGCGTTTGC	AATGCACCAG	GTCATCATTG
10951	ACCGAGCGGT	GTTCCACCAG	GCCGCTGCCT	CGCAACTCTT	CGCAGGCTTC
11001	GCCGACCTGC	TCGCGCCACT	TCCTCACGCG	GGTGGAAATC	GATCCGCACA
11051	TGAGGCGGAA	GGTTCCAGC	TTGAGCGGGT	ACGGCTCCCG	GTGCGAGCTG
11101	AAATAGTCGA	ACATCCGTCG	GGCCGTCCGC	GACAGCTTGC	GGTACTTCTC
11151	CCATATGAAT	TTCTGTAGT	GGTCGCCAGC	AAACAGCAGC	ACGATTTCTT
11201	CGTCGATCAG	GACCTGGCAA	CGGGACGTTT	TCTTGCCACG	GTCCAGGACG
11251	CGGAAGCGGT	GCAGCAGCGA	CACCGATTCC	AGGTGCCCAA	CGCGGTCCGA
11301	CGTGAAGCCC	ATCGCCGTCG	CCGTGAGGCG	CGACAGGCAT	TCCTCGGCC
11351	TCGTGTAATA	CCGGCCATTG	ATCGACCAGC	CCAGTTCCTG	GCAAAGCTCG
11401	TAGAACGTGA	AGGTGATCCG	CTCGCCGATA	GGGGTCCGCT	TCGCTACTC
11451	CAACACCTGC	TGCCACACCA	GTTCGTATC	GTCCGGCCGC	AGCTCGACGC
11501	CGGTGTAGGT	GATCTTCACG	TCCTTGTGTA	CGTGGAAAAT	GACCTTGTTC
11551	TGCAGCGCCT	CGCGCGGGAT	TTTTCTGTTG	CGCGTGGTGA	ACAGGGCAGA
11601	GCGGGCCGTG	TCGTTTGGCA	TCGCTCGCAT	CGTGTCCGGC	CACGGCCGAA
11651	TATCGAACAA	GGAAAGCTGC	ATTTCCCTGA	TCGTCTGCTT	CGTGTGTTTC
11701	AGCAACGCGG	CCTGCTTGGC	CTCGCTGACC	TGTTTTGCCA	GGTCTCTGCC
11751	GGCGGTTTTT	CGCTTCTTGG	TCGTCATAGT	TCCTCGCGTG	TCGATGGTCA
11801	TCGACTTCGC	CAAACCTGCC	GCCTCCTGTT	CGAGACGACG	GAAACGCTCC
11851	ACGGCGCGCG	ATGGCGCGGG	CAGGGCAGGG	GGAGCCAGTT	GCACGCTGTC
11901	GCGTCCGATC	TTGGCCGTAG	CTTGCTGGAC	CATCGAGCCG	ACGGACTGGA
11951	AGGTTTCGCG	GGGCGCACGC	ATGACGCTGC	GGCTTGCATG	GGTTTCGGCA
12001	TCCTCGCGCG	AAAACCCCGC	GTCCGATCAGT	TCTTGCCCTGT	ATGCCTTCCG
12051	GTCAAACGTC	CGATTCATTG	ACCCTCCTTG	CGGGATTGCC	CCGACTCAGC
12101	CCGGGGCAAT	GTGCCCTTAT	TCCTGATTTG	ACCCGCTCGG	TGCCTTGGTG
12151	TCCAGATAAT	CCACCTTATC	GGCAATGAAG	TCGGTCCCGT	AGACCGTCTG
12201	GCCGTCCCTC	TCGTACTTGG	TATTCGGAAT	CTTGCCCTGC	ACGAATACCA
12251	GCGACCCCTT	GCCCAAATAC	TTGCCGTGGG	CCTCGGCCTG	AGAGCCAAAA
12301	CACCTGATGC	GGAAAGATGC	GGTGCCTCC	TGCTTGTCCG	CGTCTGTGGC
12351	CGCGCCAAAC	TTTTCCGATCC	GCAAGCGCGC	GGTCCGATC	TTACGCTGG
12401	AACAGTACGC	CGAGGCGGGC	ATCATGACCC	GCGAGCAATA	CGAGGTCATT
12451	AAAAGCCCGC	TGATGTATGA	TATAGCGGCC	CGGCTGCTCC	TGTTTCTCGC
12501	GCACCGAAT	GGGTGACTTC	ACCCCGCGCT	CTTTGATCGT	GGCACCATT
12551	TCCGCGATGC	TCTCCGGGGA	AAAGCCGGGG	TTGTCGGCCG	TCCGCGGCTG
12601	ATGCGGATCT	TCGTTCGATCA	GGTCCAGGTC	CAGCTCGATA	GGGCCGGAAC
12651	CGCCCTGAGA	CGCCGCAGGA	CGCTCCAGGA	GGCTCGACAG	GTCCCGGATG
12701	CTATCCAACC	CCAGGCGCGA	CGGCTGCGCC	GCGCCGCGG	CTTCTGAGC
12751	GGCCGACGCG	GTGTTTTTCT	TGGTGGTCTT	GGCTTGAGCC	CGAGTCATTG
12801	GGAAATCTCC	ATCTTCGTGA	ACACGTAATC	AGCCAGGGCG	GAACTCTT
12851	TCGATGCCTT	GCGCGCGGCC	GTTTTCTTGA	TCTTCAGAC	CGGCACACCG
12901	GATGCGAGGG	CATCGGCGAT	GCTGCTGCGC	AGGCCAACGG	TGGCCGGAAT
12951	CATCATCTTG	GGGTACGCGG	CCAGCAGCTC	GGCTTGGTGG	CGCGCGTGGC
13001	GCGGATTCGG	CGCATCGACC	TTGCTGGGCA	CCATGCCAAG	GAATTGACGC
13051	TTGGGCTTCT	TCTGGCGCAC	GTTCCGAATG	GTCTGACCA	TCTTCTGTAT
13101	GCCCTGGATG	CTGTACGCC	CAAGCTCGAT	GGGGACAGC	ACATAGTCGG
13151	CCGCGAAGAG	GGCGCCCGCC	AGGCCGACGC	CAAGGTCGCG	GGCCGTGTCG
13201	ATCAGGCACA	CGTCGAAGCC	TTGGTTCGCC	AGGCCCTTGA	TGTTCCGCCC
13251	GAACAGCTCG	CGGGCGTCGT	CCAGCGACAG	CCGTTCCGGG	TTCCGCGATA
13301	CCGGGTTGGA	CTCGATGAGG	GCGAGGCGCG	CGGCTGGCC	GTCCGCGGCT
13351	GCGGTCGCGG	TTTCGGTCCA	GCCGCGGCA	GGGACAGCGC	GAAACAGCTT
13401	GCTTGCATGC	AGGCCGGTAG	CAAAGTCTTT	GAGCGTGTAG	GACGCATTGC
13451	CCTGGGGGTC	CAGGTCGATC	ACGGCAACCC	GCAAGCCCGG	CTCGAAAAAG
13501	TCGAAGCCAA	GATGCACAAG	GGTCGAAGTC	TTGCCGACGC	CGCCTTCTG
13551	GTTGGCCGTG	ACCAAAGTTT	TCATCGTTTG	GTTTCCTGTT	TTTTCTTGGC

-continued

13601 GTCCGCTTCC CACTTCCGGA CGATGTACGC CTGATGTTCC GGCAGAACCG
 13651 CCGTTACCCG CGCGTACCCC TCGGGCAAGT TCTGTCTTC GAACGCGGCC
 13701 CACACCGGAT GCACCGCTTG CGACACTGCG CCCCTGGTCA GTCCACGGGA
 13751 CGTTGCGAAC GTCGCCGTGT GCTTCCATC GACTAAGACG CCCCAGCCTA
 13801 TCTCGATGGT CTGCTGCCCC ACTTCCAGCC CCTGGATCGC CTCCTGGAAC
 13851 TGGGTTTCGG TAAGCCGTTT CTTCATGGAT AACACCCATA ATTTGCTCCG
 13901 CGCCTTGGTT GAACATAGCG GTGACAGCCG CCAGCACATG AGAGAAGTIT
 13951 AGCTAAACAT TTCTCGCAGC TCAACACCTT TAGCCGCTAA AACTCGTCTC
 14001 TGGCGTAACA AAACAAAAGC CCGAAACCG GGCTTTCGTC TCTTGCCGCT
 14051 TATGGCTCTG CACCCGGCTC CATCACCAAC AGGTCGCGCA CGCGCTTCC
 14101 TCGGTTGCGG ATCGACACTG CCAGCCCAAC AAAGCCGGTT GCCGCCCGC
 14151 CCAGGATCGC GCCGATGATG CCGGCCACAC CGGCCATCGC CCACCAGGTC
 14201 GCCGCTTCC GGTTCATTC CTGCTGGTAC TGCTTCGCAA TGCTGGACCT
 14251 CGGCTCACCA TAGGCTGACC GCTCGATGGC GTATGCCGCT TCTCCCTTG
 14301 GCGTAAAACC CAGCGCCGCA GCGGCATTG CCATGCTGCC CGCCGCTTTC
 14351 CCGACCAGA CGCGCGCAC AGGCTTGGC TCCAGACCTT CGGCCACGGC
 14401 GAGTGGCGCA AGGACATAAT CAGCCGCCGA CTTGGCTCCA CGCGCCTCGA
 14451 TCAGCTCTTG CACTCGCGCG AAATCCTTGG CCTCCACGGC CGCCATGAAT
 14501 CGCCACGCGC GCGAAGGCTC CGCAGGCGC

PRKSSHT1DPUF.SEQ

1 CCACCCAGGC CGCGCCCTC ACTGCCGGC ACCTGGTCG TGAATGTCGA
 51 TGCCAGCACC TGGGGCACGT CAATGCTTCC GGGGTCGCG CTCGGGCTGA
 101 TCGCCCATCC CGTTACTGCC CCATCCCGG CAATGGCAAG GACTGCCAGC
 151 GCTGCCATTT TTGGGGTGAG GCCGTTGCGG GCCGAGGGGC GCAGCCCTTG
 201 GGGGGATGGG AGGCCCGCGT TAGCGGGCGG GGAGGGTTCG AGAAGGGGGG
 251 GCACCCCTCC TCGGCGTGCG CGGTCACGCG CACAGGGCGC AGCCCTGGTT
 301 AAAAACAAGG TTTATAAATA TTGGTTTAAA AGCAGGTTAA AAGACAGGTT
 351 AGCGGTGGCC GAAAAACGGG CGGAAACCTT TGCAATGCT GGATTTTCG
 401 CCTGTGGACA GCCCCTCAA TGTCAATAGG TGCGCCCTC ATCTGTGAGC
 451 ACTCTGCCCC TCAAGTGTCA AGGATCGCGC CCCTCATGTG TCAGTAGTCG
 501 CGCCCTCAA GTGTCAATAC CGCAGGGCAC TTATCCCCAG GCTTGTCCAC
 551 ATCATCTGTG GAAACTCGC GTAAAAACAG GCGTTTTCG CGATTTGCGA
 601 GGCTGGCCAG CTCCACGTCG CCGGCCGAAA TCGAGCCTGC CCCTCATCTG
 651 TCAACGCCGC GCCGGGTGAG TCGGCCCTC AAGTGTCAAC GTCCGCCCTC
 701 CATCTGTGAG TGAGGGCCAA GTTTTCCGCG AGGTATCCAC AACGCCGGC
 751 GCCGCGGTGT CTCGCACACG GCTTCGACGG CGTTTCTGGC GCCTTGTGAG
 801 GGCATAGAG GGCGCCAGC CCAGCGCGA GGGCAACCAG CCCGTTGAGC
 851 GTCGAAAGG CGCTCTTCCG CTTCCTGCT CACTGACTCG CTGCGCTCGG
 901 TCGTTCGGCT CGCGCGAGC GTATCAGCTC ACTCAAGGC GGTAAATACG
 951 TTATCCACAG AATCAGGGGA TAACGCAGGA AAGAATGCT GAGCAAAAGG
 1001 CCAGCAAAAG GCCAGGAACC GTAAAAAGGC CGCGTTGCTG GCGTTTTC
 1051 ATAGGCTCCG CCCCCTGAC GAGCATCACA AAAATCGAGC CTCAAGTCAG
 1101 AGGTGGCGAA ACCCGACAGG ACTATAAAGA TACCAGGCGT TTCCCTTGG
 1151 AAGTCCCTC GTGCGCTCTC CTGTCCGAC CCTGCCGCTT ACCGGATACC
 1201 TGTCCGCTT TCTCCCTTCG GGAAGCGTGG CGCCATTCG CATTGAGGCT
 1251 CGCAACTGT TGGGAAGGGC GATCGGTGCG GGCCTCTTCG CTATTACGCC
 1301 AGCTGGCGAA AGGGGATGT GCTGCAAGGC GATTAAGTTG GGTAAACGCC
 1351 GGGTTTTCCC AGTCACGACG TTGTAAAACG ACGGCCAGTG AATTGCGCCG
 1401 CGGGCTGGCC GAGGTGCTGG GCAAGCCCTA CCTCCAGGCC CCCATCGGGG
 1451 TCGAGAGCAC GACCGCTTC CTGCCGCGC TGGGGAGAT TCTGGCCCTC
 1501 GATCCGGAGC CCTTCATCGA GCGCGAGAAG CACTCGACGC TGAAGCCGCT
 1551 GTGGGATCTG TGGCGGAGTG TCACGCAGGA CTTCCTCGGG ACGGCCAATT
 1601 TCGGAATCGT GCGGACCGAA ACTTATGCAA GAGGCATCCG AAACATATCT
 1651 GAAGCGGATC TCGGGCTGCC CTGCGCCTTC GCCGTGGCCC GCAAGAGGGG

-continued

5101 CGGCTGGCCG ACGAAGTGCG TTCCGAGGTC AGTGATAGGG GTAGTTTCTT
5151 ATTTTAGGCA GTTTATATGA AATTAAGACA TGCAGATGTC ACAGTGGATA
5201 TTGAACTGGT CTCGAAAGCT CAATATCCCC CAAAGCACAA GCACAAACTT
5251 CGACATCATG CAGAAGCGTT TCCGAAACCG CGTCTTCGAC GTGGGCATCG
5301 CCGAGCAGCA TGCCGTGACC TTCGCGGCCG GCCTCGCCGG GCGCGGGATG
5351 AAGCCCTTCT GCGCGATCTA TTCTCTGTTT CTGCAACCGG GTTACGACCA
5401 GATCGCCCAT GACGTGGCGC TGCAGAACCT TCCCGTCCGC TTCGTGATCG
5451 ACCGGGGGGG GCTCGTGGGG GCCGATGGCG CGACCCATGC GGGGGCCTTC
5501 GACGTTGGCT TCATCACTTC GCTGCCAAC ATGACCCTGA TGGCCGCGGC
5551 CGAGGAGGCC GAGCTCATCC ACATGATCGC CACCCCGGTG GCCTTCGAGC
5601 AGGGCCCATC CGCTTCCCG TTCCCGCGGG GCGAGGGGGT GGGCGTCGAG
5651 ATGCCCGAGC GCGGGACGGT GCTGGAGCCC GCGCGGGGCC GCGTGGTTCG
5701 CGAAGGGACG GATGTCGCGA TCCTCTCCCT CGGCGCGCAT CTGCACGAGG
5751 CCTTGCAGGC GCGGAAACTT CTCGAGGCCG AGGGGGTGAG CGTACCCTG
5801 GCCGACGCCG GCTTCTCGCG CCCGCTCGAC ACGGGCTCA TCGACCAGCT
5851 CGTGGCCCAT CACGCGGCCG TGATAACGGT GGAGCAGGGG GCCATGGGGC
5901 GCTTCGCGCG CCATGTCATG CACTATCTCG CCAATTCGGG CGGCTTCGAC
5951 GGGGGCCCTCG CGCTCCGGGT CATGACGCTG CCGACCCGCT TCATCGAGCA
6001 GCGGAGCCCG GAGGACATGT ATGCCGATGC GGGGCTGCGG GCCGAGGATA
6051 TCAAGCTTGG CGTAATCATG GTCATAGCTG TTTCTGTGT GAAATTTGTA
6101 TCCGCTCACA ATTCACACA ACATACGAGC CGGAAGCATA AAGTGTAAAG
6151 CCTGGGGTGC CTAATGAGTG AGCTAACTCA CATTAATTGC GTTGCCTCA
6201 CTGCCCCGTT TCCAGTCCGG AAACCTGTGCG TGCCAGCTGC ATTAATGAAT
6251 CCGCCAACGC GCGGGGAGAG GCGGTTTGGC TATTGGGGCG TCGGCTTTGC
6301 CTTGCTCGTC GGTGATGTAC TTCACCAGCT CCGCGAAGTC GCTCTTCTTG
6351 ATGAGCCGCA TGGGACGCTG CTTGGCAATC ACGCGCACCC CCGGCCGTT
6401 TTAGCGGCTA AAAAAGTCAT GGCTCTGCC TCGGGCGGAC CACGCCCATC
6451 ATGACCTTGC CAAGCTCGTC CTGCTTCTCT TCGATCTTCG CCAGCAGGGC
6501 GAGGATCGTG GCATCACCGA ACCGCGCCGT GCGCGGGTCG TCGGTGAGCC
6551 AGAGTTTTCAG CAGGCCGCC AGCGCGCCCA GTTCGCCATT GATGCGGGCC
6601 AGCTCGCGGA CGTGCTCATA GTCCACGACG CCGTGATT TGTAGCCCTG
6651 GCCGACGGCC AGCAGGTAGG CCGACAGGCT CATGCCGGCC GCCGCCGCT
6701 TTTCTCAAT CGCTCTTCGT TCGCTGGAA GGCAGTACAC CTTGATAGGT
6751 GGGCTGCCCT TCCTGGTTGG CTTGGTTTCA TCAGCCATCC GCTTGCCCTC

6801 ATCTGTTACG CCGGCGGTAG CCGGCCAGCC TCGCAGAGCA GGATTCCCGT
6851 TGAGCACCCG CAGGTGCGAA TAAGGGACAG TGAAGAAGGA ACACCCGCTC
6901 CCGGTTGGGC CTACTTACC TATCTTCCC GGCTGACGCC GTTGATACA
6951 CCAAGGAAAG TCTACACGAA CCCTTTGGCA AAATCCTGTA TATCGTGCGA
7001 AAAAGGATGG ATATACCGAA AAAATCGCTA TAATGACCC GAAGCAGGGT
7051 TATGCAGCGG AAAAGCGCCA CGCTTCCC GAAGGAGAAAG GCGGACAGGT
7101 ATCCGGTAAG CCGCAGGGTC GGAACAGGAG AGCGCACGAG GGAGCTTCCA
7151 GGGGAAACG CCTGGTATCT TTATAGTCTT TCGGGTTTC GCCACCTCG
7201 ACTTGAGCGT CGATTTTGT GATGCTCGTC AGGGGGCGG AGCCTATGGA
7251 AAAACGCCAG CAACCGGCC TTTTACGGT TCCTGGCCTT TFGCTGGCCT
7301 TTTGCTACA TGTTCCTTCC TCGTTATCC CCTGATTCTG TGGATAACCG
7351 TATTACCGCC TTTGAGTGAG CTGATACCCG TCGCCGACG CGAACGACCG
7401 AGCGCAGCGA GTCAGTGAGC GAGGAAGCGG AAGAGCGCCA GAAGCCCGCC
7451 AGAGAGGCCG AGCGCGCCG TGAGGCTTGG ACGCTAGGCC AGGGCATGAA
7501 AAAGCCCGTA GCGGGCTGCT ACGGGCTGCT GACCGGGTGG AAAGGGGAG
7551 GGGATGTTGT CTACATGGCT CTGCTGTAGT GAGTGGGTTG CGCTCCGGCA
7601 CCGTCTCGA TCAATCGTCA CCCTTCTCG GTCTTCAAC GTTCTGACA
7651 ACGAGCCTCC TTTTCGCCAA TCCATCGACA ATCACCGCGA GTCCCTGCTC
7701 GAACGCTGCG TCCGGACCGG CTTCTCGGAA GCGCTCTATC GCGGCCCGCA
7751 ACAGCGCGCA GAGCGAGCC TGTTCACCG TGCCGCGCGG CTCGCGGCA
7801 TCGCTGTGCG CCGCCTGCTC CTCAAGCAGC GCCCAACAG TGAAGTAGCT
7851 GATTGTATC AGCGCATGTA CGGCTGCCCC GGCCGAAAAA CCCGCTCGC
7901 AGAGGAAGCG AAGCTGCGCG TCGCCGTTT CCATCTGCGG TCGCCCGGT
7951 CCGTGGCCG CATGGATGCG CCGCCATCG CGGTAGGCGA GCAGCCCTG
8001 CCTGAAGCTG CCGGCATTCC CGATCAGAAA TGAGCGCCAG TCGTCTCGG
8051 CTCTCGGCAC CGAATGCGTA TGATTTCTCC CCAGCATGCC TTCGGCCAGT
8101 GCGTCGAGCA GCGCCCGCTT GTTCTGAAG TGCCAGTAAA GCGCCGGCTG
8151 CTGAACCCCC AACCGTTCCG CCAGTTTCCG TGTCTCAGA CCGTCTACGC
8201 CGACCTCGTT CAACAGGTCC AGGGCGGCAC GGATCACTGT ATTCTGCTG
8251 AACTTTGTCA TGCTTGACAC TTTATCACTG ATAAACATAA TATGTCCACC
8301 AACTTATCAG TGATAAAGAA TCCGCGGCTT CAATCGGACC AGCGGAGGCT
8351 GGTCCGGAGC CCAGACATGA AACCAACAT ACCCTGATC GTAATTCTGA
8401 GCACTGTGCG GCTCGACGCT GTCGGCATCG GCCTGATTAT GCGGGTCTG
8451 CCGGGCCTCC TGGCGATCT GGTTCACCTC AACGACGTCA CCGCCACATA

-continued

8501 TGGCATTCTG CTGGCGCTGT ATGCGTTGGT GCAATTTGCC TGCGCACCTG
8551 TGCTGGGGCG GCTGTCGGAT CGTTTCGGGC GGCGGCCAAT CTGTCTCGTC
8601 TCGCTGGCCG GCGCCACTGT CACTACGCC ATCATGGCGA CAGCGCCTTT
8651 CCTTTGGGTT CTCTATATCG GGCGGATCGT GGCCGGCATT ACCGGGGCGA
8701 CTGGGGCGGT AGCCGGCGCT TATATTGCCG ATATCACTGA TGGCGATGAG
8751 CGCGCGCGGC ACTTCGGCTT CATGAGCGCC TGTTCGGGT TCGGGATGTT
8801 CGCGGGACCT GTGCTCGGTG GGCTGATGGG CGGTTTCTCC CCCCACGTC
8851 CGTTCTTCGC CGCGGCAGCC TTGAACGGCC TCAATTTCTT GACGGCTGT
8901 TTCCTTTTGC CGGAGTCGCA CAAAGGCGAA CGCCGGCCGT TACGCGGGGA
8951 GGCTCTCAAC CCGCTCGCTT CGTTCCGGTG GGCCCGGGG ATGACCGTGC
9001 TCGCCGCCCT GATGGCGGTC TTCTTCATCA TGCAACTTGT CGGACAGGTG
9051 CCGCGCCGCG TTTGGGTCAT TTTCGGCGAG GATCGCTTTC ACTGGGACGC
9101 GACCACGATC GGCATTTTCG TTGCCGATT TGGCATTCTG CATTCACTCG
9151 CCCAGGCAAT GATCACCGGC CCTGTAGCCG CCCGGCTCGG CGAAAGGCGG
9201 GCACTCATGC TCGGAATGAT TGCCGACGGC ACAGGCTACA TCCTGCTTGC
9251 CTTCCGCGACA CGGGGATGGA TGCGTTCGCC GATCATGGTC CTGCTTGCTT
9301 CCGGTGGCAT CGGAATGCCG GCCTGCAAG CAATGTTGTC CAGGCAGGTG
9351 GATGAGGAAC GTCAGGGGCA GCTGCAAGGC TCACTGGCGG CGCTCACCAG
9401 CCTGACCTCG ATCGTCGGAC CCCTCCTCTT CACGGCGATC TATGGCGCTT
9451 CTATAACAAC GTGGAACGGG TGGCATGGA TTGCAGGCGC TGCCCTCTAC
9501 TTGCTCTGCC TGCCGGCGCT GCCTCGCGGG CTTTGGAGCG GCGCAGGGCA
9551 ACAGAGCCGAT CGCTGATCGT GAAAACGATA GGCCATATGC ATGCGGGTCA
9601 AGGGCACTTC CGCAAGCTA TACGCGCCCT AGGAGTGGCG TTGGAACGTT
9651 GGCCAGCCA GATACTCCCG ATCACGAGCA GGACCGCAT GATTTGAAGC
9701 GCACCTCAGC TCTGATCCA GAACAACCAT CCTAGCAACA CGGCGGTCC
9751 CCGGCTGAGA AAGCCAGTA AGAAAACAAC TGTAGTTCG AGTCGCGAGA
9801 TCCCCCGGAA CCAAAGGAA TAGGTTAAAC CCGCTCCGAT CAGGCCGAGC
9851 CACGCCAGGC CGAGAACATT GGTTCCTGTA GGCATCGGGA TTGGCGGATC
9901 AAACACTAAA GCTACTGGA CGAGCAGAAG TCCTCCGGCC GCCAGTTGCC
9951 AGGCGGTAAG GGTGAGCAGA GGCACGGGAG GTTGCACCTT GCGGTCAGC
10001 ACGGTTCCGA ACGCCATGGA AACCCGCCCC GCCAGGCCG CTGCGACGCC
10051 GACAGGATCT AGCGCTGCGT TTGGTGTCAA CACCAACAGC GCCACGCCCG
10101 CAGTTCCGCA AATAGCCCC AGGACCCGCA TCAATCGTAT CGGGCTACCT
10151 AGCAGAGCGG CAGAGATGAA CACGACCATC AGCGGCTGCA CAGCGCTTAC

10201 CGTCGCGCGG ACCCCGCCCC GCAGGCGGTA GACCGAAATA AACAAACAGC
10251 TCCAGAAATAG CGAAATATTA AGTGCGCCGA GGATGAAGAT GCGCATCCAC
10301 CAGATTTCCC TTGGAATCTG TCGGACGATC ATCACGAGCA ATAAACCCGC
10351 CGGCAACGCC CGCAGCAGCA TACCGCGGAC CCCCCTGGCT CGCTGTTGCG
10401 GCTCCACGAA AACCCCGGAC AGATGCGCCT TGTGAGCGTC CTTGGGGCCG
10451 TCCTCCTGTT TGAAGACCGA CAGCCCAATG ATCTCGCCGT CGATTAGGC
10501 CGCGAATGCC ACGGCATCTC GCAACCGTTC AGCGAACGCC TCCATGGGCT
10551 TTTTCTCCTC GTGCTCGTAA ACGGACCCGA ACATCTCTGG AGCTTCTTTC
10601 AGGGCCGACA ATCGGATCTC GCGGAAATCC TGCACGTCGG CCGCTCCAAG
10651 CCGTCAATC TGAGCCTTAA TCACAATTGT CAATTTTAAAT CCTCTGTTTA
10701 TCGGCAGTTC GTAGAGCGCG CCGTGCCTCC CGAGCGATAC TGAGCGAAGC
10751 AAGTGCCTCG AGCAGTCCCC GCTTGTCTCT GAAATGCCAG TAAAGCGCTG
10801 GCTGCTGAAC CCCAGCCGG AACTGACCCC ACAAGGCCCT AGCGTTTGCA
10851 ATGCACCAG TCATCATTGA CCCAGGCGTG TTCCACCAG CCGCTGCCCT
10901 GAACTCTTC GCAGGCTTCG CCGACCTGCT CGCGCCACTT CTTACGCCG
10951 GTGGAATCCG ATCCGCACAT GAGGCGGAAG GTTTCAGCT TGAGCGGGTA
11001 CCGTCCCGG TGCGAGCTGA AATAGTCGAA CATCCGTCGG GCCGTCCGGC
11051 ACAGCTTGGC GTACTTCTCC CATATGAATT TCGTGTAGTG GTCGCCAGCA
11101 AACAGCAGCA CGATTTCTCT GTCGATCAGG ACCTGGCAAC GGGAGCTTTT
11151 CTTGCCACGG TCCAGGACGC GGAAGCGGTG CAGCAGCGAC ACCGATTCCA
11201 GGTGCCCAAC GCGGTCGGAC GTGAAGCCCA TCGCCGTCCG CTGTAGGCGC
11251 GACAGGCATT CCTCGGCCCT CGTGTAAATC CGGCCATTGA TCGACCAGCC
11301 CAGGTCTTGG CAAAGCTCGT AGAACGTGAA GGTGATCGGC TCGCCGATAG
11351 GGTGCGCTT CCGCTACTCC AACACCTGCT GCCACACCAG TTCGTATCG
11401 TCGGCCGCA GCTCGACGCC GGTGTAGGTG ATCTTCAAGT CCTTGTGAC
11451 GTGAAAATG ACCTTGTTTT GCAGCGCCTC GCGCGGGATT TTCTTGTGTC
11501 GCGTGGTGAA CAGGGCAGAG CCGGCCGTGT CGTTTGGCAT CGCTCGCATC
11551 GTGTCCGGCC ACGGCGCAAT ATCGAACAAG GAAAGCTGCA TTTCTTGTAT
11601 CTGTGCTTTC GTGTGTTTCA GCAACGCGGC CTGCTTGGCC TCGCTGACCT
11651 GTTTTGGCAG GTCCTCGCCG GCGGTTTTTC GCTTCTTGGT GCTCATAGTT
11701 CCTCGCGTGT CGATGGTCAT CACTTTCGCC AAACCTGCCG CCTCCTGTTT
11751 GAGACGACGC GAACGCTCCA CGGCGGCCGA TGGCGCGGG GGGCAGGGG
11801 GAGCCAGTTG CACGCTGTCG CGCTCGATCT TGGCCGTAGC TTGCTGGACC
11851 ATCGAGCCGA CGGACTGGAA GGTTCGCGG GCGGCACGCA TGACGGTTCG

-continued

11901 GCTTGCAGAT GTTTCGGCAT CCTCGGCGGA AAACCCCGCG TCATCAGTT
 11951 CTTGCCTGTA TGCCCTCCGG TCAAACGTCC GATTCATTCA CCCTCCTTGC
 12001 GGGATTGCCCG CGACTCACGC CGGGGCAATG TGCCCTTAT CCTGATTTGA
 12051 CCCGCCGTGT GCCTTGTTGT CCAGATAATC CACCTTATCG GCAATGAAGT
 12101 CGGTCCCGTA GACCCTCTGG CCGTCTTCT CACTTGGT ATTCCGAATC
 12151 TTGCCCTGCA CGAATACCAG CGACCCCTTG CCCAAATACT TGCCGTGGGC
 12201 CTCGGCCTGA GAGCCAAAAC ACTTGATGCG GAAGAAGTCG GTGCGCTCCT
 12251 GCTTGTGCGC GGTCTGTGGC GCGCCAACCT TTGCGATCCG CAAGCGCGCG
 12301 GTCGCCATCT TCACGCTGGA ACAGTACGTC GAGGCGGGCA TCATGACCCG
 12351 CAGCAATAC GAGGTCATTA AAAGCGCCGT GATTGATGAT ATAGCGGCC
 12401 GGCTGCTCCT GGTCTCTCGC CACCGAAATG GGTGACTTCA CCCCAGGCTC
 12451 TTTGATCGTG GCACCGATT CCGCGATGCT CTCGGGGGAA AAGCCGGGGT
 12501 TCTCGGCCGT CCGCGGCTGA TCGGATCTT TCGGATCAG GTCCAGGTCC
 12551 AGCTCGATAG GCGCGGAACC GCCTGAGAC GCGCAGGAG CGTCCAGGAG
 12601 GCTCGACAGG TCGCCGATGC TATCCAACCC CAGGCCGAC GGCTGCGCCG
 12651 CGCTGCGGC TTCTGAGCG GCGCAGCGG TGTTTTCTT GGTGCTCTG
 12701 GCTTGAGCCG CAGTCATTGG GAAATCTCCA TCTTCGTGAA CACGTAATCA
 12751 GCCAGGGCGC GAACCTCTTT CGATGCCTTG CCGCGGGCCG TTTTCTTGAT
 12801 CTTCCAGACC GGCACACCG ATGCGAGGGC ATCGGCGATG CTGCTGGCCA
 12851 GGCCAAACGGT GCGCGAATC ATCATCTTGG GGTACGCGGC CAGCAGCTCG
 12901 GCTTGGTGGC GCGCGTGGC CGGATTCGCG GCATCGACCT TGCTGGGCAC
 12951 CATGCCAAGG AATTGCAGCT TGGCTTCTT CTGGCGCACG TTCGCAATGG
 13001 TCGTGACCAT CTTCTTGATG CCCTGGATGC TGTACGCCTC AAGCTCGATG
 13051 GGGGACAGCA CATAGTCGCG CCGAAGAGG GCGGCCGCCA GGCCGACGCC
 13101 AAGGGTGGG GCGGTGTCGA TCAGGCACAC GTCGAAGCCT TGGTTCGCCA
 13151 GGGCCTTGAT GTTCCGCCCC AACAGCTCCG GGGCGTCGTC CAGCGACAGC
 13201 CGTTCGGCGT TCGCCAGTAC CGGTTGGAC TCGATGAGG CGAGCGCGC
 13251 GGCTGCGCC TCGCCGCTC CGGTGGCGT TTCGGTCCAG CCGCCGGCAG
 13301 GGACAGCGCC GAACAGCTTG CTGCGATGCA GGCCGGTAGC AAAGTCCTTG
 13351 AGCGTGTAG ACGCATTGCC CTGGGGTCC AGTTCGATCA CCGCAACCCG
 13401 CAAGCCCGCG TCGAAAAAGT CGAAGGCAAG ATGCACAAGG GTCGAAGTCT
 13451 TGCCGACGCC GCCTTCTGG TTGGCCGTGA CCAAAGTTTT CATCGTTTGG
 13501 TTTCTGTTT TTTCTTGGC TCCGCTTCCC ACTTCCGGAC GATGTACGCC
 13551 TGATGTTCCG GCAGAACCCG CGTTACCCCG GCGTACCCTT CCGGCAAGTT
 13601 CTTGTCTCTG AACGCGGCC ACACGCGATG CACCGCTTGC GACACTGCGC
 13651 CCCTGGTCCG TCCAGCGAC GTTGCGAAGC TCGCCTGTGG CTTCCCATCG
 13701 ACTAAGACCG CCCGCGTAT CTCGATGTC TGCTGCCCA CTTCCAGCCC
 13751 CTGGATCGCC TCCTGGAAC TGGTTTCGGT AAGCCGTTT CTTATGGATA
 13801 ACACCCATAA TTTGCTCCCG GCCTTGGTTG AACATAGCGG TGACAGCCCG
 13851 CAGCACATGA GAGAAGTTTA GCTAAACATT TCTCGCACGT CAACACCTTT
 13901 AGCCGCTAAA ACTCGTCTT GCGTAACAA AACAAAAGCC CGGAAAACCG
 13951 GCTTCTGCTT CTTGCCGCTT ATGGCTCTGC ACCCGGCTCC ATCACCAACA
 14001 GGTGCGGCAC GCGCTTCACT CGGTTGCGGA TCGCACTGC CAGCCCAACA
 14051 AAGCCGGTTG CCGCCGCGC CAGGATCGCG CCGATGATGC CGGCCACACC
 14101 GGCCATCGCC CACCAGTTC CCGCTTCCG GTTCCATTCC TGCTGGTACT
 14151 GCTTCGCAAT GCTGGACCTC GGCTCACCAT AGGCTGACCG CTCGATGGCG
 14201 TATGCCGCTT CTCCCTTGG CGTAAAACCC AGCGCCGACG GCGGCATTGC
 14251 CATGCTGCC CCGCTTTCC CGACCAGAC GCGCGACCA GGCTTGGCGT
 14301 CCAGACCTTC GGCACGCGC AGCTGCGCAA GGACATAATC AGCCGCGGAC
 14351 TTGGCTCCAC GCGCTCGAT CAGCTCTGC ACTCGCGCA AATCCTTGGC
 14401 CTCCAGGCC GCCATGAATC GCACGCGCG CGAAGCTCC GAGGGCCG

DOCUMENTS CITED

- [0230] 1. Ditta, G., Schmidhauser, T., Yakobsen, E., Lu, P., Liang, X.-W., Finlay, D. R., GUiney, D., and Helinski, D. R. (1985). Plasmids related to the broad host range vector, pRK290, useful for gene cloning and for monitoring gene expression. *Plasmid* 13, 149-153.
- [0231] 2. Scott, H. N., Laible, P. D., and Hanson, D. K. (2003). Sequences of versatile broad-host-range vectors of the RK2 family. *Plasmid* 50, 74-79.
- [0232] 3. Antoine, R., and Loch, C. (1992). Isolation and molecular characterization of a novel broad-host-range plasmid from *Bordetella bronchiseptica* with sequence similarities to plasmids from Gram-positive organisms. *Molecular Microbiology* 6, 1785-1799.
- [0233] 4. Kovach, M. E., Phillips, R. W., Elzer, P. H., Roop, R. M., and Peterson, K. M. (1994). pBBR1MCS: Broad host range cloning vector. *Biotechniques* 16, 800-802.
- [0234] 5. Kovach, M. E., Elzer, P. H., Hill, D. S., Robertson, G. T., Farris, M. A., Roop II, R. M., and Peterson, K. M. (1995). Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* 166, 175-176.
- [0235] 6. DeShazer, D., Woods, D. E. (1996). Broad-host-range cloning and cassette vectors based on the R388 trimethoprim resistance gene. *Biotechniques* 20, 762-764.
- [0236] 7. J. K. Lee et al. (1989) Post-transcriptional control of puc operon expression of B800-850 light-harvesting complex formation in *Rhodobacter sphaeroides* J. *Bacteriol.* 171, 3391-3405.
- [0237] 8. J. O. Goldsmith, and S. G. Boxer, (1996) *Biochim. Biophys. Acta*, 1276, 171-175.
- [0238] 9. C. Aslanidis et al., (1990) *Nucleic Acids Res. Ligation-independent cloning of PCR products (LIC-PCR)* 18, 6069-6074
- [0239] 10. Yoon, J. R., et al. (2002) Express primer tool for high-throughput gene cloning and expression. *BioTechniques* 33: 1-5.
- [0240] 11. Pokkuluri et al. (2002) The structure of a mutant photosynthetic reaction center shows unexpected changes in main chain orientations and quinone position. *Biochemistry* 41: 5998-6007.

[0241] 12. H. Myllykallio, F. E. Jenney, Jr., C. R. Moomaw, C. A. Slaughter, and F. Daldal. Cytochrome c_y of *Rhodobacter capsulatus* is attached to the cytoplasmic membrane by an uncleaved signal sequence-like anchor. J. Bacteriol. 179:2623-2631 (1997).

[0242] 13. B. J. MacGregor and T. J. Donohue. Evidence for two promoters for the cytochrome C_2 gene (*cycA*) of *Rhodobacter sphaeroides*. J. Bacteriol. 173:3949-3957 (1991).

[0243] 14. H. Myllykallio, F. E. Jenney, Jr., C. R. Moomaw, C. A. Slaughter, and F. Daldal. Cytochrome c_y of *Rhodobacter capsulatus* is attached to the cytoplasmic membrane by an uncleaved signal sequence-like anchor. J. Bacteriol. 179: 2623-2631 (1997).

[0244] 15. B. J. MacGregor and T. J. Donohue. Evidence for two promoters for the cytochrome C_2 gene (*cycA*) of *Rhodobacter sphaeroides*. J. Bacteriol. 173: 3949-3957 (1991).

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 43

<210> SEQ ID NO 1

<211> LENGTH: 63

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 1

tggaggatcg ccatgcacca ccaccaccac caccacgcta gcgcgggcag atcttgaggc 60

cgg 63

<210> SEQ ID NO 2

<211> LENGTH: 63

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 2

cctcaagatc tgcccgcgct agcgtggtgg tgggtggtgt ggtgcatggc gatcctccac 60

atg 63

<210> SEQ ID NO 3

<211> LENGTH: 84

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 3

tggaggatcg ccatgcacca ccaccaccac caccacgaga acctgtactt ccagtccgct 60

agctagggaa gatcttgagg ccgg 84

<210> SEQ ID NO 4

<211> LENGTH: 84

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 4

cctcaagatc ttccctagct agcggactgg aagtacaggt tctcgtggtg gtgggtggtg 60

tgggtgcatgg cgatcctcca catg 84

-continued

<210> SEQ ID NO 5
<211> LENGTH: 62
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 5

ctagttccat atgatagatc tcaccaccac caccaccacc accaccacca ctaataggcc 60
gg 62

<210> SEQ ID NO 6
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 6

aaggatatact atctagagtg gtggtggtgg tgggtggtgg ggtggtgatt atcc 54

<210> SEQ ID NO 7
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 7

aaccacgccc accagtaggc aggaggaaca cgtgctcgtcc ggtgg 45

<210> SEQ ID NO 8
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 8

gaccaccgga cgacacgtgt tcctcctgcc ta 32

<210> SEQ ID NO 9
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 9

gcctattcca atcctacgta gaaggaaga tc 32

<210> SEQ ID NO 10
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

-continued

<400> SEQUENCE: 10
aagatcttcc cttctacgta ggattggaat agg 33

<210> SEQ ID NO 11
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 11
gagaacctgt acttccaatc ctttacgtag aaataggaa gatc 44

<210> SEQ ID NO 12
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 12
aagatcttcc ctatttctac gtaaaggatt ggaagtacag gttct 45

<210> SEQ ID NO 13
<211> LENGTH: 14386
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
plasmid

<400> SEQUENCE: 13
ccaccaggc cgccgccctc actgcccgc acctggtcgc tgaatgtcga tgccagcacc 60
tgcggcacgt caatgcttcc gggcgctcgc ctcgggctga tcgccatcc cgttactgcc 120
ccgatcccgc caatggcaag gactgccagc gctgccattt ttggggtgag gccgttcgcg 180
gccgaggggc gcagcccctg ggggatggg aggcccgctg tagcgggccg ggagggttcg 240
agaagggggg gcacccccct tcggcgtcgc cggtcacgcg cacagggcgc agccctggtt 300
aaaaacaagg ttataaata ttggtttaa agcaggtaa aagacaggtt agcggtggcc 360
gaaaaacggg cgaaaccct tgcaaatgct ggattttctg cctgtggaca gccctcaaa 420
tgtcaatagg tgcgccctc atctgtcagc actctgccc tcaagtgtca aggatcgcg 480
ccctcatctg tcagtagtcg cccccctcaa gtgtcaatac cgcagggcac ttatccccag 540
gcttgtccac atcatctgtg gaaaactcgc gtaaaatcag gcgttttcgc cgatttgcca 600
ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgccgc 660
gccgggtgag tcggcccctc aagtgtcaac gtccgccctc catctgtcag tgagggccaa 720
gttttccgcg aggtatccac aacgcggcg gccgcggtgt ctcgcacacg gcttcgacgg 780
cgtttctggc gcgtttgcag ggccatagac ggccgccagc ccagcggcga gggcaaccag 840
cccgtgagc gtccgaaagg cgctcttcg cttcctcgt cactgactcg ctgcgctcgg 900
tcgttcggct gcggcgagcg gtatcagctc actcaaaggc ggaataacgg ttatccacag 960
aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc 1020

-continued

gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg cccccctgac gagcatcaca	1080
aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt	1140
ttccccctgg aagctcccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc	1200
tgctcgcctt tctcccttcg ggaagcgtgg cgccattcgc cattcaggct gcgcaactgt	1260
tgggaagggc gatcggtgcg ggcctcttcg ctattacgcc agctggcgaa aggggatgt	1320
gctgcaaggc gattaagtgg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg	1380
acggccagtg aattcggccg cgggctggcc gaggtgctgg gcaagcccta cctccaggcc	1440
cccatcgggg tcgagagcac gaccgccttc ctgcgccgcc tggcgagat tctgggcctc	1500
gatccggagc cttctatcga gcgcgagaag cactcgacgc tgaagcccgt gtgggatctg	1560
tggcggagtg tcacgcagga cttcttcggg acggccaatt tcggaatcgt ggcgaccgaa	1620
acttatgcaa gaggcattccg aaactatctc gaaggcgatc tcgggctgcc ctgcgccttc	1680
gccgtggccc gcaagagggg ctcgaaagac gacaacgaag cggtgccgcg actgatccgc	1740
cagcaccgtc cgctcgtgct catggggtcg atcaacgaga agatttacct tgcggaactg	1800
aaagccggtc acggcccga accctcttcc atcgtgcctt ctttcccggg tgcggcgatc	1860
cggcgcgcta ccggaacgcc cgttatggga tatgcaggtg ctacgtgggtt actgcaggaa	1920
gtttgcaacg ccctgttcga cgcctgttcc cacattctgc ccctcgggac ggagatggac	1980
agcgcgcgcc ccacaccgac gacactgcgc gcgcacttcc cgtgggatgc cgatgcgcaa	2040
gcggccctgg accgcacgtg agaggagcat ccggttctca cccggatcag cgcgcgcgct	2100
gccttgccgc acgcccgcga gaaggtgcc ctctgatgcc gtgccgagag ggtcgtgaga	2160
gagactgtcg aagccctgcg tgggcccggc ttcggcgaga ggaagggaga gaaccaatga	2220
gcgatcatgc cgtcaacacg ccggtccatg ccgccagggc ccacgggac cagcaccac	2280
gtcccgagtt ctacgtctac ttcgcccgtc ttctgtggg cgccttcccg gtggccttcg	2340
tgagctggat cgtctcgacg atcccacc gcaggcttcc caagcgcggc cccttcgctg	2400
ccgctgggtt cgatgccaa ggcgatcacc cgctgatttt ccgcccctga ccgcaggtea	2460
ggttgcgaca cgcattcgt cgtctcccca aggggcccgc gattaatcg gagggcatgg	2520
tgcttaccg taaccaccgc caccagcatg tggaggatcg ccatgcacca ccaccaccac	2580
caccacgcct attccaatcc tacgtagaag ggaagatctt gagggccggc cctccgtcgc	2640
gggggcacc cagcccgcga tcgattccaa ggttcagcca ttgagacggc tccgcttcgc	2700
gcgcaagcgc gggttgggcc gactgcaagc ggagaggaa gcatggcact gctcagcttc	2760
gagcgaaaat atcgcgtgcc gggggcacg ctggtcggcg gaaacctgtt cgacttctgg	2820
gtcggccctt tctatgtcgg cttcttcggg gttgcgacgt ttttcttcgc ggcctgggt	2880
atcattctga ttgcctggag tgccgtactc cagggtacct ggaaccccca actcatctct	2940
gtctaccgc cggcccttga atatggcctg ggaggtgcac ccctcgaaa aggcgggctg	3000
tggcagatca tcacgatctg cgcactgggt gccttcgtca gctgggcgct gcgogaagtc	3060
gaaatctgcc gtaagctggg catcgggtac cacatcccgt tcgccttcgc gttcccatc	3120
ctggcctacc tgacgtggt gctgttccgc ccggtgatga tgggcccctg gggctatgcc	3180
ttcccctacg gatctggac gcacctcgac tgggtgtcga acacgggcta cacctacggc	3240
aacttccact acaacctcgc ccacatgatc gccatctcgt tcttcttcac gaacgcgctg	3300

-continued

gctctggcgc tgcacggcgc ccttgtgctc tccgcggcca accccgagaa gggcaaggaa	3360
atgcgacgc cggatcacga ggatacgttc ttccgcgac tggtcggcta ctcgatcggg	3420
acgctcggca tccaccgcct cggcctgctg ctctcgtga gcgccgtctt cttcagcgc	3480
ctctgcatga tcattaccgg caccatctg ttcgatcagt gggtcgactg gtggcaatgg	3540
tgggtgaagc tgccgtggtg ggcgaacatc ccgggaggca tcaatggctg agtatcagaa	3600
catcttctcc caggtccagg tccgcggacc ggccgacctg gggatgaccg aagacgtcaa	3660
cctggccaac cgttcggggc tcggtccctt ctcgacctg ctcggtggt tcggcaacgc	3720
ccagctcggc ccgatctatc tcggctcgtc cggcgtcctg tccctcttct cgggcctgat	3780
gtggttcttc accatcggga tctggtctg gtatcaggcg ggtggaacc cggccgtctt	3840
cctgcgcgac ctgttcttct tctcgtcga gccgccggca cccgaatacg gtctgtcctt	3900
cgcggtccg ctgaaggaa gcggtctgtg gctgatcgcg tcgttcttca tgttcgtcgc	3960
ggtctggtcc tgggtggggc gcacatctc ccgcgctcag gcgctgggca tgggcaagca	4020
caccgcctg gcgttctct cggccatctg gctgtgatg gtgctgggct tcatccgtcc	4080
gatcctcatg ggtcctggt cggaaagcgt tccctacggc atcttctcgc acctcactg	4140
gacgaacaac ttctcgtcgc tccacggcaa cctgttctac aacccttcc acggtctctc	4200
gatgccttc ctctacgggt cggcctgct cttcgcgatg cacggtgca ccatcctgc	4260
ggtctccgc ttcggcggc agcgcgagct ggagcagatc gccgaccgcg ggacggcagc	4320
ggagcgggc gccctcttct ggcgctggac catgggttc aacgccacga tggaaagcat	4380
ccaccgctg gccatctgga tggcggctc cgtgacctc accggcggca tcgggatcct	4440
gctctcggc acggtcgtg acaactgta cgtctggggc cagaaccacg gcatggcgc	4500
gctgaactga ggagcgtca caatggctga caagaccatc ttcaacgatc acctcaacac	4560
caatccgaag accaacttc gcctctgggt cgctttccag atgatgaagg gtgcgggctg	4620
ggctggcggc gtgttcttc ggacgctcct tctcatcggg ttcttccggg tggtcgggcg	4680
gatgcttcc atccaggaga accaggctcc ggcgccgaac atcaccggcg ctctggagac	4740
cgggatcag ctgatcaagc atctcgtctg agacaagtct cggggcaggg cggcgcgag	4800
ccgccgctc ctccaagtcc gggccatc gccggcccgc gtcggggcg acaccacagc	4860
ccggttccct tctgttggc gacagggacc tggtgccgtg tggaaagccg caccggcacc	4920
tttgacatt caccggagc tctgatgacc aatcccacc cgcgaccoga aaccctctt	4980
ttggatcgc tctgctgcc ggccgacatg aaggcgtga gtgacccga actggagcgg	5040
ctggccgac aagtgcgttc cgaggcagt gatagggta gtttcttatt ttaggcagtt	5100
tatatgaaat taagacatgc agatgtcaca gtggatattg aactggtctc gaaagctcaa	5160
tatccccaa agcacaagca caaacttcga catcatgcag aagcgttcc cgaaccgcgt	5220
cttcgacgt ggcatcgcg agcagcatgc cgtgacctc gcggccggc tcgccggggc	5280
cgggatgaag cccttctgc gatctattc ctgctcctg caaccgggtt acgaccagat	5340
cgccatgac gtggcgtgc agaaccttc cgtccgcttc gtgatcgacc gggcggggct	5400
cgtgggggc gatggcgcga cccatgcggg gcccttcgac gttggcttca tcacttcgct	5460
gcccacatg accgtgatg ccgcggccga cgaggccgag ctcatccaca tgatgccac	5520
cgccgtggc ttcgacgag gccccatcgc cttccgcttc ccgcggggcg aggggtggg	5580

-continued

cgtcgagatg cccgagcgcg ggacgggtgct ggagcccggc cggggccgcg tgggtgcgca	5640
agggacggat gtcgcatcc tctccttcg cgcgcacatcg cacgaggcct tgcaggcggc	5700
gaaacttctc gaggccgagg gggtagcgt gaccgtggcc gaccccgcct tctcgcgcc	5760
gctcgacacg gggctcatcg accagctcgt gcgccatcac gcggcgttg taacggtgga	5820
gcagggggcc atgggcggct tcggcgccca tgtoatgcac tatctcgcca attccggcgg	5880
cttcgacggg ggccctcgcg tccgggtcat gacgtgccc gaccgcttca tcgagcaggc	5940
gagccccgag gacatgtatg ccgatgcggg gctgcggggc gaggatatca agcttggcgt	6000
aatcatggtc atagctgttt cctgtgtgaa attgttatcc gtcacaatt ccacacaaca	6060
tacgagccgg aagcataaag tgtaaagcct ggggtgccta atgagtgagc taactcacat	6120
taattgcggt gcgctcactg cccgctttcc agtcgggaaa cctgtcgtgc cagctgcatt	6180
aatgaatcgg ccaacgcgcg gggagaggcg gtttgctat tgggcgctcg gtcttgccct	6240
gctcgtcggg gatgtacttc accagctcgg cgaagtcgct cttcttgatg gagcgcattg	6300
ggacgtgctt ggcaatcacg cgcaccccc ggccgtttta gcggctaaaa aagtcattggc	6360
tctgccctcg gcggaccac gcccatcatg accttgccaa gctcgtcctg cttctcttcg	6420
atcttcgcca gcaggcgag gatcgtggca tcaccgaacc gcgccgtgcg cgggtcgtcg	6480
gtgagccaga gtttcagcag gccgcccag cggcccaggt cgcattgat gcgggcccagc	6540
tcgggacgt gctcatagtc cacgacgccc gtgattttgt agccctggcc gacggccagc	6600
aggtaggccg acaggtcat gccggccc gccgcctttt cctcaatcgc tcttcgttcg	6660
tctggaagcg agtacacct gataggtggg ctgcccttcc tggttggctt ggtttcatca	6720
gccatccgct tgcctcctc tgttacgccc gcggtagccc gccagcctcg cagagcagga	6780
ttccggtga gcaccgccag gtgcgaataa gggacagtga agaaggaaca cccgctcgcg	6840
ggtggcccta cttcacctat cctgcccggc tgacgccgtt ggatacacca aggaaagtct	6900
acacgaacct tttggcaaaa tcctgtatat cgtgcgaaaa aggatggata taccgaaaaa	6960
atcgtataa tgaccccgaa gcagggttat gcagcggaaa agcggccagc ttcccgaagg	7020
gagaaaggcg gacaggtatc cggtaagcgg cagggtcggg acaggagagc gacagaggga	7080
gcttccaggg gaaacgcct ggtatctta tagtctgctc gggtttcgccc acctctgact	7140
tgagcgtcga tttttgtgat gctcgtcagg ggggcggagc ctatggaaaa acgccagcaa	7200
cgcgcccttt ttacggttcc tggccttttg ctggcctttt gctcacatgt tctttcctgc	7260
gttatccctt gattctgtgg ataaccgtat taccgccttt gactgagctg ataccgctcg	7320
ccgcagccga acgaccgagc gcagcagtc agtgagcag gaagcggaa agcggccagaa	7380
ggcggccaga gaggccgagc gcggcgtga ggcttgagc ctagggcagg gcatgaaaaa	7440
gcccgtagcg ggctgctacg ggcgtctgac gcggtgaaa gggggagggg atgttgctca	7500
catggctctg ctgtagtgag tgggttcgctc tccggcagcg gtcctgatca atcgtcacc	7560
ttctcggctc cttcaacgct cctgacaacg agcctccttt tcgccaatcc atcgacaatc	7620
accgagagtc cctgctcga cgctgcgtcc ggaccggctt cgtcgaaggc gtctatcgcg	7680
gcccgaaca gcggcgagag cggagcctgt tcaacggtgc cgcgcgctc gccggcatcg	7740
ctgtcggcg cctgctcctc aagcaogccc ccaacagtga agtagctgat tgtoatcagc	7800
gcattgacgg cgtccccgcg cgaaaaacc gcctcgcaga ggaagcgaag ctgcgcgtcg	7860

-continued

gccgtttcca	tctgcggtgc	gcccggtcgc	gtgccggcat	ggatgcgcgc	gccatcgcg	7920
taggcgagca	gcgcctgcct	gaagctgcgg	gcattcccga	tcagaaatga	gcgccagtcg	7980
tcgtcggtc	tcggcaccga	atgcgtatga	ttctccgcca	gcatggcttc	ggccagtgcg	8040
tcgagcagcg	cccgttgtt	cctgaagtgc	cagtaaagcg	ccggtctctg	aacccccaac	8100
cgttccgcca	gtttgcgtgt	cgtcagaccg	tctacgccga	cctcgttcaa	caggtccag	8160
gcggcagga	tactgtatt	cggctgcaac	tttgtcatgc	ttgacacttt	atcactgata	8220
aacataatat	gtccaccaac	ttatcagtga	taaagaatcc	gcgcggtcaa	tcggaccagc	8280
ggaggctggt	ccggaggcca	gacatgaaac	ccaacatacc	cctgatcgta	attctgagca	8340
ctgtcgcgct	cgacgctgc	ggcatcgccc	tgattatgcc	ggtgctgccg	ggcctcctgc	8400
gcgatctggt	tactcgaac	gacgtcaccg	cccactatgg	cattctgctg	gcgctgtatg	8460
cgttggtgca	atttgcctgc	gcacctgtgc	tgggcgcgct	gtcggatcgt	ttcgggcggc	8520
ggccaatctt	gctcgtctcg	ctggccggcg	ccactgtcga	ctacgccatc	atggcgacag	8580
cgcctttcct	ttgggttctc	tatatcgggc	ggatcgtggc	cggcatcacc	ggggcgactg	8640
ggcggtagc	cgcgcttat	attgcgata	tactgatgg	cgatgagcgc	gcgcggcact	8700
tcggcttcat	gagcgcctgt	ttcgggttcg	ggatggctgc	gggacctgtg	ctcggtgggc	8760
tgatggggcg	tttctcccc	cacgtcccgt	tcttcgccgc	ggcagccttg	aacggcctca	8820
atttctgac	ggctgtttc	cttttgccgg	agtgcacaaa	aggcgaacgc	cgcccgttac	8880
gcggggaggc	tctcaaccgg	ctcgtctcgt	tccggtgggc	ccggggcatg	accgtcgtcg	8940
ccgcctgat	ggcggctctc	ttcatcatgc	aacttgtcgg	acaggtgccg	gccgcgcttt	9000
gggtcatttt	cgcgaggat	cgctttcact	gggacgcgac	cacgatcggc	atttctgctg	9060
ccgcatttgg	cattctgcat	tactcgcgcc	aggcaatgat	caccggccct	gtagccgcc	9120
ggctcggcga	aaggcgggca	ctcatgctcg	gaatgattgc	cgacggcaca	ggctacatcc	9180
tgcttgccct	cgcgacacgg	ggatggatgg	cgttcccgat	catggctcctg	cttgcttcgg	9240
gtggcatcgg	aatgccggcg	ctgcaagcaa	tgttgtccag	gcaggtggat	gaggaacgtc	9300
aggggcagct	gcaaggctca	ctggcggcgc	tcaccagcct	gacctcgtc	gtcggacccc	9360
tcctcttac	ggcgatctat	gcggcttcta	taacaacgtg	gaacgggtgg	gcatggattg	9420
caggcgtgc	cctctacttg	ctctgcctgc	cggcgtcgcg	tcgcgggctt	tgagcggcg	9480
cagggaacg	agccgatcgc	tgatcgtgga	aacgataggc	ctatgccatg	cgggtcaagg	9540
cgacttccgg	caagctatac	gcgccctagg	agtgcggttg	gaacggtggc	ccagccagat	9600
actcccgatc	acgagcagga	gcgccgatgat	ttgaagcgca	ctcagcgtct	gatccaagaa	9660
caaccatcct	agcaaacagc	cggctcccgg	gctgagaaag	cccagtaagg	aaacaactgt	9720
aggttcgagt	cgcgagatcc	cccgaacca	aaggaagtag	gttaaaccgg	ctccgatcag	9780
gccgagccac	gccagggcca	gaacattggt	tcctgtaggc	atcgggattg	gcggatcaaa	9840
cactaaagct	actggaacga	gcagaagtcc	tccggccgcc	agttgccagg	cggtaaaggt	9900
gagcagagcg	acgggagggt	gccacttgcg	ggtcagcacg	gttccgaacg	ccatggaaac	9960
cgccccgcgc	aggcccgtcg	cgacgccgac	aggatctagc	gctgcgtttg	gtgtcaaac	10020
caacagcgcc	acgcccgag	ttccgcaaat	agcccccagg	accgccatca	atcgtatcgg	10080
gctacctagc	agagcggcag	agatgaacac	gaccatcagc	ggctgcacag	cgccctaccgt	10140

-continued

cgccgcgacc	ccgcccggca	ggcggtagac	cgaaataaac	aacaagctcc	agaatagcga	10200
aatattaagt	gcgccgagga	tgaagatgcg	catccaccag	attcccgttg	gaatctgtcg	10260
gacgatcatc	acgagcaata	aacccgccgg	caacgcccgc	agcagcatac	cggcgacccc	10320
tcggcctcgc	tgctcgggct	ccacgaaaac	gccggacaga	tcgccttgtg	gagcgtcctt	10380
ggggccgtcc	tcctgtttga	agaccgacag	cccaatgac	tcgccgtcga	tgtaggcgcc	10440
gaatgccacg	gcatctcgca	accgttcagc	gaacgcctcc	atgggctttt	tctcctcgtg	10500
ctcgtaaacg	gaccgcaaca	tctctggagc	tttcttcagg	gccgacaatc	ggatctcgcg	10560
gaaatcctgc	acgtcggccg	ctccaagccg	tcgaatctga	gccttaatca	caattgtcaa	10620
ttttaatcct	ctgtttatcg	gcagttcgta	gagcgcgccc	tgcgctccga	gcgatactga	10680
gcgaagcaag	tcgctcgagc	agtgcgccct	tgcttctgaa	atgccagtaa	agcgtctggct	10740
gctgaacccc	cagccggaac	tgaccccaca	aggccctagc	gtttgcaatg	caccagggtca	10800
tcattgacct	aggcgtgttc	caccaggccg	ctgcctcgca	actcttcgca	ggcttcgccc	10860
acctgctcgc	gccacttctt	cacgcgggtg	gaatccgac	cgcacatgag	gcggaaggtt	10920
tccagcttga	gcgggtacgg	ctcccgggtc	gagctgaaat	agtcgaacat	ccgtcgggcc	10980
gtcggcgaca	gcttgcggtg	cttctcccat	atgaatttcg	tgtagtggtc	gccagcaaac	11040
agcagcagca	tttctctcgc	gatcaggacc	tggcaacggg	acgttttctt	gccacggctc	11100
aggacgcgga	agcgggtcag	cagcgacacc	gattccagg	gcccacgcg	gtcggacgtg	11160
aaagccatcg	ccgtcgcctg	taggcgcgac	aggcattcct	cggccttcgt	gtaataccgg	11220
ccattgatcg	accagcccag	gtcctggcaa	agctcgtaga	acgtgaaggt	gatcggctcg	11280
ccgatagggg	tcgcttccgc	gtactccaac	acctgctgcc	acaccagttc	gtcatcgtcg	11340
gcccgcagct	cgacgcgggt	gtaggtgatc	ttcacgtcct	tgctgacgtg	gaaaatgacc	11400
ttgttttgca	gcgcctcgcg	cgggattttc	ttgttgccgc	tggtgaacag	ggcagagcgg	11460
gccgtgtcgt	ttggcatcgc	tcgcatcgtg	tccggccacg	gcgcaatc	gaacaaggaa	11520
agctgcatct	ccttgatctg	ctgcttcgtg	tgcttcagca	acgcggcctg	cttggcctcg	11580
ctgacctgtt	ttgccaggtc	ctgcgcggcg	gtttttcgtc	tcttggctgt	catagttcct	11640
cgcggtcga	tggtcatcga	cttcgccaaa	cctgcccctc	cctgttcgag	acgacgcgaa	11700
cgctccacgg	cggccgatgg	cgcgggcaag	gcagggggag	ccagttgcac	gctgtcgcgc	11760
tcgatcttgg	ccgtagcttg	ctggaaccac	gagccgacgg	actggaaggt	ttcgcggggc	11820
gcacgcatga	cgggtcggct	tcgcatggtt	tcggcatcct	cggcgaaaaa	cccgcgctcg	11880
atcagttcct	gcctgtatgc	cttccgggtc	aacgtccgat	tcattcaacc	tccttgccgg	11940
attgccccga	ctcacgccgg	ggcaatgtgc	cottattcct	gatttgaacc	gcctgggtgc	12000
ttggtgtcca	gataatccac	cttatcggca	atgaagtcgg	tcccgtagac	cgtctggccc	12060
tccttctcgt	acttgggtatt	ccgaatcttg	ccctgcacga	ataccagcga	ccccttgccc	12120
aaatacttgc	cgtgggcctc	ggcctgagag	ccaaaacact	tgatgcggaa	gaagtcgggtg	12180
cgctcctgct	tgctcggcgt	cgtggcccgc	ccaacctttg	cgatccgcaa	gcgcgcggtc	12240
gccatcttca	cgctggaaca	gtacgtcgag	gcgggcatca	tgaccgcgga	gcaatacagag	12300
gtcattaaaa	gcgccgtgat	tgatgatata	goggccccgc	tgctcctggt	tctcgcgcac	12360
cgaaatgggt	gacttcaccc	cgcgctcttt	gatcgtggca	ccgatttcgg	cgatgctctc	12420

-continued

```

cggggaaaag ccggggttgt cggccgtccg cggtgatgc ggatcttcgt cgatcaggtc 12480
cagggtccagc tcgatagggc cggaaaccgc ctgagacgcc gcaggagcgt ccaggaggct 12540
cgacaggtcg ccgatgctat ccaaccccag gccggacggc tgcgccgcgc ctgcggcttc 12600
ctgagcggcc gcagcggtgt tttcttgggt ggtcttggt tgagccgcag tcattgggaa 12660
atctccatct tcgtgaacac gtaatcagcc agggcgcgaa cctctttcga tgccttgccg 12720
gcgccggttt tcttgatctt ccagaccggc acaccggatg cgagggcacg gccgatgctg 12780
ctgcgcaggc caacgggtgc cggaatcacc atcttggggg acgcggccag cagctcggct 12840
tggtggcgcg cgtggcgcgg attccgcgca tcgacctgc tgggcacccat gccaaaggaat 12900
tgacgcttgg cgttcttctg gcgcacgttc gcaatggtcg tgacctctt cttgatgccc 12960
tggtgctgt acgcctcaag ctcgatgggg gacagacat agtcggccgc gaagagggcg 13020
gccgccaggc cgacgccaa ggtcggggcc gtgtcgatca ggcacacgtc gaagccttgg 13080
ttcgccaggc cttgatgtt cccccgaa agctcgcggg cgtcgtccag cgacagccgt 13140
tcggcgcttc ccagtaccg gttgactcg atgagggcga ggcgcgcggc ctggccgctg 13200
ccggctgcgg gtgcggttcc ggtccagccg ccggcaggga cagcgcgaa cagcttgctt 13260
gcatgcaggc cggtagcaaa gtccttgagc gtgtaggacg cattgccctg ggggtccagg 13320
tcgatcagc caaccgcaa gccgcgctc aaaaagtcga aggcaagatg cacaagggtc 13380
gaagtcttgc cgacgcccc tttctggtg gccgtgacca aagtttcat cgtttggtt 13440
cctgtttttt cttggcgctc gcttcccact tccggacgat gtacgcctga tgttccggca 13500
gaaccgccgt taccgcgcg taccctcgg gcaagttctt gtcctcgaac gcggcccaca 13560
cgcgatgcac cgttgcgac actgcgccc tggtcagtcc cagcgcggtt gcgaacgctg 13620
cctgtggctt cccatcgact aagacgccc gcgctatctc gatggtctgc tgcccactt 13680
ccagcccctg gatcgctcc tggaaactgg tttcggtgaa cgtttcttc atggataaca 13740
cccataattt gtcgccgccc ttggttgaac atagcgtgta cagccgccag cacatgagag 13800
aagtttagct aacatcttct cgcacgtcaa cacctttagc cgtaaaaact cgtccttggc 13860
gtaacaaaac aaaagcccg aaaccgggt ttcgtctctt gccgcttatg gctctgcacc 13920
cggtccatc accaacaggt cgcgcacgc cttactcgg ttgcggatcg aactgccag 13980
cccaacaaa cgggttgccg ccgccgccg gatcgcgccc atgatgccg ccacaccggc 14040
catcgcccac caggtcgcg ccttcgggtt ccattcctgc tggtaactgt tcgcaatgct 14100
ggacctcggc tcacatagc ctgaccgctc gatggcgat gccgcttctc cccttggcgt 14160
aaaaccagc gccgcaggcg gcattgccat gctgccgcc gctttccga ccacgacgcg 14220
cgcaccaggc ttgcggtcca gacctcggc cacggcgagc tgcgcaagga cataatcagc 14280
cgccgacttg gtcacgcg cctcgatcag ctcttgact cgcgcgaaat ccttggcctc 14340
cacggccgcc atgaatcgcg cagcggcgga aggetccgca gggccg 14386

```

<210> SEQ ID NO 14

<211> LENGTH: 14370

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic plasmid

<400> SEQUENCE: 14

-continued

cccccaggc cgcccccctc actgcccgc acctggctgc tgaatgtcga tgccagcacc	60
tgcggcacgt caatgcttcc gggcgtcgcg ctcgggctga tcgcccattc cgttactgcc	120
ccgatcccgg caatggcaag gactgccagc gctgccattt ttgggggtgag gccgttcgcg	180
gccgaggggc gcagcccctg gggggatggg aggcccgctg tagcggggcg ggagggttcg	240
agaagggggg gcacccccct tcggcgtcgc cggtcacgcg cacagggcgc agccctggtt	300
aaaaacaagg ttataaata ttggtttaa agcaggtaa aagacaggtt agcggtggcc	360
gaaaaacggg cggaaacctc tgcaaatgct ggattttctg cctgtggaca gcccctcaa	420
tgtaaatag tgcgcccctc atctgtcagc actctgcccc tcaagtgtca aggatcgcgc	480
ccctcatctg tcagtagtcg cggccctcaa gtgtcaatac cgcagggcac ttatcccag	540
gcttgccac atcatctgtg ggaaactcgc gtaaaatcag gcgttttcgc cgatttgca	600
ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgccgc	660
gccgggtgag tcggcccctc aagtgtcaac gtccgcccct catctgtcag tgagggccaa	720
gttttcgcg aggtatccac aacgcggcg gcccggtgt ctgcacacg gcttcgacgg	780
cgtttctggc gcgtttgca ggccatagac ggccgccagc ccagggcga gggcaaccag	840
cccgtgagc tcggaagg cgctcttcgc cttcctcgt cactgactcg ctgcgctcgg	900
tcgttcggct gcggcagcgt gtatcagctc actcaaaggc ggtaatacgg ttatccacag	960
aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc	1020
gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg cccccctgac gagcatcaca	1080
aaaaatcgac ctcaagttag aggtggcgaa acccgacagg actataaaga taccagcgct	1140
ttccccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc	1200
tgtccgcctt tctcccttcg ggaagcgtgg cgccattcgc cattcaggct gcgcaactgt	1260
tgggaagggc gatcggctgc gccctcttcg ctattacgcc agctggcgaa aggggatgt	1320
gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg	1380
acggccagtg aattcggccc cgggctggcc gaggtgctgg gcaagcccta cctccaggcc	1440
cccatcgggg tcgagagcac gaccgccttc ctgcgcccgc tgggcgagat tctgggcctc	1500
gatccggagc cttcatcga gcgcgagaag cactcgacgc tgaagccctg gtgggatctg	1560
tggcggagtg tcacgcagga cttcttcggg acggccaatt tcggaatcgt ggcgaccgaa	1620
acttatgcaa gaggcattcc aaactatctc gaaggcagtc tcgggctgcc ctgcgccttc	1680
gccgtggccc gcaagagggg ctggaagacc gacaacgaag cgggtgcgcg actgatccgc	1740
cagcacctgc cgtctgtgct catggggtcg atcaacgaga agatttacct tgcggaactg	1800
aaagccggtc acggcccga accctcttcc atcgtgctc ctttcccggg tgcggcgatc	1860
cggcgcgcta ccggaaccgc cgttatggga tatgcagtg ctacgtggtt actgcaggaa	1920
gtttgcaacg ccctgttcga cggcctgttc cacattctgc ccctcgggac ggagatggac	1980
agcgcggcgc ccacaccgac gacctcgcgc cgcgacttcc cgtgggatgc cgatgcgcaa	2040
gcggccctgg accgcatcgt agaggagcat ccggttctca cccggatcag cgcgcgcgt	2100
gccttgccgc acgcccgcga gaaggctgcc ctcgatgccg gtgccgagag ggtcgtgaga	2160
gagactgtcg aagccctgcg tgggcccggc ttcggcgaga ggaagggaga gaaccaatga	2220
gcgatcatgc cgtcaacacg ccggtccatg ccgccagggc ccacgggcac ctagcaccac	2280

-continued

gtgccgagtt ctacgtctac ttcgccgtca ttctgctggg cgccttcccg gtggccttcg	2340
tgagctggat cgtctcgacg atccgccacc gcaggcttcc caagcgcggc cccttcgagt	2400
ccgcctgggt cgatgccaaag gcgatcacgc cgctgatttt ccgcgccctga ccgcaggctca	2460
ggttgcgaca cgccattcgt cgtctcccca aggggcggcg gattaatcgg gagggcatgg	2520
tgccctaccg taaccacagc caccagcatg tggaggatcg ccatgcacca ccaccaccac	2580
caccacgcta gcgcgggcag atcttgaggc cggccctccg tcgcgggcgg caccacgccc	2640
cgcatcgatt ccaaggttca gccattgaga cggctccgct tcgcgcgcaa gcgcggttg	2700
ggccgactgc aagcggagag ggaagcatgg cactgctcag cttcgagcga aaatatcgcg	2760
tgccgggggg cagcctggtc ggcgaaacc tgttcgactt ctgggtcggc cctttctatg	2820
tcggcttctt cggggttgcg acgtttttct tcgcgccctt gggatcatt ctgattgcct	2880
ggagtccgt actccagggt acctggaacc cccaactcat ctctgtctac ccgccggccc	2940
ttgaatatgg cctgggaggt gcaccctcg caaaaggcgg gctgtggcag atcatcacga	3000
tctcgccac tggtgccttc gtcagctggg cgctgcgca agtcgaaatc tgccgtaagc	3060
tgggcatcgg gtaccacatc ccgttcgct tcgcttcgc catcctggcc tacctgacgc	3120
tggtgctggt ccgccggtg atgatggcg cctggggcta tgccttccc tacgggatct	3180
ggacgcacct cgactgggtg tcgaacacgg gctacaccta cggcaacttc cactacaacc	3240
ctgccacat gatcgccatc tcgttcttct tcacgaacgc gctggctctg gcgctgcacg	3300
gcgcccttgt gctctcccg gccaaacccg agaaggcga ggaatgcgg accccgatc	3360
acgaggatac gttcttccgc gatctggtcg gctactcgat cgggacgctc ggcattccacc	3420
gcctcggcct gctgctctcg ctgagcggc tcttcttcag cgcctctgc atgatcatta	3480
ccggcaccat ctggttcgat cagtggtcg actggtggca atggtgggtg aagctgccgt	3540
ggtggcgaa catcccggga ggcattcaat gctgagtac agaacatctt ctcccaggtc	3600
caggtcgcg gaccggcga cctggggatg accgaagacg tcaacctggc caaccgttcg	3660
ggcgtcggtc cttctcgc cctgctcggc tggttcggca acgccagct cggcccgatc	3720
tatctcggct cgctcggcct cctgtccctc ttctcgggcc tgatgtggtt cttcaccatc	3780
gggatctggt tctggtatca ggcgggtcg aaccggcgc tcttctcgc cgaactgttc	3840
ttcttctcgc tcgagccgcc ggcacccgaa tacggtctgt ccttcgccc tccgctgaag	3900
gaagcggggc tgtggtgat cgcgtcgttc ttcatgttcg tcgcggtctg gtcctggtg	3960
ggcgcacct atctccgcgc tcaggcgtg ggcattggca agcacaccgc ctgggcgttc	4020
ctctcggcca tctggtgtg gatggtcgt ggcttcatcc gtcgatcct catgggttc	4080
tggtcggaa cggttcccta cggcatcttc tcgcacctcg actggacgaa caacttctcg	4140
ctcgtccacg gaaacctgtt ctacaacccc ttccacggtc tctcgatcgc ctctctctac	4200
gggtcggccc tgctcttcgc gatgcacggt gcgaccatcc tcgcggtctc ccgcttcggc	4260
ggcgagcgc agctggagca gatcggcgc cgcgggacgg cagcggagcg ggcgccttc	4320
ttctggcgt ggaccatggg tttcaacgcc acgatggaag gcattcaacc ctgggccatc	4380
tgatggcgg tcctcgtgac cctcaccggc ggcattcggga tcctgctctc gggcacggtc	4440
gtggacaact ggtacgtctg gggccagaac cacggcatgg cgcgctgaa ctgaggagcg	4500
atcacaatgg ctgacaagac catcttcaac gatcacctca acaccaatcc gaagaccaac	4560

-continued

cttcgcctct	gggtcgcttt	ccagatgatg	aaggggtcgg	gctgggctgg	cggcgtgttc	4620
ttcgggacgc	tccttctcat	cgggttcttc	cgggtggtcg	ggcggatgct	tccgatccag	4680
gagaaccagc	ctccggcgcc	gaacatcacc	ggcgtcttgg	agaccgggat	cgagctgatc	4740
aagcatctcg	tctgagacaa	gtctcggggc	agggcggcgc	gaggccgccc	gctcctccaa	4800
gtccgggcca	tatcgccggc	ccgggtccgg	ggcgacacca	cagcccgggt	cccttctgt	4860
tggcgacagc	gacctggtgc	cgtgtggaag	accgcacggc	acccttttga	cattcacggg	4920
aggctctgat	gaccaatccc	accccgcgac	ccgaaacccc	gcttttggat	cgcgtctgct	4980
gcccggccga	catgaaggcg	ctgagtgcgc	ccgaactgga	gcggctggcc	gacgaagtgc	5040
gttccgaggt	cagtgatagg	ggtagtcttc	tatttttaggc	agtttatatg	aaattaagac	5100
atgcagatgt	cacagtggat	attgaaactgg	tctcgaaagc	tcaatatccc	caaagcaca	5160
agcacaacct	tcgacatcat	gcagaagcgt	ttcccgaacc	gcgtcttcga	cgtgggcatac	5220
gccgagcagc	atgccgtgac	cttcgogggc	ggcctcgcgg	gggcccggat	gaagcccttc	5280
tgcgcatct	attcctcgtt	cctgcaacgg	ggttacgacc	agatcgccca	tgacgtggcg	5340
ctgcagaacc	ttcccgtccg	cttcgtgatc	gaccgggccc	ggctcgtggg	ggccgatggc	5400
gcgacccatg	cggggccctt	cgacgttggc	ttcatcactt	cgctgcccaa	catgaccgtg	5460
atggcccggc	ccgacgaggc	cgagctcacc	cacatgatcg	ccaccgcccgt	ggccttcgac	5520
gagggcccca	tcgcttccg	cttcccggcg	ggcgaggggg	tgggcgtcga	gatgcccgag	5580
cgcgggacgc	tgctggagcc	cggccggggc	cgcgtggtgc	gcgaagggac	ggatgtcgcg	5640
atcctctcct	tcggcgcgca	tctgcacgag	gccttgacgg	cggcgaaact	tctcgaggcc	5700
gaggggtgga	gcgtgaccgt	ggccgacgcc	cgttctcgc	gcccgtcga	cacggggctc	5760
atcgaccagc	tcgtgcgcca	tcacgogggc	ctggtaacgg	tggagcaggg	ggccatgggc	5820
ggcttcggcg	cccatgtcat	gcactatctc	gccaatccg	gcggcttcga	cgggggctc	5880
gcgctccggg	tcatgacgct	gcccgaccgc	ttcatcgagc	aggcgagccc	cgaggacatg	5940
tatgccgatg	cggggctcgc	ggccgaggat	atcaagcttg	gcgtaatcat	ggtcatagct	6000
gtttcctgtg	tgaattgtt	atccgctcac	aattccacac	aacatacgag	ccggaagcat	6060
aaagtgtaaa	gcctgggggtg	cctaagtgtg	gagctaactc	acattaattg	cgttgcgctc	6120
actgcccgct	ttccagtcgg	gaaacctgtc	gtgccagctg	cattaatgaa	tcggccaacg	6180
cgcggggaga	ggcggtttgc	gtattggggc	ctcggctctg	ccttgctcgt	cggatgatgta	6240
cttcaccagc	tcgcggaagt	cgctcttctt	gatggagcgc	atggggacgt	gcttgccaat	6300
cacgcgcacc	ccccggccgt	tttagcggct	aaaaaagtca	tggctctgcc	ctcgggcgga	6360
ccacgcccat	catgaccttg	ccaagctcgt	cctgcttctc	ttcgatcttc	gccagcaggg	6420
cgaggatcgt	ggcatcaccg	aaccgogccg	tgccggggtc	gtcggtgagc	cagagtttca	6480
gcagggccgc	cagggcgccc	aggtcgccat	tgatgcgggc	cagctcggcg	acgtgctcat	6540
agtccacgac	gcccgtgatt	ttgtagccct	ggccgacggc	cagcaggtag	gccgacaggc	6600
tcattgccgc	cgcgcggcc	ttttcctcaa	tcgctcttcg	ttcgtctgga	aggcagtaca	6660
ccttgatagc	tgggctgccc	ttcctgggtg	gcttgggttc	atcagccatc	cgttgccct	6720
catctgttac	gcccggcgta	gcccggccagc	ctcgcagagc	aggattcccg	ttgagcaccg	6780
ccaggtcgca	ataaggggaca	gtgaagaagg	aacacccgct	cgcgggtggg	cctacttcac	6840

-continued

ctatcctgcc	cggtgacgc	cgttggatac	accaaggaaa	gtctacacga	accctttggc	6900
aaaaatcctgt	atatacgtgcg	aaaaaggatg	gataataccga	aaaaatcgct	ataatgaccc	6960
cgaagcaggg	ttatgcagcg	gaaaagcgcc	acgcttcccg	aagggagaaa	ggcggacag	7020
tatccggtaa	gcggcagggt	cggaacagga	gagcgcacga	gggagcttcc	aggggaaac	7080
gcctggtatc	tttatagtcc	tgctcgggtt	cgccacctct	gacttgagcg	tcgatttttg	7140
tgatgctcgt	cagggggcg	gagcctatgg	aaaaacgcca	gcaacgcggc	ctttttacgg	7200
ttcctggcct	tttgctggcc	ttttgctcac	atgttctttc	ctgcgttatac	ccctgattct	7260
gtggataacc	gtattaccgc	ctttgagtga	gctgataccg	ctcgcgcag	ccgaacgacc	7320
gagcgcagcg	agtcagttag	cgaggaagcg	gaagagcgcc	agaaggccgc	cagagagggc	7380
gagcgcggcc	gtgagccttg	gacgctaggg	cagggcatga	aaaagcccgt	agcgggctgc	7440
tacggcgctc	tgacgcgggt	gaaaggggga	ggggatgttg	tctacatggc	tctgctgtag	7500
tgagtgggtt	gcgctccggc	agcggtcctg	atcaatcgtc	accctttctc	ggctcctcaa	7560
cgttcctgac	aacgagcctc	cttttcgcca	atccatcgac	aatcaccgcg	agtccttget	7620
cgaacgctgc	gtccggaccg	gcttcgctga	aggcgtctat	cgcgcccgcc	aacagcggcg	7680
agagcggagc	ctgttcaacg	gtgcccccgc	gctcgcggc	atcgtgtctg	ccggcctgct	7740
cctcaagcac	ggccccaaaca	gtgaagtagc	tgattgtcat	cagcgcattg	acggcgtccc	7800
cgccgaaaa	acccgcctcg	cagaggaagc	gaagctgcgc	gtcggccgtt	tccatctgcg	7860
gtgcgcccgg	tcgctgcccg	gcatggatgc	gcgcgccatc	gcgtagggcg	agcagcgcct	7920
gcctgaagct	gcgggcatcc	ccgatcagaa	atgagcgcca	gtcgtcgtcg	gctctcggca	7980
ccgaatcgct	atgattctcc	gccagcatgg	cttcggccag	tcgctcgagc	agcggcccgt	8040
tgttcctgaa	gtgccagtaa	agcgcggct	gctgaacccc	caaccgttcc	gccagtttgc	8100
gtgtcgtcag	accgtctacg	ccgacctcgt	tcaacagtc	cagggcggca	cggatcactg	8160
tattcgctg	caactttgtc	atgcttgaca	ctttatcact	gataaacata	atatgtccac	8220
caacttatca	gtgataaaga	atccgcgcgt	tcaatcggac	cagcggaggc	tggtccggag	8280
gccagacatg	aaaccaaca	taccctgat	cgtaattctg	agcactgtcg	cgctcgacgc	8340
tgtcggcatc	ggcctgatta	tgccgggtct	gccgggcctc	ctgcgcgac	tggttctctc	8400
gaacgacgct	accgcccact	atggcattct	gctggcgcgtg	tatgcgttgg	tgcaatttgc	8460
ctgcgcacct	gtgctggggc	cgctgtcgga	tcgtttcggg	cgcgggccaa	tcttgctcgt	8520
ctcgtggcc	ggcgccactg	tcgactacgc	catcatggcg	acagcgcctt	tcctttgggt	8580
tctctatatac	ggcgggatcg	tgccggcat	caccggggcg	actggggcgg	tagccggcgc	8640
ttatattgcc	gatatacactg	atggcgtatga	gcgcgcgcgg	cacttcggct	tcatgagcgc	8700
ctgtttcggg	ttcgggatgg	tcgcgggacc	tgtgctcgggt	gggctgatgg	gcggtttctc	8760
ccccacgct	ccgttctctg	ccgcggcagc	cttgaacggc	ctcaatttcc	tgacgggctg	8820
ttctcttttg	ccggagtcgc	acaaaggcga	acgcgggcgg	ttacggccggg	aggctctcaa	8880
cccgcctcgt	tcgttccgggt	ggccocgggg	catgaccgct	gtcgcgcgcc	tgatggcgggt	8940
cttctctcatc	atgcaacttg	tcggacaggt	gccggccgcg	ctttgggtca	ttttcggcga	9000
ggatcgcctt	cactgggagc	cgaccacgat	cggcatttctg	cttgccgcct	ttggcattct	9060
gcattcactc	gcccaggcaa	tgatcaccgg	ccctgtagcc	gcccggctcg	gcgaaaggcg	9120

-continued

ggcactcatg	ctcggaatga	ttgccgacgg	cacaggctac	atcctgcttg	ccttcgcgac	9180
acggggatgg	atggcgcttc	cgatcatggt	cctgcttgct	tcgggtggca	tcggaatgcc	9240
ggcgctgcaa	gcaatgttgt	ccaggcaggt	ggatgaggaa	cgtcaggggc	agctgcaagg	9300
ctcactggcg	gcgctcacca	gcctgacctc	gatcgtegga	cccctcctct	tcacggcgat	9360
ctatgctgct	tctatacaaa	cgtagaacgg	gtgggcatgg	attgcaggcg	ctgccctcta	9420
cttgctctgc	ctgccggcgc	tgcgtcgcgg	gctttggagc	ggcgagggc	aacgagccga	9480
tcgctgatcg	tggaaacgat	aggcctatgc	catgctgggc	aaggcgactt	ccggcaagct	9540
atacgcgccc	taggagtgcg	gttggaacgt	tggcccagcc	agatactccc	gatcacgagc	9600
aggacgccga	tgatttgaag	cgcactcagc	gtctgatcca	agaacaacca	tcctagcaac	9660
acggcggtcc	ccgggctgag	aaagcccagt	aaggaaacaa	ctgtaggttc	gagtcgagag	9720
atcccccgga	accaaaggaa	gtaggttaaa	cccgtccgga	tcaggccgag	ccacgccagg	9780
ccgagaacat	tggttcctgt	aggcatcggg	attggcggat	caaacactaa	agctactgga	9840
acgagcagaa	gtcctccggc	cgccagttgc	caggcggtaa	aggtgagcag	aggcacggga	9900
ggttgccact	tgcgggtcag	cacggttccg	aacgccatgg	aaaccgcccc	cgccaggccc	9960
gctgcgacgc	cgacaggatc	tagcgtcgcg	tttggtgtca	acaccaacag	cgccacgccc	10020
gcagttccgc	aaatagcccc	caggaccgcc	atcaatcgta	tcgggctacc	tagcagagcg	10080
gcagagatga	acacgaccat	cagcggtcgc	acagcgccta	ccgtcgcgcg	gacccccgcc	10140
ggcagcgcgt	agaccgaaat	aaacaacaag	ctccagaata	gcaaatatt	aagtgcgcgc	10200
aggatgaaga	tgcgcatcca	ccagattccc	gttggaatct	gtcggacgat	catcacgagc	10260
aataaacccc	ccggcaacgc	ccgcagcagc	ataccggcga	cccctcggcc	tcgctgttcg	10320
ggctccacga	aaacgcggga	cagatgcgcc	ttgtgagcgt	ccttggggcc	gtcctcctgt	10380
ttgaagaccg	acagcccaat	gatctcgcgc	tcgatgtagg	cgccgaatgc	cacggcatct	10440
cgcaaccggt	cagcgaacgc	ctccatgggc	tttttctcct	cgtgctcgta	aacggaccgc	10500
aacatctctg	gagctttctt	cagggccgac	aatcggatct	cgcggaatc	ctgcacgtcg	10560
gccgctccaa	gccgtcgaat	ctgagcctta	atcacaattg	tcaattttaa	tcctctgttt	10620
atcggcagtt	cgtagagcgc	gccgtgcgtc	ccgagcgata	ctgagcgaag	caagtgcgtc	10680
gagcagtgcc	cgcttgttcc	tgaatgcca	gtaaagcgc	ggctgctgaa	ccccagccc	10740
gaactgaccc	cacaaggccc	tagcgtttgc	aatgcaccag	gtcatcattg	accagggcgt	10800
gttcaccag	gccgctgcct	cgcaactctt	cgcaggcttc	gccgacctgc	tcgogccact	10860
tcttcacgcg	ggtggaatcc	gatccgcaca	tgaggcggaa	ggtttccagc	ttgagcgggt	10920
acggctcccg	gtgcgagctg	aaatagtcga	acatccgtcg	ggcgtcggc	gacagcttgc	10980
ggtacttctc	ccatataaat	ttcgtgtagt	ggtcgcacgc	aaacagcagc	acgatttcct	11040
cgctgatcag	gacctggcaa	cgggacgttt	tcttgccacg	gtccaggacg	cggaagcgggt	11100
gcagcagcga	caccgattcc	aggtgcccga	cgcggtcggga	cgtgaagccc	atcgccgtcg	11160
cctgtaggcg	cgacaggcat	tcctcggcct	tcgtgtaata	ccggccattg	atcgaccagc	11220
ccaggctcctg	gcaaagctcg	tagaacgtga	aggtgatcgg	ctcgccgata	ggggtgcgct	11280
tcgctactc	caacacctgc	tgccacacca	gttcgtcctc	gtcggcccgc	agctcgacgc	11340
cggtgtaggt	gatcttcacg	tccttgttga	cgtggaaaat	gacctgtttt	tcgagcgcct	11400

-continued

cgcgcgggat	tttcttgttg	cgcgtggtga	acagggcaga	gcgggccgtg	tcgtttggca	11460
tcgctcgc	cgtgtccgc	cacggcgcaa	tatcgaacaa	ggaagctgc	atttccttga	11520
tctgctgctt	cgtgtgtttc	agcaacgcgg	cctgcttggc	ctcgtgacc	tgttttgcca	11580
ggctctcgcc	ggcggttttt	cgcttcttgg	tcgtcatagt	tcctcgcgtg	tcgatggtca	11640
tcgacttcgc	caaacctgcc	gcctcctggt	cgagacgacg	cgaacgctcc	acggcggccg	11700
atggcgcggg	cagggcaggg	ggagccagtt	gcacgctgtc	gcgctcgatc	ttggccgtag	11760
cttgctggac	catcgagccg	acggactgga	aggtttcgcg	ggcgccacgc	atgacggtgc	11820
ggcttgcgat	ggtttcggca	tcctcggcgg	aaaaccccgc	gtcgatcagt	tcttgccctgt	11880
atgccttccg	gtcaaacgtc	cgattcattc	accctccttg	cgggattgcc	ccgactcacg	11940
ccggggcaat	gtgcccttat	tcctgatttg	acccgcctgg	tgccctgggtg	tccagataat	12000
ccaccttata	ggcaatgaag	tcggtcccgt	agaccgtctg	gccgtccttc	tcgtacttgg	12060
tattccgaat	cttgccctgc	acgaatacca	gcgaccctt	gccaaatac	ttgccgtggg	12120
cctcggcctg	agagcaaaa	cacttgatgc	ggaagaagtc	ggtgcgctcc	tgcttgtcgc	12180
cggctgtggc	cgcgccaaac	tttgcgatcc	gcaagcgcgc	ggtcgccatc	ttcacgctgg	12240
aacagtacgt	cgaggcgggc	atcatgaccc	gcgagcaata	cgaggtcatt	aaaagcggcg	12300
tgattgatga	tatagcggcc	cggtgctcc	tggttctcgc	gcaccgaaat	gggtgacttc	12360
accccgcgct	ctttgatcgt	ggcaccgatt	tcgcgatgc	tctccgggga	aaagccgggg	12420
ttgtcggccg	tcgcggcgtg	atgcggatct	tcgtcgatca	ggtccaggtc	cagctcgata	12480
gggcccgaac	cgccctgaga	cgccgcagga	gcgtccagga	ggctcgacag	gtcggccgatg	12540
ctatccaacc	ccaggccgga	cggtgcgcc	gcgcctgcgg	cttctgagc	ggccgcagcg	12600
gtgtttttct	tggtggtcct	ggcttgagcc	gcagtcattg	ggaatctcc	atcttcgtga	12660
acacgtaate	agccagggcg	cgaacctctt	tcgatgcctt	gcgcgcggcc	gttttcttga	12720
tcttccagac	cgccacaccg	gatgcgaggg	catcggcgat	gctgctgcgc	aggccaacgg	12780
tgcccggaat	catcatcttg	gggtacgcgg	ccagcagctc	ggcttggtgg	cgcgcggtggc	12840
gcggttccg	cgcatcgacc	ttgctgggca	ccatgccaaag	gaattgcagc	ttggcgttct	12900
tctggcgca	gttcgcaatg	gtcgtgacca	tcttcttgat	gccctggatg	ctgtacgcct	12960
caagctcgat	gggggacagc	acatagtcgg	ccgcgaagag	ggcgcccgcc	aggccgacgc	13020
caagggtcgg	ggccgtgtcg	atcaggcaca	cgtcgaagcc	ttggttcgcc	agggccttga	13080
tgttcggccc	gaacagctcg	cgggcgtcgt	ccagcagacg	ccgttcggcg	ttcggcagta	13140
ccgggttgg	ctcgatgag	gcgagggcgc	cggcctggcc	gtcgcggct	gcgggtgcgg	13200
tttcggtcca	gcccgcggca	gggacagcgc	cgaacagctt	gcttgcatgc	aggccggtag	13260
caaagtcctt	gagcgtgtag	gacgcattgc	cctgggggtc	caggtcgatc	acggcaaccc	13320
gcaagccgcg	ctcgaaaaag	tcgaaggcaa	gatgcacaag	ggtcgaagtc	ttgccgacgc	13380
cgctttctg	ggtggccgtg	accaaagttt	tcctcgtttg	gtttcctggt	ttttcttggc	13440
gtcgccttcc	cacttccgga	cgatgtacgc	ctgatgttcc	ggcagaaccc	ccgttaccgg	13500
cgctacccc	tcgggcaagt	tcttgtcctc	gaacgcggcc	cacacgcgat	gcaccgcttg	13560
cgacactgcg	cccctggtca	gtcccagcga	cgttgcgaac	gtcgcctgtg	gcttcccatc	13620
gactaagacg	ccccgcgcta	tctcgatggt	ctgctgcccc	acttccagcc	cctggatcgc	13680

-continued

```

ctcctggaac tggctttcgg taagccggtt cttcatggat aacaccata atttgctccg 13740
cgcttgggtt gaacatagcg gtgacagccg ccagcacatg agagaagttt agctaaacat 13800
ttctcgcacg tcaacacctt tagccgctaa aactcgtcct tggcgtaaca aaacaaaagc 13860
cgggaaacgg ggctttcgtc tcttgccgct tatggctctg caccggctc catcaccaac 13920
aggtcgcgca cgcgcttcac tcggttgccg atcgacactg ccagcccaac aaagccggtt 13980
gccgcccggc ccaggatcgc gccgatgatg ccggccacac cggccatcgc ccaccaggtc 14040
gccgcttcc ggttccattc ctgctggtag tgcttcgcaa tgctggacct cggctcacca 14100
taggtgacc gctcgatggc gtatgccgct tctccccttg gcgtaaaacc cagcggccgca 14160
ggcggcattg ccatgctgcc cgcgcttcc cggaccacga cgcgcccacc aggcttgccg 14220
tccagacctt cggccacggc gagctgcgca aggacataat cagccgccga cttggctcca 14280
cgcgctcga tcagctcttg cactcgcgcg aaatccttgg cctccacggc cgccatgaat 14340
cgcgcaacgg gcgaaggctc cgcagggcgg 14370

```

<210> SEQ ID NO 15

<211> LENGTH: 14398

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic plasmid

<400> SEQUENCE: 15

```

ccaccaggc cgcggccctc actgcccggc acctggtcgc tgaatgtcga tgccagcacc 60
tgccgcacgt caatgcttcc gggcgtcgcg ctcgggctga tcgccatcc cgttactgcc 120
ccgatcccgg caatggcaag gactgccagc gctgccattt ttgggggtgag gccgttcgcg 180
gccgaggggg gcagcccctg ggggatggg aggcccgcgt tagcgggccc ggagggttcg 240
agaagggggg gcacccccct tcggcgtcgc cggtcacgcg cacagggcgc agccctggtt 300
aaaaacaagg ttataaata ttggtttaa agcaggttaa aagacaggtt agcgggtggc 360
gaaaaacggg cggaaacctt tgcaaatgct ggattttctg cctgtggaca gccctcaaa 420
tgtcaatagc tcgcccctc atctgtcagc actctgccc tcaagtgtca aggatcgcgc 480
ccctcatctg tcagtagtgc cccccctcaa gtgtcaatac cgcagggcac ttatccccag 540
gcttgctcac atcatctgtg gaaaactcgc gtaaaatcag gcgttttcgc cgatttgcca 600
ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgccgc 660
gccgggtgag tcggcccctc aagtgtcaac gtcccctcct catctgtcag tgagggccaa 720
gttttccgcg aggtatccac aacgcggcg gcccggtgt ctcgcacacg gcttcgacgg 780
cgtttctggc gcgtttgcag ggccatagac ggccgcagc ccagcggcga gggcaaccag 840
cccgtgagc gtccgaaagg cgtctctcgg cttcctcgtc cactgactcg ctgcgctcgg 900
tcgttcggct gcggcgagcg gtatcagctc actcaaaggc ggaataacgg ttatccacag 960
aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc 1020
gtaaaaaggc cgcgttgctg gcgtttttcc ataggctcgg cccccctgac gagcatcaca 1080
aaaaatcgac ctcaagttag aggtggcgaa acccgacagg actataaaga taccaggcgt 1140
ttccccctgg aagctccctc gtgcgctcct ctgttccgac cctgccgctt accggatacc 1200

```

-continued

tgtccgcctt tctcccttcg ggaagcgtgg cgccattcgc cattcaggct gcgcaactgt	1260
tgggaagggc gatcgggtgcg ggcctcttcg ctattacgcc agctggcgaa aggggatgt	1320
gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaacg	1380
acggccagtg aattcggccg cgggtctggc gaggtgctgg gcaagcccta cctccaggcc	1440
cccatcgggg tcgagagcac gaccgccttc ctgcgccgcc tgggcgagat tctgggcctc	1500
gatccggagc ctttcacgca gcgcgagaag cactcgacgc tgaagcccgt gtgggatctg	1560
tggcggagtg tcacgcagga cttcttcggg acggccaatt tcggaatcgt ggcgaccgaa	1620
acttatgcaa gaggcacccg aaactatctc gaaggcgatc tcgggctgcc ctgcgccttc	1680
gccgtggccc gcaagagggg ctcgaaagacc gacaacgaag cgggtgcgcg actgatccgc	1740
cagcaccgtc cgctcgtgct catggggtcg atcaacgaga agatttacct tgcggaactg	1800
aaagccggtc acggcccgca accctcttcc atcgctgcct ctttcccggg tgcggcgatc	1860
cggcgcgcta ccggaacgcc cgttatggga tatgcaggtg ctacgtgggt actgcaggaa	1920
gtttgcaacg ccctgttcga cgccctgttc cacattctgc ccctcgggac ggagatggac	1980
agcgcgcg ccacaccgac gacactgcgc gcgcacttcc cgtgggatgc cgatgcgcaa	2040
gcggccctgg accgcacgt agaggagcat ccggttctca cccggatcag cgcgcgcgct	2100
gccttgccgc acgcccgca gaaggtgcc ctcgatgccg gtgccgagag ggtcgtgaga	2160
gagactgtcg aagccctgcg tgggcggggc ttcggcgaga ggaagggaga gaaccaatga	2220
gcgatcatgc cgtcaacacg ccggtccatg ccgccagggc ccacgggac cgagcaccac	2280
gtgcgagtt ctacgtctac ttcgcgctca ttctgctggg cgccttcccg gtggccttcg	2340
tgagctggat cgtctcgacg atccgccacc gcaggcttcc caagcgcggc cccttcgctg	2400
ccgctgggtt cgatgccaa ggcgatcagc cgctgatttt ccgcgcctga ccgcaggtca	2460
ggttgcgaca cgccattcgt cgtctcccca agggcgggcg gattaatcgg gagggcatgg	2520
tgccctaccg taaccacgc caccagcatg tggaggatcg ccatgcacca ccaccaccac	2580
caccacgaga acctgtactt ccaatccttt acgtagaaat agggaagatc ttgagggccg	2640
gccctccgtc gcggcgcca cccacgccg catcgattcc aaggttcagc cattgagacg	2700
gctccgcttc gcgcgcaagc gcgggttggg ccgactgcaa gcgagaggg aagcatggca	2760
ctgctcagct tcgagcgaaa atatcgctg ccggggggca cgctggctcg cggaaacctg	2820
ttcgacttct gggtcggccc tttctatgtc ggcttcttcg gggttgcgac gtttttcttc	2880
gcggccctgg gtatcattct gattgcctgg agtgccgtac tccagggtac ctggaacccc	2940
caactcatct ctgtctaccg gccggccctt gaatatggcc tgggaggtgc acccctcgca	3000
aaaggcgggc tgtggcagat catcaogatc tgcgccactg gtgccttcgt cagctgggcg	3060
ctgcgcgaag tcgaaatctg ccgtaagctg ggcacggtg accacatccc gttcgccttc	3120
gcgttcgcca tcctggccta cctgacgctg gtgctgttcc gcccggtgat gatggcgccc	3180
tggggctatg ctttccccta cgggatctgg acgcacctcg actgggtgtc gaacacgggc	3240
tacacctacg gcaacttcca ctacaacctt gccacatga tcgccatctc gttcttcttc	3300
acgaacgcgc tggctctggc gctgcacggc gcccttgtgc tctccgcggc caaccccgag	3360
aagggaag aaatcgcgac gccggatcac gaggatacgt tcttccgca tctggtcggc	3420
tactcgatcg ggacgctcg catccaccgc ctgcgcctgc tgctctcgt gagcgcgctc	3480

-continued

ttcttcagcg ccctctgcat gatcattacc ggcacccatct ggttcgatca gtgggtcgac	3540
tggtggcaat ggtgggtgaa gctgcccgtgg tgggcgaaca tcccgggagg catcaatggc	3600
tgagtatcag aacatcttct cccaggtcca ggtccgcgga ccggccgacc tggggatgac	3660
cgaagacgtc aacctggcca accgttcggg cgtcgggtccc ttctcgaccc tgctcggctg	3720
gttcggcaac gccacgctcg gcccgatcta tctcggctcg ctcggcgtcc tgtccctctt	3780
ctcgggcctg atgtggttct tcaccatcgg gatctggttc tggtatcagg cgggctggaa	3840
cccggccgtc ttctcgcgag acctgttctt cttctcgtc gagccgccgg caccogaata	3900
cggctctgtcc ttcgcggctc cgctgaagga aggcgggctg tggctgatcg cgtcgttctt	3960
catgttcgtc gcggtctggt cctggtgggg ccgcacctat ctcgcgctc aggcgctggg	4020
catgggcaag cacaccgctt gggcgttctt ctcggccatc tggctgtgga tgggtctggg	4080
cttcatccgt ccgatcctca tggggtcctg gtcggaagcg gttccctacg gcatcttctc	4140
gcacctcgac tggacgaaca acttctcgtt cgtccacggc aacctgttct acaaccctt	4200
ccacggctc tcgatcgcct tcctctacgg gtcggccctg ctcttcgcga tgcacgggtc	4260
gaccatcctc gcggtctccc gcttcggcgg cgagcgcgag ctggagcaga tcgccgaccg	4320
cgggacggca gcggagcggg ccgccctctt ctggcgtgag accatgggtt tcaacgccac	4380
gatggaagcg atccaccgct gggccatctg gatggcggtc ctcgtgaccc tcaccggcgg	4440
catcgggatc ctgctctcgg gcacggctgt ggacaactgg tacgtctggg gccagaacca	4500
cggcatggcg ccgctgaact gaggagcgt cacaatggct gacaagacca tcttcaacga	4560
tcacctcaac accaatccga agaccaacct tcgctctggt gtcgcttctc agatgatgaa	4620
gggtgcgggc tgggctggcg gcgtgttctt cgggacgctc cttctcatcg ggttcttccg	4680
ggtggtcggg cggatgcttc cgatccagga gaaccaggct ccggcgcga acatcaccgg	4740
cgctctggag accgggatcg agctgatcaa gcatctcgtc tgagacaagt ctcggggcag	4800
ggcggcgcga ggccgccgc tcctccaagt ccgggccata tcgccgccc gggtcgggg	4860
cgacaccaca gcccggttcc cttctggtg gcgacaggga cctggtgccg tgtggaagac	4920
cgacggcac cttttgaca ttcacgggag gctctgatga ccaatcccac ccgcgaccc	4980
gaaacccgc ttttgatcg cgtctgctgc ccggccgaca tgaaggcgt gagtgacgcc	5040
gaaactggag ggctggccga cgaagtgcgt tccgaggtca gtgataggg tagtttctta	5100
ttttaggcag tttatatgaa attaagacat gcagatgtca cagtggatat tgaactggtc	5160
tcgaaagctc aatatcccc aaagcacaag cacaacttc gacatcatgc agaagcgttt	5220
cccgaaccgc gtcttcgacg tgggcatcgc cgagcagcat gccgtgacct tcgcggccgg	5280
cctcgcggg gccgggatga agccctctg cgcgatctat tcctcgttcc tgcaacgggg	5340
ttacgaccag atcgcccatg acgtggcgt gcagaacctt cccgtccgct tcgtgatcga	5400
ccgggcggg ctcgtggggg ccgatggcgc gacctatcg gggcctctc acgttggtt	5460
catcaactcg ctgcccaaca tgaccgtgat ggcgcggcc gacgagccg agctcatcca	5520
catgatcgcc accgcctggt ccttcgacga gggcccatc gccttccgct tcccgcggg	5580
cgaggggggt ggcgtcgaga tgcccagcg cgggacggtg ctggagccc gcccgggccc	5640
cgtggtcgc gaaggacgag atgtcgcgat cctctcctt ggccgcgcat tgcacgaggc	5700
cttcgagcgc gcgaaacttc tcgagggcga gggggtgagc gtgacctgg ccgacgccg	5760

-continued

cttctcgcgc	ccgctcgaca	cggggctcat	cgaccagetc	gtgcgccatc	acgcggcgct	5820
ggtaacggtg	gagcaggggg	ccatggggcg	cttcggcgcc	catgtcatgc	actatctcgc	5880
caattccggc	ggcttcgacg	ggggcctcgc	gctccgggtc	atgacgctgc	ccgaccgctt	5940
catcgagcag	gcgagccccg	aggacatgta	tgccgatgcy	gggctgcggg	ccgaggatat	6000
caagcttggc	gtaatcatgg	tcatagctgt	ttcctgtgtg	aaattgttat	ccgctcacia	6060
ttccacacia	catacgagcc	ggaagcataa	agtgtaaagc	ctgggggtgc	taatgagtga	6120
gctaactcac	attaattgcy	ttgcgctcac	tgcccgtttt	ccagtcggga	aacctgtcgt	6180
gccagctgca	ttaatgaatc	ggccaacgcy	cggggagagg	cggtttgcgt	attgggcgct	6240
cggctctgcc	ttgctcgtcg	gtgatgtact	tcaccagctc	cgcgaagtcg	ctcttcttga	6300
tgagcgcgat	ggggacgtgc	ttggcaatca	cgcgcacccc	ccggccgttt	tagcggctaa	6360
aaaagtcatg	gctctgccct	cgggcggacc	acgcccatac	tgaccttgcc	aagctcgtcc	6420
tgcttctctt	cgatcttcgc	cagcagggcg	aggatcgtgg	catcaccgaa	ccgcgcgctg	6480
cgcgggtcgt	cggtagacca	gagtttcagc	aggccgcccc	ggcggcccag	gtcgcattg	6540
atgcgggcca	gctcgcggac	gtgctcatag	tccacgacgc	ccgtgatttt	gtagccctgg	6600
ccgacggcca	gcaggtaggc	cgacaggctc	atgccggccg	ccgccgctt	ttcctcaatc	6660
gctcttcggt	cgtctggaag	gcagtacacc	ttgatagtg	ggctgccctt	cctggttggc	6720
ttggtttcat	cagccatccg	cttgccctca	tctgttacgc	cggcggtagc	cgccagcct	6780
cgcagagcag	gattcccgtt	gagcaccgcc	agggtgcgaat	aaggacagct	gaagaaggaa	6840
caccgcctcg	cgggtgggcc	tacttcacct	atcctgcccg	gctgacgccg	ttggatacac	6900
caagaaaagt	ctacacgaac	cctttggcaa	aatcctgtat	atcgtgcgaa	aaaggatgga	6960
tataccgaaa	aaatcgctat	aatgaccocg	aagcagggtt	atgcagcgga	aaagcgcac	7020
gcttcccgaa	gggagaaagg	cggacaggta	tccggtaagc	ggcagggctc	gaacaggaga	7080
gcgcacgagg	gagcttccag	gggaaacgc	ctggtatctt	tatagtcctg	tcgggtttcg	7140
ccacctctga	cttgagcgtc	gatttttgtg	atgctcgtca	ggggggcgga	gcctatggaa	7200
aaagccagc	aacgcggcct	ttttaoggtt	cctggccttt	tgctggcctt	ttgctcacat	7260
gttctttcct	gcgttatccc	ctgattctgt	ggataaccgt	attaccgect	ttgagtgagc	7320
tgataccgct	cgcccgagcc	gaacgaccga	gcgcagcgag	tcagtgcgcy	aggaagcgga	7380
agagcggcag	aaggccgcca	gagagggcca	gcgcggccgt	gaggcttggc	cgctagggca	7440
gggcatgaaa	aagcccgtag	cgggctgcta	cgggcgtctg	acgcggtgga	aagggggagg	7500
ggatgttgtc	tacatggctc	tgctgtagt	agtgggttgc	gctccggcag	cggctcctgat	7560
caatcgtcac	cctttctcgg	tccttcaacg	ttcctgacaa	cgagcctcct	tttgcctaat	7620
ccatcgacia	tcaccgagag	tccttgcctg	aacgctgcgt	ccggaccggc	ttcgtcgaag	7680
gcgtctatcg	cggcccgcga	cagcggcgag	agcggagcct	gttcaacggt	gcccgcgcgc	7740
tcgocggcat	cgctgtcgc	ggcctgctcc	tcaagcacgg	ccccaacagct	gaagtagctg	7800
attgtcatca	gcgattgac	ggcgtccccg	gccgaaaaac	ccgctcgcga	gaggaagcga	7860
agctgcgcgt	cggccgtttc	catctgcggt	gcgcccggtc	gcgtgccggc	atggatgcgc	7920
gcgcatcgc	ggtaggcgag	cagcgcctgc	ctgaagctgc	gggcattccc	gatcagaaat	7980
gagcggcag	cgctcgtcgc	tctcggcacc	gaatgcgtat	gattctcgcg	cagcatggct	8040

-continued

tcggccagtg	cgctcgagcag	cgccccgcttg	ttcctgaagt	gccagtaaag	cgccggctgc	8100
tgaaccccc	accgttccgc	cagtttgctg	gtcgtcagac	cgtctacgcc	gacctcgttc	8160
aacaggccca	ggggcgccag	gatcactgta	ttcggctgca	actttgtcat	gcttgacact	8220
ttatcactga	taaacataat	atgtccacca	acttatcagt	gataaagaat	ccgcgcgttc	8280
aatcggacca	gcggaggctg	gtccggaggc	cagacatgaa	acccaacata	cccctgatcg	8340
taattctgag	cactgtcgcg	ctcgcagctg	tcggcatcgg	cctgattatg	ccgggtcctg	8400
cgggcctcct	gcgcgatctg	gttcaactcga	acgacgtcac	cgcccactat	ggcattctgc	8460
tgggcgtgta	tcggttggtg	caatttgctt	gcgcacctgt	gctgggcgcg	ctgtcggatc	8520
gtttcggggc	gcggccaatc	ttgctcgtct	cgctggccgg	cgccactgtc	gactacgcca	8580
tcatggcgac	agcgcctttc	ctttgggttc	tctatatcgg	gcggatcgtg	gccggcatca	8640
ccggggcgac	tggggcggta	gccggcgctt	atattgccga	tatcactgat	ggcgatgagc	8700
gcgcgcggca	cttcgccttc	atgagcgcct	gtttcgggtt	cgggatggtc	gcgggacctg	8760
tgctcgggtg	gctgatgggc	ggtttctccc	cccacgctcc	gttcttcgcc	gcggcagcct	8820
tgaacggcct	caatttcctg	acgggctggt	tccttttgcc	ggagtcgcac	aaaggcgaac	8880
gcccggcgtt	acgcggggag	gctctcaacc	cgctcgtctc	gttccggtgg	gcccggggca	8940
tgaccgtcgt	cgccgcctct	atggcggctt	tcttcatcat	gcaactgttc	ggacaggtgc	9000
cgcccgcgct	ttgggtcatt	ttcggcgagg	atcgctttca	ctgggacgcg	accacgatcg	9060
gcatttcgct	tgccgcattt	ggcattctgc	attcactcgc	ccaggcaatg	atcaccggcc	9120
ctgtagccgc	ccggctcggc	gaaaggcggg	cactcatgct	cggaatgatt	gccgacggca	9180
caggctacat	cctgcttgcc	ttcgcgacac	gggatggat	ggcgttcccg	atcatggtcc	9240
tgcttgcttc	gggtggcctc	ggaatgccgg	cgctgcaagc	aatgtgttcc	aggcaggtgg	9300
atgaggaacg	tcaggggcag	ctgcaaggct	cactggcggc	gctcaccagc	ctgacctcga	9360
tcgtcggacc	cctcctcttc	acggcgatct	atgcggcttc	tataacaacg	tggaacgggt	9420
gggatggatg	tgccagcgct	gccctctact	tgctctgctt	gccggcgctg	cgctcggggc	9480
tttgagcgcg	cgccagggca	cgagccgatc	gctgatcgtg	gaaacgatag	gcctatgcca	9540
tcgggtcaaa	ggcgacttcc	ggcaagctat	acgcgcctca	ggagtgcggg	tggaacggtt	9600
gcccagccag	atactcccga	tcacgagcag	gacgccgatg	atttgaagcg	cactcagcgt	9660
ctgatccaag	aacaaccatc	ctagcaaac	ggcggctccc	gggctgagaa	agcccagtaa	9720
ggaaacaact	gtaggttcga	gtcgcgagat	cccccggaac	caaaggaagt	aggttaaacc	9780
cgctccgatc	aggccgagcc	acgccaggcc	gagaacattg	gttctctgat	gcatcgggat	9840
tggggatca	aacactaaag	ctactggaac	gagcagaagt	cctccggccc	ccagttgcca	9900
ggcggtaaa	gtgagcagag	gcacgggagg	ttgccacttg	cggttcagca	cggttccgaa	9960
cgccatggaa	accgcccccg	ccaggccccg	tgccgacccg	acaggatcta	gcgctcgttt	10020
tggtgtcaac	accaacagcg	ccacgccccg	agttccgcaa	atagccccca	ggaocgccat	10080
caatcgtatc	gggctaccta	gcagagcggc	agagatgaac	acgaccatca	gcggctgcac	10140
agcgcctacc	gtcgcgcgca	ccccgccccg	caggcggtag	accgaaataa	acaacaagct	10200
ccagaatagc	gaaatattaa	gtgcgcccag	gatgaagatg	cgcatccacc	agattcccgt	10260
tggaatctgt	cggacgatca	tcacgagcaa	taaaccggcc	ggcaacgccc	gcagcagcat	10320

-continued

accggcgacc cctcggcctc gctgttcggg ctccacgaaa acgccggaca gatgcgccctt	10380
gtgagcgtcc ttggggccgt cctcctgttt gaagaccgac agcccaatga tctcgcctgc	10440
gatgtaggcg ccgaatgcc a cggcatctcg caaccgttca gcgaacgcct ccatgggctt	10500
ttctcctcgc tgctcgtaaa cggaccgaa catctctgga gctttcttca gggccgacaa	10560
tcggatctcg cggaaatcct gcacgtcggc cgctccaagc cgtcgaatct gagccttaat	10620
cacaattgtc aattttaatc ctctgtttat cggcagttcg tagagcgcgc cgtgctccc	10680
gagcgatact gagcgaagca agtgcgtcga gcagtgcccc cttgttctcg aaatgccagt	10740
aaagcgtgg ctgctgaacc cccagccgga actgaccca caaggcccta gcgtttgcaa	10800
tgaccagggt catcattgac ccaggcgtgt tccaccaggc cgctgcctcg caactcttcg	10860
caggcttcgc cgacctgctc gcgccacttc ttcacgcggg tggaatccga tccgcacatg	10920
aggcgaaggt ttccagctt gagcgggtac ggctcccggg gcgagctgaa atagtcgaac	10980
atccgtcggg ccgtcggcga cagcttgcgg tacttctccc atatgaattt cgtgtagtgg	11040
tcgccagcaa acagcagcgc gatttctcgc tcgatcagga cctggcaacg ggacgttttc	11100
ttgccacggt ccaggacgcg gaagcgggtc agcagcgaca ccgattccag gtgcccaacg	11160
cggtcggacg tgaagcccat cgcctcgcgc tgtaggcgcg acaggcattc ctccgccttc	11220
gtgtaatacc ggccattgat cgaccagccc aggtcctggc aaagctcgtg gaacgtgaag	11280
gtgatcggct cgcgataggt ggtgcgcttc gcgtactcca acacctgctg ccacaccagt	11340
tcgtcatcgt cggcccgcag ctgcagccgc gtgtaggtga tcttcacgtc cttggtgacg	11400
tggaaaatga ccttgttttg cagcgcctcg cgcgggattt tcttggtgcg cgtggtgac	11460
agggcagagc gggccgtgct gtttgcatc gctcgcctcg tgcctggcca cggcgcaata	11520
tcgaaacaag aaagctgcat ttccttgatc tgctgcttcg tgtgtttcag caacgcggcc	11580
tgcttgccct cgtgacctg ttttgccagg tccctgcggc cggtttttcg cttcttggtc	11640
gtcatagttc ctgcgctgct gatggctatc gacttcgcca aacctgcgc ctctgttcg	11700
agacgacgcg aacgtccac ggcggccgat ggcgcggca gggcagggg agccagttgc	11760
acgtgtcgc gctcgtatct ggcgtagct tgctggacca tcgagccgac ggaactggaag	11820
gtttcgcggg gcgcacgcat gacggtgcgg cttgcgatgg tttcggcatc ctccgcggaa	11880
aaccccgcgt cgatcagttc ttgcctgtat gccttcgggt caaacgtccg attcattcac	11940
cctccttgcc ggattgcccc gactcacgcc ggggcaatgt gcccttattc ctgatttgac	12000
ccgcctgggt ccttggtgct cagataatcc acctatcgg caatgaagtc ggtcccgtag	12060
accgtctggc cgtccttctc gtacttggtt tccgaatct tgccctgcac gaataccagc	12120
gacccttgcc caaataactt gccgtgggcc tcggcctgag agccaaaaca cttgatgccc	12180
aagaagtcgg tcgctcctg cttgtcgcgc gtcgtggccg cgccaacctt tcgcatccgc	12240
aagcgcggc tcgccatctt cacgctggaa cagtacgtcg aggcgggcat catgaccgc	12300
gagcaatacg aggtcattaa aagcgcctg attgatgata tagcggcccg gctgctcctg	12360
gttctcgcgc accgaaatgg gtgacttcac cccgcgctct ttgatcgtgg caccgatctc	12420
cgcgatgctc tccggggaaa agccggggtt gtcggccgct cgcggctgat gcggatcttc	12480
gtcgtacagc tccaggtcca gctcgtatag gccggaaccg ccctgagacg ccgcaggagc	12540
gtccaggagg ctgcacaggt cgcctgatct atccaacccc aggcgggacg gctgcgcgc	12600

-continued

```

gcctgcgget tcctgagcgg ccgcagcggg gtttttcttg gtggtcttgg cttgagccgc 12660
agtcattggg aaatctccat cttcgtgaac acgtaatcag ccagggcgcg aacctctttc 12720
gatgccttgc gcgcggccgt tttcttgatc ttccagaccg gcacaccgga tgcgagggca 12780
tcggcgatgc tgctgcgcag gccaacgggt gccggaatca tcactcttggg gtacgcggcc 12840
agcagctcgg cttggtggcg cgcgtggcgc ggattccgcg catcgacctt gctgggcacc 12900
atgccaagga attgcagcct gccgttcttc tggcgcacgt tcgcaatggg cgtgaccatc 12960
ttcttgatgc cctggatgct gtacgcctca agctcgtatg gggacagcac atagtccggc 13020
gcgaagaggg cggccgccag gccgacgcca agggtcgggg ccgtgtcgat caggcacacg 13080
tcgaagcctt ggttcgccag gcccttgatg ttcgccccga acagctcgcg ggcgtcgtcc 13140
agcgacagcc gttcggcgtt cgccagtacc gggttggact cgatgagggc gaggcgcgcg 13200
gcctggccgt cgcggctgc gggtcgggtt tcggtccagc cgcgcgcagg gacagcgcg 13260
aacagcttgc ttgcatgcag gccggtagca aagtccttga gcgtgtagga cgcattgccc 13320
tgggggtcca ggtgatcac ggcaaccgc aagccgcgct cgaaaaagtc gaaggcaaga 13380
tgcaaaagg tcgaagtctt gccgacgcg cctttctggt tggccgtgac caaagttttc 13440
atcgtttggt ttctgtttt ttcttggcgt ccgcttccca cttccggacg atgtacgcct 13500
gatgttccgg cagaaccgcc gttaccgcg cgtaccctc gggcaagttc ttgtcctcga 13560
acgcggccca cagcgatgc accgcttgcg acactgcgcc cctggtcagt cccagcgcg 13620
ttgcgaacgt cgcctgtgac ttcccatcga ctaagacgcc ccgcgctatc tcgatgttet 13680
gctgccccac ttccagcccc tggatgcct cctggaactg gctttcggtg agccgtttct 13740
tcatggataa caccataat ttgctccgcg ccttgggtga acatagcggg gacagccgcc 13800
agcatatgag agaagtttag ctaaaccatt ctcgcacgtc aacacctta gccgctaaaa 13860
ctcgtccttg gcgtaaaaa acaaaagccc ggaaccggg cttctgtctc ttgccgctta 13920
tggctctgca cccggtcca tcaccaacag gtcgcgcagc cgcttcactc ggttgcggat 13980
cgacactgcc agcccaaaa agccggttgc cgcgcgcccc aggatcgcgc cgatgatgcc 14040
ggccacaccg gccatgcgcc accaggtcgc cgccttccgg ttccattcct gctggtactg 14100
cttcgcaatg ctggacctcg gctcaccata ggctgaccgc tcgatggcgt atgcccttc 14160
tccccttggc gtaaaaccca gcgccgagg cggcattgcc atgctgcgcg ccgctttccc 14220
gaccacgacg cgcgcaccag gcttgcggtc cagaccttcg gccacggcga gctgcgcaag 14280
gacataatca gccgccgact tggctccacg cgcctcgcac agctcttgcg ctcgcgcgaa 14340
atccttggcc tccacggcgc ccatgaatcg cgcacgcggc gaaggctcgc cagggcgcg 14398

```

<210> SEQ ID NO 16

<211> LENGTH: 14391

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic plasmid

<400> SEQUENCE: 16

```

cccccaggc cgcgccctc actgcccgc acctggtcgc tgaatgtcga tgccagcacc 60
tgcggcagct caatgcttcc gggcgtcgcg ctcgggctga tcgcccatcc cgttactgcc 120
ccgatcccg caatggcaag gactgccagc gctgccattt ttggggtgag gccgttcgcg 180

```

-continued

gccgaggggc gcagcccctg ggggatggg aggcccgct tagcgggccc ggagggttcg	240
agaagggggg gcacccccct tcggcgtgcg cggtcacgcg cacagggcgc agccctggtt	300
aaaaacaagg ttataaata ttggtttaa agcaggttaa aagacaggtt agcggtggcc	360
gaaaaacggg cggaaaccct tgcaaatgct ggattttctg cctgtggaca gcccctcaaa	420
tgtaaatagg tgcgccctc atctgtcagc actctgcccc tcaagtgtca aggatcgcgc	480
ccctcatctg tcagtagtcg cgcacctcaa gtgtcaatac cgcagggcac ttatccccag	540
gcttgccac atcatctgtg gaaaactcgc gtaaaatcag gcgttttcgc cgatttgca	600
ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgccgc	660
gccgggtgag tcggcccctc aagtgtcaac gtccgccctc catctgtcag tgagggccaa	720
gttttcgcgc aggtatccac aacgccggcg gcccggtgt ctcgcacacg gcttcgacgg	780
cgtttctggc gcgtttgcaq ggccatagac ggccgccagc ccagggcga gggcaaccag	840
cccggtgagc gtcgaaagg cgtctctccg cttcctcgtc cactgactcg ctgcgctcgg	900
tcgttcggct gcggcagcg gtatcagctc actcaaaggc ggtaatacgg ttatccacag	960
aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc	1020
gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg cccccctgac gagcatcaca	1080
aaaaatcgac ctcaagtcaq aggtggcgaa acccgacagg actataaaga taccaggcgt	1140
ttccccctgq aagctccctc gtgcctctc ctgttccgac cctgccgctt accggatacc	1200
tgccgcctt tctcccttcg ggaagcgtgg cgcattcgc cattcaggct gcgcaactgt	1260
tgggaagggc gatcggtgcg gccctcttcg ctattacgcc agctggcgaa aggggatgt	1320
gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg	1380
acggccagtg aattcggccc cgggctggcc gagtgctgg gcaagcccta cctccaggcc	1440
cccatcgggg tcgagagcac gaccgccttc ctgcgccgcc tgggcgagat tctgggcctc	1500
gatccggagc cttcatcga gcgcgagaag cactcgacgc tgaagccctg gtgggatctg	1560
tggcggagtg tcacgcagga cttcttcggg acggccaatt tcggaatcgt ggcgaccgaa	1620
acttatgcaa gaggcacccg aaactatctc gaaggcagtc tcgggctgcc ctgcgccttc	1680
gccgtggccc gcaagagggg ctgaaagacc gacaacgaag cggtgccggg actgatccgc	1740
cagcacccgc cgtcgtgctc catggggtcg atcaacgaga agatttacct tgcggaactg	1800
aaagccggtc acggcccga accctcttcc atcgtgcct ctttcccggg tgcggcgatc	1860
cggcgcgcta ccggaaccgc cgttatggga tatgcagtg ctacgtggtt actgcaggaa	1920
gtttgcaacg ccctgttcga cgcctgttc cacattctgc ccctcgggac ggagatggac	1980
agcgcggcgc ccacaccgac gacctgcgc cgcgacttcc cgtgggatgc cgatgcgcaa	2040
gcggccctgq accgcatcgt agaggagcat ccggttctca cccggatcag cgcgcgcgt	2100
gccttgccgc acgcccga gaaggctgcc ctcgatgcc gtgcccagag ggtcgtgaga	2160
gagactgtcg aagccctgcg tgggcccggc ttcggcgaga ggaagggaga gaaccaatga	2220
gcgatcatgc cgtcaacacg ccggtccatg ccgccagggc ccacgggcac cgagcaccac	2280
gtgccgagtt ctacgtctac ttcgcgtca ttctgctggg cgccttcccg gtggccttcg	2340
tgagctggat cgtctcagc atccgccacc gcaggcttcc caagcgcggc cccttcgcgt	2400
ccgctcgtt cgatgccaaq gcgatcacgc cgtgatctt ccgcccctga ccgaggtca	2460

-continued

ggttgcgaca	cgccattcgt	cgtctcccca	aggggcggcg	gattaatcgg	gagggcatgg	2520
tgcottaccg	taaccacagc	caccagcatg	tggagatcg	ccatgcacca	ccaccaccac	2580
caccacgaga	acctgtactt	ccagtccgct	agctaggaa	gatcttgagg	ccggccctcc	2640
gtcgcgggcg	gcaccacagc	ccgcatcgat	tccaaggttc	agccattgag	acggctccgc	2700
ttcgcgcgca	agcgcgggtt	gggccgactg	caagcggaga	gggaagcatg	gcactgctca	2760
gcttcgagcg	aaaataticg	gtgccggggg	gcacgctggt	cggcggaaac	ctgttcgact	2820
tctgggtcgg	ccctttctat	gtcggcttct	tcgggggtgc	gacgtttttc	ttcgcggccc	2880
tgggtatcat	tctgattgcc	tggagtgcgg	tactccaggg	tacctggaac	ccccaaactca	2940
tctctgtcta	ccgcgccggc	cttgaatatg	gcctgggagg	tgacccctc	gcaaaaggcg	3000
ggctgtggca	gatcatcacg	atctgcgcca	ctggtgcctt	cgtcagctgg	gcgctgcgcg	3060
aagtcgaaat	ctgccgtaag	ctgggcatcg	ggtaccacat	cccgttcgcc	ttcgcgttcg	3120
ccatcctggc	ctacctgacg	ctggtgctgt	tccgcccggg	gatgatgggc	gcctggggct	3180
atgccttccc	ctacgggacg	tggacgcacc	tcgactgggt	gtcgaacacg	ggctacacct	3240
acggcaactt	ccactacaac	cctgcccaca	tgatcgccat	ctcgttcttc	ttcacgaacg	3300
cgctggctct	ggcgtgcac	ggcgccttg	tgctctccgc	ggccaacccc	gagaagggca	3360
aggaaatgcg	gacgcccgat	cacgaggata	cgttcttcgg	cgatctggtc	ggctactcga	3420
tcgggacgct	cgcatccac	cgctcggcc	tgctgctctc	gctgagcgc	gtcttcttca	3480
gcgccctctg	catgatcatt	accggcacca	tctggttcga	tcagtgggtc	gactggtggc	3540
aatggtgggt	gaagctgcgg	tgggtggcga	acatcccggg	aggcatcaat	ggctgagtat	3600
cagaacatct	tctcccaggt	ccaggtccgc	ggaccggccg	acctggggat	gaccgaagac	3660
gtcaacctgg	ccaaccgttc	ggcgtcggg	cccttctcga	ccctgctcgg	ctggttcggc	3720
aacgccacg	tcggcccgat	ctatctcggc	tcgctcggcg	tcctgtccct	cttctcgggc	3780
ctgatgtggt	tcttcacat	cgggatctgg	ttctggtatc	aggcgggctg	gaaccggcc	3840
gtcttcctgc	gcgacctgtt	cttcttctcg	ctcagccgc	cggcaccoga	atacggctctg	3900
tccttcggcg	ctccgctgaa	ggaaggcggg	ctgtggctga	tcgctcgtt	cttcatgttc	3960
gtcgcggtct	ggtcctggtg	ggccgcacc	tatctccgg	ctcagggcgt	ggcatgggc	4020
aagcacaccg	cctgggcggt	cctctcggcc	atctggctgt	ggatggtgct	gggcttcac	4080
cgccgatcc	tcatggggtc	ctggtcggaa	gcggttcct	acggcatctt	ctcgcacctc	4140
gactggacga	acaacttctc	gctcgtccac	ggcaacctgt	tctacaaacc	cttccacggt	4200
ctctcgatcg	ccttctctca	cgggtcggcc	ctgctctcgg	cgatgcacgg	tgcgaccatc	4260
ctcgcggtct	cccgtctcgg	cggcagcgc	gagctggagc	agatcgcoga	ccgcgggacg	4320
gcagcgggag	ggcccccctt	cttctggcgc	tggaccatgg	gtttcaacgc	cacgatggaa	4380
ggcatccacc	gctgggcca	ctggatggcg	gtcctcgtga	ccctcaccgg	cggcatcggg	4440
atctgctct	cgggcacggt	cgtggacaac	tggtagctct	ggggccagaa	ccacggcatg	4500
gcgcgcgctg	actgaggagc	gatcacaatg	gctgacaaga	ccatcttcaa	cgatcacctc	4560
aacaccaatc	cgaagaccaa	ccttcgcctc	tgggtcgtct	tccagatgat	gaagggtgcg	4620
ggctgggctg	gcggcgtggt	cttcgggacg	ctccttctca	tcgggttctt	ccgggtggtc	4680
ggcgggatgc	ttccgatcca	ggagaaccag	gctccggcgc	cgaacatcac	cggcgtctctg	4740

-continued

gagaccggga tcgagctgat caagcatctc gtctgagaca agtctcgggg cagggcggcg 4800
cgaggccgcc cgctcctcca agtccgggcc atatcgccgg cccgggtccg gggcgacacc 4860
acagcccggg tcccttcctg ttggcgacag ggacctggtg ccgtgtggaa gaccgcacgg 4920
cacccttttg acattcacgg gaggetctga tgaccaatcc caccocgca cccgaaaccc 4980
cgcttttgga tcgctctgctg tgcggggccg acatgaaggc gctgagtgc gccgaactgg 5040
agcggctggc cgacgaagtg cgttcgagg tcagtgatag gggtagttc ttattttagg 5100
cagtttatat gaaattaaga catgcagatg tcacagtga tattgaactg gtctcgaag 5160
ctcaatatcc cccaaagcac aagcacaac ttcgacatca tgcagaagcg tttcccgaac 5220
cgctcttcg acgtgggcat cgcgagcag catgccgtga ccttcgccc cggcctcgcc 5280
ggggccggga tgaagccct ctgcgcgac tattcctcgt tcctgcaacg gggttacgac 5340
cagatcgccc atgacgtggc gctgcagaac cttcccgtcc gcttcgtgat cgaccggggc 5400
gggctcgtgg gggccgatgg cgcgacccat gcggggcct tcgacgttg cttcatcact 5460
tcgctgcccc acatgaccgt gatggcccg gccgacgagg ccgagctcat ccacatgatc 5520
gccaccgcc tggccttcga cgaggcccc atgccttcc gcttcccgcg gggcgagggg 5580
gtgggcgtcg agatgcccga gcgcgggacg gtgctggagc ccggccgggg ccgctgtgtg 5640
cgcgaaggga cggatgtcgc gatcctctcc ttcggcgcgc atctgcacga ggccttgca 5700
gcgccgaaac ttctcgaggc cgaggggggtg agcgtgaccg tggccgacgc ccgcttctcg 5760
cgcccgtcgc acacggggct catcgaccag ctcgtgcgcc atcacgcggc gctggtaacg 5820
gtggagcagg gggccatggg cggttcggc gcccatgtca tgcactatct cgccaattcc 5880
ggcgcttcg acggggcct cgcgtcccg gtcatgacgc tgcggaccg cttcatcgag 5940
caggcgagcc ccgaggacat gtatgccgat gcggggctgc gggccgagga tatcaagcct 6000
ggcgtaatca tggatcatag tgttctcgt gtgaaattgt tatccgctca caattccaca 6060
caacatacga gccggaagca taaagtgtaa agcctggggg gcctaagtag tgagctaac 6120
cacattaatt gcgttgctc cactgcccgc tttccagtgc gaaacctgt cgtgccagct 6180
gcattaatga atcgccaac gcgcggggag aggcggtttg cgtattgggc gctcgtctt 6240
gccttgctcg tcggtgatgt acttcaccag ctcccggaag tcgctcttct tgatggagcg 6300
catggggacg tgcctggcaa tcacgcgcac ccccggccg ttttagcggc taaaaagtc 6360
atggctctgc cctcggcgcc accacgccc tcatgacct gccaaactcg tctgcttct 6420
cttogatctt cgccagcagg gcgaggatcg tggcatcacc gaaccgcgcc gtgcgcgggt 6480
cgtcgtgtag ccagagtttc agcaggccgc ccaggcggcc caggctgcca ttgatgcggg 6540
ccagctcgcg gacgtgctca tagtccacga cgcccgatgat tttgtagccc tggccgacgg 6600
ccagcaggta gggcgacagg ctcatgccgg ccgcccggc ctttctctca atcgtcttc 6660
gttcgtctgg aaggcagtac accttgatag gtgggctgcc cttcctggtt ggcttggtt 6720
catcagccat ccgcttgccc tcatctgta cgccggcggg agccggccag cctcgcagag 6780
caggattccc gttgagcacc gccaggtgcg aataagggac agtgaagaag gaaccccgc 6840
tcgcggtggg gcctacttca cctatcctgc ccgctgacg ccgttgata caccaaggaa 6900
agtctacacg aaccctttgg caaaatcctg tatactctgc gaaaaaggat ggatataccg 6960
aaaaaatcgc tataatgacc ccgaagcagg gttatgcagc gaaaagcgc cacgcttccc 7020

-continued

gaaggagaaa aggcggacag gtatccggta agcggcaggg tcggaacagg agagcgcacg 7080
agggagcttc cagggggaaa cgccctggat ctttatagtc ctgtcggggt tcgccacctc 7140
tgacttgagc gtcgattttt gtgatgctcg tcaggggggc ggagcctatg gaaaaacgcc 7200
agcaacgcgg cctttttacg gttcctggcc ttttgctggc cttttgctca catgttcttt 7260
cctcgcgttat cccctgattc tgtggataac cgtattaccg cctttgagtg agctgatacc 7320
gctcgcgcga gccgaacgac cgagcgcagc gagtcagtga gcgaggaagc ggaagagcgc 7380
cagaaggccg ccagagaggc cgagcgcggc cgtgaggctt ggacgctagg gcagggcatg 7440
aaaaagcccg tagcgggctg ctacgggctg ctgacgcggg ggaagggggg aggggatgtt 7500
gtctacatgg ctctgctgta gtgagtgggt tgcgctccgg cagcggctct gatcaatcgt 7560
caccctttct cggctcctca acgctcctga caacgagcct ccttttcgcc aatccatcga 7620
caatcaccgc gactccctgc tcgaaacgctg cgtccggacc ggcttcgctg aaggcgtcta 7680
tcgcgcccg caacagcggc gagagcggag cctgttcaac ggtgccgccg cgctcgcgg 7740
catcgcgtgc gccggcctgc tcctcaagca cggccccaac agtgaagtag ctgattgtca 7800
tcagcgcatt gacggcgtcc ccggccgaaa aaccgcctc gcagaggaag cgaagctgcg 7860
cgtcggccgt ttccatctgc ggtgcgcccg gtcgcgtgcc ggcatggatg cgcgcgccat 7920
cgcggtaggc gagcagcgc tcctgaagc tgcgggcatt cccgatcaga aatgagcgc 7980
agtcgtcgtc ggctctcggc accgaatcgc tatgattctc cgcagcatg gcttcggcca 8040
gtgcgtcagc cagcgcgccg ttgttctcga agtgccagta aagcgcggc tgctgaacc 8100
ccaaccgttc cgcagtttg cgtgctgta gaccgtctac gccgacctc ttcaacaggt 8160
ccagggcggc acggatcact gtattcggct gcaactttgt catgcttgac actttatcac 8220
tgataaacat aatattgtcca ccaacttata agtgataaag aatccgcgcg ttcaatcga 8280
ccagcggagg ctggtccgga gccagacat gaaacccaac ataccctga tcgtaattct 8340
gagcactgtc gcgctcagc ctgctggcat cggcctgatt atgcccgtgc tgcgggcct 8400
cctgcgcgat ctggttcaact cgaacgacgt caccgcccac tatggcattc tgcggcgcct 8460
gtatgcgttg gtgcaatttg cctgcgcacc tgtgctggg gcgctgctgg atcgtttcgg 8520
gcccgggcca atcttctcgc tctcgtggc cggcgcact gtcgactacg ccatcatggc 8580
gacagccctc ttcttttggg ttctctatata cggcgggata gtcggcggca tcaccggggc 8640
gactggggcg gtagccggcg cttatattgc cgatatcact gatggcgatg agcgcgcgcg 8700
gcacttcggc ttcatgagcg cctgtttcgg gttcgggatg gtcgcccggc ctgtgctcgg 8760
tgggctgatg ggcggtttct cccccacgc tccgttcttc gccgcggcag ccttgaacgg 8820
cctcaatttc ctgacgggct gtttcctttt gccggagtgc cacaaaggcg aacgccggcc 8880
gttacgccgg gaggctccta accgcctcgc ttcgttccgg tggcccggg gcatgaccgt 8940
cgtcgcgccg ctgatggcgg tcttcttcat catgcaactt gtcggacagg tgcggccgcg 9000
gctttggggtc attttcggcg aggatcgcct tcaactgggac gcgaccacga tcggcatttc 9060
gcttgccgca tttggcattc tgcattcact cggccaggca atgatcaacc gccctgtagc 9120
cgcccggctc gccgaaagc gggcactcat gctcggaatg attgccgacg gcacaggcta 9180
catcctgctt gccttcgcga cacggggatg gatggcgttc ccgatcatgg tcctgcttgc 9240
ttcgggtggc atcggaatgc cggcgtgca agcaatgttg tccaggcagg tggatgagga 9300

-continued

acgtcagggg cagctgcaag gctcactggc ggcgctcacc agcctgacct cgatcgctcg 9360
acccctcctc ttcacggcga tctatgcggc ttctataaca acgtggaacg ggtgggcatg 9420
gattgcaggc gctgccctct acttgetctg cctgccggcg ctgctgctcg ggctttggag 9480
cggcgcaagg caacgagccg atcgetgato gtgaaacga taggcctatg ccatgcccgt 9540
caaggcgact tccggcaagc tatacgcgcc ctaggagtgc ggttggaacg ttggcccagc 9600
cagatactcc cgatcacgag caggacggcg atgatttgaa gcgactcag cgtctgatcc 9660
aagaacaacc atcctagcaa cacggcggtc cccgggctga gaaagcccag taaggaaaca 9720
actgtaggtt cgagtgcga gatcccccg aaccaaagga agtaggttaa acccgctccg 9780
atcaggccga gccacgccag gccgagaaca ttggttcctg taggcatcgg gattggcggg 9840
tcaaactact aagctactgg aacgagcaga agtccctcgg ccgccagttg ccaggcggta 9900
aaggtgagca gaggcacggg aggttgccac ttgctgggtc gcacggttcc gaacgccatg 9960
gaaaccgccc ccgccaggcc cgctgcgacg ccgacaggat ctagcctgct gtttggtgct 10020
aacaccaaca gcgccacgcc cgcagttccg caaatagccc ccaggaccgc catcaatcgt 10080
atcgggctac ctacgagagc gccagagatg aacacgacca tcagcggctg cacagcgcct 10140
accgtcgccg cgaccccgcc cggcaggcgg tagaccgaaa taaacaaca gctccagaat 10200
agcgaatat taagtgcgcc gaggatgaag atgcgcatcc accagattcc cgttggaatc 10260
tgtcggacga tcatcacgag caataaacc gccggcaacg cccgcagcag cataccggcg 10320
acccctcggc ctgctgttcc gggctccacg aaaacgcccg acagatgcgc cttgtgagcg 10380
tccttggggc cgtcctcctg tttgaagacc gacagcccaa tgatctcgc gtcgatgtag 10440
gcgcgcaatg ccacggcacc tcgcaaccgt tcagcgaacg cctccatggg ctttttctcc 10500
tcgtgctcgt aaacggaccc gaacatctct ggagctttct tcagggccga caatcggatc 10560
tcgcgaaat cctgcacgct gccgctcca agcctgcgaa tctgagcctt aatcacaatt 10620
gtcaatttta atcctctgtt tatcggcagt tcgtagagcg gcgctgctg cccgagcgat 10680
actgagcgaa gcaagtgcgt cgagcagtc cgcctgttcc ctgaaatgcc agtaaagcgc 10740
tggtgctga accccagacc ggaactgacc ccacaaggcc ctagcgtttg caatgcacca 10800
ggtoatcatt gaccagggc tgttccacca ggcgctgccc tcgcaactct tcgcaggctt 10860
cgccgacctg ctgcgcccac ttcttcacgc gggtggaatc cgatccgcac atgagcggga 10920
aggtttccag cttgagcggg tacggctccc ggtgcgagct gaaatagtcg aacatccgct 10980
gggocgtcgg cgacagcttg cggtacttct cccatatgaa tttcgtgtag tggtcgccag 11040
caaacagcac gacgatttcc tcgtcgatca ggacctggca acgggacggt ttcttgccac 11100
ggtccaggac gcggaagcgg tgacgagcgc acaccgattc caggtgccc aacgggtcgg 11160
acgtgaagcc catcgcgcgc gcctgtaggc gogacaggca ttcctcggcc ttcgtgtaat 11220
accggccatt gatcgaccag cccaggtcct ggcaaagctc gtagaacgtg aaggtgatcg 11280
gctcgcgat aggggtgcgc ttcgcgtact ccaacacctg ctgccacacc agttcgtcat 11340
cgtcggcccg cagctcgacg ccggtgtagg tgatcttcac gtcctgttg acgtggaaaa 11400
tgacctgtt ttgacgccc tcgcgcgga tttcttgtt gcgctggtg aacaggcag 11460
agcggccggt gtcgtttggc atcgetcgca tcgtgtccgg ccacggcgca atatcgaaca 11520
aggaaagctg catttccctg atctgctgct tcgtgtgtt cagcaacgcg gcctgcttgg 11580

-continued

cctcgtgac ctgttttgcc aggtcctcgc cggcggtttt tcgcttcttg gtcgtcatag 11640
ttcctcgcgt gtcgatggtc atcgacttcg ccaaacctgc cgcctcctgt tcgagacgac 11700
gcgaacgctc cacggcggcc gatggcgcgg gcagggcagg gggagccagt tgcacgctgt 11760
cgcgctcgat cttggccgta gcttgctgga ccatcagacc gacggactgg aaggtttcgc 11820
ggggcgcacg catgacggtg cggcttcgca tggtttcggc atcctcggcg gaaaaccccg 11880
cgtcgatcag ttcttgcttg tatgccttcc ggtcaaactg ccgattcatt caccctcctt 11940
gcgggattgc cccgactcac gccggggcaa tgtgccctta ttcctgattt gacccgcctg 12000
gtgccttggt gtccagataa tccaccttat cggcaatgaa gtcggtcccg tagaccgtct 12060
ggcgcctcct ctcgtacttg gtattccgaa tcttgccctg cacgaatacc agcgaccctt 12120
tgcccaaata cttgcctggg gcctcggcct gagagccaaa acacttgatg cggagaagt 12180
cggtcgcctc ctgcttctgc ccggtcgtgg ccgcgccaac ctttgcgatc cgcaagcgcg 12240
cggtcgccat cttcacgctg gaacagtacg tcgagggcgg catcatgacc cgcgagcaat 12300
acgaggtcat taaaagcgcg gtgattgatg atatagcggc ccgctgctc ctggttctcg 12360
cgcaccgaaa tgggtgactt caccocgcgc tctttgatcg tggcaccgat ttcgcgatg 12420
ctctccgggg aaaagccggg gttgtcggcc gtccgcggct gatgcggatc ttcgctgatc 12480
aggtccaggt ccagctcgat agggccggaa ccgccctgag acgccgcagg agcgtccagg 12540
aggctcgaca ggtcgcgat gctatccaac cccagggcgg acgctgcgc cgcgcctgcg 12600
gcttctgag cggccgcagc ggtgttttct ttggtgtct tggcttgagc cgcagtcatt 12660
gggaaatctc catcttcgct aacacgtaat cagccagggc gcgaaacctt ttcgatgcct 12720
tgcgccggcg cgttttcttg atcttccaga ccggcacacc ggatgcgagg gcatcggcga 12780
tgctgctgcg caggccaacg gtggccggaa tcatcatctt ggggtacgcg gccagcagct 12840
cggcttggtg gcgcgcgtgg cgcgattcc gcgcatcgac cttgctgggc accatgcca 12900
ggaaattcag cttggcgttc tcttgggcga cgttcgcaat ggtcgtgacc atcttcttga 12960
tgccctggat gctgtacgcc tcaagctcga tgggggacag cacatagtcg gccgcgaaga 13020
ggggggccgc caggccgacg ccaagggtcg gggccgtgtc gatcaggcac acgtcgaagc 13080
cttggttcgc cagggccttg atgttgcgcc cgaacagctc cggggcgtcg tccagcgaca 13140
gccgttcggc gttcgcagat accgggttg actcgtgag ggcgagcgc gcggcctggc 13200
cgtcgcggc tcggggtgcg gtttcggtcc agccgcggc agggacagcg ccgaacagct 13260
tgcttgcgat caggccggtg gcaaagtctt tgagcgtgta ggacgattg ccctgggggt 13320
ccaggtcgat cacggcaacc cgcaagccgc gctcgaaaaa gtcgaaggca agatgcacaa 13380
gggtcgaagt cttgcgcagc ccgcctttct ggttggccgt gaccaaagt ttcacgttt 13440
ggttctctgt ttttcttgg cgtccgcttc ccaactccgg acgatgtacg cctgatgttc 13500
cggcagaacc gccgttaccg gcgcgtaccg ctcgggcaag ttcttgcct cgaacgcggc 13560
ccacacgca tgcaccgctt gcgacactgc gccctggtc agtccagcg acgttcgaa 13620
cgtcgcctgt ggcttccat cgaactaac gcccgcgct atctcgatgg tctgctgcc 13680
cacttccagc ccctggatcg cctcctgaa ctggctttcg gtaagccgtt tcttcatgga 13740
taacacccat aatttctcc gcgccttgg tgaacatagc ggtgacagcc gccagcact 13800
gagagaagtt tagctaaaca tttctgcac gtcaaacct ttagccgcta aaactcgtcc 13860

-continued

```

ttggcgtaac aaaacaaaag cccgaaaacc gggctttcgt ctcttgccgc ttatggctct 13920
gcacccggct ccatcaccaa caggctgcgc acgcgcttca ctcggttgcg gatcgacact 13980
gccagcccaa caaagccggt tgccgccgc gccaggatcg cggcgatgat gccggccaca 14040
ccggccatcg cccaccaggt cggcccttc cggttccatt cctgctggta ctgcttcgca 14100
atgctggacc tcggtcacc ataggctgac cgctcgatgg cgtatgccgc ttctcccctt 14160
ggcgtaaaac ccagcgcgc aggcggcatt gccatgctgc ccgcccgttt cccgaccacg 14220
acgcgcgcac caggcttgcg gtccagacct tcggccacgg cgagctgcgc aaggacataa 14280
tcagccgccg acttggtctc acgcgcctcg atcagctctt gcactcgcgc gaaatccttg 14340
gcctccacgg ccgcatgaa tcgcgcacgc ggcgaaggct ccgcaaggcc g 14391

```

<210> SEQ ID NO 17

<211> LENGTH: 14377

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic plasmid

<400> SEQUENCE: 17

```

ccaccaggc cggcccttc actgcccgc acctggtcgc tgaatgtcga tgccagcacc 60
tgcggcacgt caatgcttcc gggcgtcgc ctcgggtga tcgccatcc cgttactgcc 120
ccgatcccg caatgcaag gactgccagc gctgccattt ttggggtgag gccgttcgcg 180
gccgaggggc gcagcccctg ggggatggg aggcgccgc tagcgggccg ggagggttcg 240
agaaggggg gcacccccct tcggcgtcgc cggtcacgc cacaggggc agccctggtt 300
aaaaacaag ttataaata ttggtttaa agcaggttaa aagacagggt agcggtggcc 360
gaaaaacgg cgaaacct tgcaaatgct ggattttctg cctgtggaca gccctcaa 420
tgtcaatag tcgcccctc atctgtcagc actctgccc tcaagtgtca aggatcgcgc 480
ccctcatctg tcagtagtcg cccccctcaa gtgtcaatac cgcagggcac ttatccccag 540
gcttgtccac atcatctgtg gaaaactcgc gtaaaatcag gcgttttcgc cgatttgca 600
ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgccgc 660
gccgggtgag tcggcccctc aagtgtcaac gtccgccct catctgtcag tgagggccaa 720
gttttcgcg aggtatccac aacgcggcg gcccggtgt ctcgcacacg gcttcgacgg 780
cgtttctggc gcgtttgcag ggccatagac ggccgccagc ccagcggcga gggcaaccag 840
cccgtgagc gtccgaaagg cgtcttccg cttcctcgt cactgactcg ctgcgctcgg 900
tcgttcggct gcggcgagc gtatcagctc actcaaaggc ggaataacgg ttatccacag 960
aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc 1020
gtaaaaaggc cgcgttgctg gcgttttcc ataggctccg cccccctgac gagcatcaca 1080
aaaaatcagc ctcaagtca aggtggcgaa acccgacagg actataaaga taccaggcgt 1140
ttccccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc 1200
tgtccgctt tctcccttc ggaagcgtgg cgccattcgc cattcaggct gcgcaactgt 1260
tggaagggc gatcggtgcg gccctcttc ctattacgcc agctggcgaa aggggatgt 1320
gctgcaaggc gattaagttg ggtaacgcca gggtttccc agtcacgacg ttgtaaacg 1380

```

-continued

acggccagtg aattcggccg cgggctggcc gaggtgctgg gcaagcccta cctccaggcc	1440
cccatcgggg tcgagagcac gaccgccttc ctgcgccgcc tgggcgagat tctgggcctc	1500
gatccggagc ctttcacga gcgcgagaag cactcgacgc tgaagcccgt gtgggatctg	1560
tggcggagtg tcacgcagga cttcttcggg acggccaatt tcggaatcgt ggcgaccgaa	1620
acttatgcaa gaggcacccg aaactatctc gaaggcgatc tcgggctgcc ctgcgccttc	1680
gcctggcccc caagaggggc tcgaagaccg acaacgaagc ggtgcgcgga ctgatccgcc	1740
agcaccgtcc gctcgtgctc atggggctca tcaacgagaa gatttacctt gcggaactga	1800
aagccggtca cggcccggca accctcttcc atcgtgcctt ctttcccggg tgcggcgatc	1860
cggcgcgcta ccggaaccgc cgttatggga tatgcagggt ctacgtgggt actgcaggaa	1920
gtttgcaacg ccctgttcga cgcctgttcc cacattctgc ccctcgggac ggagatggac	1980
agcgcgcgcc ccacaccgac gacactgcgc cgcgacttcc cgtgggatgc cgatgcgcaa	2040
cgggcctcgg accgcacgtg agaggagcat ccggttctca cccggatcag cgcgcgcgct	2100
gccttgccgc acgcccgca gaaggcttgc cctcgatgcc ggtgccgaga gggctgtgag	2160
agagactgtc gaagccctga cgtgggccgg gcttcggcga gaggaaggga gagaaccaat	2220
gagcgatcat gccgtcaaca cgcgggtcca tgcgccagg gccacgggc accgagcacc	2280
tcgtgccgag ttctacgtct acttcgccgt cattctgctg ggcgccttcc cgttgccctt	2340
cgtgagctgg atcgtctcga cgatccgcca ccgcaggctt ccaagcgcg gcccttcgc	2400
gtccgcctgg ttcgatgcca aggcgatcac gccgctgatt ttccgcgcct gaccgcagg	2460
caggttgca caccgccattc gtcgtctccc caagggcggg cggattaatc gggagggcat	2520
ggtgccttac cgtaaccac gccaccagta ggcaggagga acacgtgtcg tccggtggtc	2580
accaccacca ccaccaccac caccaccact aataggccgg ccctccgctg cgggcggcac	2640
ccacgcccg atcgattcca aggttcagcc attgagacgg ctccgcttcc cgcgcaagcg	2700
cgggttgggc cgactgcaag cggagagggga agcatggcac tgctcagctt cgagcgaaaa	2760
tatcgcgtgc cggggggcac gctggtcggc ggaaacctgt tcgacttctg ggtcggccct	2820
ttctatgtcg gcttctctcg ggttgogacg tttttcttcc cggccctggg tatcattctg	2880
attgcctgga gtgccgtact ccagggtacc tggaaacccc aactcatctc tgtctaccg	2940
ccggcccttg aatattggcct gggaggtgca cccctcga aaggcgggct gtggcagatc	3000
atcacgatct gcgccactgg tgccttcgct agctgggccc tgcgcgaagt cgaaatctgc	3060
cgtaagctgg gcatcgggta ccacatcccg ttcgccttcc cgttcgcat cctggcctac	3120
ctgacgtcgg tgcgttccg cccggtgatg atgggcgcct ggggctatgc cttcccctac	3180
gggatctgga cgcacctcga ctgggtgctg aacacgggct acacctacgg caactccac	3240
tacaaccctg cccacatgat cgccatctcg ttcttcttca cgaacgcgct ggctctggcg	3300
ctgcacggcg cccttggtct ctccgcggcc aaccccgaga agggcaagga aatgcggacg	3360
ccggatcacg aggatacgtt cttcccgcat ctggtcggct actcgatcgg gacgctcggc	3420
atccaccgcc tcggcctgct gctctcgtg agcgcctctt tctcagcgc cctctgcatg	3480
atcattaccg gcaccatctg gttcgatcag tgggtcgaact ggtggcaatg gtgggtgaa	3540
ctgccgtggt gggcgaacat cccgggaggc atcaatggct gagtatcaga acatcttctc	3600
ccaggctccag gtccgcggac cggccgacct ggggatgacc gaagacgtca acctggccaa	3660

-continued

ccggtcgggc gtcggtcctt tctcgacct gctcggctgg ttcggcaacg cccagctcgg	3720
cccgatctat ctcggtctgc tcggcgtcct gtccctcttc tcgggcctga tgtggttctt	3780
caccatcggg atctggttct ggtatcaggc gggctggaac ccggccgtct tcctgcgcga	3840
ctctgttctt ttctcgtctg agccgcgggc acccgaatac ggtctgtcct tcgctggctcc	3900
gctgaaggaa ggcgggctgt ggctgatcgc gtcgttcttc atgttcgtcg cggctctggtc	3960
ctggtggggc cgcacctatc tccgcgtca ggcctgggc atgggcaagc acaccgcctg	4020
ggcgttcctc tcggccatct ggtctgtgat ggtctgggc ttcattccgtc cgtacctcat	4080
ggggtcctgg tcggaagcgg ttcctactcg catcttctcg cacctcgact ggacgaacaa	4140
cttctcgtct gtccacggca acctgttcta caacccttc cacggtctct cgtatgcctt	4200
cctctacggg tcggccctgc tcttcgcat gcacggtgcg accatcctcg cggctctccg	4260
cttcggcggc gagcgcgagc tggagcagat gcgcgaccgc gggacggcag cggagcgggc	4320
cgccctcttc tggcgttga ccatgggttt caacgccacg atggaaggca tccaccgctg	4380
ggccatctgg atggcgttc tcgtgacct caccggcggc atcgggatcc tgctctcggg	4440
cacggtcgtg gacaactggt acgtctgggg ccagaaccac ggcatggcgc cgtgaactg	4500
aggagcgate acaatggctg acaagacct cttcaacgat cacctcaaca ccaatccgaa	4560
gaccaacctt cgctctggtg tcgctttcca gatgatgaag ggtcgggctt gggctggcgg	4620
cgtgttcttc gggacgtcc ttctcatcgg gttcttcgg gtggtcgggc ggtatgcttc	4680
gatccaggag aaccaggctc cggcgcgaa catcaccggc gctctggaga ccgggatcga	4740
gctgatcaag catctcgtct gagacaagtc tcggggcagg gcggcgcgag gccgcccgct	4800
cctccaagtc cgggccatat cgcgcgcccg ggtccggggc gacaccacag cccggttccc	4860
ttctgttg cgacaggac ctggtgccgt gtggaagacc gcacggcacc cttttgacat	4920
tcacgggag ctctgatgac caatcccacc ccgcgaccgg aaaccccgct tttggatcgc	4980
gtctgctgcc cggccgacat gaaggcgtg agtgacgcc aactggagcg gctggccgac	5040
gaagtgcgtt ccgaggtcag tgataggggt agtttcttat tttaggcagt ttatatgaa	5100
ttaagacatg cagatgtcac agtgatatt gaactggtct cgaagctca atatccccca	5160
aagcacaagc aaaaacttcg acatcatgca gaagcgttcc ccgaaccgcg tcttcgacgt	5220
gggcacgcc gagcagcatg ccgtgacct cgcggccggc ctcgccgggg ccgggatgaa	5280
gcccttctgc gcgatctatt cctcgttctt gcaacggggt tacgaccaga tcgccatga	5340
cgtggcgtc cagaaccttc ccgtccgctt cgtgatcgac cgggcggggc tcgtgggggc	5400
cgatggcgcg acctatgcgg gggccttcga cgttggttc atcacttcgc tgcccaacat	5460
gaccgtgatg gccgcggccg acgagggcga gctcatccac atgatcgcca ccgccgtggc	5520
cttcgacgag ggccccatcg ccttcogctt cccgcggggc gagggggtgg gcgtcgagat	5580
gcccagcgc gggacggtgc tggagcccgg ccggggccgc gtggtgcgcg aagggacgga	5640
tgctcgate ctctccttcg gcgcgatct gcacgagcc ttgcaggcgg cgaacttct	5700
cgaggccgag ggggtgagcg tgaccgtgac gcacgccgc ttctcgcgc cgtcgcacac	5760
ggggtcctc gaccagctcg tgcgcacat cgcggcgtg gtaacggtgg agcagggggc	5820
catggggcgc ttcggcgcct atgtcatgca ctatctcgc aattccggcg gcttcgacgg	5880
gggcctcgcg ctccgggtca tgacgtgcc gcaccgcttc atcgagcagg cgagccccga	5940

-continued

ggacatgtat gccgatgctg ggctgcgggc cgaggatata aagcttgccg taatcatggt	6000
catagctggt tcctgtgtga aattgttata cgctcacaat tccacacaac atacgagccg	6060
gaagcataaa gtgtaaaagg tgggtgctct aatgagtgag ctaactcaca ttaattgcgt	6120
tgcgctcaact gcccgccttc cagtcgggaa acctgtctgt ccagctgcat taatgaatcg	6180
gccaacgcgc ggggagaggc ggtttgctga ttgggcgctc ggtcttgccct tgctcgtcgg	6240
tgatgtactt caccagctcc gcgaagtcgc tcttcttgat ggagcgcag gggacgtgct	6300
tggcaatcac gcgcaccccc cgccgctttt agcggctaaa aaagtcatgg ctctgccctc	6360
gggaggacca cgcccatcat gaccttgcca agctcgtcct gcttctcttc gatcttcgcc	6420
agcagggcga ggatcgtggc atcaccgaac cgcgccgtgc gcgggtcgtc ggtgagccag	6480
agtttcagca ggccgcccag gcggcccagc tcgccattga tgcgggccag ctccgggacg	6540
tgctcatagt ccacgacgcc cgtgatattg tagccctggc cgacggccag caggtaggcc	6600
gacaggctca tgccggccgc cgcgcctttt tcctcaatcg ctcttcgttc gtctggaagg	6660
cagtacacct tgatagtggt gctgccttc ctggttggtc tggtttcac agccatccgc	6720
ttgccctcat ctgttacgcc ggcggtagcc ggccagcctc gcagagcagg attcccgttg	6780
agcaccgcca ggtgcgaata agggacagtg aagaaggaac acccgctcgc gggggggcct	6840
acttcaccta tcctgcccgc ctgacgccgt tggatacacc aaggaaagtc tacacgaacc	6900
ctttggcaaa atcctgtata tcgtgcgaaa aaggatgat ataccgaaaa aatcgtata	6960
atgacccga agcaggggta tcagcgggaa aagcggccag cttcccgaag ggagaaaggc	7020
ggacagggat ccggtaaagc gcagggtcgc aacaggagag cgcacgaggg agcttcacag	7080
gggaaacgcc tggtatcttt atagtcctgt cgggtttcgc cacctctgac ttgagcgtcg	7140
atttttgtga tgctcgtcag gggggcggag cctatggaaa aacgccagca acgcccctt	7200
tttacgggtc ctggcctttt gctggccttt tgctcacatg ttctttcctg cgttatcccc	7260
tgattctgtg gataaccgta ttaccgcctt tgagttagct gataccgctc gccgcagccg	7320
aacgaccgag cgcagcaggt cagttagcga ggaagcggaa gagcggcaga aggccgccag	7380
agaggccgag cgcggccgtg aggcttgac gctagggcag ggcatgaaaa agcccgtagc	7440
gggctgctac gggcgtctga cgcggtgaa agggggaggg gatgtgtct acatggetct	7500
gctgtagtga gtgggttgcc ctccggcagc ggtcctgac aatcgtcacc ctttctcgg	7560
ccttcaacgt tcctgacaac gagcctcctt ttcgccaatc catcgacaat cacgcgagt	7620
ccctgctcga acgctgcgtc cggaccggct tcgtcgaagg cgtctatcgc ggcccgaac	7680
agcggcgaga gcggagcctg ttcaacggtg ccgcccgcct cgcggcctc gctgtcggc	7740
gcctgctcct caagcagcgc cccaacagtg aagtagctga ttgtcatcag cgcattgacg	7800
gcgtccccgc ccgaaaaacc cgcctcgcag aggaagcga gctgcgcgtc ggccgtttcc	7860
atctgctgtg cgcgggtgct cgtgcgggca tggatgctgc cgcctcgcg gtaggcgagc	7920
agcgcctgcc tgaagctgct ggcattcccg atcagaaaat agcggcagtc gtcgtcggct	7980
ctcggcaccg aatgcgtatg attctccgcc agcatggctt cggccagtc gtcgagcagc	8040
gcccgttgt tcctgaagtg ccagtaaaag gccggctgct gaacccccaa ccgttccgcc	8100
agtttgctg tcgtcagacc gtctaocgg acctcgttca acaggtccag ggcggcagc	8160
atcactgtat tcggctgcaa ctttgtcatg cttgacactt taccactgat aaacataata	8220

-continued

tgtccaccaa	cttatcagtg	ataaagaatc	cgcgcggttca	atcggaccag	cgagggtctgg	8280
tccggaggcc	agacatgaaa	cccaacatac	ccctgatcgt	aattctgagc	actgtcgcgc	8340
tcgacgtgt	cggcacgcgc	ctgattatgc	cggtgctgcc	ggcctcctg	cgcgatctgg	8400
ttcactcgaa	cgacgtcacc	gcccactatg	gcattctgct	ggcgtgtat	gcgttggtgc	8460
aatttgctg	cgacacctg	ctgggcgcgc	tgctggatcg	tttcgggcgg	cgccaatct	8520
tgctcgtctc	gctggccgc	gccactgtcg	actacgcat	catggcgaca	gcgcccttcc	8580
tttgggttct	ctatatcggg	cggatcgtgg	ccggcatcac	cggggcgact	ggggcggtag	8640
ccggcgctta	tattgccgat	atcactgatg	gcgatgagcg	cgcgcgccac	ttcggcttca	8700
tgagcgctg	tttcgggttc	gggatggtcg	cgggacctgt	gctcgggtgg	ctgatggcg	8760
gtttctcccc	ccacgctccg	ttcttcgccc	cggcagcctt	gaacggcctc	aatttcctga	8820
cgggctgttt	ccttttgccc	gagtcgcaca	aaggcgaacg	ccggccgcta	cgccgggag	8880
ctctcaacc	gctcgtctcg	ttccgggtgg	cccggggcat	gaccgtcgtc	gccgcctga	8940
tggcgtctt	cttcatcatg	caactgtcg	gacaggtgcc	ggccgcgctt	tgggtcattt	9000
tcggcgagga	tcgctttcac	tgggacgcga	ccacgatcgg	catttcgctt	gccgcatttg	9060
gcattctgca	ttcactcgcc	caggcaatga	tcaccggccc	tgtagccgcc	cggtcggcg	9120
aaaggcggg	actcatgctc	ggaatgattg	ccgacggcac	aggctacatc	ctgcttgctt	9180
tcgcgacacg	gggatggatg	gcgttcccga	tcatggctct	gcttgcttcg	ggtggcatcg	9240
gaatgcggc	gctgcaagca	atggtgtcca	ggcaggtgga	tgaggaacgt	caggggcagc	9300
tgcaaggctc	actggcggcg	ctcaccagcc	tgacctcgat	cgctcggaacc	ctcctcttca	9360
cggcgatcta	tcgcgcttct	ataacaactg	ggaacgggtg	ggcatggatt	gcagggcctg	9420
ccctctactt	gctctgctg	ccggcgctgc	gtcgcgggct	ttggagcggc	gcagggcaac	9480
gagccgatcg	ctgatcgtgg	aaacgatagg	cctatgccat	cggggtcaag	gcgacttccg	9540
gcaagctata	cgcgccctag	gagtgcggtt	ggaacgttgg	cccagccaga	tactcccgat	9600
cacgagcag	acgccgatga	tttgaagcgc	actcagcgtc	tgatccaaga	acaaccatcc	9660
tagcaaacac	gcggtccccg	ggctgagaaa	gcccagtaag	gaaacaactg	taggttcgag	9720
tcgcgagatc	ccccggaacc	aaaggaagta	ggttaaacc	gctccgatca	ggccgagcca	9780
cgccagggcg	agaacattgg	ttcctgtagg	catcgggatt	ggcggatcaa	acactaaagc	9840
tactggaac	agcagaagtc	ctccggccgc	cagttgccag	cggttaaagg	tgagcagagg	9900
cacgggaggt	tgccacttgc	gggtcagcac	ggttccgaac	gccatggaaa	ccgccccgc	9960
cagggccgct	gcgacgccga	caggatctag	cgctgcgttt	ggtgtcaaca	ccaacagcgc	10020
cacgcccgca	gttccgcaaa	tagccccag	gaccgccatc	aatcgtatcg	ggctacctag	10080
cagagcggca	gagatgaaca	cgaccatcag	cggtgcaca	gcgctaccg	tcgccgcgac	10140
cccgccggc	aggcggtaga	ccgaaataaa	caacaagctc	cagaatagcg	aaatattaag	10200
tgcccgagg	atgaagatgc	gcacccacca	gattcccgtt	ggaatctgtc	ggaogatcat	10260
cacgagcaat	aaaccgcgcc	gcaacgcccg	cagcagcata	ccggcgaccc	ctcggcctcg	10320
ctgttcgggc	tccacgaaaa	cgccggacag	atgcgccttg	tgagcgtcct	tggggccgctc	10380
ctcctgtttg	aagaccgaca	gcccacatgat	ctcgcctcg	atgtaggcgc	cgaatgccac	10440
ggcatctcgc	aaccgttcag	cgaacgcctc	catgggcttt	ttctcctcgt	gctcgtaaac	10500

-continued

ggacccgaac atctctggag ctttcttcag ggccgacaat cggatctcgc ggaatcctg	10560
cacgtcggcc gctccaagcc gtcgaatctg agccttaatc acaattgtca attttaatcc	10620
tctgtttatc ggacgttcgt agagcgcgcc gtgcgtcccg agcgatactg agcgaagcaa	10680
gtgcgtcag cagtgcccgc ttgttctga aatgccagta aagcgtggc tgctgaacct	10740
ccagccggaa ctgacccacc aaggccctag cgtttgcaat gcaccaggtc atcattgacc	10800
caggcgtggt ccaccaggcc gctgcctcgc aactcttcgc agccttcgcc gacctgctg	10860
cgccacttct tcacgcgggt ggaatccgat ccgcacatga ggcggaagg ttccagcttg	10920
agcgggtacg gctcccgggt cgagctgaaa tagtcgaaca tccgtcgggc cgtcggcgac	10980
agcttgcggt acttctccca tatgaatttc gtgtagtggt ccgacgaaa cagcacgacg	11040
atttcctcgt cgatcaggac ctggcaacgg gacgttttct tgccacggtc caggacgagg	11100
aagcgggtca gcagcgacac cgattccagg tgcccaacgc ggtcggacgt gaagcccatc	11160
gccgtcgcct gtaggcgca caggcattcc tcggccttcg tgtaataacc gccattgatc	11220
gaccagccca ggtcctggca aagctcgtag aacgtgaagg tgatcggctc gccgataggg	11280
gtgcgcttcg cgtactccaa cacctgctgc cacaccagtt cgtcatcgtc ggcccgcagc	11340
tcgacgccgg ttaggtgat ctccacgtcc ttggtgacgt ggaaaatgac cttgttttgc	11400
agcgcctcgc gcgggatttt ctgtttgcgc gtggtgaaca ggcagagcg gcccgctgcg	11460
tttggcatcg ctgcacatgt gtcggccac ggcgcaatat cgaacaagga aagctgcatt	11520
tccttgatct gctgcttcgt gtgtttcagc aacgcggcct gcttggcctc gctgacctgt	11580
tttggcagg tctcgcgggc ggtttttcgc ttcttggctg tcatagttcc tcgctgtcgc	11640
atggtcatcg acttcgcaa acctgccgc tcctgttcga gacgacgca acgctccacg	11700
gcggccgatg gcgcgggag gccaggggga gccagttgca cgctgtcgcg ctcgatcttg	11760
gccgtagctt gctggaccat cgagccgacg gactggaagg tttcgcgggg cgcacgcatg	11820
acggtgcggc ttgcgatggt ttcggcatcc tcggcggaaa accccgcgtc gatcagttct	11880
tgctgtatg ctttccggtc aaacgtccga ttcattcacc ctcttgcgg gattgccccg	11940
actcacgccg gggaatgtg cccttattcc tgatttgacc cgcctggtgc cttggtgtcc	12000
agataatcca cttatcgcg aatgaagtcg gtcccgtaga ccgtctggcc gtccttctcg	12060
tacttggtat tccgaatctt gccctgcacg aataccagcg accccttgc caaatacttg	12120
ccgtgggctt cggcctgaga gccaaaacac ttgatgcgga agaagtcggt gcgctcctgc	12180
ttgtcgcgg tcgtggccgc gccaaccttt gcgatccgca agcgcgcggt cggcatcttc	12240
acgctggaac agtacgtcga gccgggcatc atgacccgcg agcaatacga ggtcattaaa	12300
agcgcctgga ttgatgatat agcggcccgg ctgctcctgg ttctcgcgca ccgaaatggg	12360
tgacttcacc ccgcctctt tgatcgtggc accgatttcc gcgatgctct ccggggaaaa	12420
gccggggttg tcggccttcc gcggtgatg cggatcttcg tcgatcaggc ccaggctccag	12480
ctcgataggg ccggaaccgc cctgagacgc cgcaggagcg tccaggaggc tcgacaggtc	12540
gccgatgcta tccaacccca gcccggacgg ctgcgccgcg cctgcggctt cctgagcggc	12600
cgagcgggtg tttttcttgg tggctttggc ttgagccgca gtcattggga aatctccatc	12660
ttcgtgaaca cgtaatcagc cagggcgcga acctcttctg atgccttgcg cgcggccggt	12720
ttcttgatct tccagaccgg cacaccggat gcgagggcat cggcgatgct gctgcgcagg	12780

-continued

```

ccaacgggtgg ccggaatcat catcttgggg tacgcggcca gcagctcggc ttggtggcgc 12840
gcgtggcgcg gattccgcgc atcgacctg ctgggcacca tgccaaggaa ttgcagcttg 12900
gcgttcttct ggcgcacggt cgcaatggtc gtgacctct tcttgatgcc ctggatgctg 12960
tacgcctcaa gctcgatggg ggacagcaca tagtcggccg cgaagagggc ggccgccagg 13020
ccgacgcca gggtcggggc cgtgtcgatc aggcacacgt cgaagccttg gttcgccagg 13080
gccttgatgt tcgccccgaa cagctcgcgg gcgtcgtcca gcgacagccg ttcggcggtc 13140
gccagtaccg gtttgactc gatgagggcg aggcgcgcgg cctggccgtc gccggctgcg 13200
ggtgcggttt cggtcacgcc gccgcaggg acagcgcga acagcttgct tgcatgcagg 13260
ccggtagcaa agtccttgag cgtgtaggac gcattgcctt ggggtccag gtcgatcacg 13320
gcaaccgcga agccgcgctc gaaaaagtcg aaggcaagat gcacaagggt cgaagtcttg 13380
ccgacgccc ctttctgggt ggccgtgacc aaagtttca tcgtttggtt tcctgttttt 13440
tcttgcgctc cgctccacc ttccggacga tgtacgctg atgttccggc agaaccgccc 13500
ttaccgccc gtaccctcgg gcaagttct tgcctcga cgcggcccac acgcatgca 13560
ccgcttgcca cactgcgccc ctggtcagtc ccagcagcgt tgcgaacgtc gcctgtggct 13620
tcccctcgac taagacgccc cgcgctatct cgatggctg ctgcccact tccagcccct 13680
ggatcgctc ctggaactgg ctttcggtaa gccgtttctt catggataac acccataatt 13740
tgctccgccc cttggttgaa catagcggtg acagccgcca gcacatgaga gaagttagc 13800
taaacatttc tcgcacgtca acaccttag ccgctaaac tcgtccttgg cgtaacaaaa 13860
caaaagccc gaaaccgggc tttcgtctct tgccgcttat ggctctgcac ccggctccat 13920
caccaacagg tcgcgcacgc gcttactcgg gttgcggatc gacctgcca gccaacaaa 13980
gccggttgc gccgcgcca ggatcgccc gatgatgccc gccacaccgg ccatcgcca 14040
ccaggtcgcc gccttccggt tccattcctg ctggtactgc ttcgcaatgc tggacctgg 14100
ctcaccatag gctgaccgct cgatggcgta tgccgcttct ccccttggcg taaaaccag 14160
cgccgaggg gccattgcca tgctgcccgc cgctttcccg accacgacgc ggcaccagg 14220
cttgcggtcc agacctcgg ccacggcgag ctgcaagc acataatcag ccgcccactt 14280
ggctccacgc gcctcgatca gctcttgac tcgcgcgaaa tccttggcct ccacggccgc 14340
catgaatcgc gcacgcggcg aaggctccc agggccg 14377

```

<210> SEQ ID NO 18

<211> LENGTH: 14386

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic plasmid

<400> SEQUENCE: 18

```

ccaccaggc cgccgcctc actgcccgc acctggtcgc tgaatgtcga tgccagcacc 60
tgcgacacgt caatgcttcc gggcgctcgc ctcgggctga tcgccatcc cgtaactgcc 120
ccgatccccg caatggcaag gactgccagc gotgccattt ttgggggtgag gccgttcgcg 180
gccgaggggg gcagcccctg gggggatggg aggccgcgt tagcgggccc ggagggttcg 240
agaagggggg gcacccccct tcggcgtgcg cggtcacgcg cacagggcgc agccctgggt 300
aaaaacaagg ttataaata ttggtttaa agcaggttaa aagacaggtt agcggtgccc 360

```

-continued

gaaaaacggg	cgaaaacct	tgcaaatgct	ggattttctg	cctgtggaca	gcccctcaaa	420
tgtaaatagg	tgcgcccctc	atctgtcagc	actctgcccc	tcaagtgtca	aggatcgcgc	480
ccctcatctg	tcagtagtcg	cgcccctcaa	gtgtcaatac	cgcagggcac	ttatccccag	540
gcttgccac	atcatctgtg	ggaaactcgc	gtaaaatcag	gcgttttcgc	cgatttgcca	600
ggctggccag	ctccacgtcg	ccggccgaaa	tcgagcctgc	ccctcatctg	tcaacgccgc	660
gccgggtgag	tcggcccctc	aagtgtcaac	gtccgcccct	catctgtcag	tgagggccaa	720
gttttcgcgc	aggatccac	aacgcggcgc	gcccggtgt	ctcgcacacg	gcttcgacgc	780
cgtttctggc	gcgtttgca	ggccatagac	ggccgccagc	ccagcggcga	gggcaaccag	840
cccggtgagc	gtcggaaaag	cgctcttcgc	cttcctcgc	cactgactcg	ctgcgctcgc	900
tcgttcggct	gcggcgagcg	gtatcagctc	actcaaaggc	ggtaatacgc	ttatccacag	960
aatcagggga	taacgcagga	aagaacatgt	gagcaaaaag	ccagcaaaag	gccaggaacc	1020
gtaaaaaggc	cgcgttgctg	gcgtttttcc	ataggctccg	ccccctgac	gagcatcaca	1080
aaaaatcgac	ctcaagtctg	aggtagcgaa	acccgacagg	actataaaga	taccaggcgt	1140
ttccccctgg	aagctccctc	gtgcgctctc	ctgttccgac	cctgccgctt	accggatacc	1200
tgccgcctt	tctcccttcg	ggaagcgtgg	cgccattcgc	cattcaggct	gcgcaactgt	1260
tggaagggc	gatcggtagc	ggcctcttcg	ctattacgcc	agctggcgaa	aggggatgt	1320
gctgcaaggc	gattaagttg	ggtaacgcca	gggttttccc	agtcacgacg	ttgtaaaacg	1380
acggccagtg	aattcggccc	cgggctggcc	gagtgctggt	gcaagcccta	cctccaggcc	1440
cccatcgggg	tcgagagcac	gaccgccttc	ctgcgcgcc	tggcgagat	tctgggcctc	1500
gatccggagc	ccttcatcga	gcgcgagaag	cactcgacgc	tgaagcccgt	gtgggatctg	1560
tgcgagagtg	tcacgcagga	cttcttcggg	acggccaatt	tcggaatcgt	ggcgaccgaa	1620
acttatgcaa	gaggcatccg	aaactatctc	gaaggcgatc	tcgggctgcc	ctgcgccttc	1680
gcctggcccc	caagaggggc	tcgaagaccg	acaacgaagc	ggtgcgcgga	ctgatccgcc	1740
agcaccgtcc	gctcgtgctc	atggggtcga	tcaacgagaa	gatttacctt	gcggaactga	1800
aagccggtca	cgcccggcga	accctcttcc	atcgtgcctt	cttcccggg	tgcggcgatc	1860
cgggcgctca	ccggaaccgc	cgttatggga	tatgcaggtg	ctacgtggtt	actgcaggaa	1920
gtttgcaacg	ccctgttcga	cgccctgttc	cacattctgc	ccctcgggac	ggagatggac	1980
agcgcggcgc	ccacaccgac	gacctgcgc	cgcgacttcc	cgtgggatgc	cgatgcgcaa	2040
cgggccctgg	accgcatcgt	agaggagcat	cggttctc	cccggatcag	cgccgcgcgt	2100
gccttgccgc	acgccgccga	gaaggcttgc	cctcgatgcc	ggtgccgaga	gggtcgtgag	2160
agagactgtc	gaagccctga	cgtgggccgg	gcttcggcga	gaggaaagga	gagaaccaat	2220
gagcgatcat	gccgtcaaca	cgccggtcca	tgccgccagg	gcccacgggc	accgagcacc	2280
tcgtgcccag	ttctactctt	acttcgccgt	cattctgctg	ggcgccttcc	cggtggcctt	2340
cgtgagctgg	atcgtctcga	cgatccgcca	ccgcaggctt	cccgaagcgc	gccccttcgc	2400
gtccgcctgg	ttcgatgcca	aggcgatcac	gocgctgatt	ttccgcgcct	gacgcaggt	2460
caggttgcca	cacgccatc	gtcgtctccc	caagggcgg	cggattaatc	gggagggcat	2520
ggtgccttac	cgtaaccac	gccaccagta	ggcaggagga	acacgtgtcg	tccgggtggtc	2580
accaccacca	ccaccaccac	caccaccacc	accaccacta	ataggccggc	cctccgtcgc	2640

-continued

gggcgccacc	cacgcccga	tcgattccaa	ggttcagcca	ttgagacggc	tccgcttcgc	2700
gcgcaagcgc	gggttgggcc	gactgcaagc	ggagaggaa	gcatggcact	gctcagcttc	2760
gagcgaaaat	atcgcgctgc	ggggggcacg	ctggctcgcg	gaaacctggt	cgacttctgg	2820
gtcggccctt	tctatgtcgg	cttcttcggg	gttgcgacgt	ttttcttcgc	ggccctgggt	2880
atcattctga	ttgcctggag	tgccgtactc	cagggtaact	ggaaccccca	actcatctct	2940
gtctaccgcg	cggcccttga	atatggcctg	ggaggtgcac	ccctcgaaa	aggcgggctg	3000
tggcagatca	tcacgatctg	cgccactggt	gccttcgtea	gctgggcgct	gcgcaagtc	3060
gaaatctgcc	gtaagctggg	catcgggtac	cacatcccgt	tcgcttcgcg	gttcgccatc	3120
ctggcctacc	tgacgtggtg	gctgttccgc	ccggtgatga	tggcgccctg	gggctatgcc	3180
ttcccctacg	ggatctggac	gcacctcgac	tgggtgtcga	acacgggcta	cacctacggc	3240
aacttccact	acaacctcgc	ccacatgatc	gccatctcgt	tcttcttcac	gaacgcgctg	3300
gctctggcgc	tgacggcgcg	ccttgtgctc	tccgcggcca	accccgagaa	gggcaaggaa	3360
atcgggacgc	cggatcacga	ggatacgttc	ttccgcgatc	tggtcggcta	ctcgatcggg	3420
acgctcggca	tccaccgcct	cggcctgctg	ctctcgtga	gcgccgtctt	cttcagcgcc	3480
ctctgcatga	tcattaccgg	caccatctgg	ttcgatcagt	gggtcgaactg	gtggcaatgg	3540
tgggtgaagc	tgccgtggtg	ggcgaacatc	ccgggaggca	tcaatggctg	agtatcagaa	3600
catcttctcc	caggtccagg	tcccgggacc	ggccgacctg	gggatgaccg	aagacgtcaa	3660
cctggccaac	cgttcggggc	tcggtccttt	ctcgaccctg	ctcggctggt	tcggcaacgc	3720
ccagctcggc	ccgatctatc	tcggctcgtc	cggcgtctg	tccctcttct	cgggcctgat	3780
gtggttcttc	accatcggga	tctggttctg	gtatcaggcg	ggctggaacc	cggccgtctt	3840
cctgcgcgac	ctgttcttct	tctcgtctga	gccgcggca	cccgaatacg	gtctgtctct	3900
cgcggctccg	ctgaaggaa	gcgggctgtg	gctgatcgcg	tcgttcttca	tgttcgtcgc	3960
ggtctggtcc	tgggtgggcc	gcacctatct	ccgcgctcag	gcgctgggca	tgggcaagca	4020
caccgcctgg	gcgttctctc	cggccatctg	gctgtggatg	gtgctgggct	tcacccgtcc	4080
gatcctcatg	gggtcctggt	cggaaagcgt	tccctacggc	atcttctcgc	acctcgactg	4140
gacgaacaac	ttctcgtcgc	tccacggcaa	cctgttctac	aacctcttcc	acggtctctc	4200
gatcgccttc	ctctacgggt	cggccctgct	cttcgcgatg	cacggtgcga	ccatcctcgc	4260
ggtctcccgc	ttcggcggcg	agcgcgagct	ggagcagatc	gccgaccgcg	ggacggcagc	4320
ggagcgggcc	gcctcttctc	ggcgtggac	catgggtttc	aacgccacga	tggaaagcat	4380
ccaccgctgg	gccatctgga	tggcggctct	cgtgaccctc	accggcggca	tcgggatcct	4440
gctctcgggc	acggtcgtgg	acaactggta	cgtctggggc	cagaaccacg	gcatggcgcc	4500
gctgaactga	ggagcagatc	caatggctga	caagaccatc	ttcaacgatc	acctcaacac	4560
caatccgaag	accaaccttc	gcctctgggt	cgctttccag	atgatgaagg	gtcggggctg	4620
ggctggcggc	gtgttcttgc	ggacgctcct	tctcatcggg	ttcttccggg	tggctggggc	4680
gatgttctcc	atccaggaga	accaggtccc	ggcggcgaac	atcaccggcg	ctctggagac	4740
cgggatcag	ctgatcaagc	atctcgtctg	agacaagtct	cggggcaggg	cggcgcgagg	4800
ccgcccgcctc	ctccaagtcc	gggccaatc	gccggcccgg	gtccggggcg	acaccacagc	4860
ccggttccct	tcctgttggc	gacagggacc	tgggtcccgtg	tggaaagacc	cacggcacc	4920

-continued

ttttgacatt	cacgggagcg	tctgatgacc	aatccccacc	cgcgaccgca	aaccccgctt	4980
ttggatcgcg	tctgctgccc	ggccgacatg	aaggcgctga	gtgacgccga	actggagcgg	5040
ctggccgacg	aagtgcgttc	cgaggtcagt	gataggggta	gtttcttatt	ttaggcagtt	5100
tatatgaaat	taagacatgc	agatgtcaca	gtggatattg	aactggcttc	gaaagctcaa	5160
tatcccccaa	agcacaagca	caaacttcga	catcatgcag	aagcgtttcc	cgaaccgctt	5220
cttcgacgtg	ggcatcgccg	agcagcatgc	cgtgaccttc	gcggccggcc	tcgccggggc	5280
cgggatgaag	cccttctgcg	cgatctattc	ctcgttctcg	caacgggggt	acgaccagat	5340
cgcccatgac	gtggcgctgc	agaaccttcc	cgtccgcttc	gtgatcgacc	gggcggggct	5400
cgtgggggcc	gatggcgcg	cccatgcggg	ggccttcgac	gttggcttca	tcacttcgct	5460
gccccaatg	accgtgatgg	ccgcggccga	cgaggccgag	ctcatccaca	tgatcgccac	5520
cgccgtggcc	ttcgacgagg	gccccatcgc	cttccgcttc	ccgcggggcg	aggggggtgg	5580
cgtcgagatg	cccagcgcg	ggacggtgct	ggagcccggc	cggggccgcg	tggtgcgca	5640
agggacggat	gtcgcgatcc	tctccttcgg	cgcgcatctg	cacgaggcct	tgacggcgcc	5700
gaaacttctc	gagggcgagg	gggtgagcgt	gaccgtggcc	gacgcccgct	tctcgcgcc	5760
gctcgacacg	ggctcactcg	accagctcgt	gcgccatcac	gcggcgctgg	taacggtgga	5820
gcagggggcc	atggggcgct	tcggcgccca	tgtcatgcac	tatctcgcca	attccggcgg	5880
cttcgacggg	ggcctcgcgc	tccgggtcat	gacgctgccc	gaccgcttca	tcgagcaggc	5940
gagccccgag	gacatgtatg	ccgatgcggg	gctgcgggcc	gaggatatca	agcttgcgct	6000
aatcatggtc	atagctgttt	cctgtgtgaa	attgttatcc	gctcacaatt	ccacacaaca	6060
tacgagccgg	aagcataaag	tgtaaagcct	ggggtgccta	atgagtgagc	taactcacat	6120
taattgcggt	gcgctcactg	cccgccttcc	agtcgggaaa	cctgtcgtgc	cagctgcatt	6180
aatgaatcgg	ccaacgcgcg	gggagaggcg	gtttgcgtat	tgggcgctcg	gtcttgccct	6240
gctcgtcggg	gatgtacttc	accagctccg	cgaagtcgct	cttcttgatg	gagcgcagtg	6300
ggacgtgctt	ggcaatcacg	cgcaccccc	ggcgttttta	gcggctaaaa	aagtcatggc	6360
tctgccctcg	ggcggaccac	gcccacatcg	accttgccaa	gctcgtcctg	cttctcttcg	6420
atcttcgcca	gcagggcgag	gatcgtggca	tcaccgaacc	gcgccgtgcg	cgggtcgtcg	6480
gtgagccaga	gtttcagcag	gcccggcagg	cggcccagggt	cgccattgat	gcgggcccagc	6540
tcgggacgct	gctcatagtc	cacgacgccc	gtgattttgt	agccctggcc	gacggccagc	6600
aggtaggccc	acaggctcat	gcccggccgc	gcccctttt	cctcaatcgc	tcttcgctcg	6660
tctggaagcg	agtacacctt	gataggtggg	ctgcccttcc	tggttggett	ggtttcatca	6720
gccatccgct	tgccctcctc	tgttacgccc	gcggtagccg	gccagcctcg	cagagcagga	6780
ttcccgttga	gcaccgccag	gtgcgaataa	gggacagtga	agaaggaaca	cccgcctcgc	6840
ggtgggccta	cttccactat	cctgcccggc	tgacgcccgt	ggatacacca	aggaaagtct	6900
acacgaaccc	tttgcaaaa	tcctgtatat	cgtgcgaaaa	aggatggata	taccgaaaaa	6960
atcgtataaa	tgaccccgaa	gcagggttat	gcagcggaaa	agcggcagcg	ttccgaaagg	7020
gagaaaaggcg	gacaggtatc	cggtaagcgg	cagggtcggg	acaggagagc	gcacgagggg	7080
gcttccaggg	ggaaacgcct	ggtatcttta	tagtctctgc	gggtttcgcc	acctctgact	7140
tgagcgtcga	tttttggatg	gctcgtcagg	ggggcggagc	ctatggaaaa	acgccagcaa	7200

-continued

cgcgcccttt ttacggttcc tggccttttg ctggcctttt gctcacatgt tctttcctgc 7260
gttatccctt gattctgtgg ataaccgtat taccgccttt gagtgagctg ataccgctcg 7320
ccgcagccga acgaccgagc gcagcgagtc agtgagcgag gaagcggaag agcggcagaa 7380
ggccgcagca gaggccgagc gcggccgtga ggcttggacg ctagggcagg gcatgaaaaa 7440
gcccgtagcg ggctgctacg ggcgtctgac gcgggtgaaa gggggagggg atgttgtcta 7500
catggctctg ctgtagtgag tgggtgctgc tccggcagcg gtcctgatca atcgtcacc 7560
tttctcggtc cttcaacgtt cctgacaacg agcctccttt tcgccaatcc atcgacaatc 7620
accgcgagtc cctgctgcaa cgtctcgtcc ggaccggctt cgtcgaaggc gtcctatcgcg 7680
gcccgcaca gggcgagag cggagcctgt tcaacggtgc cgcgcgctc gccggcatcg 7740
ctgtcgcggc cctgctcctc aagcacggcc ccaacagtga agtagctgat tgtcatcagc 7800
gcattgacgg cgtccccggc cgaaaaacc cctcgcaga ggaagcggaag ctgcgcgtcg 7860
gccgtttcca tctgcggtgc gcccggtcgc gtgccggcat gtagtgcgcg gccatcgcgg 7920
taggcgagca gcgcctgcct gaagctgcgg gcattcccga tcagaaatga gcgccagtcg 7980
tcgtcggctc tcggcaccga atgcgtatga ttctccgcca gcatggcttc gccagtgcg 8040
tcgagcagcg cccgcttgtt cctgaagtgc cagtaaagcg ccggctgctg aacccccaac 8100
cgttccgcca gtttgcgtgt cgtcagaccg tctacgcgca cctcgttcaa caggtccagg 8160
gcccgcagga tcaactgtatt cggctgcaac tttgtcatgc ttgacacttt atcaactgata 8220
aacataatat gtccaccaac ttatcagtga taaagaatcc gcgcgttcaa tcggaccagc 8280
ggaggctggt ccggaggcca gacatgaaac ccaacatacc cctgatcgta attctgagca 8340
ctgtcgcgct cgacgctgtc ggcacgtgcc tgattatgcc ggtgctgccg gccctcctgc 8400
gcgatctggt tcaactgaac gacgtcaccg cccactatgg cattctgctg gcgctgtatg 8460
cgttggtgca atttgcctgc gcacctgtgc tgggcgcgct gtcggatcgt ttcgggcggc 8520
ggccaatctt gctcgtctcg ctggccggcg ccaactgtcga ctacgccatc atggcgacag 8580
cgctttcct ttgggttctc tatatcgggc ggatcgtggc cggcatcacc ggggcgactg 8640
gggggttagc cggcgttat attgcgata tcaactgatgg cgatgagcgc gcgcggcaact 8700
tcggcttcat gagcgcctgt ttcgggttcg ggatggtcgc gggacctgtg ctcggtgggc 8760
tgatggcgcg tttctcccc cacgctccgt tcttcgcccg ggcagccttg aacggcctca 8820
atctcctgac gggctgtttc cttttgccgg agtcgcacaa agcggaacgc cggccgttac 8880
gcccggagcg tctcaaccgg ctcgcttcgt tccggtgggc ccggggcatg accgtcgtcg 8940
ccgccctgat ggcggtcttc ttcacatcgc aacttgctcg acaggtgccg gccgcgcttt 9000
gggtcatttt cggcgaggat cgctttcact gggacgcgac cacgatcggc atttcgcttg 9060
ccgcatttgg cattctgcat tcaactgccc aggcaatgat caccggcctt gtagccgcc 9120
ggctcggcga aaggcgggca ctcatgctcg gaatgattgc cgacggcaca ggtacatcc 9180
tgcttgcctt cgcgacacgg ggatggatgg cgttcccgat catggtcctg cttgcttcgg 9240
gtggcatcgg aatgccggcg ctgcaagcaa tgttgtccag gcaggtggat gaggaacgtc 9300
agggcagct gcaaggctca ctggcggcgc tcaccagcct gacctcgtc gtcggacccc 9360
tcctcttca cggcatctat gcggcttcta taacaacgtg gaacgggttg gcatggattg 9420
caggcgtcgc cctctacttg ctctgctgc cggcgtcgc tcgcgggctt tggagcggcg 9480

-continued

cagggcaacg agccgatcgc tgatcgtgga aacgataggc ctatgccatg cgggtcaagg 9540
cgacttcccg caagctatac gcgccctag agtgccgttg gaacgttggc ccagccagat 9600
actcccgatc acgagcagga cgccgatgat ttgaagcga ctcagcgtct gatccaagaa 9660
caaccatcct agcaacacgg cgggtcccg gctgagaaag cccagtaagg aaacaactgt 9720
aggttcgagt cgcgagatcc cccggaacca aaggaagtag gttaaacccg ctccgatcag 9780
gccgagccac gccaggccga gaacattggt tcctgtaggc atcgggattg gcggatcaaa 9840
cactaaagct actggaacga gcagaagtcc tccggccgcc agttgccagg cggtaaaggt 9900
gagcagaggc acgggaggtt gccacttgcg ggtcagcacg gttccgaacg ccatggaaac 9960
cgcccccgcc agggccgctg cgacgccgac aggatctagc gctgcgtttg gtgtcaacac 10020
caacagcgcc acggcccgag ttccgcaaat agccccagg accgccatca atcgtatcgg 10080
gctacctagc agagcggcag agatgaacac gaccatcagc ggctgcacag cgcctaccgt 10140
cgccgcgacc ccggccggca ggcggtagac cgaaataaac aacaagctcc agaatagcga 10200
aatattaagt gcggcgagga tgaagatgcg catccaccag attcccgttg gaatctgtcg 10260
gacgatcacc acgagcaata aacccgccgg caacggccgc agcagcatac cggcgacccc 10320
tcggcctcgc tgttcgggct ccacgaaaac gccggacaga tgcgccttgt gagcgtcctt 10380
ggggccgtcc tcctgtttga agaccgacag cccaatgatc tcgccgtcga ttagggcgcc 10440
gaatgccacg gcatctcgca accgttcagc gaacgcctcc atgggctttt tctcctcgtg 10500
ctcgtaaacg gaccggaaca tctctggagc tttcttcagg gccgacaatc gcatctcgcg 10560
gaaatcctgc acgtcggccg ctccaagcgg tcgaatctga gccttaatca caattgtcaa 10620
ttttaatcct ctgtttatcg gcagttcgta gagcgcgccc tgcgtcccga gcgatactga 10680
gcgaagcaag tgcgtcgagc agtgcccgtc tgttctgaa atgccagtaa agcgtgggct 10740
gctgaacccc cagccggaac tgacccaca aggccctagc gtttgcaatg caccaggtea 10800
tcattgacct agggcgtgtc caccaggccg ctgcctcgca actcttcgca ggcttcgccc 10860
acctgctcgc gccacttctt cacgcgggtg gaatccgatc cgcacatgag gcggaagggt 10920
tcagcttga gcgggtacgg ctcccgggtc gagctgaaat agtcgaacat ccgtcgggcc 10980
gtcggcgaca gtttgcggtt cttctccat atgaatttcg ttagtggtc gccagcaaac 11040
agcacgacga tttcctcgtc gatcaggacc tggcaacggg acgttttctt gccacggtcc 11100
aggacggga agcgggtcag cagcgacacc gattccagggt gcccaacgcg gtcggacgtg 11160
aagccatcg ccgtcgcctg taggcgagc aggcattcct cggccttcgt gtaataccgg 11220
ccattgatcg accagcccag gtcctggcaa agctcgtaga acgtgaaggat gatcggctcg 11280
ccgatagggg tgcgcttcgc gtactccaac acctgctgcc acaccagttc gtcacgtcg 11340
gcccgagctc cgacggcgtg gtaggtgatc ttcacgtcct tgttgacgtg gaaaatgacc 11400
ttgttttgca gcgcctcgc cgggatttct ttgttgcgcg tgggaacag ggcagagcgg 11460
gccgtgtcgt ttggcatcgc tcgcatcgtg tccggccacg gcgcaatata gaacaaggaa 11520
agctgcattt ccttgatctg ctgcttcgtg tgtttcagca acgcgccctg cttggcctcg 11580
ctgacctggt ttgccaggtc ctgcggcg gtttttcgct tcttggtcgt catagttcct 11640
cgctgtcga tggatcagca cttcgccaaa cctgccgect cctgttcgag acgacgcgaa 11700
cgctccacgg cggccgatgg cgcgggacgg gcagggggag ccagttgcac gctgtcggc 11760

-continued

tcgatcttg ccgtagcttg ctggaccatc gagccgacgg actggaaggt ttcgcggggc 11820
gcacgcatga cgggtgcgct tgcgatggtt tcggcatcct cggcggaaaa ccccgcgctc 11880
atcagttctt gcctgtatgc cttccggta aacgtccgat tcattcacc ccttgcggg 11940
attgccccga ctacagccgg ggcaatgtgc ccttattcct gatttgacc gcctgtgccc 12000
ttggtgtcca gataatccac cttatcggca atgaagtccg tcccgtagac cgtctggccg 12060
tccttctcgt acttggattt ccgaatcttg ccctgcacga ataccagcga ccccttgccc 12120
aaatacttgc cgtggccctc gccctgagag ccaaaacact tgatgcgga gaagtccggtg 12180
cgctcctgct tgtcgcgggt cgtggcccg ccaacctttg cgatccgcaa gcgcgcggtc 12240
gccatcttca cgctggaaca gtacgtcag ggggcatca tgaccgcga gcaatacag 12300
gtcattaaaa gcgcctgat tgatgatata gcggccggc tgcctctggt tctcgcgac 12360
cgaaatgggt gaactcacc cgctctctt gatcgtggca ccgatttccg cgatgctctc 12420
cgggaaaaag ccggggttgt cggccgtccg cggctgatgc ggatctcgt cgatcaggtc 12480
caggccagc tcgataggc cggaaaccgc ctgagacgcc gcaggagcgt ccaggaggct 12540
cgacaggctc ccgatgctat ccaaccccag gccggacggc tgcgcgcgc ctgcccctc 12600
ctgagcggcc gcagcgggt tttcttgggt ggtcttggct tgagccgag tcattgggaa 12660
atctccatct tcgtgaacac gtaatcagcc agggcgcaa cctcttctga tgccttgcc 12720
gcggccgtt tcttgatctt ccagaccggc acaccggatg cgaggcctc gccgatgctg 12780
ctgcgcaggc caacggtg gcgaaatc atcttgggt acgcggccag cagctcggct 12840
tgggtggcgc cgtggcgcg attccgcgca tcgacctgc tgggcacat gccaaaggaat 12900
tgacgcttg cgttctctg gcgacgttc gcaatggtc tgaccatctt cttgatgcc 12960
tggatgctg acgcctcaag ctcgatggg gacagcacat agtcggcgc gaagagggcg 13020
gccgccaggc cgacgcaag gtcggggcc gtgctgatca ggcacacgtc gaagccttg 13080
ttcgcaggc ccttgatggt cgcgccgaac agctcgcgg cgtcgtccag cgacagccgt 13140
tcggcgctc ccagtagcgg gttggactc atgagggcga ggcgcgcggc ctggccgctc 13200
ccggctcgc gtgcggttcc ggtccagcc cggcaggga cagcgcgaa cagcttgcct 13260
gcatgcaggc cggtagcaaa gtccttgagc gtgtaggac cattgccctg ggggtccag 13320
tcgatcacgg caaccgcaa gccgcgctc aaaaagtcga agcaagatg cacaagggtc 13380
gaagtcttgc cgacgccc tttctggtt gccgtgacca aagtttctat cgtttggtt 13440
cctgtttttt cttggcgtcc gcttccact tccggacgat gtacgctga tgttccgca 13500
gaaccgccc taccgcgcg taccctcgc gcaagtctt gtcctgaac gcggcccaca 13560
cgcgatgac cgttgcgac actgcgccc tggctcagtc cagcagcgtt gcgaacgctc 13620
cctgtggctt cccatcgact aagacgccc gcgctatctc gatggtctgc tgcccactt 13680
ccagcccctg gatcgcctc tggaaactgc tttcggtaag ccgtttctt atggataaca 13740
cccataattt gctccgccc ttggttgaac atagcgtga cagccgcaag cacatgagag 13800
aagtttagct aaacatttct cgcacgtcaa caccttagc cgtaaaaact cgtccttggc 13860
gtaacaaaac aaaagcccgc aaaccgggt ttcgtctctt gccgcttatg gctctgcacc 13920
cggctccatc acaaacaggt cgcgcacgc cttcactcgg ttgcggatc aactgcag 13980
cccaacaaag ccggttgcg ccgccgcaag gatcgcgcg atgatgcgg ccacaccggc 14040

-continued

```

catcgccac caggtcgccg ccttcgggtt ccattcctgc tggactgct tcgcaatgct 14100
ggacctcggc tcaccatagc ctgaccgctc gatggcgat gccgcttctc cccttggcgt 14160
aaaaccagc gccgcaggcg gcattgccat gctgcccgcc gctttccga ccacgacgcg 14220
cgcaccagc ttgcggtcca gacctcggc cacggcgagc tgcgcaagga cataatcagc 14280
cgccgacttg gctccacgcg cctcgatcag ctcttgcact cgcgcgaaat ccttggcctc 14340
cacggccgcc atgaatcgcg cacgcggcga aggctccgca gggccg 14386

```

<210> SEQ ID NO 19

<211> LENGTH: 14368

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic plasmid

<400> SEQUENCE: 19

```

ccaccagc cgcgcctc actgcccgc acctggtc tgaatgtcga tgccagcacc 60
tgcggcacgt caatgcttc gggcgtcgc ctcgggtga tcgccatcc cgttactgcc 120
ccgatcccg caatgcaag gactgccagc gctgccattt ttggggtag gccgttcgag 180
gccgaggggc gcagcccctg ggggatggg agggccgct tagcgggccc ggagggttcg 240
agaaggggg gcacccccct tcggcgtcgc cggtcacgc cacagggcg agccctggtt 300
aaaaacaag ttataaata ttggtttaa agcaggtta aagacaggtt agcggtgcc 360
gaaaaacgg cgaaacct tgcaaatgct ggattttctg cctgtggaca gccctcaa 420
tgtcaatag tgcgccctc atctgtcagc actctgccc tcaagtgtca aggatcgcg 480
ccctcatctg tcagtagtcg cccccctcaa gtgtcaatac cgcagggcac ttatccccag 540
gctgtcccac atcatctgt ggaaactcgc gtaaaatcag gcgttttcgc cgatttgca 600
ggctggccag ctccacgtc cggccgaaa tcgagcctgc ccctcatctg tcaacgccgc 660
gccgggtgag tcggcccctc aagtgtcaac gtccgccct catctgtcag tgagggccaa 720
gttttcgag aggtatccac aacgcggcg gcccggtgt ctcgcacacg gcttcgacg 780
cgtttctggc gcgtttgag gccatagac ggccgcagc ccagcggcga gggcaaccag 840
cccgtgagc gtcgaaag cgctcttcg cttcctcgt cactgactc ctgcgctcg 900
tcgttcgct gcggcgagc gtatcagctc actcaaagg ggaatacgg ttatccacag 960
aatcagggga taacgcagga aagaacatgt gagcaaaag ccagcaaaag gccaggaacc 1020
gtaaaaagg cgcgttgctg gcgttttcc ataggctccg cccccctgac gagcatcaca 1080
aaaaacgag ctcaagtca aggtggcga acccgacag actataaaga taccaggcgt 1140
ttccccctg aagctccctc gtgcgctc ctgttccgac cctgccgctt accggatacc 1200
tgtccgctt tctcccttc ggaagcgtg cgccattcgc cattcaggct gcgcaactgt 1260
tgggaagggc gatcggtagc gccctcttc ctattacgcc agctggcga aggggatgt 1320
gctgcaaggc gattaagttg ggtaacgcca gggtttccc agtcacgacg ttgtaaacg 1380
acggccagtg aattcggccg cgggtggcc gaggtgctgg gcaagcccta cctccaggcc 1440
cccacgggg tcgagagc gaccgcttc ctgcccgc tggcgagat tctggcctc 1500
gatccggagc ccttcacga gcgcgagaag cactcgacgc tgaagccgt gtgggatctg 1560

```

-continued

tggcggagtg	tcacgcagga	cttcttcggg	acggccaatt	tcggaatcgt	ggcgcaccgaa	1620
acttatgcaa	gaggcatccg	aaactatctc	gaaggcgatc	tcgggctgcc	ctgcgccttc	1680
gcctggcccc	caagaggggc	tcgaagaccg	acaacgaagc	ggtgcgcgga	ctgatccgcc	1740
agcaccgtcc	gctcgtgctc	atggggctca	tcaacgagaa	gatttacctt	gcggaactga	1800
aagccggtca	cgccccgca	accctcttcc	atcgcctgct	ctttcccggg	tcgggcgatc	1860
cggcgcgcta	ccggaaccgc	cgttatggga	tatgcagggt	ctacgtgggt	actgcaggaa	1920
gtttgcaacg	ccctgttcga	cgccctgttc	cacattctgc	ccctcgggac	ggagatggac	1980
agcgcgcgcc	ccacaccgac	gacactgcgc	cgcgacttcc	cgtgggatgc	cgatgcgcaa	2040
cgggcctcgg	accgcacatg	agaggagcat	ccggttctca	cccggatcag	cgccgcgcgt	2100
gccttgccgcg	acgccccgca	gaaggcttgc	cctcgcgatc	ggtgccgaga	gggtcgtgag	2160
agagactgtc	gaagccctga	cgtgggcccg	gcttcggcga	gaggaaggga	gagaaccaat	2220
gagcgatcat	gccgtcaaca	cgccggctca	tgccgccagg	gcccaacggc	accgagcacc	2280
tcgtgccgag	ttctacgtct	acttcgccgt	cattctgctg	ggcgccctcc	cggtgccctt	2340
cgtgagctgg	atcgtctcga	cgatccgccca	ccgcaggctt	ccaagcgcg	gccccttcgc	2400
gtccgcctgg	ttcgcgatcca	aggcgatcac	gccgctgatt	ttccgcgcct	gaccgcagggt	2460
caggttgcga	cacgccatcc	gtcgtctccc	caaggggccc	cggattaatc	gggagggcat	2520
ggtgccttac	cgtaaccacc	gccaccagta	ggcaggagga	acacgtgtcg	tccggtggtc	2580
accaccacca	ccaccaccac	taataggccg	gccctccgtc	gcgggcggca	cccacgcccc	2640
catcgattcc	aaggttcagc	cattgagacg	gctccgcttc	gcgcgcaagc	gcgggttggg	2700
ccgactgcaa	gcggagaggg	aagcatggca	ctgctcagct	tcgagcgaaa	atatcgcgtg	2760
ccggggggca	cgctggtcgg	cggaaacctg	ttcgcactct	gggtcggccc	ttctatgtc	2820
ggcttcttcg	gggttcgcag	gtttttcttc	gcggccctgg	gtatcattct	gattgcctgg	2880
agtgccttac	tccagggtac	ctggaacccc	caactcatct	ctgtctaacc	gccggccctt	2940
gaatatggcc	tgggaggtgc	acccctcgca	aaaggcgggc	tgtggcagat	catcaccgatc	3000
tgccccactg	gtgcctctgt	cagctggggc	ctgcgcgaag	tcgaaatctg	ccgtaagctg	3060
ggcatcgggt	accacatccc	gttcgccttc	gcgttcgcca	tcctggccta	cctgacgctg	3120
gtgctgttcc	gcccgggtgat	gatggggccc	tggggctatg	ccttccccta	cgggatctgg	3180
acgcacctcg	actgggtgtc	gaacaogggc	tacacctacg	gcaacttcca	ctacaaccct	3240
gcccacatga	tcgccatctc	gttcttcttc	acgaacgcgc	tggctctggc	gctgcacggc	3300
gcccttgtgc	tctccgcggc	caaccocgag	aagggcaagg	aaatgcggac	gccggatcac	3360
gaggatacgt	tcttccgcga	tctggtcggc	tactcgatcg	ggacgctcgg	catccaccgc	3420
ctcggcctgc	tgtctctcgt	gagcgccgct	ttcttcagcg	ccctctgcat	gatcattacc	3480
ggcaccatct	ggttcgatca	gtgggtcgac	tgggtggcaat	ggtgggtgaa	gctgccgtgg	3540
tggggcaaca	tcccgggag	catcaatggc	tgagtatcag	aacatcttct	cccaggtcca	3600
ggtccgcgga	ccggccgacc	tggggatgac	cgaagacgct	aacctggcca	accgttcggg	3660
cgctcgtccc	ttctcgaccc	tgctcggctg	gttcggcaac	gcccgactcg	gcccgatcta	3720
tctcggctcg	ctcggcgtcc	tgtccctctt	ctcgggcctg	atgtggttct	tcaccatcgg	3780
gatctgggtc	tggtatcag	cgggctgaa	cccggccgct	ttcctgcgcg	acctgttctt	3840

-continued

cttctcgctc gagccgccgg caccggaata cggctctgtcc ttcgcggtc cgtggaagga	3900
aggcgggctg tggctgatcg cgtcgttctt catgttcgtc gcggtctggt cctggtgggg	3960
ccgcacctat ctccgcgctc aggcgctggg catgggcaag cacaccgctt gggcgttctt	4020
ctcggccatc tggctgtgga tgggtctggg cttcatccgt ccgacccca tggggtcctg	4080
gtcggaaagc gttccctacg gcattctctc gcacctcgc tggacgaaca acttctcgct	4140
cgtccacggc aacctgttct acaacccctt ccacggctc tcgatcgctt tcctctacgg	4200
gtcggccctg ctcttcgca tgcacggtgc gaccatctc gcggtctccc gcttcggcgg	4260
cgagcgcgag ctggagcaga tcgccgaccg cgggacggca gcggagcggg ccgccctctt	4320
ctggcgctgg accatgggtt tcaacgccac gatggaaggc atccaccgct gggccatctg	4380
gatggcggtc ctcgtgacct tcaccggcgg catcgggacg ctgctctcgg gcacggtcgt	4440
ggacaactgg tacgtctggg gccagaacca cggcatggcg ccgctgaact gaggagcgat	4500
cacaatggct gacaagacca tcttcaacga tcacctcaac accaatccga agaccaacct	4560
tcgctctcgg gtcgctttcc agatgatgaa gggtgcgggc tgggctggcg gcgtgttctt	4620
cgggacgctc cttctcatcg ggttcttccg ggtggtcggg cggatgctt ccatccagga	4680
gaaccaggct ccggcgccga acatcaccgg cgctctggag accgggatcg agctgatcaa	4740
gcatctcgtc tgagacaagt ctcggggcag ggcggcgcga ggcggcccgc tcctccaagt	4800
ccgggccata tcgccggccc gggtcggggg cgacaccaca gcccggttcc cttctgttg	4860
gcgacaggga cctggtgccg tgtggaagac cgcacggcac ccttttgaca ttcacgggag	4920
gctctgatga ccaatccac ccgcgaccc gaaacccgc ttttgatcg cgtctgctgc	4980
ccggccgaca tgaaggcgt gagtgaccgc gaactggagc ggctggcga cgaagtgcgt	5040
tccgaggtca gtgatgggg tagtttcta ttttaggcag tttatatgaa attaagacat	5100
gcagatgta cagtggatat tgaactggtc tcgaaagctc aatatcccc aaagcacaag	5160
cacaaacttc gacatcatgc agaagcgttt cccgaaccgc gtcttcgacg tgggcatcgc	5220
cgagcagcat gccgtgacct tcgcgccggg cctcgccggg gccgggatga agcccttctg	5280
cgcgatctat tcctcgttcc tgcaacgggg ttacgaccag atcgccatg acgtggcgt	5340
gcagaacctt cccgtccgct tcgtgatcga ccgggcgggg ctcgtggggg ccgatggcgc	5400
gaccatgcg ggggccttcg acgttggctt catcacttcg ctgcccaaca tgaccgtgat	5460
ggccgcggcc gacgagccg agctcatcca catgatcgc accgccgtgg ccttcgacga	5520
gggccccatc gccttcgct tcccgcgggg cgagggggtg ggcgtcgaga tgcccgagcg	5580
cgggacggtg ctggagcccg gccggggcgg cgtggtgcgc gaagggacgg atgtcgcgat	5640
cctctcttc ggcgcgcatc tgcacgaggc cttgcaggcg gcgaaacttc tcgaggccga	5700
gggggtgagc gtgaccgtgg ccgacgcccg cttctcgcgc ccgctcgaca cggggtcat	5760
cgaccagctc gtgcgccatc acgcgcgct ggtaacggtg gagcaggggg ccatgggcgg	5820
cttcggcgc catgtcatgc actatctcgc caattccggc ggcttcgacg gggcctcgc	5880
gctccgggtc atgacgctgc ccgaccgctt catcgagcag gcgagccccg aggacatgta	5940
tgccgatgcg gggctcgggg ccgaggatat caagcttggc gtaatcatgg tcatagctgt	6000
ttcctgtgtg aaattgttat ccgctcaca ttccacaca catacgagcc ggaagcataa	6060
agtgtaaagc ctggggtgcc taatgagtga gctaactcac attaatgctg ttgcgctcac	6120

-continued

tgcccgcttt ccagtcggga aacctgtcgt gccagctgca ttaatgaatc ggccaacgcg	6180
cggggagagc cggtttgcgt attgggcgct cggctctgcc ttgctcgtcg gtgatgtact	6240
tcaccagctc cgcgaagtcg ctcttcttga tggagcgcac ggggacgtgc ttggcaatca	6300
cgcgcacccc ccggccgctt tagcggctaa aaaagtcatg gctctgccct cgggcggacc	6360
acgcccatac tgaccttgcc aagctcgtcc tgcttctctt cgatcttcgc cagcaggcgc	6420
aggatcgtgg catcaccgaa ccgcgccgtg cgcgggctgt cggtagacca gagtttcagc	6480
aggccgcccc ggcggccagc gtcgccattg atgcgggcca gctcgcggac gtgctcatag	6540
tccacgacgc ccgtgatttt gtagccctgg ccgacggcca gcaggtaggc cgacaggctc	6600
atgccggcgc ccgccgcctt ttctctcaatc gctcttcggt cgtctggaag gcagtacacc	6660
ttgataggtg ggctgcctt cctggttggc ttggtttcat cagccatccg cttgccctca	6720
tctgttacgc cggcggtagc cggccagcct cgcagagcag gattcccgtt gaccaccgcc	6780
aggtgcgaat aagggacagt gaagaaggaa caccgcctcg cgggtgggcc tacttcacct	6840
atctgcccc gctgacgcc ttggatacac caaggaaagt ctacacgaac cctttggcaa	6900
aatcctgtat atcgtgcgaa aaaggatgga tataccgaaa aaatcgctat aatgacccc	6960
aagcagggtt atgcacggga aaagcggcac gcttcccga gggagaaagg cggacaggta	7020
tccgtaagc ggcagggctg gaacaggaga gcgcacgagg gagcttccag ggggaaacgc	7080
ctggtatctt tatagtcctg tcgggtttcg ccacctctga cttgagcgtc gatttttgtg	7140
atgctcgtca gggggcgga gcctatggaa aaacgccagc aacgggcct ttttacggtt	7200
cctggccttt tgctggcctt ttgctcacat gttctttcct gcgttatccc ctgattctgt	7260
ggataaccgt attacgcct ttgagtgagc tgataccgct cgcgcagacc gaacgaccga	7320
gcgcagcgag tcagtgagc aggaagcgga agagcgcagc aagggccca gagaggccga	7380
gcgcggcctg gagccttggc cgcctaggga gggcatgaaa aagcccgtag cgggctgcta	7440
cgggctctg acgcggtgga aagggggagg gtaggttgtc tacatggctc tgctgtagtg	7500
agtgggttg ctcggcgag cggctcctgat caatcgctac cctttctcgg tccttcaacg	7560
ttctgacaa cagcctcct tttcgccaat ccctgacaa tcaccgcgag tccctgctcg	7620
aacgctgcgt ccggaccggc ttcgtogaag gcgtctatcg cggcccga cagcggcgag	7680
agcggagcct gttcaacggt gccgcgcgc tcgcccgcac cgtgctgcc ggcctgctc	7740
tcaagcacgg ccccaacagt gaagtagctg attgtcatca gcgcattgac ggcgtcccc	7800
gccgaaaaac ccgcctcgca gaggaagcga agctgcgcgt cggccgtttc catctgcggt	7860
gcgcccggtc gcgtgcccgc atggatgcgc gcgccatgc ggtaggcgag cagcgcctgc	7920
ctgaagctgc gggcattccc gatcagaaat gagcgcagc cgtcgtcggc tctcggcacc	7980
gaatgcgtat gattctccgc cagcatggct tcggccagtg cgtcgcagc cgcggcttg	8040
ttctgaagt gccagtaaag ccggcgctgc tgaacccca accgttccgc cagtttgcgt	8100
gtcgtcagac cgtctacgcc gacctgttc aacaggtcca gggcggcac gatcactgta	8160
ttcggctgca actttgtcat gcttgacct ttatcactga taaacataat atgtccacca	8220
acttatcagt gataaagaat ccgcgcgttc aatcgacca gcggaggctg gtcggaggc	8280
cagacatgaa acccaacata ccctgatcg taattctgag cactgtcgcg ctcgacgctg	8340
tcggcatcgc cctgattatg ccggtgctgc cgggcctcct gcgcgatctg gttcactcga	8400

-continued

acgacgtcac	cgcccactat	ggcattctgc	tggcgctgta	tgcgttggtg	caatttgctt	8460
gcgcacctgt	gctgggcgcg	ctgtcggatc	gtttcgggcg	gcgccaatc	ttgctcgtct	8520
cgctggccgg	cgccactgtc	gactacgcca	tcatggcgac	agcgcctttc	ctttgggttc	8580
tctatatcgg	gcgatcgtg	gccggcatca	ccggggcgac	tggggcggtg	gccggcgctt	8640
atattgccga	tatactgat	ggcagtagc	gcgcgcgca	cttcggcttc	atgagcgctt	8700
gtttcgggtt	cggtatggtc	gcgggacctg	tgctcgggtg	gctgatgggc	ggtttctccc	8760
ccccagctcc	gttcttcgcc	gcggcagcct	tgaacggcct	caatttctctg	acgggctggt	8820
tccttttgcc	ggagtcgcac	aaaggcgaac	gccggccggt	acgccgggag	gctctcaacc	8880
cgctcgtctc	gttccgggtg	gccccgggca	tgaccgtcgt	cgccgcctctg	atggcggtct	8940
tcttcatcat	gcaactgtc	ggacaggtgc	cggccgcgct	ttgggtcatt	ttcggcgagg	9000
atcgctttca	ctgggacgcg	accacgatcg	gcatttcgct	tgccgcattt	ggcattctgc	9060
attcactcgc	ccaggcaatg	atcaccggcc	ctgtagccgc	ccggctcggc	gaaaggcggg	9120
cactcatgct	cggaatgatt	gccgacggca	caggctacat	cctgcttgc	ttcgcgacac	9180
gggatggat	ggcgttcccg	atcatggtcc	tgcttgcttc	gggtggcatc	ggaatgcccg	9240
cgctgcaagc	aatgttctcc	aggcaggtgg	atgaggaacg	tcaggggag	ctgcaaggct	9300
cactggcggc	gctcaccagc	ctgacctcga	tcgtcggacc	cctcctcttc	acggcgatct	9360
atgcgcttc	tataacaacg	tggaaacggg	gggcatggat	tgacggcgtc	gccctctact	9420
tgctctgcct	gccggcgctg	cgctcggggc	tttgagcggg	cgcagggcaa	cgagccgatc	9480
gctgatcgtg	gaaacgatag	gcctatgcc	tgcgggtcaa	ggcgacttcc	ggcaagctat	9540
acgcgcccta	ggagtgcggt	tggaaacgtg	gccagccag	atactcccga	tcacgagcag	9600
gacgccgatg	atttgaagcg	cactcagcgt	ctgatccaag	aacaaccatc	ctagcaacac	9660
ggcggtcccc	gggtgagaa	agcccagtaa	ggaacaact	gtaggttcga	gtcgcgagat	9720
cccccggaac	caaaggaagt	aggttaaacc	cgctccgatc	aggccgagcc	acgccaggcc	9780
gagaacattg	gttctctgag	gcctcgggat	tggcggatca	aacctaaag	ctactggaac	9840
gagcagaagt	cctccggccc	ccagttgcc	ggcggtaaag	gtgagcagag	gcaocgggag	9900
ttgccacttg	cggttcagca	cggttccgaa	cgccatggaa	accgcccccg	ccaggcccgc	9960
tgcgacgccg	acaggatcta	gcgctgcgtt	tgggtgcaac	accaacagcg	ccacgccccg	10020
agttccgcaa	atagccccca	ggaccgccat	caatcgtatc	gggctacctc	gcagagcggc	10080
agagatgaac	acgaccatca	gcggctgcac	agcgcctacc	gtcgcgcgca	ccccgcccg	10140
caggcggtag	accgaaataa	acaacaagct	ccagaatagc	gaaatattaa	gtcgcgcgag	10200
gatgaagatg	cgcatccacc	agattcccgt	tggaatctgt	cggacgatca	tcaogagcaa	10260
taaacccgcc	ggcaacgccc	gcagcagcat	accggcgacc	cctcggcctc	gctgttcggg	10320
ctccacgaaa	acgccggaca	gatgcgcctt	gtgagcgtcc	ttggggccgt	cctcctgttt	10380
gaagaccgac	agcccaatga	tctcgcgctc	gatgtaggcg	ccgaatgcc	cggcatctcg	10440
caaccgttca	gcgaacgcct	ccatgggctt	tttctcctcg	tgctcgtaaa	cggacccgaa	10500
catctctgga	gctttcttca	gggccgacaa	tcggatctcg	cggaaatcct	gcacgtcggc	10560
cgctccaagc	cgtcgaatct	gagccttaat	cacaattgtc	aattttaatc	ctctgtttat	10620
cggcagttcg	tagagcgcgc	cgtgcgtccc	gagcgatact	gagcgaagca	agtgcgtcga	10680

-continued

gcagtgcccc	cttgttcctg	aaatgccagt	aaagcgtgg	ctgctgaacc	cccagccgga	10740
actgacccca	caagcccta	gcgtttgcaa	tgcaccaggt	catcattgac	ccagggcgtgt	10800
tccaccaggc	cgctgcctcg	caactcttcg	caggcttcgc	cgacctgctc	gcgccacttc	10860
ttcacgcggg	tggaatccga	tccgcacatg	aggcggaaag	tttccagctt	gagcgggtac	10920
ggctcccggg	gcgagctgaa	atagtcgaac	atccgtcggg	ccgtcggcga	cagcttgcg	10980
tacttctccc	atatgaatth	cgtgtagtgg	tcgccagcaa	acagcacgac	gatttctctg	11040
tcgatcagga	cctggcaacg	ggacgttttc	ttgccacggg	ccaggacgcg	gaagcgggtg	11100
agcagcgaca	ccgattccag	gtgcccaacg	cggtcggacg	tgaagcccat	cgccgtcgcc	11160
tgtaggcgcg	acaggcattc	ctcggccttc	gtgtaatacc	ggccattgat	cgaccagccc	11220
aggtcctggc	aaagctcgta	gaacgtgaag	gtgatcggct	cgccgatagg	ggtgcgcttc	11280
gcgtaactca	acacctgctg	ccacaccagt	tcgtcatcgt	cggcccgcag	ctcgacgccc	11340
gtgtaggtga	tcttcacgtc	cttggtgacg	tggaaaatga	ccttgttttg	cagcgccctg	11400
cgcggggatt	tcttggtgcg	cgtggtgaac	agggcagagc	gggcccgtgc	gtttggcctc	11460
gctcgcatcg	tgtccggcca	cggcgcaata	tcgaacaagg	aaagctgcat	ttccttgatc	11520
tgctgcttcg	tgtgtttcag	caacgcggcc	tgcttgccct	cgctgacctg	ttttgccagg	11580
tcctcgccgg	cggttttctg	cttcttggtc	gtcatagtcc	ctcgcgtgct	gatggctcctc	11640
gacttcgcca	aaactgccgc	ctcctgcttc	agacgacgcg	aacgctccac	ggcggccgat	11700
ggcgcgggca	gggcaggggg	agccagttgc	acgctgctgc	gctcgatcct	ggcgttagct	11760
tgctggacca	tcgagccgac	ggactggaag	gtttcgccgg	gcgcacgcat	gacggtgccc	11820
cttgcgatgg	tttcggcctc	ctcggcgcaa	aaccccgctg	cgatcagttc	ttgcctgtat	11880
gccttcgggt	caaacgtccg	attcattcac	cctccttgcg	ggattgcccc	gactcacgccc	11940
ggggcaatgt	gcccttattc	ctgatttgac	ccgcctggtg	ccttggtgct	cagataatcc	12000
accttatcgg	caatgaagtc	ggtcccgtag	accgtctggc	cgctcctctc	gtacttggtg	12060
ttccgaatct	tgccctgcac	gaataccagc	gacccttgc	ccaaatactt	gccgtgggccc	12120
tcggcctgag	agccaaaaca	cttgatgcgg	aagaagtcgg	tcgctcctg	cttgcgcgcc	12180
gtcgtggccc	cgccaacctt	tgcgatccgc	aagcgcgcgg	tcgccatcct	cacgctggaa	12240
cagtacgtcg	aggcgggcat	catgaccgcc	gagcaatagc	aggtcattaa	aagcgcgctg	12300
attgatgata	tagcggcccc	gctgctcctg	gttctcgcgc	accgaaatgg	gtgacttcac	12360
cccgcgctct	ttgatcgtgg	caccgatttc	cgcgatgctc	tccggggaaa	agccgggggt	12420
gtcggccgct	cgcggctgat	gcggatcttc	gtcgatcagg	tccaggtoaca	gctcgatagg	12480
gccggaaccg	ccctgagacg	ccgcaggagc	gtccaggagg	ctcgacaggt	cgccgatgct	12540
atccaacccc	aggccggacg	gctgcgccgc	gcctgcggct	tcctgagcgg	ccgcagcggg	12600
gtttttcttg	gtggtcttgg	cttgagccgc	agtcattggg	aaatctccat	cttcgtgaa	12660
acgtaatcag	ccagggcgcg	aacctcttcc	gatgccttgc	gcgcggccgt	tttcttgatc	12720
ttccagaccg	gcacaccgga	tgcgagggca	tcggcgatgc	tgctgcgcag	gccaaccggtg	12780
gccggaatca	tcactcttgg	gtacgcggcc	agcagctcgg	cttggtggcg	cgcgtggcgc	12840
ggattccgcg	catcgacctt	gctgggcacc	atgccaagga	attgcagctt	ggcgttcttc	12900
tgggcacagt	tcgcaatggt	cgtgaccatc	ttcttgatgc	cctggatgct	gtacgcctca	12960

-continued

```

agctcgatgg gggacagcac atagtcggcc gcgaagaggg cggccgccag gccgacgcca 13020
agggtcgggg ccgtgtcgat caggcacacg tcgaagcctt ggttcgccag gcccttgatg 13080
ttcgccccga acagctcgcg ggcgtcgtcc agcgacagcc gttcgcggtt cgccagtacc 13140
gggttggaact cgatgagggc gaggcgcgcg gcctggccgt cgcgggctgc gggtgcggtt 13200
tcggtccagc cgccggcagg gacagcgccg aacagcttgc ttgcatgcag gccggtagca 13260
aagtccctga gcgtgtagga cgcattgccc tgggggtcca ggtcgatcac ggcaaccgca 13320
aagccgcgct cgaaaaagtc gaaggcaaga tgcacaaggg tcgaagtctt gccgacgccc 13380
cctttctggt tggccgtgac caaagtttc atcgtttggt ttctgtttt ttcttgcggt 13440
ccgcttcccc cttccggagc atgtacgcct gatgttccgg cagaaccgcc gttaccgcg 13500
cgtacccttc gggcaagttc ttgtcctcga acgcgcccca cacgcgatgc accgcttgcg 13560
aactgcgcc cctggtcagt cccagcgagc ttgcgaacgt cgcctgtggc tccccatcga 13620
ctaagagccc ccgcgctatc tcgatggtct gctgccccac ttccagcccc tggatcgctt 13680
cctggaactg gctttcggtg agccgtttct tcatggataa caccataat ttgctccgcg 13740
ccttggttga acatagcggg gacagcccgc agcacatgag agaagttag ctaaaccattt 13800
ctcgacgctc aacaccttta gccgtaaaa ctcgtccttg gcgtaaaaa aaaaaagccc 13860
ggaaaccggg ctttctgtct ttgccgtta tggctctgca cccggctcca tcaccaacag 13920
gtcgcgcacg cgcttactc ggttgcgat cgacactgcc agcccaaaa agccggttgc 13980
cgccgcgccc aggatcgcg cgatgatgcc ggccacaccg gccatcgccc accaggtcgc 14040
cgcttccgg ttccattcct gctggtactg cttcgcaatg ctggacctcg gctcaccata 14100
ggctgaccgc tcgatggcgt atgccgcttc tccccttggc gtaaaaccca gcgccgagc 14160
cggcattgcc atgctgccc cgctttccc gaccacgagc cgcgcaccag gcttgcggtc 14220
cagaccttcg gccacggcga gctgcgcaag gacataatca gccgccgact tggctccacg 14280
cgctcgatc agctcttgca ctcgcgcgaa atccttgccc tccacggccg ccatgaatcg 14340
cgcacgccc gaaggtccc cagggccc 14368

```

<210> SEQ ID NO 20

<211> LENGTH: 14375

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic plasmid

<400> SEQUENCE: 20

```

ccaccagc cgccgcctc actgcccgc acctggtcgc tgaatgtcga tgccagcacc 60
tcgcgacagt caatgcttc gggcgtcgc ctcgggctga tcgccatcc cgtaactgcc 120
ccgatcccc caatggcaag gactgccagc gctgccattt ttggggtgag gccgttcgcg 180
gccgaggggc gcagcccctg ggggatggg aggccgcgt tagcgggccc ggagggttcg 240
agaaggggg gcacccccct tcggcgtcgc cggtcacgc cacagggcgc agccctggtt 300
aaaaacaag ttataaata ttggtttaa agcaggttaa aagacaggtt agcggtgccc 360
gaaaaacgg cgaaaccct tgcaaatgct ggattttctg cctgtggaca gccctcaaa 420
tgtcaatagg tgcgccctc atctgtcagc actctgccc tcaagtgtca aggatcgcgc 480
ccctcatctg tcagtactc cgcctcaca gtgtcaatac cgcagggcac ttatccccag 540

```

-continued

gcttgtccac atcatctgtg ggaaactcgc gtaaaatcag gcgttttcgc cgatttgcca	600
ggctggccag ctcccagctg ccggccgaaa tcgagcctgc ccctcatctg tcaacgccgc	660
gccgggtgag tcggcccctc aagtgtcaac gtccgcccct catctgtcag tgagggccaa	720
gttttcgcgc aggtatccac aacgccggcg gcccggtgt ctcgcacacg gcttcgacgg	780
cgtttctggc gcgtttgcaq ggccatagac ggccgccagc ccagggcga gggcaaccag	840
cccggtgagc gtcgaaagg cgctcttcgc cttcctcgt cactgactcg ctgcgctcgg	900
tcgttcggct gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg ttatccacag	960
aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc	1020
gtaaaaaggc cgcgttgctg gcgtttttc ataggctccg cccccctgac gagcatcaca	1080
aaaaatcgac ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt	1140
ttccccctg agctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc	1200
tgccgcctt tctcccttcg ggaagcgtgg cgccattcgc cattcaggct gcgcaactgt	1260
tgggaagggc gatcggctgc gccctcttcg ctattacgcc agctggcga aggggatgt	1320
gctgcaaggc gattaagttg ggtaacgca gggttttccc agtcacgacg ttgtaaaacg	1380
acggccagtg aattcggccg cgggctggcc gagtgctgg gcaagcccta cctccaggcc	1440
cccatcgggg tcgagagcac gaccgccttc ctgcgccgcc tggcgagat tctgggcctc	1500
gatccggagc cttcatcga gcgcgagaag cactcgacgc tgaagcccgt gtgggatctg	1560
tggcgagtg tcacgcagga cttcttcgg acggccaatt tcggaatcgt ggcgaccgaa	1620
acttatgcaa gaggcacccg aaactatctc gaaggcgatc tcgggctgcc ctgcgccttc	1680
gccgtggccc gcaagagggg ctcgaagacc gacaacgaag cggtgccgg actgatccgc	1740
cagcacccgc cgcctgtgct catggggtcg atcaacgaga agatttaect tgcggaactg	1800
aaagccggtc acggcccga accctcttc atcgtgcct ctttccggg tgcggcgatc	1860
cggcgcgcta ccggaaccgc cgttatggga tatgcagtg ctacgtggtt actgcaggaa	1920
gtttgcaacg ccctgttcga cgccctgttc cacattctgc ccctcgggac ggagatggac	1980
agcggccgc ccacaccgac gacactgcgc cgcgacttcc cgtgggatgc cgatgcgcaa	2040
gcggccctg accgcatcgt agaggagcat ccggttctca cccggatcag cgcgcgcgt	2100
gccttgccgc acgccccga gaaggctgcc ctcgatgcc gtgccgagag ggtcgtgaga	2160
gagactgtgc aagccctgcg tgggcccggc ttcggcgaga ggaagggaga gaaccaatga	2220
gcgatcatgc cgtcaacacg ccggtccatg ccgccagggc ccacgggcac cgagcaccac	2280
gtgccgagtt ctacgtctac ttcgcgctca ttctgctggg cgccttcccg gtggccttcg	2340
tgagctggat cgtctcgacg atccgccacc gcaggcttcc caagcgcggc cccttcgcgt	2400
ccgctggtt cgatgccaaq gcgatcacgc cgtgatctt cgcgcctga ccgcaggctca	2460
ggttgcgaca cgccattcgt cgtctccca agggcgggc gattaatcgg gagggcatgg	2520
tgcttaccg taaccaccgc caccagcatg tggcgactag ttccatatga tagatctac	2580
caccaccacc accaccacca ccaccactaa taggccggcc ctcgctcgcg ggcggcacc	2640
acggccgcat cgattccaag gttcagccat tgagacggct ccgcttcgcg cgcaagcgcg	2700
ggttgggccc actgcaagcg gagagggag catggcactg ctcagcttcg agcgaaaata	2760
tcgctgccc gggggcacgc tggctggcg aaacctgttc gacttctggg tcggccctt	2820

-continued

ctatgtcggc	ttcttcgggg	ttgcgacggt	tttcttcgcg	gccttgggta	tcattctgat	2880
tgcttgaggt	gccgtactcc	agggtaacctg	gaacccccaa	ctcatctctg	tctacccgcc	2940
ggccctgaa	tatggcctgg	gaggtgcacc	cctcgcaaaa	ggcgggctgt	ggcagatcat	3000
cacgatctgc	gccactggtg	ccttcgctcag	ctgggcgctg	cgcgaagtcg	aaatctgccg	3060
taagctgggc	atcgggtacc	acatcccgtt	cgccttcgcg	ttcgccatcc	tggcctacct	3120
gacgctgggt	ctgttcgcc	cggatgatgat	gggcgcctgg	ggctatgect	tcccctacgg	3180
gatctggacg	cacctcgact	gggtgtcgaa	cacgggctac	acctacggca	acttccacta	3240
caaccctgcc	cacatgatcg	ccatctcgtt	cttcttcacg	aacgcgctgg	ctctggcgct	3300
gcacggcgcc	ctgtgtctct	ccgcggccaa	ccccgagaag	ggcaaggaaa	tcgggacgcc	3360
ggatcacgag	gatacgttct	tccgcgatct	ggtcggctac	tcgatcggga	cgctcggcat	3420
ccaccgcctc	ggcctgctgc	tctcgtctgag	cgcctcttc	ttcagcgcgc	tctgcatgat	3480
cattaccggc	accatctggt	tcgatcagtg	ggtcgactgg	tggcaatggt	gggtgaagct	3540
gccgtgggtg	gcgaacatcc	cgggaggcat	caatggctga	gtatcagaac	atcttctccc	3600
aggtccaggt	ccgcggaccg	gccgaacctg	ggatgaccga	agacgtcaac	ctggccaacc	3660
gttcgggctg	cggctccttc	tcgacctcgc	tcgctgggtt	cggcaacgcc	cagctcggcc	3720
cgatctatct	cggctcgtct	ggcgtcctgt	ccctcttctc	ggcctgatg	tggttcttca	3780
ccatcgggat	ctggttctgg	tatcaggcgg	gctggaaccc	ggcctcttc	ctgcgcgacc	3840
tggtcttctt	ctcgtctgag	ccgcggcac	ccgaatacgg	tctgtccttc	gcggctccgc	3900
tgaaggaagg	cgggtctggt	ctgatcgcgt	cgttcttcat	gttcgtcgcg	gtctggtcct	3960
gggtggggccg	cacctatctc	cgcgctcagg	cgttgggcat	gggcaagcac	accgcctggg	4020
cgttctctct	ggccatctgg	ctgtggatgg	tgctgggctt	catccgtccg	atcctcatgg	4080
ggctcgtgtc	ggaagcgggt	ccctacggca	tcttctcgca	cctcgcactg	acgaacaact	4140
tctcgtcgt	ccacggcaac	ctgttttaca	accccttcca	cggctctctg	atcgccttcc	4200
tctacgggtc	ggccctgctc	ttcgcgatgc	acggtgcgac	catcctcgcg	gtctcccgcct	4260
tcggcgcgca	gcgcgagctg	gagcagatcg	ccgaccgcgg	gacggcagcg	gagcggggccg	4320
ccctcttctg	gcgctggacc	atgggtttca	acgccacgat	ggaaggcatc	cacgcctggg	4380
ccatctggat	ggcggctcct	gtgacctca	ccgcggcat	cgggatcctg	ctctcgggca	4440
cggctcgtgga	caactggtac	gtctggggcc	agaaccacgg	catggcgccg	ctgaactgag	4500
gagcgatcac	aatggctgac	aagaccatct	tcaacgatca	cctcaacacc	aatccgaaga	4560
ccaaccttcg	cctctgggtc	gctttccaga	tgatgaaggg	tcggggctgg	gctggcggcg	4620
tggtctctcg	gacgctcctt	ctcatcgggt	tcttcgggt	ggtcggggcg	atgcttccga	4680
tccaggagaa	ccaggctccg	gcgccgaaca	tcaccggcgc	tctggagacc	gggatcggagc	4740
tgatcaagca	tctcgtctga	gacaagtctc	ggggcagggc	ggcgcgaggc	cgcccgcctc	4800
tccaagtccg	ggcatatctg	ccggcccggg	tccggggcga	caccacagcc	cggttccctt	4860
cctgttggtg	acagggacct	ggtgcogtgt	ggaagaccgc	acggcacctt	tttgacattc	4920
acgggaggct	ctgatgacca	atcccacccc	gcgaccgaa	accccgtttt	tgatcgcgct	4980
ctgctgcccg	gccgacatga	aggcgcctgag	tgacgccgaa	ctggagcggc	tgcccgacga	5040
agtgcgttcc	gaggtcagtg	ataggggtag	tttcttattt	taggcagttt	atatgaaatt	5100

-continued

aagacatgca gatgtcacag tggatattga actggtctcg aaagctcaat atccccaaa 5160
gcacaagcac aaacttcgac atcatgcaga agcgtttccc gaaccgctc ttcgacgtgg 5220
gcatcgccga gcagcatgcc gtgaccttcg cggccggcct cggcggggcc gggatgaagc 5280
ccttctgcgc gatctattcc tcgttctcgc aacggggtta cgaccagatc gcccatgacg 5340
tggcgctgca gaaccttccc gtccgcttcg tgatcgaccg ggcggggctc gtgggggccc 5400
atggcgcgac ccgatcgggg gccttcgacg ttggcttcat cacttcgctg cccaacatga 5460
ccgtgatggc cgcggccgac gaggcgacg tcattcccat gatcgccacc gccgtggcct 5520
tcgacgaggg ccccatcgcc ttccgcttcc cgcggggcga gggggtgggc gtcgagatgc 5580
ccgagcgcgg gacggtgctg gagcccgcc gggcccgct ggtgcgcgaa gggacggatg 5640
tcgcatcct ctccctcggc gcgcatctgc acgaggcctt gcaggcggcg aaacttctcg 5700
agggcaggg ggtgagcgtg accgtggccg acgcccgtt ctccgcccgc ctcgacacgg 5760
ggctcatcga ccagctcgtg cgcctcacg cggcgctggt aacggtggag cagggggcca 5820
tggcggtt cggcgccat gtcatgcact atctcgcaa ttccggcggc ttcgacgggg 5880
gcctcgcgct ccgggtcatg acgctgccg accgcttcat cgagcaggcg agccccgag 5940
acatgtatgc cgatcgggg ctgcccggc aggatataca gcttggcgta atcatggtca 6000
tagctgtttc ctgtgtgaaa ttgttatccg ctcaaatc cacacaacat acgagccgga 6060
agcataaagt gtaaagcctg ggggtcctaa tgagtgagct aactcacatt aattgcgttg 6120
cgctcactgc ccgctttcca gtcgggaaac ctgtcgtgcc agctgcatta atgaatcggc 6180
caacgcgcgg ggagagggcg ttgctgatt gggcgctcgg tcttgccttg ctgctcgtg 6240
atgtacttca ccagctccgc gaagtgcctc ttcttgatgg agcgcagggg gacgtgcttg 6300
gcaatcacgc gcacccccg gccgttttag cggctaaaa agtcatggct ctgccctcgg 6360
gcggaaccg cccatcatga ccttgccaag ctgcctcgc ttctcttoga tcttcgccag 6420
cagggcgagg atcgtggcat caccgaaccg cgcctgctgc gggctcctgg tgagccagag 6480
ttcagcagg ccgcccagg gccccaggc gccattgat cgggccagct cgcggacgtg 6540
ctcatagtcc acgaccccg tgattttgta gccctggccg acggccagca ggtaggccga 6600
caggctcatg ccggccggc ccgcttttc ctcaatcgt ctctgctcgt ctggaaggca 6660
gtacaccttg ataggtgggc tgcccttctt ggttggtctg gtttcatcag ccattccgctt 6720
gccctcatct gttacgccg cggtagccg ccagcctcgc agagcaggat tcccgttgag 6780
cacgccagc tcgcaataag ggacagtga gaaggaacac ccgctcgcgg gtgggcctac 6840
ttcacctatc ctgcccggc gacgcgctg gatacaccaa gaaagteta cacgaacct 6900
ttggcaaat cctgtatc gtgcgaaaa ggtgggat accgaaaaa tcgctataat 6960
gaccccgaag cagggtatg cagcgaaaa gcgccacgct tcccgaaggg agaaggcgg 7020
acaggtatcc ggtaagcgc agggctcgaa caggagagc cacgaggag cttccagggg 7080
gaaacgctg gtatctttat agtctgtcg ggttctgcca cctctgactt gagcgtgat 7140
ttttgtgat ctgctcagg gggcggagcc tatggaaaa cgcagcaac gcggccttt 7200
tacggttctt gcccttttgc tggcctttg ctccatggt ctttctcgc ttatccccctg 7260
attctgtgga taaccgtatt accgctttg agtgagctga taccgctcgc cgcagccgaa 7320
cgaccgagc cagcgagtca gtgagcagg aagcggaaaga gcgccagaag gccgccagag 7380

-continued

aggccgagcg	cggccgtgag	gcttgacgc	tagggcaggg	catgaaaaag	cccgtagcgg	7440
gctgctacgg	gcgtctgacg	cggtgaaaag	ggggagggga	tgttgtctac	atggctctgc	7500
tgtagtgagt	gggttgcgct	ccggcagcgg	tcctgatcaa	tcgtcacccct	ttctcggctc	7560
ttcaacggtc	ctgacaacga	gcctcctttt	cgccaatcca	tcgacaatca	ccgcgagctc	7620
ctgctcgaac	gctgcgtccg	gaccggcttc	gtcgaaggcg	tctatcgccg	cccgaacag	7680
cggcgagagc	ggagcctggt	caacgggtgc	gccgcgctcg	ccggcatcgc	tgtcgcggc	7740
ctgctcctca	agcacggccc	caacagtga	gtagctgatt	gtcatcagcg	cattgacggc	7800
gtccccggcc	gaaaaaccgg	cctcgcagag	gaagcgaagc	tgcgcgctcg	ccgtttccat	7860
ctgcggtgcg	cccggctcgg	tgccggcatg	gatgcgcgcg	ccatcgcggg	aggcgagcag	7920
cgcctgcctg	aagctcgggg	cattcccgat	cagaaatgag	cgccagtcgt	cgctcggctc	7980
cggcaccgaa	tcgctatgat	tctccgccag	catggcttcg	gccagtgcgt	cgagcagcgc	8040
ccgcttggtc	ctgaagtgcc	agtaaagcgc	cggctgctga	acccccaacc	gttccgccag	8100
tttgctgtc	gtcagaccgt	ctacgccgac	ctcgttcaac	aggccaggg	cggcacggat	8160
cactgtattc	ggctgcaact	ttgtcatgct	tgacacttta	tactgataa	acataatag	8220
tccaccaact	tatcagtgat	aaagaatccg	cgcgttcaat	cggaccagcg	gaggctggtc	8280
cggaggccag	acatgaaacc	caacataccc	ctgatcgtaa	ttctgagcac	tgtcgcgctc	8340
gacgctgctg	gcatcggcct	gattatgccg	gtgctgccgg	gcctcctcgg	cgatctgggt	8400
cactcgaacg	acgtcaccgc	ccactatggc	attctgctgg	cgtgtgatgc	gttggtgcaa	8460
tttgctgctg	cacctgtgct	gggcgcgctg	tcggatcgtt	tcgggcggcg	gccaatcttg	8520
ctcgtctcgc	tgcccgcgcc	cactgtcgac	tacgccatca	tggcgacagc	gcctttcctt	8580
tgggttctct	atatcggggc	gatcgtggcc	ggcatcaccg	ggcgactgg	ggcggtagcc	8640
ggcgttata	ttgccgatat	cactgatggc	gatgagcgcg	cgcggcactt	cggcttcatg	8700
agcgcctggt	tcgggttcgg	gatggtcgcg	ggacctgtgc	tcggtgggct	gatgggcggg	8760
ttctcccccc	acgctccggt	cttcgcccg	gcagccttga	acggcctcaa	ttcctgacg	8820
ggctgtttcc	ttttgccgga	gtcgcacaaa	ggcgaacgcc	ggcgttacg	ccgggagggc	8880
ctcaaccgcg	tcgcttcggt	ccggtgggcc	cggggcatga	ccgtcgtcgc	cgccctgatg	8940
ggcgtcttct	tcatcatgca	acttgtcggg	caggtgccgg	ccgcgctttg	ggcattttc	9000
ggcgaggatc	gctttcactg	ggacgcgacc	acgatcggca	tttcgcttgc	cgcatttggc	9060
attctgcatt	cactcgccca	ggcaatgatc	accggccctg	tagccgcocg	gctcggcgaa	9120
aggcgggcac	tcatgctcgg	aatgattgcc	gacggcacag	gctacatcct	gcttgccttc	9180
gcgacacggg	gatggatggc	gttcccgatc	atggtcctgc	ttgcttcggg	tggcatcgga	9240
atgccggcgc	tgcaagcaat	gttgtccagg	caggtggatg	aggaacgtca	ggggcagctg	9300
caaggctcac	tggcggcgct	caccagcctg	acctcgatcg	tcggaccocct	cctcttcacg	9360
gcgatctatg	cggcttctat	aacaacgtgg	aacgggtggg	catggattgc	aggcgtgccc	9420
ctctaactgc	tctgcctgcc	ggcgtcgcgt	cgcgggcttt	ggagcggcgc	agggcaacga	9480
gccgatcgcg	gatcgtggaa	acgataggcc	tatgccatgc	gggtcaaggc	gacttccggc	9540
aagctatacg	cgccctagga	gtgcggttgg	aacgttggcc	cagccagata	ctcccgatca	9600
cgagcaggac	gccgatgatt	tgaagcgcac	tcagcgtctg	atccaagaac	aaccatccta	9660

-continued

gcaacacggc ggtccccggg ctgagaaaagc ccagtaagga aacaactgta ggttcgagtc 9720
gcgagatccc ccggaaccaa aggaagtagg ttaaaccgcg tccgatcagg ccgagccacg 9780
ccaggccgag aacattggtt cctgtaggca tcgggattgg cggatcaaac actaaagcta 9840
ctggaacgag cagaagtccct ccggccgcca gttgccaggc ggtaaagggtg agcagaggca 9900
cgggagggtt ccaactgctg gtcagcacgg ttccgaacgc catggaacc gcccccgcca 9960
ggcccgtgc gacgccgaca ggatctagcg ctgcgtttg tgtcaacacc aacagcgcca 10020
cgcccgcagt tccgcaata gcccccagga ccgccatcaa tcgtatcggg ctacctagca 10080
gagcggcaga gatgaacacg accatcagcg gctgcacagc gcctaccgtc gccgcgacct 10140
cgcccggcag gcggtagacc gaaataaaca acaagctcca gaatagcga atattaagt 10200
cgccgaggat gaagtgcgc atccaccaga ttcccgttg aatctgctgg acgatcatca 10260
cgagcaataa acccgccggc aacgcccga gcagcatacc ggcgaccct cggcctcgt 10320
gttcgggctc cacgaaaacg ccggacagat gcgccttg agcgtcctt ggccgctcct 10380
cctgtttgaa gaccgacagc ccaatgatct gcgcgtcgat gtaggcgccc aatgccacgg 10440
catctcga aacgctcagc aacgcctcca tgggctttt ctcctcgtgc tcgtaaacgg 10500
acccgaacat ctctggagct ttcttcaggc ccgacaatcg gatctcggc aaatcctgca 10560
cgtcggccgc tccaagccgt cgaatctgag ccttaatcac aattgtcaat ttaaatcctc 10620
tgtttatcgg cagttcgtag agcgcgccgt gcgtcccag cgatactgag cgaagcaagt 10680
gcgtcagca gtgcccgtt gttcctgaaa tgccagtaaa gcgctggctg ctgaaccccc 10740
agccggaact gacccccaaa ggcccctagc tttgcaatgc accaggatc cattgacca 10800
ggcgtgttcc accaggccgc tgccctgcaa ctcttcgag gcttcgcca cctgctcgg 10860
ccacttctc acgcggtggt aatccgatcc gcacatgagg cgaaggttt ccagcttgag 10920
cgggtacggc tcccgtgctg agctgaaata gtcgaacatc cgtcggccg tcggcgacag 10980
cttgcggtac ttctccata tgaatttctg gtagtgtcgc ccagcaaca gcaagcagat 11040
ttcctcgtc atcagacact ggcaacggga cgttttctt ccacggtcca ggacgaggaa 11100
gcggtgcagc agcagaccg attccagggt cccaacgagg tcggacgtga agccatcgc 11160
cgtgccctgt aggcgcgaca ggcattcctc ggccttcgt taataccggc cattgatcga 11220
ccagcccagc tcctggcaaa gctcgtagaa cgtgaagggt atcggctcgc cgaatagggt 11280
gcgcttcgag tactccaaca cctgtgcca caccagttcg tcctcgtcgg ccgcagctc 11340
gacgcccgtg tagtgatct tcacgtcctt gttgacgtg aaaatgaact tgttttgag 11400
cgctcgcgc gggattttct gttgctcgt ggtgaacagg gcagagcggg ccgtgctgtt 11460
tggcatcgtc cgcctcgtt ccggccacgg cgcaatatc aacaaggaaa gctgcatttc 11520
cttgatctgc tgcttcgtg gtttcagcaa cgcggcctgc ttggcctcgc tgaactgttt 11580
tgccaggtcc tcgcccggcg tttttcgtt cttggtcgtc atagttcctc gcgtgctgat 11640
ggtcatcgac ttgcgcaaac ctgcccctc ctggtcgtg cgcagcgaac gctccacggc 11700
ggcagatggc gcgggacggc cagggggagc cagttgcacg ctgtcgcgt cgaatctggc 11760
cgtagcttgc tggaccatcg agccgacgga ctggaagggt tcgcggggcg cacgcatgac 11820
ggtgcggctt gcgatggtt ccgcatcctc ggcggaaaac cccgcgtcga tcagttctt 11880
cctgatgccc ttccggtcaa acgtcogatt cattcaccct ccttgcggga ttgcccgcac 11940

-continued

tcacgccggg gcaatgtgcc cttattcctg atttgaccgg cctgggtcct tgggtgccag	12000
ataatccacc ttatcggcaa tgaagtcggt cccgtagacc gtctggccgt ccttctcgta	12060
cttggatatt cgaatcttgc cctgcacgaa taccagcgac cccttgccca aatacttgcc	12120
gtgggctctg gctgtagagc caaaacactt gatgcggaag aagtcgggtg gctcctgctt	12180
gtcgcgggtc gtggccgctc caacctttgc gatccgcaag cgcgcggtcg ccatcttcac	12240
gctggaacag tacgtcgagg cgggcatcat gaccgcgag caatacgagg tcattaaaag	12300
cgccgtgatt gatgatatag cggcccggtc gctcctgggt ctcgcgcacc gaaatgggtg	12360
acttcacccc gcgctctttg atcgtggcac cgatttcgc gatgctctcc ggggaaaagc	12420
cggggtgtgc ggccgtccgc ggctgatgag gatcttcgtc gatcaggctc aggtccagct	12480
cgatagggcc ggaaccgccc tgagacgccc caggagcgtc caggaggctc gacaggctgc	12540
cgatgctatc caaccccgag ccggacggct cgcgcgcgcc tgcggcttcc tgagcggccg	12600
cagcgggtgt tttcttgggt gtcttgggtt gagccgcagt cattgggaaa tctccatctt	12660
cgtgaacacg taatcagcca gggcgcgaac ctctttcgat gccttgccgc cggccgtttt	12720
cttgatcttc cagaccggca caccggatgc gagggcatcg gcgatgctgc tgcgcaggcc	12780
aacgggtggc ggaatcatca tcttggggta cgcggccagc agctcggctt ggtggcgcgc	12840
gtggcgcgga ttccgcgcat cgacctgtct gggcaccatg ccaaggaatt gcagcttggc	12900
gttctctctg cgcacgttgc caatggctct gaccatcttc ttgatgccct ggatgctgta	12960
cgctcaagc tcgatggggg acagcacata gtcggccgag aagaggcggg ccgccaggcc	13020
gacgccaagc gtcggggcgg tgcgatcag gcacacgtcg aagccttggg tcgccagggc	13080
cttgatgttc gccccgaaca gctcgcgggc gtcgtccagc gacagccgtt cggcgttcgc	13140
cagtaccggg ttggactcga tgaggcgcag cgcgcgcgcc tggccgtcgc cggctgcggg	13200
tgcggtttcg gtccagccgc cggcagggac agcgcgcaac agcttgcttg catgcaggcc	13260
ggtagcaaa tccttgagcg ttaggagcgc attgccctgg gggccagggt ccatcacggc	13320
aacccgcaag ccgcgctcga aaaagtcgaa ggcaagatgc acaagggtcg aagtcttggc	13380
gacgcccctt ttctggttgg ccgtgaccaa agttttcatc gtttggtttc ctgttttttc	13440
ttggcgtccg cttcccactt ccggaogatg tacgctgat gttccggcag aacgcgctt	13500
acccgcgcgt acccctcggg caagttcttg tcctcgaacg cggccacac gcgatgcacc	13560
gcttgcgaca ctgcgcccct ggtcagtccc agcgcgcttg cgaacgtcgc ctgtggcttc	13620
ccatcgacta agacgccccg cgctatctcg atggtctgct gccccacttc cagcccctgg	13680
atgcctcctt ggaactggct ttcggtaagc cgtttcttca tggataacac ccataatttg	13740
ctccgcgctt tggttgaaca tagcgggtgac agccgccagc acatgagaga agtttagcta	13800
aacatttctc gcacgtcaac acctttagcc gctaaaactc gtccttggcg taacaaaaca	13860
aaagcccgga aaccgggctt tcgctctctg ccgcttatgg ctctgcaacc ggctccatca	13920
ccaacaggtc gcgcacgcgc ttcaactcggg tgcggatcga cactgccagc ccaacaaagc	13980
cggttgcccg cgcgcgagg atcgcgcccga tgatgccggc cacaccggcc atcgcgcc	14040
aggtcgcgcg cttccggctc cattcctgct ggtactgctt cgcaatgctg gacctcggct	14100
caccataggc tgaccgctcg atggcgtatg ccgcttctcc ccttggcgta aaaccagcg	14160
ccgcaggcgg cattgccatg ctgccgcgg ctttcccagc cacgacgcgc gcaccaggct	14220

-continued

```

tgcggtccag accttcggcc acggcgagct gcgcaaggac ataatcagcc gccgacttgg 14280
ctccacgcgc ctcgatcagc tcttgcactc gcgcgaaatc cttggcctcc acggccgcca 14340
tgaatcgcgc acgcgcgcaa ggctccgcag ggcccg                               14375

```

```

<210> SEQ ID NO 21
<211> LENGTH: 14384
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        plasmid

```

```

<400> SEQUENCE: 21

```

```

ccaccaggc cgcgccctc actgcccgc acctggtcgc tgaatgtcga tgccagcacc      60
tgcggcacgt caatgcttcc gggcgtcgcg ctcgggctga tcgccatcc cgttactgcc    120
ccgatccccg caatggcaag gactgccagc gctgccattt ttgggggtgag gccgttcgcg    180
gccgaggggg gcagcccctg ggggatggg aggcccgctg tagcgggccc ggagggttcg    240
agaagggggg gcacccccct tcggcgtcgc cggtcacgcg cacagggcgc agccctggtt    300
aaaaacaagg ttataaata ttggtttaa agcaggttaa aagacagggt agcggtgggc    360
gaaaaacggg cgaaaccct tgcaaatgct ggattttctg cctgtggaca gccctcaaa    420
tgtcaatagc tcgccccctc atctgtcagc actctgccc tcaagtgtca aggatcgcgc    480
ccctcatctg tcagtagtcg cccccctcaa gtgtcaatac cgcagggcac ttatccccag    540
gcttgtccac atcatctgtg gaaaactcgc gtaaaatcag gcgttttcgc cgatttgcca    600
ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgccgc    660
gccgggtgag tcggccccct aagtgtcaac gtccgccctc catctgtcag tgagggccaa    720
gttttccgcg aggtatccac aacgcggcgc gcccggtgt ctcgcacacg gcttcgacgg    780
cgtttctggc gcgtttgcag ggccatagac ggccgcagc ccagcggcga gggcaaccag    840
cccgtgagc gtcggaagg cgtcttccg cttcctcgt cactgactcg ctgcgctcgg    900
tcgttcggct gcggcgagc gtatcagctc actcaaaggc ggaataacgg ttatccacag    960
aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc   1020
gtaaaaaggc cgcgttgctg gcgttttcc ataggctcgc cccccctgac gagcatcaca   1080
aaaaatcgac ctcaagtca aggtggcgaa acccgacagg actataaaga taccaggcgt   1140
ttccccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc   1200
tgtccgctt tctcccttcg ggaagcgtgg cgccattcgc cattcaggct gcgcaactgt   1260
tggaagggc gatcgggtgc gccctcttcg ctattacgcc agctggcga aggggatgt   1320
gctgcaaggc gattaagttg ggtaacgcca gggtttccc agtcacgacg ttgtaaacg   1380
acggccagt aattcggccg cgggtggcc gaggtgctgg gcaagcccta cctccaggcc   1440
cccacgggg tcgagagcac gaccgcttc ctgcgccgc tggcgagat tctgggctc   1500
gatccggagc cttcatcga gcgcgagaag cactcgacgc tgaagcccgt gtgggatctg   1560
tgcgagagt tcacgcagga cttcttcgg acggccaatt tcggaatcgt ggcgaccgaa   1620
acttatgcaa gaggcacccg aaactatctc gaaggcagc tcgggctgcc ctgoccttc   1680
gccgtggccc gcaagagggg ctcaagacc gacaacgaag cggtgccgcg actgatccgc   1740

```

-continued

cagcaccgtc	cgctcgtgct	catggggctcg	atcaacgaga	agatttacct	tgcggaactg	1800
aaagccggtc	acggcccgca	accctctttc	atcgtgcct	ctttcccggg	tgcggcgatc	1860
cggcgcgcta	ccggaaccgc	cgttatggga	tatgcagggtg	ctacgtgggt	actgcaggaa	1920
gtttgcaacg	ccctgttcga	cgccctgttc	cacattctgc	ccctcgggac	ggagatggac	1980
agcgcgcg	ccacaccgac	gacactgcgc	cgcgacttcc	cgtgggatgc	cgatgcgcaa	2040
gcggccctg	accgcatcgt	agaggagcat	ccggttctca	cccggatcag	cgccgcgct	2100
gccttgccg	acgccccga	gaaggctgcc	ctcgtatccg	gtgccgagag	ggctgtgaga	2160
gagactgtcg	aagccctg	tgggcggggc	ttcggcgaga	ggaagggaga	gaaccaatga	2220
gcgatcatgc	cgtcaacacg	ccggtccatg	ccgccagggc	ccacgggac	cgagcaccac	2280
gtgcccagtt	ctacgtctac	ttcgcgttca	ttctgctggg	cgccttccc	gtggccttcg	2340
tgagctggat	cgtctcgacg	atcccacc	gcaggcttcc	caagcgcggc	cccttcgct	2400
ccgctgggt	cgatgccaa	gcgatcacgc	cgtgatctt	ccgcgcctga	ccgcaggtea	2460
ggttgcgaca	cgccattcgt	cgtctcccca	agggcgggcg	gattaatcgg	gagggcatgg	2520
tgcttaccg	taaccacgc	caccagcatg	tggcgactag	ttccatatga	tagatctcac	2580
caccaccacc	accaccacca	ccaccaccac	caccactaat	aggccggccc	tccgtcggg	2640
gcggcacc	cgcccgcatc	gattccaag	ttcagccatt	gagacggctc	cgcttcggc	2700
gcaagcgcg	gttgggccga	ctgcaagc	agagggaagc	atggcactgc	tcagcttcga	2760
gcgaaaatat	cgcgtgccg	ggggcacgct	ggtcggcgga	aacctgttcg	acttctgggt	2820
cgccctttc	tatgtcggct	tcttcgggg	tgcgacgtt	ttcttcggg	ccctgggtat	2880
cattctgatt	gcctggagtg	ccgtactcca	gggtacctgg	aacccccaac	tcactctctg	2940
ctaccgccg	gcccctgaat	atggcctggg	aggtgcaccc	ctcgcaaaag	gcgggctgtg	3000
gcagatcatc	acgatctg	ccactggtgc	cttcgtcagc	tgggcgctgc	gcgaagtcca	3060
aatctgccgt	aagctgggca	tcgggtacca	catcccgttc	gccttcggct	tcgccatcct	3120
ggcctacctg	acgctggtgc	tgcttcgccc	ggtgatgatg	ggcgcctggg	gctatgcctt	3180
cccctacggg	atctggacgc	acctcagctg	ggtgtcgaac	acgggctaca	cctacggcaa	3240
cttccactac	aaccctgccc	acatgatcgc	catctcgttc	ttcttcacga	acgcgctggc	3300
tctggcgctg	cacggcgccc	ttgtgctctc	cgcggccaac	cccgagaag	gcaaggaaat	3360
gcggacgccg	gatcacgagg	atagttctt	cgcgatctg	gtcggctact	cgatcgggac	3420
gctcggcatc	caccgcctcg	gcctgctgct	ctcgtgagc	gccgtcttct	tcagcgcct	3480
ctgcatgate	attaccggca	ccatctggtt	cgatcagtgg	gtcgcactgg	ggcaatgggt	3540
ggtgaagctg	ccgtggtggg	cgaacatccc	gggaggcac	aatggctgag	tatcagaaca	3600
tcttctccca	ggtccaggtc	cgcggaccgg	ccgacctggg	gatgaccgaa	gacgtcaacc	3660
tggccaaccg	ttcggggcgc	ggtcccttct	cgacctgct	cggtggttc	ggcaacgccc	3720
agctcggccc	gatctatctc	ggctcgtcgc	gogtctctgc	cctcttctcg	ggcctgatgt	3780
ggttctctac	catcgggata	tggttctggt	atcaggcggg	ctggaaccgg	gccgtcttcc	3840
tgcgcgacct	gttcttcttc	tcgctcagac	cgcgggcacc	cgaatacgg	ctgtccttcg	3900
cggctccgct	gaaggaagc	gggctgtggc	tgatcgcgtc	gttcttcatg	ttcgtcggg	3960
tctggtcctg	gtggggccc	acctatctcc	cgcctcaggc	gctgggcatg	ggcaagcaca	4020

-continued

ccgctgggc gttcctctcg gccatctggc tgtggatggt gctgggcttc atccgtccga	4080
tcctcatggg gtccctggcg gaagcgggtc cctacggcat cttctcgcac ctcgactgga	4140
cgaacaactt ctcgctcgtc cacggcaacc tgttctacaa ccccttcac ggtctctcga	4200
tcgcttctct ctacgggtcg gccctgctct tcgcatgca cggtgcgacc atcctcggg	4260
tctcccgtt cggcggcgag cgcgagctgg agcagatcgc cgaccgagg acggcagcgg	4320
agcgggccgc cctctctcgg cgtctggacca tgggtttcaa cgccacgatg gaaggcatcc	4380
accgctgggc catctggatg gcggtcctcg tgaccctcac cggcggcacc gggatcctgc	4440
tctcgggac ggtcgtggac aactggtagc tctggggcca gaaccacggc atggcgccgc	4500
tgaactgagg agcagtcaca atggctgaca agaccatctt caacgatcac ctcaacacca	4560
atccgaagac caacctcgc ctctgggtcg ctttccagat gatgaagggt gcgggctggg	4620
ctggcggcgt gttcttcggg acgctccttc tcatcgggtt cttccgggtg gtcgggcgga	4680
tgcttcgat ccaggagaac caggctccgg cgccgaacat caccggcgtc ctggagaccg	4740
ggatcgagct gatcaagcat ctgctctgag acaagtctcg gggcaggcgg gcgagagcc	4800
gcccgtcct ccaagtccgg gccatatacgc cggcccgggt ccggggcgac accacagccc	4860
ggttcccttc ctgttggcga cagggacctg gtgcccgtgt gaagaccgca cggcacctt	4920
ttgacattca cgggaggtc tgatgacaa tcccaccccg cgaccgaaa ccccgctttt	4980
ggatcgcgtc tgctgcccg ccgacatgaa ggcgtgagt gacgccgaac tggagcggct	5040
ggccgacgaa gtgcttccg aggtcagtga tagggtagt ttcttatttt aggcagtta	5100
tatgaaatta agacatgca atgtcacagt ggatattgaa ctggtctcga aagctcaata	5160
tccccaaaag cacaagcaca aacttcgaca tcatgcagaa gcgtttcccg aaccgctct	5220
tcgacgtggg catcgcggag cagcatgccg tgacctcgc ggcggcctc gccggggccg	5280
ggatgaagcc cttctgcgag atctattcct cgttctcga acggggttac gaccagatcg	5340
cccatgacgt ggcgctgcag aacctcccg tccgcttcgt gatcgaccgg gcggggctcg	5400
tgggggccga tggcgcgacc catgccccg ccttcgacgt tggcttcac acttcgctgc	5460
ccaacatgac cgtgatggc gcggcgcag aggcgagct catccacatg atcggcaccg	5520
ccgtggcctt cgacgagggc cccatcgcct tccgcttccc gcggggcgag ggggtggcg	5580
tcgagatgcc cgagcgggg acggtgctgg agcccggcg gggccgctg gtgcgcgag	5640
ggacggatgt cgcgacctc tccttcggcg cgcattcga cgaggcctt caggcggcga	5700
aacttctcga ggccgaggg gtgagcgtga ccgtggccga cgcccgttc tcgccccgc	5760
tcgacacggg gctcatcgac cagctcgtgc gccatcacgc ggcgctgta acggtggagc	5820
agggggccat gggcggcttc ggcgcocatg tcatgacta tctcgccaat tccggcggct	5880
tcgacgggg cctcgcgctc cgggtcatga cgtgcccga ccgcttcac gagcagcga	5940
gccccagga catgtatgcc gatgccccg tgcgggccga ggatatcaag cttggcgtaa	6000
tcatggtcat agctgtttcc tgtgtgaaat tgttatccgc tcacaattcc acacaacata	6060
cgagccggaa gcataaagt taaagcctg ggtgcctaata gtaggacta actcacatta	6120
attgcgttc gctcactgcc cgtttccag tcgggaaacc tgtcgtgcca gctgcattaa	6180
tgaatcgcc aacgcgggg gagaggggt ttgctgattg ggcgctcgg cttgccttgc	6240
tcgtcgtga tgtacttcac cagctcccg aagtcgctct tcttgatgga gcgcatggg	6300

-continued

acgtgcttgg caatcacgcg ccccccccg cggtttttagc ggctaaaaaa gtcatggctc	6360
tgccctcggg cggaccacgc ccatcatgac cttgccaagc tcgtcctgct tctcttcgat	6420
cttcgccagc agggcgagga tcgtggcatc accgaaccgc gccgtgcgcg ggtcgtcggg	6480
gagccagagt ttcagcagcg cgcaccggcg gcccaggctg ccattgatgc gggccagctc	6540
gcggaagtgc tcatagtcca cgacgcccg gatttttagc ccctggccga cggccagcag	6600
gtaggccgac aggctcatgc cggccgccg cgccttttcc tcaatcgctc ttcgttcgctc	6660
tggaaggcag tacaccttga taggtgggct gcccttctcg gttggcttgg tttcatcagc	6720
catccgcttg ccctcatctg ttacgccggc ggtagccggc cagcctcgca gagcaggatt	6780
cccgttgagc accgccaggt gcgaataagg gacagtgaag aaggaacacc cgctcgcggg	6840
tgggcctact tcacctatcc tgcccggctg acgccgttgg atacaccaag gaaagtctac	6900
acgaaccctt tggcaaaatc ctgtatatcg tgcgaaaaag gatggatata ccgaaaaaat	6960
cgctataatg accccgaagc agggttatgc agcggaaaag cggcacgctt cccgaaggga	7020
gaaaggcggg caggtatccg gtaagcggca gggtcggaac aggagagcgc acgagggagc	7080
ttccaggggg aaacgcctgg tatctttata gtcctgtcgg gtttcgccac ctctgacttg	7140
agcgtcgatt tttgtgatgc tcgtcagggg ggcggagcct atggaaaaac gccagcaacg	7200
cggccttttt acggttctcg gccttttctg ggccttttgc tcacatgttc tttcctgcgt	7260
tatcccctga ttctgtggat aaccgtatta ccgcctttga gtgagctgat accgctcgcc	7320
gcagccgaac gaccgagcgc agcgagttag tgagcgagga agcggaaagc cgcagaagg	7380
ccgccagaga ggcgagcgc ggcctgagc cttggacgct agggcagggc atgaaaaagc	7440
ccgtagcggg ctgctacggg cgtctgacgc ggtggaaaag gggaggggat gttgtctaca	7500
tggtctgctg tagtgtagtg ggttgcgctc cggcagcggg cctgatcaat cgtcacctt	7560
tctcgttctc tcaacgttcc tgacaacgag cctccttttc gccaatccat cgacaatcac	7620
cgcgagtccc tgctcgaacg ctgctgccgg accggcttcc tcgaaggcgt ctatcgcggc	7680
ccgcaacagc ggcgagagcg gagcctgttc aacggtgccg ccgcgctcgc cggcatcgct	7740
gtcgcggccc tgctcctcaa gcacggcccc aacagtgaag tagctgattg tcatcagcgc	7800
attgacggcg tccccggccg aaaaaccgcg ctccgagagg aagcgaagct gcgcgtcggc	7860
cgtttccatc tcggtgcgca ccggtcgcgt gccggcatgg atgcgcgcgc catcgcggta	7920
ggcgagcagc gcctgcctga agctcggggc attcccgatc agaaatgagc gccagtcgctc	7980
gtcggctctc ggcaccgaat gcgtatgatt ctccgccagc atggcttcgg ccagtgcgctc	8040
gagcagcggc cgcttgttcc tgaagtgccg gtaaagcggc ggtgctgtaa cccccaacg	8100
ttccgccagt ttgcgtgtcg tcagaocgct tacgccgacc tcgttcaaca ggtccagggc	8160
ggcacggatc actgtattcg gctgcaactt tgtcatgctt gacactttat cactgataaa	8220
cataatatgt ccaccaactt atcagtgata aagaatccgc gcgttcaatc ggaccagcgg	8280
aggctggtcc ggaggccaga catgaaaccc aacatacccc tgatcgtaat tctgagcact	8340
gtcgcgctcg acgctgtcgc catcggcctg attatgccgg tgctgccggg cctcctgcgc	8400
gatctggttc actcgaacga cgtcaccgcc cactatggca tttctgctggc gctgtatgcg	8460
ttggtgcaat ttgcctgcgc acctgtgctg ggcgcgctgt cggatcgttt cgggcggcgg	8520
ccaatcttgc tcgtctcgtg ggcggcgccc actgtcgact acgccatcat ggcgacagcg	8580

-continued

cctttccttt gggttcteta tategggcgg atcgtggccg gcatcacccg ggcgactggg	8640
gcggtagccg gcgcttatat tgccgatac actgatggcg atgagcgcgc gcggcacttc	8700
ggcttcatga gcgcctgttt cgggttcggg atggtcgcgg gacctgtgct cggtgggctg	8760
atgggcggtt tctccccca cgtccggttc ttgcgcggc cagccttga cggcctcaat	8820
ttcctgacgg gctgtttcct ttgcccggag tcgcacaaag gcgaacgcc gccgttacgc	8880
cgggaggctc tcaaccgcct cgttcggttc cggtgggccc gggcatgac cgtcgtcgc	8940
gccctgatgg cggctctctt catcatgcaa cttgtcggac aggtgccggc cgcgctttgg	9000
gtcattttcg gcgaggatcg ctttactg gacgcgacca cgtcggcat ttcgcttgcc	9060
gcatttggca ttctgcatc actcgcacc gcaatgatca ccggccctgt agccgcccg	9120
ctcggcga ggcgggact catgctcgg atgattgcc acggcacagg ctacatcctg	9180
cttgccttcg cgacacggg atggatggc ttcccgatca tggctctgct tgcttcgggt	9240
ggcatcggaa tgcccgcgt gcaagcaatg ttgtccaggc aggtggatga ggaacgtcag	9300
ggcgactgc aaggctcact ggcggcgtc accagcctga cctcgtcgt cggacccctc	9360
ctcttcacgg cgatctatgc ggcttctata acaacgtgga acgggtggc atggattgca	9420
ggcgtgccc tctactgtct ctgcctgccc gcgctgctc gcgggctttg gagcggcgca	9480
gggcaacgag ccgatcgtg atcgtgaaa cgataggcct atgccatgcg ggtcaaggcg	9540
acttccggca agctatacgc gccctaggag tgcggttgg acgttgccc agccagatac	9600
tccgatcac gagcaggac ccgatgattt gaagcgcact cagcgtctga tccaagaaca	9660
accatcctag caacacggcg gtccccggc tgagaaagcc cagtaaggaa acaactgtag	9720
gttcgagtcg cgagatcccc cggaacaaa ggaagtaggt taaaccgct ccgatcaggc	9780
cgagccacgc caggccgaga acattggttc ctgtaggcat cgggattggc ggatcaaaca	9840
ctaaagctac tggaacgagc agaagtcctc cggccgccag ttgccaggcg gtaaaggatg	9900
gcagaggcac gggagggttc cacttgccgg tcagcacggt tccgaacgcc atggaaaccg	9960
ccccgccag gcccgctgac acgcccagc gatctagcgc tgcgtttggt gtcaacacca	10020
acagcggcc gcccgagtt ccgcaaatag ccccaggac cggcatcaat cgtatcgggc	10080
tacctagcag agcggcagag atgaacacga ccatcagcgg ctgcacagcg cctaccgtcg	10140
cccgacccc gcccggcagg cggtagaccg aaataaaca caagctccag aatagcga	10200
tattaagtgc gccgaggatg aagatgcgca tccaccagat tcccgttgg atctgtcgg	10260
cgatcatcac gagcaataa cccgcccgc acgcccgcag cagcataccg gcgacccctc	10320
ggcctcgtg ttccgggtcc acgaaaacgc cggacagatg cgccttgtga gcgtccttg	10380
ggcgtcctc ctggttgaag accgacagcc caatgatctc gccgtcgtg taggcgccg	10440
atgccacggc atctcgaac cgttcagcga acgctccat gggctttttc tcctcgtgct	10500
cgtaaacgga ccgcaacatc tctggagctt tcttcagggc cgacaatcgg atctcggga	10560
aatcctgcac gtcggccgct ccaagccgct gaatctgagc cttaatcaca attgtcaatt	10620
ttaatcctct gtttatcgc agttcgtaga gcgcgccgtg cgtcccagc gatactgagc	10680
gaagcaagtg cgtcagcag tgcccgttg ttcctgaaat gccagtaaag cgtggtgctg	10740
tgaaccccc gccggaactg accccacaag gccctagcgt ttgcaatgca ccaggctc	10800
attgacccag gcgtgttcca ccaggccgct gcctcgaac tcttcgagc cttgcggc	10860

-continued

ctgctcgcgc	cactttctca	cgcggtgga	atccgatccg	cacatgaggc	ggaaggtttc	10920
cagcttgagc	gggtacggct	cccggtgca	gctgaaatag	tcgaacatcc	gtcgggccgt	10980
cgcgacagc	ttgcggtact	tctcccatat	gaatttcgtg	tagtggtcgc	cagcaaacag	11040
cacgacgatt	tcctcgtcga	tcaggacctg	gcaacgggac	gttttcttgc	cacggtccag	11100
gacgcggaag	cggtgcagca	gcgacaccga	ttccaggtgc	ccaacgcggg	cggacgtgaa	11160
gcccacgcgc	gtcgcctgta	ggcgcgacag	gcattcctcg	gccttcgtgt	aataccggcc	11220
attgatcgac	cagcccaggt	cctggcaaa	ctcgtagaac	gtgaaggtga	tcggctcggc	11280
gataggggtg	cgcttcgcgt	actccaacac	ctgctgccac	accagttcgt	catcgtcggc	11340
ccgcagctcg	acgcccgtgt	aggtgatcct	cacgtccttg	ttgacgtgga	aaatgacctt	11400
gttttcgagc	gcctcgcgcg	ggattttctt	gttgccgctg	gtgaacaggg	cagagcgggc	11460
cgtgctgtht	ggcctcgcct	gcctcgtgct	cggccacggc	gcaatatcga	acaagaaaag	11520
ctgcatttcc	ttgatctgct	gcttcgtgtg	tttcagcaac	gcggcctgct	tggcctcget	11580
gacctgtttt	gccaggtcct	cgccggcggt	ttttcgttc	ttggtcgtca	tagttcctcg	11640
cgtgctgatg	gtcctcgcct	tcgcaaaacc	tgccgcctcc	tgctcagagc	gacgcgaaag	11700
ctccacggcg	gccgatggcg	cgggcagggc	agggggagcc	agttgcacgc	tgtcgcgctc	11760
gatcttggtc	gtagctgctg	ggaccatcga	gccgacggac	tggaaagttt	cgcgggggcg	11820
acgcatgacg	gtgcgcttg	cgatggttcc	ggcctcctcg	gcgaaaacc	ccgcgctgat	11880
cagttcttgc	ctgtatgcct	tccggtcaaa	cgcccgattc	attcaccttc	cttgccgggat	11940
tgccccgact	cacgcccggg	caatgtgccc	ttattcctga	tttgaccgcg	ctggtgcctt	12000
ggtgtccaga	taatccacct	tatcggcaat	gaagtcggtc	ccgtagaccg	tctggccgctc	12060
cttctcgtac	ttggtattcc	gaatcttgcc	ctgcacgaat	accagcgacc	ccttgcccaa	12120
atacttgccg	tggccctcgg	cctgagagcc	aaaacacttg	atgcggaaga	agtcggtgcg	12180
ctcctgcttg	tcgcccgtcg	tggccgcgcc	aacctttgcg	atccgcaagc	gcgcccgtcgc	12240
catcttcacg	ctggaacagt	acgtcagggc	gggcatcatg	accgcgagc	aataccaggt	12300
cattaaaagc	gccgtgattg	atgatatagc	ggcccggctg	ctcctggttc	tcgocgaccg	12360
aaatgggtga	cttacccccg	cgctctttga	tcgtggcacc	gatttccgcg	atgctctccg	12420
gggaaaagcc	ggggttgctg	gccgtccgcg	gctgatgcgg	atcttcgctg	atcaggttcca	12480
ggtccagctc	gatagggccg	gaaccgccct	gagacgccgc	aggagcgtcc	aggaggctcg	12540
acaggtcgcg	gatgctatcc	aaccccagcg	cggacggctg	cgccgcgect	gcggcttccct	12600
gagcggcccg	agcgggtgtt	ttcttggtgg	tcttggtctg	agccgcagtc	attgggaaat	12660
ctccatcttc	gtgaacacgt	aatcagccag	ggcgcgaacc	tctttcgtatg	ccttgccgcg	12720
ggccgttttc	ttgatcttcc	agaccggcac	accggtgctg	agggcatcgg	cgatgctgct	12780
gcgcaggcca	acggtggccg	gaatcatcat	cttgggttac	gcggccagca	gctcggcttg	12840
gtggcgcgcg	tggcgcggat	tccgcgcatc	gaccttgctg	ggcaccatgc	caaggaattg	12900
cagcttgggc	ttcttctggc	gcacgttcgc	aatggtcgtg	accatcttct	tgatgccctg	12960
gatgctgtac	gcctcaagct	cgatggggga	cagcacatag	tcggcccgca	agagggcggc	13020
cgccaggccc	acgccaaggg	tcggggccgt	gtcgatcagg	cacacgtcga	agccttggtt	13080
cgccagggcc	ttgatgttgc	ccccgaacag	ctcgcgggcg	tcgtccagcg	acagccgttc	13140

-continued

```

ggcgttcgcc agtaccgggt tggactcgat gagggcgagg cgcgcggcct ggcgcgcgcc 13200
ggctgcgggt gcggtttcgg tccagccgcc ggcagggaca gcgccgaaca gcttgcttgc 13260
atgcaggccg gtagcaaaagt ccttgagcgt gtaggacgca ttgccctggg ggtccaggtc 13320
gatcacggca acccgcaagc cgcgctcgaa aaagtcgaag gcaagatgca caagggtcga 13380
agtcttgccg acgccgcctt tctggttggc cgtgaccaa gtttcatcg tttggtttcc 13440
tgttttttct tggcgtccgc tcccacttc cggacgatgt acgcctgatg ttccggcaga 13500
accgccgta cccgcgcgta cccctcgggc aagttcttgt cctcgaacgc ggcccacacg 13560
cgatgcaccg cttgcgacac tgcgcccctg gtcagtcca gcgacgttgc gaacgtcgcc 13620
tgtggcttcc catcgactaa gacgccccgc gctatctcga tggctctgctg ccccacttcc 13680
agcccctgga tcgcctcctg gaactggctt tcggtaagcc gtttcttcat ggataacacc 13740
cataatttgc tccgcgcctt ggttgaacat agcggtgaca gccgccagca catgagagaa 13800
gtttagctaa acatttctcg cacgtcaaca ccttagccg ctaaaactcg tccttggcgt 13860
aacaaaaaaa aagcccggaa accgggcttt cgtctcttgc cgcttatggc tctgcaccgc 13920
gctccatcac caacaggtcg cgcacgcgct tcaactcgtt gcggtcgcac actgccagcc 13980
caacaaagcc ggttgccgcc gccgccagga tcgcgcgat gatgccggcc acaccggcca 14040
tcgccacca ggtgcgcgcc ttcgggttcc attcctgctg gtaactgttc gcaatgctgg 14100
acctcggctc accataggtt gaccgctcga tggcgtatgc cgcttctccc cttggcgtaa 14160
aaccagcgc cgcaggcgcg attgcatgc tgcccgcgcg tttcccgacc acgacgcgcg 14220
caccaggctt gcggtccaga ccttcggcca cggcgagctg cgcaaggaca taatcagccg 14280
ccgacttggc tccacgcgcc tcgatcagct cttgcaactc cgcgaaatcc ttggcctcca 14340
cggccgccat gaatcgcgca cgcggcgaag gctccgcagg gccg 14384

```

<210> SEQ ID NO 22

<211> LENGTH: 14530

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic plasmid

<400> SEQUENCE: 22

```

ccaccagcgc cgcgcgcctc actgcccgcg acctggtcgc tgaatgtcga tgccagcacc 60
tgcggcacgt caatgcttcc gggcgtcgcg ctcgggctga tcgccatcc cgttactgcc 120
ccgatcccg caatggcaag gactgccagc gctgccattt ttggggtag gccgttcgcg 180
gccgaggggc gcagcccctg ggggatggg aggccgcgct tagcgggccc ggagggttcg 240
agaagggggg gcacccccct tcggcgtcgc cggtcacgcg cacagggcgc agccctggtt 300
aaaaacaagg ttataaata ttggtttaa agcaggttaa aagacaggtt agcggtggcc 360
gaaaaacggg cggaaacctc tgcaaatgct ggattttctg cctgtggaca gcccccaaaa 420
tgtcaatagg tgcgcccctc atctgtcagc actctgcccc tcaagtgtca aggatcgcgc 480
ccctcatctg tcagtatcgc cgcgccctca gtgtcaatac cgcagggcac ttatccccag 540
gcttgtccac atcatctgtg ggaaactcgc gtaaaatcag gcgttttcgc cgatttgcga 600
ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgcgcg 660
gccgggtgag tcggcccctc aagtgtcaac gtccgccctc catctgtcag tgagggccaa 720

```

-continued

gttttcgcg aggtatccac aacgccggcg gcccggtgt ctgcacacg gcttcgacg	780
cgtttctggc gcgtttgcaq ggccatagac ggccgccagc ccagcggcga gggcaaccag	840
cccggtgagc gtcgaaagg cgctcttccg cttcctcgct cactgactcg ctgcgctcgg	900
tcgttcggct gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg ttatccacag	960
aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc	1020
gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg cccccctgac gagcatcaca	1080
aaaaatcgac ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt	1140
ttccccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc	1200
tgccgcctt tctcccttcg ggaagcgtgg cgccattcgc cattcaggct gcgcaactgt	1260
tgggaagggc gatcggtgcg gccctcttcg ctattacgcc agctggcga aggggatgt	1320
gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg	1380
acggccagtg aattcggccg cgggctggcc gagtgctgg gcaagcccta cctccaggcc	1440
cccatcgggg tcgagagcac gaccgccttc ctgcgcgcc tgggcgagat tctgggcctc	1500
gatccggagc cttcatcga gcgcgagaag cactcgacgc tgaagcccg gtgggatctg	1560
tggcggagtg tcacgcagga cttcttcggg acggccaatt tcggaatcgt ggcgaccgaa	1620
acttatgcaa gaggcacccg aaactatctc gaaggcgatc tcgggctgcc ctgcgccttc	1680
gccgtggccc gcaagagggg ctcgaagacc gacaacgaag cggtgccgg actgatccgc	1740
cagcacccgc cgctcgtgct catggggctg atcaacgaga agatttacct tgcggaactg	1800
aaagccggtc acggcccga accctcttcc atcgtgcct ctttccggg tgcggcgatc	1860
cggcgcgcta ccggaaccgc cgttatggga tatgcagtg ctacgtggtt actgcaggaa	1920
gtttgcaacg ccctgttcga cgccctgttc cacattctgc ccctcgggac ggagatggac	1980
agcgcgcccg ccacaccgac gacactgcgc cgcgacttcc cgtgggatgc cgatgcgcaa	2040
gcggccctgg accgcatcgt agaggagcat cggttctca cccggatcag cgcgcgcgt	2100
gccttgccgc acgcccccga gaaggctgcc ctcgatgccg gtgccgagag ggtcgtgaga	2160
gagactgtcg aagccctgcg tgggcggggc ttcggcgaga ggaagggaga gaaccaatga	2220
gcgatcatgc cgtcaacacg ccggtccatg ccgccagggc ccacgggcac cgagcaccac	2280
gtgccgagtt ctacgtctac ttcgcgctca ttctgctggg cgccttcccg gtggccttcg	2340
tgagctggat cgtctcgacg atccgccacc gcaggcttcc caagcgcggc cccttcgctg	2400
ccgctgggtt cgatgccaaq gcgatcacgc cgtgatctt cgcgcctga ccgcaggtca	2460
ggttgcgaca cgccattcgt cgtctcccga agggcgggcg gattaatcgg gagggcatgg	2520
tgcttaccg taaccaccgc caccagcatg tggcgactag gggaggatag taatgctcgt	2580
caagacgcac atcacgaaga ttggggctac gctgttcgcc gtggctctgt tctacgggtt	2640
catctacatg ctttccaact cgctctttgc gacgcggccg gccaccgccg ttgcggtcgg	2700
tgcggacggg aaagccctga ctagtccat atgatagatc tcaccaccac caccaccacc	2760
actaataggc cggccctccg tcgcggggcg caccacgcc cgcacgatt ccaaggttca	2820
gccattgaga cggctccgct tcgcgcgcaa gcgcgggttg ggccgactgc aagcggagag	2880
ggaagcatgg cactgctcag cttcgagcga aaatatcgcg tgccgggggg cagcgtggtc	2940
ggcggaaacc tgctcgactt ctgggtcggc ctttctatg tcggcttctt cggggttcgg	3000

-continued

acgtttttct	tcgcggccct	gggtatcatt	ctgattgcct	ggagtgccgt	actccagggt	3060
acctggaacc	cccaactcat	ctctgtctac	cgcgggccc	ttgaatatgg	cctgggaggt	3120
gcaccctcgc	caaaaggcgg	gctgtggcag	atcatcacga	tctgcgccac	tggtgccttc	3180
gtcagctggg	cgctgcgcga	agtcgaaatc	tgccgtaagc	tgggcatcgg	gtaccacatc	3240
ccgttcgcct	tcgctgtcgc	catcctggcc	tacctgacgc	tgggtctggt	ccgcccggtg	3300
atgatgggcg	cctggggcta	tgcctcccc	tacgggatct	ggacgcacct	cgactgggtg	3360
tcgaacacgg	gtacaccta	cggcaacttc	cactacaacc	ctgccacat	gatcgccatc	3420
tcgttcttct	tcacgaacgc	gctggctctg	gcgctgcacg	gcgcccttgt	gctctccgcg	3480
gccaaacccg	agaagggcaa	ggaaatgcgg	acgccggtc	acgaggatac	gttcttccgc	3540
gatctggteg	gtactcgcg	cgggacgctc	ggcatccacc	gcctcggcct	gctgctctcg	3600
ctgagcgccg	tcttctcag	cgccctctgc	atgatcatta	ccggcaccat	ctggttcgat	3660
cagtgggtcg	actggtggca	atggtgggtg	aagctgccgt	ggtgggcgaa	catcccggga	3720
ggcatcaatg	gctgagtatc	agaacatctt	ctcccaggtc	caggctccgcg	gaccggccga	3780
cctgggggatg	accgaagacg	tcaacctggc	caaccgttcg	ggcgtcggtc	ccttctcgac	3840
cctgctcggc	tggttcggca	acgcccagct	cggcccgatc	tatctcggct	cgctcggcgt	3900
cctgtccctc	ttctcgggcc	tgatgtggtt	cttcaccatc	gggatctggt	tctggtatca	3960
ggcgggtcgg	aacccggccg	tcttctcgcg	cgacctgttc	ttcttctcgc	tcgagccgcc	4020
ggcaccgcaa	tacggtctgt	ccttcgcggc	tccgctgaag	gaaggcgggc	tgtggctgat	4080
cgcgtcgttc	ttcatgttcg	tcgcggtctg	gtcctggtgg	ggccgcacct	atctccgcgc	4140
tcaggcctcg	ggcatgggca	agcacaccgc	ctgggcgttc	ctctcggcca	tctggctgtg	4200
gatggtgctg	ggcttcatcc	gtccgatcct	catggggtcc	tggctcggaa	cggttcccta	4260
cggcatcttc	tcgcacctcg	actggacgaa	caacttctcg	ctcgtccacg	gcaacctggt	4320
ctacaacccc	ttccacggtc	tctcgatcgc	cttctcttac	gggtcggccc	tgctcttcgc	4380
gatgcacggt	gcgaccatcc	tcgcggtctc	cgccttcggc	ggcgagcgcg	agctggagca	4440
gatcgcgcac	cgcgggacgg	cagcggagcg	ggccgccttc	ttctggcgcg	ggaccatggg	4500
tttcaacgcc	acgatggaag	gcacccaccg	ctgggccatc	tggatggcgg	tcctcgtgac	4560
cctcaccggc	ggcatcggga	tcctgctctc	gggcacggtc	gtggacaact	ggtacgtctg	4620
gggccagaac	cacggcatgg	cgccgctgaa	ctgaggagcg	atcacaatgg	ctgacaagac	4680
catcttcaac	gatcacctca	acaccaatcc	gaagaccaac	cttcgcctct	gggtcgcctt	4740
ccagatgatg	aaggggtcgg	gctgggctgg	cggcgtgttc	ttcgggacgc	tccttctcat	4800
cgggttcttc	cgggtggtcg	ggcggatgct	tccgatccag	gagaaccagg	ctcggcgccc	4860
gaacatcacc	ggcgtctcgg	agaccgggat	cgagctgac	aagcatctcg	tctgagacaa	4920
gtctcggggc	agggcggcgc	gagggcggcc	gctcctccaa	gtccgggcca	tatcgcgggc	4980
ccgggtccgg	ggcgacacca	cagcccgggt	cccttctctg	tggcgacagg	gacctggtgc	5040
cgtgtggaag	accgcacggc	acccttttga	cattcacggg	aggctctgat	gaccaatccc	5100
accccgcgac	ccgaaacccc	gcttttggat	cgcgtctgct	gcccggccga	catgaaggcg	5160
ctgagtgcgc	ccgaaactgga	gcggctggcc	gacgaagtgc	gttccgaggt	cagtgatagg	5220
ggtagtttct	tattttaggc	agtttatatg	aaattaagac	atgcagatgt	cacagtggat	5280

-continued

attgaactgg tctcgaagc tcaatatccc ccaaagcaca agcacaaact tcgacatcat 5340
gcagaagcgt ttcccgaacc gcgtcttcga cgtgggcacg gccgagcagc atgccgtgac 5400
cttcgcggcc ggccctcggc gggccgggat gaagcccttc tgcgcatctt attcctcgtt 5460
cctgcaacgg ggttacgacc agatcgccca tgacgtggcg ctgcagaacc ttcccgtccg 5520
cttcgtgacg gaccggggcg ggctcgtggg ggccgatggc gcgaccatg cgggggcctt 5580
cgacgttggc ttcatcactt cgctgcccga catgaccgtg atggccggcg ccgacgagc 5640
cgagctcacc cacatgatcg ccaccgccgt ggccctcgac gagggcccga tcgccttccg 5700
cttcccggcg ggcgaggggg tgggcgtcga gatgcccagc gcggggacgg tgctggagcc 5760
cgcccgggcg cgcgtggtgc gcgaaggac ggatgtcgcg atcctctcct tcggcgcgca 5820
tgtgcacgag gccttgacgg cggcgaact tctcggagcc gagggggtga gcgtgaccgt 5880
ggcgacgccc cgcttctcgc gcccgcctga cacggggctc atcgaccagc tcgtgcgcca 5940
tcacgcggcg ctggtaacgg tggagcaggg ggccatggc ggcttcggcg cccatgtcat 6000
gcactatctc gccaattccg gcggcttcga cgggggcctc gcgctccggg tcatgacgct 6060
gcccagccgc ttcatcgagc aggcgagccc cgaggacatg tatgccgatg cggggctgcg 6120
ggccgaggat atcaagcttg gcgtaatcat ggtcatagct gtttcctgtg tgaattggtt 6180
atccgctcac aattccacac aacatagcag ccggaagcat aaagtgtaaa gcctgggggtg 6240
cctaagtagt gagctaaact acattaattg cgttgcgctc actgcccgct ttccagtcgg 6300
gaaacctgct gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttg 6360
gtattgggcg ctccgctctg ccttgctcgt cgggtgatga cttaccagc tccgcgaagt 6420
cgctcttctt gatggagcgc atggggacgt gcttgcaat cacgcgcacc ccccgccgt 6480
tttagcggct aaaaaagtca tggctctgcc ctcgggcgga ccacgcccac catgacctt 6540
ccaagctcgt cctgcttctc ttcgatcttc gccagcaggg cgaggatcgt ggcatcaccg 6600
aacgcgcgcc tcgcccgggtc gtcggtgagc cagagtttca gcaggccgcc caggcggccc 6660
aggctgccat tgatcgggc cagctcggcg acgtgctcat agtccacgac gcccgtgatt 6720
ttgtagccct ggccgacgcg cagcaggtag gccgacagc tcatgccggc cgcgcgcgcc 6780
tttctctcaa tcgctcttcg ttcgtctgga aggcagtaca ccttgatagg tgggctgcc 6840
ttcctggttg gcttggttcc atcagccatc cgcttgccct catctgttac gccggcggt 6900
gccggccagc ctgcgagagc aggattcccg ttgagcaccg ccagggtcga ataagggaca 6960
gtgaagaagg aacacccgct cgcgggtggg cctacttcac ctatcctgcc cggctgacgc 7020
cgttgatac accaaggaaa gtctacacga accctttggc aaaatcctgt atatcgtgcg 7080
aaaaaggatg gatataccga aaaaatcgct ataatgaccc cgaagcaggg ttatgcagcg 7140
gaaaagcggc acgcttcccg aaggagaaaa ggcggacagc tatccggtaa gcggcaggg 7200
cggaacagga gagcgcacga gggagcttcc agggggaac gcctggtatc tttatagttc 7260
tgtcgggttt cgccacctct gacttgagcg tcgatttttg tgatgctcgt cagggggcg 7320
gagcctatgg aaaaacgcca gcaacggcg ctttttacgg ttcctggcct tttgctggcc 7380
ttttgctcac atgttcttcc ctgcgttacc ccctgattct gtggataacc gtattaccgc 7440
ctttgagtga gctgataccg ctgcgcgag ccgaacgacc gagcgcagcg agtcagtgag 7500
cgaggaagcg gaagagcgcc agaaggccgc cagagaggcc gagcgcggcc gtgaggcttg 7560

-continued

gacgctaggg cagggcatga aaaagcccgt agcgggctgc tacgggctgc tgacgcgggtg 7620
gaaaggggga ggggatgttg tctacatggc tctgctgtag tgagtggggt gcgctccggc 7680
agcggctctg atcaatcgtc accctttctc ggtccttcaa cgttcctgac aacgagcctc 7740
cttttcgcca atccatcgac aatcaccgcg agtcctctgct cgaacgctgc gtccggaccg 7800
gcttcgtcga agggctctat cgcggcccgc aacagcggcg agagcggagc ctgttcaacg 7860
gtgccgccgc gctcgcggcg atcgtgtcgc ccggcctgct cctcaagcac ggccccaaaca 7920
gtgaagtagc tgattgtcat cagcgcattg acggcgtccc cggccgaaaa acccgctctg 7980
cagaggaagc gaagctgcgc gtcggccgtt tccatctgcg gtgcgcccgg tcgctgcccg 8040
gcatgtagc gcgcgccatc gcggtaggcg agcagcgcct gcctgaagct gcgggcattc 8100
ccgatcagaa atgagcgcca gtcgtcgtcg gctctcggca ccgaatcgct atgattctcc 8160
gccagcatgg cttcggccag tgcgtcagac agcgcgccgt tgttcctgaa gtgccagtaa 8220
agcgcgggct gctgaacccc caaccgttcc gccagtttgc gtgtcgtcag accgtctacg 8280
ccgacctcgt tcaacaggtc cagggcggca cggatcactg tattcggctg caactttgtc 8340
atgcttgaca ctttatcact gataaacata atatgtccac caacttatca gtgataaaga 8400
atccgcgcgt tcaatcggac cagcggaggc tggccggag gccagacatg aaaccaaca 8460
taccctgat cgtaattctg agcactgtcg cgctcgaagc tgcggcattc ggctgatta 8520
tgccggtgct gccgggctc ctgcgcgac tggttcactc gaacgacgtc accgccact 8580
atggcattct gctggcgtg tatgcgttg tgcaatttgc ctgcgcacct gtgctgggcg 8640
cgctgtcggg tcgtttcggg cggcgccaa tcttgctcgt ctgctggcc ggcggcactg 8700
tcgactacgc catcatggcg acagcgcctt tcctttgggt tctctatata gggcggatcg 8760
tgcccgcat caccggggcg actggggcg tagccggcgc ttatattgcc gatatactg 8820
atggcgatga gcgcgcggc cacttcggct tcatgagcgc ctgtttcggg ttcgggatgg 8880
tcgcccggacc tgtgctcgtt gggctgatgg gcggtttctc ccccccacgt ccgttcttcg 8940
ccgcggcagc cttgaacggc ctcaatttcc tgacgggctg tttccttttg ccggagtcgc 9000
acaaaggcga acgcccggcg ttacgcccgg aggctctcaa ccgctcgtc tcgttccggt 9060
gggcccgggg catgaccgct gtcgcgccc tgatggcggg cttcttcac atgcaacttg 9120
tcggacaggt gccggccgcg ctttggttca ttttcggcga ggatcgctt cactgggacg 9180
cgaccacgat cggcatttcc cttgcgcgat ttggcattct gcattcactc gcccaggcaa 9240
tgatcaccgg ccctgtagcc gcccggctcg gcgaaaggcg ggactcatg ctcggaatga 9300
ttgccgacgg cacaggctac atcctgcttg ccttcgcgac acggggatgg atggcgttcc 9360
cgatcatggt cctgctgtct tcgggtggca tcggaatgcc ggcgctgcaa gcaatgttgt 9420
ccaggcaggt ggatgaggaa cgtcaggggc agctgcaagg ctcactggcg gcgctacca 9480
gcctgacctc gatcgtcggg cccctcctct tcacggcgat ctatgcccgt tctataaaca 9540
cgtggaacgg gtgggcatgg attgcaggcg ctgcctctc cttgctctgc ctgccggcg 9600
tcgctcggcg gctttggagc ggcgcagggc aacgagccga tcgctgatcg tggaaacgat 9660
aggcctatgc catgcgggtc aaggcgactt ccggcaagct atacgcgccc taggagtgcg 9720
gttggaacgt tggcccagcc agatactccc gatcacgagc aggacgcca tgatttgaag 9780
cgcactcagc gctctatcca agaacaacca tcctagcaac acggcgggtcc ccgggctgag 9840

-continued

aaagcccagt aaggaaacaa ctgtaggttc gagtcgcgag atcccccgga accaaaggaa 9900
gtaggttaaa cccgctccga tcaggccgag ccacgccagg ccgagaacat tggttcctgt 9960
aggcatcggg attggcggat caaacactaa agctactgga acgagcagaa gtcctccggc 10020
cgccagttgc caggcggtaa aggtgagcag aggcacggga ggttgccact tgcgggtcag 10080
cacggttccg aacgccatgy aaaccgcccc cgccaggccc gctgcgacgc cgacaggatc 10140
tagcgtcgcg ttgtgtgca acaccaacag cgccacgccc gcagttccgc aaatagcccc 10200
caggaccgcc atcaatcgta tcgggctacc tagcagagcg gcagagatga acacgacccat 10260
cagcggctgc acagcgccta ccgtcgcgcg gaccccgccc ggcaggcggg agaccgaaat 10320
aaacaacaag ctccagaata gcgaaatatt aagtgcgccg aggatgaaga tgcgcatcca 10380
ccagattccc gttggaatct gtcggacgat catcacgagc aataaacccg ccggcaacgc 10440
ccgcagcagc ataccggcga cccctcggcc tcgctgttcg ggtccacga aaacgcccga 10500
cagatgcgcc ttgtgagcgt ccttggggcc gtcctcctgt ttgaagaccg acagcccaat 10560
gatctcgcgc tcgatgtagg cgccgaatgc cacggcatct cgcaaccggt cagcgaacgc 10620
ctccatgggc tttttctcct cgtgctcgta aacggaccgg aacatctctg gagctttctt 10680
cagggccgac aatcggatct cgcggaatc ctgcacgtcg gccgctcaa gccgtcgaat 10740
ctgagcctta atcacaattg tcaattttaa tcctctgttt atcggcagtt cgtagagcgc 10800
gccgtgcgct ccgagcgata ctgagcgaag caagtgcgct gagcagtgcc cgcttgttcc 10860
tgaaatgcca gtaaagcgct ggctgctgaa cccccagccg gaactgaccc cacaaggccc 10920
tagcgtttgc aatgcaccag gtcacattg acccaggcgt gttccaccag gccgctgcct 10980
cgcaactctt cgcagcttc gccgacctgc tcgcccact tcttcacgcg ggtggaatcc 11040
gatccgcaca tgaggcggaa ggtttccagc ttgagcgggt acggctcccg gtgcgagctg 11100
aaatagtcga acatccgtcg gccgctcggc gacagcttgc ggtacttctc ccatatgaat 11160
ttcgtgtagt ggtcgcagc aaacagcagc acgatttcct cgtcgatcag gacctggcaa 11220
cgggacgttt tcttgccagc gtccaggagc cggaagcggg gcagcagcga caccgattcc 11280
aggtgccca cgcggtcgga cgtgaagccc atcgcctcgc cctgtaggcg cgacaggcat 11340
tcctcggcct tcgtgtaata ccggcattg atcgaccagc ccaggctctg gcaaagctcg 11400
tagaacgtga aggtgatcgg ctccgagata ggggtgcgct tcgctgactc caacacctgc 11460
tgccacacca gttcgtcatc gtcggcccgc agctcgacgc cgggtgaggt gatcttccag 11520
tccttgttga cgtggaaaat gacctgtttt tgcagcgcct cgcgaggat tttcttgtt 11580
cgctggtgta acagggcaga gcgggcccgt tcgtttgca tcgctcgcac cgtgtccggc 11640
cacggcgcga tatcgaacaa ggaaagctgc atttccttga tctgctgctt cgtgtgttcc 11700
agcaacgcgg cctgcttggc ctccgtgacc tgttttgcca ggtcctcgc gccggttttt 11760
cgcttcttgg tcgtcatagt tcctcgcgct tcgatggta tcgacttcgc caaacctgcc 11820
gcctcctggt cgagacgagc cgaacgctcc acggcggccg atggcgcggg cagggcaggg 11880
ggagccagtt gcacgctgct gcctcogac ttggccgtag cttgctggac catcgagccg 11940
acggactgga aggtttcgcg gggcgcacgc atgacggtgc ggcttgatg ggtttcggca 12000
tcctcggcgg aaaaccccgc gtcgatcagt tcttgcctgt atgccttcgc gtaaacgctc 12060
cgattcattc acctccttg cgggattgcc ccgactcacg ccggggcaat gtgoccttat 12120

-continued

tcctgatttg acccgcttg tgccttggtg tccagataat ccaccttata ggcaatgaag 12180
tcggtcccgt agaccgtctg gccgtccttc tegtacttgg tattccgaat cttgccctgc 12240
acgaatacca gcgaccctt gcccaaatac ttgccgtggg cctcggcctg agagccaaaa 12300
cacttgatgc ggaagaagtc ggtgcgctcc tgcttgctgc cggtcgtggc cgcgccaacc 12360
tttgcgatcc gcaagcgcgc ggtcgccatc ttcacgctgg aacagtacgt cgaggcgggc 12420
atcatgacct gcgagcaata cgaggctatt aaaagcgcgc tgattgatga tatagcggcc 12480
cggctgctcc tggttctcgc gcaccgaaat gggtgacttc acccgcgct ctttgatcgt 12540
ggcaccgatt tccgcgatgc tctccgggga aaagccgggg ttgtcggccg tccgcggctg 12600
atgcgatct tcgctgatca ggtccaggtc cagctcgata gggccggaac cgcctgaga 12660
cgccgcagga gcgtccagga ggctcgacag gtcgccgatg ctatccaacc ccaggccgga 12720
cggctgcgcc gcgctgcggc cttcctgagc ggccgcagcg gtgttttct tggtggtctt 12780
ggcttgagcc gcagtcattg gaaaatctcc atcttcgtga acacgtaac agccaggcg 12840
cgaaacctct tcgatgcctt gcgcgcgcc gttttcttga tcttccagac cggcacaccg 12900
gatgagagg catcgcgat gctgctgcgc aggccaacgg tggccggaat catcatctt 12960
gggtacgcgg ccagcagctc ggcttggtgg cgcgcgtggc gcgattccg cgcacgacc 13020
ttgctggca ccatgccaag gaattgcagc ttggcgttct tctggcgcac gttcgcaatg 13080
gtcgtgacca tcttcttgat gccctggatg ctgtacgcct caagctcgat gggggacagc 13140
acatagtcgg ccgcgaagag ggcggccgcc aggccgacgc caagggtcgg ggccgtgtcg 13200
atcaggcaca cgtcgaagcc ttggttcgcc agggccttga tgttcgccc gaacagctcg 13260
cggcgctcgt ccagcagac ccggtcggcg ttcgccagta ccgggttggc ctcgatgagg 13320
gcgagggcgc cggcctggcc gtcgcggtt gcgggtgcgg tttcggcca gccgccgca 13380
gggacagcgc cgaacagctt gcttgcatgc aggccgtag caaagtctt gagcgttag 13440
gacgattgc cctgggggtc caggtogatc acggcaacc gcaagccgc ctcgaaaaag 13500
tcgaaggcaa gatgcacaag ggtcgaagtc ttgccgacgc cgcctttctg gttggccgtg 13560
accaaagttt tcatcgtttg gtttctgtt ttttcttggc gtcgcttcc cacttccgga 13620
cgatgtacgc ctgatgttcc ggcagaaccg cgttaccgc cgcgtacccc tcgggcaagt 13680
tcttgcctc gaacgcggcc cacacgcgat gcaccgcttgc gcacactgcg cccctggtca 13740
gtccagcga cgttcgaac gtcgcctgtg gcttccatc gactaagacg ccccgctca 13800
tctogattg ctgctgccc acttccagcc cctggatcgc ctcctggaac tggcttctcg 13860
taagccgtt cttcatggat aacaccata atttgctccg cgccttggtt gaacatagcg 13920
gtgacagccg ccagcacatg agagaagttt agctaaacat ttctcgcacg tcaacacctt 13980
tagccgctaa aactcgtcct tggcgtaaca aaacaaaagc ccgaaaaccg ggcttctcgc 14040
tcttgcgct tatggtctg caccggctc catcaccaac aggtcgcgca cgcgcttcc 14100
tcggttgcgg atcgacactg ccagcccaac aaagccggtt gccgccgccc ccaggatcgc 14160
gccgatgat ccggccacac cggccatcgc ccaccagtc gccgccttcc ggttccattc 14220
ctgctggtac tgcttcgcaa tgctggacct cggctcacca taggctgacc gctcogattg 14280
gtatgcgct tctccccttgc gctgaaaacc cagcgcgca ggcggcattg ccatgctgccc 14340
cgccgcttcc ccgaccagca cgcgcgacc aggttgcgg tccagacctt cggccacggc 14400

-continued

```

gagctgcgca aggacataat cagccgccga cttggctcca cgcgcctcga tcagctcttg 14460
cactcgcgcg aaatccttgg cctccacggc cgccatgaat cgcgcacgcg gcaaggctc 14520
cgcagggccg                                     14530

```

```

<210> SEQ ID NO 23
<211> LENGTH: 14448
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        plasmid

```

```

<400> SEQUENCE: 23

```

```

ccaccaggc cgcgccctc actgcccgc acctggtcgc tgaatgtcga tgccagcacc 60
tgccgcacgt caatgcttcc gggcgtcgcg ctcgggtga tcgccatcc cgttactgcc 120
ccgatccccg caatggcaag gactgccagc gctgccattt ttggggtgag gccgttcgcg 180
gccgaggggg gcagcccctg ggggatggg aggcccgctg tagcgggccc ggagggttcg 240
agaagggggg gcacccccct tcggcgtcgc cggtcacgcg cacagggcgc agccctggtt 300
aaaaacaagg ttataaata ttggtttaa agcaggttaa aagacagggt agcggtgggc 360
gaaaaacggg cggaaacct tgcaaatgct ggattttctg cctgtggaca gccctcaaa 420
tgtcaatagc tcgcccctc atctgtcagc actctgccc tcaagtgtca aggatcgcgc 480
ccctcatctg tcagtagtcg cccccctcaa gtgtcaatac cgcagggcac ttatccccag 540
gcttgtccac atcatctgtg gaaaactcgc gtaaaatcag gcgttttcgc cgatttgcca 600
ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgccgc 660
gccgggtgag tcggcccctc aagtgtcaac gtccgccctc catctgtcag tgagggccaa 720
gttttccgcg aggtatccac aacgcggcg gcccggtgt ctcgcacacg gcttcgacgg 780
cgtttctggc gcgtttgcag ggccatagac ggccgccagc ccagcggcga gggcaaccag 840
ccccgtgagc gtcggaagg cgtcttccg cttcctcgtc cactgactcg ctgcgctcgg 900
tcgttcggct gcggcgagcg gtatcagctc actcaaaggc ggaataacgg ttatccacag 960
aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc 1020
gtaaaaaggc cgcgttgctg gcgttttcc ataggctcgg cccccctgac gagcatcaca 1080
aaaaatcgac ctcaagtca aggtggcgaa acccgacagg actataaaga taccaggcgt 1140
ttccccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc 1200
tgtccgctt tctcccttcg ggaagcgtgg cgccattcgc cattcaggct gcgcaactgt 1260
tggaagggc gatcgggtcg gccctcttcg ctattacgcc agctggcgaa aggggatgt 1320
gctgcaaggc gattaagttg ggtaacgcca gggtttccc agtcacgacg ttgtaaacg 1380
acggccagtg aattcggccg cgggtggcc gaggtgctgg gcaagcccta cctccaggcc 1440
cccacgggg tcgagagcac gaccgcttc ctgcgccgcc tggcgagat tctgggctc 1500
gatccggagc cttcatcga gcgcgagaag cactcgacgc tgaagcccgt gtgggatctg 1560
tgccggagtg tcacgcagga cttcttcggg acggccaatt tcggaatcgt ggcgaccgaa 1620
acttatgcaa gaggcacccg aaactatctc gaaggcgatc tcgggctgcc ctgoccttc 1680
gccgtggccc gcaagagggg ctcgaagacc gacaacgaag cggtgccgcg actgatccgc 1740

```

-continued

cagcaccgtc	cgctcgtgct	catggggtcg	atcaacgaga	agatttacct	tgcggaactg	1800
aaagccggtc	acggcccgca	accctctttc	atcgctgcct	ctttcccggg	tgcggcgatc	1860
cggcgcgcta	ccggaacgcc	cgttatggga	tatgcagggtg	ctacgtgggtt	actgcaggaa	1920
gtttgcaacg	ccctgttcga	cgccctgttc	cacattctgc	ccctcgggac	ggagatggac	1980
agcgcgcg	ccacaccgac	gacactgcgc	cgcgacttcc	cgtgggatgc	cgatgcgcaa	2040
gcggccctgg	accgcatcgt	agaggagcat	ccggttctca	cccggatcag	cgcgcgcgct	2100
gccttgccgcg	acgccccgca	gaaggctgcc	ctcgatgccg	gtgccgagag	ggtcgtgaga	2160
gagactgtcg	aagccctgcg	tgggcggggc	ttcggcgaga	ggaagggaga	gaaccaatga	2220
gcgatcatgc	cgtcaacacg	ccggtccatg	ccgccagggc	ccacggggcac	cgagcaccac	2280
gtgccgagtt	ctacgtctac	ttcgcgctca	ttctgctggg	cgccttcccg	gtggccttcg	2340
tgagctggat	cgtctcgacg	atccgccacc	gcaggcttcc	caagcgcggc	cccttcgcgt	2400
ccgcctgggt	cgatgccaa	gcgatcacgc	cgctgatttt	ccgcgcctga	ccgcaggtca	2460
ggttgcgaca	cgccattcgt	cgtctcccca	agggggggcg	gattaatcgg	gagggcatgg	2520
tgccctaccg	taaccacacg	caccagcatg	tggcgactag	gggaggatag	taatgaagtt	2580
ccaagtcaag	gccctcgccg	ccatcgccgc	attcgcggcg	ctgccggcgc	tcgcgcagac	2640
tagttccata	tgatagatct	caccaccacc	accaccacca	ctaataggcc	ggccctccgt	2700
cgcgggcggc	acccacgccc	gcacgcatc	caaggttcag	ccattgagac	ggctccgctt	2760
cgcgcgcaag	cgcgggttgg	gccgactgca	agcggagagg	gaagcatggc	actgctcagc	2820
ttcgcgcgaa	aatatcgcgt	gccggggggc	acgctggtcg	gcgaaaacct	gttcgacttc	2880
tgggtcggcc	ctttctatgt	cggcttcttc	ggggttcgca	cgttttctt	cgcggccctg	2940
ggtatcattc	tgattgcctg	gagtgccgta	ctccagggta	cctggaacct	ccaactcatc	3000
tctgtctacc	cgcggccctc	tgaatatggc	ctgggaggtg	caccctcgcg	aaaaggcggg	3060
ctgtggcaga	tcacacgat	ctgcgccact	ggtgccttcg	tcagctgggc	gctgcgcgaa	3120
gtcgaaatct	gccgtaagct	gggcatcggg	taccacatcc	cgttcgcctt	cgcgttcgcc	3180
atcctggcct	acctgacgct	ggtgctgttc	cgcccggtga	tgatgggcgc	ctggggctat	3240
gccttcccct	acgggatctg	gacgcacctc	gactgggtgt	cgaacacggg	ctacacctac	3300
ggcaacttcc	actacaacct	tgcccacatg	atcgccatct	cgttcttctt	cacgaacgcg	3360
ctggctctgg	cgctgcacgg	cgccttgtg	ctctccgcgg	ccaacctcga	gaagggcaag	3420
gaaatgcgga	cgcggatca	cgaggatacg	ttcttccgcg	atctggtcgg	ctactcgatc	3480
gggacgctcg	gcatccaccg	cctcggcctg	ctgctctcgc	tgagcgcgct	cttcttcagc	3540
gccctctgca	tgatcattac	cggcaaccatc	tggttcgcac	agtgggtcga	ctggtggcaa	3600
tgggtgggtga	agctgcgctg	gtgggcgaac	atcccgggag	gcatcaatgg	ctgagtatca	3660
gaacatcttc	tcccaggctc	aggtcgcgcg	accggccgac	ctggggatga	ccgaagacgt	3720
caacctggcc	aaccgttcgg	gcgtcggctc	cttctcgacc	ctgctcggct	ggttcggcaa	3780
cgccacgctc	ggcccgatct	atctcggctc	gctcggcgtc	ctgtccctct	tctcggccct	3840
gatgtgggtc	ttaccatcgc	ggatctgggt	ctggtatcag	gcgggctgga	accgcggcgt	3900
cttctcgcgc	gacctgttct	tcttctcgtc	cgagccgccc	gcacctgaat	acggtctgtc	3960
cttcgcggct	ccgctgaagg	aaggcgggct	gtggctgac	gcgtcgttct	tcagtctcgt	4020

-continued

cgcggtctgg	tcttgggtgg	gccgcaccta	tctccgcgct	caggcgctgg	gcatgggcaa	4080
gcacaccgcc	tggcggttcc	tctcggccat	ctggctgtgg	atggtgctgg	gcttcatccg	4140
tccgatcctc	atggggctct	ggtcggaaag	ggttccctac	ggcatcttct	cgcacctcga	4200
ctggagcaac	aacttctcgc	tcgtccacgg	caacctgttc	tacaaccctc	tccacgggtc	4260
ctcgatcgcc	ttcctctacg	ggtcggccct	gctcttcgcg	atgcacggtg	cgaccatcct	4320
cgcggtctcc	cgcttcggcg	gcgagcgcga	gctggagcag	atcgccgacc	gcgggacggc	4380
agcggagcgg	gccgccctct	tctggcgctg	gaccatgggt	ttcaacgcca	cgatggaagg	4440
catccaccgc	tgggccatct	ggatggcggg	cctcgtgacc	ctcaccggcg	gcatcgggat	4500
cctgctctcg	ggcacggctg	tggacaactg	gtacgtctgg	ggccagaacc	acggcatggc	4560
gccgctgaac	tgaggagcga	tcacaatggc	tgacaagacc	atcttcaacg	atcacctcaa	4620
caccaatccg	aagaccaacc	ttcgctctg	ggtcgccttc	cagatgatga	agggcgcggg	4680
ctgggctggc	ggcgtgttct	tcgggacgct	ccttctcatc	gggttcttcc	gggtggtcgg	4740
gcggtatgct	ccgatccagg	agaaccaggc	tccggcgccg	aacatcaccg	gcgctctgga	4800
gaccgggatc	gagctgatca	agcatctcgt	ctgagacaag	tctcggggca	ggcgggcgcg	4860
aggccgcccc	ctcctccaag	tccgggccat	atcgccggcc	cgggtccggg	gcgacaccac	4920
agcccggttc	ccttctctgt	ggcgacaggg	acctggtgcc	gtgtggaaga	ccgcacggca	4980
cccttttgac	attcacggga	ggctctgatg	accaatccca	ccccgcgacc	cgaaaccccg	5040
cttttgatc	gcgtctgctg	cccggccgac	atgaaggcgc	tgagtgcgcg	cgaactggag	5100
cggctggccg	acgaagtgcg	ttccgaggtc	agtgataggg	gtagtcttct	attttaggca	5160
gtttatatga	aattaagaca	tgacgatgct	acagtggata	ttgaactggg	ctcgaaagct	5220
caatatcccc	caaagcacia	gcacaaactt	cgacatcatg	cagaagcggt	tcccgaaaccg	5280
cgtcttcgac	gtgggcatcg	ccgagcagca	tgccgtgacc	ttcggggccg	gcctcgcccg	5340
ggccgggatg	aagcccttct	gcgcgatcta	ttcctcgttc	ctgcaacggg	gttacgacca	5400
gatcgcccat	gacgtggcgc	tgacgaacct	tcccgctcgc	ttcgtgatcg	accgggcccgg	5460
gctcgtgggg	gccgatggcg	cgaccatgct	gggggccttc	gacgttggct	tcataccttc	5520
gctgcccaac	atgacctgta	tggccgcggc	cgacgaggcc	gagctcatcc	acatgatcgc	5580
caccgcgctg	gccttcgacg	agggccccat	cgccttcgcg	ttcccgcggg	gcgagggggg	5640
ggcgctcgag	atgcccgagc	gcgggacggg	gctggagccc	ggccggggcc	gcgtggtcgc	5700
cgaagggacg	gatgtcgcga	tcctctcctt	cggcgcgcat	ctgcacgagg	ccttgacggc	5760
ggcgaaactt	ctcgaggccg	aggggggtgag	cgtgaccgtg	gccgacgccc	gcttctcgcg	5820
cccgtctgac	acggggctca	tcgaccagct	cgtgcgccat	cacgcggcgc	tggtaacggt	5880
ggagcagggg	gccatggggc	gcttcggcgc	ccatgtcatg	cactatctcg	ccaattcccg	5940
cggcttcgac	gggggcctcg	cgctccgggt	catgacgctg	cccgaccgct	tcacgagca	6000
ggcgagcccc	gaggacatgt	atgccgatgc	gggctgcgg	gccgaggata	tcaagcttgg	6060
cgtaatcatg	gtcatagctg	tttctgtgtg	gaaattgtta	tccgctcaca	attccacaca	6120
acatacgagc	cggaagcata	aagtgtaaag	cctgggggtgc	ctaagtgtg	agctaactca	6180
cattaattgc	gttgcgctca	ctgccogctt	tccagtcggg	aaacctgtcg	tgccagctgc	6240
attaatgaat	cggccaacgc	gcggggagag	gcggtttcgc	tattgggcgc	tcggtcttgc	6300

-continued

cttgctcgtc ggtgatgtac ttcaccagct ccgcgaagtc gctcttcttg atggagcgca	6360
tggggacgtg cttggcaatc acgcgcaccc cccggccggt ttagcggeta aaaaagtcac	6420
ggctctgccc tggggcgac cacgcccac atgacctgc caagctcgtc ctgcttctct	6480
tcgatcttcg ccagcagggc gaggatcgtg gcacaccga accgcgccgt gcgcgggtcg	6540
tcggtgagcc agagtttcag caggccgccc aggcggccca ggtcgccatt gatgcgggcc	6600
agctcgcgga cgtgctcata gtccacgacg cccgtgattt ttagcctctg gccgacggcc	6660
agcaggtagc ccgacaggct catgcgggcc gccgccgct tttcctcaat cgtctctctg	6720
tcgtctgga ggcagtacac cttgataggt gggctgccct tcctggttg cttggtttca	6780
tcagccatcc gcttgccctc atctgttacg ccggcggtag ccggccagcc tcgcagagca	6840
ggattcccg ttagcaccgc cagggtgcga taaggacag tgaagaagga acacccgctc	6900
gcgggtgggc ctacttcacc taccctgccc ggctgacgcc gttggataca ccaagaaaag	6960
tctacacgaa ccctttggca aaatcctgta tatcgtgcga aaaaggatgg atataccgaa	7020
aaaaatcgta taatgacccc gaagcagggt tatgcagcgg aaaagcgcga cgttcccga	7080
agggagaaa gcggacaggt atccggtaag cggcagggtc ggaacaggag agcgcacgag	7140
ggagcttcca ggggaaacg cctggtatct ttatagtcct gtcgggttcc gccacctctg	7200
actgagcgt cgatttttgt gatgctcgtc agggggggcg agcctatgga aaaacgccag	7260
caacgcggcc tttttacggt tcctggcctt ttgctggcct tttgctcaca tgttctttcc	7320
tgcgttatcc cctgattctg tggataaccg tattaccgcc tttgagttag ctgataccgc	7380
tcgccgcagc cgaacgaccg agcgcagcga gtcagttagc gaggaagcgg aagagcgcga	7440
gaaggccgcc agagaggccg agcgcggccg tgaggccttg acgctagggc agggcatgaa	7500
aaagcccgtg gcgggctgct acgggcgtct gacgcggtgg aaagggggag gggatgtgt	7560
ctacatggct ctgctgtagt gagggtgttg cgctccggca gcggtcctga tcaatcgtca	7620
ccctttctcg gtccctcaac gttcctgaca acgagcctcc ttttcgcaa tccatcgaca	7680
atcacccgca gtccctgctc gaacgctcgc tccggaccgg cttcgtcga ggcgtctatc	7740
gcggcccgca acagcggcga gagcggagcc tgttcaacgg tgcgcccgcg ctgcgccgca	7800
tcgctgtcgc cggcctgctc ctcaagcacg gcccacaacg tgaagtagct gattgtcatc	7860
agcgcattga cggcgtcccc ggccgaaaaa cccgcctcgc agaggaagcg aagctgcgcg	7920
tcggccggtt ccattctcgg tgcgcccgtt cgcgtgccgg catggatgcg cgcgccatcg	7980
cggtagcgca gcagcgcctg cctgaagctg cgggcattcc cgatcagaaa tgagcgcag	8040
tcgtcgtcgc ctctcggcac cgaatgcgta tgattctcgg ccagcatggc ttcggccagt	8100
gcgtcagca gcgcccgctt gttcctgaag tgccagtaaa gcgccgctg ctgaaccccc	8160
aaccgttccg ccagtttgcg tgcgtcaga ccgtctacgc cgacctggt caacaggtcc	8220
agggcgac ggatcactgt attcggctgc aactttgtca tgcttgacac tttatcactg	8280
ataaacataa tatgtccacc aacttatcag tgataaagaa tccgcgcgtt caatcggacc	8340
agcggaggct ggtccggagc ccagacatga aaccaacat acccctgatc gtaattctga	8400
gcactgtcgc gctcagcgt gtcggcatcg gcctgattat gccggtgctg ccgggcctcc	8460
tgcgcgatct ggttcaactc aacgaogtca ccgcccacta tggcattctg ctggcgtgt	8520
atgctgtggt gcaatttgcg tgcgcacctg tgcgtggcgc gctgctggat cgtttcgggc	8580

-continued

ggcgccaat cttgctcgtc tcgctggccg gcgccactgt cgactacgcc atcatggcga	8640
cagcgccttt cctttgggtt ctctatatcg ggcggatcgt ggccggcatc accggggcga	8700
ctggggcggt agccggcgct tatattgcg atatactga tggcgatgag cgcgcgcggc	8760
acttcggctt catgagcgcc tgtttcgggt tcgggatggt cgcgggacct gtgctcggtg	8820
ggctgatggg cggtttctcc ccccaocgct cgttcttcgc cgcggcagcc ttgaacggcc	8880
tcaatttctt gacgggctgt ttcttttgc cggagtcgca caaaggcga cgcggccct	8940
tacgccggga ggctctcaac ccgctcgtt cgttcgggtg ggcccggggc atgaccgtcg	9000
tcgccccct gatggcggtc ttcttcatca tgcaacttgt cggacaggtg ccggccgcgc	9060
ttgggtcat ttccggcgag gatcgttctt actgggacgc gaccacgac gccatttctc	9120
ttgccgcat tggcattctg cttcactcg cccaggaat gatcaccggc cctgtagccg	9180
cccggctcgg cgaaggcgg gcaactcag tcggaatgat tgccgacggc acaggctaca	9240
tcctgcttgc cttcgcgaca cggggatgga tggcgttccc gatcatggtc ctgcttgctt	9300
cgggtggcat cggaatgcc gcgctgcaag caatgttctc caggcaggtg gatgaggaac	9360
gtcaggggca gctgcaagg tcaactggcg cgctcaccag cctgacctc atcgtcggac	9420
ccctcctctt cacggcgatc tatgcggctt ctataacaac gtggaacggg tgggcatgga	9480
ttgcaggcgc tgcctctac ttgctctgcc tgcggcgct cgtcgcggg ctttgagcg	9540
gcgcagggca acgagccgat cgctgatcgt ggaacgata gccctatgcc atgcgggtca	9600
aggcgacttc cggcaagcta tacgcgccct aggagtgcgg ttggaacgtt ggcccagcca	9660
gatactccc atcacagca ggacgccgat gatttgaagc gcaactcagc tctgatccaa	9720
gaacaacct cctagcaaca cggcggctcc cgggctgaga aagccagta aggaaacaac	9780
tgtaggttcg agtcgcgaga tccccggaa ccaagggaag taggttaaac ccgctccgat	9840
caggccgagc cacgccaggc cgagaacatt ggttctgta ggcatcggga ttggcggatc	9900
aaacactaaa gctactggaa cgagcagaag tcctccggcc gccagttgcc aggcggtaaa	9960
ggtgagcaga ggcacgggag gttgccactt cggggtcagc acggttccga acgccatgga	10020
aaccgcccc gccaggcccc ctgacgccc gacagatct agcgtcgcgt ttggtgtcaa	10080
caccaacagc gccacgccc cagttccgca aatagcccc aggaccgcca tcaatcgtat	10140
cgggtacct agcagagcgg cagagatgaa cacgaccatc agcggctgca cagcgcctac	10200
cgctgccgag accccgccc gcaggcggta gaccgaaata aacaacaagc tccagaatag	10260
cgaatatta agtgcgccga ggatgaagat gcgcatccac cagattccc ttggaatctg	10320
tcggacgac atcacagca ataaacccgc cggcaacgcc cgcagcagca tacggcgac	10380
ccctcggcct cgctgttcgg gctccacgaa aacccggac agatgcgcct tgtgagcgtc	10440
cttggggccg tcctcctgtt tgaagaccga cagcccaatg atctcgcgt cgatgtaggc	10500
gccgaatgcc acggcatctc gcaaccgttc agcgaacgcc tccatgggct ttttctctc	10560
gtgctcgtaa acggaccgca acatctctgg agctttctt agggccgaca atcggatctc	10620
gcggaatcc tgcacgtcgg ccgctccaag ccgctcgaatc tgagccttaa tcacaattgt	10680
caattttaat cctctgttta tcggcagttc gttagagcgg ccgtgcgtcc cgagcgatac	10740
tgagcgaagc aagtgcgtcg agcagtgccc gottgttctt gaaatgccag taaagcgtg	10800
gctgctgaac cccagccgg aactgacccc acaaggccct agcgtttgca atgcaccagg	10860

-continued

tcacattga	cccagcgtg	ttccaccagg	ccgctgcctc	gcaactcttc	gcaggcttcg	10920
ccgacctgct	cgcgccactt	cttcacgcgg	gtggaatccg	atccgcacat	gaggcggaa	10980
gtttccagct	tgagcgggta	cggtccccg	tgcgagctga	aatagtcgaa	catccgtcgg	11040
gccgtcggcg	acagcttgcg	gtacttctcc	catatgaatt	tcgtgtagtg	gtcgccagca	11100
aacagcacga	cgatttcctc	gtcgatcagg	acctggcaac	gggacgtttt	cttgccacgg	11160
tccaggacgc	ggaagcggg	cagcagcgac	accgattcca	ggtgcccaac	gcggctcggc	11220
gtgaagccca	tcgccgtcgc	ctgtaggcgc	gacaggcatt	cctcggcctt	cgtgtaatac	11280
cggccattga	tcgaccagcc	caggtcctgg	caaagctcgt	agaacgtgaa	ggtgatcggc	11340
tcgcccagatg	gggtgcgctt	cgcgtactcc	aacacctgct	gccacaccag	ttcgatcatc	11400
tcggcccgca	gtcgcagcc	ggtgtaggtg	atcttcacgt	cctgttgac	gtggaaaatg	11460
acctgtttt	gcagcgcctc	gcgcgggatt	ttctgttg	gcgtggtgaa	cagggcagag	11520
cgggcctgt	cgtttgcat	cgctcgcac	gtgtccggcc	acggcgcaat	atcgaacaag	11580
gaaagctgca	ttccttgat	ctgctgcttc	gtgtgtttca	gcaacgcggc	ctgcttgcc	11640
tcgctgacct	gttttgccag	gtcctcgcg	gcggttttc	gcttcttgg	cgatcatag	11700
cctcgcgtgt	cgatggtcat	cgacttcgcc	aaacctgcg	cctcctgttc	gagacgacgc	11760
gaaacctcca	cgggcgccga	tgccgcgggc	gggcaggggg	agccagttgc	acgctgtcgc	11820
gctcgatcct	ggccgtagct	tgctggacca	tcgagccgac	ggactggaag	gtttcgcggg	11880
gcgcacgcat	gacggtcgg	cttgcatg	tttcggcatc	ctcggcgga	aaccccgct	11940
cgatcagttc	ttgcctgat	gccttcgg	caaacgtccg	attcattcac	cctccttgc	12000
ggattgccc	gactcacgcc	gggcaatgt	gcccttattc	ctgattgac	ccgcctggtg	12060
ccttggtg	cagataatcc	accttatc	caatgaagtc	ggtcccgtag	accgtctg	12120
cgctcttctc	gtacttggt	ttccgaatct	tgccctgcac	gaataccagc	gaccccttgc	12180
ccaaatactt	gccgtggg	tcggcctgag	agccaaaaca	cttgatgcgg	aagaagtcgg	12240
tgcgctcctg	ctgtgcgcg	gtcgtggc	cgccaacctt	tgcatccgc	aagcgcgcgg	12300
tcgcatcctt	cacgctgaa	cagtaoctc	aggcgggcat	catgaccgc	gagcaatac	12360
aggtcattaa	aagcgcctg	attgatgata	tagcggccc	gctgctcctg	gttctcgcgc	12420
accgaaatg	gtgacttcac	cccgcctct	ttgatcgtg	caccgatttc	cgcatgctc	12480
tcgggggaaa	agccggggtt	gtcggcgc	cgcgctgat	gcgatcttc	gtcgtcag	12540
tccaggtcca	gtcgtatag	gccggaacc	ccctgagac	ccgaggagc	gtccaggagg	12600
ctcgacaggt	cgccgatgct	atccaacccc	aggccggac	gctgcgcgc	gctgcggct	12660
tcctgagcgg	ccgacgggt	gttttcttg	gtggtcctg	cttgagcgc	agtcattgg	12720
aaatctccat	cttcgtgaa	acgtaatcag	ccagggcgc	aacctcttc	gatgccttgc	12780
gcgcggccgt	ttcttgatc	ttccagacc	gcacaccgga	tgcgagggca	tcggcgatgc	12840
tgctgcgcag	gccaacggt	gccggaatca	tcactctgg	gtacgcggcc	agcagctcg	12900
cttggtggcg	cgctggcgc	ggattccgc	catcgacctt	gctgggcacc	atgccaagga	12960
attgcagctt	ggcgttcttc	tgccgcacgt	tcgcaatggt	cgtgaccatc	ttcttgatgc	13020
cctgatgct	gtacgcctca	agctcgtg	gggacagcac	atagtcggcc	gcgaagagg	13080
cgcccgccag	gccgacgcca	agggtcggg	ccgtgctgat	caggcacacg	tcgaagcctt	13140

-continued

```

ggttcgccag ggccttgatg ttcgccccga acagctcgcg ggcgtcgtcc agcgcacagcc 13200
gttcggcggt cgccagtacc gggttggact cgatgagggc gaggcgcgcg gcctggccgt 13260
cgccggctgc gggtgcggtt tcggtccagc cgccggcagg gacagcgcgc aacagcttgc 13320
ttgatgcag gccggtagca aagtccttga gcgtgtagga cgcattgccc tgggggtcca 13380
ggtcgatcac ggcaaccgc aagccgcgct cgaaaaagtc gaaggcaaga tgcacaaggg 13440
tcgaagtctt gccgacgcgc cctttctggt tggccgtgac caaagttttc atcgtttgg 13500
ttctgtttt ttcttggcgt ccgcttccca cttccggagc atgtacgcct gatgttccgg 13560
cagaaccgcc gttaccgcgc cgtaccctc gggcaagttc ttgtcctcga acgcgccca 13620
cacgcgatgc accgcttgcg aactcgcgc cctggtcagt cccagcgcgc ttgcgaacgt 13680
cgctgtggc ttcccatcga ctaagacgcc ccgcgctatc tcgatggtct gctgccccac 13740
ttccagcccc tggatcgcct cctggaactg gcttccgta agccgtttct tcatggataa 13800
caccataat ttgctccgcg ccttggttga acatagcggg gacagccgcc agcacatgag 13860
agaagttag ctaaacttt ctcgcacgtc aacacctta gccgctaaaa ctgctccttg 13920
gcgtaacaaa aaaaagccc ggaaaccggg ctttctctc ttgccgctta tggctctgca 13980
cccggctcca tcaccaacag gtcgcgcacg cgcttcactc ggttgcggat cgacactgcc 14040
agcccaacaa agccggttgc cgccgcgcgc aggatcgcgc cgatgatgcc ggccacaccg 14100
gccatcgccc accaggtcgc cgccttccgg ttccattcct gctggtaactg cttegcaatg 14160
ctggactcgc gctcaccata ggctgaccgc tcgatggcgt atgccgttc tccccttggc 14220
gtaaaaccca gcgcgcagg cggcattgcc atgctgcccg ccgctttccc gaccacgacg 14280
cgcgaccag gttgcggtc cagaccttc gccacggcga gctgcgcaag gacataatca 14340
gcgcgcgact tggctccagc cgcctcagtc agctcttcca ctgcgcgcaa atccttggcc 14400
tccacggcgc ccatgaatcg cgcacgcggc gaaggctccg cagggccg 14448

```

```

<210> SEQ ID NO 24
<211> LENGTH: 48
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
primer

```

```

<400> SEQUENCE: 24

```

```

ctgtactagt ggaggatagt aatgaagcgc aaagtattgt taattccg 48

```

```

<210> SEQ ID NO 25
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
primer

```

```

<400> SEQUENCE: 25

```

```

tgtaagatct ttgtcggcc tccttactgt atttctccc 39

```

```

<210> SEQ ID NO 26
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

```

-continued

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 26

aggcaggagg aacctatga agcgaaagt attgttaatt ccg 43

<210> SEQ ID NO 27
<211> LENGTH: 48
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 27

gaccaccgga cgacacttgt tgtgcgcct cttactgta tttctccc 48

<210> SEQ ID NO 28
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 28

aggcaggagg aacctatg 19

<210> SEQ ID NO 29
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 29

caagtgtcgt ccggtggtc 19

<210> SEQ ID NO 30
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 30

catagtgttc ctctgcct 19

<210> SEQ ID NO 31
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 31

gaccaccgga cgacacttg 19

<210> SEQ ID NO 32
<211> LENGTH: 19
<212> TYPE: DNA

-continued

<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 32

aggcaggagg aacctatg 19

<210> SEQ ID NO 33
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 33

gaccaccgga cgaccttg 19

<210> SEQ ID NO 34
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
peptide

<400> SEQUENCE: 34

Asn Ser His His His His His His His
1 5

<210> SEQ ID NO 35
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 35

gatctcacca ccaccaccac caccactaat ag 32

<210> SEQ ID NO 36
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 36

ctattagtgg tgggtgtggt ggtgtgga 28

<210> SEQ ID NO 37
<211> LENGTH: 11
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 37

ctagtcgcca c 11

-continued

<210> SEQ ID NO 38
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 38

Val Ser Ser Gly Gly His His
1 5

<210> SEQ ID NO 39
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 39

gtgtcgtccg gtggtcacc 19

<210> SEQ ID NO 40
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 40

gtgttctctcc tgctact 18

<210> SEQ ID NO 41
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 41

His His His His His His His
1 5

<210> SEQ ID NO 42
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 42

His
1 5 10

<210> SEQ ID NO 43
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

-continued

<400> SEQUENCE: 43

His
 1 5 10

We claim:

1. A versatile broad host-range heterologous protein expression vector comprising:

- (a) a promoter nucleic acid sequence operable in a photosynthetic bacteria;
- (b) a nucleic acid sequence encoding an extended purification tag;
- (c) a cloning cassette comprising a multiple cloning site; and
- (d) a selection marker to select in the photosynthetic bacteria.

2. The vector of claim 1, wherein the photosynthetic bacteria is *Rhodobacter*.

3. The vector of claim 1, wherein the extended purification tag is N-terminal to the heterologous protein.

4. The vector of claim 1, wherein the extended purification tag is C-terminal to the heterologous protein.

5. The vector of claim 1, wherein the extended purification tag is a histidine tag comprising about 7 to about 13 contiguous histidine residues.

6. The vector of claim 1, wherein the purification tag comprises a linker sequence.

7. The vector of claim 6, wherein the linker sequence comprises about 1 to about 20 amino acids.

8. The vector of claim 1 further comprises a cleavable signal sequence.

9. The vector of claim 1 further comprises a membrane anchor domain.

10. The vector of claim 1, wherein the cloning cassette facilitates ligation independent cloning.

11. The vector of claim 1, wherein the heterologous protein is a membrane protein.

12. The vector of claim 1, wherein the heterologous protein is a soluble protein.

13. The vector of claim 1 further comprises a nucleic acid sequence encoding a component of an intracytoplasmic membrane of *Rhodobacter*.

14. The vector of claim 1, wherein the promoter is inducible.

15. A method of producing a heterologous protein in a photosynthetic organism, the method comprising:

- (a) cloning a nucleic acid sequence encoding the heterologous protein into a vector of claim 1;
- (b) expressing the heterologous protein in a photosynthetic bacteria;
- (c) purifying the heterologous protein using an extended purification tag; and

(d) obtaining heterologous protein from the photosynthetic organism.

16. The method of claim 15, wherein the photosynthetic bacteria is *Rhodobacter*.

17. The method of claim 15, wherein the extended purification tag is a histidine tag comprising about 7 to about 13 contiguous histidine residues.

18. The method of claim 15, wherein the extended purification tag is a histidine tag comprising about 7 to about 13 contiguous histidine residues and further comprising about 1 to about 20 linker amino acids.

19. The method of claim 15, wherein the heterologous protein is a membrane protein.

20. The method of claim 15, wherein the heterologous protein is a soluble protein.

21. A method of producing a heterologous membrane protein in *Rhodobacter*, the method comprising:

- (a) cloning a nucleic acid sequence encoding the heterologous membrane protein into a vector comprising a promoter sequence operable in *Rhodobacter*, an N- or C-terminal extended purification tag comprising about 7 to about 30 amino acids in length;

(b) expressing the heterologous membrane protein in the *Rhodobacter*;

(c) sequestering and compartmentalizing the heterologous membrane protein into an intracytoplasmic membrane (ICM) complex;

(d) purifying the heterologous membrane protein using the extended purification tag; and

(e) obtaining heterologous membrane protein from the *Rhodobacter*.

22. A method of producing a heterologous soluble protein in *Rhodobacter*, the method comprising:

- (a) cloning a nucleic acid sequence encoding the heterologous soluble protein into a vector comprising a promoter sequence operable in *Rhodobacter*, an N- or C-terminal extended purification tag comprising about 7 to about 30 amino acids in length, and a membrane anchor or linker sequence;

(b) expressing the heterologous soluble protein in the *Rhodobacter*;

(c) purifying the heterologous soluble protein using the extended purification tag; and

(d) obtaining heterologous soluble protein from the *Rhodobacter*.

* * * * *